
Simulationsstudien zur ortsspezifischen Biokonjugation maßgeschneiderter Polymere



*Dissertation zur Erlangung des naturwissenschaftlichen
Doktorgrades der Julius-Maximilians-Universität Würzburg*

vorgelegt von

Josef Kehrein

aus Bad Kreuznach

Würzburg 2022



Eingereicht bei der Fakultät für Chemie und Pharmazie am

Gutachter der schriftlichen Arbeit

1. Gutachter: _____

2. Gutachter: _____

Prüfer des öffentlichen Promotionskolloquiums

1. Prüfer: _____

2. Prüfer: _____

3. Prüfer: _____

Datum des öffentlichen Promotionskolloquiums

Doktorurkunde ausgehändigt am

nil admirari

Sämtliche Arbeiten dieser Dissertation wurden im Zeitraum von Februar 2018 bis einschließlich Dezember 2021 unter der Aufsicht von Prof. Dr. Christoph Sotriffer am Institut für Pharmazie und Lebensmittelchemie der Julius-Maximilians-Universität Würzburg durchgeführt. Die hier vorgestellten Ergebnisse wurden teilweise in wissenschaftlichen Publikationen veröffentlicht:

1. N. Hauptstein, P. Pouyan, J. Kehrein, M. Dirauf, M. D. Driessen, M. Raschig, K. Licha, M. Gottschaldt, U. S. Schubert, R. Haag, L. Meinel, C. Sotriffer und T. Lühmann, „Molecular Insights into Site-Specific Interferon- α 2a Bioconjugates Originated from PEG, LPG, and PEtOx“, *Biomacromolecules*, Jg. 22, Nr. 11, S. 4521-4534, 2021

Bei folgenden Posterbeiträgen wurde mitgewirkt:

1. J. Kehrein und C. Sotriffer, „Simulation Studies on the Site-Specific Bioconjugation of Polymers“, 14th German Conference on Cheminformatics (Mainz), 2018
2. J. Kehrein und C. Sotriffer, „Simulation Studies on the Site-Specific Bioconjugation of Polymers“, 12th European Workshop in Drug Design (Siena), 2019
3. J. Kehrein und C. Sotriffer, „Simulationsstudien zur ortsspezifischen Biokonjugation maßgeschneiderter Polymere“, ProMatLeben - Polymere: 1. Doktoranden- und PostDoc-Konferenz (Berlin), 2019
4. J. Kehrein, C. B. Nkemeto, T. Lühmann, L. Meinel und C. Sotriffer, „Der Effekt PEG-basierter Biokonjugationen auf die konformative Proteinstabilität: Ein simulationsbasiertes Vorhersagemodell“, ProMatLeben - Polymere: 2. Doktoranden- und PostDoc-Konferenz (Berlin, virtuell), 2021
5. N. Hauptstein, P. Pouyan, M. Dirauf, J. Kehrein, M. Raschig, K. Licha, M. Gottschaldt, U. S. Schubert, L. Meinel, C. Sotriffer, R. Haag und T. Lühmann, „Site-selective decoration of Interferon- α 2a with PEG alternatives and their physical, chemical, and biological characterisation“, ProMatLeben - Polymere 2. Doktoranden- und PostDoc-Konferenz (Berlin, virtuell), 2021
6. J. Kehrein, C. B. Nkemeto, T. Lühmann, L. Meinel und C. Sotriffer, „Effects of PEG-based bioconjugation on protein conformational stability: A simulation-based model“, ProMatLeben - Polymere: Midtermkonferenz (Berlin, virtuell), 2021

7. N. Hauptstein, P. Pouyan, M. Dirauf, J. Kehrein, M. Raschig, K. Licha, M. Gottschaldt, U. S. Schubert, L. Meinel, C. Sotriffer, R. Haag und T. Lühmann, „Correlation of IFN- α 2a bioconjugate stability and polymer hydrophilicity“, ProMatLeben - Polymere: Midtermkonferenz (Berlin, virtuell), 2021
8. N. Hauptstein, P. Pouyan, K. Wittwer, J. Kehrein, M. Dirauf, C. Gizem, M. Raschig, M. Tully, I. Nischang, U. S. Schubert, C. Sotriffer, C. Pfaller, R. Haag, T. Lühmann und L. Meinel, „Site-selective decoration of Interferon- α 2a with PEG alternatives and their physical, chemical, and biological characterisation“, ProMatLeben - Polymere: 3. Doktoranden- und PostDoc-Konferenz (Berlin, virtuell), 2021

Bei folgenden Präsentationen wurde mitgewirkt:

1. M. Tully, N. Engel, M. Dirauf, N. Hauptstein, J. Kehrein, C. B. Nkemeto, P. Pouyan, K. Wittwer, M. Gottschaldt, K. Licha, T. Lühmann, L. Meinel, C. Pfaller, U. S. Schubert, C. Sotriffer und R. Haag, „PEG-alternative Polymere für die Biokonjugation an Proteine - Next-PEG“, ProMatLeben - Polymere: 1. Doktoranden- und PostDoc-Konferenz (Berlin), 2019
2. C. Sotriffer und J. Kehrein, „Simulation-driven model building: Towards prediction of site-specific bioconjugation“, 7th Strasbourg Summer School in Chemoinformatics (Straßburg, virtuell), 2020
3. N. Hauptstein, M. Dirauf, N. Engel, J. Kehrein, C. B. Nkemeto, P. Pouyan, M. Tully, K. Wittwer, R. Haag, K. Licha, M. Gottschaldt, L. Meinel, C. Pfaller, U. S. Schubert, C. Sotriffer und T. Lühmann, „PEG-alternative Polymere für die Biokonjugation an Proteine - Next-PEG“, ProMatLeben - Polymere: 2. Doktoranden- und PostDoc-Konferenz (Berlin, virtuell), 2021
4. N. Hauptstein, P. Pouyan, M. Dirauf, J. Kehrein, M. Raschig, K. Licha, M. Gottschaldt, U. S. Schubert, C. Sotriffer, L. Meinel, R. Haag und T. Lühmann, „Correlation of IFN- α 2a bioconjugate stability and polymer hydrophilicity“, ProMatLeben - Polymere: Midtermkonferenz (Berlin, virtuell), 2021
5. K. Wittwer, N. Hauptstein, M. Dirauf, N. Engel, J. Kehrein, C. B. Nkemeto, P. Pouyan, M. Tully, R. Haag, K. Licha, M. Gottschaldt, L. Meinel, C. Pfaller, U. S. Schubert, C. Sotriffer und T. Lühmann, „PEG-alternative Polymere für die Biokonjugation an Proteine - Next-PEG“, ProMatLeben - Polymere: 3. Doktoranden- und PostDoc-Konferenz (Berlin, virtuell), 2021

Folgende Arbeiten, denen andere Forschungsthemen zugrunde liegen, wurden im genannten Zeitraum ebenfalls publiziert:

1. C. Kersten, E. Fleischer, J. Kehrein, C. Borek, E. Jaenicke, C. Sotriffer und R. Brenk, „How To Design Selective Ligands for Highly Conserved Binding Sites: A Case Study Using *N*-Myristoyltransferases as a Model System“, *Journal of Medicinal Chemistry*, Jg. 63, Nr. 5, S. 2095–2113, 2020
2. J. Wiest, J. Kehrein, M. Saedtler, K. Schilling, E. Cataldi, C. A. Sotriffer, U. Holzgrabe, T. Rasmussen, B. Böttcher, M. Cronin-Golomb, M. Lehmann, N. Jung, M. Windbergs und L. Meinel, „Controlling Supramolecular Structures of Drugs by Light“, *Molecular Pharmaceutics*, Jg. 17, Nr. 12, S. 4704–4708, 2020
3. S. Eltschkner, J. Kehrein, T. A. Le, S. Davoodi, B. Merget, S. Basak, J. D. Weinrich, J. Schiebel, P. J. Tonge, B. Engels, C. Sotriffer und C. Kisker, „A Long Residence Time Enoyl-Reductase Inhibitor Explores an Extended Binding Region with Isoenzyme-Dependent Tautomer Adaptation and Differential Substrate-Binding Loop Closure“, *ACS Infectious Diseases*, Jg. 7, Nr. 4, S. 746–758, 2021
4. L. Hahn, T. Zorn, J. Kehrein, T. Kielholz, A.-L. Ziegler, B. Sochor, E. S. Litsysyna, N. A. Durandin, T. Laaksonen, V. Aseyev, C. Sotriffer, M. Windbergs, A.-C. Pöppler und R. Luxenhofer, „Unravelling a novel mechanism in polymer self-assemblies: An order-order transition based on molecular interactions between hydrophilic and hydrophobic polymer blocks“, *ChemRxiv Preprint*, 2021

Hinweise zu Tabellen und Abbildungen, die aus Publikationen entnommen, adaptiert oder mithilfe von Rohdaten reproduziert wurden:

- Tab. 1.4: Reproduziert und mit neuen Referenzen erweitert

Nachdruck mit Genehmigung aus: S. L. Baker, A. Munasinghe, H. Murata, P. Lin, K. Matyjaszewski, C. M. Colina und A. J. Russell, „Intramolecular Interactions of Conjugated Polymers Mimic Molecular Chaperones to Stabilize Protein–Polymer Conjugates“, *Biomacromolecules*, Jg. 19, Nr. 9, S. 3798–3813, 2018. Copyright 2018 American Chemical Society (Link: <https://pubs.acs.org/doi/10.1021/acs.biomac.8b00927>).

- Abb. 3.6, 3.7, 3.8, 3.9: Graphen auf der linken Seite entnommen

Nachdruck mit Genehmigung aus: P. B. Lawrence, Y. Gavrilov, S. S. Matthews, M. I. Langlois, D. Shental-Bechor, H. M. Greenblatt, B. K. Pandey, M. S. Smith, R. Paxman, C. D. Torgerson, J. P. Merrell, C. C. Ritz, M. B. Prigozhin, Y. Levy und J. L. Price, „Criteria for Selecting PEGylation Sites on Proteins for Higher Thermodynamic and Proteolytic Stability“, *Journal of the American Chemical Society*, Jg. 136, Nr. 50, S. 17 547–17 560, 2014. Copyright 2014 American Chemical Society (Link: <https://pubs.acs.org/doi/10.1021/ja5095183>).

- Abb. 3.17: komplette Abbildung entnommen
- Abb. 5.1: Grafiken zu den 10-kDa-Varianten entnommen
- Abb. 5.12: komplette Abbildung entnommen
- Abb. 5.20: Grafik in (B) entnommen
- Abb. 5.23: Balkendiagramm reproduziert
- Abb. 5.24: Graphen reproduziert, Grafik in (D) entnommen
- Abb. 5.27: Graphen reproduziert, Grafiken in (B) und (C) entnommen
- Abb. 5.29: komplette Abbildung entnommen

Nachdruck mit Genehmigung aus: N. Hauptstein, P. Pouyan, J. Kehrein, M. Dirauf, M. D. Driessen, M. Raschig, K. Licha, M. Gottschaldt, U. S. Schubert, R. Haag, L. Meinel, C. Sotriffer und T. Lühmann, „Molecular Insights into Site-Specific Interferon- α 2a Bioconjugates Originated from PEG, LPG, and PEtOx“, *Biomacromolecules*, Jg. 22, Nr. 11, S. 4521-4534, 2021. Copyright 2021 American Chemical Society (Link: <https://pubs.acs.org/doi/10.1021/acs.biomac.1c00775>).

- Abb. 5.30, 5.32: Balkendiagramme reproduziert
- Abb. 5.33: Korrelation in (A) adaptiert

Daten entnommen aus: N. Hauptstein, P. Pouyan, K. Wittwer, G. Cinar, O. Scherf-Clavel, K. Licha, T. Lühmann, I. Nischang, U. S. Schubert, C. Pfaller, R. Haag und L. Meinel, „Polymer selection impacts the pharmaceutical profile of site specifically conjugated Interferon- α 2a“, *ChemRxiv Preprint*. Veröffentlicht unter der Creative Commons Attribution License (Link: <https://doi.org/10.26434/chemrxiv-2022-j85br>).

Ein Teil der in dieser Arbeit vorgestellten Molekulardynamik-Simulationen wurde auf HPC-Clustern verschiedener Rechenzentren durchgeführt. Im Folgenden sind die auf Basis der NAMD-Logdateien aufsummierten Simulations- und Rechenzeiten relevanter Produktionsläufe aufgelistet:

- Curta-Cluster der ZEDAT der FU Berlin

11.85 μ s bzw. 339.82 Tage Rechenzeit.

Für die zur Verfügung gestellte Rechenzeit möchte ich mich beim HPC-Dienst der ZEDAT, Freie Universität Berlin bedanken.

Link: <http://dx.doi.org/10.17169/refubium-26754>

- Julia-Cluster des Rechenzentrums der JMU Würzburg

34.43 μ s bzw. 593.21 Tage Rechenzeit.

Für die zur Verfügung gestellte Rechenzeit möchte ich mich beim Rechenzentrum der JMU Würzburg bedanken.

Link: <https://www.rz.uni-wuerzburg.de/dienste/rzserver/high-performance-computing/>

- Linux-Cluster des Leibniz-Rechenzentrums der Bayerischen Akademie der Wissenschaften (Garching bei München)

18.40 μ s bzw. 145.74 Tage Rechenzeit.

Für die zur Verfügung gestellte Rechenzeit möchte ich mich beim Leibniz-Rechenzentrum bedanken.

Link: www.lrz.de

Danksagung

Ich danke Prof. Dr. Christoph Sotriffer recht herzlich für seine Betreuung und den regelmäßigen Austausch in persönlichen Gesprächen. Ebenso danke ich Prof. Dr. Dr. Lorenz Meinel und Prof. Dr. Tessa Lühmann für meine Einbindung in das Verbundprojekt *PEG-alternative Polymere für die Biokonjugation an Proteine (Next-PEG)* und weitere Forschungsprojekte der pharmazeutischen Technologie.

Weiterhin danke ich dem BMBF und dem VDI für die Finanzierung und Organisation dieses Projekts, der ZEDAT der FU Berlin, dem Leibniz-Rechenzentrum in Garching bei München und dem Rechenzentrum der JMU Würzburg für ihre Cluster-Ressourcen, sowie allen Kooperationspartnern (JMU Würzburg, FU Berlin, FSU Jena, PEI) für die Zusammenarbeit: Prof. Dr. Michael Gottschaldt, Prof. Dr. Rainer Haag, Prof. Dr. Kai Licha, Prof. Dr. Tessa Lühmann, Prof. Dr. Dr. Lorenz Meinel, Prof. Dr. Michael Schirner, Prof. Dr. Ulrich S. Schubert, Prof. Dr. Christoph Sotriffer, Dr. Christian Pfaller, Dr. Michael Dirauf, Nora Engel, Dr. Paria Pouyan, Dr. Michael Tully, Kevin Wittwer, sowie ganz besonders Niklas Hauptstein und Cynthia Bessem Nkemeto.

Ich bedanke mich bei allen Kollegen des Arbeitskreises für ihre Unterstützung und die schöne Zeit in Würzburg: Dr. Sebastian Bothe, Stephan Böhler, Mathias Diebold, Erik Endres, Carina Herbst, Gerald Keller, Dr. Maximilian Kuhn, Dr. Yogesh Narkhede, Dr. Yesid Ramirez, Natalia Yuan Chen, Markus Zehe und Thomas Zimmermann. Ein Dank gilt ebenso allen Assistenten des ersten Semesters für ihr tatkräftiges Mitwirken: Dr. Sebastian Bothe, Stephan Böhler, Sara Endres, Dr. Paul Güntzel, Gerald Keller, Mohamed Marzouk, Dr. Ruben Pawellek, Dr. Joseph Skaf, Natalia Yuan Chen und Markus Zehe.

Zusätzlich danke ich meiner ehemaligen Praktikantin Katarzyna Miśkiewicz und allen weiteren Personen, die mir geholfen haben, insbesondere Dorothee Haas, Eduardo Merino und Magdalena Nowak. Ebenfalls danke ich Prof. Dr. Ruth Brenk, Dr. Christian Kersten und Dr. Raphael Klein, die mein Interesse für diesen Forschungsbereich bereits während meines Pharmaziestudiums gefördert haben.

Allen Personen, mit denen ich in zusätzlichen Projekten zusammenarbeiten durfte, danke ich für die Kooperation: Prof. Dr. Robert Luxenhofer, Jun.-Prof. Dr. Ann-Christin Pöppler, Dr. Sandra Eltschkner, Lukas Hahn, Dr. Johannes Wiest, PD Dr. Frank Waller, Simon Hanio, Jonas Schlauersbach und Markus Zehe.

Zuletzt gilt ein ganz besonderer Dank meinen Freunden, vor allem Markus, Mohamed und Paul, sowie meiner Familie für den nötigen Beistand in guten sowie schlechten Zeiten.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Polymer-Biokonjugate	1
1.1.1	Einführung	1
1.1.2	Herstellungsverfahren	5
1.1.2.1	Grafting-Methoden	5
1.1.2.2	Polymersynthese	6
1.1.2.3	Konjugationsstrategien	7
1.1.3	Polymerarten	9
1.1.3.1	Polyethylenglykol	9
1.1.3.2	Polyvinylpyrrolidon	10
1.1.3.3	Polyglycerole	10
1.1.3.4	Polyoxazoline	11
1.1.3.5	Weitere Polymere	11
1.1.4	Therapeutische Beispiele	12
1.1.4.1	Adagen®	12
1.1.4.2	Neulasta®	13
1.1.4.3	Pegasys®	13
1.1.4.4	Macugen®	13
1.1.5	Thermodynamische Charakterisierung	14
1.1.5.1	Temperaturabhängige CD-Spektroskopie	14
1.1.5.2	Thermal Shift Assay	16
1.1.6	Stabilisierungsmechanismen	17
1.1.6.1	Bisherige Hypothesen	17
1.1.6.2	Arbeiten von Joshua L. Price	18
1.1.6.3	Zusammenfassende Übersicht	19
1.2	Molekulardynamik-Simulationen	21
1.2.1	Grundlagen	21
1.2.1.1	Abbildung atomarer Wechselwirkungen	21
1.2.1.2	Bewegungsgleichungen	21
1.2.1.3	Wahl des Zeitschritts	23
1.2.1.4	Strukturaufbereitung und Equilibrierung	23
1.2.1.5	Periodische Randbedingungen	24
1.2.2	Biokonjugat-Simulationen	25
1.2.3	Konvergenzkriterien	26

1.2.3.1	Self-consistency checks	27
1.2.3.2	Weitere Konvergenztests	28
1.2.4	Gaußbeschleunigte Molekulardynamik-Simulationen	29
1.2.5	Wohltemperierte Metadynamik-Simulationen	30
1.3	QSPR-Modelle zur Stabilitätsvorhersage	33
1.3.1	Grundlagen	33
1.3.1.1	Hintergrund	33
1.3.1.2	Deskriptoren	33
1.3.2	Statistische Validierungsparameter	34
1.3.2.1	Regressionsparameter	34
1.3.2.2	Klassifikationsparameter	35
1.3.3	Modellarten dieser Studie	37
1.3.3.1	Multiple lineare Regression	37
1.3.3.2	Logistische, binomiale Regression	38
1.3.3.3	Künstliches neuronales Netz	39
1.3.3.4	XGBoost-Modell	39
1.3.4	Limitierungen	40
1.4	Constrained-Network-Analyse	42
1.4.1	Theoretischer Hintergrund	42
1.4.2	Anwendung des CNA-Programms	43
1.5	Untersuchte Strukturen	44
1.5.1	Modellproteine	44
1.5.1.1	Pin 1 WW-Domäne	44
1.5.1.2	Her2-Affibody	45
1.5.1.3	Src SH3-Domäne	46
1.5.1.4	GCN4 und coil-V _a L _d	47
1.5.2	Biologika	48
1.5.2.1	Interferon- α 2a	48
2	Zielsetzung	49
3	Materialien und Methoden	51
3.1	Modellproteine	51
3.1.1	Wildtyp-Simulationen und unkonjugierte Mutanten	51
3.1.2	Simulierte Konjugate	51
3.1.2.1	Konjugate mit publizierten Stabilitätsdaten	51
3.1.2.2	Zusätzliche Konjugate	52
3.1.3	Ermittlung geeigneter Simulationsparameter	54
3.1.3.1	PyRED-Methode	54
3.1.3.2	antechamber-Methode	55
3.1.3.3	Monomerbasierte Parametrisierung	57
3.1.3.4	Problemstellungen	57
3.1.3.5	Vergleich der Simulationen mit Referenz	58

3.1.4	Modellierung der Strukturen	63
3.1.4.1	Aufbereitung der Peptidstrukturen	63
3.1.4.2	Einbau modifizierter Aminosäuren	63
3.1.4.3	Solvatation der Strukturen	64
3.1.5	Simulationsprotokoll	64
3.1.5.1	Equilibrierung und erster Produktionslauf	64
3.1.5.2	Fortführung des Produktionslaufs	65
3.1.6	QSPR-Vorhersagemodelle	66
3.1.6.1	Deskriptoren	66
3.1.6.2	Modellgenerierung: Ansatz 1	66
3.1.6.2.1	Aufbereitung der Daten	67
3.1.6.2.2	Modellgenerierung	69
3.1.6.2.3	Evaluation	70
3.1.6.3	Modellgenerierung: Ansatz 2	70
3.1.6.3.1	Ausgeschlossene Konjugate	71
3.1.6.3.2	Definition unterschiedlicher Datensätze	71
3.1.6.3.3	Clustering der Daten	72
3.1.6.3.4	Modellgenerierung	73
3.1.6.3.5	Evaluation	75
3.1.6.3.6	Finales Modell	75
3.1.7	Metadynamik-Simulationen	76
3.1.7.1	Untersuchte Systeme	76
3.1.7.2	Kollektivvariablen	76
3.1.7.3	Simulationsparameter	76
3.1.8	Vergleich der Polymere	77
3.2	Biologika	78
3.2.1	Wildtyp-Simulationen	78
3.2.2	Simulierte Konjugate	78
3.2.3	Parameterfindung	80
3.2.4	Modellierung der Strukturen	80
3.2.5	Gaußbeschleunigte Molekulardynamik-Simulationen	82
3.2.6	Coarse-grained-Simulation	82
3.2.7	Experimentelle Arbeiten	83
4	Ergebnisse: Modellproteine	85
4.1	QSPR-Vorhersagemodelle	85
4.1.1	Ansatz 1	86
4.1.1.1	Referenzmodelle	86
4.1.1.2	Modelle C1 und D1	87
4.1.1.3	Modelle E und F	89
4.1.1.4	Modelle G und H	91
4.1.1.5	Leistung der Modelle	91

4.1.1.6	Verwendete Deskriptoren	93
4.1.1.7	MD-Analysen	94
4.1.1.7.1	Interaktionen und Solvatation	94
4.1.1.7.2	Linker verschiedener Länge	95
4.1.1.7.3	D-/L-Asparagin	97
4.1.1.7.4	Linker verschiedener Rigidität	98
4.1.1.7.5	Desolvatation apolarer Flächen	99
4.1.1.8	Fazit	99
4.1.2	Ansatz 2	100
4.1.2.1	Referenzmodelle	101
4.1.2.2	Regressionsmodell A2-R6	101
4.1.2.2.1	Allgemeine Leistung	101
4.1.2.2.2	Einzelne Konjugate	102
4.1.2.2.3	Applikationsdomäne	103
4.1.2.2.4	Vergleich zur strukturbasierten Richtlinie	104
4.1.2.2.5	Modelldeskriptoren	105
4.1.2.2.6	Alternative Polymerarten	106
4.1.2.2.7	Modell A2-R6v2	107
4.1.2.2.8	Fazit	108
4.1.2.3	Klassifikationsmodell A2-C1	109
4.1.2.3.1	Allgemeine Leistung	109
4.1.2.3.2	Einzelne Konjugate	110
4.1.2.3.3	Modelldeskriptoren	110
4.1.2.3.4	Alternative Polymerarten	111
4.1.2.3.5	Fazit	111
4.1.2.4	Zusätzliche Modelle	111
4.1.2.4.1	Parametervergleich	111
4.1.2.4.2	Modell A2v2-R1	113
4.1.2.4.3	Modell A2v2-C1	114
4.1.2.5	Fazit	115
4.2	Metadynamik-Simulationen	117
4.2.1	Energiediagramme	117
4.2.2	Konvergenz	117
4.2.3	Energiedifferenzen	119
4.2.4	Polymerdynamik	120
4.2.5	Fazit	122
4.3	Vergleich der Polymere	123
4.3.1	Pin 1-Konjugate	124
4.3.1.1	Konjugationsstelle 14	124
4.3.1.2	Konjugationsstelle 16	126
4.3.1.3	Konjugationsstelle 17	128
4.3.1.4	Konjugationsstelle 18	130

4.3.1.5	Konjugationsstelle 19	132
4.3.1.6	Konjugationsstelle 21	134
4.3.1.7	Konjugationsstelle 23	136
4.3.1.8	Konjugationsstelle 26	138
4.3.1.9	Konjugationsstelle 27	140
4.3.1.10	Konjugationsstelle 28	142
4.3.1.11	Konjugationsstelle 29	144
4.3.1.12	Konjugationsstelle 30	146
4.3.1.13	Konjugationsstelle 32	148
4.3.2	Her2-Konjugate	150
4.3.2.1	Konjugationsstelle 7	150
4.3.2.2	Konjugationsstelle 8	152
4.3.2.3	Konjugationsstelle 15	154
4.3.2.4	Konjugationsstelle 21	156
4.3.2.5	Konjugationsstelle 23	158
4.3.2.6	Konjugationsstelle 25	160
4.3.2.7	Konjugationsstelle 29	162
4.3.2.8	Konjugationsstelle 33	164
4.3.2.9	Konjugationsstelle 39	166
4.3.2.10	Konjugationsstelle 43	168
4.3.2.11	Konjugationsstelle 46	170
4.3.2.12	Konjugationsstelle 47	172
4.3.2.13	Konjugationsstelle 49	174
4.3.2.14	Konjugationsstelle 50	176
4.3.2.15	Konjugationsstelle 53	178
4.3.3	Zusammenfassender Vergleich	180
4.3.3.1	Pin 1 WW-Domäne	180
4.3.3.1.1	Vergleich der berechneten Eigenschaften	180
4.3.3.1.2	PEG und LPG	184
4.3.3.1.3	Effekt des Linkers	185
4.3.3.1.4	PMeOx und PEtOx	186
4.3.3.2	Her2-Affibody	187
4.3.3.2.1	Vergleich der berechneten Eigenschaften	187
4.3.3.2.2	PEG	192
4.3.3.2.3	LPG	193
4.3.3.2.4	PMeOx und PEtOx	194
4.4	Zusammenfassung	195
5	Ergebnisse: Biologika	199
5.1	Gaußbeschleunigte Molekulardynamik-Simulationen	199
5.1.1	Polymerdichten	199
5.1.2	Kontaktprofile	201

5.1.2.1	IFN- α 2a-K ₃₁ -BCN-10-kDa	202
5.1.2.1.1	BCN-PEG-10-kDa	202
5.1.2.1.2	BCN-LPG-10-kDa	203
5.1.2.1.3	BCN-PEtOx-10-kDa	206
5.1.2.2	IFN- α 2a-K ₃₁ -DBCO-10-kDa	206
5.1.2.2.1	DBCO-PEG-10-kDa	206
5.1.2.2.2	DBCO-LPG-10-kDa	208
5.1.2.2.3	DBCO-PEtOx-10-kDa	209
5.1.2.3	Acyl-PEG-10-kDa	210
5.1.2.4	Größere BCN-Konjugate (20 - 50 kDa)	211
5.1.2.4.1	20 kDa	211
5.1.2.4.2	30 - 50 kDa	212
5.1.3	Fazit	216
5.2	Coarse-grained-Simulation	217
5.2.1	Vergleich mit GaMD-Simulationen	217
5.2.2	Verwendetes Kraftfeld	219
5.3	Untersuchungen zur thermischen Stabilität	220
5.3.1	10-kDa-Varianten	220
5.3.1.1	CN-Analysen	221
5.3.1.1.1	Generierte Netzwerke	223
5.3.1.1.2	Perkolationsindizes	223
5.3.1.2	Native Kontakte	223
5.3.1.3	RMSF-Verhältnis	227
5.3.1.4	Vergleich mit LiP-Assays	227
5.3.2	Größere Konjugate	228
5.4	Untersuchungen zur Bioaktivität	231
5.5	Zusammenfassung	234
6	Fazit und Ausblick	237
6.1	Fazit	237
6.2	Ausblick	240
6.2.1	Studien zu Modellproteinen	240
6.2.1.1	Vergrößerung des Datensatzes	240
6.2.1.2	Weitere QSPR-Modelle	240
6.2.1.3	Sampling der ungefalteten Varianten	241
6.2.1.4	Analyse anderer Eigenschaften	241
6.2.1.5	Experimentelle Ansätze	242
6.2.1.6	Alternative Simulationsansätze	242
6.2.1.6.1	Metadynamik-Simulationen	242
6.2.1.6.2	Erhöhte Temperaturen	242
6.2.1.6.3	Gerichtete MD-Simulationen	243
6.2.1.6.4	Entfaltung über Gō-Modelle	243

6.2.1.7	MD-Alternativen	243
6.2.2	Größere Konjugate	243
6.2.2.1	Verbesserung des Samplings	243
6.2.2.1.1	Abbildung des Solvens	243
6.2.2.1.2	GaMD-Alternativen	244
6.2.2.1.3	Vergrößerte Kraftfelder	244
6.2.2.2	MD-Alternativen	244
Zusammenfassung		245
Summary		247
Anhang		249
A Ausgleichsrechnungen		251
A.1	Zugrundeliegende Gleichungen	251
A.2	Fehlerfortpflanzung	253
B Deskriptoren		255
B.1	Mittlere Abweichung (RMSD)	255
B.2	Mittlere Fluktuation (RMSF)	255
B.3	Solvenzzugängliche Oberfläche (SASA)	256
B.4	Radiale Verteilungsfunktion (RDF)	256
B.5	Gyrationsradius	256
B.6	HB/Lebensdauer-Analyse	257
B.7	Interaktionsenergie	257
B.8	Rückgrat-Diederwinkel	257
B.9	Ordnungsparameter	258
B.10	Native Kontakte	259
B.11	Diffusionskoeffizient	259
B.12	HB-Autokorrelation	259
B.13	CNA-basierte Deskriptoren	260
C Konjugate		263
D Vorhersagemodelle		273
D.1	Berechnete Deskriptoren	273
D.2	Ansatz 1	278
D.2.1	Modelltypen	278
D.2.2	Modelleleistungen	280
D.2.3	ADAN-Analysen	281
D.2.4	Pin 1-unspezifische Deskriptoren	281
D.3	Ansatz 2	283
D.3.1	Pin 1-unspezifische Deskriptoren	283

D.3.2	Modelltypen	284
D.3.3	Modellleistungen	284
D.3.4	Analysen	286
	Literatur	291
	Abbildungsverzeichnis	323
	Tabellenverzeichnis	329
	Abkürzungsverzeichnis	331

Kapitel 1

Einleitung

1.1 Polymer-Biokonjugate

1.1.1 Einführung

Durch die Optimierung und Weiterentwicklung neuer biotechnologischer Verfahren seit der Markteinführung rekombinant hergestellten Humaninsulins im Jahre 1982 [1] haben makromolekulare Arzneistoffe in der Pharmazie stetig an Bedeutung hinzugewonnen. Biopharmazeutika bzw. Biologika und die davon abgeleiteten Biosimilars stellen Peptide, Proteine oder nukleinsäurebasierte Strukturen wie Aptamere dar und sind vielseitig einsetzbar. Besonders monoklonale Antikörper dienen oftmals zur Therapie von Autoimmunerkrankungen wie Morbus Crohn oder rheumatoider Arthritis, können aber auch gegen Krebs eingesetzt werden [2]. Zur Klasse der Biologika zählen ebenso Impfstoffe [3]. Acht der zehn weltweit umsatzstärksten Medikamente des Jahres 2019 stellten Biopharmazeutika dar [4] und es ist zu erwarten, dass dieser Markt noch weiter wachsen wird. Aufgrund ihrer komplexen Struktur binden diese makromolekularen Substanzen selektiv an ihr Angriffsziel [3, 5].

Die Nutzung biotechnologisch hergestellter Makromoleküle, die parenteral appliziert werden müssen [6], ist jedoch keineswegs unproblematisch. Durch enzymatische Degradation und renale Filtration wird ihre Wirkung *in vivo* deutlich beeinträchtigt. Auch hinsichtlich der Wirkstoffformulierung ergeben sich Fragen zur Stabilität, da die Funktion dieser Arzneistoffe wesentlich von ihrer dreidimensionalen Struktur abhängt, welche in der Regel sensibel auf Änderungen des pH-Werts, der Temperatur und andere äußere Einflüsse reagiert [3, 5]. Der aufwendige und kostspielige zellbasierte Herstellungsprozess definiert maßgeblich das Produkt, weshalb Biosimilars auch nicht mit Generika gleichzusetzen sind [2, 3, 7]. Ergänzend zu pharmakokinetischen und pharmazeutisch-technologischen Problematiken besitzen diese Wirkstoffe ein gewisses immunogenes Potenzial, wodurch es zu unerwünschten Abwehrreaktionen kommen kann [3, 8].

Durch kovalente Verknüpfung biologischer Makromoleküle mit Polymeren können neuartige Hybride erhalten werden, die im Idealfall die Vorteile beider Materialwelten miteinander vereinen. Dieser Prozess wird Biokonjugation genannt und kann die physikochemischen und pharmakokinetischen Eigenschaften von Arzneistoffen wesentlich verbessern [9]. Während besonders die Medizin von solchen Konjugaten profitiert hat, sollte nicht unerwähnt bleiben, dass diese Verbindungen bspw. auch in der Lebensmittel- und Kosmetikindustrie zunehmend Anwendung finden [10]. Verbesserte Biopharmazeutika werden indes auch als Biobetters bezeichnet [11]. Unter smarten Biokonjugaten versteht man gar Systeme, bei denen die Konformation des Polymers durch äußere Stimuli wie Licht, Temperatur oder pH-Wertänderungen kontrolliert werden kann [12].

Die ersten Arbeiten zu Polymer-Biokonjugaten (PBK) wurden bereits in den Jahren 1952 und 1953 publiziert, bei denen Chymotrypsin und Albumin mit Polyglycinen verknüpft wurden [13–15]. Das eigentliche medizinische Potential wurde jedoch erst durch die 1977 von Abuchowski et al. hergestellten Konjugate deutlich [16, 17]. Zur Biokonjugation ist in der Regel eine spezielle Linkerstruktur nötig, über die das Polymer mit dem Protein verknüpft werden kann. In den Arbeiten von Abuchowski et al. wurde Polyethylenglykol (PEG) unterschiedlichen Molekulargewichts, hier und in den meisten anderen Fällen in Form von Monomethoxy-PEG (mPEG), mittels toxischem Cyanurchlorid als Linker kovalent an Aminogruppen des bovinen Serumalbumins und der bovinen Leberkatalase geknüpft. Erst 1990 wurde dann das erste Konjugat von der amerikanischen Behörde für Lebens- und Arzneimittel (*Food and Drug Administration*, FDA) zugelassen. Es handelt sich um Adagen®, eine mit 5 kDa PEG konjugierte Adenosin-Desaminase zur Behandlung des schweren kombinierten Immundefekts (SCID) [18]. Aktuell werden 18 von der FDA zugelassene PBK in den USA vermarktet (Tab. 1.1, Stand: 2020), alle auf Basis von PEG [9, 19, 20]. Daneben können aber auch kleinere Moleküle mit Polymeren verknüpft werden. Ein Beispiel hierfür ist Movantik®, ein als Naloxegol bezeichnetes, PEGyliertes Naloxon. Die Konjugation soll hier den Übertritt über die Blut-Hirn-Schranke verhindern [21]. Ebenfalls wurden bereits mit PEG beschichtete, niedermolekulare Wirkstoffe enthaltende Liposomen von der FDA zugelassen: Onivyde® mit Irinotecan und Doxil®/Caelyx® mit Doxorubicin als aktive Komponenten [20, 22]. PEG kann hierbei wesentlich die Protein-Corona des Partikels beeinflussen, d. h. das Ausmaß der Adsorption von Proteinen *in vivo* [23, 24]. 1993 wurde in Japan mit SMANCS/Lipiodol auch ein PBK gegen Leberzellkarzinome zugelassen, welches nicht auf PEG basiert. Es handelt sich um ein an jeweils zwei Aminosäuren mit einem Styrol-Maleimid-Copolymer konjugiertes Neokarzinostatin, welches in einer lipophilen Formulierung intraarteriell appliziert wird, um sich im Tumorgewebe anzureichern [25, 26]. Eine weiterführende Liste zu aktuell in der Entwicklung befindlichen PBK für unterschiedlichste Anwendungsgebiete ist in [27] zu finden.

Tabelle 1.1: Aktuell in den USA vermarktete und von der FDA zugelassene Biokonjugate (Stand: 2020). Handelsnamen, unPEGylierte Wirkstoffe, Indikationsbeispiele und das Jahr der Zulassung, nach dem die Substanzen geordnet sind, sind aufgeführt. In der dritten Spalte ist die Art der PEGylierung näher beschrieben: einfach (E) oder mehrfach (M), sowie linear (L) oder verzweigt (V), inklusive Angabe des Molekulargewichts der PEG-Ketten in kDa [9, 19, 20].

Handelsname	Wirkstoff	PEG	Indikation	Zulassung
Jivi®	Faktor VIII	E/V/60	Hämophilie A	2018
Asparlas™	L-Asparaginase	M/L/5	Akute lymphoblastische Leukämie	2018
Revcovi™	Adenosin-Desaminase	M/L/5.6	SCID	2018
Palynziq®	Phenylalanin-Ammoniak-Lyase	M/V/20	Phenylketonurie	2018
Rebiny®	Faktor IX	E/V/40	Hämophilie B	2017
Adynovate®	Faktor VIII	M/V/20	Hämophilie A	2015
Plegridy®	Interferon- β 1a	E/L/20	Multiple Sklerose	2014
Sylatron™	Interferon- α 2b	E/L/12	Melanome	2011
Krystexxa®	Uricase	M/L/10	Urikopathie	2010
Cimzia®	Certolizumab	E/V/40	Morbus Crohn	2008
Mircera®	Epoetin beta	E/L/30	Renale Anämie	2007
Macugen®	Aptamer gegen vaskulären endothelialen Wachstumsfaktor (VEGF)	E/V/40	Makuladegeneration	2004
Somavert®	Somatropin-Antagonist	M/L/5	Akromegalie	2003
Neulasta®	Granulozyten-Kolonie-stimulierender Faktor (G-CSF)	E/L/20	Neutropenie	2002
Pegasys®	Interferon- α 2a	E/V/40	Hepatitis B und C	2001
Pegintron®	Interferon- α 2b	E/L/20	Hepatitis C	2000
Oncaspar®	L-Asparaginase	M/L/5	Akute lymphoblastische Leukämie	1994
Adagen®	Adenosin-Desaminase	M/L/5	SCID	1990

Im Prinzip stellen PBK mit ihrer direkten Modifikation des betreffenden Arzneistoffs eine mögliche Alternative zur Nutzung von Nanovesikeln dar, deren Bedeutung in den letzten Jahren ebenso stieg. Beide Ansätze verfolgen im Grunde ein ähnliches Ziel: das Erreichen des Zielorts schlecht bioverfügbarer oder besonders instabiler Substanzen zu erleichtern. Somit kann der für die Pharmazie relevante chemische Raum, welcher lange Zeit lediglich durch einige wenige zentrale Dogmen geprägt war (Induced-Fit-Theorie nach Koshland [28], Lipinski's Rule of Five [29, 30]), mit Fokus auf eine verbesserte Pharmakokinetik erweitert werden. Vor allem aber können bereits bestehende Therapien bzgl. dieser Problematiken optimiert werden. Während

Nanovesikel durch den EPR (*Enhanced permeability and retention*)-Effekt die Applikation besonders zytotoxischer Substanzen im Zuge einer Chemotherapie erlauben, wird durch eine Biokonjugation der Abbau sowie die Elimination der Moleküle *in vivo* deutlich verlangsamt [23, 31].

In den vergangenen Jahren haben sich unterschiedliche Polymerisations- und Konjugationsmethoden zur Synthese von PBK etabliert. Während frühere Konjugate unspezifisch an verschiedenen Stellen PEGyliert waren und große Polydispersitäten aufwiesen, existieren heute Verfahren zur ortsspezifischen Konjugation mit maßgeschneiderten Polymeren definierter Länge. Diese Techniken sorgen chargenübergreifend für gleichmäßigere Eigenschaften der Konjugate, was besonders hinsichtlich einer zu vermeidenden Variabilität der Bioverfügbarkeit von Vorteil sein kann [9, 18]. Gleichzeitig stärken diese neuen Methoden zunehmend auch den Wunsch nach systematischeren Vorgehensweisen zur Synthese von Konjugaten mit bestmöglichen Eigenschaften. Letztere sind grundsätzlich abhängig vom Konjugationsort, der Anzahl an verknüpften Polymerketten, der Linkerstruktur und dem Aufbau sowie der Länge der Polymere. Rationale Ansätze zur Herstellung von PBK profitieren vor allem von computertechnischen Methoden, mit denen das Verhalten auf molekularer Ebene näher analysiert werden kann [32]. Die bisherigen Befunde dieses Forschungsgebiets werden in Kap. 1.1.6.1 näher erläutert.

PBK weisen in aller Regel eine bessere Löslichkeit, ein vermindertes Aggregationsverhalten und eine erhöhte thermische und proteolytische Stabilität auf und werden durch ihren größeren hydrodynamischen Radius schlechter renal filtriert, sodass sie länger im Blutkreislauf verbleiben. Ebenso wird das immunogene Potenzial des Wirkstoffs gesenkt [33]. Es wird vermutet, dass PEG wie ein Schutzschild fungiert und das Protein vor proteolytischen Enzymen und Antikörpern, die an Epitope des Arzneistoffs binden könnten, schützt. Dahingehend wird in der Literatur auch oft von einem sogenannten Stealth-Effekt geschrieben, der im Hinblick auf Nanovesikel auch einer möglichen Bindung von Albumin durch PEG zugeschrieben wird [23]. Zusätzlich soll das Polymer wie eine Art Sieb wirken können: Kleinere Moleküle diffundieren bis zur Proteinoberfläche, während größere Moleküle, wie bspw. Antikörper, bereits an der Polymeroberfläche abgeblockt werden [9]. Dieser Effekt wurde an α -Chymotrypsin, das mit Dendrimeren verschiedener Generationen konjugiert worden war, studiert [9, 34].

Neben den genannten Vorteilen zeigen PBK jedoch zumeist eine verminderte Aktivität. Vermutlich beruht dies auf einer sterischen Abschirmung der Binderegionen des Makromoleküls durch das Polymer [35]. Es wurde jedoch auch bereits eine durch die Konjugation modulierte Proteindynamik für die Aktivitätsminderung verantwortlich gemacht, weshalb die Nähe der Konjugationsstelle zum aktiven Zentrum nicht zwangsläufig der entscheidende Faktor für die Aktivitätsmodulation ist [36]. Ebenfalls berichten einige Forscher von anti-PEG-Antikörpern. Dies kann zur beschleunigten Filtration der Wirkstoffe aus dem Blutkreislauf und zu Hypersensibilitätsreaktionen führen.

Omontys®[®], ein PEGyliertes Erythropoeitin-Mimetikum, wurde aufgrund schwerer Immunreaktionen vom Markt zurückgezogen [20, 37]. Dennoch werden Biokonjugate, vornehmlich aufgrund der durch die verringerte Clearance deutlich verlängerten Halbwertszeit im Blutkreislauf, zunehmend eingesetzt. Hierdurch kann die Frequenz der parenteralen Applikationen verringert werden, wodurch die Therapie für den Patienten deutlich erleichtert wird [18]. Die Vor- und Nachteile einer Konjugation sind in Tab. 1.2 zusammengefasst.

Tabelle 1.2: Mögliche Vor- und Nachteile einer Polymer-Biokonjugation an therapeutisch verwendete Biologika [18, 20, 33, 37].

Vorteile	Nachteile
Erhöhte Löslichkeit	Verringerte Bioaktivität
Verminderte Immunogenität	Immunogenes Potenzial
Verminderte Aggregation	Hohe Produktionskosten
Verminderte Clearance	Variable Bioverfügbarkeit
Erhöhte proteolytische Stabilität	Potenzielle Toxizität
Thermische Stabilisierung	Thermische Instabilisierung

1.1.2 Herstellungsverfahren

1.1.2.1 Grafting-Methoden

Grundsätzlich wird zwischen drei verschiedenen Ansätzen der Biokonjugation unterschieden (Abb. 1.1): Die *Grafting to*-Methode ist die am häufigsten angewandte. Hierbei wird das Polymer separat synthetisiert und anschließend über eine geeignete Linkerstruktur an das Protein geknüpft. Dadurch sind Polymersynthesen auch in apolaren Lösungsmitteln möglich und die Makromoleküle können vor der Konjugation charakterisiert werden, jedoch ist die Ausbeute der Konjugation in der Regel gering, da zwei Makromoleküle miteinander reagieren müssen. Bei der *Grafting from*-Methode wird ein geeigneter Initiator in das zu modifizierende Protein integriert, sodass die Polymerisation direkt am Makromolekül stattfinden kann. Sie muss daher in einem für das Protein geeigneten Milieu stattfinden, die Ausbeute ist jedoch größer als bei der *Grafting to*-Methode und die Produkte können leicht von den niedermolekularen Monomeren abgetrennt werden. Die letzte Variante stellt die *Grafting through*-Methode dar und wird aufgrund des geringen Polymerisationsgrades nur selten genutzt. Hierbei werden die Proteine vor der Polymerisation mit Monomeren verknüpft. Es bilden sich dadurch Polymerketten mit mehreren kovalent gebundenen Proteinen [9, 12, 32, 38].

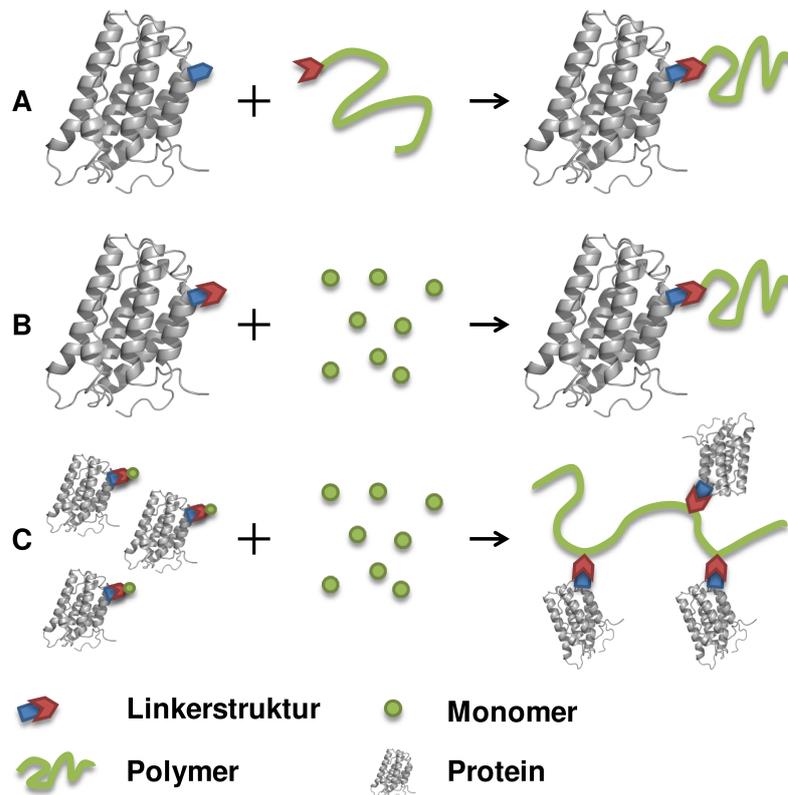


Abbildung 1.1: Schematische Darstellung der (A) *Grafting to*-, (B) *Grafting from*- und (C) *Grafting through*-Methoden [12].

1.1.2.2 Polymersynthese

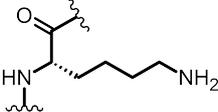
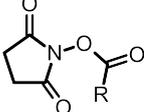
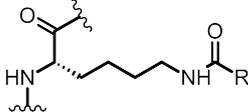
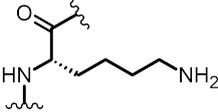
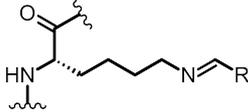
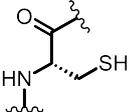
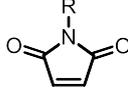
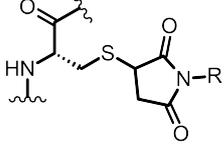
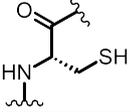
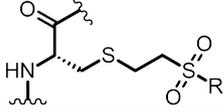
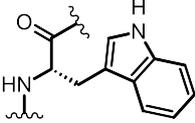
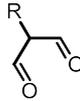
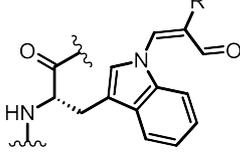
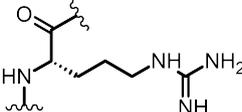
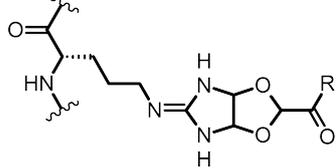
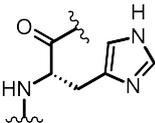
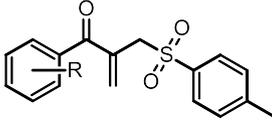
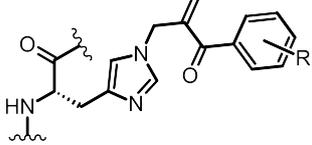
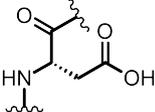
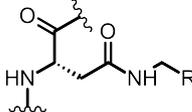
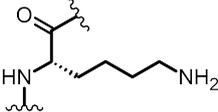
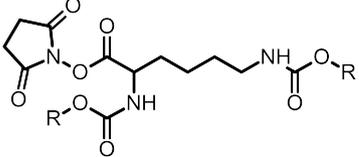
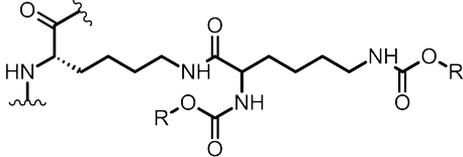
Zur Synthese der Polymere werden vor allem bei der *Grafting from*-Methode oft kontrolliert radikalische Polymerisationen (CRP) durchgeführt, welche die Bildung äußerst komplexer makromolekularer Architekturen erlauben. Bei der Atomtransferferradikalpolymerisation (ATRP) wird ein organisch gebundenes Halogenid an einen Cu^{I} -Liganden, z. B. Tris(2-pyridylmethyl)amin (TPMA), mittels Redoxreaktion übertragen. Dadurch entsteht ein zur Polymerisation geeignetes Radikal, welches durch Bindung des Halogenids des Cu^{II} -Liganden wieder deaktiviert wird. Es herrscht ein gut kontrollierbares Gleichgewicht zwischen der aktiven und der inaktiven Polymerisation. Als eine Alternative zur ATRP kann die reversible Additions-Fragmentierungs-Kettenübertragungs-Polymerisation (RAFT) angesehen werden. Hierbei stellt sich ein Gleichgewicht unter Verwendung eines Kettentransferreagenzes (CTA) ein, wofür in der Regel eine Thiocarbonylthioverbindung genutzt wird. CTAs oder ATRP-Initiatoren können z. B. mittels N-Hydroxysuccinimid (NHS)-aktivierter Carbonsäuren an das Makromolekül geknüpft werden [9, 12, 32, 38].

1.1.2.3 Konjugationsstrategien

Mithilfe verschiedener Linkerstrukturen und variabler Reaktionsbedingungen können die Polymerketten bei der *Grafting to*-Methode an unterschiedliche Aminosäuren eines Proteins konjugiert werden. Eine Auswahl ist als Übersicht in Tab. 1.3 dargestellt. Besonders Aminogruppen von Lysinen dienen z. B. über Bildungen von Schiffischen Basen mit Aldehyden oder NHS-vermittelte Amidierungen als Konjugationsstellen. Die Aminogruppe des N-Terminus kann spezifisch bei pH-Werten adressiert werden, bei denen Lysine noch protoniert vorliegen. Cysteine stellen gute Nukleophile und, aufgrund ihres relativ seltenen Vorkommens in freier Form, zumeist auch spezifische Konjugationsorte in Proteinen dar, die über die Bildung eines Thioethers bspw. mit Maleimiden oder Vinylsulfonen reagieren können. Alternativ können auch Disulfidbrücken spezifisch angegriffen werden und zu überbrückten PEGylierungen führen. Möglich sind ebenfalls Konjugationen über Histidine, die in manchen Fällen auch zu einer Art Prodrug führen können, da instabile Acyl-Histidin-Bindungen *in vivo* in einer allmählichen Freisetzung des unkonjugierten Wirkstoffs resultieren (wie bspw. bei Pegintron®). Zu erwähnen sind schließlich auch komplexere Linkerstrukturen, die es ermöglichen, verzweigte PEG-Ketten mit größerem Stealth-Effekt zu konjugieren [12, 18, 19, 39].

Auch wenn die 20 natürlich vorkommenden Aminosäuren bereits eine breite Palette an möglichen Konjugationen erlauben, kann eine ortsspezifische Verknüpfung leichter über nicht-natürliche Aminosäuren erreicht werden. Diese können mittels genetischer Modifizierung über ein Amber-Codon während der Translation an der gewünschten Position in die Aminosäuresequenz eingebaut werden [40]. Alternativ existieren für kleinere Peptide auch Einbaumöglichkeiten über Festphasenpeptidsynthesen. Nicht-natürliche Aminosäuren ermöglichen den Einsatz eleganter und selektiver Click-Chemie-Verfahren [12, 41]. Ein Beispiel stellt N_ϵ -[(2-Azidoethoxy)carbonyl]-L-lysin (AZK, auch NAEK genannt) dar, welches über seine Azidfunktion mittels 1,3-dipolarer Huisgen-Cycloaddition an Alkingruppen geknüpft werden kann. In umgekehrter Weise kann auch N_ϵ -[(Allyloxy)carbonyl]-L-lysin verwendet werden [40, 42, 43]. Alternativ zur zumeist kupferkatalysiert durchgeführten Addition kommen Cyclooctinderivate ohne toxikologisch bedenkliche Metallkatalysatoren aus, indem bioorthogonale, ringspannungsgeförderte Azid-Alkin-Cycloadditionen (*Strain-promoted azide-alkyne click chemistry reaction*, SPAAC) durchgeführt werden [41]. Beispiele für solche Derivate sind der Dibenzocyclooctin (DBCO)- und der Bicyclononin (BCN)-Linker (Abb. 1.2) [12]. Erst kürzlich wurden als Alternative zu bereits etablierten ringspannungsgeförderten Cycloadditionen auch Konjugationen mittels trans-Cyclohepten über inverse Elektronenbedarfs-Diels-Alder-Reaktionen beschrieben [44].

Tabelle 1.3: Beispiele möglicher Kupplungsreagenzien und daraus entstehende Produkte für verschiedene Aminosäuren. R stellt dabei den Polymerrest dar [12, 19].

Aminosäure	Kupplungsreagenz	Kupplungsprodukt	Referenzen
			[45–48]
			[49, 50]
			[51, 52]
			[53]
			[54]
			[55]
			[56]
			[57, 58]
			[19]

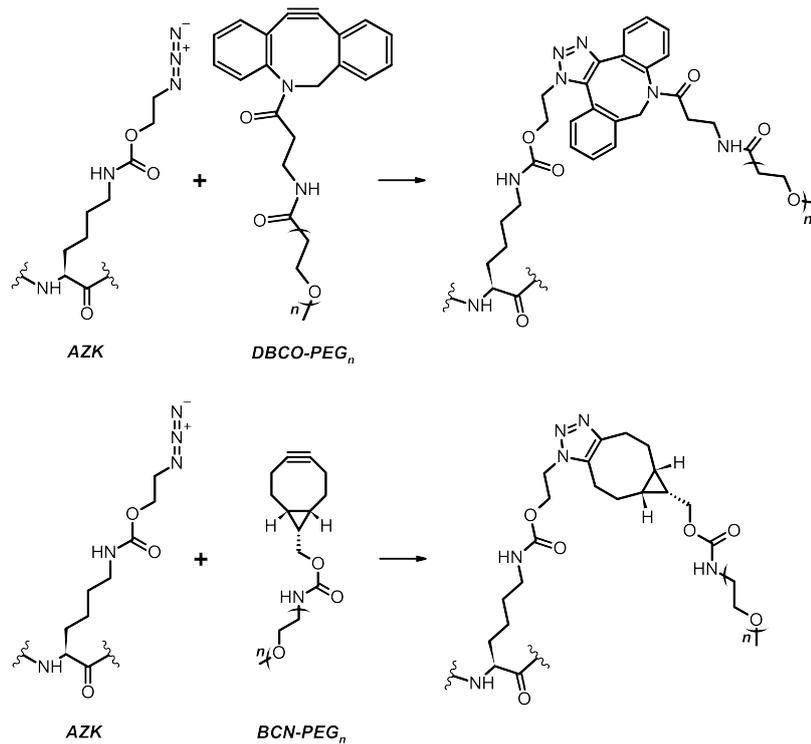


Abbildung 1.2: Kupplungsreaktionen für PEGylierte DBCO- und BCN-Linker mit AZK.

Neben kovalenten Verknüpfungen von Polymeren an Makromoleküle kann eine ortsspezifische Bindung auch nichtkovalent über Biotin-Streptavidin- oder HisTag-Ni²⁺-Nitrilotriessigsäure-Wechselwirkungen erfolgen [12]. Schließlich können auch enzymatische Kupplungsreaktionen verwendet werden, die aufgrund ihrer Sequenzabhängigkeit zur ortsspezifischen Konjugation geeignet sind. Hierfür werden bspw. Transglutaminasen oder auch Tyrosinasen genutzt [18, 41, 59]. Berücksichtigt man auch kovalente Konjugationen peptidischer Elemente an Proteine, so sollte schließlich auch die SpyTag/SpyCatcher-Methode erwähnt werden, bei der eine Isopeptidbindung zwischen einer Asparaginsäure des kleinen SpyTag-Peptids und einer Lysinseitenkette des SpyCatcher-Proteins gebildet wird. Diese Technologie wird z. B. zur Herstellung von Vakzinen und Protein-Hydrogelen genutzt [60].

1.1.3 Polymerarten

Die Welt der Polymere bietet schier unendliche Kombinationen möglicher biokompatibler Strukturen, die zur Konjugation verwendet werden können. Im Folgenden sollen nur einige kurz erwähnt werden.

1.1.3.1 Polyethylenglykol

PEG ist das für Biokonjugationen am häufigsten genutzte Polymer. Es handelt sich um ein, je nach Molekulargewicht, in der Regel gut wasserlösliches Makromolekül mit simpler, linearer Struktur, welches in der Pharmazie schon seit langer Zeit eine breite Anwendung findet. Es kann unter anderem als Cosolvens, Bindemittel, Gelbildner

oder auch Wirkstoff gegen Obstipation verwendet werden [61]. Aufgrund der bereits langen Nutzung wird es von der FDA allgemein als biokompatibel, nicht toxisch und Goldstandard zur Biokonjugation angesehen. Die Etherstrukturen resultieren in einer starken Hydratation und damit in einem größeren hydrodynamischen Radius, wodurch die in Kap. 1.1.1 schon erwähnten Vorteile entsprechender PBK realisiert werden. Der amphiphile Charakter des Polymers ermöglicht es ebenfalls, Konjugate für apolare Lösungsmittel zur Durchführung organischer Synthesen herzustellen [26, 62, 63].

Durch die intensive Anwendung dieses Polymers sind jedoch auch die Nachteile dieses Strukturtyps hinlänglich bekannt. In Patienten nachgewiesene anti-PEG-Antikörper vom Typ IgG und IgM können bei wiederholten Gaben zur schnelleren Clearance des Wirkstoffs (*Accelerated blood clearance*, ABC-Phänomen) und schweren Immunreaktionen führen [20, 26, 37]. Eine kürzlich publizierte Kristallstruktur illustriert die Bindung des flexiblen Polymers an einen Antikörper [64]. Anaphylaktische Reaktionen aufgrund einer PEG-Allergie wurden auch bei Arzneimitteln beobachtet, in denen das Polymer als Hilfsstoff, z. B. zur Stabilisation von Lipidnanopartikeln, eingesetzt wird, so auch im Falle des gegen das Coronavirus SARS-CoV-2 verwendeten mRNA-Vakzins Tozinameran. Auch handelsübliche Pflegeprodukte wie Shampoos enthalten oft PEG [65]. Eine Prophylaxe durch Gabe von freien PEG-Ketten könnte den Titer an Antikörpern jedoch für eine gewisse Zeit verringern [66]. Ebenso sind die zur Biokonjugation genutzten PEG-Ketten nicht biologisch abbaubar, wodurch sich Polymerrückstände in Leber-, Nieren- und Milzzellen ansammeln und Vakuolen bilden können [26]. Aufgrund dieser Nachteile wurden bereits verschiedene biokompatible PEG-Alternativen für die Konjugation in Betracht gezogen. Sie lassen sich in biologisch nicht abbaubare und natürlich abbaubare Polymere einteilen. Ausgewählte Strukturen sind in Abb. 1.3 gezeigt. Bisher existieren zu diesen jedoch deutlich weniger *in vivo*-Studien.

1.1.3.2 Polyvinylpyrrolidon

Polyvinylpyrrolidon (PVP) stellt eine hydrophile, nicht toxische Alternative zu PEG dar. Es kann durch eine RAFT-Polymerisation synthetisiert werden. Ein Vergleich der Biodistributionen konjugierter Tumornekrosefaktoren mit unterschiedlichen Polymerarten ergab für das PVP-Konjugat eine im Vergleich zur PEG-Variante erhöhte Halbwertszeit im Blutkreislauf [63, 67]. Allerdings zeigte PVP bei der Untersuchung konjugierter Uricasen auch eine im Vergleich zu PEG erhöhte Immunogenität [68, 69].

1.1.3.3 Polyglycerole

Polyglycerol (PG) ist ein hydrophiles PEG-Derivat mit zusätzlichen Hydroxymethylseitenketten, das in Studien eine sehr gute Zell- und Hämokompatibilität und eine erhöhte Halbwertszeit im Vergleich zu PEG zeigte [63, 70, 71]. Bisher wurde auch *in vivo* keine nennenswerte Toxizität beobachtet [72]. Ein weiterer Vorteil zu PEG liegt

in der geringeren Viskosität [73]. PG kann mittels Glycidol über eine Ringöffnungspolymerisation erhalten werden. Durch die freie Hydroxymethylseitenkette des linearen Polyglycerols (LPG) können im Gegensatz zu PEG auch weitere Derivatisierungen erfolgen [70]. Hyperverzweigte Polyglycerole stellen kompakte, äußerst komplexe Architekturen dar und zeigen eine längere Halbwertszeit im Blutkreislauf als LPG [71]. Allerdings sind auch PG-Ketten nicht bioabbaubar und können sich in der Leber und der Milz anreichern [63, 72].

1.1.3.4 Polyoxazoline

Ebenfalls nicht bioabbaubar sind Poly(2-alkyl-2-oxazoline) (POx), welche durch eine kationische Ringöffnungspolymerisation (CROP) erhalten werden können [74]. Sie sind schon seit über 50 Jahren bekannt, ihre Bedeutung im Bereich der Biowissenschaften nahm jedoch erst im letzten Jahrzehnt deutlich zu [75]. Wie bei PG ergibt sich hier durch die Derivatisierung der Oxazolinseitenkette eine große strukturelle Vielfalt. Es können bspw. Poly(2-methyl-2-oxazolin) (PMeOx) und Poly(2-ethyl-2-oxazolin) (PEtOx) synthetisiert werden und mittels dieser Variation der Seitenkettenlänge die Polarität des Polymers beeinflusst werden [76]. POx wird ein ähnlicher Stealth-Effekt wie PEG zugerechnet, und stark verlängerte Halbwertszeiten im Blutkreislauf wurden bereits für POxylierte G-CSF-Konjugate beschrieben [26, 63]. Auch wenn das Polymer nicht bioabbaubar ist, so wurde bereits von einer kompletten Ausscheidung eines radionuklidmarkierten Polymers berichtet [77], und es wurden in Hasen eine geringere Immunogenität als PEG und keine histopathologischen Veränderungen von Geweben festgestellt [78]. PEtOx-Ketten können eine geringere Viskosität als PEG aufweisen und sind in verschiedensten Medien löslich. Aufgrund der Vorteile wurden bereits eine Reihe von Biokonjugaten auf Basis von POx hergestellt [76]. PMeOx gilt als etwas hydrophiler, PEtOx hingegen als etwas lipophiler als PEG. Entsprechende Strukturen wurden vor einigen Jahren erstmals auf atomarer Ebene mittels Molekulardynamik-Simulationen untersucht [79].

1.1.3.5 Weitere Polymere

Ein Beispiel für ein synthetisches, bioabbaubares PEG-Derivat stellt das Pfropfcopolymer Poly(oligo(ethylenglykol)methacrylat) (POEGMA) dar, dessen PEG-Ketten *in vivo* aufgrund der Esterbindung an die Hauptkette abgebaut werden können. Konjugate dieses Polymers zeigen eine stark verlängerte Halbwertszeit im Blutkreislauf und eine geringfügig verminderte Aktivität [26, 63]. Auch die Immunogenität der PEG-Strukturen kann bei diesem Strukturtyp vermindert sein [80].

Schließlich ist auch zu erwähnen, dass, so wie auch Glykosylierungen von Proteinen in Zellen stattfinden, viele zuckerbasierte Konjugationsmöglichkeiten existieren. Bspw. wurden bereits Konjugate von Erythropoetin, G-CSF und Interferon- α mit Hydroxyethylstärke (HES) hergestellt. Diese zeigten verbesserte pharmakokinetische Eigenschaften. HES kann als Plasmaexpander genutzt und durch α -Amylase abgebaut

werden. Allerdings mehren sich Berichte zu toxischen Akkumulationen von HES in Leber und Nieren [81]. Eine Alternative zu dieser Struktur stellt bspw. Dextrin dar, welches bereits mit endothelialen Wachstumsfaktoren konjugiert wurde. Aber auch Polysialinsäuren sind als mögliche Alternativen zu nennen [26].

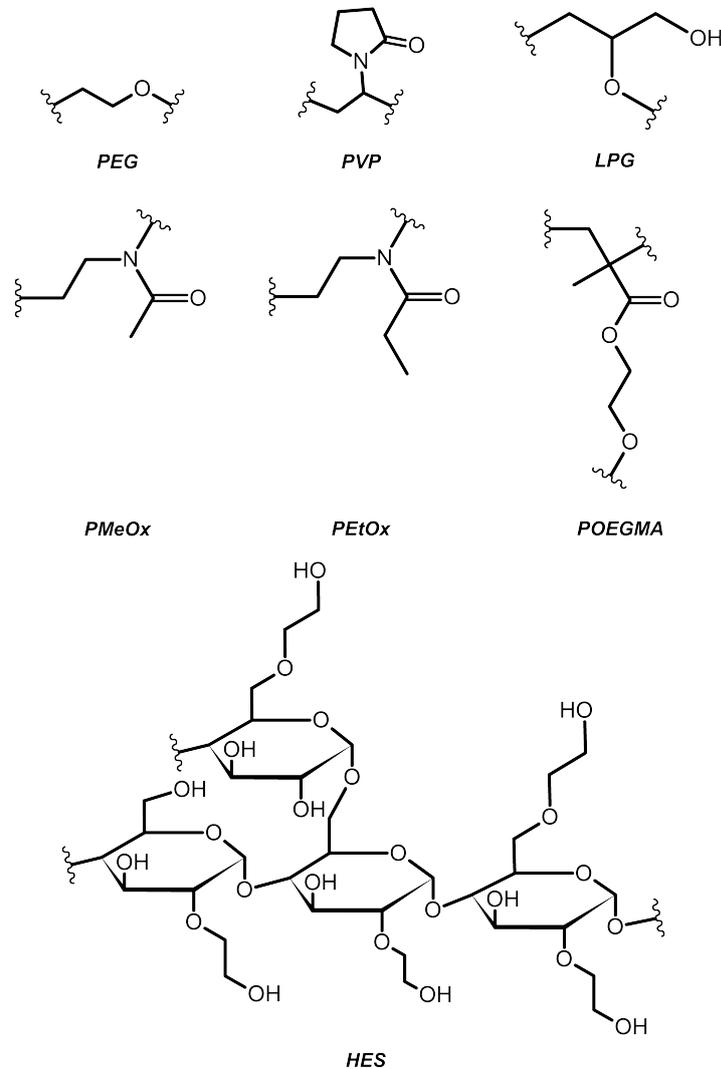


Abbildung 1.3: Strukturformeln ausgewählter Polymere (Bezeichnungen gemäß Text).

1.1.4 Therapeutische Beispiele

Im Folgenden sollen einige der in Tab. 1.1 aufgeführten PBK kurz beschrieben werden.

1.1.4.1 Adagen®

Wie in Kap. 1.1.1 erwähnt stellt Adagen® das erste vermarktete PBK dar. Mutationen des für den Purinmetabolismus wichtigen Proteins Adenosin-Desaminase können zu einem SCID führen. Eine mit 5 kDa PEG konjugierte bovine Variante des Proteins dient zur Therapie. In Adagen® sind zwischen 11 und 17 Polymere, die zuvor mit einer Succinimidylsuccinatgruppe für die Konjugation aktiviert wurden, unspezifisch

an Lysinseitenketten geknüpft. Durch die deutlich erhöhte Halbwertszeit sind mehrmalige Bluttransfusionen nicht mehr notwendig und das Biokonjugat muss lediglich einmal wöchentlich appliziert werden [18, 19].

1.1.4.2 Neulasta®

Neulasta® stellt das erste vermarktete Konjugat dar, welches ortsspezifisch konjugiert ist. Es handelt sich um PEGyliertes, rekombinant hergestelltes, humanes G-CSF, das bei Neutropenie eingesetzt wird. Zur Synthese wird 20 kDa PEG mit einer terminalen Aldehydfunktion selektiv an die N-terminale Aminogruppe von Methionin über eine reduktive Aminierung unter sauren Bedingungen geknüpft. Auch hier ergeben sich durch die PEGylierung verbesserte pharmakokinetische Eigenschaften und damit eine erleichterte Therapie: Das Arzneimittel muss im Gegensatz zur unPEGylierten Form nur einmalig subkutan appliziert werden [18, 19].

1.1.4.3 Pegasys®

Pegasys® ist ein Beispiel für ein Konjugat mit verzweigter PEG-Kette. Es handelt sich um Interferon- α 2a, das gegen Hepatitis B und C eingesetzt wird und an seinen solvenszugänglichen Lysinseitenketten (K31, K121, K131 und K134, selten auch an K70 und K83) mit der in Tab. 1.3 am Ende aufgeführten PEG₂-NHS-Struktur, die aus einem an beiden Aminogruppen verknüpften Lysin besteht, konjugiert ist. Die Konjugate an den Positionen K31 und K134 zeigen die größte Aktivität, die für Pegasys® insgesamt dennoch nur 7 % beträgt (bezogen auf die unPEGylierte Variante). Trotzdem ist der antivirale Effekt durch eine einmal wöchentliche Applikation aufgrund der längeren Halbwertszeit stärker ausgeprägt als bei dem unkonjugierten Protein, welches dreimal wöchentlich appliziert werden muss. Die Clearance der unPEGylierten Variante beträgt 6000 ml/h, jene von Pegasys® 60 bis 100 ml/h [18, 19].

1.1.4.4 Macugen®

Zuletzt soll noch das intravitreal applizierte Macugen® erwähnt werden. Es handelt sich hierbei um ein Anti-VEGF-Aptamer aus 28 Nukleotiden, das gegen eine Makuladegeneration eingesetzt werden kann. Es stellt somit das bisher einzige nukleinsäurebasierte, in den USA zugelassene Biokonjugat dar. Aptamere können *in vivo* schnell von Nukleasen abgebaut werden. Auch hier wird die Halbwertszeit durch die PEGylierung mit einer verzweigten 40 kDa PEG-Struktur deutlich erhöht. Die beiden linearen PEG-Ketten sind, wie im Falle von Pegasys®, zunächst an die Aminogruppen eines Lysins geknüpft. Dieses bildet hier jedoch kein Amid mit der Seitenkette eines weiteren Lysins, sondern ist über eine kurze Alkylkette mit der Phosphatgruppe des 5'-Ende des Aptamers verknüpft [18, 19].

1.1.5 Thermodynamische Charakterisierung

Wie in Kap. 1.1.1 angemerkt ist die Funktion eines biologischen Makromoleküls maßgeblich von dessen dreidimensionaler Struktur abhängig. Insofern ist die thermodynamische Stabilität von besonderer Bedeutung für therapeutisch genutzte Biologika und sollte einen wichtigen Aspekt in der Entwicklung neuer Biokonjugate darstellen. Zumeist wird eine simple Zwei-Phasen-Kinetik angenommen, bei der das ungefaltete bzw. denaturierte Protein beim Faltungsprozess einen Übergangszustand hin zur energetisch stabileren, gefalteten bzw. nativen Konformation durchläuft (Abb. 1.4). In der Realität kann das Energiediagramm durch viele verschiedene Übergangszustände oder auch metastabile Konformationen eine deutlich größere Komplexität aufweisen.

Die thermodynamische Stabilität kann durch die Differenz der Gibbs-Energie zwischen beiden Endzuständen beschrieben werden. Sie steht in unmittelbarem Zusammenhang mit der Equilibriumskonstanten K , sowie den dazugehörigen kinetischen (Ent-)faltungskonstanten k_f und k_u (Glg. 1.1). Schon geringe Änderungen in der Proteinsequenz können die Stabilität des Wildtyps verändern und dadurch entscheidende biologische und physikochemische Eigenschaften beeinflussen. So basiert das Vorkommen verschiedener Krankheiten bspw. auf Punktmutationen in Proteinen. Gleichzeitig erlauben Modifikationen, die die thermodynamische Stabilität von Proteinen optimieren, das Vorkommen thermophiler Bakterien bei erhöhten Temperaturen. Zur Charakterisierung dieser Stabilität existieren verschiedenste experimentelle Methoden, von denen zwei im Folgenden kurz vorgestellt werden sollen [82–85].

$$\Delta G^0 = -RT \ln K = -RT \ln\left(\frac{k_f}{k_u}\right) \quad (1.1)$$

1.1.5.1 Temperaturabhängige CD-Spektroskopie

Eine temperaturabhängige Circular dichroismus (CD)-Spektropolarimetrie kann zum Studium des Effekts der Konjugation auf die thermodynamische Stabilität des Proteins herangezogen werden. Zur Untersuchung der Sekundärstrukturen von Proteinen wird hierbei zirkular polarisiertes Licht unterschiedlicher Wellenlängen, welches aus einer links- und einer rechtsdrehenden Komponente besteht, durch eine Lösung des Makromoleküls gestrahlt. Optisch aktive Strukturen absorbieren beide Komponenten in unterschiedlichem Maße. Dies kann durch chirale Moleküle, aber auch durch asymmetrische Sekundärstrukturen wie α -Helices und β -Faltblätter verursacht werden [86]. Die Differenz ΔA zwischen den Absorptionen der links- und rechtsdrehenden Lichtanteile $A_{L/R}$ ist gemäß dem Lambert-Beer'schen Gesetz proportional zu den Extinktionskoeffizienten $\epsilon_{L/R}$ (Einheit: $\frac{dm^3}{mol \cdot cm}$), multipliziert mit der Konzentration m in $\frac{mol}{dm^3}$ und der Küvettenlänge l in cm:

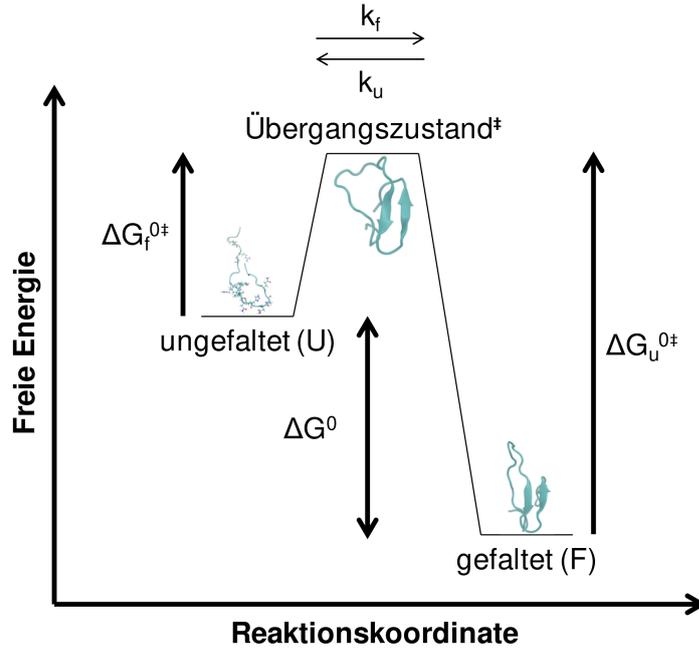


Abbildung 1.4: Schematische Darstellung des Faltungsprozesses von Proteinen auf Basis einer Zwei-Phasen-Kinetik.

$$(A_L - A_R) = (\epsilon_L - \epsilon_R) m l \quad (1.2)$$

Die Elliptizität $[\theta]_\lambda$ der Wellenlänge λ (gleiche Einheit wie l), angegeben in Radianten, kann wiederum unter Nutzung der entsprechenden Absorptionsindizes $n'_{L/R}$ aus dem Verhältnis der Achsenlängen der aus der unterschiedlichen Absorption resultierenden Ellipse des polarisierten Lichts berechnet werden. Angegeben in Grad mithilfe der Absorptionen $A_{L/R}$ ergibt sich schließlich ein Umrechnungsfaktor von 32.98 [87].

$$[\theta]_\lambda = \tan^{-1}\left(\frac{b}{a}\right) \quad (1.3)$$

$$[\theta]_\lambda \text{ (Radianten)} = \frac{\pi l}{\lambda} (n'_L - n'_R) \quad (1.4)$$

$$= \frac{2.303 l m}{4} (\epsilon_L - \epsilon_R) \quad (1.5)$$

$$[\theta]_\lambda \text{ (Grad)} = 32.98 m l (\epsilon_L - \epsilon_R) \quad (1.6)$$

$$= 32.98 (A_L - A_R) \quad (1.7)$$

$$= 32.98 \Delta A \quad (1.8)$$

Dabei stellt b die kürzere und a die längere Achse der Ellipse dar. Die durchschnittliche Elliptizität pro Aminosäure $[\theta]_{mrw,\lambda}$ (*Mean residue ellipticity*, Angabe in $\frac{\text{deg cm}^2}{\text{dmol res}}$) und die molare Elliptizität $[\theta]_{molar,\lambda}$ (Angabe in $\frac{\text{deg cm}^2}{\text{dmol}}$) können durch folgende Umrechnungen erhalten werden:

$$[\theta]_{mrw,\lambda} = \frac{[\theta]_{\lambda} MRW}{10 l C} = \frac{[\theta]_{\lambda} \frac{M}{N-1}}{10 l C} \quad (1.9)$$

$$[\theta]_{molar,\lambda} = \frac{100 [\theta]_{\lambda}}{m l} \quad (1.10)$$

Dabei sind M das Molekulargewicht in g/mol, N die Anzahl an Aminosäuren, $N - 1$ die Anzahl an Peptidbindungen, $\frac{M}{N-1}$ das davon abgeleitete mittlere Aminosäuregewicht (*Mean residue weight*, MRW), l die Küvettenlänge in cm, C die Konzentration in g/ml und m die molare Konzentration in mol/l.¹ Typische Sekundärstrukturen ergeben im Bereich von etwa 190 bis 240 nm charakteristische Banden in einem CD-Spektrum, welches zur schnellen Überprüfung der Proteinstabilität oder einer ersten Information im Rahmen einer Strukturaufklärung dienen kann. Für die Absorption unter 240 nm sind vor allem $n \rightarrow \pi^*$ (220 nm)- und $\pi \rightarrow \pi^*$ (190 nm)-Übergänge der Peptidbindungen verantwortlich [86, 88].

Da die Elliptizität maßgeblich durch die Sekundärstrukturen bestimmt wird, kann die Proteinentfaltung mit einer temperaturabhängigen CD-Spektroskopie untersucht werden. Hierzu werden $[\theta]_{mrw,\lambda}$ oder $[\theta]_{molar,\lambda}$ für eine festgelegte Wellenlänge (für α -helikale Strukturen normalerweise bei 222 nm, für β -faltblattstrukturreiche Proteine bei 218 nm) gegen die Temperatur aufgetragen. Die Messungen werden jeweils nach einer Equilibrierungszeit von bspw. 2 Minuten für jeden Temperaturwert in einem festgelegten Intervall (z. B. von 1 bis 95 °C, mit 2 °C-Sprüngen) durchgeführt. Erhalten wird eine Entfaltungskurve, mit deren Wendepunkt die Schmelz- bzw. Entfaltungstemperatur T_m des Makromoleküls bestimmt werden kann. Über eine Van-'t-Hoff-Analyse können daraus wichtige thermodynamische Parameter abgeleitet werden, so z. B. die freie Faltungsenergie ΔG_f^0 . Weitere Details zu den hierfür notwendigen Ausgleichsrechnungen sind in Anhang A aufgeführt.

1.1.5.2 Thermal Shift Assay

Eine Alternative zur temperaturabhängigen CD-Spektroskopie stellt ein fluoreszenzbasierter Thermal Shift Assay (TSA) bzw. die Differential Scanning Fluorimetry-Methode (DSF) dar. Hierbei wird die Temperatur einer Lösung des Proteins ebenfalls sukzessive erhöht und währenddessen die allmählich ansteigende Fluoreszenzintensität, die aus einem geeigneten proteinbindenden Farbstoff wie bspw. SYPRO™ Orange resultiert, detektiert. Über den Wendepunkt, der über die erste Ableitung ermittelt werden kann, wird die Schmelztemperatur erhalten. Alternativ kann hierzu die Eigenfluoreszenz von Tryptophanen oder Tyrosinen genutzt werden [89–91].

¹In diesem Abschnitt wurden die Bezeichnungen aus [86] und [87] verwendet. Üblicherweise wird die Stoffmengenkonzentration mit c gekennzeichnet, die Massenkonzentration hingegen mit β . Gemäß der angegebenen Literatur ist die Angabe der SI-Einheit *mol* in Form von Dezimol (*dmol*) und die Verwendung von Klammern für die Größen $[\theta]_{\lambda}$, $[\theta]_{mrw,\lambda}$ und $[\theta]_{molar,\lambda}$ in der CD-Spektroskopie üblich.

1.1.6 Stabilisierungsmechanismen

1.1.6.1 Bisherige Hypothesen

Eine Vielzahl an publizierten Arbeiten beschäftigt sich mit der Frage, wie es zur Erhöhung der thermischen bzw. konformativen Stabilität eines Proteins durch die Konjugation mit PEG kommen kann. 2008 bestätigten Rodríguez-Martínez et al. mithilfe von Wasserstoff-Deuterium (H/D)-Austauschexperimenten unter Verwendung von Fourier-Transform-Infrarotspektroskopie an konjugiertem α -Chymotrypsin die bereits bestehende Hypothese, dass eine PEGylierung mit einer Reduktion der Proteindynamik einhergehen kann. Ebenfalls wurde beobachtet, dass der Stabilisierungseffekt ein Plateau ab vier Monomeren erreicht und bei längeren Polymerketten stagniert [92]. Molekulardynamik (MD)-Simulationen aus dem Jahre 2011 von Yang et al. mit konjugiertem Insulin zeigten eine stabilere Sekundärstruktur und eine bei erhöhten Temperaturen niedrigere mittlere Abweichung (RMSD) von der nativen Proteinstruktur, sowie einen Rückgang an solvenszugänglicher Oberfläche (SASA) des Proteins. In diesen Simulationen interagierte das Polymer hauptsächlich über hydrophobe Wechselwirkungen mit dem Protein, während es eine Vielzahl an Wasserstoffbrückenbindungen (HB) zum Solvens ausbildete [93]. Eine Reduktion der SASA ist zu erwarten und wurde auch in anderen Publikationen beschrieben. Ein solcher Effekt könnte die Proteinerückgrat-HB durch Abschirmung des Solvens verstärken [94]. Weng et al. studierten 2012 die Stabilität PEGylierter SH3-Domänen mithilfe von Guanidiumchlorid als Denaturierungsreagenz und leiteten hiervon ein physikalisches Modell zur Erklärung der Stabilisierung ab. Dieses Modell beschäftigt sich mit dem Verhältnis zwischen proteinoberflächennahen Polymer- und Wasseratomen und basiert auf der Annahme, dass die PEGylierung die SASA der entfalteten und der nativen Proteinstruktur moduliert. Da das für das Solvens zugängliche Volumen in unmittelbarer Umgebung des Proteins bei der Entfaltung steigt (*Excluded volume*-Effekt bei Faltungsvorgängen), das eingenommene Volumen der Polymerkette jedoch gleich bleibt, wird die SASA des Proteins im gefalteten Zustand durch die PEG-Kette stärker beeinflusst als im denaturierten. Die Autoren schlussfolgerten, nach Analyse des sogenannten *m*-Werts, der die Faltungskooperativität des Proteins beschreibt und mit der SASA-Modulation korreliert, dass die Stabilisierung maßgeblich aus einer Senkung der Entfaltungsgeschwindigkeit (ausgedrückt durch die kinetische Konstante k_u) resultiert. Auch diese Arbeit zeigte auf, dass der Stabilisierungseffekt vom Molekulargewicht der größeren PEG-Ketten unbeeinflusst blieb, jedoch Konjugationen an mehreren Stellen gleichzeitig diesen erhöhten [95]. Eine weitere Studie zum Verhalten von Biokonjugaten mit PEG-Alternativen bei verschiedenen pH-Werten stellte vor kurzer Zeit die Hypothese auf, dass Polymere in ähnlicher Weise wie Chaperone in der Lage seien, die Rückfaltung partiell ungefalteter Proteine zu begünstigen [96].

1.1.6.2 Arbeiten von Joshua L. Price

Weitere Befunde auf diesem Forschungsgebiet sind vor allem den Arbeiten von Joshua L. Price (Brigham Young University, Utah) zu verdanken. Auf Basis zahlreicher systematischer Untersuchungen zu kleinen Oligomer-Konjugaten der WW-Domäne der 34 Aminosäuren langen, β -faltblattstrukturreichen, humanen Peptidyl-Prolyl-cis-trans-Isomerase Pin 1 wurden verschiedene Hypothesen aufgestellt, wobei sich auch hier zeigte, dass PEG-Tetramere die ideale Länge zur Stabilisierung besitzen [97]. PEG führte bei verschiedenen Mutationen in Konjugationsnähe im Gegensatz zur N-Glykosylierung zur Stabilisierung [98]. Während zunächst Sekundärstrukturänderungen und PEG-Wechselwirkungen mit spezifischen Aminosäuren für die Stabilisierung verantwortlich gemacht worden waren [99], zeigten weiterführende Simulationen und CD-basierte Stabilitätsmessungen, auch unter Nutzung von Deuteriumoxid als Solvens, dass die Desolvatation von in der Nähe der Konjugationsstelle befindlichen Hydroxylseitenketten (von Serin, Threonin oder Tyrosin) zu einem entscheidenden entropischen Gewinn führen könnte. Faltungskinetiken, die durch DSF-Messungen mithilfe der Fluoreszenz von Tryptophan ermittelt wurden, deuteten bei stabilisierten Varianten, wie zuvor auch schon in [97], auf eine Erniedrigung der Entfaltungskonstante k_u ohne Beeinträchtigung der Faltungskonstante k_f hin. Auch wurde die konformative Stabilität in dieser Arbeit erfolgreich mit der Proteolysestabilität und damit mit verbesserten pharmakokinetischen Eigenschaften korreliert. Aufgrund dieser Beobachtungen definierten Price et al. einen kristallstrukturbasierten Winkel θ_{Price} .² Dieser wird zwischen zwei Vektoren a und b gemessen. Ersterer verläuft vom C_α -Atom der zu konjugierenden Aminosäure durch deren Seitenkettenmassenschwerpunkt. Der zweite wird von diesem Massenschwerpunkt aus zum nächstgelegenen Sauerstoffatom einer Hydroxylseitenkette gezogen (Abb. 1.5). Kleinere Werte für θ_{Price} verdeutlichen die Orientierung der Konjugationsstelle zur Hydroxylgruppe und korrelierten mit einer Stabilisierung bzw. negativeren $\Delta\Delta G_f^0$ -Werten (Differenz freier Faltungsenthalpien zwischen PEGylierter und unPEGylierter Variante, siehe Kap. 1.1.5.1 und Anhang A). Getestet wurde dieses kristallstrukturbasierte Vorhersagemodell an einer PEGylierten SH3-Domäne, welche korrekt als stabilisiert klassifiziert wurde [100].

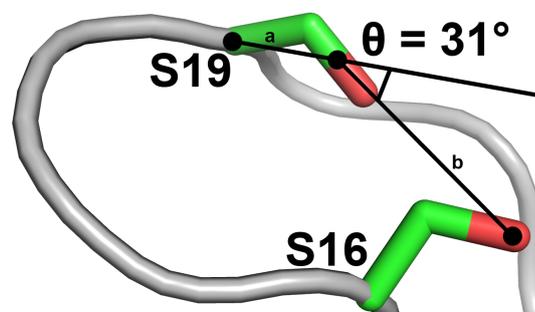


Abbildung 1.5: Illustration des von Price et al. definierten Winkels θ_{Price} am Beispiel der zu konjugierenden Aminosäure S19 der WW-Domäne von Pin 1 [100].

²Zur Vermeidung von Verwechslungen mit der zuvor beschriebenen Elliptizität wurde hier die Bezeichnung *Price* ergänzt.

Zusätzliche Studien von Price et al. deuten daraufhin, dass rigide, planare Linkerstrukturen mit einer ganz bestimmten Länge am besten geeignet sind, um den Desolvationseffekt zu maximieren [101]. Die Desolvatation kann jedoch nicht nur in der Nähe von Hydroxylseitenketten zum Stabilitätsgewinn führen. Ebenso scheint PEG durch eine Abschirmung des Solvens intramolekulare Salzbrücken zwischen verschiedenen geladenen Aminosäuren verstärken zu können (getestet an der WW-Domäne von Pin 1 und den Homo-(Di-/Tri-)meren 1CW und GCN4, siehe auch Kap. 1.5) [102]. Auch ist es in der Lage, vornehmlich exponierte, hydrophobe Areale zu bedecken [103] und NH- π -Wechselwirkungen zwischen Aminosäuren zu verstärken [104]. Arbeiten zu einem α -helikalen Trimer deuten zudem darauf hin, dass PEG-Konjugationen mittels eines Cys-Maleimid-Linkers an der Position i zu benachbarten Lysininen an den Stellen $i + 3$ und $i + 4$ am besten geeignet sind, um eine Stabilisierung α -helikaler Strukturen zu erreichen (Abb. 1.6). Während MD-Simulationen generell auf Interaktionen der PEG-Ketten mit Lysininen hindeuten [105–107], wird eine Stabilisierung nach Price et al. in einer solchen Position vermutlich durch Interaktionen des Maleimids mit den Lysininen erreicht. Konjugierte PEG-Ketten führen dagegen zu einer Destabilisierung dieser Wechselwirkungen, wobei eine Stabilisierung in Relation zur reinen Cystein-Mutante dennoch erreicht wird [108].

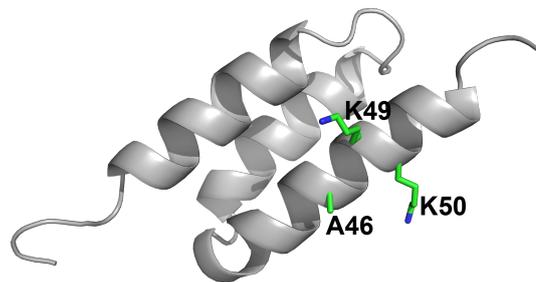


Abbildung 1.6: Potentielle Konjugationsstelle A46 in der Struktur des Her2-Affibodys. K49 und K50 sind an den Positionen $i+3$ und $i+4$ in der Helix lokalisiert.

1.1.6.3 Zusammenfassende Übersicht

Insgesamt scheint die PEG-basierte Erhöhung der konformativen Stabilität also aus mehreren Effekten zu resultieren. Da in der Vergangenheit von sich im Gleichgewicht befindenden proteinnahen und solvensexponierten PEG-Konformationen berichtet wurde, besteht vermutlich eine feine Balance zwischen enthalpisch günstigen Wirkungen (spezifische Protein-PEG-Interaktionen und Verstärkung intramolekularer Salzbrücken) und einem entropischen Gewinn, der anscheinend eine noch bedeutendere Rolle spielt (Verdrängung schwach koordinierten Wassers um Hydroxylgruppen und um solvensexponierte, apolare Oberflächen). Diese Effekte sind umgebungsabhängig und können durch einzelne Mutationen beeinflusst werden [99]. Eine noch weiterführende, zusammenfassende Übersicht möglicher Stabilisierungsmechanismen bzw. Beobachtungen aus der Literatur zu verschiedenen PBK ist in Tab. 1.4 gezeigt.

Tabelle 1.4: Mechanismen zur thermischen / chemischen PBK-Stabilisierung (Reproduzierung mit Genehmigung aus [96], Copyright 2018 American Chemical Society).³

Protein	Polymer	Stabilisierungsmechanismus / Beobachtung	Ref.
1CW / GCN4	PEG	Abschirmung von Rückgrat-HB, Abschirmung von intramolekularen Salzbrücken, Interaktionen mit Lysin an Positionen $i+3$, $i+4$	[94, 102]
Cellulase	Am, DMAm, DMAm-AAm, DMAm-DMAEMA	Keine Verbesserung thermischer / chemischer Stabilität	[109]
Chymotrypsin	PEG	Bindung hydrophober Regionen, reduzierte Proteinflexibilität	[92]
Chymotrypsin	Spermidin	Hydrophobe Wechselwirkungen und HB	[110]
Chymotrypsin	PCBAm, POEGMA, PDMAEMA, PAM, PSMA	Begünstigte Rückfaltung durch Stabilisierung partiell entfalteter Proteine	[96]
Chymotrypsin	PCBAm, PQA, PSMA, POEGMA	Säurestabil durch Reduktion elektrostatischer Interaktionen, Strecken des Polymers ins Solvens	[111]
Cytochrom C	PEG	Thermisch instabil, aber Erhalt einer metastabilen, bioaktiven Konformation	[112]
Faktor VIIa	GlycoPEG	PEG verlangsamt Aggregation	[113]
Insulin	PEG	Hydrophobe Wechselwirkungen, Wasserverdrängung	[93]
Lysozym	PEG	Wechselwirkung über HB oder keine Verbesserung thermischer Stabilität	[114, 115]
Lysozym	Am, DMAm, OEOA, Am/PCMA, Am/AA, Am/DMAEMA, AGA	Thermisch instabil durch Entropieänderung beim Faltungsprozess, chemisch stabile ionische Polymere interagieren mit Protein	[116]
Methionyl-G-CSF	PEG	Reduktion der Aggregation	[117]
Pin 1 WW-Domäne	PEG	Entropisch günstige Desolvatation, Stabilisierung des Übergangs- und nativen Zustands während (Ent)faltung durch Wechselwirkungen, verringerte Entfaltungskonstante und Proteinflexibilität, Abschirmung hydrophober Flächen, intramolekularer Salzbrücken und NH- π -Interaktionen	[33, 97, 99–101] [102–104, 118]
Pyrophosphatase	POEGMA, PNIPAAm	Polymer sollte länger als Distanz zwischen Konjugations- und Bindestelle sein, um durch eine Hydrathülle nahe letzterer die Aggregation des Proteins zu senken	[119, 120]
Staphylokinase	PEG	Flexibles Polymer bildet Hydrathülle zur sterischen Abschirmung	[121]
Src SH3-Domäne	PEG	Erniedrigte Entfaltungskonstante k_f , SASA-Modulation stärker in nativer Struktur	[95]
Trypsin	PEG	Wechselwirkung über HB	[122]
Trypsin	Dextrin, ST-HPMA	Thermische Stabilität und Schutz gegen Autolyse durch sterische Abschirmung und HB	[123]

³AA: Acrylsäure, AGA: N-Acryloyl-D-glucosamin, Am: Acrylamid, DMAEMA: Dimethylaminoethoxymethacrylat, DMAm: Dimethylacrylamid, OEOA: Oligo(ethylenoxid)methyletheracrylat, PAM: Poly(ammoniummethacrylat), PCBAm: Poly(carboxybetainacrylamid), PCMA: Phosphorylcholinmethacrylat, PEG: Poly(ethylenglykol), PNIPAAm: Poly(N-isopropylacrylamid), POEGMA: Poly(oligo(ethylenglykol)methacrylat), PSMA: Poly(sulfonatmethacrylat), ST-HPMA: semitelecholisches Poly[N-(2-hydroxypropyl)-methacrylamid].

1.2 Molekulardynamik-Simulationen

1.2.1 Grundlagen

Zum grundlegenden Verständnis des Verhaltens inhärent flexibler und dynamischer Biomoleküle auf atomarer Ebene stellt die in den 1950er- entwickelte und in den 1970er-Jahren erstmals auf Proteine angewandte Technik der MD-Simulationen ein wichtiges Werkzeug der pharmazeutischen Forschung dar [124–127], dessen Hintergrund im Folgenden beschrieben werden soll.

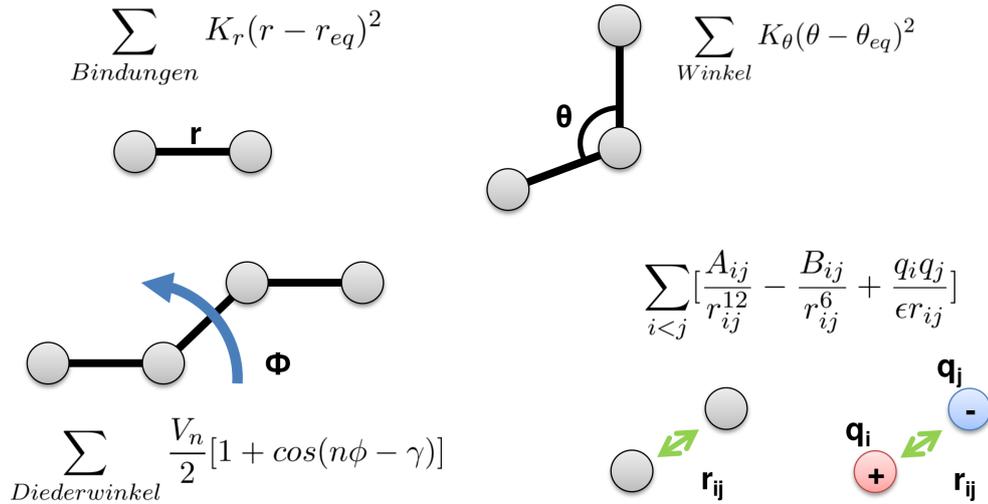
1.2.1.1 Abbildung atomarer Wechselwirkungen

Da es sich bei Makromolekülen in kondensierten Phasen zumeist um Systeme mit tausenden Atomen handelt und somit quantenmechanische (QM) Methoden mit expliziter Berücksichtigung elektronischer Wellenfunktionen durch unterschiedliche Basissätze äußerst zeit- und ressourcenintensive Berechnungen darstellen, werden für solche Simulationen klassische molekularmechanische Ansätze verfolgt. Ganze Atome werden hierbei durch einzelne, in der Regel nicht polarisierbare Partikel mit definierter Masse und Ladung in einer Simulationsbox, die das Makromolekül und das unmittelbar umgebene Solvens beinhaltet, repräsentiert. Sowohl intra- als auch intermolekulare Wechselwirkungen der Partikel werden über einen Satz verschiedener Terme beschrieben, der in seiner Summe als Kraftfeld bezeichnet wird und die Gesamtenergie E_{total} des Systems definiert. Dieses umfasst Konstanten zu jedem darin definierten Atomtyp und basiert in aller Regel auf einer Vielzahl festgelegter empirischer Parameter, die zuvor durch experimentelle oder quantenmechanische Verfahren ermittelt wurden. Üblicherweise werden zur Beschreibung von intramolekularen Bindungen und Winkeln harmonische Potentiale und für Diederwinkel periodische Kosinusfunktionen genutzt. Elektrostatische Wechselwirkungen werden durch das Coulombsche Gesetz und Van-der-Waals (VdW)-Interaktionen mithilfe des Lennard-Jones-Potentials berechnet (Abb. 1.7). Zusätzliche Terme, z. B. für Out-of-Plane-Schwingungen, können ebenfalls im Kraftfeld integriert sein [128–130]. Kraftfelder für biologische Systeme sind bspw. AMBER, CHARMM, GROMOS und OPLS [131]. Sie können mit Programmen wie AMBER, GROMACS und NAMD genutzt werden [132].

1.2.1.2 Bewegungsgleichungen

Mithilfe des verwendeten Kraftfelds kann die Kraft $F_i(t)$, die auf das Atom i mit der Masse m_i und der Position $x_i(t)$ im euklidischen Raum in einer bestimmten Konfiguration des Systems zur Zeit t wirkt, bestimmt werden. $F_i(t)$ stellt die Vektorsumme der Interaktionen mit anderen Partikeln im System dar. $U(x(t))$ ist die gesamte potentielle Energie, die wiederum eine Funktion der Positionen $x(t)$ der Atome ist:

$$F_i(t) = -\frac{\partial U(x(t))}{\partial x_i(t)} = m_i \frac{\partial^2 x_i(t)}{\partial t^2} = m_i \frac{\partial v_i(t)}{\partial t} = m_i a_i(t) \quad (1.11)$$



$$E_{total} = \sum_{\text{Bindungen}} K_r (r - r_{eq})^2 + \sum_{\text{Winkel}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{Diederwinkel}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} [\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon r_{ij}}]$$

Abbildung 1.7: Illustration exemplarischer Kraftfeld-Terme (am Beispiel von AMBER [133]) zur Beschreibung inter- und intramolekularer Wechselwirkungen und damit der Gesamtenergie E_{total} des Systems. Bindungslängen werden mit einem harmonischen Potential unter Verwendung einer Konstanten K_r beschrieben. Hierfür ist die Abweichung der aktuellen Bindungslänge r von der Equilibriumsbindungslänge r_{eq} entscheidend. In analoger Weise werden die Winkel θ zwischen drei benachbarten Atomen beschrieben. Eine Torsion wird über eine periodische Kosinusfunktion definiert, mit der energetischen Barriere V_n , der Periodizität n und der Phase γ . Nichtkovalente Wechselwirkungen werden über das Lennard-Jones-Potential mit den Konstanten A_{ij} und B_{ij} , sowie elektrostatische Interaktionen mittels des coulombischen Gesetzes (Ladungen q_i und q_j und Dielektrizitätskonstante ϵ) beschrieben [129].

Durch die Kraft $F_i(t)$ können über das zweite Newtonsche Gesetz die Beschleunigung $a_i(t)$ und, in Kombination dieser mit der Position $x_i(t)$ und der Geschwindigkeit $v_i(t)$ des Partikels zur Zeit t , auch die Position und Geschwindigkeit zur Zeit $t + \delta t$ ermittelt werden. Zur Integration der Bewegungsgleichungen dienen verschiedene Methoden auf Basis von Taylorreihen, wie z. B. der Verlet-Algorithmus, der die Position im vorangegangenen Zeitschritt $t - \delta t$ zur Berechnung nutzt:

$$x_i(t + \delta t) = 2 x_i(t) - x_i(t - \delta t) + \delta t^2 a_i(t) \quad (1.12)$$

$$v_i(t) = \frac{x_i(t + \delta t) - x_i(t - \delta t)}{2 \delta t} \quad (1.13)$$

Gemäß Glg. 1.13 werden Geschwindigkeiten bei diesem Ansatz nach Erhalt der Positionen des nächsten Zeitschritts berechnet. Alternativ kann das Leapfrog-Verfahren genutzt werden, welches die Geschwindigkeiten und Positionen asynchron ermittelt. Das Programm NAMD nutzt als Integrator eine Form der Velocity-Verlet-Methode, bei der zunächst die neuen Positionen zum Zeitpunkt $t + \delta t$, im Anschluss daran die Geschwindigkeiten zum Zeitpunkt $t + \frac{1}{2}\delta t$, danach über die Kraft $F_i(t + \delta t)$ die Beschleunigung $a_i(t + \delta t)$ und schließlich die Geschwindigkeit zum Zeitpunkt $t + \delta t$ kalkuliert werden:

$$x_i(t + \delta t) = x_i(t) + \delta t v_i(t) + \frac{1}{2} \delta t^2 a_i(t) \quad (1.14)$$

$$v_i(t + \frac{1}{2} \delta t) = v_i(t) + \frac{1}{2} \delta t a_i(t) \quad (1.15)$$

$$a_i(t + \delta t) = \frac{F_i(t + \delta t)}{m_i} \quad (1.16)$$

$$v_i(t + \delta t) = v_i(t + \frac{1}{2} \delta t) + \frac{1}{2} \delta t a_i(t + \delta t) \quad (1.17)$$

Initiale Momente $p_i(0)$ und damit Geschwindigkeiten $v_i(0)$ aller Partikel werden in der Regel zu Beginn auf Basis der Maxwell-Boltzmann-Verteilung zufällig entsprechend der gewünschten Temperatur $T(t)$ generiert:

$$T(t) = \frac{1}{(3N - n) k_B} \sum_{i=1}^N \frac{|p_i(t)|^2}{m_i} \quad (1.18)$$

wobei N die Anzahl an Atomen im System, k_B die Boltzmann-Konstante und n die Menge an eingeschränkten Freiheitsgraden (z. B. durch den unten erwähnten SHAKE-Algorithmus) darstellen [134–136].

1.2.1.3 Wahl des Zeitschritts

Durch die Berechnungen mit dem vorher definierten Schritt δt kann somit eine Trajektorie erhalten werden, die die Momente und Positionen aller Partikel der Simulationsbox über die Zeit hinweg beschreibt. Um jedoch ein stabiles System zeiteffizient simulieren zu können, ist die Wahl des Zeitschritts entscheidend. Dieser sollte nicht größer als die höchste Frequenz sein, welche den Schwingungen der Schweratom-Wasserstoff-Bindungen zuzuordnen ist (etwa 0.1 fs). Daher werden die Schweratom-Wasserstoff-Bindungslängen in der Regel fixiert, zumeist mit dem sogenannten SHAKE- [137] oder auch, im Falle von NAMD, mit dem davon abgeleiteten RATTLE-Algorithmus [138]. Auf Basis dieser Eliminierung hochfrequenter Schwingungen wird der übliche Zeitschritt für Simulationen biomolekularer Systeme auf 2 fs gesetzt [134, 136].

1.2.1.4 Strukturaufbereitung und Equilibrierung

Startkoordinaten des zu untersuchenden Proteins werden für eine MD-Simulation zumeist auf Basis von Kristall- oder Nuklearmagnetresonanz (NMR)-Strukturen aus der Proteindatenbank (PDB) [139] aufbereitet. Das System wird zunächst in einem geeigneten Protonierungszustand solvatisiert und für einige Schritte energieminiert. Dies bedeutet, dass die Struktur auf Basis des verwendeten Kraftfelds in das, je nach eingestelltem Gradienten, nächstgelegene Energieminimum des hochdimensionalen Konformationsraums gelangt. In einer dem eigentlichen Produktionslauf vorangestellten Equilibrierung wird sichergestellt, dass zentrale physikalische Kenngrößen (v. a. die kinetische und potentielle Energie, sowie Temperatur, Volumen und Druck) und Eigenschaften des Systems, welche relevant für die zu untersuchende Fragestellung sind, einen stabilen Fluktuationsbereich erreichen, bevor es zur Probennahme

(dem sogenannten *Sampling*) für die Analyse kommt. Hierfür wird das System im mikrokanonischen Ensemble (NVT), d. h. bei konstanter Anzahl an Partikeln (N), sowie bei konstantem Volumen (V) und einer an ein externes Wärmebad gekoppelten, konstanten Temperatur (T) simuliert. Da es sich meistens um Kristallstrukturen handelt, die unter Kryobedingungen aufgenommen wurden, wird die Temperatur anfangs sukzessive von 100 K auf den gewünschten Wert erhöht (meistens 300 K). Nachdem sich ein stabiler Energiewert E_{total} und eine konstante Temperatur im NVT-Ensemble eingestellt haben, wird für kurze Zeit eine Equilibrierung im isothermal-isobarischen NPT-Ensemble zum Erreichen der gewünschten Dichte durchgeführt. Hierbei werden die Partikelanzahl, der Druck (P) und die Temperatur konstant gehalten. Der Druck kann über ein Barostat kontrolliert werden. Anschließend wird der Produktionslauf im NPT-Ensemble gestartet, wobei Momentaufnahmen des Systems in regelmäßigen Schritten herausgeschrieben werden (üblicherweise jede Pikosekunde). Je nach zu untersuchendem System bieten sich verschiedene Wärmebäder zum Erhalt einer stabilen Temperatur an, so z. B. das Andersen-, das Langevin-, das Nosé-Hoover- oder das Berendsen-Thermostat. Analog hierzu können das Andersen-, das Parrinello-Rahman- oder auch das Martyna-Tuckerman-Tobias-Klein-Barostat zur Einstellung eines konstanten Drucks dienen [129, 130].

1.2.1.5 Periodische Randbedingungen

Zur Verhinderung von Artefakten an der Oberfläche der Simulationsbox und zur generellen Minimierung von Effekten, die aus der Analyse eines endlichen Systems resultieren, werden Simulationen biomolekularer Strukturen üblicherweise unter periodischen Randbedingungen (*Periodic boundary conditions*, PBC) durchgeführt, die eine Annäherung an Bulk-Bedingungen darstellen. Dies bedeutet, dass das System von Abbildern seinesgleichen umgeben ist. Um intermolekulare Wechselwirkungen mit dem selben Partikel aus einer Nachbarbox zu unterbinden, werden bestimmte Limitierungen (Cut-off-Werte) gesetzt, die nicht größer sein sollten als die halbe Länge der Box. Im Falle von Van-der-Waals-Interaktionen und Coulomb-Wechselwirkungen auf kurzer Distanz wird üblicherweise eine Switching-Funktion ab einer Reichweite von 10 Å eingesetzt, mit der die Wechselwirkungen bis 12 Å graduell auf Null abfallen. Da es im Falle von elektrostatischen Interaktionen aber zu großen Artefakten durch eine vollständige Annullierung ab einer bestimmten Reichweite käme (diese fallen nur mit einem Faktor von r^{-1} ab, das Lennard-Jones-Potential mit einem Faktor von r^{-6}), werden solche Wechselwirkungen über weite Distanzen hin (Long-Range-Interaktionen) mithilfe recheffizienter, gitterbasierter Methoden angenähert. Hierfür werden vor allem die Particle-Mesh-Ewald-Methode [140], bei der eine schnelle Fourier-Transformation zum Einsatz kommt, oder entsprechende Abwandlungen hiervon genutzt (Particle-Particle-Mesh- oder Smooth-Particle-Mesh-Ewald-Methode) [130].

1.2.2 Biokonjugat-Simulationen

Seit der ersten MD-Simulation eines PEG-Protein-Konjugats auf atomarer Ebene [141] steigt die Menge pro Jahr publizierter Studien, die Simulationen PEGylierter Proteine beschreiben. Dagegen verbleibt die Anzahl an Publikationen, die sich mit der Entwicklung bzw. Optimierung geeigneter Kraftfelder beschäftigen, auf einem relativ niedrigen Niveau [142]. Die Simulation eines Biokonjugats stellt insofern eine besondere Herausforderung dar. Kraftfelder werden weiterhin nur mit Fokus auf bestimmte Molekülarten entwickelt. AMBER14ffSB [133] basiert bspw. auf quantenmechanischen Berechnungen mit dem Hartree-Fock (HF)-6-31G*-Basissatz und einer Ladungsverteilung mittels der RESP (*Restrained electrostatic potential*)-Methode [143] und ist speziell für Proteine konzipiert. Daneben existieren auch Kraftfelder für DNA (OL15 [144]), Kohlenhydrate (GLYCAM_06j [145]), Lipide (LIPID14 [146]) oder kleinere organische Moleküle (GAFF2 [147]). Erste Parameter für PEG-Ketten wurden in den frühen 1990er Jahren entwickelt [148]. Für Hybrid-Moleküle mit kovalent gebundenen Polymeren gibt es jedoch noch keine weitreichend etablierten Parameter. Für PEG-Protein-Konjugate wird oft das CHARMM-Kraftfeld C35r [149] für Etherstrukturen genutzt, welches mit Experimenten übereinstimmende Werte bzgl. hydrodynamischer Radien und Persistenzlängen zeigte [32]. Alternativen stellen auch das TraPPE-UA- und eine modifizierte Version des GAFF-Kraftfelds dar [150–152]. Doch auch andere Kraftfelder wurden in der Vergangenheit für Simulationen von PEG-Konjugaten genutzt (z. B. AMBER99SB-ILDN [153] für die WW-Domäne von Pin 1 [100] oder das United-Atom (UA)-Kraftfeld GROMOS96 43a1 [154] für Insulin [93]).

Weiterhin stellt die Größe der Biokonjugate ein Problem dar. Da in der Regel Polymere mit einem Molekulargewicht von 5 bis 40 kDa konjugiert werden, stellen Simulationen dieser Strukturen auf atomarer Ebene auch heute noch einen immensen Zeitaufwand dar. Diese Konjugate repräsentieren Systeme mit zahlreichen Freiheitsgraden und ein ausreichendes Sampling ist dadurch nur bedingt zu erreichen. Um dieses Problem zu lösen, werden entweder erweiterte Sampling-Verfahren (z. B. gesteuerte [155] oder beschleunigte [156] MD-Simulationen) oder aber, sehr viel öfter, vergrößerte (*Coarse-grained*, CG) Kraftfelder eingesetzt [32]. Bei letzterer Methode wird der Detailgrad der atomaren Struktur durch einen Zusammenschluss mehrerer Atome zu einzelnen Partikeln bzw. Kugeln (*Beads*) drastisch gesenkt. Bei UA-Kraftfeldern, wie dem oben genannten GROMOS96 43a1, werden Wasserstoff-Atome mit den jeweiligen Schweratomen zusammengeführt. Deutlich gröber ist das MARTINI-Kraftfeld [157]. Das zunächst für Lipide entwickelte CG-Kraftfeld wurde auf Proteine [158], Kohlenhydrate [159] und Nukleinsäuren [160, 161] erweitert und vereint in der Regel vier Atome zu einem von 20 möglichen Partikeltypen (je nach Polarität, Ladung und Größe). Durch CG-Ansätze wird die Anzahl an Dimensionen des Konformationsraums verringert und Simulationen großer Konjugate werden so in annehmbarer

Zeit durchführbar (Abb. 1.8). Dies führt jedoch auch zu einem deutlichen Detailverlust. Zusätzlich müssen CG-Methoden für neue Moleküle (bspw. Polymere) je nach zu analysierender Eigenschaft auf Basis experimenteller Messungen (z. B. über Gyrationradien) neu definiert und iterativ über einen Vergleich mit ausreichend langen, klassischen MDs mit atomarem Detail optimiert werden [162]. Für PBK existieren noch keine allgemein etablierten CG-Ansätze [32], wobei erst kürzlich ein Simulationsprotokoll für das MARTINI 3-Kraftfeld veröffentlicht wurde [163].

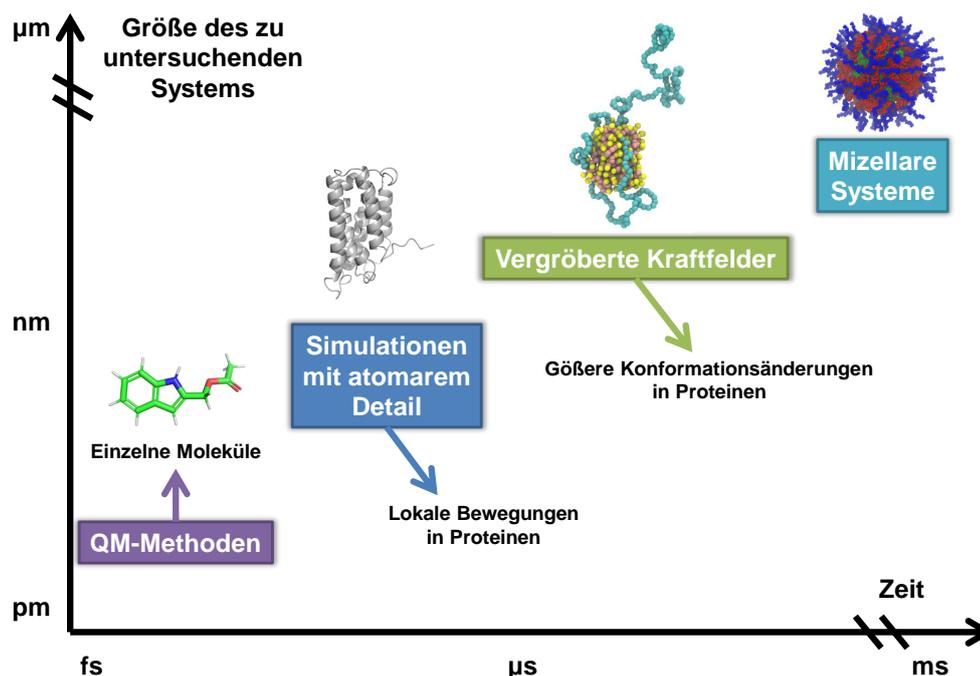


Abbildung 1.8: Geeignete Simulationsansätze und üblicher Zeitraum zur Beobachtung bestimmter Dynamiken in Systemen unterschiedlicher Größe. Rechenintensive QM-Methoden eignen sich für Atome bis hin zu einzelnen Molekülen. Klassische MD-Simulationen auf atomarer Ebene können lokale Bewegungen von Proteinen registrieren. Vergrößerte Kraftfelder reduzieren die Anzahl möglicher Freiheitsgrade und ermöglichen größere Konformationsänderungen. Angelehnt an Abb. 1 aus [162].

1.2.3 Konvergenzkriterien

Ein ausreichendes Sampling der MD-Simulationen ist nötig, um zuverlässige Analysen der Trajektorien durchführen zu können. Klare Richtlinien für minimale Simulationszeiten molekularer Systeme existieren jedoch nicht. Eine Simulation gilt üblicherweise als konvergiert, wenn aus dem möglichen Konformations- bzw. Phasenraum (*Phase space*) ausreichende Probenahmen gezogen wurden, die einer Boltzmann-Gewichtung entsprechen. Nur dann ist gemäß der Ergodenhypothese garantiert, dass die Eigenschaften über die Zeit hinweg mit jenen eines makroskopischen Systems gleichzusetzen sind [136, 164]. Es gilt also, fernab von Nichtgleichgewichtssystemen, die Annahme:

$$\langle G \rangle_{eq} = \lim_{t \rightarrow \infty} \langle G \rangle_t \quad (1.19)$$

wobei $\langle G \rangle_{eq}$ den Wert der betreffenden Eigenschaft im Equilibrium beschreibt und $\langle G \rangle_t$ das entsprechende Ensemblemittel. Diese Annahme ist generell kritisch zu sehen, da sie streng genommen nur für unendlich lange Simulationen gegeben ist. Alternativ können ausreichend viele Replika zur Berechnung der gesuchten makroskopischen Eigenschaften durchgeführt werden (*Ensemble averages*) [164]. Eine Konvergenz ist ultimativ nicht beweisbar, da, egal nach welcher Simulationszeit, nie eruiert werden kann, ob es noch weitere, unbesuchte Konformationen im Phasenraum gibt.

1.2.3.1 Self-consistency checks

Auch wenn ein ausreichendes Sampling nicht klar bestimmbar ist, so existieren dennoch Methoden, um zu überprüfen, ob zumindest der bisher besuchte Konformationsraum ein in sich konsistentes System abbildet (sogenannte *Self-consistency checks*, SCC). Zur Evaluation einzelner Variablen aus Simulationen ist eine Berücksichtigung der Korrelation zeitlich aufeinanderfolgender Momentaufnahmen von Bedeutung. Der Standardfehler (*Standard error*, SE) des Mittelwerts kann folgendermaßen berechnet werden:

$$SE = \frac{\sigma}{\sqrt{N_{ind}}} \sim \sqrt{\frac{\tau}{t_{sim}}} \quad (1.20)$$

wobei σ die Standardabweichung, N_{ind} die Anzahl an unabhängigen Messungen, τ die aus der Autokorrelationsfunktion berechenbare Korrelationszeit und t_{sim} die gesamte Simulationszeit darstellen [136, 165, 166]. Im Folgenden sollen drei SCC-Kriterien beschrieben werden, die eine allgemeine, strukturelle Konvergenz signalisieren [167]. Beim SDVG-Kriterium (benannt nach den Autoren Smith, Daura und van Gunsteren [168]) wird ein Clustering durchgeführt, z. B. auf Basis des RMSD-Werts der C_α -Atome. Die Anzahl an gefundenen Clustern sollte sich hierbei gegen Ende der Simulation hin nicht mehr verändern. Dies deutet darauf hin, dass keine neuen Bereiche des möglichen Konformationsraums mehr erreicht werden. Dieses Kriterium kann jedoch nicht detektieren, ob die Struktur in einem metastabilen Zustand gefangen ist. Als ergänzende Analyse kann daher die Clusterentropiekonstanz (*Constancy of cluster entropy*, CCE) dienen, definiert nach:

$$CCE = \sum p_i \log(p_i) \quad (1.21)$$

Sie beschreibt die Wahrscheinlichkeit p_i , die Struktur zu gegebener Zeit in einem der bereits gefundenen Cluster i wiederzufinden. Im Falle einer in einem Zustand gefangenen Struktur sinkt die CCE stetig. Für eine ausreichende Konvergenz sollte sich dieser Wert gegen Simulationsende hin daher nicht mehr stark ändern [167]. Ein drittes Kriterium stellt die BBCOM/BCOM-Methode von Romo und Grossfield dar (*Bootstrapped block covariance overlap/block covariance overlap*) [169]. Hierbei

werden Hauptkomponentenanalysen (*Principal component analysis*, PCA), bspw. auf Basis der C_α -Atome, durchgeführt, um deren Kovarianz zu bestimmen. Für eine PCA wird für N Atome im System eine $3N \times 3N$ -Korrelationsmatrix erstellt:

$$C_{ij} = \langle x_i - \bar{x}_i \rangle \langle x_j - \bar{x}_j \rangle \quad (1.22)$$

wobei x_i die x, y oder z-Koordinate des Atoms i darstellt und \bar{x}_i den Wert der durchschnittlichen Struktur repräsentiert. Über eine Diagonalisierung der erhaltenen Matrix werden Eigenvektoren erhalten, die die zentralen Bewegungsrichtungen der Struktur im Raum beschreiben, sowie Eigenwerte, die die durchschnittliche Fluktuation entlang dieser Vektoren beschreiben [168]. Anschließend wird das Verhältnis zwischen der Kovarianzübereinstimmung von Simulationsblöcken ansteigender Größe mit der gesamten Trajektorie und der Kovarianzübereinstimmung von Simulationsblöcken, die zunächst einem Bootstrapping-Verfahren unterzogen wurden, mit der gesamten Trajektorie berechnet. Im Idealfall sollte das Verhältnis einen Wert von 1 erreichen. Beim Bootstrapping-Verfahren werden zufällige Momentaufnahmen aus der Trajektorie ausgewählt, wodurch vermieden wird, dass ausschließlich Strukturen mit hoher Korrelation zueinander selektiert werden. Die Kovarianzübereinstimmung zwischen zwei Hauptkomponenten mit den Eigenwerten λ_i und λ_j und den Eigenvektoren \vec{v}_i und \vec{v}_j der Ensembles A und B ist definiert nach [169]:

$$\Omega_{A,B} = 1 - \sqrt{\frac{\sum_i^{N_{Modi}} (\lambda_i^A + \lambda_i^B) - 2 \sum_i^{N_{Modi}} \sum_j^{N_{Modi}} \sqrt{\lambda_i^A \lambda_j^B} (\vec{v}_i^A \cdot \vec{v}_j^B)^2}{\sum_i^{N_{Modi}} (\lambda_i^A + \lambda_i^B)}} \quad (1.23)$$

1.2.3.2 Weitere Konvergenztests

Für Markow-Chain-Monte-Carlo (MCMC)-Verfahren existieren weiterhin verschiedene Tests, um eine stationäre Verteilung einzelner Parameter zu evaluieren. Hierbei wird geprüft, ob die Probennahmen aus einer sich nicht mehr ändernden und damit zeitinvarianten Verteilung stammen. Dies bedeutet also, dass die Übergangsmatrix der Markow-Kette über die Zeit hinweg unverändert bleibt [170, 171]. Während es sich bei diesen Monte-Carlo-Verfahren um Prozesse handelt, bei denen im Gegensatz zu molekulardynamischen Ansätzen zufällige Stichproben ohne zeitliche Abfolge generiert werden, so sollte mit diesen Tests analog auch eine Stationarität einzelner Variablen, die aus einer Zeitreihe wie einer MD-Simulation gewonnen wurden, überprüfbar sein.

Zwei dieser Konvergenztests wurden in dieser Arbeit getestet: Beim Heidelberger-Welch-Test wird die Stationarität mit einem Cramér-von-Mises-Test über die gesamte Trajektorie geprüft. Wird dieser Test nicht erfüllt, werden die ersten 10 % der Zeitreihe verworfen und die Konvergenz erneut geprüft. Dies wird so lange durchgeführt, bis die Stationarität entweder erfüllt wird oder aber 50 % der Zeitreihe verworfen wurden,

wobei die Nullhypothese, dass es sich um eine stationäre Verteilung handelt, in letzterem Fall nicht erfüllt wurde. Wurde dieser erste Teil des Heidelberger-Welch-Tests erfüllt, so wird in einem zweiten Verfahren geprüft, ob das 95%-Konfidenzintervall des Mittelwerts der nicht verworfenen Zeitreihe kleiner ist als 10 % dieses Mittelwerts [172–174]. Beim Geweke-Test werden hingegen die Mittelwerte der ersten und der zweiten Hälfte der Zeitreihe miteinander verglichen, um einen Z-Wert für deren Übereinstimmung zu erhalten. Eine Stationarität ist wahrscheinlicher, je näher dieser Wert bei 0 liegt (idealerweise unter 2) [174, 175].

1.2.4 Gaußbeschleunigte Molekulardynamik-Simulationen

Eine Methode des erweiterten Samplings stellen gaußbeschleunigte MD-Simulationen dar (*Gaussian accelerated MDs*, GaMDs), eine Weiterentwicklung der beschleunigten MD-Simulationen (*Accelerated MDs*, aMDs) [176, 177]. Bei dieser Methode wird ein harmonisches Zusatzpotential $\Delta V(\vec{r})$ zum zugrundeliegenden Potential $V(\vec{r})$ addiert, welches die energetischen Barrieren aller Konformationsänderungen verringert. Somit wird eine modifizierte Energielandschaft mit dem Potential $V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r})$ erhalten. Nur selten beobachtbare Übergänge in herkömmlichen MD-Simulationen (*Conventional MDs*, cMDs) werden so schneller realisiert. Im Gegensatz zum aMD-Ansatz folgt dieses Zusatzpotential jedoch einer Gauß-Verteilung, wodurch eine verbesserte energetische Neugewichtung mithilfe einer Kumulanten zweiter Ordnung ermöglicht wird. Neugewichtungen mit geringen Fehlern können so auch für größere Systeme durchgeführt werden, im Falle von aMD-Simulationen liegt die Grenze für eine mögliche Neugewichtung typischerweise bei Proteinen mit 40 Aminosäuren. Das Potential $\Delta V(\vec{r})$ wird nur addiert, sofern das Grundpotential geringer als die Referenzenergie E ist:

$$\Delta V(\vec{r}) = \begin{cases} \frac{1}{2} \frac{k_0}{V_{max} - V_{min}} (E - V(\vec{r}))^2, & V(\vec{r}) < E \\ 0, & V(\vec{r}) \geq E \end{cases} \quad (1.24)$$

V_{max} und V_{min} stellen die Maxima und Minima des Potentials dar und k_0 ist eine Kraftkonstante zwischen 0 und 1, wobei größere Werte eine stärkere Modifizierung der Energielandschaft bedeuten. Die Referenzenergie E kann Werte zwischen V_{max} und $V_{min} + \frac{1}{k}$ annehmen, wobei die eigentliche Kraftkonstante k über folgende Gleichung definiert ist:

$$k = \frac{k_0}{V_{max} - V_{min}} \quad (1.25)$$

Standardmäßig ist E jedoch als V_{max} festgesetzt. Die Kraftkonstanten und die Referenzenergie werden somit automatisch bestimmt. Lediglich das obere Limit der Standardabweichung des Potentials σ_0 wird vor der Simulation durch den Nutzer eingestellt. Es steht mit den Parametern k und E in folgendem Zusammenhang:

$$\sigma_{\Delta V} = k(E - V_{av})\sigma_V \leq \sigma_0 \quad (1.26)$$

Dabei stellen $\sigma_{\Delta V}$ die Standardabweichung des Zusatzpotentials, σ_V jene des Potentials und V_{av} das durchschnittliche Potential dar. Mit $E = V_{max}$ gilt, dass k_0 einen Maximalwert von 1 erreichen kann [176, 177]:

$$k_0 = \min(1.0, \frac{\sigma_0}{\sigma_V} \frac{V_{max} - V_{min}}{V_{max} - V_{av}}) \quad (1.27)$$

Ein Zusatzpotential kann über den GaMD-Ansatz zum Diederpotential im System, zum gesamten Potential, sowie zu beiden Elementen gleichzeitig addiert werden (Dual Boost-Schema). Die Übereinstimmung von ΔV mit einer tatsächlichen Gaußschen Verteilung kann an der Anharmonizität γ der Verteilung des Zusatzpotentials bemessen werden, welche bei größeren Werten keine genaue Neugewichtung erlaubt. Hierzu wird die Entropie $S_{\Delta V}$ von ΔV von der jeweiligen maximalen Entropie S_{max} einer idealen Gauß-Verteilung subtrahiert [178, 179]:

$$\gamma = S_{max} - S_{\Delta V} = \frac{1}{2} \ln(2\pi e \sigma_{\Delta V}^2) + \int_0^{\infty} p(\Delta V) \ln(p(\Delta V)) d\Delta V \quad (1.28)$$

Der Parameter $p(\Delta V)$ stellt die normierte Wahrscheinlichkeitsverteilung von ΔV für die Momentaufnahmen in einer der Klassen (*Bins*) der zur Neugewichtung gewählten Reaktionskoordinate(n) dar, e ist die Eulersche Zahl. Zur Simulation wird üblicherweise eine kurze Phase einer cMD-Simulation vorangeschaltet. Nachdem sich das System in dieser über einige Zeitschritte hinweg stabilisiert hat, werden Werte für die Parameter V_{max} , V_{min} , V_{avg} und σ_V gesammelt. Am Ende dieser Phase werden ebenfalls E und k_0 ermittelt. In einer nachgeschalteten, kurzen Phase der GaMD-Equilibrierung wird das Zusatzpotential dann entsprechend dieser Parameter zum Grundpotential addiert. Nach dieser Equilibrierung werden E und k_0 dann im eigentlichen Produktionslauf gemäß eines vordefinierten Zeitintervalls regelmäßig aktualisiert. Nach abgeschlossener Simulation können mithilfe des *PyReweighting*-Skripts Neugewichtungen auf Basis von Kumulanten, einer Maclaurin-Serie oder des exponentiellen Durchschnitts erfolgen [178]. Durch die, im Gegensatz zu Metadynamik-Simulationen (*vide infra*), ungerichtete Beschleunigung bei einer GaMD-Simulation, können ohne Vorwissen geeigneter Kollektivvariablen das Sampling generell verbessert und anspruchsvolle Konformationsänderungen wie bspw. Protein(ent-)faltungen oder Ligand-Protein-Bindungsprozesse beobachtet werden [176, 177]. Auch eine Modifikation des Algorithmus zur Berechnung von Ligand-Bindungskinetiken wurde kürzlich publiziert [180].

1.2.5 Wohltemperierte Metadynamik-Simulationen

Erweitertes Sampling kann auch gerichtet über vorher definierte Kollektivvariablen (*Collective variables*, CV), bspw. bestimmte Distanzen oder Winkel im System, erfolgen. Dies eignet sich besonders im Falle von im Vorfeld bekannten Konformationsänderungen, die mit der Simulation beobachtet werden sollen und bei denen die Variablen, die zu einem Übergang führen, bekannt sind. Eine Methode, die diesen

Ansatz verfolgt, ist jene der wohltemperierten Metadynamik-Simulationen mit multiplen „Walkern“ (*Multiple walkers well-tempered metadynamics simulations*, MW-wtMetaMDs). Bei MetaMDs wird ein kontextabhängiges Zusatzpotential entlang der CV zum Grundpotential des Systems addiert. Dies bedeutet, dass zunächst v. a. die Minima der Energielandschaft mit einem Zusatzpotential „aufgefüllt“ werden, in welcher sich die Struktur anfangs befindet. Durch das Zusatzpotential werden energetische Barrieren überwunden und alle Minima im Hinblick auf die definierten CV sukzessive erreicht [181–183]. Das Potential des Systems ist hier definiert nach:

$$V_G(s, t) = \omega \sum_{t'=\tau_G, 2\tau_G, \dots (t' < t)} e^{-\frac{(s(t)-s(t'))^2}{2\sigma^2}} \quad (1.29)$$

Dabei ist $s(t)$ der Wert der Kollektivvariablen s zur Zeit t . Die Gaußsche Höhe ω repräsentiert die Größe des hinzugefügten Potentials, welches alle τ_G Zeitschritte zum Potential addiert wird. Zusätzlich wird die Breite der Gauß-Kurve entlang der betrachteten Variable mit dem Parameter σ definiert. Im Falle mehrerer CV (Anzahl d) ergibt sich folgende modifizierte Gleichung:

$$V_G(s, t) = \omega \sum_{t'=\tau_G, 2\tau_G, \dots (t' < t)} e^{-\sum_{\alpha=1}^d \frac{(s_{\alpha}(t)-s_{\alpha}(t'))^2}{2\sigma_{\alpha}^2}} \quad (1.30)$$

Durch ein ausreichendes Sampling kann vom modifizierten Potential auf die zugrundeliegende freie Energie $F(s)$ der unmodifizierten Energielandschaft geschlossen werden:

$$\lim_{t \rightarrow \infty} V_G(s, t) \sim -F(s) \quad (1.31)$$

Das Sampling entlang der CV wird maßgeblich durch höhere Werte von ω , σ und niedrigere τ_G -Werte begünstigt. Allerdings sollten diese Werte mit Vorsicht justiert werden, da geringere Energiebarrieren durch große repulsive Gauß-Kurven mit hoher Frequenz gleichzeitig auch die Qualität des rekonstruierten, zugrundeliegenden Energieprofils beeinträchtigen [181–183]. Im Falle von wtMDs wird die Konvergenz des Energieprofils deutlich begünstigt [184]. Bei dieser Abwandlung wird ein Bias-Faktor $\gamma = \frac{T+\Delta T}{T}$ definiert (oft mit Werten zwischen 6 und 15), der den Wert ΔT bestimmt, durch welchen wiederum die Gaußschen Kurven mit der Anfangshöhe ω_0 über die Simulationszeit hinweg ihre Höhe reduzieren:

$$\omega = \omega_0 e^{-\frac{V_G(s,t)}{k_b \Delta T}} \quad (1.32)$$

Dabei würde $\Delta T = 0$ einer reinen cMD und $\Delta T \rightarrow \infty$ einer normalen MetaMD entsprechen. Durch diese Modifikation ergibt sich folgende Beziehung:

$$\lim_{t \rightarrow \infty} V_G(s, t) \frac{T + \Delta T}{\Delta T} \sim -F(s) \quad (1.33)$$

Durch multiple „Walker“, d. h. mehrere parallel geschaltete, miteinander „kommunizierende“ Replika, wird das Sampling weiter verbessert [185]. Diese „Walker“ (Anzahl N_w) füllen die Minima gleichzeitig auf und fließen lediglich in einem gemeinsamen Potential V_G in folgender Beziehung zusammen:

$$V_G(s, t) = \frac{\omega}{\tau_G} \sum_{i=1}^{N_w} \int_{i=0}^t dt' e^{-\frac{|s(t)-s(t')|^2}{2|\sigma|^2}} \quad (1.34)$$

Durch eine höhere Anzahl an multiplen „Walkern“ wird die Qualität der freien Energielandschaft nicht verbessert, jedoch wird die nötige Simulationszeit zum Erreichen einer Konvergenz erniedrigt.

In der Praxis werden üblicherweise zusätzliche, hohe energetische Barrieren als Minima bzw. Maxima entlang der Reaktionskoordinaten, den CV, definiert. Dies dient dazu, dass die CV nur den für die Fragestellung relevanten Konformationsraum erreichen und es in diesem zu einem ausreichenden Sampling kommt. Für geeignete Werte von ω , σ , τ_G und γ existieren prinzipiell keine Richtwerte und es sollten für jedes spezifische System verschiedene Kombinationen getestet werden. Je nach gewählter Kollektivvariable finden sich in der Literatur jedoch immer wiederkehrende, typische Bereiche zur Untersuchung von Proteinen, die als erster Anhaltspunkt dienen können. Als Richtwert für σ wird bspw. ein Drittel oder die Hälfte der Standardabweichung der Kollektivvariablen in einer klassischen MD verwendet [186]. Als beliebtes MetaMD-Modul hat sich bisweilen PLUMED (*Plugin for molecular dynamics*) etabliert, welches in Kombination mit verschiedenen Simulationsprogrammen genutzt werden kann [187, 188].

1.3 QSPR-Modelle zur Stabilitätsvorhersage

1.3.1 Grundlagen

1.3.1.1 Hintergrund

Modelle zur Beschreibung von quantitativen Struktur-Wirkungs/Eigenschafts-Beziehungen (*Quantitative structure-activity/property relationships*, QSAR/QSPR) werden in der medizinischen Chemie schon seit mehreren Jahrzehnten eingesetzt. Sie stellen eine Schnittstelle zwischen der Informatik und der Chemie dar und haben sich v. a. durch die immer noch rasant steigenden Mengen an verfügbaren Messdaten als hilfreiches Werkzeug zur Aufklärung möglicher Korrelationen zwischen Moleküleigenschaften und gewünschten biologischen oder physikochemischen Eigenschaften von Molekülen und darauf aufbauende Wirkstoffentwicklungen etabliert [189]. Die Studien von Hansch und Fujita zu Effekten des Octanol-Wasser-Verteilungskoeffizienten pflanzenwachstumsbeeinflussender Moleküle werden üblicherweise als zentraler Grundstein dieses Forschungsbereichs angesehen [189–193]. Durch eine steigende Rechenleistung und die Etablierung von Hochdurchsatz-Screenings in der Chemie haben sich die methodischen Ansätze stark weiterentwickelt, sodass nicht nur simple lineare Korrelationen Anwendung finden, sondern, unter Zuhilfenahme von Werkzeugen aus der von Wold begründeten Chemometrik [194] und dem maschinellen Lernen (*Machine learning*), auch andere Methoden wie bspw. Support Vector Machines oder künstliche neuronale Netze [195].

1.3.1.2 Deskriptoren

Als Variablen zur Modellgenerierung werden sogenannte Deskriptoren verwendet, die bestimmte Eigenschaften eines Moleküls beschreiben. Diese werden mit einer oder mehreren Antwortvariablen, zumeist der biologischen Aktivität, korreliert. Deskriptoren lassen sich in 1D (z. B. Summenformeln), 2D (Strukturformeln), 3D (Formeln mit Informationen zur räumlichen Orientierung bzw. Konformation), 4D (z. B. über eine Berücksichtigung eines zeitlichen Verlaufs oder mehrerer Konformationen oder Protonierungszustände gleichzeitig), 5D (Berücksichtigung der induzierten Passformen von Protein-Ligand-Komplexen) oder gar 6D (Berücksichtigung verschiedener Solvensmodelle) einteilen [189, 196]. Während 4/5/6D-Deskriptoren noch sehr selten genutzt werden, haben sich 3D-QSAR-Methoden besonders durch die Verwendung der PLS-Methode (*Partial least squares*) [197] nach Erscheinen des CoMFA (*Comparative molecular field analysis*)-Verfahrens [198] weithin in diesem Forschungsbereich etabliert [189]. Theoretische Hintergründe zu in dieser Arbeit relevanten Deskriptoren werden in Anhang B aufgezeigt.

1.3.2 Statistische Validierungsparameter

Zur Evaluation der prädiktiven Leistung von QSPR-Modellen haben sich verschiedene Parameter etabliert, von denen im folgenden Kapitel einige erläutert werden sollen.

1.3.2.1 Regressionsparameter

In [199] werden die bekanntesten Validierungsparameter für regressionsbasierte Fragestellungen zusammengefasst. Der am häufigsten genutzte Parameter ist der Determinationskoeffizient R^2 , auch Bestimmtheitsmaß genannt:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (1.35)$$

$$= 1 - \frac{RSS}{TSS} \quad (1.36)$$

Hierbei stellen RSS (*Residual sum of squares*) und TSS (*Total sum of squares*) die Residuenquadratsumme und die Quadratsumme der Abweichungen zum Mittelwert dar. y_i sind die einzelnen Messwerte, \bar{y} der Mittelwert der Messwerte, sowie \hat{y}_i die Vorhersagen. Zusätzlich kann ein adjustiertes Bestimmtheitsmaß R_{adj}^2 ermittelt werden, welches die Anzahl verwendeter Deskriptoren p und die Datensatzgröße n berücksichtigt [200]:

$$R_{adj}^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} \quad (1.37)$$

Mit R^2 wird das Bestimmtheitsmaß zumeist für die Messwerte, auf denen die Modellgenerierung basiert, bezeichnet. Wird der Determinationskoeffizient für Werte berechnet, die nicht Teil des Trainingsdatensatzes sind (n Datenpunkte eines externen Sets EXT), wird dieser zumeist mit Q^2 gekennzeichnet:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y})^2} \quad (1.38)$$

$$= 1 - \frac{PRESS}{TSS} \quad (1.39)$$

Die entsprechende RSS wird hier als $PRESS$ (*Predicted residual sum of squares*) bezeichnet. Hier kann entweder die TSS des Trainingsdatensatzes (Q_{F1}^2 nach Shi [201]) oder, damit der Parameter tatsächlich nur auf den externen Daten basiert, die des Testdatensatzes genutzt werden (Q_{F2}^2 nach Schüürmann [202]). Beide Q^2 -Varianten zeigen unter Umständen unrealistische Abschätzungen der prädiktiven Leistung des Modells, wenn die Verteilung des externen Datensatzes nicht vergleichbar mit der des Trainingsdatensatzes ist [203]. Ein sehr viel robusterer Validierungsparameter stellt hingegen der Konkordanzkorrelationskoeffizient (*Concordance correlation coefficient*,

CCC) nach Lin dar [204, 205], welcher folgendermaßen definiert ist:

$$CCC = \frac{2 \sum_{i=1}^{n_{EXT}} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y})^2 + \sum_{i=1}^{n_{EXT}} (\hat{y}_i - \bar{\hat{y}})^2 + n_{EXT}(\bar{y} - \bar{\hat{y}})^2} \quad (1.40)$$

Die Variable $\bar{\hat{y}}$ stellt den Mittelwert der Vorhersagewerte dar. Der Validierungsparameter CCC berücksichtigt sowohl die Residuen der Werte zur Regressionslinie als auch die Übereinstimmung letzterer mit den experimentellen Werten (gekennzeichnet durch eine annähernde Steigung von 1 beim Auftragen von experimentellen gegen prädiktive Werte) [199]. Daneben existieren ebenso verschiedene Rangkorrelationskoeffizienten, die besonders dann bedeutend sind, wenn es vorwiegend darum geht, ein Modell zu generieren, welches, ungeachtet der Abweichungen von den experimentellen Messwerten, die korrekte Reihenfolge der Datenpunkte vorhersagt. Zu erwähnen ist hierbei das Kendall'sche Tau τ_b [206, 207]:

$$\tau_b = \frac{C - D}{\sqrt{(C + D + X_0)(C + D + Y_0)}} \quad (1.41)$$

C ist die Anzahl an übereinstimmenden Paaren, D jene an nicht übereinstimmenden und X_0 sowie Y_0 die Anzahl an punktgleichen Paaren entlang der x- bzw. y-Achse [208]. Zur Beschreibung des Modellfehlers kann entweder der mittlere absolute Fehler (*Mean absolute error*, MAE) oder die Wurzel der Summe der quadrierten Fehler (*Root-mean-square error*, RMSE) verwendet werden, wobei letztere größere Fehler stärker gewichtet:

$$MAE = \frac{\sum_{i=1}^{n_{EXT}} |y_i - \hat{y}_i|}{n_{EXT}} \quad (1.42)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_{EXT}} (y_i - \hat{y}_i)^2}{n_{EXT}}} \quad (1.43)$$

1.3.2.2 Klassifikationsparameter

Zur Beurteilung der Leistung eines Klassifikationsmodells mit zwei verschiedenen Klassen kann die korrekte Klassifizierungsrate (*Correct classification rate*, CCR), die auch der Richtigkeit (*Accuracy*, Acc) entspricht, herangezogen werden:

$$CCR = \frac{N^{corr}}{N^{tot}} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1.44)$$

Hier wird die Anzahl an korrekt klassifizierten Fällen N^{corr} mit der Gesamtzahl N^{tot} in Relation gesetzt. Dies entspricht also der Summe an richtig positiven (*True positives*, TP) und negativen Fällen (*True negatives*, TN), geteilt durch die gesamte Anzahl an Fällen, inklusive falsch positiver (*False positives*, FP) und falsch negativer (*False negatives*, FN). Handelt es sich um relativ ungleiche Größen zweier Klassen, so sollte eine entsprechende Gewichtung berücksichtigt werden, z. B. über die ausgeglichene Richtigkeit (*Balanced accuracy*, BAcc), welche definiert ist als

$BAcc = \frac{Spezifität + Sensitivität}{2}$ [209]. Die Spezifität beschreibt den Anteil an tatsächlich negativen Fällen, welche korrekt als negativ eingestuft wurden (*True negative rate*, TNR). Die Falsch-positiv-Rate (FPR) ist entsprechend definiert als $1 - Spezifität$. Die Sensitivität stellt dagegen die Richtig-positiv-Rate (*True positive rate* (TPR) oder *recall*) dar und beschreibt den Anteil an tatsächlich positiven Fällen, der korrekt als positiv eingestuft wurde. Zusätzlich existiert die Präzision (*Positive predictive value*, PPV). Sie setzt die Anzahl an korrekt positiv eingestuften Fällen in Relation zur Gesamtzahl an positiv eingestuften Fällen. Das Gegenteil der Präzision ist die Segreganz (*Negative predictive value*, NPV) [210, 211]. Die beschriebenen Kriterien können aus der entsprechenden Konfusionsmatrix abgeleitet werden (Abb. 1.9).

$$TPR = \frac{TP}{TP + FN} \quad (1.45)$$

$$TNR = \frac{TN}{TN + FP} \quad (1.46)$$

$$FPR = 1 - TNR = \frac{FP}{TN + FP} \quad (1.47)$$

$$PPV = \frac{TP}{TP + FP} \quad (1.48)$$

$$NPV = \frac{TN}{TN + FN} \quad (1.49)$$

		Voraussage		
		Positiv	Negativ	
Klasse	Positiv	Richtig positiv	Falsch negativ	Sensitivität
	Negativ	Falsch positiv	Richtig negativ	Spezifität

Präzision
Segreganz

Abbildung 1.9: Darstellung einer Konfusionsmatrix, bei der die tatsächlichen Klassifizierungen den Vorhersagen gegenübergestellt werden. Es ergeben sich richtig positive, falsch positive, richtig negative und falsch negative Fälle. Die im Text beschriebenen Kriterien nutzen jeweils einen bestimmten Anteil an Fällen, der in der Abbildung farblich markiert ist.

Aufbauend auf diesen Definitionen stellt die Berechnung der Fläche unter der Grenzwertoptimierungskurve (ROC) eine etwas weiterführende Beschreibung der Modellleistung dar (*Area under the receiver operating characteristic curve*, AUROC bzw. AUC). Zur Erstellung einer ROC-Kurve wird die ermittelte TPR gegen die FPR aufgetragen. Eine früh und stark ansteigende Kurve weit über der Diagonalen des Graphen deutet auf ein Modell mit guter Klassifikationsleistung hin. Dagegen repräsentiert eine Kurve nahe der Diagonalen eine Modelleistung, die einer Zufallsklassifizierung entspricht. Die Kurve verdeutlicht somit die Leistung des Modells bei verschiedenen Grenzwerten.

Handelt es sich um einen Datensatz mit vielen negativen und wenigen positiven Fällen, kann es von besonderem Interesse sein, die Leistung des Modells im Hinblick auf die Detektion letzterer beurteilen zu wollen. Bei einer solch ungleichen Verteilung kann daher auch die AUPRC (*Area under the precision-recall curve*) berechnet werden,

die die Fläche unter der Kurve darstellt, wenn die TPR gegen die PPV aufgetragen wird. Hierbei fließt die Anzahl an tatsächlich negativen Fällen nicht mit ein. Im Gegensatz zu einer ROC-Kurve, bei der die Diagonale einer Zufallswahrscheinlichkeit mit dem AUC-Wert von 0.5 entspricht, stimmt der Basiswert der AUPRC mit dem Anteil an positiven Fällen an der Gesamtzahl überein. Je nach Situation kann daher auch ein niedriger AUPRC-Wert eine gute Detektion der tatsächlich positiven Fälle signalisieren. Die Segreganz spielt dagegen eine bedeutende Rolle bei der Beurteilung der Detektion negativer Fälle [212–214].

1.3.3 Modellarten dieser Studie

Da in dieser Arbeit QSPR-Modelle auf Basis sehr kleiner Datensätze generiert wurden, wurde der Fokus v. a. auf multilineare Regressions- und logistische Klassifikationsmodelle gelegt. Durch Nutzung von Datenbanken wie ChEMBL [215] liegen die Datensatzgrößen für QSAR-/QSPR-Ansätze heutzutage oftmals in einem Bereich zwischen 10^4 und 10^6 Molekülen [216], dagegen kommen Datensätze im Bereich um lediglich 100 Verbindungen, wie in dieser Arbeit, deutlich seltener vor [217, 218]. Im zuerst durchgeführten Ansatz 1 in Kap. 3.1.6.2 wurden jedoch auch eine Vielzahl anderer, komplexerer Modelle getestet. Hierbei lieferten neuronale Netze und XGBoost-Modelle erwähnenswerte Resultate. Der theoretische Hintergrund der vier erwähnten Modellarten soll im Folgenden kurz erläutert werden.

1.3.3.1 Multiple lineare Regression

Bei einer multiplen linearen bzw. multilinearen Regression werden die N Messergebnisse $y_1, \dots, y_i, \dots, y_N$ mit den entsprechenden Werten $x_1, \dots, x_i, \dots, x_p$ der p Prediktoren bzw. Deskriptoren in einer linearen Beziehung korreliert. Es wird also davon ausgegangen, dass die Antwortvariable über eine bestimmte Funktion $f(x)$ mit den Deskriptoren in Zusammenhang steht. Letztere werden hierzu mit den Koeffizienten β_1, \dots, β_p gewichtet (β_0 repräsentiert den Ordinatenabschnitt). Eine auf einer Gauß-Verteilung basierende, grundsätzlich vorhandene Abweichung (Mittelwert = 0) der tatsächlichen Werte von den Messwerten der Antwortvariablen wird zusätzlich mit dem Fehler ϵ berücksichtigt. Dieser schließt also unbekanntes, durch das Modell prinzipiell nicht abdeckbares, zufällige Messabweichungen ein.

$$f(x) = \beta_0 + \sum_{j=1}^p x_j \beta_j + \epsilon \quad (1.50)$$

Zur Ermittlung optimaler Koeffizienten während der Modellgenerierung wird üblicherweise die Methode der kleinsten Quadrate (*least squares*) angewandt. Es wird die RSS der Messpunkte y_i und deren Vorhersagen $f(x_i)$ minimiert:

$$RSS(\beta) = \sum_{i=1}^N (y_i - f(x_i))^2 \quad (1.51)$$

$$= \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 \quad (1.52)$$

Dies geschieht durch Nullsetzen der ersten Ableitung $\frac{\delta RSS}{\delta \beta}$. Es wird eine Hyperebene im \mathbb{R}^{p+1} -dimensionalen Raum konstruiert, bei der die Distanzen der Messpunkte zu dieser minimiert werden und deren orthogonale Projektionen die jeweiligen Vorhersagen $f(x_i)$ darstellen [219, 220].

1.3.3.2 Logistische, binomiale Regression

Über eine logistische, binomiale Regression kann die Wahrscheinlichkeit $p(X) = Pr(Y = 1|X)$ modelliert werden, welche Auskunft darüber gibt, ob eine Messung Y auf Basis der Eigenschaft X der Klasse 1 anstatt der Klasse 0 zugeordnet werden kann. Die Eigenschaft X resultiert hierbei aus einem oder mehreren Deskriptoren. Um Werte von $p(X) < 0$ und $p(X) > 1$ zu vermeiden, wird im Gegensatz zur linearen Regression eine logistische Funktion genutzt, die ein sigmoidales Verhalten aufweist:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (1.53)$$

Für alle Werte von X wird eine Wahrscheinlichkeit der Klassenzugehörigkeit $p(X)$ zwischen 0 und 1 erhalten. Es besteht hier ein linearer Zusammenhang von X mit der logarithmierten Chance $\frac{p(X)}{1-p(X)}$, das sogenannte Logit:

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X \quad (1.54)$$

Daraus folgt, dass eine Änderung von X zu $(X + 1)$ einer Multiplikation der Chance mit e^{β_1} entspricht. Die Änderung der eigentlichen Wahrscheinlichkeit $p(X)$ bei einer solchen Zunahme von X hängt jedoch vom vorherigem Wert von X ab, da zwischen Wahrscheinlichkeit und Deskriptor keine lineare Beziehung herrscht. Im Gegensatz zur Methode der kleinsten Quadrate bei einer multilinenen Regression wird zur Bestimmung von β_0 und β_1 die Methode der maximalen Mutmaßlichkeit (*Maximum likelihood*) angewandt. Es wird folgende Likelihood-Funktion optimiert:

$$\ell(\beta_0, \beta_1) = \prod_{i=1}^n p(x_i)^{y_i} (1 - p(x_i))^{1-y_i} \quad (1.55)$$

Auf Basis einzelner Bernoulli-Verteilungen resultiert diese Funktion für n Fälle mit den tatsächlichen Klassenzugehörigkeiten $y = 1$ oder $y = 0$ also aus dem Produkt aller Wahrscheinlichkeiten $p(x)$. Die Koeffizienten werden hierbei so gewählt, dass die maximale Anzahl an Messpunkten des Datensatzes durch die sigmoidale Funktion in die korrekte Klasse eingeteilt werden [219, 221].

1.3.3.3 Künstliches neuronales Netz

Bei einem künstlichen neuronalen Netz handelt es sich um eine nichtlineare Modellart, die für Klassifikationen und Regressionen genutzt werden kann. Die Deskriptoren werden, ähnlich der Vorstellung eines Netzes aus Nervenbahnen, über Neuronen (Linearkombinationen der Variablen) mit der Antwortvariablen $f(x)$ korreliert. Letztere steht in der Regel in einer linearen Beziehung zu den Werten der Neuronen h_1, \dots, h_H (mit den Gewichtungen γ_k und dem Ordinatenabschnitt γ_0).

$$f(x) = \gamma_0 + \sum_{k=1}^H \gamma_k h_k \quad (1.56)$$

Jedoch stehen diese H Neuronen, die zur Generierung der Vorhersagefunktion $f(x)$ dienen, zumeist über nichtlineare Funktionen mit den ursprünglichen P Deskriptoren in Beziehung. Steht der Wert des k -ten Neurons h_k bspw. in einer logistischen Beziehung zu den P Deskriptoren, so gilt für diesen (mit der Gewichtung β_j des j -ten Deskriptors):

$$h_k(x) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^P x_i \beta_i)}} \quad (1.57)$$

Für jedes Neuron werden die Deskriptoren unterschiedlich stark gewichtet. Durch die nichtlineare Beziehung der Deskriptoren mit den Neuronen, einem drastischen Anstieg an veränderbaren Parametern schon bei einem geringen Anstieg der Neuronenanzahl und der Möglichkeit mehrerer hintereinandergeschalteter Neuronen ist diese Modellart sehr anfällig für eine Überanpassung. Diese Modelle eignen sich daher in der Regel nur für sehr große Datensätze [211].

1.3.3.4 XGBoost-Modell

Der Begriff XGBoost steht für *Extreme gradient boosting* und stellt eine noch relativ neue Modellart dar, die 2016 von Chen und Guestrin entwickelt wurde [222] und in den darauffolgenden Jahren an Popularität gewann, da hiermit viele Machine Learning-Wettbewerbe gewonnen wurden [223]. Zunehmend findet diese Modellart auch in der Chemie Anwendung [224, 225]. Es handelt sich um ein parallelisiertes Verfahren, welches zu einem Ensemble an nacheinander generierten Entscheidungsbäumen führt. Dieses Ensemble wird über ein Gradientenverfahren zur Minimierung einer Zielfunktion $obj^{(t)}$ genutzt. Beim überwachten Lernen (*Supervised learning*) besteht letztere aus einer Verlustfunktion ℓ , die den Fehler zwischen Messwert y_i und Vorhersage \hat{y}_i beschreibt, und einem Regularisierungsterm Ω , der die Modellkomplexität berücksichtigt:

$$obj^{(t)} = \sum_{i=1}^n \ell(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^t \Omega(f_i) \quad (1.58)$$

Die Art der Verlustfunktion (z. B. der RMSE-Wert) kann bei diesem Modelltyp frei gewählt werden. Die Vorhersage $\hat{y}_i^{(t)}$ stellt den prädiktiven Wert nach t Trainingsritten mit zusätzlichen Entscheidungsbaumgenerierungen dar. Der Regularisierungsterm Ω bestraft komplexere Modelle und soll damit automatisch einer Überanpassung entgegenwirken (ähnlich der Regularisierung bei einer Ridge- oder Lasso-Regression):

$$\Omega(f_i) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (1.59)$$

T repräsentiert die Anzahl an endständigen Ästen im jeweiligen Baum, w_j deren Bepunktung, λ und γ sind vom Nutzer einstellbare Hyperparameter. Die Bepunktung der Äste hängt von der Anzahl an Messwerten ab, die diese jeweiligen Enden des Baums erreichen. Die Komplexität der Bäume wird ebenso durch ein Beschneiden (*Pruning*) beschränkt [226].

1.3.4 Limitierungen

Die breite Anwendung von QSPR-Modellen in verschiedensten Forschungsbereichen durch Nicht-Statistiker birgt einige Gefahren. Eine falsche Interpretation der prädiktiven Leistung solcher Modelle kann vielfältige Ursachen haben: eine zu hohe Heterogenität des Datensatzes, ein unvollständiger Datensatz, Nutzung nicht interpretierbarer oder kollinearere Deskriptoren, zu hohe Fehlerbereiche der verwendeten Deskriptoren, eine inadäquate Applikationsdomäne (*Applicability domain*, AD) und nicht zuletzt eine Überanpassung (*Overfitting*) an den Datensatz durch Verwendung von Modelltypen mit vielen Freiheitsgraden oder zu vielen Deskriptoren in Relation zur Datensatzgröße [189].

Das Problem, dass durch ein komplexes Modell im Prinzip aus jedem arbiträren Datensatz eine scheinbar vielversprechend wirkende Korrelation generiert werden kann, die jedoch keinen tatsächlich gegebenen Zusammenhang der Deskriptoren mit der Größe der Antwortvariablen begründen muss, wird durch folgendes Zitat von Unger und Hansch treffend beschrieben:

„One must rely heavily on statistics in formulating a quantitative model but, at each critical step in constructing the model, one must set aside statistics and ask questions. [...] Without such a qualitative perspective, one is apt to generate statistical unicorns, beasts that exist on paper but not in reality [...] it has recently become all too clear that one can correlate a set of dependent variables using random numbers as independent variables. Such correlations meet the usual criteria of high significance“ [227].

Aufgrund der geschilderten Problematiken wird empfohlen, sich bei der Modellgenerierung an gewisse Richtlinien, z. B. jene von Tropsha [228], zu halten. Nicht zuletzt hat auch die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (*Organisation for Economic Co-operation and Development*, OECD) solche Richtlinien publiziert:

„To facilitate the consideration of a QSAR model for regulatory purposes, it should be associated with (1) a defined end point, (2) an unambiguous algorithm, (3) a defined domain of applicability, (4) appropriate measures of goodness-of-fit, robustness, and predictivity, (5) a mechanistic interpretation, if possible“ [229].

Auch aufgrund einer oftmals nicht gegebenen Publikation aller wesentlichen Modellparameter streben wissenschaftliche Journale immer öfter gewisse Grundsätze für QSPR-Modelle an, um eine Reproduzierbarkeit sicherzustellen [230]. Dennoch ist dies ein schweres Unterfangen, da solche Modelle, wie bereits oben erwähnt, in verschiedensten Forschungsgebieten für unterschiedliche Problemstellungen Anwendung finden.

1.4 Constrained-Network-Analyse

1.4.1 Theoretischer Hintergrund

Eine sogenannte Constrained Network-Analyse (CNA) kann dazu dienen, die Stabilität von Strukturen auf Basis ihrer Rigidität zu beurteilen [231–239]. Diese basiert auf dem FIRST-Programm (*Floppy inclusions and rigid substructure topography*) und dessen Pebble Game-Algorithmus zur Identifizierung rigider Cluster und flexibler Regionen in Biomakromolekülen. Die Atome des Moleküls werden hierbei zur Generierung eines Netzwerks als Körper mit zunächst jeweils sechs Freiheitsgraden gehandhabt. Durch jede kovalente und nichtkovalente Wechselwirkung (letzteres umfasst HB, Salzbrücken und hydrophobe Interaktionen) werden diese Körper bzw. Eckpunkte über Stäbe (*Bars*) mit anderen Atomen des Moleküls im Netzwerk verbunden und somit in ihrer Flexibilität und ihren Freiheitsgraden eingeschränkt. Nach Generierung eines solchen *Body-and-bar*-Netzwerks aus Körpern und Stäben werden in diesem rigide Areale determiniert, welche sich dadurch kennzeichnen, dass aufgrund der Vernetzung keine relativen Bewegungen zwischen den Körpern ermöglicht wird [234].

Das Prinzip basiert auf der Graph-Theorie [236] und findet seinen Ursprung in den Arbeiten von Maxwell im 19. Jahrhundert [240]. Interne Freiheitsgrade F werden auch als *Floppy modes* bezeichnet und deren Anzahl kann in einem d -dimensionalen Raum über folgende Beziehung beschrieben werden:

$$F = dN - (N_c - N_r) - \frac{d(d+1)}{2} \quad (1.60)$$

Dabei stellen N die Anzahl an Eckpunkten im Netzwerk, N_c die Anzahl an Stäben bzw. Resktriktionen (*Constraints*) und N_r die redundanten Constraints dar. Als redundant werden Resktriktionen aufgefasst, wenn sie nicht zur Stabilisierung des Netzwerks beitragen [241]. Zur Anwendung der Theorie auf Proteine werden Einfachbindungen mit jeweils fünf Stäben, nichtrotierbare Bindungen mit sechs Stäben repräsentiert. HB werden ebenso mit fünf Constraints, hydrophobe Wechselwirkungen mit zwei Constraints abgebildet. Letztere werden berücksichtigt, sofern die Distanz eines Kohlenstoff- oder Schwefelatoms zu einem anderen geringer ist als die Summe ihrer VdW-Radii, addiert mit einem spezifischen Cut-off-Wert (normalerweise bei $D = 0.25 \text{ \AA}$). HB werden in ihrer Stärke E_{HB} über folgende Gleichung definiert:

$$E_{HB} = D_0 \left\{ 5 \left(\frac{R_0}{R} \right)^{12} - 6 \left(\frac{R_0}{R} \right)^{10} \right\} f(\theta, \Phi, \phi) \quad (1.61)$$

wobei $D_0 = 8 \text{ kcal/mol}$ das Minimum des Interaktionspotentials darstellt, $R_0 = 2.8 \text{ \AA}$ die Equilibriumsdistanz zwischen Donor und Akzeptor, sowie R die tatsächliche Distanz. Die Funktion $f(\theta, \Phi, \phi)$ hängt von der Hybridisierung der beteiligten Atome ab. Bspw. nimmt die Funktion die Form $f = \cos^2(\theta) \cos^2(\Phi - 109.5)$ im Falle von sp^3 -hybridisierten Donoren und Akzeptoren an. Die Variablen θ , Φ und ϕ beschreiben

dabei die Winkel zwischen den beteiligten Atomen (Donor, Akzeptor und Wasserstoffatom). Die Einschränkung der internen Freiheitsgrade wird schließlich über den 3D-(6,6)-Pebble Game-Algorithmus definiert. Hierbei werden die Freiheitsgrade als $6N - 6$ Steine (*Pebbles*) visualisiert und zunächst auf alle Eckpunkte gleichmäßig verteilt. Gemäß gewissen Regeln werden diese dann auf die bestehenden Stäbe zwischen den Körpern verteilt. Die Anzahl an verbleibenden Pebbles auf den Eckpunkten gibt schließlich Auskunft über die Rigidität des Systems und beschreibt die Freiheitsgerade [242–244].

1.4.2 Anwendung des CNA-Programms

Das CNA-Programm erweitert die Funktionen des FIRST-Programms, indem es thermale Entfaltungssimulationen auch auf Basis von Strukturensamples ermöglicht und während diesen auch die Verstärkung hydrophober Wechselwirkungen mit zunehmender Temperatur realisieren kann. Durch die Nutzung von Ensembles, die bspw. aus MD-Simulationen gewonnen werden, können die Rigiditätsanalysen der Entfaltungssimulationen deutlich verbessert werden. Bei diesen Simulationen wird die Stärke E_{HB} von HB im Netzwerk über die oben beschriebene Glg. 1.61 [236] bestimmt. Es handelt sich um die Cut-off-Werte, bei denen diese aufgelöst werden (in vergleichbarer Weise können Salzbrücken zerstört werden). Über eine zuvor empirisch bestimmte lineare Beziehung ($T_m = 300K - 20 E_{HB}$) werden diese Cut-off-Werte im CNA-Programm mit spezifischen Temperaturwerten gleichgesetzt. Die genannte Beziehung basiert auf einem systematischen Vergleich der mit dem CNA-Programm ermittelten Entfaltungstemperaturen mit experimentellen Daten zu 20 verschiedenen Proteinen von mesophilen Bakterien und homologen thermophilen Varianten [231]. Das anfangs generierte Netzwerk wird somit steigenden Temperaturen ausgesetzt, indem der Cut-off-Wert der Energie sukzessive geändert wird und schwache HB nach und nach aus dem System entfernt werden. Nach jedem Temperatursprung werden das Netzwerk und rigide Cluster neu evaluiert. Das CNA-Programm ermöglicht im Anschluss eine Beurteilung der thermischen Stabilität auf Basis verschiedener globaler und lokaler Rigiditätsparameter. Eine nähere Beschreibung ist in Anhang B.13 gegeben.

Auch Wassermoleküle und Liganden können bei der CN-Analyse berücksichtigt werden (wobei das Ergebnis in der Regel durch eine Einbeziehung von Wassermolekülen auf der Proteinoberfläche nicht signifikant beeinflusst wird). Findet das Programm keine Informationen zu Bindungsordnungen der Heteroatome in der PDB-Datenbank (basierend auf dem dreistelligen PDB-Code des Moleküls), so werden diese mithilfe von OpenBabel [245, 246] über eine distanzbasierte Konnektivität errechnet und die Hybridisierungen der Atome somit bestimmt [234]. Eine Visualisierung der gefundenen Cluster sowie der nichtkovalenten Wechselwirkungen des Netzwerks und der beschriebenen globalen und lokalen Parameter kann über das kostenlos zur Verfügung stehende PyMOL-Plugin *visualCNA* [246] erfolgen.

1.5 Untersuchte Strukturen

Im Folgenden werden die für diese Arbeit relevanten Peptide und Proteine beschrieben. Dabei sollen zunächst die kleineren Modellproteine vorgestellt werden, deren Konjugate in dieser Arbeit zum Studium des Effekts von PEG auf die konformative Proteinestabilität, sowie zum Vergleich der alternativen Polymere LPG, MeOx und PEtOx dienen (Kap. 4). Danach wird das größere, therapeutisch bedeutsame Protein Interferon- α 2a (IFN- α 2a) beschrieben, dessen PBK in Kap. 5 hinsichtlich favorisierter Protein-Polymer-Wechselwirkungen, thermischer Stabilität und Bioaktivität näher untersucht wurden.

1.5.1 Modellproteine

1.5.1.1 Pin 1 WW-Domäne

Die humane Peptidyl-prolyl-cis-trans-Isomerase bzw. Rotamase Pin 1 spielt eine wichtige Rolle in der G2/M-Phase des Zellzyklus. In Eukaryoten ist dieses Protein hochkonserviert. Es besteht aus einer N-terminalen, 39 Aminosäuren langen WW-Domäne und einer größeren C-terminalen Domäne. Während letztere für die Funktion des Proteins als Rotamase verantwortlich ist, dient die kleinere WW-Domäne zur Etablierung von Protein-Protein-Interaktionen über prolinreiche Areale von Substraten (Erkennung von PPXY-Sequenzen). Der Name resultiert aus den zwei Tryptophanen der Domäne (W11 und W34 im Falle der humanen Variante). Das Protein wurde 1997 kristallisiert, die Struktur besitzt eine Auflösung von 1.35 Å (PDB: 1PIN) und ist in Abb. 1.10-A dargestellt. Ein PEG-Pentamer findet sich hier zwischen den Aminosäuren Y23, S32 und W34 der WW-Domäne und N90 und K97 der C-terminalen Domäne. Im Bereich zwischen L22, C113 und M130 letzterer ist weiterhin ein Dipeptid kristallisiert [247, 248].

Die WW-Domäne besteht aus drei antiparallelen β -Faltblättern, die über zwei Schleifen (*Reverse turns*) miteinander verbunden sind (Abb. 1.10-B). Die Stabilität des Proteins wird maßgeblich durch zwei gegenüberliegende, hydrophobe Zentren bestimmt. Weiterhin sind eine Salzbrücke zwischen E12 und R14, sowie ein ausgedehntes HB-Netzwerk zwischen den Aminosäuren P9, E12, N26, H27, I28 und T29 als stabilisierende Elemente beschrieben [249]. Die Aminosäuren M1 bis E6 wurden in der Kristallstruktur nicht aufgelöst und auch nicht von Price et al. in deren Studien berücksichtigt (siehe Kap. 1.1.6.2), sodass ein Konstrukt aus 34 Aminosäuren verwendet wurde.

Aufgrund der geringen Größe und einer anzunehmenden, simplen Zwei-Phasen-Kinetik während der Entfaltung (was allerdings in einer Infrarotspektroskopiestudie infrage gestellt wurde [250]) wurde die β -faltblattreiche WW-Domäne bereits in verschiedenen Mutations- und Faltungsstudien als Modellprotein verwendet. Die Entfaltung findet vermutlich zunächst durch eine Loslösung des kleineren, dritten β -Faltblatts

von der Tertiärstruktur statt, wohingegen eine Denaturierung des ersten β -Faltblatts eine höhere Energiebarriere darstellt [247, 251–260].

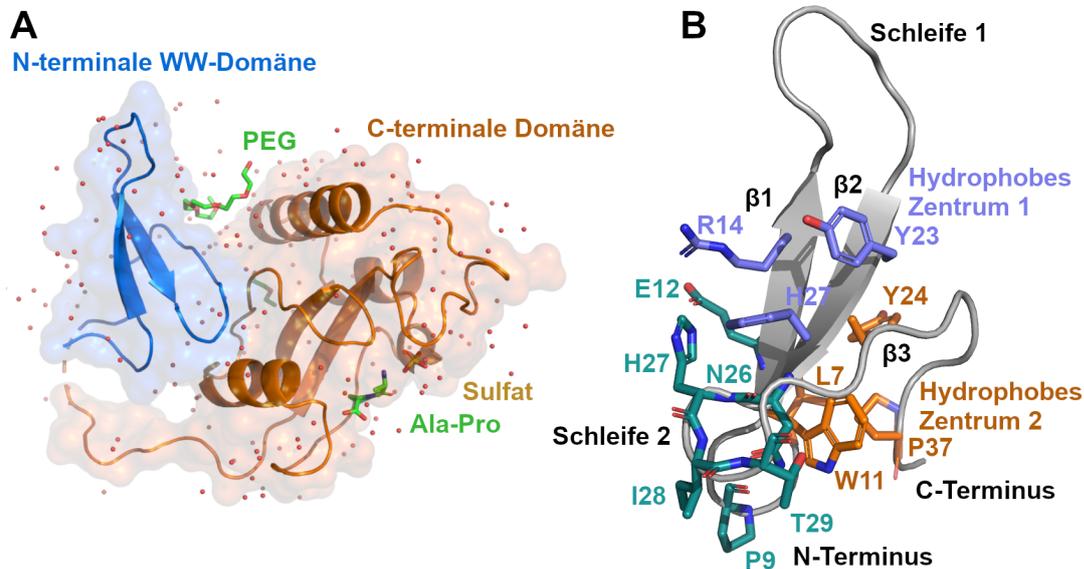


Abbildung 1.10: (A) Darstellung der Pin 1-Kristallstruktur 1PIN [247] mit unterschiedlicher Färbung der beiden Domänen (blaue WW-Domäne: Aminosäuren 1 bis 39, orangene C-terminale Domäne: Aminosäuren 45 bis 163). (B) Illustration der WW-Domäne. Hydrophobe Zentren sind violett und orange gefärbt, am besagten HB-Netzwerk teilnehmende Aminosäuren grün.

1.5.1.2 Her2-Affibody

Im Gegensatz zur β -faltblattreichen Struktur der Pin 1 WW-Domäne stellt der Her2-Affibody ein α -helikales Peptid dar. Als antigenbindende Alternativen zu Antikörpern leiten sich die wesentlich kleineren, 58 Aminosäuren langen Affibodies von der IgG-bindenden Domäne des Proteins A aus *Staphylococcus aureus* ab. Durch einige Modifikationen dieser B-Domäne wurde 1987 die stabilere Z-Domäne erhalten. Randomisierte Mutationen von 13 Aminosäuren der ersten zwei Helices brachten mittels eines Phagen-Displays anschließend Peptide hervor, die mit einer hohen Affinität z. B. an Insulin, TNF (Tumor-Nekrose-Faktor)- α oder auch den humanen, epidermalen Wachstumsfaktor 2 (*Human epidermal growth factor receptor 2*, Her2) binden. Bei einer Überexpression dieses Faktors, v. a. im Falle von Mammakarzinomen, kann ein solcher Affibody als ein wichtiges, diagnostisches Werkzeug eingesetzt werden. Es bindet an ein anderes Epitop als therapeutisch verwendete Antikörper wie bspw. Trastuzumab und bietet aufgrund seiner geringen Größe im Vergleich zu Antikörpern einige Vorteile, bspw. eine verbesserte Löslichkeit. Da Her2-Affibodies von der Z-Domäne abstammen, werden sie auch Z_{Her2}-Affibodies genannt [261].

Für diese Studie diente eine NMR-Struktur der Z_{Her2:342}-Variante (PDB: 2KZJ) als Ausgangspunkt (Abb. 1.11). Die bei der Strukturaufklärung erhaltenen Messdaten erlauben ebenso eine alternative Struktur dieses Peptids mit einer nicht durchgängigen

ersten Helix (PDB: 2KZI). In Lösung kommt es möglicherweise zu einer Interkonversion beider Zustände, wobei in der Kristallstruktur nach Bindung an Her2 (PDB: 3MZW) die durchweg helikale Konformation aufgelöst wurde [262].

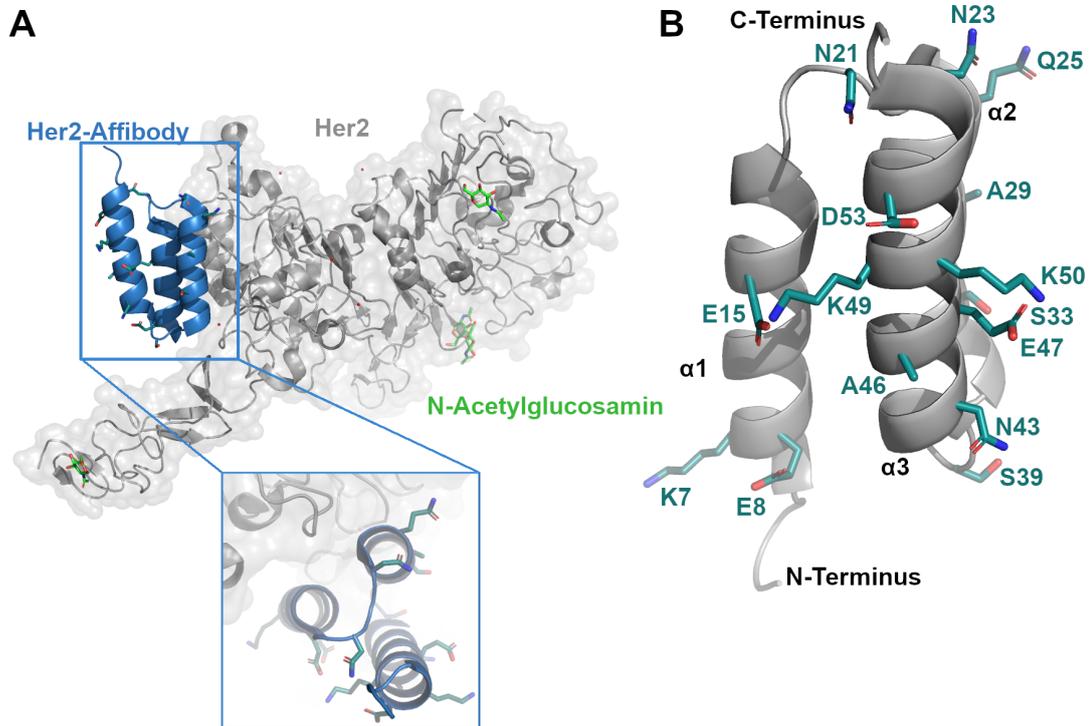


Abbildung 1.11: (A) Darstellung des Her2-Affibodys in der Kristallstruktur 3MZW, gebunden an die extrazelluläre Domäne von Her2. Die Bindestelle ist in einem Ausschnitt unterhalb der Struktur vergrößert dargestellt. In (B) ist der Affibody ohne Her2 vergrößert dargestellt (NMR-Struktur 2KZJ). In dieser Arbeit untersuchte Konjugationsstellen sind in beiden Bildern türkis gefärbt.

1.5.1.3 Src SH3-Domäne

Während in dieser Studie primär Konjugate der Pin 1 WW-Domäne und des Her2-Affibodys untersucht werden sollten, wurde ein PBK der Src SH3-Domäne des Bankivahuhns (*Gallus gallus*) von Price et al. als eine Validierungsmöglichkeit für seine kristallstrukturbasierte Vorhersagerichtlinie zur PEG-basierten Stabilisierung genutzt [100]. Es handelt sich um ein der Pin1 WW-Domäne strukturell sehr ähnliches Protein, welches ebenso über viele β -Faltblattstrukturbereiche verfügt (Abb. 1.12). Eine NMR-Struktur wurde 1993 veröffentlicht (PDB: 1SRL) [263]. Price et al. charakterisierten eine T20N-Mutante und die entsprechend konjugierte Asn(PEG)₄-Variante. Eine starke Stabilisierung mit einem $\Delta\Delta G_f^0$ -Wert von -1.20 ± 0.10 kcal/mol wurde bestimmt [100]. Wie auch die WW-Domäne dient diese Struktur in der Zelle zur Erkennung prolinreicher Elemente [247].

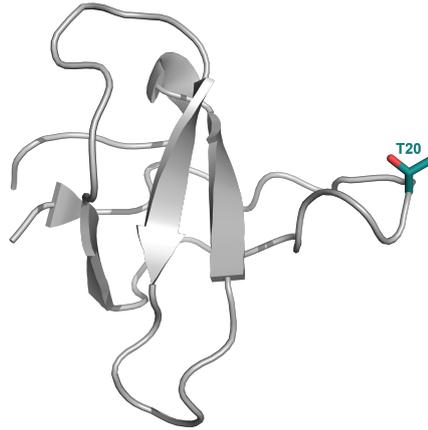


Abbildung 1.12: Darstellung der Src SH3-NMR-Struktur mit der von Price et al. untersuchten Mutationsstelle T20.

1.5.1.4 GCN4 und coil-V_aL_d

In ihren neuesten Publikationen [102, 264] charakterisierten Price et al. ebenfalls Konjugate der α -helikalen, Doppelwendeln (*Coiled coils*) bildenden Peptide GCN4 und coil-V_aL_d (Abb. 1.13). Während es sich bei GCN4 um einen homodimeren Leucin-Zipper der Hefe (*Saccharomyces cerevisiae*) handelt, welcher als Transkriptionsfaktor fungiert (PDB: 2ZTA) [265], stellt das Trimer coil-V_aL_d ein artifizielles bzw. entworfenes Konstrukt dar, das von Price et al. auch als 1CW bezeichnet wurde (PDB: 1COI [266] und das strukturell ähnliche Trimer mit dem Code 5UXT [267]).

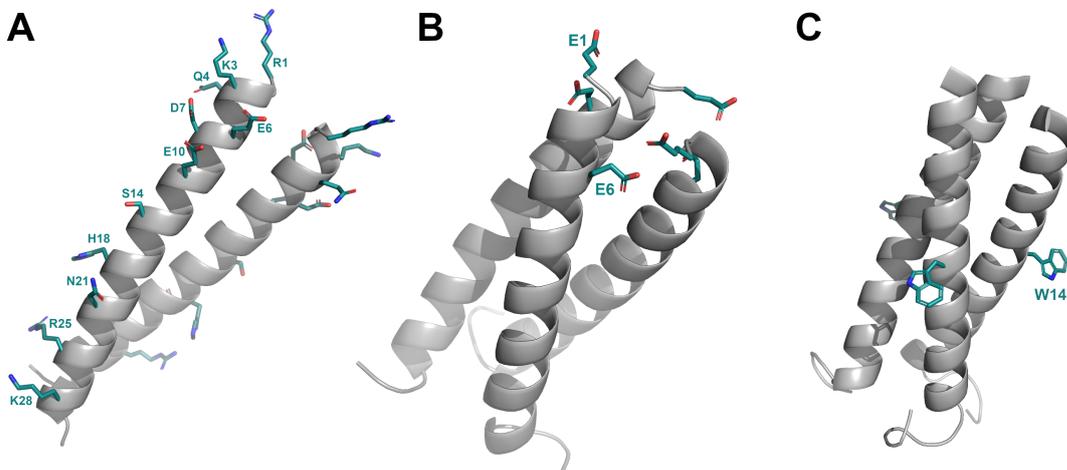


Abbildung 1.13: Strukturen (A) GCN4, sowie von (B) coil-V_aL_d und (C) dem davon abgeleiteten Trimer mit dem PDB-Code 5UXT. Die von Price et al. untersuchten Konjugationsstellen sind grün gefärbt und jeweils für ein Monomer gekennzeichnet.

1.5.2 Biologika

1.5.2.1 Interferon- α 2a

Das Protein IFN- α 2a bindet an die Interferon- α -Rezeptoren IFNAR1 und IFNAR2. Über eine dadurch induzierte Signalkaskade mit anschließender Gentranskription besitzt das therapeutisch verwendete Protein eine immunmodulierende und antivirale Wirkung. Aus diesem Grund wird es gegen unterschiedliche Tumoren, sowie Hepatitis B und C eingesetzt (in letzterem Falle zumeist in Kombination mit Ribavirin). Die konventionelle interferonbasierte Therapie gegen Hepatitis C spielt in aktuellen Leitlinien im Gegensatz zu interferonfreien DAA-Therapien (*Direct acting agents*) mit verschiedenen Protease/NS5A/NS5B-Inhibitoren nur noch eine untergeordnete Rolle. Die unPEGylierte Variante (Handelsname: Roferon-A®) besitzt eine äußerst geringe Halbwertszeit von lediglich etwa sechs bis neun Stunden, weshalb die PEGylierte Variante Peginterferon- α 2a mit dem Handelsnamen Pegasys® (siehe Kap. 1.1.4.3) entwickelt wurde, welche deutliche pharmakokinetische Vorteile bietet [18, 268–270]. Untersuchungen in Bezug auf Hepatozyten deuten allerdings darauf hin, dass eine längere Aktivierung des Jak/STAT-Signalwegs nicht den alleinigen Grund für eine verbesserte Therapie darstellen könnte. Durch ein verändertes pharmakodynamisches Profil könnten weitere immunmodulierende Gene aktiviert werden [271]. In Abb. 1.14-A ist die Kristallstruktur mit dem PDB-Code 3SE3 abgebildet, in der das Protein an seine Rezeptoren gebunden ist [272]. Die NMR-Struktur 1ITF des isolierten Proteins wurde für die in dieser Arbeit durchgeführten Simulationen genutzt (Abb. 1.14-B). Es besteht aus fünf α -helikalen Abschnitten mit insgesamt 165 Aminosäuren [273].

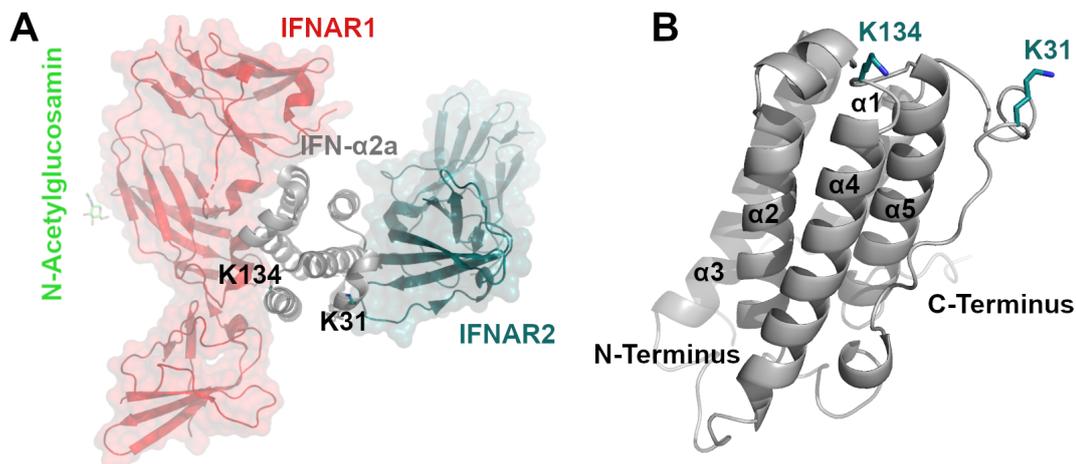


Abbildung 1.14: (A) Kristallstruktur 3SE3 mit IFN- α 2a und den beiden Rezeptoren IFNAR1 und IFNAR2. Die beiden potentiellen, in Kap. 3.2.2 genannten Konjugationsstellen K31 und K134 sind gekennzeichnet. In (B) ist die isolierte IFN- α 2a-Struktur aus der PDB-Datei 1ITF gezeigt.

Kapitel 2

Zielsetzung

Die in Kap. 1.1.6 geschilderten Hypothesen zu Stabilisierungsmechanismen von PBK verdeutlichen, dass das Verhalten von Konjugaten auf molekularer Ebene noch nicht gänzlich erforscht ist, erst recht nicht im Falle der selten genutzten PEG-Alternativen. Ein besseres Verständnis der Polymerdynamiken nach kovalenter Knüpfung an Proteine kann dazu beitragen, Veränderungen der pharmakokinetischen, pharmakodynamischen und thermodynamischen Eigenschaften von PBK in Relation zu ihren unkonjugierten Varianten zu erklären. Übergeordnetes Ziel dieser Studie ist es, diese Dynamiken mithilfe von MD-Simulationen näher zu analysieren und einen rationalen Ansatz zur Auswahl der geeignetsten Konjugationsstellen und Polymerarten zu entwickeln, ohne experimentell kostspielige und zeitaufwendige Screenings durchführen zu müssen. Speziell sollen die Polymere LPG, PMeOx und PEtOx als potenzielle Alternativen mit dem Goldstandard PEG verglichen werden.

Hierfür sollen in dieser Arbeit in adäquater Zeit simulierbare Oligomer-Konjugate zweier kleinerer Modellproteine, der humanen Pin 1 WW-Domäne und des Her2-Affibodys, analysiert werden. Die bereits von Price et al. experimentell charakterisierten Varianten der WW-Domäne (siehe Kap. 1.1.6.2) eignen sich dabei als Ausgangspunkt zum Studium der PEG-basierten Proteinstabilisierung. Das Verhalten der Konjugate in den Simulationen soll mit deren (De-)stabilisierungen in Zusammenhang gebracht werden, bspw. mittels QSPR-Modellen. Experimentelle Daten zu Her2-Affibody-, LPG-, PMeOx- und PEtOx-Konjugaten stehen im Vorfeld nicht zur Verfügung, sodass hier lediglich Schlussfolgerungen auf Basis der *in silico*-Daten gezogen werden können. Zusätzliche Synthesen weiterer, von Price et al. noch nicht charakterisierter Konjugate sind im Rahmen der Kooperation mit Prof. Dr. Dr. Lorenz Meinel und Prof. Dr. Tessa Lüthmann jedoch denkbar.

Die Untersuchungen sollen schließlich für das therapeutisch relevante Protein IFN- α 2a fortgesetzt werden. Die Arbeiten der Verbundpartner in diesem vom Bundesministerium für Bildung und Forschung finanzierten Projekt mit dem Titel *PEG-alternative Polymere für die Biokonjugation an Proteine (Next-PEG)* fokussieren sich auf die Synthese und Charakterisierung von Biokonjugaten dieses Therapeutikums (mit Polymerketten im Größenbereich zwischen 5 und 50 kDa). In der Arbeitsgruppe von

Prof. Dr. Rainer Haag sollen an der Freien Universität Berlin die Polymere PEG und LPG synthetisiert werden, im Arbeitskreis von Prof. Dr. Ulrich S. Schubert an der Friedrich Schiller-Universität in Jena die Polymere PMeOx und PEtOx.

Im Arbeitskreis von Prof. Dr. Dr. Lorenz Meinel und Prof. Dr. Tessa Lühmann an der Julius-Maximilians-Universität Würzburg werden diese Polymere mittels Click-Chemie-Verfahren an Proteine konjugiert und hinsichtlich ihrer Größe, thermischen Stabilität und Bioaktivität charakterisiert. Unter der Aufsicht von Dr. Christian Pfaller am Paul-Ehrlich-Institut sollen die Arbeiten durch *in vivo*-Studien der bereitgestellten Konjugate weitergeführt werden.

Simulationsstudien zu IFN- α 2a-Konjugaten sollen in der hier vorliegenden Arbeit durchgeführt werden, um die Ergebnisse der experimentellen Charakterisierungen der Verbundpartner zur thermischen Stabilität und Bioaktivität auf molekularer Ebene zu interpretieren.

Zusammengefasst soll in dieser Arbeit mithilfe von Simulationsverfahren und chemoinformatischen Methoden die Dynamik unterschiedlicher PBK studiert werden. Der Effekt der Konjugation auf pharmazeutisch relevante Eigenschaften soll anhand der Ergebnisse dieser Studien gedeutet und darauf aufbauend die Entwicklung optimierter Konjugate mit rationalen Ansätzen vorangetrieben werden. Übergeordnetes Ziel des Verbundprojekts ist die Charakterisierung und Etablierung von PEG-Alternativen für die Biokonjugation, wozu die vorliegende Dissertation mit Hilfe computergestützter Modelle und Simulationen auf atomarer Ebene einen Beitrag leisten will.

Kapitel 3

Materialien und Methoden

3.1 Modellproteine

Im Folgenden sollen die simulierten Konjugate der Modellproteine aus Kap. 1.5.1.1 vorgestellt (Kap. 3.1.2), sowie die Methoden zur Parametrisierung (Kap. 3.1.3), Modellierung (Kap. 3.1.4) und Simulation (Kap. 3.1.5) dieser Strukturen beschrieben werden. Im Anschluss werden die Vorgehensweisen zur Generierung von QSPR-Modellen (Kap. 3.1.6), zur Durchführung von Metadynamik-Simulationen (Kap. 3.1.7) und zum systematischen Vergleich der Dynamik der Polymere (Kap. 3.1.8) geschildert. Die Ergebnisse werden in Kap. 4 vorgestellt und diskutiert.

3.1.1 Wildtyp-Simulationen und unkonjugierte Mutanten

Die Wildtypen (WT) der Modellproteine aus Kap. 1.5.1.1 wurden mit den unten beschriebenen Einstellungen für Konjugate simuliert (Kap. 3.1.4 und 3.1.5).

Unkonjugierte Mutanten der in Kap. 3.1.2 beschriebenen Konjugate mit entsprechender Linkerstruktur (zumeist also Asparagin) und ohne Polymer wurden analog simuliert. Zur Messung des von Price et al. postulierten Winkels diente die Struktur ohne Mutation der Konjugationsstelle (d. h. ohne Linker), was in den meisten Fällen dem Wildtyp entsprach. Für einige Varianten, die neben der Konjugationsstelle zusätzliche Punktmutationen enthielten (z. B. 19(S16A)-1, vgl. Namen in Tab. C.2) wurden die entsprechenden Mutanten des Wildtyps zusätzlich simuliert.

3.1.2 Simulierte Konjugate

3.1.2.1 Konjugate mit publizierten Stabilitätsdaten

Von den in Kap. 1.5.1 vorgestellten Konstrukten wurden, mit Ausnahme der Konjugate von 1CW mit Asn(PEG)₄₅-Struktur aus [108], alle von Price et al. experimentell charakterisierten Konjugate der Pin 1 WW-Domäne, der Src SH3-Domäne und der Doppelwendeln GCN4 und coil-V_aL_d simuliert (insgesamt 131 Varianten). Es handelte sich hierbei zum größten Teil um PEG-Tetramere mit einer endständigen Methoxygruppe, welche kovalent an Seitenkettenamide von Asparaginen geknüpft wurden.

Eine komplette Übersicht aller in dieser Arbeit simulierten Konjugate der Modellproteine geben die Tab. C.1 (chemische Formeln der konjugierten Aminosäuren) und C.2 (Auflistung aller PBK mit Sequenzen) in Anhang C. In Tab. C.2 wurden die Bio-konjugate nach einem einheitlichen Schema benannt. Diese Namen werden in dieser Dissertation durchweg verwendet, wobei die Pin 1-Konjugate ohne PDB-Code und entsprechende unkonjugierte Varianten lediglich mit der Position der Konjugationsstelle (also ohne Kennzeichen für den Linker) bezeichnet werden. Die untersuchten Konjugationsstellen für die Src SH3-Domäne und die Doppelwendeln GCN4 und coil-V_aL_d wurden bereits in den Abb. 1.12 und 1.13 gezeigt. Im Falle der Pin 1 WW-Domäne sind die analysierten Varianten in Abb. 3.1 illustriert.

3.1.2.2 Zusätzliche Konjugate

Zusätzlich zu den von Price et al. charakterisierten Varianten der Pin 1 WW-Domäne wurden weitere PBK mit Asn(PEG)₄-Struktur simuliert (Abb. 3.2). Diese könnten in zukünftigen Arbeiten analog synthetisiert und mittels temperaturabhängiger CD-Spektroskopie charakterisiert werden. Ebenso wurden zum Vergleich der alternativen Polymere LPG, PMeOx und PEtOx jeweils 13 Asn(LPG)₅-, Asn(PMeOx)₅- und Asn(PEtOx)₅-Strukturen modelliert (siehe Kennzeichen ϵ , λ und ψ in Tab. C.1). Eine Anzahl von fünf Monomeren wurde hierbei ausgewählt, da zu diesem Zeitpunkt der Arbeit zur Aussicht stand, Konjugate mit einer solchen Struktur im Rahmen der Kooperation mit den Verbundpartnern analog zu vermessen. Hierzu sollten experimentelle Daten von niedermolekularen Polyoxazolingemischen mit einem Polymerisationsgrad von vier bis sechs Monomeren gewonnen werden. Die Länge der LPG-Variante wurde dementsprechend analog auf fünf Monomere gesetzt.

Da bisher keine PBK des Her2-Affibodys experimentell charakterisiert wurden, wurden in diesem Falle 15 verschiedene Konjugationsstellen ausgewählt und ebenso mit (PEG)₄-, (LPG)₅-, (PMeOx)₅- und (PEtOx)₅-Oligomeren modelliert. Hierbei wurde jedoch eine Cys-Maleimid-Struktur als Linker verwendet (siehe Kennzeichen δ , ϕ , γ und ω in Tab. C.1), da von den Verbundpartnern geplant war, diese Varianten ebenfalls zu charakterisieren. Diese sollten nicht, wie im Falle der WW-Konjugate von Price et al., mittels Einbau von PEGylierten Aminosäuren über eine Festphasenpeptidsynthese synthetisiert werden, sondern mithilfe eines Click-Chemie-Verfahrens nach erfolgter Expression des Affibodys mit entsprechender Cystein-Mutation. Die Stellen sind über das gesamte Peptid verteilt, jedoch nicht im unmittelbaren Bereich der Binderegion zu Her2 (vergleiche Vergrößerung in Abb. 1.11-A). Ein weiteres Auswahlkriterium hierbei waren die zwei von Price et al. bereits publizierten Richtlinien, beschrieben in Kap. 1.1.6.2. Lysine an den Positionen $i + 3$ und $i + 4$ sollten zu einer Stabilisierung gegenüber der Cystein-Mutante führen [108], weshalb die Aminosäuren A46 und D53 ausgewählt wurden. Zusätzlich wurden Konjugationsstellen mit relativ kleinen oder großen Werten für den Winkel θ_{Price} bei der Auswahl priorisiert. PBK

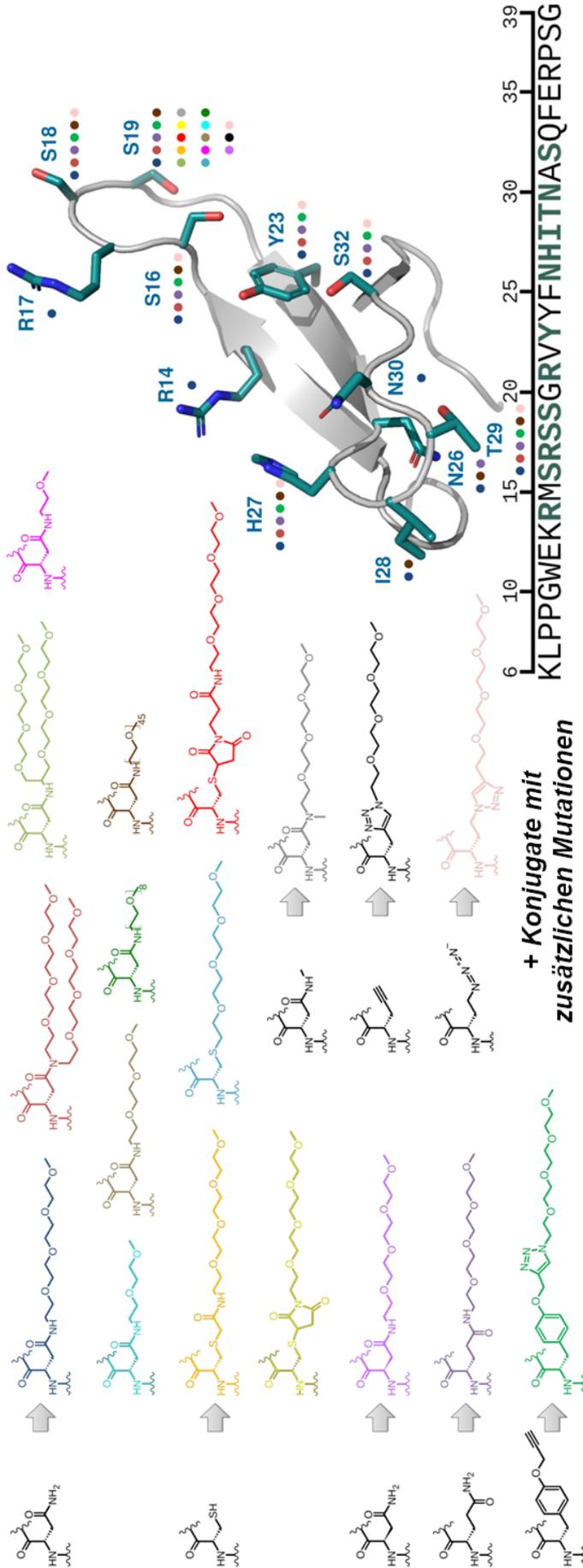


Abbildung 3.1: Untersuchte Konjugationsstellen der Pin 1 WW-Domäne mit bereits publizierten Stabilitätsdaten. Diese sind in der Struktur rechts benannt und türkis gefärbt. Die dazu aufgelisteten, farblichen Punkte symbolisieren die an der jeweiligen Position simulierten Varianten, deren Strukturen links in den selben Farben dargestellt sind. Die jeweils unkonjugierte Aminosäure ist ebenfalls illustriert. Zusätzlich zu den hier gezeigten Varianten des Wildtyps wurden eine Vielzahl an Konjugaten mit zusätzlichen Punktmutationen untersucht (siehe hierzu Tab. C.2).

ohne publizierte Stabilitätsdaten wurden in Tab. C.2 an entsprechender Stelle mit der Abkürzung *NA* (*Not available*) gekennzeichnet.

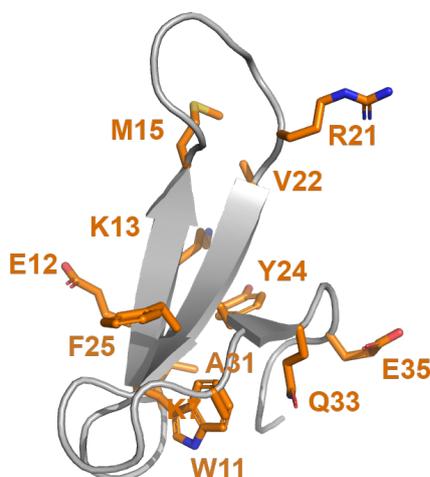


Abbildung 3.2: Zusätzlich simulierte, nicht von Price et al. charakterisierte Konjugationsstellen der WW-Domäne.

3.1.3 Ermittlung geeigneter Simulationsparameter

3.1.3.1 PyRED-Methode

Zur Durchführung klassischer MD-Simulationen wurden für proteinogene Aminosäuren die Parameter des Amberff14SB-Kraftfeldes [133] genutzt. Für die PBK jedoch mussten zusätzliche Elemente definiert werden, wobei zur Modellierung MOE 2018.01, 2019.01 und 2020.01 (Molecular Operating Environment) [274] verwendet wurde. Hierzu wurde, mit Ausnahme der Asn(PEG)₁-, Asn(PEG)₂-, Asn(PEG)₃-, Asn(PEG)₈- und Asn(PEG)₄₅-Strukturen (siehe hierfür Kap. 3.1.3.3), die konjugierte Aminosäure mitsamt der (PEG)₄-Kette als eine einzelne Aminosäure beschrieben. Da es sich um ein an das Protein kovalent gebundenes Strukturelement handelte, wurde im Falle der Asn(PEG)₄-Struktur (entspricht Kennzeichen 1 in Tab. C.1) zunächst der PyRED-Server verwendet [275–278]. Dieser eignet sich aufgrund der möglichen Festsetzung von Ladungen und anschließender Entfernung bestimmter Atome besonders, um Fragmente mit offenen Valenzen für größere Strukturen zu definieren (z. B. Zucker in glykosylierten Proteinen, Monomere in Polymeren, nichtproteinogene Aminosäuren in Peptiden). Die konjugierte Aminosäure wurde hierzu mit aminosäurerückgratähnlichen Endgruppen modelliert: eine Acetylgruppe am N-Terminus, sowie eine Aminomethylgruppe am C-Terminus (Abb. 3.3). Die entsprechende PDB-Struktur wurde dann auf Basis des Amber14:EHT-Kraftfeldes [133, 279] energieminiert (Gradient: 0.001 kcal/mol/Å) und mitsamt den nötigen Einstellungen auf den Server hochgeladen: Festsetzung der Gesamtladung der Termini auf Null (mithilfe der INTRA-MCC1-Einstellung), Zuweisung von Amber14ffSB-Atomtypen und Nutzung von Gaussian 16 Rev. B.01 [280] zur Bestimmung von RESP-Ladungen [143]. Die

vom Server erhaltenen Kraftfelddateien geben an, in welchen Fällen von den im Amberff14SB-Kraftfeld festgesetzten Parametern abweichende Werte festgelegt wurden, wobei sich diese zumeist minimal von den Standardwerten unterschieden (Tab. 3.1).

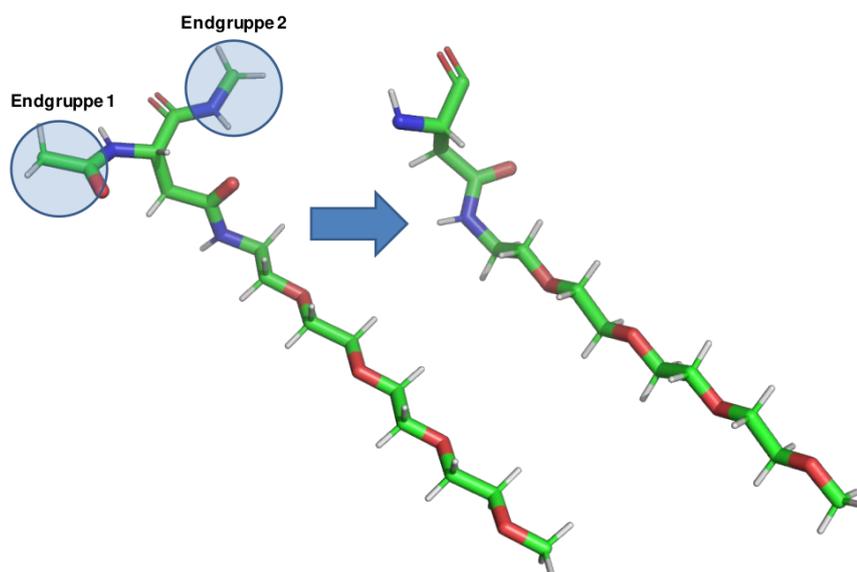


Abbildung 3.3: Beschreibung des PyRED-Verfahrens am Beispiel von Asn(PEG)_4 . Die Gesamtladung der terminalen Acetyl- und Aminomethylgruppen wurde auf Null festgesetzt. Nach Entfernung wurden geeignete RESP-Ladungen für das mittlere Fragment, mit einer Nettoladung von Null, erhalten.

Tabelle 3.1: Darstellung eines Auszugs aus der von PyRED generierten Parameterdatei für die Asn(PEG)_4 -Struktur. Abweichungen von den Amber14ffSB-Standardparametern sind in der rechten Spalte angegeben (rot markiert).

Bindung	k [kcal/mol/Å ²]	Distanz [Å]	Ursprung
C-CT	315	1.522	<i>adapted from parm10.dat 317.0</i>
C-N	490	1.335	<i>taken from parm10.dat</i>
C-O	570	1.229	<i>taken from parm10.dat</i>

3.1.3.2 antechamber-Methode

Da der PyRED-Server nach erfolgter Definition der Asn(PEG)_4 -Struktur für fast zwei Jahre aufgrund von Wartungsarbeiten nicht erreichbar war, wurde für alle weiteren Varianten ein ähnlicher Ansatz lokal durchgeführt. Hierzu wurden RESP-Ladungen für die vorbereitete PDB-Struktur der konjugierten Aminosäure mittels Gaussian 09 Rev. C.01 [281] ermittelt und geeignete Amberff14SB-Atomtypen über antechamber von AmberTools18 [282] zugewiesen. Im Anschluss wurde die Struktur in xleap geladen und die Endgruppen wurden entfernt. Die Ladungen des Proteinrückgrats wurden dann den üblichen Werten im Amber14ffSB-Kraftfeld angeglichen. Diese unterschieden sich zumeist nur minimal von den lokal mit Gaussian 09 Rev. C.01 erhaltenen

Werten. Die Gesamtladung wurde anschließend berechnet und ein Überschuss an positiver oder negativer Ladung wurde schließlich gleichmäßig auf die restlichen Atome der konjugierten Aminosäure verteilt. Da dieser Überschuss insgesamt in der Regel in einem Bereich von -0.2 bis $+0.2$ lag, wurden die zugewiesenen Ladungen einzelner Atome durch diesen Vorgang nur in einem äußerst geringen Maße modifiziert und nicht in ihrem Vorzeichen verändert (Abb. 3.4). Während hierbei lokal durch antechamber wie beim PyRED-Ansatz zunächst Amberff14SB-Atomtypen zugewiesen wurden, wurden bei dieser Methode jedoch anschließend mit diesem Kraftfeld kompatible GAFF2-Parameter [147] hinzugezogen, was in den erhaltenen Dateien einsehbar war (Tab. 3.2). Sogenannte Strafpunkte (*Penalty scores*) gaben dabei an, wie stark die ermittelten Parameter auf Extrapolationen bekannter Kraftfeldwerte beruhten [283].

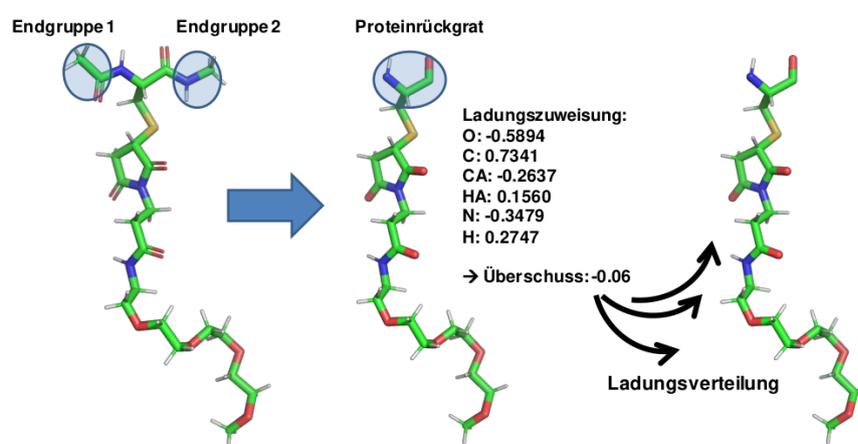


Abbildung 3.4: Illustration der alternativ durchgeführten, lokalen antechamber-Methode zur Parametrisierung der konjugierten Aminosäuren, am Beispiel einer Cys-Maleimid-(PEG)₄-Struktur (Kennzeichen 4 in Tab. C.1). Nach Entfernung der Endgruppen und Zuweisung der üblichen Ladungen des Proteinrückgrats wurde die Nettoladung berechnet und der Überschuss möglichst gleichmäßig auf die übrigen Atome verteilt.

Tabelle 3.2: Darstellung eines Auszugs aus der von antechamber generierten Parameterdatei für eine Cys-Maleimid-(PEG)₄-Struktur (Kennzeichen 4 aus Tab. C.1). Die zugrundeliegenden GAFF2-Parameter werden in der rechten Spalte angezeigt, zusammen mit entsprechenden Strafpunkten.

Bindung	k [kcal/mol/Å ²]	Distanz [Å]	Ursprung
C-CT	313.00	1.524	same as c-c3, penalty score= 0.0
C-N	427.60	1.379	same as c- n, penalty score= 0.0
C-O	637.70	1.218	same as c- o, penalty score= 0.0

Auch wenn für manche Parameter je nach verwendeter Methode leicht unterschiedliche Werte erhalten wurden, so ergab ein Vergleich von Simulationen der Asn(PEG)₄-Struktur mit den von PyRED zugewiesenen Ladungen und Atomtypen und jenen mit

diesem lokalen antechamber-Ansatz keine Unterschiede im Interaktionsprofil des Polymers, weshalb diese alternative Methode für alle weiteren konjugierten und nichtproteino-genen Aminosäuren verwendet wurde, solange der PyRED-Server nicht zur Verfügung stand. Später konnten die Varianten Asn(LPG)₅, Asn(PEtOx)₅, Asn(PMeOx)₅, Cys-Maleimid-(LPG)₅, Cys-Maleimid-(PEtOx)₅ und Cys-Maleimid-(PMeOx)₅ wieder analog zu Asn(PEG)₄ über den PyRED-Server parametrisiert werden.

3.1.3.3 Monomerbasierte Parametrisierung

Der Fokus dieser Arbeit lag zunächst auf den von Price et al. studierten PEG-Tetrameren, weshalb die Oligomere, wie bei Price et al. auch [100], als einzelne Fragmente parametrisiert wurden. Da jedoch im Zuge der Erweiterung des Datensatzes an Modellprotein-Konjugaten auch Asn(PEG)_x-Strukturen mit variablen Kettenlängen simuliert werden sollten, wurde für die Varianten Asn(PEG)₁, Asn(PEG)₂, Asn(PEG)₃, Asn(PEG)₈, Asn(PEG)₄₅ und alle größeren PEG-Konjugate aus Kap. 5 in analoger Weise zum lokalen antechamber-Ansatz ein PEG-Monomer (Fragment CH₂-CH₂-O, Abb. 3.5) als vielseitig einsetzbares Strukturelement parametrisiert. Entsprechend wurden hierbei zusätzliche PEG-Monomere als Endgruppen auf beiden Seiten während der Ladungsbestimmung genutzt. Zur Parametrisierung einer terminalen PEG-Methoxygruppe (Fragment CH₂-CH₂-O-CH₃) wurde analog mit einer Endgruppe auf einer einzelnen Seite verfahren.

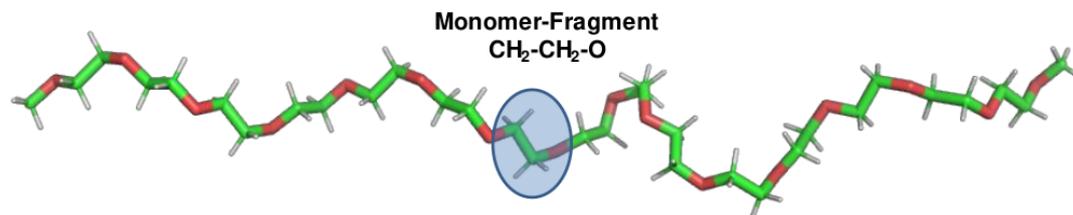


Abbildung 3.5: Darstellung der zur Parametrisierung eines einzelnen PEG-Monomers verwendeten Struktur. Die endständigen Monomere außerhalb des markierten Bereichs wurden entfernt und überschüssige Ladungen analog zum beschriebenen antechamber-Ansatz verteilt.

3.1.3.4 Problemstellungen

Für die von Price et al. charakterisierten Cys-Maleimid-Strukturen sind in den Referenzpublikationen keine stereochemischen Informationen angegeben, sodass davon ausgegangen werden kann, dass hierbei experimentell ein Racemat untersucht worden ist. Für die Simulationen wurde hierfür jeweils die (S)-Konfiguration modelliert.

Ein ähnliches Problem stellen die LPG-Strukturelemente in dieser Arbeit dar. Durch die Hydroxymethylseitenketten ergeben sich Stereozentren. Da die Taktizität dieses Polymers während der Synthese zumeist nicht kontrolliert wird [70], wurden in diesen Fällen alternierende (S)- und (R)-Konfigurationen modelliert (als Baustein der größeren Konjugate wurde ein Dimer mit einem (S)- und einem (R)-Monomer parametrisiert).

Schließlich stellte auch die unkonjugierte Aminosäure Azidohomoalanin ein Problem bei der Atomtypenzuweisung dar (siehe Kennzeichen § aus Tab. C.1). Der PyRED-Server konnte den endständigen Stickstoffatomen keinen Typ zuweisen (diese wurden vom Server mit N? gekennzeichnet), den endständigen Atomen der Azidgruppe wurde durch den lokalen Ansatz mittels antechamber ebenso der Atomtyp DU (*Dummy atom*) zugewiesen. Alternative Parameter aus älteren Publikationen auf Basis der Arbeiten zum Arzneistoff Azidothymidin, in denen neue Atomtypen definiert worden waren (um eine möglichst realistische Mischung aus sp- und sp²-Hybridisierungsanteilen der Stickstoffatome wiederzugeben), wurden im Anschluss getestet, führten jedoch durchweg zu instabilen Simulationen und wurden daher wieder verworfen [284, 285]. Schließlich wurde den terminalen Stickstoffatomen der GAFF2-Atomtyp n1 zugewiesen. Dies entspricht einer sp-Hybridisierung, wodurch die reale Verteilung der Elektronen möglicherweise nicht ideal abgebildet wird. Eine zu erwartende annähernd lineare Struktur der Azidgruppe in den Simulationen wurde dadurch allerdings beibehalten. Da die Parameter des Azidohomoalanins jedoch lediglich für einige wenige unkonjugierte Varianten des gesamten Datensatzes genutzt wurden, kann mit hoher Wahrscheinlichkeit ein entscheidender Einfluss der eventuell unzulänglichen Hybridisierung dieser endständigen Gruppe auf die berechneten Deskriptoren ausgeschlossen werden.

3.1.3.5 Vergleich der Simulationen mit Referenz

Price et al. untersuchten in einer ihrer Studien [100] die Interaktionsprofile zehn verschiedener simulierter WW-Konjugate, sowie die RDF-Werte (*Radial distribution function*, radiale Verteilungsfunktion) des Solvens um bestimmte Seitenkettenatome für vier WW-Konjugate. Daher wurden zum Vergleich analoge Analysen mit den in dieser Arbeit durchgeführten Simulationen der relevanten Asn(PEG)₄-Konjugate vorgenommen (Abb. 3.6, 3.7, 3.8 und 3.9).

Die Interaktionspartner der Polymere entsprechen in den meisten Fällen qualitativ jenen aus der Referenzpublikation; Unterschiede sind im Falle der Konjugate 27-1 und 29-1 zu erkennen. Diese Differenzen können einerseits darauf zurückzuführen sein, dass nicht bekannt ist, mit welchem Analysenprogramm die Wechselwirkungsenergien in der Referenzpublikation berechnet wurden (in der hier vorliegenden Arbeit wurde der *lie*-Befehl aus cpptraj [282] genutzt, welcher das Solvens lediglich implizit berücksichtigt). Vor allem aber nutzten Price et al. Mulliken-Ladungen [286] in Zusammenhang mit dem älteren Amberff99SB-ILDN-Kraftfeld [153, 287], was im Vergleich zur Nutzung von RESP-Ladungen [275] mit dem neueren Amberff14SB-Kraftfeld [133] mit hoher Wahrscheinlichkeit zu einer abweichenden Polymerdynamik führt.

Die RDF-Werte stimmen mit den Ergebnissen von Price et al. überein. Dies verdeutlicht, dass die generierten Parameter geeignet sind, um die von Price et al. beschriebenen Desolvatationsvorgänge um spezifische Aminosäuren herum wiederzugeben.

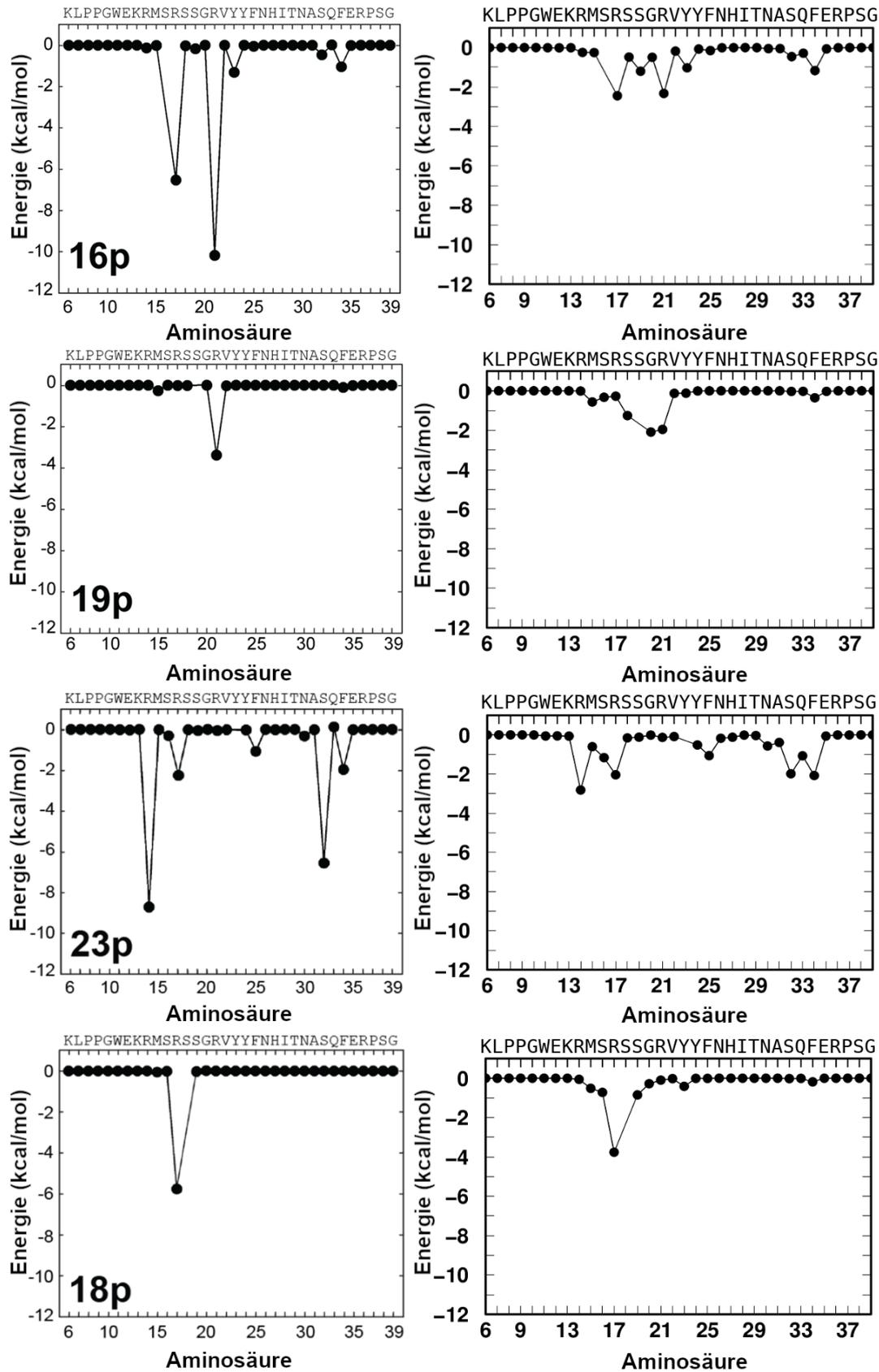


Abbildung 3.6: Vergleich der Interaktionsprofile von Price et al. Die Namen entsprechen jenen aus der Originalpublikation (vgl. Tab. C.2). Auf der linken Seite (Nachdruck mit Genehmigung aus [100], Copyright 2014 American Chemical Society) sind die Ergebnisse von Price et al. dargestellt, auf der rechten Seite analog jene aus dieser Arbeit.

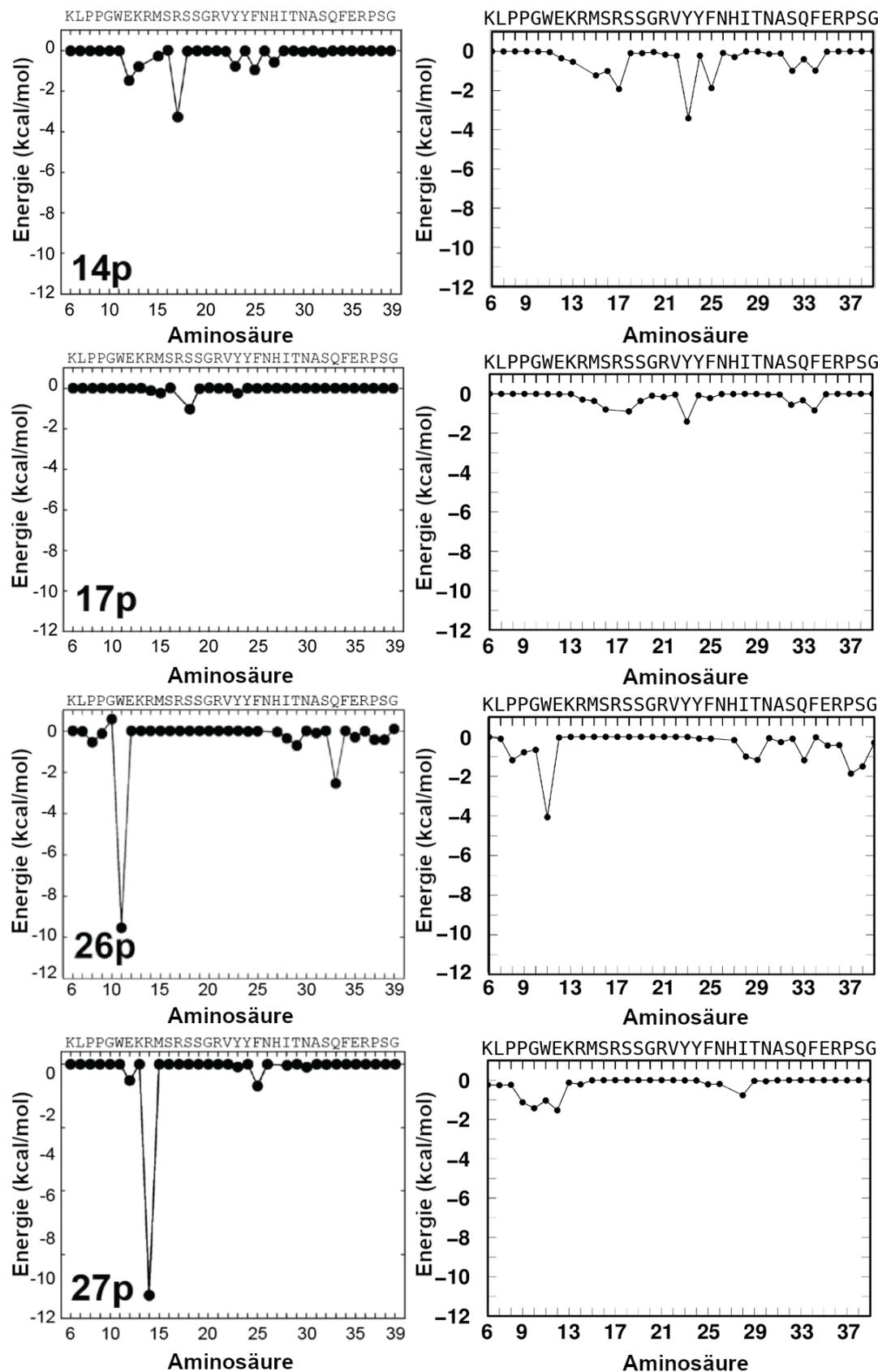


Abbildung 3.7: Vergleich der Interaktionsprofile von Price et al. (Fortführung zu Abb.3.6). Die Namen entsprechen jenen aus der Originalpublikation (vgl. Tab. C.2). Auf der linken Seite (Nachdruck mit Genehmigung aus [100], Copyright 2014 American Chemical Society) sind die Ergebnisse von Price et al. dargestellt, auf der rechten Seite analog jene aus dieser Arbeit.

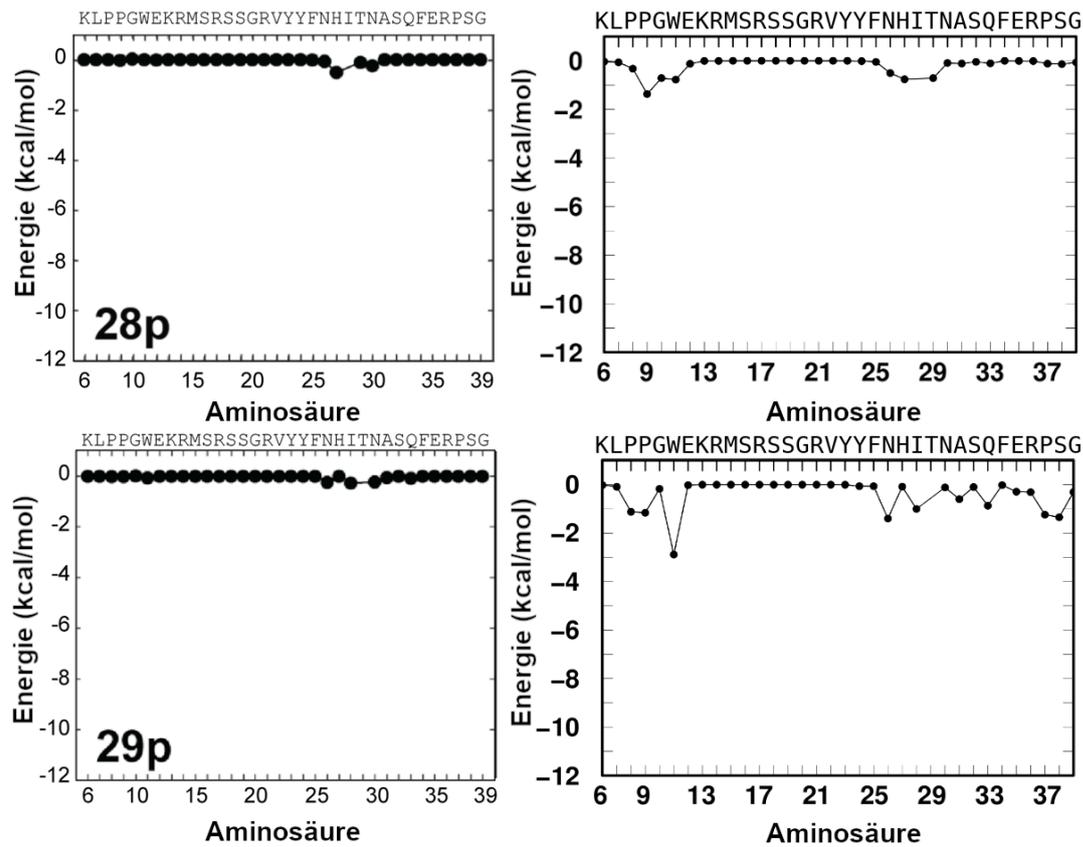


Abbildung 3.8: Vergleich der Interaktionsprofile von Price et al. (Fortführung zu Abb.3.6). Die Namen entsprechen jenen aus der Originalpublikation (vgl. Tab. C.2). Auf der linken Seite (Nachdruck mit Genehmigung aus [100], Copyright 2014 American Chemical Society) sind die Ergebnisse von Price et al. dargestellt, auf der rechten Seite analog jene aus dieser Arbeit.

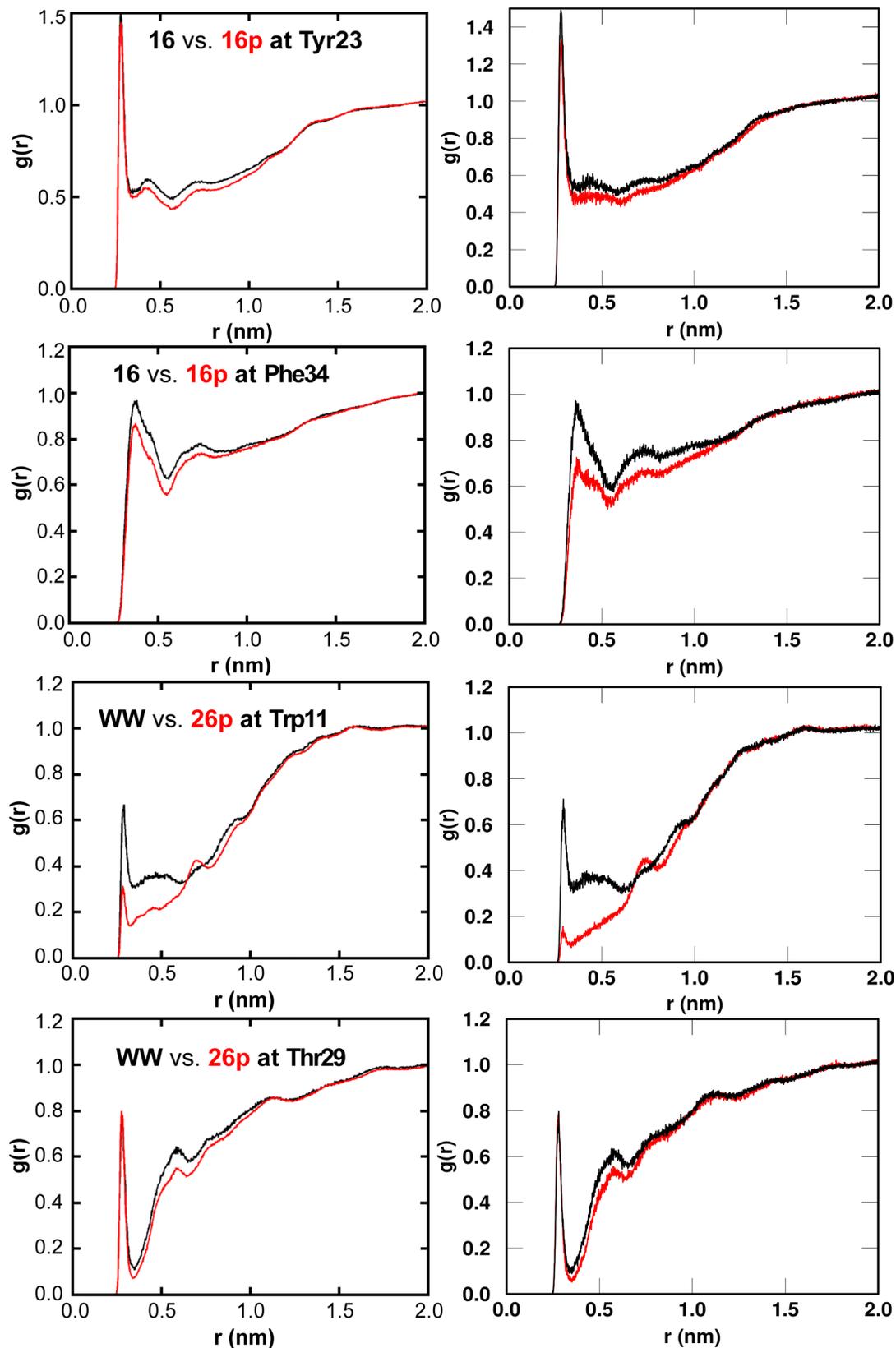


Abbildung 3.9: Vergleich der RDF-Profile um bestimmte Atome von Aminosäureseitenketten (OH für Tyr, C_z für Phe, $N_{\epsilon 1}$ für Trp, $O_{\gamma 1}$ für Thr). Auf der linken Seite sind die Ergebnisse von Price et al. (Nachdruck mit Genehmigung aus [100], Copyright 2014 American Chemical Society) dargestellt. Dabei zeigen die schwarzen Linien die RDF-Werte der unkonjugierten, die roten jene der konjugierten Varianten. Die Namen entsprechen jenen aus der Originalpublikation, WW steht hierbei für den Wildtyp. Auf der rechten Seite sind die Ergebnisse aus den in dieser Arbeit durchgeführten Simulationen analog dargestellt.

3.1.4 Modellierung der Strukturen

3.1.4.1 Aufbereitung der Peptidstrukturen

Die verwendeten Peptide wurden im Falle von Kristallstrukturen in MOE bei einem pH-Wert von 7 protoniert (Pin 1 WW-Domäne 1PIN, sowie die Doppelwendeln mit den PDB-Strukturen 2ZTA, 1COI und 5UXT), nachdem das enthaltene Kristallwasser und sonstige Liganden entfernt worden waren. In die Pin 1 WW-Domäne (Aminosäuren K6 bis G39) wurde vor der Protonierung eine W34F-Mutation eingefügt und das entsprechende Phenylalanin lokal energiminimiert. Diese Mutation wurde in allen von Price et al. charakterisierten Varianten dieses Peptids verwendet [99], da die Tryptophan-Variante, so die Erklärung von Price auf Anfrage, „eigenartige“ Signale während der DSF-Messungen verursachte.¹ Es wurde durch MOE keine Protonierung der Seitenkette von K6 durchgeführt, diese fungierte hierbei über ihr Stickstoffatom als HB-Akzeptor für die Hydroxylgruppe von S38. Da es sich jedoch um die Termini handelt, welche in ersten Simulationen ein sehr flexibles Verhalten zeigten und keine dauerhafte Koordination von K6 zu S38 zu beobachten war, wurde diese Seitenkette nachträglich manuell protoniert.

Der Protonierungszustand der verwendeten NMR-Strukturen (Her2 Affibody 2KZJ und Src SH3-Domäne 1SRL) wurde beibehalten, da eine Protonierung in MOE bei pH 7 (2KZJ wurde bei pH 6 aufgenommen) zu keinen Veränderungen führte. Die Termini aller Strukturen wurden bei der WW-Domäne, dem Her2-Affibody und der Src SH3-Domäne, analog zu Price et al., nicht mit zusätzlichen Endgruppen versehen, sondern mit geladenem Zustand modelliert. Im Falle der Doppelwendel aus der Kristallstruktur 1COI wurden die Termini beibehalten: Es handelt sich um eine C-terminale Acetylgruppe und ein N-terminales Amid. Um Sequenzen zu modellieren, die den von Price et al. charakterisierten Varianten entsprechen, mussten im Falle der Kristallstrukturen 2ZTA und 5UXT, die lediglich eine Acetylgruppe enthielten, die N-terminalen Amide nachträglich eingefügt werden. Ebenso mussten dazu in der 1COI-Struktur die C-terminalen Aminosäuren H28 und G29 entfernt und durch ein lokal energiminimiertes Tyrosin ersetzt werden. Möglicherweise wurde dieses Tyrosin von Price et al. zur Durchführung von DSF-Messungen eingebaut.

3.1.4.2 Einbau modifizierter Aminosäuren

Die konjugierten Aminosäuren wurden mittels MOE an den jeweiligen Mutationsstellen eingefügt. Diese wurde lokal mit dem Amber14:EHT-Kraftfeld energiminimiert.

¹Erklärung von Prof. Joshua L. Price in einer E-Mail-Korrespondenz: „*Years ago, Prof. Jeff Kelly's lab at Scripps observed some unusual behavior in fluorescence-monitored temperature jump kinetic experiments on the WW variant with Trp at position 34, possibly due to the different environments of the two Trp residues during the unfolding process. In any case, they found that replacing Trp34 with Phe eliminated this unusual behavior and resulted in a variant that could be readily analyzed via kinetic and thermodynamic experiments based on a two-state folding model.*“

Anschließend wurde zur Ermittlung einer geeigneten Startstruktur des Polymers jeweils eine stochastische Konformationssuche durchgeführt, wobei die restlichen Aminosäuren fixiert wurden. Es wurden folgende Einstellungen verwendet und die stabilste Konformation nach erfolgter Suche gespeichert: Ablehnungslimit = 100, Iterationslimit = 2000, Gradient = 0.001 kcal/mol/Å und Konformationslimit = 1000. Im Falle der unPEGylierten Varianten wurde die betreffende Aminosäure mutiert und lokal energieminiert.

3.1.4.3 Solvation der Strukturen

Zur Simulation wurden die Peptide mit dem Programm `sander` aus dem AMBER18-Programmpaket [282] initial energieminiert und anschließend mit TIP3P-Wasser [288] durch `tleap` von AmberTools18 [282] solvatisiert. Um die Anzahl an Wassermolekülen bei diesem Schritt möglichst gering zu halten, wurden die Hauptachsen der Strukturen zuvor mit den Achsen des Koordinatensystems aligniert, sodass die Proteine in der Simulationsbox optimal ausgerichtet waren. Es wurde jeweils eine rechteckige Solvensbox mit einer minimalen Distanz von 10 Å der Peptidatome zu den Grenzflächen der Box generiert. Im Falle der länglichen Doppelwendeln wurden jedoch kubische Boxen zur Solvation genutzt, da diese Peptide nicht als annähernd globulär angesehen werden können und sie dadurch während der Simulationen und nach Alignierung der Hauptachsen über Rotationen evtl. mit den periodischen Nachbarn wechselwirken konnten. Natrium- oder Chloridionen wurden nach erfolgter Solvation zum Ladungsausgleich des Systems hinzugefügt.

3.1.5 Simulationsprotokoll

3.1.5.1 Equilibrierung und erster Produktionslauf

Die solvatisierten Peptide wurden im Anschluss mit NAMD 2.12 oder 2.13 simuliert [289]. Ein Zeitschritt von 2 fs wurde gewählt. Es wurden periodische Randbedingungen unter Verwendung des RATTLE-Algorithmus [138] genutzt. Zur Abbildung von Interaktionen auf längere Distanz wurde die Particle-Mesh-Ewald-Methode [140] mit einem Cut-off-Wert von 1.2 nm angewandt [140]. Zunächst wurden 10000 Minimierungsschritte durchgeführt, bevor das System in einer ersten Equilibrierung von 100 auf 300 K über ein Zeitintervall von 500 ps langsam aufgeheizt wurde. Eine Kraftkonstante von 0.5 kcal/mol/Å² wurde verwendet, um die Peptidatome zunächst in ihrer Beweglichkeit zu begrenzen. Auf diese Weise konnten sich die durch `tleap` artifiziell gesetzten Wassermoleküle zuerst reorganisieren. Die Kraftkonstante wurde über weitere 1.6 ns sukzessive bis zum Wert Null verkleinert. Eine konstante Temperatur von 300 K im NVT-Ensemble wurde durch Nutzung eines Langevin-Thermostats erreicht. Nach vollständiger Annullierung der Kraftkonstante wurde für eine weitere Nanosekunde in diesem Ensemble simuliert, bevor eine zweite, 1 ns lange Equilibrierung im NPT-Ensemble durchgeführt wurde, um die Dichte des Systems zu equilibrieren.

Dieses Ensemble wurde durch Festlegung des Drucks mit dem Nosé-Hoover-Langevin-Piston-Barostat auf 1.01325 bar erreicht. Nach dieser Equilibrierung wurde ein erster Produktionslauf im NPT-Ensemble für 100 ns vorgenommen, wobei Momentaufnahmen für jede Pikosekunde herausgeschrieben wurden.

3.1.5.2 Fortführung des Produktionslaufs

Der oben beschriebene Produktionslauf wurde für weitere 100 ns fortgeführt, sofern bestimmte Konvergenzkriterien nicht erfüllt wurden. Hierzu wurden die drei in Kap. 1.2.3.1 beschriebenen SCC angewandt. Folgende Limitierungen wurden gesetzt: Gemäß dem SDVG-Kriterium durfte kein neuer Cluster in den letzten 25 ns der Simulation gefunden werden. Hierzu diente eine hierarchische RMSD-Clusteranalyse (Average-Linkage) mittels cpptraj [282]. Mit $\epsilon = 1$ wurde festgelegt, dass das Clustering bis zu einer Minimaldistanz von 1 Å zwischen den gefundenen Clustern durchgeführt wurde. Die Clusterkonfigurationsentropie durfte sich in den letzten 25 ns höchstens um 5 % ändern. Schließlich musste das BBCOM/BCOM-Verhältnis, welches idealerweise einen Wert von 1 erreichen sollte, bei ansteigenden Blocklängen bis zur Hälfte der Simulationszeit einen Wert von 1.20 erreichen. Dieses Verhältnis wurde mithilfe des LOOS-Pakets berechnet [290, 291]. Die drei beschriebenen SCC wurden einerseits für die C_α -Atome der Peptide und in einem zweiten Durchlauf für die Sauerstoffatome der Polymerketten evaluiert. Exemplarische Plots zur Veranschaulichung aller drei Kriterien sind in Abb. 3.10 gezeigt. Zwei der drei gewählten Kriterien mussten erfüllt werden (sowohl für die C_α - als auch für die Polymer-Sauerstoffatome), ansonsten wurde der Produktionslauf für 100 ns fortgeführt und alle Kriterien wurden erneut untersucht. Waren die Bedingungen dann immer noch nicht erfüllt, wurden wiederum 100 ns simuliert. Im Falle der unPEGylierten Varianten dienten lediglich die C_α -Atome als Test. Letztendlich wurden die meisten WW-Konjugate für 200 oder 300 ns simuliert. Die größeren Her2-Konjugate benötigten deutlich mehr Zeit (zumeist 600 bis 800 ns), um die geforderten Kriterien zu erfüllen. Asn(PEG)₄₅-Varianten wurden jeweils für 800 ns simuliert, zeigten aber auch danach keine ausreichende Konvergenz bzgl. der Polymer-Sauerstoffatome. Diese wurden daher für die QSPR-Modellgenerierungen als nicht konvergiert eingestuft (siehe Kap. 3.1.6.2 und 3.1.6.3).

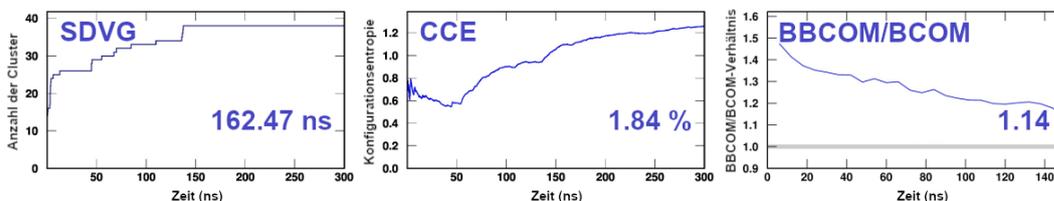


Abbildung 3.10: Exemplarische Darstellung des Verlaufs der drei SCC-Kriterien einer Simulation. Die violette Zahl im SDVG-Plot (links) gibt die Zeit an, ab der kein weiterer Cluster gefunden wurde. Im CCE-Plot gibt die Zahl die Veränderung des Werts in den letzten 25 ns an. Im rechten BBCOM/BCOM-Plot stellt die Zahl das Verhältnis bei der größten Blocklänge dar. Letztere erreicht als Maximum die Hälfte der Simulationszeit, da ansonsten nicht mehr als ein einziger Block generiert werden kann.

3.1.6 QSPR-Vorhersagemodelle

Im Verlauf der Arbeit wurde der Datensatz an Simulationen und daraus berechneter Deskriptoren sukzessive erweitert. Ein erster erweiterter Modellierungsansatz wurde auf Basis des R-Pakets *caret* (*Classification And REgression Training* [211, 292]) entwickelt. Dieses R-Paket von Max Kuhn (Entwickler von RStudio, ehemaliger Biostatistiker bei Pfizer [293]) vereint verschiedenste Modellierungspakete aus der Data Science-Community von R mit einer einheitlichen Syntax und zusätzlichen Möglichkeiten zur Datenaufbereitung und Evaluation. Es gehört neben *scikit-learn* und *TensorFlow* zu einem der beliebtesten Programme für Prozesse des maschinellen Lernens. In diesem Ansatz, der in Kap. 3.1.6.2 im Detail beschrieben werden soll, wurde der Fokus v. a. darauf gelegt, eine Vielzahl der in diesem Paket verfügbaren Modellarten des maschinellen Lernens zu testen. Da die simulierten Konjugate jedoch einen für Machine Learning-Verfahren relativ kleinen Datensatz darstellten, wurde im Anschluss ein hiervon abgewandelter zweiter Ansatz entwickelt (Kap. 3.1.6.3), der sich auf die Generierung von multilinearen Regressions- und binomialen, logistischen Klassifikationsmodellen beschränkte.

3.1.6.1 Deskriptoren

Die Trajektorien der untersuchten Modellproteine wurden genutzt, um daraus eine Vielzahl an Deskriptoren zu berechnen. Dies wurde für die PBK und deren unkonjugierten Varianten durchgeführt (im Falle des Winkels von Price et al. ebenso für die mutierten Wildtyp-Varianten ohne Konjugationsstelle). Hierzu wurde hauptsächlich *cpptraj* von *AmberTools18* [282] genutzt, aber auch das *mdanalysis*-Modul (Version 1.0.0) und das *CNA*-Programm (Version 2.0). Der theoretische Hintergrund zu diesen Deskriptoren wird in Anhang B beleuchtet. Eine Gesamtübersicht aller kalkulierten Variablen ist in Tab. D.1 gezeigt. Im Falle der *CNA*-Parameter wurden Momentaufnahmen ohne Wassermoleküle und Natrium- bzw. Chloridionen aus den Trajektorien für jede Nanosekunde als PDB-Datei herausgeschrieben. Die Deskriptoren wurden anschließend in zwei unterschiedlichen Ansätzen (siehe Kap. 3.1.6.2 und 3.1.6.3) verwendet, um QSPR-Vorhersagemodelle zur PEG-basierten Stabilisierung zu generieren. Für den ersten Ansatz (*vide infra*) wurden, wo möglich, Medianwerte berechnet, für den zweiten Mittelwerte.

3.1.6.2 Modellgenerierung: Ansatz 1

Für diesen Ansatz wurden die bis zu diesem Zeitpunkt simulierten 85 WW-Konjugate verwendet (siehe Tab. C.2 zur Auswahl). Eine zusammenfassende Übersicht über den Ablauf ist in Abb. 3.11 gegeben, die Resultate werden in Kap. 4.1.1 aufgeführt und diskutiert.

3.1.6.2.1 Aufbereitung der Daten

Filterung der Konjugate Zunächst wurden sieben Konjugate, die einen experimentellen Messfehler von ≥ 0.10 kcal/mol aufwiesen, aus der Modellgenerierung ausgeschlossen: 14-1, 23-1, 32(Y23F)-1, 19(S16F:R21T)-1, 23-2, 19(R21A)-1 und 19(R21L)-1. Im Falle von 14-1 und 23-1 handelte es sich um Konjugate, die von Price et al. ebenso aus weiteren Analysen ausgeschlossen wurden, da ihre CD-Spektren deutliche Umstrukturierungen des Proteins nicht ausschließen konnten [100]. Die sieben Asn(PEG)₄₅-Konjugate wiesen keine strukturelle Konvergenz der Trajektorien auf (siehe Kap. 3.1.5.2) und wurden daher auch entfernt.

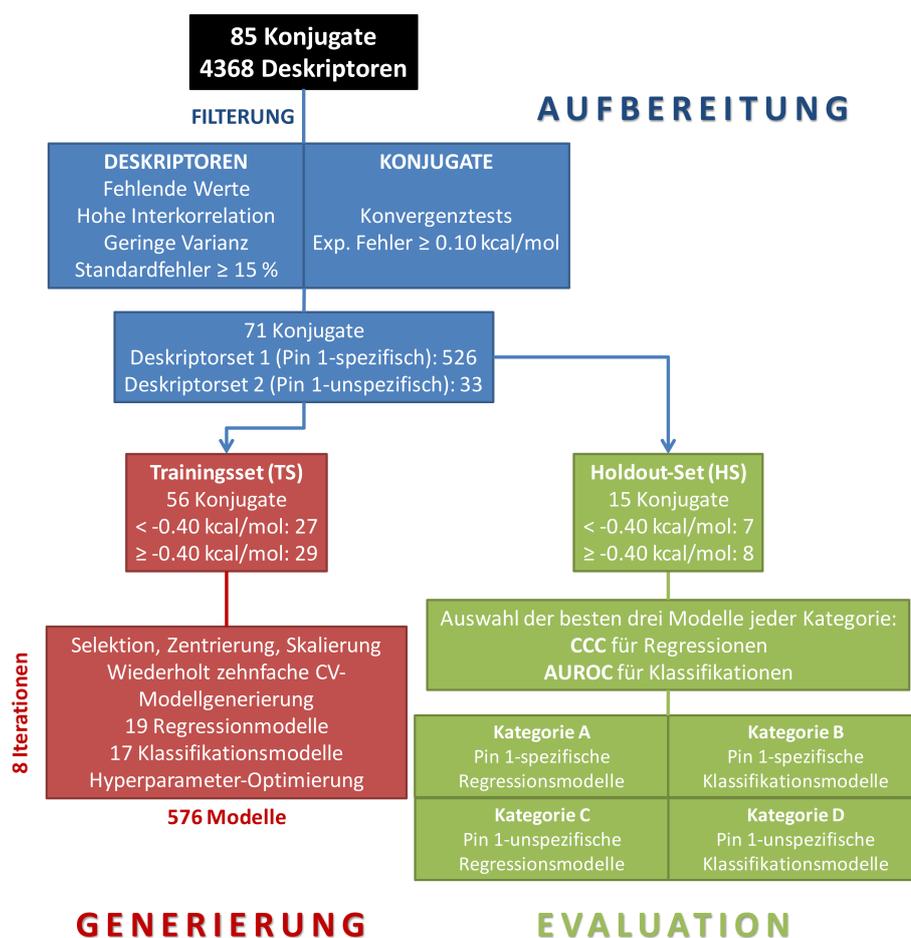


Abbildung 3.11: Übersicht über den ersten Modellgenerierungsansatz. Die Schritte wurden eingeteilt in eine Datenaufbereitungs-, eine Modellgenerierungs- und eine Evaluationsphase.

Filterung der Deskriptoren 4368 Deskriptoren wurden berechnet (Tab. D.1), von denen jene mit fehlenden Werten für einzelne Konjugate und solche, die identisch oder sehr ähnlich für den gesamten Datensatz waren, entfernt wurden. Ebenso wurden Variablen mit hoher Interkorrelation exkludiert (Limit: 0.85). Dabei wurden diejenigen Deskriptoren herausgefiltert, die eine durchschnittliche Korrelation von über 0.85 zu jedem anderen Deskriptor aufwiesen (*mean absolute correlation*), wobei diese Korrelationen nach jeder Filterung neu evaluiert wurden. Die Samplingqualität der

einzelnen Deskriptoren wurde als unzureichend eingestuft und führte damit zu einer Filterung, sofern der Standardfehler des Mittelwerts $\geq 15\%$ des entsprechenden Deskriptorwerts für mehr als 10 % der Konjugate im Datensatz zeigte. Die Berechnung des Standardfehlers wurde mithilfe des R-Paketes `coda` [174] durchgeführt, welches die effektive Stichprobengröße unter Berücksichtigung der Autokorrelation aufeinanderfolgender Momentaufnahmen berechnete (siehe Glg. 1.20). Weitere Prüfungen auf Stationarität (Heidelberger-Welch- und Geweke-Tests gemäß Kap. 1.2.3.2) wurden anfänglich ebenfalls als Filter verwendet, entfernten jedoch einen relativ großen Anteil der berechneten Variablen und wurden daher als zu stringent angesehen und aus dem Ansatz wieder herausgenommen.

Finale Datensätze Die genannten Filterschritte führten zu zwei Deskriptorsets für 71 WW-Konjugate. Ein Deskriptorset enthielt die 526 selektierten Deskriptoren, ein davon abgeleitetes, reduziertes Set lediglich die 33 Pin 1 WW-unspezifischen. Letztere dienen zur Generierung von Modellen, welche prinzipiell auf andere Proteine angewandt werden können. Hierfür wurden lediglich jene Deskriptoren ausgewählt, die die Eigenschaften der Polymerketten oder das Verhältnis zwischen der PEGylierten und der unPEGylierten Struktur beschreiben (und keine Eigenschaften einzelner, spezifischer Aminosäuren oder Strukturbereiche). Diese 33 Deskriptoren sind in Anhang D.2.4 aufgelistet. Der aufbereitete Datensatz wurde anschließend in einen Trainingsdatensatz (TS) von 56 Konjugaten (79 %) und einen zufällig ausgewählten, externen Validierungsdatensatz (Holdout-Set, HS) mit den restlichen 15 Varianten aufgeteilt. Die Verteilung der Faltungsenthalpien beider Datensätze mit Auflistung der HS-Konjugate ist in Abb. 3.12 gezeigt.

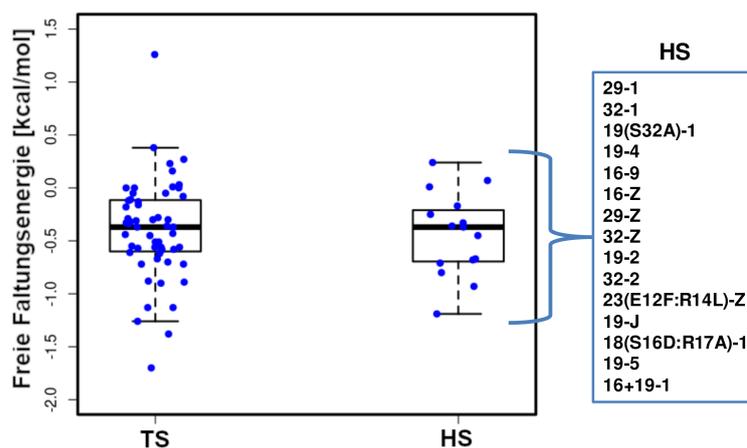


Abbildung 3.12: Verteilung der freien Faltungsenthalpien im TS und im HS des ersten Modellgenerierungsansatzes (Median-Wert jeweils -0.37 kcal/mol). Rechts daneben sind alle 15 Konjugate aufgelistet, die dem HS angehören.

3.1.6.2.2 Modellgenerierung

Ausgewählte Modellarten Mithilfe des Trainingsdatensatzes wurden anschließend 19 verschiedene Regressions- und 17 unterschiedliche Klassifikationsmodelle erstellt (siehe Tab. D.2). Für letztere wurden die Konjugate in zwei Klassen eingeteilt, je nachdem, ob die Differenz der freien Faltungsenergie unter -0.40 kcal/mol lag oder nicht. Diese Grenze wurde gewählt, da der Median im gesamten Datensatz einen Wert von -0.37 kcal/mol betrug. Im Validierungsset fanden sich sieben Konjugate aus der Klasse 1 (< -0.40 kcal/mol) und acht aus der Klasse 2 (> -0.40 kcal/mol).

Deskriptor-Selektionsmethoden Drei Selektierungsmethoden zur Auswahl bedeutender Deskriptoren aus dem Trainingsdatensatz wurden angewandt: (1) Rekursive Feature-Eliminierung (RFE) [294], (2) simuliertes Annealing [295], sowie (3) genetische Algorithmen [296].

Bei der RFE handelt es sich um eine Backward-Eliminationsmethode, bei der zunächst ein Modell (eingestellt wurde hierbei die Anwendung von kreuzvalidierten Random Forest-Modellen und Entscheidungsbäumen mit Bagging-Methode) mit allen Deskriptoren erstellt wird und von diesem jener Deskriptor mit der geringsten Bedeutung für das Modell entfernt wird. Anschließend wird mit dem verbliebenen Datensatz ein neues Modell generiert. Für jede mögliche Anzahl an Deskriptoren (in Ansatz 1 eingestelltes Maximum: alle, Minimum: 2) wird die Auswahl zur darauffolgenden Modellgenerierung herausgeschrieben.

Beim simulierten Annealing werden sukzessive zufällig generierte Deskriptorsubsets generiert. Verbessert sich die Leistung des Modells gegenüber der vorherigen randomisierten Auswahl, so wird das neue Deskriptorset beibehalten. Mit einer gewissen Wahrscheinlichkeit, die mit Zunahme der Iterationen allmählich absinkt, können dabei ebenso Modelle mit etwas schlechterer Leistung als das aktuell beste Modell ausgewählt werden.

Genetische Algorithmen stellen ebenso wie das simulierte Annealing zufallsbasierte Methoden dar. Hierbei wird die Deskriptorauswahl durch das Generieren von Tochtergenerationen und Mutationen, ausgehend von zufällig erstellten Chromosomen, durchgeführt. Chromosomen stellen dabei Binärvektoren dar, die für jedes Gen (die Deskriptoren) die Information enthalten, ob dieses für die Modellgenerierung genutzt wird. Jene Chromosomen mit der besten Leistung werden an zufälligen Positionen geschnitten und miteinander vermischt, um die Tochtergenerationen zu bilden [211].

Weitere Datentransformationen Die Deskriptoren des Datensatzes wurden zur Modellgenerierung ebenfalls zentriert (Subtraktion des jeweiligen Mittelwerts) und anschließend skaliert (Division durch die Standardabweichung).

Generierung der Modelle Danach wurden Modelle mit wiederholt zehnfacher Kreuzvalidierung und Testung unterschiedlicher Hyperparameter-Werte erstellt (entsprechend den in caret standardmäßig eingestellten Intervallen). Der gesamte Modellbildungsprozess wurde mit verschiedenen Einstellungen der oben genannten Selektionsmethoden wiederholt, um verschiedene Deskriptorsubsets zu testen, und in einem der insgesamt acht Durchläufe auch ohne jegliche Selektion durchgeführt. Dies führte zu insgesamt 576 verschiedenen Modellen, deren Leistung anschließend mithilfe des Validierungsdatensatzes überprüft wurde.

3.1.6.2.3 Evaluation

Modelleistung Die drei besten Modelle aus vier verschiedenen Kategorien (A bis D) wurden anschließend für eine nähere Evaluation ausgewählt, basierend auf dem Konkordanzkorrelationskoeffizienten für Regressions- und dem AUROC-Wert für Klassifikationsmodelle in Bezug auf das HS. Kategorie A entsprach hierbei WW-spezifischen Regressions-, Kategorie B WW-spezifischen Klassifikations-, Kategorie C WW-unspezifischen Regressions- und Kategorie D WW-unspezifischen Klassifikationsmodellen. Die ausgewählten Modelle werden in Kap. 4.1.1 evaluiert und interpretiert.

Applikationsdomäne Mit dem ADAN-R-Paket (*Applicability domain analysis*) [297] wurden die Applikationsdomänen (AD) der Regressionsmodelle über sechs verschiedene Kriterien evaluiert. Über eine Dimensionsreduktion mittels der PLS-Methode wurden die Modelldeskriptoren des Trainingsdatensatzes hierbei in einen Raum der entsprechenden latenten Variablen gesetzt. In diesem Raum wurden dann ebenfalls die Konjugate des Validierungsdatensatzes lokalisiert und deren (1) euklidische Distanz zum Zentroid, sowie (2) zum nächstgelegenen Punkt des Trainingsdatensatzes evaluiert. Ebenso wurden Ausreißer des PLS-Modells über einen auf den X-Residuen basierenden (3) Distanz-Score detektiert. Die Vorhersage des untersuchten Testkonjugats für die Differenz der freien Faltungsenergie $\Delta\Delta G_f^0$ wurde weiterhin (4) mit dem Durchschnittswert und (5) dem nächstgelegenen Wert im Trainingsdatensatz verglichen. Schließlich wurde ebenso (6) die Leistung der Prädiktion für die fünf nächstgelegenen Konjugate im PLS-Raum beurteilt. Basierend auf diesen sechs Kriterien wurde evaluiert, ob sich das Konjugat innerhalb der Applikationsdomäne befand (max. 1 Kriterium missachtet), mit einem erhöhten Vorhersagefehler zu rechnen war (2 - 3 Kriterien nicht erfüllt) oder sich außerhalb der AD des Modells befand (4 - 6 Kriterien nicht erfüllt). Ein Kriterium gilt als nicht erfüllt wenn der hierfür ermittelte Wert über dem 95%-Perzentil des Trainingsdatensatzes liegt.

3.1.6.3 Modellgenerierung: Ansatz 2

Ein modifizierter Ansatz wurde entwickelt, der sich auf Regressions- und Klassifikationsmodelle beschränkt, welche auch für kleinere QSPR-Datensätze geeignet sein sollten: multivariate lineare Regressionen und binomiale, logistische Klassifikationen

(Tab. D.5). Dieser Ansatz wurde an das Prinzip der wiederholt doppelten Kreuzvalidierung (*Repeated double cross validation*, rdCV) angelehnt. Dieses wurde von Filzmoser et al. [298] für PLS-Modelle konzipiert und beruht darauf, dass die optimale Anzahl an latenten Variablen in einer inneren For-Schleife determiniert wird, die prädiktive Leistung in einer äußeren. Dieser Prozess wird für verschiedene Kalibrierungsdatensätze wiederholt, um ultimativ eine Verteilung der zur erwartenden Modelleistung zu erhalten [298, 299].

3.1.6.3.1 Ausgeschlossene Konjugate

5UXT-Strukturen (siehe Kap. 1.5.1.4) wurden generell von der Modellgenerierung ausgeschlossen, da diese über keine Aminosäuren verfügen, mit denen der von Price et al. postulierte Winkel berechenbar gewesen wäre. Ebenso wurden Daten zu vier weiteren Konjugaten (26(W11F)-1, 26(W11 ν)-1, 26(W11Y)-1 und 26-9) erst nach Erstellung der Modelle veröffentlicht [104], sodass diese auch nicht berücksichtigt wurden.

3.1.6.3.2 Definition unterschiedlicher Datensätze

Im Gegensatz zu Ansatz 1 wurden neun verschiedene Datensätze unterschiedlicher Größe untersucht (Tab. 3.3). Dies sollte dazu dienen, die Leistung von Modellen zu beurteilen, die jeweils nur bestimmte Arten von Proteinen und / oder Linkertypen berücksichtigten. Ausgehend vom in Ansatz 1 gewählten Datensatz (dieser entspricht Datensatz 1 aus diesem zweiten Ansatz, nur ohne 1SRL.20-1) wurde ein erweiterter Datensatz 2 mit zu diesem Zeitpunkt neu simulierten WW-Konjugaten definiert. Datensatz 3 schließt zusätzlich die Konjugate der Doppelwendeln (PDB-Strukturen 2ZTA und 1COI) ein. Datensatz 4 beschränkt sich dagegen allein auf jene 1PIN/1SRL/2ZTA/1COI-Konjugate mit Asparagin-Linker. Der noch weiter reduzierte Datensatz 5 beinhaltet lediglich die WW-Konjugate aus Datensatz 4. Da in diesem Ansatz 2 weiterhin ein clusterbasiertes Verfahren (*vide infra*) zur Definition der HS angewandt wurde, wurde dieses für den ursprünglichen Datensatz 1 auch genutzt, was zu Datensatz 6 führte (es handelt sich also um Datensatz 1 mit einem veränderten HS). Die weiteren Datensätze 7 und 8 entsprechen Datensätzen 2 und 1 ohne 1SRL.20-1, mit der Ausnahme, dass hierbei auch alle Pin 1-spezifischen Deskriptoren berücksichtigt wurden. Datensatz 9 entspricht Datensatz 6 mit einem leicht modifizierten Clustering-Ansatz. Eine Liste Pin 1-unspezifischer Deskriptoren in Ansatz 2 ist in Anhang D.3.1 gezeigt.

Tabelle 3.3: Auflistung der getesteten Datensätze im zweiten Modellgenerierungsansatz mit der jeweiligen Anzahl an verwendeten Konjugaten nach Filterung (TS + HS).

Nr.	Beschreibung	Konjugate	HS
1	Datensatz ohne Doppelwendeln, Konjugate mit Kennzeichen § und Cyclohexylalanin-Mutationen. Das HS wurde hier noch ohne Clustering-Ansatz zufällig ausgewählt und eine exp. Fehlergrenze von 0.10 kcal/mol wurde genutzt.	61 + 11	17-1, 27-1, 16(S32A)-1, 27-9, 16-Z, 19-2, 23(E12L:R14L)-Z, 23(E12L:R14A)-Z, 19(S16F)-1, 16+26-1, 1SRL.20-1
2	Alle Konjugate mit β -Faltblattstrukturen.	78 + 11	26-1, 29(S32A)-1, 19-4, 23-9, 29-Z, 16-2, 23(E12A:R14F)-Z, 19(S16F)-1, 23(E12A:R14CHA)-Z, 23-§, 1SRL.20-1
3	Alle Konjugate.	100 + 14	27-1, 32-9, 16-Z, 27-2, 23(E12L:R14A)-Z, 19-U, 16+26-1, 23-1, 19(R21A)-1, 23(E12CHA:R14CHA)-Z, 23-§, 1SRL.20-1, 2ZTA.10-1, 2ZTA.18(R25A)-1
4	Nur Asn-Konjugate.	52 + 7	18(S16D:R17A)-1, 18(R17A)-1, 23-1, 19(R21A)-1, 1SRL.20-1, 2ZTA.7-1, 2ZTA.10-1
5	Nur Asn-Konjugate der WW-Domäne.	25 + 4	16-1, 27-1, 23-1, 19(R21A)-1
6	Analog zu Datensatz 1, aber mit beschriebenem Clustering-Ansatz für das HS.	61 + 11	18-1, 28-1, 30-1, 23-9, 19-3, 16-Z, 32-Z, 19-2, 23(E12L:R14A)-Z, 19-5, 1SRL.20-1
7	Pin 1-spezifische Deskriptoren mit allen Konjugaten der WW-Domäne.	78 + 10	27-1, 19(Y23F)-1, 27-9, 19-3, 19-2, 23(E12F:R14A)-Z, 23(E12A)-Z, 16+26-1, 32(Y23F)-1, 23-§
8	Analog zu Datensatz 7 mit der Auswahl von Datensatz 1.	61 + 10	17-1, 27-1, 16(S32A)-1, 27-9, 16-Z, 19-2, 23(E12L:R14L)-Z, 23(E12L:R14A)-Z, 19(S16F)-1, 16+26-1
9	Analog zu Datensatz 6, wobei die Anzahl an Konjugaten der „extremen“ Cluster lediglich mit 2 multipliziert wurde.	61 + 11	16-1, 27-1, 19-8, 19-9, 19-3, 16-Z, 29-Z, 18-2, 18(R17A:R21A)-1, 23(R14A)-Z, 1SRL.20-1

3.1.6.3.3 Clustering der Daten

Für die Datensätze 2 bis 9 wurde ein klar definierter Ablauf zur randomisierten Auswahl eines geeigneten HS erstellt. Der zu untersuchende Datensatz wurde hierzu über einen k-Means-Algorithmus mit $k = 5$ auf Basis der $\Delta\Delta G_f^0$ -Werte geclustert. Fünf Cluster wurden ausgewählt, da bei Nutzung des gesamten Datensatzes 3 (114 Konjugate) bei dieser Anzahl visuell ein Kurvenknick beobachtet werden konnte (Abb. 3.13), gemäß der oft genutzten „Ellenbogen“-Methode [300]. Anschließend wurde eine Zufallsauswahl getroffen, die folgende Bedingungen erfüllen musste: Die relativen Anteile (1) der definierten Cluster, (2) der Linkertypen (Kennzeichen nach Tab. C.1) und (3) Strukturen, die aus anderen Proteinen als aus der WW-Domäne entstammen

(1SRL, 1COI, 2ZTA), mussten etwa jenen im gesamten Datensatz entsprechen. Für die Datensätze 1 bis 4, sowie 6 und 9, in denen die Src SH3-Domäne enthalten ist, musste diese zur Validierung in jedem Falle in das HS einsortiert werden.

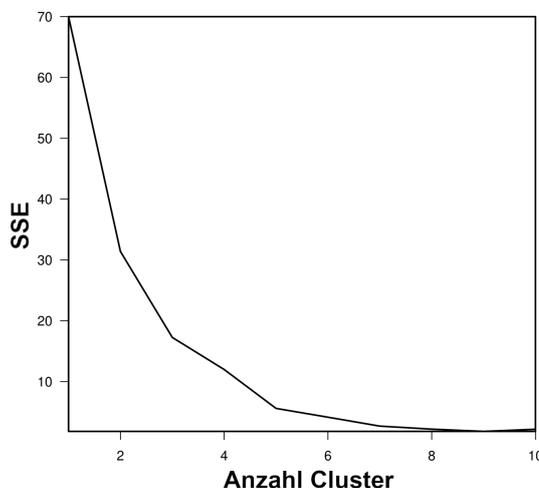


Abbildung 3.13: SSE (*Sum of squared errors*)-Werte in Abhängigkeit des Werts von k beim Clustering des gesamten Datensatzes an 114 Konjugaten. Diese SSE-Werte repräsentieren die Summe der quadrierten Abweichungen aller $\Delta\Delta G_f^0$ -Werte von den dazugehörigen Cluster-Zentroiden.

Die Anzahl an ausgewählten Strukturen wurde jedoch für die zwei „extremen“ Cluster mit einem Faktor von 3 multipliziert, um besonders viele stark (de-)stabilisierte Varianten im HS zu erhalten. Darunter sind die beiden Cluster zu verstehen, die die am stärksten (de-)stabilisierten Konjugate beinhalten. Entsprechend wurde die Anzahl in den übrigen Clustern gleichmäßig reduziert. Dies wurde durchgeführt, da im Datensatz besonders viele Konjugate in einem engen Bereich zwischen etwa -0.70 und -0.20 kcal/mol präsent waren, die eine Modellbildung deutlich erschwerten. Im Fokus der Modellgenerierung stand also die Detektion besonders (de-)stabilisierter Konjugate. Der Clustering-Ansatz ist am Beispiel des Datensatzes 3 in Abb. 3.14 exemplarisch geschildert.

3.1.6.3.4 Modellgenerierung

Aufbereitung der Daten Der durchgeführte Ansatz ist in Abb. 3.15 für den Fall des Datensatzes 6 (relevant für Kap. 4.1.2.2) zusammengefasst. Im Gegensatz zu Ansatz 1 wurde der Datensatz hierbei noch vor Filterung der Variablen in TS und HS aufgetrennt, um einen Einfluss des letzteren auf den Modellbildungsprozess komplett zu verhindern. Es wurden dann wie im vorherigen Ansatz Konjugate gefiltert, die den Konvergenztest nicht bestanden haben. Die Filterung auf Basis der experimentellen Fehler wurde mit einem neuen Cut-off-Wert von 0.16 kcal/mol durchgeführt. So wurde lediglich das Konjugat 23-1 mit einem Fehler von 0.40 kcal/mol entfernt. Deskriptoren wurden, wie in Ansatz 1 bereits beschrieben, gefiltert. Eine Ausnahme bildete das gewählte Interkorrelationslimit, was von 0.85 auf 0.99 erhöht wurde.

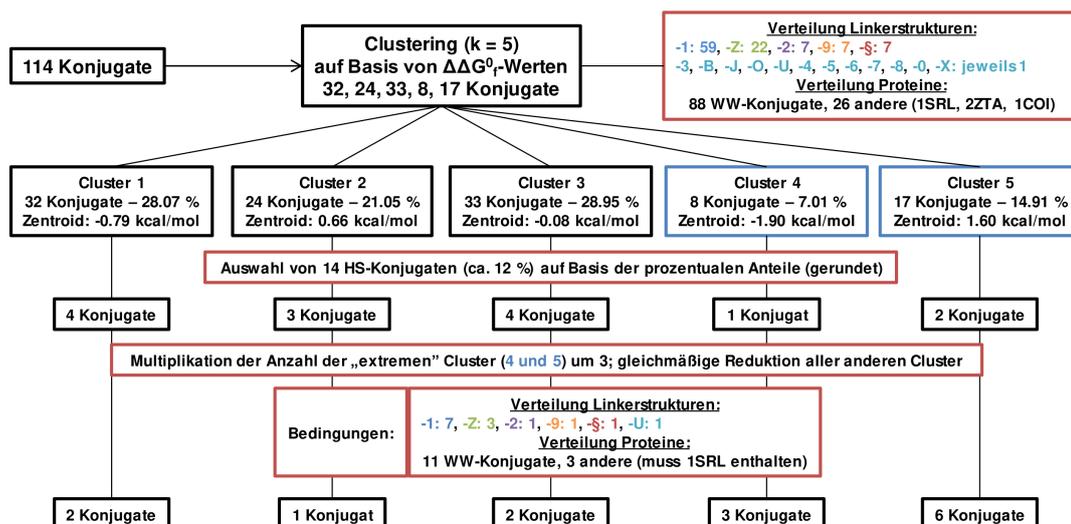


Abbildung 3.14: Durchführung des Clustering-Ansatzes am Beispiel von Datensatz 3. Die Linkerstrukturen sind gemäß Tab. C.1 benannt.

Segmentierungsprozess Der Trainingsdatensatz wurde dann zufällig in gleichgroße Segmente aufgeteilt, sodass in jedem Segment etwa zehn Konjugate vorzufinden waren. Ein Segment wurde nun als Testset ausgewählt. Die restlichen Segmente fungierten als Kalibrierungsset und wurden zur Modellgenerierung verwendet.

Deskriptor-Selektionsmethoden Hierzu wurden verschiedene Subsets der Deskriptoren ausgewählt: Im Falle der Regressionen durch RFE, sowie per Backward-, Forward- und Sequential-Feature-Elimination [220], im Falle der Klassifikationsmodelle nur über RFE. Hierbei wurde die Anzahl an Deskriptoren auf max. 10 beschränkt (in Ansatz 1 gab es keine Limitierung). Backward-, Forward- und Sequential-Feature-Eliminationen wurden auf Basis von linearen Regressionen mithilfe des R-Pakets leaps durchgeführt. Im Gegensatz zum bereits in Ansatz 1 beschriebenen Konzept der Backward-Feature-Elimination, werden bei einer Forward-Elimination sukzessive Deskriptoren zum Modell hinzugefügt. Bei jedem Schritt wird hierbei die Inklusion jedes verbliebenen Deskriptors getestet und im Anschluss jener mit dem geringsten p -Wert bzw. dem stärksten Einfluss auf das Modell hinzugefügt. Der Prozess wird fortgeführt, bis keine Variable mit einem signifikanten Einfluss mehr zum Modell hinzugefügt werden kann oder das Limit von 10 Deskriptoren erreicht wird. Die Sequential-Feature-Elimination stellt eine Kombination der beiden anderen Ansätze dar. Zunächst werden Deskriptoren gemäß der Forward-Methode sukzessive hinzugefügt, nach jedem Schritt wird jedoch ein Entfernen bereits inkludierter Variablen evaluiert [211].

Generierung der Modelle Für jedes generierte Subset wurden anschließend eine wiederholt zehnfach kreuzvalidierte Modellgenerierung durchgeführt und statistische Parameter bzgl. des zum Training genutzten Kalibrierungssets als auch des Testsets ermittelt. Der Prozess der Deskriptorselektierung, der Modellgenerierung und

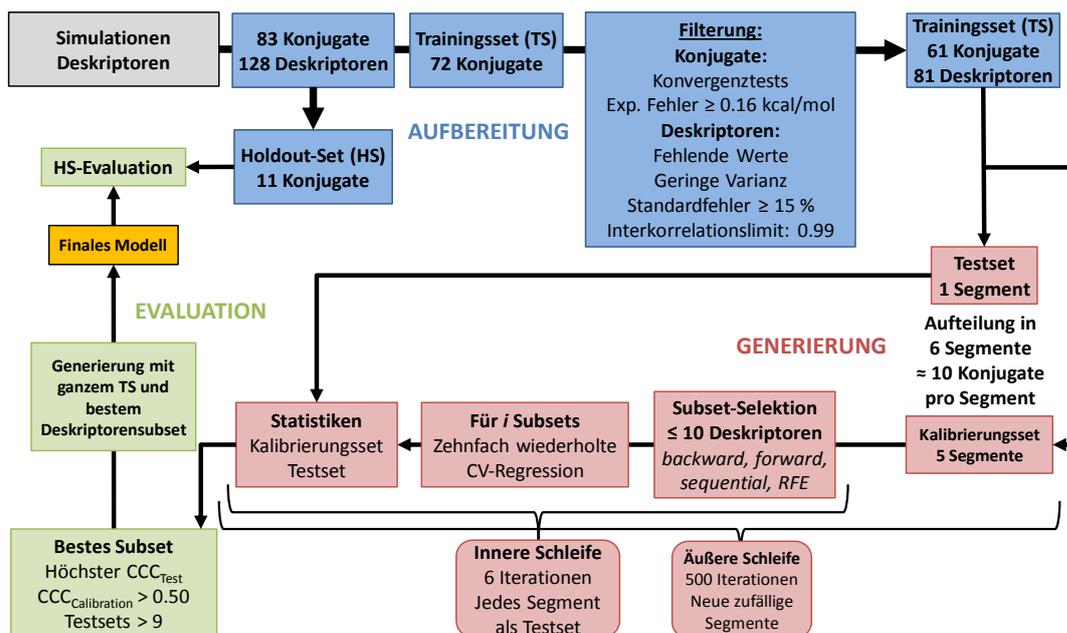


Abbildung 3.15: Übersicht über den zweiten Modellgenerierungsansatz am Beispiel von Datensatz 6. Die Schritte wurden eingeteilt in eine Datenaufbereitungs-, eine Modellgenerierungs- und eine Evaluationsphase.

der statistischen Evaluation (innere For-Schleife) wurde dann für mehrere Durchläufe wiederholt, da jedes Segment einmal als Testset fungierte. Anschließend wurde das Trainingsset in neue Segmente aufgeteilt und der Prozess wurde wiederholt. Schließlich wurde diese äußere For-Schleife für 500 Iterationen durchgeführt.

3.1.6.3.5 Evaluation

Das beste Deskriptorensuperset wurde anschließend ausgewählt. Im Falle der Regressionen wurde jenes selektiert, welches den höchsten Median-Wert des Konkordanzkoeffizienten in Bezug auf die Testsets aufwies, unter der Voraussetzung, dass das Subset auch einen Median-Wert dieses Koeffizienten von über 0.50 für die Kalibrierungssets zeigte und, für eine ausreichende Statistik, mindestens mit zehn Testsets verwendet wurde. Für die Klassifikationsmodelle wurde die gewichtete Genauigkeit als Auswahlkriterium genutzt, wobei das Subset mindestens einen Median-Wert von 0.80 für die Kalibrierungssets zeigen musste.

3.1.6.3.6 Finales Modell

Das auf diese Weise jeweils ausgewählte, beste Deskriptorensuperset wurde anschließend genutzt, um auf Basis des gesamten Trainingsdatensatzes mit wiederholt zehnfacher Kreuzvalidierung ein finales Modell zu generieren. Dieses wurde dann anschließend mit dem externen Holdout-Set (HS) geprüft. Abwandlungen des Deskriptorensupersets durch verschiedene Grenzwerte eines anschließenden Interkorrelationslimits (0.75 bis 0.99) wurden ebenfalls getestet. Die Ergebnisse werden in Kap. 4.1.2 geschildert. Wie

in Ansatz 1 wurden die Applikationsdomänen der Modelle mit dem ADAN-R-Paket untersucht.

3.1.7 Metadynamik-Simulationen

3.1.7.1 Untersuchte Systeme

Um den Unterschied in der konformativen Stabilität zwischen einem PEGylierten und einem unPEGylierten Peptid *in silico* erfassen zu können, wurden wohltemperierte Metadynamik-Simulationen mit multiplen „Walkern“ durchgeführt. Als Fallbeispiele zur Untersuchung dienten hier die WW-Konjugate 16-1 und 27-1 (sowie deren unPEGylierte Mutanten). Mit einer starken Stabilisierung von -0.90 ± 0.03 kcal/mol bei einer PEGylierung an Position 16 und einer relativ starken Destabilisierung von 0.38 ± 0.04 kcal/mol an Position 27 eigneten sich diese Varianten hierfür sehr gut.

3.1.7.2 Kollektivvariablen

Auf Basis der Ergebnisse aus [255] wurden zwei Kollektivvariablen (CV 1 und CV 2) unter Nutzung des PLUMED-Plugins (Version 2.5.0) definiert: die Summe der Proteinerückgrat-HB-Distanzen zwischen (1) dem ersten und zweiten, sowie (2) dem zweiten und dritten β -Faltblatt. Dies entsprach folgenden Werten: für CV 1 die Summe aus den Distanzen G10_O-H27_H, F25_O-E12_H, E12_O-F25_H, Y23_O-R14_H, R14_O-Y23_H und R21_O-S16_H, für CV 2 die Summe aus den Distanzen A31_O-N26_H, Y24_O-Q33_H und Q33_O-Y24_H (Abb. 3.16).

3.1.7.3 Simulationsparameter

Die vorbereiteten Modellproteine wurden wie in den Kap. 3.1.4.1 und 3.1.5 beschrieben solvatisiert und equilibriert. Für den Produktionslauf wurden acht „Walker“ mit der Endkonformation der NPT-Equilibrierung als Startposition gestartet. Diese wurden für jeweils 300 ns simuliert, was einer Gesamtzeit von 2.4 μ s pro Protein entsprach. Folgende Einstellungen wurden zur Kontrolle der CVs genutzt: Kurvenbreite $\sigma_s = 0.05$, Höhe $\omega = 0.05$ und Frequenz $\tau_G = 10$ ps. Eine Temperierung der MetaMDs wurde mithilfe eines Bias-Faktors von 8 erreicht. Der erlaubte Sampling-Bereich der CV 1 lag bei 0 bis 5.2 nm, jener von CV 2 bei 0 bis 2.8 nm. Um keinen Konformationsraum außerhalb des definierten Gitters zu erreichen, wurden die oberen Energiebarrieren 4.8 nm für CV 1 und 2.4 nm für CV 2 mit der Konstanten κ mit einem Wert von 1000 kJ/mol eingestellt. Diese Konstante wird genutzt, um die Stärke des zurückhaltenden Potentials V_{CV} zu definieren, das auf das System ausgeübt wird, sobald dieses die genannten Gittergrenzen überschreitet:

$$V_{CV} = \kappa \left(\frac{x - a + o}{s} \right)^e \quad (3.1)$$

Dabei ist s ein Skalierungsfaktor mit Standardwert 1, e definiert das Potenzgesetz mit Standardwert 2, a ist das vom Nutzer eingestellte Limit der CV und o (*offset*)

eine mögliche Verschiebung dieses Limits, die jedoch standardmäßig auf Null gesetzt ist [301]. Die Dateien mit den Kurvenhöhen (*Hills*) wurden durch die Walker alle 100 Zeitschritte gelesen. Die CVs wurden alle 500 Zeitschritte herausgeschrieben. Im Anschluss wurden die Simulationen mithilfe des R-Pakets *metadynminer* analysiert [302]. Die Resultate der Simulationen werden in Kap. 4.2 beschrieben.

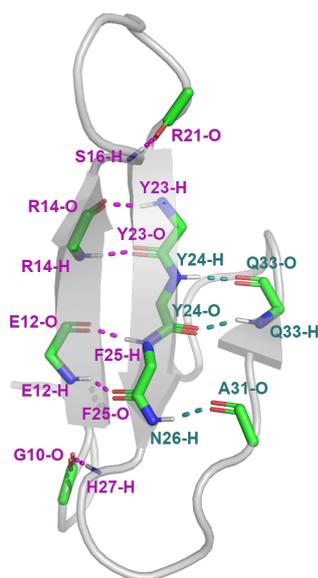


Abbildung 3.16: Darstellung der für die Kollektivvariablen genutzten Distanzen (CV 1 in magenta, CV 2 in türkis). Die relevanten Atome sind in den selben Farben gekennzeichnet.

3.1.8 Vergleich der Polymere

Für einen systematischen Vergleich des *in silico*-Verhaltens der Polymere PEG, LPG, PMeOx und PEtOx wurden verschiedene Eigenschaften für eine Reihe von WW- und Her2-Affibody-Konjugaten näher untersucht: (1) die Interaktionsenergien der Polymeratome mit jeder Aminosäure und (2) dem gesamten Protein (jeweils mit -1 multipliziert), (3) die vom Polymer bedeckte Proteinoberfläche und (4) der Anteil der Polymeroberfläche, die bedeckt wird, (5) Protein-Polymer HB pro Polymeroberfläche in \AA^2 , (6) die Anzahl an Wassermolekülen innerhalb von 5\AA der Polymeratome pro Oberfläche in \AA^2 , (7) das massengewichtete RMSF (*Root-mean-square fluctuation*)-Verhältnis des Proteinrückgrats pro Aminosäure im Vergleich zur Simulation der unPEGylierten Variante, sowie schließlich (8) die massengewichteten RMSF-Werte der Polymerschweratome.

Im Falle der Pin 1 WW-Domäne wurden die Asn-Konjugate XX-1, XX- ϵ , XX- λ und XX- ψ für die 13 Konjugationsstellen $XX = 14, 16, 17, 18, 19, 21, 23, 26, 27, 28, 29, 30$ und 32 miteinander verglichen. Für den Her2-Affibody wurden die 15 Stellen $XX = 7, 8, 15, 21, 23, 25, 29, 33, 39, 43, 46, 47, 49, 50$ und 53 hinsichtlich der Cys-Maleimid-Konjugate XX-4, XX- ϕ , XX- γ und XX- ω untersucht. Die Ergebnisse sind in Kap. 4.3 aufgeführt.

3.2 Biologika

Im Folgenden werden die simulierten Konjugate von IFN- α 2a beschrieben (Kap. 3.2.2). Analog zur Vorgehensweise bei den Modellproteinen sollen die Parametrisierung (Kap. 3.2.3), die Modellierung (Kap. 3.2.4) und das Simulationsprotokoll (Kap. 3.2.5) beschrieben werden. Schließlich werden die Durchführung einer Coarse-grained-Simulation eines PEG-Konjugats (Kap. 3.2.6) und die experimentellen Arbeiten der Verbundpartner zur Charakterisierung der thermischen Stabilität und Bioaktivität (Kap. 3.2.7) geschildert. Ergebnisse und Diskussion sind in Kap. 5 zu finden.

3.2.1 Wildtyp-Simulationen

Im Falle von IFN- α 2a wurden die in der PDB-Datei 1ITF aufgelisteten, ersten zehn der 24 Konformationen für jeweils 150 ns simuliert (insgesamt 1.5 μ s), mit den in den Kap. 3.1.4.3 und 3.1.5.1 bereits geschilderten Einstellungen für konventionelle MD-Simulationen der Modellproteine.

3.2.2 Simulierte Konjugate

IFN- α 2a wurde experimentell an der Position K31 mit den Polymeren PEG, LPG und PEOx konjugiert, indem an dieser Stelle die nicht-natürliche Aminosäure AZK eingebaut wurde und diese über eine SPAAC mit einem DBCO- oder BCN-Linker konjugiert wurde (siehe Kap. 1.1.2.3 und 3.2.7). Diese Konjugationsstelle stellt eine der ebenfalls in Pegasys® adressierten Positionen dar und eignete sich, unter Berücksichtigung der Binderegionen des Proteins zu den Rezeptoren IFNAR1 und IFNAR2 (vgl. Abb. 1.14-A), neben K134 als Angriffsort des Linkers. K31 und K134 stellen die Konjugationsstellen in Pegasys für PEG dar, die vermutlich zur geringsten Verminderung der Bioaktivität führen [303]. Ebenso zeigen die in dieser Arbeit durchgeführten WT-Simulationen, dass die Seitenketten beider Aminosäuren weit entfernt von den Binderegionen des Proteins vorzufinden sind und sich nicht zu diesen hin ausrichten (Abb. 3.17). K134 lieferte experimentell bei der Expression allerdings geringere Ausbeuten, sodass der Fokus auf K31 gelegt wurde [304].

In silico wurden BCN-Varianten mit unterschiedlich langen Polymeren (10 - 50 kDa) simuliert. DBCO-Konjugate wurden lediglich mit 10-kDa-Polymerketten simuliert, da auch nur für Konjugate dieser Größe experimentelle Daten erhoben wurden (Tab. 3.4).

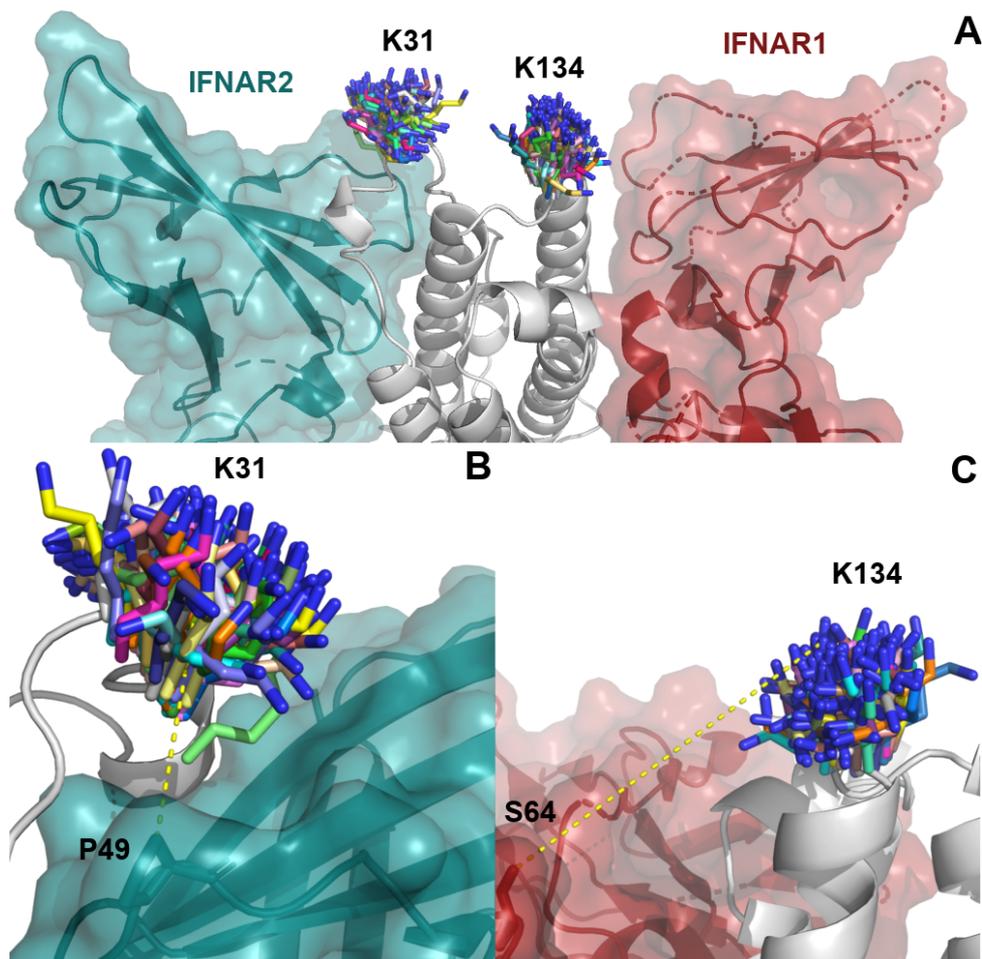


Abbildung 3.17: (A) Darstellung der Seitenketten von K31 und K134 für jede Nanosekunde eines Replikons der WT-Simulationen, nach Alignierung des Proteins auf die Kristallstruktur 3SE3. (B) Die mittlere Distanz von K31 zum nächstgelegenen Atom P49-C β von IFNAR2 beträgt 11.48 Å, (C) jene von K134 zu S64-O γ von IFNAR1 16.69 Å (Nachdruck mit Genehmigung aus [304], Copyright 2021 American Chemical Society).

Tabelle 3.4: Auflistung aller Konjugate von IFN- α 2a, die für die Arbeiten in Kap. 5 simuliert wurden. Dabei sind die zum Zeitpunkt der Modellierung experimentell angestrebten Molekulargewichte, die Anzahlen modellierter Monomere und die entsprechenden Massen letzterer aufgelistet.

Linker	Polymer	Zielgewicht [kDa]	Modellierte Monomere	Masse [kDa]
BCN	PEG	10	227	10.02
BCN	LPG	10	136	10.07
BCN	PEtOx	10	101	10.03
DBCO	PEG	10	227	10.02
DBCO	LPG	10	136	10.07
DBCO	PEtOx	10	101	10.03
BCN	PEG	20	452	19.92
BCN	LPG	20	290	21.45
BCN	PEtOx	20	200	19.84
BCN	PEG	40	903	39.80
BCN	LPG	50	650	48.15
BCN	PEtOx	30	300	29.76

3.2.3 Parameterfindung

Die Parametrisierung von PEG-Monomeren wurde bereits in Kap. 3.1.3.3 beschrieben. Mit den LPG-, PMeOx- und PEtOx-Monomeren wurde analog verfahren. Da die LPG-Strukturen Stereozentren enthielten, wurde hierbei ein Dimer als einzelner Baustein definiert, welches über ein (S)- und ein (R)-konfiguriertes Monomer verfügte. Die konjugierte Lysinseitenkette wurde mitsamt des konjugierten DBCO- oder BCN-Linkers als eine neue Aminosäure definiert. Während für das Proteinerückgrat dieselben Endgruppen wie in Kap. 3.1.3.1 und 3.1.3.2 verwendet wurden, wurde die jeweils endständige Amid-Funktion, die mit der Polymerkette verbunden war, mit PEG-Monomeren während der Ladungsverteilung über die in Kap. 3.1.3.2 beschriebene antechamber-Methode versehen (Abb. 3.18).

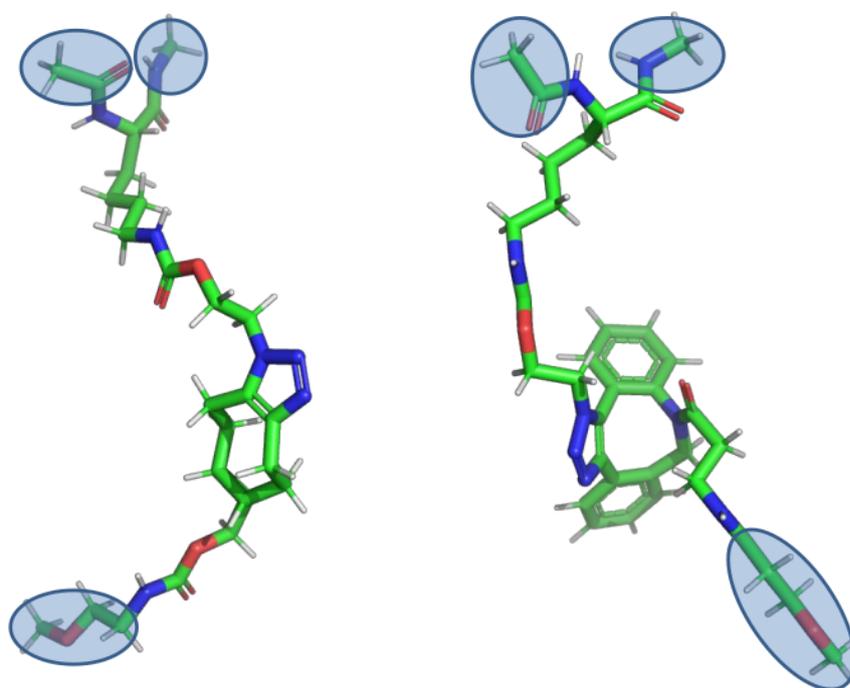


Abbildung 3.18: Parametrisierte BCN- (links) und DBCO-Linker (rechts). Die Endgruppen, die dabei gemäß der beschriebenen antechamber-Methode in xleap entfernt wurden, sind blau gekennzeichnet.

3.2.4 Modellierung der Strukturen

Die NMR-Struktur von IFN- α 2a, die bei pH 6 aufgenommen worden war, wurde ohne Änderung der Protonierung verwendet. MOE deprotonierte den N-Terminus und änderte die Tautomerie einiger solvens exponierter Histidine bei pH 7, Maestro 2019.1 von Schrödinger [305] hingegen nicht. Die Polymerketten mit den in Tab. 3.4 aufgeführten Größen wurden zunächst mittels tleap von AmberTools18 in einer ausgestreckten Konformation generiert. Diese wurden dann an den BCN-Linker konjugiert und in die Position K31 der NMR-Struktur des IFN- α 2a eingefügt. DBCO-Varianten wurden analog modelliert.

Zunächst wurden die Aminosäuren mit Ausnahme der Position K31 in MOE fixiert. Die potentielle Energie des Systems wurde minimiert (Amber14:EHT-Kraftfeld, Gradient: 0.001 kcal/mol/Å), bevor eine kurze Simulation mit implizitem Solvens (1 ns) durchgeführt wurde. Hierfür wurde der Noisé-Poincare-Andersen-Ansatz [306, 307] unter Anwendung des generalisierten Born-Solvensmodells [308] in MOE genutzt. Dies führte zu komprimierten Startstrukturen, bei denen sich das Polymer auf der Proteinoberfläche befand (Abb. 3.19). Erste klassische MD-Simulationen mit langgestreckten Polymerketten als Startpositionen wurden getestet, führten jedoch zu keiner Annäherung des Polymers auf die Proteinoberfläche in annehmbarer Zeit. Auf diese Weise konnte ebenso die Anzahl nötiger Solvensmoleküle deutlich verringert werden.

Zehn verschiedene Startstrukturen wurden für jedes Konjugat generiert, wobei sich das Polymer dem Protein in den Simulationen mit implizitem Solvens jeweils aus verschiedenen Richtungen annäherte und somit in den darauffolgenden gaußbeschleunigten Molekulardynamik-Simulationen (Kap. 3.2.5) Wechselwirkungen mit unterschiedlichen Arealen abgebildet werden konnten. Erste Simulationen mit rechteckigen Solvensboxen wiesen Interaktionen des Polymers mit Proteinen der periodischen Nachbarn auf. Daher wurden die generierten Konjugate in isometrischen, trunkierten Oktaedern mit einem Mindestabstand von 15 Å des Konjugats bis zum Rande der Box mit TIP3P-Wasser solvatisiert und mit Natrium- oder Chloridionen das System neutralisiert.

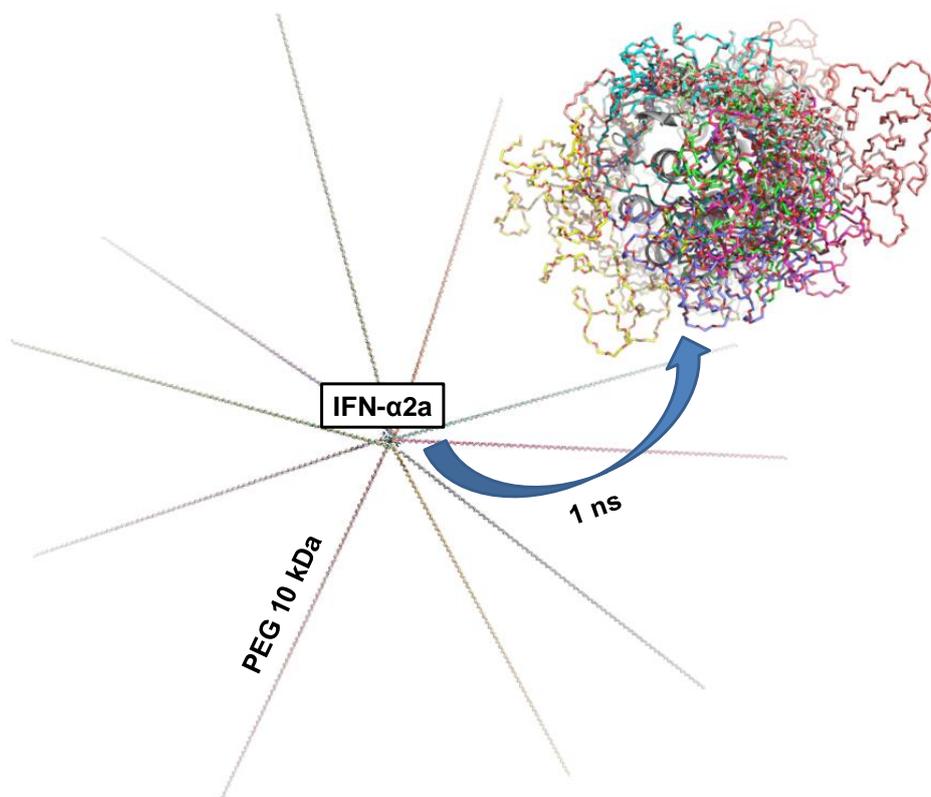


Abbildung 3.19: Langgestreckte Strukturen des Polymers PEG 10 kDa nach Generierung durch tleap und Einfügung an Position K31. Rechts sind die Startkonformationen der Simulationen dargestellt, die nach initialer Simulation in implizitem Solvens erhalten wurden.

3.2.5 Gaußbeschleunigte Molekulardynamik-Simulationen

Für die zehn generierten Startkonformationen jedes Konjugats wurde eine NVT-Equilibration analog zu jener der Modellproteine (siehe Kap. 3.1.5.1) durchgeführt. Anschließend wurde eine auf Basis des GaMD-Ansatzes [176, 177] 20 ns lange NPT-Equilibration getätigt. Hierbei wurden zunächst 2 ns einer konventionellen MD-Simulation durchgeführt. In der zweiten Hälfte dieser Phase (1 ns) wurden Werte für V_{max} , V_{min} , V_{avg} und σ_V gesammelt und entsprechend E und k_0 berechnet. Die letzten Werte für E und k_0 wurden dann genutzt, um ein konstantes, gaußbeschleunigtes Zusatzpotenzial für eine weitere Nanosekunde anzuschalten. Dieses Zusatzpotenzial wurde dann in den darauffolgenden, restlichen 17 ns der Equilibration durch Änderung von E und k_0 nach jedem Zeitschritt verändert. Für den GaMD-Ansatz wurde die duale Boost-Methode gewählt, bei der das Gesamt- als auch das Diederpotential im System durch das harmonische Zusatzpotenzial angehoben werden. Das obere Limit der Standardabweichung σ_0 der beiden Potenziale wurde auf 6 kcal/mol gesetzt, dem Standardwert in NAMD. Momentaufnahmen wurden alle zehn Pikosekunden herausgeschrieben. Im Anschluss an diese Equilibrationsphase wurde dann der Produktionslauf mit randomisierten Startgeschwindigkeiten für insgesamt 150 ns durchgeführt, in welchem die Werte E und k_0 weiter verändert werden konnten.

Jede Polymervariante wurde mit zehn verschiedenen Startkonformationen und somit für 1.5 μ s simuliert. Dies ähnelt den Simulationsansätzen aus [107] und [309], wobei die zehn unterschiedlichen Startkonformationen in diesen Referenzstudien mittels der pycsim-Software [310, 311] über das Prinzip eines selbstmeidenden Pfades generiert wurden. Dabei wurden sukzessive neue Monomere an die Konjugationsstelle gebaut und die entsprechende Kette nach jedem Schritt minimiert. Wenn jedoch bedacht wird, dass die Kette experimentell über eine *Grafting to*-Methode mit ihrer ganzen Länge an das Protein konjugiert wurde, sollte der in dieser Arbeit angewandte Ansatz über eine Energieminimierung und eine Simulation der ausgestreckten Polymerkette eine mögliche Alternative darstellen. Die Dynamik der PBK wird in Kap. 5.1 geschildert.

3.2.6 Coarse-grained-Simulation

Eine an Position K31 mit 10 kDa PEG konjugierte Variante von IFN- α 2a wurde mithilfe des Kraftfelds MARTINI 2 für 10 μ s simuliert [158, 312]. Geeignete Parameter und Startstrukturen für PEG wurden mithilfe des PolyPly-Programms (Version 1.1.0) vorbereitet [313, 314]. Zum Coarse-graining des Proteins wurde zunächst das martinize.py-Skript genutzt. Die erhaltene Struktur wurde in eine kubische Simulationsbox mit 8 nm Mindestabstand des Proteins zum Rand gesetzt. Eine initiale Energieminimierung (100 Schritte) der vergrößerten Struktur im Vakuum wurde dann durchgeführt. Anschließend wurde die Polymerkette mithilfe von PolyPly sukzessive in einer randomisierten Konformation ohne Linker an die terminale Seitenkettenkugel von K31 konjugiert und verlängert (227 Monomere). Vor und nach

einer anschließenden Solvataion mit nicht polarisierbarem MARTINI-Wasser und Neutralisation durch Natrium- oder Chloridionen wurde das System jeweils nochmals energieminiert, bevor eine Equilibrierung für 10000 Schritte mit 2 fs-Intervallen durchgeführt wurde. Schließlich folgte der eigentliche Produktionslauf für 10 μ s mit 20-fs-Zeitschritten. Coulomb-Interaktionen wurden über die Reaktionsfeld-Methode und VdW-Wechselwirkungen gemäß eines Verlet-Schemas jeweils mit einer Cut-off-Distanz von 1.1 nm berücksichtigt. Zur Temperaturkontrolle wurde die Velocity-Rescale-Methode, zur Druckkontrolle das Parrinello-Rahman-Barostat angewandt. Eine Struktur des Konjugats vor Start des Produktionslaufes ist in Abb. 3.20 gezeigt. Die Beobachtungen werden in Kap. 5.2 beschrieben.

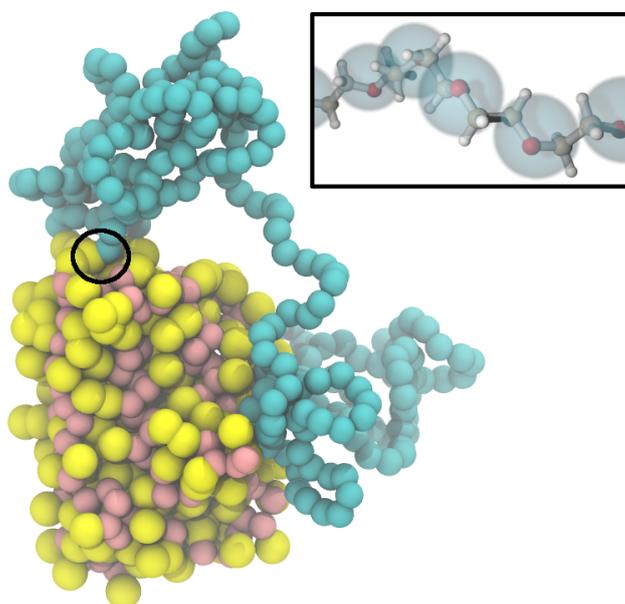


Abbildung 3.20: Startstruktur der MARTINI-Simulation vor dem Produktionslauf. PEG-Kugeln sind in cyan dargestellt, Proteinrückgratkugeln in rot und Seitenkettenkugeln in gelb. Die Konjugationsstelle der K31-Seitenkettenkugel ist markiert. Rechts daneben ist eine schematische Darstellung der Vergrößerung der PEG-Monomere illustriert.

3.2.7 Experimentelle Arbeiten

LPG wurde an der FU Berlin über eine anionische und PEtOx an der FSU Jena über eine kationische ringöffnende Polymerisation gewonnen, entsprechende mPEG-Ketten konnten kommerziell erworben werden (Abb. 3.21-A, B und C). Alle weiteren Arbeiten in Kap. 3.2.7 wurden durch Niklas Hauptstein an der JMU Würzburg durchgeführt. IFN- α 2a und funktionalisierte Varianten wurden durch Nutzung einer Amber-Codon-Mutation in BL21 StarTM(DE3)-Zellen exprimiert, erhaltene Einschlusskörperchen isoliert und aufbereitet, das Protein rückgefaltet und über Ionenaustausch- und Größenausschlusschromatographien aufgereinigt. Anschließend wurde die SPAAC durchgeführt (Abb. 3.21-D), die Proteine wurden durch verschiedene Methoden weiter charakterisiert (CD-Spektroskopie, Polyacrylamid-Gelelektrophorese, dynamische Lichtstreuung und Massenspektrometrie).

Um die Bioaktivität der Konjugate zu überprüfen, wurden SEAP (*Secreting alkaline phosphatase*)-Assays mit HEK (*Human embryonic kidney*)-IFN α/β -Zellen durchgeführt. Durch Bindung aktiven Interferons setzen diese Zellen über den JAK-STAT-Signalweg SEAP in das umgebende Medium frei. Die Aktivität der Phosphatase kann wiederum durch eine QUANTI-Blue™-Lösung detektiert werden [315].

Zur Untersuchung der thermischen Stabilitäten wurden Thermal Shift Assays durchgeführt, bei denen die Entfaltung über die Bindung des Farbstoffs SYPRO™ Orange bei erhöhten Temperaturen beobachtet wird. Durch die Bindung des hydrophoben Farbstoffs kommt es zur Änderung der Fluoreszenzintensität, wodurch die Schmelz- bzw. Entfaltungstemperaturen im Anschluss über den Wendepunkt der erhaltenen Fluoreszenzkurve ermittelt werden können [89, 90].

Zusätzlich wurde die Proteolysestabilität der Konjugate untersucht, indem sogenannte Limited Proteolysis (LiP)-Assays durchgeführt wurden [316]. Hierbei wurden die PBK mit einer Proteinase K versetzt. Anschließend wurden zu verschiedenen Zeitpunkten (nach 1, 3, 5, 10 und 30 min) Proben entnommen und aufbereitet. Schließlich wurden die durch die Proteolyse erhaltenen Peptidfragmente mit massenspektrometrischen Methoden charakterisiert. Die Ergebnisse werden in den Kap. 5.3 und 5.4 mit den Beobachtungen der Simulationen in Verbindung gebracht.

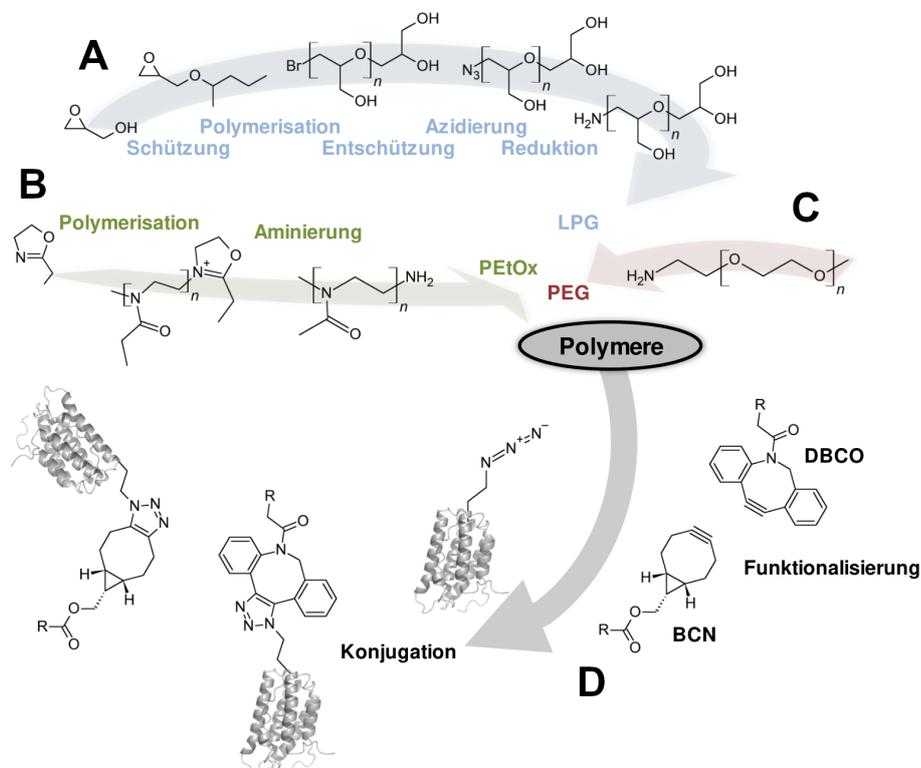


Abbildung 3.21: (A) Synthese von LPG über eine anionische ringöffnende Polymerisation. (B) Synthese von PEtOx über eine kationische ringöffnende Polymerisation. (C) Struktur der verwendeten mPEG-Kette. (D) Funktionalisierung der gewonnenen Polymere durch die Linker BCN und DBCO, sowie anschließende SPAAC-Reaktion über eine Azidgruppe des Proteins zur Konjugation der synthetisierten Polymere.

Kapitel 4

Ergebnisse: Modellproteine

4.1 QSPR-Vorhersagemodelle

Die systematischen Vorarbeiten von Price et al. zum Modellprotein Pin 1-WW, welches mit PEG-Oligomeren verknüpft wurde, deuten darauf hin, dass die PEG-basierte Stabilisierung durch deutlich einfachere Systeme als die üblicherweise therapeutisch genutzten Biologika mit Polymerketten von 5 bis 40 kDa beurteilt werden kann. Solche Modellproteine sind auch durch *in silico*-Ansätze in größerem Maßstab analysierbar. Während erweiterte Simulationsmethoden prinzipiell zur Ermittlung von Energiedifferenzen genutzt werden können (vgl. Kap. 4.2), stellen diese zum einen sehr rechenintensive Ansätze zur Untersuchung mehrerer Systeme dar, zum anderen handelt es sich im vorliegenden Fall um äußerst geringe energetische Differenzen. Im Hinblick auf die Genauigkeit computerchemischer Methoden zur Ermittlung von Energiewerten kleinerer Moleküle wird üblicherweise von einem Mindestfehler von 1 kcal/mol gesprochen, im Falle makromolekularer Systeme wie PBK ist dieser sicherlich noch größer [317, 318].

Daher kann davon ausgegangen werden, dass simulationsbasierte Energieberechnungen für die hier vorliegenden Konjugate nur schwer zu realisieren sind. QSPR-Ansätze stellen in diesem Falle einen recheneffizienteren Ansatz dar und können zusätzlich Einblick in die wichtigsten Stabilisierungsfaktoren zur Vorhersage weiterer Konjugationseffekte geben. Da sich die Polymerketten von PBK jedoch sehr flexibel verhalten und verschiedene Areale des Proteins adressieren können, ist ein wie in dieser Arbeit gewählter simulationsbasierter Ansatz unter Berücksichtigung der Dynamik zur Generierung geeigneter Deskriptoren sinnvoll.

Im Folgenden sollen die Ergebnisse der in Kap. 3.1.6 beschriebenen Modellgenerierungen geschildert werden. Zunächst werden die QSPR-Modelle aus Ansatz 1 beschrieben (Kap. 4.1.1), in welchem unterschiedliche Methoden des maschinellen Lernens auf Basis der Vorhersagen für ein einzelnes Holdout-Set getestet wurden. Im Anschluss werden die Resultate zu Ansatz 2 beschrieben (Kap. 4.1.2), in welcher ein erweitertes Kreuzvalidierungsverfahren angewandt wurde, um multilineare Regressions- und binomiale logistische Klassifikationsmodelle zu erstellen.

4.1.1 Ansatz 1

Im ersten Modellierungsansatz zur Vorhersage der PEG-basierten Stabilisierung von PBK wurden verschiedene Modellarten getestet. Dabei wurden Modelle mithilfe eines Trainingssets (TS) erstellt und auf Basis der Leistung bzgl. eines einzelnen Holdout-Sets (HS) selektiert. Dies wurde gemäß Kap. 3.1.6.2 für vier verschiedene Kategorien durchgeführt: (A) Pin 1-spezifische und (C) unspezifische (auf andere Systeme theoretisch transferierbare) Regressionsmodelle, sowie (B) Pin 1-spezifische und (D) unspezifische Klassifikationsmodelle. In den Kategorien A und B zeigten die Pin 1-spezifischen Modelle jedoch eine vollständige Überanpassung (*Overfitting*) an das TS und nutzten dabei eine in Relation zur Datensatzgröße viel zu große Anzahl an Deskriptoren. Die Resultate dieser Kategorien sollen daher im Folgenden nicht weiter diskutiert werden. Eine Übersicht der drei ausgewählten Modelle jeder Kategorie ist dennoch in Tab. D.3 zu finden.

Auch im Falle der hier nun näher diskutierten Pin 1-unspezifischen Modelle mit deutlich weniger Deskriptoren sollte angemerkt werden, dass es sich um sehr komplexe Modelltypen mit vielen Hyperparametern handelt, die bei solch kleinen Datensätzen auch bei einer geringen Anzahl an Deskriptoren zum Overfitting neigen können. Weiterhin wurden die besten Modelle bzgl. der Vorhersageleistung für ein einzelnes Holdout-Set selektiert, weshalb es sich um kein ausreichend externes Validierungsset handelt. Ansatz 1 soll wegen dieser Limitierungen lediglich aufgrund folgender Aspekte präsentiert werden: (1) Die Grenzen der Nutzbarkeit erweiterter Methoden des maschinellen Lernens in Bezug zum zur Verfügung stehenden Datensatz sollen aufgezeigt werden, (2) die Eignung der berechneten Deskriptoren zur Wiedergabe der in der Literatur postulierten Stabilisierungsmechanismen soll an exemplarischen Simulationen evaluiert werden, (3) der Ansatz dient als Vorarbeit zu den besser validierten Modellen im hierzu abgewandelten Ansatz 2 (Kap. 4.1.2).

4.1.1.1 Referenzmodelle

Zunächst ist zur Beurteilung der Leistung der generierten Regressionsmodelle entscheidend, wie gut ein Modell abschneiden würde, welches jedem Konjugat des HS den Medianwert des TS (-0.37 kcal/mol) zuordnen würde. Dieses Modell soll als Referenz dienen und liefert einen $RMSE_{HS}$ -Wert von 0.39 kcal/mol. Somit sind nur Vorhersagemodelle mit einem deutlich kleineren Wert als 0.39 kcal/mol für eine weitere Analyse von Relevanz.

Analog hierzu kann ein Referenzmodell für die Klassifikationsmodelle generiert werden. Hierfür wurde jedes Konjugat des HS einmalig zufällig einer der beiden Klassen zugeordnet (also mit einer Wahrscheinlichkeit von jeweils 50 %, da die Verteilung im TS auch annähernd 1:1 ist). Für eine solche Einteilung wird eine CCR_{HS} von 0.67 erhalten. Geeignete Vorhersagemodelle sollten also zu einer deutlich besseren CCR_{HS} führen.

4.1.1.2 Modelle C1 und D1

Im Folgenden sollen die beiden besten Modelle der Kategorien C und D beschrieben werden (genannt C1 und D1). C1 stellt ein entscheidungsbaumbasiertes XGBoost-Modell dar, mit RMSE_{HS} - und CCC_{HS} -Werten von 0.24 kcal/mol bzw. 0.75 (Tab. 4.1). Beim Blick auf die Vorhersagen wird deutlich, dass eine korrekte Rangordnung der Konjugate v. a. in der Nähe des Medians des Datensatzes nur bedingt möglich ist (Abb. 4.1). Außerhalb der ADAN-Applikationsdomäne (siehe Kap. 3.1.6.2) befinden sich 4 der 15 HS-Konjugate: 19-4, 16-Z, 19-J und 16+19-1 (siehe Tab. D.4 für die Einteilungen pro Konjugat).

Das Konjugat der Src SH3-Domäne, welches von Price et al. zur Validierung genutzt wurde (genannt 1SRL.20-1), wird durch das Modell korrekterweise als stabilisiert eingestuft. Mit dem kristallstrukturbasierten Vorhersagemodell von Price et al. (Funktion der Regressionsgeraden: $\Delta\Delta G_f^0 = -1.05 + 0.0073 \theta_{\text{Price}} [100]$) wird eine Faltungsenergie von -0.49 kcal/mol erhalten. Das Modell C1 enthält zusätzliche Informationen von Konjugaten mit anderen Linkerstrukturen und erzielt eine ähnliche Leistung, es sagt eine Differenz der freien Faltungsenergie von -0.78 kcal/mol voraus. Dies ist zwar immer noch relativ weit vom experimentellen Wert von -1.20 kcal/mol entfernt, 1SRL.20-1 stellt mit einem solch negativen Wert und einem experimentellen Fehler von 0.10 kcal/mol jedoch auch ein Konjugat dar, welches im Vergleich zu den anderen PBK im Datensatz eine sehr große Stabilisierung zeigt und theoretisch durch den Filterschritt in Ansatz 1 herausgefiltert werden würde. ADAN klassifiziert dieses Konjugat weiterhin in Kategorie 3, was bedeutet, dass diese Vorhersage als eher unsicher einzustufen ist.

Das Modell nutzt 11 verschiedene Deskriptoren. Die bedeutendsten sind *protpeghbonds* (Anzahl an Protein-PEG-HB), *protprothbondsratio* (Verhältnis der Anzahl an Protein-Protein-HB zwischen PEGylierter und unPEGylierter Variante), *shelloratio* (Verhältnis der Anzahl an Wassern in der zweiten Hydrathülle zwischen PEGylierter und unPEGylierter Variante), *rmsdp* (RMSD-Wert der Polymerschweratome) und *rdffpeg_200* (RDF-Wert um PEG innerhalb von 3 - 4 Å).

Das beste Modell der Kategorie D, ein neuronales Netz, erzielt einen CCR_{HS} -Wert von 0.80 und einen AUROC_{HS} -Wert von 0.91. Es nutzt eine verborgene Schicht (*Hidden layer*) mit 5 Neuronen und 2 zusätzlichen Bias-Neuronen (Abb. 4.2). 1SRL.20-1 wird mit einer Wahrscheinlichkeit von 58 % korrekterweise der Klasse 1 zugeordnet. Das Modell nutzt jedoch andere Deskriptoren als das Regressionsmodell C1, u. a. *pgbrg* (PEG-PEG-HB über überbrückende Wassermoleküle), *rdfoh_200ratio* (RDF-Verhältnisse um S/T/Y), *radgyrpeg* (Gyrationsradius der Polymerkette) und *rmsdratio* (Proteinrückgrat-RMSD-Verhältnisse zwischen PEGylierter und unPEGylierter Variante).

Tabelle 4.1: Statistische Validierungsparameter für die Modelle C1 und D1, sowie deren verwendete Deskriptoren (geordnet nach der Bedeutung für das Modell). Für das XGBoost-Modell wird der Gewinn (*Gain*) dargestellt. Dieser repräsentiert den Bruchteil des Beitrags dieses Deskriptors für alle im Prozess generierten Entscheidungsbäume (Summe = 1) [226]. Für das neuronale Netz wird eine Methode genutzt, die die Absolutwerte aller Gewichtungen im Netzwerk bzgl. der einzelnen Deskriptoren kombiniert [319].

Modell C1									
#	Art	RMSE _{TS}	MAE _{TS}	CCC _{TS}	R ² _{TS}	RMSE _{HS}	MAE _{HS}	CCC _{HS}	R ² _{HS}
C1	xgbTree	0.14	0.10	0.95	0.92	0.24	0.20	0.75	0.63
Deskriptor					Gain				
<i>protpeghbonds</i>					0.23				
<i>protprothbondsratio</i>					0.20				
<i>shelloratio</i>					0.12				
<i>rmsdp</i>					0.12				
<i>rdfpeg_200</i>					0.07				
<i>rdf_oh400ratio</i>					0.06				
<i>rmsdratio</i>					0.06				
<i>totalint</i>					0.06				
<i>pgbrg</i>					0.05				
<i>radgyrpeg</i>					0.02				
<i>entropysideratio</i>					0.01				
Modell D1									
#	Art	AUROC _{TS}		CCR _{TS}		AUROC _{HS}		CCR _{HS}	
D1	nnet	0.95		0.93		0.91		0.80	
Deskriptor					Gewichtungen				
<i>pgbrg</i>					17.70				
<i>rdfoh_200ratio</i>					16.56				
<i>radgyrpeg</i>					16.04				
<i>rmsdratio</i>					15.74				
<i>rmsdp</i>					11.10				
<i>shellohratio</i>					11.04				
<i>pg4dist</i>					7.22				
<i>entropysideratio</i>					4.62				

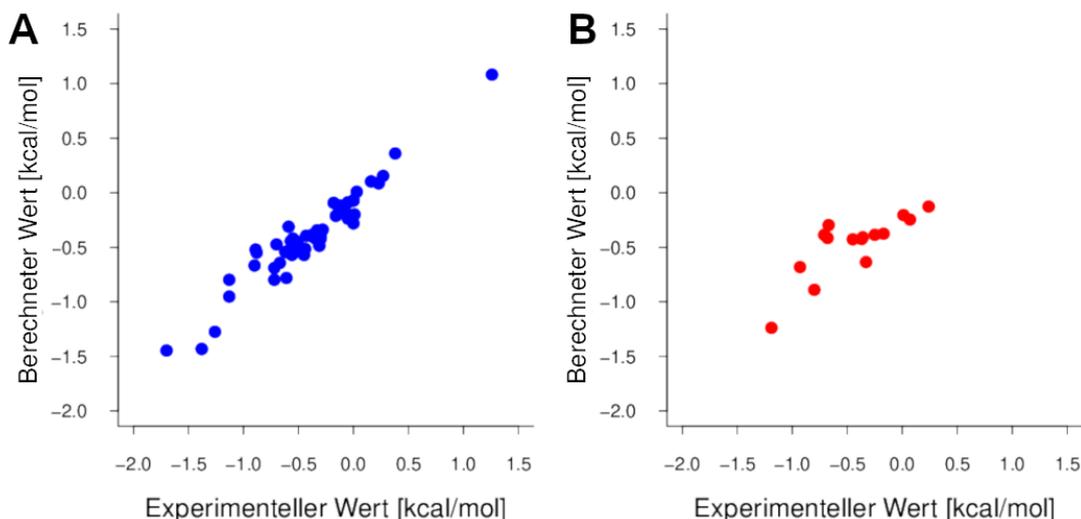


Abbildung 4.1: Auftragung der durch das Modell C1 vorhergesagten Differenzen der freien Faltungsentgien gegen die experimentellen Werte der (A) TS- und (B) HS-Konjugate.

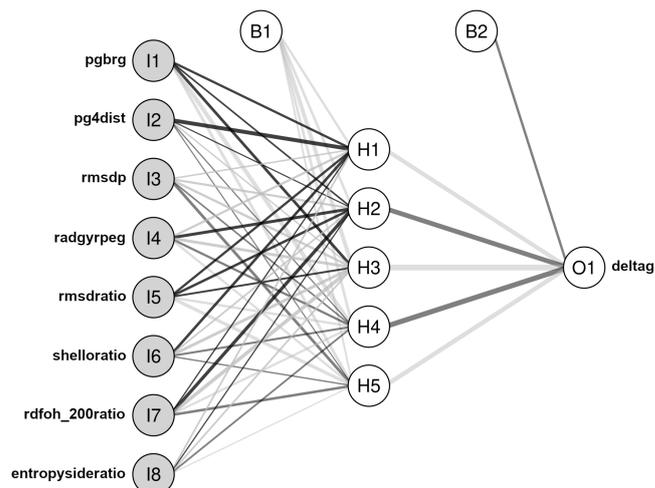


Abbildung 4.2: Visualisierung des Netzwerks von Modell D1. Die Neuronen der Eingabeschicht (*Input layer*, links) sind grau gefärbt, in der Mitte sind die Neuronen der verborgenen Schicht (H) und der Bias-Neuronen (B) dargestellt. Die Ausgabeschicht (*Output layer*, O) ist rechts zu sehen. Die Neuronen sind mit Linien vernetzt, deren Breite die Größe und Farbe die Vorzeichen der Gewichtungen symbolisieren (schwarz: +, grau: -).

4.1.1.3 Modelle E und F

Im Anschluss wurden weitere Modellgenerierungen mithilfe verschiedener Kombinationen der bedeutendsten Pin 1-unspezifischen Modelldeskriptoren aus den Kategorien A bis D getestet. Dabei wurden die in Ansatz 1 erfolgreichsten Modelltypen genutzt: Regressionen mithilfe des XGBoost-Ansatzes und neuronale Netze zur Klassifikation. Daraus ergaben sich ultimativ die Modelle E und F (Tab. 4.2, sowie Abb. 4.3 und 4.4).

Modell E zeigt im Vergleich zu C1 einen leicht verbesserten $RMSE_{HS}$ -Wert (0.21 kcal/mol), ein $RMSE_{TS}$ -Wert von nur 0.02 signalisiert allerdings auch, vergleichbar mit Modellen aus der Kategorie A, eine deutliche Überanpassung. Die Vorhersage von -0.63 kcal/mol für 1SRL.20-1 liegt im Bereich jener des Modells C1, doch auch hierbei wird das Konjugat bei der ADAN-Analyse einer erhöhten Kategorie 2 zugeordnet, was für eine unsichere Vorhersage spricht. 5 der 15 Konjugate des HS werden ebenso höheren ADAN-Kategorien zugeordnet (siehe Tab. D.4 für die Einteilungen pro Konjugat), was vergleichbar mit den Resultaten zu Modell C1 ist (4/15). Im Gegensatz zu letzterem stellt *shelloratio* für das Modell E den bedeutendsten Deskriptor dar.

Verglichen mit D1 zeigt Modell F eine etwas schlechtere Leistung bzgl. des TS, jedoch einen verbesserten CCR_{HS} -Wert von 0.87. Ebenso klassifiziert es 1SRL.20-1 richtig. Der wichtigste Deskriptor ist hierbei *radgyrpeg*. Eine Visualisierung des Modells zeigt, dass das Netzwerk aus einer verborgenen Schicht mit 5 Neuronen und 2 zusätzlichen Bias-Neuronen besteht (Abb. 4.4). In Relation zur Anzahl genutzter Deskriptoren stellt dies eine relativ große Anzahl dar. Zur Vermeidung einer möglichen Überanpassung sollte die Größe der verborgenen Schicht in der Regel kleiner als die Größe der Eingabeschicht (*Input layer*) sein [320].

Tabelle 4.2: Statistische Validierungsparameter für die Modelle E und F, sowie deren verwendete Deskriptoren (geordnet nach der Bedeutung für das Modell). Für das XGBoost-Modell wird der Gewinn (*Gain*) dargestellt. Dieser repräsentiert den Bruchteil des Beitrags dieses Deskriptors für alle im Prozess generierten Entscheidungsbäume (Summe = 1) [226]. Für das neuronale Netz wird eine Methode genutzt, die die Absolutwerte aller Gewichtungen im Netzwerk bzgl. der einzelnen Deskriptoren kombiniert [319].

Modell E									
#	Art	RMSE _{TS}	MAE _{TS}	CCC _{TS}	R ² _{TS}	RMSE _{HS}	MAE _{HS}	CCC _{HS}	R ² _{HS}
E	xgbTree	0.02	0.02	1.00	1.00	0.21	0.18	0.78	0.72
Deskriptor					Gain				
<i>shelloratio</i>					0.21				
<i>rmsdratio</i>					0.16				
<i>molbur</i>					0.13				
<i>pgbrg</i>					0.11				
<i>radgyrpeg</i>					0.11				
<i>rdfoh_200ratio</i>					0.11				
<i>protpeghbonds</i>					0.09				
<i>protprothbonds</i>					0.09				
Modell F									
#	Art	AUROC _{TS}		CCR _{TS}		AUROC _{HS}		CCR _{HS}	
F	nnet	0.85		0.75		0.91		0.87	
Deskriptor					Gewichtungen				
<i>radgyrpeg</i>					29.47				
<i>rmsdratio</i>					27.58				
<i>rdfoh_200ratio</i>					24.64				
<i>pgbrg</i>					18.31				

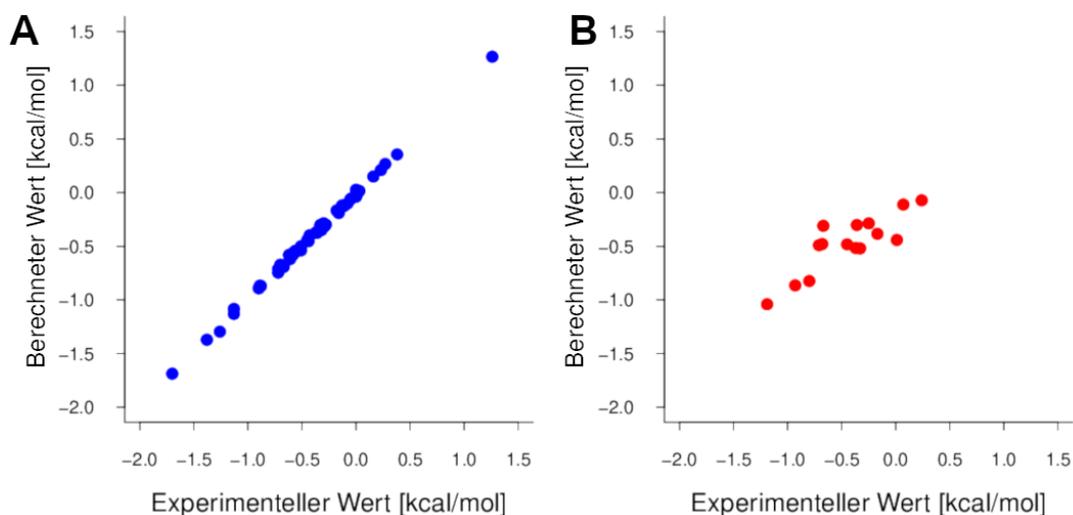


Abbildung 4.3: Auftragung der durch das Modell E vorhergesagten Differenzen der freien Faltungsennergien gegen die experimentellen Werte der (A) TS- und (B) HS-Konjugate.

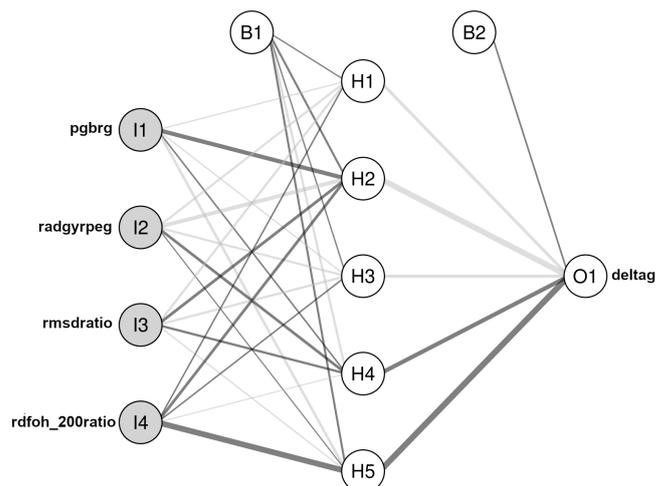


Abbildung 4.4: Visualisierung des Netzwerks von Modell F. Die Neuronen der Eingabeschicht (links) sind grau gefärbt, in der Mitte sind die Neuronen der verborgenen Schicht (H) und der Bias-Neuronen (B) dargestellt. Die Ausgabeschicht (O) ist rechts zu sehen. Die Neuronen sind mit Linien vernetzt, deren Breite und Farbe die Vorzeichen der Gewichtungen symbolisieren (schwarz: +, grau: -).

4.1.1.4 Modelle G und H

Aufgrund einer möglichen Überanpassung durch die Nutzung sehr komplexer Modellarten mit einem relativ kleinen Datensatz wurden die Deskriptoren der Modelle E und F weiterhin verwendet, um eine multivariate lineare Regression (Modell G) und eine binomiale logistische Regression (Modell H) zu generieren. Die Ergebnisse sind in Tab. 4.3 und in Abb. 4.5 gezeigt. Die deutlich schlechtere Leistung bzgl. des TS im Vergleich zu den Modellen E und F legt die Vermutung nahe, dass auch die komplexen Modelltypen dieses ersten Ansatzes außerhalb der Kategorien A und B zu einer Überanpassung neigen. Gleichzeitig zeigen die Modelle G und H eine bessere Leistung bzgl. des HS im Vergleich zum TS. Dies kann daraus resultieren, dass die Modelle bei diesem ersten Ansatz auf Basis der Leistung bzgl. des HS selektiert wurden.

Beide erwähnten Probleme, die mit einer hohen Modellkomplexität einhergehende Überanpassung und der Bias bzgl. des HS, sollen daher in Ansatz 2 mit einem erweiterten Kreuzvalidierungsverfahren gelöst werden (siehe Kap. 4.1.2).

4.1.1.5 Leistung der Modelle

Die in diesem ersten Ansatz generierten Pin 1-spezifischen Modelle aus den Kategorien A und B nutzen eine Vielzahl an Deskriptoren und zeigen damit ein komplettes Overfitting. Zu erwähnen sind lediglich die Pin 1-unspezifischen Modelle der Kategorien C, D, E und F, die alle überzeugende Statistiken aufweisen. Dennoch signalisieren die Leistungen der Modelle G und H, dass es auch in diesen Kategorien, v. a. bei den Modellen E und F, zu einer gewissen Überanpassung kommt. Somit sind die statistischen Werte kritisch zu sehen.

Tabelle 4.3: Statistische Validierungsparameter für die Modelle G (Art *lm* = lineare Regression) und H (Art *glm* = logistische Regression), sowie deren verwendete Deskriptoren (geordnet nach der Bedeutung für das Modell, mit Angabe der Regressionskoeffizienten für den skalierten Datensatz). Die Bedeutung basiert auf den t- (Modell G) bzw. z-Werten (Modell H) der einzelnen Koeffizienten, die aus dem Verhältnis des jeweiligen Koeffizienten und des dazugehörigen Standardfehlers resultieren.

Modell G									
#	Art	RMSE _{TS}	MAE _{TS}	CCC _{TS}	R ² _{TS}	RMSE _{HS}	MAE _{HS}	CCC _{HS}	R ² _{HS}
G	lm	0.34	0.27	0.64	0.47	0.27	0.23	0.76	0.58
Deskriptor					Koeffizient				
<i>pgbrg</i>					0.24				
<i>rdfoh_200ratio</i>					-0.11				
<i>molbur</i>					-0.21				
<i>protprothbonds</i>					-0.08				
<i>rmsdratio</i>					-0.06				
<i>protpeghbonds</i>					-0.09				
<i>radgyrpeg</i>					-0.05				
<i>shellohratio</i>					-0.04				
Modell H									
#	Art	AUROC _{TS}	CCR _{TS}	AUROC _{HS}	CCR _{HS}				
H	glm	0.73	0.71	0.93	0.93				
Deskriptor					Koeffizient				
<i>radgyrpeg</i>					1.54				
<i>pgbrg</i>					-13.50				
<i>rdfoh_200ratio</i>					1.81				
<i>rmsdratio</i>					14.24				

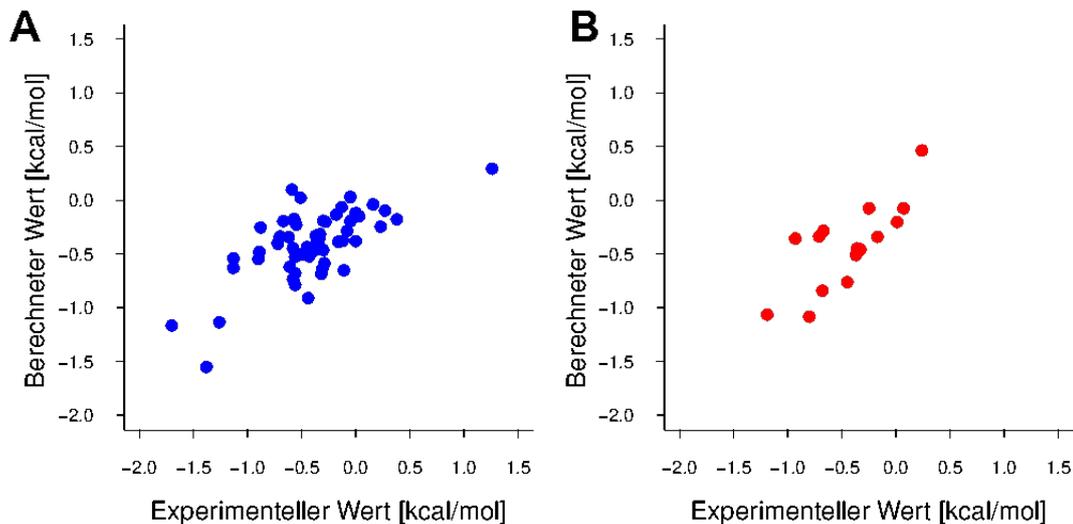


Abbildung 4.5: Auftragung der durch das Modell G vorhergesagten Differenzen der freien Faltungsenergien gegen die experimentellen Werte der (A) TS- und (B) HS-Konjugate.

Ungeachtet dessen enthielten die Modelle aus den Kategorien C, D, E und F Informationen von PBK mit diversen Linkerstrukturen und konnten dennoch teilweise bessere Stabilitätsvorhersagen erbringen als die von Price et al. postulierte, kristallstrukturbasierte Methode. Diese basierte lediglich auf einem einzigen Deskriptor und einer kleinen Auswahl an 10 Asn(PEG)₄-Konjugaten. ADAN-Analysen deuten jedoch auf große Unsicherheiten bzgl. der Vorhersage zu 1SRL.20-1 und damit eine

limitierte Anwendbarkeit auf andere Proteine hin. Im Vergleich zum Referenzmodell, welches jedem HS-Konjugat den Medianwert des TS von -0.37 kcal/mol zuweisen würde ($\text{RMSE}_{\text{HS}} = 0.39$ kcal/mol), zeigen die erstellten Regressionsmodelle bessere Leistungen. Auch das Referenzmodell zur Klassifikation zeigt mit einem CCR_{HS} -Wert von 0.67 eine geringere Leistung als die Modelle D1, F und H.

4.1.1.6 Verwendete Deskriptoren

Die Auswahl der Deskriptoren *shelloratio*, *rdfoh_200ratio* und *rdfoh_400ratio* durch die Modelle aus den Kategorien C und D unterstützen die Hypothese von Price et al., dass PEG vornehmlich die Solvatation von Hydroxylgruppen moduliert [100].

Während solche Solvatationseffekte zu entropisch günstigen Wasserverdrängungen auf der Proteinoberfläche führen können, beschreibt der wichtigste Deskriptor von Modell C1, *protpeghbonds*, die Anzahl an Protein-PEG-HB und damit eher spezifische, enthalpisch günstige Wechselwirkungen. Dieser Parameter zeigt auch eine schwache Korrelation (Abb. 4.6-A) mit der Antwortvariablen. Ein weiterer HB-Deskriptor, der die Protein-Protein-Wechselwirkungen beschreibt (*protprothbondsratio*), zeigt, dass PEG durch eine Abschirmung möglicherweise intramolekulare HB stärkt (vgl. [94]).

Das Modell E gewichtet hingegen den Solvatationsparameter *shelloratio* deutlich stärker. Weiterhin nutzt es den Deskriptor *molbur*, mit welchem eine schwache Korrelation der bedeckten Proteinoberfläche mit der Differenz der freien Faltungsenthalpie festgestellt werden kann (Abb. 4.6-B). Modell F nutzt den Deskriptor *radgyrpeg*, was vornehmlich auf doppelt PEGylierte Konjugationsstellen zurückzuführen ist, da hierbei alle PEG-Atome zur Berechnung berücksichtigt wurden. Dies spiegelt jedoch auch die Hypothese von Price et al. wieder, dass eine weitere PEGylierung an einer weit entfernten Position einen synergistischen Effekt haben kann (16+26-1, 19+26-1), während bei einer zweiten PEGylierung in unmittelbarer Nähe die Stabilisierung durch beide Konjugationen eher beeinträchtigt wird (26+29-1, 16+19-1). In letzteren Fällen adressieren beide Ketten ähnliche Areale [100].

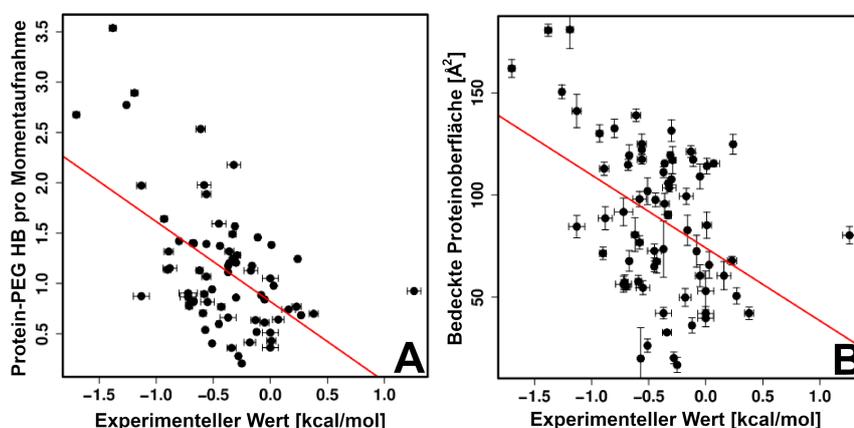


Abbildung 4.6: Lineare Regressionen ausgewählter Deskriptoren. Dargestellt sind *protpeghbonds* (A, $R^2 = 0.31$) und *molbur* (B, $R^2 = 0.18$).

4.1.1.7 MD-Analysen

Zur weiteren Interpretation wurden relevante Deskriptoren aller Modellkategorien in exemplarischen Simulationen näher untersucht. Dadurch soll auch die Eignung dieser berechneten Deskriptoren zur Wiedergabe der von Price et al. postulierten Effekte einer PEG-basierten Konjugation demonstriert werden. Diese Analysen sollen im Folgenden vorgestellt werden.

4.1.1.7.1 Interaktionen und Solvation

16-1 (-0.90 ± 0.03 kcal/mol) und 27-1 (0.38 ± 0.04 kcal/mol) stellen stark stabilisierte (16-1) bzw. stark destabilisierte (27-1) Varianten dar. Im Falle von 16-1 wird mehr Proteinoberfläche durch PEG bedeckt ($molbur = 71.42 \pm 3.08 \text{ \AA}^2$ für 16-1, $42.13 \pm 3.09 \text{ \AA}^2$ für 27-1) und die Anzahl an Protein-PEG-HB ist entsprechend höher ($protpeghbonds = 1.14 \pm 0.01$ für 16-1, 0.70 ± 0.01 für 27-1).

Beides kann in den Simulationen über die Zeit hinweg beobachtet werden (Abb. 4.7 und 4.8): Im Falle von 27-1 interagiert der Asparagin-Linker mit E12, während die PEG-Kette ins Solvens hinausragt. Teilweise adressiert diese den N-Terminus oder das Rückgrat von G10. Am Ende der Simulation ist 27-1 zwischen G10 und I28 positioniert, was zu einer Erhöhung der bedeckten Proteinoberfläche führt. Im Gegensatz hierzu wendet sich das Polymer von 16-1 nur sehr selten von der Oberfläche ab. Es wechselwirkt entweder mit S32 und F34 des dritten β -Faltblattes oder mit Aminosäuren der ersten Schleife (R17, S18, S19, R21).

Um die Solvation der Proteine zu vergleichen, wurden RDF-Differenz-Plots zwischen der PEGylierten und der unPEGylierten Mutante erzeugt (Abb. 4.8). Eine Desolvation ist deutlicher bei der stabilisierten Variante 16-1 zu erkennen, v. a. um die Aminosäuren R17, S19, Y23 und F34. Im Falle des destabilisierten Konjugats 27-1 erhöht sich der RDF-Wert um G20. Analoge Plots weiterer stark (de-)stabilisierter PBK zeigen einen ähnlichen Trend (Abb. D.1). Auch wenn die Solvation von G20 maßgeblich durch die Konformation der ersten Schleife bestimmt wird, spielte der Deskriptor *glycinflip* (ϕ -Diederwinkel von G20, dessen Änderung eine Umorientierung signalisieren kann) keine bedeutende Rolle für die Pin 1-spezifischen Modelle. Eine Umorientierung bzw. ein Flip von G20 wurde in einigen Simulationen visuell beobachtet (Abb. D.2).

Durch diesen beispielhaften Vergleich zwischen 16-1 und 27-1 werden die Unterschiede in der Desolvation und den Deskriptoren *protpeghbonds* und *molbur* aufgezeigt und auf molekularer Ebene veranschaulicht.

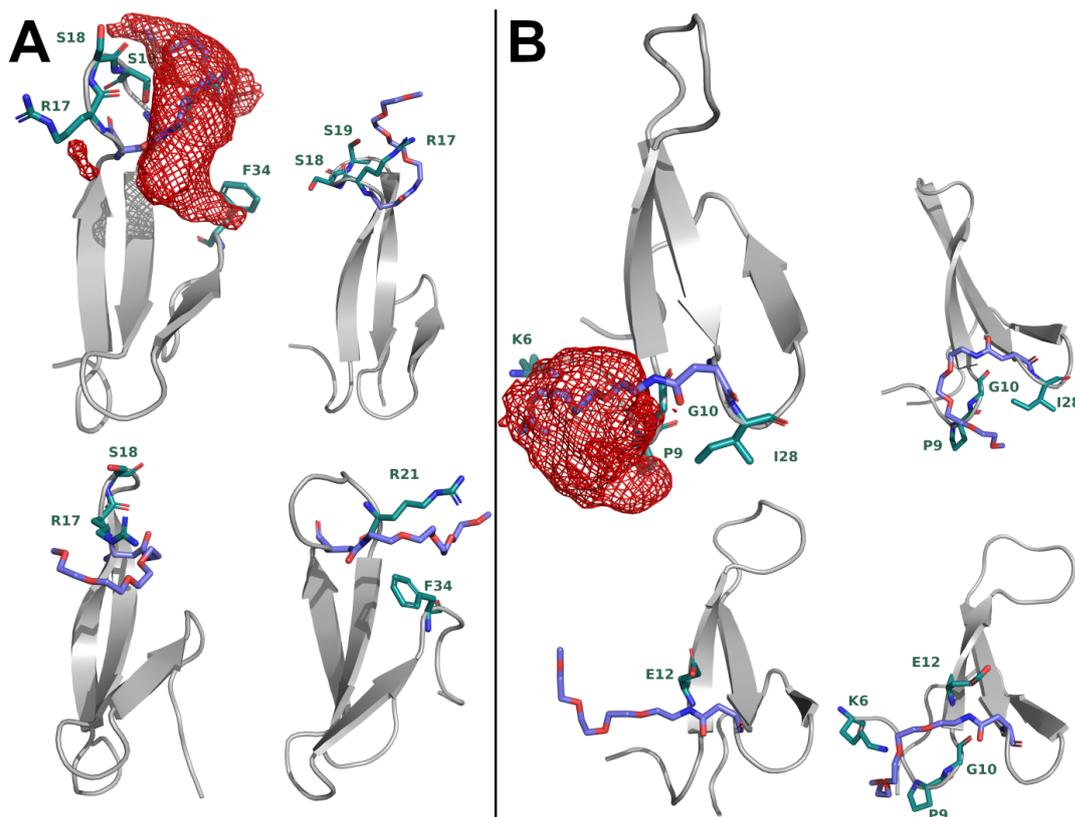


Abbildung 4.7: Momentaufnahmen aus den Simulationen von (A) 16-1 und (B) 27-1. In der ersten Momentaufnahme, der Startposition, sind Dichten der PEG-Atome (Isowert: 0.02) für die gesamte Simulation mit roten Gittern illustriert. Drei kleinere Darstellungen zeigen jeweils häufig beobachtete bzw. repräsentative Konformationen des Polymers in den Simulationen. Aminosäuren in der Nähe sind grün dargestellt und gekennzeichnet, die Konjugationsstelle ist in violett gezeigt.

4.1.1.7.2 Linker verschiedener Länge

Eine weitere Hypothese von Price et al. besagt, dass rigide, planare Linker mit einer gewissen Distanz für eine optimale Desolvatation benötigt werden [101]. Asparagin positioniert PEG durch die geringere Länge näher an die Hydrathüllen des Proteins als Glutamin. Zum Vergleich wurden die Konjugate 29-1 (-0.36 ± 0.04 kcal/mol) und 29-9 (0.16 ± 0.06 kcal/mol) untersucht. Beide PEG-Ketten sind auf der unteren Seite der β -Faltblätter lokalisiert, aber im Falle von 29-9 weiter weg vom Protein positioniert. Dies wird durch den Deskriptor *pg4dist* deutlich (2.99 ± 0.21 Å für 29-1, 3.71 ± 0.24 Å für 29-9). Die Aminosäuren von 29-1 sind stärker desolvatisiert (Abb. 4.9), v. a. W11, A31, Q33, P37 und S38.

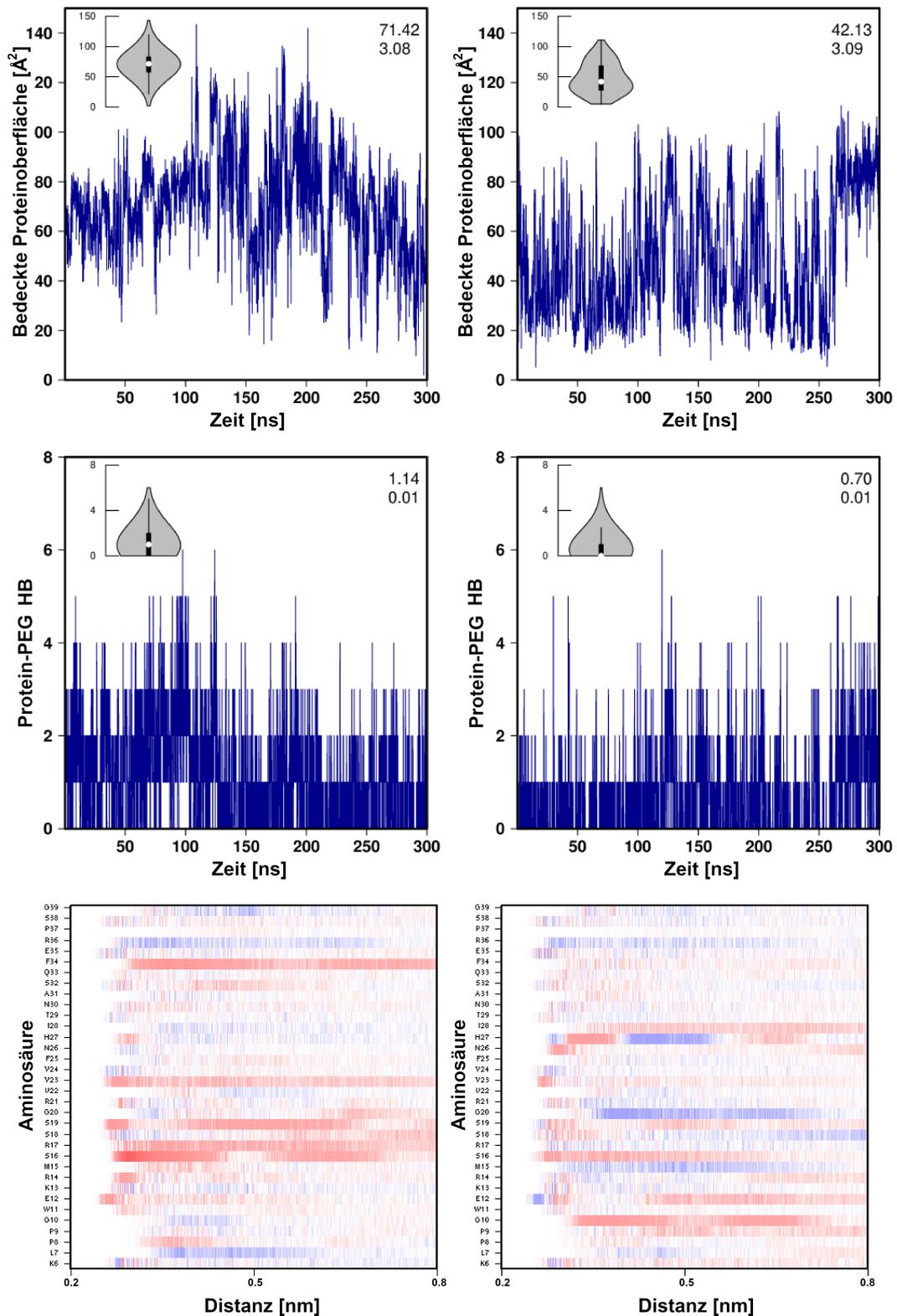


Abbildung 4.8: Vergleich zwischen den Simulationen von 16-1 (links) und 27-1 (rechts). Die erste Reihe zeigt die durch das Polymer bedeckte Proteinoberfläche über die Zeit, während die zweite die Anzahl an beobachteten Protein-PEG-HB darstellt. Zusätzlich dargestellt sind jeweils Violin-Plots. Die dazugehörigen Deskriptorwerte und deren Standardfehler sind jeweils oben rechts dargestellt. Die letzte Reihe zeigt RDF-Differenz-Plots (PEGyliert - unPEGyliert) für jede Aminosäure zwischen 0.2 und 0.8 nm. Blaue Areale (Werte zwischen -0.1 und -2) repräsentieren eine stärkere Solvatation der PEGylierten Variante, rote eine der unPEGylierten (Werte von 0.1 bis 2).

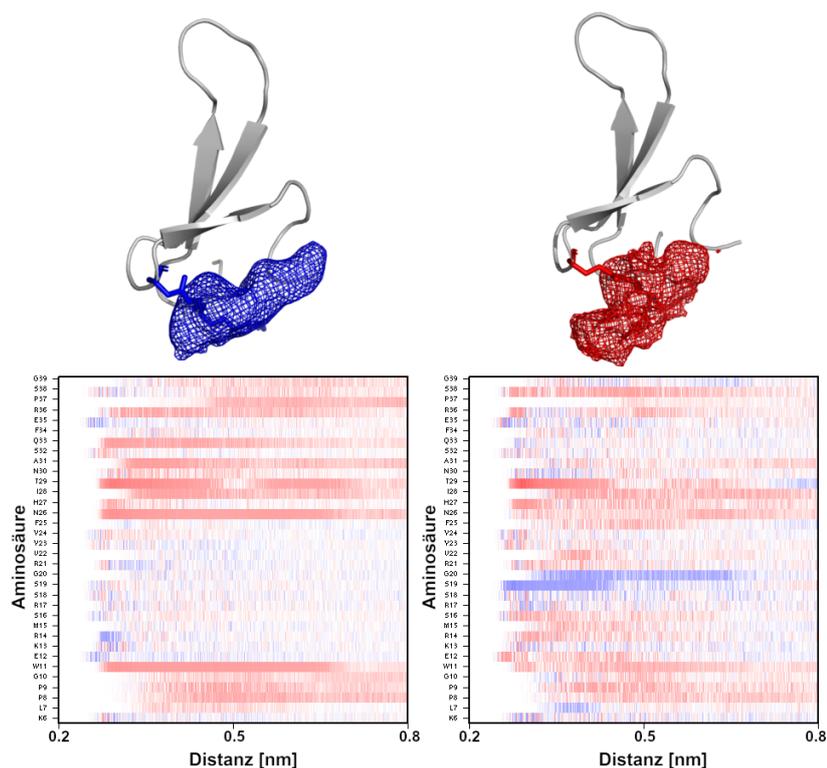


Abbildung 4.9: Darstellung von Dichten der PEG-Atome für 29-1 (blau) und 29-9 (rot), analog zu Abb. 4.7. Darunter sind die RDF-Differenz-Plots dargestellt (links: 29-1, rechts: 29-9).

4.1.1.7.3 D-/L-Asparagin

Ein weiterer Vergleich kann zwischen den Varianten 19-1 (-0.70 ± 0.04 kcal/mol) und 19-8 (0.01 ± 0.04 kcal/mol) hergestellt werden. 19-8 verfügt über ein D-Asparagin anstatt eines L-Asparagins als Linker. Price et al. schlussfolgerten, dass die Orientierung der PEG-Kette durch die Änderung der Linkerkonfiguration entscheidend verändert wird: die PEG-Kette sollte bei 19-8 vom Protein abgewandt sein. In den entsprechenden Simulationen orientiert sich das Polymer von 19-1 in Richtung der Aminosäuren von Schleife 1, was zu einer Desolvatation um M15, S16 und G20 führt (Abb. 4.10). Im Gegensatz hierzu wechselwirkt PEG bei 19-8 mit S32 und F34, das Polymer ist also nicht von der Proteinoberfläche abgewandt.

Vorherige Simulationen von Price et al. deuteten auf eine Stabilisierung der Schleifenkonformation im Falle von 19-1 hin [99]. In den hier durchgeführten Simulationen kommt es auch nur im Falle von 19-8 zu einer deutlichen Konformationsänderung der ersten Schleife. Vielleicht ist diese Änderung auf die Integration der D-Aminosäure zurückzuführen. Allerdings sollte angemerkt werden, dass in Kraftfeldern für MD-Simulationen (in diesem Falle AMBERff14SB) prinzipiell keine von den L-Konfigurationen abweichenden Parameter für die D-Varianten genutzt werden, sodass hier lediglich die Startkonformation der Konjugationsstelle zu Unterschieden in den Simulationen geführt haben dürfte. Da die Polymerkette im Falle von 19-8 zur Proteinoberfläche geneigt ist, zeigt diese Variante eine größere bedeckte Proteinoberfläche

($molbur = 55.04 \pm 1.49 \text{ \AA}^2$ für 19-1, $85.19 \pm 6.42 \text{ \AA}^2$ für 19-8). Diese Beobachtung zeigt, dass eine von Price et al. postulierte Abkehr der PEG-Kette vom Protein durch eine D-Konfiguration des Asparagins nicht bestätigt werden kann. Allerdings ist die Anzahl an Protein-PEG-HB im Falle von 19-1 tatsächlich höher ($protpeghbonds = 0.81 \pm 0.01$ für 19-1, 0.43 ± 0.01 für 19-8), was eine Destabilisierung von 19-8 im Vergleich zu 19-1 begründen könnte.

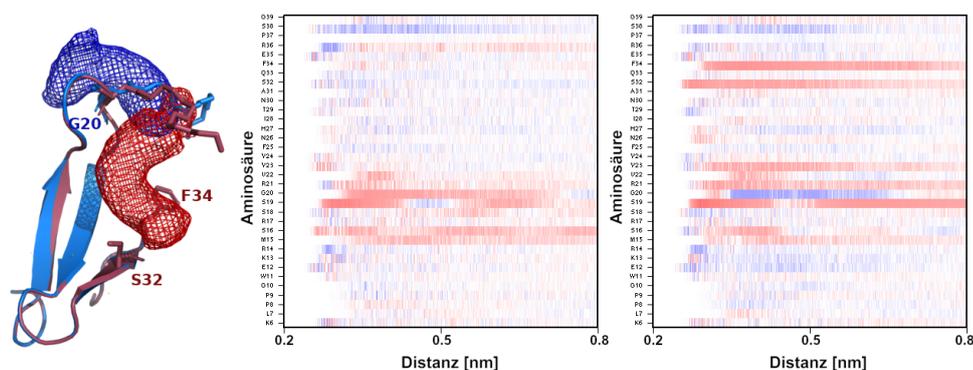


Abbildung 4.10: Darstellung der PEG-Dichten für 19-1 (rot) und 19-8 (blau), analog zu Abb. 4.7. Daneben sind die RDF-Differenz-Plots dargestellt (links: 19-1, rechts: 19-8).

4.1.1.7.4 Linker verschiedener Rigidität

Zum Vergleich eines rigiden mit einem flexibleren Linker wurden die Trajektorien von 19-1 ($-0.70 \pm 0.04 \text{ kcal/mol}$) und 19-4 ($0.24 \pm 0.02 \text{ kcal/mol}$) verglichen (Abb. 4.11). 19-4 enthält einen sterisch anspruchsvolleren Cystein-Maleimid-Linker. Bei diesem stark destabilisierten Konjugat ist eine deutliche Konformationsänderung der ersten Schleife zu erkennen. M15 orientiert sich hierbei weg von V22, letztere Aminosäure ist dadurch stark solvens exponiert. Ein deutlich höherer Wert für $rmsfp$ wurde bestimmt (25.52 \AA für 19-1, 93.84 \AA für 19-4).

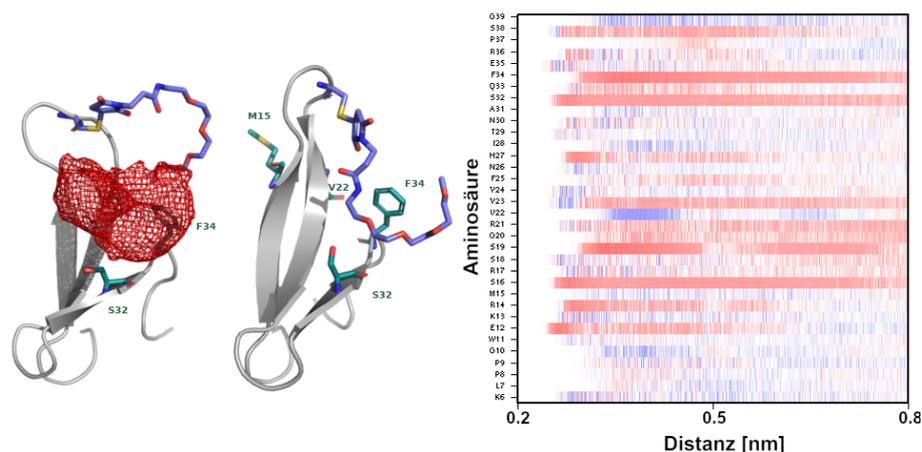


Abbildung 4.11: Darstellung der PEG-Dichte für 19-4, analog zu Abb. 4.7. Daneben zeigt eine weitere Momentaufnahme die veränderte Ausrichtung von M15. Rechts ist der entsprechende RDF-Differenz-Plot gezeigt.

4.1.1.7.5 Desolvation apolarer Flächen

Zuletzt wurde zur Untersuchung der Wasserverdrängung um solvensexponierte apolare Aminosäuren [103] die Simulation von 23-Z (-0.29 ± 0.03 kcal/mol) mit jenen der beiden Mutanten 23(E12L:R14L)-Z (-0.89 ± 0.04 kcal/mol) und 23(E12F:R14F)-Z (-0.33 ± 0.02 kcal/mol) verglichen (Abb. 4.12). Letztere besitzen jeweils zwei zusätzliche apolare Aminosäuren im Bereich des ersten β -Faltblatts. Gemäß der Hypothese einer Desolvation solvensexponierter hydrophober Areale ist die PEG-Kette im Falle der beiden Mutanten eher zum ersten β -Faltblatt orientiert. Der Parameter *rdfapol_200ratio*, der die Solvation in der ersten Hydrathülle um apolare Aminosäuren beschreibt, signalisiert eine Desolvation dieser Mutanten (0.774 ± 0.002 für 23(E12L:R14L)-Z, 0.884 ± 0.003 für 23(E12L:R14F)-Z).

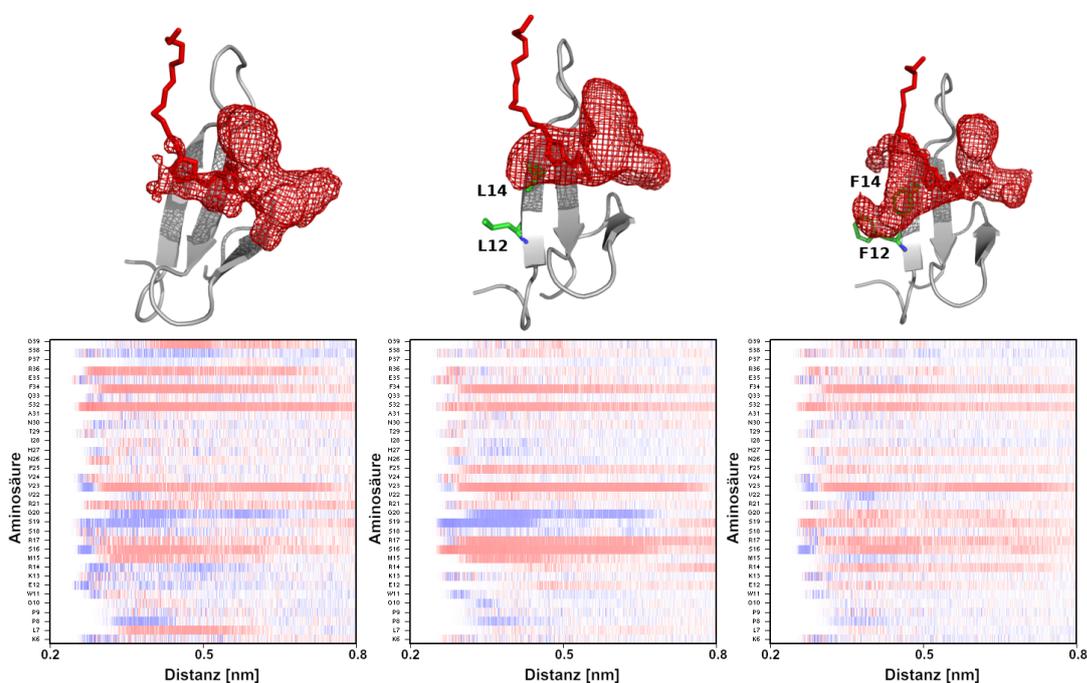


Abbildung 4.12: Darstellung der PEG-Dichten und der dazugehörigen RDF-Differenz-Plots für 23-Z (links), 23(E12L:R14L)-Z (Mitte) und 23(E12F:R14F)-Z (rechts).

4.1.1.8 Fazit

Die Pin 1-unspezifischen Regressionsmodelle C1 und E lieferten zu Price et al. vergleichbare Vorhersagen für 1SRL.20-1 und die Klassifikationsmodelle D1 und F zufriedenstellende statistische Werte, ADAN-Analysen und die Modelle G und H deuten jedoch auf Vorhersagen mit erhöhten Fehlern und mögliche Überanpassungen hin. Die Modelldeskriptoren unterstreichen die Bedeutung von Solvenseffekten durch eine PEGylierung, berücksichtigen aber auch enthalpisch günstige Wirkungen, bspw. über die Anzahl an Protein-PEG-HB. Verschiedene Hypothesen von Price et al. und damit zusammenhängende, bedeutende Modelldeskriptoren wurden in exemplarischen Trajektorien verschiedener Konjugate untersucht, was eine Eignung klassischer MD-Simulationen zur Wiedergabe der postulierten Effekte unterstreicht.

Zur weiteren Optimierung und einer Vermeidung von Überanpassungen ist jedoch ein deutlich größerer Datensatz an Konjugaten zur Nutzung der hier angewandten Methoden des maschinellen Lernens nötig. Ein solcher Datensatz sollte eine vergleichbare Anzahl α -helikaler und β -faltblattreicher Strukturen enthalten (zur Etablierung generalisierbarer Vorhersagemodelle). Teilweise enthielt der hier genutzte Datensatz Linkerstrukturen, die nur in einem einzelnen Konjugat vorkamen und im HS vertreten waren (z. B. 19-J, 19-4 oder 19-5). Allerdings würde der Datensatz durch eine Restriktion auf Asparagin-Linker-basierte Konjugate deutlich verringert werden. Zusätzlich sollte erwähnt werden, dass der Effekt auf die Stabilität des Konjugats natürlich auch aus einer Veränderung der Linkerstruktur während der PEGylierung resultieren kann. So existieren gewisse Konjugationsstrategien (z. B. bei den Konjugaten mit den Kennzeichen Z oder X), die zu einer deutlichen strukturellen Modifizierung führen. So wird nicht nur der Effekt der PEGylierung berücksichtigt.

Wie am Anfang dieses Kapitels bereits angemerkt, wurden diese Modelle auf Basis eines einzigen HS selektiert, und es kamen sehr komplexe Modellarten zum Einsatz, die in der Regel nur bei größeren Datensätzen angewandt werden sollten, um eine Überanpassung aufgrund der Hyperparameter zu vermeiden. Die erhaltenen statistischen Werte der beschriebenen Modelle sind daher kritisch zu sehen. Der beschriebene Ansatz dient jedoch als Grundlage für den angepassten zweiten Ansatz (*vide infra*). Weiterhin kann er als Grundlage für mögliche zukünftige Arbeiten dienen, auf die die in diesem Kapitel erläuterten Modelltypen (neuronale Netze und XGBoost-Modelle) angewandt werden könnten.

4.1.2 Ansatz 2

Im Folgenden sollen das beste Regressions- und das beste Klassifikationsmodell aus Ansatz 2 vorgestellt werden. In diesem Ansatz wurde ein erweitertes Kreuzvalidierungsverfahren genutzt, sodass das zugehörige HS im Gegensatz zu Ansatz 1 keinen Einfluss auf die Modellauswahl hatte und es sich somit um eine externe Validierung handelt. Weiterhin wurden lediglich multivariate lineare Regressions- und binomiale logistische Klassifikationsmodelle erstellt. Gemäß Kap. 3.1.6.3 wurden hierbei 9 verschiedene Datensätze untersucht (siehe Tab. 3.3), die sich in der Art und der Anzahl der Konjugate, sowie der Nutzung Pin 1-spezifischer und -unspezifischer Deskriptoren unterschieden. Die statistischen Parameter der finalen Modelle zu jedem Datensatz sind in den Tab. D.6 und D.7 zu finden. Hier sollen das Regressionsmodell zum Datensatz 6 (genannt A2-R6) und das Klassifikationsmodell zu Datensatz 1 (genannt A2-C1) mit den Korrelationslimits von 0.99 vorgestellt werden, die die jeweils besten Ergebnisse erzielten.

Die Modelle der anderen Konjugate zeigen eher mäßige Leistungen und sollen hier nicht näher diskutiert werden. Dies ist vermutlich auf die Zusammensetzungen der jeweiligen Datensätze zurückzuführen. Der von Price et al. experimentell ermittelte Effekt von Cyclohexylalanin- und Azidohomoalanin-Mutationen (siehe Kap. 3.1.3.4)

in den Datensätzen 2 und 3 konnte durch die Simulationen evtl. nicht korrekt abgebildet werden. Weiterhin stellen die α -helikalen Strukturen der Doppelwendeln in Datensatz 2 von den WW-Domänen strukturell deutlich abweichende Peptide dar. Hierbei wird der Effekt der PEGylierung auf Di- und Trimere berücksichtigt. Die Datensätze 4 und 5 bestehen wiederum grundsätzlich aus einer sehr geringen Anzahl an Konjugaten. Dies führt unter anderem dazu, dass Regressionsmodelle zu Datensatz 5 bzgl. des HS mit nur vier Konjugaten zwar recht hohe R^2 -Werte zeigen (Tab. D.6), diese jedoch lediglich auf jeweils zwei im Hinblick auf ihre Faltungsenergie nah beieinander liegenden stark stabilisierten und destabilisierten Varianten beruhen. Die Leistungen für das TS sind deutlich schlechter. Für Datensätze 8 und 9 wurden proteinspezifische Deskriptoren genutzt, welche nicht für generalisierte Modelle verwendet werden können und auch in Ansatz 1 bereits zu unbrauchbaren Modellen führten.

4.1.2.1 Referenzmodelle

Wie auch in Ansatz 1 sollten zunächst die Leistungen von Referenzmodellen für die jeweiligen Datensätze untersucht werden. Ein Modell, welches für jedes Konjugat des TS von Datensatz 6 den Median- oder den Mittelwert vorhersagen würde, zeigt einen $RMSE_{TS}$ -Wert von 0.44 kcal/mol und einen $RMSE_{HS}$ -Wert von 0.53 kcal/mol.

Für das Klassifikationsmodell A2-C1 wurden zunächst die Anteile an Konjugaten der stabilisierten Klasse 1 (61 von 72, bzw. 84.72 %) und der un- bzw. destabilisierten Klasse 2 (11 von 72, bzw. 15.28 %) ermittelt. Bei einer zufälligen Einteilung der TS- und der HS-Konjugate des Datensatzes 1 in eine der beiden Klassen mit den genannten Wahrscheinlichkeiten würden ein CCR_{TS} -Wert von 0.74, ein $BACC_{TS}$ -Wert von 0.44, ein CCR_{HS} -Wert von 0.64 und ein $BACC_{HS}$ -Wert von 0.44 erhalten werden.

4.1.2.2 Regressionsmodell A2-R6

4.1.2.2.1 Allgemeine Leistung

Die Leistung des Regressionsmodells A2-R6 ist in Tab. 4.4 für das TS und das HS zusammengefasst. Eine Illustration für das TS und das HS ist in Abb. 4.13 dargestellt. Das Modell erzielt mit einem $RMSE_{TS}$ -Wert von 0.25 kcal/mol, einem R^2_{TS} -Wert von 0.68 und einem CCC_{TS} -Wert von 0.81 eine gute Leistung für das TS. Der Wert für $\tau_{B(TS)}$ ist jedoch mit 0.45 sehr gering. Dies resultiert aus der hohen Anzahl an Konjugaten im Bereich zwischen etwa -0.70 und -0.20 kcal/mol. Die Fähigkeit des Modells, diese Varianten in einer sehr dichten „Punktwolke“ korrekt zu ordnen, ist, wie im ersten Ansatz, sehr schlecht. Dieser enge Bereich repräsentiert einen für QSPR-Modelle nur limitiert zugänglichen Raum. Die Leistung ist für das HS etwas geringer, aber immer noch akzeptabel ($RMSE_{HS}$ -Wert von 0.37 kcal/mol und CCC_{HS} -Wert von 0.74). Von entscheidender Bedeutung ist, dass das Modell extrem stabilisierte und destabilisierte Varianten, welche aufgrund des bei der Erstellung angewandten Cluster-Ansatzes (siehe Kap. 3.1.6.3) im HS überrepräsentiert sind, korrekt einordnet. Durch deren Überrepräsentation zeigen R^2_{HS} und $\tau_{B(HS)}$ sogar bessere Werte

als im TS. Sowohl im Hinblick auf das TS (0.25 kcal/mol) als auch das externe Validierungsset (0.37 kcal/mol) zeigt das Modell A2-R6 bessere RMSE-Werte als das zu Beginn beschriebene Referenzmodell ($\text{RMSE}_{\text{TS}} = 0.44$ kcal/mol, $\text{RMSE}_{\text{HS}} = 0.53$ kcal/mol).

Tabelle 4.4: Statistische Validierungsparameter für das Modell A2-R6, sowie dessen verwendete Deskriptoren (geordnet nach der Bedeutung für das Modell, mit Angabe der Regressionskoeffizienten für den skalierten Datensatz). Die Bedeutung basiert auf den t-Werten der einzelnen Koeffizienten, die aus dem Verhältnis des jeweiligen Koeffizienten und des dazugehörigen Standardfehlers resultieren.

Modell A2-R6							
RMSE_{TS}	$\tau_B(\text{TS})$	CCC_{TS}	R^2_{TS}	RMSE_{HS}	$\tau_B(\text{HS})$	CCC_{HS}	R^2_{HS}
0.25	0.45	0.81	0.68	0.37	0.82	0.74	0.76
Deskriptor				Koeffizient			
<i>molbur</i>				-0.96			
<i>rdf_500ratio</i>				1.76			
<i>rdfpeg_500</i>				-0.67			
<i>rdf_700ratio</i>				-0.83			
<i>pgbrg</i>				0.20			
<i>shellpego</i>				0.50			
<i>shellpolratio</i>				0.13			
<i>rdf_400ratio</i>				-0.81			
<i>rdf_200ratio</i>				-0.29			
<i>pegangle</i>				-0.08			

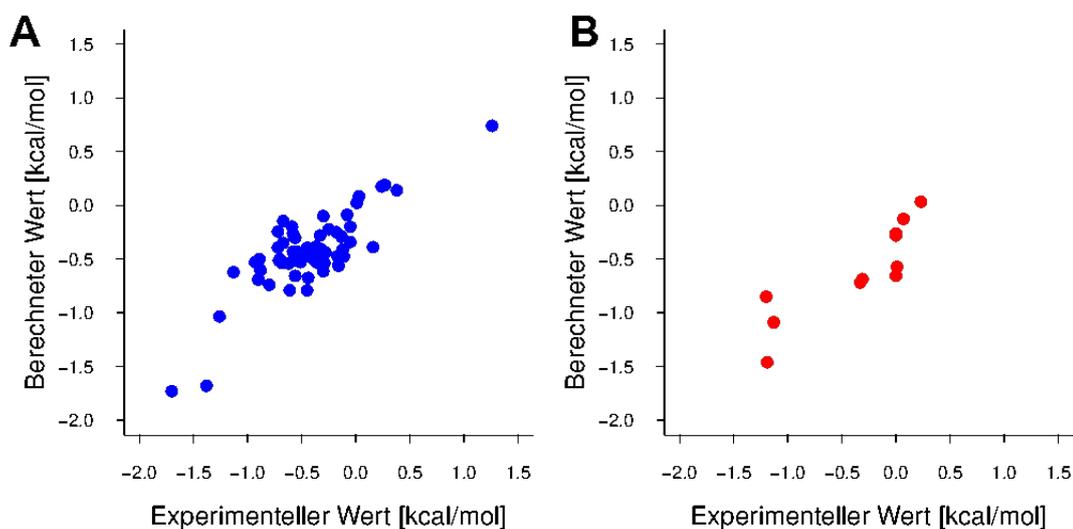


Abbildung 4.13: Auftragung der durch das Modell A2-R6 vorhergesagten Differenzen der freien Faltungsenthalpien gegen die experimentellen Werte der (A) TS- und (B) HS-Konjugate.

4.1.2.2.2 Einzelne Konjugate

Die Vorhersagen zu den einzelnen Konjugaten des HS sind in Tab. 4.5 aufgelistet. Die entsprechenden ADAN-Kategorien sollen weiter unten diskutiert werden. Die drei besonders stabilisierten Konjugate 19-2, 19-3 und 1SRL.20-1 ($\Delta\Delta G_f^0 < -1.00$ kcal/mol) werden vom Modell korrekterweise als die stabilisiertesten Varianten im HS eingestuft. Ebenso wird ein positiver Wert für die destabilisierte Variante 23-9 berechnet.

Diese Beobachtungen deuten darauf hin, dass das Modell prinzipiell gut geeignet ist, deutlich (de-)stabilisierte Konjugate zu detektieren. Schaut man sich die Differenzen zu den experimentellen Werten an (Tab. 4.5), so zeigt sich, dass das Modell den Stabilisierungseffekt der Konjugate generell überbewertet. Dies führt dazu, dass die Varianten 18-1, 28-1 und 30-1 als schwach stabilisiert eingestuft werden, obwohl deren PEGylierungen auf Basis der experimentellen Befunde keinen Effekt auf die thermische Stabilität ausüben. Unter Berücksichtigung der von der ADAN-Analyse ausgegebenen 95%-Konfidenzintervalle und den experimentellen Fehlern stimmen die Vorhersagen des Modells für die Konjugate des HS mit den Messungen überein, die ADAN-Konfidenzintervalle für 28-1, 16-Z und 32-Z decken jedoch auch einen Großteil des Energiebereichs, in welchem die experimentellen Befunde zu verorten sind, ab. Im Falle eines kleineren Konfidenzintervalls von ± 0.25 kcal/mol gäbe es keine Übereinstimmung bei 32-Z.

Tabelle 4.5: Vorhersagen von Modell A2-R6 zu den einzelnen Konjugaten des HS. Dazu ist die Differenz zum experimentell ermittelten Wert aufgelistet. Die Vorhersagen sind mit den 95%-Konfidenzintervallen versehen, die aus der ADAN-Kategorisierung resultieren. Dabei führen niedrige Kategorien (0 - 1, grün) zu einem geringen Wert (± 0.25 kcal/mol) und höhere (2 - 3, rot) zu einem größeren (± 0.50 kcal/mol). Die Differenzen sind grün gefärbt, wenn das ADAN-Konfidenzintervall den Bereich des exp. Werts einschließt, andernfalls sind sie rot gefärbt.

Konjugat	Experimenteller Wert [kcal/mol]	Vorhersage [kcal/mol]	Differenz [kcal/mol]	ADAN
18-1	0.00 ± 0.07	-0.66 ± 0.25	-0.66	0/6
28-1	0.00 ± 0.07	-0.26 ± 0.50	-0.26	2/6
30-1	0.00 ± 0.07	-0.28 ± 0.25	-0.28	0/6
23-9	0.23 ± 0.04	0.03 ± 0.25	-0.20	0/6
19-3	-1.13 ± 0.04	-1.09 ± 0.25	0.04	0/6
16-Z	0.07 ± 0.05	-0.13 ± 0.50	-0.20	3/6
32-Z	-0.33 ± 0.03	-0.72 ± 0.50	-0.39	2/6
19-2	-1.19 ± 0.02	-1.46 ± 0.25	-0.27	1/6
23(E12L:R14A)-Z	-0.31 ± 0.02	-0.69 ± 0.25	-0.38	0/6
19-5	0.01 ± 0.02	-0.57 ± 0.25	-0.58	1/6
1SRL.20-1	-1.20 ± 0.10	-0.85 ± 0.25	0.35	0/6

4.1.2.2.3 Applikationsdomäne

Gemäß der ADAN-Analyse liegen fast alle Konjugate des HS innerhalb der Applikationsdomäne des Modells A2-R6. 3 von 11 HS-Konjugaten (28-1, 16-Z und 32-Z) wurden höheren Kategorien zugeordnet, diese weisen damit größere Fehlerbereiche auf (± 0.50 kcal/mol statt ± 0.25 kcal/mol). Dies deutet darauf hin, dass das generierte Modell die Wertebereiche der relevanten Deskriptoren dieser Konjugate prinzipiell abdeckt und somit für eine Vorhersage geeignet sein sollte, trotz des Vorkommens relativ vieler Konjugate mit extremen $\Delta\Delta G_f^0$ -Werten im HS. Auch 1SRL.20-1 wird im Gegensatz zu den Modellen aus Ansatz 1 (Kap. 4.1.2) nicht in eine höhere ADAN-Kategorie eingestuft, was auf eine Anwendbarkeit des Modells auf andere Proteine hinweist.

4.1.2.2.4 Vergleich zur strukturbasierten Richtlinie

Ein direkter Vergleich mit der strukturbasierten Richtlinie von Price et al. [100] ist schwierig, da diese lediglich auf einer kleinen Anzahl (10) an Konjugaten mit Asparagin-Linkern beruht (16-1, 17-1, 18-1, 19-1, 26-1, 27-1, 28-1, 29-1, 30-1 und 32-1). Sie wurde außerdem nur mit einem einzigen zusätzlichen Konjugat getestet (1SRL.20-1). In Tab. 4.6 sollen zunächst die bereits beschriebenen statistischen Werte des Modells A2-R6 zum TS und HS (Spalte 1) den korrespondierenden Werten zur Richtlinie von Price et al. mit den dort untersuchten 10 Konjugaten (Spalte 2) gegenübergestellt werden. Dabei zeigt sich, dass die $RMSE_{TS}$ -, CCC_{TS} und R^2_{TS} -Werte des Modells A2-R6 vergleichbar mit jenen der strukturbasierten Richtlinie sind. Letztere zeigt jedoch eine deutlich bessere Rangordnung (τ_B). Der von Price et al. postulierte Winkel wird auf Basis der Kristallstruktur ermittelt. Wird der Mittelwert dieses Winkels in der Simulation des WT der WW-Domäne berechnet, so werden die Werte in Spalte 3 erhalten. Mit Ausnahme eines leicht erniedrigten τ_B -Werts werden die statistischen Werte leicht verbessert, was auf verringerte Residuen für 16-1 und 19-1 zurückzuführen ist (Regression nicht gezeigt). Dies verdeutlicht, dass die Dynamik der Strukturen hierbei eine wichtige Rolle spielen kann, was prinzipiell für einen MD-basierten Ansatz zur Beschreibung des Polymerverhaltens spricht.

Berechnet man θ_{Price} jedoch für alle Konjugate (43), bei denen der WT zur Winkelmessung genutzt werden kann, so zeigt sich, dass die Leistung deutlich abfällt (Spalte 4). Diese Berechnung schließt also alle WW-Konjugate des Datensatzes 6 ein, bei denen keine zusätzlichen Mutationen eingefügt wurden und der WT somit die unPEGylierte Variante ohne Asparagin-Linker darstellt (dies bedeutet: z. B. nicht die Konjugate 1PIN.19(S16A)-1, 1PIN.19(Y23F)-1 oder 1PIN.23(E12L:R14L)-Z). Werden auch entsprechende Simulationen zu WT-Varianten dieser Mutanten (vgl. Kap. 3.1.1) zur Winkelmessung genutzt, können die statistischen Werte für den gesamten Datensatz 6 evaluiert werden. Unter Nutzung dieser mutierten Kristallstrukturen als zusätzliche Referenzen ist die Leistung jedoch noch niedriger (Spalte 5). Zu Price et al. vergleichbare Winkel, bei dem der Vektor b (siehe Kap. 1.1.6.2) nicht zur nächstgelegenen Hydroxylgruppe, sondern zur nächstgelegenen polaren, apolaren oder aromatischen Aminosäure gezogen wird, wurden auch berechnet, führten jedoch ebenfalls zu keinen vielversprechenden, univariaten Korrelationen.

Insgesamt zeigen diese Beobachtungen, dass die strukturbasierte Richtlinie von Price et al. bei einer kleinen Auswahl an Konjugaten auf Basis eines Asparagin-Linkers mit in der Nähe verorteten Hydroxylgruppen durchaus eine gute Korrelation ergibt, sie jedoch keineswegs auf eine größere Vielfalt an Strukturen anwendbar ist (auch nicht auf erweiterte Datensätze, welche lediglich Konjugate mit Asparagin-Linkern enthalten). Das Modell A2-R6 erzielt im Hinblick auf den zugrundeliegenden, größeren Datensatz mit verschiedenen Linkertypen eine bessere Leistung. Price et al. postulierten in den nachfolgenden Studien viele zusätzliche Effekte einer PEGylierung, die durch

Tabelle 4.6: Vergleichende Gegenüberstellung der Modelleleistung von A2-R6 mit der von Price et al. entwickelten, strukturbasierten Richtlinie. Die Leistung von A2-R6 bzgl. des TS und des HS ist in der ersten Spalte gegeben (HS/TS). Es folgt die Leistung der strukturbasierten Richtlinie, wie sie in der Referenzpublikation [100] (Ref.) aufgeführt ist. In Spalte 3 (WT MD) sind die Werte unter Verwendung der aus der WT-Simulation entnommenen Winkel aufgeführt. In Spalte 4 (erweitert) wird die Leistung bei Nutzung aller Konjugate, bei denen die WT-Kristallstruktur als Referenz genutzt werden kann, aufgeführt. In der letzten Spalte (D6) ist schließlich die Leistung für den Datensatz 6 (HS/TS) aufgelistet. Die Anzahl verwendeter Konjugate ist jeweils in Klammern angegeben.

Parameter	A2-R6 (61/11)	Ref. (10)	WT MD (10)	Erweitert (43)	D6 (61/11)
RMSE	0.25/0.37	0.22	0.19	0.47	0.43/0.45
CCC	0.81/0.74	0.78	0.85	0.23	0.06/0.43
τ_B (TS)	0.45/0.82	0.78	0.69	0.23	0.04/0.21
R^2	0.68/0.76	0.64	0.74	0.23	0.03/0.27

diesen Winkel nicht abgebildet werden (Desolvation apolarer Flächen [103], Stärkung intramolekularer Salzbrücken [102] und NH- π -Bindungen [104]). Insofern ist es nicht verwunderlich, dass diese strukturbasierte Richtlinie keine zufriedenstellenden Vorhersagen für erweiterte Datensätze liefert. Durch die Verwendung verschiedener Deskriptoren können im angewandten QSPR-Modellierungsansatz dagegen prinzipiell unterschiedliche Aspekte berücksichtigt werden, was aus den ausgewählten Modelldeskriptoren ersichtlich ist (*vide infra*).

Die einzige zur Validierung zusätzlich von Price et al. untersuchte Struktur auf Basis dieses Winkels stellt 1SRL.20-1 dar. Dieses Konjugat wurde erfolgreich als stabilisiert klassifiziert, der vorhergesagte $\Delta\Delta G_f^0$ -Wert weicht mit -0.49 kcal/mol jedoch sehr weit vom experimentell bestimmten (-1.20 ± 0.10 kcal/mol) ab. Das Modell A2-R6 liefert dagegen eine Vorhersage von -0.85 ± 0.25 kcal/mol und liegt damit näher am Messwert.

4.1.2.2.5 Modelldeskriptoren

Den wichtigsten Deskriptor stellt *molbur* dar, dieser war bereits für die Modelle in Ansatz 1 sehr bedeutend. Es zeigt sich somit eine Stabilisierung durch eine erhöhte Abdeckung der Proteinoberfläche mit den Polymeratomen. Dieser Deskriptor korreliert ebenso mit *protpeghbonds* (Anzahl an Protein-PEG-HB). Univariate Regressionen ergaben im Vergleich zu den übrigen Deskriptoren relativ gute R^2 -Werte (siehe Abb. 4.6 für Regressionen aus Ansatz 1). In allen Fällen jedoch sind die Trends maßgeblich durch die stark stabilisierten Varianten 16-2, 16+26-1 und 19+26-1 bestimmt. Weitere univariate Regressionen der anderen Modelldeskriptoren sind in Abb. D.3 gezeigt. Der Deskriptor *pgbrg* beschreibt die Anzahl an überbrückten HB zwischen PEG-Atomen, die über dazwischenliegende Wassermoleküle ausgebildet werden. Entsprechende Konformationen der PEG-Ketten, bei denen sich diese um ein einzelnes Wassermolekül unmittelbar auf der Proteinoberfläche winden, wurden in einigen Simulationen beobachtet. Eine erniedrigte Anzahl dieser HB korreliert mit einer größeren Stabilisierung, was damit erklärt werden könnte, dass das Polymer stattdessen mit dem Protein HB ausbilden kann.

Mit Ausnahme des am wenigsten bedeutenden Deskriptors *pegangle* beschreiben alle übrigen Deskriptoren die Solvatation des Konjugats oder der Polymerkette. Die Deskriptoren *rdf_200ratio*, *rdf_400ratio* und *shellpego* beschreiben die Solvatation in der ersten oder zweiten Hydrathülle des Konjugats bzw. der PEG-Kette, *rdfpeg_500*, *rdf_500ratio* und *rdf_700ratio* dagegen berücksichtigen Solvenseffekte von noch weiter entfernt gelegenen Regionen. Price et al. stellten die Hypothese auf, dass die Modulation äußerer Hydrathüllen für den stabilisierenden Effekt ausschlaggebend sein könnte [100]. Es ist jedoch anzumerken, dass die Koeffizienten der Deskriptoren *rdf_200ratio*, *rdf_400ratio*, *rdfpeg_500* und *rdf_700ratio* negativ sind und somit eine erhöhte Hydratation favorisiert wird. Die Deskriptoren *rdf_XXXratio* beschreiben jedoch die Solvatation um alle Protein- und Polymeratome herum, im Gegensatz zu den Deskriptoren *rdfwopeg_XXXratio*. Letztere wurden jedoch größtenteils durch den ersten Filterschritt im Modellgenerierungsansatz entfernt. Im Gegensatz dazu wird *rdf_500ratio* ein sehr großer Koeffizient zugeordnet. Beim Blick auf die univariaten Regressionen in Abb. D.3 zeigen all diese RDF-Deskriptoren einen leichten Trend hin zu einer Desolvatation bei negativeren $\Delta\Delta G_f^0$ -Werten. Dies unterstreicht die Desolvatationseffekte, die von Price et al. beschrieben wurden. Die entsprechenden Korrelationen sind jedoch sehr schwach ausgeprägt.

Der Deskriptor *shellpolratio* zeigt ebenso eine leichte Korrelation der Proteinstabilisation mit der Desolvatation polarer Aminosäuren, was die durch den Winkel θ_{Price} beschriebene, entropisch günstige Wasserverdrängung wiedergeben könnte. Die Bedeutung dieses Deskriptors für das Modell ist jedoch gering. Der Deskriptor *pegangle* stellt den Winkel zwischen den Massenzentren des Proteins, des Linkers und der Polymerkette dar (angelehnt an [35]) und kann somit als Maß dafür dienen, wie sehr das Polymer zur Proteinoberfläche hin orientiert ist. Mit einem Koeffizienten von -0.08 (geringste relative Gewichtung) hat dieser jedoch keine große Bedeutung.

Insgesamt werden durch die Modelldeskriptoren direkte Protein-PEG-Interaktionen (*molbur*) als auch Solvenseffekte über zahlreiche RDF-/Hydrathüllen-Variablen beschrieben. Letzteres könnte die in der Literatur beschriebenen, entropisch günstigen Wasserverdrängungen wiedergeben, der wichtigste Parameter *molbur* ebenso enthalpisch günstige Protein-PEG-Interaktionen. Die ausgewählten Variablen bzgl. der Solvatation berücksichtigen Effekte in vielen verschiedenen Hydrathüllen. Dies macht es schwierig, die wichtigsten Areale einer Desolvatation genauer zu spezifizieren.

Eine exemplarische Untersuchung des wichtigsten Deskriptors *molbur* wurde bereits für die Modelle des ersten Ansatzes anhand eines besonders stabilisierten (16-1) und eines destabilisierten Konjugats (27-1) durchgeführt (siehe Kap. 4.1.1.7).

4.1.2.2.6 Alternative Polymerarten

Die Deskriptoren des Modells A2-R6 sind ebenso für andere Polymerarten berechenbar. Daher wurden für die in Kap. 4.3 beschriebenen Pin 1-Varianten mit LPG,

PMeOx und PEtOx analoge Vorhersagen erstellt (siehe Tab. D.8). Für die allermeisten Konjugate werden jedoch sehr starke Destabilisationen vorhergesagt (Ausnahme bildet das Konjugat 19- ψ mit -1.25 kcal/mol). Dies legt die Vermutung nahe, dass für diese Polymere separate Vorhersagemodelle generiert werden müssen, was erst nach zukünftigen, experimentellen Charakterisierungen erfolgen kann. Vergleicht man die Mittelwerte der Modelldeskriptoren für die PEG-Alternativen mit jenen der Konjugate des TS, so zeigt sich, dass die Vorhersage einer Destabilisierung für LPG, PMeOx und PEtOx besonders durch größere Werte von *pgbrg* bestimmt wird (siehe auch Abb. D.5). Vor allem LPG kann aufgrund der Hydroxymethylseitenketten sehr viel mehr überbrückende HB ausbilden, weshalb dieser Deskriptor größere Werte zeigt.

4.1.2.2.7 Modell A2-R6v2

Im Anschluss an die Analyse des Modells A2-R6 wurde schließlich die Leistung von Modellen mit einer reduzierten Anzahl an Deskriptoren untersucht. Hierzu wurden sukzessive die für das Modell A2-R6 unbedeutendsten Deskriptoren entfernt und die finale Modellgenerierung wiederholt. Durch Entfernung der beiden Deskriptoren *pegangle* und *rdf_200ratio*, die eine sehr geringe Bedeutung für das Modell A2-R6 zeigten, wurde das abgewandelte Modell A2-R6v2 erhalten (Tab. 4.7, Abb. 4.14), nur geringfügig schlechtere statistische Werte zeigt, jedoch 8 statt 10 Deskriptoren nutzt. Die adjustierten Bestimmtheitsmaße dieses Modells, die die Anzahl verwendeter Deskriptoren berücksichtigen, sind dennoch niedriger als die des Modells A2-R6 ($R^2_{\text{adj}}(\text{TS}) = 0.67$ und $R^2_{\text{adj}}(\text{HS}) = 0.73$ für Modell A2-R6; $R^2_{\text{adj}}(\text{TS}) = 0.63$ und $R^2_{\text{adj}}(\text{HS}) = 0.63$ für Modell A2-R6v2).

Tabelle 4.7: Statistische Validierungsparameter für das Modell A2-R6v2, sowie dessen verwendete Deskriptoren (geordnet nach der Bedeutung für das Modell, mit Angabe der Regressionskoeffizienten für den skalierten Datensatz). Die Bedeutung basiert auf den t-Werten der einzelnen Koeffizienten, die aus dem Verhältnis des jeweiligen Koeffizienten und des dazugehörigen Standardfehlers resultieren.

Modell A2-R6v2							
RMSE _{TS}	$\tau_B(\text{TS})$	CCC _{TS}	R^2_{TS}	RMSE _{HS}	$\tau_B(\text{HS})$	CCC _{HS}	R^2_{HS}
0.26	0.40	0.78	0.64	0.40	0.67	0.69	0.66
Deskriptor				Koeffizient			
<i>molbur</i>				-0.94			
<i>rdfpeg_500</i>				-0.65			
<i>rdf_500ratio</i>				1.37			
<i>pgbrg</i>				0.20			
<i>rdf_700ratio</i>				-0.61			
<i>shellpego</i>				0.52			
<i>rdf_400ratio</i>				-0.91			
<i>shellpolratio</i>				0.08			

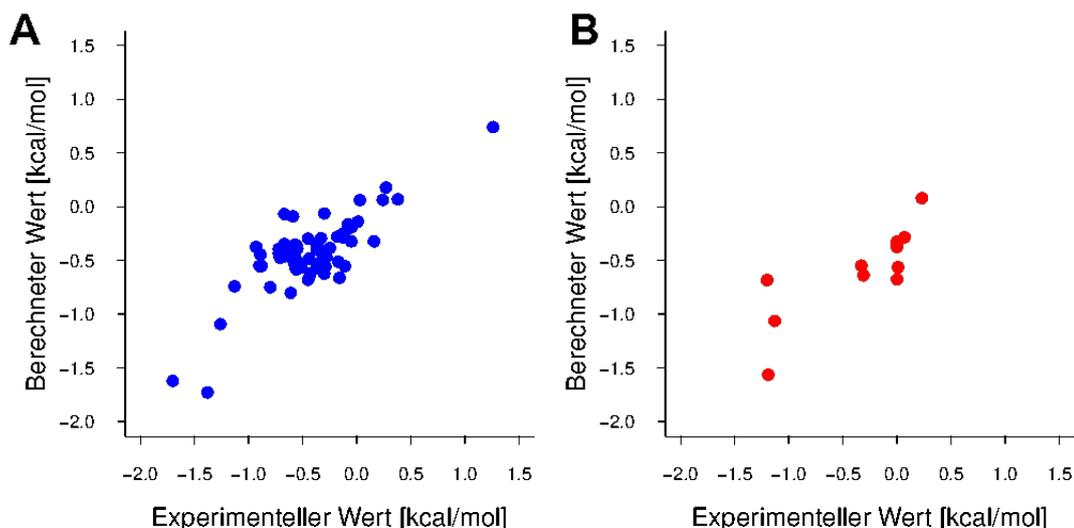


Abbildung 4.14: Auftragung der durch das Modell A2-R6v2 vorhergesagten Differenzen der freien Faltungsenergien gegen die experimentellen Werte der (A) TS- und (B) HS-Konjugate.

4.1.2.2.8 Fazit

Die besten Modelle für jeden Datensatz wurden in Ansatz 2 im Gegensatz zu Ansatz 1 in Kap. 4.1.1 nicht auf Basis der Leistung bzgl. des externen HS, sondern im Hinblick auf die aus dem TS gebildeten Testsets während des Generierungsprozesses ausgewählt. Mithilfe dieser erweiterten Kreuzvalidierung fungieren die HS in diesem Ansatz als externe Validierungssets, gegenüber jenen die Modelle unvoreingenommen sind. Dies spiegelt sich auch in den, je nach Datensatz, geringen Leistungen gemäß Tab. D.6 und D.7 wieder. Modell A2-R6 konnte die beste Leistung für das HS erzielen, wobei die Leistung für das TS vergleichbar mit den Modellen anderer Datensätze ist.

Modell A2-R6 konnte besonders stabilisierte und destabilisierte Konjugate des entsprechenden HS erfolgreich einordnen, hat jedoch Probleme, die Rangfolge von Konjugaten in einem engen Bereich zwischen etwa -0.70 bis -0.20 kcal/mol korrekt wiederzugeben. Es erzielt für 1SRL.20-1 eine bessere Vorhersage als die publizierte, kristallstrukturbasierte Richtlinie von Price et al. [100]. Im Vergleich zu dieser schneidet das Modell bzgl. des untersuchten Datensatzes auch insgesamt deutlich besser ab (vgl. Tab. 4.6).

Wie auch beim Modell E in Ansatz 1 sind hierbei sowohl Modulationen der Hydrathüllen als auch direkte Protein-PEG-Interaktionen bedeutend. Letztere werden durch den Deskriptor *molbur* wiedergegeben, der auch für das Modell E bedeutend ist. Wie jedoch auch schon für die Modelle in Ansatz 1 erwähnt zeigen sich hier einige Limitierungen. Eine breitere Verteilung an $\Delta\Delta G_f^0$ -Werten und größere Datensätze sind mit einer vergleichbaren Verteilung an Linker- und Proteinstrukturen notwendig, um eine verlässliche Modellierung zu gewährleisten.

4.1.2.3 Klassifikationsmodell A2-C1

Zur Unterscheidung stabilisierter und destabilisierter Varianten bietet sich alternativ ein klassifikationsbasierter Ansatz an. Hierbei stellt sich natürlich die Frage nach einem geeigneten Schwellenwert zur Einteilung, da die meisten Konjugate eine leichte Stabilisierung aufweisen. Während die Konjugate im ersten Modellgenerierungsansatz in stark stabilisierte (< 0.40 kcal/mol) und schwach stabilisierte / unbeeinflusste / destabilisierte (≥ 0.40 kcal/mol) Varianten eingeteilt wurden, wurde für den zweiten Ansatz die Trennung bei 0 kcal/mol vorgenommen (stabilisiert vs. unbeeinflusst / destabilisiert). Hier soll das Klassifikationsmodell für den ersten Datensatz, genannt A2-C1, näher beschrieben werden.

4.1.2.3.1 Allgemeine Leistung

Das Klassifikationsmodell A2-C1 erreicht eine gewichtete Genauigkeit $BAcc_{TS}$ von 0.84 und einen AUC_{TS} -Wert von 0.93 (siehe Tab. 4.8). Auch bzgl. des HS sind die Werte besser als das anfangs vorgestellte Referenzmodell: $BAcc_{HS}$ beträgt 0.77 und AUC_{HS} 0.83. ROC-Kurven sind in Abb. 4.15 dargestellt.

Tabelle 4.8: Statistische Validierungsparameter für das Modell A2-C1, sowie dessen verwendete Deskriptoren (geordnet nach der Bedeutung für das Modell, mit Angabe der Regressionskoeffizienten für den skalierten Datensatz). Die Bedeutung basiert auf den z-Werten der einzelnen Koeffizienten, die aus dem Verhältnis des jeweiligen Koeffizienten und des dazugehörigen Standardfehlers resultieren.

Modell A2-C1							
$AUROC_{TS}$	$AUPRC_{TS}$	CCR_{TS}	$BAcc_{TS}$	$AUROC_{HS}$	$AUPRC_{HS}$	CCR_{HS}	$BAcc_{HS}$
0.93	0.96	0.93	0.84	0.83	0.81	0.81	0.77
Deskriptor				Koeffizient			
<i>pegwathbonds</i>				-5.16			
<i>rdfpeg_500</i>				-13.69			
<i>rdfpeg_400</i>				-14.17			
<i>rdfwopeg_400ratio</i>				-2.07			
<i>molbur</i>				5.05			
<i>rdfpeg_200</i>				2.99			
<i>rdfpeg_300</i>				2.09			

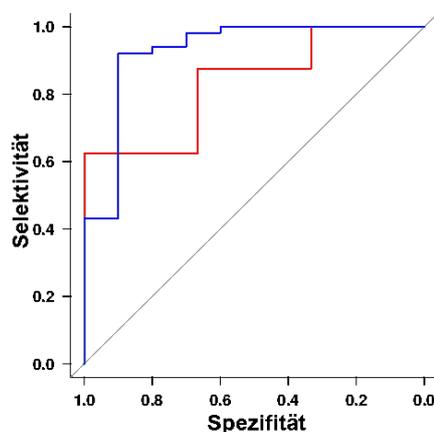


Abbildung 4.15: ROC-Kurven des Modells A2-C1 für das TS (blau) und das HS (rot).

4.1.2.3.2 Einzelne Konjugate

Die Wahrscheinlichkeit, der Klasse der stabilisierten Konjugate zugewiesen zu werden, ist in Tab. 4.9 wiederum für jedes Konjugat des HS separat aufgelistet. Es lässt sich erkennen, dass die meisten Wahrscheinlichkeiten nahe oder genau bei 100 % liegen und es wenige Fälle nahe des Schwellenwerts von 0.50 gibt. Den deutlich destabilisierten Konjugaten des HS (27-9 und 27-1) werden jedoch korrekterweise Wahrscheinlichkeiten von unter 50 % zugewiesen. Das Konjugat 16-Z liegt, unter Berücksichtigung der Fehlergrenze, sehr nahe an der Grenze um 0 kcal/mol. Diesem wird jedoch eine Wahrscheinlichkeit von nahezu 100 % zugewiesen. Falsch ist ebenso die Zuweisung des Konjugats 17-1 zur destabilisierten Klasse. Diese Variante zeigt auch lediglich eine relativ schwache Stabilisierung von -0.18 ± 0.05 kcal/mol. Das Konjugat 1SRL.20-1 wird aber korrekterweise als stabilisiert eingestuft.

Tabelle 4.9: Voraussagen (Wahrscheinlichkeiten, der Klasse der stabilisierten Konjugate zugewiesen zu werden) bzgl. des HS für jedes einzelne Konjugat. Dabei ist in Klammern angegeben, ob das Konjugat auf Basis dieser Werte als stabilisiert (S) oder nicht stabilisiert (N) einzustufen ist.

Konjugat	Experimenteller Wert [kcal/mol]	Vorhersage
17-1	-0.18 ± 0.05 (S)	0.26 (N)
27-1	0.38 ± 0.04 (N)	0.01 (N)
16(S32A)-1	-0.58 ± 0.03 (S)	0.96 (S)
27-9	0.27 ± 0.02 (N)	0.43 (N)
16-Z	0.07 ± 0.05 (N)	1.00 (S)
19-2	-1.19 ± 0.02 (S)	1.00 (S)
23(E12L:R14L)-Z	-0.89 ± 0.04 (S)	1.00 (S)
23(E12L:R14A)-Z	-0.31 ± 0.02 (S)	1.00 (S)
19(S16F)-1	-1.13 ± 0.07 (S)	1.00 (S)
16+26-1	-1.38 ± 0.03 (S)	1.00 (S)
1SRL.20-1	-1.20 ± 0.10 (S)	0.94 (S)

4.1.2.3.3 Modelldeskriptoren

Der wichtigste Deskriptor ist *pegwathbonds*. Eine geringere Anzahl an HB des Polymers zum Solvens korreliert somit mit einer größeren Wahrscheinlichkeit einer Stabilisierung. Ist diese Anzahl geringer, so ist zu erwarten, dass das Polymer stattdessen mit dem Protein wechselwirkt. Ebenso ist abermals der Deskriptor *molbur* in der Auswahl der Variablen vorhanden.

Interessanterweise beschreiben vier der fünf übrigen Deskriptoren die Solvatation der PEG-Kette in verschiedenen Hydrathüllen. Wie auch im Falle der Deskriptoren von Modell A2-R6, die eine Modulation der Solvatation beschreiben, zeigen sich hier teilweise gegensätzliche Koeffizienten. Dem Deskriptor *rdfpeg_400* wird ein negativer Koeffizient zugewiesen, den Deskriptoren *rdfpeg_200*, *rdfpeg_300*, *rdfpeg_500* ein positiver. Zusätzlich zu Variablen, die die Solvatation des Polymers beschreiben, beschreibt der Deskriptor *rdfopeg_400ratio* das Verhältnis der Solvatation des Proteinteils der Struktur zwischen PEGylierter und unPEGylierter Variante. Diese ist

im Falle der stabilisierten Varianten niedriger als bei destabilisierten Konjugaten. Eine Darstellung der univariaten Regressionen ist in Abb. D.4 gezeigt, wobei sich keine eindeutigen Trends erkennen lassen.

4.1.2.3.4 Alternative Polymerarten

Wie auch schon für das Regressionsmodell A2-R6 vorgestellt können Vorhersagen für die alternativen Polymere erhalten werden. Diese sind in Tab D.9 gezeigt. Ähnlich zu Modell A2-R6 werden die allermeisten Konjugate hierbei als destabilisiert eingestuft. Lediglich die Konjugate 17- λ , 19- ϕ und 23- ϕ werden als stabilisiert klassifiziert. Die vom Modell ausgegebenen Wahrscheinlichkeiten, der Klasse an stabilisierten Konjugaten zugewiesen zu werden, liegen zumeist nahe oder genau bei 0 % oder 100 %, mit Ausnahme von 17- λ (67 %) und 32- λ (43 %). Wie bei Modell A2-R6 zeigt ein Blick auf die Verhältnisse zwischen den Mittelwerten der PEG-Alternativen und der Konjugate des TS, weshalb die meisten LPG-/POx-Konjugate als destabilisiert eingestuft werden (siehe Abb. D.6). Diese PEG-Alternativen weisen deutlich mehr HB zum Solvens auf, wodurch größere Werte für den Deskriptor *pegwathbonds* erhalten werden. Hier wäre evtl. eine Normierung dieses Deskriptorwerts auf die Polymeroberfläche geeignet (ähnlich zu den Analysen in Kap. 4.3).

4.1.2.3.5 Fazit

Das Klassifikationsmodell A2-C1 zeigt eine gute Leistung bzgl. des TS und des HS. Der Deskriptor *molbur* ist abermals einer der bedeutendsten. Im Hinblick auf die Modulation der Hydratation spielen jedoch, im Gegensatz zu Modell A2-R6, Deskriptoren, die die Solvenshülle des Polymers beschreiben, eine größere Rolle. Das von Price et al. zur Validierung genutzte Konjugat (1SRL.20-1) wird korrekterweise als stabilisiert eingestuft. Die deutlich destabilisierten Konjugate 27-1 und 27-9 wurden ebenso erfolgreich klassifiziert.

4.1.2.4 Zusätzliche Modelle

4.1.2.4.1 Parametervergleich

Da zur Simulation der unterschiedlichen Konjugate letztlich drei verschiedene Parametrisierungsmethoden zum Einsatz kamen (PyRED, antechamber, monomerbasiert, siehe Kap. 3.1.3), stellt sich die Frage, ob diese zu einer unterschiedlichen Dynamik des Polymers führten und damit die Werte der verwendeten Deskriptoren beeinflussten. Zur Untersuchung dessen wurden diese Parametrisierungen exemplarisch am Beispiel der Konjugationsstelle 16, ausgehend von einer identischen Startkonformation des PEG-Tetramer-Konjugats, verglichen (Simulationsprotokoll analog zu Kap. 3.1.5). Die per PyRED parametrisierte Struktur, deren Simulation in dieser Arbeit zur Generierung von Vorhersagemodellen diente, wurde hierfür ein weiteres Mal simuliert, um ebenfalls die Fluktuationen zwischen verschiedenen Replika mit identischen Parametern zu untersuchen.

Abb. 4.16-A und 4.16-B zeigen, dass die Dynamik der Tetramere zum großen Teil vergleichbar ist, die Polymerdichte der antechamber-Variante ist identisch mit jener der PyRED-Referenzsimulation. Lediglich die monomerbasierte Parametrisierung scheint eine geringere Tendenz aufzuzeigen, die Aminosäuren der ersten Schleife im Bereich um S18 - G20 zu adressieren. Dagegen ist eine etwas größere Dichte im Bereich des dritten β -Faltblatts um S32 zu finden.

Zusätzlich wurden relevante Pin 1-unspezifische Deskriptoren aus Ansatz 2 berechnet. Abb. 4.16-C verdeutlicht, dass die Unterschiede zwischen der PyRED- und der antechamber-Methode minimal sind und sich im Bereich der Standardfehler der Deskriptoren und / oder der Unterschiede zwischen zwei PyRED-Replika befinden. Kein Deskriptor zeigt eine Abweichung von $\geq 25\%$ in Relation zum Wert der PyRED-Referenztrajektorie. Im Gegensatz hierzu zeigt die Simulation mit monomerbasierten Parametern für einzelne Deskriptoren etwas stärker abweichende Werte. Hier finden sich insgesamt vier Parameter (vgl. Namen in Tab. D.1) mit Abweichungen von über 25 %: *intoh* (134 %), *pegwathbonds* (152 %), *rmsfp* (155 %) und *pgbrg* (243 %). Die monomerbasierten Parameter wurden allerdings hauptsächlich für die größeren Systeme in Kap. 5 verwendet. Im Hinblick auf die generierten Vorhersagemodelle wurden sie für die Konjugate 19-B (-0.34 ± 0.04 kcal/mol), 19-J (-0.37 ± 0.05 kcal/mol), 19-O (-0.72 ± 0.07 kcal/mol) und 19-U (-0.61 ± 0.04 kcal/mol) genutzt. Es handelt sich somit um einen sehr geringen Anteil an Konjugaten und Deskriptoren des gesamten Datensatzes und diese Varianten stellen keine besonders stark (de-)stabilisierten Mutanten dar, sodass kein entscheidender Einfluss der Deskriptorwerte dieser vier Konjugate bei der Modellgenerierung anzunehmen ist.

Zur weiteren Untersuchung wurden diese vier genannten Konjugate im Anschluss auch nochmal mit der PyRED-Methode neu parametrisiert und simuliert. Im Zuge einer erneuten Deskriptorberechnung wurden ebenfalls die in Kap. 4.1.2.2 zu Modell A2-R6 beschriebenen, verschiedenen Abwandlungen des Price-Winkels (zu nächstgelegenen (a)polaren oder aromatischen Aminosäuren), die in univariaten Regressionen keine überzeugenden Korrelationen zeigten, als zusätzliche Deskriptoren in diesen aktualisierten Datensatz integriert (vgl. Tab. D.1). Mit letzterem wurden dann gemäß Ansatz 2 nochmal neue Modelle generiert. Für den Datensatz 1 gemäß Tab. 3.3 konnten zu den Modellen A2-R6 und A2-C1 vergleichbar gute Regressions- als auch Klassifikationsmodelle erhalten werden. Diese wurden als A2v2-R1 und A2v2-C1 bezeichnet und sollen im Folgenden kurz beschrieben werden.

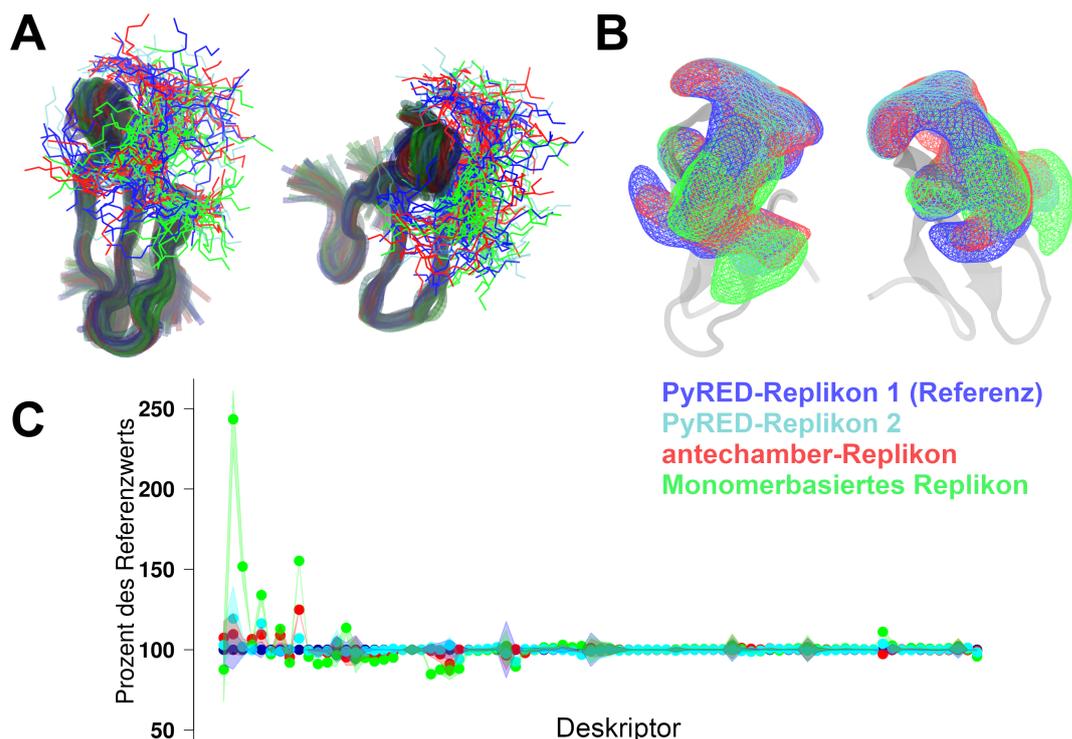


Abbildung 4.16: (A) Illustration der Polymerschweratome (alle 5 ns) und des Proteinrückgrats (alle 20 ns) für vier verschiedene Simulationen des Konjugats 16-1: per Parametrisierung über die PyRED-Methode (dunkelblau, Referenzsimulation), per Parametrisierung über die antechamber-Methode (rot), per fragmentbasierter Parametrisierung mittels antechamber (grün), sowie ein weiteres Replikon der mit PyRED parametrisierten Struktur (cyan). (B) Dichten der Polymerschweratome (Durchschnitt aller Momentaufnahmen, Isowert: 0.04). (C) Prozentuale Abweichung der Deskriptorwerte (relevante Deskriptoren des Modells A2-R6, siehe Kap. 4.1.2.2) vom Wert, der für die Referenzsimulation berechnet wurde. Dabei sind die Mittelwerte der vier verschiedenen Simulationen als Punkte gezeigt, die dazugehörigen Standardfehler sind durch entsprechend eingefärbte Bereiche gekennzeichnet.

4.1.2.4.2 Modell A2v2-R1

Die statistischen Werte von Modell A2v2-R1 sind in Tab. 4.10 aufgeführt und Regressionen sind in Abb. 4.17 gezeigt. Für das Konjugat 1SRL.20-1 wird ein Wert von -0.98 ± 0.22 kcal/mol berechnet (lediglich 27-1, 19(S16F)-1 und 16+26-1 werden in höhere ADAN-Kategorien eingestuft). Interessanterweise werden vier verschiedene Price-Winkel als Deskriptoren verwendet. Die Koeffizienten von *price1* und *price3* (Varianten des ursprünglichen Winkels) deuten, ähnlich wie bei vorherigen Modellen, auf gegeneinander ausgleichende Effekte hin. Dabei können die Koeffizienten der Deskriptoren *price3pol* und *price2arom* auf eine Orientierung des Polymers weg von polaren hin zu apolaren Aminosäureseitenketten gedeutet werden, was wiederum die Hypothese einer Desolvatation apolarer Proteinoberflächen unterstützt [103]. Der Koeffizient des zusätzlichen Deskriptors *pegwathbonds* deutet, ähnlich wie zuvor *molbur*, abermals auf eine Stabilisierung durch eine Bedeckung des Proteins durch das Polymer hin. Während univariate Regressionen der Price-Winkel also keine guten Korrelationen zeigten, führte eine Kombination zu einem relativ guten multivariaten Modell.

Tabelle 4.10: Statistische Validierungsparameter für das Modell A2v2-R1, sowie dessen verwendete Deskriptoren (geordnet nach der Bedeutung für das Modell, mit Angabe der Regressionskoeffizienten für den skalierten Datensatz). Die Bedeutung basiert auf den t-Werten der einzelnen Koeffizienten, die aus dem Verhältnis des jeweiligen Koeffizienten und des dazugehörigen Standardfehlers resultieren.

Modell A2v2-R1							
RMSE _{TS}	τ_B (TS)	CCC _{TS}	R ² _{TS}	RMSE _{HS}	τ_B (HS)	CCC _{HS}	R ² _{HS}
0.22	0.58	0.84	0.72	0.35	0.75	0.85	0.77
Deskriptor				Koeffizient			
<i>price3pol</i>				-0.33			
<i>price2arom</i>				0.34			
<i>pegwathbonds</i>				0.35			
<i>price3</i>				0.44			
<i>totalint</i>				0.22			
<i>shellpego</i>				-0.30			
<i>price1</i>				-0.33			
<i>rdfoh_300ratio</i>				-0.11			

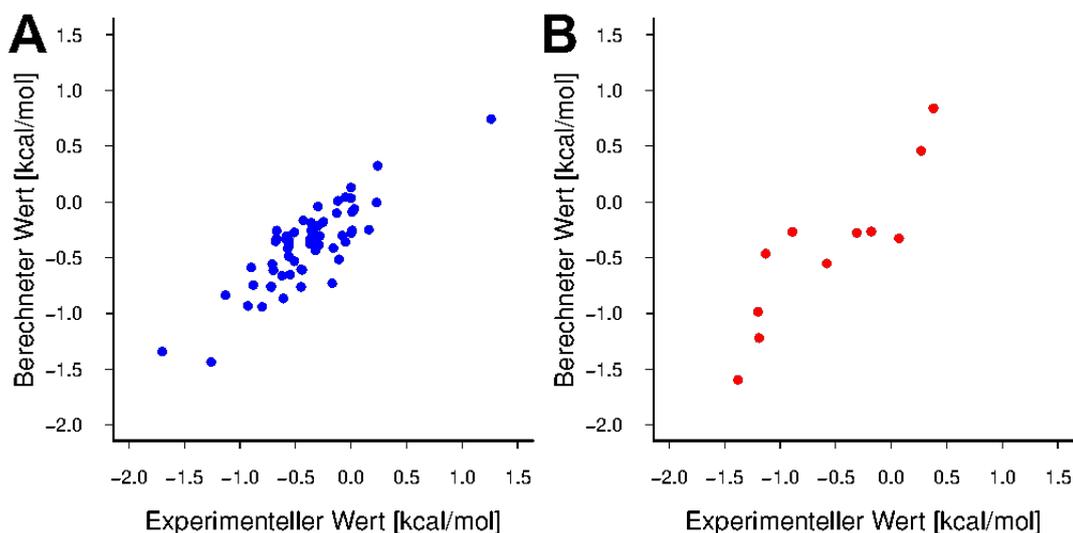


Abbildung 4.17: Auftragung der durch das Modell A2v2-R1 vorhergesagten Differenzen der freien Faltungsentropien gegen die experimentellen Werte der (A) TS- und (B) HS-Konjugate.

4.1.2.4.3 Modell A2v2-C1

Die Statistiken und Deskriptoren zu Modell A2v2-C1 sind in Tab. 4.11 und die ROC-Kurven in Abb. 4.18 gezeigt. Auch in diesem Falle kommt der Winkel *price3pol* zur Anwendung, zusätzlich zu verschiedenen RDF-Deskriptoren und, wie bei Modell A2-R1, dem Deskriptor *pegwathbonds*. Für 1SRL.20-1 wird eine Wahrscheinlichkeit von 99 % der Zugehörigkeit zur Klasse stabilisierter Konjugate vorausgesagt.

Tabelle 4.11: Statistische Validierungsparameter für das Modell A2v2-C1, sowie dessen verwendete Deskriptoren (geordnet nach der Bedeutung für das Modell, mit Angabe der Regressionskoeffizienten für den skalierten Datensatz). Die Bedeutung basiert auf den z-Werten der einzelnen Koeffizienten, die aus dem Verhältnis des jeweiligen Koeffizienten und des dazugehörigen Standardfehlers resultieren.

Modell A2v2-C1							
AUROC _{TS}	AUPRC _{TS}	CCR _{TS}	BAcc _{TS}	AUROC _{HS}	AUPRC _{HS}	CCR _{HS}	BAcc _{HS}
0.91	0.95	0.93	0.80	0.88	0.83	0.82	0.67
Deskriptor				Koeffizient			
<i>pegwathbonds</i>				-2.15			
<i>price3pol</i>				-1.36			
<i>rdfpeg_200</i>				1.26			
<i>rdfpeg_300</i>				-14.86			
<i>rdfpeg_400</i>				14.23			
<i>rdfwopeg_400ratio</i>				-2.52			

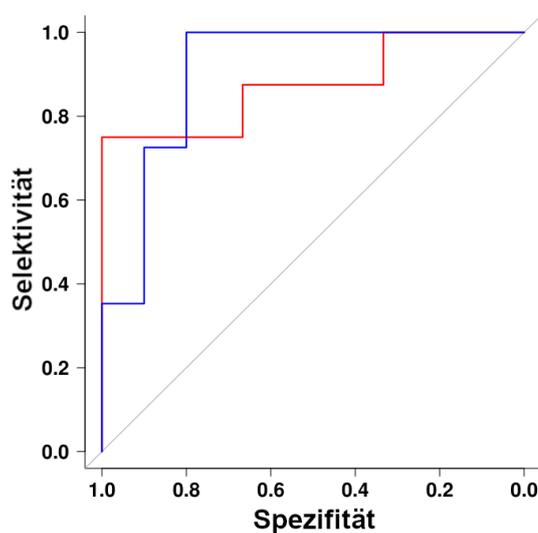


Abbildung 4.18: ROC-Kurven des Modells A2v2-C1 für das TS (blau) und das HS (rot).

4.1.2.5 Fazit

Der abgewandelte Modellgenerierungsansatz 2 führte zu den finalen Modellen A2-R6, A2-C1, A2v2-R1 und A2v2-C1. Auch wenn diese auf den ersten Blick leicht verschlechterte statistische Werte im Vergleich zu den komplexeren Modellen E und F des ersten Ansatzes zeigen, so basieren sie auf für kleinere Datensätze besser geeigneten Modellarten und wurden mittels eines erweiterten Kreuzvalidierungsverfahrens und besseren Holdout-Sets genauer auf ihre prädiktive Leistung in Bezug auf externe Konjugate hin untersucht. Abermals wurden v. a. *molbur* sowie viele verschiedene Solvatationsdeskriptoren zur Modellierung genutzt.

Alle vier genannten Modelle des Ansatzes 2 stuften die von Price et al. zur Validierung genutzte Src SH3-Domäne als stabilisiert ein, die Regressionsmodelle A2-R6 und A2v2-R1 zeigten zudem im Vergleich zur strukturbasierten Richtlinie verbesserte Vorhersagen. Eine bessere Leistung des Modells A2-R6 gegenüber letzterer Richtlinie wurde durch analoge Winkelmessungen für den gesamten Datensatz 6 verdeutlicht.

Aus diesen Ergebnissen lässt sich schließen, dass die von Price et al. publizierte Richtlinie lediglich für einen kleinen Teil der Asn-Konjugate geeignet ist, die in dieser Arbeit erstellten Modelle jedoch bessere Leistungen für deutlich erweiterte Datensätze zeigen. Vergleichbare Leistungen zwischen den Modellen A2-R6 und A2v2-R1 deuten darauf hin, dass einerseits die monomerbasierte Parametrisierung einiger weniger Konjugate für das Modell A2-R6 keinen entscheidenden Einfluss auf die Leistung der finalen Modelle zeigte, und dass andererseits eine Kombination verschiedener Abwandlungen des Price-Winkels zu einem recht guten Modell führen kann. Durch solche Abwandlungen (bspw. *price2arom*) können von Price et al. in späteren Publikationen erwähnte Hypothesen abgebildet werden (so z. B. eine PEG-Orientierung hin zu apolaren Bereichen). Für die PEG-Alternativen LPG, PMeOx und PEtOx könnten durch weitere Messungen analoge Modelle erstellt werden.

Für die zusätzlich simulierte WW-Konjugate, zu denen von Price et al. noch keine experimentellen Charakterisierungen vorgenommen wurden (siehe Kap. 3.1.2.2), sind die Vorhersagen der Modelle aus Ansatz 2 in Tab. 4.12 gezeigt. Zukünftige experimentelle Charakterisierungen dieser Varianten könnten zur weiteren Evaluation aller in dieser Arbeit generierten Modelle herangezogen werden.

Tabelle 4.12: Vorhersagen der Modelle A2-R6, A2-C1, A2v2-R1 und A2v2-C1 zu den zusätzlich simulierten, von Price et al. nicht experimentell charakterisierten Konjugaten. Im Falle der Regressionsmodelle sind die Vorhersagen mit den ADAN-Konfidenzintervallen aufgelistet (*NA* = außerhalb der AD), für die Klassifikationsmodelle sind die Klassifizierungen mit der Wahrscheinlichkeit zur Zuordnung zur Klasse der stabilisierten Konjugate gezeigt.

Konjugat	A2-R6 [kcal/mol]	A2-C1	A2v2-R1 [kcal/mol]	A2v2-C1
7-1	-1.01 ± 0.25	S (1.00)	<i>NA</i>	S (1.00)
11-1	-0.93 ± 0.25	S (1.00)	-0.92 ± 0.22	S (0.98)
12-1	-0.29 ± 0.25	S (0.96)	-0.38 ± 0.22	S (0.99)
13-1	-0.24 ± 0.25	S (0.84)	-0.89 ± 0.22	S (0.69)
15-1	-0.31 ± 0.25	N (0.35)	-0.17 ± 0.22	S (0.67)
21-1	-0.33 ± 0.25	S (0.71)	-0.65 ± 0.22	S (0.87)
22-1	-1.10 ± 0.25	S (1.00)	-0.84 ± 0.22	S (1.00)
24-1	-0.77 ± 0.25	N (0.01)	1.07 ± 0.44	N (0.03)
25-1	-0.79 ± 0.25	S (0.99)	-0.62 ± 0.22	S (0.99)
31-1	-0.40 ± 0.25	S (0.37)	-0.38 ± 0.44	S (0.98)
33-1	-0.15 ± 0.25	N (0.24)	0.29 ± 0.22	S (0.81)
35-1	-0.77 ± 0.25	S (0.99)	-0.64 ± 0.22	S (0.98)

4.2 Metadynamik-Simulationen

Wohltemperierte Metadynamik-Simulationen mit jeweils acht „Walkern“, bei denen Distanzen zwischen den β -Faltblättern als CV dienten, wurden für die exemplarischen Konjugate 16-1 (-0.90 ± 0.03 kcal/mol) und 27-1 (0.38 ± 0.04 kcal/mol) sowie deren unkonjugierte Varianten durchgeführt. Hierdurch sollte untersucht werden, ob der von Price et al. gemessene energetische Unterschied *in silico* durch eine (partielle) Entfaltung wiedergegeben werden kann. CV 1 stellt dabei die Entfernung zwischen den β -Faltblättern 1 und 2 dar, CV 2 dagegen jene zwischen den β -Faltblättern 2 und 3.

4.2.1 Energiediagramme

In Abb. 4.19 ist die freie Energie als Funktion der beiden CV für alle vier untersuchten Systeme dargestellt. Dabei sind die Energieminima eingezeichnet, die durch das R-Paket metadynminer erkannt wurden. Das 2D-Energiediagramm wird standardmäßig in 8×8 Bins unterteilt und in jedem dieser Abschnitte wird das Energieminimum bestimmt. Alle gefundenen Minima werden dann nach ihrer Energie sortiert und mit Buchstaben gekennzeichnet. Die Anzahl an Bins zur Unterteilung ist prinzipiell veränderbar.

Diese Diagramme zeigen zunächst, dass eine volle Entfaltung entlang der CV 2 für 16-1 wahrscheinlicher als im Falle der unkonjugierten Variante ist (vgl. Minimum C von 16 mit Minimum F von 16-1). Dagegen zeigt sich bei dem destabilisierten Konjugat 27-1 eine höhere Entfaltungstendenz im Vergleich zur unkonjugierten Mutante (vgl. fehlendes Minimum für 27 mit Minimum D von 27-1). Im Falle des Konjugats 27-1 ist ebenso eine größere Entfaltungstendenz entlang CV 1 zu beobachten (Minimum E von 27-1). Allerdings lässt sich dies auch beim stabilisierten Konjugat 16-1 beobachten (Minimum D von 16-1). Wie in Kap. 1.5.1.1 erwähnt ist eine Entfaltung entlang der CV 2 gemäß [255] für eine strukturell ähnliche WW-Domäne wahrscheinlicher als entlang der CV 1, da ein Übergang entlang CV 2 eine geringere Energiebarriere darstellt. Es kann daher angenommen werden, dass die CV 2 initiale Vorgänge einer kompletten Proteinfaltung beschreibt. In keiner der Energiediagramme ist eine komplette Entfaltung zu beobachten.

4.2.2 Konvergenz

Im Zuge einer Berechnung energetischer Differenzen sollte die Konvergenz der Systeme evaluiert werden, da die oben beschriebenen Unterschiede in den erkannten Minima von einem unzureichenden Sampling herrühren können. Für Metadynamik-Simulationen kann hierzu die Energiedifferenz der registrierten Minima zum globalen Minimum über die Simulationszeit hinweg untersucht werden. Ändert sich diese Differenz gegen Ende der Stichprobennahme nicht mehr deutlich, so kann eine Konvergenz angenommen werden.

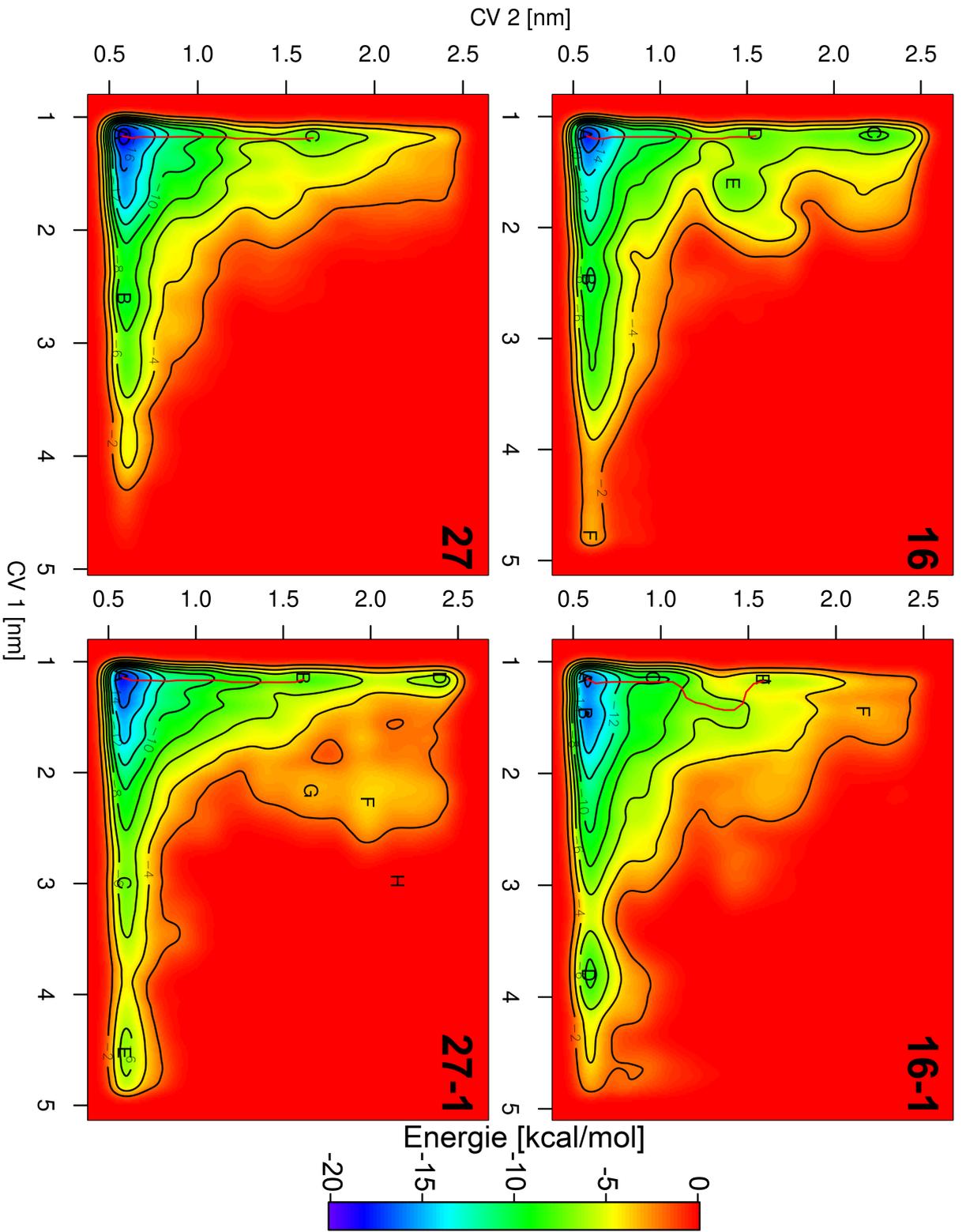


Abbildung 4.19: Energiediagramme für die Simulationen von 16, 16-1, 27 und 27-1. Die durch das metadynamer-Paket erfassten Energieminima sind durch Buchstaben gekennzeichnet. Zusätzlich sind die weiter unten beschriebenen Übergänge als rote Linien eingezeichnet (siehe Kap. 4.2.3).

In Abb. 4.20 sind die Energiedifferenzen der Minima über die Zeit hinweg illustriert. Einige detektierte Minima zeigen gegen Ende der Simulationen weiterhin große Fluktuationen, so z. B. Minimum E der Variante 16 oder Minimum F der Variante 16-1. Für einen energetischen Vergleich (*vide infra*) wurde eine partielle Entfaltung entlang der CV 2 (bis 1.5 - 1.6 Å) untersucht, da diese Minima bzgl. der Energiedifferenzen gegen Ende relativ geringe Fluktuationen zeigen.

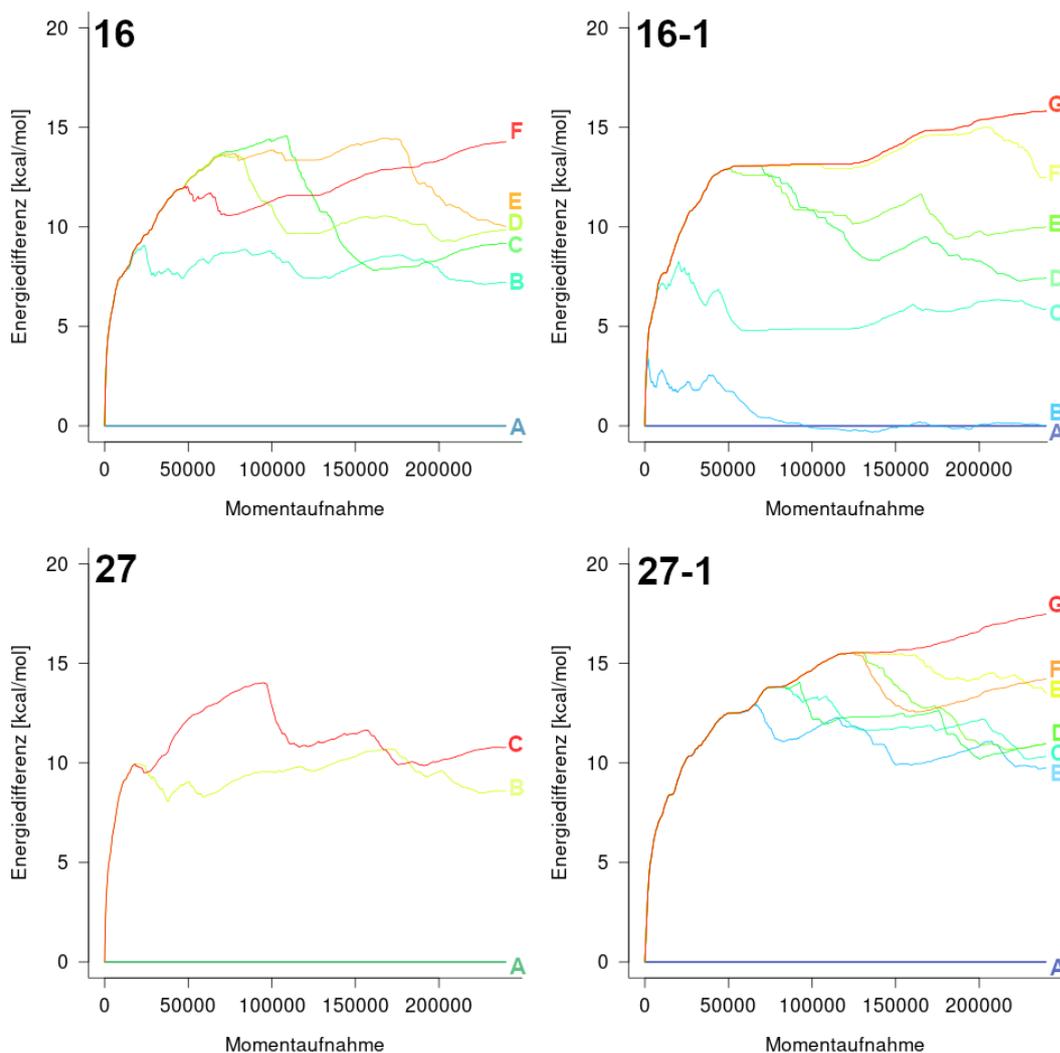


Abbildung 4.20: Verlauf der energetischen Differenzen zwischen den registrierten Minima und den globalen Minima über die Simulationszeit hinweg.

4.2.3 Energiedifferenzen

Es stellt sich die Frage, ob die von Price et al. erkannten (De-)stabilisierungen über die registrierten Energieminima wiedergegeben werden können. Eine vollständige Entfaltung wurde nicht beobachtet, und die Energielandschaften zeigen teils unterschiedliche Minima, was einen Vergleich erschwert. In allen vier Simulationen jedoch lässt sich ein Energieminimum im Hinblick auf CV 2 zwischen 1.5 und 1.6 Å erkennen.

In Tab. 4.13 sind daher die Differenzen dieser Minima zum jeweils globalen Minimum aufgeführt. Die dazu angegebenen $\Delta\Delta G$ -Werte vergleichen die energetischen Unterschiede der Differenzen in der unPEGylierten und der PEGylierten Variante. In Übereinstimmung mit einer Destabilisierung zeigt sich im Falle des Konjugats 27-1 ein geringerer energetischer Unterschied des untersuchten Übergangszustands zur nativen Struktur im Vergleich zur unPEGylierten Variante. Das stabilisierte Konjugat 16-1 zeigt dagegen eine leicht höhere Differenz. Die angegebenen Fehler basieren auf einem Vergleich der minimalen und maximalen Energiedifferenzen zum globalen Minimum in den letzten 10, 15 und 20 ns der „Walker“. Unter Berücksichtigung dieser letzten Nanosekunden lassen sich relativ stabile Energiewerte mit geringen Fehlerbereichen ermitteln.

Tabelle 4.13: Berechnung der energetischen Differenzen der Übergangszustände zum globalen Minimum (ΔG). Der angegebene Fehler stellt die Differenz des größten Energieunterschieds zum niedrigsten Unterschied innerhalb der letzten X ns dar (für X = 10 / 15 / 20 ns). Der Wert $\Delta\Delta G$ stellt den Unterschied dieser Differenzen dar (Energieunterschied_{unPEGyliert} - Energieunterschied_{PEGyliert}). Dazugehörige Fehler wurden über das Gaußsche Fehlerfortpflanzungsgesetz ermittelt.

Variante	Übergang	ΔG [kcal/mol]	Fehler [kcal/mol]	$\Delta\Delta G$ [kcal/mol]	Fehler [kcal/mol]
16	A ->D	9.96	0.06 / 0.10 / 0.20	-0.11	0.07 / 0.16 / 0.26
16-1	A ->E	10.07	0.04 / 0.12 / 0.16		
27	A ->C	10.85	0.01 / 0.06 / 0.11	1.02	0.13 / 0.19 / 0.21
27-1	A ->B	9.83	0.13 / 0.18 / 0.18		

Die berechneten energetischen Differenzen berücksichtigen lediglich die Unterschiede zwischen den Minima, für eine weitere Evaluation der Energiebarrieren zwischen diesen Zuständen wurde die NEB (*Nudged elastic band*)-Methode genutzt, die den minimalen Pfad (Abb. 4.21) zwischen zwei Minima berechnet [321]. Die ermittelten Energiebarrieren (Aktivierungsenergien ΔG^\ddagger) für die in Tab. 4.13 aufgeführten Übergänge sind in Tab. 4.14 aufgelistet. Die Energiebarriere ist im Falle der destabilisierten Variante 27-1 geringer als bei der unkonjugierten Mutante.

4.2.4 Polymerdynamik

Ungeachtet energetischer Differenzen können diese Simulationen dazu dienen, die Dynamik der Polymere während des Faltungsprozesses zu interpretieren. Generell scheint die Entfaltung entlang der CV 1 zunächst über einen Bindungsbruch der HB in der Nähe der ersten Schleife zu geschehen (also beginnend bei R21_O-S16_H), während die Entfaltung entlang der CV 2 nicht von der zweiten Schleife aus sondern mit einer Auflösung der HB am Ende des dritten β -Faltblatts (Q33_O-Y24_H) beginnt. Wie kann PEG an der Konjugationsstelle 16 zu einer Stabilisierung führen? In Übereinstimmung mit der in Kap. 4.1.1.7 beschriebenen höheren Tendenz zur Bedeckung der Proteinoberfläche im Vergleich zu 27-1 kann die Polymerkette auch während des Entfaltungsprozesses mit den Aminosäuren wechselwirken (Abb. 4.22).

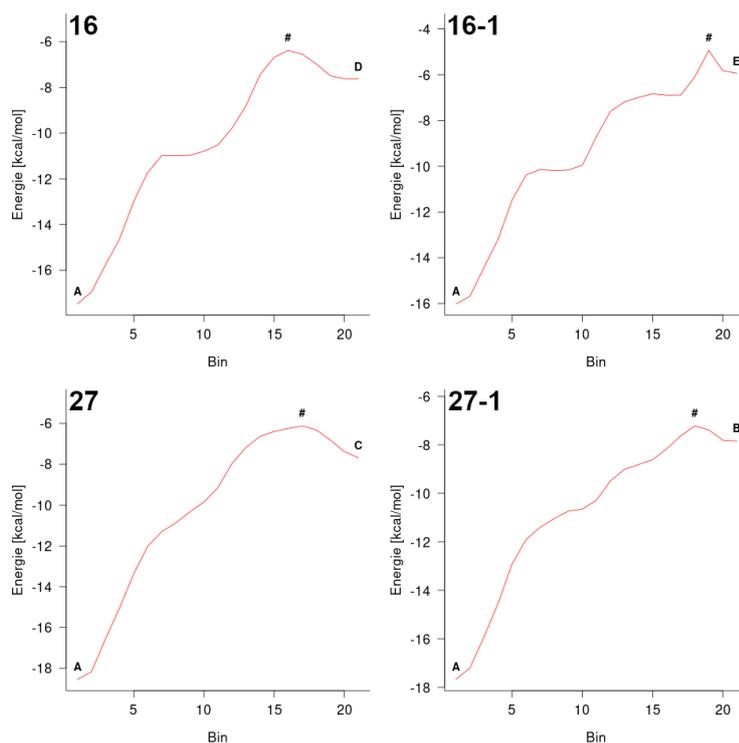


Abbildung 4.21: Schematische Darstellung der geeignetsten Pfade auf Basis der NEB-Methode zwischen den globalen Minima und den in Tab. 4.13 beschriebenen Energieminima. Das Kennzeichen # gibt die dabei registrierte Energiebarriere an. Die Pfade sind ebenso in Abb. 4.19 als rote Linien gekennzeichnet.

Tabelle 4.14: Berechnung der Energiebarrieren zum Zeitpunkt der jeweils letzten Momentaufnahme in Bezug auf die aufgelisteten Übergänge. Ebenso wird die Differenz dieser ΔG^\ddagger -Werte zwischen der unPEGylierten und der PEGylierten Variante gezeigt.

Variante	Übergang	ΔG^\ddagger [kcal/mol]	$\Delta\Delta G^\ddagger$ [kcal/mol]
16	A ->D	11.11	-0.04
16-1	A ->E	11.15	
27	A ->C	12.41	1.97
27-1	A ->B	10.44	

Das Polymer wechselwirkt mit S19 und R21 und legt sich dabei auf das bei der Entfaltung dem Solvens zugänglich werdende Areal. Diese Dynamik könnte den Entfaltungsprozess modulieren. Auch bei einer Loslösung des dritten β -Faltblatts könnte die Entfaltung durch PEG an dieser Konjugationsstelle erschwert werden, da das Polymer, wie in Kap. 4.3.1 weiter unten beschrieben, mit S32 dieses Faltblatts interagiert. Im Gegensatz hierzu zeigt sich bei dem destabilisierten Konjugat 27-1, dass die Polymerkette keine Areale des Entfaltungsprozesses adressiert. Wie auch im Falle der entsprechenden klassischen MD (vgl. Kap. 4.1.1.7) liegt PEG hauptsächlich solvensexponiert vor. Die relativ stabile Wechselwirkung des Linker-Amids mit der Seitenkette von E12 wird beim Entfaltungsprozess aufgehoben.

Diese Beobachtungen deuten also darauf hin, dass sich die PEG-Ketten bei Entfaltungsprozessen auf die freiwerdenden Areale legen und so zu einer Stabilisierung beitragen können. Dies könnte die Hypothese aus [96] stützen. Hierin wurde, in Analogie zu Chaperonen, eine polymerbasierte erleichterte Rückfaltung zur nativen Struktur durch eine Etablierung stabilerer partiell entfalteter Strukturen postuliert.

4.2.5 Fazit

Die erhaltenen Energiediagramme und die daraus berechneten Differenzen spiegeln die Stabilisierung von 16-1 und die Destabilisierung von 27-1 rein qualitativ wider, jedoch deuten die Fluktuationen der meisten Energiedifferenzen in Abb. 4.20 auch nach 2.4 μs Simulationszeit auf eine mangelnde Konvergenz hin, weshalb der Fokus auf eine partielle Entfaltung entlang CV 2 gelegt wurde. Es wäre dahingehend interessant, deutlich längere Simulationen der Konjugate 16-2 (-1.70 ± 0.03 kcal/mol) und 27-2 (1.26 ± 0.06 kcal/mol) durchzuführen, da hierbei der experimentell ermittelte energetische Unterschied noch größer ist.

Unabhängig von der Eignung der Simulationen zur Berechnung der Energiedifferenzen konnte beobachtet werden, dass PEG im Falle der destabilisierten Variante entfaltete Areale adressiert. Dies könnte den weiteren Entfaltungsprozess verhindern. Im Gegensatz dazu interagierte die Polymerkette im Falle der destabilisierten Variante nicht mit diesen Bereichen.

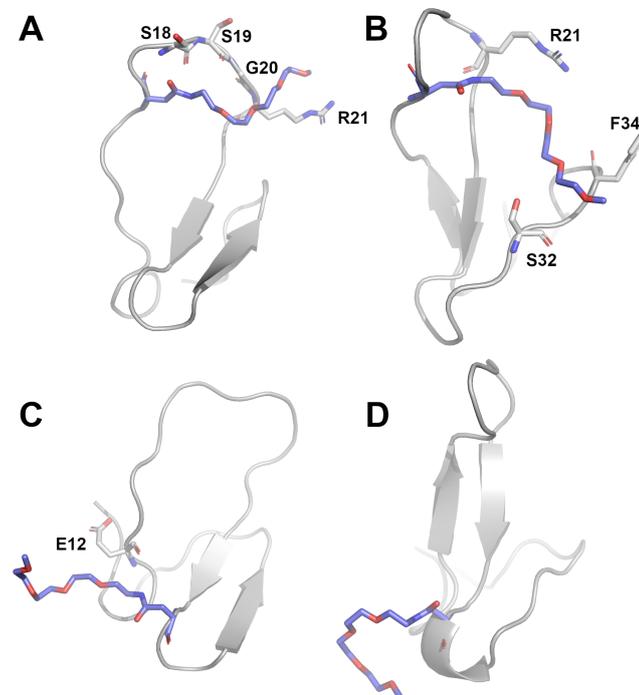


Abbildung 4.22: Exemplarische PEG-Konformationen im Falle einer partiellen Entfaltung. Gezeigt sind Entfaltungen entlang (A) der CV 1 und (B) der CV 2 für 16-1, sowie entlang (C) der CV 1 und (D) der CV 2 für 27-1. Im Falle des Konjugats 16-1 streckt sich die PEG-Kette entlang des entfalteten Areals, während bei 27-1 keine Interaktionen beobachtet werden können.

4.3 Vergleich der Polymere

Im Folgenden wird die Dynamik der vier Polymervarianten PEG, LPG, PMeOx und PEtOx in den Modellprotein-Konjugaten beschrieben und vergleichend gegenübergestellt. Jede Konjugationsstelle wird auf einer Doppelseite mithilfe der in Kap. 3.1.8 beschriebenen Eigenschaften vorgestellt. Diese Untersuchungen umfassen Interaktionsenergien (mit -1 multipliziert), sowie Informationen zum Bedeckungsgrad der Protein- und Polymeroberflächen, zu der Anzahl an Protein-Polymer HB, der Solvation der Polymerkette und den Fluktuationen des Proteinrückgrats und des Polymers. Zusätzlich wurden die Simulationen visuell analysiert. Exemplarische Momentaufnahmen, die auf Basis dieser Analysen als charakteristisch für die Trajektorien mit Fokus auf mögliche Unterschiede hinsichtlich der Protein-Polymer-Wechselwirkungen angesehen wurden, werden ergänzend zu einer entsprechenden Beschreibung der Polymerdynamik präsentiert.

Es werden 13 Konjugationsstellen der Pin 1 WW-Domäne mit Asparagin-Linker vorgestellt: 14, 16, 17, 18, 19, 21, 23, 26, 27, 28, 29, 30 und 32. Dabei handelt es sich um Positionen aus den verschiedenen Strukturelementen (β -Faltblätter und Schleifen 1 und 2), welche von Price et al. bereits in Bezug zu PEG experimentell charakterisiert wurden und für die entsprechende zusätzliche Simulationen zu den alternativen Polymeren durchgeführt wurden (vgl. Kap. 3.1.2.2). Im Falle des Her2-Affibodys werden insgesamt 15 Konjugationsstellen mit Cystein-Maleimid-Linker vorgestellt: 7, 8, 15, 21, 23, 25, 29, 33, 39, 43, 46, 47, 49, 50 und 53. Diese Auswahl basierte auf bei der Bindung des Affibodys an Her2 freiliegenden Proteinarealen, sowie den strukturbasierten Richtlinien von Price et al. (vgl. Kap. 3.1.2.2).

Ein Einschub dieser systematisch dargestellten Analysen soll hier gezeigt werden, da sie als Datengrundlage für die im Anschluss für beide Proteine aufgeführten, zusammenfassenden Vergleiche (siehe Kap. 4.3.3) dienen. Hierbei werden sowohl Unterschiede zwischen den Polymerarten als auch den Konjugationsstellen beleuchtet.

4.3.1 Pin 1-Konjugate

4.3.1.1 Konjugationsstelle 14

Die Interaktionsprofile der Polymere sind vergleichbar (Abb. 4.24), wobei die POx-Varianten stärkere Interaktionen mit Y23 und F25 zeigen, PMeOx speziell auch mit S32 und F34. Die größeren POx-Polymere führen zur Bedeckung einer größeren Proteinoberfläche. LPG ist am flexibelsten, kann aber im Vergleich zu den POx-Varianten aufgrund der Hydroxylgruppen mehr HB zum Protein aufbauen. PEG orientiert sich vornehmlich zum dritten β -Faltblatt hin und wechselwirkt dadurch mit Y23, F25 und S32 (Abb. 4.23-A). Es ist zwischen den aromatischen Gruppen von Y23 und F25 positioniert, teils auch in der Nähe von R17. Andererseits kann es sich auch in die andere Richtung umorientieren und mit K13 und M15 wechselwirken. Die Hydroxylgruppen von LPG interagieren v. a. mit dem Solvens und dem Polymerrückgrat. Die Struktur befindet sich hauptsächlich im selben Areal wie PEG, Interaktionen mit R17 sind aber häufiger zu beobachten (Abb. 4.23-B). Weiterhin sind Konformationen zu registrieren, in denen sich LPG entlang des ersten und zweiten β -Faltblatts in Richtung H27 streckt. PMeOx verbleibt sehr stabil zwischen den Aromaten von Y23 und F25, wechselwirkt hierbei auch mit den Seitenketten von R17 und S32 (Abb. 4.23-C). Während der Simulation kommt es auch zur Umlagerung der endständigen Monomere hin zum Aromaten von F34, wobei die Seitenkette des Monomers zumeist parallel auf dem aromatischen Ring aufliegt. Im Falle von PEtOx kommt es selten zur Ausstreckung zum dritten β -Faltblatt, es interagiert v. a. mit dem Rückgrat von R17 und dem Aromaten von Y23 (Abb. 4.23-D). Teilweise befindet es sich auch über F25 Richtung H27 und N30.

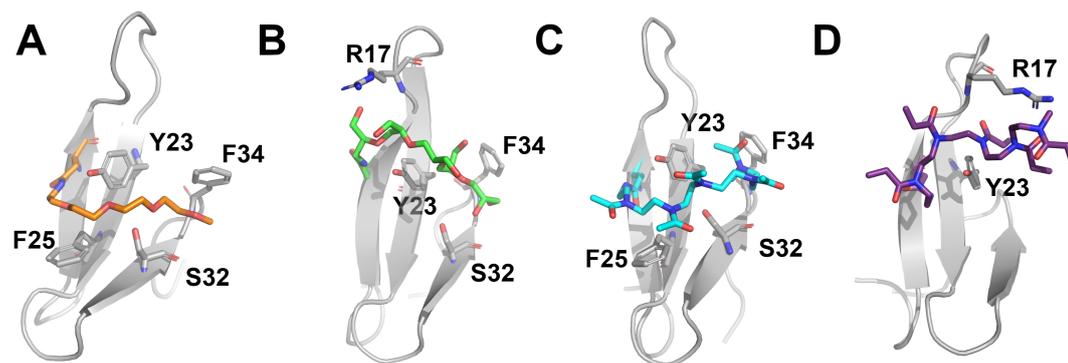


Abbildung 4.23: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 14 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

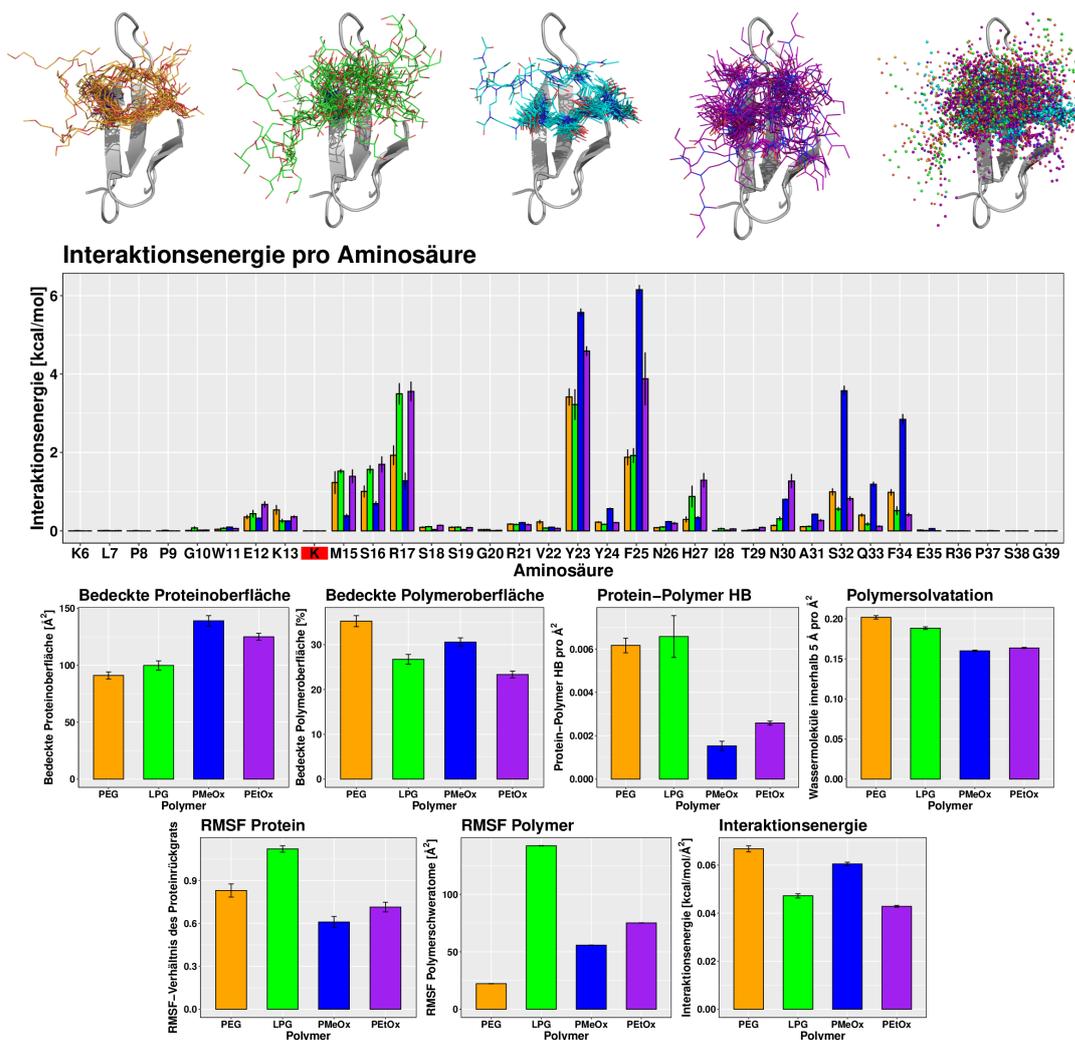


Abbildung 4.24: Protein-Polymer-Dynamiken an Konjugationsstelle 14 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.2 Konjugationsstelle 16

Im Gegensatz zur Konjugationsstelle 14 zeigen die POx-Polymere größere RMSF-Werte (Abb. 4.26), da die Monomere nicht zwischen den Aromaten von Y23, F25 und F34 positioniert sind. PEG windet sich um die erste Schleife zwischen den Seitenketten von R17 und R21 zum Rückgrat von G20 hin (Abb. 4.25-A). Teilweise orientiert es sich auch zu S32 oder auf die andere Seite der ersten Schleife zu M15 und R17. LPG befindet sich ebenso hauptsächlich neben R17 in unmittelbarer Nähe der ersten Schleife, in einer geknäuelten oder gestreckten Konformation (Abb. 4.25-B). Mit dieser Aminosäure zeigt es die im Vergleich zu allen Polymerarten größte Interaktionsenergie. Die ersten Monomere wechselwirken auch mit Y23 und F34. Während die PMeOx-Monomere initial auch in der Nähe von R17 vorzufinden sind, wechseln sie ihre Ausrichtung hin zu den Aminosäuren S32 und F34 (Abb. 4.25-C). Dabei sind wiederum zu Y23 und F34 parallel ausgerichtete Polymerseitenketten zu beobachten. PEtOx bildet eine ausgestreckte Konformation um die erste Schleife. Alternativ wechselwirkt es mit Y23, N30, S32 und F34 (Abb. 4.25-D).

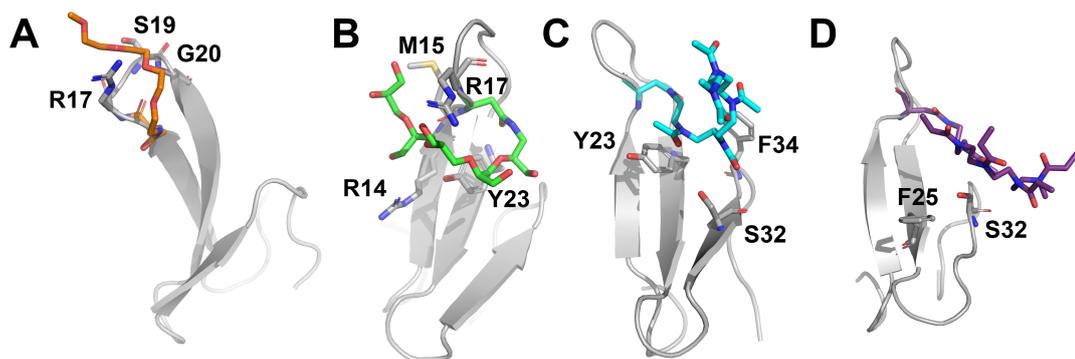


Abbildung 4.25: Beispielfotografien von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 16 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

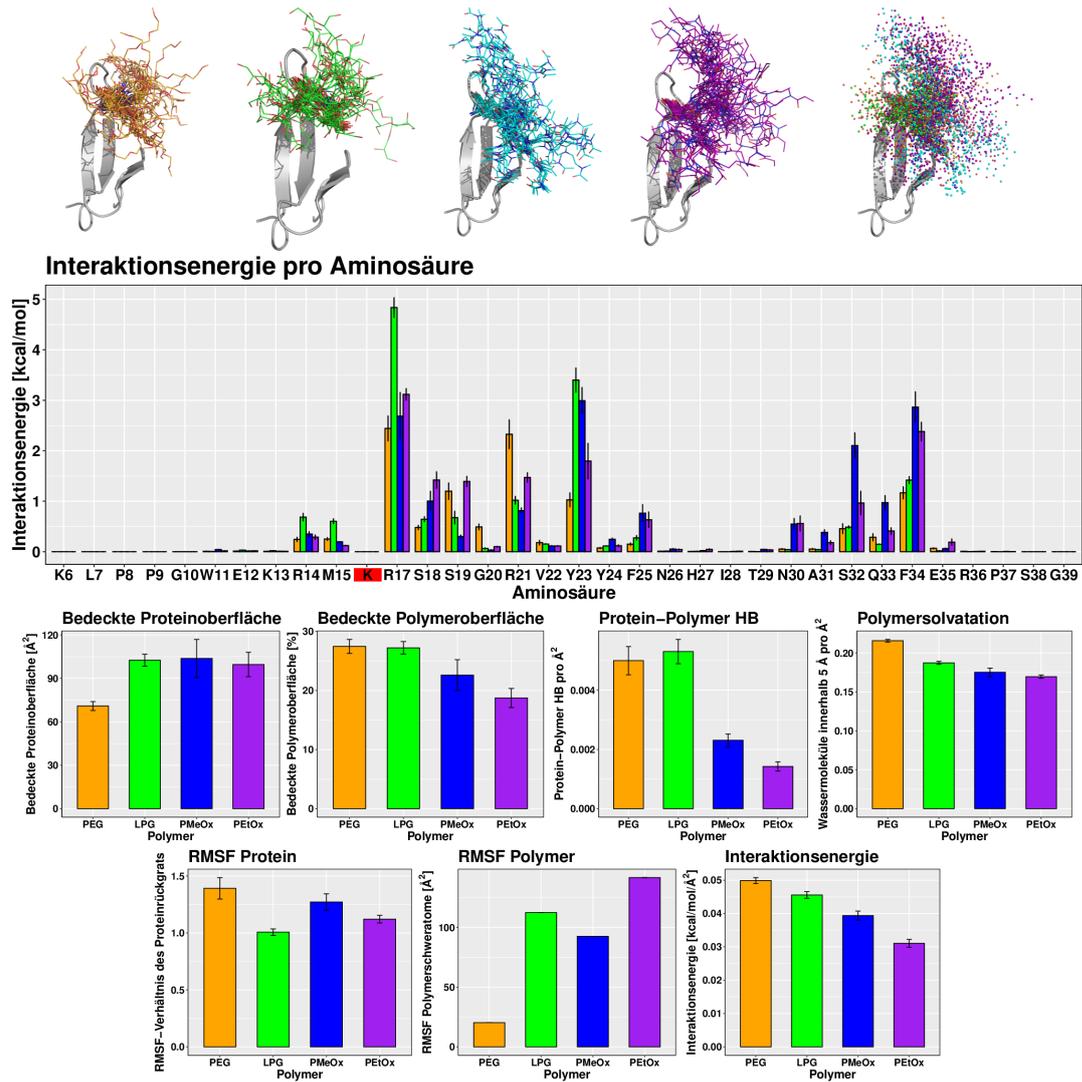


Abbildung 4.26: Protein-Polymer-Dynamiken an Konjugationsstelle 16 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.3 Konjugationsstelle 17

Nur PMeOx verweilt hauptsächlich auf der Proteinoberfläche, was in einer deutlich größeren Bedeckung des Proteins und stärkeren Interaktionen resultiert (Abb. 4.28). PEG ändert seine Konformation im Laufe der Simulation sehr häufig, einzig zwischen S32 und F34 entlang der Oberfläche von Y23 lassen sich relativ stabile Wechselwirkungen beobachten (Abb. 4.27-A). Ähnlich zu PEG ist auch LPG recht flexibel und interagiert vornehmlich mit sich selbst oder dem Solvens (Abb. 4.27-B). Daher kann keiner stabile Konformation hin zu F34 beobachtet werden. PMeOx fluktuiert hingegen nicht im Bereich der Konjugationsstelle um die erste Schleife, sondern bildet sehr schnell stabile Wechselwirkungen zu Y23, F25, sowie mit den terminalen Monomeren zu den Rückgraten von N30 und A31 (Abb. 4.27-C) und den initialen Monomeren zur Seitenkette von R14. PEtOx zeigt andere Dynamiken, da es, ähnlich wie PEG und LPG, im Bereich der ersten Schleife fluktuiert. Hier wechselwirkt es teilweise mit der Carbonylgruppe von S18 (Abb. 4.27-D). Entsprechend weist es im Vergleich zu den anderen Konjugationsstellen recht hohe RMSF-Werte auf.

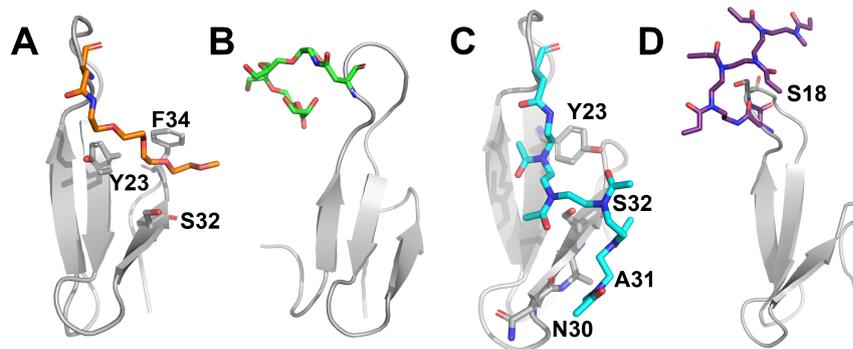


Abbildung 4.27: Beispielaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 17 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

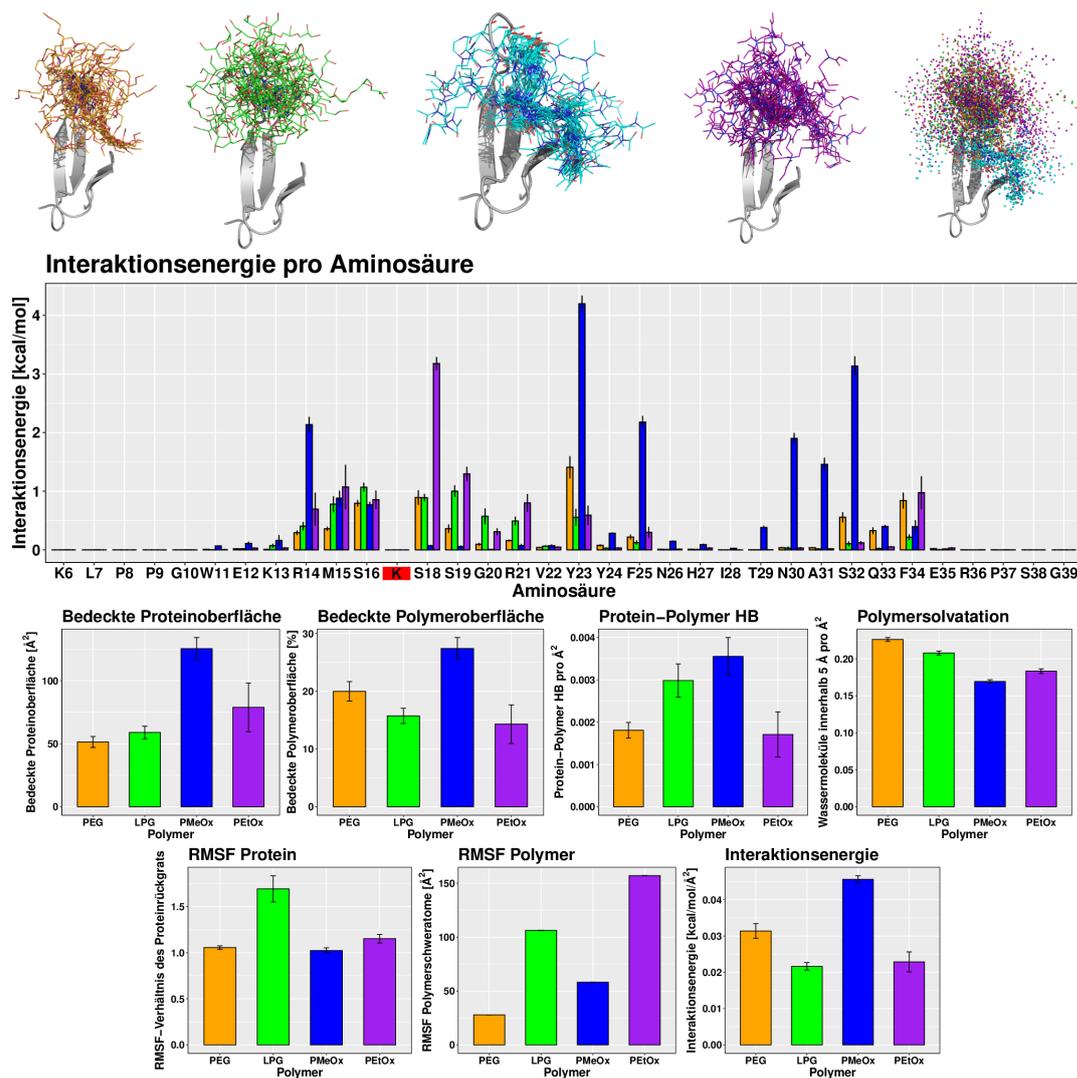


Abbildung 4.28: Protein-Polymer-Dynamiken an Konjugationsstelle 17 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.4 Konjugationsstelle 18

Position 18 zeigt die größten Unterschiede hinsichtlich der Polymerdynamik, wodurch auch die Interaktionsprofile deutliche Differenzen aufweisen (Abb. 4.30). PMeOx ist das einzige Polymer mit Wechselwirkungen zu Y23, S32, Q33 und F34, PEtOx dagegen zeigt Interaktionen mit K13, R14 und M15. LPG adressiert lediglich R17. Die PEG-Alternativen zeigen wesentlich höhere RMSF-Werte. PEG orientiert sich zu M15 und R17 hin, mit Wechselwirkungen zum eigenen Asparagin-Linker. Öfters aber zeigt es Konformationen um die andere Seite der Schleife herum, abgewandt von M15 (Abb. 4.29-A). Zwischenzeitlich fluktuiert das Polymer im Solvens. LPG adressiert nur selten die Seite der ersten Schleife mit M15. Es ist hauptsächlich solvens exponiert und interagiert mit sich selbst oder der Seitenkette von R17 (Abb. 4.29-B). In manchen Momentaufnahmen zeigt es in Richtung des zweiten und dritten β -Faltblatts. PMeOx liegt nicht solvens exponiert vor, sondern streckt sich hin zum Aromaten von Y23, ebenso zu F34 und S32 (Abb. 4.29-C). Im Gegensatz hierzu orientiert sich PEtOx v. a. auf die proteinabgewandte Seite der Schleife und interagiert mit dem hydrophoben Areal von K13, M15 und V22 (Abb. 4.29-D). Die terminale Carbonylgruppe kann mit der positiv geladenen Seitenkette von K13 wechselwirken. Zu Beginn fluktuiert das Polymer im Solvens, abgewandt vom Protein.

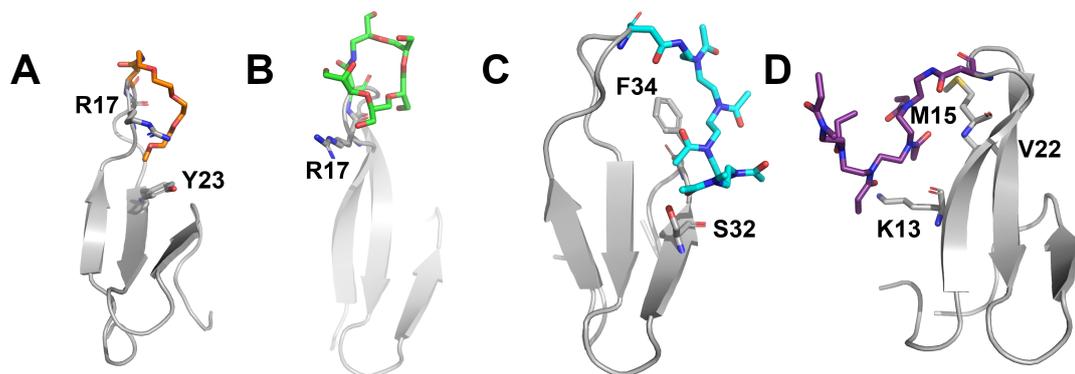


Abbildung 4.29: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 18 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

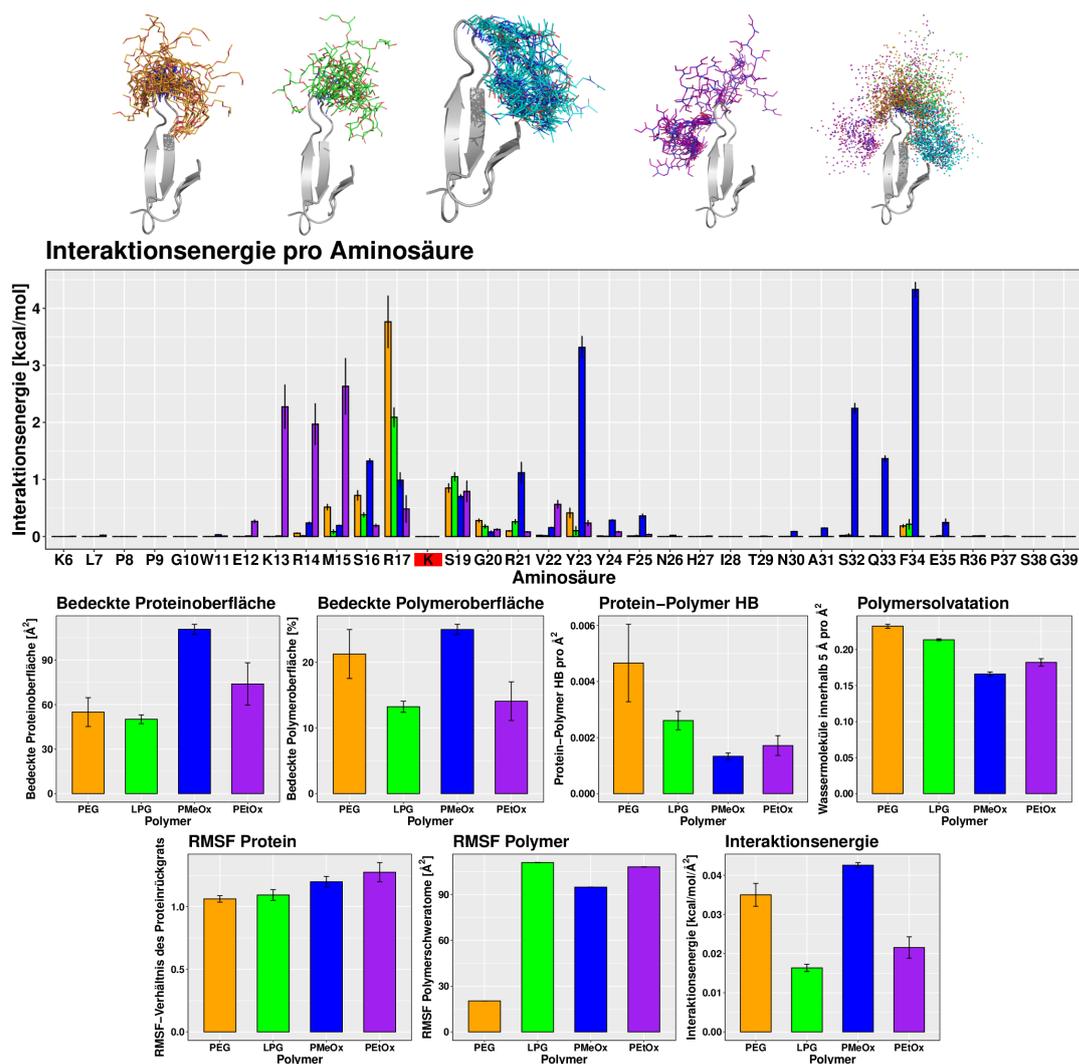


Abbildung 4.30: Protein-Polymer-Dynamiken an Konjugationsstelle 18 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.5 Konjugationsstelle 19

An dieser Stelle zeigt PEG die größtmögliche Oberflächenbedeckung, den geringsten RMSF-Wert und die stärkste Interaktionsenergie (Abb. 4.32), hauptsächlich mit Aminosäuren der ersten Schleife. PEG windet sich um die erste Schleife und interagiert mit dem Amidrückgrat von G20, sowie mit R21 mithilfe des Asparagin-Linkers (Abb. 4.31-A). Eine Orientierung zu S32 und F34 ist nur selten zu beobachten. LPG liegt solvensexponierter, abgewandt vom Protein, vor. Es interagiert mit der Seitenkette von R21 und, für kürzere Zeit, mit der entfernt gelegenen Aminosäure E35 (Abb. 4.31-B). PMeOx ist zur Rückseite der ersten Schleife orientiert und wechselwirkt mit dem Rückgrat von R17, sowie mit M15 (Abb. 4.31-C). Zu Beginn ist das Polymer in der Nähe von S32 und F34 positioniert. PEG verbleibt auf dem Aromaten von Y23 und interagiert mit S32 über die terminale Carbonylgruppe, während die Seitenkette von R21 ebenfalls in der Nähe des Polymers verbleibt (Abb. 4.31-D).

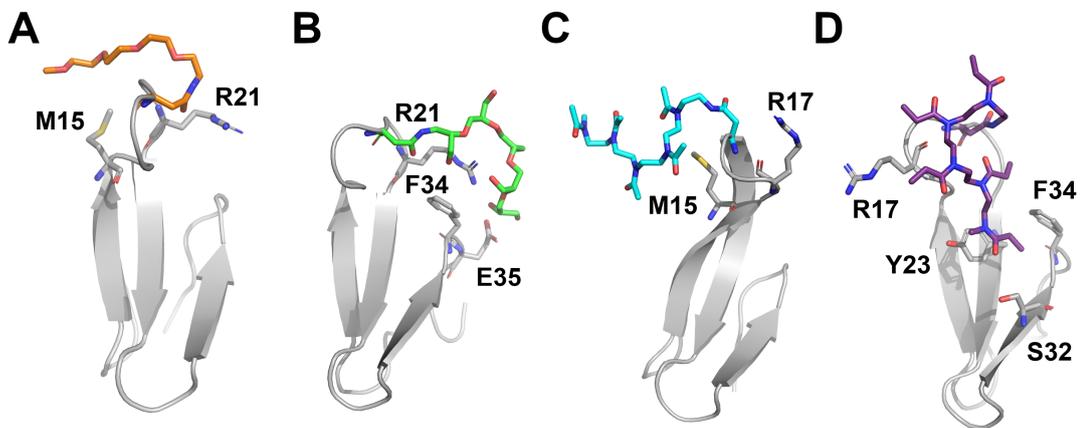


Abbildung 4.31: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEG an der Konjugationsstelle 19 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEG violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

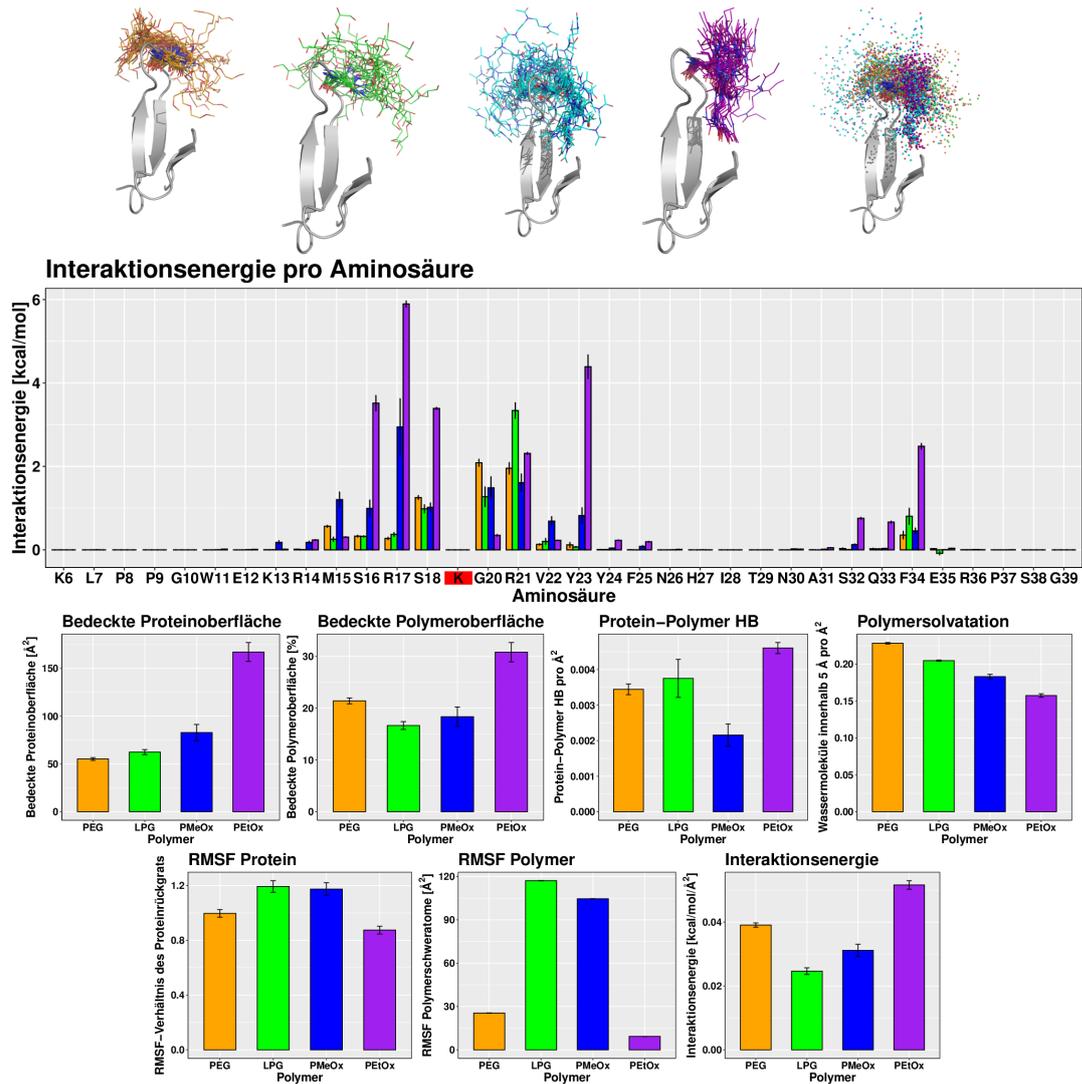


Abbildung 4.32: Protein-Polymer-Dynamiken an Konjugationsstelle 19 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.6 Konjugationsstelle 21

Das Interaktionsprofil zeigt wiederum starke Interaktionen von P_{Me}Ox mit Y23 und P_{Et}Ox mit F34 (Abb. 4.34). PEG verbleibt abermals zwischen S32 und F34 (Abb. 4.33-A) und wechselwirkt mit S16 über seinen Linker. Alternativ windet es sich um R17. LPG orientiert sich ebenso zum zweiten und dritten β -Faltblatt hin (Abb. 4.33-B), manchmal auch zur ersten Schleife. P_{Me}Ox verbleibt auf dem Aromaten von Y23 und erreicht mit seinem terminalen Monomer A31 (Abb. 4.33-C). Im Gegensatz hierzu interagiert P_{Et}Ox mit F34 in der kompletten Simulation (Abb. 4.33-D). Der Rest des Polymers ist vom Protein abgekehrt, die terminale Carbonylgruppe kann mit der Seitenkette von S19 wechselwirken.

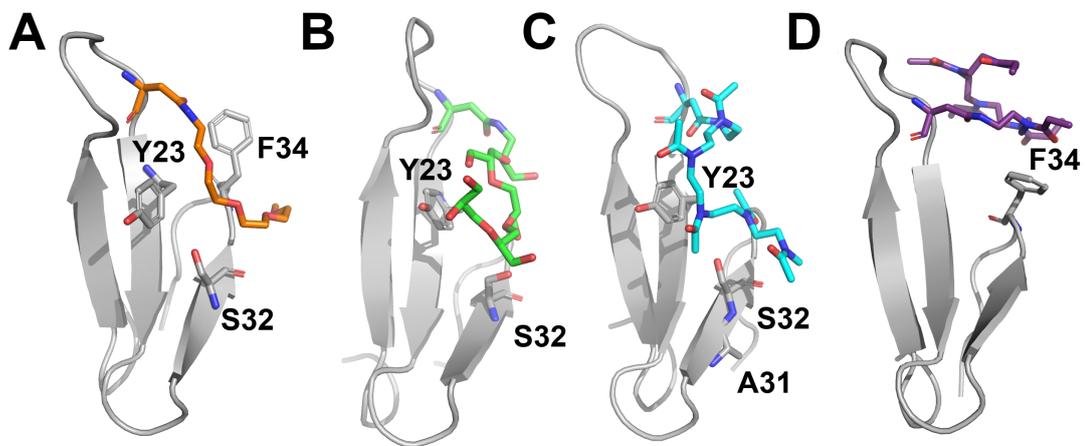


Abbildung 4.33: Beispielfotografien von (A) PEG, (B) LPG, (C) P_{Me}Ox und (D) P_{Et}Ox an der Konjugationsstelle 21 von Pin 1 WW. PEG ist orange, LPG grün, P_{Me}Ox blau und P_{Et}Ox violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

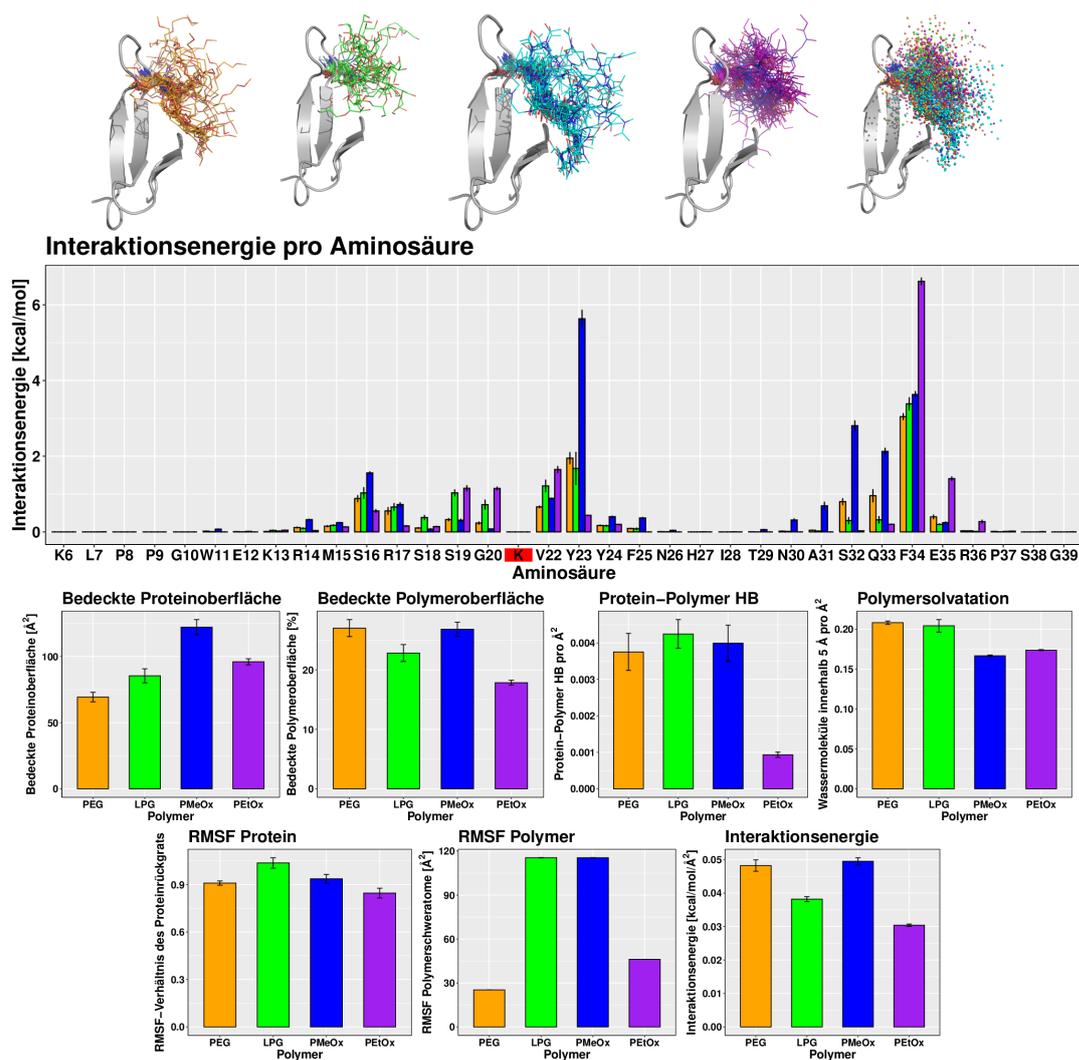


Abbildung 4.34: Protein-Polymer-Dynamiken an Konjugationsstelle 21 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEIOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.7 Konjugationsstelle 23

PEG interagiert hier mit R14, R17 und S32 (Abb. 4.35-A). LPG zeigt ähnliche Wechselwirkungen, die Konjugation führt hier jedoch zu einer stärkeren Desolvatation um F25 (Abb. 4.35-B). PMeOx fluktuiert stark, was ebenso in einem recht großen RMSF-Wert resultiert (Abb. 4.36). Zu Beginn ist es zwischen R14 und R17 positioniert und kann M15 mit seinen hydrophoben Strukturen erreichen (Abb. 4.35-C). Im zweiten Teil der Simulation streckt es sich ins Solvens aus, direkt über dem dritten β -Faltblatt. PEtOx ist nicht zwischen R14 und R17 vorzufinden, sondern verbleibt auf dem zweiten und dritten β -Faltblatt und wechselwirkt mit F25, N30, S32 und F34 (Abb. 4.35-D). Es zeigt dadurch auch die größte Proteinoberflächenbedeckung und einen sehr geringen RMSF-Wert (Abb. 4.36).

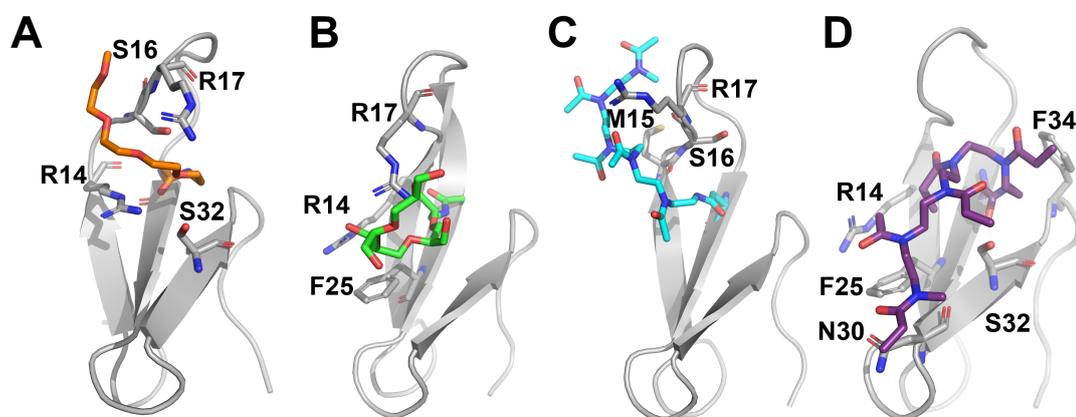


Abbildung 4.35: Beispielfotografien von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 23 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

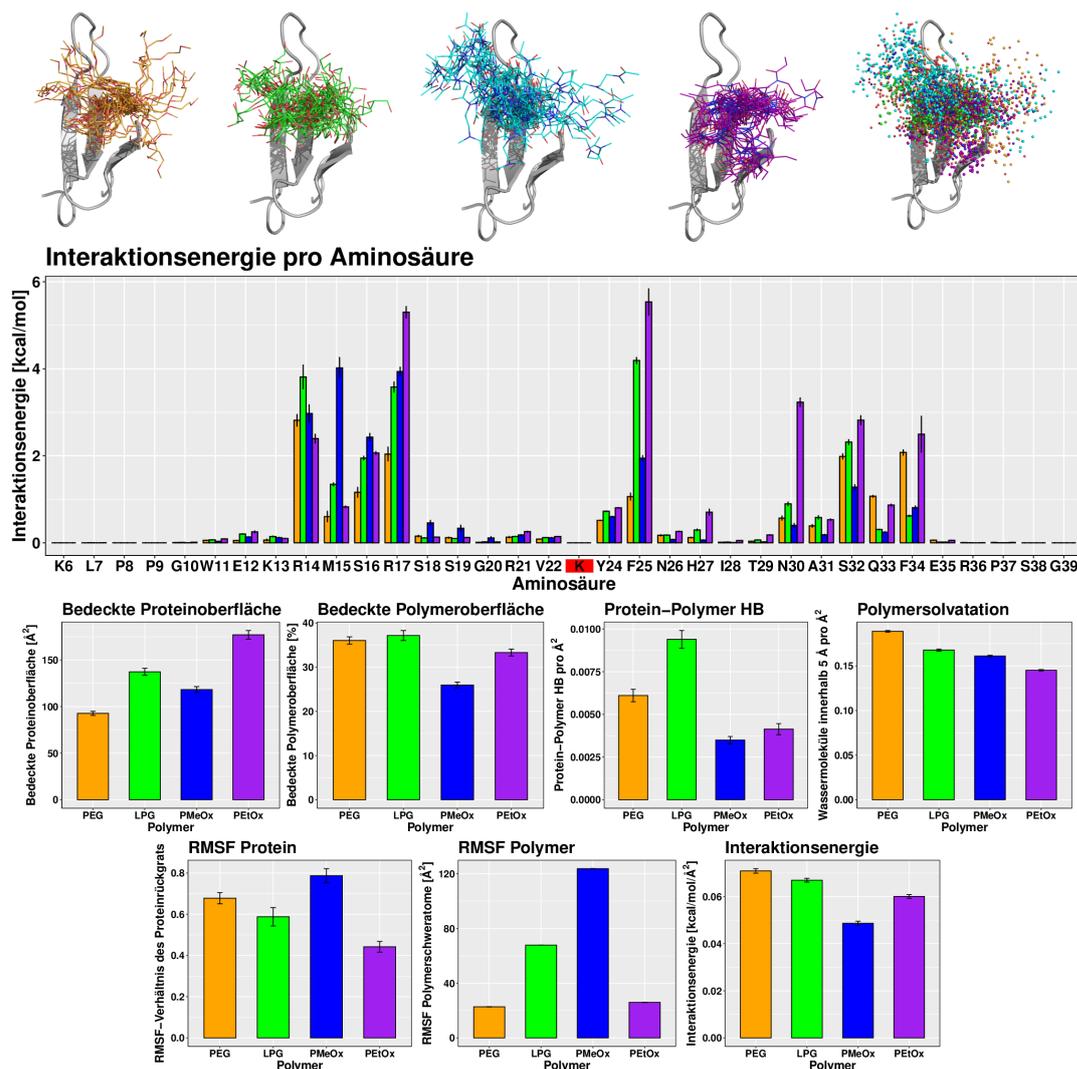


Abbildung 4.36: Protein-Polymer-Dynamiken an Konjugationsstelle 23 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.8 Konjugationsstelle 26

Mit Ausnahme einiger weniger Momentaufnahmen verharrt PEG in der Nähe des hydrophoben Zentrums um W11. Stabile HB mit den Seitenketten von W11 und Q33 sind registrierbar (Abb. 4.37-A). Auch wenn LPG ein flexibleres Verhalten zeigt, so nimmt es doch vergleichbare Konformationen ein. HB zu W11, T29, Q33, S38 und zur C-terminalen Carboxylgruppe sind zu beobachten (Abb. 4.37-B). PMeOx ist nicht so nah zu Q33 wie PEG und LPG positioniert, kann jedoch auch W11 desolvatisieren (Abb. 4.37-C). Der Amidwasserstoff des Linkers bildet eine stabile Orientierung zum Aromaten von W11, während die entsprechende Carbonylgruppe mit dem Rückgrat von I28 und T29 wechselwirkt. Das Polymer bleibt hingegen relativ flexibel und streckt sich manchmal zum C-Terminus aus, teilweise auch zu den hydrophoben Aminosäuren P8, P9 und I28. An dieser Konjugationsstelle verhält sich PEtOx identisch (Abb. 4.37-D).

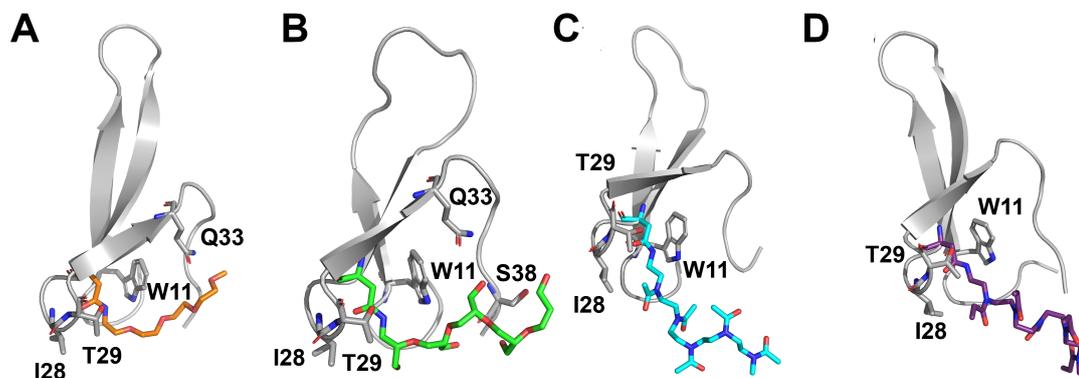


Abbildung 4.37: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 26 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

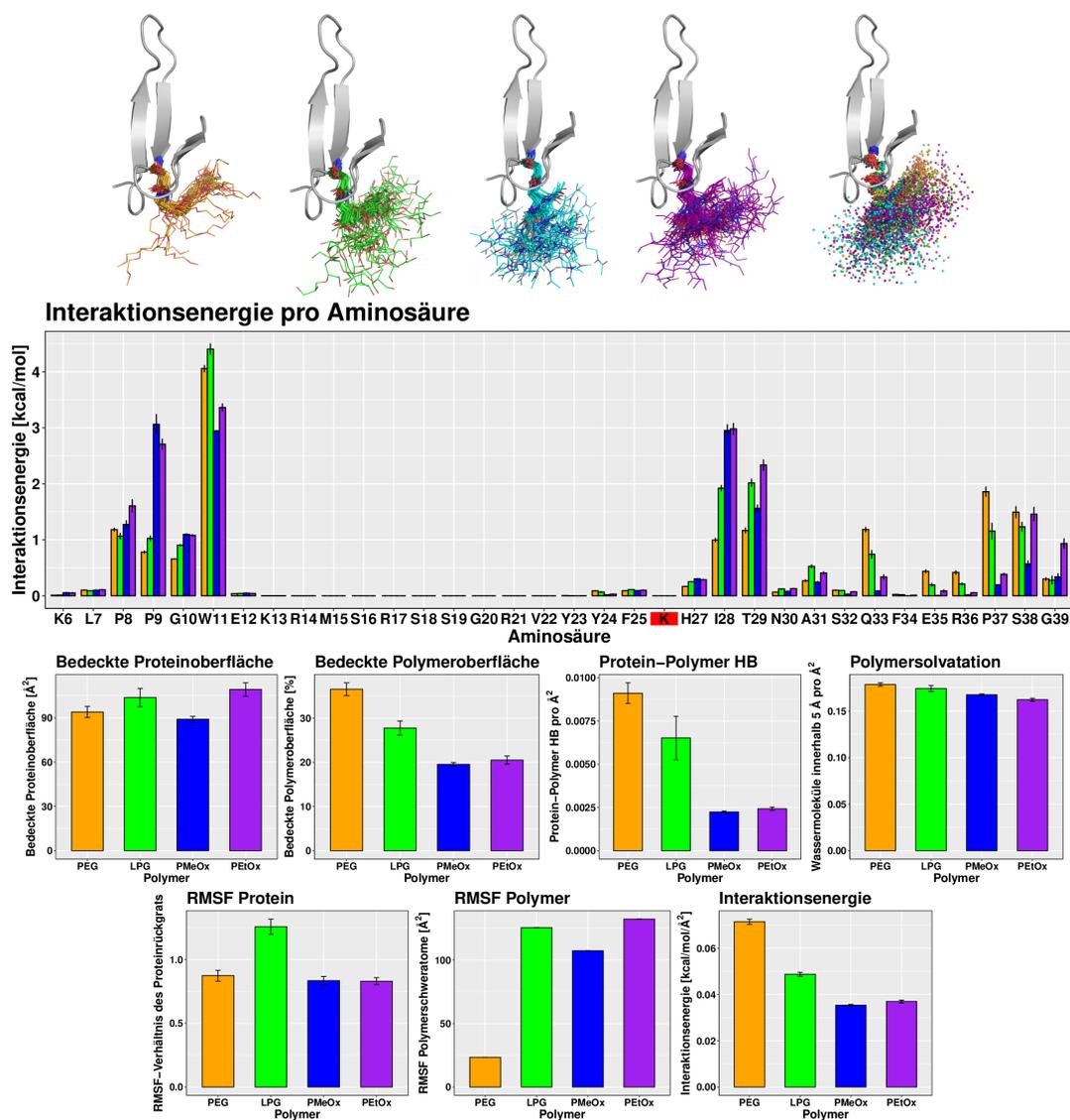


Abbildung 4.38: Protein-Polymer-Dynamiken an Konjugationsstelle 26 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.9 Konjugationsstelle 27

PEG interagiert zunächst mit R14, anschließend über seinen Linker mit der Carboxylgruppe von E12 und über die Ethersauerstoffatome mit dem Rückgratamid von P9 (Abb. 4.39-A). Die Wechselwirkung mit E12 ist über den Großteil der Zeit hinweg zu beobachten, während die Polymerkette sehr flexibel bleibt und teilweise I28 adressiert. Im Falle von LPG etabliert der Linker dieselbe Interaktion, während sich die Polymerkette zum N-Terminus ausstreckt und hauptsächlich solvensexponiert bleibt (Abb. 4.39-B). Teilweise sind jedoch auch HB zwischen den Hydroxylgruppen von LPG und E12 zu erkennen. Der Linker wechselwirkt auch im Falle von PMeOx mit E12 (Abb. 4.39-C). Das Polymer adressiert zum Teil P9. Im Gegensatz zu allen anderen Polymeren scheint PEtOx mit seiner Länge Y23 und F25 des oberen hydrophoben Zentrums erreichen zu können. Dabei wechselwirkt es auch mit R14 und R17 (Abb. 4.39-D). Die Bedeckung der Proteinoberfläche ist somit bei diesem Polymer am größten (Abb. 4.40).

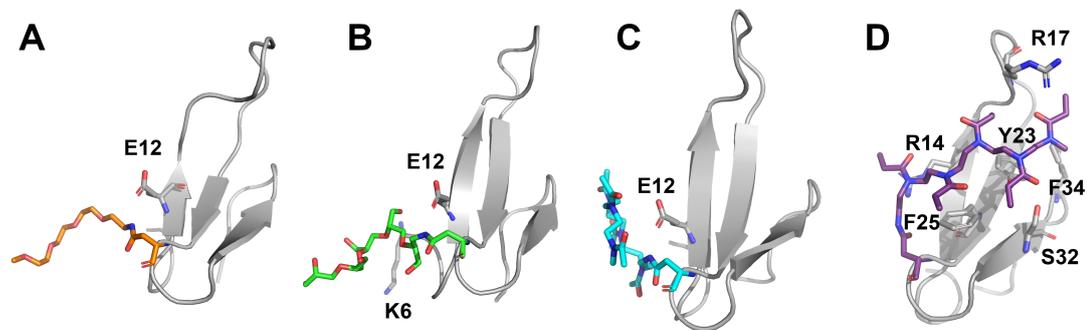


Abbildung 4.39: Beispielfotografien von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 27 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymerernähe sind ebenso dargestellt.

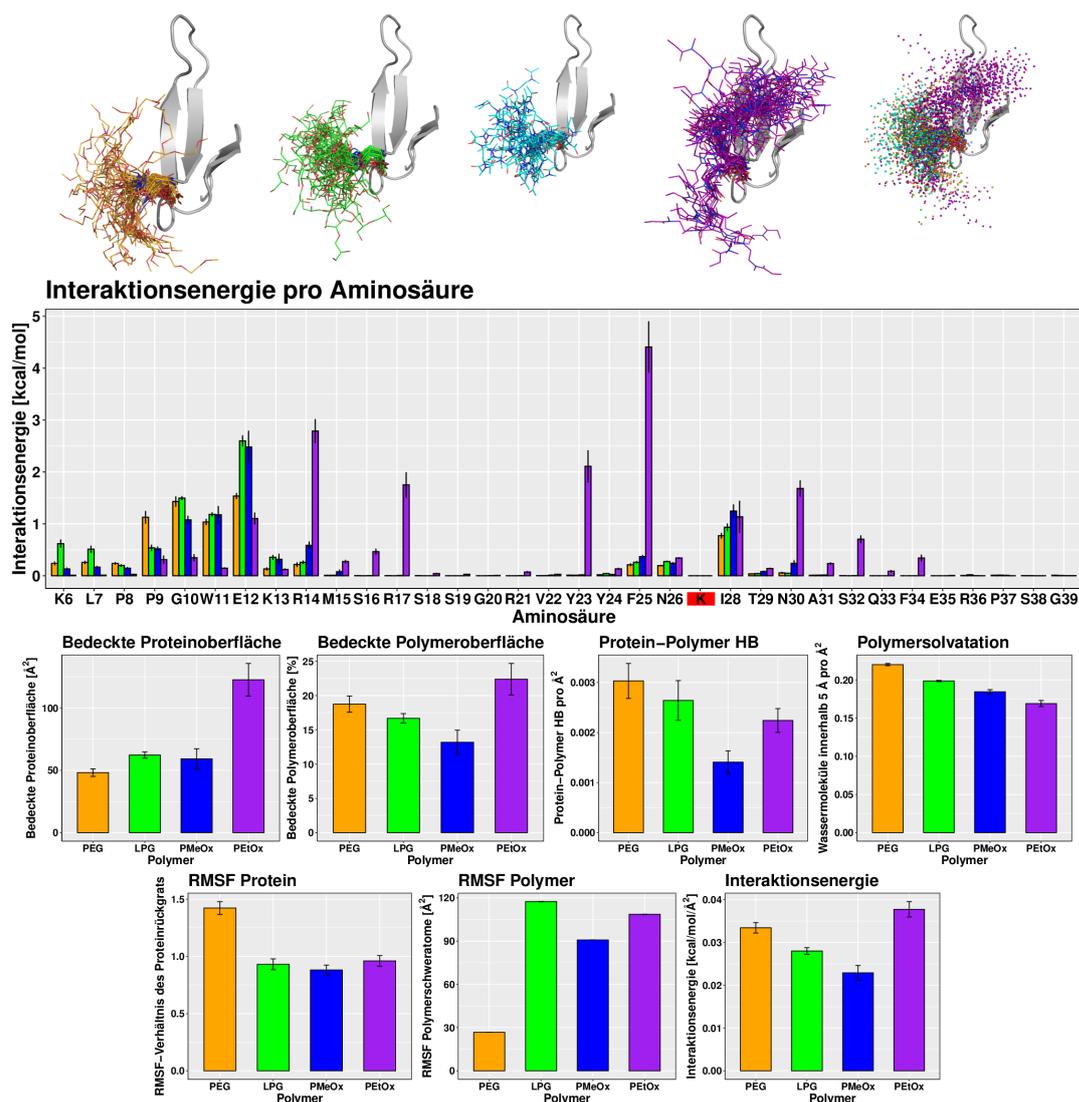


Abbildung 4.40: Protein-Polymer-Dynamiken an Konjugationsstelle 27 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.10 Konjugationsstelle 28

PEG orientiert sich entweder zum Rückgratamid von P9 oder, was häufiger der Fall ist, in Richtung W11 und N26 (Abb. 4.41-A). LPG verhält sich ähnlich (Abb. 4.41-B), kann jedoch stärkere Interaktionen mit H27 etablieren. PMeOx bildet keine Wechselwirkungen mit dem Protein aus, mit Ausnahme der Seitenkette der benachbarten Aminosäure T29 (Abb. 4.41-C). PEtOx verbleibt solvensexponiert und sehr flexibel (Abb. 4.41-D). Das Interaktionsprofil verdeutlicht stärkere Wechselwirkungen von PEG und LPG mit den N-terminalen Aminosäuren im Vergleich zu den POx-Polymeren.

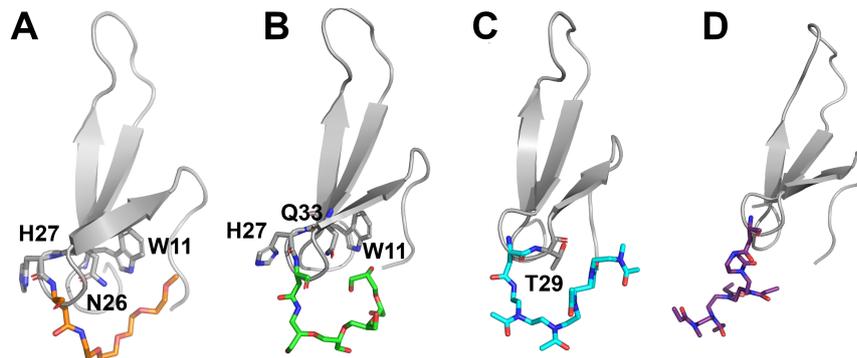


Abbildung 4.41: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 28 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

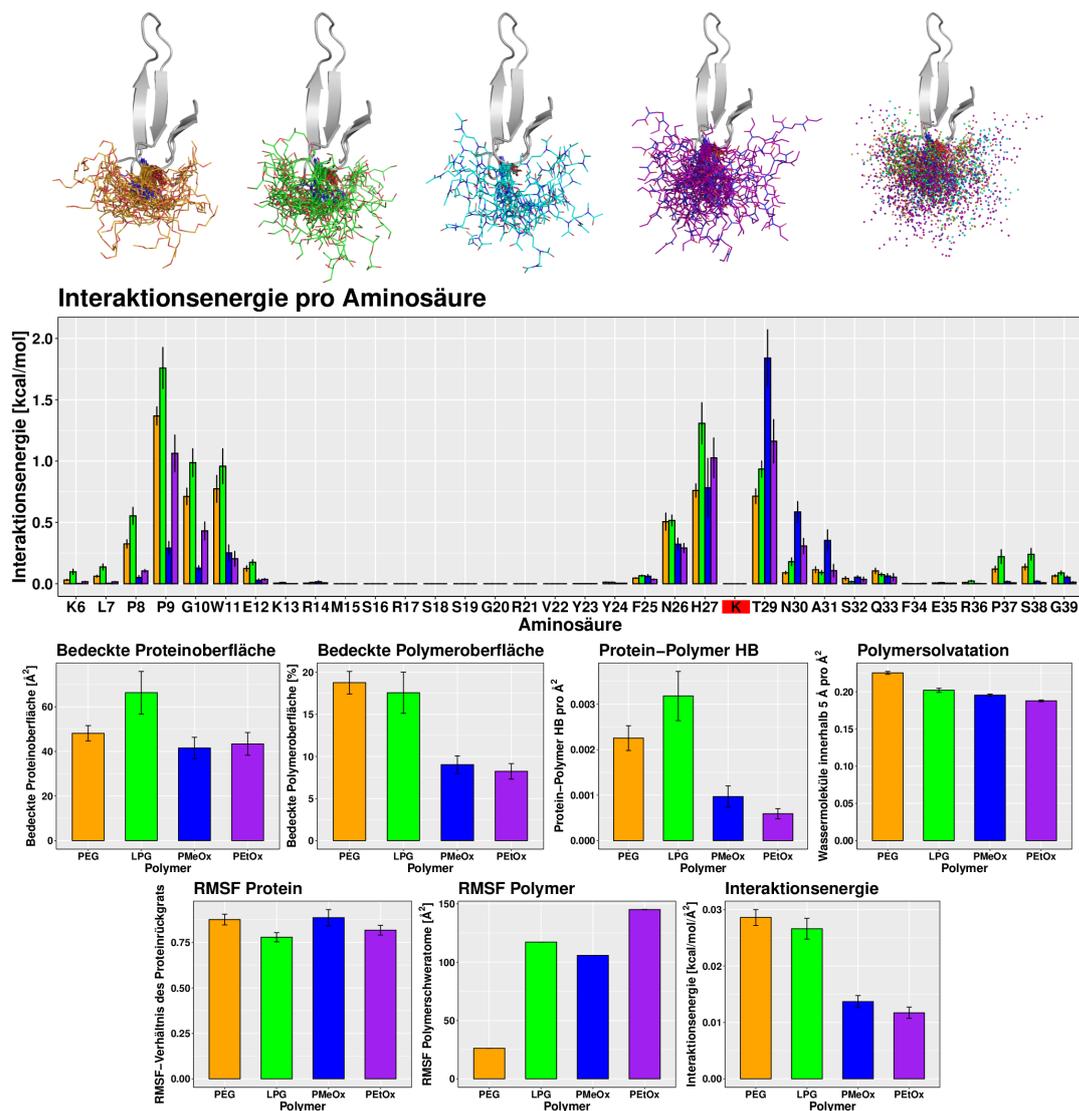


Abbildung 4.42: Protein-Polymer-Dynamiken an Konjugationsstelle 28 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.11 Konjugationsstelle 29

Mit Ausnahme einiger weniger Momentaufnahmen etabliert PEG abermals Interaktionen mit W11 und Q33 des unteren hydrophoben Zentrums (Abb. 4.43-A). LPG zeigt ein ähnliches Verhalten (Abb. 4.43-B). PMeOx ist flexibler und wechselwirkt seltener mit W11 und Q33. Dagegen adressiert es oft die hydrophoben Seitenketten P9 und I28 (Abb. 4.43-C). Im Gegensatz hierzu verbleibt PEtOx in einer stabilen, geknäuelten Konformation in der Nähe von W11 und N26 (Abb. 4.43-D), was auch am RMSF-Wert erkennbar ist (Abb. 4.44).

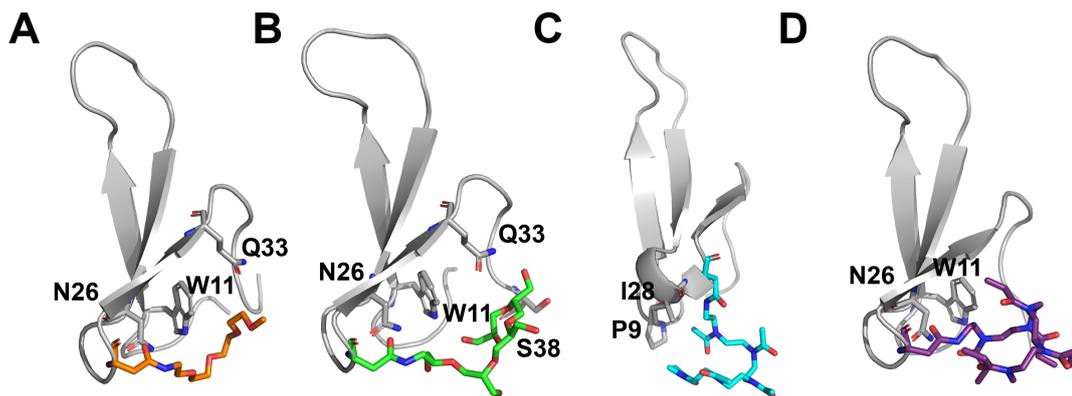


Abbildung 4.43: Beispielfähige Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 29 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

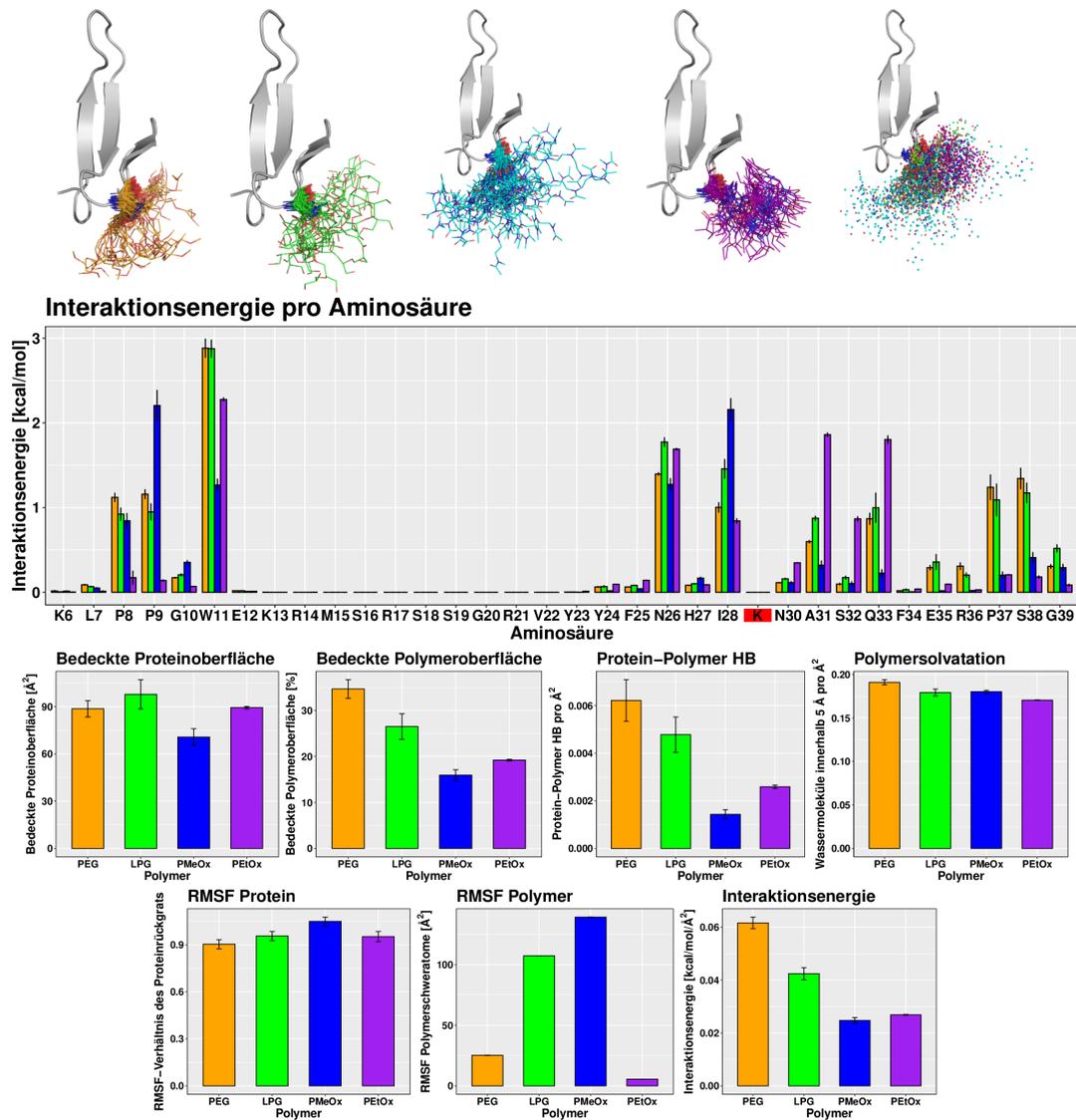


Abbildung 4.44: Protein-Polymer-Dynamiken an Konjugationsstelle 29 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.12 Konjugationsstelle 30

PEG zeigt ein äußerst flexibles Verhalten und etabliert lediglich mit S32 länger andauernde Interaktionen (Abb. 4.45-A). LPG fluktuiert ebenso, bevor es Wechselwirkungen mit S16, R17, Y23 und S32 in der zweiten Hälfte der Trajektorie zeigt (Abb. 4.45-B). PMeOx verbleibt auf der Oberfläche von Y23 und F25, interagiert jedoch, im Gegensatz zu LPG, nicht mit Aminosäuren der ersten Schleife (Abb. 4.45-C). Zusätzlich zu Y23 und Y25 interagiert es auch über eine Carbonylgruppe mit R14. Im Gegensatz zu den anderen Polymeren richtet sich PEtOx nicht entlang der Proteinoberfläche aus, sondern um I28 und den Aromaten von H27 (Abb. 4.45-D). PMeOx zeigt, basierend auf diesen Beschreibungen, die größte Oberflächenbedeckung (Abb. 4.46).

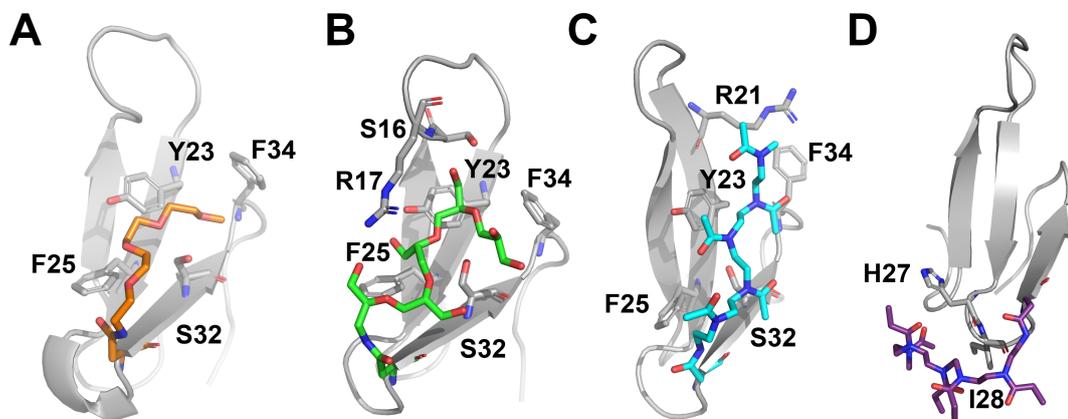


Abbildung 4.45: Beispielfotografien von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 30 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

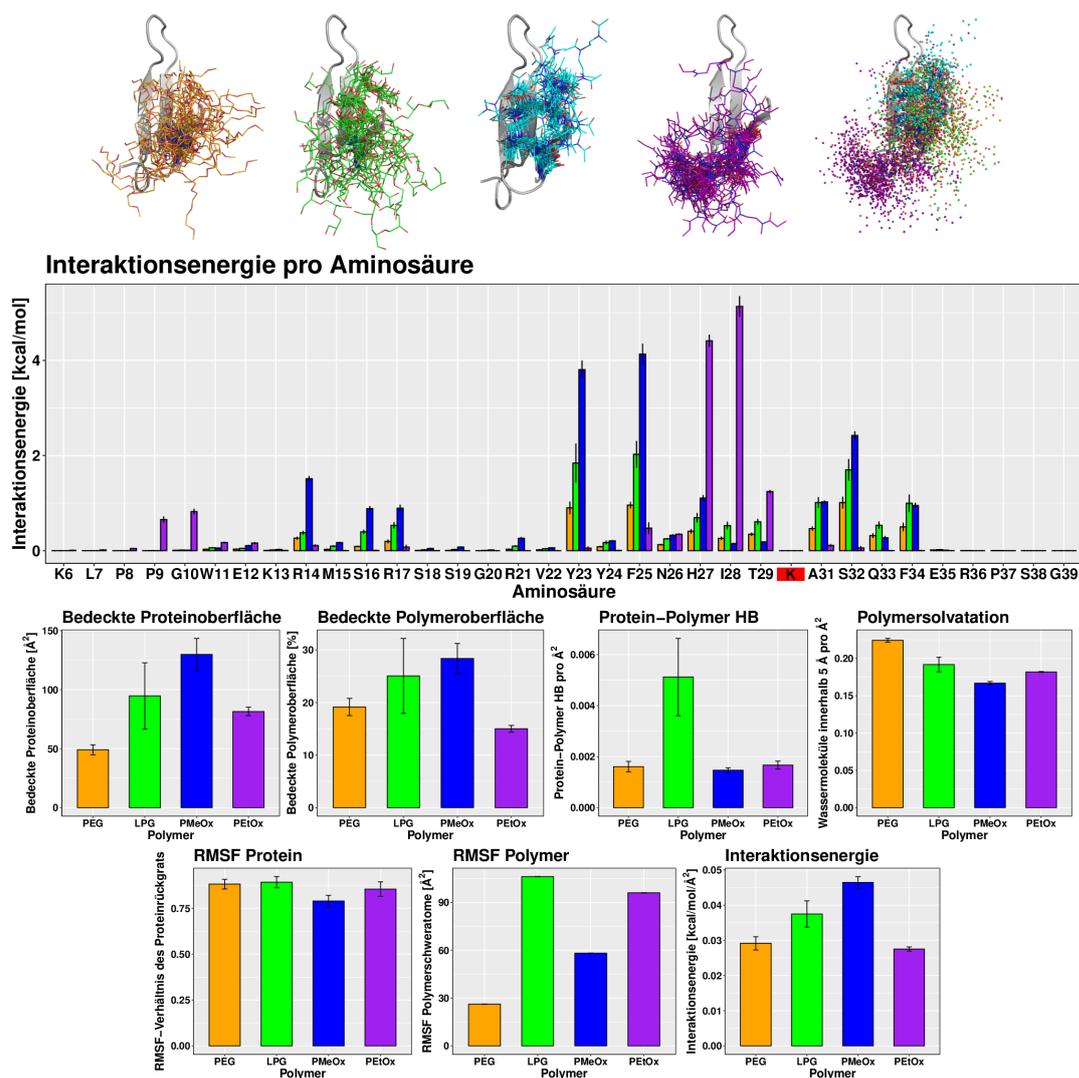


Abbildung 4.46: Protein-Polymer-Dynamiken an Konjugationsstelle 30 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.1.13 Konjugationsstelle 32

PEG ist zwischen den Aminosäuren Y23 und F34 positioniert und streckt sich zu S16 und R17 aus (Abb. 4.47-A). In manchen Momentaufnahmen wechselt es seine Konformation in Richtung des dritten β -Faltblatts und interagiert mit dem Amidrückgrat von F34. Ähnliche Konformationen hin zur ersten Schleife können im Falle von LPG beobachtet werden (Abb. 4.47-B), wobei auch Momentaufnahmen mit kompakten Polymerkonformationen zu registrieren sind. PMeOx streckt sich zu R17 aus und ist imstande, die Aromaten von Y23 und F34 über die gesamte Simulationszeit hinweg zu adressieren (Abb. 4.47-C). PEtOx ist rigider, etabliert aber ähnliche Interaktionen (Abb. 4.47-D). Beide POx-Varianten befinden sich auf der Oberfläche des zweiten und dritten β -Faltblatts und führen somit zu einer starken Bedeckung der Proteinoberfläche (Abb. 4.48).

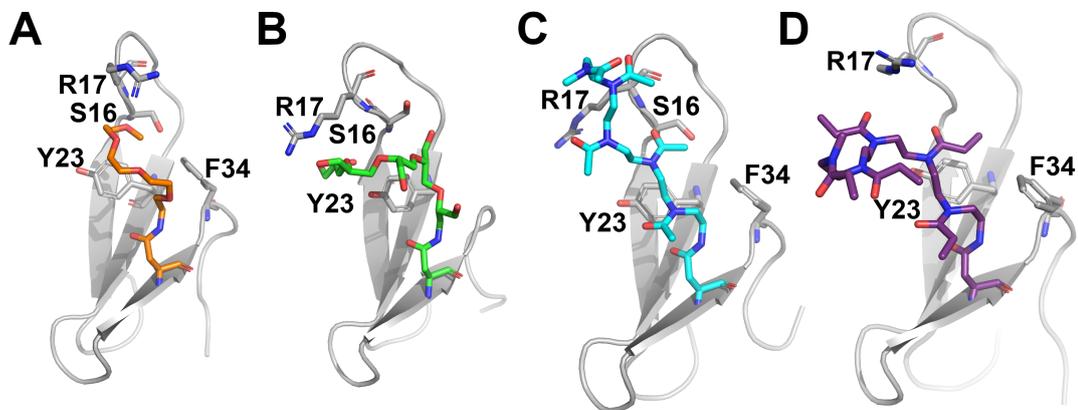


Abbildung 4.47: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 32 von Pin 1 WW. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

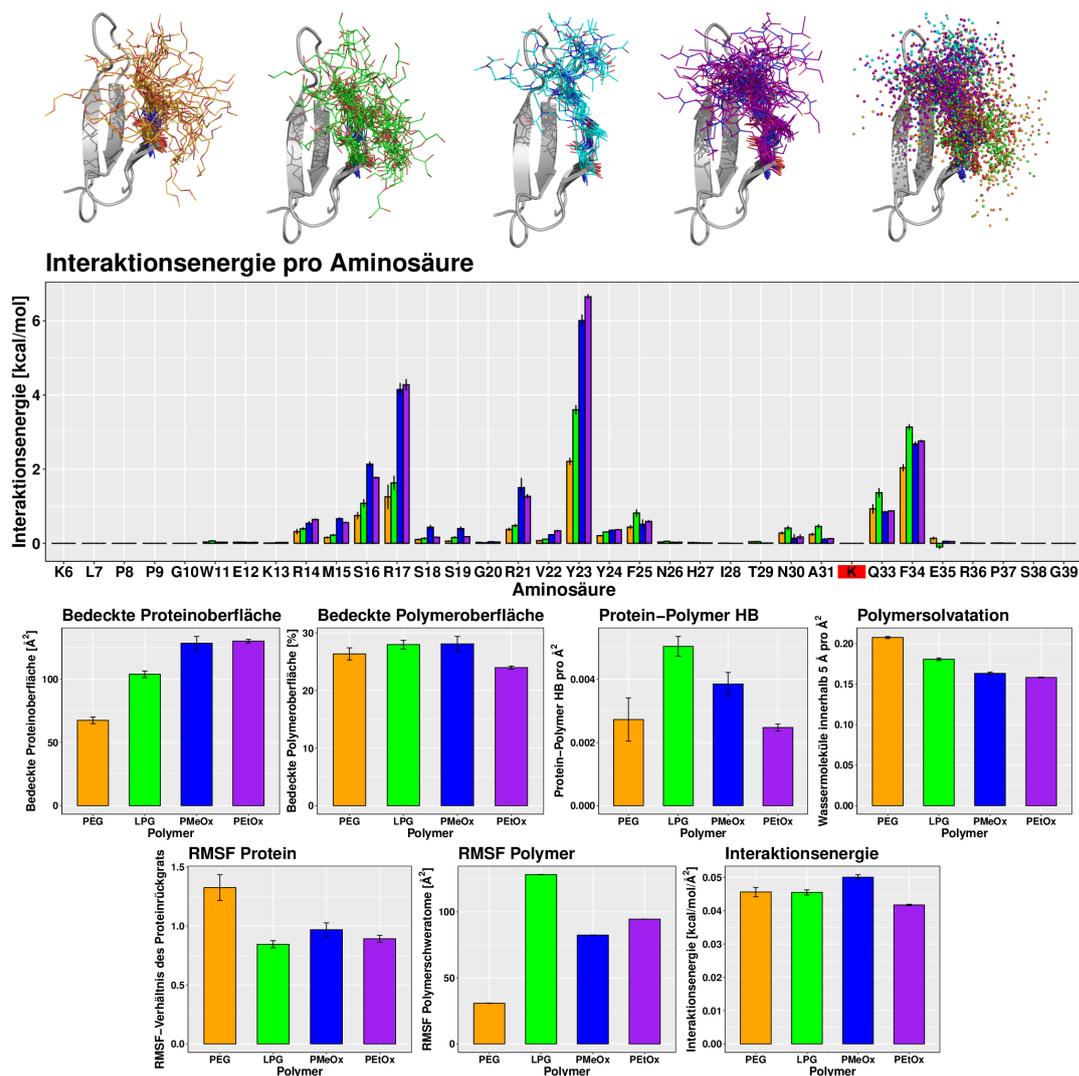


Abbildung 4.48: Protein-Polymer-Dynamiken an Konjugationsstelle 32 von Pin 1 WW. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.2 Her2-Konjugate

4.3.2.1 Konjugationsstelle 7

PEG orientiert sich zunächst um R10 herum und interagiert ebenso mit Y13, W14 und Y35 (Abb. 4.49-A). In manchen Momentaufnahmen verbleibt das Polymer solvensexponiert oder adressiert V1 und N3 am N-Terminus. LPG ist nicht um R10 herum positioniert, stattdessen wechselwirkt es mit der Seitenkette von E8 und der Rückgratcarbonylgruppe von S39 (Abb. 4.49-B). Während der zweiten Hälfte der Trajektorie wird R10 durch eine Carbonylgruppe des Linkers adressiert, während sich das Polymer zu N11 und W14 hin umorientiert. PMeOx ist zunächst oberhalb der dritten α -Helix zu N43 hin ausgestreckt. Danach findet es sich relativ solvensexponiert im Bereich des N-Terminus wieder, bevor es, ähnlich wie PEG, wiederum R10, Y13, W14 und Y35 adressiert (Abb. 4.49-C). PEtOx verbleibt über die gesamte Simulationszeit hinweg in der Nähe des Aromaten von W14 parallel zur α -Helix, während das Amid des Linkers teilweise mit der Carbonylgruppe von N11 wechselwirkt (Abb. 4.49-D). Mit W14 zeigt es die mit Abstand größte Interaktionsenergie.

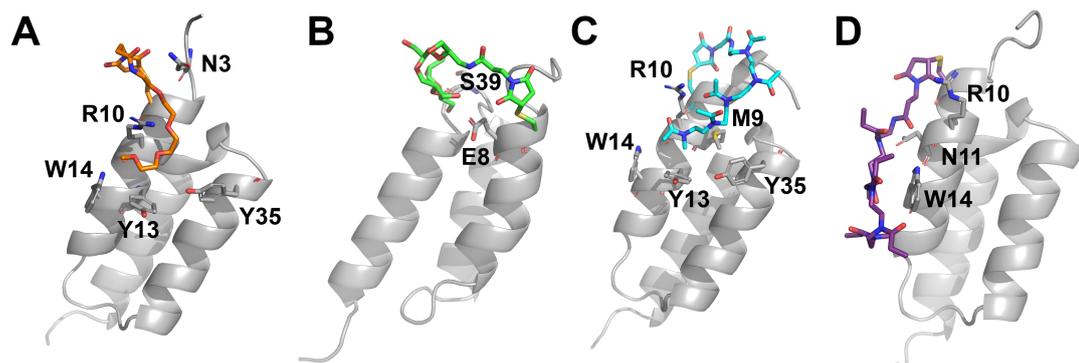


Abbildung 4.49: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 7 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

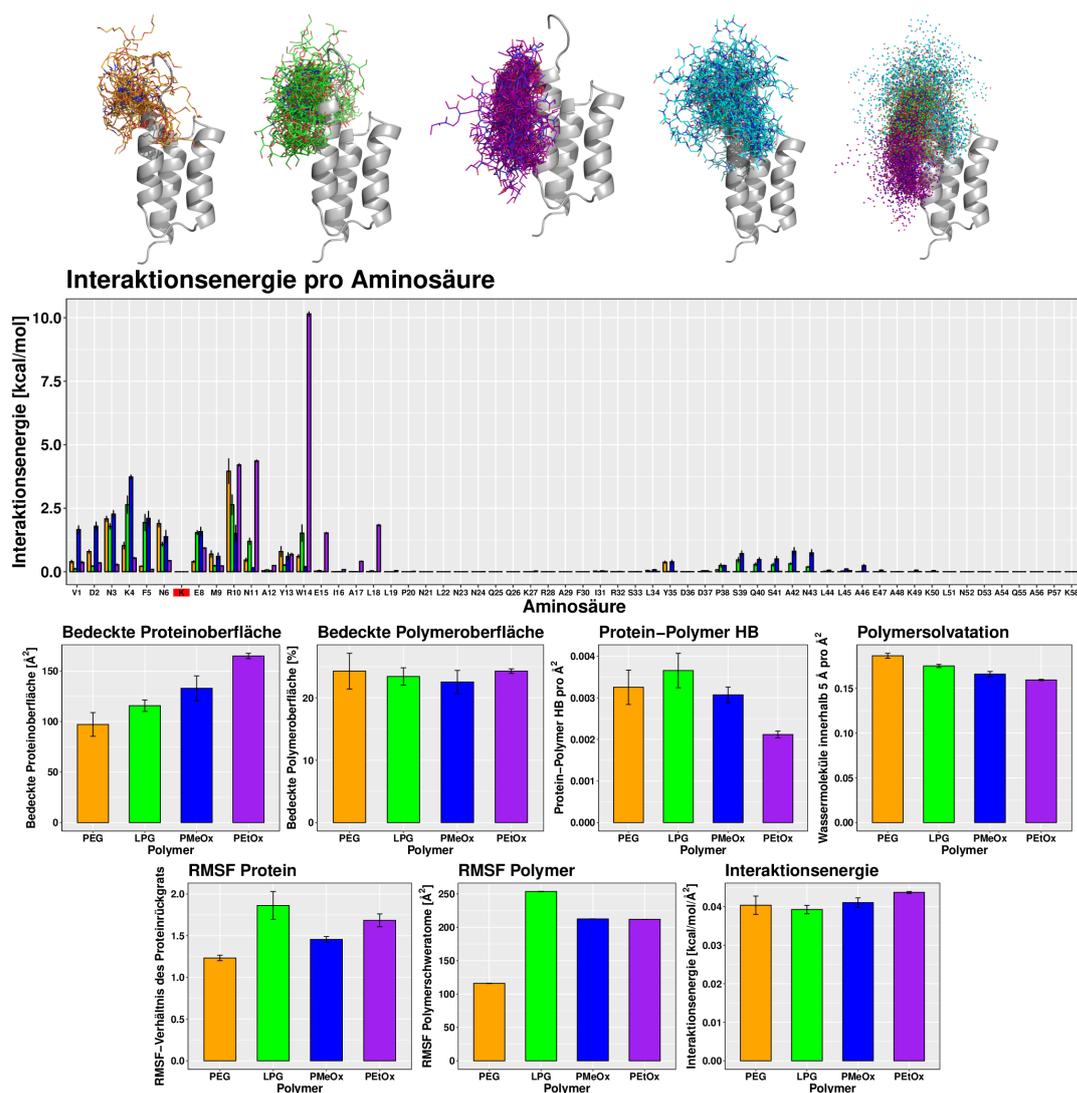


Abbildung 4.50: Protein-Polymer-Dynamiken an Konjugationsstelle 7 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als K auf rotem Grund gekennzeichnet.

4.3.2.2 Konjugationsstelle 8

Zunächst liegt PEG geknäuelte vor und adressiert die Seitenkette von N11 mit dem terminalen Monomer. Danach streckt es sich zu K49 und K50 aus (Abb. 4.51-A). Letzteres interagiert entweder mit dem Polymer oder dem Linker des Konjugats. LPG ist sehr flexibel und adressiert entweder eigene Strukturbereiche oder Aminosäuren in der Nähe des N-Terminus, v. a. V1, D2 und K4 (Abb. 4.51-B). Gegen Ende der Simulation streckt es sich nach unten zur Helix aus, in die Nähe von N11, E15 und K49. Es wechselt dann die Richtung und wechselwirkt mit der Seitenkette von D37 über seinen Linker und dem Rückgrat von Y35 über eine der Hydroxylgruppen. PMeOx verbleibt zunächst nicht zielgerichtet auf der Proteinoberfläche, hauptsächlich nahe dem N-Terminus. Später jedoch etabliert es eine relativ stabile Konformation um die Seitenkette von N43 herum und adressiert mit den terminalen Monomeren hydrophobe Areale von A29, F30 und L44 (Abb. 4.51-C). PEtOx wechselwirkt mit dem Seitenkettenamid von N11 in einer geknäuelten Konformation. Über einige ausgestreckte Orientierungen hinweg windet es sich anschließend um den Aromaten von F5 und wechselwirkt mit N3 über seine terminale Carbonylgruppe, was zu einer Deformation der ersten Helix führt (Abb. 4.51-D). Entsprechend ist die Interaktionsenergie mit F5 an dieser Stelle am größten.

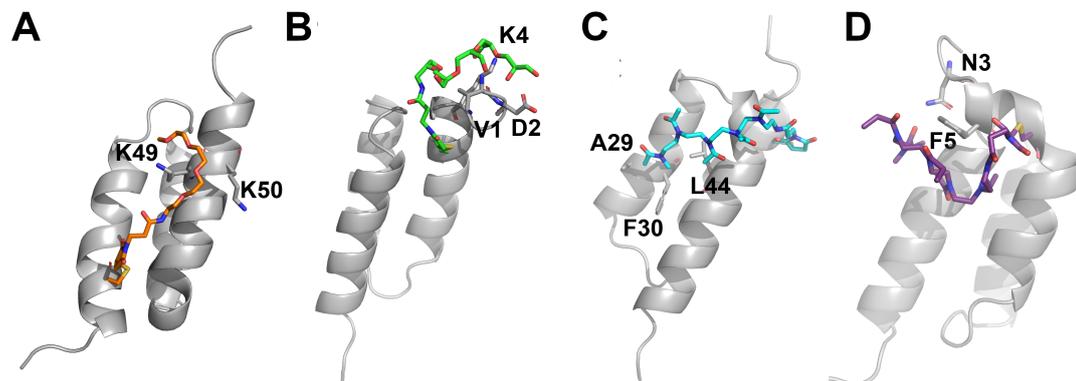


Abbildung 4.51: Beispielfotografien von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 8 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

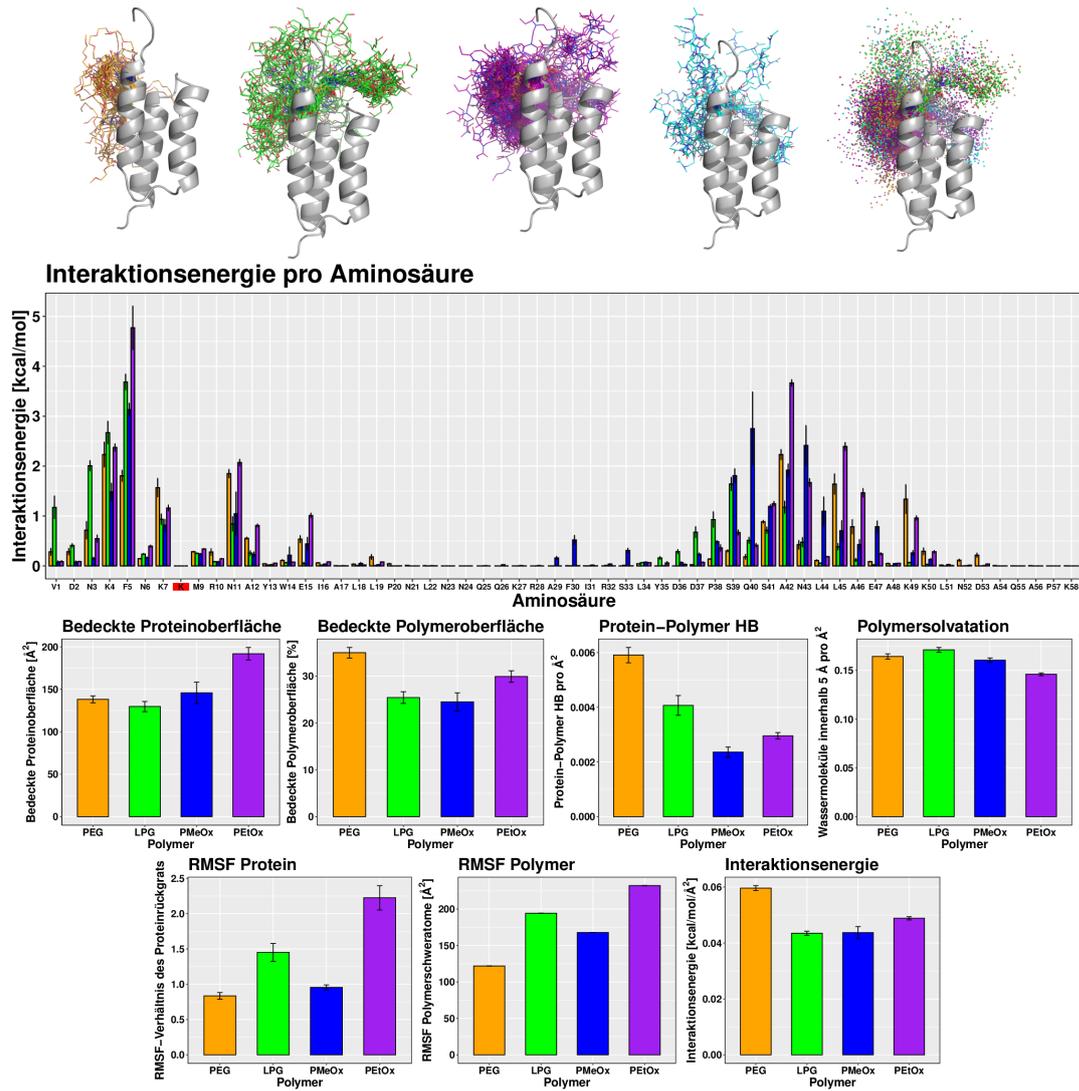


Abbildung 4.52: Protein-Polymer-Dynamiken an Konjugationsstelle 8 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEIOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.2.3 Konjugationsstelle 15

PEG wechselwirkt zunächst mit K7 und R10 (Abb. 4.53-A). In der zweiten Hälfte der Trajektorie verbleibt das Polymer vornehmlich in einer geknäuelten oder solvens-exponierten Konformation, interagiert aber auch teilweise mit den Seitenketten von K49 und N52. LPG ist zunächst in der Nähe des C-Terminus um die Aminosäuren N21, N52, D53 und K58 positioniert. Dann adressiert es K7, R10, N11 und vor allem E8 (Abb. 4.53-B). Am Ende orientiert es sich wieder zum C-Terminus hin. PMeOx verbleibt zumeist in der Nähe von N11 mit seinem Linker, sowie um R10 und W14 mit den Monomeren (Abb. 4.53-C). Gegen Ende streckt es sich aus zu einem hydrophoben Areal, wobei es mit K7, M9, Y13 und Y35 adressiert. In ähnlicher Weise interagiert auch der Linker von PEtOx mit N11, während der Polymer-Part oberhalb der Aromaten von Y13 und W14 liegt, dabei wechselwirkt es auch mit Y35 (Abb. 4.53-D).

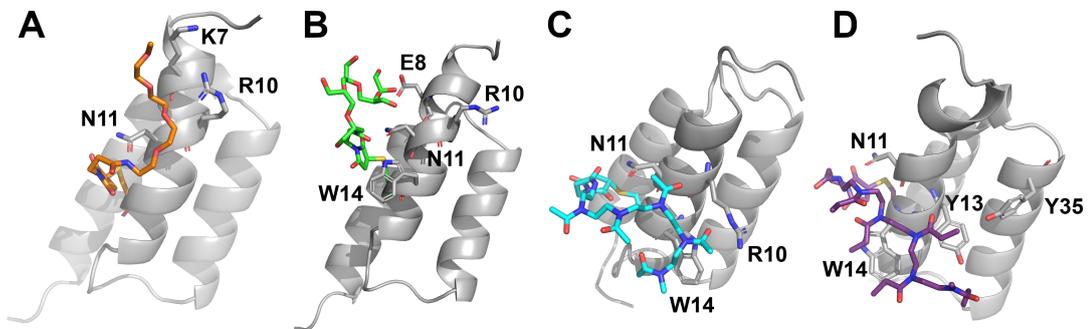


Abbildung 4.53: Beispielfotografien von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 15 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

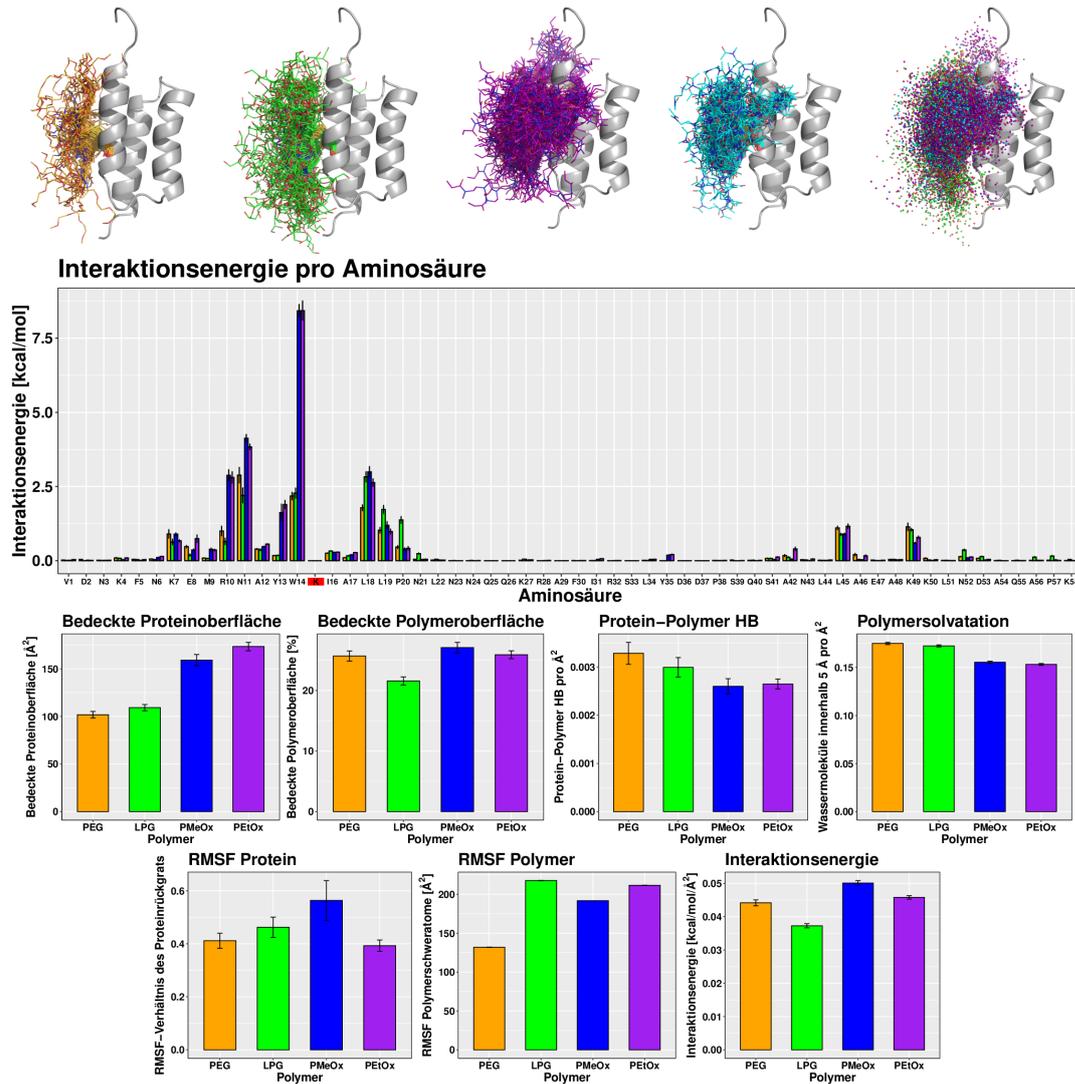


Abbildung 4.54: Protein-Polymer-Dynamiken an Konjugationsstelle 15 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als K auf rotem Grund gekennzeichnet.

4.3.2.4 Konjugationsstelle 21

PEG windet sich zunächst um K49. Nach nur 30 ns dreht sich der Linker um, was zu einer Polymerkonformation fernab der Helices führt. Interaktionen mit den Seitenketten von N23, Q26 und N55 können beobachtet werden (Abb. 4.55-A). Ähnlich zu PEG hält sich auch LPG hauptsächlich unterhalb des helikalen Parts des Proteins auf und adressiert dabei N23, Q26, N24 und das Rückgrat von L22 (Abb. 4.55-B). Es streckt sich auch entlang des C-Terminus aus und wechselwirkt mit K58. PMeOx scheint über lange Zeit hinweg keine spezifische Aminosäure zu adressieren (Abb. 4.55-C). Es hält sich vornehmlich in einer geknäuelten Konformation unterhalb des helikalen Parts des Proteins auf. Manchmal streckt sich das Polymer zu den hydrophoben Aminosäuren L18 und L19 sowie zur Seitenkette von N52 aus. PEtOx zeigt ebenso geknäuelte Konformationen nahe N23, Q26 und Q55 (Abb. 4.55-D).

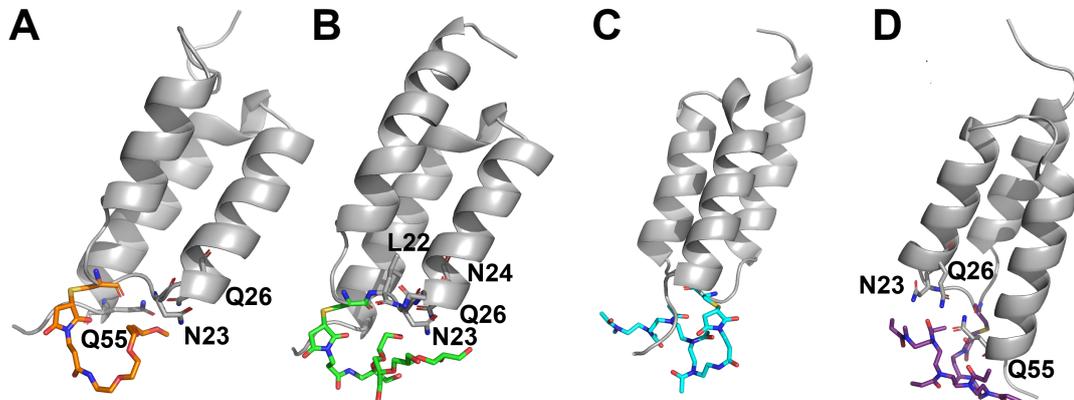


Abbildung 4.55: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 21 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäure in Polymernähe sind ebenso dargestellt.

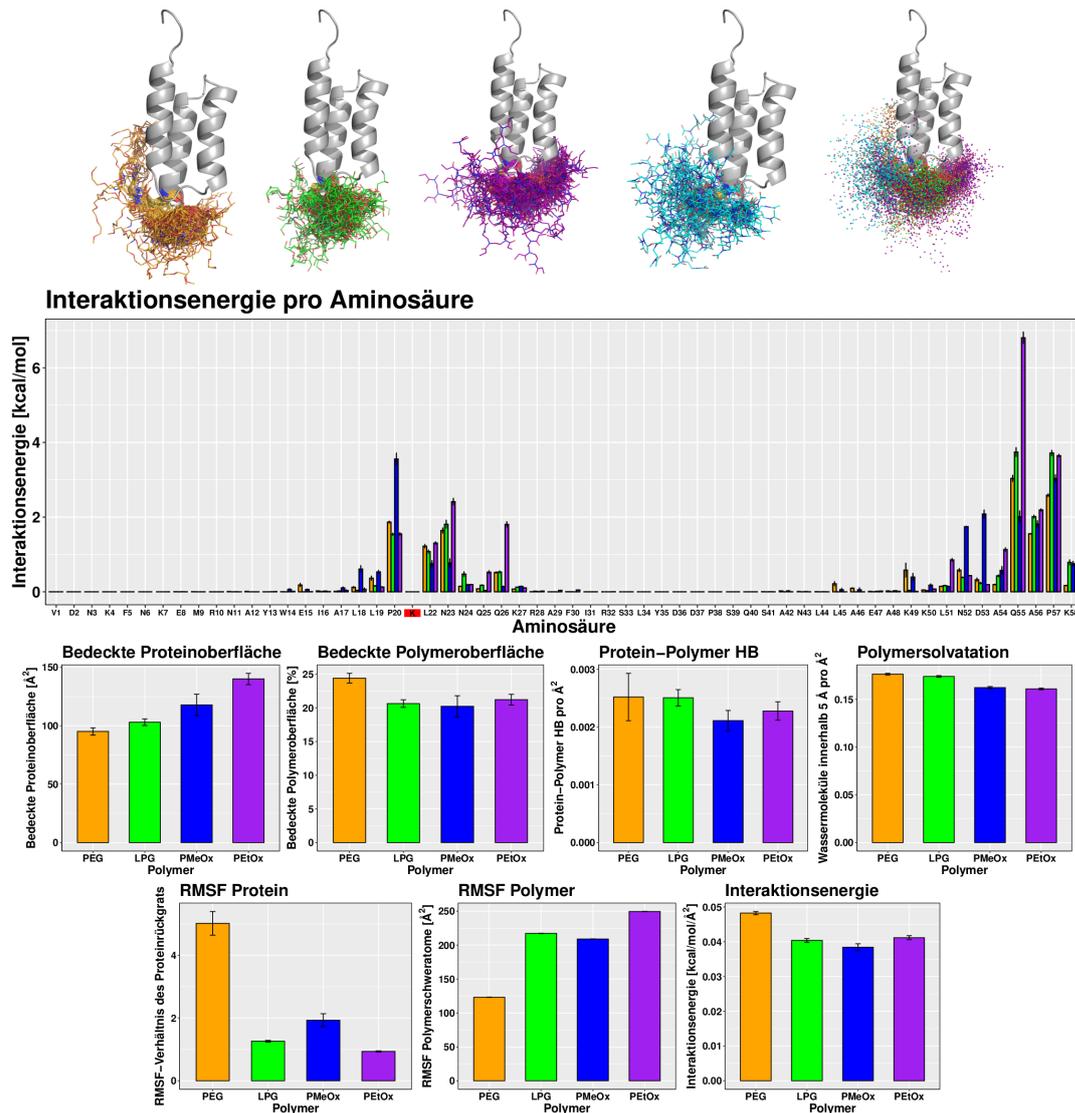


Abbildung 4.56: Protein-Polymer-Dynamiken an Konjugationsstelle 21 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als K auf rotem Grund gekennzeichnet.

4.3.2.5 Konjugationsstelle 23

PEG schützt Q25, Q26 und Q55 in einer geknäuelten Konformation (Abb. 4.57-A). Erst gegen Ende der Simulation adressiert es F30 und L44. LPG wechselwirkt mit einer Vielzahl von Aminosäuren: A17, Q25, Q26 und S33 (Abb. 4.57-B), gegen Ende auch mit Q47. Im Falle von PMeOx werden zunächst auch Aminosäuren in der Nähe des C-Terminus adressiert (Abb. 4.57-C), am Ende die hydrophoben Aminosäuren F30 und L44. Das größere PEOx hält sich im selben Areal in der Nähe von N23, Q26 und Q55 auf (Abb. 4.57-D) und zeigt starke Wechselwirkungen mit den hydrophoben Aminosäuren A29 und F30.

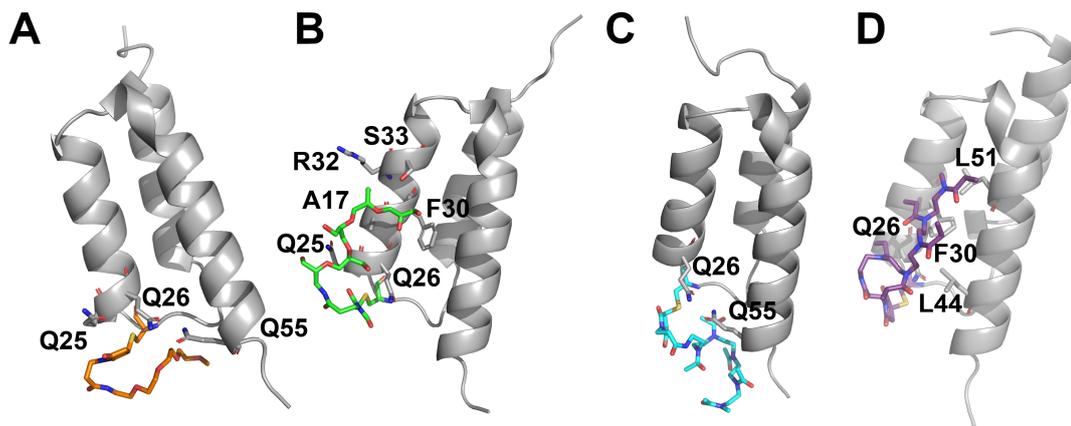


Abbildung 4.57: Beispielfotografien von (A) PEG, (B) LPG, (C) PMeOx und (D) PEOx an der Konjugationsstelle 23 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

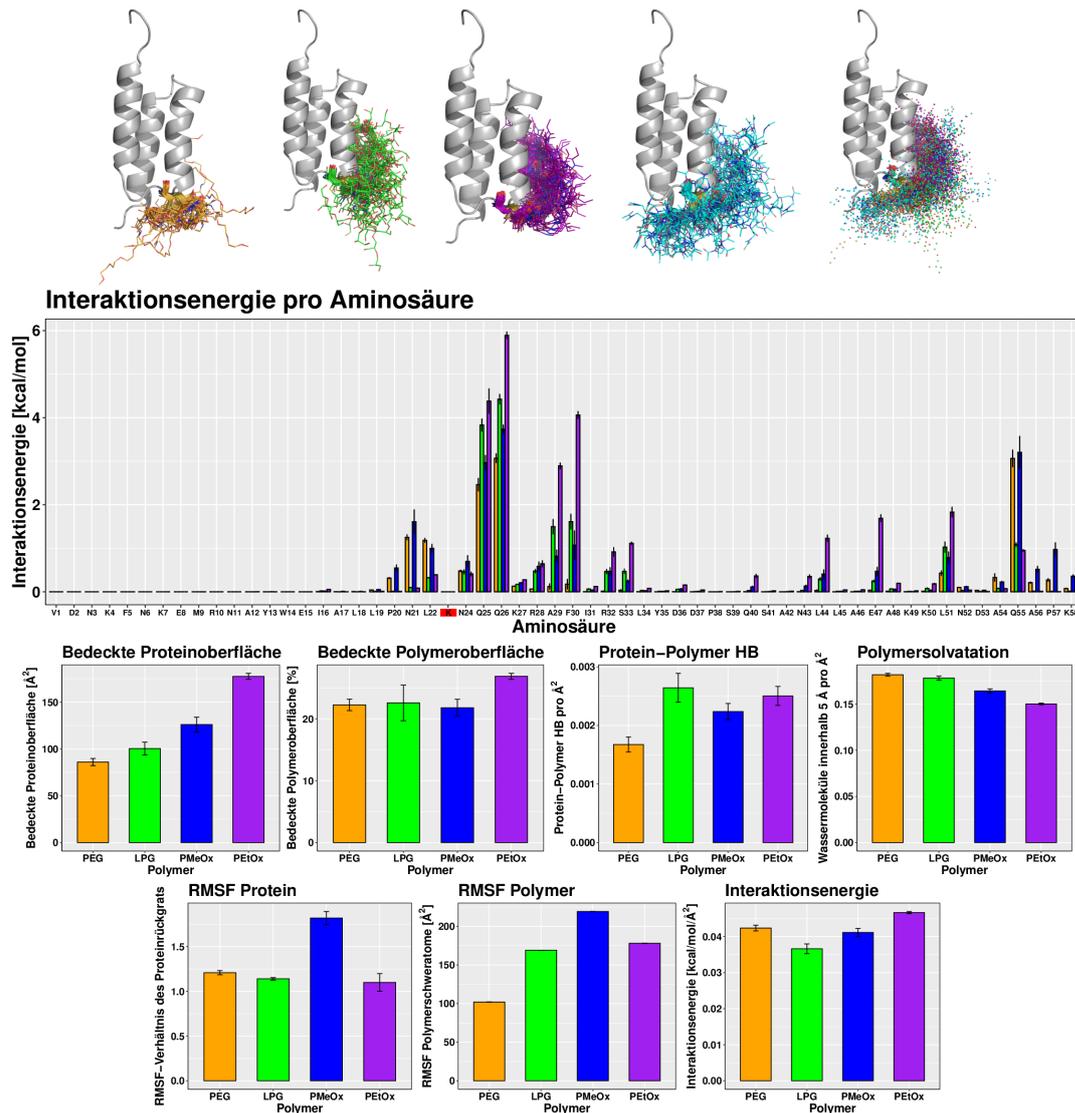


Abbildung 4.58: Protein-Polymer-Dynamiken an Konjugationsstelle 23 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als K auf rotem Grund gekennzeichnet.

4.3.2.6 Konjugationsstelle 25

PEG hält sich im selben hydrophoben Areal wie bei Konjugationsstelle 23 auf: über Q26 hin zu F30 und L44 (Abb. 4.59-A). LPG verbleibt unterhalb der Helices und schützt N23, Q26 und Q55 (Abb. 4.59-B), streckt sich aber auch manchmal aus und interagiert mit K58. Genau wie PEG verbleibt auch PMeOx nahe F30 und L44 (Abb. 4.59-C). Im Gegensatz hierzu streckt sich das größere PEtOx über diese Aminosäuren hinweg zu einem weiteren hydrophoben Areal aus: M9, Y13, I31 und Y35 (Abb. 4.59-D).

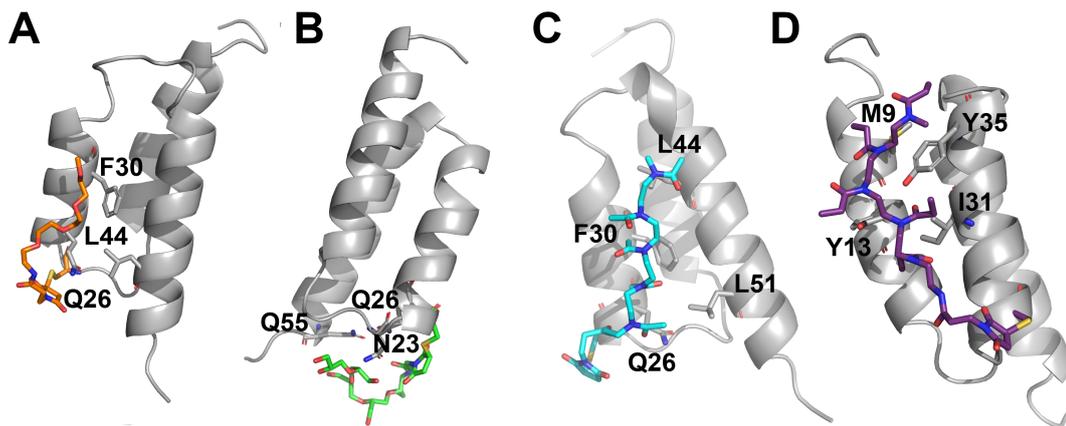


Abbildung 4.59: Beispielfotografien von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 25 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

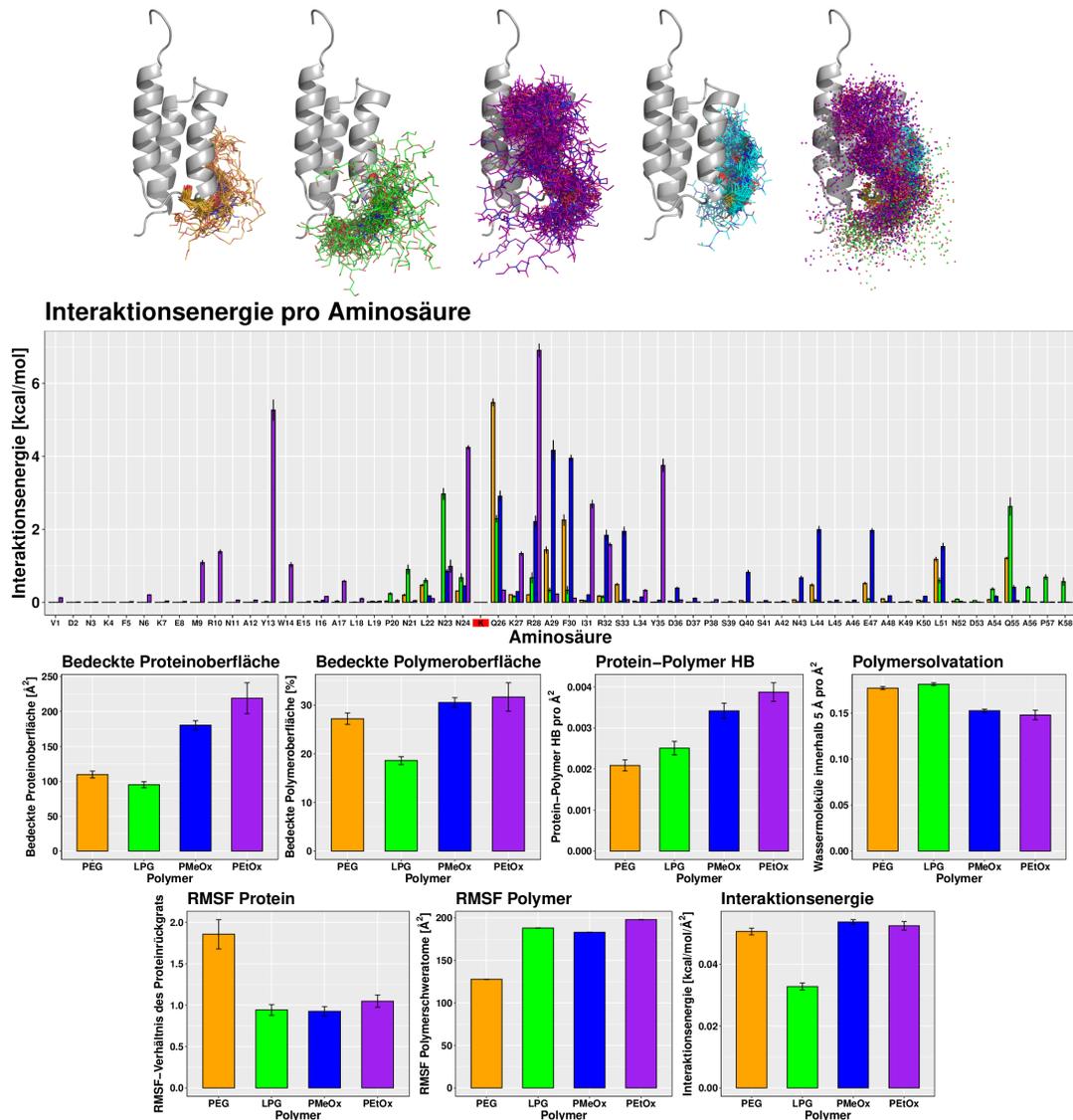


Abbildung 4.60: Protein-Polymer-Dynamiken an Konjugationsstelle 25 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als K auf rotem Grund gekennzeichnet.

4.3.2.7 Konjugationsstelle 29

PEG verbleibt in einer geknäuelten Konformation und interagiert mit R28 und R32 (Abb. 4.61-A). Am Ende der Simulation adressiert es F30 und L44. LPG befindet sich ebenso auf den Aminosäuren F30 und L44, wobei die terminalen Monomere HB mit den Seitenketten von D36 und D37 ausbilden können (Abb. 4.61-B). PMeOx verbleibt ebenso auf F30, L44, sowie L51 (Abb. 4.61-C). PEtOx zeigt ausgestrecktere Konformationen in der Nähe von N23, Q25, Q55 und A56 (Abb. 4.61-D). Ähnlich zu allen anderen Polymeren zeigt es jedoch auch Konformationen nahe F30 und L44.

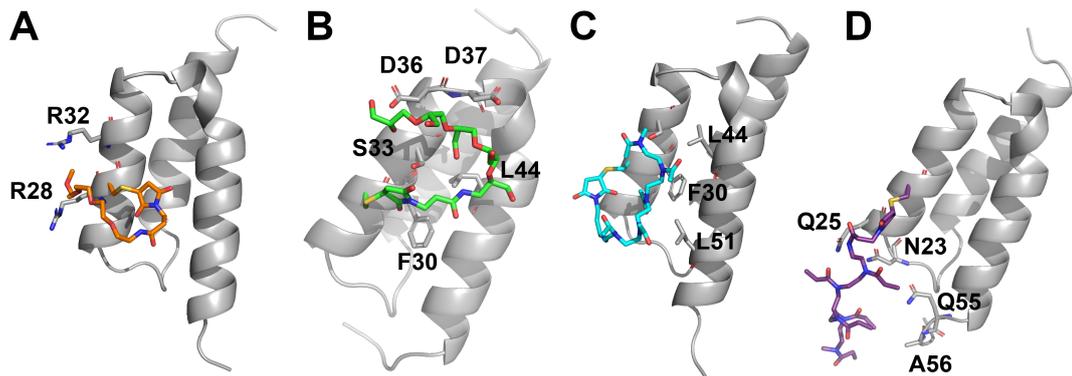


Abbildung 4.61: Beispielfotografien von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 29 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

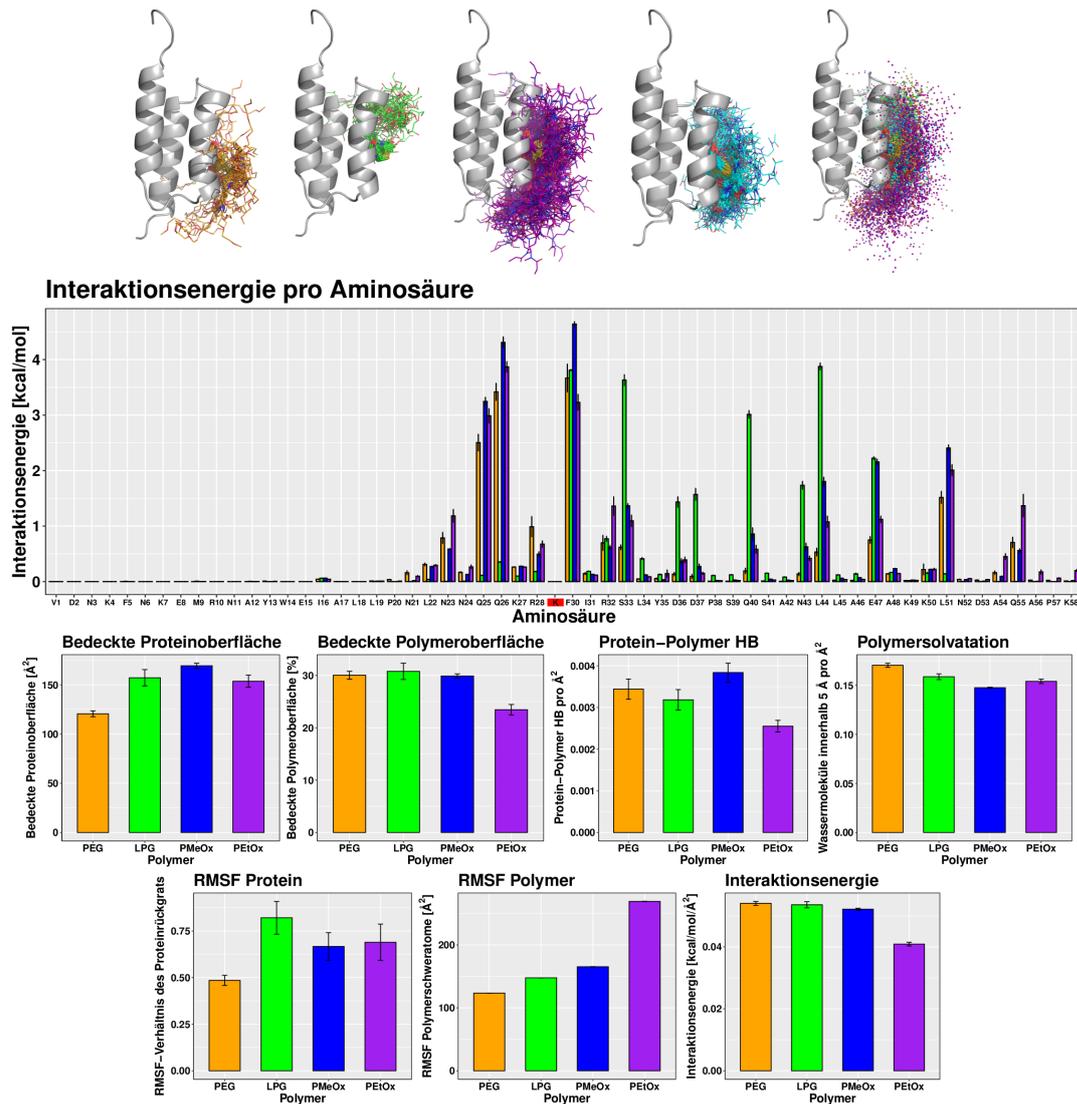


Abbildung 4.62: Protein-Polymer-Dynamiken an Konjugationsstelle 29 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEIOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.2.8 Konjugationsstelle 33

PEG interagiert entweder mit N23, Q26, F30, K50, L51 und Q55 (Abb. 4.63-A) oder windet sich um A29. LPG wechselwirkt mit den Aminosäuren F30, L51 und A54, sowie mit E47 mit seinen ersten Hydroxylgruppen (Abb. 4.63-B). PMeOx und PEtOx verhalten sich ähnlich zu PEG und strecken sich nach unten zur Helix (N23, Q26, F30, K50, L51 und Q55) oder verbleiben in einer geknäuelten Konformation (Abb. 4.63-C und Abb. 4.63-D).

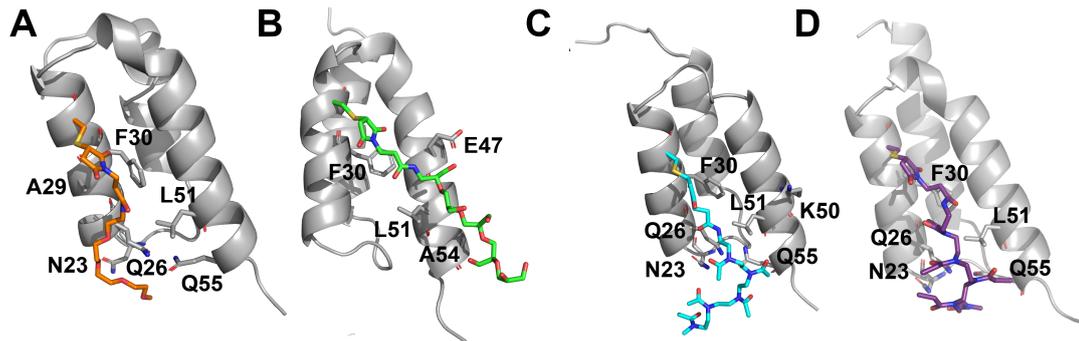


Abbildung 4.63: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 33 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

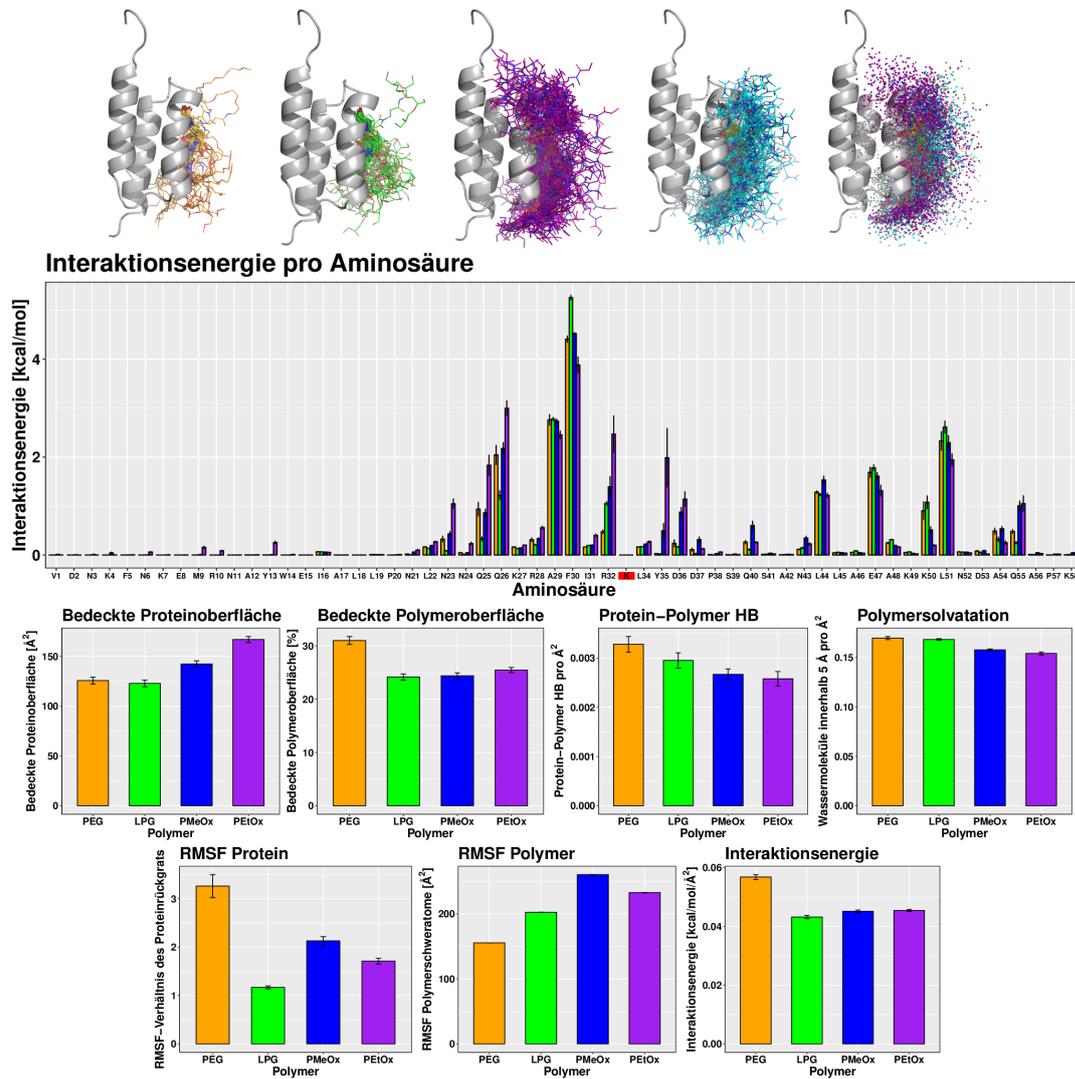


Abbildung 4.64: Protein-Polymer-Dynamiken an Konjugationsstelle 33 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.2.9 Konjugationsstelle 39

Während PEG zunächst zwischen Y15 und Y35 positioniert ist, streckt es sich relativ früh ins Solvens aus. Es verbleibt dann in der Nähe von F5 und windet sich zum Teil um K4 (Abb. 4.65-A). LPG verbleibt ebenso solvensexponiert und abgewandt von den Helices. Oft interagiert es mit der Seitenkette von D2 über eine HB (Abb. 4.65-B). PMeOx hält sich ebenso in der Nähe von F5 des C-Terminus auf (Abb. 4.65-C), richtet sich jedoch auch teilweise nach unten zu Y15 und Y35 aus. PEtOx ist sehr flexibel, Interaktionen mit F5 sind jedoch auch zu beobachten (Abb. 4.65-D). Dabei hält sich das Polymer auf einer Seite der Helices auf, die von den anderen Polymeren nicht adressiert wird (nahe von A42 und L45).

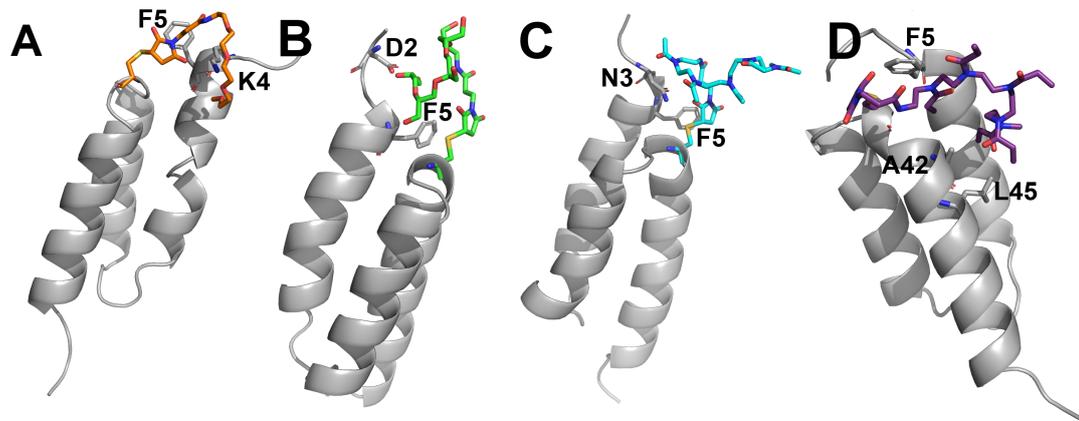


Abbildung 4.65: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 39 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

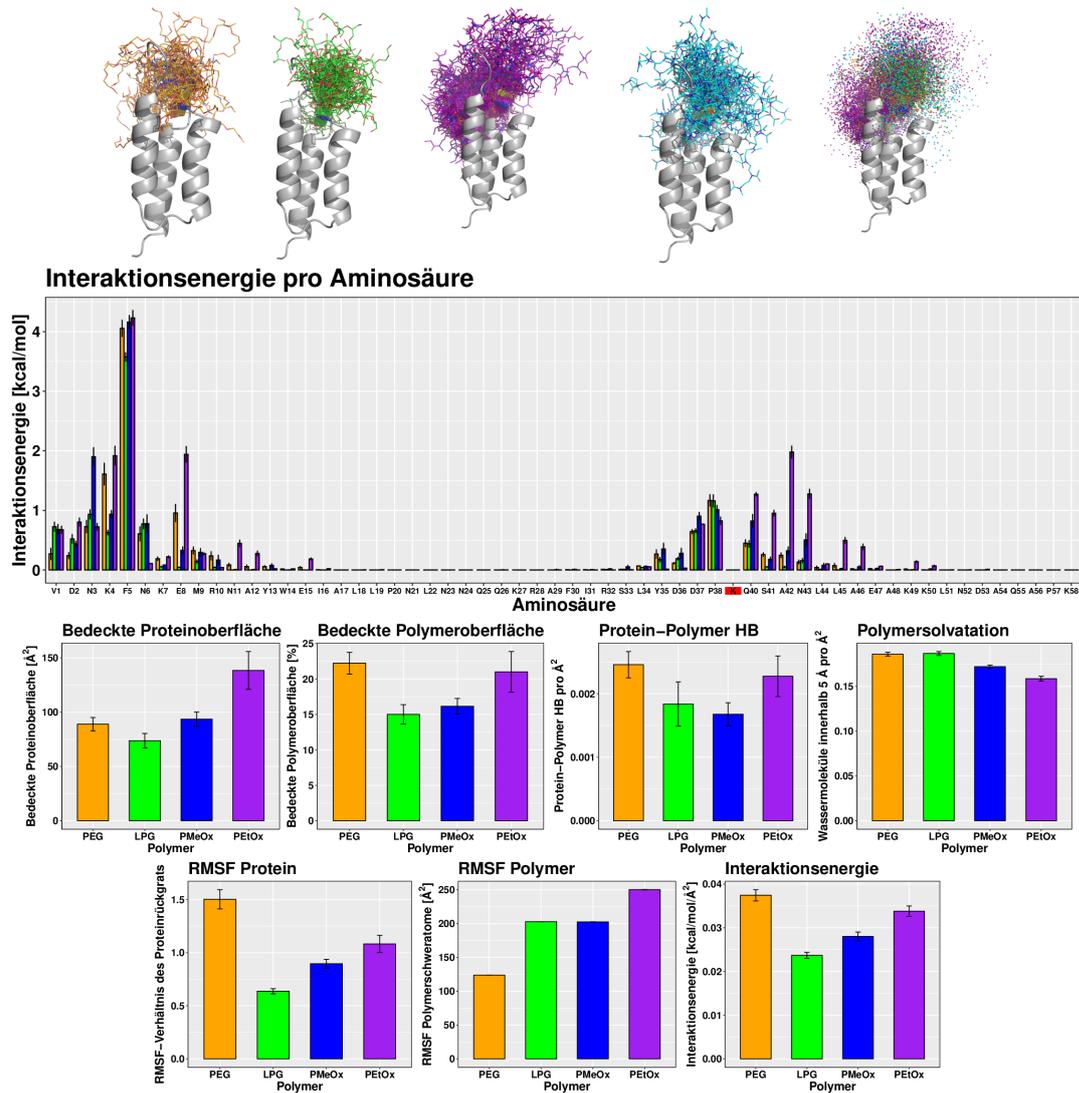


Abbildung 4.66: Protein-Polymer-Dynamiken an Konjugationsstelle 39 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.2.10 Konjugationsstelle 43

Nachdem es zunächst nach unten zu E47, K50 und L51 gerichtet ist (Abb. 4.67-A), etabliert PEG eine geknäuelte Konformation in der zweiten Hälfte der Trajektorie. LPG wechselwirkt über seinen Linker und die ersten Hydroxylgruppen mit E47, über die terminalen Monomere mit K49. Danach etabliert es zum Teil HB mit D36 (Abb. 4.67-B) oder richtet sich nach unten zum hydrophoben Areal von F30 und L51. PMeOx fluktuiert zunächst stark, adressiert später jedoch dieselben Aminosäuren (Abb. 4.67-C). Dieser Bereich wird auch von PEtOx geschützt (Abb. 4.67-D). Ähnlich zur Konjugationsstelle 39 wechselwirkt das Polymer später auch mit der anderen Seite der Helix (F5, A42 und L45).

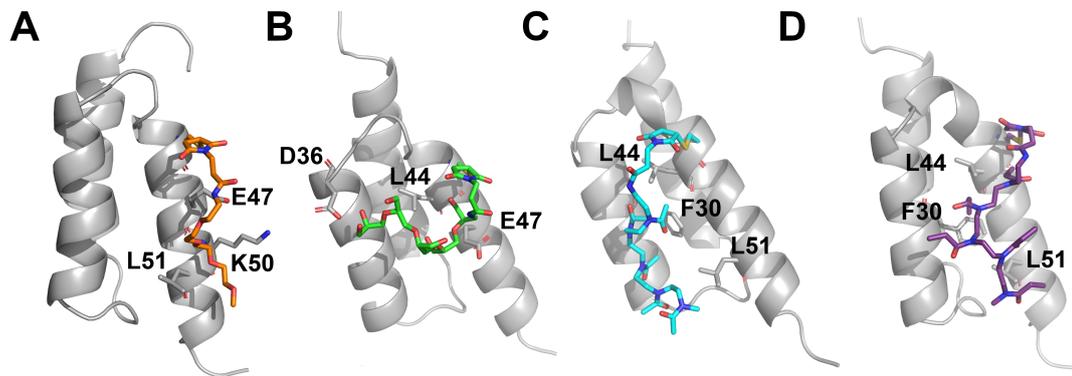


Abbildung 4.67: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 43 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

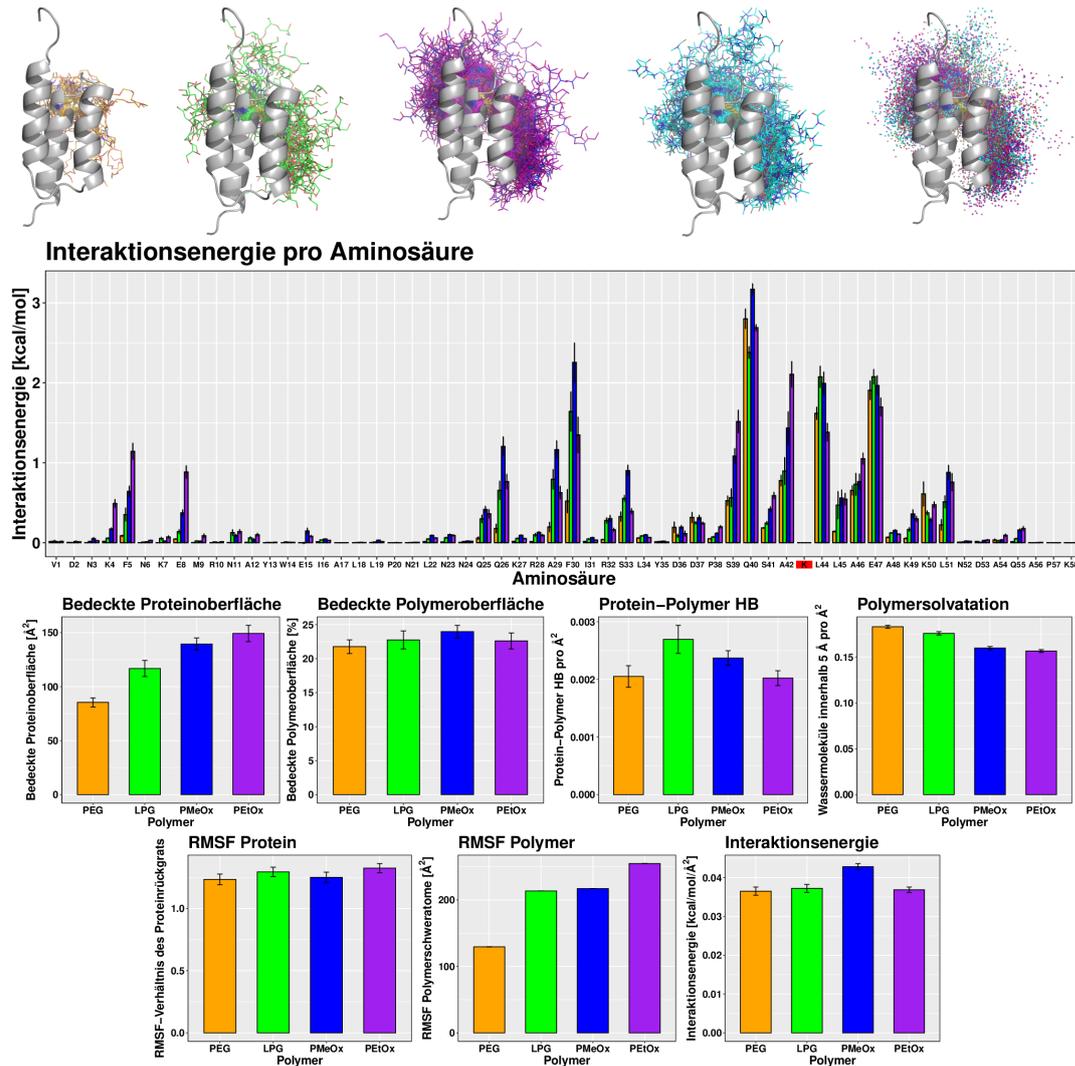


Abbildung 4.68: Protein-Polymer-Dynamiken an Konjugationsstelle 43 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.2.11 Konjugationsstelle 46

Nach einiger Zeit in einer geknäuelten Form und Wechselwirkungen mit N43 richtet sich die PEG-Kette in Richtung F30 aus, während die Amid-Struktur des Linkers mit E47 und die Carbonylgruppe des Linkers mit K50 interagieren kann (Abb. 4.69-A) verbleibt PEG, abgewandt vom Protein, in einer geknäuelten Form. LPG scheint ebenso geringe Fluktuationen an dieser Stelle aufzuweisen. Es wechselwirkt durchgängig mit K49 und D53 (Abb. 4.69-B). Im Gegensatz hierzu wechselt P_{Me}Ox häufig die Position. Nur nahe zu F30 und L44 lassen sich teilweise stabile Konformationen beobachten (Abb. 4.69-C). Ähnlich zu den Dynamiken an den Konjugationsstellen 39 und 43 verbleibt P_{Et}Ox in der Nähe von F5, A42 und L45 zwischen den Helices 1 und 3 (Abb. 4.69-D). Gegen Ende streckt es sich auch zu N11 von Helix 1 aus.

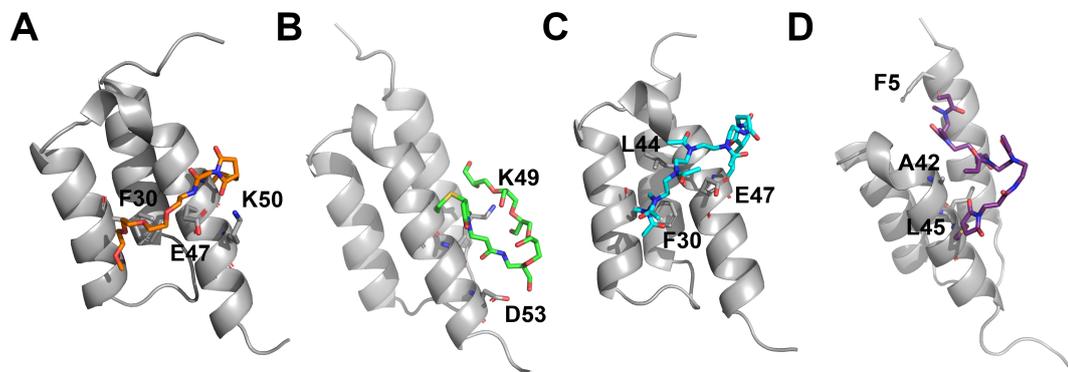


Abbildung 4.69: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) P_{Me}Ox und (D) P_{Et}Ox an der Konjugationsstelle 46 des Her2-Affibodys. PEG ist orange, LPG grün, P_{Me}Ox blau und P_{Et}Ox violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

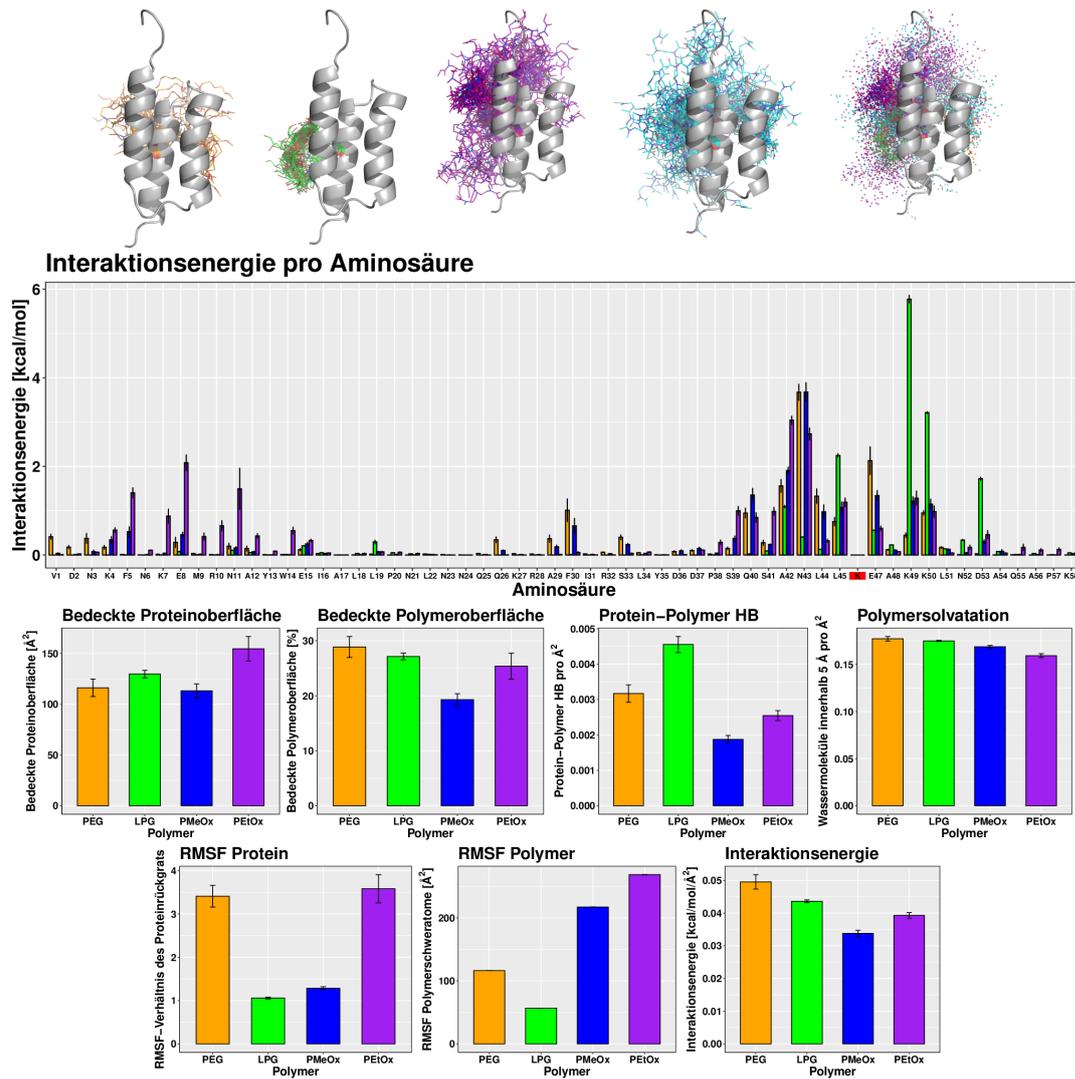


Abbildung 4.70: Protein-Polymer-Dynamiken an Konjugationsstelle 46 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.2.12 Konjugationsstelle 47

PEG adressiert den hydrophoben Bereich von F30 und L44 (Abb. 4.71-A). LPG verbleibt oberhalb von K49 und bildet, wie bei Konjugationsstelle 46, HB zu D53 aus (Abb. 4.71-B). PMeOx interagiert zunächst über seine terminale Carbonylgruppe mit N23 (Abb. 4.71-C) und wechselwirkt dann gegen Ende mit Q26, F30 und L54. PEtOx interagiert mit ähnlichen Aminosäuren. Nach der Hälfte der Simulationszeit lassen sich Konformationen beobachten, bei denen das Polymer zwischen R32 und D36 positioniert ist (Abb. 4.71-D).

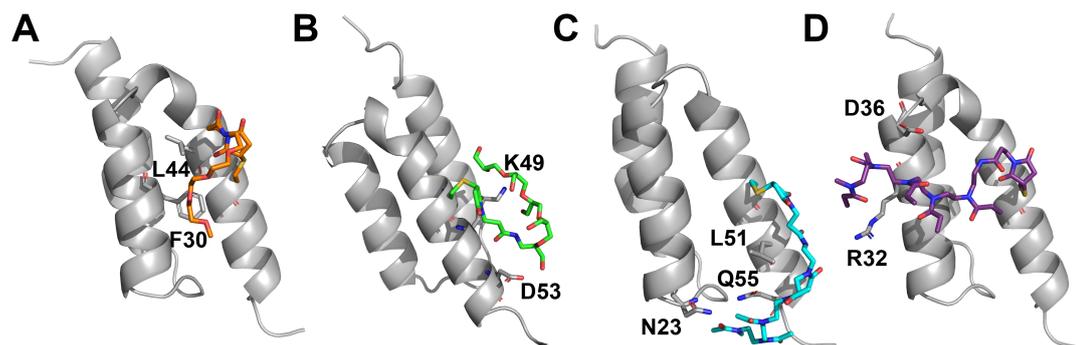


Abbildung 4.71: Beispielfotografien von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 47 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

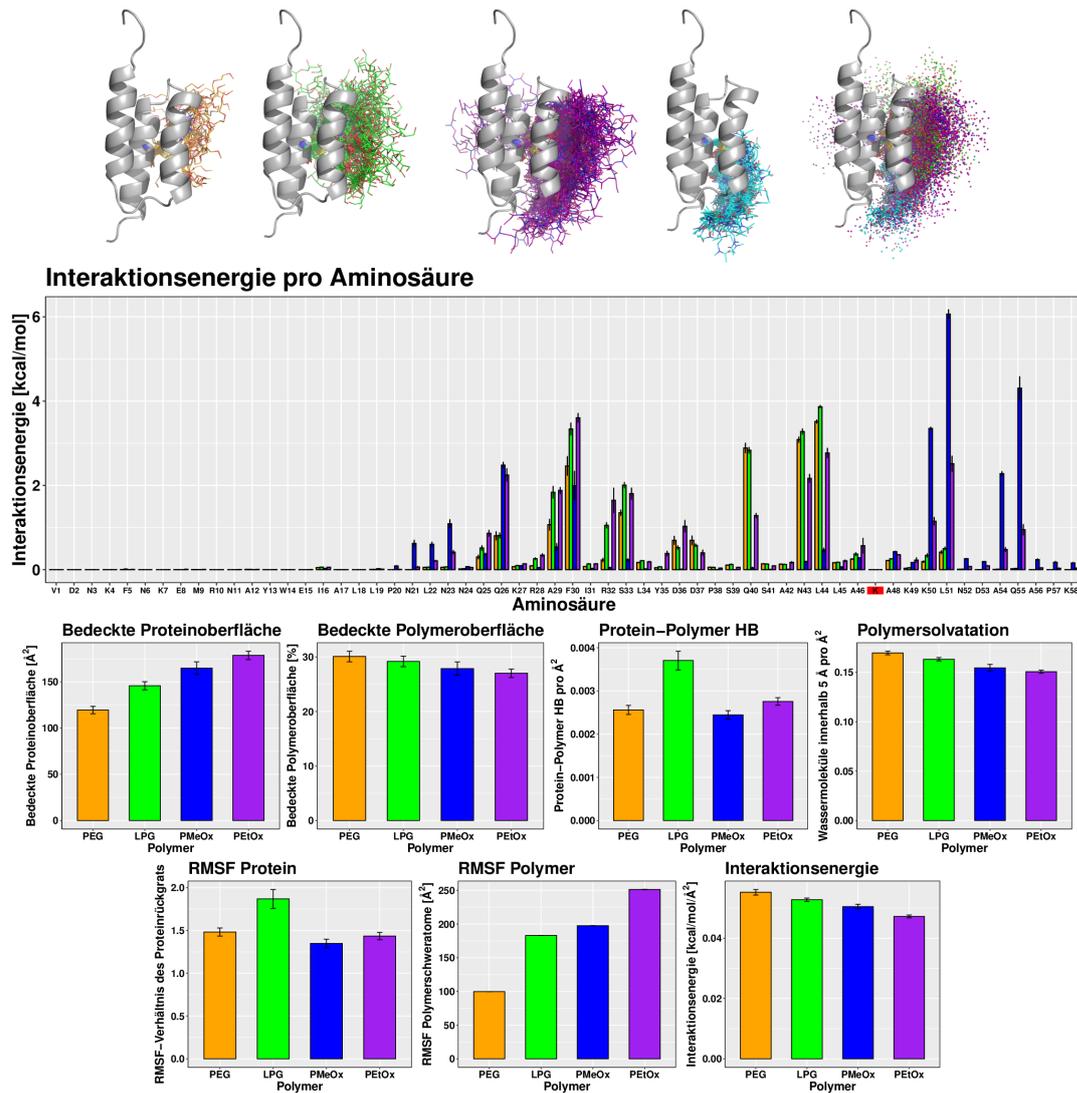


Abbildung 4.72: Protein-Polymer-Dynamiken an Konjugationsstelle 47 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.2.13 Konjugationsstelle 49

Neben einigen geknäuelten Konformationen hält sich PEG entweder neben N43 oder oberhalb von A42 und L45 auf (Abb. 4.73-A). LPG bildet mit seinen Hydroxylgruppen zumeist HB mit E47, während sich die Ethersauerstoffatome um K50 winden. Später adressiert das Polymer die Carboxylgruppen von E8 und E15 (Abb. 4.73-B). PMeOx hält sich zunächst zwischen den Aminosäuren R14, E15, S41 und L45 auf, bevor es eine stabile Konformation über eine Carbonylgruppe des Linkers mit N3 etabliert. Ebenso bildet eine Carbonylgruppe des Polymer Interaktionen mit dem Rückgrat von N43 aus (Abb. 4.73-C). Im Falle von PEtOx ist die Proteinkonformation deutlich modifiziert, weil das Polymer sich um den N-Terminus windet. HB zu N3 und F5 können beobachtet werden. Die Seitenkette von F5 wird durch das Polymer ebenso geschützt (Abb. 4.73-D).

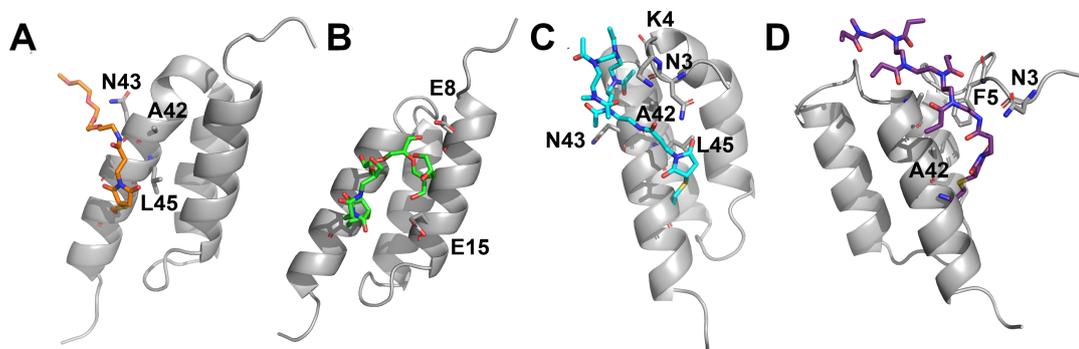


Abbildung 4.73: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 49 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

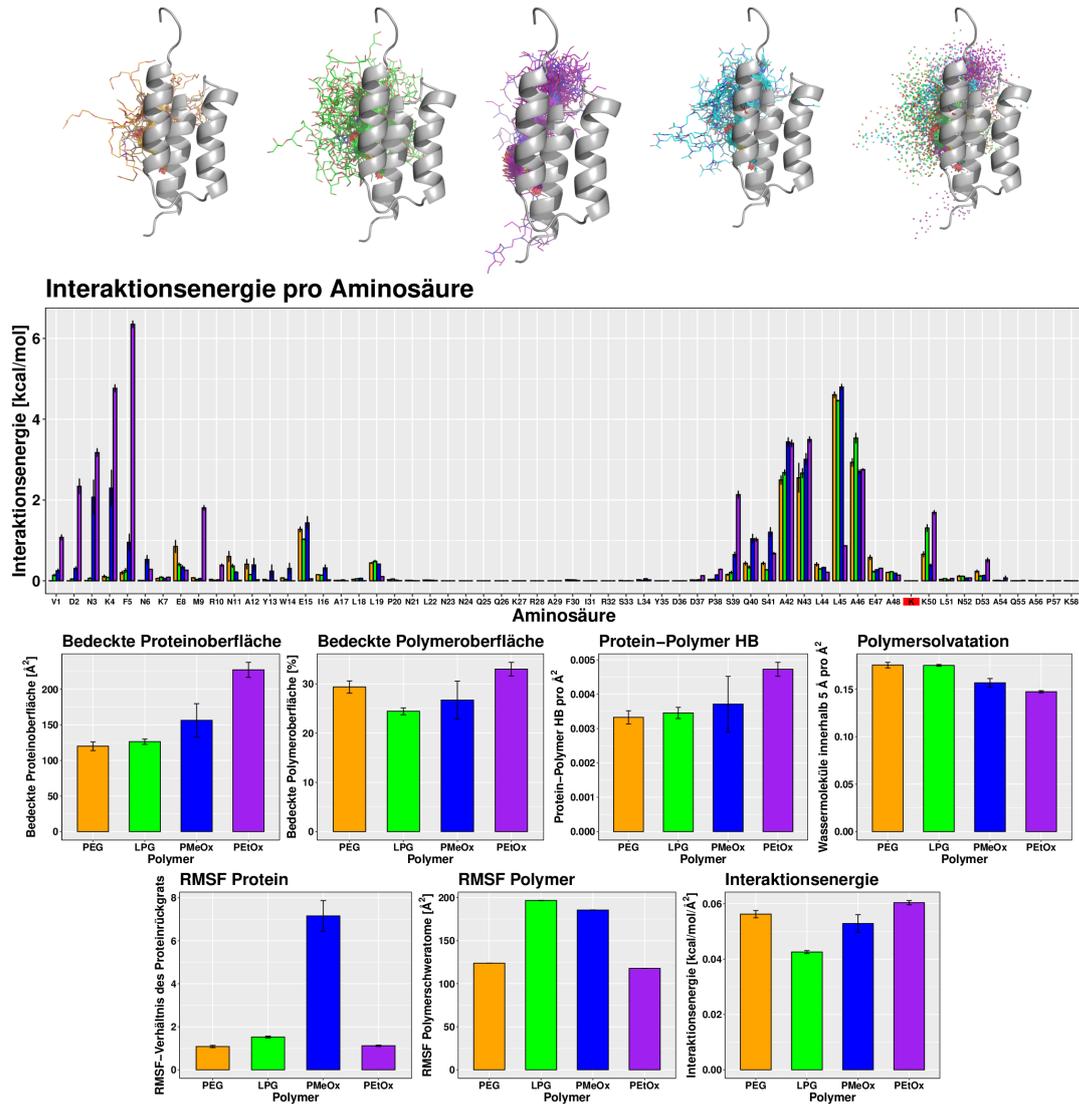


Abbildung 4.74: Protein-Polymer-Dynamiken an Konjugationsstelle 49 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEIOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.2.14 Konjugationsstelle 50

PEG fluktuiert stark, wobei stabile Konformationen in der Nähe von A29, F30, S33 und L51 (Abb. 4.75-A), sowie von Q40 und N43 beobachtet werden können. LPG wechselwirkt mit einer Vielzahl an Seitenketten in der Nähe des hydrophoben Bereichs: S33, E36, N40 und D47 (Abb. 4.75-B). Am Ende der Simulation streckt sich das Polymer entlang der Helix aus, ähnlich zu PEG. Im Falle von PMeOx werden zunächst A42 und N43 adressiert (Abb. 4.75-C), bevor sich das Polymer, wie auch die anderen Varianten, nach unten zu F30 ausrichtet. Letztere Aminosäure wird auch von PEG zu Beginn der Simulation adressiert. Danach orientiert sich das Polymer parallel zur Helix aus. Dabei kann K58 mit dem Linker wechselwirken (Abb. 4.75-D).

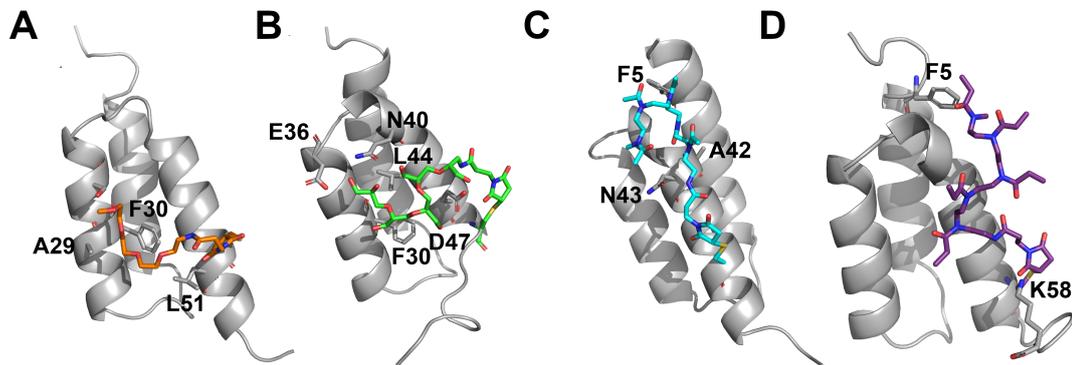


Abbildung 4.75: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEG an der Konjugationsstelle 50 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEG violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

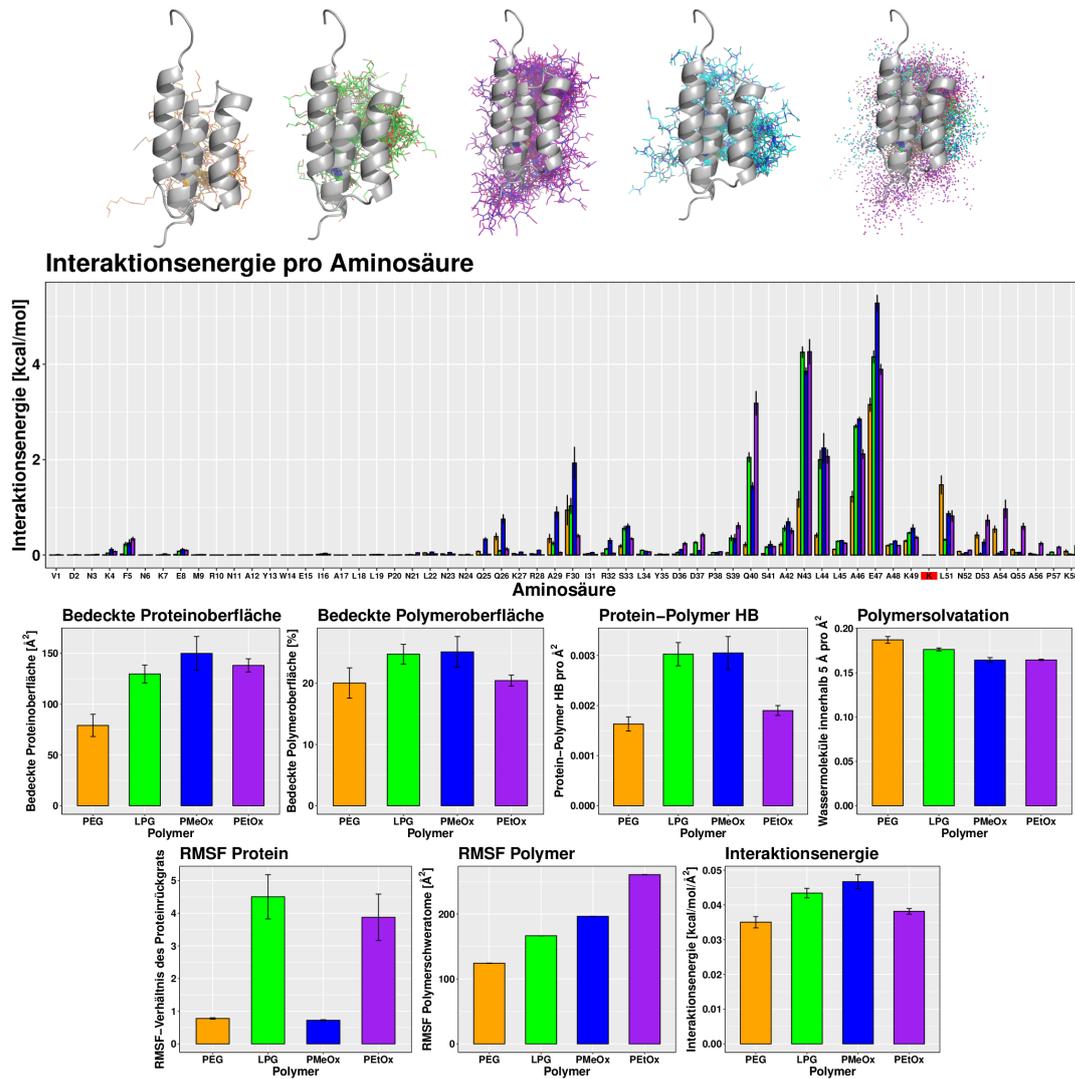


Abbildung 4.76: Protein-Polymer-Dynamiken an Konjugationsstelle 50 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.2.15 Konjugationsstelle 53

Nach wenigen Nanosekunden windet sich PEG um K49 und verändert seine Konformation nicht mehr (Abb. 4.77-A). In vergleichbarer Weise interagiert LPG mit K49 über seine Ethergruppen, während die Hydroxylgruppen E15 adressieren (Abb. 4.77-B). Gegen Ende streckt sich das Polymer nach oben und interagiert mit einer Carbonylgruppe des Linkers mit K50 und den Hydroxylgruppen mit E8. Beide POx-Varianten bleiben solvensexponiert und richten sich teilweise parallel zur Helix aus, wobei F5 adressiert werden kann (Abb. 4.77-C und 4.77-D).

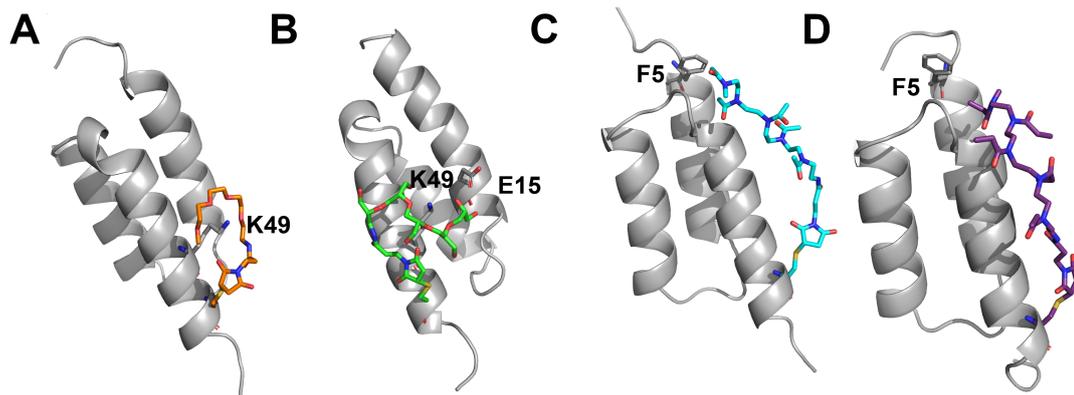


Abbildung 4.77: Beispielhafte Momentaufnahmen von (A) PEG, (B) LPG, (C) PMeOx und (D) PEtOx an der Konjugationsstelle 53 des Her2-Affibodys. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Ausgewählte zusätzliche Aminosäuren in Polymernähe sind ebenso dargestellt.

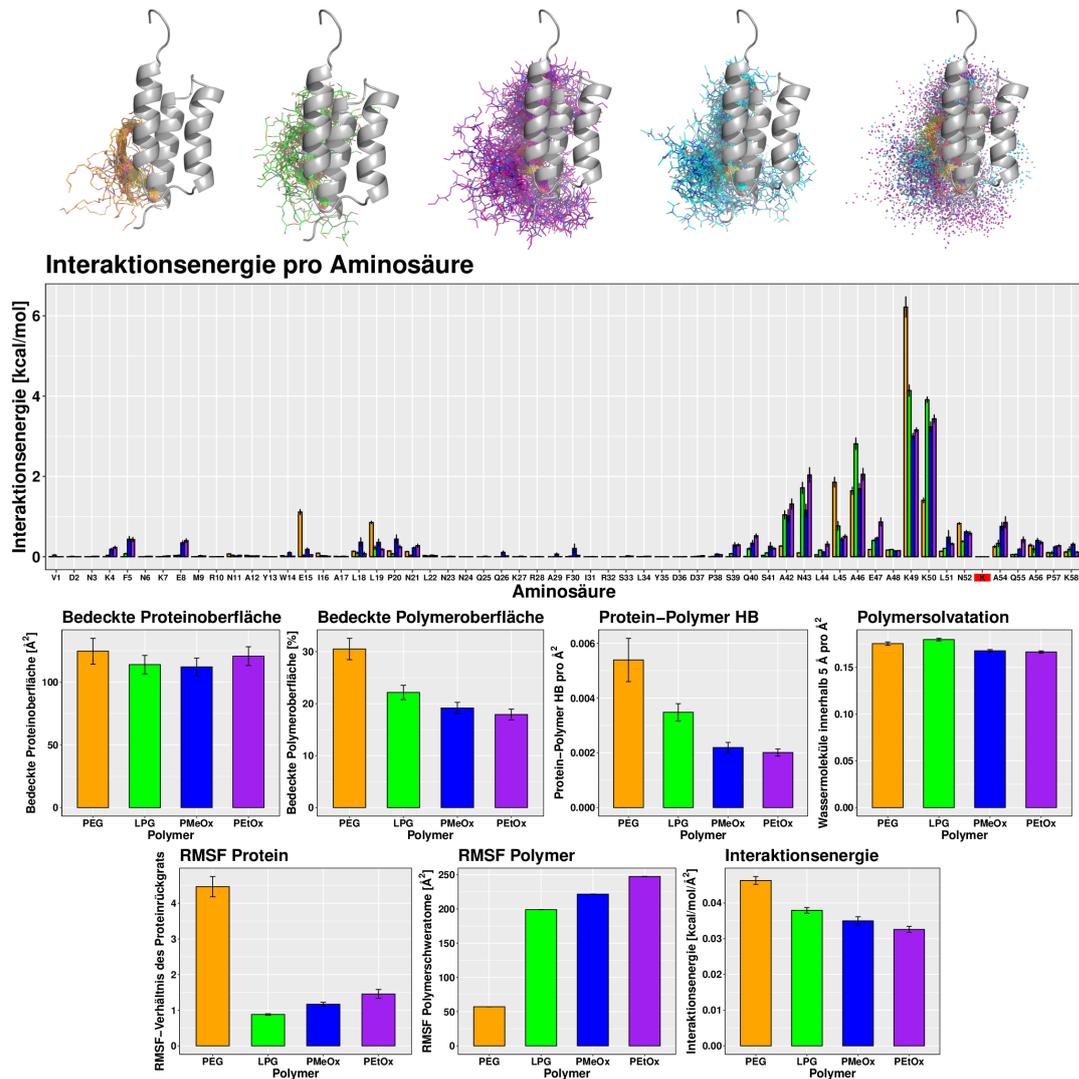


Abbildung 4.78: Protein-Polymer-Dynamiken an Konjugationsstelle 53 des Her2 Affibody. Die oberste Reihe zeigt Momentaufnahmen aus den Trajektorien jedes Polymers im Intervall von 10 ns, die erste Momentaufnahme des Proteins ist als Cartoon dargestellt. PEG ist orange, LPG grün, PMeOx blau und PEtOx violett gefärbt. Im letzten Bild der obersten Reihe sind diese Momentaufnahmen als Sphären übereinandergelagert. Darunter sind alle Eigenschaften dargestellt, die in Kap. 3.1.8 geschildert wurden. Die Konjugationsstelle ist im Falle der aminosäurespezifischen Interaktionsenergien (mit -1 multipliziert) als *K* auf rotem Grund gekennzeichnet.

4.3.3 Zusammenfassender Vergleich

4.3.3.1 Pin 1 WW-Domäne

4.3.3.1.1 Vergleich der berechneten Eigenschaften

Für eine zusammenfassende Übersicht der ermittelten Eigenschaften sind die Mittelwerte aller Konjugationsstellen in Abb. 4.79 für die Pin 1 WW-Domäne dargestellt. Zum Vergleich einzelner Konjugate sind die in Kap. 4.3.1 gezeigten Durchschnittswerte weiterhin in Abb. 4.80 aufgeführt. Eine tabellarische Übersicht der ermittelten Interaktionsenergien ist für jede Simulation in der Abb. 4.81 zu finden.

Beim Blick auf die Mittelwerte in Abb. 4.79 lässt sich zunächst anmerken, dass einige Eigenschaften wie bspw. die Interaktionsenergie pro Polymeroberfläche erwartungsgemäß größere Werte für das kürzere PEG-Polymer zeigen. Durch die geringere Länge ist ein größerer Anteil innerhalb der Proteinnähe zu verorten, während die Wahrscheinlichkeit proteinferner Strukturen bei den anderen Polymeren mit jeweils fünf Monomeren deutlich größer ist. Insofern stellt sich grundsätzlich die Frage, in welcher Form die Polymerketten verschiedener Länge vergleichend gegenübergestellt werden sollten. Im Falle größerer Konjugate (siehe Kap. 5) werden in der Regel Polymere ähnlicher Molekulargewichte miteinander verglichen, die aufgrund der unterschiedlichen Konstitutionen ebenso verschiedene Längen besitzen.

Ungeachtet dessen zeigt PEG auf Basis dieser Berechnungen die insgesamt größte Solvatation (Anzahl der in der Nähe befindlichen Wassermoleküle). Der Anteil an bedeckter Proteinoberfläche ist am größten im Falle der sterisch anspruchsvolleren POx-Polymere. Diese führen ebenso insgesamt zu einer leichten Erniedrigung der Rückgratflexibilität des Proteins. Dagegen zeigen die etherbasierten Polymervarianten PEG und LPG deutlich mehr HB zum Protein. PEG zeigt weiterhin deutlich geringere RMSF-Werte als die übrigen Polymerarten. LPG stellt dagegen das Polymer mit den größten Fluktuationen dar.

Beim Vergleich der einzelnen Konjugationsstellen (Abb. 4.80) zeigt sich, dass alle Polymerarten bei den in den Faltblättern zu verortenden Stellen (14, 23, 26, 32) eine größere Bedeckung aufweisen, als es bei den Konjugationen an den flexiblen Schleifen der Fall ist (Abb. 4.80-A, 4.80-B). Für PEG und LPG ist die Anzahl an HB zum Protein besonders für die im zweiten Faltblatt gelegenen Konjugationsstellen 23 und 26 relativ groß (Abb. 4.80-C). Im Gegensatz zu den Varianten, die an den Faltblättern konjugiert sind, weisen jene mit Konjugationen an der ersten und zweiten Schleife eine stärkere Solvatation von PEG und LPG auf (Abb. 4.80-D). Erhöhte Fluktuationen des Proteinrückgrats im Gegensatz zu den unkonjugierten Varianten sind besonders für Konjugate an den Stellen 16, 17 und 18 zu finden (Abb. 4.80-E). PEG mit den geringsten Fluktuationen zeigt keine großen Unterschiede für die RMSF-Werte der verschiedenen Konjugationsstellen (Abb. 4.80-F). Besonders große Unterschiede in

Abhängigkeit der Konjugationsstelle sind dagegen für PEG zu finden. Dieses Polymer zeigt für die Stellen 19, 21, 23 und 29 relativ geringe, für die Stellen 16, 17, 26, 27 und 28 relativ hohe Fluktuationen. Übereinstimmend mit den Beobachtungen zur Bedeckung der Polymere bzw. des Proteins sind die Interaktionsenergien zwischen Protein und Polymer für die Konjugationsstellen an den Faltblättern relativ groß (Abb. 4.80-G).

Im Hinblick auf die favorisierten Wechselwirkungsbereiche der Polymere (Abb. 4.81) interagiert PEG bei den Konjugationsstellen 14, 16, 17, 18, 19 und 21 mit Aminosäuren der ersten Schleife und des dritten Faltblatts (vgl. Beschreibungen in Kap. 4.3.1). Konjugationen im zweiten Faltblatt und der zweiten Schleife führen dagegen zu Wechselwirkungen im Bereich von W11 und E12. Das Profil von LPG ähnelt jenem von PEG. Die POx-Varianten zeigen starke Wechselwirkungen mit Aminosäuren des zweiten und dritten Faltblatts (Y23, F25, S32, F34), nur Konjugate der zweiten Schleife zeigen, wie bei PEG und LPG, erhöhte Wechselwirkungen mit Bereichen des N-Terminus und des ersten Faltblatts (W11 und E12). Die Protein-Polymer-Wechselwirkungen und die Dynamik der einzelnen Polymerarten sollen im Folgenden noch näher charakterisiert werden.

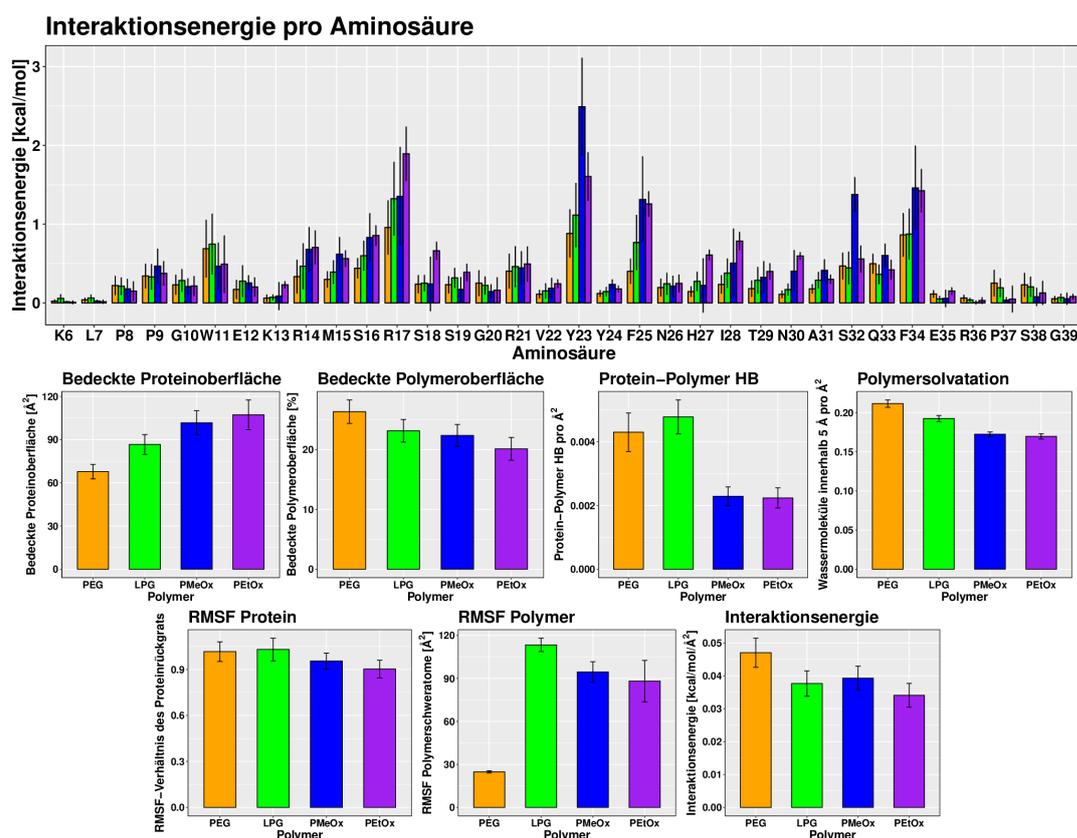


Abbildung 4.79: Mittelwerte untersuchter Eigenschaften für die Pin 1-Konjugate (aus Werten der Abbildungen in Kap. 4.3.1). Daten für PEG sind orange, für LPG grün, für PMeOx blau und für PEG violett gefärbt. Fehlerbalken zeigen den Standardfehler des Mittelwerts.

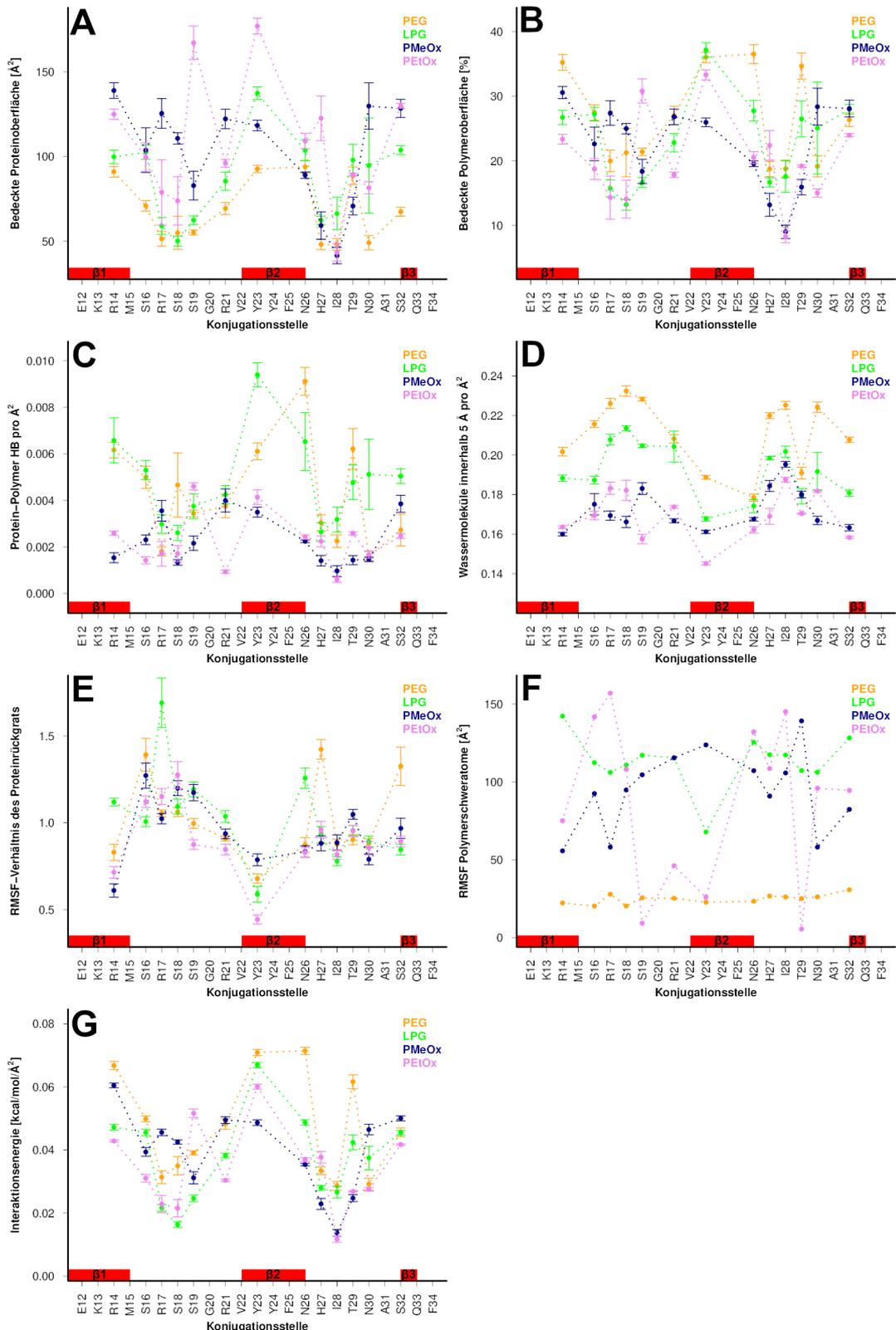


Abbildung 4.80: Übersicht über die Mittelwerte der berechneten Eigenschaften der Pin 1 WW-Konjugate, aufgetragen für alle vier Polymere für jede Konjugationsstelle. Fehlerbalken zeigen den Standardfehler des Mittelwerts.

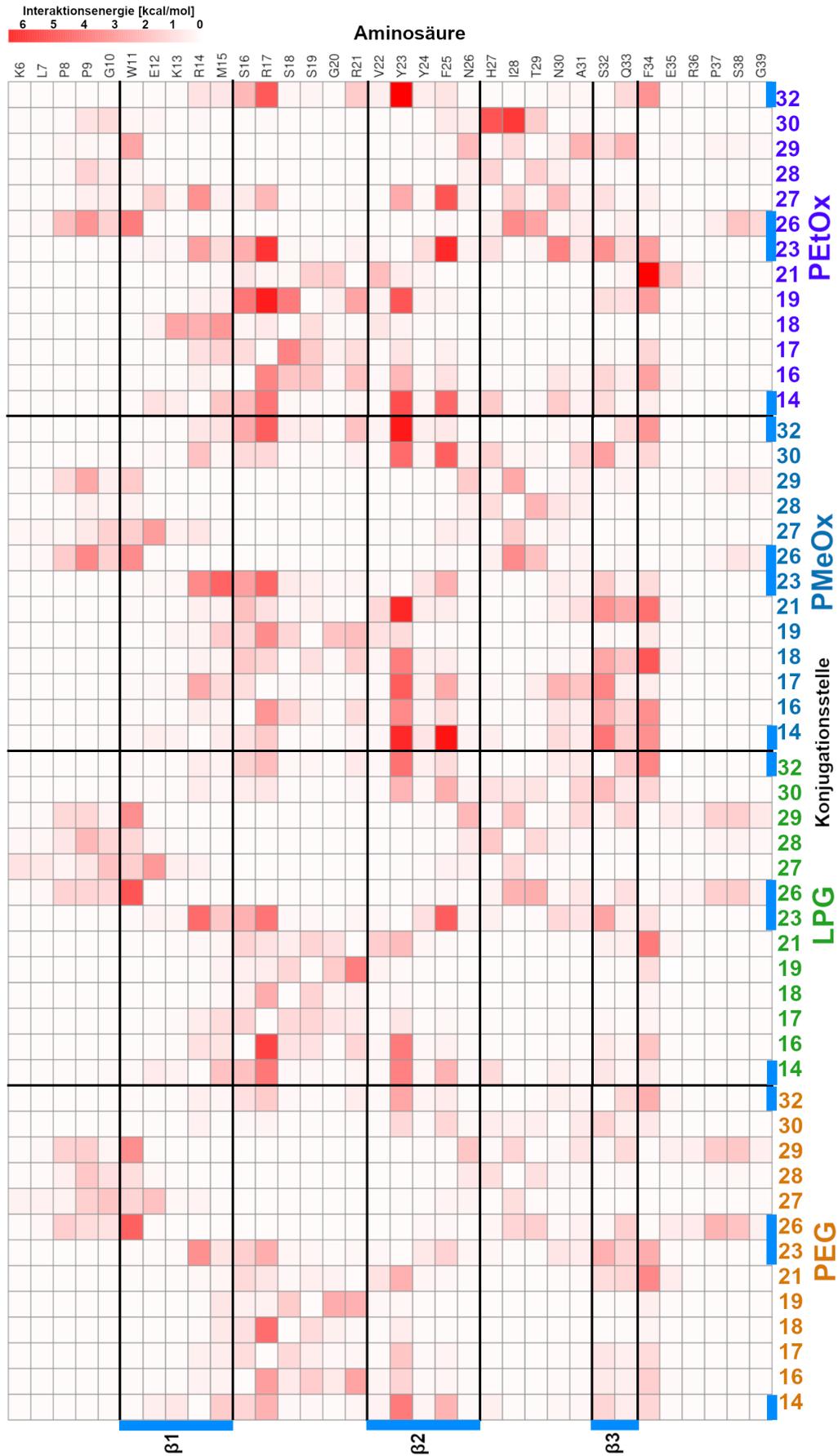


Abbildung 4.81: Durchschnittliche Interaktionsenergien (mit -1 multipliziert, gefärbt von weiß nach rot) zwischen den Polymeren an jeder Konjugationsstelle (Reihen) und jeder einzelnen Aminosäure der Pin 1 WW-Domäne (Spalten). Sekundärstrukturelemente sind blau gekennzeichnet. Die Interaktionsenergie für die Konjugationsstelle wurde jeweils auf Null gesetzt.

4.3.3.1.2 PEG und LPG

Die Polymere PEG und LPG bestehen aus denselben Etherrückgraten und zeigen insgesamt sehr ähnliche Wechselwirkungsprofile. An vielen Konjugationsstellen (14, 17 und 21) zeigen sich, besonders bei PEG, bevorzugte Konformationen in der Nähe der hydrophoben Aminosäuren Y23 und F25 unter gleichzeitiger Ausbildung von HB mit der Seitenkette von S32 (Abb. 4.82-A). Durch das Etherrückgrat werden somit hydrophobe Areale adressiert, in denen gleichzeitig einzelne HB ausgebildet werden können (mit Y23 oder S32). Dies entspricht den Beobachtungen aus [107]. Sind die Polymere nicht in diesem Bereich vorzufinden, so lassen sich Interaktionen mit Argininen (R17, R21) oder Serinen der ersten Schleife (S18, S19) beobachten (Abb. 4.82-B).

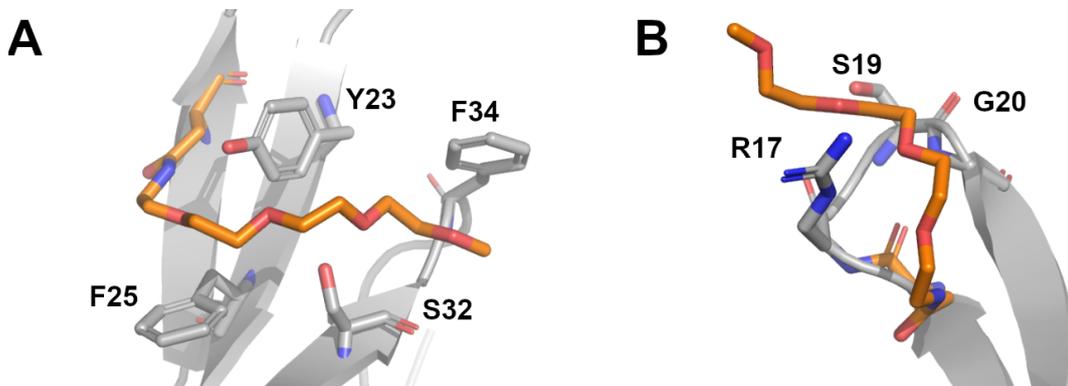


Abbildung 4.82: (A) Ausrichtung der PEG-Kette an der Konjugationsstelle 14 zwischen hydrophoben Aminosäuren hin zu S32. (B) Interaktion von PEG mit benachbartem Arginin und Serin an Konjugationsstelle 16.

Im Falle des hydrophileren LPG, welches über eine größere Anzahl an HB ausbildenden Strukturelementen verfügt, sind die Interaktionen mit R17 teils etwas ausgeprägter. Auch stärkere intrapolymerere Wechselwirkungen und Interaktionen mit dem Solvens sind in vielen Simulationen über die zusätzlichen Hydroxylgruppen beobachtbar (Abb. 4.83-A). Aufgrund des gleichen Polymerrückgrats ist jedoch nur selten ein deutlich unterschiedliches Verhalten von PEG zu LPG zu erkennen. So streckt sich letzteres im Falle der Konjugationsstelle 19 über F34 hinaus zur Carboxylgruppe von E35, ersteres hingegen adressiert diesen Bereich nicht. Über seine endständigen Hydroxylgruppen kann LPG, wenn auch nur sehr kurz, mit dieser Aminosäure wechselwirken (Abb. 4.83-B). Dies stellt jedoch, neben E12, die einzige Carboxylgruppe innerhalb der WW-Domäne dar. Letztere wird an der benachbarten Konjugationsstelle 27 ebenfalls adressiert. Im Falle der Her2-Simulationen sind solche Interaktionen häufiger beobachtbar (*vide infra*).

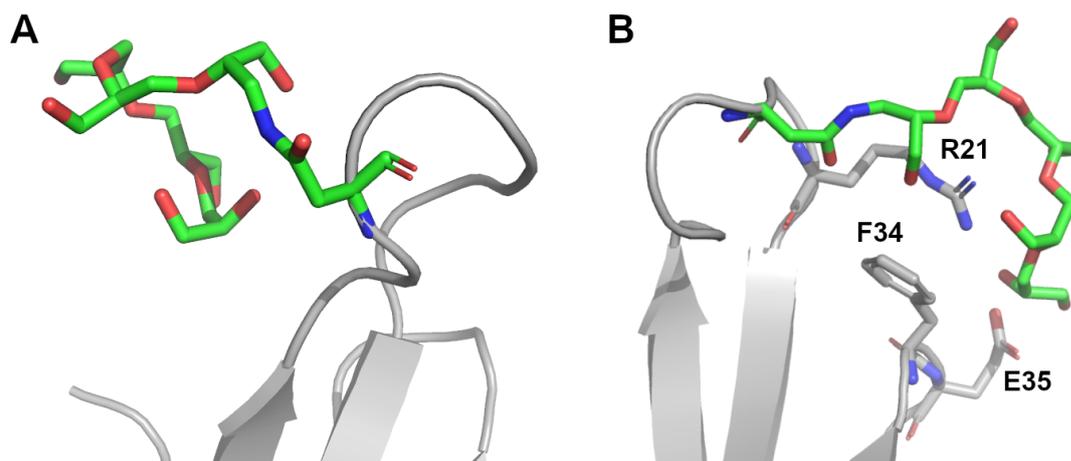


Abbildung 4.83: (A) Interaktionen der LPG-Kette mit den eigenen Monomeren (Konjugationsstelle 17). (B) Interaktionen von LPG mit R21 und E35 an Konjugationsstelle 19.

4.3.3.1.3 Effekt des Linkers

An einigen Konjugationsstellen wird auch deutlich, inwiefern die Dynamik des jeweiligen Polymers von den Interaktionen des Linkers bestimmt wird. So ist der Asparagin-Linker entscheidend an Wechselwirkungen mit E12 (Abb. 4.84-A) und W11 (Abb. 4.84-B) an den Positionen 26 bzw. 27 beteiligt und bestimmt somit die initiale Ausrichtung der Polymerkette. Die Auswahl des Linkers kann also durchaus einen großen Effekt auf die Polymerdynamik haben. Im Falle der Konjugationsstelle 26 ist die NH- π -Wechselwirkung des Linkers mit W11 für alle Polymervarianten beobachtbar. Diese Interaktion wurde von Price et al. in deren neuester Studie näher beleuchtet [104] und konnte in der hier vorliegenden Arbeit somit wiedergefunden werden, wobei eine polymerbasierte Desolvatation in diesem Areal den Autoren zufolge zu einer Stärkung der Wechselwirkung führen sollte.

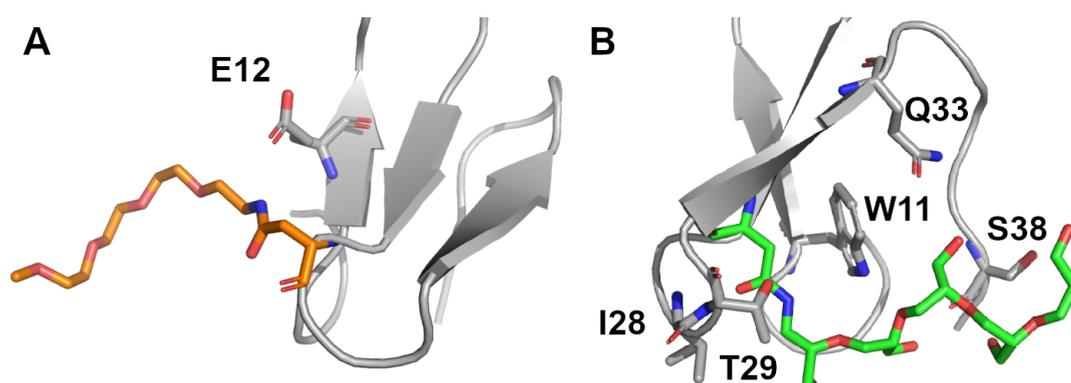


Abbildung 4.84: (A) Interaktionen des Asparagin-Linkers mit der Seitenkette von E12 (PEG-Konjugationsstelle 27). (B) Interaktionen des Asparagin-Linkers mit W11 an LPG-Konjugationsstelle 19.

4.3.3.1.4 PMeOx und PEtOx

Die POx-Varianten zeigen eine teils deutlich von PEG und LPG abweichende Dynamik, da sie strukturell grundlegend verschieden sind. Es zeigt sich insgesamt eine stärkere Tendenz zu Wechselwirkungen mit aromatischen Strukturen, was auch dem Interaktionsprofil in Abb. 4.79 zu entnehmen ist. Dabei sind Konformationen zu erkennen, bei denen die Seitenketten der POx-Polymere auf den aromatischen Ringen, besonders von Y23, F25 und F34, aufliegen (Abb. 4.85). Diese Wechselwirkungen basieren höchstwahrscheinlich auf den hydrophoben Alkylgruppen der Seitenketten. Ähnliche Konformationen für PMeOx wurden kürzlich auch in einer Simulation einer Poly(2-methyl-2-oxazolin)-*b*-poly(2-phenyl-2-oxazolin)-*b*-poly(2-methyl-2-oxazolin)-basierten Wurmmizelle beobachtet, wobei hierbei ebenfalls die Hypothese möglicher $n_{\text{Amid}} \rightarrow \pi^*_{\text{Aromat}}$ - und $\pi_{\text{Amid}} \cdots \pi_{\text{Aromat}}$ -Wechselwirkungen aufgestellt wurde [322]; diese spielen jedoch vermutlich, im Vergleich zu hydrophoben Effekten, bei diesen Oligomeren und einer geringen Anzahl möglicher aromatischer Wechselwirkungspartner nur eine untergeordnete Rolle. Nichtkovalente $n \rightarrow \pi^*$ -Wechselwirkungen wurden in der Vergangenheit bereits bzgl. strukturell ähnlicher Peptide und des prolinreichen Proteins Kollagen beschrieben [323, 324].

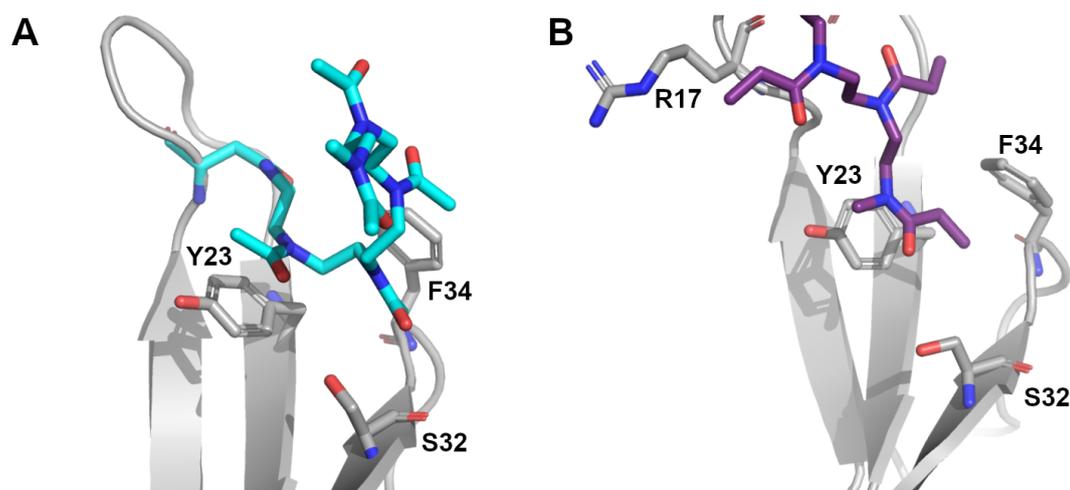


Abbildung 4.85: Beispiele für POx-Konformationen, bei denen die Monomere mit aromatischen Aminosäuren wechselwirken. **(A)** Interaktionen von PMeOx an der Konjugationsstelle 16. **(B)** Interaktionen von PEtOx an Konjugationsstelle 19.

Auch wenn diese Interaktionen an einigen Konjugationsstellen für beide POx-Varianten beobachtet werden (14, 16, 21, 32), so zeigen PMeOx und PEtOx teilweise auch eine deutlich unterschiedliche Dynamik. Exemplarisch sollten hier die Stellen 18 und 19 hervorgehoben werden (Abb. 4.86-A und 4.86-B). Während sich PMeOx an der Stelle 18 zu F34 hin orientiert, wendet sich PEtOx zu den Aminosäuren K13 und M15 auf der Rückseite der ersten Schleife hin. An der Stelle 19 ist teilweise das Gegenteil zu beobachten: PMeOx zeigt Konformationen in der Nähe von M15, dagegen werden hier Y23 und F34 deutlich öfter von PEtOx adressiert. Weiterhin kann PMeOx das obere hydrophobe Zentrum im Falle der Konjugationsstelle 30 erreichen, PEtOx verbleibt im Bereich der zweiten Schleife in der Nähe von H27. Dagegen adressiert PEtOx das

hydrophobe Zentrum an der Konjugationsstelle 27, PMeOx hingegen verbleibt hier, durch Wechselwirkungen des Linkers mit E12, in der Nähe der zweiten Schleife (Abb. 4.86-C und 4.86-D).

Diese beiden Gegenüberstellungen (Konjugationsstelle 18 vs. 19, 27 vs. 30) deuten darauf hin, dass die Distanz der Konjugationsstelle zu favorisierten Wechselwirkungspartnern das Verhalten beider POx-Derivate, zumindest im Falle kürzerer Oligomere, entscheidend beeinflussen kann. Die Rückgratlängen beider Varianten sind gleich, die PEtOx-Seitenketten sind jedoch hydrophober und sterisch anspruchsvoller. Im Falle der Stellen 18 und 30 könnte die Konjugationsstelle für das größere PEtOx-Oligomer aufgrund sterischer Effekte zu nah an den Aminosäuren Y23 und F34 positioniert sein, während diese im Falle der etwas weiter entfernt gelegenen Stellen 19 und 27 durch die größeren PEtOx-Seitenketten besser erreicht werden können als es bei PMeOx der Fall ist.

Neben Wechselwirkungen mit aromatischen Gruppen zeigen sich auch viele Interaktionen über die Carbonylgruppen der POx-Derivate, die prinzipiell als HB-Akzeptoren zur Verfügung stehen (z. B. bei PMeOx für R17 an den Konjugationsstellen 14 und 23, bei PEtOx für R14, R17 und N30 an Konjugationsstelle 27). So lassen sich auch HB zu Hydroxylgruppen von bspw. S32 und Y23 beobachten. Das Wechselwirkungspotenzial der eher polaren Amid-Strukturelemente der Poly(2-alkyl-2-oxazoline) sollte daher nicht vernachlässigt werden. Die Relevanz dieser Interaktionen wurde in der Literatur im Falle POx-basierter, wirkstoffbeladener Mizellen bereits erwähnt [325].

4.3.3.2 Her2-Affibody

4.3.3.2.1 Vergleich der berechneten Eigenschaften

Analog zu den Konjugaten der Pin 1 WW-Domäne sind die Mittelwerte der Eigenschaften in Abb. 4.87 gezeigt, jene für die einzelnen Konjugationsstellen in Abb. 4.88, sowie die Interaktionsenergien der Polymere mit den Aminosäuren in einer tabellarischen Form in der Abb. 4.89.

Wie bei der Pin 1 WW-Domäne zeigen die POx-Varianten im Hinblick auf die Mittelwerte der Konjugationsstellen (Abb. 4.87) die größte Proteinbedeckung und Wechselwirkungen mit aromatischen Aminosäuren (F5, W14, F30). Die kürzere PEG-Kette zeigt abermals die größte prozentuale Bedeckung. PEG und LPG zeigen mehr HB mit dem Protein und eine im Vergleich zu den POx-Varianten höhere Solvatation. Alle Polymervarianten zeigen bei diesem Peptid größere RMSF-Werte im Verhältnis zu den Simulationen unkonjugierter Varianten. Auch hier zeigt PEG deutlich geringere RMSF-Werte im Vergleich zu den Polymeralternativen und die größte Interaktionsenergie in Relation zur verfügbaren Polymeroberfläche.

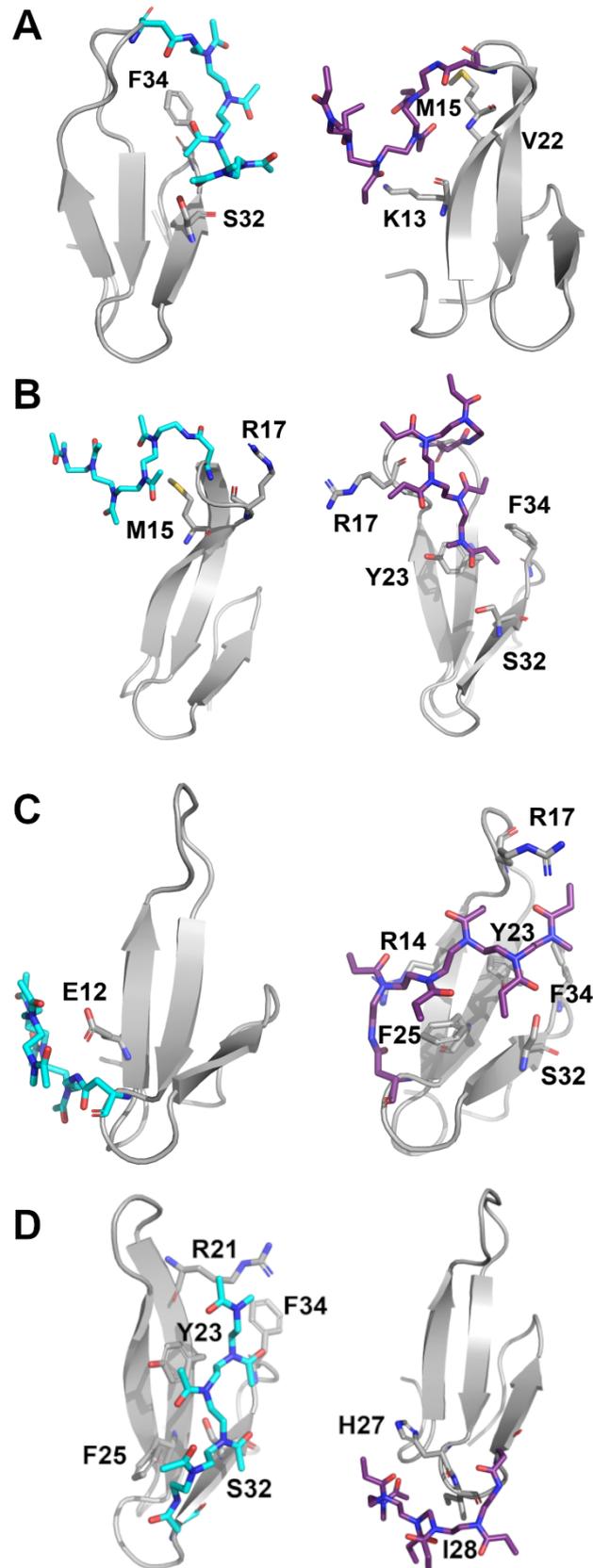


Abbildung 4.86: Vergleich der POx-Interaktionen mit aromatischen Aminosäuren an den Konjugationsstellen (A) 18, (B) 19, (C) 27 und (D) 30. PMeOx ist türkis gefärbt, PEtOx violett.

Vergleicht man die einzelnen Konjugationsstellen miteinander (Abb. 4.88), so lässt sich, in Relation zur Pin 1 WW-Domäne (vgl. Abb. 4.80), keine Abhängigkeit der Eigenschaften von der Sekundärstruktur des Konjugationsbereichs erkennen. Der Bedeckungsgrad des Proteins und des jeweiligen Polymers (Abb. 4.88-A und 4.88-B), sowie die entsprechende Anzahl an HB (Abb. 4.88-C) und die Protein-Polymer-Interaktionsenergie (Abb. 4.88-G) sind aber im Falle der Stelle 39 zwischen Helices 2 und 3 relativ gering. Bei PEG kommt es vermehrt zu erhöhten Fluktuationen des Proteinrückgrats (Abb. 4.88-E, Stellen 21, 33, 46, 53), dies lässt sich für die anderen Polymere nur vereinzelt beobachten (für LPG an Stelle 50, für PMeOx an Stelle 49, für PEtOx an den Stellen 46 und 50).

Die Tabelle zu den Protein-Polymer-Wechselwirkungen (Abb. 4.89) zeigt hauptsächlich erhöhte Interaktionen mit Aminosäuren in der Nähe der Konjugationsstelle (d. h. Konjugationen an Helix 1 führen bspw. zu Wechselwirkungen mit der ersten Helix). Die POx-Varianten zeigen teilweise starke Wechselwirkungen mit W14, sowie favorisierte Interaktionen mit F5, auch bei Konjugationen an der dritten Helix. Auch für den Affibody soll das Verhalten der einzelnen Polymere im Folgenden näher beschrieben werden.

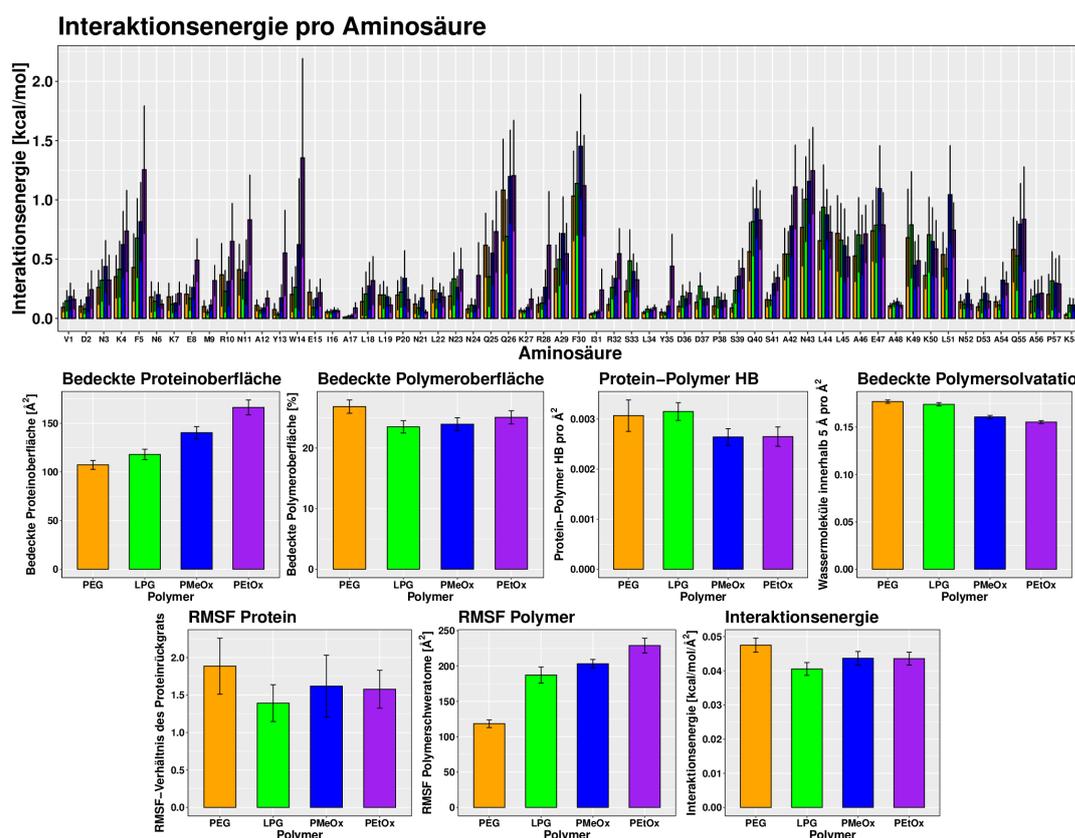


Abbildung 4.87: Mittelwerte untersuchter Eigenschaften für die Her2-Konjugate (aus Werten der Abbildungen in Kap. 4.3.2). Daten für PEG sind orange, für LPG grün, für PMeOx blau und für PEtOx violett gefärbt. Fehlerbalken zeigen den Standardfehler des Mittelwerts.

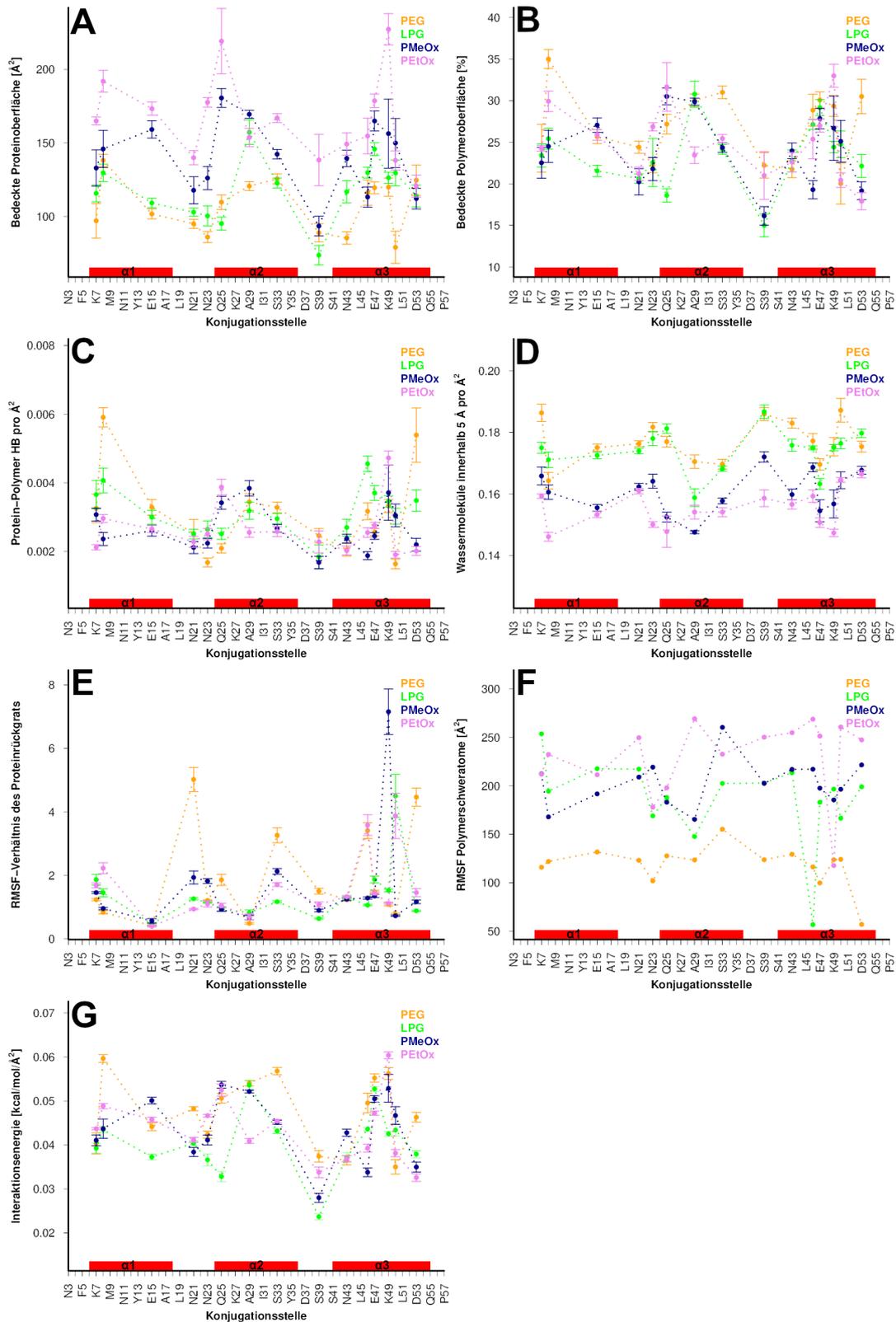


Abbildung 4.88: Übersicht über die Mittelwerte der berechneten Eigenschaften der Affibody-Konjugate, aufgetragen für alle vier Polymere für jede Konjugationsstelle. Fehlerbalken zeigen den Standardfehler des Mittelwerts.

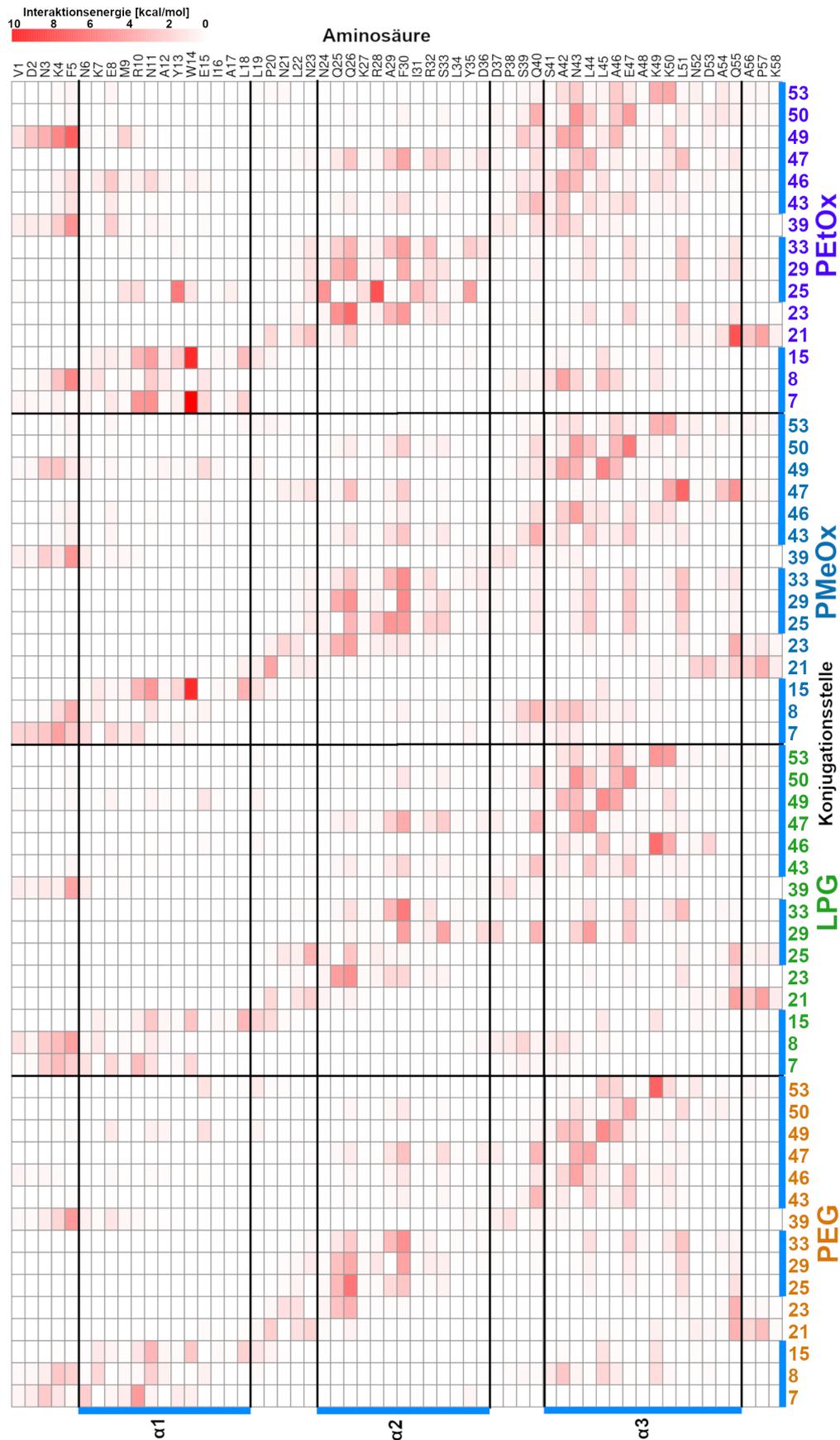


Abbildung 4.89: Durchschnittliche Interaktionsenergien (mit -1 multipliziert, gefärbt von weiß nach rot) zwischen den Polymeren an jeder Konjugationsstelle (Reihen) und jeder einzelnen Aminosäure des Her2-Affibody (Spalten). Sekundärstrukturelemente sind blau gekennzeichnet. Die Interaktionsenergie für die Konjugationsstelle wurde jeweils auf Null gesetzt.

4.3.3.2.2 PEG

Durch die deutlich größere strukturelle Vielfalt des Her2-Affibodys im Vergleich zur kleineren WW-Domäne ergeben sich teils andere Wechselwirkungen. Können, in Anlehnung an Beobachtungen aus [105–108], Interaktionen der PEG-Kette und / oder des Maleimid-Linkers mit Lysinen, besonders benachbarter an den Positionen $i + 3$ und $i + 4$, beobachtet werden? Wechselwirkungen von PEG mit Lysininen lassen sich in der Tat sehr viel öfter als im Falle der WW-Domäne beobachten. In dieser lassen sich auch nur zwei Lysine finden, von denen eines am N-Terminus (K6) und das andere auf der dem oberen hydrophoben Zentrum abgewandten Seite positioniert ist. Interaktionen mit diesen wurden daher erwartungsgemäß wenige gefunden. Im Falle des Her2-Affibodys jedoch lassen sich an vielen Konjugationsstellen entsprechende Wechselwirkungen finden. So interagiert PEG bspw. mit K49 und K50 an den Stellen 7 und 8 (Abb. 4.90-A), mit K4 an Position 39 und mit K50 an Position 43.

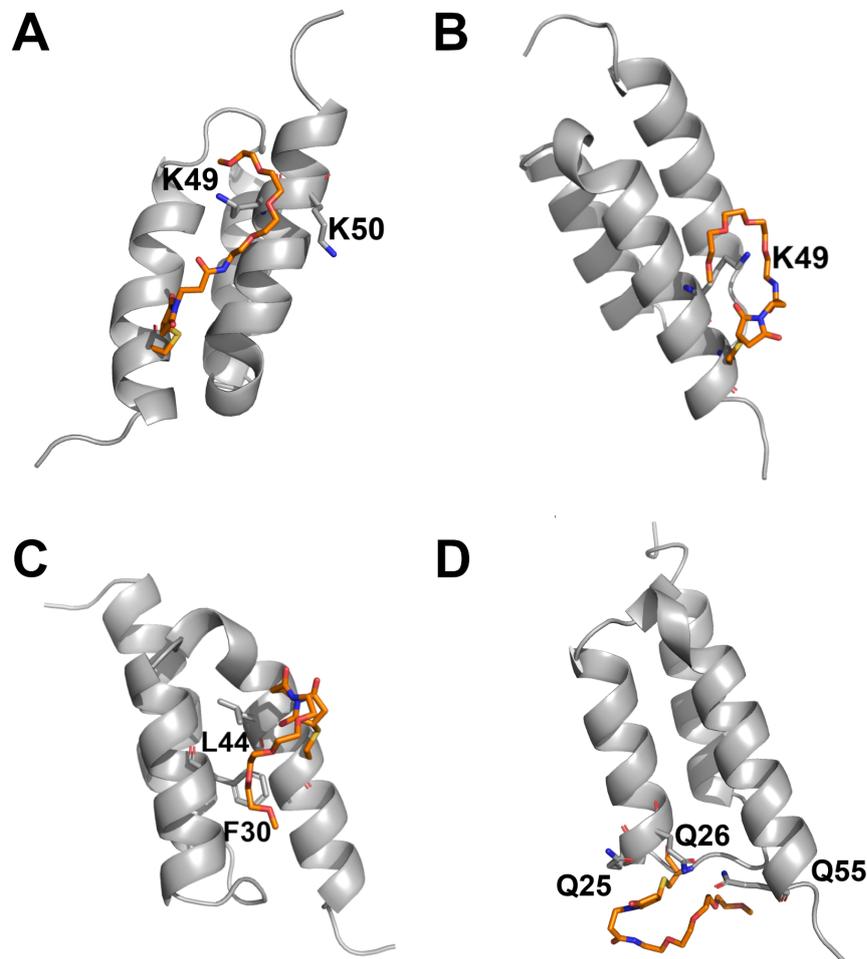


Abbildung 4.90: (A) PEG-Interaktion mit K49 an Konjugationsstelle 8. (B) Interaktionen von PEG und den Carbonylgruppen des Maleimid-Linkers mit K49 an Konjugationsstelle 53. (C) Bedeckung des hydrophoben Areals um F30 und L44 durch die PEG-Kette an Konjugationsstelle 47. (D) Exemplarische PEG-Konformation in der Nähe der Aminosäuren Q25, Q26 und Q55 im Bereich der Schleife zwischen Helices 1 und 2.

Lysine an den Stellen $i + 3$ und $i + 4$ sind im Falle der Konjugationsstellen 46 und 53 vorzufinden (K49 und K50). An ersterer sind keine starken Wechselwirkungen der PEG-Kette mit Lysinen zu beobachten, an letzterer jedoch interagiert das Polymer mit K49. Interaktionen der Carbonylgruppen des Maleimid-Linkers, wie von Price et al. postuliert, werden sowohl an Konjugationsstelle 46 als auch an Stelle 53 beobachtet (Abb. 4.90-B). Auch an Position 39 interagiert eine Carbonylgruppe des Linkers teilweise mit einem nahegelegenen Lysin (K4). Inwiefern diese Interaktionen durch eine PEGylierung beeinträchtigt werden, müsste durch zusätzliche Simulationen der jeweiligen Cys-Maleimid-Strukturen untersucht werden. Dies zeigt jedoch, dass der Cys-Maleimid-Linker, wie auch der Asparagin-Linker der WW-Domäne, bedeutende Wechselwirkungen mit dem Peptid ausbilden kann, die bei einer Konjugation berücksichtigt werden müssen.

An besonders vielen Stellen wird das Areal um F30 und L44 adressiert (Abb. 4.90-C). Ähnlich zu PEG-Konformationen nahe Y23 und F34 der WW-Domäne werden somit einige apolare Proteinbereiche desolvatisiert. Es existieren weiterhin auch viele Interaktionen mit positiv geladenen Argininen (z. B. R10 an Position 7, R28 und R32 an Position 29), ähnlich zu jenen mit R17 und R21 der WW-Domäne. Viele Konjugationsstellen (21, 23, 25) befinden sich nahe der Schleife zwischen der ersten und zweiten Helix. Hier finden sich eine Vielzahl an Glutaminen und Asparaginen (N23, N24, Q25, Q26, Q55). Diese können als HB-Donatoren fungieren, entsprechend sind starke Interaktionsenergien registrierbar (Abb. 4.90-D). Gleichzeitig könnte die Stabilität des Peptids ebenso durch eine desolvatationsbasierte Stärkung der intramolekularen HB, bspw. zwischen Q26 und Q55, moduliert werden, ähnlich wie es Price et al. im Falle von Salzbrücken postulierten [102].

4.3.3.2.3 LPG

Auch im Falle von LPG lassen sich einige Interaktionen mit Lysinen finden, so z. B. an Position 46 mit K49 (wobei hier ebenso der Maleimid-Linker mit dieser Aminosäure wechselwirkt) oder an Position 8 mit K4 (Abb. 4.91-A). Ebenso wird das hydrophobe Areal F30 / L44 adressiert. Am deutlichsten jedoch lassen sich Wechselwirkungen der Hydroxylgruppen mit Aspartaten und Glutamaten finden. Die Anzahl an Carboxylgruppen dieser Aminosäuren ist in der WW-Domäne sehr begrenzt. Im Falle des Her2-Affibodys lassen sich jedoch zahlreiche Interaktionen an verschiedenen Konjugationsstellen beobachten: so z. B. an Position 8 mit D2 und E15, an Position 33 mit E47 und an Position 43 mit D36 und E47 (Abb. 4.91-B). Die funktionellen Gruppen der Aminosäuren N23, N24, Q25, Q26 und Q55 können, ähnlich wie bei PEG, erwartungsgemäß auch als Wechselwirkungspartner dienen.

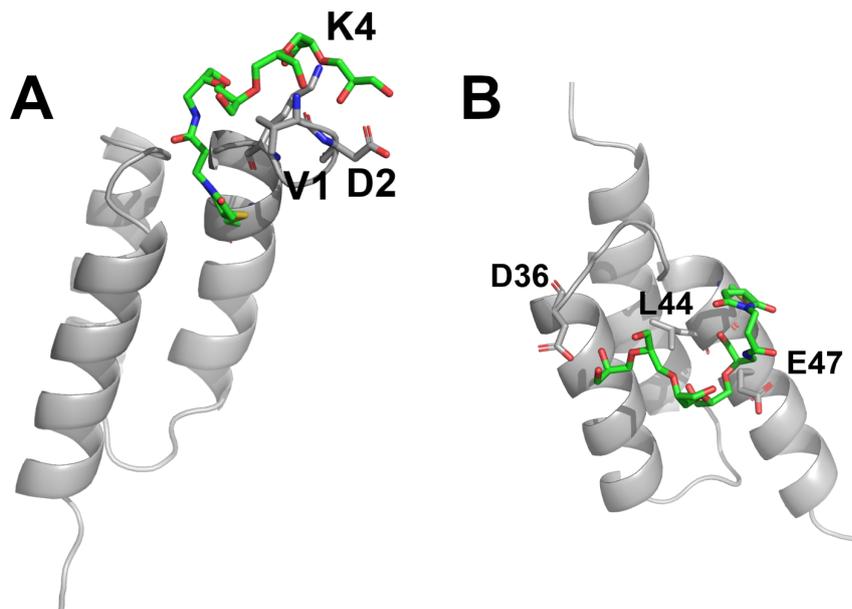


Abbildung 4.91: (A) LPG-Interaktionen mit K4 an Konjugationsstelle 8. (B) LPG-Interaktionen mit D36 und E47 an Position 43.

4.3.3.2.4 PMeOx und PEtOx

Wie bei der WW-Domäne lassen sich zahlreiche Interaktionen mit aromatischen Gruppen beobachten, was auch dem Interaktionsprofil in Abb. 4.87 zu entnehmen ist. So werden bspw. F5 durch Konjugation an Position 8, Y13 und W14 durch Konjugation an Position 7, sowie F30 und Y35 durch Konjugation an der Position 25 adressiert (Abb. 4.92). Es lassen sich, wie bei den anderen Polymervarianten, viele Konformationen in der Nähe des solvensexponierten hydrophoben Areal von F30 / L44 / L45 finden. Analog zu den Konjugaten der WW-Domäne können die Polymere über ihre Carbonylgruppen mit HB-ausbildenden funktionellen Gruppen des Peptids wechselwirken, so z. B. mit N23 an Position 33 oder Q55 an Position 47.

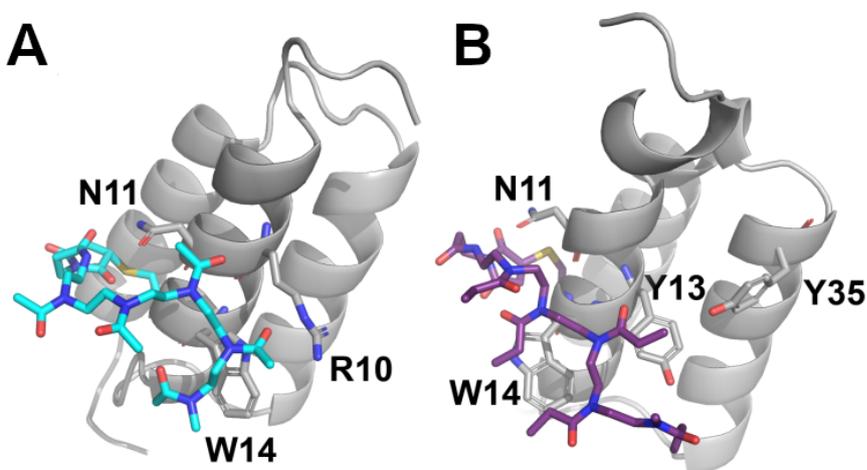


Abbildung 4.92: (A) PMeOx- und (B) PEtOx-Interaktionen an Konjugationsstelle 15.

4.4 Zusammenfassung

In Kap. 4.1.1 wurden verschiedene Modellarten des maschinellen Lernens zur Generierung von Vorhersagemodellen zur PEG-basierten thermodynamischen Stabilisierung getestet. Eine fehlende Limitierung der Anzahl genutzter Deskriptoren führte bei Pin 1-spezifischen Modellen zum kompletten Overfitting. Auch wenn die Pin 1-unspezifischen Modelle mit einer deutlich geringeren Anzahl an Deskriptoren eine sehr gute Leistung bzgl. eines Validierungssets zeigten, so deuten die Beobachtungen auch hier teilweise auf ein Overfitting hin. Weiterhin wurde das jeweils beste Modell auf Basis des Validierungssets ausgewählt, sodass dieses nicht als unvoreingenommen genug angesehen werden kann. Der entwickelte Ansatz diente jedoch als Vorarbeit zu den Modellen in Kap. 4.1.2 und konnte aufzeigen, dass die Nutzbarkeit solcher komplexerer Modellarten bei der vorliegenden Datensatzgröße sehr begrenzt ist. Modelldeskriptoren beschreiben sowohl spezifische Protein-PEG-Interaktionen als auch Modulationen der Solvation in verschiedenen Hydrathüllen. Verschiedene Modelldeskriptoren wurden zum exemplarischen Vergleich strukturell unterschiedlicher Konjugate (z. B. mit Linkern verschiedener Rigidität) mithilfe visueller Analysen der Polymerdynamik und der Proteinsolvatation näher untersucht. Dies signalisiert die Eignung von MDs als Methode zur Abbildung des Verhaltens der Konjugate auf molekularer Ebene.

Der Modellgenerierungsansatz in Kap. 4.1.2, welcher sich auf simplere Modelltypen begrenzte, führte schließlich zu den Modellen A2-R6 und A2-C1. Modell A2-R6 zeigte eine bei einem deutlich erweiterten Datensatz im Vergleich zur kristallstrukturbasierten Richtlinie von Price et al. allgemein verbesserte Leistung. Auch für ein Konjugat der strukturell der WW-Domäne ähnlichen Src SH3-Domäne konnte eine optimierte Vorhersage erhalten werden. Modell A2-C1 stufte dieses Konjugate ebenfalls als stabilisiert ein. Während Modell A2-R6 ähnliche Deskriptoren zu den Pin 1-unspezifischen Regressionsmodellen in Kap. 4.1.1 nutzte, legte Modell A2-C1 den Fokus auf die Solvation der PEG-Kette. Die im Anschluss zusätzlich generierten Modelle A2v2-R1 und A2v2-C1 mit einem aktualisierten Datensatz 1 zeigten vergleichbar gute Statistiken und nutzten verschiedene Abwandlungen des Price-Winkels als Deskriptoren. In nahezu allen Modellen aus Kap. 4.1 wurde der Deskriptor *molbur*, der die durch PEG bedeckte Proteinoberfläche beschreibt, als besonders bedeutend eingestuft. Es stellt sich jedoch die Frage, ob gewisse Einstellungen im Modellgenerierungsansatz (z. B. die Filterschritte bzgl. der Deskriptoren) weiter optimiert werden können. Vorhersagen zu alternativen Polymerarten deuten darauf hin, dass für diese separate Modelle entwickelt werden müssten.

Generell stellt die für QSPR-Ansätze extrem kleine Anzahl an Konjugaten in einem äußerst engen Energieintervall die für diese Studie entscheidende Limitation dar. Deutlich mehr experimentelle Untersuchungen sind notwendig, um optimierte Modelle zu entwickeln. Eine QSPR-Modellgenerierung wurde dennoch vorgenommen, da die

Bedeutung verschiedener Deskriptoren anhand der Modelle interpretiert werden kann und die energetischen Differenzen für eine *in silico*-basierte Berechnung mithilfe deutlich rechenintensiverer, erweiterter Simulationsmethoden (siehe Kap. 4.2) sehr gering sind bzw. im Bereich der für solche Methoden grundsätzlich angenommenen Fehlergrenzen liegen. Ein Fokus auf andere, prinzipiell aus Simulationen sehr viel leichter zugängliche Eigenschaften, wie bspw. die sterische Abschirmung der Binderegionen von Proteinen durch PEG, wäre denkbar. Allerdings standen für diese Studie lediglich die von Price et al. publizierten $\Delta\Delta G_f^0$ -Werte und keine Messwerte zu Bioaktivitäten zur Verfügung.

In Kap. 4.2 wurde ein zu den QSPR-Modellen grundlegend anderer Ansatz gewählt, um den Effekt auf die konformative Stabilität zu untersuchen: Es wurden für zwei exemplarische, besonders stabilisierter und destabilisierter Konjugate Metadynamik-Simulationen durchgeführt. Die berechneten Energiedifferenzen stimmten qualitativ mit den experimentellen Beobachtungen überein, für eine quantitative Aussage sind die Profile jedoch nicht ausreichend konvergiert. Ebenso wurde beobachtet, dass sich die PEG-Ketten im Falle von stabilisierten Varianten auf partiell entfaltete Areale legen, wodurch der Entfaltungsprozess möglicherweise beeinflusst wird. Eine größere Anzahl an Konjugaten mit deutlich längeren Simulationszeiten müsste untersucht werden, um die Eignung dieses Metadynamik-Ansatzes zur Abschätzung der PEG-basierten Stabilisierung evaluieren zu können.

Ein systematischer Vergleich der Dynamik verschiedener Polymervarianten (PEG, LPG, PMeOx und PEtOx) zeigte im Falle von PEG bevorzugte Konformationen in der Nähe solvensexponierter hydrophober Areale (in Übereinstimmung mit [103]), die gleichzeitig HB-Donatoren zur Etablierung von Wechselwirkungen mit den Ethersauerstoffatomen des Polymers bieten. Starke Interaktionen mit Lysininen und Argininen wurden im Falle der strukturell deutlich komplexeren Her2-Affibody-Konjugate registriert. Eine bevorzugte Wechselwirkung von PEG mit Lysininen wurde bereits in der Literatur beschrieben [107]. Ebenso konnten in diesen Simulationen Interaktionen des Maleimid-Linkers mit Lysininen beobachtet werden, in Übereinstimmung mit den Beschreibungen aus [108]. LPG besitzt das selbe Etherrückgrat wie PEG, aber zusätzliche Hydroxymethylseitenketten. Im Falle der WW-Konjugate wurden zu PEG sehr ähnliche Interaktionsprofile registriert. In den Affibody-Strukturen konnten jedoch deutlich mehr Wechselwirkungen mit den hier vermehrt vorkommenden Carboxylgruppen von Aspartaten und Glutamaten beobachtet werden. Desolvationen hydrophober Areale wurden ebenso beobachtet. Die POx-Varianten zeigten deutlich andere Wechselwirkungsprofile. Die hydrophoben Alkylseitenketten interagierten mit aromatischen Gruppen von Tyrosinen und Phenylalaninen. Zusätzlich konnten die Carbonylgruppen der tertiären Amid-Strukturen als HB-Akzeptoren mit verschiedenen Aminosäuren wechselwirken, bspw. Asparaginen oder Glutaminen. An einzelnen Konjugationsstellen ließen sich deutliche Unterschiede zwischen dem Verhalten

von PMeOx und PEtOx erkennen, die mit der sterisch anspruchsvolleren Beschaffenheit des letzteren Polymers erklärt werden können. Im Vergleich zum Her2-Affibody konnte für die Pin 1 WW-Domäne eine größere Abhängigkeit der untersuchten Eigenschaften von der Sekundärstruktur des Konjugationsbereichs beobachtet werden. Konjugationsstellen in den unstrukturierten Bereichen führten bei der WW-Domäne zu stärker solvensexponierten Polymeren, geringeren Wechselwirkungen zum Protein und größeren Fluktuationen des Proteinrückgrats.

Insgesamt konnten in diesem Kapitel QSPR-Modelle zur PEG-basierten Stabilisierung generiert werden, die eine größere strukturelle Vielfalt abdeckten und im Vergleich zur von Price et al. publizierten kristallstrukturbasierten Richtlinie deutlich bessere Vorhersagen lieferten. Für eine weitere Optimierung stellt die vorliegende Datensatzgröße jedoch die entscheidende Limitierung dar. Exemplarische Metadynamik-Simulationen konnten die experimentell ermittelten energetischen Differenzen qualitativ erfassen und zeigten eine mögliche Dynamik der PEG-Ketten beim Entfaltungsprozess auf. Systematische Vergleiche der Dynamik verschiedener Polymerarten zeigten favorisierte Interaktionspartner auf.

Kapitel 5

Ergebnisse: Biologika

5.1 Gaußbeschleunigte Molekulardynamik-Simulationen

Durch gaußbeschleunigte Molekulardynamik-Simulationen zehn verschiedener Startkonformationen (insgesamt $1.5 \mu\text{s}$ pro Konjugat) konnten lokale Wechselwirkungen mit allen Proteinbereichen näher studiert werden. Zur Analyse der Interaktionen wurden Polymerdichten und entsprechende Kontaktprofile berechnet. Diese Aspekte sollen im vorliegenden Kapitel interpretiert werden, zusätzliche Analysen zur experimentell bestimmten thermischen Stabilität sind in Kap. 5.3, jene zur Bioaktivität in Kap. 5.4 aufgeführt.

5.1.1 Polymerdichten

Es wurden Polymerdichten berechnet, die die Gitterpunkte repräsentieren, für die ein Wert von mindestens 30 % der in der Simulation höchsten gefundenen Aufenthaltswahrscheinlichkeit der Polymeratome registriert wurde (Abb. 5.1). Diese Analysen gewähren einen ersten Eindruck von der Verteilung der Polymeratome auf der Proteinoberfläche und der bevorzugten Wechselwirkungsareale. Es zeigt sich, dass in allen Fällen die Binderegionen für die Rezeptoren IFNAR1 und IFNAR2 teilweise oder vollständig von den Dichten bedeckt werden, was für eine generell verringerte Bioaktivität im Vergleich zum Wildtyp spricht (vgl. Kap. 5.4). Die Dichten nehmen bei größeren Polymerketten (20 - 50 kDa) ein vergleichbar großes Volumen um das Protein ein, auch bei den größten Konjugaten (30 - 50 kDa) sind jedoch unbedeckte Proteinbereiche zu finden. Vor allem im Falle der PEOx-Konjugate zeigen sich größere Areale mit freiliegender Proteinoberfläche, was darauf hindeutet, dass die Interaktionen des Proteins mit dieser Polymerart am geringsten ausgeprägt sind. Die Aufenthaltspunkte der Polymere der DBCO-Konjugate sind im Vergleich zu jenen der BCN-10-kDa-Varianten leicht verändert. So zeigt sich bspw. für PEG eine größere Dichte im oberen Bereich unmittelbar um die Konjugationsstelle. PEOx zeigt beim DBCO-Konjugat eine geringere Tendenz der Adressierung dieses Areals. Für einen detaillierteren Blick auf mögliche Protein-Polymer-Wechselwirkungen sollen die nachfolgend beschriebenen Kontaktprofile dienen.

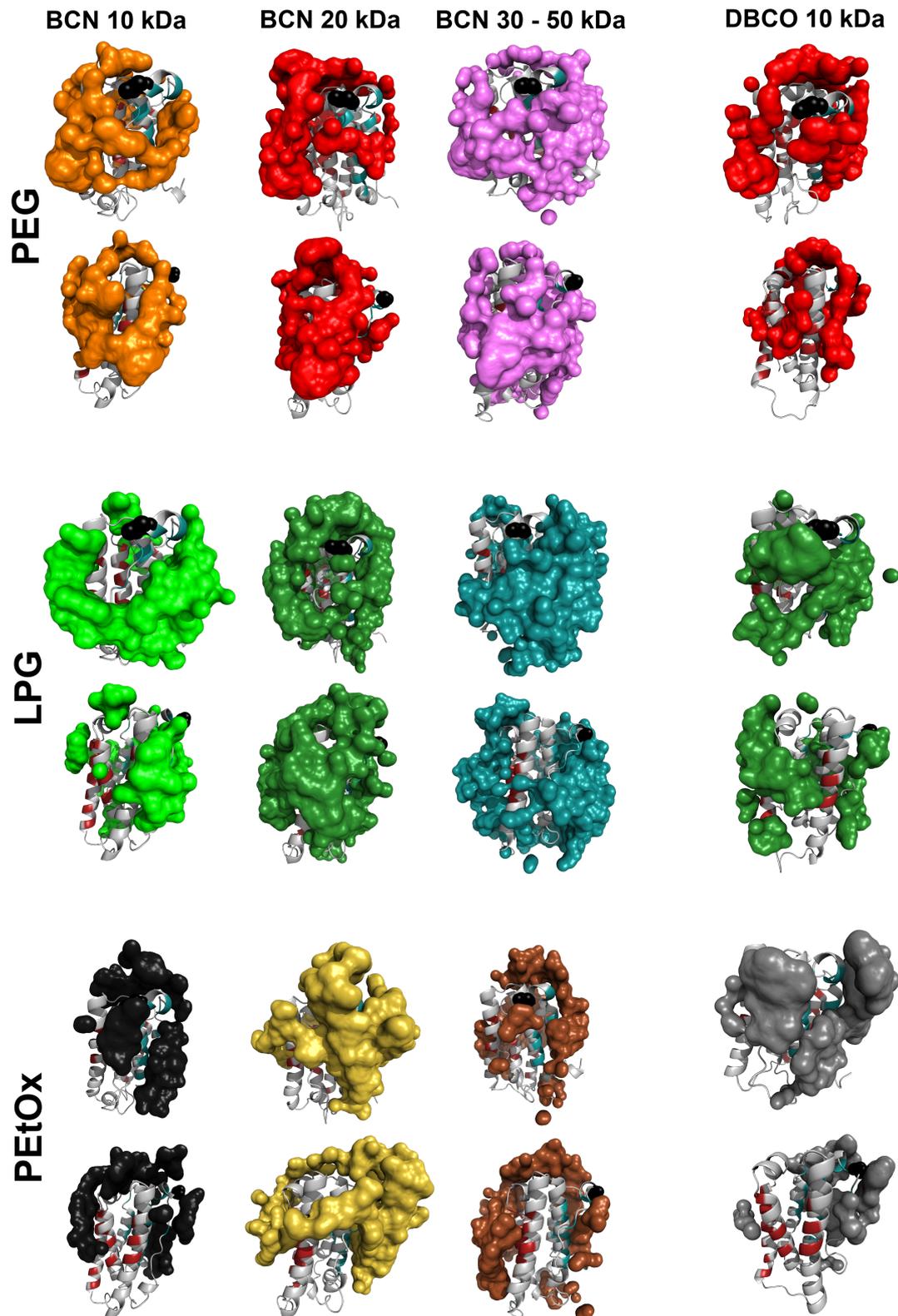


Abbildung 5.1: Polymerdichten für die Simulationen der Konjugate mit BCN-Linker für PEG (10 kDa: orange, 20 kDa: rot, 40 kDa: pink), LPG (10 kDa: hellgrün, 20 kDa: grün, 50 kDa: cyan) und PEtOx (10 kDa: schwarz, 20 kDa: sandfarben, 30 kDa: braun), sowie ganz rechts für jene mit DBCO-Linker (PEG: rot, LPG: grün, PEtOx: grau). Diese sind aus jeweils zwei Blickrichtungen dargestellt und repräsentieren alle Gitterpunkte mit mindestens 30 % der höchsten ermittelten Polymerdichte in der jeweiligen Simulation. Die erste Momentaufnahme ist als Cartoon dargestellt, in welchem die Bindungsstellen von IFNAR1 (rot) und IFNAR2 (grün) koloriert sind. Die Konjugationsstelle ist zusätzlich mit schwarzen Kugeln markiert (Nachdruck mit Genehmigung aus [304], Copyright 2021 American Chemical Society).

5.1.2 Kontaktprofile

Zur Bestimmung favorisierter Wechselwirkungsbereiche auf der Proteinoberfläche wurde zunächst der durchschnittliche Anteil jeder Aminosäure an der solvenszugänglichen Oberfläche des Proteins in den zehn Simulationen des Wildtyps über den *sasa*-Befehl in VMD [326] mit einer Solvenssphäre von 1.4 Å bestimmt (Abb. 5.2). Anschließend wurde jeweils der prozentuale Anteil jeder einzelnen Aminosäure an der gesamten Bedeckung der Polymeroberfläche für alle 0.1 ns in den Simulationen bestimmt. Die Prozente wurden ebenso für alle 20 verschiedenen proteinogenen Aminosäuretypen aufsummiert. Im Falle einer komplett ungerichteten Interaktion mit dem Protein sollte der ermittelte Anteil mit dem prozentualen Anteil der Aminosäure an der solvenszugänglichen Proteinoberfläche übereinstimmen, da sich das Polymer gleichmäßig über das Protein verteilen würde und lang genug ist, um jeden Bereich des Proteins erreichen zu können. Liegt der Anteil an der Bedeckung des Polymers aber sehr viel höher, so spricht dies für eine favorisierte Wechselwirkung. Bei dieser Analyse wurden die Werte für die Konjugationsstelle K31 auf Null gesetzt bzw. herausgefiltert.

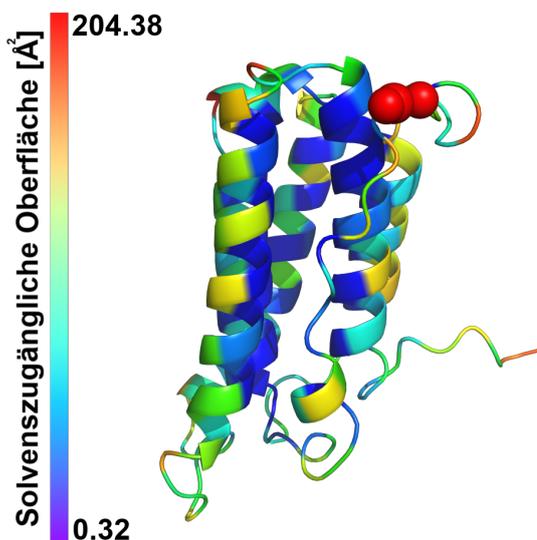


Abbildung 5.2: Illustration der IFN- α 2a-Struktur, gefärbt nach den per VMD ermittelten Absolutwerten der solvenszugänglichen Oberflächen (die Konjugationsstelle ist zusätzlich als Kugeln dargestellt). Für die Kontaktprofile wurden die prozentualen Anteile genutzt, wobei der Wert für K31 auf Null gesetzt wurde.

Es ist zu erwarten, dass bei einer solchen Analyse hydrophobe Interaktionen, bei welchen die Polymeratome große Flächen der Aminosäuren (bspw. F, Y, L) adressieren, zu größeren Anteilen an bedeckter Polymeroberfläche führen. Interagiert das Polymer hingegen über eine HB spezifisch mit lediglich einem Bereich der Aminosäureseitenkette (z. B. einem Amid-Stickstoffatom von N/Q), so wird die restliche Oberfläche dieser Aminosäure nicht zwingend adressiert. Weiterhin fließen in die Ermittlung der mittleren Anteile an der bedeckten Polymeroberfläche auch wiederkehrende transiente Annäherungen des Polymers an das Protein ein, die über die Zeit hinweg nicht unbedingt lange bestehen bleiben.

Als eine ergänzende Analyse der stabilsten, spezifischen Wechselwirkungen wurden daher ebenfalls die längsten konsekutiven Aufenthaltsdauern des Polymers um die Aminosäuren herum ermittelt (sozusagen die maximalen Überlebensdauern der Kontakte). Dies entspricht der Vorgehensweise aus [107]. Hierfür wurde ein Kontakt in einer Momentaufnahme registriert, sobald sich ein Polymeratom (Wasserstoffatome eingeschlossen) innerhalb von 3 Å eines spezifischen Seitenkettenschweratoms einer Aminosäure befand (im Falle von Glycin das C $_{\alpha}$ -Atom). Es wurde dann für jede Aminosäure die größte Anzahl an konsekutiven Momentaufnahmen ermittelt, in denen der Kontakt bestand (max. 150 ns, was der Länge eines Replikons entspricht). Auch hier wurden die Kontakte mit der Konjugationsstelle K31, die zu jeder Zeit bestehen, herausgefiltert.

5.1.2.1 IFN- α 2a-K₃₁-BCN-10-kDa

5.1.2.1.1 BCN-PEG-10-kDa

PEG adressiert v. a. Aminosäuren im Bereich zwischen R120 und Y135 (Abb. 5.3-A). Dieses Areal der vierten Helix besteht aus einer Vielzahl positiv geladener Seitenketten (R120, K121, R125, K131, K134). Beim Vergleich des Anteils der Bedeckung der Polymeroberfläche mit der solvenzugänglichen Oberfläche zeigen die hydrophoben Aminosäuren L, Y, F, A und die positiv geladenen Aminosäuren R und K die größte Präferenz (Abb. 5.3-B und 5.3-C). Die teils recht großen Standardfehler sind darauf zurückzuführen, dass unterschiedliche Areale des Proteins in den verschiedenen Replika adressiert werden. Eine favorisierte Interaktion von PEG mit Lysininen und Argininen zeigt sich auch im Hinblick auf die maximalen Kontaktzeiten (Abb. 5.3-D). Bei visueller Betrachtung der Simulationen lässt sich eine gehäufte Windung der Polymerketten um die positiv geladenen Seitenketten von R und K beobachten (Abb. 5.4-A und 5.4-B). Die drei Lysine mit der größten maximalen Kontaktzeit (K70, K131, K134) liegen dicht beieinander in unmittelbarer Nähe der Konjugationsstelle (Abb. 5.4-C).

Eine bevorzugte Windung von PEG-Ketten um positiv geladene Lysine herum wurde in einer neueren Publikation, in der Simulationen von PEG-Biokonjugaten durchgeführt wurden, bereits beschrieben [107]. In Übereinstimmung mit dieser Publikation deutet die bevorzugte Bedeckung der Aminosäuren L, Y, F und A darauf hin, dass bei der Interaktion mit Lysininen und Argininen gleichzeitig nahegelegene hydrophobe Areale adressiert werden (z. B. L117 und L128 in Helix 4). Die Tendenz zur Interaktion von PEG über Ethersauerstoffatome mit Kationen wurde bzgl. positiv geladener Ionen bereits früh in der Literatur beschrieben [327, 328] und ebenso im Falle von PEGylierten Lipidmembranen beobachtet, wobei ein Kation in der Regel durch fünf Ethersauerstoffatome koordiniert wurde [23]. Der häufige Einsatz von PEG bei Proteinkristallisationen resultiert weiterhin in einer Vielzahl an PDB-Strukturen, in denen PEG-Oligomere in der Nähe von Lysininen zu finden sind [107].

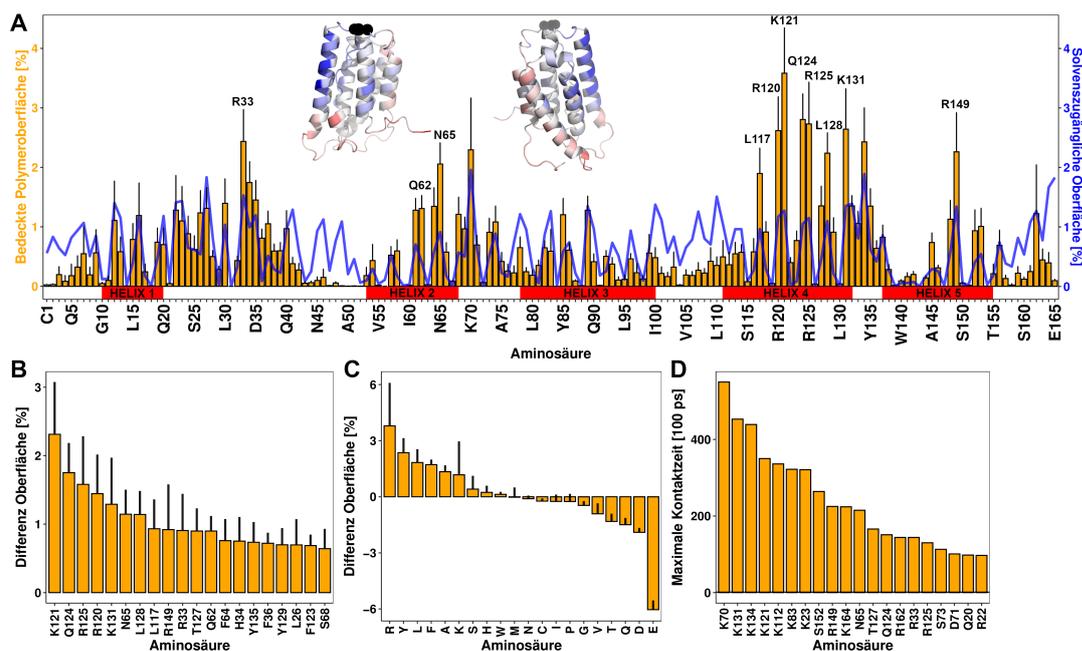


Abbildung 5.3: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des 10-kDa-BCN-PEG-Konjugats in Relation zum Anteil an der solvenszugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenszugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächendifferenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

5.1.2.1.2 BCN-LPG-10-kDa

Im Falle von LPG, welches sich aus dem selben Rückgrat mit Ethersauerstoffatomen zusammensetzt, werden ebenso einige Arginine adressiert (R22, R125, R149), es lässt sich jedoch eine geringere Tendenz zur Adressierung der vierten Helix beobachten (Abb. 5.5-A). Insgesamt zeigt sich eine Verteilung der Anteile an der bedeckten Polymeroberfläche, die eher jener der solvenszugänglichen Oberflächen entspricht. Wechselwirkungen mit Aminosäuren im Bereich um die erste Helix (M16, R22) sind häufiger zu beobachten. Die aufsummierten Werte zeigen wiederum Wechselwirkungen mit hydrophoben Aminosäuren wie Y, F, L, A, aber in diesem Falle auch Methioninen und Serinen (Abb. 5.5-C). Arginine zeigen eine stark positive Differenz, dagegen stehen Lysine in der Rangfolge im Vergleich zu den Resultaten der PEG-Simulationen deutlich weiter hinten. Erhöhte Bedeckungen des unteren unstrukturierten Bereichs zwischen Helix 1 und Helix 2 resultieren weiterhin in stärkeren Interaktionen mit Glutaminen (v. a. Q40).

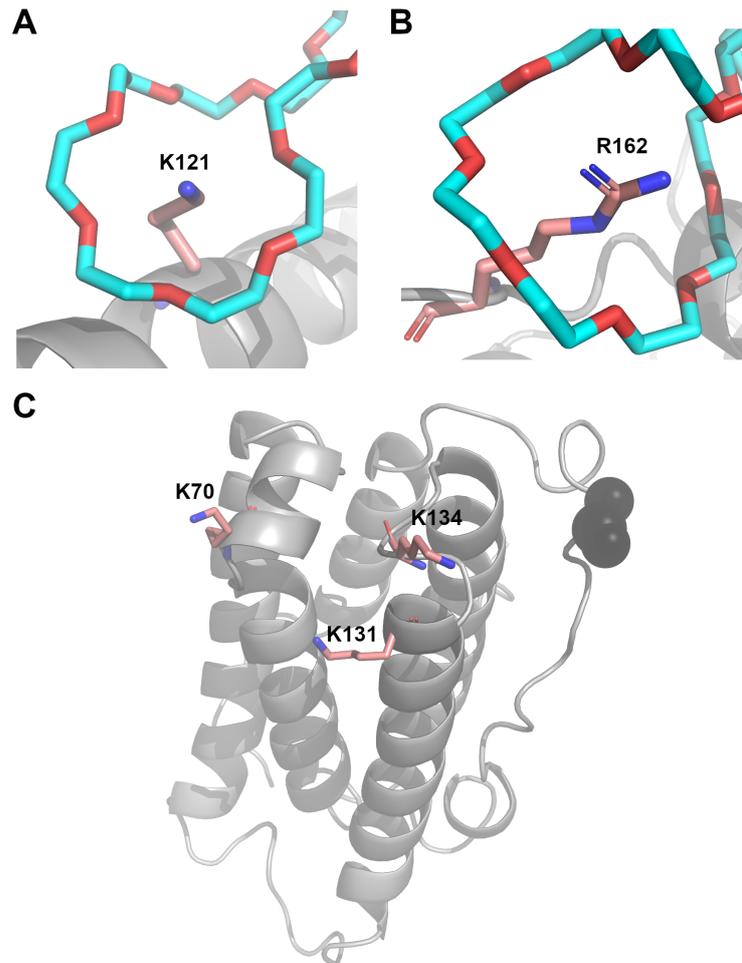


Abbildung 5.4: Darstellungen exemplarischer Interaktionen mit (A) Lysin und (B) Arginin in der Simulation der 10-kDa-PEG-Variante. Eine wie im Text beschriebene Koordination über fünf Ethersauerstoffatome ist hier gegeben. Wasserstoffatome sind ausgeblendet. In (C) ist die Verteilung der am stärksten adressierten Lysine illustriert. Dabei ist die Konjugationsstelle mit schwarzen Kugeln gekennzeichnet.

Die maximalen Kontaktzeiten zeichnen ein anderes Bild. Neben Argininen und Lysinen sind hier einige Aspartate und Glutamate vorzufinden (E51, D71, D82, E87), die über ihre Carboxylgruppen die HB-Donatoren der LPG-Hydroxymethylseitenketten adressieren können (Abb. 5.6). Aspartat steht in der Rangfolge in Abb. 5.5-C im Vergleich zu den PEG-Simulationen entsprechend weiter vorne, Glutamat dagegen verbleibt hier an letzter Stelle (die Differenz ist jedoch nicht so negativ wie in Abb. 5.3-C). Dies ist v. a. darauf zurückzuführen, dass E165 den C-Terminus und damit einen sehr großen Anteil der solvenszugänglichen Oberfläche darstellt. Der Anteil an bedeckter Polymeroberfläche ist für diese Aminosäure im Vergleich zu Abb. 5.3-C sehr viel höher, liegt aber dennoch unterhalb der entsprechenden SASA. Generell sind die maximalen Kontaktzeiten in Abb. 5.5-D im Vergleich zu Abb. 5.3-D deutlich länger, sodass davon ausgegangen werden kann, dass LPG in der Lage ist, stabilere Wechselwirkungen mit dem Protein zu etablieren.

Es gibt aufgrund des gemeinsamen Polymerrückgrats also insgesamt große Überschneidungen mit dem Verhalten von PEG (Wechselwirkungen mit hydrophoben Aminosäuren und Argininen), allerdings führen die Seitenketten von LPG zu einem erweiterten Interaktionsprofil, indem v. a. zusätzliche HB-Akzeptoren (Carboxylgruppen von D/E) adressiert werden können. Weiterhin scheint die Tendenz der Bedeckung von Serinen (z. B. S163) und Methioninen (M16, M148) etwas erhöht zu sein.

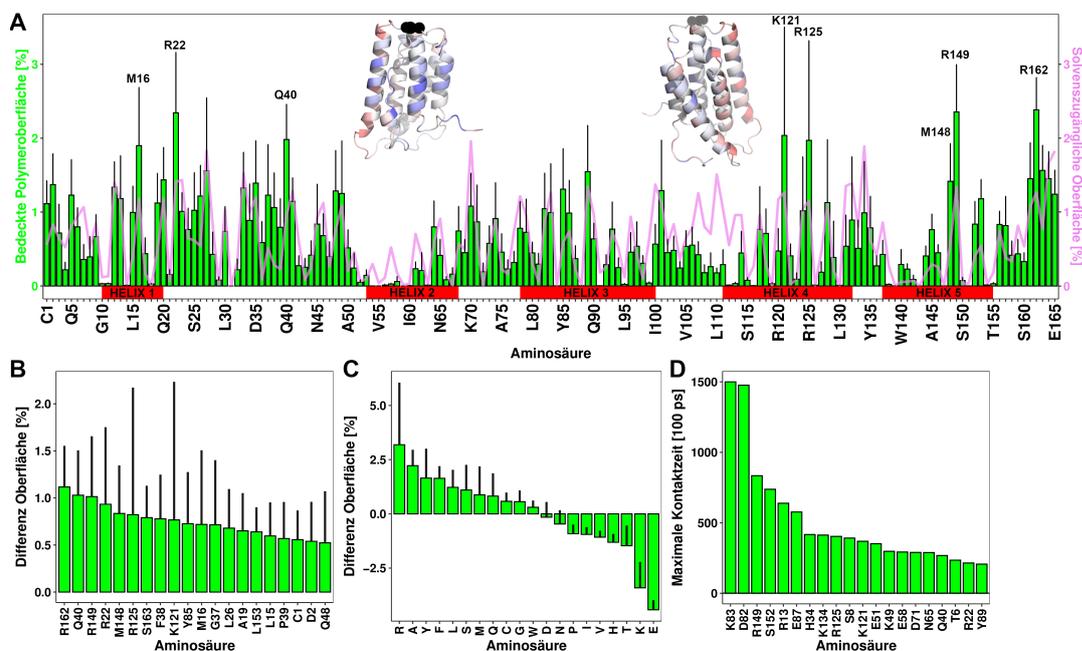


Abbildung 5.5: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des 10-kDa-BCN-LPG-Konjugats in Relation zum Anteil an der solvenszugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenszugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächendifferenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

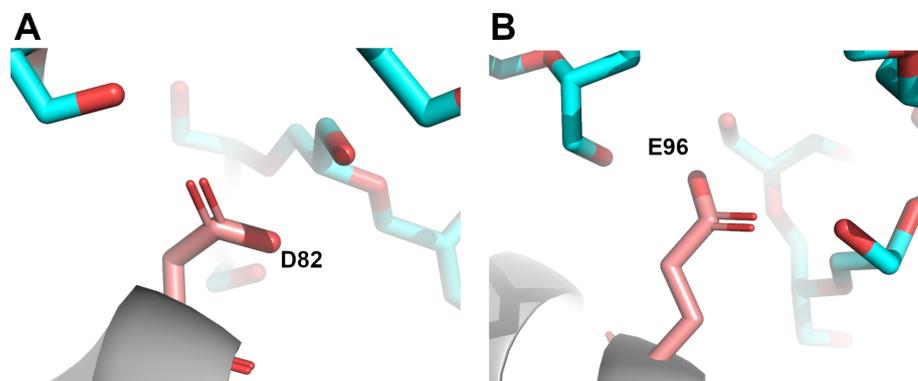


Abbildung 5.6: Darstellungen exemplarischer Interaktionen mit (A) Aspartaten und (B) Glutamaten in der Simulation der 10-kDa-LPG-Variante. Wasserstoffatome sind ausgeblendet.

5.1.2.1.3 BCN-PEtOx-10-kDa

PEtOx mit seinen größeren, teils hydrophoben Seitenketten zeigt wiederum ein deutlich anderes Kontaktprofil. Es werden v. a. die Aminosäuren des flexiblen Bereichs zwischen der ersten und der zweiten Helix adressiert, in welchem sich auch die Konjugationsstelle befindet. Die Carbonylgruppe der Seitenkette kann als potenzieller HB-Akzeptor fungieren. Dies ist aus Interaktionen mit Hydroxylgruppen von Serinen (z. B. S72, S152), sowie Amidgruppen von Glutaminen (z. B. Q40, Q46, Q61) und Asparaginen (z. B. N45, N65, N156) ersichtlich (Abb. 5.7 und 5.8-A). Weiterhin scheinen v. a. Wechselwirkungen mit Tyrosinen (Y85, Y89, Y135) und Phenylalaninen (z. B. F27, F64) vorzukommen. Hierbei lassen sich Konformationen erkennen, bei denen die hydrophobe Seitenkette auf diesen Arealen aufliegt (Abb. 5.8-B).

Diese Beobachtungen resultieren in einer im Vergleich zu den PEG/LPG-Simulationen anderen Rangfolge der Flächendifferenzen (Abb. 5.7-C). F und Y stehen in dieser weiter vorne. Auch die Differenzen der Serine und Asparagine sind sehr viel positiver als bei PEG und LPG. Wie auch im Falle des LPG werden ebenfalls Methionine adressiert (M16 und M148). Insgesamt ist das Wechselwirkungsprofil also von Interaktionen mit aromatischen Aminosäuren (Y, F) und HB-Donatoren (S, N, Q, R) für die Carbonylgruppe der Polymerseitenketten geprägt. Die Absolutwerte der längsten Kontaktzeiten sind vergleichbar mit jenen in den PEG-Simulationen.

5.1.2.2 IFN- α 2a-K₃₁-DBCO-10-kDa

5.1.2.2.1 DBCO-PEG-10-kDa

Wie beeinflusst die Nutzung des DBCO-Linkers das Interaktionsprofil? Analoge Kontaktprofile zu den BCN-Konjugaten sind in den Abb. 5.9, 5.10 und 5.11 gezeigt. Im Falle von PEG zeigen sich gehäufte Wechselwirkungen mit Aminosäuren in der Nähe der Konjugationsstelle (R22, K23, R33) und dem C-terminal gelegenen R162. Interaktionen mit den Aminosäuren der vierten Helix (K121, R125, K134) sind weiterhin zu beobachten, jedoch weitaus seltener (Abb. 5.9-A). Wie auch im Falle der BCN-Variante zeigen sich v. a. Wechselwirkungen mit Argininen und Lysinen, sowie hydrophoberen Aminosäuren (L, F und A). Ähnlich zu den BCN-LPG- und BCN-PEtOx-Varianten, die eine Tendenz zur Adressierung des unstrukturierten Bereichs um die Konjugationsstelle aufweisen, sind hier Polymerbedeckungen durch M16 erkennbar. Weiterhin werden Serine deutlich öfter adressiert (S25, S136, S152). Während sich also das bevorzugte Proteinareal in diesen Simulationen des DBCO-PEG-Konjugats von jenem des BCN-PEG-Konjugats unterscheidet, handelt es sich, mit Ausnahme der Serine, zum größten Teil um die selben Aminosäuretypen.

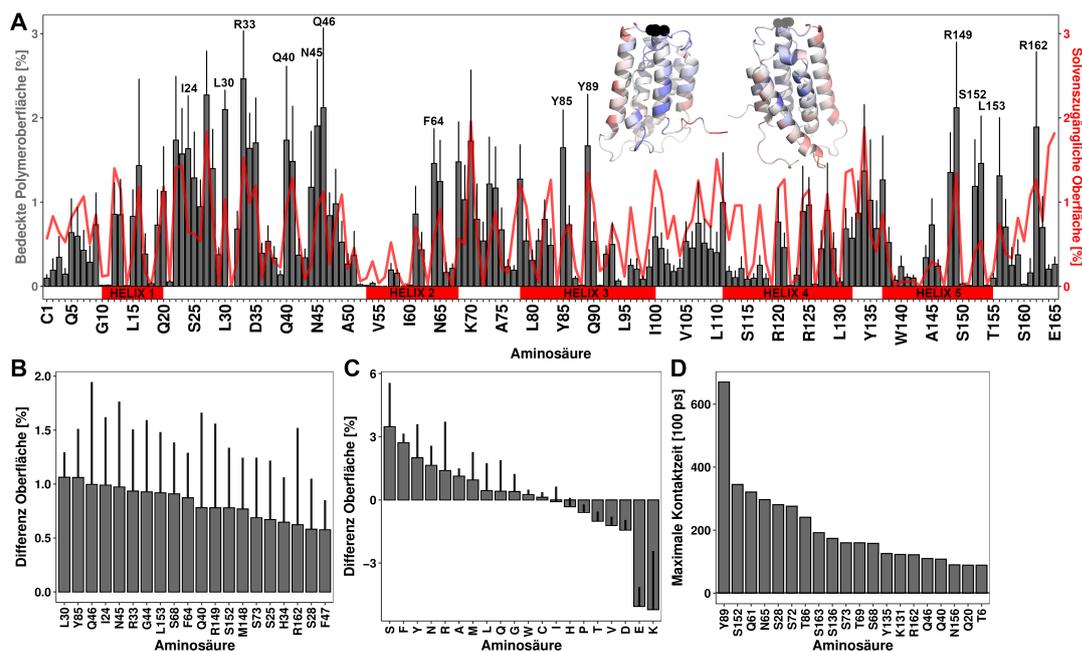


Abbildung 5.7: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des 10-kDa-BCN-PEtOx-Konjugats in Relation zum Anteil an der solvenz zugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenz zugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächendifferenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

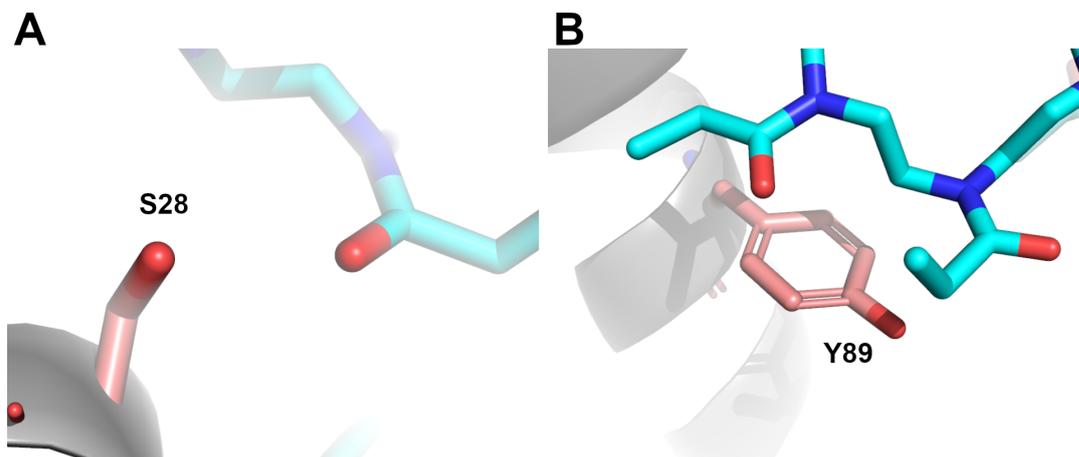


Abbildung 5.8: Darstellungen exemplarischer Interaktionen mit (A) Serinen und (B) Tyrosinen in der Simulation der 10-kDa-PEtOx-Variante. Wasserstoffatome sind ausgeblendet.

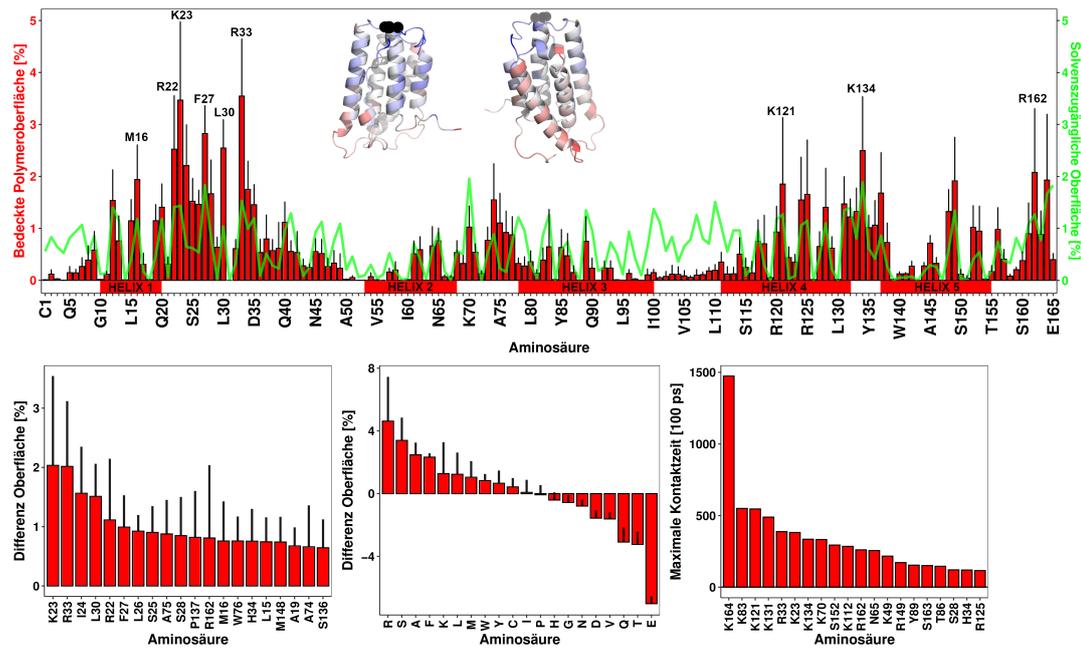


Abbildung 5.9: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des 10-kDa-DBCO-PEG-Konjugats in Relation zum Anteil an der solvenszugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenszugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächendifferenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

5.1.2.2.2 DBCO-LPG-10-kDa

Auch die DBCO-LPG-Variante zeigt eine im Vergleich zur BCN-LPG-Variante geringere Tendenz, die vierte Helix zu adressieren (Abb. 5.10-A). Stattdessen zeigt das Polymer eine höhere Wahrscheinlichkeit, mit dem unteren Bereich um Q40 - Q48 zwischen Helix 1 und Helix 2 zu wechselwirken. Im Falle des BCN-Derivats wurden eher Aminosäuren am Beginn dieses unstrukturierten Bereichs oder innerhalb der ersten Helix adressiert (L15, M16, R22). Somit wird auch bei LPG durch den Austausch des Linkers ein anderes Areal des Proteins adressiert, im Gegensatz zu BCN-LPG scheinen hierbei Asparagine, Glutamine und Serine öfter adressiert zu werden (vgl. Abb. 5.5-C und Abb. 5.10-C). Wie auch bei BCN-LPG zeigen die maximalen Kontaktzeiten dagegen eine Präferenz mit Aspartaten und Glutamaten. Mit Ausnahme der stabilen Wechselwirkungen der BCN-LPG-Variante mit D82 und K83 sind die Absolutwerte der Kontaktzeiten der DBCO- vergleichbar mit jener der BCN-Variante.

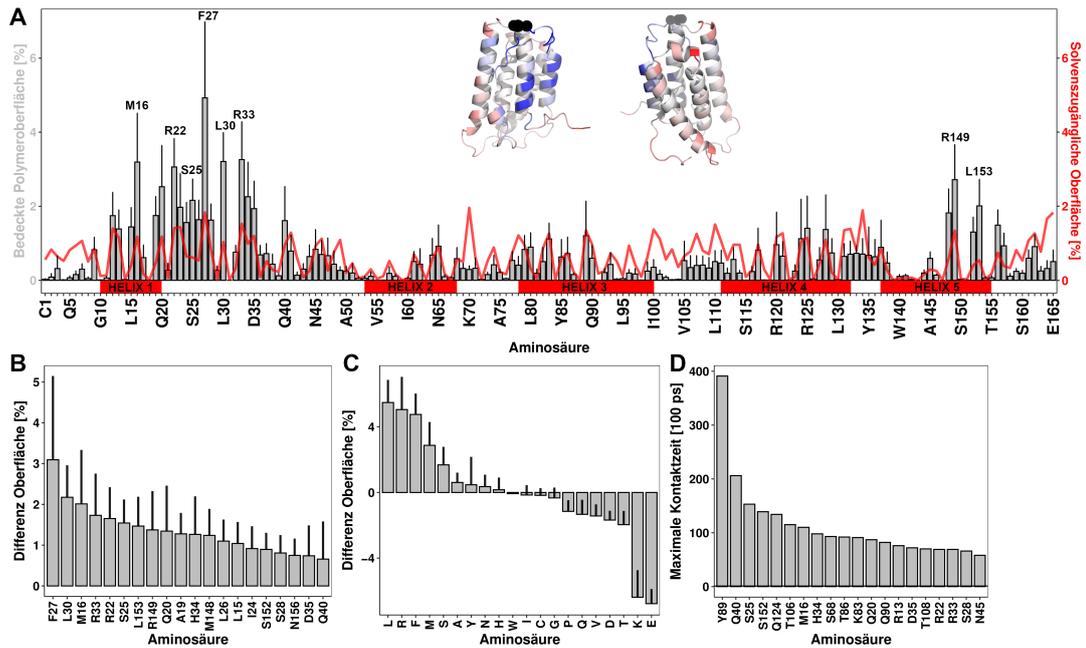


Abbildung 5.11: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des 10-kDa-DBCO-PETox-Konjugats in Relation zum Anteil an der solvenszugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenszugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächendifferenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

5.1.2.3 Acyl-PEG-10-kDa

Zusätzlich zum Vergleich zwischen BCN- und DBCO-Linker wurde ein Replikon der BCN-PEG-10-kDa-Variante für 150 ns simuliert, in der der Linker in der Startkonformation durch eine einfache Peptidbindung bzw. Acylgruppe ausgetauscht wurde. Eine solche wird oft bei klassischen Kupplungsreaktionen ohne Click-Chemie-Verfahren unter Nutzung eines Succinimidylesters erhalten (siehe erstes Beispiel in Tab. 1.3). Sehr ähnliche Proteinoberflächen werden durch PEG mit dieser Linkerart, verglichen mit BCN, adressiert (Abb. 5.12). Während das acylierte Lysin teilweise ins Solvens gestreckt ist (im Gegensatz zum BCN-Linker, welcher zumeist direkt auf der Proteinoberfläche aufliegt), wechselwirken die ersten PEG-Monomere mit den selben Aminosäuren (K121 und R149). Die von PEG bedeckte Proteinoberfläche beträgt im Falle der K31-BCN-Variante 21.98 ± 0.05 % und beim acylierten Derivat 20.84 ± 0.10 %. Dieser exemplarische Vergleich zeigt damit auf, dass der in diesen Studien genutzte BCN-Linker trotz seiner Größe nicht zu einer drastischen Verringerung der vom Polymer adressierten Oberfläche führt [304].

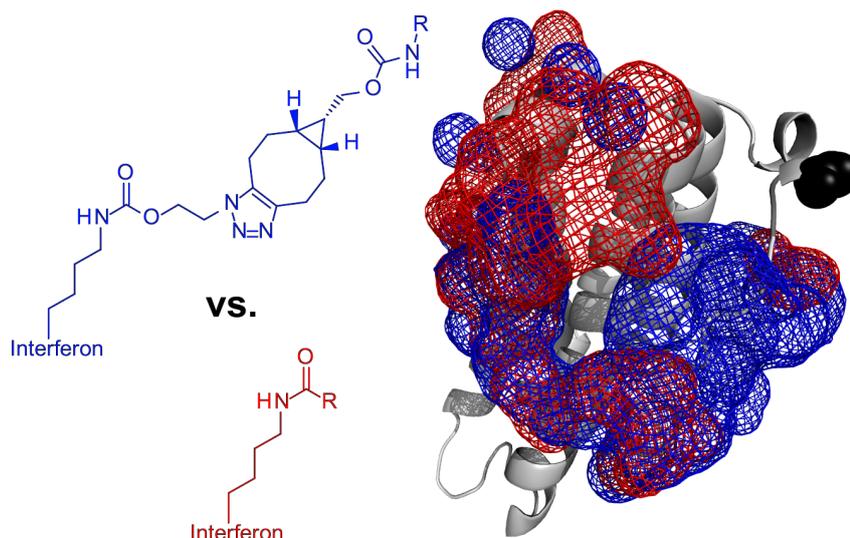


Abbildung 5.12: Polymerdichten (Berechnung analog zu Abb. 5.1) für die Simulationen der K31-BCN-PEG- (blaues Gitter) und K31-Acyl-PEG-Varianten (rotes Gitter). Rückgrat-atome der Konjugationsstelle sind als schwarze Kugeln dargestellt (Nachdruck mit Genehmigung aus [304], Copyright 2021 American Chemical Society).

5.1.2.4 Größere BCN-Konjugate (20 - 50 kDa)

Werden die Kontaktprofile durch Nutzung längerer Polymerketten beeinflusst? Im Folgenden sollen auch die Kontaktprofile der BCN-Konjugate mit längeren Polymerketten (20 - 50 kDa) kurz beschrieben werden. Dabei sollte angemerkt werden, dass diese Varianten aus Zeitgründen genauso lange wie die 10-kDa-Varianten simuliert wurden, obwohl diese Systeme deutlich mehr Freiheitsgrade beinhalten.

5.1.2.4.1 20 kDa

Im Falle von BCN-PEG-20-kDa zeigt sich ein zur 10-kDa-Variante nahezu identisches Profil: Das Polymer wechselwirkt bevorzugt mit Aminosäuren der Helix 4, v. a. Argininen und Lysinen (Abb. 5.13). Abermals zeigen sich positive Flächendifferenzen für die hydrophoben Aminosäuren Tyrosin, Phenylalanin und Alanin. R22, K23 und R33 in der Nähe der Konjugationsstelle werden allerdings häufiger adressiert. Auch die maximalen Kontaktzeiten zeigen generell etwas größere Werte.

BCN-LPG-20-kDa zeigt ein Profil, welches auf eine gleichmäßigere Bedeckung der Proteinoberfläche hindeutet, als es bei der 10-kDa-Variante der Fall ist (Abb. 5.14-A). So werden im Vergleich zu letzterer auch vermehrt Aminosäuren im Bereich der zweiten Helix und zwischen Helix 3 und 4 adressiert. Nur noch die Termini stellen größere Bereiche mit negativen Flächendifferenzen dar. Viele Arginine werden adressiert (R22, R120, R125), diese fallen in der Rangfolge dennoch zurück (Abb. 5.14-C). Dagegen rückt Cystein (C29, C138) nach vorne. Weiterhin kommt es auch zur Bedeckung von I24, welches sich in unmittelbarer Nähe zu R22 und K23 befindet. Dies führt zu insgesamt sehr positiven Flächendifferenzen für Isoleucin. Serine zeigen die positivste Flächendifferenz. Weiterhin zeigen sich (noch ausgeprägter als bei der

10-kDa-Variante) stabile Wechselwirkungen mit Glutamaten und Aspartaten (z. B. D32, E58, D77, E78, E132, Abb. 5.14-D), die Absolutwerte der Kontaktzeiten sind vergleichbar mit jenen aus Abb. 5.5-D.

BCN-PEtOx-20-kDa zeigt eine im Vergleich zum 10-kDa-Derivat geringere Tendenz zur Bedeckung der Aminosäuren im unteren Areal des unstrukturierten Bereichs zwischen Helix 1 und 2 (Q40 - A50, Abb. 5.15). Entsprechend rücken Asparagine und Glutamine, die in diesem Bereich vermehrt vorzufinden sind, in der Rangfolge bzgl. der Flächendifferenzen nach hinten. Aminosäuren um F27, die auch schon von der 10-kDa-Variante adressiert wurden, zeigen hingegen einen noch größeren Anteil an der Polymerbedeckung.

5.1.2.4.2 30 - 50 kDa

Auch für das BCN-PEG-40-kDa-Konjugat zeigt sich eine Präferenz für Arginine und Lysine, sowie hydrophobe Aminosäuren Leucin, Methionin, Alanin und Phenylalanin. Tyrosin rückt in der Rangfolge nach hinten. Im Gegensatz zur 10-kDa- und 20-kDa-Variante werden die Termini stärker bedeckt, v. a. das C-terminal gelegene R162 wird öfter adressiert.

Im Falle von BCN-LPG-50-kDa werden Aminosäuren der flexiblen Termini und in unmittelbarer Nähe zur Konjugationsstelle im Vergleich zur 10-kDa-Variante vermehrt adressiert. Wie bei der LPG-20-kDa-Variante auch zeigt sich eine Bedeckung der Polymeroberfläche, die sehr der Verteilung der solvenszugänglichen Oberfläche ähnelt. Die Absolutwerte der maximalen Kontaktzeiten sind für diese Größenordnung der Konjugate wiederum für LPG am größten. Diese signalisieren abermals stabile Wechselwirkungen mit Aspartaten (D2, D71) und vielen Glutamaten (E107, E113, E146).

Wie auch im Falle der 20-kDa-Variante sind Bedeckungen beim BCN-PEtOx-30-kDa-Konjugat im unteren unstrukturierten Bereich zwischen Helix 1 und 2 (Q40 - A50) im Gegensatz zur 10-kDa-Variante erniedrigt. Analog zu den größten PEG- und LPG-Konjugaten zeigen sich leicht erhöhte Bedeckungen im Bereich der beiden Termini. Wechselwirkungen mit Argininen rücken in der Rangfolge der Flächendifferenzen nach hinten. Im Hinblick auf die maximalen Kontaktzeiten zeigen sich ähnliche Präferenzen wie bei der 10-kDa-Variante (Interaktionen mit Tyrosinen und Serinen). Durch Interaktionen mit Q61 rückt Glutamin an die Spitze der Rangfolge der Aminosäuren mit den positivsten Flächendifferenzen.

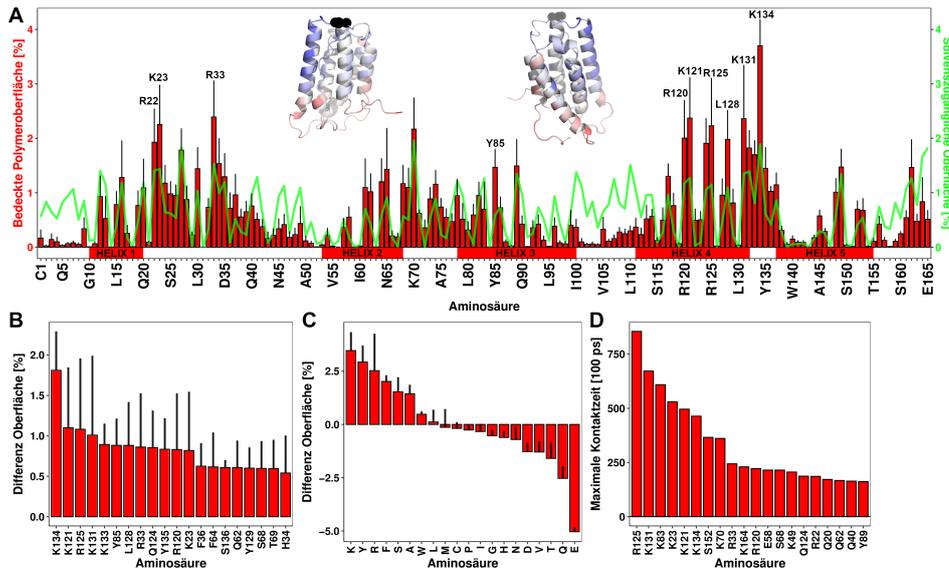


Abbildung 5.13: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des 20-kDa-BCN-PEG-Konjugats in Relation zum Anteil an der solvenszugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenszugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächen-differenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

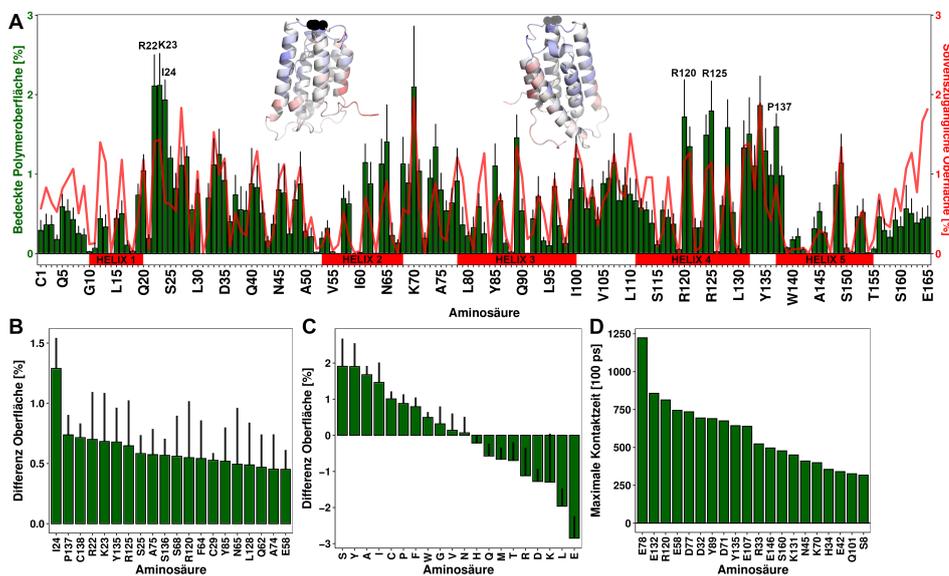


Abbildung 5.14: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des 20-kDa-BCN-LPG-Konjugats in Relation zum Anteil an der solvenszugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenszugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächen-differenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

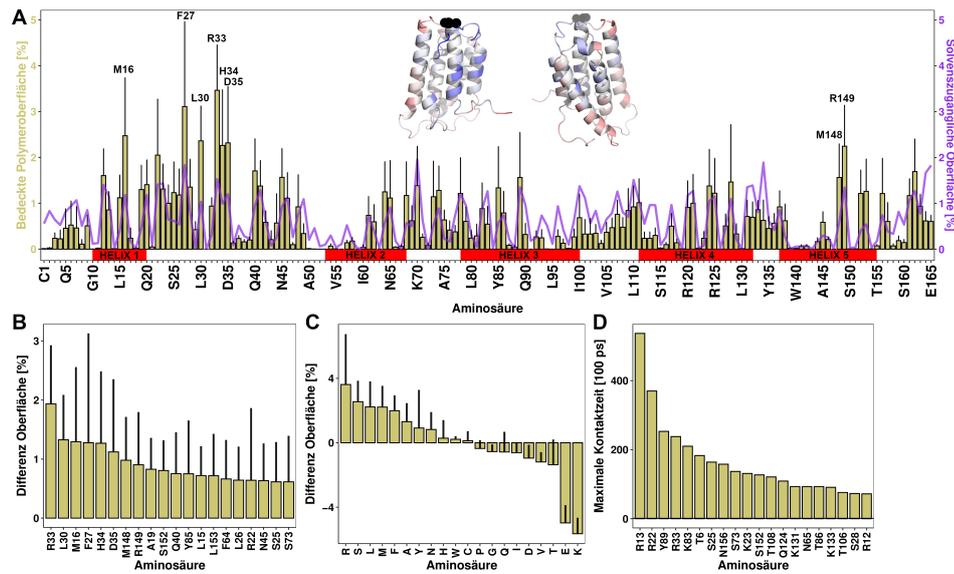


Abbildung 5.15: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des 20-kDa-BCN-PEtOx-Konjugats in Relation zum Anteil an der solvenszugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenszugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächendifferenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

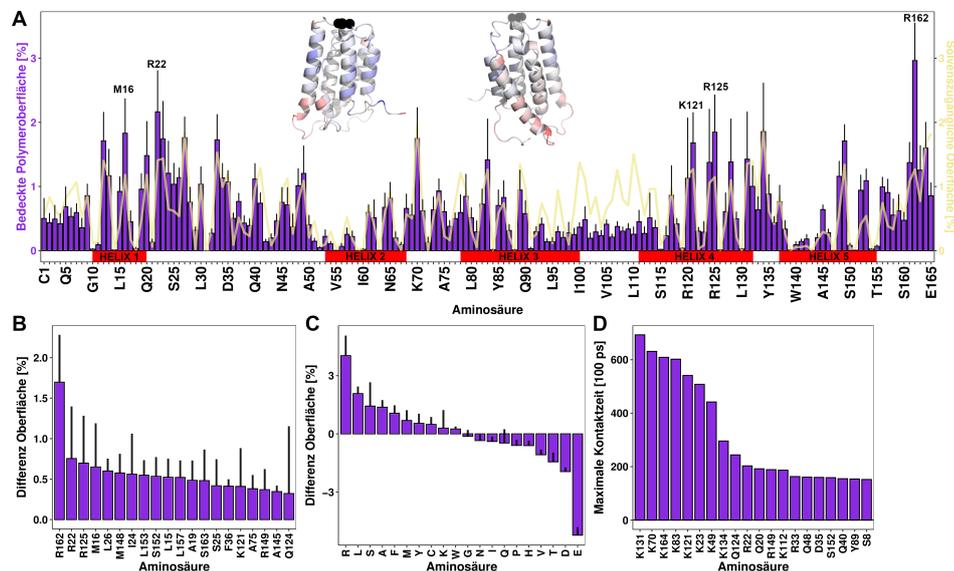


Abbildung 5.16: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des 40-kDa-BCN-PEG-Konjugats in Relation zum Anteil an der solvenszugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenszugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächendifferenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

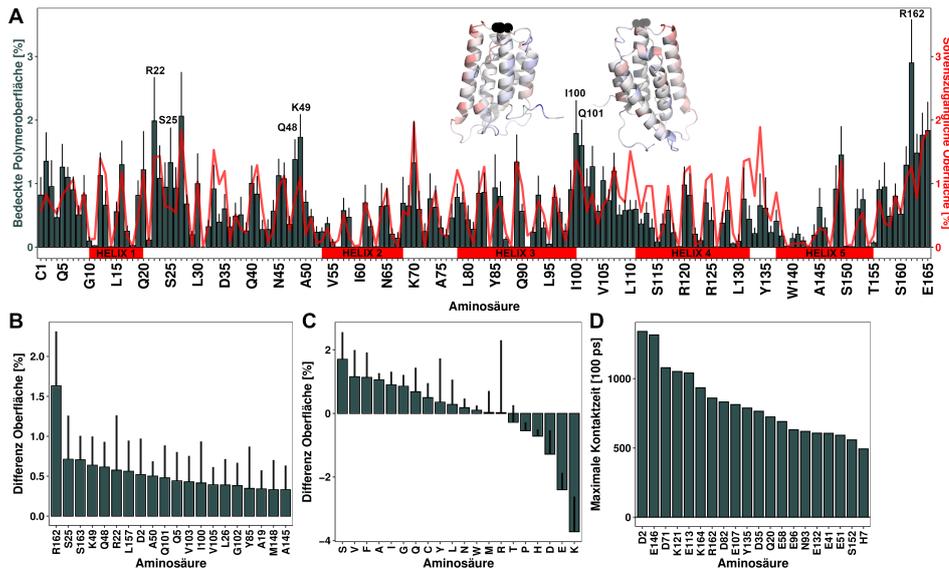


Abbildung 5.17: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des 50-kDa-BCN-LPG-Konjugats in Relation zum Anteil an der solvenszugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenszugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächendifferenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

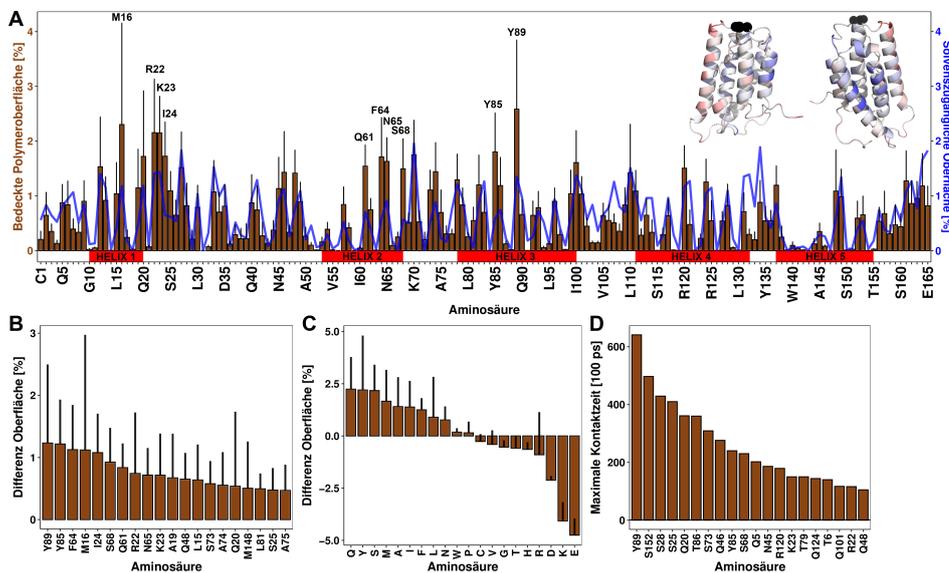


Abbildung 5.18: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des 30-kDa-BCN-PETox-Konjugats in Relation zum Anteil an der solvenszugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenszugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächendifferenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

5.1.3 Fazit

Die berechneten Kontaktprofile der 10-kDa-BCN-Varianten deuten auf eine bevorzugte Interaktion von PEG mit positiven Argininen und Lysinen hin, um die es sich mit seinen Ethersauerstoffatomen windet. Diese Aminosäuren sind v. a. im Bereich der vierten Helix vorzufinden. Ergänzend hierzu werden hydrophobe Aminosäuren (F, Y, L, A) adressiert, was die Beobachtungen aus [107] stützt. LPG mit zusätzlichen Hydroxymethylseitenketten zeigt ein erweitertes Interaktionsprofil mit geringerer Bedeckung von Lysinen. Bedeckungen von Serinen, Methioninen und Glutaminen werden vermehrt registriert. Der größte Unterschied zu PEG findet sich in den maximalen Kontaktzeiten des Polymers mit dem Protein. LPG zeigt hierbei die stabilsten Interaktionen aller drei Polymerarten und eine Präferenz zu den Aminosäuren Aspartat und Glutamat, mit denen es HB ausbilden kann. PEtOx zeigt bevorzugte Interaktionen mit den aromatischen Aminosäuren Tyrosin (v. a. Y89) und Phenylalanin. Zusätzlich können über die Carbonylgruppe der Seitenkette verschiedene HB-Donatoren, v. a. Glutamine, Asparagine und Serine, adressiert werden.

Insgesamt bewirkt der Austausch des BCN- durch den DBCO-Linker durchaus eine Veränderung des vom Polymer bedeckten Proteinareals. Interaktionen mit Aminosäuren der Helix 4 werden verringert, dagegen zeigen sich größere Wechselwirkungen mit Aminosäuren, die sich in der Helix 1 oder im unstrukturierten Bereich zwischen der ersten und zweiten Helix befinden. Die zur Interaktion bevorzugten Aminosäuretypen ändern sich jedoch nur marginal. Im Falle des DBCO-LPG-Derivats, welches das einzige DBCO-Konjugat darstellt, das eine im Vergleich zur BCN-Variante deutlich erniedrigte Schmelztemperatur aufweist (vgl. Kap. 5.3.1), zeigen sich u. a. geringere Bedeckungen von Argininen und eine erhöhte Tendenz zu Interaktionen mit Asparaginen und Serinen. Weiterhin scheinen stabile Wechselwirkungen mit D82 und K83, wie sie bei der BCN-Variante gefunden wurden, beim DBCO-Derivat nicht ausgebildet zu werden.

Im Hinblick auf die größeren Konjugate (20 - 50 kDa) zeigen die Polymere generell eine höhere Tendenz, Aminosäuren in der Nähe der Konjugationsstelle zu bedecken. Die größten Derivate (30 - 50 kDa) weisen zudem deutlich mehr Wechselwirkungen mit den beiden flexiblen Termini des Proteins auf.

5.2 Coarse-grained-Simulation

5.2.1 Vergleich mit GaMD-Simulationen

Die CG-Simulation einer IFN- α 2a-K31-PEG-10-kDa-Variante ohne Linker zeigt eine im Vergleich zu den GaMD-Ansätzen deutlich andere Polymerdynamik. Durch die längere Simulationszeit von 10 μ s mit einer sehr viel größeren Solvensbox und einer starken Reduktion an Freiheitsgraden kommt es zum Sampling vieler Konformationen mit solvensexponierten Polymerketten. Diese Verteilung ist aus den Punktwolken in Abb. 5.19 ersichtlich. Eine Polymerdichte in Analogie zu jenen aus den GaMD-Ansätzen zeigt jedoch lediglich im Bereich um R120 - R125 erhöhte Aufenthaltswahrscheinlichkeiten (Abb. 5.20). Dies entspricht dem Areal der Helix 4, welches auch in den GaMDs durch PEG am häufigsten adressiert wird.

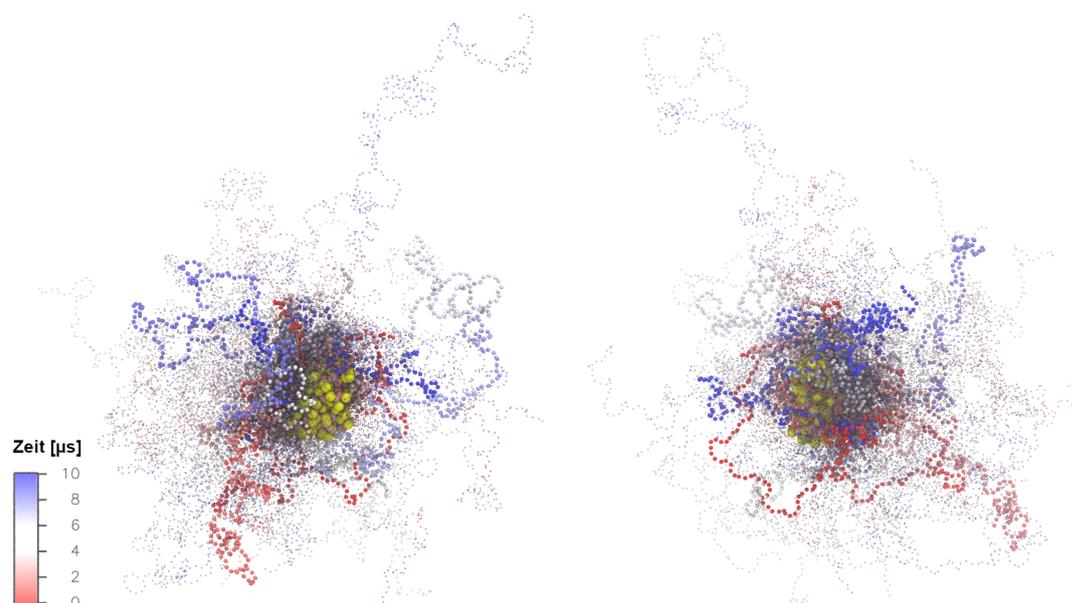


Abbildung 5.19: Darstellung von Momentaufnahmen der CG-Simulation aus zwei Blickrichtungen (Rückgratkugeln lachsfarben, Seitenkettenkugeln gelb, Konjugationsstelle schwarz). Der Polymer-Anteil des Konjugats ist für alle 100 ns als kleinere, transparente, sowie für jede Mikrosekunde zusätzlich als opake Kugeln gezeigt. Diese Kugeln sind von rot über weiß hin zu blau, basierend auf dem Zeitpunkt der Momentaufnahme, koloriert.

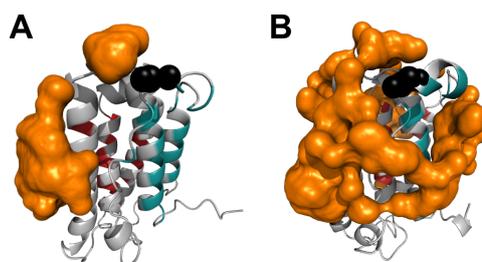


Abbildung 5.20: (A) Polymerdichte analog zu Abb. 5.1 aus der CG-MD. (B) Polymerdichte aus Abb. 5.1 zur GaMD des BCN-PEG-10-kDa-Konjugats (Nachdruck mit Genehmigung aus [304], Copyright 2021 American Chemical Society).

Ein entsprechendes Kontaktprofil wurde ebenfalls erstellt und ist in Abb. 5.21 einsehbar. Auch wenn hier wie bei der GaMD der BCN-PEG-10-kDa-Variante viele Arginine und Lysine bedeckt werden (R120, K121, R125, K133), so scheinen Bedeckungen der hydrophoben Aminosäuren Leucin und v. a. Phenylalanin deutlich ausgeprägter zu sein. Im Gegensatz zur GaMD zeigt sich eine geringe Adressierung von R33, K70 und R149, dagegen interagiert das Polymer mit F27 und F36 zwischen Helix 1 und 2. Eine deutliche Präferenz für Phenylalanine zeigt sich auch beim Blick auf die maximalen Kontaktzeiten. Die Absolutwerte dieser Zeiten sind im Vergleich zur GaMD deutlich kleiner. Hierbei wurde ein Kontakt registriert, sofern eine Polymerkugel innerhalb von 8 Å der äußersten Aminosäureseitenkettenkugel zu finden war. Ein größerer Grenzwert im Vergleich zu den GaMDs (3 Å) wurde gewählt, da die CG-Kugeln größere Radii als die Atome in den GaMDs aufweisen. Geringere Grenzwerte als 8 Å wurden getestet, führten jedoch zu kaum registrierbaren Kontakten.

Insgesamt zeigt sich, dass PEG analog zur GaMD der BCN-PEG-10-kDa-Variante Bereiche der vierten Helix adressiert. Im Unterschied zur GaMD zeigen sich jedoch schwächere Wechselwirkungen mit Argininen und Lysinen und stärkere mit Leucinen und Phenylalaninen. Diese Unterschiede resultieren vermutlich aus dem MARTINI-2-Kraftfeld, möglicherweise werden PEG-Wechselwirkungen mit polaren Proteinbereichen in diesem unterschätzt (*vide infra*). Andererseits kommt es zum Sampling proteinferner Polymerkonformationen, was bei den GaMDs nicht beobachtet wurde.

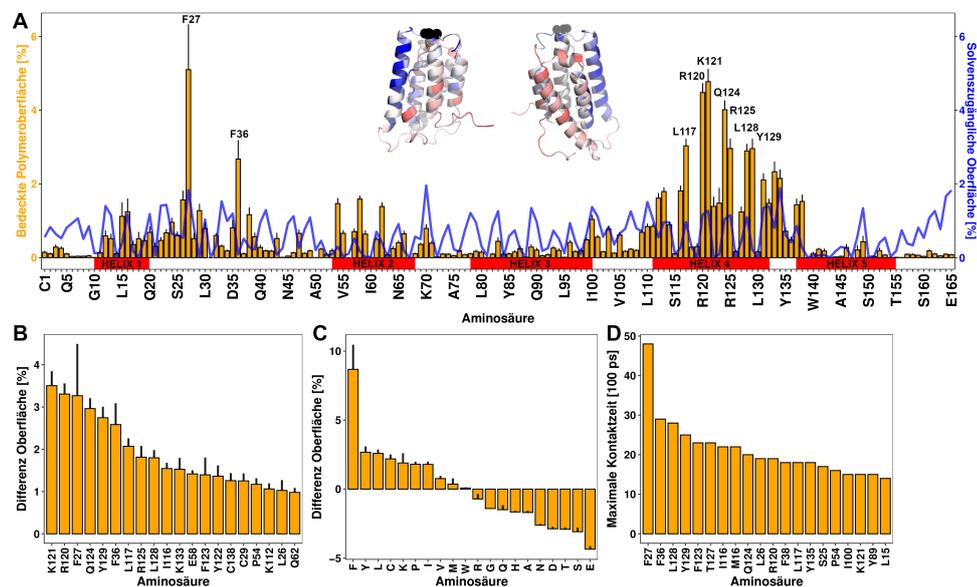


Abbildung 5.21: (A) Anteil jeder Aminosäure am bedeckten Polymerbereich des CG-10-kDa-BCN-PEG-Konjugats in Relation zum Anteil an der solvenzzugänglichen Oberfläche des Wildtyps (Berechnung der Flächen für alle 0.1 ns). Dazu ist die IFN- α 2a-Struktur aus zwei Blickrichtungen gezeigt, mit schwarzen Kugeln an der Konjugationsstelle K31. Die Färbung zeigt die Differenz der Prozentanteile an (Polymeroberfläche - solvenzzugängliche Oberfläche, von blau (+2) über weiß bis rot (-2)). (B) Top 20 Aminosäuren mit der positivsten Oberflächendifferenz aus (A). (C) Für die 20 proteinogenen Aminosäuren aufsummierte Werte aus (A). (D) Top 20 der längsten konsekutiven Kontaktzeiten mit dem Polymer. In (A), (B) und (C) sind die Standardfehler des Mittelwerts (Mittelwert + Fehler) gekennzeichnet.

5.2.2 Verwendetes Kraftfeld

Es wurden die PEG-Parameter des MARTINI-2-Kraftfelds verwendet, welche auch im PolyPly-Programm implementiert sind [313]. Diese wurden mit Berücksichtigung der Dynamik reiner PEG-Ketten, PEGylierter Lipid-Bilayer, PEG-basierter nichtionischer Tenside und Polystyren-PEG-Blockcopolymeren entwickelt. Die Fähigkeit älterer MARTINI-Parameter für PEG, Protein-Interaktionen richtig wiederzugeben, wurde in einer Publikation kürzlich infrage gestellt und in diesem Zuge ein abgewandeltes Parameterset entwickelt [329]. Unter diesen Gesichtspunkten könnte eine Implementierung dieser modifizierten MARTINI-2-Parameter oder eine Nutzung des MARTINI-3-Kraftfelds entlang eines kürzlich publizierten Protokolls [163] zu einer größeren Übereinstimmung mit den Ergebnissen aus den atombasierten Simulationen führen.

Dennoch zeigen die Ergebnisse dieser Simulation exemplarisch, dass die GaMDs lediglich proteinnahe Konformationen der Polymere abbilden, die aber unter Umständen nur einen sehr geringen Anteil des in der Wirklichkeit besuchten Phasenraums darstellen. Im Hinblick auf PEG wird hierbei zwischen einem Shroud (Hülle) und einem Dumbbell (Hantel)-Modell unterschieden (Abb. 5.22). Im Dumbbell-Modell bildet die PEG-Kette eine eigenständige Entität (*random coil*) fernab der Proteinoberfläche, während das Shroud-Modell ein Umwinden des Polymers um diese herum beschreibt. Je nach System und Polymerlänge werden beide Vorstellungen als mögliche Polymeranordnungen in Betracht gezogen [107, 330]. Entsprechende weiterführende CG-Simulationen könnten daher dabei helfen, den Anteil an Konformationen, in denen sich das Polymer vom Protein abgewandt befindet, richtig wiederzugeben und damit die Relevanz der in den GaMD-Simulationen beobachteten Interaktionen besser einstufen zu können. Eine solche Gewichtung wäre auch im Hinblick auf die in den Kap. 5.3 und 5.4 durchgeführten Berechnungen zur Bestimmung der thermischen Stabilität und Bioaktivität von Bedeutung.

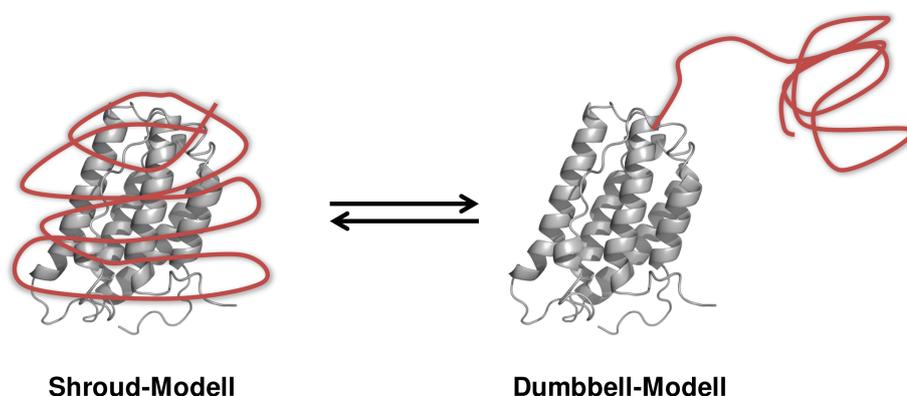


Abbildung 5.22: Schematische Darstellungen des Shroud (links) und des Dumbbell-Modells (rechts). Das Polymer kann das Protein, je nach Größe, umhüllen und mit diesem wechselwirken oder als eher unabhängige Entität im Solvens vorliegen. Die durchgeführten GaMD-Simulationen können lediglich proteinnahe Konformationen wiedergeben.

5.3 Untersuchungen zur thermischen Stabilität

In den letzten Kapiteln wurden die bevorzugten Wechselwirkungspartner der verschiedenen Polymerarten näher analysiert. Hier soll nun untersucht werden, ob die experimentell mittels DSF-Messungen (siehe Kap. 3.2.7) ermittelten Schmelztemperaturen verschiedener Konjugate mit aus den Simulationen ableitbaren Parametern in Verbindung gebracht werden können, um somit den Effekt der Polymere auf die thermische Stabilität interpretieren zu können.

5.3.1 10-kDa-Varianten

Folgende Schmelzpunkte wurden experimentell zunächst ermittelt:

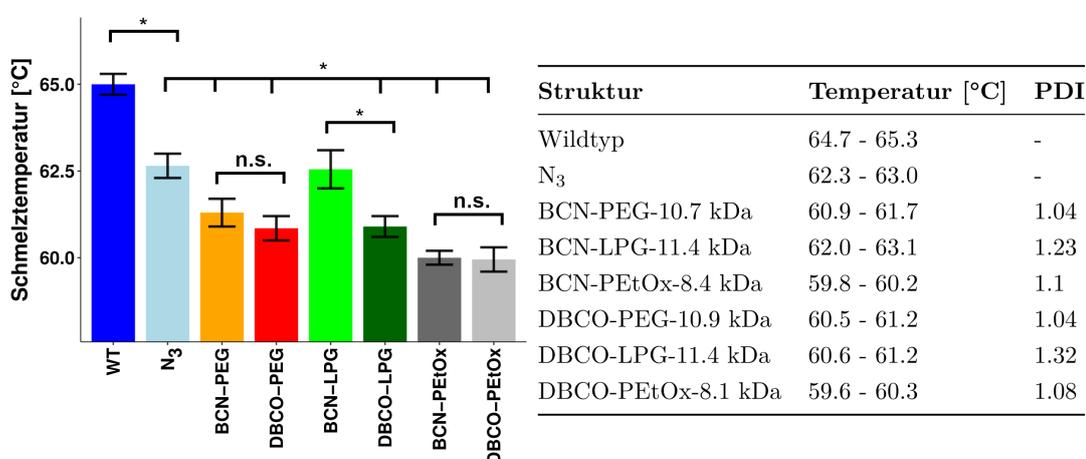


Abbildung 5.23: Per DSF ermittelte Schmelztemperaturen (95%-Konfidenzintervalle; $n = 5$; signifikante Unterschiede mit p -Werten von <0.05 durch Sterne angedeutet). Gelistete Massen basieren auf MALDI-TOF (Matrix-assistierte Laser-Desorption-Ionisierung-Time-of-flight)-Messungen. Die Polydispersität (PDI) wurde über eine Größenausschlusschromatographie ermittelt (Reproduzierung mit Genehmigung aus [304], Copyright 2021 American Chemical Society).

Es handelt sich um einen äußerst engen Temperaturbereich. Der Wildtyp weist eine gegenüber allen anderen aufgelisteten Strukturen signifikant höhere thermische Stabilität auf. Alle Konjugate weisen dementsprechend eine Destabilisierung auf, wobei diese im Falle des BCN-LPG-Derivats am geringsten ausfällt. Dieses Konjugat ist das einzige, welches sich signifikant von allen anderen unterscheidet und eine vergleichbare Schmelztemperatur zur unkonjugierten Struktur mit Azidgruppe aufweist (hier mit N₃ gekennzeichnet). Es handelt sich auch um den einzigen Fall, in dem der Austausch des BCN- mit einem DBCO-Linker einen deutlichen Effekt auf die Schmelztemperatur hat. Die PEG-Derivate zeigen im Vergleich zu den LPG-Konjugaten etwas niedrigere Schmelztemperaturen, die PEtOx-Konjugate sind die instabilsten PBK.

Zunächst wurden aus den entsprechenden GaMD-Simulationen der BCN-10-kDa-Varianten Parameter ermittelt, die Protein-Polymer-Wechselwirkungen beschreiben, z. B. die Anzahl an Protein-Polymer-HB, die durch das jeweilige Polymer bedeckte

Proteinoberfläche oder die Interaktionsenergie mit allen Aminosäuren des Proteins. Diese Untersuchungen zeigten jedoch zunächst keine Korrelation mit den experimentellen Befunden. Es wurden schließlich drei Parameter gefunden, die alle in gewisser Weise die Proteinflexibilität beschreiben: (1) die durch eine CNA vorhergesagte Schmelztemperatur, (2) die Anzahl an nativen Kontakten, (3) sowie das Verhältnis der RMSF-Werte zum Wildtyp. Die nachträglich simulierten DBCO-Varianten wurden mithilfe dieser Parameter korrekt bzgl. ihrer thermischen Stabilität eingestuft. Entsprechende Korrelationen sind in den Abb. 5.24 gezeigt und sollen im Folgenden näher beschrieben werden.

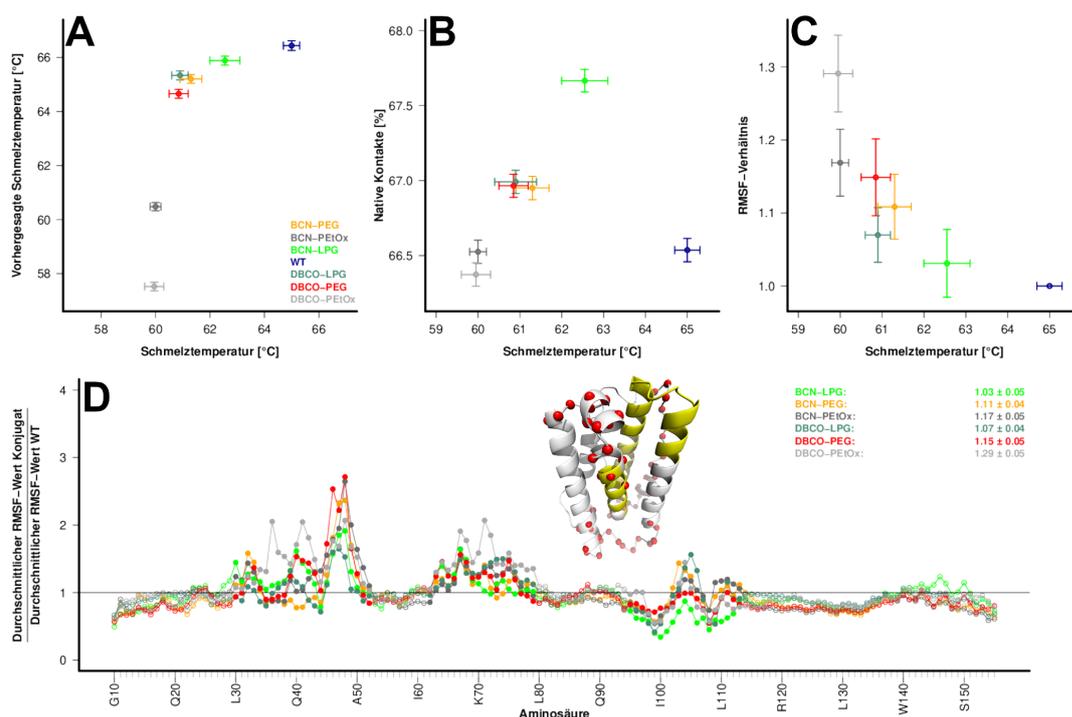


Abbildung 5.24: (A) Korrelation zwischen den CNA-basierten und den experimentell bestimmten Schmelztemperaturen. Farben entsprechen den Dichtedarstellungen der Abb. 5.1: WT = blau, BCN-LPG = grün, DBCO-LPG = dunkelgrün, BCN-PEG = orange, DBCO-PEG = rot, BCN-PEtOx = dunkelgrau, DBCO-PEtOx = grau. (B) Korrelation zwischen den experimentell bestimmten Schmelztemperaturen und der Anzahl nativer Kontakte. (C) Korrelation zwischen den experimentell bestimmten Schmelztemperaturen und den RMSF-Verhältnissen, die in (D) oben rechts aufgelistet sind. (D) Aminosäurespezifische RMSF-Verhältnisse zwischen den Biokonjugaten und dem Wildtyp (Verhältnis der Durchschnittswerte aller Replika). Dargestellt sind die Aminosäuren G10 bis T155 ohne die flexiblen Termini. Die größten Unterschiede sind für die Aminosäuren der Schleifen und einige Bereiche der Helices 2 und 3 zu erkennen (L30 - T52, Q62 - T79 und L95 - E113). Diese sind als rote Kugeln in der dargestellten Struktur markiert. Letztere ist gelb in Arealen mit großen Unterschieden bzgl. der in Abb. 5.27 weiter unten dargestellten Perkolationsindizes gefärbt (Reproduzierung bzw. Nachdruck mit Genehmigung aus [304], Copyright 2021 American Chemical Society).

5.3.1.1 CN-Analysen

Zur Durchführung der CN-Analysen der in Kap. 5 behandelten IFN- α 2a-Konjugate mit dem Programm CNA 2.0 wurden für alle 0.1 ns PDB-Strukturen der Simulationen gespeichert. Dieses Analysenprogramm wurde bisher nur für Proteine oder Systeme

mit kleineren, nichtkovalenten Liganden verwendet. Die Monomere wurden entsprechend als Ligand-Heteroatome definiert, während der BCN- bzw. DBCO-Linker als Teil der Aminosäure K31 beschrieben wurde. Somit konnte der Einfluss nichtkovalenter Wechselwirkungen des Polymers mit dem Protein auf die Rigidität des Proteins untersucht werden.

Ein Energiebereich zwischen -10 und 0 kcal/mol mit Schritten von 0.05 kcal/mol wurde analysiert ($E = -10 \text{ } 0 \text{ } 0.05$). Zur Beurteilung hydrophober Effekte wurde der Standardgrenzwert von 0.25 Å ($c = 0.25$) verwendet. Befanden sich zwei Kohlenstoff-, ein Kohlenstoff- und ein Schwefel-, oder zwei Schwefelatome innerhalb dieses Grenzwerts (inklusive der jeweiligen VdW-Radien) wurde dies als hydrophober Kontakt eingestuft ($H = 1$). Deren Stärke wurde mit zunehmenden Temperaturen bei der Entfaltungssimulation des Programms nicht moduliert ($TUS = 1$). Diese Einstellungen in Kombination mit einem auf der Konfigurationsentropie vom Typ 2 basierenden Spline-Fit ergab die beste Korrelation (Abb. 5.24-A).

Der Wildtyp wurde korrekt als stabilste Variante eingestuft, gefolgt vom BCN-LPG-Konjugat. Auch die übrige Rangfolge stimmt mit den experimentellen Befunden überein, wobei beide PEOx-Varianten als deutlich destabilisiert beurteilt wurden. Die vom Programm ausgegebene Schmelztemperatur basiert auf einer Umrechnung der jeweiligen Energie-Grenzwerte, bei denen eine Entfaltung anzunehmen ist. Im Falle der Konfigurationsentropie vom Typ 2 ist dies an jenem Grenzwert gegeben, an dem sich der größte Rigiditätscluster des Netzwerks anfängt aufzulösen und das System nicht mehr dominiert. Die Konfigurationsentropie H vom Typ 2 basiert auf der Informationstheorie nach Shannon und beschreibt den Grad der Unordnung in Bezug zur Realisation eines bestimmten Zustands:¹

$$H = - \sum_s w_s \ln w_s \quad (5.1)$$

$$w_s = \frac{s^2 n_s}{\sum_s s^2 n_s} \quad (5.2)$$

$$n_s = \frac{\text{Anzahl Cluster der Größe } s}{\text{Anzahl Atome } N} \quad (5.3)$$

Die finale Entropie wird hierbei aus der Summe der Werte für alle im System gefundenen Clustergrößen s gebildet. Dabei beschreibt w_s die Wahrscheinlichkeit eines Atoms, sich in einem Cluster der Größe s zu befinden. Diese Wahrscheinlichkeit resultiert wiederum aus einer Verrechnung mit der normierten Clustergröße n_s , bei der die Anzahl der Cluster mit der Größe s in Relation zur Anzahl an Atomen N im gesamten Netzwerk gesetzt wird. Zu Beginn besteht lediglich ein rigider Cluster im System, die Entropie besitzt daher den Wert Null. Fällt der größte Cluster in sich zusammen, so kommt es zum Anstieg möglicher Systemzustände (verschiedene Clustergrößen s) und damit der Entropie H . Verändert sich die Entropie also an einem

¹Die eher unübliche Nutzung des Kennzeichens H für eine Entropie richtet sich hierbei nach den publizierten Arbeiten zu CNA-basierten Deskriptoren [234, 235].

gegebenen Energie-Grenzwert sehr stark, so wird ein Phasenübergang vermutet. Die ausgegebenen absoluten Temperaturwerte sollten allerdings lediglich relativ zwischen vergleichbaren Systemen beurteilt werden [231, 234, 235, 331].

5.3.1.1.1 Generierte Netzwerke

Die Stabilität der PEG- und LPG-Varianten scheint bei der thermalen Entfaltungssimulation durch das CNA-Programm überbewertet zu werden. Dies könnte daraus resultieren, dass diese Strukturelemente Ethersauerstoffatome besitzen, deren Fähigkeit zur Ausbildung von HB vom Programm überschätzt wird. Eine Visualisierung exemplarischer Ensemble-Netzwerke per visualCNA zeigt, dass LPG besonders viele HB ausbildet, PEtOx das Protein dagegen hauptsächlich über hydrophobe Constraints beeinflusst (Abb. 5.25). PEG zeigt deutlich weniger HB als LPG. Die Ethersauerstoffatome scheinen bei beiden Varianten jedoch maßgeblich zur Rigidität beizutragen und viele HB-Constraints, v. a. über Lysine und Arginine, ausbilden zu können (Abb. 5.26).

5.3.1.1.2 Perkulationsindizes

Neben globalen Rigiditätsindizes können mithilfe des CNA-Programms auch lokale Effekte über verschiedene Parameter analysiert werden (siehe Kap. 1.4). Eine nähere Inspektion der lokalen Indizes zeigt auf, dass die destabilisierten PEtOx-Konjugate besonders im Falle der Perkulationsindizes vom Typ 2 deutlich positivere Energiegrenzwerte in bestimmten Bereichen aufweisen (um V55 - N65 und R125 - T155, entspricht Bereichen der Helices 2, 4 und 5), verglichen mit allen übrigen Strukturen (Abb. 5.27-A). Diese Indizes geben die Grenzwerte an, bei denen sich die Aminosäuren vom aktuell größten rigiden Cluster loslösen. Geringe Absolutwerte deuten auf eine während der Entfaltung frühere Flexibilisierung der entsprechenden Bereiche hin. Beim Blick auf die instabilen Areale in Kombination mit den gefundenen Polymerdichten (Abb. 5.27-B und 5.27-C) zeigt sich, dass diese Aminosäuren durch PEG und LPG bedeckt sind und damit rigidisierende Wechselwirkungen ermöglicht werden, während sie im Falle der PEtOx-Konjugate zu einem stärkeren Grad solvens-exponiert bleiben.

5.3.1.2 Native Kontakte

Die Anzahl nativer Kontakte wird oft zur Untersuchung von Entfaltungssimulationen genutzt [332]. In diesem Falle wurden alle Aminosäurekontakte (für die individuellen Atome berechnet), basierend auf der jeweiligen NMR-Struktur mit dem standardmäßigen Grenzwert von 4.5 Å, berücksichtigt. Ausgenommen wurden die Kontakte von K31 und der sehr flexiblen Termini C1 - L9 und N156 - E165. Daraus wurde der durchschnittliche Anteil an Momentaufnahmen berechnet, in denen diese Kontakte in

den Simulationen registriert wurden. Trotz der äußerst geringen prozentualen Unterschiede kann eine Korrelation im Hinblick auf die untersuchten Konjugate beobachtet werden (Abb. 5.24-B).

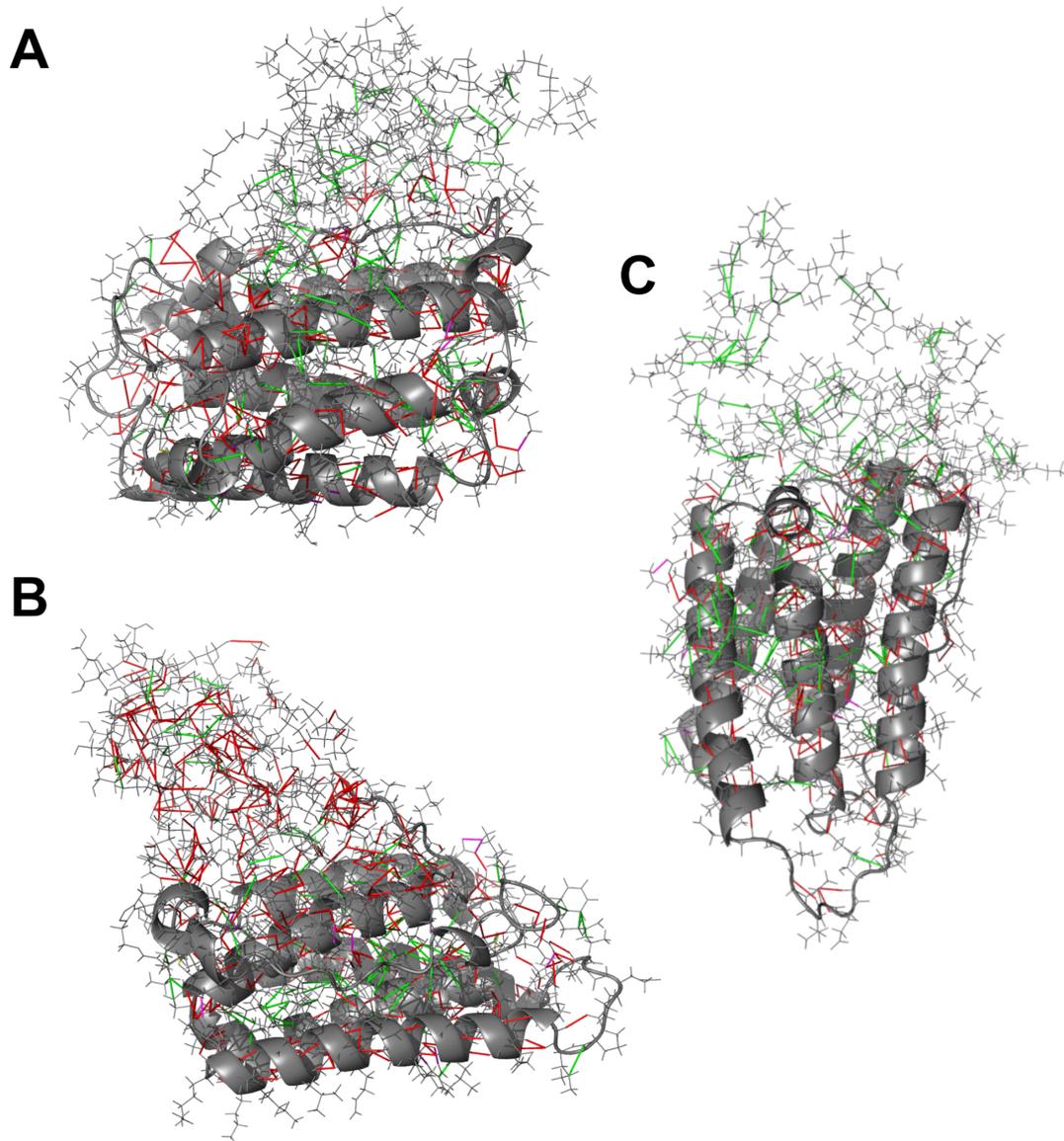


Abbildung 5.25: Visualisierung exemplarischer CNA-Netzwerke der 10-kDa-BCN-Konjugate von (A) PEG, (B) LPG und (C) PEtOx mittels visualCNA. Rote Stäbe illustrieren HB, violette Salzbrücken und grüne hydrophobe Wechselwirkungen.

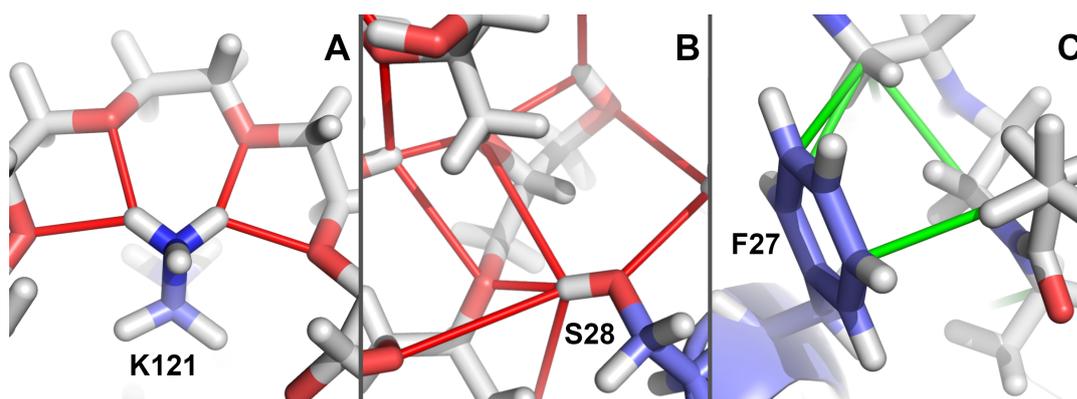


Abbildung 5.26: Visualisierung exemplarischer CNA-Interaktionen der 10-kDa-BCN-Konjugate. Die betreffenden Aminosäuren sind violett gefärbt, die Polymerkohlenstoffatome weiß. Constraints sind analog zu Abb. 5.25 illustriert. **(A)** PEG bildet viele HB-basierte Constraints über seine Ethersauerstoffatome aus, z. B. mit Lysinen. **(B)** LPG bildet ebenso viele HB-Constraints über sein Rückgrat aus, aber auch über die zusätzlichen Hydroxymethylseitenketten. **(C)** PETox bildet überwiegend hydrophobe Constraints mit dem Protein aus.

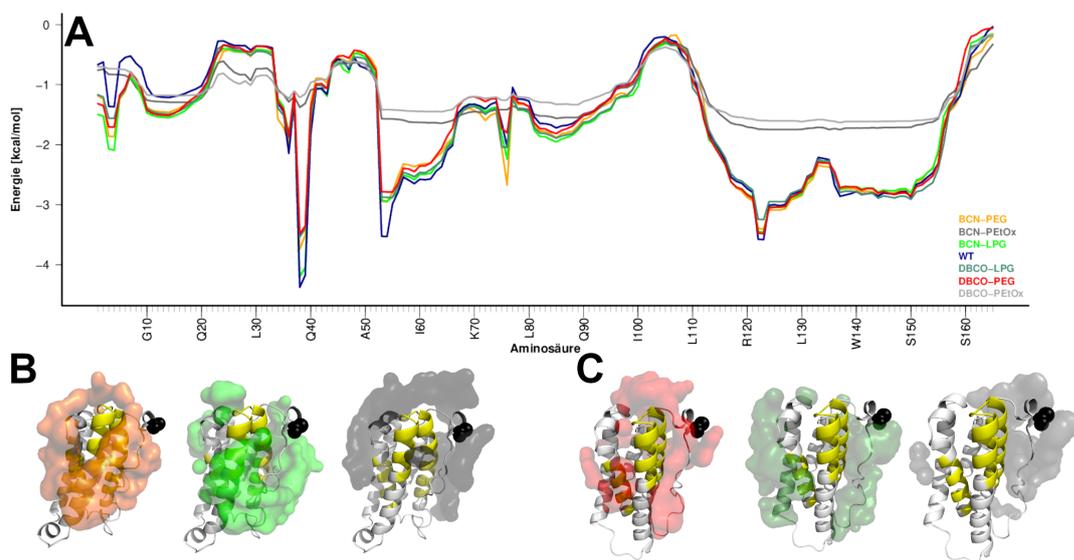


Abbildung 5.27: **(A)** Energie-Grenzwerte für die aminosäurespezifischen Perkolationsindizes vom Typ 2, mit großen Differenzen für die PETox-Konjugate in den Bereichen um V55 - N65 und R125 - T155. **(B)** Darstellung der Polymerdichten der BCN-Konjugate aus Abb. 5.1, mit den relevanten Bereichen aus (A) in gelb markiert. **(C)** Darstellung der Polymerdichten der DBCO-Konjugate aus Abb. 5.1, mit den relevanten Bereichen aus (A) in gelb markiert (Reproduzierung bzw. Nachdruck mit Genehmigung aus [304], Copyright 2021 American Chemical Society).

Korrekterweise wird bei dieser Kontakte-Analyse auch lediglich im Falle der LPG-Derivate ein signifikanter Unterschied zwischen der BCN- und der DBCO-Variante erkannt. Es ist denkbar, dass der Anteil durch die Polymere auf verschiedene Weisen beeinflusst werden kann: Durch Protein-Polymer-Interaktionen können üblicherweise im Protein vorliegende Wechselwirkungen ersetzt und damit die Prozentzahlen verringert werden. Dagegen kann ein Aufliegen des Polymers auf der Proteinoberfläche auch zum Erhalt bzw. zur Stärkung solcher intramolekularen Interaktionen beitragen

(vgl. [102]), was besonders im Falle der sehr flexiblen Schleifenbereiche vorstellbar ist, die sich ohne Gegenwart der Polymere im Solvens leichter voneinander abwenden können. Daher ist es durchaus auch nachvollziehbar, dass der Wildtyp eine niedrigere Anzahl an Kontakten aufweisen kann.

Bei dieser Analyse kann ebenso eine relative „Kontaktstärke“ für jedes Atom ausgegeben werden, wobei der stärkste native Kontakt (ein Kontakt, der durchweg erhalten bleibt) den Wert 100 erhält (somit handelt es sich um die Prozentzahl an Momentaufnahmen, in denen der Kontakt gefunden wird). Für den exemplarischen Vergleich zwischen der BCN-LPG- und der DBCO-LPG-Variante wurde die Differenz dieser „Kontaktstärken“ ermittelt und visualisiert (Abb. 5.28-A). Es zeigt sich, dass Unterschiede in der Anzahl der nativen Kontakte vermehrt aus Bereichen der unteren unstrukturierten Bereiche und der Termini resultieren und nicht ausschließlich am Ort der Konjugationsstelle zu verorten sind. Um letztere zeigt sich lediglich eine größere Kontaktstärke für das benachbarte I24 im Falle des DBCO-Konjugats, erhöhte Kontaktstärken der BCN-Variante sind v. a. im Bereich um R125 zu finden. Der sterisch anspruchsvollere DBCO-Linker zeigt eine im Vergleich zum BCN-Linker größere Anzahl an Atomen, die sich weiter entfernt von der Proteinoberfläche befinden (Abb. 5.28-B), was die Ausrichtung der Polymerkette beeinflussen kann. Insgesamt existieren sowohl Bereiche, in denen ein Erhalt der nativen Kontakte im Falle der BCN-Variante wahrscheinlicher ist, als auch Areale, in denen das DBCO-Konjugat größere „Kontaktstärken“ zeigt. Somit lässt sich kein einzelner Proteinbereich bestimmen, der für den Unterschied beider Konjugate hinsichtlich ihrer nativen Kontakte als entscheidend eingestuft werden kann.

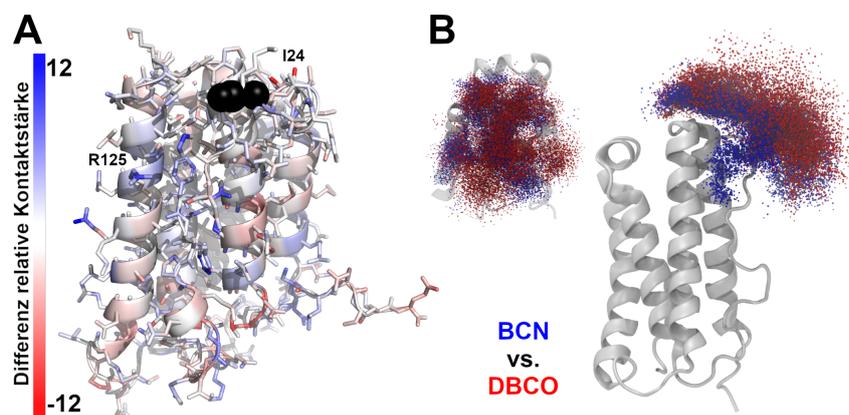


Abbildung 5.28: (A) Atombasierte Differenzen der relativen „Kontaktstärken“ zwischen der BCN-LPG- und der DBCO-LPG-Variante ($\text{Kontaktstärke}_{\text{BCN}} - \text{Kontaktstärke}_{\text{DBCO}}$). Rote Areale zeigen Bereiche, in denen in den DBCO-Simulationen native Kontakte öfter erhalten blieben, blaue repräsentieren jene Atome, die ihre nativen Kontakte in den BCN-Simulationen länger erhalten konnten (die Konjugationsstelle ist mit schwarzen Kugeln gekennzeichnet). (B) Verteilung der BCN- (blau) und DBCO-Atome (rot) in den Simulationen (Illustration aller 5 ns als Punkte).

5.3.1.3 RMSF-Verhältnis

RMSF-Werte der Rückgratschweratome aller Aminosäuren wurden berechnet. Dabei wurde die Fluktuation um die durchschnittliche Struktur der jeweiligen Simulation berechnet und zum Erhalt von Pseudo-B-Faktoren mit $\frac{8\pi^2}{3}$ multipliziert. Der Durchschnittswert aller Replika wurde jeweils mit jenem der Wildtyp-Simulationen in Relation gesetzt (Abb. 5.24-C und 5.24-D). Größere Einflüsse der Polymere sind v. a. im Bereich der flexiblen Schleifen und einiger Aminosäuren der Helices 2 und 3 zu beobachten (L30 - T52, Q62 - T79 und L95 - E113). Die Änderungen in diesen Bereichen korrelieren mit den experimentell bestimmten Schmelztemperaturen, wobei eine Destabilisierung durch eine größere Fluktuation dieser Bereiche gekennzeichnet ist.

5.3.1.4 Vergleich mit LiP-Assays

Wie in Kap. 3.2.7 erwähnt wurden sogenannte LiP-Assays durchgeführt, bei denen die Konjugate, die einer Proteolyse durch die Proteinase K unterzogen wurden, nach bestimmten Zeitintervallen massenspektrometrisch charakterisiert wurden. Die detektierten Peptidfragmente gaben Rückschlüsse auf die Schnittstellen durch die Proteinase und damit auf mögliche Schutzmechanismen der konjugierten Polymere. Die Ergebnisse zu den 10-kDa-BCN-Varianten sind in Abb. 5.29 grafisch zusammengefasst. Es lassen sich keine bedeutenden Unterschiede zwischen den Schnittmustern der unterschiedlichen Polymervarianten erkennen, auch nicht im Vergleich zum Wildtyp. Dies deutet darauf hin, dass die Polymere lediglich schwache Interaktionen zum Protein ausbilden und somit proteinabgewandte Konformationen (Dumbbell-Modell, siehe Kap. 5.2) dominieren.

Die Resultate sollten jedoch nicht direkt mit den Ergebnissen aus Kap. 5.1 in Verbindung gebracht werden. Während hierbei die Proteolysen zu bestimmten Zeitpunkten im Minutenbereich untersucht wurden, konnten durch die Simulationen lediglich jeweils $1.5 \mu\text{s}$ jedes Biokonjugats simuliert werden. Dadurch konnten ausschließlich proteinnaher Konformationen wiedergegeben werden. Die Beobachtungen unterstreichen jedoch die in Kap. 5.2 bereits erwähnte Notwendigkeit eines alternativen bzw. ergänzenden Multiscale-Modeling-Simulationsansatzes, der die proteinfernen Polymerkonformationen in geeigneter Weise berücksichtigen kann. Weiterhin ist zu erwähnen, dass der Proteinverdau natürlich ebenso von den gewählten Zeitpunkten und v. a. der verwendeten Proteinase abhängt. Eine Analyse mit verschiedenen Proteasen, welche mehr Schnittstellen abdecken, könnte ein differenzierteres Bild liefern.

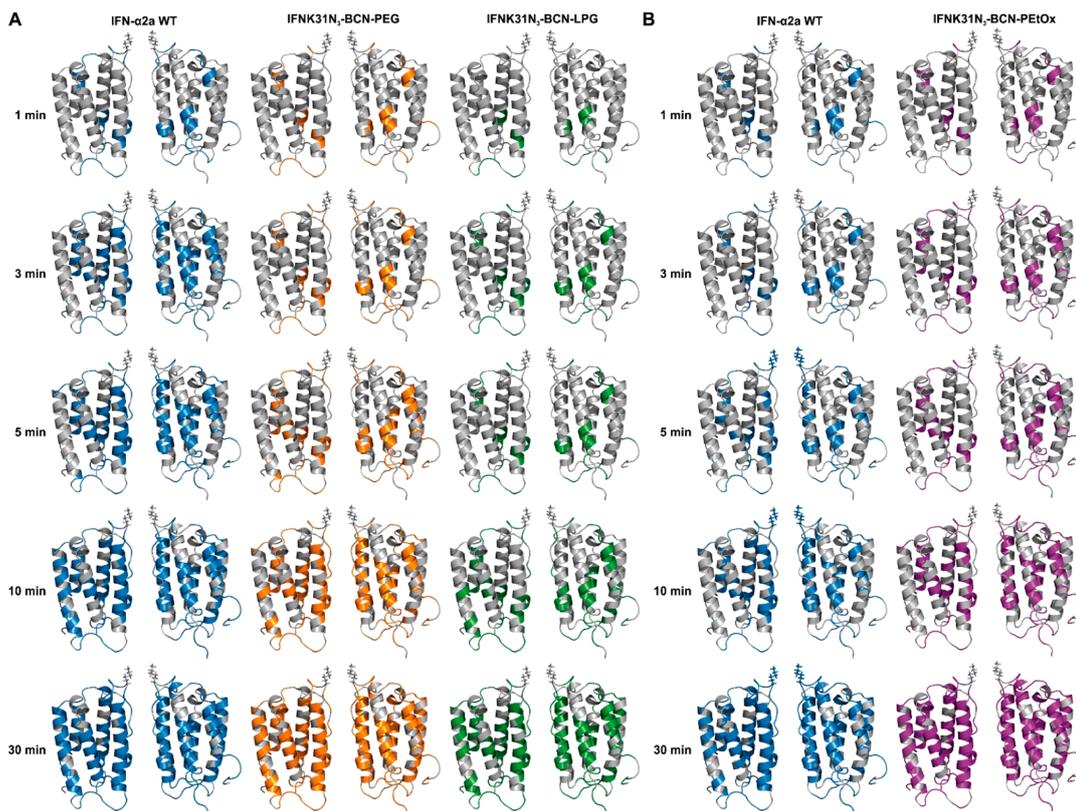


Abbildung 5.29: Ergebnisse der limitierten Proteolyse der 10-kDa-BCN-Konjugate und des Wildtyps mit Proteinase K. Die Aminosäuren an Position P1 und P1' des detektierten Verdaus sind zu gegebenen Zeitpunkten farblich markiert. Dabei wurden in (A) die Fragmente des Wildtyps (blau), sowie der BCN-PEG- (orange) und BCN-LPG-Konjugate (grün) zur Detektion mit Ethylacetat extrahiert; in (B) wurden jene des Wildtyps (blau) und des BCN-PEtOx-Konjugats (violett) mit Chloroform extrahiert (Nachdruck mit Genehmigung aus [304], Copyright 2021 American Chemical Society).

5.3.2 Größere Konjugate

Von den Verbundpartnern wurden für die größeren Varianten (20 - 50 kDa) zusätzliche Schmelztemperaturmessungen vorgenommen (Abb. 5.30). Die für die 10-kDa-Konjugate beschriebenen Parameter wurden analog in den Simulationen dieser Strukturen (siehe Kap. 5.1.2.4) untersucht. Price et al. konnten bei einer Vergrößerung einer konjugierten PEG-Kette an der WW-Domäne über vier Monomere hinaus hin zu einer 2 kDa-PEG-Kette (entspricht ca. 33 % Massenanteil am Konjugat, ähnlich zu therapeutisch verwendeten Strukturen) keine bedeutenden Unterschiede in der thermischen Stabilität registrieren [100]. Auch bei den hier vorliegenden IFN- α 2a-PEG-Konjugaten ist keine Änderung der Schmelztemperaturen mit ansteigender Polymermasse (10, 20, 40 kDa) erkennbar.

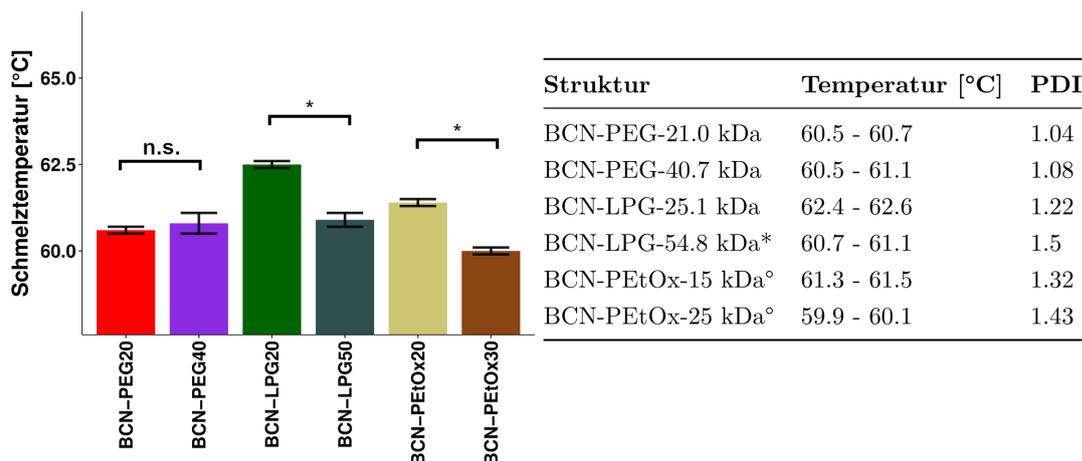


Abbildung 5.30: Per DSF ermittelte Schmelztemperaturbereiche (Mittelwert \pm Standardabweichung; $n \geq 4$; signifikante Unterschiede mit p -Werten von <0.05 durch Sterne angedeutet). Die gelisteten molaren Massen basieren auf MALDI-TOF-Messungen, dynamischer Lichtstreuung ($^{\circ}$) oder einer Größenausschlusschromatographie (*). Die Polydispersität (PDI) wurde über eine Größenausschlusschromatographie ermittelt (PEG/LPG-Daten publiziert unter Creative Commons Attribution License [333], PEtOx-Daten eingereicht [334]).

Die Ergebnisse analoger Parameterberechnungen zu Abb. 5.24 und 5.27 sind in Abb. 5.31 zusammengefasst. Im Falle der größeren PEtOx-Konjugate wird abermals eine Destabilisierung mit relativ positiven Perkolationsindizes in den Bereichen V55 - N65 und R125 - T155 durch das CNA-Programm vorhergesagt. Hier kann jedoch für keinen der oben beschriebenen Parameter eine überzeugende Korrelation mit der Schmelztemperatur festgestellt werden, besonders das größte LPG-Konjugat zeigt bei den RMSF-Verhältnissen große Abweichungen zu den übrigen Varianten.

Da diese größeren Konjugate genauso lange wie die 10-kDa-Varianten simuliert wurden, jedoch noch weitaus mehr Freiheitsgrade besitzen und somit deutliche Konformationsänderungen der Polymere vermutlich auch nur in weitaus größeren Solvensboxen realisierbar sind, ist die Annahme eines unzureichenden Samplings als primäre Ursache für diese Diskrepanz gerechtfertigt. Zu diesem Aspekt kommt hinzu, dass sich die Molekulargewichte der simulierten Ketten im Falle der größeren Konjugate deutlicher von den experimentell untersuchten Strukturen unterscheiden, als es bei den 10-kDa-Varianten der Fall ist (die modellierten Molekulargewichte stellen die anfangs experimentell geplanten Zielgrößen dar). So macht es vermutlich einen Unterschied, ob ein IFN- α 2a-K31-BCN-PEtOx_{25 kDa}- oder ein IFN- α 2a-K31-BCN-PEtOx_{30 kDa}-Derivat modelliert wird. Generell stellt die gegebene Polydispersität dieser Konjugate, welche besonders im Falle der LPG-Derivate relativ stark vom Wert 1 abweicht, eine Varietät an tatsächlich vorliegenden Kettenlängen dar, welche durch die hier vorgenommenen Modellierungen nicht wiedergegeben werden kann.

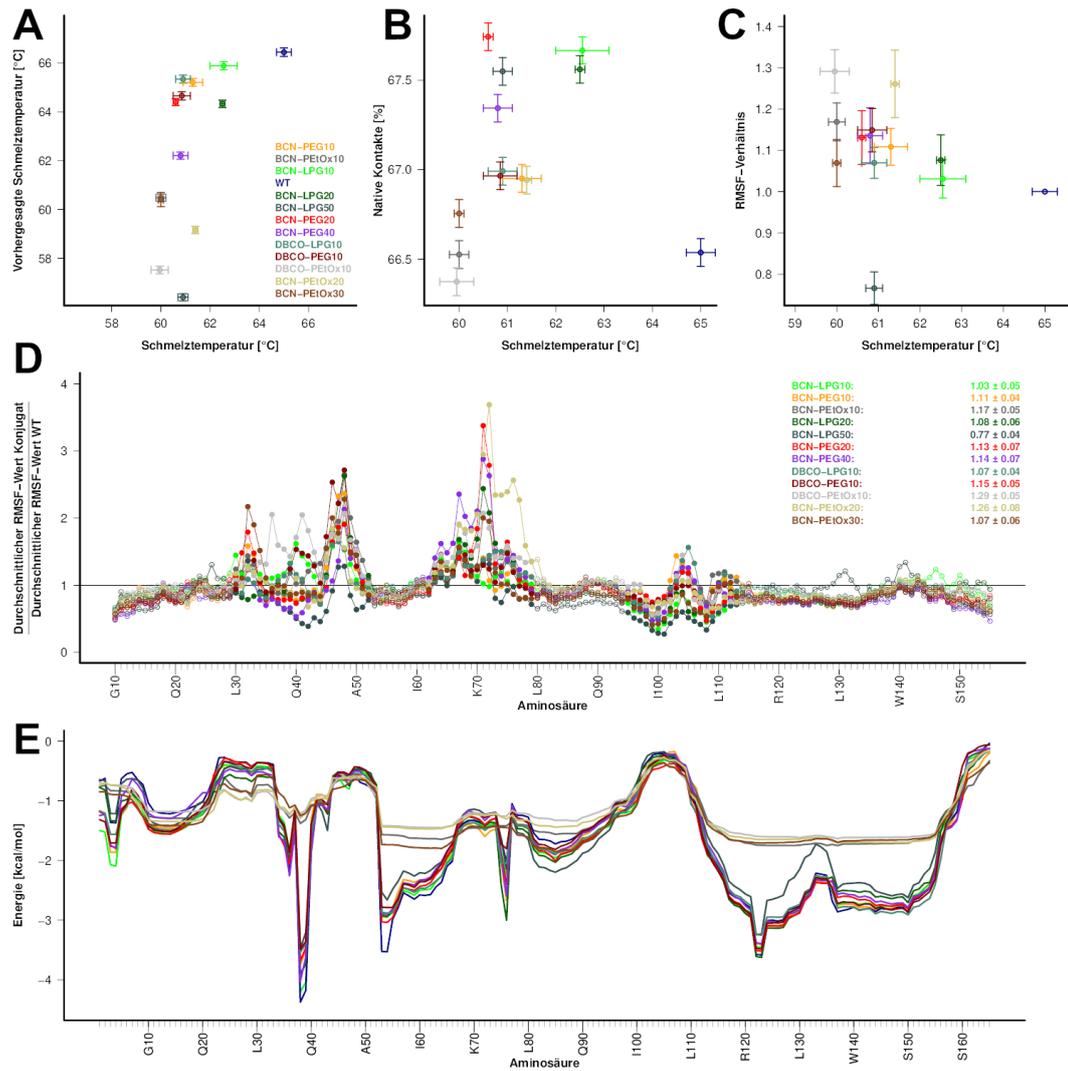


Abbildung 5.31: Analoge Darstellungen der Berechnungen aus Abb. 5.24 und 5.27 mit den zusätzlich experimentell bestimmten Schmelztemperaturen aller IFN- α 2a-BCN-Konjugate aus Abb. 5.23 und 5.30.

5.4 Untersuchungen zur Bioaktivität

Zusätzlich zu DSF-Messungen wurden die Bioaktivitäten der BCN-Konjugate experimentell über SEAP-Assays (siehe Kap. 3.2.7) bestimmt (Abb. 5.32). Diese Aktivitäten korrelierten mit den per dynamischer Lichtstreuung (DLS) ermittelten hydrodynamischen Durchmessern der Konjugate (5.33-A). Ein größerer hydrodynamischer Durchmesser kann die Bindung des Proteins an IFNAR1 und IFNAR2 beeinträchtigen, da die Polymere somit eine Annäherung des Proteins an die Rezeptoren unterbinden. Dabei zeigt das PEtOx-20-kDa-Konjugat experimentell einen Durchmesser, der eher jenem der 10-kDa-Konjugaten ähnelt, und die entsprechende 30-kDa-Variante einen Durchmesser, der jenem der 20-kDa-PEG/LPG-Konjugaten ähnelt.

Da dieser Durchmesser mit einem bestimmten Faktor mit dem jeweiligen Gyrationdurchmesser korrelieren sollte, wurde dieser für alle BCN-Konjugate *in silico* ermittelt und mit den experimentellen Daten in Verbindung gesetzt (Abb. 5.33-B und 5.33-C). Analog zu den hydrodynamischen Durchmessern zeigt sich eine Korrelation, sodass der Gyrationdurchmesser zur Abschätzung der Bioaktivität genutzt werden kann.

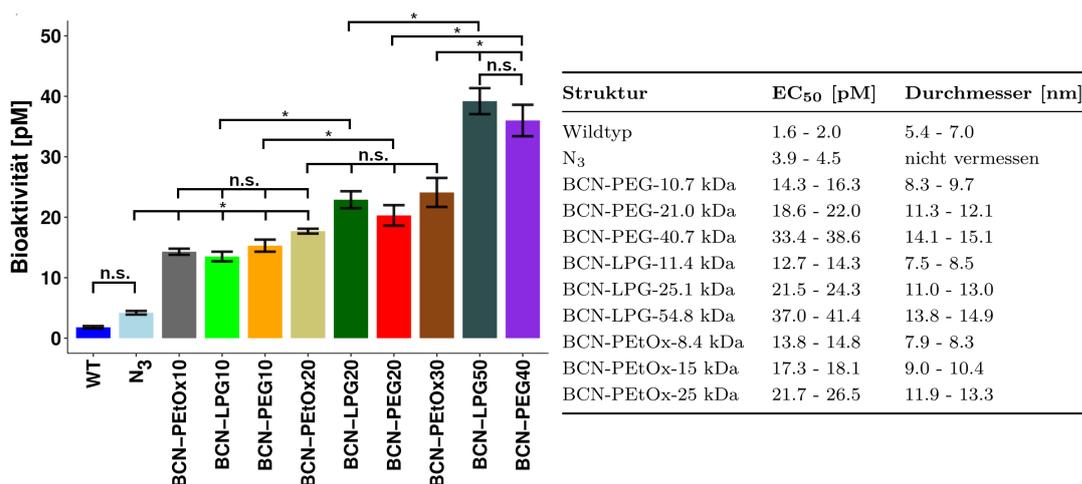


Abbildung 5.32: Per SEAP-Assay und DLS ermittelte Bioaktivitäten und hydrodynamische Durchmesser (Mittelwert \pm Standardabweichung; $n \geq 3$; signifikante Unterschiede mit p -Werte von <0.05 angedeutet; PEG/LPG-Daten unter Creative Commons Attribution License [333], PEtOx-Daten eingereicht [334]).

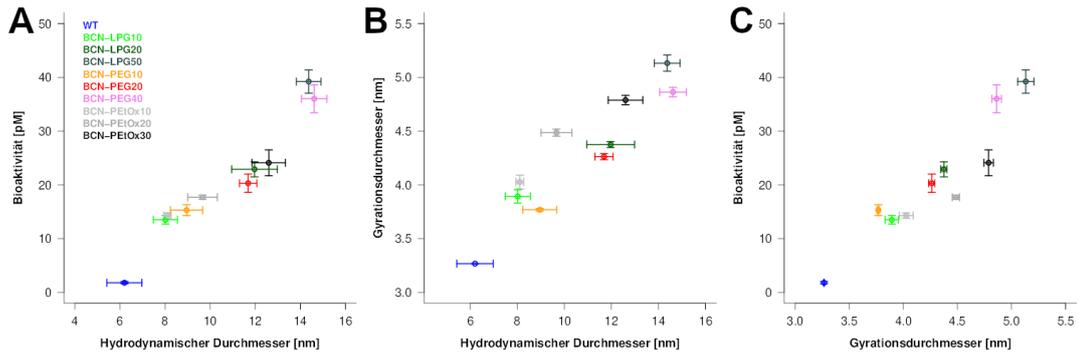


Abbildung 5.33: (A) Korrelation der Bioaktivitäten mit den hydrodynamischen Durchmessern der Konjugate ($R^2 = 0.92$, $\tau_B = 0.96$, PEG/LPG-Daten unter Creative Commons Attribution License [333], PEtOx-Daten eingereicht [334]). (B) Korrelation der hydrodynamischen Durchmesser mit berechneten Gyrationdurchmessern ($R^2 = 0.85$, $\tau_B = 0.78$). (C) Korrelation der Gyrationdurchmesser mit der Bioaktivität ($R^2 = 0.88$, $\tau_B = 0.82$).

Sofern die Bioaktivität direkt durch die Länge des ausgestreckten Polymers bestimmt werden kann, sollte bei selbem Molekulargewicht (z. B. 10 kDa) und vergleichbarer Rückgratlänge von jeweils drei Schweratomen ein ansteigender Durchmesser in der Reihenfolge PEtOx (98.13 Da pro Monomer), LPG (74.08 Da pro Monomer) und PEG (44.05 Da pro Monomer) zu erwarten sein. Die PEtOx-20-kDa-Variante zeigt erwartungsgemäß einen im Vergleich zu PEG und LPG mit selbem Molekulargewicht signifikant geringeren hydrodynamischen Durchmesser, die hydrodynamischen Durchmesser der PEG- und LPG-Varianten gleicher Größenordnung unterscheiden sich jedoch nicht signifikant.

Aufgrund der Beobachtungen stellt sich die Frage, ob analoge Korrelationen mit der Bioaktivität auch mithilfe einfacher ermittelbarer Eigenschaften wie der Monomeranzahl und / oder der Länge des ausgestreckten Polymers gefunden werden können, ohne hydrodynamische Durchmesser experimentell bestimmen oder längere Simulationen durchführen zu müssen. In Abb. 5.34-A und 5.34-B sind entsprechende Korrelationen gezeigt, die verdeutlichen, dass eine geringfügig schlechtere Korrelation direkt mithilfe der Monomeranzahlen erhalten wird. Zusätzlich wurde die Korrelation aus Abb. 5.33-C lediglich mit den 10 Startkonformationen jedes Konjugats wiederholt, welche nach den initialen, 1 ns langen Simulationen in implizitem Solvens erhalten wurden (Abb. 5.34-C). Hier wurden zu den GaMD-Analysen nahezu identische Mittelwerte erhalten, deren Korrelation leicht verbesserte statistische Werte aufweist. Somit stellt die Generierung dieser Startkonformationen eine ausreichende und gleichzeitig im Vergleich zu den GaMDs und experimentellen Messungen der hydrodynamischen Durchmessern deutlich schnellere Methode zur ersten Abschätzung der Bioaktivität dar.

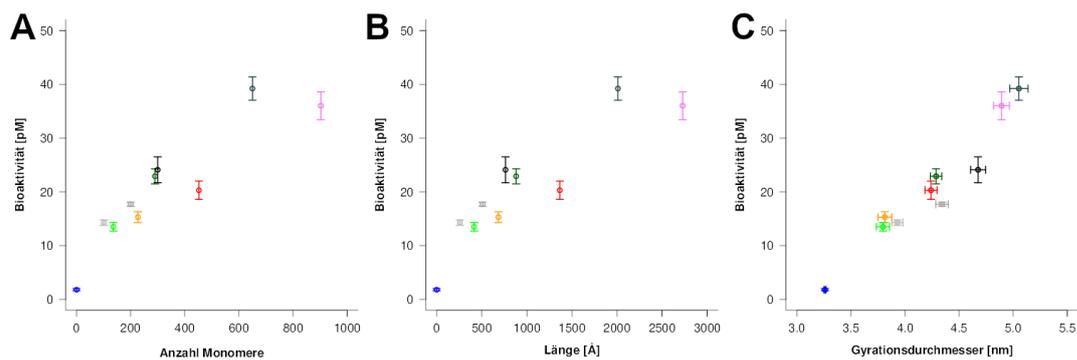


Abbildung 5.34: (A) Korrelation der Bioaktivitäten mit den für die Simulationen modellierten Anzahlen an Monomeren ($R^2 = 0.83$, $\tau_B = 0.78$) und (B) den in MOE vermessenen Längen der ausgestreckten Polymerketten ($R^2 = 0.81$, $\tau_B = 0.73$). (C) Korrelation der Bioaktivität mit den Gyrationdurchmessern der Startkonformationen der GaMDs ($R^2 = 0.93$, $\tau_B = 0.87$). Farbschema siehe 5.33-A.

5.5 Zusammenfassung

In Kap. 5.1 wurden GaMD-Simulationen für eine Vielzahl von Konjugaten unterschiedlicher Größe und Zusammensetzung durchgeführt. Durch nähere Analysen der Protein-Polymer-Kontaktprofile wurden bevorzugte Wechselwirkungspartner der Polymerarten ermittelt. Im Falle der BCN-PEG-10-kDa-Variante wurden bevorzugte Interaktionen des Polymers mit Argininen und Lysinen in der Nähe hydrophober Aminosäuren (F, Y, L, A) detektiert, was den Beobachtungen aus der Literatur entspricht [107]. LPG zeigte ein erweitertes Interaktionsprofil, indem die zusätzlichen Hydroxymethylseitenketten HB-Akzeptoren adressierten, Lysine wurden geringer bedeckt. LPG wies die stabilsten Wechselwirkungen aller Polymerarten auf, vornehmlich mit Aspartaten und Glutamaten. PEtOx wechselwirkte über die HB-Akzeptorfähigkeit der Carbonylsauerstoffatome der Seitenketten mit HB-Donatoren wie Serinen, Glutaminen und Asparaginen, sowie mit den hydrophoben Alkylseitenketten auch über aromatische Aminosäuren, v. a. Tyrosine. Diese Beobachtungen decken sich mit den systematischen Analysen aus Kap. 4.3. DBCO-10-kDa-Varianten zeigten verstärkte Bedeckungen der Aminosäuren in der Nähe der Konjugationsstelle, favorisierte Aminosäuretypen stimmten allerdings zu einem großen Teil mit den Beobachtungen zu den BCN-Konjugaten überein. Größere BCN-Konjugate (20 - 50 kDa) zeigten ebenso verstärkte Wechselwirkungen mit Aminosäuren der ersten Helix und des unstrukturierten Bereichs zwischen Helix 1 und 2. Die größten Varianten (30 - 50 kDa) wiesen zudem stärkere Bedeckungen beider Termini auf.

Die CG-Simulation eines PEG-Konjugats in Kap. 5.2 verdeutlicht, dass die Polymere durchaus auch zum großen Teil solvensexponierte, proteinabgewandte Konformationen einnehmen können, die durch die Limitationen der GaMD-Simulationen nicht beobachtet werden können. Eine passende Gewichtung zwischen proteinnahen und -fernen Konformationen durch geeignete CG-Simulationen könnte die Ergebnisse in Kap. 5.3 verbessern, da hierbei lediglich erstere berücksichtigt werden konnten.

In Kap. 5.3 konnten mögliche Ursachen für eine Destabilisierung der 10-kDa-Konjugate von IFN- α 2a ermittelt werden. Die Parameter beschreiben in verschiedener Weise die Flexibilität unterschiedlicher Proteinbereiche. Ein Effekt von PEG auf die Stabilität von Proteinen über eine Modulation der Flexibilität wurde bereits in der Literatur beschrieben [92]. Vermutlich aufgrund mangelnder Stichprobengrößen wurden im Falle der größeren IFN- α 2a-Konjugate keine analogen Korrelationen gefunden. Effekte, die in einer (De-)stabilisierung resultieren, repräsentieren vermutlich proteinindividuelle, multifaktorielle Phänomene. Es wären voraussichtlich, ähnlich zu den Modellproteinstudien in Kap. 4, weiterführende multivariate Korrelationen notwendig, um diese ausreichend abbilden zu können.

Kap. 5.4 zeigte, analog zu experimentell ermittelten hydrodynamischen Durchmessern, eine Korrelation der Gyrationdurchmesser mit den Bioaktivitäten der Konjugate auf. Dabei konnte eine vergleichbare Korrelation mithilfe der Startkonformationen der GaMDs erhalten werden. Diese können relativ zeiteffizient modelliert werden, sodass die Ermittlung dieser Gyrationdurchmesser eine wertvolle Alternative zur experimentellen Bestimmung hydrodynamischer Durchmesser darstellen kann. Da Korrelationen mit der Anzahl an Monomeren oder den Polymerlängen zu etwas schlechteren Ergebnissen führten, hängt die erniedrigte Bioaktivität nicht ausschließlich von der Länge der ausgestreckten Polymerketten ab.

Insgesamt konnte durch dieses Kapitel demonstriert werden, dass Simulationen größerer Konjugate auf atomarer Ebene durchaus wertvolle Beiträge zur Interpretation experimenteller Befunde liefern können. Sie können weiterhin dazu dienen, favorisierte Protein-Polymer-Wechselwirkungspartner zu detektieren. Die genannten Limitierungen machen es jedoch unabdingbar, für weitergehende Analysen solcher Systeme in Zukunft deutlich zeitintensivere Simulationen durchzuführen und / oder zur Abdeckung des möglichen Konformationsraums geeignete Parameter für CG-Kraftfelder zu entwickeln. Letztere können jedoch spezifische Wechselwirkungen mit den Biologika unter Umständen nicht akkurat wiedergeben. Nur mit einem geeigneten Multiscale-Modeling-Ansatz [335] könnten also detaillierte Polymer-Protein-Interaktionen wie in Kap. 5.1.2 unter gleichzeitiger Berücksichtigung proteinabgewandter Polymerkonformationen über längere Simulationszeiten hinweg getreu abgebildet werden.

6. Fazit und Ausblick

6.1 Fazit

PBK stellen im Vergleich zu klassischen Proteinsystemen aufgrund der kovalent gebundenen, in der Regel sehr großen Polymerketten äußerst flexible Strukturen dar. Computergestützte Verfahren zur Evaluation von Protein-Protein-Interaktionen, bspw. über eigens dafür entwickelte Docking-Programme mit Scoring-Funktionen, erhielten in den letzten Jahren große Aufmerksamkeit [336–339]. Proteine verfügen über charakteristische Sekundär-/Tertiär-/Quartärstrukturen, aufgebaut aus den 20 bekannten proteinogenen Aminosäuren mit unterschiedlichen physikochemischen Eigenschaften, die die Entwicklung solcher generalisierter Ansätze erlauben. Dagegen bilden die zur Biokonjugation genutzten Polymere, vornehmlich PEG, in der Regel keine geordneten Strukturbereiche aus. Die simplen Etherstrukturen von PEG lassen die Vermutung eher unspezifischer, schwacher Interaktionen mit Proteinen zu. Entsprechend ist neben der Nutzung von Docking-Ansätzen auch die Anwendung experimenteller Methoden, bspw. der Röntgenstrukturanalyse, zur Strukturaufklärung dieser Systeme extrem limitiert.

Die Charakterisierung von PBK profitierte in der Vergangenheit jedoch von molekulardynamischen Simulationsmethoden, die die Dynamik dieser Systeme abbilden können. Weithin etablierte Kraftfelder zur Abbildung solcher hybrider Systeme stehen jedoch noch nicht zur Verfügung, bisher wurden unterschiedlichste Parametersätze zur Simulation genutzt. Dies macht einen Vergleich dieser Studien schwierig. Vergrößerte Kraftfelder erlauben das Abbilden von Konformationsänderungen über größere Zeitintervalle hinweg und können hier sinnvoll sein, resultieren jedoch auch in einem deutlichen Detailverlust, der ein Studium spezifischer Protein-Polymer- oder Solvens-Polymer-Wechselwirkungen verhindert. Es stellt sich die Frage, ob charakteristische Wechselwirkungen über deutlich kleinere Oligomer-Modellprotein-Konjugate studiert werden können. Diese stellen Systeme dar, die auch in annehmbarer Zeit mit atomarem Detail simuliert werden können. Systematische Studien aus dem Labor von J. L. Price deuten darauf hin, dass die thermische Stabilität größerer Konjugate von den Eigenschaften solcher Systeme abgeleitet werden kann. Hinweise auf den Einfluss entropisch günstiger Wasserverdrängungen untermauern die Vermutung, dass vergrößerte Kraftfelder hierfür unzulänglich sein könnten.

In Kap. 4 wurde daher untersucht, ob für die thermische Stabilität entscheidende Eigenschaften aus MD-Simulationen dieser Systeme extrahiert werden können. Die von Price et al. experimentell determinierten $\Delta\Delta G_f^0$ -Werte wurden in multivariaten QSPR-Vorhersagemodellen als Antwortvariablen genutzt, verschiedenste Eigenschaften aus den MD-Simulationen als Deskriptoren. Die ersten Modellgenerierungen in Kap. 4.1.1 zeigten zu optimistische Leistungen, da die besten Modelle auf Basis der Leistung des Holdout-Sets ausgewählt wurden, sowie sehr komplexe Modelltypen mit einer Vielzahl an Deskriptoren für einen relativ kleinen Datensatz verwendet wurden. Ansätze in Kap. 4.1.2 wurden auf lineare Regressions- und logistische Klassifikationsmodelle beschränkt, mit einem Holdout-Set, welches nicht zur Auswahl des besten Modells diente. Dies führte zu Modellen, die eine zur kristallstrukturbasierten Richtlinie von Price et al. verbesserte Leistung zeigten und dabei sehr viel mehr Konjugate mit größerer struktureller Vielfalt abdeckten. Es wurden Deskriptoren genutzt, die sowohl spezifische Protein-PEG-Wechselwirkungen als auch die Solvation der Strukturen beschreiben.

Die Modelle untermauern das Gesamtbild, welches sich aus der bisherigen Literatur ergibt: PEG kann über verschiedene Mechanismen zur Erhöhung der konformativen Stabilität führen. Hierzu können spezifische Protein-Polymer-Wechselwirkungen [99], Stärkungen intramolekularer Salzbrücken [102] oder NH- π -Bindungen [104] oder auch Desolvationen um Hydroxylgruppen [100] oder apolare Flächen [103] beitragen. Ein multivariates Vorhersagemodell sollte daher zur Ermittlung der PEG-basierten Stabilisierung geeignet sein. Die entscheidende Limitierung dieser Studie stellte jedoch der zur Verfügung stehende Datensatz dar. Zur Etablierung generalisierter Modelle müsste dieser in Zukunft noch deutlich erweitert werden. Die in Kap. 4.2 gezeigten Metadynamik-Simulationen können als Startpunkt für weiterführende Studien dienen und zeigen einen möglichen Mechanismus einer PEG-basierten Stabilisierung während des Entfaltungsprozesses auf.

Ungeachtet einer Abschätzung der thermodynamischen Stabilität können die Analysen zu favorisierten Wechselwirkungspartnern aus den Kap. 4.3 und 5.1 dabei helfen, die Dynamik der PEG-Alternativen LPG, PMeOx und PEtOx abzuschätzen. PEG interagiert, in Übereinstimmung mit bisherigen Publikationen, besonders mit Lysininen und Argininen in der Nähe apolarer Aminosäuren. LPG adressierte über zusätzliche Hydroxylgruppen auch Aspartate und Glutamate. Die POx-Varianten wechselwirkten einerseits mit HB-Donatoren über ihre Carbonylgruppen und andererseits mit den aromatischen Aminosäuren Tyrosin und Phenylalanin. Die Simulationen zu Konjugaten von IFN- α 2a in Kombination mit Untersuchungen zur thermischen Stabilität und Bioaktivität demonstrieren, dass *in silico*-Ansätze bei der Interpretation experimenteller Beobachtungen helfen können. Die CG-Simulation verdeutlicht schließlich: Zur Untersuchung therapierelevanter Biokonjugate ist ein Multiscale-Modeling-Ansatz notwendig. So könnten spezifische Interaktionen evaluiert und gleichzeitig in Relation zu proteinabgewandten Polymerkonformationen gewichtet werden.

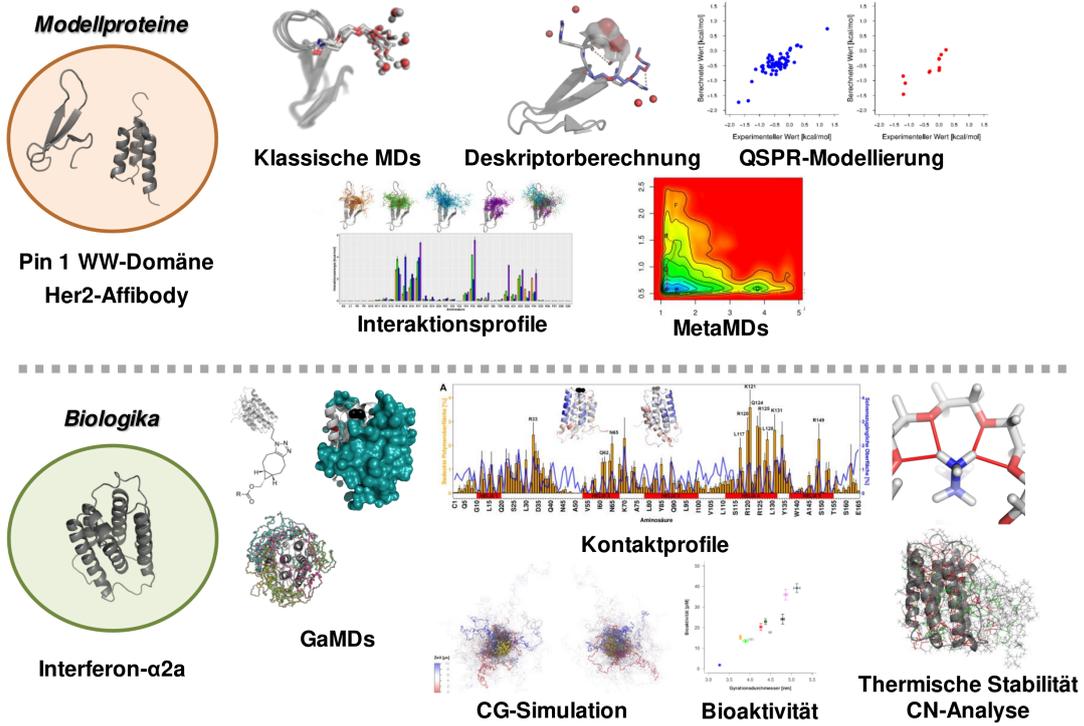


Abbildung 6.1: Zusammenfassende Illustration der Arbeit.

6.2 Ausblick

Ausgehend von den Ergebnissen dieser Arbeit lassen sich eine Vielzahl an denkbaren Fortführungen der Studien zu PBK festhalten. Ein Teil soll im Folgenden kurz aufgeführt werden.

6.2.1 Studien zu Modellproteinen

6.2.1.1 Vergrößerung des Datensatzes

Die größte Limitierung bzgl. der Generierung geeigneter Vorhersagemodelle in Kap. 4 stellte die geringe Anzahl an Konjugaten mit geeigneten experimentellen Daten dar. Eine deutlich größere Menge an charakterisierten Varianten mit einer breiteren Verteilung der $\Delta\Delta G_f^0$ -Werte wäre essenziell, möglichst mit einem gleichmäßigen Vorkommen der verwendeten Linkertypen und Konjugationsbereiche (β -Faltblätter, α -Helices, unstrukturierte Elemente). Zur Erstellung generalisierter Modelle wären daher ebenso mehr Konjugate anderer Proteine nötig.

Zusätzlich könnten die simulierten Oligomer-Konjugate der Polymeralternativen analog charakterisiert werden und so deren Vorhersagen überprüft werden. Für diese könnten zusätzliche Modelle entwickelt werden. Durch einen anschließenden Vergleich bedeutender Modelldeskriptoren könnte überprüft werden, ob für diese PEG-Alternativen andere Eigenschaften einen wesentlich bedeutenderen Einfluss auf die thermische Stabilisierung zeigen als es bei PEG der Fall ist. Bspw. könnten, ausgehend von den Beobachtungen zur Polymerdynamik, die Anzahl an Protein-Polymer-HB im Falle von LPG oder hydrophobe Interaktionen im Falle der POx-Varianten einen stärkeren Effekt haben.

6.2.1.2 Weitere QSPR-Modelle

Bei der Vielzahl an gewonnenen Deskriptoren wäre eine Nutzung von Modellarten denkbar, die mehrere kollineare Parameter in einigen wenigen latenten Variablen vereinen. Der entwickelte Modellgenerierungsansatz in Kap. 4.1.2 könnte mit Modellarten auf Basis der PLS-Methode getestet werden. Die Generierung von PLS-Modellen könnte dabei weiterhin von zusätzlichen 3D-Deskriptoren profitieren. Diese Informationen könnten aus Dichtekarten gewonnen werden. So könnte z. B. die Übereinstimmung des Aufenthalts der Polymerketten oder bestimmter Desolvatations-Hotspots neu simulierter Konjugate mit jenen stark stabilisierter Varianten verglichen werden. Differenzdichtekarten zwischen den Simulationen PEGylierter und unPEGylierter Strukturen wurden bspw. für das Solvens erstellt und darauf aufbauend die Anzahl an Gitterpunkten zwischen den beiden Konjugaten 16-1 und 27-1 verglichen, an denen starke Desolvatationen registriert wurden. Detailliertere Arbeiten mit vergleichbaren Dichtekarten könnten zur Gewinnung geeigneterer Deskriptoren führen.

6.2.1.3 Sampling der ungefalteten Varianten

Wie einige in Kap. 1.1.6.1 aufgeführte Publikationen suggerieren [95], könnte eine PEGylierung ebenso einen starken Einfluss auf die ungefaltete Variante eines Proteins haben. In der hier vorliegenden Arbeit wurden solche Konformationen nicht untersucht. Ein großes Ensemble an jeweils 50 verschiedenen, ungefalteten Konformationen der unPEGylierten Varianten von 16-1 und 23-1 wurde testweise über den ProtSA-Webserver generiert [340] und, mitsamt PEGylierter Versionen, in kurzen Replika (je 2 ns) simuliert, um die solvenszugänglichen Oberflächen mit jenen der gefalteten Varianten zu vergleichen. Die Idee hierbei war, dass der abschirmende Effekt des Polymers bei stabilisierten Varianten deutlich größer für die gefaltete Variante im Vergleich zur ungefalteten ist (und somit die gefaltete Struktur im Vergleich zur ungefalteten stärker stabilisiert wird). Die Ergebnisse zeigten jedoch keine Unterschiede zwischen den Konjugaten. Eventuell könnten weitere Deskriptoren zwischen den (un)gefalteten Konformationen verglichen werden. Solche Ensembles könnten ebenso dazu dienen, mittels erweiterter Sampling-Methoden die Faltungsgeschwindigkeit des Konjugats zu ermitteln.

6.2.1.4 Analyse anderer Eigenschaften

Vergleichbare Simulationsstudien zu Affinitäts- oder Proteolysedaten kleinerer Modellproteine sind denkbar. Solche Daten standen für diese Arbeit nicht zur Verfügung und konnten nur in sehr begrenztem Maße in Bezug auf größere Konjugate untersucht werden. Während die thermodynamische Stabilität von Konjugaten nicht direkt aus den beobachteten Wechselwirkungen des Polymers mit dem Protein ableitbar ist, wären Schlussfolgerungen zu Affinitäten der Konjugate bei bekannter Binderegion auf Basis solcher Simulationen einfacher. Eine verringerte Affinität sollte mit dem Ausmaß der sterischen Abschirmung der Binderegion in den Simulationen korrelieren. In ähnlicher Weise sollte die Abschirmung bekannter Schnittstellen von Proteasen mit Proteolysedaten übereinstimmen.

Eine weitere pharmazeutisch wichtige Eigenschaft stellt die Aggregationstendenz von Biologika dar. Durch den AGGRESCAN3D-Webserver wurden bspw. potenzielle Hotspots der untersuchten Proteine detektiert, die maßgeblich zu einer Aggregation führen könnten. Hilfsstoffe in pharmazeutischen Formulierungen können dazu beitragen, diese Areale zu maskieren [341–343]. In ähnlicher Weise könnte eine *in silico*-Studie zur polymerbasierten Abschirmung dieser Proteinbereiche zur Abschätzung der Aggregationstendenz von Konjugaten beitragen.

Im Hinblick auf die thermische Stabilität könnte der Fokus bspw. auch von den $\Delta\Delta G_f^0$ -Werten weg, hin zu möglichen Entfaltungszentren der Proteine gerichtet werden. So wurden bspw. mit dem CNA-Programm mögliche Weak spots der Proteine identifiziert. Es könnte näher untersucht werden, inwiefern eine Abschirmung dieser Bereiche mit einer Stabilisierung der Konjugate einhergeht.

6.2.1.5 Experimentelle Ansätze

Die hier präsentierten Modellierungsarbeiten basierten auf publizierten Stabilitätsdaten, die über eine temperaturabhängige CD-Spektroskopie ermittelt wurden. Zur Etablierung von Datensätzen wären auch andere Messmethoden zur Charakterisierung der Proteininstabilität denkbar, bspw. über eine chemisch induzierte Entfaltung mittels Guanidiniumchlorid [95] oder DSF-Messungen analog zu jenen der größeren Konjugate in dieser Arbeit. Eine Alternative zum beschriebenen LiP-Assay zum Studium der Protein-Polymer-Interaktionen könnten NMR-Messungen darstellen. Erst kürzlich wurde ein solcher Ansatz mit $^1\text{H}/^{15}\text{N}$ -HSQC (Heteronuclear Single Quantum Coherence)-Messungen beschrieben [15], wobei jedoch fraglich ist, inwiefern die 2,2,6,6-Tetramethylpiperidinyloxyl (TEMPO)-Strukturen, die hierbei als Spin labels eingesetzt wurden, die Polymerdynamik beeinflussen.

6.2.1.6 Alternative Simulationsansätze

6.2.1.6.1 Metadynamik-Simulationen

Der in dieser Arbeit entwickelte Ansatz zu Metadynamik-basierten Entfaltungssimulationen könnte auf eine größere Auswahl an Konjugaten mit deutlich längeren Simulationszeiten erweitert werden. So könnte die Eignung zur Abschätzung PEG-basierter Stabilisierungen evaluiert werden. Die engen Energieintervalle erlauben aber vermutlich keinen verlässlichen quantitativen Vergleich.

6.2.1.6.2 Erhöhte Temperaturen

Neben klassischen MD-Simulationen bei 300 K wurden zu Beginn der Arbeit einige wenige Konjugationsstellen ebenso bei erhöhten Temperaturen simuliert (330 K und 360 K). Dieser Ansatz wurde jedoch nicht weiter verfolgt, da die Anzahl benötigter Simulationen bei analoger Untersuchung des gesamten Datensatzes immens ansteigen würde. Dennoch wäre es retrospektiv gesehen interessant zu studieren, ob bestimmte berechnete Eigenschaften (z. B. der Sekundärstrukturanteil und die Anzahl nativer Kontakte) bei erhöhten Temperaturen im Falle besonders destabilisierter Varianten deutlich stärker moduliert werden als es bei stabilisierten Konjugaten der Fall ist. In den Anfängen dieser Arbeit wurde für einige wenige Varianten bspw. die Temperaturabhängigkeit der Rückgratorparameter untersucht, da diese in einer Studie über Ribonucleasen bereits sehr gut mit den experimentell bestimmten Schmelztemperaturen korrelierten [344]. Nach zusätzlichen Simulationen weiterer Konjugate ergab sich aber keine vielversprechende Korrelation mehr. Es müsste im Einzelfall überprüft werden, in welcher Art und Weise die jeweiligen Eigenschaften eine Temperaturabhängigkeit zeigen. Vergleichbare Simulationen könnten auch zur Beobachtung einer kompletten Entfaltung des Proteins dienen.

6.2.1.6.3 Gerichtete MD-Simulationen

Neben Metadynamik-Simulationen existieren andere Simulationsansätze zur Untersuchung von Nichtgleichgewichtsbedingungen. Gerichtete MD-Simulationen können bspw. genutzt werden, um mithilfe der Jarzynski-Gleichung die Bindungskinetiken von Liganden zu studieren [345, 346]. Durch Anlegen einer konstanten Kraft zur Entfaltung einer einfachen, PEGylierten Helix wurde bereits der schützende Effekt des Polymers auf die Rückgrat-HB untersucht [94]. Ein solcher Ansatz wurde in dieser Arbeit nicht verfolgt, da die zugrundeliegenden Modellproteine wesentlich komplexere Strukturen aufweisen und die Auswahl eines möglichen Angriffspunkts zur Entfaltung somit keineswegs trivial ist.

6.2.1.6.4 Entfaltung über Gō-Modelle

Zum Studium der Proteinentfaltung können ebenso extrem vereinfachte Gō-Modelle dienen, die eine Art der Vergrößerung des Kraftfelds repräsentieren [253, 256, 347]. Dabei werden Aminosäuren als einzelne Kugeln dargestellt. So wurde bspw. ein Gō-Modell für einen Replica Exchange-Ansatz zum Studium der nativen Kontakte von T4-Lysozym-PEG-Konjugaten mit DBCO-Linker verwendet [348]. Vergleichbare Gō-basierte Simulationsansätze wurden in dieser Arbeit mit dem MMTSB-Toolset [349] für CHARMM [350], sowie mit dem Programm GENESIS [351] für unPEGylierte Varianten getestet. Hier besteht jedoch grundsätzlich das Problem einer geeigneten Parameterentwicklung für den Polymer- und Linker-Part der Konjugate.

6.2.1.7 MD-Alternativen

Weiterhin existieren weitaus recheneffizientere Alternativen zu MDs zur schnellen Beurteilung der Stabilität und Dynamik biomolekularer Strukturen. Verschiedene Ansätze sind bspw. über das Python-Paket PolyDy [352] möglich: Analysen der Normalmodi, sowie anisotropische oder auch Gaußsche Netzwerkmodelle. Diese Methoden könnten direkt zur Beurteilung von Stabilitäten dienen oder aber zur Gewinnung möglicher Deskriptoren für die Modellgenerierung. Dabei stellt sich aber die Frage nach einer geeigneten Implementierung der PEGylierten Aminosäure im Netzwerk.

6.2.2 Größere Konjugate

6.2.2.1 Verbesserung des Samplings

6.2.2.1.1 Abbildung des Solvens

Eine entscheidende Limitierung dieser Studie stellte die begrenzte Stichprobennahme der größeren Konjugate dar, die, gerade aufgrund des expliziten Solvens, Systeme mit zahlreichen Freiheitsgraden darstellen. Denkbar wären hierzu Simulationen mit implizitem Solvens, Hybrid-Simulationen (z. B. mit dem PACE-Kraftfeld [353]) oder eine Nutzung des neu entwickelten Solvation Shell-Modells [354].

6.2.2.1.2 GaMD-Alternativen

In dieser Studie wurden zur Unterstützung des Samplings GaMDs durchgeführt. Auch wenn die generierten Daten grundsätzlich zur energetischen Regewichtung entlang ausgewählter Reaktionskoordinaten (z. B. Gyrationenradien, SASA) zur Verfügung stehen, könnten klassische beschleunigte MD-Simulationen oder auch andere Ansätze wie Replica Exchange-Methoden das Sampling weiter verbessern, auch wenn eine genaue Regewichtung entlang bestimmter Variablen dadurch evtl. nicht mehr gewährleistet ist.

6.2.2.1.3 Vergrößerte Kraftfelder

Die Etablierung geeigneter CG-Parameter könnte die Analyse größerer Konjugate mithilfe längerer Simulationen erleichtern. Ein Protokoll zur Simulation PEGylierter Biologika mit dem neuen MARTINI 3-Kraftfeld wurde kürzlich publiziert [163]. Im Falle der alternativen Polymere müssten geeignete Parameter vorerst entwickelt werden. In einer kürzlich publizierten Studie zu CG-Modellen von Polyethylenimininen wurde darauf hingewiesen, dass diese als Startpunkt zur Entwicklung von Parametern für Poly(2-alkyl-2-oxazoline) dienen könnten [355]. Diese wurden mit einer zusätzlichen Kugel für die Seitenkette testweise im PolyPly-Programm implementiert. Ebenso wurden für LPG die zur Verfügung stehenden PEG-Parameter mit einer kleinen Kugel ähnlich jener der Seitenkette von Serin erweitert. Die generierten Modelle führten in ersten CG-Simulationen jedoch zur vollständigen Aggregation der Polymere auf die Proteinoberfläche, was auf unzureichende Parameter hindeutet. Weiterführende Studien unter iterativem Abgleich mit Ergebnissen aus klassischen MD-Simulationen und Experimenten wären notwendig. Zuletzt stellt aber auch die korrekte Parametrisierung der komplexen Cyclooctin-Linker eine große Herausforderung dar.

Wären vergrößerte Kraftfelder für die entsprechenden Systeme etabliert, so könnte weiterhin ein Backmapping-Ansatz eingesetzt werden, um geeignete Konformationen für weitergehende Analysen gemäß Kap. 5.3.1 durchzuführen [356]. So könnten die Ergebnisse der CN-Analysen eventuell verbessert werden, da der Anteil an proteinabgewandten Konformationen berücksichtigt werden würde.

6.2.2.2 MD-Alternativen

Neben molekulardynamischen Ansätzen bieten bspw. auch Methoden, die auf dem Prinzip des maschinellen Lernens basieren, Möglichkeiten, den Konformationsraum von Systemen zu untersuchen, so z. B. der Ansatz des Generative Topographic Mapping [357, 358] oder jener der Boltzmann-Generatoren [359]. Auch rigiditätsbasierte, geometrische Simulationen existieren, um den Konformationsraum größerer Proteinsysteme zu studieren [360].

Zusammenfassung

Polymer-Biokonjugationen, vornehmlich mit dem Goldstandard PEG, führen zu einer verbesserten Pharmakokinetik, beeinflussen aber auch die konformative Stabilität von Proteinen. Bisherige Mutationsstudien, in denen überwiegend (Asn)PEG₄-Konjugate der β -faltblattstrukturreichen, humanen Pin 1 WW-Domäne untersucht wurden, postulieren auf einer Proteindesolvatation beruhende Stabilisierungsmechanismen: eine Stärkung intramolekularer Salzbrücken und NH- π -Bindungen, sowie entropisch günstige Wasserverdrängungen um apolare Aminosäuren und Hydroxylgruppen.

Ziel dieser Arbeit ist es, die Protein-Polymer-Dynamik auf molekularer Ebene zu charakterisieren, um damit rationale Ansätze zum Design neuer Biokonjugate voranzutreiben und mögliche PEG-Alternativen zu etablieren. Hierzu wurde eine Vielzahl an Deskriptoren mittels Molekulardynamik-Simulationen der WW-Konjugate gewonnen und mit publizierten Stabilitätsdaten ($\Delta\Delta G_f^0$ -Werte) in multivariaten Regressions- und logistischen Klassifikationsmodellen korreliert. Die gewonnenen QSPR-Modelle decken im Vergleich zu einer bereits publizierten, kristallstrukturbasierten Richtlinie einen größeren und strukturell vielfältigeren Datensatz an Konjugaten ab und zeigen gleichzeitig, auch für ein Konjugat der Src SH3-Domäne, eine deutlich verbesserte Leistung. Die Modelldeskriptoren beschreiben sowohl eine Modulation der Solvation als auch Protein-Polymer-Interaktionen. Metadynamik-Simulationen zeigten zudem die Polymerdynamik während einer partiellen Proteinfaltung auf. Mithilfe weiterer Simulationen von Konjugaten des α -helikalen Her2-Affibodys wurde die Dynamik von PEG und verschiedener Alternativen (LPG, PEtOx, PMeOx) systematisch studiert. PEG interagiert mit positiv geladenen Lysin und Arginin in der Nähe hydrophober Aminosäuren. LPG zeigte zusätzliche Wechselwirkungen der Hydroxylgruppen mit Aspartat und Glutamat. POx-Polymere interagierten mit Phenylalanin, Tyrosin und über Carbonylgruppen mit HB-Donatoren.

Größere Konjugate (10 - 50 kDa PEG/LPG/PEtOx) des antiviralen Biologikums Interferon- α 2a wurden mittels gaußbeschleunigter MDs und einer CG-Simulation analysiert. Charakteristische Wechselwirkungspartner stimmten mit den Beobachtungen zu Oligomer-Konjugaten überein. In Einklang mit experimentellen Daten der Kooperationspartner zu den 10-kDa-Varianten deuteten zusätzliche Constrained-Network-Analysen, welche die Proteinflexibilität evaluieren, auf eine thermische Destabilisierung hin. Die Bioaktivität der untersuchten Konjugate wurde weiterhin erfolgreich mit den Gyrationdurchmessern der modellierten Strukturen korreliert.

Summary

Bioconjugation of polymers, mainly the gold standard PEG, can improve pharmacokinetic properties but also modulate conformational stability of proteins. Mutation studies on (Asn)PEG₄ conjugates of the β -sheet rich human Pin 1 WW domain suggest various desolvation effects playing a crucial role: strengthening of intramolecular salt-bridges and NH- π bonds, as well as entropically favorable water expulsion around hydrophobic patches and hydroxyl groups.

The goal of this study is to characterize protein-polymer dynamics on a molecular level to drive forward rational design of new bioconjugates and establish viable PEG alternatives. A variety of descriptors was calculated from molecular dynamics simulations of WW conjugates and correlated with published stability data ($\Delta\Delta G_f^0$ values) generating multivariate regression and logistic classification models. Compared to a previously published crystal structure-based guideline, QSPR models covered a structurally more diverse and bigger dataset and showed significantly improved predictions, including for a conjugate of the Src SH3 domain. Model descriptors captured modulations of solvation as well as protein-polymer interactions. Metadynamics simulations depicted PEG dynamics upon partial protein unfolding. Combined with simulations for conjugates of the α -helical Her2 affibody, data was further used to systematically dissect the dynamics of PEG and its alternatives LPG, PEtOx and PMeOx. PEG interacted with lysines and arginines near hydrophobic patches. LPG additionally addressed aspartates and glutamates via its hydroxyl groups. POx variants interacted with phenylalanines, tyrosines, as well as hydrogen bond donors via carbonyl groups.

Larger conjugates (10 - 50 kDa PEG/LPG/PEtOx) of antiviral biologic Interferon- α 2a were analyzed via Gaussian accelerated MDs and an exemplary CG simulation. Interaction patterns agreed with observations for oligomer conjugates. In accordance with experimental data of collaboration partners for 10 kDa variants, constrained network analyses, assessing protein flexibility, suggested a thermal destabilization upon bioconjugation. Bioactivity of conjugates was further successfully correlated with diameters of gyration of modeled structures.

Anhang

A. Ausgleichungsrechnungen

A.1 Zugrundeliegende Gleichungen

Die Elliptizität $[\theta]_\lambda$ steht, unter Annahme einer klassischen Zwei-Phasen-Kinetik der Entfaltung, mit der temperaturabhängigen Faltungskonstante K_f , die das Gleichgewicht zwischen der gefalteten ($[F]$ mit $[\theta_f]$) und der ungefalteten ($[U]$ mit $[\theta_u]$) Struktur bzw. den Anteil des ersteren (f) beschreibt, in folgendem Zusammenhang [100, 102, 104]:

$$[\theta] = (1 - f) [\theta_u] + f[\theta_f] \quad (\text{A.1})$$

$$[\theta_f] = N_0 + N_1 T \quad (\text{A.2})$$

$$[\theta_u] = D_0 + D_1 T \quad (\text{A.3})$$

$$K_f = \frac{[F]}{[U]} \quad (\text{A.4})$$

$$= \frac{f}{1 - f} \quad (\text{A.5})$$

$$[\theta] = \frac{(D_0 + D_1 T) + K_f(N_0 + N_1 T)}{1 + K_f} \quad (\text{A.6})$$

wobei D_0 und N_0 die Ordinatenabschnitte, sowie D_1 und N_1 die Steigungen der Regressionsgeraden für das denaturierte (D , Basislinie nach Entfaltung) bzw. das native (N , Basislinie vor Entfaltung) Protein darstellen. Aus der Faltungskonstante kann wiederum die freie Faltungsenergie ΔG_f^0 ermittelt werden:¹

$$K_f = e^{\frac{-\Delta G_f^0}{RT}} \quad (\text{A.7})$$

ΔG_f^0 steht unmittelbar in Zusammenhang mit der Enthalpieänderung $\Delta H_f(T_m)$ und der Änderung der Wärmekapazität ΔC_p :

$$\Delta G_f^0 = \frac{\Delta H_f(T_m)}{T_m} (T_m - T) + \Delta C_p [T - T_m - T \ln(\frac{T}{T_m})] \quad (\text{A.8})$$

¹In den Referenzpublikationen von Price et al. wird die Angabe zu Standardbedingungen teilweise nicht aufgeführt (Bezeichnung als ΔG_f anstatt ΔG_f^0).

Durch Kombination der Glg. A.1 - A.8 ergibt sich folgende Bedingung:

$$[\theta] = \frac{(D_0 + D_1 T) + e^{-\left[\frac{\Delta H_f(T_m)}{T_m}(T_m - T) + \Delta C_p \ln\left(\frac{T}{T_m}\right)\right]}}{RT} (N_0 + N_1 T)}{1 + e^{-\left[\frac{\Delta H_f(T_m)}{T_m}(T_m - T) + \Delta C_p \ln\left(\frac{T}{T_m}\right)\right]}} \quad (\text{A.9})$$

Eine Ausgleichsrechnung (*Fitting*) kann somit mit den experimentell ermittelten Werten für $[\theta]$ basierend auf dieser Beziehung durchgeführt werden. Hierbei sind D_0 , N_0 , D_1 , N_1 , $\Delta H_f(T_m)$, ΔC_p und T_m variable Parameter. Glg. A.8 basiert auf folgenden Beziehungen:

$$\Delta G^0 = \Delta H^0 - T \Delta S^0 \quad (\text{A.10})$$

$$\Delta H^0 = \Delta H_f(T_m) + \Delta C_p (T - T_m) \quad (\text{A.11})$$

$$\Delta S^0 = \Delta S_f(T_m) + \Delta C_p \ln\left(\frac{T}{T_m}\right) \quad (\text{A.12})$$

$$\Delta S_f(T_m) = \frac{\Delta H_f(T_m)}{T_m} \quad (\text{A.13})$$

Die Ergebnisse der Fitting-Parameter, besonders $\Delta H_f(T_m)$, hängen bei dieser Methode bei einer einzigen Denaturierungskurve in der Regel stark von der gewählten Anzahl an Punkten zur Erstellung der Regressionslinien der nativen und denaturierten Varianten ab [361]. Durch Kombination aller drei Messreihen einer Variante in einem einzigen Fitting-Prozess können die Variablen aber mit einer ausreichenden Sicherheit bestimmt werden. Dies wird erreicht, indem die Messreihen $[\theta]_1$, $[\theta]_2$ und $[\theta]_3$ einer Triplika-Messung in einem einzigen Vektor $[\theta]_{ges}$ vereinigt werden und für die rechte Seite der Glg. A.9 dann jeweils die für den Messwert passenden Basislinien-Variablen verwendet werden (d. h., für Messwerte aus $[\theta]_1$ werden z. B. N_{01} , N_{11} , D_{01} und D_{11} genutzt). Die Variablen T_m , $\Delta H_f(T_m)$ und ΔC_p kommen dagegen bei jedem Messwert von $[\theta]_{ges}$ zum Einsatz. Die Ausgleichsrechnung basiert also auf der Minimierung der Residuen aller Messpunkte aus $[\theta]_{ges}$ (Glg. A.15). Es handelt sich somit, bei Triplika-Messungen, um einen Fitting-Prozess mit insgesamt 15 Variablen: T_m , $\Delta H_f(T_m)$, ΔC_p , N_{01} , N_{11} , D_{01} , D_{11} , N_{02} , N_{12} , D_{02} , D_{12} , N_{03} , N_{13} , D_{03} und D_{13} [104].

$$[\theta]_{ges} = \left[[\theta]_1, [\theta]_2, [\theta]_3 \right] \quad (\text{A.14})$$

$$[\theta] = \left\{ \begin{array}{l} \frac{(D_{01}+D_{11} T)+e^{-\left[\frac{\Delta H_f(T_m)}{T_m}(T_m-T)+\Delta C_p \left[T-T_m-T \ln\left(\frac{T}{T_m}\right)\right]\right]}}{RT} (N_{01}+N_{11} T)}{1+e^{-\left[\frac{\Delta H_f(T_m)}{T_m}(T_m-T)+\Delta C_p \left[T-T_m-T \ln\left(\frac{T}{T_m}\right)\right]\right]}} \quad [\theta] \in [\theta]_1 \\ \\ \frac{(D_{02}+D_{12} T)+e^{-\left[\frac{\Delta H_f(T_m)}{T_m}(T_m-T)+\Delta C_p \left[T-T_m-T \ln\left(\frac{T}{T_m}\right)\right]\right]}}{RT} (N_{02}+N_{12} T)}{1+e^{-\left[\frac{\Delta H_f(T_m)}{T_m}(T_m-T)+\Delta C_p \left[T-T_m-T \ln\left(\frac{T}{T_m}\right)\right]\right]}} \quad [\theta] \in [\theta]_2 \\ \\ \frac{(D_{03}+D_{13} T)+e^{-\left[\frac{\Delta H_f(T_m)}{T_m}(T_m-T)+\Delta C_p \left[T-T_m-T \ln\left(\frac{T}{T_m}\right)\right]\right]}}{RT} (N_{03}+N_{13} T)}{1+e^{-\left[\frac{\Delta H_f(T_m)}{T_m}(T_m-T)+\Delta C_p \left[T-T_m-T \ln\left(\frac{T}{T_m}\right)\right]\right]}} \quad [\theta] \in [\theta]_3 \end{array} \right. \quad (\text{A.15})$$

Erhalten werden schließlich die relevanten Parameter T_m , $\Delta H_f(T_m)$ und ΔC_p für die PEGylierte und die unPEGylierte Variante, mit dazugehörigen Standardfehlern des Fits. Aus diesen wird mit Glg. A.8 zunächst ΔG_f^0 bei der Schmelztemperatur T_m und daraus abgeleitet $-T\Delta S_f^0$ der unPEGylierten Variante berechnet. ΔG_f^0 der unPEGylierten Struktur beträgt bei dieser Temperatur 0 kcal/mol, K_f beträgt 1 und f 0.5. Anschließend werden ΔG_f^0 , ΔH_f^0 und $-T\Delta S_f^0$ der PEGylierten Variante bei der Schmelztemperatur T_m der unPEGylierten Variante berechnet und somit die gesuchten energetischen Differenzen zwischen der PEGylierten und der unPEGylierten Variante erhalten, die als Indikator für eine veränderte thermodynamische Stabilität dienen: die Differenz der freien Faltungsenthalpien $\Delta\Delta G_f^0$ und den damit zusammenhängenden Faltungsenthalpien $\Delta\Delta H_f^0$ und -entropien $-T\Delta\Delta S_f^0$ [86, 88, 100, 104].

Für die nichtlinearen Kurven werden je nach verwendetem Analysenprogramm (z. B. GraphPad, SigmaPlot oder Psi-Plot) unterschiedliche Optimierungsalgorithmen zur Minimierung der Residuen genutzt, wobei zumeist der Levenberg-Marquardt-Algorithmus genutzt wird [86, 88, 362].

A.2 Fehlerfortpflanzung

Zum Erhalt der Standardabweichung einer Variable x , die über eine Funktion f mit den Parametern a , b und c in Zusammenhang steht, werden die Fehler s_a , s_b , s_c gemäß Gaußscher Fehlerfortpflanzung mit den partiellen Ableitungen verrechnet.

$$s_f = \sqrt{\left(\frac{\delta f(x)}{\delta a} s_a\right)^2 + \left(\frac{\delta f(x)}{\delta b} s_b\right)^2 + \left(\frac{\delta f(x)}{\delta c} s_c\right)^2} \quad (\text{A.16})$$

Auf Basis der Glg. A.8 gilt für den Fehler von ΔG_f^0 somit [102]:

$$\frac{d\Delta G_f^0}{d\Delta H_f(T_m)} = \frac{T_m - T}{T_m} \quad (\text{A.17})$$

$$\frac{d\Delta G_f^0}{d\Delta C_p} = T - T_m - T \ln\left[\frac{T}{T_m}\right] \quad (\text{A.18})$$

$$\frac{d\Delta G_f^0}{d\Delta T_m} = \Delta C_p \left[\frac{T}{T_m} - 1\right] - \frac{\Delta H_f(T_m)}{T_m} \left[\frac{T_m - T}{T_m} - 1\right] \quad (\text{A.19})$$

$$\sigma_{\Delta G_f^0} = \sqrt{\left(\frac{d\Delta G_f^0}{d\Delta H_f(T_m)} s_{\Delta H_f(T_m)}\right)^2 + \left(\frac{d\Delta G_f^0}{d\Delta C_p} s_{\Delta C_p}\right)^2 + \left(\frac{d\Delta G_f^0}{d\Delta T_m} s_{T_m}\right)^2} \quad (\text{A.20})$$

B. Deskriptoren

Viele verschiedene Eigenschaften, die für Vorhersagemodelle als Deskriptoren genutzt werden können, können aus MD-Trajektorien gewonnen werden. Im Folgenden sollen einige zentrale Konzepte kurz erklärt werden, die in der vorliegenden Arbeit eine Anwendung finden. Fast alle genannten Parameter können mithilfe von `cpptraj` von `AmberTools18` berechnet werden [282, 363]. Die zu verwendenden Befehle sind mit aufgeführt. Wenn nicht anders referenziert stammen die entsprechenden Informationen aus dem Handbuch von `AMBER18` [282].

B.1 Mittlere Abweichung (RMSD)

`cpptraj: rmsd`

Der RMSD-Wert (*Root-mean-square deviation*) beschreibt die mittlere Abweichung einer Struktur von einer Referenzstruktur. In der Regel wird als Referenz die zugrundeliegende Kristallstruktur genutzt. Der Wert lässt sich nach folgender Gleichung berechnen, wobei N die Anzahl der Atome, x_i die Position des i -ten Atoms im Raum und x_i^{ref} die Position dieses Atoms in der Referenzstruktur darstellen [364]:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x_i^{ref})^2} \quad (\text{B.1})$$

B.2 Mittlere Fluktuation (RMSF)

`cpptraj: rmsf`

Während der RMSD-Wert einen Mittelwert zu einem bestimmten Zeitpunkt angibt, stellt der RMSF-Wert (*Root-mean-square fluctuation*) die Fluktuation über ein gewisses Zeitintervall dar [365]:

$$RMSF = \sqrt{\frac{1}{T} \sum_{j=1}^T (x_j - x_j^{ref})^2} \quad (\text{B.2})$$

Hierbei wird die Abweichung eines einzigen Partikels von der Referenzstruktur also über das Zeitintervall j bis T gemittelt. Als Referenz dient hier in der Regel die

gemittelte Position des jeweiligen Partikels. Der RMSF-Wert für eine Gruppe von Atomen kann massengewichtet über folgende Gleichung berechnet werden:

$$RMSF_{Gesamt} = \frac{\sum Atomfluktuation_i Masse_i}{\sum Masse_i} \quad (\text{B.3})$$

Die Fluktuationen können weitergehend auch quadriert und mit dem Faktor $\frac{8}{3}\pi^2$ gewichtet werden, um einen zu den B-Faktoren von Kristallstrukturanalysen vergleichbaren Wert zu erhalten.

B.3 Solvenz zugängliche Oberfläche (SASA)

cpptraj: surf, molsurf

Über den Befehl *surf* kann der Beitrag einer Atomauswahl zur Gesamtoberfläche des jeweiligen Moleküls gemäß des LCPO-Algorithmus (*Linear combination of pairwise overlaps*) berechnet werden [366]. Hierbei wird der Van-der-Waals-Radius mit einem Faktor von 1.4 multipliziert. Im Gegensatz hierzu berechnet der *molsurf*-Befehl die absolute Connolly-Oberfläche (standardmäßiger Sondenradius: 1.4 Å) [367] einer Auswahl ohne Berücksichtigung der Bindung zu übrigen Atomen. Die Auswahl wird hierbei also als isoliertes System betrachtet.

B.4 Radiale Verteilungsfunktion (RDF)

cpptraj: rdf

Über die radiale Verteilungsfunktion (*Radial distribution function*) wird die Wahrscheinlichkeit wiedergegeben, mit der eine Atomauswahl A innerhalb eines bestimmten Abstands R von der Atomauswahl B lokalisiert ist. Dieser Wert wird aus einem normalisierten Histogramm berechnet, mit der Formel:

$$RDF = Dichte * \left(\left[\frac{4\pi}{3}(R + dR)^3 \right] - \left[\frac{4\pi}{3}dR^3 \right] \right) \quad (\text{B.4})$$

Dabei entspricht dR der Länge eines Intervalls des Histogramms. In der Regel wird diese Berechnung für Solvensmoleküle um bestimmte Proteinatome genutzt, wobei charakteristische Maxima für die unterschiedlichen Hydrathüllen erkennbar werden. In diesem Falle sollte die RDF-Funktion bei größeren Distanzen im Bereich des Bulk-Wassers einen Wert von 1 annehmen.

B.5 Gyrationradius

cpptraj: radgyr

Der Gyrationradius R_g (oder Streumassenradius) ist eine bekannte Größe in der Polymerforschung und stellt die mittlere Distanz aller Atome zum Massen- oder geometrischen Zentrum der Auswahl dar. Nach IUPAP (*Internationale Union für reine*

und angewandte Physik) ist der Radius definiert nach [368]:

$$R_g = \sqrt{\frac{\sum_{i=1}^n m_i r_i^2}{\sum_{i=1}^n m_i}} \quad (\text{B.5})$$

Dabei stellen m_i die Masse und r_i die Distanz des i -ten Atoms zum Massenzentrum der gesamten Auswahl dar. Für bestimmte Formen (z. B. Zylinder oder Kugeln) existieren auch Umrechnungsmethoden zur Ermittlung des hydrodynamischen Radius R_h der Struktur [369].

B.6 HB/Lebensdauer-Analyse

cpptraj: hbond, lifetime

Über den *hbond*-Befehl können die Wasserstoffbrückenbindungen (HB) pro Momentaufnahme, mit definierten Grenzen bzgl. Distanz und Winkel, aus einer Trajektorie erhalten werden. Diese Daten können anschließend mit dem *lifetime*-Befehl genutzt werden, um die durchschnittlichen und maximalen Zeitspannen zu berechnen, über die eine HB beibehalten wird.

B.7 Interaktionsenergie

cpptraj: lie

Die Interaktionsenergie zwischen zwei Atomauswahlen kann über einen vereinfachten Ansatz mithilfe von *lie* mit einem standardmäßigen Cut-off-Wert von 12 Å für Van-der-Waals- und elektrostatische Wechselwirkungen berechnet werden. Es wird eine simple Shifting-Funktion genutzt, wobei die Coulomb-Interaktionen über folgende Gleichung berechnet werden:

$$E = \frac{kq_i q_j}{r_{ij}} \left(1 - \frac{r_{ij}^2}{r_{cut}^2}\right)^2 \quad (\text{B.6})$$

Dabei wird die Konstante k durch die vorher definierbare Dielektrizitätskonstante bestimmt (normalerweise jene des verwendeten Wassermodells), q_i und q_j stellen die Potentialladungen der Atome, r_{ij} die Distanz und r_{cut} die Cut-off-Distanz dar. Es handelt sich somit um eine Berechnung ohne Berücksichtigung expliziter Solvensmoleküle.

B.8 Rückgrat-Diederwinkel

cpptraj: multidihedral

Über *multidihedral* können die Φ - und Ψ -Winkel jeder Aminosäure bestimmt werden, über die die Konformation des Proteins beschrieben werden kann. Diese Diederwinkel sind bekannt aus Ramachandran-Plots [370]. Der Diederwinkel Φ wird über die Atome C_{i-1} , N_i , $C\alpha_i$ und C_i definiert, Ψ über die Atome N_i , $C\alpha_i$, C_i und N_{i+1} (Abb. B.1).

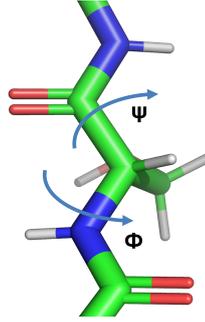


Abbildung B.1: Illustration der Winkel Φ und Ψ .

B.9 Ordnungsparameter

cpptraj: ired, matrix, diagmatrix, vector

Die Bestimmung generalisierter Ordnungsparameter S^2 beruht im Zuge des iRED-Ansatzes (*Isotropic reorientational eigenmode dynamics*) auf einer Diagonalisierung einer aus einer Hauptkomponentenanalyse erhaltenen Kovarianzmatrix [371]. Aus MDs können diese Parameter auf Basis der x -, y - und z -Komponenten der Einheitsvektoren entlang der zu untersuchenden Bindung über folgende Gleichung erhalten werden [372]:

$$S^2 = \frac{3}{2}[(x^2)^2 + (y^2)^2 + (z^2)^2 + 2(xy^2)^2 + 2(xz^2)^2 + 2(yz^2)^2] - \frac{1}{2} \quad (\text{B.7})$$

Dieser Parameter beschreibt die interne Dynamik der Bindung und bewegt sich im Bereich zwischen 0 und 1. Kleine Werte signalisieren eine sehr flexible, hohe eine rigide Bindung. Diese stehen mit der molekularen Korrelationsfunktion $C_M(t)$ in Zusammenhang, die eine im Raum frei bewegliche Bindung beschreibt. Um die Rotationskorrelationsfunktion $C_T(t)$ zu definieren, ist weiterhin eine von der molekularen Korrelation unabhängige, zusätzliche Korrelationsfunktion $C_E(t)$ notwendig, die die partielle Rigidität aufgrund der kovalenten Bindung zum Makromolekül berücksichtigt [373–375]:

$$C_T(t) = C_M(t) C_E(t) \quad (\text{B.8})$$

$$= e^{-\frac{t}{\tau_M}} (S^2 + (1 - S^2) e^{-\frac{t}{\tau_E}}) \quad (\text{B.9})$$

$$= S^2 e^{-\frac{t}{\tau_M}} + (1 - S^2) e^{-\frac{t}{\tau_T}} \quad (\text{B.10})$$

Hierbei stellen τ_M , τ_E und τ_T die entsprechenden Korrelationszeiten dar. Experimentell können Orderparameter über NMR-Messungen der Spektraldichtefunktion $J(\omega)$ ermittelt werden. Werden S^2 -Werte für bestimmte Seitenkettenbindungen oder für die N-H-Bindungen des Proteinerückgrats ermittelt, so können über diese die Konformationsentropien S auf Basis der Konstanten M , A und B berechnet werden:

$$S = k_B M [A + B(1 - S^2)] \quad (\text{B.11})$$

M stellt dabei die Anzahl an Diederwinkeln dar und A und B sind Fitting-Parameter, die sich je nach zu untersuchender Aminosäure unterscheiden. Diese Fits wurden auf Basis einer Korrelation von S_{NMR}^2 -Messwerten mit Konformationsentropien S publiziert, wobei letztere aus MD-Simulationen mithilfe von Diederwinkeln über einen von-Mises-Kernel-Ansatz stammen [376].

B.10 Native Kontakte

cpptraj: nativecontacts

Die Stabilität der nativen Konformation kann über eine Messung ihrer Atom-Atom-Kontakte erfolgen. Diese Methode wird bspw. in Studien zur thermischen Stabilität von Proteinen [332] oder Faltungen von $G\bar{o}$ -Modellen [258] genutzt. Zunächst werden alle Atom-Atom-Kontakte in der Referenzstruktur (üblicherweise der Kristallstruktur) über einen Cut-off-Wert von 4.5 Å definiert [377]. Für jede Momentaufnahme der Simulation mit einer Distanz unterhalb dieses Werts wird ein Kontakt registriert. So kann der Anteil an Momentaufnahmen bestimmt werden, in der die Kontakte der Referenzstruktur wiedergefunden wurden.

B.11 Diffusionskoeffizient

cpptraj: diffusion

Der Diffusionskoeffizient D entlang der Dimensionen n steht mit der mittleren quadratischen Verschiebung (MSD) über die Zeit t über die Einstein-Beziehung in Zusammenhang:

$$2nD = \lim_{t \rightarrow \infty} \frac{MSD}{t} \quad (\text{B.12})$$

Diese Kalkulation sollte für ein statistisch valides Ergebnis über eine Vielzahl an Atomen durchgeführt werden und eignet sich daher in der Regel nur für Solvensmoleküle.

B.12 HB-Autokorrelation

Als Alternative zur weiter oben genannten Lebensdauer-Analyse in *cpptraj* kann mithilfe des *mdanalysis*-Moduls [378, 379] die Autokorrelationsfunktion $C(t)$ und daraus die Überlebenszeit τ von HB h_{ij} zwischen den Atomen i und j ab dem Zeitpunkt t_0 berechnet werden [380]:

$$\tau = \int_0^\infty C(t) = \left\langle \frac{h_{ij}(t_0)h_{ij}(t_0+t)}{h_{ij}(t_0)^2} \right\rangle \quad (\text{B.13})$$

Hier kann die kontinuierliche oder die intermittierende Überlebenszeit berechnet werden. Bei letzterer sind zwischenzeitliche Bindungsbrüche erlaubt.

B.13 CNA-basierte Deskriptoren

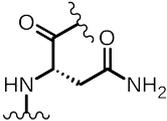
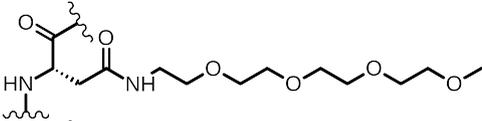
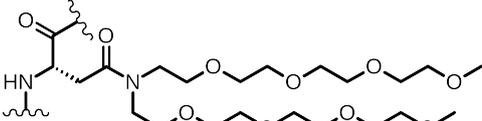
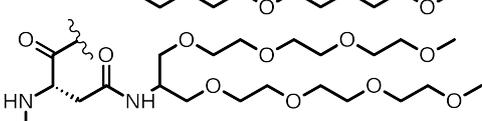
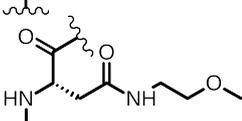
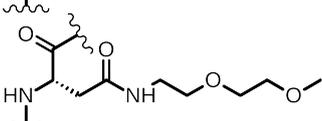
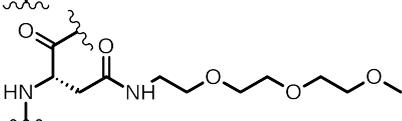
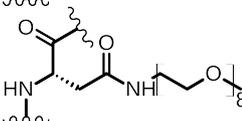
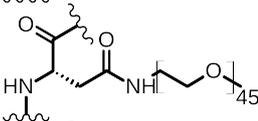
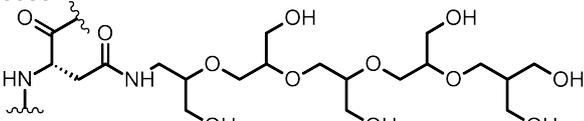
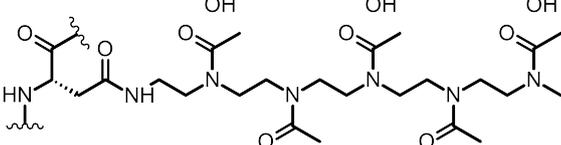
Mit dem CNA-Programm kann die Entfaltungstemperatur über einen sigmoidalen oder Spline-Fit bestimmt werden. Hierzu eignen sich die verschiedenen globalen Parameter (die Clusterkonfigurationsentropie H , der Rigiditätsordnungsparameter P_∞ , die Floppy Mode-Dichte Φ , die durchschnittliche Rigiditätsclustergröße S) zur Detektierung unterschiedlicher Entfaltungsmomente. Die Entfaltung eines Protein erfolgt in der Regel in mehreren Schritten. So detektiert die Clusterkonfigurationsentropie vom Typ 1 (CCE1) bspw. besonders den ersten Übergang von einer rigiden zu einer flexiblen Struktur, während jene vom Typ 2 (CCE2) eine höhere Sensitivität für beim Entfaltungsprozess später ablaufende Flexibilisierungen aufweist (eine genaue Beschreibung aller Parameter ist im SI-Material von [235] zu finden). Die Entfaltungstemperatur wird somit durch den Übergang der Struktur von einem überwiegend rigiden zu einem flexiblen Netzwerk definiert. Da sich die Temperatur auf Basis einer bisherigen Studie zu 20 Proteinen aus dem Cut-off-Wert der Energie ableitet (siehe Kap. 1.4.2), sollte diese Temperatur nicht als absoluter Wert beurteilt werden, sie kann jedoch als Vergleich der Stabilität zweier vergleichbarer Systeme dienen. Zusätzlich zu diesen globalen Parametern berechnet das Programm auch lokale Parameter (Perkolationsindizes p_i , Rigiditätsindizes r_i und Stabilitätskarten rc_{ij}), die die Rigidität einzelner Aminosäuren beschreiben. Schließlich können daraus sogenannte *Weak spots* erfasst werden. Diese stellen Aminosäuren dar, die sich beim Erreichen der Entfaltungstemperatur vom größten Cluster lösen und somit potenzielle Mutationsorte zur Verbesserung der thermischen Stabilität des Makromoleküls repräsentieren. Eine Beschreibung aller berechenbarer CNA-Parameter ist in Tab. B.1 gegeben.

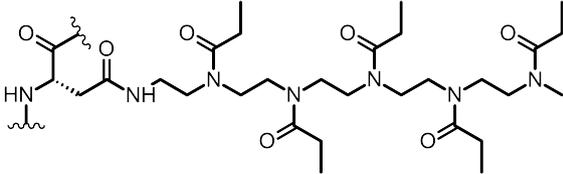
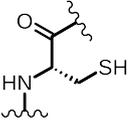
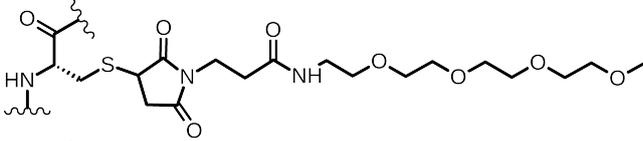
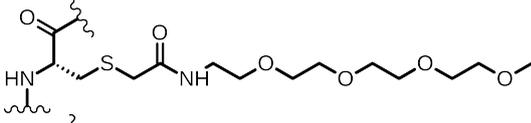
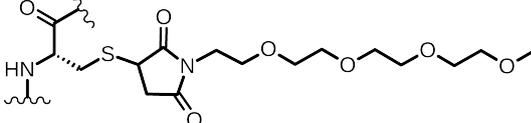
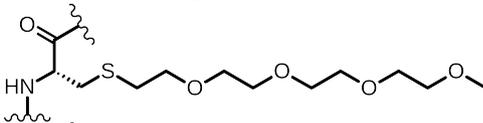
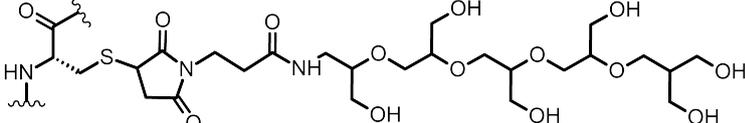
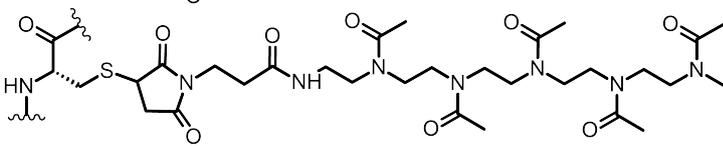
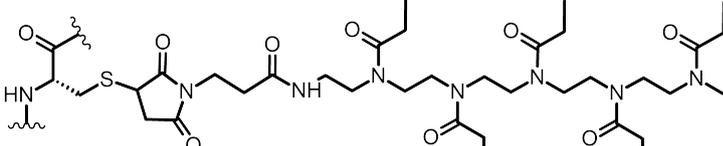
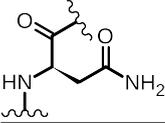
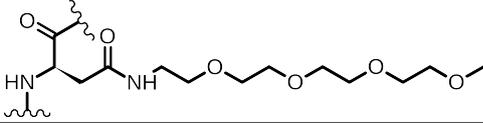
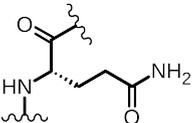
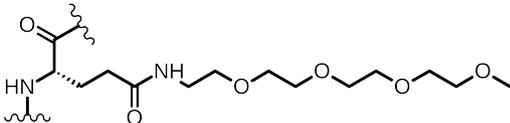
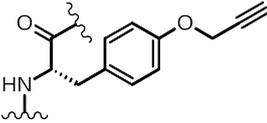
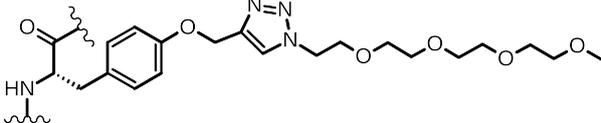
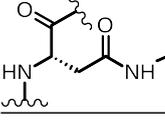
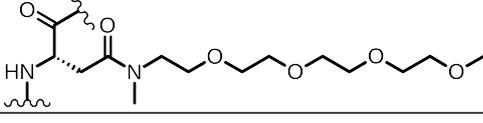
Tabelle B.1: Bei der CN-Analyse berechenbare lokale und globale Parameter, die die Rigidität des Biomakromoleküls beschreiben. Ebenso sind die vier verschiedenen Arten von Weak spots aufgeführt, die auf Basis der Phasenübergangstemperatur erfasst werden [235].

Parameter	Art	Beschreibung
Rigiditätsordnungsparameter P_∞	global	Anteil des Netzwerks, welches Teil des aktuell größten Clusters (Typ 1) bzw. des ursprünglich größten Clusters ist (Typ 2).
Clusterkonfigurationsentropie H	global	Wahrscheinlichkeit, dass sich ein Atom innerhalb eines Clusters mit einer bestimmten Größe befindet (Grad der Unordnung des Netzwerks). Bei CCE2 fließt die Clustergröße mit ihrem Quadrat in die Berechnung ein.
Durchschnittliche Rigiditätsclustergröße S	global	Durchschnittliche Clustergröße, wobei der größte Cluster nicht in die Berechnung einfließt.
Floppy Mode-Dichte Φ	global	Anzahl an Floppy Modes (Freiheitsgraden) in Relation zur Anzahl an Atomen.
Perkolationsindizes p_i	lokal	Lokales Analog zu P_∞ . Beschreibt die Wasserstoffbrückenbindungsenergie, bei der sich eine Bindung vom größten ursprünglichen Cluster (Typ 1) oder vom aktuell größten Cluster (Typ 2) löst.
Rigiditätsindizes r_i	lokal	Ähnlich wie p_i , wobei hier die Energie bestimmt wird, bei der sich die Bindung von jeglichem Cluster löst.
Seitenkettenrigiditätsindizes r_{sc}	lokal	Energie, bei der sich C_α und C_β nicht mehr im selben Cluster befinden.
Stabilitätskarten rc_{ij}	lokal	Energie, bei der sich die C_α -Atome zweier Aminosäuren nicht mehr im selben Cluster befinden.
Weak spot Typ 1	lokal	Die Aminosäure, die Teil des größten Clusters ist, wird gebräuchlich nach der Phasenübergangstemperatur flexibel.
Weak spot Typ 2	lokal	Die Aminosäure jeglichen Clusters (mind. 30 Atome) wird nach der Phasenübergangstemperatur flexibel.
Weak spot Typ 3	lokal	Aminosäuren, die Wasserstoffbrückenbindungen bilden, welche nach der Phasenübergangstemperatur gelöst werden.
Weak spot Typ 4	lokal	Aminosäuren innerhalb von 5 Å der Aminosäuren vom Weak spot Typ 3.

C. Konjugate

Tabelle C.1: Chemische Strukturformeln aller in den Modellproteinen untersuchten, konjugierten Aminosäuren. In der rechten Spalte ist jeweils ein Kennzeichen für die jeweilige Struktur gegeben, welches in Tabelle C.2 verwendet wird.

unkonjugiert	konjugiert	Kennzeichen
		1
		2
		3
		B
		J
		O
		U
		\$
		ε
		λ

		ψ
		4
		5
		6
		7
		ϕ
		γ
		ω
		8
		9
		Z
		0

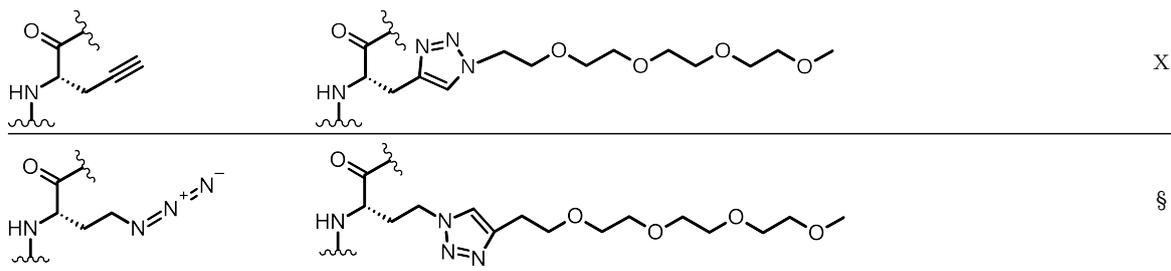


Tabelle C.2: Liste aller simulierten Konjugate der Modellproteine. In der ersten Spalte wurde jedem Biokonjugat ein Name nach einem einheitlichen Schema *AAAA.XX(YY)-Z* gegeben. Der Name beginnt mit dem vierstelligen PDB-Code *AAAA* der jeweiligen Struktur. Danach steht *XX* für die Konjugationsstelle (im Falle doppelt konjugierter Varianten ergänzt durch *+XX*). Der in manchen Fällen enthaltene Term (*YY*) beschreibt im Vergleich zum Wildtyp zusätzlich vorkommende Punktmutationen. *Z* entspricht einem Kennzeichen aus Tabelle C.1 und beschreibt die chemische Struktur der konjugierten Aminosäure. Unkonjugierte Varianten werden im Haupttext lediglich mit der Konjugationsstelle *XX* bezeichnet. Der in der jeweiligen originalen Publikation für das entsprechende Konjugat verwendete Name ist in der dritten Spalte gegeben. In der vierten Spalte ist die Aminosäuresequenz gegeben, wobei die konjugierte Aminosäure mit dem Kennzeichen aus C.1 in Blau dargestellt ist. Etwaige zusätzliche Mutationen (*YY*) sind orange gefärbt. Die jeweils experimentell ermittelte freie Faltungsenergie ist in der fünften Spalte aufgeführt. Die sechste Spalte *R* gibt die Referenzpublikation an. In der letzten Spalte *A* ist gekennzeichnet, für welchen QSPR-Modellansatz das betreffende Konjugat verwendet wurde (1, 2 oder 1+2). Varianten, welche aus keiner Referenzpublikation stammen, enthalten die Kennzeichnungen *NA* (*not available*). Terminale Acetylgruppen und Amide sind jeweils mit *A-* bzw. *-N* in der Sequenz gegeben. Die Sequenz ist für Di- und Trimere jeweils nur für ein Monomer dargestellt. Der Buchstabe χ symbolisiert eine Cyclohexylalanin-Mutation, ν eine Naphthylalanin-Mutation.

Nr.	Name	Name in Referenz	Sequenz	$\Delta\Delta G_f^0$ [kcal/mol]	R	A
1	1PIN.19-B	6PEG ₁	KLPPGW EK R M S R S B GRVYYFNHITNASQFERPSG	-0.34 ± 0.04	[97]	1+2
2	1PIN.19-J	6PEG ₂	KLPPGW EK R M S R S J GRVYYFNHITNASQFERPSG	-0.37 ± 0.05	[97]	1+2
3	1PIN.19-O	6PEG ₃	KLPPGW EK R M S R S O GRVYYFNHITNASQFERPSG	-0.72 ± 0.07	[97]	1+2
4	1PIN.19-U	6PEG ₈	KLPPGW EK R M S R S U GRVYYFNHITNASQFERPSG	-0.61 ± 0.04	[97]	1+2
5	1PIN.14-1	14p	KLPPGW EK 1 M S R S SGRVYYFNHITNASQFERPSG	0.00 ± 0.40	[100]	1+2
6	1PIN.16-1	16p	KLPPGW EK R M 1 R S SGRVYYFNHITNASQFERPSG	-0.90 ± 0.03	[100]	1+2
7	1PIN.17-1	17p	KLPPGW EK R M S 1 S SGRVYYFNHITNASQFERPSG	-0.18 ± 0.05	[100]	1+2
8	1PIN.18-1	18p	KLPPGW EK R M S R 1 S GRVYYFNHITNASQFERPSG	0.00 ± 0.07	[100]	1+2
9	1PIN.19-1	19p	KLPPGW EK R M S R S 1 GRVYYFNHITNASQFERPSG	-0.70 ± 0.04	[100]	1+2
10	1PIN.23-1	23p	KLPPGW EK R M S R S SGRV 1 YFNHITNASQFERPSG	0.40 ± 0.10	[100]	1+2
11	1PIN.26-1	26p	KLPPGW EK R M S R S SGRVYYF 1 HITNASQFERPSG	-0.58 ± 0.06	[100]	1+2
12	1PIN.27-1	27p	KLPPGW EK R M S R S SGRVYYF 1 ITNASQFERPSG	0.38 ± 0.04	[100]	1+2
13	1PIN.28-1	28p	KLPPGW EK R M S R S SGRVYYFNH 1 ITNASQFERPSG	0.00 ± 0.07	[100]	1+2
14	1PIN.29-1	29p	KLPPGW EK R M S R S SGRVYYFNH 1 ITNASQFERPSG	-0.36 ± 0.04	[100]	1+2
15	1PIN.30-1	30p	KLPPGW EK R M S R S SGRVYYFNHIT 1 ASQFERPSG	0.00 ± 0.07	[100]	1+2
16	1PIN.32-1	32p	KLPPGW EK R M S R S SGRVYYFNHITN 1 AQFERPSG	-0.45 ± 0.02	[100]	1+2
17	1PIN.19-8	D-19p	KLPPGW EK R M S R S 8 GRVYYFNHITNASQFERPSG	0.01 ± 0.04	[100]	1+2
18	1PIN.19(S16A)-1	19p-S16A	KLPPGW EK R M A 1 S R VYYFNHITNASQFERPSG	-0.51 ± 0.02	[100]	1+2
19	1PIN.19(Y23F)-1	19p-Y23F	KLPPGW EK R M S R S 1 GRV F YFNHITNASQFERPSG	-0.43 ± 0.03	[100]	1+2
20	1PIN.19(S32A)-1	19p-S32A	KLPPGW EK R M S R S 1 GRVYYFNHITN A QFERPSG	-0.71 ± 0.03	[100]	1+2
21	1PIN.19(S16A:Y23F)-1	19p-S16A,Y23F	KLPPGW EK R M A 1 S R V F YFNHITNASQFERPSG	-0.72 ± 0.08	[100]	1+2

22	1PIN.16(Y23F)-1	16p-Y23F	KLPPGWEKRM1RSSGRVFFYNHITNASQFERPSG	-0.45 ± 0.06	[100]	1+2
23	1PIN.16(S32A)-1	16p-S32A	KLPPGWEKRM1RSSGRVFFYNHITNAQFERPSG	-0.58 ± 0.03	[100]	1+2
24	1PIN.26(T29A)-1	26p-T29A	KLPPGWEKRMSRSSGRVFF1HITANASQFERPSG	-0.32 ± 0.06	[100]	1+2
25	1PIN.29(S32A)-1	29p-S32A	KLPPGWEKRMSRSSGRVFFNH1INAQFERPSG	-0.88 ± 0.06	[100]	1+2
26	1PIN.32(Y23F)-1	32p-Y23F	KLPPGWEKRMSRSSGRVFFYNHITNA1QFERPSG	-0.61 ± 0.11	[100]	1+2
27	1PIN.16+26-1	16p/26p	KLPPGWEKRM1RSSGRVFF1HITNASQFERPSG	-1.38 ± 0.03	[100]	1+2
28	1PIN.19+26-1	19p/26p	KLPPGWEKRMSRS1GRVFF1HITNASQFERPSG	-1.26 ± 0.02	[100]	1+2
29	1PIN.26+29-1	26p/29p	KLPPGWEKRMSRSSGRVFF1HI1NASQFERPSG	-0.56 ± 0.04	[100]	1+2
30	1PIN.16+19-1	16p/19p	KLPPGWEKRM1RS1GRVFFYNHITNASQFERPSG	-0.61 ± 0.11	[100]	1+2
31	1PIN.16-§	16p45	KLPPGWEKRM§RSSGRVFFYNHITNASQFERPSG	-0.39 ± 0.03	[100]	1+2
32	1PIN.18-§	18p45	KLPPGWEKRMSR§SSGRVFFYNHITNASQFERPSG	0.02 ± 0.08	[100]	1+2
33	1PIN.19-§	19p45	KLPPGWEKRMSR§SSGRVFFYNHITNASQFERPSG	-0.67 ± 0.05	[100]	1+2
34	1PIN.26-§	26p45	KLPPGWEKRMSRSSGRVFF§HITNASQFERPSG	-0.27 ± 0.06	[100]	1+2
35	1PIN.27-§	27p45	KLPPGWEKRMSRSSGRVFFYN§ITNASQFERPSG	0.65 ± 0.05	[100]	1+2
36	1PIN.28-§	28p45	KLPPGWEKRMSRSSGRVFFYNH§TNASQFERPSG	0.36 ± 0.05	[100]	1+2
37	1PIN.29-§	29p45	KLPPGWEKRMSRSSGRVFFYNHIT§NASQFERPSG	-0.36 ± 0.04	[100]	1+2
38	1PIN.19-2	1-nbp	KLPPGWEKRMSRS2GRVFFYNHITNASQFERPSG	-1.19 ± 0.02	[101]	1+2
39	1PIN.19-3	1-cbp	KLPPGWEKRMSRS3GRVFFYNHITNASQFERPSG	-1.13 ± 0.04	[101]	1+2
40	1PIN.19-0	m1p	KLPPGWEKRMSRS0GRVFFYNHITNASQFERPSG	-0.30 ± 0.01	[101]	1+2
41	1PIN.19-7	2p-1	KLPPGWEKRMSRS7GRVFFYNHITNASQFERPSG	-0.16 ± 0.02	[101]	1+2
42	1PIN.19-6	2p-2	KLPPGWEKRMSRS6GRVFFYNHITNASQFERPSG	-0.08 ± 0.02	[101]	1+2
43	1PIN.19-5	2p-3	KLPPGWEKRMSRS5GRVFFYNHITNASQFERPSG	0.01 ± 0.02	[101]	1+2
44	1PIN.19-4	2p-4	KLPPGWEKRMSRS4GRVFFYNHITNASQFERPSG	0.24 ± 0.02	[101]	1+2
45	1PIN.19-Z	4p	KLPPGWEKRMSRSZGRVFFYNHITNASQFERPSG	-0.04 ± 0.02	[101]	1+2
46	1PIN.19-9	5p	KLPPGWEKRMSRS9GRVFFYNHITNASQFERPSG	-0.37 ± 0.01	[101]	1+2
47	1PIN.19-X	6p	KLPPGWEKRMSRSXGRVFFYNHITNASQFERPSG	-0.59 ± 0.03	[101]	1+2
48	1PIN.19-§	3p	KLPPGWEKRMSRS§GRVFFYNHITNASQFERPSG	-0.36 ± 0.01	[101]	2
49	1PIN.18(S16D:R17A)-1	pβ18	KLPPGWEKRMDA1SGRVFFYNHITNASQFERPSG	-0.25 ± 0.02	[102]	1+2
50	1PIN.18(S16D:R17A:R21A)-1	pβ18-DA	KLPPGWEKRMDA1SGAVVFFYNHITNASQFERPSG	-0.28 ± 0.01	[102]	1+2
51	1PIN.18(R17A)-1	pβ18-SR	KLPPGWEKRMSA1SGRVFFYNHITNASQFERPSG	-0.51 ± 0.01	[102]	1+2
52	1PIN.18(R17A:R21A)-1	pβ18-SA	KLPPGWEKRMSA1SGAVVFFYNHITNASQFERPSG	-0.57 ± 0.02	[102]	1+2
53	1PIN.23(R14A)-Z	pβ23-EA	KLPPGWEKAMRSSGRVZYFNHITNASQFERPSG	-0.11 ± 0.02	[102]	1+2
54	1PIN.23(E12A)-Z	pβ23-AR	KLPPGWAKMRSSGRVZYFNHITNASQFERPSG	-0.05 ± 0.02	[102]	1+2
55	1PIN.23(E12A:R14A)-Z	pβ23-AA	KLPPGWAKAMRSSGRVZYFNHITNASQFERPSG	-0.30 ± 0.02	[102]	1+2
56	1PIN.23(E12χ:R14χ)-Z	pXX	KLPPGWχKχMRSSGRVZYFNHITNASQFERPSG	-1.30 ± 0.02	[102]	2
57	1PIN.23(E12χ:R14A)-Z	pXA	KLPPGWχKAMRSSGRVZYFNHITNASQFERPSG	-0.40 ± 0.02	[102]	2
58	1PIN.23(E12χ:R14L)-Z	pXL	KLPPGWχKLMRSSGRVZYFNHITNASQFERPSG	-0.90 ± 0.02	[102]	2
59	1PIN.23(E12A:R14χ)-Z	pAX	KLPPGWAKχMRSSGRVZYFNHITNASQFERPSG	-0.51 ± 0.02	[102]	2
60	1PIN.23(E12L:R14L)-Z	pLL	KLPPGWLKLMRSSGRVZYFNHITNASQFERPSG	-0.89 ± 0.04	[103]	1+2
61	1PIN.23(E12L:R14A)-Z	pLA	KLPPGWLKAMRSSGRVZYFNHITNASQFERPSG	-0.31 ± 0.02	[103]	1+2
62	1PIN.23(E12A:R14L)-Z	pAL	KLPPGWAKLMRSSGRVZYFNHITNASQFERPSG	-0.56 ± 0.04	[103]	1+2
63	1PIN.23(E12F:R14F)-Z	pFF	KLPPGWFKFMRSSGRVZYFNHITNASQFERPSG	-0.33 ± 0.02	[103]	1+2
64	1PIN.23(E12F:R14A)-Z	pFA	KLPPGWFKAMRSSGRVZYFNHITNASQFERPSG	-0.36 ± 0.02	[103]	1+2
65	1PIN.23(E12A:R14F)-Z	pAF	KLPPGWAKFMRSSGRVZYFNHITNASQFERPSG	-0.37 ± 0.02	[103]	1+2
66	1PIN.23(E12F:R14L)-Z	pFL	KLPPGWFKLMRSSGRVZYFNHITNASQFERPSG	-0.68 ± 0.02	[103]	1+2
67	1PIN.23(E12L:R14F)-Z	pLF	KLPPGWLFKMRSSGRVZYFNHITNASQFERPSG	-0.56 ± 0.02	[103]	1+2
68	1PIN.16-9	16Qp	KLPPGWEKRM9RSSGRVFFYNHITNASQFERPSG	-0.67 ± 0.02	[118]	1+2
69	1PIN.16-Z	16Zp	KLPPGWEKRMZRSSGRVFFYNHITNASQFERPSG	0.07 ± 0.05	[118]	1+2
70	1PIN.16-2	16Nbp	KLPPGWEKRM2RSSGRVFFYNHITNASQFERPSG	-1.70 ± 0.03	[118]	1+2

71	1PIN.16-§	16Xp	KLPPGWEKRM§RSSGRVYFNHITNASQFERPSG	-0.39 ± 0.04	[118]	2
72	1PIN.18-9	18Qp	KLPPGWEKMSR§9SGRVYFNHITNASQFERPSG	-0.12 ± 0.02	[118]	1+2
73	1PIN.18-Z	18Zp	KLPPGWEKMSRZ2SGRVYFNHITNASQFERPSG	-0.05 ± 0.04	[118]	1+2
74	1PIN.18-2	18Nbp	KLPPGWEKMSR2§SGRVYFNHITNASQFERPSG	0.03 ± 0.01	[118]	1+2
75	1PIN.18-§	18Xp	KLPPGWEKMSR§§SGRVYFNHITNASQFERPSG	0.00 ± 0.02	[118]	2
76	1PIN.23-9	23Qp	KLPPGWEKMSR§§9YFNHITNASQFERPSG	0.23 ± 0.04	[118]	1+2
77	1PIN.23-Z	23Zp	KLPPGWEKMSR§§VZVFNHITNASQFERPSG	-0.29 ± 0.03	[118]	1+2
78	1PIN.23-2	23Nbp	KLPPGWEKMSR§§V2VFNHITNASQFERPSG	0.18 ± 0.10	[118]	1+2
79	1PIN.23-§	23Xp	KLPPGWEKMSR§§§YFNHITNASQFERPSG	0.43 ± 0.16	[118]	2
80	1PIN.27-9	27Qp	KLPPGWEKMSR§§§VYFN9ITNASQFERPSG	0.27 ± 0.02	[118]	1+2
81	1PIN.27-Z	27Zp	KLPPGWEKMSR§§§VYFN2ITNASQFERPSG	-0.62 ± 0.03	[118]	1+2
82	1PIN.27-2	27Nbp	KLPPGWEKMSR§§§VYFN2ITNASQFERPSG	1.26 ± 0.06	[118]	1+2
83	1PIN.27-§	27Xp	KLPPGWEKMSR§§§VYFN§ITNASQFERPSG	0.68 ± 0.04	[118]	2
84	1PIN.29-9	29Qp	KLPPGWEKMSR§§§VYFNHI9NASQFERPSG	0.16 ± 0.06	[118]	1+2
85	1PIN.29-Z	29Zp	KLPPGWEKMSR§§§VYFNHI2NASQFERPSG	-0.17 ± 0.06	[118]	1+2
86	1PIN.29-2	29Nbp	KLPPGWEKMSR§§§VYFNHI2NASQFERPSG	-0.67 ± 0.02	[118]	1+2
87	1PIN.29-§	29Xp	KLPPGWEKMSR§§§VYFNHI§NASQFERPSG	0.32 ± 0.07	[118]	2
88	1PIN.32-9	32Qp	KLPPGWEKMSR§§§VYFNHI9QFERPSG	-0.44 ± 0.02	[118]	1+2
89	1PIN.32-Z	32Zp	KLPPGWEKMSR§§§VYFNHI2AZQFERPSG	-0.33 ± 0.03	[118]	1+2
90	1PIN.32-2	32Nbp	KLPPGWEKMSR§§§VYFNHI2AZQFERPSG	-0.93 ± 0.03	[118]	1+2
91	1PIN.32-§	32Xp	KLPPGWEKMSR§§§VYFNHI2ASQFERPSG	0.01 ± 0.03	[118]	2
92	1PIN.19(S16F:R21T)-1	6PEG ₄ -F,T	KLPPGWEKRMFR§1GTVVYFNHITNASQFERPSG	-0.81 ± 0.10	[98]	1+2
93	1PIN.19(S16F)-1	6PEG ₄ -F	KLPPGWEKRMFR§1GRVYFNHITNASQFERPSG	-1.13 ± 0.07	[98]	1+2
94	1PIN.19(R21T)-1	6PEG ₄ -T	KLPPGWEKMSR§1GTVVYFNHITNASQFERPSG	-0.55 ± 0.06	[98]	1+2
95	1PIN.19(R21A)-1	6PEG ₄ -A	KLPPGWEKMSR§1GAVVYFNHITNASQFERPSG	-0.96 ± 0.12	[98]	1+2
96	1PIN.19(R21L)-1	6PEG ₄ -L	KLPPGWEKMSR§1GLVVYFNHITNASQFERPSG	-0.68 ± 0.11	[98]	1+2
97	1SRL.20-1	SH3T20NPEG	TFVALYDYESR1ETDLSFKKGERLQIVNTEGDW WLAHSLTGTGYIPSNVYVAPS	-1.20 ± 0.10	[100]	2
98	2ZTA.1-1	p2 α 1	A-1MKQLEDKVEELLSKNYHLENEVARLKKLVG-N	0.16 ± 0.04	[102]	2
99	2ZTA.3-1	p2 α 3	A-RM1QLEDKVEELLSKNYHLENEVARLKKLVG-N	0.24 ± 0.04	[102]	2
100	2ZTA.4-1	p2 α 4	A-RMK1LEDKVEELLSKNYHLENEVARLKKLVG-N	0.70 ± 0.04	[102]	2
101	2ZTA.6-1	p2 α 6	A-RMKQL1DKVEELLSKNYHLENEVARLKKLVG-N	0.46 ± 0.02	[102]	2
102	2ZTA.7-1	p2 α 7	A-RMKQLE1KVEELLSKNYHLENEVARLKKLVG-N	0.72 ± 0.02	[102]	2
103	2ZTA.10-1	p2 α 10	A-RMKQLEDKV1ELLSKNYHLENEVARLKKLVG-N	1.34 ± 0.04	[102]	2
104	2ZTA.14-1	p2 α 14	A-RMKQLEDKVEELL1KNYHLENEVARLKKLVG-N	-0.14 ± 0.06	[102]	2
105	2ZTA.18-1	p2 α 18	A-RMKQLEDKVEELLSKNY1LENEVARLKKLVG-N	-0.22 ± 0.02	[102]	2
106	2ZTA.21-1	p2 α 21	A-RMKQLEDKVEELLSKNYHLE1EVARLKKLVG-N	0.48 ± 0.02	[102]	2
107	2ZTA.25-1	p2 α 25	A-RMKQLEDKVEELLSKNYHLENEVA1LKKLVG-N	0.34 ± 0.02	[102]	2
108	2ZTA.28-1	p2 α 28	A-RMKQLEDKVEELLSKNYHLENEVARL1LVG-N	-0.26 ± 0.02	[102]	2
109	2ZTA.4(E11A)-1	p2 α 4-KA	A-RMK1LEDKVEALLSKNYHLENEVARLKKLVG-N	-0.04 ± 0.01	[102]	2
110	2ZTA.4(K8A)-1	p2 α 4-AE	A-RMK1LEDAVEELLSKNYHLENEVARLKKLVG-N	-0.12 ± 0.02	[102]	2
111	2ZTA.4(K8A:E11A)-1	p2 α 4-AA	A-RMK1LEDAVEALLSKNYHLENEVARLKKLVG-N	-0.70 ± 0.02	[102]	2
112	2ZTA.18(R25A)-1	p2 α 18-EA	A-RMKQLEDKVEELLSKNY1LENEVAALLKKLVG-N	0.56 ± 0.02	[102]	2
113	2ZTA.18(E22A)-1	p2 α 18-AR	A-RMKQLEDKVEELLSKNY1LENAVARLKKLVG-N	0.28 ± 0.02	[102]	2
114	2ZTA.18(E22A:R25A)-1	p2 α 18-AA	A-RMKQLEDKVEELLSKNY1LENAVALKKLVG-N	-0.32 ± 0.01	[102]	2
115	1COI.1-1	p3 α 1	Ace-9VEALEKKVALESKVQALEKKVEALEY-N	0.48 ± 0.01	[102]	2
116	1COI.1(E3A)-1	p3 α 1-AK	A-9VALEKKVALESKVQALEKKVEALEY-N	0.45 ± 0.01	[102]	2
117	1COI.1(K7A)-1	p3 α 1-EA	A-9VEALEAKVALESKVQALEKKVEALEY-N	0.51 ± 0.01	[102]	2
118	1COI.1(E3A:K7A)-1	p3 α 1-AA	A-9VALEAKVALESKVQALEKKVEALEY-N	0.33 ± 0.01	[102]	2

119	1COI.6-1	p3 α 6	A-EVEAL9KKVAALLESKVQALEKKVEALEY-N	0.96 \pm 0.01	[102]	2
120	1COI.6(K8A)-1	p3 α 6-AE	A-EVEAL9KAVAALLESKVQALEKKVEALEY-N	0.42 \pm 0.01	[102]	2
121	1COI.6(E13A)-1	p3 α 6-KA	A-EVEAL9KKVAALASKVQALEKKVEALEY-N	0.39 \pm 0.01	[102]	2
122	1COI.6(K8A:E13A)-1	p3 α 6-AA	A-EVEAL9KAVAALASKVQALEKKVEALEY-N	-0.51 \pm 0.01	[102]	2
123	5UXT.14-1	Np	A-EVEALEKKVEALE1KVQKLEKKVEALEHGWDGR-N	0.50 \pm 0.02	[264]	
124	5UXT.14(E10A)-1	Np1	A-EVEALEKKVALE1KVQKLEKKVEALEHGWDGR-N	-0.24 \pm 0.02	[264]	
125	5UXT.14(K18A)-1	Np2	A-EVEALEKKVEALE1KVQALEKKVEALEHGWDGR-N	0.22 \pm 0.02	[264]	
126	5UXT.14(E10A:K18A)-1	Np3	A-EVEALEKKVALE1KVQALEKKVEALEHGWDGR-N	-0.62 \pm 0.02	[264]	
127	1PIN.26(W11F)-1	FNp	KLPPGF ^e EKRMSRSSGRVYF1HITNASQFERPSG	-0.34 \pm 0.02	[104]	
128	1PIN.26-(W11 ν)-1	ZNp	KLPPG ^v EKRMSRSSGRVYF1HITNASQFERPSG	-0.62 \pm 0.03	[104]	
130	1PIN.26(W11Y)-1	YNp	KLPPGY ^e EKRMSRSSGRVYF1HITNASQFERPSG	-0.47 \pm 0.05	[104]	
131	1PIN.26-9	WQp	KLPPGW ^e EKRMSRSSGRVYF9HITNASQFERPSG	0.14 \pm 0.02	[104]	
132	1PIN.7-1	NA	K1PPGWEKRMSRSSGRVYFNHITNASQFERPSG	NA	NA	
133	1PIN.11-1	NA	KLPPG1 ^e KRMSRSSGRVYFNHITNASQFERPSG	NA	NA	
134	1PIN.12-1	NA	KLPPGW1 ^e KRMSRSSGRVYFNHITNASQFERPSG	NA	NA	
135	1PIN.13-1	NA	KLPPGWE1 ^e RSMSRSSGRVYFNHITNASQFERPSG	NA	NA	
136	1PIN.15-1	NA	KLPPGWEKRS1 ^e RSMSRSSGRVYFNHITNASQFERPSG	NA	NA	
137	1PIN.21-1	NA	KLPPGWEKRMSRSSG1 ^e VYFNHITNASQFERPSG	NA	NA	
138	1PIN.22-1	NA	KLPPGWEKRMSRSSG1 ^e YFNHITNASQFERPSG	NA	NA	
139	1PIN.24-1	NA	KLPPGWEKRMSRSSGRVY1 ^e FNHITNASQFERPSG	NA	NA	
140	1PIN.25-1	NA	KLPPGWEKRMSRSSGRVY1 ^e NHITNASQFERPSG	NA	NA	
141	1PIN.31-1	NA	KLPPGWEKRMSRSSGRVYFNHITN1 ^e SQFERPSG	NA	NA	
142	1PIN.33-1	NA	KLPPGWEKRMSRSSGRVYFNHITNAS1 ^e FERPSG	NA	NA	
143	1PIN.35-1	NA	KLPPGWEKRMSRSSGRVYFNHITNASQF1 ^e RPSG	NA	NA	
144	1PIN.14- ϵ	NA	KLPPGWEK ^e MSRSSGRVYFNHITNASQFERPSG	NA	NA	
145	1PIN.16- ϵ	NA	KLPPGWEKR ^e MSRSSGRVYFNHITNASQFERPSG	NA	NA	
146	1PIN.17- ϵ	NA	KLPPGWEKRMS ^e SSGRVYFNHITNASQFERPSG	NA	NA	
147	1PIN.18- ϵ	NA	KLPPGWEKRMSR ^e SGRVYFNHITNASQFERPSG	NA	NA	
148	1PIN.19- ϵ	NA	KLPPGWEKRMSR ^e GRVYFNHITNASQFERPSG	NA	NA	
149	1PIN.21- ϵ	NA	KLPPGWEKRMSRSSG ^e VYFNHITNASQFERPSG	NA	NA	
150	1PIN.23- ϵ	NA	KLPPGWEKRMSRSSGRV ^e YFNHITNASQFERPSG	NA	NA	
151	1PIN.26- ϵ	NA	KLPPGWEKRMSRSSGRVYF ^e HITNASQFERPSG	NA	NA	
152	1PIN.27- ϵ	NA	KLPPGWEKRMSRSSGRVYFN ^e ITNASQFERPSG	NA	NA	
153	1PIN.28- ϵ	NA	KLPPGWEKRMSRSSGRVYFNH ^e TNASQFERPSG	NA	NA	
154	1PIN.29- ϵ	NA	KLPPGWEKRMSRSSGRVYFNHIT ^e NASQFERPSG	NA	NA	
155	1PIN.30- ϵ	NA	KLPPGWEKRMSRSSGRVYFNHIT ^e ASQFERPSG	NA	NA	
156	1PIN.32- ϵ	NA	KLPPGWEKRMSRSSGRVYFNHITNA ^e QFERPSG	NA	NA	
157	1PIN.14- λ	NA	KLPPGWEK ^{\lambda} MSRSSGRVYFNHITNASQFERPSG	NA	NA	
158	1PIN.16- λ	NA	KLPPGWEKRM ^{\lambda} RSRSSGRVYFNHITNASQFERPSG	NA	NA	
159	1PIN.17- λ	NA	KLPPGWEKRMS ^{\lambda} SSGRVYFNHITNASQFERPSG	NA	NA	
160	1PIN.18- λ	NA	KLPPGWEKRMSR ^{\lambda} SGRVYFNHITNASQFERPSG	NA	NA	
161	1PIN.19- λ	NA	KLPPGWEKRMSRS ^{\lambda} GRVYFNHITNASQFERPSG	NA	NA	
162	1PIN.21- λ	NA	KLPPGWEKRMSRSSG ^{\lambda} VYFNHITNASQFERPSG	NA	NA	
163	1PIN.23- λ	NA	KLPPGWEKRMSRSSGRV ^{\lambda} YFNHITNASQFERPSG	NA	NA	
164	1PIN.26- λ	NA	KLPPGWEKRMSRSSGRVYF ^{\lambda} HITNASQFERPSG	NA	NA	
165	1PIN.27- λ	NA	KLPPGWEKRMSRSSGRVYFN ^{\lambda} ITNASQFERPSG	NA	NA	
166	1PIN.28- λ	NA	KLPPGWEKRMSRSSGRVYFNH ^{\lambda} TNASQFERPSG	NA	NA	
167	1PIN.29- λ	NA	KLPPGWEKRMSRSSGRVYFNHIT ^{\lambda} NASQFERPSG	NA	NA	
168	1PIN.30- λ	NA	KLPPGWEKRMSRSSGRVYFNHIT ^{\lambda} ASQFERPSG	NA	NA	

169	1PIN.32- λ	NA	KLPPGWEKMRSSGRVYFNHITNA λ QFERPSG	NA	NA
170	1PIN.14- ψ	NA	KLPPGWEK ψ MSRSSGRVYFNHITNASQFERPSG	NA	NA
171	1PIN.16- ψ	NA	KLPPGWEKRM ψ RSSGRVYFNHITNASQFERPSG	NA	NA
172	1PIN.17- ψ	NA	KLPPGWEKRM ψ SSGRVYFNHITNASQFERPSG	NA	NA
173	1PIN.18- ψ	NA	KLPPGWEKMR ψ SGRVYFNHITNASQFERPSG	NA	NA
174	1PIN.19- ψ	NA	KLPPGWEKMR ψ SRVYFNHITNASQFERPSG	NA	NA
175	1PIN.21- ψ	NA	KLPPGWEKMRSSG ψ VYFNHITNASQFERPSG	NA	NA
176	1PIN.23- ψ	NA	KLPPGWEKMRSSGRV ψ YFNHITNASQFERPSG	NA	NA
177	1PIN.26- ψ	NA	KLPPGWEKMRSSGRVYF ψ HITNASQFERPSG	NA	NA
178	1PIN.27- ψ	NA	KLPPGWEKMRSSGRVYFN ψ ITNASQFERPSG	NA	NA
179	1PIN.28- ψ	NA	KLPPGWEKMRSSGRVYFNH ψ TNASQFERPSG	NA	NA
180	1PIN.29- ψ	NA	KLPPGWEKMRSSGRVYFNHI ψ NASQFERPSG	NA	NA
181	1PIN.30- ψ	NA	KLPPGWEKMRSSGRVYFNHIT ψ ASQFERPSG	NA	NA
182	1PIN.32- ψ	NA	KLPPGWEKMRSSGRVYFNHITNA ψ QFERPSG	NA	NA
183	2KZJ.7-4	NA	VDNKFN ϕ EMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
184	2KZJ.8-4	NA	VDNKFN ϕ MRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
185	2KZJ.15-4	NA	VDNKFNKEMRNAYW ϕ IALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
186	2KZJ.21-4	NA	VDNKFNKEMRNAYWEIALLP ϕ LNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
187	2KZJ.23-4	NA	VDNKFNKEMRNAYWEIALLPNL ϕ NQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
188	2KZJ.25-4	NA	VDNKFNKEMRNAYWEIALLPNLN ϕ QKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
189	2KZJ.29-4	NA	VDNKFNKEMRNAYWEIALLPNLNNQK ϕ R4FIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
190	2KZJ.33-4	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIR ϕ 4L YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
191	2KZJ.39-4	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDP ϕ 4QSANLLAEAKKLNDQAQPK	NA	NA
192	2KZJ.43-4	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSA ϕ 4LLAEAKKLNDQAQPK	NA	NA
193	2KZJ.46-4	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLL ϕ 4EAKKLNDQAQPK	NA	NA
194	2KZJ.47-4	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLA ϕ 4AKKLNDQAQPK	NA	NA
195	2KZJ.49-4	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAE ϕ 4KLNDAQPK	NA	NA
196	2KZJ.50-4	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAK ϕ 4LNDQAQPK	NA	NA
197	2KZJ.53-4	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKLN ϕ 4AQPK	NA	NA
198	2KZJ.7- ϕ	NA	VDNKFN ϕ EMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
199	2KZJ.8- ϕ	NA	VDNKFN ϕ MRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
200	2KZJ.15- ϕ	NA	VDNKFNKEMRNAYW ϕ IALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
201	2KZJ.21- ϕ	NA	VDNKFNKEMRNAYWEIALLP ϕ LNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
202	2KZJ.23- ϕ	NA	VDNKFNKEMRNAYWEIALLPNL ϕ NQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
203	2KZJ.25- ϕ	NA	VDNKFNKEMRNAYWEIALLPNLN ϕ QKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
204	2KZJ.29- ϕ	NA	VDNKFNKEMRNAYWEIALLPNLNNQK ϕ R4FIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA

205	2KZJ.33- ϕ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIR ϕ L YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
206	2KZJ.39- ϕ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDP ϕ QSANLLAEAKKLNDQAQPK	NA	NA
207	2KZJ.43- ϕ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSA ϕ LLAEAKKLNDQAQPK	NA	NA
208	2KZJ.46- ϕ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLL ϕ EAKKLNDQAQPK	NA	NA
209	2KZJ.47- ϕ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLA ϕ AKKLNDQAQPK	NA	NA
210	2KZJ.49- ϕ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAE ϕ KLNDQAQPK	NA	NA
211	2KZJ.50- ϕ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAK ϕ LNDQAQPK	NA	NA
212	2KZJ.53- ϕ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLN ϕ AQAPK	NA	NA
213	2KZJ.7- γ	NA	VDNKFN γ EMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
214	2KZJ.8- γ	NA	VDNKFNK γ MRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
215	2KZJ.15- γ	NA	VDNKFNKEMRNAYW γ IALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
216	2KZJ.21- γ	NA	VDNKFNKEMRNAYWEIALLP γ LNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
217	2KZJ.23- γ	NA	VDNKFNKEMRNAYWEIALLPNL γ NQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
218	2KZJ.25- γ	NA	VDNKFNKEMRNAYWEIALLPNLNN γ QKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
219	2KZJ.29- γ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRFIR γ FIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
220	2KZJ.33- γ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIR γ L YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
221	2KZJ.39- γ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDP γ QSANLLAEAKKLNDQAQPK	NA	NA
222	2KZJ.43- γ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSA γ LLAEAKKLNDQAQPK	NA	NA
223	2KZJ.46- γ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLL γ EAKKLNDQAQPK	NA	NA
224	2KZJ.47- γ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLA γ AKKLNDQAQPK	NA	NA
225	2KZJ.49- γ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAE γ KLNDQAQPK	NA	NA
226	2KZJ.50- γ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAK γ LNDQAQPK	NA	NA
227	2KZJ.53- γ	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLN γ AQAPK	NA	NA
228	2KZJ.7- ω	NA	VDNKFN ω EMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
229	2KZJ.8- ω	NA	VDNKFNK ω MRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
230	2KZJ.15- ω	NA	VDNKFNKEMRNAYW ω IALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
231	2KZJ.21- ω	NA	VDNKFNKEMRNAYWEIALLP ω LNNQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
232	2KZJ.23- ω	NA	VDNKFNKEMRNAYWEIALLPNL ω NQKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
233	2KZJ.25- ω	NA	VDNKFNKEMRNAYWEIALLPNLNN ω QKRAFIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
234	2KZJ.29- ω	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRFIR ω FIRSL YDDPSQSANLLAEAKKLNDQAQPK	NA	NA
235	2KZJ.33- ω	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIR ω L YDDPSQSANLLAEAKKLNDQAQPK	NA	NA

236	2KZJ.39- ω	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDP ω QSANLLAEAKKLNDQAQPK	NA	NA
237	2KZJ.43- ω	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSA ω LLAEAKKLNDQAQPK	NA	NA
238	2KZJ.46- ω	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLL ω EAKKLNDQAQPK	NA	NA
239	2KZJ.47- ω	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLA ω AKKLNDQAQPK	NA	NA
240	2KZJ.49- ω	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAE ω KLNDQAQPK	NA	NA
241	2KZJ.50- ω	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAK ω LNDQAQPK	NA	NA
242	2KZJ.53- ω	NA	VDNKFNKEMRNAYWEIALLPNLNNQKRAFIRSL YDDPSQSANLLAEAKKL ω AQAPK	NA	NA

D. Vorhersagemodelle

D.1 Berechnete Deskriptoren

Tabelle D.1: Übersicht über die untersuchten Deskriptoren. In der ersten Spalte ist der jeweilige bei der Modellgenerierung verwendete Name dargestellt (bei zwei verschiedenen Namen wurde der jeweilige Deskriptor im zweiten Modellansatz umbenannt). Die zweite Spalte stellt jeweils eine Beschreibung des Deskriptors bereit. Die Variablen wurden für alle einzelnen Aminosäuren ($Rs = residues$), für die unterschiedlichen Proteinbereiche ($Rg = regional$) oder auch das gesamte Protein ($Pr = protein$) berechnet. Variablen, die für die fünf unterschiedlichen Regionen kalkuliert wurden, wurde der Suffix 1 für $\beta 1$, 2 für $\beta 2$, 3 für $\beta 3$, 4 für Schleife 1 und 5 für Schleife 2 zugeschrieben (z. B. beschreibt *sscontent1* den durchschnittlichen Sekundärstrukturanteil für alle Aminosäuren des ersten β -Faltblatts). In ähnlicher Weise erhalten Deskriptoren, die für alle einzelnen Aminosäuren berechnet wurden, den jeweiligen Suffix 6 – 39 (z. B. beschreibt *int14* die Interaktionsenergie des Polymers mit R14). Im Falle des Ansatzes 1 wurden, wo möglich Medianwerte berechnet, im Falle von Ansatz 2 Mittelwerte. Die letzte Spalte gibt an, bei welchem Modellansatz der jeweilige Deskriptor verwendet wurde (1, 2 oder 1+2).

Name	Beschreibung	Rs	Rg	Pr	A
bbhbonds	Anzahl an Proteinrückgrat-HB pro Momentaufnahme (Cutoff: 3.5 Å und 150°).			X	1+2
bbwathbonds	Anzahl an Proteinrückgrat-HB zu Wassermolekülen pro Momentaufnahme.			X	1+2
contacts	Anzahl an nativen Kontakten des Proteinrückgrats (Cutoff: 4.5 Å).			X	1+2
contacts2	Anzahl an nativen Kontakten aller Proteinatome (Cutoff: 4.5 Å).			X	2
crigidity	Rigide-Kontakte-Index (CNA-Parameter) für jede Aminosäure X (<i>crigidityX</i>) und Region X (<i>crigiditymeanX</i>), sowie für das ganze Protein (<i>crigiditymean</i>).	X	X	X	1+2
entropyside	Konfigurationsentropie, berechnet nach den Seitenkettenorderparametern gemäß Brüschweiler (kein Wert für nicht proteinogene Aminosäuren kalkulierbar) [376].	X	X	X	1
glycinflip	Prozent an Momentaufnahmen, in denen G20 innerhalb der ersten Schleife einen positiven ϕ -Wert zeigt, was einer Drehung des Proteinrückgrats entspricht. Dieses Phänomen wurde in den ersten Simulationen beobachtet (kein Median / Durchschnitt berechenbar).			X	1+2
int	Interaktionsenergie der Polymeratome zu jeder Aminosäure.	X			1+2
intoh	Interaktionsenergie der Polymeratome zu jedem Serin, Threonin und Tyrosin.			X	1+2

intX	Interaktionsenergie der Polymeratome zu bestimmten proteinogenen Aminosäure (X = Gly, Ser, Tyr, Phe, Asn, Glu oder Pro).	X				1+2
intX_Y	Interaktionsenergie jeder Aminosäure X zu jeder anderen Aminosäure Y im Protein.	X				1+2
lifetime / lifeavg	Durchschnittliche Überlebenszeit der HB zwischen Proteinrückgratatomen und Wasser.	X				1+2
lifenum	Totale Überlebenszeit der HB zwischen Proteinrückgratatomen und Wasser.	X				2
lifepegavg	Durchschnittliche Überlebenszeit der HB des Polymers zum Solvens.	X				2
lifepegnum	Totale Überlebenszeit der HB des Polymers zum Solvens.	X				2
lifescavg	Durchschnittliche Überlebenszeit der HB des Polymers zum Solvens.	X				2
lifescnum	Totale Überlebenszeit der HB des Polymers zum Solvens.	X				2
molbur	Proteinoberfläche, welche durch Polymeratome bedeckt ist.	X				1+2
orderback	Proteinrückgratamid-Orderparameter. Der Durchschnittswert für alle 10 ns-Trajektorienblöcke wurde hierzu berechnet, angelehnt an die Methoden aus [381].	X	X	X		1+2
orderside	Seitenkettenorderparameter (keine Werte für nicht proteinogene Aminosäuren). Der Durchschnittswert für alle 10 ns-Trajektorienblöcke wurde hierzu berechnet.	X	X	X		1+2
p1	Perkolationsindex vom Typ 1 (CNA-Parameter) für jede Aminosäure X ($p1_X$) und Region X ($p1mean_X$), sowie für das ganze Protein ($p1mean$).	X	X	X		1+2
pegangle	Winkel zwischen dem Massenzentrum des Protein, dem Massenzentrum des Aminosäurebereichs der Konjugationsstelle und der Polymerkette (beschrieben in [35]) als ein Indikator für die Ausrichtung des Polymers zur Proteinoberfläche.	X				1+2
pegdcom	Distanz zwischen dem Massenzentrum des Proteins und dem Massenzentrum der konjugierten Aminosäure (beschrieben in [35]).	X				1+2
pegohhbonds	Anzahl an HB der Polymerkette zu Hydroxylseitenketten (S, T und Y).	X				2
pegwathbonds	Anzahl an HB der Polymerkette zu Wassermolekülen pro Momentaufnahme.	X				1+2
pg4dist	Ansatz 1: Minimaldistanz des letzten Polymer-Sauerstoffatoms zu jeglicher oder einer bestimmten Aminosäure. Ansatz 2: Maximaldistanz jeglicher Polymeratome zu Proteinatomen.	X		X		1+2
pgbrg	Anzahl an HB pro Momentaufnahme, die zwischen Polymeratomen über überbrückende Wassermoleküle gebildet wurden. Dieses Phänomen wurde in den ersten Simulationen beobachtet.	X				1+2
phi	ϕ -Winkel aller Aminosäuren X.	X				1+2
price1	Winkel θ_{Price} in der Kristallstruktur, gemäß [100], zwischen dem C_α -Atom der konjugierten Aminosäure, dem Massenzentrum der entsprechenden Seitenkette (vor Konjugation) und dem nächstgelegenen Serin, Threonin oder Tyrosin (kein Durchschnitt berechenbar).	X				1+2
price2	Winkel analog zu $price1$, jedoch handelt es sich um den Durchschnitt aus der Simulation der entsprechenden unkonjugierten Mutante.	X				1+2

price3	Winkel analog zu <i>price2</i> , jedoch handelt es sich um den Durchschnitt aus der Simulation der entsprechenden konjugierten Mutante.	X	1+2
price4	Winkel analog zu <i>price3</i> , jedoch wird bei der Berechnung des Massenzentrums der Seitenkette der Polymerbereich ignoriert.	X	1+2
price1apol	Wie <i>price1</i> für apolare Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
price2apol	Wie <i>price2</i> für apolare Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
price3apol	Wie <i>price3</i> für apolare Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
price4apol	Wie <i>price4</i> für apolare Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
price1arom	Wie <i>price1</i> für aromatische Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
price2arom	Wie <i>price2</i> für aromatische Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
price3arom	Wie <i>price3</i> für aromatische Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
price4arom	Wie <i>price4</i> für aromatische Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
price1pol	Wie <i>price1</i> für polare Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
price2pol	Wie <i>price2</i> für polare Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
price3pol	Wie <i>price3</i> für polare Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
price4pol	Wie <i>price4</i> für polare Aminosäureseitenketten. Nur im Falle der Modelle A2v2-R1 und A2v2-C1 genutzt.	X	(2)
protpeghb	Anzahl an HB des Polymers pro Momentaufnahme zu jeder einzelnen Aminosäure X.	X	1+2
protpeghbonds	Durchschnittliche Anzahl an HB des Polymers pro Momentaufnahme zum gesamten Protein.	X	1+2
protprothbonds	Durchschnittliche Anzahl an Protein-Protein-HB pro Momentaufnahme.	X	1+2
protwathbonds	Durchschnittliche Anzahl an Protein-Wasser-HB pro Momentaufnahme.	X	1+2
psi	ψ -Winkel aller Aminosäuren X.	X	1+2
radgyrpeg / rg-peg	Gyrationsradius der Polymer-Schweratome.	X	1+2
rdf_Y	RDF-Wert von Wassermolekülen um Aminosäuren herum in einem Bereich von 2-4 Å, 4-6 Å und 6-8 Å (benannt als Y = 200, 400, 600). Im Ansatz 2 wurden dagegen die Bereiche 2-3 Å, 3-4 Å, 4-5 Å, 5-6 Å, 6-7 Å, 7-8 Å und 9-10 Å berechnet.	X	X X 1+2
rdfapol_Y	RDF-Wert von Wassermolekülen um apolare Aminosäuren herum in einem Bereich von 2-4 Å, 4-6 Å und 6-8 Å (benannt als Y = 200, 400, 600). Im Ansatz 2 wurden dagegen die Bereiche 2-3 Å, 3-4 Å, 4-5 Å, 5-6 Å, 6-7 Å, 7-8 Å und 9-10 Å berechnet.	X	1+2

rdfoh_Y	RDF-Wert von Wassermolekülen um die Aminosäuren Serin, Threonin und Tyrosin herum in einem Bereich von 2-4 Å, 4-6 Å und 6-8 Å (benannt als Y = 200, 400, 600). Im Ansatz 2 wurden dagegen die Bereiche 2-3 Å, 3-4 Å, 4-5 Å, 5-6 Å, 6-7 Å, 7-8 Å und 9-10 Å berechnet.	X			1+2
rdfpeg_Y	RDF-Wert von Wassermolekülen um die Polymeratome herum in einem Bereich von 2-4 Å, 4-6 Å und 6-8 Å (benannt als Y = 200, 400, 600). Im Ansatz 2 wurden dagegen die Bereiche 2-3 Å, 3-4 Å, 4-5 Å, 5-6 Å, 6-7 Å, 7-8 Å und 9-10 Å berechnet.	X			1+2
rdfpegtores_Y	RDF-Wert von Polymeratomen um das Protein herum in einem Bereich von 2-4 Å, 4-6 Å und 6-8 Å (benannt als Y = 200, 400, 600). Im Ansatz 2 wurden dagegen die Bereiche 2-3 Å, 3-4 Å, 4-5 Å, 5-6 Å, 6-7 Å, 7-8 Å und 9-10 Å berechnet.	X	X		1+2
rdfpol_Y	RDF-Wert von Wassermolekülen um polare Aminosäuren herum in einem Bereich von 2-4 Å, 4-6 Å und 6-8 Å (benannt als Y = 200, 400, 600). Im Ansatz 2 wurden dagegen die Bereiche 2-3 Å, 3-4 Å, 4-5 Å, 5-6 Å, 6-7 Å, 7-8 Å und 9-10 Å berechnet.	X			1+2
rdfwopeg_Y	RDF-Wert von Wassermolekülen um das Protein (ohne Polymer) herum in einem Bereich von 2-4 Å, 4-6 Å und 6-8 Å (benannt als Y = 200, 400, 600). Im Ansatz 2 wurden dagegen die Bereiche 2-3 Å, 3-4 Å, 4-5 Å, 5-6 Å, 6-7 Å, 7-8 Å und 9-10 Å berechnet.	X			1+2
rdfX_Y	RDF-Wert von Wassermolekülen um bestimmte Aminosäuren herum in einem Bereich von 2-4 Å, 4-6 Å und 6-8 Å (benannt als Y = 200, 400, 600). Dies wurde für X = Gly, Ser, Tyr, Phe, Asn, Glu und Pro berechnet. Im Ansatz 2 wurden dagegen die Bereiche 2-3 Å, 3-4 Å, 4-5 Å, 5-6 Å, 6-7 Å, 7-8 Å und 9-10 Å berechnet.	X			1+2
rdfX_Y	RDF-Wert von Wassermolekülen um alle Aminosäuren X herum, in einem Bereich von 2-3 Å, 3-4 Å, 4-5 Å, 5-6 Å, 6-7 Å, 7-8 Å und 9-10 Å.	X			2
rg	Gyrationsradius der Protein- und Polymer-Schweratome.	X	X	X	1+2
rgbb	Gyrationsradius des Proteinrückgrats.			X	1+2
rgwopeg	Gyrationsradius der Protein-Schweratome (ohne Polymer).		X	X	1+2
rhb	Anzahl an HB jeder Aminosäure X zu Wassermolekülen pro Momentaufnahme.	X			1+2
rigidity	Rigiditätsindex (CNA-Parameter) für jede Aminosäure X (<i>crigidityX</i>) und Region X (<i>rigiditymeanX</i>), sowie für das ganze Protein (<i>rigiditymean</i>).	X	X	X	1+2
rmsd	RMSD-Wert der Schweratome des Proteinrückgrats (<i>Rg</i> und <i>Pr</i> -Kalkulationen), bzw. der Schweratome der gesamten Aminosäure (<i>Rs</i>). Als Referenz diente die aufbereitete Kristallstruktur.	X	X	X	1+2
rmsdp	RMSD-Wert der Polymer-Schweratome (Startstruktur als Referenz).			X	1
rmsf	Massengewichtete RMSF-Werte der Schweratome des Proteinrückgrats pro Aminosäure (<i>Rg</i> und <i>Pr</i> -Kalkulationen), bzw. der Schweratome der gesamten Aminosäure (<i>Rs</i>). Als Referenz diente die durchschnittliche Struktur der Simulation. Kein Durchschnitt kalkulierbar.	X	X	X	1+2
rmsfp	Massengewichtete RMSF-Werte der Schweratome des Polymers. Als Referenz diente die durchschnittliche Struktur der Simulation. Kein Durchschnitt kalkulierbar.			X	1+2
shell	Anzahl an Wassermolekülen in der ersten Hydrathülle (<3 Å), Polymeratome eingeschlossen.	X	X	X	1+2

shellapol	Anzahl an Wassermolekülen in der ersten Hydrathülle (<3 Å) von apolaren Aminosäuren.	X			1+2
shellapolo	Anzahl an Wassermolekülen in der ersten und zweiten Hydrathülle (<5 Å) von apolaren Aminosäuren.	X			1+2
shello	Anzahl an Wassermolekülen in der ersten und zweiten Hydrathülle (<5 Å), Polymeratome eingeschlossen.	X	X	X	1+2
shellloh	Anzahl an Wassermolekülen in der ersten Hydrathülle (<3 Å) der Aminosäuren Serin, Threonin und Tyrosin.	X			1+2
shelloho	Anzahl an Wassermolekülen in der ersten und zweiten Hydrathülle (<5 Å) der Aminosäuren Serin, Threonin und Tyrosin.	X			1+2
shellpeg	Anzahl an Wassermolekülen in der ersten Hydrathülle (<3 Å) der Polymeratome.	X			1+2
shellpegapol	Anzahl an Polymeratomen in der ersten Hydrathülle (<3 Å) apolarer Proteinatome.	X			2
shellpego	Anzahl an Wassermolekülen in der zweiten Hydrathülle (<5 Å) der Polymeratome.	X			1+2
shellpegpola	Anzahl an Polymeratomen in der ersten Hydrathülle (<3 Å) polarer Proteinatome.	X			2
shellpegoh	Anzahl an Polymeratomen in der ersten Hydrathülle (<3 Å) von Hydroxylseitenketten (S, T und Y).	X			2
shellpegapolo	Anzahl an Polymeratomen in der zweiten Hydrathülle (<5 Å) apolarer Proteinatome.	X			2
shellpegpola	Anzahl an Polymeratomen in der zweiten Hydrathülle (<5 Å) polarer Proteinatome.	X			2
shellpegoho	Anzahl an Polymeratomen in der zweiten Hydrathülle (<5 Å) von Hydroxylseitenketten (S, T und Y).	X			2
shellpol	Anzahl an Wassermolekülen in der ersten Hydrathülle (<3 Å) von polaren Aminosäuren.	X			1+2
shellpola	Anzahl an Wassermolekülen in der ersten und zweiten Hydrathülle (<5 Å) von polaren Aminosäuren.	X			1+2
shellwopeg	Anzahl an Wassermolekülen in der ersten Hydrathülle (<3 Å), Polymeratome ausgenommen.	X			1+2
shellwopego	Anzahl an Wassermolekülen in der ersten Hydrathülle (<5 Å), Polymeratome ausgenommen.	X			1+2
srigidity	Seitenkettenrigiditätsindex (CNA-Parameter) für jede Aminosäure X (<i>srigidityX</i>) und Region X (<i>srigiditymedianX</i>), sowie für das ganze Protein (<i>srigiditymean</i>).	X	X	X	1+2
sscontent	Durchschnittlicher Sekundärstrukturanteil des Proteins.	X	X	X	1+2
surf	Solvenzzugängliche Oberfläche.	X	X	X	1+2
surfapol	Solvenzzugängliche Oberfläche apolarer Atome.	X	X	X	1+2
surfapr	Verhältnis zwischen SASA apolarer und polarer Atome.			X	1+2
surfpeg	Solvenzzugängliche Oberfläche der Polymeratome.			X	1+2
surfpola	Solvenzzugängliche Oberfläche polarer Atome.	X	X	X	1+2
tau1	Kontinuierliche HB-Überlebenszeit der Seitenketten zu Wassermolekülen, berechnet mit dem mdanalysis-Modul.			X	2
tau2	Intermittierende HB-Überlebenszeit der Seitenketten zu Wassermolekülen, berechnet mit dem mdanalysis-Modul.			X	2

totalint		Interaktionsenergie zwischen Polymer- und Proteinatomen.	X	1+2
trphcce1 / trphcex1		Phasenübergangstemperatur (CNA-Parameter), basierend auf der Clusterkonfigurationsentropie vom Typ 1.	X	1+2
trphcce2 / trphcex2		Phasenübergangstemperatur (CNA-Parameter), basierend auf der Clusterkonfigurationsentropie vom Typ 2.	X	1+2
trphcces / trphcexs		Phasenübergangstemperatur (CNA-Parameter), basierend auf der Clusterkonfigurationsentropie vom Typ 2 mit doppelt sigmoidalem Fit.	X	1+2
trphrop1 / trphropx1	/	Phasenübergangstemperatur (CNA-Parameter), basierend auf dem Rigiditätsorderparameter vom Typ 1.	X	1+2
trphrop2 / trphropx2	/	Phasenübergangstemperatur (CNA-Parameter), basierend auf dem Rigiditätsorderparameter vom Typ 2.	X	1+2

D.2 Ansatz 1

D.2.1 Modelltypen

Tabelle D.2: Übersicht über die getesteten Modelltypen im ersten QSPR-Ansatz (Kap. 3.1.6.2) gemäß der caret-Webseite [382]. Der Name der Modellart und eine kurze Beschreibung sind aufgeführt. Zusätzlich sind die zugrundeliegenden R-Pakete aufgelistet. Die letzte Spalte gibt an, ob hierbei Regressions- (R) und / oder Klassifikationsmodelle (C) generiert wurden.

Name	Beschreibung	R-Paket	Typ
avNNet	Model averaged neural network	nnet	C
C5.0Rules	Single C5.0 ruleset	C50	C
C5.0Tree	Single C5.0 tree	C50	C
cubist	Cubist model	cubist	R
gaussprLinear	Gaussian process model	kernlab	R
gaussprPoly	Gaussian process model with polynomial kernel	kernlab	R
gpls	Generalized partial least squares model	gpls	C
mlp	Multi-layer perceptron	RSNNS	C
mlpML	Multi-layer perceptron with multiple layers	RSNNS	C
mlpWeightDecay	Multi-layer perceptron using weight decay	RSNNS	C
mlpWeightDecayML	Multi-layer perceptron using weight decay, with multiple layers	RSNNS	C
nnet	Neural network	nnet	C, R
ORFlog	Oblique random forest (using logistic regression)	obliqueRF	C
ORFpls	Oblique random forest (using partial least squares)	obliqueRF	C
ORFRidge	Oblique random forest (linear combinations created using L2 regularization)	obliqueRF	C
ORFsvm	Oblique random forest (using support vector machines)	obliqueRF	C

parRF	Parallel random forest	e1071, randomForest, foreach, import	C, R
pcaNNet	Neural networks with feature extraction	nnet	C
pcr	Principal component analysis	pls	R
pls	Partial least squares	pls	R
regLogistic	Regularized logistic regression	LiblineaR	C
rf	Random forest	rf	C, R
ridge	Ridge regression	elasticnet	R
rqlasso	Quantile regression with LASSO penalty	rqPen	R
rqnc	Non-convex penalized quantile regression	rqPen	R
RRFglobal	Regularized random forest	RRF	R
rvmRadial	Relevance vector machines with radial basis function kernel	kernlab	R
svmLinear3	Support vector machines with linear kernel	LiblineaR	R
svmPoly	Support vector machines with polynomial kernel	kernlab	R
svmRadial	Support vector machines with radial basis function kernel	kernlab	R
svmRadialSigma	Support vector machines with radial basis function kernel (tunes over cost and sigma)	kernlab	R
xgbLinear	eXtreme gradient boosting (linear learners)	xgboost	R
xgbTree	eXtreme gradient boosting (tree learners)	xgboost, plyr	R

D.2.2 Modellleistungen

Tabelle D.3: Validierungsparameter für die besten Modelle aus den Kategorien A bis D im ersten Modellgenerierungsansatz. Der Modelltyp ist angegeben (vgl. Tab. D.2), dazu zu jedem Datensatz RMSE, MAE, CCC und R^2 (für Regressionsmodelle), bzw. AUROC und CCR (Klassifikationsmodelle). Die jeweilige Anzahl genutzter Deskriptoren (Des.) ist in der letzten Spalte angegeben. Die Modelle sind auf Basis der $RMSE_{HS}$ - bzw. CCR_{HS} -Werte geordnet. Die Kategorien A und B weisen ausschließlich Modelle auf, die, in Relation zum Datensatz, zu viele Deskriptoren verwenden und damit ein komplettes Overfitting aufweisen.

Pin 1-spezifische Regressionsmodelle (Kategorie A)												
#	Art	RMSE _{TS}	MAE _{TS}	CCC _{TS}	R ² _{TS}	RMSE _{HS}	MAE _{HS}	CCC _{HS}	R ² _{HS}	Des.		
A1	xgbLinear	0.02	0.00	1.00	1.00	0.25	0.19	0.72	0.59	166		
A2	xgbLinear	0.02	0.00	1.00	1.00	0.26	0.21	0.70	0.59	526		
A3	cubist	0.04	0.03	1.00	1.00	0.28	0.23	0.69	0.52	76		
Pin 1-spezifische Klassifikationsmodelle (Kategorie B)												
#	Art	AUROC _{TS}			CCR _{TS}	AUROC _{HS}			CCR _{HS}	Des.		
B1	pcaNNet	0.99			0.99	1.00			0.87	49		
B2	avNNet	1.00			1.00	1.00			0.87	190		
B3	ORFsvm	1.00			1.00	1.00			0.80	49		
Pin 1-unspezifische Regressionsmodelle (Kategorie C)												
#	Art	RMSE _{TS}	MAE _{TS}	CCC _{TS}	R ² _{TS}	RMSE _{HS}	MAE _{HS}	CCC _{HS}	R ² _{HS}	Des.		
C1	xgbTree	0.14	0.10	0.95	0.92	0.24	0.20	0.75	0.63	11		
C2	svmLinear3	0.36	0.26	0.55	0.40	0.25	0.20	0.76	0.61	12		
C3	gaussprLinear	0.34	0.28	0.62	0.46	0.26	0.20	0.74	0.56	12		
Pin 1-unspezifische Klassifikationsmodelle (Kategorie D)												
#	Art	AUROC _{TS}			CCR _{TS}	AUROC _{HS}			CCR _{HS}	Des.		
D1	nnet	0.95			0.93	0.91			0.80	8		
D2	regLogistic	0.76			0.71	0.91			0.73	8		
D3	mlpWeightDecayML	1.00			1.00	0.94			0.67	23		

D.2.3 ADAN-Analysen

Tabelle D.4: Kategorisierung der HS-Konjugate in eine von sechs ADAN-Kategorien, dargestellt für die Modelle A1 bis A3, C1 bis C3 und E. Je höher die Kategorie, desto schlechter ist das Modell zur Einschätzung der Faltungsenergie für das jeweilige HS-Konjugat geeignet (Kategorien 0 - 1 in grün = innerhalb der Applikationsdomäne, 2 - 3 in orange = unsichere Vorhersage, 4 - 6 in rot = außerhalb der Applikationsdomäne).

Konjugat	Modelle						
	A1	A2	A3	C1	C2	C3	E
29-1	3	1	2	0	0	0	0
32-1	3	1	2	0	0	0	0
19(S32A)-1	3	1	3	0	0	0	0
19-4	3	1	2	2	3	2	2
16-9	4	1	4	0	0	0	0
16-Z	3	1	3	3	4	4	2
29-Z	4	1	1	0	0	0	1
32-Z	3	1	3	0	0	0	1
19-2	4	2	4	0	0	0	3
32-2	4	1	3	0	0	0	0
23(E12F:R14L)-Z	3	1	3	1	1	1	3
19-J	3	1	4	2	2	2	0
18(S16D:R17A)-1	4	1	2	0	0	0	0
19-5	3	1	2	0	0	0	3
16+19-1	3	1	3	2	2	2	0
1SRL.20-1	-	-	-	3	3	3	2

D.2.4 Pin 1-unspezifische Deskriptoren

protpeghbonds, *pgbrg*, *totalint*, *pg4dist*, *rmsdp*, *molbur*, *surfpeg*, *radgyrpeg*, *rdfpeg_200*, *price2*, *pegdcom*, *pegangle*, *protprothbondsratio*, *bbhbondsratio*, *bbwathbondsratio*, *lifetime*, *rmsdratio*, *shelloratio*, *shellwopegoratio*, *shellohoratio*, *shellpoloratio*, *shellapoloratio*, *rdfoh_200ratio*, *rdfoh_400ratio*, *rdfoh_600ratio*, *rdfpol_200ratio*, *rdfpol_400ratio*, *rdfpol_600ratio*, *rdfapol_200ratio*, *rdfapol_400ratio*, *rdfapol_600ratio*, *entropysideratio*, *p1medianratio*

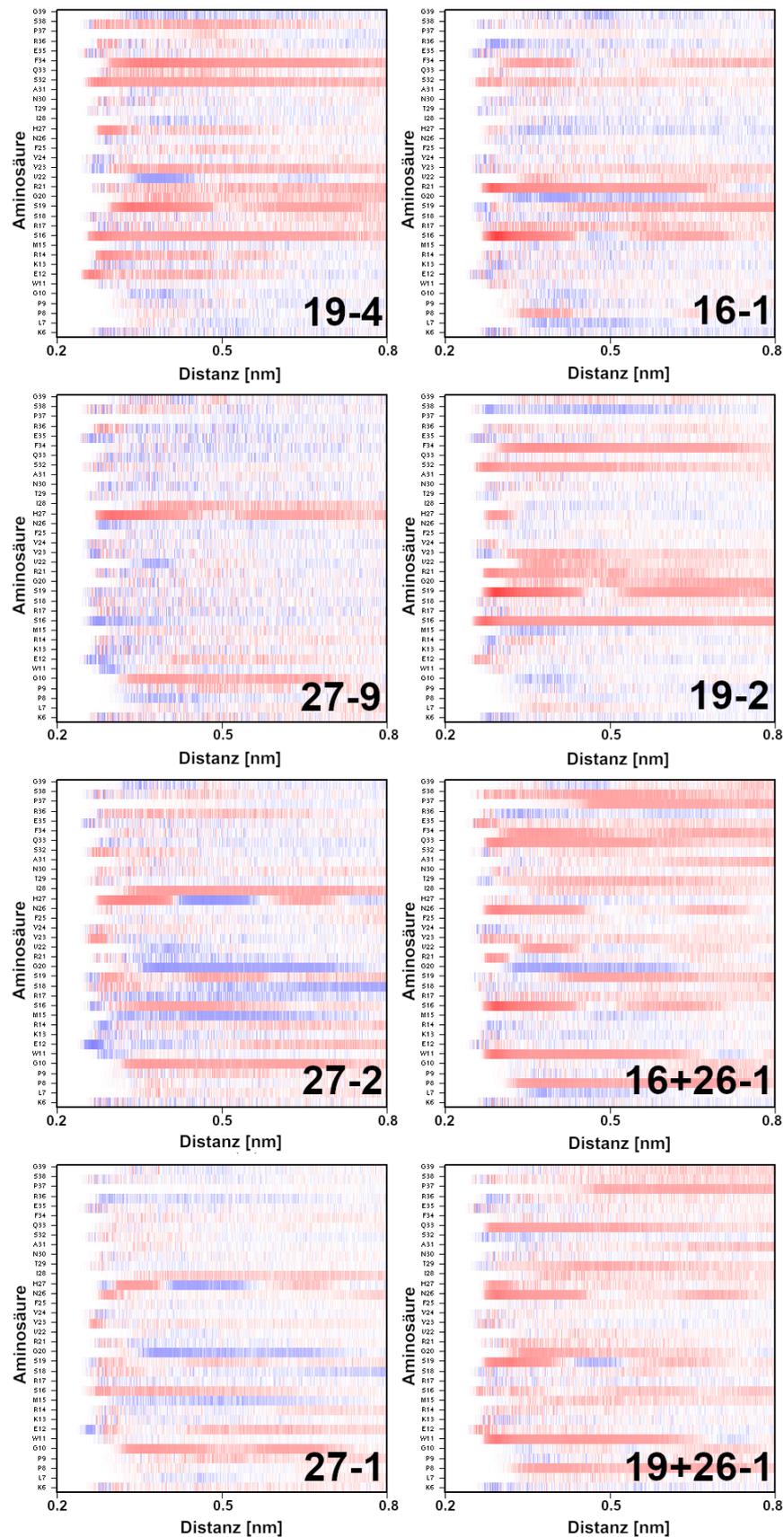


Abbildung D.1: RDF-Differenz-Plots einer Auswahl stark destabilisierter (linke Spalte) und stabilisierter Varianten (rechte Spalte).

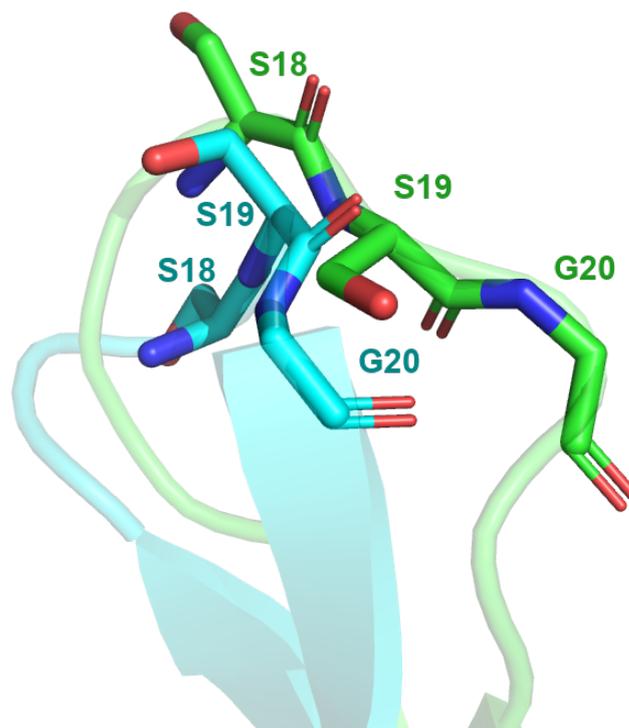


Abbildung D.2: Illustration des beobachteten Glycin-Flips. In grün ist die Startstruktur von 27-1 dargestellt, während die Konformation in cyan die Umorientierung beschreibt.

D.3 Ansatz 2

D.3.1 Pin 1-unspezifische Deskriptoren

protpeghbonds, pgbgrg, pegwathbonds, totalint, intoh, pg4dist, molbur, surfpeg, rmsfp, rgpeg, shellpeg, shellpego, shellpegapolo, shellpegpolo, rdfpeg_200, rdfpeg_300, rdfpeg_400, rdfpeg_500, rdfpeg_900, price1, price2, price3, price4, pegdcom, pegangle, lifepenum, trphccex1, trphccex2, trphropx1, trphccexs, bbhbondsratio, rmsdratio, surfaprratio, rgratio, rgbbratio, rgwopegratio, shellohratio, shellpolratio, shellapolratio, shellohoration, shellpoloratio, shellapoloratio, rdf_200ratio, rdf_400ratio, rdf_500ratio, rdf_600ratio, rdf_700ratio, rdf_900ratio, rdfwopeg_200ratio, rdfwopeg_400ratio, rdfwopeg_500ratio, rdfwopeg_600ratio, rdfwopeg_700ratio, rdfwopeg_900ratio, rdfoh_200ratio, rdfoh_300ratio, rdfoh_400ratio, rdfoh_500ratio, rdfoh_600ratio, rdfoh_700ratio, rdfoh_800ratio, rdfoh_900ratio, rdfpol_200ratio, rdfpol_300ratio, rdfpol_400ratio, rdfpol_500ratio, rdfpol_600ratio, rdfpol_700ratio, rdfpol_800ratio, rdfpol_900ratio, rdfapol_200ratio, rdfapol_300ratio, rdfapol_400ratio, rdfapol_500ratio, rdfapol_600ratio, rdfapol_700ratio, rdfapol_800ratio, rdfapol_900ratio, orderbackratio, tau1ratio, tau2ratio

D.3.2 Modelltypen

Tabelle D.5: Übersicht über die getesteten Modelltypen im zweiten QSPR-Ansatz (Kap. 3.1.6.3) gemäß der caret-Webseite [382]. Der Name der Modellart und eine kurze Beschreibung sind aufgeführt. Zusätzlich sind die zugrundeliegenden R-Pakete aufgelistet. Die letzte Spalte gibt an, ob hierbei Regressions- (R) oder Klassifikationsmodelle (C) generiert wurden.

Name	Beschreibung	R-Paket	Typ
lm	Linear regression	Basis-R-Paket	R
glm (binomial)	Generalized linear model (logistic regression)	Basis-R-Paket	C

D.3.3 Modelleleistungen

Tabelle D.6: Validierungsparameter für die erstellten Regressionsmodelle des zweiten Modellierungsansatzes für die verschiedenen Datensätze (DS). Werte für das Trainings- und das Holdout-Set sind für das standardmäßige Interkorrelationslimit von 0.99 aufgelistet. Nachgeschaltete Filterungen des finalen Deskriptorsets mit einem zusätzlichen Limit von 0.75 führten zu Modellen mit den Werten in der unteren Hälfte der Tabelle.

Korrelationslimit: 0.99								
DS	RMSE _{TS}	CCC _{TS}	$\tau_B(TS)$	R ² _{TS}	RMSE _{HS}	CCC _{HS}	$\tau_B(HS)$	R ² _{HS}
1	0.21	0.85	0.61	0.73	0.59	0.58	0.35	0.34
2	0.68	0.02	-0.05	0.00	0.36	0.51	0.34	0.34
3	0.32	0.71	0.53	0.55	0.81	0.30	0.45	0.18
4	0.26	0.87	0.69	0.77	0.94	0.03	-0.05	0.00
5	0.20	0.65	0.46	0.48	0.39	0.72	0.67	0.96
6	0.25	0.81	0.45	0.68	0.37	0.74	0.82	0.76
7	0.28	0.76	0.59	0.62	0.65	0.27	0.20	0.09
8	0.32	0.59	0.45	0.42	0.76	-0.19	-0.11	0.06
9	0.24	0.83	0.53	0.71	0.58	0.20	0.09	0.04
Korrelationslimit: 0.75								
DS	RMSE _{TS}	CCC _{TS}	$\tau_B(TS)$	R ² _{TS}	RMSE _{HS}	CCC _{HS}	$\tau_B(HS)$	R ² _{HS}
1	0.26	0.76	0.56	0.61	0.56	0.53	0.35	0.29
2	0.68	0.02	-0.05	0.00	0.36	0.51	0.34	0.34
3	0.77	0.28	0.45	0.39	0.36	0.61	0.42	0.44
4	0.26	0.87	0.69	0.77	0.94	0.03	-0.05	0.00
5	0.21	0.56	0.38	0.38	0.66	0.18	0.67	0.79
6	0.34	0.58	0.32	0.41	0.40	0.63	0.71	0.66
7	0.31	0.69	0.51	0.53	0.66	0.27	0.20	0.09
8	0.32	0.59	0.45	0.42	0.76	-0.19	-0.11	0.06
9	0.30	0.72	0.46	0.56	0.53	0.13	0.27	0.02

Tabelle D.7: Validierungsparameter für die erstellten Klassifikationsmodelle des zweiten Modellierungsansatzes für die verschiedenen Datensätze (DS). Werte für das Trainings- und das Holdout-Set sind für das standardmäßige Interkorrelationslimit von 0.99 aufgelistet. Nachgeschaltete Filterungen des finalen Deskriptorsets mit einem zusätzlichen Limit von 0.75 führten zu Modellen mit den Werten in der unteren Hälfte der Tabelle.

Korrelationslimit: 0.99						
DS	B_{Acc}_{TS}	AUC_{TS}	AUPRC_{TS}	B_{Acc}_{HS}	AUC_{HS}	AUPRC_{HS}
1	0.84	0.93	0.96	0.77	0.83	0.81
2	0.80	0.81	0.90	0.60	0.67	0.71
3	0.75	0.83	0.90	0.71	0.69	0.59
4	0.85	0.93	0.93	0.46	0.67	0.45
5	1.00	1.00	0.36	0.50	0.50	0.00
6	0.85	0.94	0.97	0.48	0.60	0.42
7	0.73	0.89	0.95	0.29	0.43	0.57
8	1.00	1.00	0.08	0.83	0.83	0.00
9	0.81	0.91	0.96	0.50	0.71	0.67
Korrelationslimit: 0.75						
DS	B_{Acc}_{TS}	AUC_{TS}	AUPRC_{TS}	B_{Acc}_{HS}	AUC_{HS}	AUPRC_{HS}
1	0.63	0.86	0.95	0.50	0.83	0.80
2	0.56	0.73	0.88	0.50	0.54	0.59
3	0.50	0.69	0.76	0.57	0.63	0.53
4	0.82	0.93	0.94	0.46	0.42	0.39
5	0.50	0.82	0.93	0.50	1.00	0.50
6	0.77	0.93	0.97	0.40	0.60	0.42
7	0.67	0.85	0.94	0.52	0.57	0.61
8	1.00	1.00	0.08	0.83	0.83	0.00
9	0.60	0.85	0.95	0.61	0.68	0.61

D.3.4 Analysen

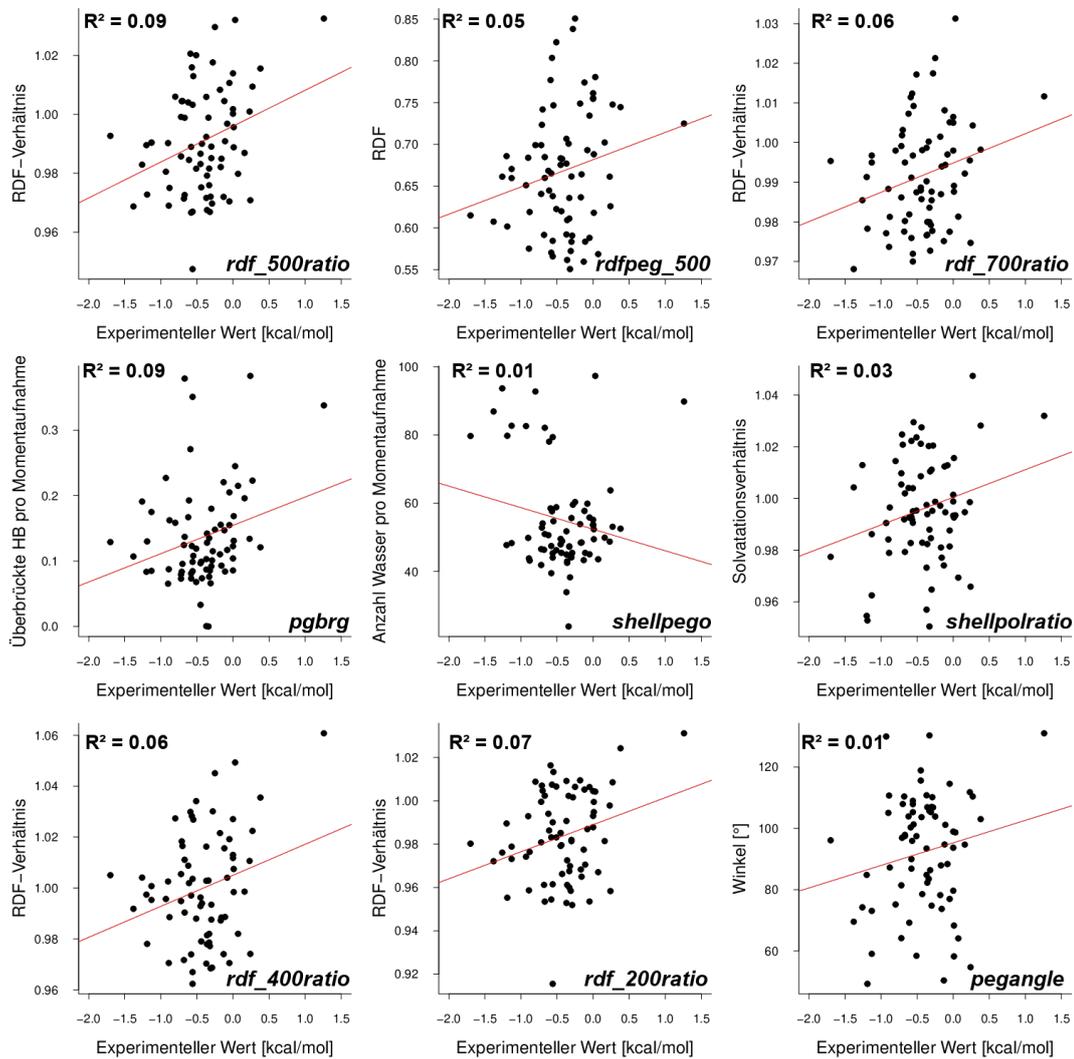


Abbildung D.3: Univariate Regressionen weiterer Modelldescriptoren von Modell A2-R6.

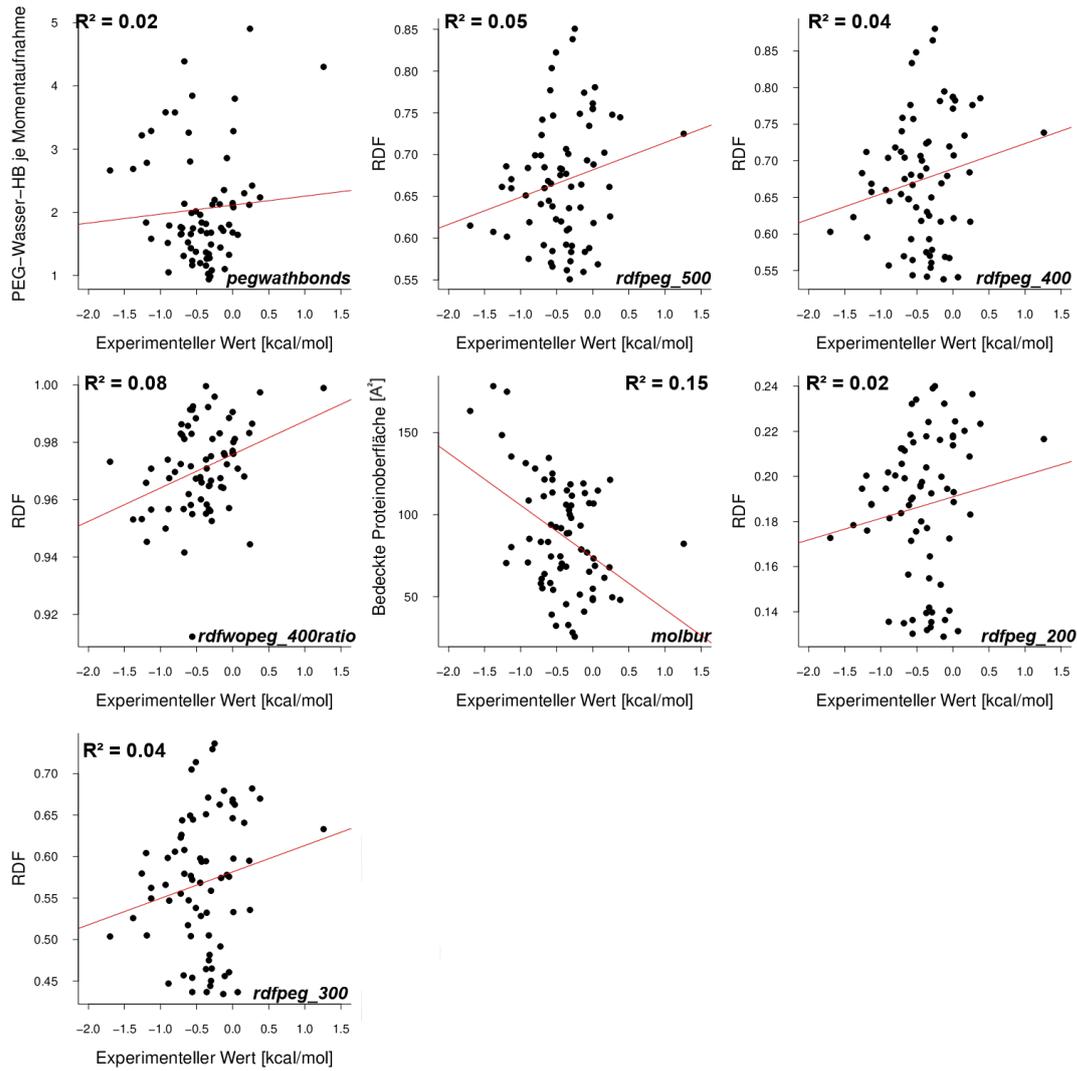


Abbildung D.4: Univariate Regressionen der Modelldescriptoren von A2-C1, analog zu Abb. D.3.

Tabelle D.8: Vorhersagen des Modells A2-R6 zu den Pin 1-Konjugaten der Polymeralternativen (vgl. Kennzeichen in Tab. C.1).

Konjugat	Vorhersage [kcal/mol]
14- λ	-0.14
16- λ	0.39
17- λ	0.10
18- λ	0.45
19- λ	0.86
21- λ	0.23
23- λ	-0.05
26- λ	0.92
27- λ	1.70
28- λ	1.10
29- λ	1.24
30- λ	0.33
32- λ	0.00
<hr/>	
14- ψ	0.67
16- ψ	0.87
17- ψ	1.05
18- ψ	1.55
19- ψ	-1.25
21- ψ	1.47
23- ψ	0.05
26- ψ	1.17
27- ψ	0.84
28- ψ	2.05
29- ψ	1.14
30- ψ	1.24
32- ψ	0.67
<hr/>	
14- ϵ	1.18
16- ϵ	0.71
17- ϵ	1.58
18- ϵ	1.18
19- ϵ	1.32
21- ϵ	1.05
23- ϵ	1.07
26- ϵ	0.90
27- ϵ	2.08
28- ϵ	1.31
29- ϵ	1.56
30- ϵ	1.01
32- ϵ	0.92

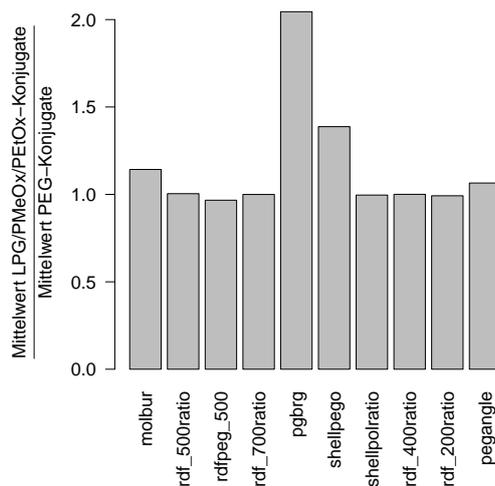


Abbildung D.5: Verhältnis zwischen den Mittelwerten der Polymere LPG, PMeOx und PEtOx und dem Mittelwert der PEG-Konjugate des TS für die relevanten Deskriptoren des Modells A2-R6.

Tabelle D.9: Vorhersagen des Modells A2-C1 zu den Pin 1-Konjugaten der Polymeralternativen (vgl. Kennzeichen in Tab. C.1). Dabei steht S für stabilisiert und N für nicht stabilisiert. In Klammern ist die Wahrscheinlichkeit aufgeführt, das jeweilige Konjugat der Klasse 1 (stabilisiert) zuzuordnen.

Konjugat	Vorhersage
14- λ	N (0.00)
16- λ	N (0.00)
17- λ	S (0.67)
18- λ	N (0.00)
19- λ	N (0.00)
21- λ	N (0.09)
23- λ	N (0.00)
26- λ	N (0.00)
27- λ	N (0.00)
28- λ	N (0.00)
29- λ	N (0.00)
30- λ	N (0.03)
32- λ	N (0.43)
14- ψ	N (0.02)
16- ψ	N (0.00)
17- ψ	N (0.00)
18- ψ	N (0.00)
19- ψ	S (0.99)
21- ψ	N (0.00)
23- ψ	S (0.99)
26- ψ	N (0.00)
27- ψ	N (0.00)
28- ψ	N (0.00)
29- ψ	N (0.00)
30- ψ	N (0.00)
32- ψ	N (0.00)
14- ϵ	N (0.08)
16- ϵ	N (0.00)
17- ϵ	N (0.00)
18- ϵ	N (0.00)
19- ϵ	N (0.00)
21- ϵ	N (0.00)
23- ϵ	N (0.00)
26- ϵ	N (0.00)
27- ϵ	N (0.00)
28- ϵ	N (0.00)
29- ϵ	N (0.00)
30- ϵ	N (0.00)
32- ϵ	N (0.00)

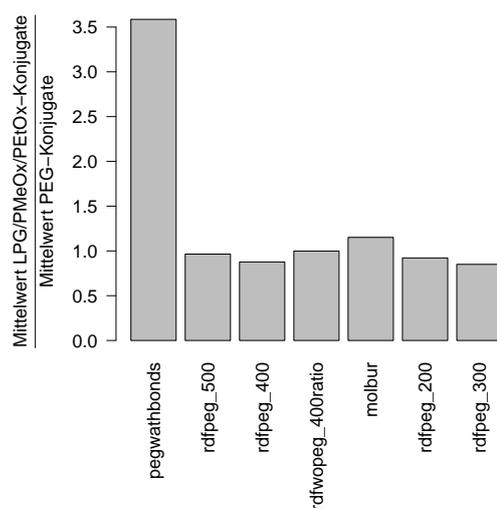


Abbildung D.6: Verhältnis zwischen den Mittelwerten der Polymere LPG, PMeOx und PEtOx und dem Mittelwert der PEG-Konjugate des TS für die relevanten Deskriptoren des Modells A2-C1.

Literatur

- [1] I. S. Johnson, „Human Insulin from Recombinant DNA Technology“, *Science*, Jg. 219, Nr. 4585, S. 632–637, 1983.
- [2] D. S. Dimitrov, „Therapeutic Proteins“, in *Therapeutic Proteins: Methods and Protocols*, Ser. Methods in Molecular Biology, V. Voynov und J. A. Caravella, Bd. 899, Humana Press, 2012, S. 1–26.
- [3] M. Kesik-Brodacka, „Progress in biopharmaceutical development“, *Biotechnology and Applied Biochemistry*, Jg. 65, Nr. 3, S. 306–322, 2018.
- [4] L. Urquhart, „Top companies and drugs by sales in 2019“, *Nature Reviews Drug Discovery*, Jg. 19, Nr. 4, S. 228–228, 2020.
- [5] S. Mitragotri, P. A. Burke und R. Langer, „Overcoming the challenges in administering biopharmaceuticals: Formulation and delivery strategies“, *Nature Reviews Drug Discovery*, Jg. 13, Nr. 9, S. 655–672, 2014.
- [6] M. Morishita und N. A. Peppas, „Is the oral route possible for peptide and protein drug delivery?“, *Drug Discovery Today*, Jg. 11, Nr. 19-20, S. 905–910, 2006.
- [7] P. J. Carter, „Introduction to current and future protein therapeutics: A protein engineering perspective“, *Experimental Cell Research*, Jg. 317, Nr. 9, S. 1261–1269, 2011.
- [8] M. Kessler, D. Goldsmith und H. Schellekens, „Immunogenicity of biopharmaceuticals“, *Nephrology Dialysis Transplantation*, Jg. 21, S. v9–v12, suppl_5 2006.
- [9] B. Kaupbayeva und A. J. Russell, „Polymer-enhanced biomacromolecules“, *Progress in Polymer Science*, Jg. 101, S. 101–194, 2020.
- [10] Y.-H. Bi, Z.-Q. Duan, X.-Q. Li, Z.-Y. Wang und X.-R. Zhao, „Introducing Biobased Ionic Liquids as the Nonaqueous Media for Enzymatic Synthesis of Phosphatidylserine“, *Journal of Agricultural and Food Chemistry*, Jg. 63, Nr. 5, S. 1558–1561, 2015.
- [11] A. Sharma, N. Kumar, B. D. Kuppermann, F. Bandello und A. Loewenstein, „Biologics, biosimilars, and biobetters: Different terms or different drugs?“, *Eye*, Jg. 33, Nr. 7, S. 1032–1034, 2019.
- [12] Y. Wang und C. Wu, „Site-Specific Conjugation of Polymers to Proteins“, *Biomacromolecules*, Jg. 19, Nr. 6, S. 1804–1825, 2018.

- [13] M. A. Stahmann und R. R. Becker, „A New Method for Adding Amino Acids and Peptides to Proteins“, *Journal of the American Chemical Society*, Jg. 74, Nr. 10, S. 2695–2696, 1952.
- [14] R. R. Becker und A. Mark, „Protein Modification by Reaction with N-Carboxy-Amino Acid Anhydrides“, *Journal of Biological Chemistry*, Jg. 204, Nr. 2, S. 745–752, 1953.
- [15] K. M. BurrIDGE, B. A. Shurina, C. T. Kozuszek, R. F. Parnell, J. S. Montgomery, J. L. VanPelt, N. M. Daman, R. M. McCarrick, T. A. Ramelot, D. Konkolewicz und R. C. Page, „Mapping protein–polymer conformations in bioconjugates with atomic precision“, *Chemical Science*, Jg. 11, Nr. 24, S. 6160–6166, 2020.
- [16] A. Abuchowski, T. Van Es, N. C. Palczuk und F. F. Davis, „Alteration of Immunological Properties of Bovine Serum Albumin by Covalent Attachment of Polyethylene Glycol“, *The Journal of Biological Chemistry*, Jg. 252, Nr. 11, S. 3578–3581, 1977.
- [17] A. Abuchowski, J. R. McCoy, N. C. Palczuk, T. Van Es und F. F. Davis, „Effect of Covalent Attachment of Polyethylene Glycol on Immunogenicity and Circulating Life of Bovine Liver Catalase“, *The Journal of Biological Chemistry*, Jg. 252, Nr. 11, S. 3582–3586, 1977.
- [18] F. M. Veronese und A. Mero, „The Impact of PEGylation on Biological Therapies“, *BioDrugs*, Jg. 22, Nr. 5, S. 315–329, 2008.
- [19] P. L. Turecek, M. J. Bossard, F. Schoetens und I. A. Ivens, „PEGylation of Biopharmaceuticals: A Review of Chemistry and Nonclinical Safety Information of Approved Drugs“, *Journal of Pharmaceutical Sciences*, Jg. 105, Nr. 2, S. 460–475, 2016.
- [20] G. Kozma, T. Shimizu, T. Ishida und J. Szebeni, „Anti-PEG antibodies: Properties, formation and role in adverse immune reactions to PEGylated nanobiopharmaceuticals“, *Advanced Drug Delivery Reviews*, 2020.
- [21] W. D. Chey, L. Webster, M. Sostek, J. Lappalainen, P. N. Barker und J. Tack, „Naloxegol for Opioid-Induced Constipation in Patients with Noncancer Pain“, *New England Journal of Medicine*, Jg. 370, Nr. 25, S. 2387–2396, 2014.
- [22] K. Kopeckova, T. Eckschlager, J. Sirc, R. Hobzova, J. Plch, J. Hrabeta und J. Michalek, „Nanodrugs Used in Cancer Therapy“, *Biomedical Papers*, Jg. 163, Nr. 2, S. 122–131, 2019.
- [23] A. Bunker, A. Magarkar und T. Viitala, „Rational design of liposomal drug delivery systems, a review: Combined experimental and computational studies of lipid membranes, liposomes and their PEGylation“, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, Jg. 1858, Nr. 10, S. 2334–2352, 2016.
- [24] P. C. Ke, S. Lin, W. J. Parak, T. P. Davis und F. Caruso, „A Decade of the Protein Corona“, *ACS Nano*, Jg. 11, Nr. 12, S. 11 773–11 776, 2017.
- [25] T. Okusaka, S. Okada, H. Ishii, M. Ikeda, H. Nakasuka, H. Nagahama, R. Iwata, H. Furukawa, K. Takayasu, Y. Nakanishi, M. Sakamoto, S. Hirohashi

- und M. Yoshimori, „Transarterial Chemotherapy with Zinostatin Stimalamer for Hepatocellular Carcinoma“, *Oncology*, Jg. 55, Nr. 4, S. 276–283, 1998.
- [26] E. M. Pelegri-O’Day, E.-W. Lin und H. D. Maynard, „Therapeutic Protein–Polymer Conjugates: Advancing Beyond PEGylation“, *Journal of the American Chemical Society*, Jg. 136, Nr. 41, S. 14 323–14 332, 2014.
- [27] P. Kiran, A. Khan, S. Neekhara, S. Pallod und R. Srivastava, „Nanohybrids as Protein-Polymer Conjugate Multimodal Therapeutics“, *Frontiers in Medical Technology*, Jg. 3, S. 676 025, 2021.
- [28] D. E. Koshland, „Application of a Theory of Enzyme Specificity to Protein Synthesis“, *Proceedings of the National Academy of Sciences*, Jg. 44, Nr. 2, S. 98–104, 1958.
- [29] C. A. Lipinski, F. Lombardo, B. W. Dominy und P. J. Feeney, „Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings IPII of original article: S0169-409X(96)00423-1. The article was originally published in *Advanced Drug Delivery Reviews* 23 (1997) 3–25. 1“, *Advanced Drug Delivery Reviews*, Jg. 46, Nr. 1-3, S. 3–26, 2001.
- [30] C. A. Lipinski, „Lead- and drug-like compounds: The rule-of-five revolution“, *Drug Discovery Today: Technologies*, Jg. 1, Nr. 4, S. 337–341, 2004.
- [31] A. Bunker und T. Róg, „Mechanistic Understanding From Molecular Dynamics Simulation in Pharmaceutical Research 1: Drug Delivery“, *Frontiers in Molecular Biosciences*, Jg. 7, 2020.
- [32] A. J. Russell, S. L. Baker, C. M. Colina, C. A. Figg, J. L. Kaar, K. Matyjaszewski, A. Simakova und B. S. Sumerlin, „Next generation protein-polymer conjugates“, *AIChE Journal*, Jg. 64, Nr. 9, S. 3230–3245, 2018.
- [33] P. B. Lawrence und J. L. Price, „How PEGylation influences protein conformational stability“, *Current Opinion in Chemical Biology*, Jg. 34, S. 88–94, 2016.
- [34] S. A. McNelles, V. M. Marando und A. Adronov, „Globular Polymer Grafts Require a Critical Size for Efficient Molecular Sieving of Enzyme Substrates“, *Angewandte Chemie International Edition*, Jg. 58, Nr. 25, S. 8448–8453, 2019.
- [35] A. Munasinghe, A. Mathavan, A. Mathavan, P. Lin und C. M. Colina, „PEGylation within a confined hydrophobic cavity of a protein“, *Physical Chemistry Chemical Physics*, Jg. 21, Nr. 46, S. 25 584–25 596, 2019.
- [36] A. Zaghmi, E. Mendez-Villuendas, A. Greschner, J. Liu, H. de Haan und M. Gauthier, „Mechanisms of activity loss for a multi-PEGylated protein by experiment and simulation“, *Materials Today Chemistry*, Jg. 12, S. 121–131, 2019.
- [37] T. T. Hoang Thi, E. H. Pilkington, D. H. Nguyen, J. S. Lee, K. D. Park und N. P. Truong, „The Importance of Poly(ethylene glycol) Alternatives for Overcoming PEG Immunogenicity in Drug Delivery and Bioconjugation“, *Polymers*, Jg. 12, Nr. 2, S. 298, 2020.

- [38] M. S. Messina, K. M. Messina, A. Bhattacharya, H. R. Montgomery und H. D. Maynard, „Preparation of biomolecule-polymer conjugates by grafting-from using ATRP, RAFT, or ROMP“, *Progress in Polymer Science*, Jg. 100, S. 101–186, 2020.
- [39] G. Pasut und F. M. Veronese, „State of the art in PEGylation: The great versatility achieved after forty years of research“, *Journal of Controlled Release*, Jg. 161, Nr. 2, S. 461–472, 2012.
- [40] G. Wandrey, J. Wurzel, K. Hoffmann, T. Ladner, J. Büchs, L. Meinel und T. Lühmann, „Probing unnatural amino acid integration into enhanced green fluorescent protein by genetic code expansion with a high-throughput screening platform“, *Journal of Biological Engineering*, Jg. 10, Nr. 1, S. 11, 2016.
- [41] A. C. Braun, M. Gutmann, T. Lühmann und L. Meinel, „Bioorthogonal strategies for site-directed decoration of biomaterials with therapeutic proteins“, *Journal of Controlled Release*, Jg. 273, S. 68–85, 2018.
- [42] S. Kim, W. Ko, B. H. Sung, S. C. Kim und H. S. Lee, „Direct protein–protein conjugation by genetically introducing bioorthogonal functional groups into proteins“, *Bioorganic & Medicinal Chemistry*, Jg. 24, Nr. 22, S. 5816–5822, 2016.
- [43] T. Lühmann, M. Schmidt, M. N. Leiske, V. Spieler, T. C. Majdanski, M. Grube, M. Hartlieb, I. Nischang, S. Schubert, U. S. Schubert und L. Meinel, „Site-Specific POxylation of Interleukin-4“, *ACS Biomaterials Science & Engineering*, Jg. 3, Nr. 3, S. 304–312, 2017.
- [44] F. Xie, X. Jia, Z. Zhu, Y. Wu, H. Jiang, H. Yang, Y. Cao, R. Zhu, B. Zhou, J. Du und Y. Tang, „Chemical trigger-enabled bioconjugation reaction“, *Organic & Biomolecular Chemistry*, 10.1039.D1OB01177D, 2021.
- [45] R. Falatach, S. Li, S. Sloane, C. McGlone, J. A. Berberich, R. C. Page, S. Averick und D. Konkolewicz, „Why synthesize protein–polymer conjugates? The stability and activity of chymotrypsin-polymer bioconjugates synthesized by RAFT“, *Polymer*, Jg. 72, S. 382–386, 2015.
- [46] G. Pasut, F. Caboi, R. Schrepfer, G. Tonon, O. Schiavon und F. Veronese, „New active poly(ethylene glycol) derivative for amino coupling“, *Reactive and Functional Polymers*, Jg. 67, Nr. 6, S. 529–539, 2007.
- [47] N. Vanparijs, S. Maji, B. Louage, L. Voorhaar, D. Laplace, Q. Zhang, Y. Shi, W. E. Hennink, R. Hoogenboom und B. G. De Geest, „Polymer-protein conjugation via a ‘grafting to’ approach – a comparative study of the performance of protein-reactive RAFT chain transfer agents“, *Polymer Chemistry*, Jg. 6, Nr. 31, S. 5602–5614, 2015.
- [48] A. L. Klibanov, M. A. Sllnkin und V. P. Torchilin, „Conjugation of proteins with chelating polymers via water-soluble carbodiimide and N-hydroxysulfosuccinimide“, *Applied Biochemistry and Biotechnology*, Jg. 22, Nr. 1, S. 45–58, 1989.

- [49] G. Pound, J. M. McKenzie, R. F. M. Lange und B. Klumperman, „Polymer–protein conjugates from α -aldehyde endfunctional poly(N-vinylpyrrolidone) synthesised via xanthate-mediated living radical polymerisation“, *Chemical Communications*, Nr. 27, S. 3193–3195, 2008.
- [50] L. Tao, G. Mantovani, F. Lecolley und D. M. Haddleton, „ α -Aldehyde Terminally Functional Methacrylic Polymers from Living Radical Polymerization: Application in Protein Conjugation “Pegylation”“, *Journal of the American Chemical Society*, Jg. 126, Nr. 41, S. 13 220–13 221, 2004.
- [51] I. Dovgan, S. Kolodych, O. Koniev und A. Wagner, „2-(Maleimidomethyl)-1,3-Dioxanes (MD): A Serum-Stable Self-hydrolysable Hydrophilic Alternative to Classical Maleimide Conjugation“, *Scientific Reports*, Jg. 6, Nr. 1, S. 30 835, 2016.
- [52] Y. Wang und C. Wu, „Quantitative Study of the Oligomerization of Yeast Prion Sup35NM Proteins“, *Biochemistry*, Jg. 56, Nr. 50, S. 6575–6584, 2017.
- [53] Z. Li, R. Huang, H. Xu, J. Chen, Y. Zhan, X. Zhou, H. Chen und B. Jiang, „Divinylsulfonamides as Specific Linkers for Stapling Disulfide Bonds in Peptides“, *Organic Letters*, Jg. 19, Nr. 18, S. 4972–4975, 2017.
- [54] A. Foettinger, M. Melmer, A. Leitner und W. Lindner, „Reaction of the Indole Group with Malondialdehyde: Application for the Derivatization of Tryptophan Residues in Peptides“, *Bioconjugate Chemistry*, Jg. 18, Nr. 5, S. 1678–1683, 2007.
- [55] M. A. Gauthier und H.-A. Klok, „Arginine-Specific Modification of Proteins with Polyethylene Glycol“, *Biomacromolecules*, Jg. 12, Nr. 2, S. 482–493, 2011.
- [56] Y. Cong, E. Pawlisz, P. Bryant, S. Balan, E. Laurine, R. Tommasi, R. Singh, S. Dubey, K. Peciak, M. Bird, A. Sivasankar, J. Swierkosz, M. Muroi, S. Heidelberger, M. Farys, F. Khayrzad, J. Edwards, G. Badescu, I. Hodgson, C. Heise, S. Somavarapu, J. Liddell, K. Powell, M. Zloh, J.-w. Choi, A. Godwin und S. Brocchini, „Site-Specific PEGylation at Histidine Tags“, *Bioconjugate Chemistry*, Jg. 23, Nr. 2, S. 248–263, 2012.
- [57] S. Yamashita, H. Katsumi, N. Hibino, Y. Isobe, Y. Yagi, Y. Tanaka, S. Yamada, C. Naito und A. Yamamoto, „Development of PEGylated aspartic acid-modified liposome as a bone-targeting carrier for the delivery of paclitaxel and treatment of bone metastasis“, *Biomaterials*, Jg. 154, S. 74–85, 2018.
- [58] Y.-A. Lu und A. M. Felix, „Pegylated peptides III. Solid-phase synthesis with pegylating reagents of varying molecular weight: Synthesis of multiply pegylated peptides“, *Reactive Polymers*, Jg. 22, Nr. 3, S. 221–229, 1994.
- [59] M. J. Lobba, C. Fellmann, A. M. Marmelstein, J. C. Maza, E. N. Kissman, S. A. Robinson, B. T. Staahl, C. Urnes, R. J. Lew, C. S. Mogilevsky, J. A. Doudna und M. B. Francis, „Site-Specific Bioconjugation through Enzyme-Catalyzed Tyrosine–Cysteine Bond Formation“, *ACS Central Science*, Jg. 6, Nr. 9, S. 1564–1571, 2020.

- [60] S. C. Reddington und M. Howarth, „Secrets of a covalent interaction for biomaterials and biotechnology: SpyTag and SpyCatcher“, *Current Opinion in Chemical Biology*, Jg. 29, S. 94–99, 2015.
- [61] H. Schiffter-Weinle, „PEG - Das Multitalent“, *Deutsche Apotheker Zeitung*, Nr. 26, 2015.
- [62] L. Radi, M. Fach, M. Montigny, E. Berger-Nicoletti, W. Tremel und P. R. Wich, „Methods of protein surface PEGylation under structure preservation for the emulsion-based formation of stable nanoparticles“, *MedChemComm*, Jg. 7, Nr. 9, S. 1738–1744, 2016.
- [63] Y. Qi und A. Chilkoti, „Protein–polymer conjugation — moving beyond PEGylation“, *Current Opinion in Chemical Biology*, Jg. 28, S. 181–193, 2015.
- [64] J. T. Huckaby, T. M. Jacobs, Z. Li, R. J. Perna, A. Wang, N. I. Nicely und S. K. Lai, „Structure of an anti-PEG antibody reveals an open ring that captures highly flexible PEG polymers“, *Communications Chemistry*, Jg. 3, Nr. 1, S. 124, 2020.
- [65] C. Hohmann-Jeddi, „PEG Als Ein Anaphylaxie-Auslöser Nach Covid-Impfung Bestätigt“. (2021), Adresse: <https://www.pharmazeutische-zeitung.de/peg-als-ein-anaphylaxie-ausloeser-nach-covid-impfung-bestaetigt-124884/> (besucht am 31.01.2021).
- [66] M. D. McSweeney, L. Shen, A. C. DeWalle, J. B. Joiner, E. C. Ciociola, D. Raghuwanshi, M. S. Macauley und S. K. Lai, „Pre-treatment with high molecular weight free PEG effectively suppresses anti-PEG antibody induction by PEG-liposomes in mice“, *Journal of Controlled Release*, Jg. 329, S. 774–781, 2021.
- [67] Y. Kaneda, Y. Tsutsumi, Y. Yoshioka, H. Kamada, Y. Yamamoto, H. Kodaira, S.-i. Tsunoda, T. Okamoto, Y. Mukai, H. Shibata, S. Nakagawa und T. Mayumi, „The use of PVP as a polymeric carrier to improve the plasma half-life of drugs“, *Biomaterials*, Jg. 25, Nr. 16, S. 3259–3266, 2004.
- [68] P. Caliceti, O. Schiavon und F. M. Veronese, „Biopharmaceutical Properties of Uricase Conjugated to Neutral and Amphiphilic Polymers“, *Bioconjugate Chemistry*, Jg. 10, Nr. 4, S. 638–646, 1999.
- [69] P. Zhang, F. Sun, S. Liu und S. Jiang, „Anti-PEG antibodies in the clinic: Current issues and beyond PEGylation“, *Journal of Controlled Release*, Jg. 244, S. 184–193, 2016.
- [70] A. Thomas, S. S. Müller und H. Frey, „Beyond Poly(ethylene glycol): Linear Polyglycerol as a Multifunctional Polyether for Biomedical and Pharmaceutical Applications“, *Biomacromolecules*, Jg. 15, Nr. 6, S. 1935–1954, 2014.
- [71] M. Imran ul-haq, B. F. Lai, R. Chapanian und J. N. Kizhakkedathu, „Influence of architecture of high molecular weight linear and branched polyglycerols on their biocompatibility and biodistribution“, *Biomaterials*, Jg. 33, Nr. 35, S. 9135–9147, 2012.

- [72] R. K. Kainthan und D. E. Brooks, „In vivo biological evaluation of high molecular weight hyperbranched polyglycerols“, *Biomaterials*, Jg. 28, Nr. 32, S. 4779–4787, 2007.
- [73] R. K. Kainthan, S. R. Hester, E. Levin, D. V. Devine und D. E. Brooks, „In vitro biological evaluation of high molecular weight hyperbranched polyglycerols“, *Biomaterials*, Jg. 28, Nr. 31, S. 4581–4590, 2007.
- [74] K. P. Luef, R. Hoogenboom, U. S. Schubert und F. Wiesbrock, „Microwave-Assisted Cationic Ring-Opening Polymerization of 2-Oxazolines“, in *Microwave-Assisted Polymer Synthesis*, R. Hoogenboom, U. S. Schubert und F. Wiesbrock, Bd. 274, Springer International Publishing, 2015, S. 183–208.
- [75] T. Lorson, M. M. Lübtow, E. Wegener, M. S. Haider, S. Borova, D. Nahm, R. Jordan, M. Sokolski-Papkov, A. V. Kabanov und R. Luxenhofer, „Poly(2-oxazoline)s based biomaterials: A comprehensive and critical update“, *Biomaterials*, Jg. 178, S. 204–280, 2018.
- [76] O. Sedlacek, V. R. de la Rosa und R. Hoogenboom, „Poly(2-oxazoline)–protein conjugates“, in *Polymer-Protein Conjugates*, Elsevier, 2020, S. 407–420.
- [77] F. C. Gaertner, R. Luxenhofer, B. Blechert, R. Jordan und M. Essler, „Synthesis, biodistribution and excretion of radiolabeled poly(2-alkyl-2-oxazoline)s“, *Journal of Controlled Release*, Jg. 119, Nr. 3, S. 291–300, 2007.
- [78] T. X. Viegas, M. D. Bentley, J. M. Harris, Z. Fang, K. Yoon, B. Dizman, R. Weimer, A. Mero, G. Pasut und F. M. Veronese, „Polyoxazoline: Chemistry, Properties, and Applications in Drug Delivery“, *Bioconjugate Chemistry*, Jg. 22, Nr. 5, S. 976–986, 2011.
- [79] A. Magarkar, T. Róg und A. Bunker, „A computational study suggests that replacing PEG with PMOZ may increase exposure of hydrophobic targeting moiety“, *European Journal of Pharmaceutical Sciences*, Jg. 103, S. 128–135, 2017.
- [80] Y. Qi, A. Simakova, N. J. Ganson, X. Li, K. M. Luginbuhl, I. Ozer, W. Liu, M. S. Hershfield, K. Matyjaszewski und A. Chilkoti, „A brush-polymer/exendin-4 conjugate reduces blood glucose levels for up to five days and eliminates poly(ethylene glycol) antigenicity“, *Nature Biomedical Engineering*, Jg. 1, Nr. 1, S. 0002, 2017.
- [81] C. J. Wiedermann und M. Joannidis, „Accumulation of hydroxyethyl starch in human and animal tissues: A systematic review“, *Intensive Care Medicine*, Jg. 40, Nr. 2, S. 160–170, 2014.
- [82] G. J. Lonergan, D. B. Cline und S. L. Abbondanzo, „Sickle Cell Anemia“, *RadioGraphics*, Jg. 21, Nr. 4, S. 971–994, 2001.
- [83] K. A. Luke, C. L. Higgins und P. Wittung-Stafshede, „Thermodynamic stability and folding of proteins from hyperthermophilic organisms: Thermodynamic stability and folding of proteins“, *FEBS Journal*, Jg. 274, Nr. 16, S. 4023–4033, 2007.

- [84] C. L. Araya, D. M. Fowler, W. Chen, I. Muniez, J. W. Kelly und S. Fields, „A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function“, *Proceedings of the National Academy of Sciences*, Jg. 109, Nr. 42, S. 16 858–16 863, 2012.
- [85] B. Li, Y. T. Yang, J. A. Capra und M. B. Gerstein, „Predicting changes in protein thermodynamic stability upon point mutation with deep 3D convolutional neural networks“, *PLOS Computational Biology*, Jg. 16, Nr. 11, P. Fariselli, e1008291, 2020.
- [86] S. M. Kelly, T. J. Jess und N. C. Price, „How to study proteins by circular dichroism“, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, Jg. 1751, Nr. 2, S. 119–139, 2005.
- [87] A. Rodger und B. Nordén, *Circular Dichroism and Linear Dichroism*, Ser. Oxford Chemistry Masters 1. Oxford ; New York: Oxford University Press, 1997, 150 S.
- [88] N. J. Greenfield, „Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions“, *Nature Protocols*, Jg. 1, Nr. 6, S. 2527–2535, 2007.
- [89] J. K. Kranz und C. Schalk-Hihi, „Protein Thermal Shifts to Identify Low Molecular Weight Fragments“, in *Methods in Enzymology*, Bd. 493, Elsevier, 2011, S. 277–298.
- [90] K. Huynh und C. L. Partch, „Analysis of Protein Stability and Ligand Interactions by Thermal Shift Assay“, *Current Protocols in Protein Science*, Jg. 79, Nr. 1, S. 28.9.1–28.9.14, 2015.
- [91] A. O. Magnusson, A. Szekrenyi, H.-J. Joosten, J. Finnigan, S. Charnock und W.-D. Fessner, „nanoDSF as screening tool for enzyme libraries and biotechnology development“, *The FEBS Journal*, Jg. 286, Nr. 1, S. 184–204, 2019.
- [92] J. A. Rodríguez-Martínez, R. J. Solá, B. Castillo, H. R. Cintrón-Colón, I. Rivera-Rivera, G. Barletta und K. Griebenow, „Stabilization of α -chymotrypsin upon PEGylation correlates with reduced structural dynamics“, *Biotechnology and Bioengineering*, Jg. 101, Nr. 6, S. 1142–1149, 2008.
- [93] C. Yang, D. Lu und Z. Liu, „How PEGylation Enhances the Stability and Potency of Insulin: A Molecular Dynamics Simulation“, *Biochemistry*, Jg. 50, Nr. 13, S. 2585–2593, 2011.
- [94] E. P. DeBenedictis, E. Hamed und S. Keten, „Mechanical Reinforcement of Proteins with Polymer Conjugation“, *ACS Nano*, Jg. 10, Nr. 2, S. 2259–2267, 2016.
- [95] W. Meng, X. Guo, M. Qin, H. Pan, Y. Cao und W. Wang, „Mechanistic Insights into the Stabilization of srcSH3 by PEGylation“, *Langmuir*, Jg. 28, Nr. 46, S. 16 133–16 140, 2012.

- [96] S. L. Baker, A. Munasinghe, H. Murata, P. Lin, K. Matyjaszewski, C. M. Colina und A. J. Russell, „Intramolecular Interactions of Conjugated Polymers Mimic Molecular Chaperones to Stabilize Protein–Polymer Conjugates“, *Biomacromolecules*, Jg. 19, Nr. 9, S. 3798–3813, 2018.
- [97] B. K. Pandey, M. S. Smith, C. Torgerson, P. B. Lawrence, S. S. Matthews, E. Watkins, M. L. Groves, M. B. Prigozhin und J. L. Price, „Impact of Site-Specific PEGylation on the Conformational Stability and Folding Rate of the Pin WW Domain Depends Strongly on PEG Oligomer Length“, *Bioconjugate Chemistry*, Jg. 24, Nr. 5, S. 796–802, 2013.
- [98] J. L. Price, E. T. Powers und J. W. Kelly, „N-PEGylation of a Reverse Turn Is Stabilizing in Multiple Sequence Contexts, unlike N-GlcNAcylation“, *ACS Chemical Biology*, Jg. 6, Nr. 11, S. 1188–1192, 2011.
- [99] S.-H. Chao, S. S. Matthews, R. Paxman, A. Aksimentiev, M. Gruebele und J. L. Price, „Two Structural Scenarios for Protein Stabilization by PEG“, *The Journal of Physical Chemistry B*, Jg. 118, Nr. 28, S. 8388–8395, 2014.
- [100] P. B. Lawrence, Y. Gavrilov, S. S. Matthews, M. I. Langlois, D. Shental-Bechor, H. M. Greenblatt, B. K. Pandey, M. S. Smith, R. Paxman, C. D. Torgerson, J. P. Merrell, C. C. Ritz, M. B. Prigozhin, Y. Levy und J. L. Price, „Criteria for Selecting PEGylation Sites on Proteins for Higher Thermodynamic and Proteolytic Stability“, *Journal of the American Chemical Society*, Jg. 136, Nr. 50, S. 17547–17560, 2014.
- [101] P. B. Lawrence, W. M. Billings, M. B. Miller, B. K. Pandey, A. R. Stephens, M. I. Langlois und J. L. Price, „Conjugation Strategy Strongly Impacts the Conformational Stability of a PEG-Protein Conjugate“, *ACS Chemical Biology*, Jg. 11, Nr. 7, S. 1805–1809, 2016.
- [102] Q. Xiao, S. R. E. Draper, M. S. Smith, N. Brown, N. A. B. Pugmire, D. S. Ashton, A. J. Carter, E. E. K. Lawrence und J. L. Price, „Influence of PEGylation on the Strength of Protein Surface Salt Bridges“, *ACS Chemical Biology*, Jg. 14, Nr. 7, S. 1652–1659, 2019.
- [103] S. R. E. Draper, D. S. Ashton, B. M. Conover, A. J. Carter, K. L. Stern, Q. Xiao und J. L. Price, „PEGylation near a Patch of Nonpolar Surface Residues Increases the Conformational Stability of the WW Domain“, *The Journal of Organic Chemistry*, Jg. 85, Nr. 3, S. 1725–1730, 2020.
- [104] S. R. E. Draper, Z. B. Jones, S. O. Earl, N. A. Dalley, D. S. Ashton, A. J. Carter, B. M. Conover und J. L. Price, „PEGylation Increases the Strength of a Nearby NH- Hydrogen Bond in the WW Domain“, *Biochemistry*, Jg. 60, Nr. 26, S. 2064–2070, 2021.
- [105] E. Hamed, T. Xu und S. Keten, „Poly(ethylene glycol) Conjugation Stabilizes the Secondary Structure of α -Helices by Reducing Peptide Solvent Accessible Surface Area“, *Biomacromolecules*, Jg. 14, Nr. 11, S. 4053–4060, 2013.
- [106] C.-Y. Liu, X. Li, W.-Y. Chen, L.-C. Chang, Y.-F. Chen, H.-L. Chen, Y.-S. Sun, H.-Y. Lai und E.-W. Huang, „PEGylation Site-Dependent Structural

- Heterogeneity Study of MonoPEGylated Human Parathyroid Hormone Fragment hPTH(1–34)“, *Langmuir*, Jg. 30, Nr. 38, S. 11 421–11 427, 2014.
- [107] A. Munasinghe, A. Mathavan, A. Mathavan, P. Lin und C. M. Colina, „Molecular Insight into the Protein–Polymer Interactions in N-Terminal PEGylated Bovine Serum Albumin“, *The Journal of Physical Chemistry B*, Jg. 123, Nr. 25, S. 5196–5205, 2019.
- [108] B. K. Pandey, M. S. Smith und J. L. Price, „Cys i –Lys $i+3$ –Lys $i+4$ Triad: A General Approach for PEG-Based Stabilization of α -Helical Proteins“, *Biomacromolecules*, Jg. 15, Nr. 12, S. 4643–4647, 2014.
- [109] T. A. Wright, M. Lucius Dougherty, B. Schmitz, K. M. Burrige, K. Makaroff, J. M. Stewart, H. D. Fischesser, J. T. Shepherd, J. A. Berberich, D. Konkolewicz und R. C. Page, „Polymer Conjugation to Enhance Cellulase Activity and Preserve Thermal and Functional Stability“, *Bioconjugate Chemistry*, Jg. 28, Nr. 10, S. 2638–2645, 2017.
- [110] S. Farhadian, B. Shareghi, A. A. Saboury, A. K. Babaheydari, F. Raisi und E. heidari, „Molecular aspects of the interaction of spermidine and α -chymotrypsin“, *International Journal of Biological Macromolecules*, Jg. 92, S. 523–532, 2016.
- [111] C. S. Cummings, A. S. Campbell, S. L. Baker, S. Carmali, H. Murata und A. J. Russell, „Design of Stomach Acid-Stable and Mucin-Binding Enzyme Polymer Conjugates“, *Biomacromolecules*, Jg. 18, Nr. 2, S. 576–586, 2017.
- [112] H. García-Arellano, B. Valderrama, G. Saab-Rincón und R. Vazquez-Duhalt, „High Temperature Biocatalysis by Chemically Modified Cytochrome *c*“, *Bioconjugate Chemistry*, Jg. 13, Nr. 6, S. 1336–1344, 2002.
- [113] B. Plesner, P. Westh und A. D. Nielsen, „Biophysical characterisation of GlycoPEGylated recombinant human factor VIIa“, *International Journal of Pharmaceutics*, Jg. 406, Nr. 1-2, S. 62–68, 2011.
- [114] Y. Nodake und N. Yamasaki, „Some Properties of a Macromolecular Conjugate of Lysozyme Prepared by Modification with a Monomethoxypolyethylene Glycol Derivative“, *Bioscience, Biotechnology, and Biochemistry*, Jg. 64, Nr. 4, S. 767–774, 2000.
- [115] J. Morgenstern, P. Baumann, C. Brunner und J. Hubbuch, „Effect of PEG molecular weight and PEGylation degree on the physical stability of PEGylated lysozyme“, *International Journal of Pharmaceutics*, Jg. 519, Nr. 1-2, S. 408–417, 2017.
- [116] M. Lucius, R. Falatach, C. McGlone, K. Makaroff, A. Danielson, C. Williams, J. C. Nix, D. Konkolewicz, R. C. Page und J. A. Berberich, „Investigating the Impact of Polymer Functional Groups on the Stability and Activity of Lysozyme–Polymer Conjugates“, *Biomacromolecules*, Jg. 17, Nr. 3, S. 1123–1134, 2016.

- [117] A. Natalello, D. Ami, M. Collini, L. D'Alfonso, G. Chirico, G. Tonon, S. Scaramuzza, R. Schrepfer und S. M. Doglia, „Biophysical Characterization of Met-G-CSF: Effects of Different Site-Specific Mono-Pegylations on Protein Stability and Aggregation“, *PLoS ONE*, Jg. 7, Nr. 8, e42511, 2012.
- [118] S. R. E. Draper, P. B. Lawrence, W. M. Billings, Q. Xiao, N. P. Brown, N. A. Bécar, D. J. Matheson, A. R. Stephens und J. L. Price, „Polyethylene Glycol Based Changes to β -Sheet Protein Conformational and Proteolytic Stability Depend on Conjugation Strategy and Location“, *Bioconjugate Chemistry*, Jg. 28, Nr. 10, S. 2507–2513, 2017.
- [119] L. Cao, X. Shi, Y. Cui, W. Yang, G. Chen, L. Yuan und H. Chen, „Protein–polymer conjugates prepared via host–guest interactions: Effects of the conjugation site, polymer type and molecular weight on protein activity“, *Polymer Chemistry*, Jg. 7, Nr. 32, S. 5139–5146, 2016.
- [120] W. Yang, L. Zhu, Y. Cui, H. Wang, Y. Wang, L. Yuan und H. Chen, „Improvement of Site-Directed Protein–Polymer Conjugates: High Bioactivity and Stability Using a Soft Chain-Transfer Agent“, *ACS Applied Materials & Interfaces*, Jg. 8, Nr. 25, S. 15 967–15 974, 2016.
- [121] Q. Mu, T. Hu und J. Yu, „Molecular Insight into the Steric Shielding Effect of PEG on the Conjugated Staphylokinase: Biochemical Characterization and Molecular Dynamics Simulation“, *PLoS ONE*, Jg. 8, Nr. 7, e68559, 2013.
- [122] H. F. Gaertner und A. J. Puigserver, „Increased activity and stability of poly(ethylene glycol)-modified trypsin“, *Enzyme and Microbial Technology*, Jg. 14, Nr. 2, S. 150–155, 1992.
- [123] B. Treetharnmathurot, L. Dieudonné, E. L. Ferguson, D. Schmaljohann, R. Duncan und R. Wiwattanapatapee, „Dextrin–trypsin and ST-HPMA–trypsin conjugates: Enzyme activity, autolysis and thermal stability“, *International Journal of Pharmaceutics*, Jg. 373, Nr. 1-2, S. 68–76, 2009.
- [124] B. J. Alder und T. E. Wainwright, „Phase Transition for a Hard Sphere System“, *The Journal of Chemical Physics*, Jg. 27, Nr. 5, S. 1208–1209, 1957.
- [125] B. J. Alder und T. E. Wainwright, „Studies in Molecular Dynamics. I. General Method“, *The Journal of Chemical Physics*, Jg. 31, Nr. 2, S. 459–466, 1959.
- [126] J. A. McCammon, B. R. Gelin und M. Karplus, „Dynamics of folded proteins“, *Nature*, Jg. 267, Nr. 5612, S. 585–590, 1977.
- [127] S. A. Hollingsworth und R. O. Dror, „Molecular Dynamics Simulation for All“, *Neuron*, Jg. 99, Nr. 6, S. 1129–1143, 2018.
- [128] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt und H. B. Yu, „Biomolekulare Modellierung: Ziele, Probleme, Perspektiven“, *Angewandte Chemie*, Jg. 118, Nr. 25, S. 4168–4198, 2006.
- [129] J. D. Durrant und J. A. McCammon, „Molecular dynamics simulations and drug discovery“, *BMC Biology*, Jg. 9, Nr. 1, S. 71, 2011.

- [130] E. Braun, J. Gilmer, H. B. Mayes, D. L. Mobley, J. I. Monroe, S. Prasad und D. M. Zuckerman, „Best Practices for Foundations in Molecular Simulations [Article v1.0]“, *Living Journal of Computational Molecular Science*, Jg. 1, Nr. 1, S. 5957, 2019.
- [131] S. Riniker, „Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview“, *Journal of Chemical Information and Modeling*, Jg. 58, Nr. 3, S. 565–578, 2018.
- [132] J. D. Hirst, D. R. Glowacki und M. Baaden, „Molecular simulations and visualization: Introduction and overview“, *Faraday Discuss.*, Jg. 169, S. 9–22, 2014.
- [133] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser und C. Simmerling, „ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB“, *Journal of Chemical Theory and Computation*, Jg. 11, Nr. 8, S. 3696–3713, 2015.
- [134] A. Leach, *Molecular Modeling: Principles and Applications*. Pearson Education Limited, 2001.
- [135] D. Frenkel und B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed, Ser. Computational Science Series 1. San Diego: Academic Press, 2002, 638 S.
- [136] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*. Wiley, 2004.
- [137] J.-P. Ryckaert, G. Ciccotti und H. J. Berendsen, „Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes“, *Journal of Computational Physics*, Jg. 23, Nr. 3, S. 327–341, 1977.
- [138] H. C. Andersen, „Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations“, *Journal of Computational Physics*, Jg. 52, Nr. 1, S. 24–34, 1983.
- [139] H. M. Berman, „The Protein Data Bank“, *Nucleic Acids Research*, Jg. 28, Nr. 1, S. 235–242, 2000.
- [140] T. Darden, D. York und L. Pedersen, „Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems“, *The Journal of Chemical Physics*, Jg. 98, Nr. 12, S. 10 089–10 092, 1993.
- [141] B. N. Manjula, A. Tsai, R. Upadhyaya, K. Perumalsamy, P. K. Smith, A. Malavalli, K. Vandegriff, R. M. Winslow, M. Intaglietta, M. Prabhakaran, J. M. Friedman und A. S. Acharya, „Site-Specific PEGylation of Hemoglobin at Cys-93(Beta): Correlation between the Colligative Properties of the PEGylated Protein and the Length of the Conjugated PEG Chain“, *Bioconjugate Chemistry*, Jg. 14, Nr. 2, S. 464–472, 2003.
- [142] H. Lee, „Molecular Simulations of PEGylated Biomolecules, Liposomes, and Nanoparticles for Drug Delivery Applications“, *Pharmaceutics*, Jg. 12, Nr. 6, S. 533, 2020.

- [143] R. Woods und R. Chappelle, „Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates“, *Journal of Molecular Structure: THEOCHEM*, Jg. 527, Nr. 1-3, S. 149–156, 2000.
- [144] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka und T. E. Cheatham, „Assessing the Current State of Amber Force Field Modifications for DNA“, *Journal of Chemical Theory and Computation*, Jg. 12, Nr. 8, S. 4114–4127, 2016.
- [145] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley und R. J. Woods, „GLYCAM06: A generalizable biomolecular force field. Carbohydrates: GLYCAM06“, *Journal of Computational Chemistry*, Jg. 29, Nr. 4, S. 622–655, 2008.
- [146] C. J. Dickson, B. D. Madej, Å. A. Skjerve, R. M. Betz, K. Teigen, I. R. Gould und R. C. Walker, „Lipid14: The Amber Lipid Force Field“, *Journal of Chemical Theory and Computation*, Jg. 10, Nr. 2, S. 865–879, 2014.
- [147] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman und D. A. Case, „Development and testing of a general amber force field“, *Journal of Computational Chemistry*, Jg. 25, Nr. 9, S. 1157–1174, 2004.
- [148] S. F. Sousa, J. Peres, M. Coelho und T. F. Vieira, „Analyzing PEGylation through Molecular Dynamics Simulations“, *ChemistrySelect*, Jg. 3, Nr. 29, S. 8415–8427, 2018.
- [149] H. Lee, R. M. Venable, A. D. MacKerell und R. W. Pastor, „Molecular Dynamics Studies of Polyethylene Oxide and Polyethylene Glycol: Hydrodynamic Radius and Shape Anisotropy“, *Biophysical Journal*, Jg. 95, Nr. 4, S. 1590–1599, 2008.
- [150] J. Fischer, D. Paschek, A. Geiger und G. Sadowski, „Modeling of Aqueous Poly(oxyethylene) Solutions: 1. Atomistic Simulations“, *The Journal of Physical Chemistry B*, Jg. 112, Nr. 8, S. 2388–2398, 2008.
- [151] N. S. V. Barbosa, Y. Zhang, E. R. A. Lima, F. W. Tavares und E. J. Maginn, „Development of an AMBER-compatible transferable force field for poly(ethylene glycol) ethers (glymes)“, *Journal of Molecular Modeling*, Jg. 23, Nr. 6, S. 1–14, 2017.
- [152] T. Ottallah, S. A. Parandian und S. W. Rick, „Analysis of Atomistic Potentials for Poly(ethylene glycol) Ethers“, *Journal of Chemical Theory and Computation*, Jg. 17, Nr. 1, S. 315–321, 2021.
- [153] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror und D. E. Shaw, „Improved side-chain torsion potentials for the Amber ff99SB protein force field: Improved Protein Side-Chain Potentials“, *Proteins: Structure, Function, and Bioinformatics*, Jg. 78, Nr. 8, S. 1950–1958, 2010.
- [154] W. F. van Gunsteren, S. Billeter, A. Eising, P. Hünenberger, P. Krüger, A. Mark, W. Scott und I. Tironi, „Biomolecular Simulation: The GROMOS96 Manual and User Guide“, *Vdf Hochschulverlag AG an der ETH Zürich, Zürich*, Jg. 86, S. 1–1044, 1996.

- [155] B. Isralewitz, M. Gao und K. Schulten, „Steered molecular dynamics and mechanical functions of proteins“, *Current Opinion in Structural Biology*, Jg. 11, Nr. 2, S. 224–230, 2001.
- [156] D. Hamelberg, J. Mongan und J. A. McCammon, „Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules“, *The Journal of Chemical Physics*, Jg. 120, Nr. 24, S. 11 919–11 929, 2004.
- [157] S. J. Marrink, A. H. de Vries und A. E. Mark, „Coarse Grained Model for Semiquantitative Lipid Simulations“, *The Journal of Physical Chemistry B*, Jg. 108, Nr. 2, S. 750–760, 2004.
- [158] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman und S.-J. Marrink, „The MARTINI Coarse-Grained Force Field: Extension to Proteins“, *Journal of Chemical Theory and Computation*, Jg. 4, Nr. 5, S. 819–834, 2008.
- [159] C. A. López, A. J. Rzepiela, A. H. de Vries, L. Dijkhuizen, P. H. Hünenberger und S. J. Marrink, „Martini Coarse-Grained Force Field: Extension to Carbohydrates“, *Journal of Chemical Theory and Computation*, Jg. 5, Nr. 12, S. 3195–3210, 2009.
- [160] J. J. Uusitalo, H. I. Ingólfsson, P. Akhshi, D. P. Tieleman und S. J. Marrink, „Martini Coarse-Grained Force Field: Extension to DNA“, *Journal of Chemical Theory and Computation*, Jg. 11, Nr. 8, S. 3932–3945, 2015.
- [161] J. J. Uusitalo, H. I. Ingólfsson, S. J. Marrink und I. Faustino, „Martini Coarse-Grained Force Field: Extension to RNA“, *Biophysical Journal*, Jg. 113, Nr. 2, S. 246–256, 2017.
- [162] S. Kmieciak, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid und A. Kolinski, „Coarse-Grained Protein Models and Their Applications“, *Chemical Reviews*, Jg. 116, Nr. 14, S. 7898–7936, 2016.
- [163] F. Grünewald, P. C. Kroon, P. C. T. Souza und S. J. Marrink, „Protocol for Simulations of PEGylated Proteins with Martini 3“, in *Structural Genomics*, Y. W. Chen und C.-P. B. Yiu, Bd. 2199, New York, NY: Springer US, 2021, S. 315–335.
- [164] P. V. Coveney und S. Wan, „On the calculation of equilibrium thermodynamic properties from molecular dynamics“, *Physical Chemistry Chemical Physics*, Jg. 18, Nr. 44, S. 30 236–30 240, 2016.
- [165] A. Grossfield und D. M. Zuckerman, „Chapter 2 Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations“, in *Annual Reports in Computational Chemistry*, Bd. 5, Elsevier, 2009, S. 23–48.
- [166] A. Grossfield, P. N. Patrone, D. R. Roe, A. J. Schultz, D. Siderius und D. M. Zuckerman, „Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]“, *Living Journal of Computational Molecular Science*, Jg. 1, Nr. 1, S. 5067, 2019.

- [167] L. Sawle und K. Ghosh, „Convergence of Molecular Dynamics Simulation of Protein Native States: Feasibility vs Self-Consistency Dilemma“, *Journal of Chemical Theory and Computation*, Jg. 12, Nr. 2, S. 861–869, 2016.
- [168] L. J. Smith, X. Daura und W. F. van Gunsteren, „Assessing equilibration and convergence in biomolecular simulations“, *Proteins: Structure, Function, and Genetics*, Jg. 48, Nr. 3, S. 487–496, 2002.
- [169] T. D. Romo und A. Grossfield, „Block Covariance Overlap Method and Convergence in Molecular Dynamics Simulation“, *Journal of Chemical Theory and Computation*, Jg. 7, Nr. 8, S. 2464–2472, 2011.
- [170] W. K. Hastings, „Monte Carlo sampling methods using Markov chains and their applications“, *Biometrika*, Jg. 57, Nr. 1, S. 97–109, 1970.
- [171] U. Krengel, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg+Teubner Verlag, 2005.
- [172] P. Heidelberger und P. D. Welch, „A spectral method for confidence interval generation and run length control in simulations“, *Communications of the ACM*, Jg. 24, Nr. 4, S. 233–245, 1981.
- [173] P. Heidelberger und P. D. Welch, „Simulation Run Length Control in the Presence of an Initial Transient“, *Operations Research*, Jg. 31, Nr. 6, S. 1109–1144, 1983.
- [174] M. Plummer, N. Best, K. Cowles und K. Vines, „CODA: Convergence Diagnosis and Output Analysis for MCMC“, *R News*, Jg. 6, Nr. 1, S. 7–11, 2006.
- [175] J. Geweke, „Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments“, Federal Reserve Bank of Minneapolis, Research Department Staff Report 148, 1991.
- [176] Y. Miao, V. A. Feher und J. A. McCammon, „Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation“, *Journal of Chemical Theory and Computation*, Jg. 11, Nr. 8, S. 3584–3595, 2015.
- [177] Y. T. Pang, Y. Miao, Y. Wang und J. A. McCammon, „Gaussian Accelerated Molecular Dynamics in NAMD“, *Journal of Chemical Theory and Computation*, Jg. 13, Nr. 1, S. 9–19, 2017.
- [178] Y. Miao, W. Sinko, L. Pierce, D. Bucher, R. C. Walker und J. A. McCammon, „Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation“, *Journal of Chemical Theory and Computation*, Jg. 10, Nr. 7, S. 2677–2689, 2014.
- [179] O. F. Lange und H. Grubmüller, „Full correlation analysis of conformational protein dynamics“, *Proteins: Structure, Function, and Bioinformatics*, Jg. 70, Nr. 4, S. 1294–1312, 2007.
- [180] Y. Miao, A. Bhattarai und J. Wang, „Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics“, *Journal of Chemical Theory and Computation*, Jg. 16, Nr. 9, S. 5526–5547, 2020.

- [181] A. Laio und F. L. Gervasio, „Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science“, *Reports on Progress in Physics*, Jg. 71, Nr. 12, S. 126 601, 2008.
- [182] A. Barducci, M. Bonomi und M. Parrinello, „Metadynamics“, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, Jg. 1, Nr. 5, S. 826–843, 2011.
- [183] L. Sutto, S. Marsili und F. L. Gervasio, „New advances in metadynamics: New advances in metadynamics“, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, Jg. 2, Nr. 5, S. 771–779, 2012.
- [184] A. Barducci, G. Bussi und M. Parrinello, „Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method“, *Physical Review Letters*, Jg. 100, Nr. 2, S. 020 603, 2008.
- [185] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti und M. Parrinello, „Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics“, *The Journal of Physical Chemistry B*, Jg. 110, Nr. 8, S. 3533–3539, 2006.
- [186] G. Bussi und D. Branduardi, „Free-Energy Calculations with Metadynamics: Theory and Practice“, in *Reviews in Computational Chemistry*, A. L. Parrill und K. B. Lipkowitz, John Wiley & Sons, Inc, 2015, S. 1–49.
- [187] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia und M. Parrinello, „PLUMED: A portable plugin for free-energy calculations with molecular dynamics“, *Computer Physics Communications*, Jg. 180, Nr. 10, S. 1961–1972, 2009.
- [188] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni und G. Bussi, „PLUMED 2: New feathers for an old bird“, *Computer Physics Communications*, Jg. 185, Nr. 2, S. 604–613, 2014.
- [189] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard und A. Tropsha, „QSAR Modeling: Where Have You Been? Where Are You Going To?“, *Journal of Medicinal Chemistry*, Jg. 57, Nr. 12, S. 4977–5010, 2014.
- [190] C. Hansch, P. P. Maloney, T. Fujita und R. M. Muir, „Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients“, *Nature*, Jg. 194, Nr. 4824, S. 178–180, 1962.
- [191] C. Hansch, „Quantitative approach to biochemical structure-activity relationships“, *Accounts of Chemical Research*, Jg. 2, Nr. 8, S. 232–239, 1969.
- [192] C. Hansch, R. M. Muir, T. Fujita, P. P. Maloney, F. Geiger und M. Streich, „The Correlation of Biological Activity of Plant Growth Regulators and Chloromycetin Derivatives with Hammett Constants and Partition Coefficients“, *Journal of the American Chemical Society*, Jg. 85, Nr. 18, S. 2817–2824, 1963.

- [193] C. Hansch und T. Fujita, „P-sigma-pi Analysis. A Method for the Correlation of Biological Activity and Chemical Structure“, *Journal of the American Chemical Society*, Jg. 86, Nr. 8, S. 1616–1626, 1964.
- [194] S. Wold, „Chemometrics; what do we mean with it, and what do we want from it?“, *Chemometrics and Intelligent Laboratory Systems*, Jg. 30, Nr. 1, S. 109–115, 1995.
- [195] E. Byvatov, U. Fechner, J. Sadowski und G. Schneider, „Comparison of Support Vector Machine and Artificial Neural Network Systems for Drug/Non-drug Classification“, *Journal of Chemical Information and Computer Sciences*, Jg. 43, Nr. 6, S. 1882–1889, 2003.
- [196] M. Damale, S. Harke, F. Kalam Khan, D. Shinde und J. Sangshetti, „Recent Advances in Multidimensional QSAR (4D-6D): A Critical Review“, *Mini-Reviews in Medicinal Chemistry*, Jg. 14, Nr. 1, S. 35–55, 2014.
- [197] S. Wold, M. Sjöström und L. Eriksson, „PLS-regression: A basic tool of chemometrics“, *Chemometrics and Intelligent Laboratory Systems*, Jg. 58, Nr. 2, S. 109–130, 2001.
- [198] R. D. Cramer, D. E. Patterson und J. D. Bunce, „Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins“, *Journal of the American Chemical Society*, Jg. 110, Nr. 18, S. 5959–5967, 1988.
- [199] P. Gramatica und A. Sangion, „A Historical Excursus on the Statistical Validation Parameters for QSAR Models: A Clarification Concerning Metrics and Terminology“, *Journal of Chemical Information and Modeling*, Jg. 56, Nr. 6, S. 1127–1131, 2016.
- [200] P. Yin und X. Fan, „Estimating R^2 Shrinkage in Multiple Regression: A Comparison of Different Analytical Methods“, *The Journal of Experimental Education*, Jg. 69, Nr. 2, S. 203–224, 2001.
- [201] L. M. Shi, H. Fang, W. Tong, J. Wu, R. Perkins, R. M. Blair, W. S. Branham, S. L. Dial, C. L. Moland und D. M. Sheehan, „QSAR Models Using a Large Diverse Set of Estrogens“, *Journal of Chemical Information and Computer Sciences*, Jg. 41, Nr. 1, S. 186–195, 2001.
- [202] G. Schüürmann, R.-U. Ebert, J. Chen, B. Wang und R. Kühne, „External Validation and Prediction Employing the Predictive Squared Correlation Coefficient — Test Set Activity Mean vs Training Set Activity Mean“, *Journal of Chemical Information and Modeling*, Jg. 48, Nr. 11, S. 2140–2145, 2008.
- [203] V. Consonni, D. Ballabio und R. Todeschini, „Comments on the Definition of the Q^2 Parameter for QSAR Validation“, *Journal of Chemical Information and Modeling*, Jg. 49, Nr. 7, S. 1669–1678, 2009.
- [204] L. I.-K. Lin, „A Concordance Correlation Coefficient to Evaluate Reproducibility“, *Biometrics*, Jg. 45, Nr. 1, S. 255, 1989.
- [205] L. I.-K. Lin, „Assay Validation Using the Concordance Correlation Coefficient“, *Biometrics*, Jg. 48, Nr. 2, S. 599–604, 1992.

- [206] M. G. Kendall, „A new measure of rank correlation“, *Biometrika*, Jg. 30, Nr. 1-2, S. 81–93, 1938.
- [207] M. G. Kendall, „The treatment of ties in ranking problems“, *Biometrika*, Jg. 33, Nr. 3, S. 239–251, 1945.
- [208] L. Puka, „Kendall’s Tau“, in *International Encyclopedia of Statistical Science*, M. Lovric, Springer Berlin Heidelberg, 2011, S. 713–715.
- [209] T. Engel und J. Gasteiger, *Chemoinformatics: Basic Concepts and Methods*, Wiley-VCH, 2018.
- [210] T. Fawcett, „An introduction to ROC analysis“, *Pattern Recognition Letters*, Jg. 27, Nr. 8, S. 861–874, 2006.
- [211] M. Kuhn und K. Johnson, *Applied Predictive Modeling*. Springer New York, 2013.
- [212] C. M. Florkowski, „Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests“, *The Clinical Biochemist Reviews*, Jg. 29, S83–S87, Suppl 1 2008.
- [213] J. Davis und M. Goadrich, „The relationship between Precision-Recall and ROC curves“, in *Proceedings of the 23rd International Conference on Machine Learning - ICML ’06*, ACM Press, 2006, S. 233–240.
- [214] T. Saito und M. Rehmsmeier, „The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets“, *PLOS ONE*, Jg. 10, Nr. 3, G. Brock, e0118432, 2015.
- [215] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani und J. P. Overington, „ChEMBL: A large-scale bioactivity database for drug discovery“, *Nucleic Acids Research*, Jg. 40, Nr. D1, S. D1100–D1107, 2012.
- [216] E. N. Muratov, J. Bajorath, R. P. Sheridan, I. V. Tetko, D. Filimonov, V. Poroikov, T. I. Oprea, I. I. Baskin, A. Varnek, A. Roitberg, O. Isayev, S. Curtalolo, D. Fourches, Y. Cohen, A. Aspuru-Guzik, D. A. Winkler, D. Agrafiotis, A. Cherkasov und A. Tropsha, „QSAR without borders“, *Chemical Society Reviews*, Jg. 49, Nr. 11, S. 3525–3564, 2020.
- [217] H. A. Martens und P. Dardenne, „Validation and verification of regression in small data sets“, *Chemometrics and Intelligent Laboratory Systems*, Jg. 44, Nr. 1-2, S. 99–121, 1998.
- [218] M. Kokabi, M. Donnelly und G. Xu, „Benchmarking Small-Dataset Structure-Activity-Relationship Models for Prediction of Wnt Signaling Inhibition“, *IEEE Access*, Jg. 8, S. 228 831–228 840, 2020.
- [219] G. James, D. Witten, T. Hastie und R. Tibshirani, *An Introduction to Statistical Learning*, Ser. Springer Texts in Statistics. Springer New York, 2013, Bd. 103.
- [220] S. Tibshirani und H. Friedman, *The Elements of Statistical Learning*, Ser. Springer Texts in Statistics. Springer New York, 2015.

- [221] R. Febrianti, Y. Widyaningsih und S. Soemartojo, „The Parameter Estimation of Logistic Regression with Maximum Likelihood Method and Score Function Modification“, *Journal of Physics: Conference Series*, Jg. 1725, S. 012 014, 2021.
- [222] T. Chen und C. Guestrin, „XGBoost: A Scalable Tree Boosting System“, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, 2016, S. 785–794.
- [223] Distributed (Deep) Machine Learning Community. „Machine Learning Challenge Winning Solutions“. (2021), Adresse: <https://github.com/dmlc/xgboost/tree/master/demo#machine-learning-challenge-winning-solutions> (besucht am 31.01.2021).
- [224] R. P. Sheridan, W. M. Wang, A. Liaw, J. Ma und E. M. Gifford, „Extreme Gradient Boosting as a Method for Quantitative Structure–Activity Relationships“, *Journal of Chemical Information and Modeling*, Jg. 56, Nr. 12, S. 2353–2360, 2016.
- [225] I. H. Rekand und R. Brenk, „DrugPred_RNA—A Tool for Structure-Based Druggability Predictions for RNA Binding Sites“, *Journal of Chemical Information and Modeling*, Jg. 61, Nr. 8, S. 4068–4081, 2021.
- [226] XGBoost developers. „Introduction to Boosted Trees“. (2020), Adresse: <https://xgboost.readthedocs.io/en/latest/tutorials/model.html> (besucht am 31.01.2021).
- [227] S. H. Unger und C. Hansch, „Model building in structure-activity relations. Re-examination of adrenergic blocking activity of Beta-Halo-Beta-arylalkylamines“, *Journal of Medicinal Chemistry*, Jg. 16, Nr. 7, S. 745–749, 1973.
- [228] A. Tropsha, „Best Practices for QSAR Model Development, Validation, and Exploitation“, *Molecular Informatics*, Jg. 29, Nr. 6-7, S. 476–488, 2010.
- [229] OECD. „OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models“. (2004), Adresse: <https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf> (besucht am 30.01.2021).
- [230] W. L. Jorgensen, „QSAR/QSPR and Proprietary Data“, *Journal of Chemical Information and Modeling*, Jg. 46, Nr. 3, S. 937–937, 2006.
- [231] S. Radestock und H. Gohlke, „Exploiting the Link between Protein Rigidity and Thermostability for Data-Driven Protein Engineering“, *Engineering in Life Sciences*, Jg. 8, Nr. 5, S. 507–522, 2008.
- [232] S. Fulle und H. Gohlke, „Analyzing the Flexibility of RNA Structures by Constraint Counting“, *Biophysical Journal*, Jg. 94, Nr. 11, S. 4202–4219, 2008.
- [233] S. Radestock und H. Gohlke, „Protein rigidity and thermophilic adaptation“, *Proteins: Structure, Function, and Bioinformatics*, Jg. 79, Nr. 4, S. 1089–1108, 2011.
- [234] C. Pflieger, P. C. Rathi, D. L. Klein, S. Radestock und H. Gohlke, „Constraint Network Analysis (CNA): A Python Software Package for Efficiently Linking

- Biomacromolecular Structure, Flexibility, (Thermo-)Stability, and Function“, *Journal of Chemical Information and Modeling*, Jg. 53, Nr. 4, S. 1007–1015, 2013.
- [235] C. Pfleger, S. Radestock, E. Schmidt und H. Gohlke, „Global and local indices for characterizing biomolecular flexibility and rigidity“, *Journal of Computational Chemistry*, Jg. 34, Nr. 3, S. 220–233, 2013.
- [236] D. J. Jacobs, A. Rader, L. A. Kuhn und M. Thorpe, „Protein flexibility predictions using graph theory“, *Proteins: Structure, Function, and Genetics*, Jg. 44, Nr. 2, S. 150–165, 2001.
- [237] H. Gohlke, L. A. Kuhn und D. A. Case, „Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach“, *Proteins: Structure, Function, and Bioinformatics*, Jg. 56, Nr. 2, S. 322–337, 2004.
- [238] C. Pfleger und H. Gohlke, „Efficient and Robust Analysis of Biomacromolecular Flexibility Using Ensembles of Network Topologies Based on Fuzzy Non-covalent Constraints“, *Structure*, Jg. 21, Nr. 10, S. 1725–1734, 2013.
- [239] P. C. Rathi, S. Radestock und H. Gohlke, „Thermostabilizing mutations preferentially occur at structural weak spots with a high mutation ratio“, *Journal of Biotechnology*, Jg. 159, Nr. 3, S. 135–144, 2012.
- [240] J. C. Maxwell, „L. On the calculation of the equilibrium and stiffness of frames“, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Jg. 27, Nr. 182, S. 294–299, 1864.
- [241] B. Hendrickson, „Conditions for Unique Graph Realizations“, *SIAM Journal on Computing*, Jg. 21, Nr. 1, S. 65–84, 1992.
- [242] A. Sljoka, „Counting for rigidity, flexibility and extensions via the pebble game algorithm“, Dissertation, York University, 2006.
- [243] S. M. Hermans, C. Pfleger, C. Nutschel, C. A. Hanke und H. Gohlke, „Rigidity theory for biomolecules: Concepts, software, and applications: Rigidity theory for biomolecules“, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, Jg. 7, Nr. 4, e1311, 2017.
- [244] A. Lee, I. Streinu und L. Theran, „Graded Sparse Graphs and Matroids“, *Journal of Universal Computer Science*, Jg. 13, Nr. 11, S. 1671–1679, 2007.
- [245] N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch und G. R. Hutchison, „Open Babel: An open chemical toolbox“, *Journal of Cheminformatics*, Jg. 3, Nr. 1, S. 1–14, 2011.
- [246] N. M. O’Boyle, C. Morley und G. R. Hutchison, „Pybel: A Python wrapper for the OpenBabel cheminformatics toolkit“, *Chemistry Central Journal*, Jg. 2, Nr. 1, S. 1–7, 2008.
- [247] R. Ranganathan, K. P. Lu, T. Hunter und J. P. Noel, „Structural and Functional Analysis of the Mitotic Rotamase Pin1 Suggests Substrate Recognition Is Phosphorylation Dependent“, *Cell*, Jg. 89, Nr. 6, S. 875–886, 1997.

- [248] E. Bayer, S. Goettsch, J. W. Mueller, B. Griewel, E. Guiberman, L. M. Mayr und P. Bayer, „Structural Analysis of the Mitotic Regulator hPin1 in Solution: Insights into Domain Architecture and Substrate Binding“, *Journal of Biological Chemistry*, Jg. 278, Nr. 28, S. 26 183–26 193, 2003.
- [249] M. Jäger, H. Nguyen, J. C. Crane, J. W. Kelly und M. Gruebele, „The folding mechanism of a beta-sheet: The WW domain“, *Journal of Molecular Biology*, Jg. 311, Nr. 2, S. 373–393, 2001.
- [250] C. M. Davis und R. B. Dyer, „WW Domain Folding Complexity Revealed by Infrared Spectroscopy“, *Biochemistry*, Jg. 53, Nr. 34, S. 5476–5484, 2014.
- [251] G. T. Ibragimova und R. C. Wade, „Stability of the Beta-Sheet of the WW Domain: A Molecular Dynamics Simulation Study“, *Biophysical Journal*, Jg. 77, Nr. 4, S. 2191–2198, 1999.
- [252] H. Nguyen, M. Jager, A. Moretto, M. Gruebele und J. W. Kelly, „Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation“, *Proceedings of the National Academy of Sciences*, Jg. 100, Nr. 7, S. 3948–3953, 2003.
- [253] J. Karanicolas und C. L. Brooks, „Integrating folding kinetics and protein function: Biphasic kinetics and dual binding specificity in a WW domain“, *Proceedings of the National Academy of Sciences*, Jg. 101, Nr. 10, S. 3432–3437, 2004.
- [254] M. Jager, Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. P. Noel, M. Gruebele und J. W. Kelly, „Structure-function-folding relationship in a WW domain“, *Proceedings of the National Academy of Sciences*, Jg. 103, Nr. 28, S. 10 648–10 653, 2006.
- [255] J. Juraszek und P. G. Bolhuis, „(Un)Folding Mechanisms of the FBP28 WW Domain in Explicit Solvent Revealed by Multiple Rare Event Simulation Methods“, *Biophysical Journal*, Jg. 98, Nr. 4, S. 646–656, 2010.
- [256] J. Karanicolas und C. L. Brooks, „The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: Lessons for protein design?“, *Proceedings of the National Academy of Sciences*, Jg. 100, Nr. 7, S. 3954–3959, 2003.
- [257] T. Sharpe, A. L. Jonsson, T. J. Rutherford, V. Daggett und A. R. Fersht, „The role of the turn in β -hairpin formation during WW domain folding“, *Protein Science*, Jg. 16, Nr. 10, S. 2233–2239, 2007.
- [258] Z. Luo, J. Ding und Y. Zhou, „Folding mechanisms of individual β -hairpins in a Gō model of Pin1 WW domain by all-atom molecular dynamics simulations“, *The Journal of Chemical Physics*, Jg. 128, Nr. 22, S. 225 103, 2008.
- [259] Z. Luo, J. Ding und Y. Zhou, „Temperature-Dependent Folding Pathways of Pin1 WW Domain: An All-Atom Molecular Dynamics Simulation of a Gō Model“, *Biophysical Journal*, Jg. 93, Nr. 6, S. 2152–2161, 2007.
- [260] H. Gelman und M. Gruebele, „Fast protein folding kinetics“, *Quarterly Reviews of Biophysics*, Jg. 47, Nr. 2, S. 95–142, 2014.

- [261] J. Löfblom, J. Feldwisch, V. Tolmachev, J. Carlsson, S. Ståhl und F. Frejd, „Affibody molecules: Engineered proteins for therapeutic, diagnostic and biotechnological applications“, *FEBS Letters*, Jg. 584, Nr. 12, S. 2670–2680, 2010.
- [262] C. Eigenbrot, M. Ultsch, A. Dubnovitsky, L. Abrahmsen und T. Hard, „Structural basis for high-affinity HER2 receptor binding by an engineered protein“, *Proceedings of the National Academy of Sciences*, Jg. 107, Nr. 34, S. 15 039–15 044, 2010.
- [263] H. Yu, M. K. Rosen und S. L. Schreiber, „ ^1H and ^{15}N assignments and secondary structure of the Src SH3 domain“, *FEBS Letters*, Jg. 324, Nr. 1, S. 87–92, 1993.
- [264] K. L. Stern, M. S. Smith, W. M. Billings, T. J. Loftus, B. M. Conover, D. Della Corte und J. L. Price, „Context-Dependent Stabilizing Interactions among Solvent-Exposed Residues along the Surface of a Trimeric Helix Bundle“, *Biochemistry*, Jg. 59, Nr. 17, S. 1672–1679, 2020.
- [265] E. O’Shea, J. Klemm, P. Kim und T. Alber, „X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil“, *Science*, Jg. 254, Nr. 5031, S. 539–544, 1991.
- [266] M. S. Weiss und D. Eisenberg, „The crystal structure of the designed trimeric coiled coil coil-V₁Ld: Implications for engineering crystals and supramolecular assemblies“, *Protein Science*, Jg. 6, Nr. 1, S. 80–88, 1997.
- [267] M. S. Smith, W. M. Billings, F. G. Whitby, M. B. Miller und J. L. Price, „Enhancing a long-range salt bridge with intermediate aromatic and nonpolar amino acids“, *Organic & Biomolecular Chemistry*, Jg. 15, Nr. 28, S. 5882–5886, 2017.
- [268] K. Aktories, U. Förstermann, F. Hofmann und K. Starke, *Allgemeine Und Spezielle Pharmakologie Und Toxikologie*. Elsevier, 2018.
- [269] L. Zinser und F. Antwerpes. „Peginterferon Alfa - DocCheck Flexikon“. (2021), Adresse: https://flexikon.doccheck.com/de/Peginterferon_alfa (besucht am 31.01.2021).
- [270] C. Sarrazin, T. Zimmermann, T. Berg, U. Neumann, P. Schirmacher, H. Schmidt, U. Spengler, J. Timm, H. Wedemeyer, S. Wirth, S. Zeuzem, Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS), Deutsche Gesellschaft für Pathologie e.V. (DGP), Deutsche Leberstiftung, Gesellschaft für Virologie e.V. (GfV), Gesellschaft für Pädiatrische Gastroenterologie und Ernährung (GPGE), Österreichische Gesellschaft für Gastroenterologie und Hepatologie (ÖGGH), Schweizerische Gesellschaft für Gastroenterologie (SGG), Deutsche Transplantationsgesellschaft e.V. (DTG), Deutsche Leberhilfe e.V., Deutsche Gesellschaft für Infektiologie e.V. (DGI), Deutsche Gesellschaft für Suchtmedizin e.V. (DGS), Deutsche AIDS-Gesellschaft e.V. (DAIG), Deutsche Arbeitsgemeinschaft niedergelassener Ärzte für die Versorgung HIV-Infizierter (DAGNÄ) und Robert Koch-Institut, „S3-Leitlinie „Prophylaxe, Diagnostik und Therapie der Hepatitis-C-Virus (HCV) -Infektion“:

- AWMF-Register-Nr.: 021/012“, *Zeitschrift für Gastroenterologie*, Jg. 56, Nr. 07, S. 756–838, 2018.
- [271] M. T. Dill, Z. Makowska, G. Trincucci, A. J. Gruber, J. E. Vogt, M. Filipowicz, D. Calabrese, I. Krol, D. T. Lau, L. Terracciano, E. van Nimwegen, V. Roth und M. H. Heim, „Pegylated IFN- regulates hepatic gene expression through transient Jak/STAT activation“, *Journal of Clinical Investigation*, Jg. 124, Nr. 4, S. 1568–1581, 2014.
- [272] C. Thomas, I. Moraga, D. Levin, P. O. Krutzik, Y. Podoplelova, A. Trejo, C. Lee, G. Yarden, S. E. Vleck, J. S. Glenn, G. P. Nolan, J. Piehler, G. Schreiber und K. C. Garcia, „Structural Linkage between Ligand Discrimination and Receptor Activation by Type I Interferons“, *Cell*, Jg. 146, Nr. 4, S. 621–632, 2011.
- [273] W. Klaus, B. Gsell, A. M. Labhardt, B. Wipf und H. Senn, „The three-dimensional high resolution structure of human interferon γ determined by heteronuclear NMR spectroscopy in solution“, *Journal of Molecular Biology*, Jg. 274, Nr. 4, S. 661–675, 1997.
- [274] MOE, *Molecular Operating Environment (MOE), 2020.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7*, 2020.
- [275] C. I. Bayly, P. Cieplak, W. Cornell und P. A. Kollman, „A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model“, *The Journal of Physical Chemistry*, Jg. 97, Nr. 40, S. 10 269–10 280, 1993.
- [276] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski und P. Cieplak, „The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building“, *Physical Chemistry Chemical Physics*, Jg. 12, Nr. 28, S. 7821–7839, 2010.
- [277] E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak und F.-Y. Dupradeau, „R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments“, *Nucleic Acids Research*, Jg. 39, W511–W517, suppl_2 2011.
- [278] F. Wang, J.-P. Becker, P. Cieplak und F.-Y. Dupradeau, „R.E.D. Python: Object Oriented Programming for Amber Force Fields“, in *Abstracts of Papers of the American Chemical Society*, Bd. 247, American Chemical Society, 2014.
- [279] P. R. Gerber und K. Müller, „MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry“, *Journal of Computer-Aided Molecular Design*, Jg. 9, Nr. 3, S. 251–268, 1995.
- [280] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F.

- Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman und D. J. Fox, *Gaussian 16 Rev. B.01*, Wallingford, CT, 2016.
- [281] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman und D. J. Fox, *Gaussian 09 Rev. C.01*, Wallingford, CT, 2009.
- [282] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York und P.A. Kollman, *AMBER 2018*, University of California, San Francisco, 2018.
- [283] J. Wang, W. Wang, P. A. Kollman und D. A. Case, „Antechamber, An Accessory Software Package For Molecular Mechanical Calculations“, *Submitted to Journal of Chemical Information and Computer Sciences*, 2000.
- [284] A. T. P. Carvalho, P. A. Fernandes und M. J. Ramos, „Parameterization of AZT—A widely used nucleoside inhibitor of HIV-1 reverse transcriptase“, *International Journal of Quantum Chemistry*, Jg. 107, Nr. 2, S. 292–298, 2007.
- [285] G. Pieffet und P. A. Petukhov, „Parameterization of aromatic azido groups: Application as photoaffinity probes in molecular dynamics studies“, *Journal of Molecular Modeling*, Jg. 15, Nr. 11, S. 1291–1297, 2009.

- [286] R. S. Mulliken, „Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I“, *The Journal of Chemical Physics*, Jg. 23, Nr. 10, S. 1833–1840, 1955.
- [287] C. Tian, K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Miguez, J. Bickel, Y. Wang, J. Pincay, Q. Wu und C. Simmerling, „ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution“, *Journal of Chemical Theory and Computation*, Jg. 16, Nr. 1, S. 528–552, 2020.
- [288] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey und M. L. Klein, „Comparison of simple potential functions for simulating liquid water“, *The Journal of Chemical Physics*, Jg. 79, Nr. 2, S. 926–935, 1983.
- [289] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé und K. Schulten, „Scalable molecular dynamics with NAMD“, *Journal of Computational Chemistry*, Jg. 26, Nr. 16, S. 1781–1802, 2005.
- [290] T. D. Romo, N. Leioatts und A. Grossfield, „Lightweight object oriented structure analysis: Tools for building tools to analyze molecular dynamics simulations“, *Journal of Computational Chemistry*, Jg. 35, Nr. 32, S. 2305–2318, 2014.
- [291] T. Romo und A. Grossfield, „LOOS: An Extensible Platform for the Structural Analysis of Simulations“, in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, MN: IEEE, 2009, S. 2332–2335.
- [292] M. Kuhn, „Building Predictive Models in R using the caret package“, *Journal of Statistical Software*, Jg. 28, Nr. 1, S. 1–26, 2008.
- [293] RStudio. „Max Kuhn - Profil“. (2021), Adresse: <https://www.rstudio.com/speakers/max-kuhn/> (besucht am 31.01.2021).
- [294] I. Guyon, J. Weston und S. Barnhill, „Gene Selection for Cancer Classification using Support Vector Machines“, *Machine Learning*, Jg. 46, Nr. 1, S. 389–422, 2002.
- [295] S. Kirkpatrick, C. D. Gelatt Jr. und M. P. Vecchi, „Optimization by Simulated Annealing“, *Science*, Jg. 220, Nr. 4598, S. 671–680, 1983.
- [296] L. Scrucca, „GA : A Package for Genetic Algorithms in R“, *Journal of Statistical Software*, Jg. 53, Nr. 1, S. 1–137, 2013.
- [297] P. Carrió, M. Pinto, G. Ecker, F. Sanz und M. Pastor, „Applicability Domain Analysis (ADAN): A Robust Method for Assessing the Reliability of Drug Property Predictions“, *Journal of Chemical Information and Modeling*, Jg. 54, Nr. 5, S. 1500–1511, 2014.
- [298] P. Filzmoser, B. Liebmann und K. Varmuza, „Repeated double cross validation“, *Journal of Chemometrics*, Jg. 23, Nr. 4, S. 160–171, 2009.
- [299] K. Varmuza und P. Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press, 2009.

- [300] R. L. Thorndike, „Who belongs in the family?“, *Psychometrika*, Jg. 18, Nr. 4, S. 267–276, 1953.
- [301] PLUMED Manual. „PLUMED: UPPER_WALLS“. (2021), Adresse: https://www.plumed.org/doc-v2.7/user-doc/html/_u_p_p_e_r__w_a_l_l_s.html (besucht am 26.11.2021).
- [302] D. Trapl und V. Spiwok, „Analysis of the Results of Metadynamics Simulations by Metadynminer and Metadynminer3d“, *arXiv Preprint (arXiv:2009.02241)*, 2020.
- [303] S. Foser, A. Schacher, K. A. Weyer, D. Brugger, E. Dietel, S. Marti und T. Schreitmüller, „Isolation, structural characterization, and antiviral activity of positional isomers of monopegylated interferon α -2a (PEGASYS)“, *Protein Expression and Purification*, Jg. 30, Nr. 1, S. 78–87, 2003.
- [304] N. Hauptstein, P. Pouyan, J. Kehrein, M. Dirauf, M. D. Driessen, M. Raschig, K. Licha, M. Gottschaldt, U. S. Schubert, R. Haag, L. Meinel, C. Sotriffer und T. Lühmann, „Molecular Insights into Site-Specific Interferon- α 2a Bioconjugates Originated from PEG, LPG, and PEtOx“, *Biomacromolecules*, Jg. 22, Nr. 11, S. 4521–4534, 2021.
- [305] *Schrödinger Release 2018-1: Epik*, Schrödinger, LLC, New York, NY, 2018.
- [306] S. D. Bond, B. J. Leimkuhler und B. B. Laird, „The Nosé–Poincaré Method for Constant Temperature Molecular Dynamics“, *Journal of Computational Physics*, Jg. 151, Nr. 1, S. 114–134, 1999.
- [307] J. B. Sturgeon und B. B. Laird, „Symplectic algorithm for constant-pressure molecular dynamics using a Nosé–Poincaré thermostat“, *The Journal of Chemical Physics*, Jg. 112, Nr. 8, S. 3474–3482, 2000.
- [308] W. C. Still, A. Tempczyk, R. C. Hawley und T. Hendrickson, „Semianalytical treatment of solvation for molecular mechanics and dynamics“, *Journal of the American Chemical Society*, Jg. 112, Nr. 16, S. 6127–6129, 1990.
- [309] A. Munasinghe, A. Mathavan, A. Mathavan, P. Lin und C. M. Colina, „Atomistic insight towards the impact of polymer architecture and grafting density on structure-dynamics of PEGylated bovine serum albumin and their applications“, *The Journal of Chemical Physics*, Jg. 154, Nr. 7, S. 075101, 2021.
- [310] M. E. Fortunato und C. M. Colina, „Pysimm: A python package for simulation of molecular systems“, *SoftwareX*, Jg. 6, S. 7–12, 2017.
- [311] M. E. Fortunato und C. M. Colina. „Pysimm“. (2021), Adresse: <https://github.com/polysimtools/pysimm>. (besucht am 31.01.2021).
- [312] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman und A. H. de Vries, „The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations“, *The Journal of Physical Chemistry B*, Jg. 111, Nr. 27, S. 7812–7824, 2007.
- [313] F. Grunewald, G. Rossi, A. H. de Vries, S. J. Marrink und L. Monticelli, „Transferable MARTINI Model of Poly(ethylene Oxide)“, *The Journal of Physical Chemistry B*, Jg. 122, Nr. 29, S. 7436–7449, 2018.

- [314] F. Grünewald, R. Alessandri, P. C. Kroon, L. Monticelli, P. C. T. Souza und S. J. Marrink, „Polyply: A Python Suite for Facilitating Simulations of (Bio-)Macromolecules and Nanomaterials“, *arXiv Preprint (arXiv:2105.05890)*, 2021.
- [315] Invivogen. „IFN-Alpha/Beta Reporter HEK 293 Cells“. (2021), Adresse: <https://www.invivogen.com/hek-blue-ifn-ab> (besucht am 19.08.2021).
- [316] Y. Feng, G. De Franceschi, A. Kahraman, M. Soste, A. Melnik, P. J. Boersema, P. P. de Lauroto, Y. Nikolaev, A. P. Oliveira und P. Picotti, „Global analysis of protein structural changes in complex proteomes“, *Nature Biotechnology*, Jg. 32, Nr. 10, S. 1036–1044, 2014.
- [317] K. M. Merz, „Limits of Free Energy Computation for Protein–Ligand Interactions“, *Journal of Chemical Theory and Computation*, Jg. 6, Nr. 5, S. 1769–1776, 2010.
- [318] B. K. Shoichet, W. P. Walters, H. Jiang und J. Bajorath, „Advances in Computational Medicinal Chemistry: A Reflection on the Evolution of the Field and Perspective Going Forward“, *Journal of Medicinal Chemistry*, Jg. 59, Nr. 9, S. 4033–4034, 2016.
- [319] M. Gevrey, I. Dimopoulos und S. Lek, „Review and comparison of methods to study the contribution of variables in artificial neural network models“, *Ecological Modelling*, Jg. 160, Nr. 3, S. 249–264, 2003.
- [320] J. Heaton, *Introduction to Neural Networks with Java*. 2008.
- [321] G. Henkelman und H. Jónsson, „Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points“, *The Journal of Chemical Physics*, Jg. 113, Nr. 22, S. 9978–9985, 2000.
- [322] L. Hahn, T. Zorn, J. Kehrein, T. Kielholz, A.-L. Ziegler, B. Sochor, E. S. Lisitsyna, N. A. Durandin, T. Laaksonen, V. Aseyev, C. Sotriffer, M. Windbergs, A.-C. Pöppler und R. Luxenhofer, „Unravelling a novel mechanism in polymer self-assemblies: An order-order transition based on molecular interactions between hydrophilic and hydrophobic polymer blocks“, *ChemRxiv Preprint (DOI: 10.33774/chemrxiv-2021-mjd8n-v2)*, 2021.
- [323] R. W. Newberry und R. T. Raines, „The $n \rightarrow *$ Interaction“, *Accounts of Chemical Research*, Jg. 50, Nr. 8, S. 1838–1846, 2017.
- [324] B. C. Gorske, B. L. Bastian, G. D. Geske und H. E. Blackwell, „Local and Tunable $n \rightarrow *$ Interactions Regulate Amide Isomerism in the Peptoid Backbone“, *Journal of the American Chemical Society*, Jg. 129, Nr. 29, S. 8928–8929, 2007.
- [325] A.-C. Pöppler, M. M. Lübtow, J. Schlauersbach, J. Wiest, L. Meinel und R. Luxenhofer, „Loading dependent Structural Model of Polymeric Micelles Encapsulating Curcumin by Solid-State NMR Spectroscopy“, *Angewandte Chemie International Edition*, 2019.
- [326] W. Humphrey, A. Dalke und K. Schulten, „VMD: Visual Molecular Dynamics“, *J. Mol. Graphics*, Jg. 14, S. 33–36, 1996.

- [327] F. Müller-Plathe und W. F. van Gunsteren, „Computer simulation of a polymer electrolyte: Lithium iodide in amorphous poly(ethylene oxide)“, *The Journal of Chemical Physics*, Jg. 103, Nr. 11, S. 4745–4756, 1995.
- [328] K. Laasonen und M. L. Klein, „Molecular dynamics simulations of the structure and ion diffusion in poly(ethylene oxide)“, *Journal of the Chemical Society, Faraday Transactions*, Jg. 91, Nr. 16, S. 2633, 1995.
- [329] F. Ramezanghorbani, P. Lin und C. M. Colina, „Optimizing Protein–Polymer Interactions in a Poly(ethylene glycol) Coarse-Grained Model“, *The Journal of Physical Chemistry B*, Jg. 122, Nr. 33, S. 7997–8005, 2018.
- [330] S. S. Pai, B. Hammouda, K. Hong, D. C. Pozzo, T. M. Przybycien und R. D. Tilton, „The Conformation of the Poly(ethylene glycol) Chain in Mono-PEGylated Lysozyme and Mono-PEGylated Human Growth Hormone“, *Bioconjugate Chemistry*, Jg. 22, Nr. 11, S. 2317–2323, 2011.
- [331] C. Andraud, A. Beghdadi und J. Lafait, „Entropic analysis of random morphologies“, *Physica A: Statistical Mechanics and its Applications*, Jg. 207, Nr. 1-3, S. 208–212, 1994.
- [332] G.-J. Bekker, B. Ma und N. Kamiya, „Thermal stability of single-domain antibodies estimated by molecular dynamics simulations“, *Protein Science*, Jg. 28, Nr. 2, S. 429–438, 2019.
- [333] N. Hauptstein, P. Pouyan, K. Wittwer, G. Cinar, O. Scherf-Clavel, K. Licha, T. Lühmann, I. Nischang, U. S. Schubert, C. Pfaller, R. Haag und L. Meinel, „Polymer selection impacts the pharmaceutical profile of site specifically conjugated Interferon-alpha2a“, *ChemRxiv Preprint (DOI: 10.26434/chemrxiv-2022-j85br)*, 2022.
- [334] N. Hauptstein, M. Dirauf, K. Wittwer, B. Sawatsky, G. Cinar, O. Scherf-Clavel, M. Raschig, T. Lühmann, I. Nischang, U. S. Schubert, C. Pfaller und L. Meinel, „Site Specific PEtOxylation of Interferon-Alpha2a“, *unpubliziert (eingereicht)*, 2022.
- [335] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno und M. Karttunen, „Multiscale modeling of emergent materials: Biological and soft matter“, *Physical Chemistry Chemical Physics*, Jg. 11, Nr. 12, S. 1869, 2009.
- [336] D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov und S. Vajda, „The ClusPro web server for protein–protein docking“, *Nature Protocols*, Jg. 12, Nr. 2, S. 255–278, 2017.
- [337] K. A. Porter, I. Desta, D. Kozakov und S. Vajda, „What method to use for protein–protein docking?“, *Current Opinion in Structural Biology*, Jg. 55, S. 1–7, 2019.
- [338] I. A. Vakser, „Protein-Protein Docking: From Interaction to Interactome“, *Biophysical Journal*, Jg. 107, Nr. 8, S. 1785–1793, 2014.
- [339] Y. Yan, H. Tao, J. He und S.-Y. Huang, „The HDock server for integrated protein–protein docking“, *Nature Protocols*, Jg. 15, Nr. 5, S. 1829–1852, 2020.

- [340] J. Estrada, P. Bernado, M. Blackledge und J. Sancho, „ProtSA: A web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble“, *BMC Bioinformatics*, Jg. 10, Nr. 1, S. 104, 2009.
- [341] O. Conchillo-Solé, N. S. de Groot, F. X. Avilés, J. Vendrell, X. Daura und S. Ventura, „AGGRESCAN: A server for the prediction and evaluation of ”hot spots“ of aggregation in polypeptides“, *BMC Bioinformatics*, Jg. 8, Nr. 1, S. 65, 2007.
- [342] R. Zambrano, M. Jamroz, A. Szczasiuk, J. Pujols, S. Kmiecik und S. Ventura, „AGGRESCAN3D (A3D): Server for prediction of aggregation properties of protein structures“, *Nucleic Acids Research*, Jg. 43, Nr. W1, W306–W313, 2015.
- [343] M. Gil-Garcia, M. Bañó-Polo, N. Varejão, M. Jamroz, A. Kuriata, M. Díaz-Caballero, J. Lascorz, B. Morel, S. Navarro, D. Reverter, S. Kmiecik und S. Ventura, „Combining Structural Aggregation Propensity and Stability Predictions To Redesign Protein Solubility“, *Molecular Pharmaceutics*, Jg. 15, Nr. 9, S. 3846–3859, 2018.
- [344] T. Zeiske, K. A. Stafford und A. G. Palmer, „Thermostability of Enzymes from Molecular Dynamics Simulations“, *Journal of Chemical Theory and Computation*, Jg. 12, Nr. 6, S. 2489–2492, 2016.
- [345] S. Park, F. Khalili-Araghi, E. Tajkhorshid und K. Schulten, „Free energy calculation from steered molecular dynamics simulations using Jarzynski’s equality“, *The Journal of Chemical Physics*, Jg. 119, Nr. 6, S. 3559–3566, 2003.
- [346] S. Decherchi und A. Cavalli, „Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation“, *Chemical Reviews*, Jg. 120, Nr. 23, S. 12 788–12 833, 2020.
- [347] S. Takada, „Gō model revisited“, *Biophysics and Physicobiology*, Jg. 16, Nr. 0, S. 248–255, 2019.
- [348] K. M. Wilding, A. K. Smith, J. W. Wilkerson, D. B. Bush, T. A. Knotts und B. C. Bundy, „The Locational Impact of Site-Specific PEGylation: Streamlined Screening with Cell-Free Protein Expression and Coarse-Grain Simulation“, *ACS Synthetic Biology*, Jg. 7, Nr. 2, S. 510–521, 2018.
- [349] M. Feig, J. Karanicolas und C. L. Brooks, „MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for applications in structural biology“, *Journal of Molecular Graphics and Modelling*, Jg. 22, Nr. 5, S. 377–395, 2004.
- [350] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York und M. Karplus, „CHARMM: The biomolecular simulation program“, *Journal of Computational Chemistry*, Jg. 30, Nr. 10, S. 1545–1614, 2009.

- [351] C. Kobayashi, J. Jung, Y. Matsunaga, T. Mori, T. Ando, K. Tamura, M. Kamiya und Y. Sugita, „GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms“, *Journal of Computational Chemistry*, Jg. 38, Nr. 25, S. 2193–2206, 2017.
- [352] A. Bakan, L. M. Meireles und I. Bahar, „ProDy: Protein Dynamics Inferred from Theory and Experiments“, *Bioinformatics*, Jg. 27, Nr. 11, S. 1575–1577, 2011.
- [353] M. D. Ward, S. Nangia und E. R. May, „Evaluation of the hybrid resolution PACE model for the study of folding, insertion, and pore formation of membrane associated peptides“, *Journal of Computational Chemistry*, Jg. 38, Nr. 16, S. 1462–1471, 2017.
- [354] V. Ovchinnikov, S. Conti, E. Y. Lau, F. C. Lightstone und M. Karplus, „Microsecond Molecular Dynamics Simulations of Proteins Using a Quasi-Equilibrium Solvation Shell Model“, *Journal of Chemical Theory and Computation*, Jg. 16, Nr. 3, S. 1866–1881, 2020.
- [355] T. A. Beu, A.-E. Ailenei und R.-I. Costinaş, „Martini Force Field for Protonated Polyethyleneimine“, *Journal of Computational Chemistry*, Jg. 41, Nr. 4, S. 349–361, 2020.
- [356] T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink und D. P. Tieleman, „Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models“, *Journal of Chemical Theory and Computation*, Jg. 10, Nr. 2, S. 676–690, 2014.
- [357] Horvath, Marcou und Varnek, „Generative Topographic Mapping of the Docking Conformational Space“, *Molecules*, Jg. 24, Nr. 12, S. 2269, 2019.
- [358] D. Horvath, I. Baskin, G. Marcou und A. Varnek, „Generative Topographic Mapping of Conformational Space“, *Molecular Informatics*, Jg. 36, Nr. 10, S. 1700036, 2017.
- [359] F. Noé, S. Olsson, J. Köhler und H. Wu, „Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning“, *Science*, Jg. 365, Nr. 6457, eaaw1147, 2019.
- [360] A. Tucs, K. Tsuda und A. Sljoka, „Probing conformational dynamics of antibodies with geometric simulations“, *arXiv Preprint (arXiv:2109.14724v1)*, S. 13, 2021.
- [361] D. L. Allen und G. J. Pielak, „Baseline length and automated fitting of denaturation data“, *Protein Science*, Jg. 7, Nr. 5, S. 1262–1263, 1998.
- [362] D. W. Marquardt, „An Algorithm for Least-Squares Estimation of Nonlinear Parameters“, *Journal of the Society for Industrial and Applied Mathematics*, Jg. 11, Nr. 2, S. 431–441, 1963.
- [363] D. R. Roe und T. E. Cheatham, „PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data“, *Journal of Chemical Theory and Computation*, Jg. 9, Nr. 7, S. 3084–3095, 2013.

- [364] B. Knapp, S. Frantal, M. Cibena, W. Schreiner und P. Bauer, „Is an Intuitive Convergence Definition of Molecular Dynamics Simulations Solely Based on the Root Mean Square Deviation Possible?“, *Journal of Computational Biology*, Jg. 18, Nr. 8, S. 997–1005, 2011.
- [365] B. Knapp, N. Lederer, U. Omasits und W. Schreiner, „vmdICE: A plug-in for rapid evaluation of molecular dynamics simulations using VMD“, *Journal of Computational Chemistry*, Jg. 31, Nr. 16, S. 2868–2873, 2010.
- [366] R. Weiser, P. S. Shenkin und W. C. Still, „Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO)“, *Journal of Computational Chemistry*, Jg. 20, Nr. 2, S. 217–230, 1999.
- [367] M. L. Connolly, „Analytical Molecular Surface Calculation“, *Journal of Applied Crystallography*, Jg. 16, Nr. 5, S. 548–558, 1983.
- [368] R. Stepto, T. Chang, P. Kratochvíl, M. Hess, K. Horie, T. Sato und J. Vohlídal, „Definitions of Terms Relating to Individual Macromolecules, Macromolecular Assemblies, Polymer Solutions, and Amorphous Bulk Polymers (IUPAC Recommendations 2014)“, *Pure and Applied Chemistry*, Jg. 87, Nr. 1, S. 71–120, 2015.
- [369] Z. Xinyi, *Synchrotron Radiation Applications*. World Scientific, 2018.
- [370] G. Ramachandran, C. Ramakrishnan und V. Sasisekharan, „Stereochemistry of polypeptide chain configurations“, *Journal of Molecular Biology*, Jg. 7, Nr. 1, S. 95–99, 1963.
- [371] J. J. Prompers und R. Brüschweiler, „General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation“, *Journal of the American Chemical Society*, Jg. 124, Nr. 16, S. 4522–4534, 2002.
- [372] C. A. MacRaild, A. H. Daranas, A. Bronowska und S. W. Homans, „Global Changes in Local Protein Dynamics Reduce the Entropic Cost of Carbohydrate Binding in the Arabinose-binding Protein“, *Journal of Molecular Biology*, Jg. 368, Nr. 3, S. 822–832, 2007.
- [373] G. Lipari und A. Szabo, „Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity“, *Journal of the American Chemical Society*, Jg. 104, Nr. 17, S. 4546–4559, 1981.
- [374] M. Clarkson. „The Model-Free Dynamics Formalism of Lipari and Szabo“. (2007), Adresse: <http://conflux.mwclarkson.com/2007/11/the-model-free-dynamics-formalism-of-lipari-and-szabo/> (besucht am 12. 12. 2020).
- [375] C. Kersten, „Selectivity Determining Features in N-Myristoyltransferases – A Model System for Drug Targets with Conserved Binding Sites“, Dissertation, Johannes Gutenberg-Universität Mainz, 2017.
- [376] D.-W. Li und R. Brüschweiler, „A Dictionary for Protein Side-Chain Entropies from NMR Order Parameters“, *Journal of the American Chemical Society*, Jg. 131, Nr. 21, S. 7226–7227, 2009.

- [377] R. B. Best, G. Hummer und W. A. Eaton, „Native contacts determine protein folding mechanisms in atomistic simulations“, *Proceedings of the National Academy of Sciences*, Jg. 110, Nr. 44, S. 17 874–17 879, 2013.
- [378] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf und O. Beckstein, „MDAnalysis: A toolkit for the analysis of molecular dynamics simulations“, *Journal of Computational Chemistry*, Jg. 32, Nr. 10, S. 2319–2327, 2011.
- [379] R. Gowers, M. Linke, J. Barnoud, T. Reddy, M. Melo, S. Seyler, J. Domański, D. Dotson, S. Buchoux, I. Kenney und O. Beckstein, „MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations“, Python in Science Conference, Austin, Texas, 2016, S. 98–105.
- [380] R. J. Gowers und P. Carbone, „A multiscale approach to model hydrogen bonding: The case of polyamide“, *The Journal of Chemical Physics*, Jg. 142, Nr. 22, S. 224 907, 2015.
- [381] S. Genheden, M. Akke und U. Ryde, „Conformational Entropies and Order Parameters: Convergence, Reproducibility, and Transferability“, *Journal of Chemical Theory and Computation*, Jg. 10, Nr. 1, S. 432–438, 2014.
- [382] M. Kuhn. „Available Models“. (2021), Adresse: <https://topepo.github.io/caret/available-models.html> (besucht am 27.08.2021).

Abbildungsverzeichnis

1.1	<i>Grafting</i> -Methoden	6
1.2	Cyclooctin-Konjugationen	9
1.3	Polymere	12
1.4	Faltungsprozess	15
1.5	Price-Winkel	18
1.6	PEG-Lysin-Wechselwirkung	19
1.7	Kraftfeld-Terme	22
1.8	Simulationsansätze	26
1.9	Konfusionsmatrix	36
1.10	Pin 1-Struktur	45
1.11	Her2-Affibody-Struktur	46
1.12	Src SH3-Struktur	47
1.13	GCN4/coilV _a L _d -Struktur	47
1.14	IFN- α 2a-Struktur	48
3.1	Pin 1-Konjugate	53
3.2	Zusätzliche Pin 1-Konjugate	54
3.3	PyRED-Methode	55
3.4	antechamber-Methode	56
3.5	Monomerbasierte Parametrisierung	57
3.6	Interaktionsprofil-Vergleich 1	59
3.7	Interaktionsprofil-Vergleich 2	60
3.8	Interaktionsprofil-Vergleich 3	61
3.9	RDF-Vergleich	62
3.10	SCC-Kriterien	65
3.11	QSPR-Modellgenerierung Ansatz 1	67
3.12	Datensatzverteilung Ansatz 1	68
3.13	k-Means-Clustering	73
3.14	Clustering-Beispiel	74
3.15	QSPR-Modellgenerierung Ansatz 2	75
3.16	Kollektivvariablen	77
3.17	Positionierung von K31 und K134	79
3.18	Cyclooctin-Linker-Strukturen	80
3.19	GaMDs - Modellierung	81
3.20	CG-Startstruktur	83

3.21 Experimentelle Arbeiten zu Interferon- α 2a	84
4.1 QSPR-Ansatz 1 - Modell C1	88
4.2 QSPR-Ansatz 1 - Modell D1	89
4.3 QSPR-Ansatz 1 - Modell E	90
4.4 QSPR-Ansatz 1 - Modell F	91
4.5 QSPR-Ansatz 1 - Modell G	92
4.6 Ansatz 1 - Univariate Regressionen	93
4.7 MD-Analyse: Dynamik von 16-1 und 27-1	95
4.8 MD-Analyse: Vergleich zwischen 16-1 und 27-1	96
4.9 MD-Analyse: Vergleich zwischen 29-1 und 29-9	97
4.10 MD-Analyse: Vergleich zwischen 19-1 und 19-8	98
4.11 MD-Analyse: Vergleich zwischen 19-1 und 19-4	98
4.12 MD-Analyse: Vergleich zwischen Mutanten von 23-Z	99
4.13 Ansatz 2 - Modell A2-R6	102
4.14 Ansatz 2 - Modell A2-R6v2	108
4.15 Ansatz 2 - ROC-Kurve Modell A2-C1	109
4.16 Parametervergleich	113
4.17 Ansatz 2 - Modell A2v2-R1	114
4.18 Ansatz 2 - ROC-Kurve Modell A2v2-C1	115
4.19 Energiediagramme	118
4.20 Energiedifferenzen	119
4.21 NEB-Pfade	121
4.22 Entfaltungsdynamik	122
4.23 Momentaufnahmen Pin 1 WW - 14	124
4.24 Eigenschaften Pin 1 WW - 14	125
4.25 Momentaufnahmen Pin 1 WW - 16	126
4.26 Eigenschaften Pin 1 WW - 16	127
4.27 Momentaufnahmen Pin 1 WW - 17	128
4.28 Eigenschaften Pin 1 WW - 17	129
4.29 Momentaufnahmen Pin 1 WW - 18	130
4.30 Eigenschaften Pin 1 WW - 18	131
4.31 Momentaufnahmen Pin 1 WW - 19	132
4.32 Eigenschaften Pin 1 WW - 19	133
4.33 Momentaufnahmen Pin 1 WW - 21	134
4.34 Eigenschaften Pin 1 WW - 21	135
4.35 Momentaufnahmen Pin 1 WW - 23	136
4.36 Eigenschaften Pin 1 WW - 23	137
4.37 Momentaufnahmen Pin 1 WW - 26	138
4.38 Eigenschaften Pin 1 WW - 26	139
4.39 Momentaufnahmen Pin 1 WW - 27	140
4.40 Eigenschaften Pin 1 WW - 27	141

4.41	Momentaufnahmen Pin 1 WW - 28	142
4.42	Eigenschaften Pin 1 WW - 28	143
4.43	Momentaufnahmen Pin 1 WW - 29	144
4.44	Eigenschaften Pin 1 WW - 29	145
4.45	Momentaufnahmen Pin 1 WW - 30	146
4.46	Eigenschaften Pin 1 WW - 30	147
4.47	Momentaufnahmen Pin 1 WW - 32	148
4.48	Eigenschaften Pin 1 WW - 32	149
4.49	Momentaufnahmen Her2 - 7	150
4.50	Eigenschaften Her2 - 7	151
4.51	Momentaufnahmen Her2 - 8	152
4.52	Eigenschaften Her2 - 8	153
4.53	Momentaufnahmen Her2 - 15	154
4.54	Eigenschaften Her2 - 15	155
4.55	Momentaufnahmen Her2 - 21	156
4.56	Eigenschaften Her2 - 21	157
4.57	Momentaufnahmen Her2 - 23	158
4.58	Eigenschaften Her2 - 23	159
4.59	Momentaufnahmen Her2 - 25	160
4.60	Eigenschaften Her2 - 25	161
4.61	Momentaufnahmen Her2 - 29	162
4.62	Eigenschaften Her2 - 29	163
4.63	Momentaufnahmen Her2 - 33	164
4.64	Eigenschaften Her2 - 33	165
4.65	Momentaufnahmen Her2 - 39	166
4.66	Eigenschaften Her2 - 39	167
4.67	Momentaufnahmen Her2 - 43	168
4.68	Eigenschaften Her2 - 43	169
4.69	Momentaufnahmen Her2 - 46	170
4.70	Eigenschaften Her2 - 46	171
4.71	Momentaufnahmen Her2 - 47	172
4.72	Eigenschaften Her2 - 47	173
4.73	Momentaufnahmen Her2 - 49	174
4.74	Eigenschaften Her2 - 49	175
4.75	Momentaufnahmen Her2 - 50	176
4.76	Eigenschaften Her2 - 50	177
4.77	Momentaufnahmen Her2 - 53	178
4.78	Eigenschaften Her2 - 53	179
4.79	Zusammenfassender Vergleich der Pin 1-Konjugate	181
4.80	Pin 1 - Übersicht Konjugationsstellen	182
4.81	Interaktionsenergien - Pin 1 WW-Domäne	183
4.82	Pin 1 - PEG-Interaktionen	184

4.83	Pin 1 - LPG-Interaktionen	185
4.84	Pin 1 - LPG-Interaktionen	185
4.85	Pin 1 - POx-Interaktionen	186
4.86	Pin 1 - POx-Vergleich	188
4.87	Zusammenfassender Vergleich der Her2-Konjugate	189
4.88	Her2 - Übersicht Konjugationsstellen	190
4.89	Interaktionsenergien - Her2-Affibody	191
4.90	Her2 - PEG-Interaktionen	192
4.91	Her2 - LPG-Interaktionen	194
4.92	Her2 - POx-Interaktionen	194
5.1	Polymerdichten	200
5.2	SASA IFN- α 2a	201
5.3	Kontaktprofil 10-kDa-BCN-PEG	203
5.4	PEG-Interaktionen mit K/R	204
5.5	Kontaktprofil 10-kDa-BCN-LPG	205
5.6	LPG-Interaktionen mit D/A	205
5.7	Kontaktprofil 10-kDa-BCN-PEtOx	207
5.8	PEtOx-Interaktionen mit S/Y	207
5.9	Kontaktprofil 10-kDa-DBCO-PEG	208
5.10	Kontaktprofil 10-kDa-DBCO-LPG	209
5.11	Kontaktprofil 10-kDa-DBCO-PEtOx	210
5.12	Vergleich BCN und Peptidbindung	211
5.13	Kontaktprofil 20-kDa-BCN-PEG	213
5.14	Kontaktprofil 20-kDa-BCN-LPG	213
5.15	Kontaktprofil 20-kDa-BCN-PEtOx	214
5.16	Kontaktprofil 40-kDa-BCN-PEG	214
5.17	Kontaktprofil 50-kDa-BCN-LPG	215
5.18	Kontaktprofil 30-kDa-BCN-PEtOx	215
5.19	Verteilung von PEG in der CG-Simulation	217
5.20	Dichtevergleich CG MD - GaMD	217
5.21	Kontaktprofil CG-MD PEG-10-kDa	218
5.22	Shroud / Dumbbell-Modell	219
5.23	Schmelzpunkte - 10 kDa	220
5.24	<i>In silico</i> -Interpretation der Schmelztemperaturen - 10 kDa	221
5.25	CNA-Netzwerke 10 kDa	224
5.26	CNA-Interaktionen	225
5.27	Perkolationsindizes IFN- α 2a	225
5.28	Native Kontakte LPG-Konjugate	226
5.29	LiP-Assay	228
5.30	Schmelzpunkte - 20 - 50 kDa	229
5.31	<i>In silico</i> -Interpretation der Schmelztemperaturen - 10 - 50 kDa	230

5.32	Bioaktivitäten	231
5.33	Korrelation Bioaktivität	232
5.34	Korrelation Bioaktivität 2	233
6.1	Übersicht der Arbeit	239
B.1	Ramachandran-Winkel	258
D.1	RDF-Differenz-Plots	282
D.2	Umorientierung der ersten Schleife	283
D.3	Korrelationen - Modell A2-R6	286
D.4	Korrelationen - Modell A2-C1	287
D.5	Verhältnis Modell A2-R6 - Polymeralternativen / PEG	288
D.6	Verhältnis Modell A2-C1 - Polymeralternativen / PEG	289

Tabellenverzeichnis

1.1	Vermarktete Biokonjugate	3
1.2	Vor- / Nachteile der Biokonjugation	5
1.3	Konjugationsbeispiele	8
1.4	Stabilisierungsmechanismen	20
3.1	PyRED-Parameter	55
3.2	antechamber-Parameter	56
3.3	Datensätze	72
3.4	Simulierte Konjugate von IFN- α 2a	79
4.1	Statistische Parameter für Modelle C1 und D1	88
4.2	Statistische Parameter für Modelle E und F	90
4.3	Statistische Parameter für Modelle G und H	92
4.4	Statistische Parameter für Modell A2-R6	102
4.5	Konjugate - Modell A2-R6	103
4.6	Strukturbasierte Richtlinie	105
4.7	Statistische Parameter für Modell A2-R6v2	107
4.8	Statistische Parameter für Modell A2-C1	109
4.9	Konjugate - Modell A2-C1	110
4.10	Statistische Parameter für Modell A2v2-R1	114
4.11	Statistische Parameter für Modell A2v2-C1	115
4.12	Zusätzliche Konjugate - Vorhersagen	116
4.13	Energiedifferenzen	120
4.14	Energiebarrieren	121
B.1	CNA-basierte Deskriptoren	261
C.1	Chemische Strukturen	263
C.2	Simulierte Modellprotein-Konjugate	265
D.1	Berechnete Deskriptoren	273
D.2	QSPR-Ansatz 1 - Modelltypen	278
D.3	QSPR-Ansatz 1 - Statistische Parameter	280
D.4	QSPR-Ansatz 1 - ADAN-Analyse	281
D.5	QSPR-Ansatz 2 - Modelltypen	284
D.6	QSPR-Ansatz 2 - Statistische Parameter der Regressionsmodelle . . .	284

D.7 QSPR-Ansatz 2 - Statistische Parameter der Klassifikationsmodelle . .	285
D.8 Polymeralternativen - Vorhersagen Modell A2-R6	288
D.9 Polymeralternativen - Vorhersagen Modell A2-C1	289

Abkürzungsverzeichnis

AA	Acrylsäure
ABC	Accelerated blood clearance
Acc	Accuracy
AD	Applicability domain
ADAN	Applicability domain analysis
AGA	N-Acryloyl-D-glucosamin
Am	Acrylamid
aMD	Accelerated MD
ATRP	Atomtransferradikalpolymerisation
AUPRC	Area under the precision-recall curve
AUROC	Area under the receiver operating characteristic curve
AZK	N _ε -[(2-Azidoethoxy)carbonyl]-L-lysin
BAcc	Balanced accuracy
BBCOM	Bootstrapped block covariance overlap method
BCN	Bicyclononin
BCOM	Block covariance overlap method
CCC	Konkordanzkorrelationskoeffizient
CCE	Constancy of cluster entropy
CCE1/2	Clusterkonfigurationsentropie Typ 1/2
CCR	Correct classification rate
CD	Circulardichroismus
CG	Coarse-grained
cMD	Conventional MD
CNA	Constrained Network-Analyse
CoMFA	Comparative molecular field analysis
CPO	Linear combination of pairwise overlaps
CROP	Kationische Ringöffnungspolymerisation
CRP	Kontrolliert radikalische Polymerisation
CTA	Kettentransferreagenz
CV	Collective variables
DAA	Direct acting agents
DBCO	Dibenzocyclooctin
DLS	Dynamische Lichtstreuung
DMAEMA	Dimethylaminoethoxymethacrylat
DMAm	Dimethylacrylamid
DSF	Differential Scanning Fluorimetry
DU	Dummy atom
EPR	Enhanced permeability and retention
FDA	Food and Drug Administration
FIRST	Floppy inclusions and rigid substructure topography
FN	False negatives
FP	False positives

FPR	False positive rate
G-CSF	Granulozyten-Kolonie-stimulierender Faktor
GaMD	Gaussian accelerated MD
HB	Wasserstoffbrückenbindung
HEK	Human embryonic kidney
Her2	Human epidermal growth factor receptor 2
HES	Hydroxyethylstärke
HS	Holdout-Set
HSQC	Heteronuclear Single Quantum Coherence
IFN- α 2a	Interferon- α 2a
IFNAR	Interferon- α -Rezeptor
iRED	Isotropic reorientational eigenmode dynamics
IUPAP	Internationale Union für reine und angewandte Physik
LiP	Limitierte Proteolyse
LPG	Lineares Polyglycerol
MAE	Mean absolute error
MALDI-TOF	Matrix-assistierte Laser-Desorption-Ionisierung-Time-of-flight
MCMC	Markow-Chain-Monte-Carlo
MD	Molekulardynamik
MetaMD	Metadynamik-Simulation
MOE	Molecular Operating Environment
mPEG	Monomethoxy-PEG
MRW	Mean residue weight
MSD	Mittlere quadratische Verschiebung
MW	Multiple walkers
MW-wtMetaMDs	Multiple walkers well-tempered metadynamics simulation
NA	Not available
NAEK	N $_{\epsilon}$ -[(2-Azidoethoxy)carbonyl]-L-lysin
NHS	N-Hydroxysuccinimid
NMR	Nuklearmagnetresonanz
NPT	Isothermal-isobarisches Ensemble
NPV	Negative predictive value
NVT	Mikrokanonisches Ensemble
OECD	Organisation for Economic Co-operation and Development
OEOA	Oligo(ethylenoxid)methyletheracrylat
PAM	Poly(ammoniummethacrylat)
PBC	Periodic boundary conditions
PBK	Polymer-Biokonjugat
PCA	Principal component analysis
PCBA _m	Poly(carboxybetainacrylamid)
PCMA	Phosphorylcholinmethacrylat
PDB	Proteindatenbank
PDI	Polydispersität
PEG	Polyethylenglykol
PEtOx	Poly(2-ethyl-2-oxazolin)
PG	Polyglycerol
PLS	Partial least squares
PMeOx	Poly(2-methyl-2-oxazolin)
PNIPAA _m	Poly(N-isopropylacrylamid)
POEGMA	Poly(oligo(ethylenglykol)methacrylat)
POx	Poly(2-alkyl-2-oxazolin)

PPV	Positive predictive value
PRESS	Predicted residual sum of squares
PSMA	Poly(sulfonatmethacrylat)
PVP	Polyvinylpyrrolidon
QM	Quantenmechanik
QSAR	Quantitative structure-activity relationships
QSPR	Quantitative structure-property relationships
RAFT	Reversible Additions-Fragmentierungs-Kettenübertragungs-Polymerisation
rdCV	Repeated double cross validation
RDF	Radial distribution function
RESP	Restrained electrostatic potential
RFE	Rekursive Feature-Eliminierung
RMSD	Root-mean-square deviation
RMSE	Root-mean-square error
RMSF	Root-mean-square fluctuation
ROC	Grenzwertoptimierungskurve
RSS	Residual sum of squares
SASA	Solvenszugängliche Oberfläche
SCC	Self-consistency check
SCID	Schwerer kombinierter Immundefekt
SDVG	Smith, Daura und van Gunsteren
SE	Standard error
SEAP	Secreting alkaline phosphatase assay
SPAAC	Strain-promoted azide-alkyne click chemistry reaction
SSE	Sum of squared errors
ST-HPMA	Semitelechelisches Poly[N-(2-hydroxypropyl)-methacrylamid
TEMPO	2,2,6,6-Tetramethylpiperidinyloxyl
TN	True negatives
TNF	Tumor-Nekrose-Faktor
TNR	True negative rate
TP	True positives
TPMA	Tris(2-pyridylmethyl)amin
TPR	True positive rate
TS	Trainingsdatensatz
TSA	Thermal Shift Assay
TSS	Total sum of squares
VdW	Van-der-Waals
VEGF	Vaskulärer endothelialer Wachstumsfaktor
wt	Well-tempered
WT	Wildtyp