

**Computational Structure-based Design Approaches:
Targeting HIV-1 Integrase and the
Macrophage Infectivity Potentiator
of *Legionella pneumophila***

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

Vorgelegt von
Martin Sippel
aus Northeim

Würzburg 2010

Eingereicht am

.....

bei der Fakultät für Chemie und Pharmazie

1. Gutachter

.....

2. Gutachter

.....

der Dissertation

1. Prüfer

.....

2. Prüfer

.....

3. Prüfer

.....

des öffentlichen Promotionskolloquiums

Datum des öffentlichen Promotionskolloquiums

.....

Doktorurkunde ausgehändigt am

.....

Danksagung

Die vorliegende Arbeit entstand im Zeitraum von November 2006 bis Januar 2010 am Institut für Pharmazie und Lebensmittelchemie der Julius-Maximilians-Universität Würzburg auf Anregung und unter Anleitung von

Herrn Prof. Dr. Christoph Sotriffer

Ihm gilt mein besonderer Dank für seine immerwährende Hilfsbereitschaft bei wissenschaftlichen Fragen sowie sein offenes Ohr für eigene Ideen zu den Projekten. Danken möchte ich ihm auch insbesondere für sein großes Engagement zur Ermöglichung meiner Aufenthalte in Italien und den USA.

Desweiteren möchte ich mich bedanken bei:

Herrn Prof. Dr. Gerhard Klebe für die Betreuung während meiner Anfangszeit an der Universität Marburg.

Dr. Mario Sechi für die herzliche Aufnahme und die wissenschaftlichen Diskussionen an der Università degli Studi di Sassari, Italien.

Dr. Nouri Neamati von der University of Southern California School of Pharmacy in Los Angeles für die Möglichkeit, in seinem Labor arbeiten und sämtliche Geräte und Materialien nutzen zu können.

Erik Serrao für seine Hilfe bei der Durchführung der Integrase-Assays.

Dem *Deutschen Akademischen Austausch-Dienst (DAAD)* für die Vergabe des Doktoranden-Stipendiums zur Finanzierung meines USA-Aufenthaltes.

Monika Nocker für die vielen guten Gespräche und den Gratis-Italienischkurs.

Meinen Kollegen *Benjamin Schaefer, Armin Welker, Ulrich Peinz, David Zilian, Christine Topf, Daniel Cappel* und *Constanze Waltenberger* für die gute Arbeitsatmosphäre.

Allen Assistenten des 1. Semesters für die gute Zusammenarbeit während der Praktika.

Christina Juli für die gute Zusammenarbeit auf dem MIP-Projekt.

Alexandra Thiele vom Max-Planck-Institut für Enzymologie in Halle für die Durchführung der MIP-Assays, *Prof. Dr. Michael Steinert* von der Universität Braunschweig für die Durchführung der Transwell-Assays und *Dr. Christian Schweimer* von der Universität Bayreuth für die Durchführung von NMR-Studien zu MIP.

Dr. Patrick Pfeffer und *Dr. Alexander Hillebrecht* für eine großartige Zeit in Marburg und darüber hinaus.

Meinen Eltern für ihre immerwährende Unterstützung.

Table of contents

1. Introduction	1
2. HIV-1 Integrase.....	3
2.1 AIDS and HIV.....	3
2.2 Integrase function and structure.....	4
2.2.1 Integrase function.....	5
2.2.2 Integrase structure	6
2.3 Integrase inhibitors.....	14
2.3.1 The diketo acid class of inhibitors.....	14
2.3.2 Other integrase inhibitor classes.....	19
2.3.3 Allosteric inhibitor binding sites.....	21
2.4 Diketo acids – Mechanism of action	24
2.4.1 Experimental basis.....	25
2.4.2 Comparative docking of free ligands and complexes.....	29
2.5 Dimerization interface binders – Design of a novel inhibitor class	41
2.5.1 Protein-protein interfaces: An emerging class of therapeutic targets	41
2.5.2 The dimerization interface: Molecular dynamics and docking studies	43
2.5.3 Dimerization inhibitor design I	61
2.5.4 Biological testing of dimerization inhibitor candidates.....	71
2.5.5 Dimerization inhibitor design II.....	73
2.5.6 The peculiarity of the dimerization interface.....	85
2.5.7 Satisfaction of hydrogen bonding groups in protein-ligand complexes	86
2.6 Summary and conclusion.....	90
3. <i>Legionella pneumophila</i> Macrophage Infectivity Potentiator.....	93
3.1 Legionnaires' Disease and MIP	93
3.2 MIP and FKBP12: Peptidyl prolyl cis/trans isomerases.....	95
3.2.1 MIP structures	97
3.2.2 FKBP12 structures	99

3.2.3 Comparison of MIP and FKBP12.....	101
3.2.4 Analysis of the MIP structural ensemble 2VCD	104
3.3 FKBP12 ligands as MIP inhibitors	106
3.4 Biological testing of compounds.....	112
3.5 DMSO as new substructure?.....	114
3.5.1 DMSO derivatives – cyclic sulfoxides.....	116
3.6 Extension of the anchor: Fragment-based design	118
3.6.1 Fragment-based drug design.....	118
3.6.2 MIP active site subpocket.....	119
3.7 Conclusion	125
4. Summary / Zusammenfassung	127
4.1 Summary	127
4.2 Zusammenfassung	130
5. Appendix.....	135
5.1 Methodologies.....	135
5.2 Abbreviations	142
5.3 Publications	145
6. Bibliography.....	147

1. Introduction

Molecular modelling in the realm of drug development (also known as computer-aided drug design) has become more and more popular in recent years. Several successful applications to compounds which have entered advanced clinical trials or even the market (e.g. captopril, saquinavir) underline the significance of molecular modelling techniques in the drug development process¹. These techniques support several stages of the drug development: They can assist in revealing the structural characteristics of proteins / protein-ligand complexes (structure elucidation), discovering a novel ligand structure (lead discovery), or optimizing a known inhibitor (lead optimization).

The scope of the present thesis is to employ molecular modelling techniques for the design of inhibitors directed against two infectious diseases: AIDS and Legionellosis. AIDS and HIV infection have become a pandemic; especially on the African continent, AIDS causes great suffering and has a drastic effect on economical development. Although an established treatment against AIDS is available, there is a clear demand for novel drugs, keeping the pipeline filled to have potent agents against HIV strains resistant to current drugs. Legionellosis is an infection transmitted by the ubiquitous aquatic bacterium *Legionella pneumophila*. Infections and major outbreaks are reported from all continents, but especially in developing countries, legionellosis is expected to be vastly underdiagnosed. A *Legionella*-tailored treatment still needs to be established.

The thesis is divided into two parts: In the first part, HIV and the target enzyme, HIV integrase, are introduced. This introduction is followed by an overview of integrase structural data and known inhibitors. One prominent integrase inhibitor class, the diketo acids, is inspected in detail: Experimental findings about a unique mode of action are further supported on a structural level by means of molecular modelling. The work enables a better understanding of how this important inhibitor class interacts with the integrase. In the next section, a novel approach towards integrase inhibition is introduced, which aims at disrupting protein-protein interactions crucial for enzymatic activity. A strategy to overcome the lack of structural data is outlined and valuable information about the integrase protein-protein interface structure is given. This information is subsequently used to generate hypotheses about ligand binding requirements, finally leading to the generation of pharmacophore models. These models are employed in virtual screenings for purchasable compounds aiming at disruption of the protein-protein interactions.

Furthermore, guidelines for the design of tailored compounds are given. Targeting protein-protein interactions is a major challenge; in this context, the peculiarity of the integrase protein-protein interface is discussed with respect to other interfaces.

In the second part of the thesis, the macrophage infectivity potentiator (MIP) is introduced as an attractive, yet unexploited target for Legionellosis treatment. Available structural data of *Legionella* MIP are critically assessed and compared to structures of a human homologous protein (FKBP12). Studies to transfer known FKBP12 ligands to MIP are outlined and differences between these ligands with respect to their potential binding to MIP are explained. In this context, the discovery of a MIP inhibitor is described. This inhibitor is currently further optimized to improve its affinity. Furthermore, light is shed on the effect of DMSO usage in MIP assays. Based on these findings, a new sulfoxide scaffold is suggested, which itself is extended in a fragment-based design approach. These studies yield valuable information for the synthesis of novel MIP inhibitor structures. The studies on MIP are a collaboration with Christina Juli (University of Wuerzburg), Alexandra Thiele (Max-Planck-Institute for Enzymology, Halle), Michael Steinert (University of Braunschweig), and Christian Schweimer (University of Bayreuth) within the framework of the SFB630 (Recognition, Preparation and Functional Analysis of Agents against Infectious Diseases).

Parts of the present thesis have been published in the scientific literature. The chapter about the diketo acid class of integrase inhibitors is reproduced in part from: Bacchi, A.; Biemmi, M.; Carcelli, M.; Carta, F.; Carlotta, C.; Fiscaro, E.; Rogolino, D.; Sechi, M.; Sippel, M.; Sottriffer, C.; Sanchez, T. W.; Neamati, N. From Ligand to Complexes. Part 2. Remarks on Human Immunodeficiency Virus Type 1 Integrase inhibition by Diketo Acid Metal Complexes. *J. Med. Chem.* **2008**, *51*, 7253-7264. Copyright 2008 American Chemical Society. Also, the studies on the integrase protein-protein interface have been published; the corresponding chapter is reproduced in part from: Sippel, M.; Sottriffer, C. Molecular Dynamics Simulations of the HIV-1 Integrase Dimerization Interface: Guidelines for the Design of a Novel Class of Integrase Inhibitors. *J. Chem. Inf. Model.* **2010**, *50*, 604-614. Copyright 2010 American Chemical Society. A manuscript dealing with the MIP inhibitor studies is currently in preparation.

2. HIV-1 Integrase

The following chapter is divided into three parts: In the first part, a general introduction to AIDS, HIV, and integrase is given. In the second part, modelling studies are presented to get a deeper understanding of the mechanism of action of a well-known class of integrase inhibitors: Light is shed on the question whether the diketo acid class of inhibitors act as metal complexes in their active form. In the third part, an approach towards the development of a whole new class of inhibitors, biasing the enzyme's multimerization, is introduced.

2.1 AIDS and HIV

AIDS (acquired immunodeficiency syndrome) is still one of the greatest challenges to mankind. According to the latest report of the WHO, the annual number of people killed by AIDS worldwide is an estimated 2.1 million, including 280,000 children. Globally, there are 33 million people living with HIV (human immunodeficiency virus), and the annual number of new HIV infections is estimated to 2.7 million (all data from the 2008 report)². Especially in sub-Saharan Africa, AIDS is epidemic. HIV prevalence among adults (age 15-49) is estimated up to 28 % in some southern African countries. Besides the individual burden, AIDS morbidity and mortality has a retarding impact on economical development in these countries. AIDS is also an emerging problem in eastern European countries.

HIV was identified as the causative agent of AIDS³. Two distinct HIV strains are known: HIV-1, which is more virulent and globally spread, causing the majority of infections, and HIV-2, which is confined to western Africa⁴. As a retrovirus, HIV encodes its genome in an RNA strand. It infects CD4 T-cells or macrophages, thereby compromising the immune system. In the advanced stage of HIV infection, the number of T-cells declines to a critical level. The breakdown of the immune system leads to opportunistic infections and tumors, eventually leading to death⁵. HIV entry is mediated through interactions of the viral envelope proteins gp120 and gp41 with CD4 and CCR5 coreceptors on the host cell surface. After successful infection of the host cell, the virion releases the enzyme reverse transcriptase (RT), which transcribes the viral RNA into its complementary DNA. This transcription is prone to errors (since reverse transcription does not exploit repairing mechanisms) and mutations can occur frequently, which may lead to drug resistance. The viral DNA is transported into the cell nucleus and integrated into the host DNA

by a second HIV enzyme, the integrase (IN). In the next step, the cell's devices for protein biosynthesis are exploited for the production of a viral polyprotein, which is subsequently cleaved into functional proteins by the third HIV enzyme, the protease (PR). The last step comprises the assembly of the separate proteins and subsequent budding, which eventually releases the new virion⁶.

Currently, there is no vaccine or cure for AIDS. Approved antiretroviral therapies address five steps of the HIV lifecycle: The fusion, the viral entry, the reverse transcription, the proteolytic maturation, and the genome integration⁷. Fusion inhibitors like the peptide enfuvirtide (brand name Fuzeon) block gp41-host cell membrane interactions. Entry inhibitors like maraviroc (brand name Celsentri), which is a non-peptidic compound, block gp120-CCR5 interactions, thereby preventing the virion from entering the host cell. Both drugs were introduced to the market quite recently, namely in 2003 and 2007, respectively⁸. For RT inhibition, several nucleoside (e.g. zidovudine) and non-nucleoside (e.g. efavirenz) inhibitors have been developed. RT inhibitors are classical antiretroviral therapeutics, the first of which was introduced in 1987 (zidovudine)⁹. The third step of the viral lifecycle addressed by antiretroviral drugs is the IN-mediated integration of the DNA. Despite a long history of IN inhibitor development, the first and so far only marketed drug was approved not before October 2007. This points out that IN still needs to be further exploited as a target in antiretroviral therapy¹⁰. The proteolytic maturation of the virion is blocked by PR inhibitors: many inhibitory compounds have been developed, e.g. saquinavir, which was approved as the first marketed PR inhibitor in 1995¹¹. RT and PR inhibitors are combined in the highly active antiretroviral therapy (HAART)¹².

2.2 Integrase function and structure

Each HIV particle contains 40-100 IN units¹³. HIV IN belongs to the protein superfamily of polynucleotidyl transferases, which cut and transfer polynucleotides like DNA. Members of this superfamily are RNaseH, RuvC resolvase, Mu transposase, and integrases of other retroviruses like ASV (avian sarcoma virus), RSV (rous sarcoma virus), and FIV (feline immunodeficiency virus). Despite their different origins, these enzymes share common structural features, suggesting a similar mechanism of catalysis¹⁴.

2.2.1 Integrase function

The IN-catalyzed insertion of viral DNA (vDNA) into the host genome is a two-step reaction: First, the vDNA is cleaved at the 3'-ends (3'-processing, 3'-P). Then, in a second step, the processed vDNA is transferred to the host genome (strand transfer, ST)¹⁵.

Following RT-catalyzed reverse transcription of the viral genome, the resulting vDNA undergoes 3'-processing within the cytoplasm. At both ends of the vDNA, recurrent base sequences are located, the so-called long terminal repeats (LTR), each comprising the U3 and U5 subunit. These subunits are recognized by IN, which then catalyzes the cleavage of two GT dinucleotides, leaving two unpaired CA dinucleotides with 3'-hydroxyl groups at both ends of the vDNA¹⁶ (Figure 1). The resulting reactive intermediate from the 3'-processing is primed for insertion into the host genome. Before this strand transfer reaction can take place, both vDNA and IN need to enter the cell nucleus to access the host DNA (hDNA). Therefore, the IN-vDNA complex is extended by cellular proteins like LEDGF/p75 (lens epithelium growth factor), INI1 (integrase interactor 1), BAF (barrier-to-autointegration factor), and HMGA1 (high mobility group chromosomal protein A1), yielding the pre-integration complex (PIC)¹⁷. Some of the cellular proteins packaged within the PIC influence IN catalytic activity, for example LEDGF/p75. Eventually, the PIC transmigrates through the nuclear envelope. Once in the nucleus, IN catalyzes the strand transfer: Both reactive 3'-OH ends of the vDNA are inserted into 5'-phosphoester groups of the target hDNA backbone through a transesterification reaction. The insertion of both vDNA ends is a concerted mechanism, with a five base-pair stagger between the junction points. Subsequently, the strand transfer is finished by trimming of the unpaired vDNA 5'-ends, gap filling of the hDNA, and finally ligation of the vDNA 5'-ends with the hDNA. These reactions are likely to be carried out by host cell DNA repair enzymes¹⁸. In contrast to the conserved-sequence requirement for the vDNA LTR recognition, the insertion site within the hDNA is rather arbitrary. Once the viral genome is fully integrated, it is also referred to as provirus. HIV-1 IN is also able to catalyze the reverse reaction of strand transfer *in vitro*, the so-called disintegration, where a short DNA substrate is split into two strands¹⁹.

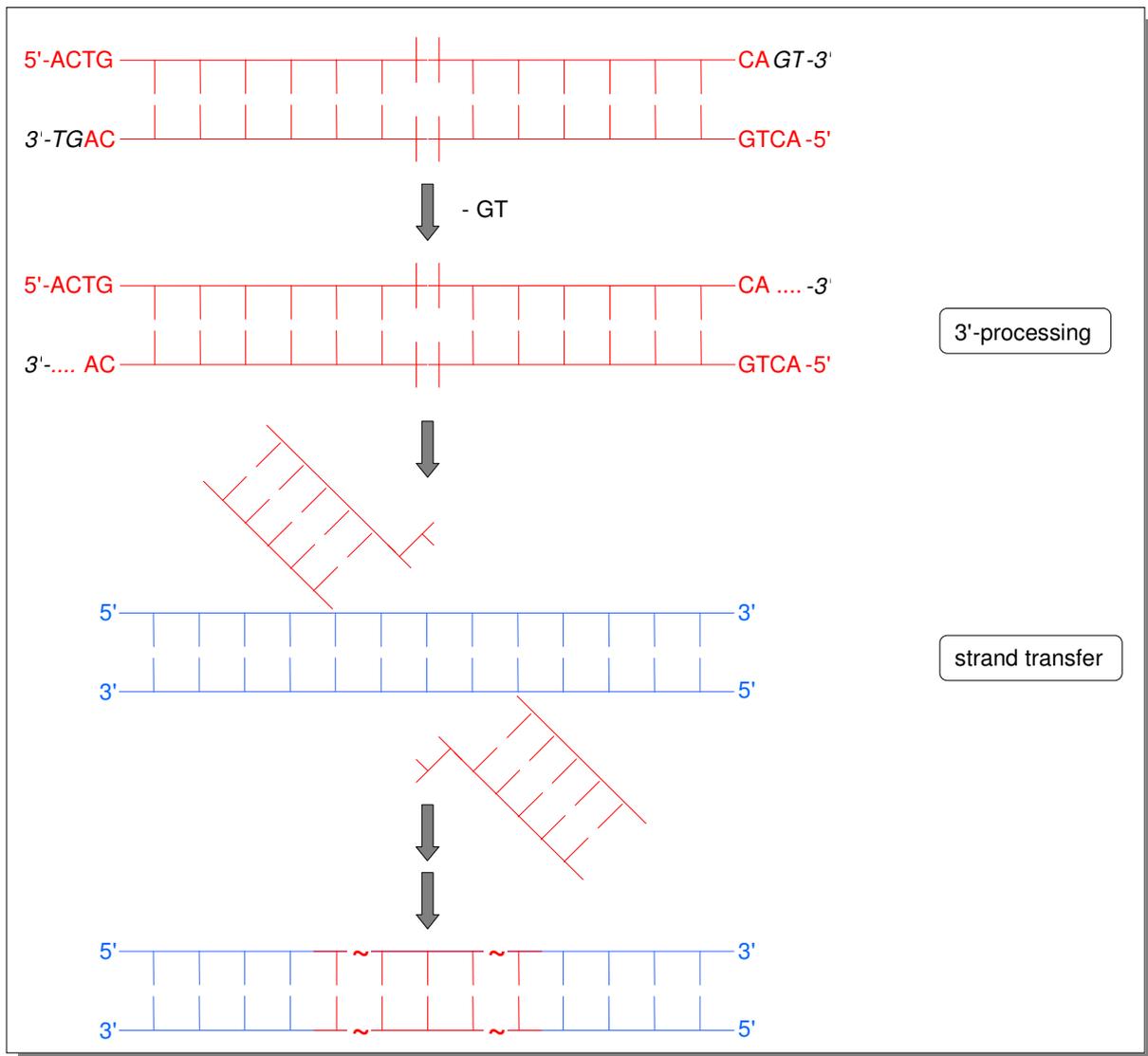


Figure 1: Mechanism of the IN-mediated integration of viral DNA (red) into host DNA (blue): 3'-processing and strand transfer is followed by DNA repair mechanisms of the cell.

2.2.2 Integrase structure

IN is a 32-kDa protein with three structural domains: The N-terminal domain (NTD, residues 1-49), the catalytic core (CCD, residues 50-212), and the C-terminal domain (CTD, residues 213-288)²⁰. For all domains, crystallographic and NMR solution structures are available. These structures are listed in Table 1a/b on page 12 and 13.

The catalytic core domain (CCD)

The CCD harbors the IN active site. A CCD monomer has six α -helices and five β -strands as secondary structure motifs, it is similar to other polynucleotidyl transferases like RNaseH, Mu Transposase, and RuvC Resolvase²¹. Two monomers associate to form a C_2 -symmetry-related dimer. Thereby, a large dimer interface of about 1400 \AA^2 per subunit is buried. The CCD harbors the active site catalytic triad D64, D116, and E152 (see Figure 2). This DDE motif is common among polynucleotidyl transferases. Any substitution of these catalytic residues abolishes IN catalytic activity²². According to the available structural data, the active site itself is rather flat, without well-defined pockets. This is not surprising, given the fact that the substrate is a DNA macromolecule.

Adjacent to the active site, there is a loop comprising residues 140-152. This structural feature is not resolved in many crystal structures (see Table 1a/b), most probably due to its inherent flexibility. Site-directed mutagenesis of residues in this loop, like G140A and G149A, leads to a reduced flexibility, which allows a better resolution of this loop, for example in structure **1B9F**²³.

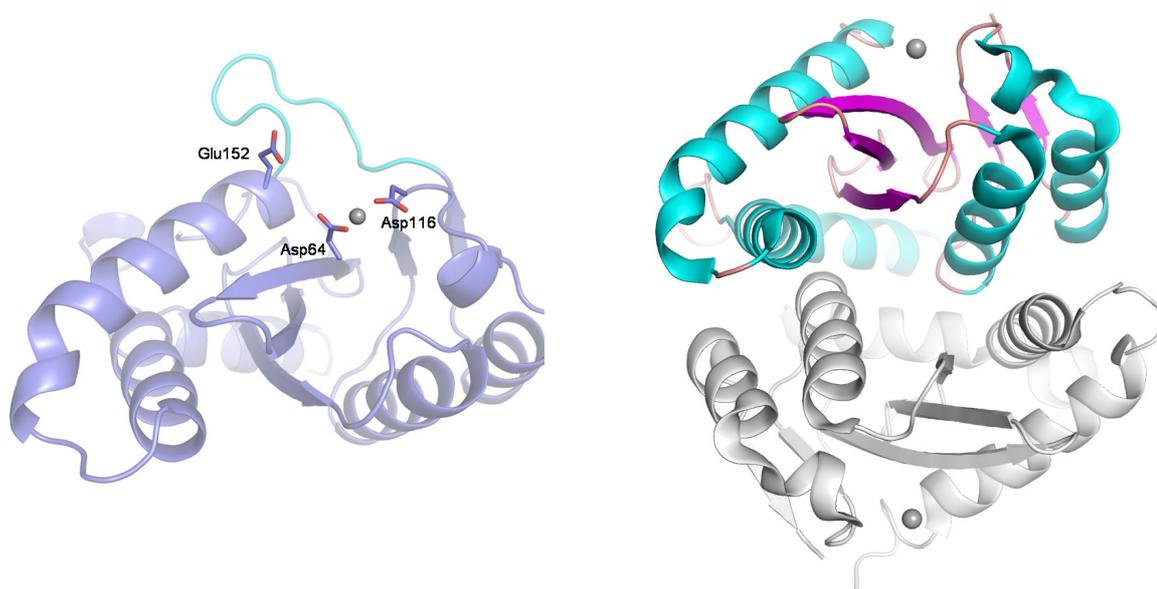


Figure 2: IN CCD. Left: Catalytic triad with adjacent loop (cyan). Right: Secondary structure motifs in the CCD dimer. Grey spheres: Active site metal ions.

However, these mutations go along with a decrease in catalytic activity, pointing out an important, yet not fully understood role of this active site loop in the DNA processing / integration reaction.

Like other polynucleotidyl transferases, IN requires divalent metal cofactors such as Mg^{2+} or Mn^{2+} for its catalytic activity²⁴ (see Figure 3). While Mg^{2+} is the more prevalent cofactor *in vivo*, Mn^{2+} is frequently used *in vitro* and leads to increased catalytic activity. IN crystal structures **1BIU**, **1BL3**, and **1QS4** reveal Mg^{2+} chelation by the catalytic aspartate residues D64 and D116^{25,26,27}. Although never observed in IN crystal structures, it has been suggested that a second metal ion can be complexed by the third catalytic residue E152 and either D64 or D116. This hypothesis is supported by two-metal-ion crystal structures of other retroviral integrases like ASV IN (PDB code 1VSH)^{28,29}. In the case of HIV IN, it is conceivable that a second metal ion is introduced by the DNA substrate, having Mg^{2+} bound to its negatively charged phosphodiester backbone.

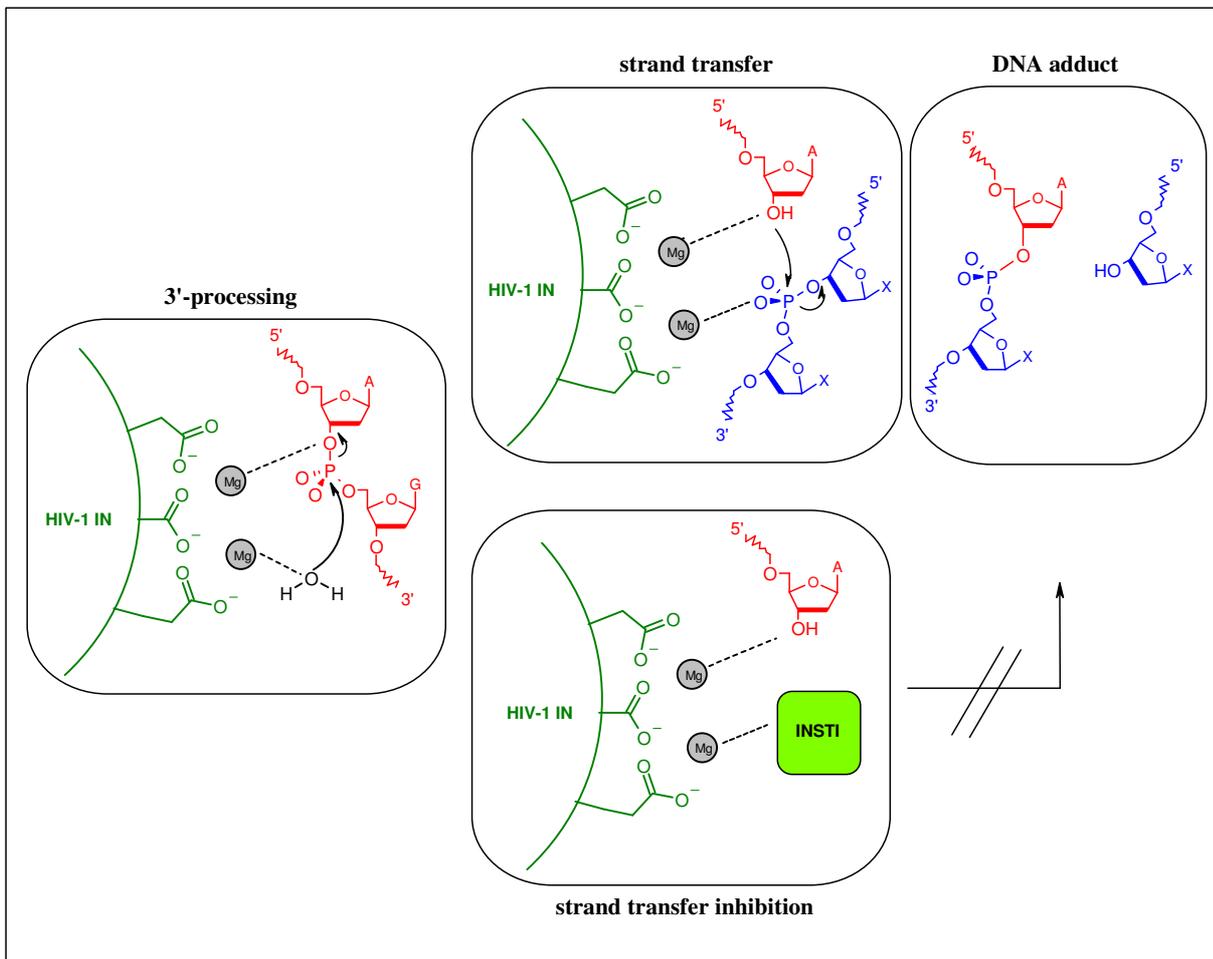


Figure 3: Metal dependency of IN catalysis. Viral DNA (blue) binds to Mg^{2+} , as does the host DNA (red) to the second metal ion. 3'-processing and strand transfer yield a DNA adduct. In presence of viral DNA substrate, integrase strand transfer inhibitors (INSTIs, see Chapter 2.3.1) bind to Mg^{2+} and block the adduct formation.

Due to the poor solubility of the CCD, the first attempts to grow wild type IN CCD crystals for X-ray experiments failed. Eventually, mutagenesis studies resulted in two mutants, F185K and F185H, having sufficient solubility for crystallisation and at the same time no significant effect on IN catalytic activity³⁰. Most of the available IN PDB entries are CCD structures (18 out of 28, see Table 1a/b). Although the IN active site is located on the CCD, both the N-terminal and the C-terminal domains are essential for 3'-processing and strand transfer. Yet, the disintegration reaction can be carried out by the CCD alone.

The N-terminal domain (NTD)

The N-terminal domain of IN comprises residues 1-49 and has a helical secondary structure (see Figure 4). Like the CCD, the NTD tends to form dimers and plays a crucial role in tetramerization of the full-length IN¹³. This tetramerization is enhanced by Zn²⁺, which itself is bound via an HHCC (histidine-histidine-cysteine-cysteine) motif within the NTD. The HHCC motif is highly conserved also among integrases of other retroviruses, like ASV, RSV, and SIV³¹.

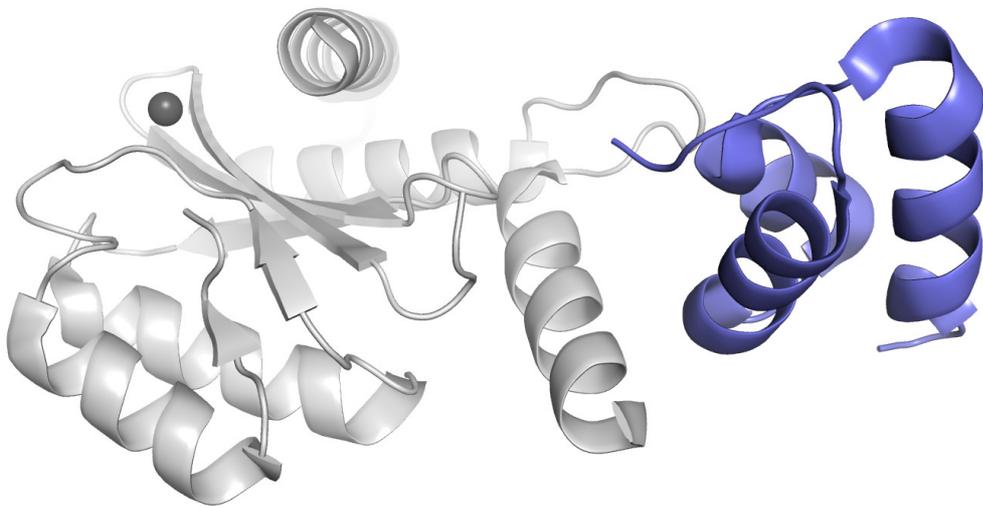


Figure 4: IN NTD (PDB code 1K6Y) with its three helices (blue), in relation to the CCD (grey).

Structural data of the NTD are available from both X-ray and NMR sources: Six PDB entries (**1WJA**, **1WJB**, **1WJC**, **1WJD**, **1WJE**, and **1WJF**)^{32,33} are NMR solution structures of the NTD alone, while **1K6Y** is a crystallographic two-domain structure of the NTD and the CCD (see Table 1a on page 12)³⁴.

The C-terminal domain (CTD)

The C-terminal domain (residues 213-288) has a barrel shape, formed by five antiparallel β -strands. Like the CCD and the NTD, also the CTD is prone to dimerization. It binds the DNA substrate non-specifically, presumably by a strip of positively charged amino acids along the CTD surface. This domain is the least conserved one, mutations do not affect its function significantly³⁵. CTD and CCD are linked by a characteristic kinked helix (see Figure 5).

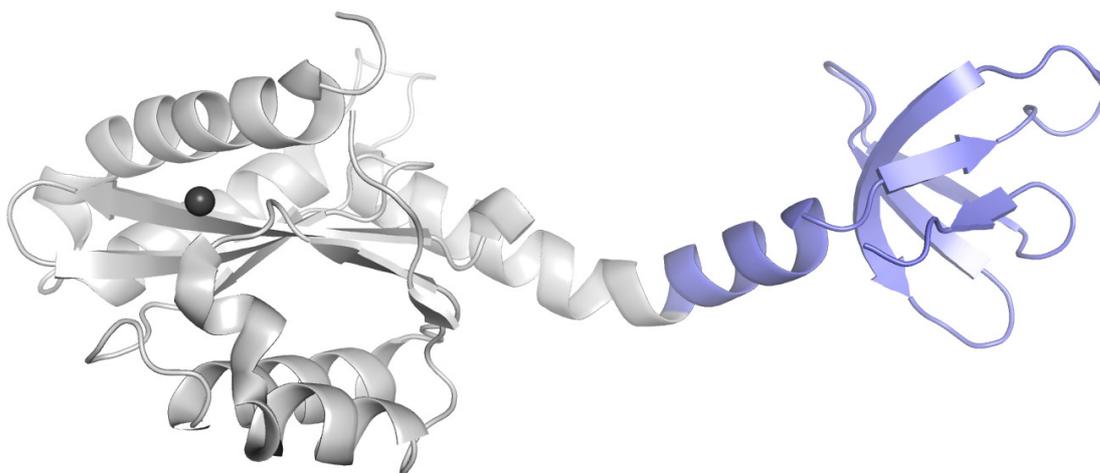


Figure 5: IN CTD structure (light blue) and CCD (grey) with the characteristic kinked helix (PDB code 1EX4).

Structural data are available from four PDB entries: **1IHV**, **1IHW**, and **1QMC** are NMR solution structures of the CTD alone, while **1EX4** is a two-domain structure of CCD and CTD derived from X-ray crystallography^{36,37}.

IN tetramerization

IN tends to form C_2 -symmetry-related dimers. In a single dimer, the two catalytic centers are on opposite sites, showing a distance of about 40 Å. Therefore, a single IN dimer can not perform the concerted integration of two vDNA ends into hDNA, since the observed five base-pair stagger would allow only a distance of about 18 Å (length of a five base stretch within B-form DNA). Thus, IN must act at least as a dimer of dimers, i.e. a “tetramer”, in a proper spatial arrangement. These considerations have also been proven experimentally^{38,39}.

Full-length IN (NTD + CCD + CTD) with bound DNA: Model creation

Until today, no full-length IN structure could be obtained, primarily due to the poor IN solubility. However, models of full-length IN complexed with viral and host DNA have been created. Such models are based on experimental and structural data of IN and other polynucleotidyl transferases. For the latter, crystal structures of the protein with DNA bound to its active site are available, e.g. Tn5 transposase cocrystallized with DNA (PDB code 1MUH)⁴⁰. Thus, vDNA binding to the IN active site can be modelled by using the coordinates of the structurally similar Tn5 transposase active site^{41,42}. The full-length IN protein can be modelled by combining the available single- or two-domain NTD, CCD, and CTD structures. For example, the model **1ZA9** was created by combining the NTD from **1WJA**, the CCD from **1BL3**, and the CTD from **1IHV** to form an IN tetramer; the vDNA was modelled into the binding sites by superposing IN with the Tn5 transposase-DNA structure 1MUH. The remaining hDNA was fitted into the model in a way that it was consistent with photo-crosslinking and mutational data. The resulting tetrameric IN-DNA model structure **1ZA9** therefore depicts the situation right before the insertion of vDNA into hDNA³⁸ (see Figure 6).

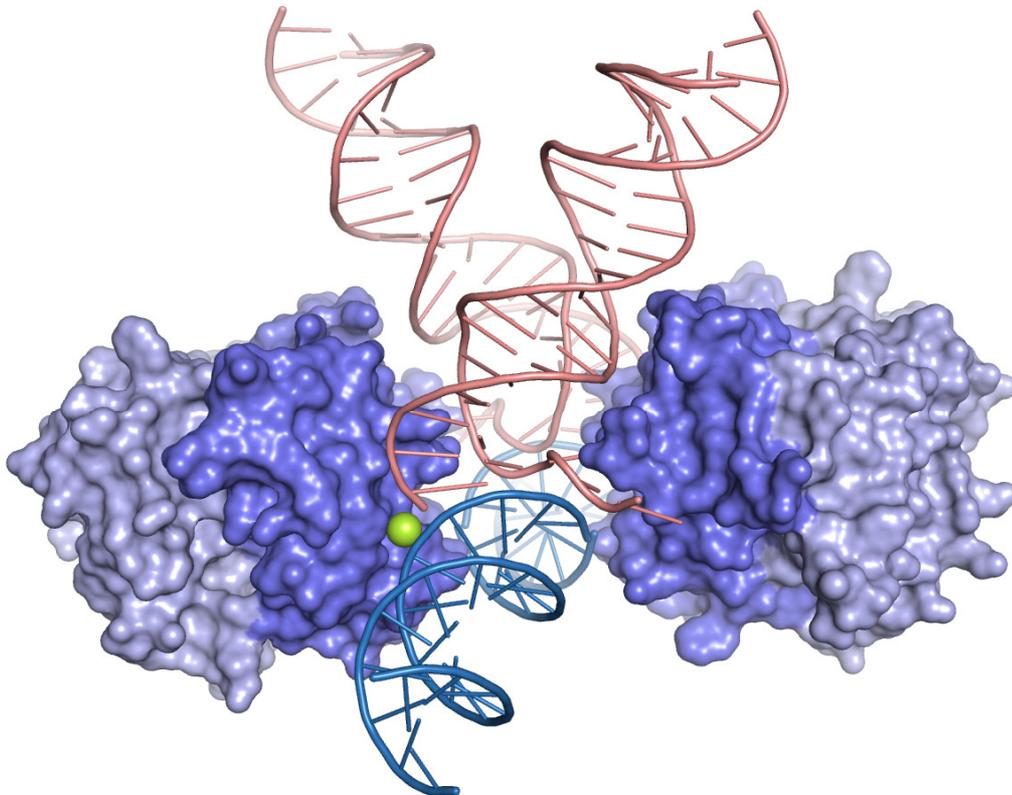


Figure 6: Full-length IN model 1ZA9 with hDNA (blue) and vDNA (red). The dimer of dimers (“tetramer”) provides two insertion sites (active site metal ion in green spheres; only one is visible).

Other, unpublished full-length IN-DNA model structures have been created using similar approaches⁴³. Liao *et al.* employed such a model structure in a structure-based virtual screening for IN inhibitors⁴⁴. Chen *et al.* created an active site model mimicking the IN-vDNA complex, neglecting NTD, CTD, and hDNA⁴⁵. This model was conceived as a template for the design of hDNA-competitive IN inhibitors (see Chapter 2.3.1), thus focusing solely on the active site. Again, Tn5 transposase served as a surrogate model for vDNA binding. Special attention was also paid to the active site loop interacting with vDNA, which was modelled by averaging over a molecular dynamics simulation-derived loop ensemble. The resulting model was employed in a docking study of recently discovered IN inhibitors to test its predictive power.

	Domain	M ²⁺	Ligand	Exp.	Res. [Å]	Mutations	CCD loop residues missing
1K6Y	NTD, CCD	-	-	X-ray	2.40	F185K, W131D, F139D	A-C: 140-148; D: 140-152
1WJA	NTD	-	-	NMR	-	-	-
1WJB	NTD	-	-	NMR	-	-	-
1WJC	NTD	-	-	NMR	-	-	-
1WJD	NTD	-	-	NMR	-	-	-
1WJE	NTD	-	-	NMR	-	H12C	-
1WJF	NTD	-	-	NMR	-	H12C	-
1EX4	CTD, CCD	-	-	X-ray	2.80	F185K, C56S, W131D, F139D, C280S	142-144; 138-149
1IHV	CTD	-	-	NMR	-	I234V, V265A	-
1IHW	CTD	-	-	NMR	-	I234V, V265A	-
1QMC	CTD	-	-	NMR	-	-	-
1ZA9	full-length model	Mg ²⁺	DNA	-	-	F185H	all present

Table 1a: IN NTD and CTD structures available in the PDB. Informations about the resolved domain, the presence of active site metal ions (M²⁺), ligands, experimental source, resolution and mutations of the IN protein are given. The last column specifies which residues of the active site loop are missing (protein chains are denoted as A, B, C, and D).

	Domain	M ²⁺	Ligand	Exp.	Res. [Å]	Mutations	CCD loop residues missing
1B92	CCD	-	-	X-ray	2.02	F185K, G149A	141-148
1B9D	CCD	-	-	X-ray	1.70	F185K	141-148
1B9F	CCD	-	-	X-ray	1.70	F185K, G140A, G149A	all present
1BHL	CCD	-	-	X-ray	2.20	F185H	138-153
1BI4	CCD	-	-	X-ray	2.46	F185H	A: 141-150; B: 140-149; C: present
1BIS	CCD	-	-	X-ray	1.95	F185K, W131E	A: 143-150, B: present
1BIU	CCD	Mg ²⁺	-	X-ray	2.50	F185K, W131E	A:141-147; B, C: 141- 148
1BIZ	CCD	-	-	X-ray	1.95	F185K, C56S	140-148
1BL3	CCD	Mg ²⁺	-	X-ray	2.00	F185H	A: 141-150; B: 140-149; C: present
1EXQ	CCD	Cd ²⁺	-	X-ray	1.60	F185K, C56S, W131D, F139D	144-150; 142-150
1ITG	CCD	-	-	X-ray	2.30	F185K	141-153
1HYV	CCD	-	[AsPh ₄] ⁺	X-ray	1.70	F185K	141-147
1HYZ	CCD	-	[As(Ph- OH) ₂ Ph ₂] ⁺	X-ray	2.30	F185K	141-147
1QS4	CCD	Mg ²⁺	5-CITEP	X-ray	2.10	F185K, W131E, A133G	A: 141-144; B: 141-142; C: 141-143
2ITG	CCD	-	-	X-ray	2.60	F185H	all present
2B4J	CCD	-	LEDGF	X-ray	2.02	F185K	A: 140-149; B: present

Table 1b: IN CCD structures available in the PDB. Informations about the resolved domain, the presence of active site metal ions (M²⁺), ligands, experimental source, resolution and mutations of the IN protein are given. The last column specifies which residues of the active site loop are missing (protein chains are denoted as A, B, C, and D).

Table 1a/b shows that a plethora of IN structures is available, yet only three are complexes with small-molecule ligands. While **1HYV** and **1HYZ** have tetraphenylarsonium and bis-(dihydroxyphenyl)-diphenyl-arsonium, respectively, as ligands in a somewhat extraordinary binding site at the dimer interface⁴⁶, only **1QS4** bears a potent inhibitor in the active site²⁷. Moreover, the observed binding mode of this ligand is most likely due to crystal packing effects (see Chapter 2.3.1). This clearly indicates the difficulties for structure-based IN inhibitor design: Since the crystallographic structures do not provide a reliable inhibitor binding mode, it is a difficult task to deduce protein-ligand interactions which are essential for inhibitory activity.

2.3 Integrase inhibitors

Due to its essential role in the HIV lifecycle, IN is a validated target in antiretroviral therapy. After more than 15 years of IN inhibitor development, the first and so far only marketed drug against IN, Merck's raltegravir (also known as MK-0518, brand name Isentress) was introduced not before October 2007^{47,48}. This highlights the difficulties in targeting this enzyme.

In the following chapter, the different classes of inhibitors will be introduced, where the focus lies on diketo acids and their derivatives, which are the most important class of inhibitors. However, not only "classical" inhibitors targeting the active site will be mentioned, but also compounds modulating the IN activity allosterically.

2.3.1 The diketo acid class of inhibitors

The class of diketo acid (DKA) inhibitors was discovered both at Merck and Shionogi by a random screening of a 250,000 compound library. DKA-like compounds are selective inhibitors of the strand transfer reaction: 3'-processing is inhibited only at much higher concentrations. They bind only to full-length integrase in association with viral DNA, which is the situation right after 3'-processing and before strand transfer. Thus, DKA-like compounds compete with hDNA for IN binding, preventing the final strand transfer. These compounds are classified as IN strand transfer inhibitors (INSTIs)⁴⁹.

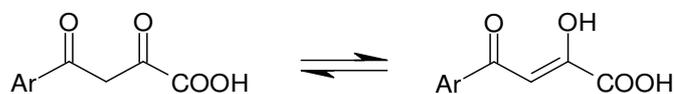


Figure 7: Tautomerization of the diketo group to the keto-enol form.

Figure 7 shows that the diketo group can tautomerize to its keto-enol form. In Table 2, examples of DKA-like compounds are listed. The carboxyl group (**L-708,906**, **L-731,988**)⁴⁹ can be replaced by bioisosteric groups like tetrazole or triazole (**5-CITEP**, **S-1360**)^{25,50}. For these compounds, the term diketo aryls is more suitable. **5-CITEP** was the first inhibitor in an IN co-crystal structure.

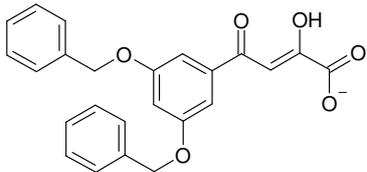
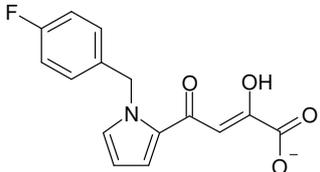
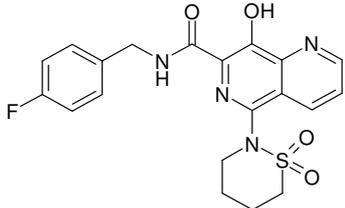
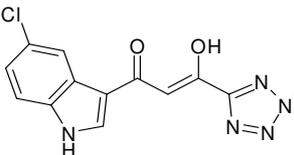
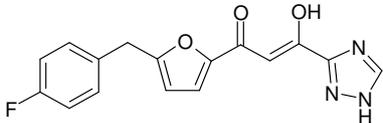
	Structure	IC ₅₀ (3'-P) [μM]	IC ₅₀ (ST) [μM]
L-708,906		27	0.06
L-731,988		6	0.08
L-870,810		0.25	0.015
5-CITEP		35	0.65
S-1360		n.a.	0.02

Table 2: Examples of DKA and DKA-related compounds. A strong selectivity for strand transfer (ST) inhibition is observed.

Systematic studies of further bioisosteric groups led to DKA derivatives like Merck's naphthyridine carboxamide (**L-870,810**)⁵¹, where the parent DKA motif is not evident anymore. Interestingly, the substitution of the carboxylic acid group is possible not only with other acidic groups like tetrazole or triazole, but also with basic heterocycles. It was proposed that the heteroaromatic nitrogen is the Lewis base equivalent of the parent carboxylate (see Figure 8)⁵².

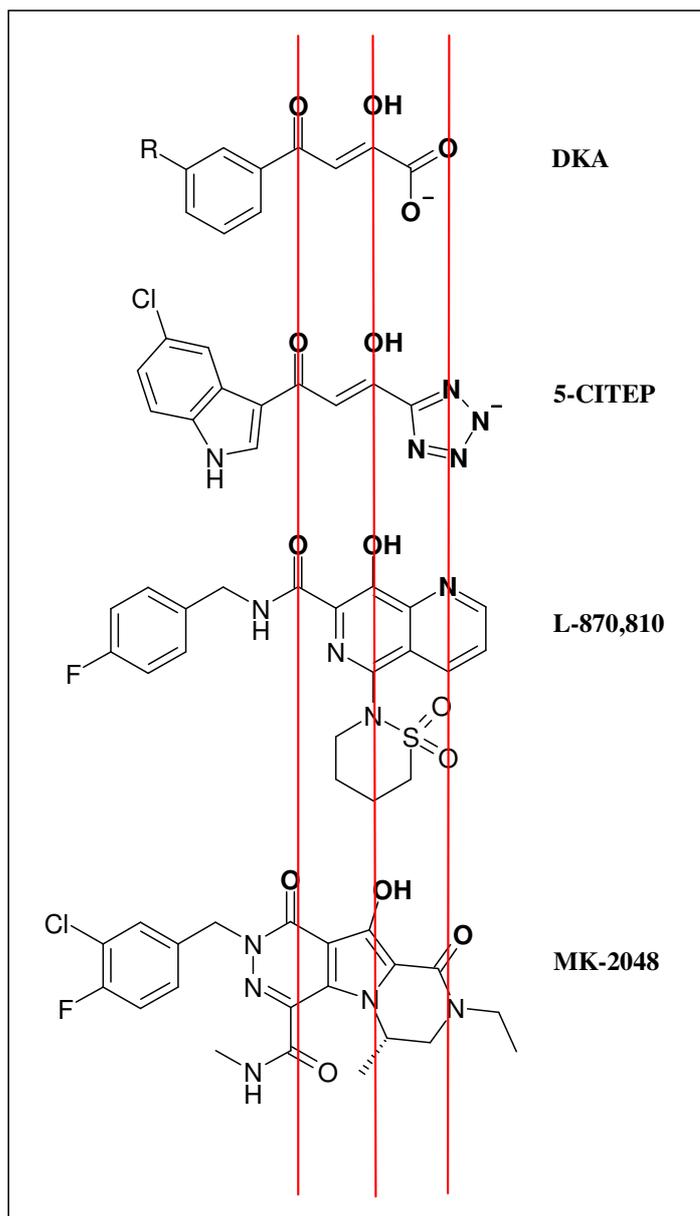


Figure 8: Modifications of the parent DKA motif leading to DKA derivatives.

Especially Merck's **L-870,810** and Shionogi's **S-1360** had been promising compounds for further development, with IC_{50} values in the nanomolar range, before both projects were stopped in clinical phase II due to toxicologic and pharmacokinetic issues.

Current drug candidates

As mentioned earlier, the first and only marketed IN inhibitor is raltegravir (**MK-0518**). This pyrimidone carboxamide shows nanomolar activity, a good pharmacokinetic profile, and is currently administered as part of HAART regimens. Three drug candidates are in advanced clinical trials⁵³.

	Structure	IC ₅₀ (ST) [μM]
MK-0518 (raltegravir)		0.016
MK-2048		n.a.
GS-9137 (elvitegravir)		0.007
GSK-364735		0.008

Table 3: Promising DKA-derived drug candidates and their inhibitory activity. Raltegravir has already been approved (shown here as a reference).

MK-2048 is a further development of **MK-0518**, yet with a different scaffold: A three-membered ring-system with a hydroxypyrrol center bears the DKA-derived pharmacophoric motif. This compound has proven to be active against IN mutants displaying raltegravir resistance⁵⁴. Gilead Sciences' **GS-9137** (also known as elvitegravir) is another DKA-derived variation: A quinolone

carboxylic acid inhibitor. Remarkably, this compound bears a monoketo instead of a diketo motif⁵⁵. **GSK-364735** is a naphthyridinone, which was developed as a follow-up to DKA **S-1360** by the involved groups; however, the parent motif is hardly apparent anymore⁵⁶.

DKA mode of action

DKAs and their derivatives are thought to chelate the IN active site metal ion cofactor. By sequestering Mg^{2+} or Mn^{2+} , respectively, hDNA binding is impeded and strand transfer is blocked (see Figure 3). This hypothesis is supported by the finding that IN inhibition by DKAs is metal-dependent: For example, the IC_{50} value of **L-708,906** is almost tenfold higher in the presence of Mn^{2+} compared to Mg^{2+} ⁵⁷. For a closer analysis of the DKA mode of action, see Chapter 2.4. Currently, **5-CITEP** is the only active site ligand observed in an IN catalytic core domain structure (PDB code **1QS4**). The inhibitor had been introduced by soaking preformed crystals, binding only to one of the three active sites within the asymmetric unit²⁵. **5-CITEP** is located between the active site residues D64, D116, and E152. The tetrazole ring acts as hydrogen bond acceptor for K156 and K159. Unexpectedly, the diketo-tetrazole moiety does not chelate the active site metal ion. It rather interacts with the identical active site inhibitor of the symmetry-related IN monomer. Thus, the observed binding mode is likely due to crystal packing effects. Docking studies by Sotriffer *et al.* revealed a binding mode unaffected by crystal packing: When **5-CITEP** was docked to the IN active site without any crystallographic neighbour, the ligand orientation was different from the crystal structure: The molecular plane is perpendicular compared to **1QS4**, lying flat in the active site, thereby providing favorable van-der-Waals interactions⁵⁸ (see Figure 9).

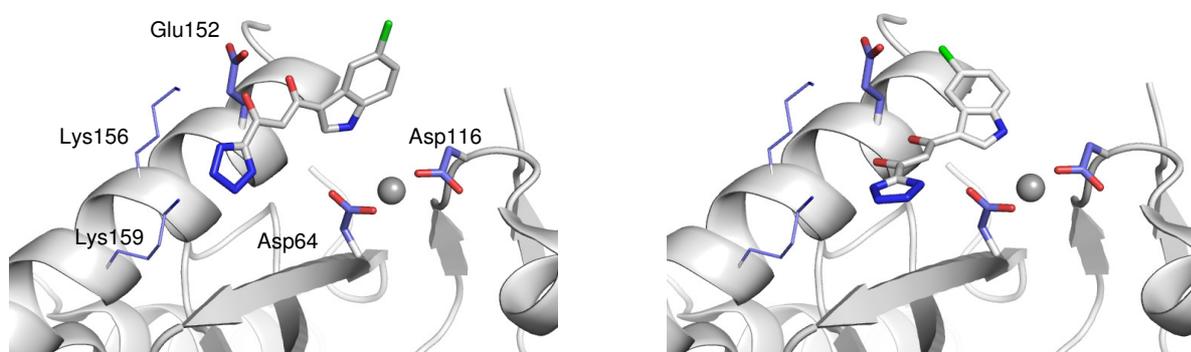


Figure 9: IN complexed with 5-CITEP. Left: Crystallographic binding mode, in which the chloroindole system is quite exposed. Right: Binding mode unaffected by crystal packing, discovered by docking, where the ligand adopts a perpendicular orientation.

Despite these structural studies, there is still very little information about the native binding mode of IN inhibitors. Especially the INSTIs pose a delicate case, because these compounds bind to an IN-DNA complex which is not straightforward to model. Some approaches towards modelling such an IN-DNA complex have recently been reported (e.g. structure **1ZA9**, see Chapter 2.2.2). Furthermore, not only the target structure, but also the ligands themselves bear some uncertainties: Sechi *et al.* suggested that DKA compounds actually act as complexes rather than free ligands, i.e., they chelate metal ions already in solution before they bind to the IN active site⁵⁹. For a detailed discussion of this topic, see Chapter 2.4.

2.3.2 Other integrase inhibitor classes

Beside the DKA derivatives, a vast number of structurally diverse compounds with inhibitory activity has been discovered. The first selective strand transfer inhibitor reported was caffeic acid phenethyl ester (**CAPE**). Like the DKA class of INSTIs, it is presumed that **CAPE** inhibits IN by sequestering an active site metal ion, most probably with its chelating dihydroxy group⁶⁰. This motif is also present in other structurally diverse IN inhibitors, which are not INSTIs, but inhibit both 3'-processing and strand transfer equally well. Examples of such polyhydroxylated aromatic compounds (see Figure 10) are quinalizarin (**QLZ**), tetramethyl-spiro-bisindan (**TMS**), styryl-quinoline (**SQL**), L-chicoric acid (**LCA**), and the natural product hypericin. They inhibit IN in the low micromolar range^{61,62}.

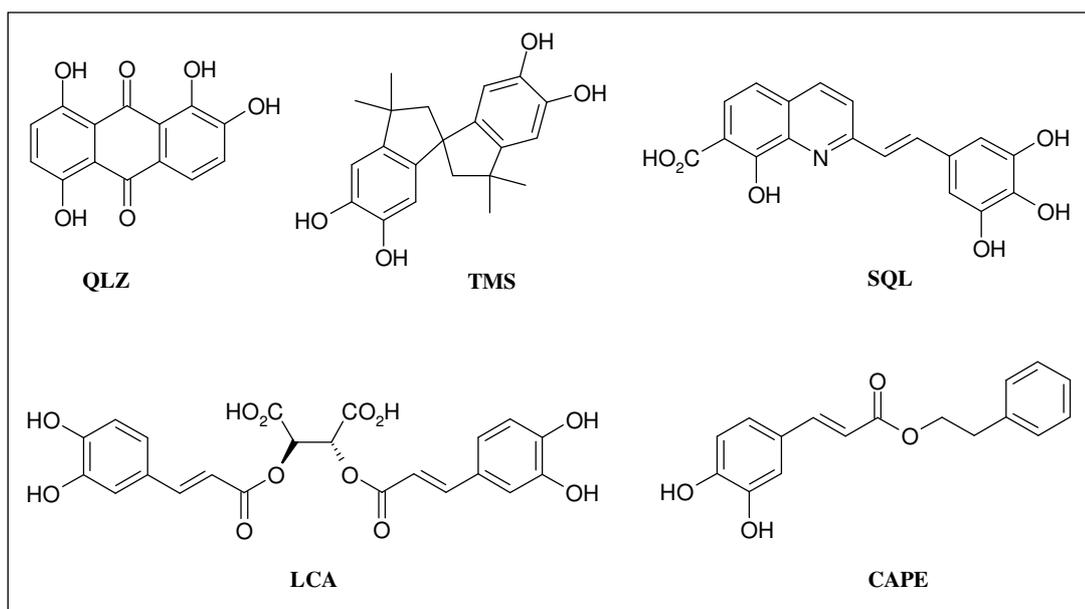


Figure 10: Representatives of the large class of polyhydroxylated IN inhibitors.

Polyamides or lexitropsins have recurrent pyrrole rings linked by amide groups. They bind DNA at AT-rich sequences, which occur on the vDNA LTR ends. Thereby, vDNA recognition by IN is hampered and IN catalytic activity is inhibited. Polyamides are therefore inhibitors which do not bind the enzyme, but the substrate. Further development of this inhibitor class is not promising, since it is virtually impossible to address the vDNA specifically⁶³.

Guanosine quartets and other oligonucleotides can inhibit both 3'-processing and strand transfer *in vitro*, but also interfere with other viral replication steps, for example with the HIV cell entry protein gp120 *in vivo*⁶⁴.

Peptides binding to an enzyme, thereby leading to inhibition, have also been reported for IN. Plasterk *et al.* were the first to find a hexameric peptide, HCKFWW, which inhibited 3'-processing and strand transfer with an IC_{50} of 2 μ M. This peptide was identified in a random screen of a 400,000 hexapeptide combinatorial library⁶⁵. Maroun *et al.* took a more rational approach: They synthesized peptides termed INH1 and INH5 which were derived from IN catalytic core sequences located at relevant structural positions. Both peptides showed moderate to strong inhibition (IC_{50} of INH5 \sim 60 nM)⁶⁶. In a similar approach, Zhao *et al.* synthesized peptides resembling IN catalytic core interfacial helices and strands. Three out of five peptides displayed activity in the low micromolar range, two of them had common amino acid sequences with Maroun's inhibitory peptides INH1 and INH5⁶⁷. Neamati *et al.* performed an extensive study with 16 oligopeptides derived from all three IN domains. In this so-called „sequence walk“ strategy, two peptides were found with IC_{50} values in the low micromolar range. Again, these peptides were similar to INH1 and INH5. Further truncation studies resulted in a hexameric peptide YFLLKL with an IC_{50} value of 20 μ M for both 3'-processing and strand transfer⁶⁸.

Today, there is a plethora of IN inhibitor classes and still new scaffolds with inhibitory activity are discovered. Yet, only few compounds have entered advanced clinical trials. New approaches, like the search for allosteric binders, are required to keep the IN inhibitor pipeline filled.

2.3.3 Allosteric inhibitor binding sites

Due to the lack of sufficient structural data of IN-inhibitor complexes, final conclusions about inhibitor binding sites can hardly be drawn. Naturally, at first one would expect inhibitors to bind to the active site. In the case of DKAs and their derivatives, this expectation is supported by numerous experimental observations, for example metal-dependency or mutational data. However, other inhibitor classes do not necessarily bind to the active site. One example is the weakly active tetraphenylarsonium inhibitor, which has been co-crystallized with IN (PDB code **1HYV**): Its binding site is located distant from the active site, right at the catalytic core dimer interface. It is presumed that structural changes within the catalytic center induced by tetraphenylarsonium binding lead to inhibition. However, uncertainties of this exposed binding position remain, since the inhibitor has contact to a sidechain of a third crystallographically related monomer, raising the possibility of crystal packing effects (see Figure 11). As explained in Chapter 2.2.1, the pre-integration complex (PIC) enters the cell nucleus to perform the final step of viral genome integration, the strand transfer reaction. Cellular proteins within the PIC exert influence on IN catalytic activity. Targeting these interactions could also pose an attractive approach towards HIV genome integration inhibition. Especially the LEDGF-IN interaction has entered IN inhibitor research recently. PDB structure **2B4J** shows that LEDGF binds to the catalytic core surface, close to W132⁶⁹ (see Figure 11).

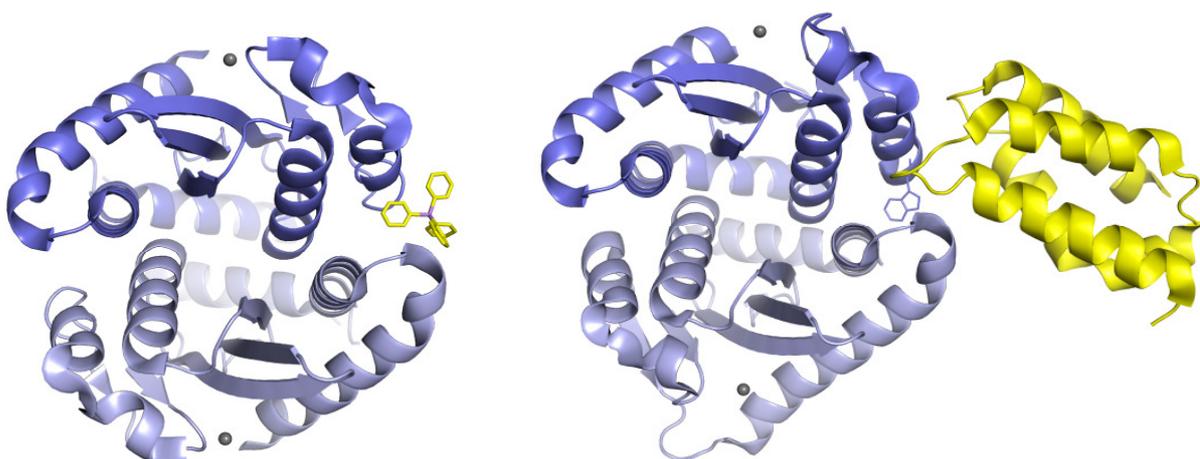


Figure 11: Left: Binding of $[\text{AsPh}_4]^+$ (yellow sticks) to the IN catalytic core dimer interface (blue), as observed in 1HYV. The active sites are located in the area of the gray spheres representing the metal ions (taken from 1QS4). Right: Binding of LEDGF (yellow) to IN catalytic core dimer (blue), as seen in 2B4J. The binding site is close to W132 (sticks) at the dimer interface.

Shkriabai *et al.* revealed the binding site of a chicoric acid derivative by means of affinity acetylation and subsequent mass spectrometry analysis⁷⁰. Their data indicate binding of this inhibitor to the dimerisation interface, close to K173, T174, and M178. Al-Mawsawi *et al.* took a similar approach: Instead of affinity acetylation, they used photoaffinity labelling with subsequent mass spectrometry to identify the binding site of hydroxycoumarines. Results showed that the inhibitor binds close to residues E131 and W132, on the surface of the IN core domain, close to the dimerization interface. This finding was supported by mutational data, i.e., mutations within this region resulted in loss of inhibitory activity⁷¹ (see Figure 12).

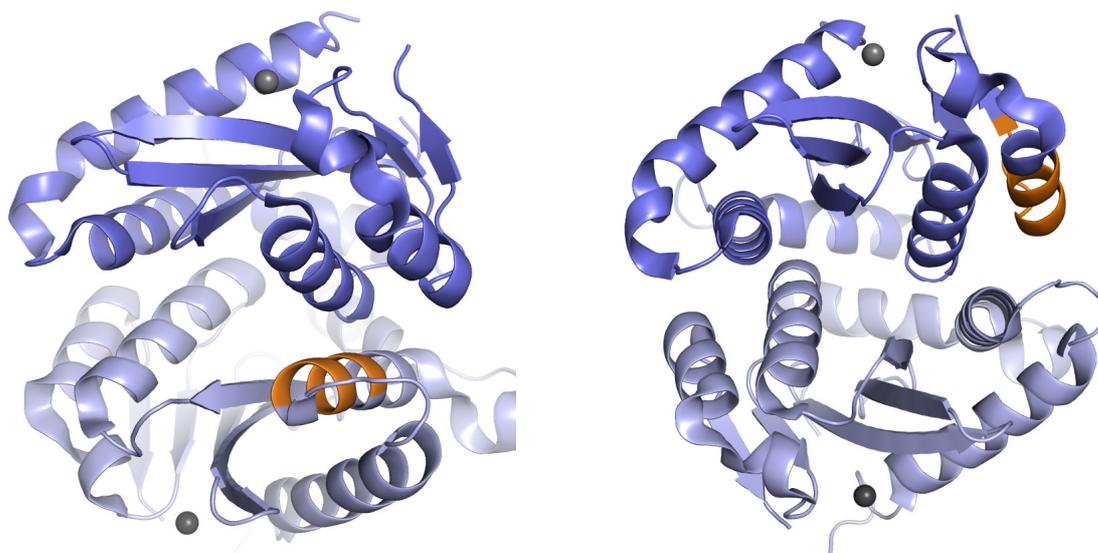


Figure 12: Left: Binding site revealed by affinity acetylation: The chicoric acid derivative inhibitor binds at the IN catalytic core dimer interface, close to residues K173, T174, and M178 (orange; active site Mg²⁺ as gray spheres). Right: Binding site revealed by photoaffinity labelling: The hydroxycoumarine inhibitor binds close to residues E131 and W132 (orange).

Maroun *et al.* analyzed IN inhibition by oligopeptides INH1 and INH5, which had been derived from the IN dimerization interface helices $\alpha 1$ and $\alpha 5$, respectively. These helices are essential for catalytic core dimerization by mutual interaction with their respective counterpart, i.e., helix $\alpha 1$ of monomer A interacts with helix $\alpha 5$ of monomer B and vice versa. Size exclusion chromatography showed that both peptides perturbed the IN monomer-dimer-tetramer equilibrium, leading to catalytically inactive monomers. Fluorescence spectroscopy and circular dichroism measurements further confirmed the hypothesis that INH1 and INH5, resembling helices $\alpha 1$ and $\alpha 5$, bind to their respective protein counterparts, thereby preventing dimerization and thus inhibiting IN⁶⁶. In contrast to the above mentioned allosteric binding sites, the INH1 / INH5 binding site is the dimerization interface of a single monomer (see Figure 13).

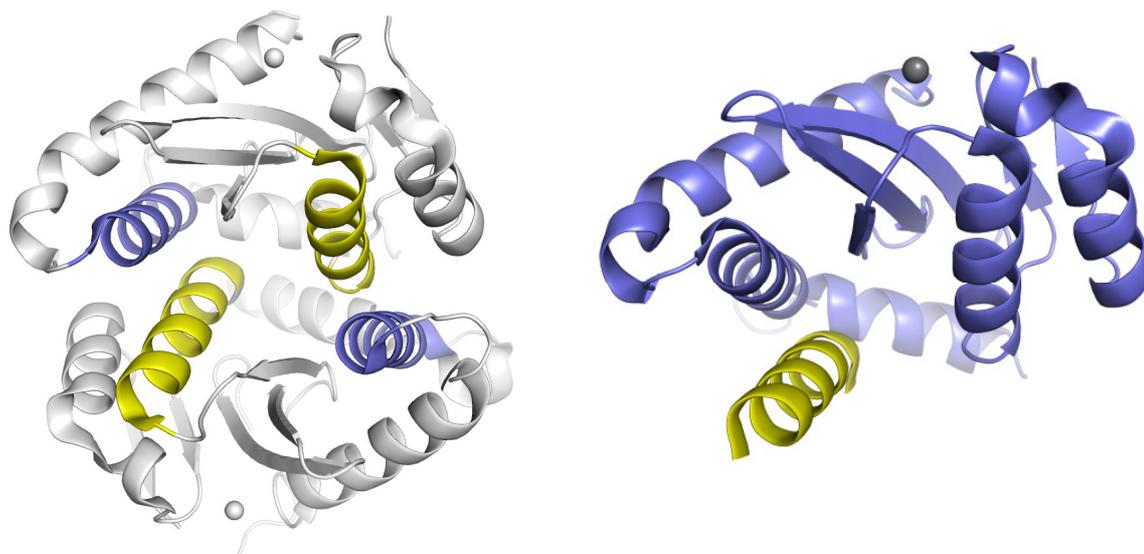


Figure 13: Left: Interaction of monomer A α 1-helix (blue) with monomer B α 5-helix (yellow) and vice versa. This interaction is crucial for IN dimerisation. Right: Presumed binding site of the inhibitory peptide INH5 (yellow, corresponding to α 5-helix). It interacts with the α 1-helix at the dimerisation interface of an IN monomer, thereby preventing binding of a second monomer.

Molecular dynamics studies of the IN core domain performed by Schames *et al.* revealed a new binding trench formed by residues H114, F139, G140, P142, and Q148, adjacent to the active site. They proposed dimeric compounds (referred to as „butterfly compounds“) with two diketo acid motifs to address both the active site and this new trench⁷².

Remarkably, the experimental studies revealed binding sites which are right at or close to the dimerization interface. Although details are not known, it indicates that this region plays a crucial role in overall IN organization. The considered binding regions are all located on the catalytic core domain, the largest IN domain. Of course, inhibitors could also bind to the terminal domains or even to the DNA substrate. Now that alternative binding sites have been discovered, the mode of action of allosteric IN inhibitors still has to be revealed. It has been hypothesized that allosteric binding interferes with the delicate ensemble of domain and substrate interactions or impedes binding of cellular proteins⁷³. Further studies need to be carried out to explore the amenability of these alternative binding sites for IN inhibitor design.

2.4 Diketo acids – Mechanism of action

The class of DKAs and DKA derivatives is currently the most prominent class of IN inhibitors (see Chapter 2.3.1). It is assumed that the diketo motif is involved in sequestering the IN active site metal ions which are essential for DNA integration. Nevertheless, the detailed mechanism of action still needs to be revealed.

Formation of M^{2+} -DKA complexes

Given the fact that the intracellular and assay concentrations of metal cofactors are significantly higher than the IN concentration ($8 \cdot 10^{-3}$ mol/l and $2 \cdot 10^{-7}$ mol/l, respectively), Sechi *et al.* suggested that the DKAs build ligand-metal complexes already in solution, i.e. prior to protein binding. To validate this hypothesis, potentiometric, structural, and biological data were collected. Indeed, different ligand-metal complexes could be detected, showing complex-dependent biological activity⁵⁹.

In Chapter 2.2.2, the role of divalent metal ions for IN-mediated DNA processing was outlined. It is hypothesized that the negatively charged DNA phosphodiester backbone interacts with the positively charged metal ions, which themselves are coordinated by the IN catalytic triad D64, D116, and E152. Although only one divalent metal ion is observed in IN crystal structures, it is widely accepted that IN-mediated DNA processing requires two metal cofactors⁷⁴. DKAs are capable of chelating metal ions in different ways (see Figure 14).

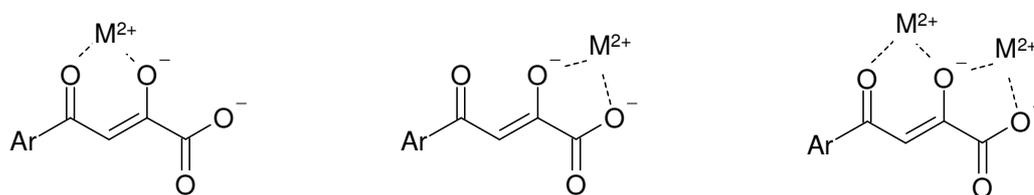


Figure 14: Possible metal chelation by a diketo acid motif (metal-to-ligand ratios 1:1, 1:1, 2:1).

Accordingly, it is assumed that DKAs inhibit IN by sequestering one or both of the metal ions required for catalysis, thereby preventing the formation of an IN-DNA complex^{75,76}. Sechi *et al.* performed a systematic study to verify this hypothesis about the DKA mode of action: Starting from DKA model ligands derived from known IN inhibitors, they provided experimental evidence for their complexing ability and synthesized new complex-based inhibitors.

To support the hypothesis on a structural level, molecular modelling studies were required, the results of which are described in Chapter 2.4.2, after an overview of the experimental data in the following section.

2.4.1 Experimental basis

To verify the assumed complexing ability of the diketo acid motif, Sechi *et al.* synthesized four model compounds (H_2L^1 , HL^2 , H_2L^3 , HL^4 in Figure 15; here, the number of hydrogens “H” indicate the protonation state of the four ligands L^{1-4}) and performed potentiometric measurements in the presence of magnesium ions. At physiological pH, the obtained data suggested formation of complexes with different metal-to-ligand ratios (Table 4).

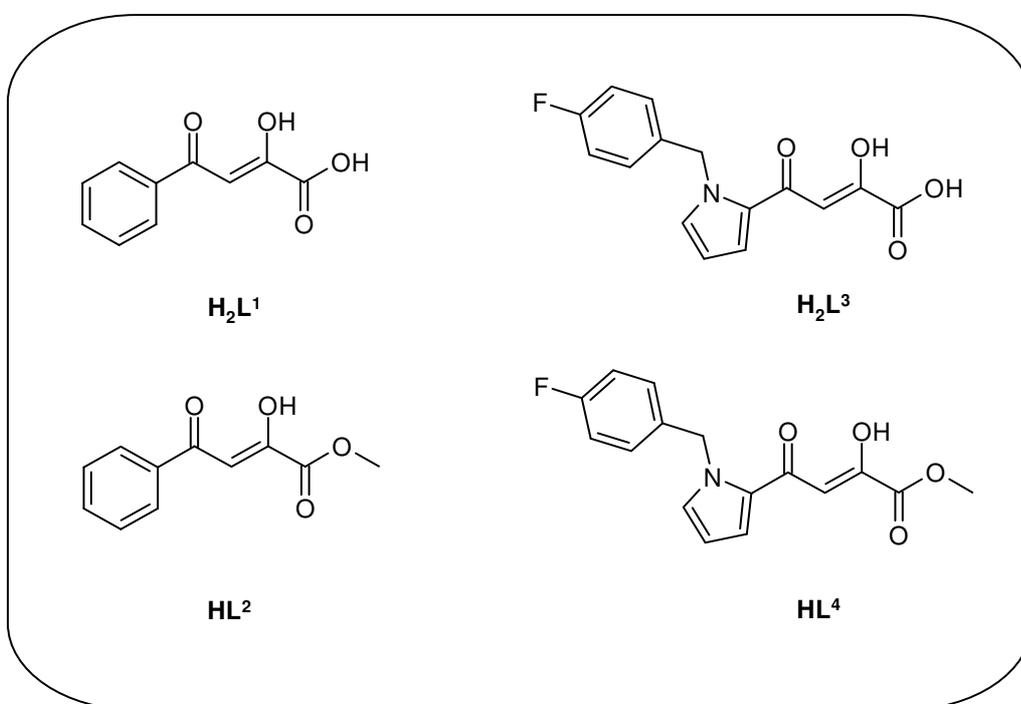


Figure 15: Model compounds for this study, comprising two diketo acids (L^1 , L^3) and their respective methyl esters (L^2 , L^4). The number of hydrogens refers to the protonation state of the hydroxyl group and the carboxyl group, respectively.

	Metal-to-ligand ratio	Species formed
H_2L^1	2:1	$[M_2L]^{2+}$
HL^2	1:1, 1:2	$[ML]^+$, ML_2
H_2L^3	2:2	M_2L_2
HL^4	1:1, 1:2	$[ML]^+$, ML_2

Table 4: Potentiometry results for the free ligands.

After confirming the metal complexing ability of diketo acids in solution by means of potentiometric measurements, a series of different metal complexes of the model ligands was synthesized. The synthesis scheme is shown in Figure 16: Here, the compound naming follows the naming convention in the original publication⁵⁹.

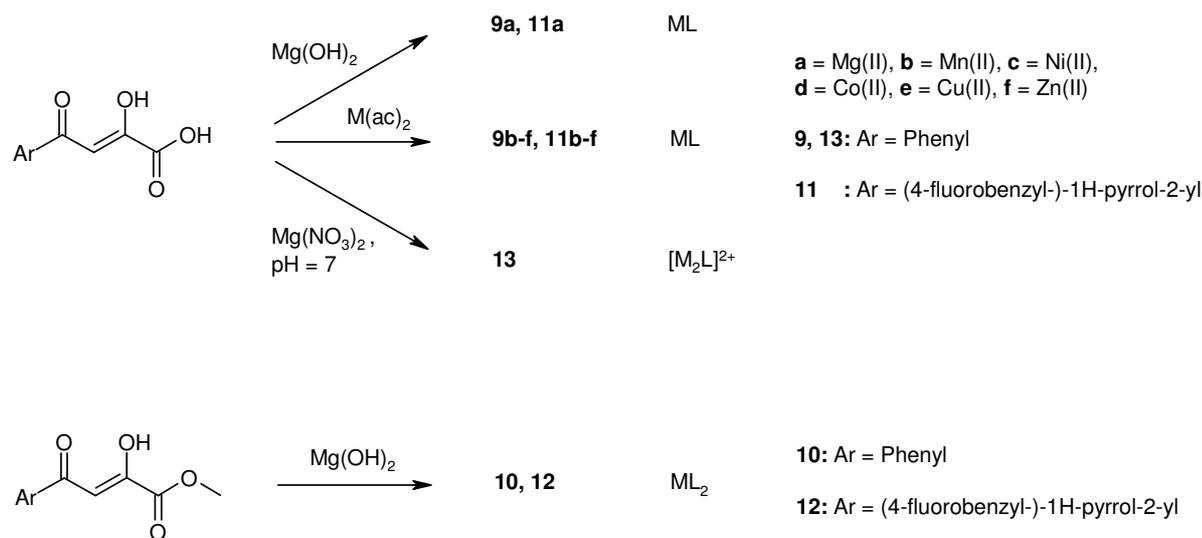


Figure 16: Synthesis of model complexes. By using basic salts, the educt ligands are deprotonated. Information about the species formed is given in the last but one column.

Series **9a-f** comprises ligand L^1 in coordination with different divalent metal ions (Mg^{2+} , Mn^{2+} , Ni^{2+} , Co^{2+} , Cu^{2+} , and Zn^{2+}), whereas series **11a-f** comprises the respective complexes of ligand L^3 . In both series, the metal-to-ligand ratio of the synthesized complexes is 1:1. However, to investigate the possible 2:1 ratio, which had been observed in the potentiometric measurements, a further complex of ligand L^1 was synthesized: **13** is a complex with two magnesium ions per ligand molecule. With respect to the esters L^2 and L^4 , only one complex per ligand was

synthesized (**10**, **12**), with a metal-to-ligand ratio of 1:2. The latter means that one metal ion is coordinated by two ligand molecules. To investigate the inhibitory activity of free ligands and their respective complexes, the compounds were subjected to biological testing in an IN inhibition assay. Apart from two exceptions, all compounds display activity in the low to medium micromolar range, with the typical strand transfer selectivity of diketo acid derivatives. For a detailed overview, see Table 5.

	M²⁺	Ligand	IC₅₀ (strand transfer) [μM]	Metal-to-ligand ratio
H₂L¹	-	L ¹	69	-
HL²	-	L ²	>333	-
H₂L³	-	L ³	0.54	-
HL⁴	-	L ⁴	48	-
9a	Mg ²⁺	L ¹	44	1:1
9b	Mn ²⁺	L ¹	42	1:1
9c	Ni ²⁺	L ¹	31	1:1
9d	Co ²⁺	L ¹	5	1:1
9e	Cu ²⁺	L ¹	11	1:1
9f	Zn ²⁺	L ¹	21	1:1
10	Mg ²⁺	L ²	>333	1:2
11a	Mg ²⁺	L ³	0.26	1:1
11b	Mn ²⁺	L ³	0.99	1:1
11c	Ni ²⁺	L ³	0.52	1:1
11d	Co ²⁺	L ³	0.72	1:1
11e	Cu ²⁺	L ³	0.62	1:1
11f	Zn ²⁺	L ³	0.74	1:1
12	Mg ²⁺	L ⁴	8	1:2
13	Mg ²⁺	L ¹	88	2:1

Table 5: Activity data for free ligands and metal complexes. The values for 3'-processing inhibition have been omitted for clarity.

The activity data can be summarized as follows: With the exception of **HL**² and **10**, all compounds show inhibitory activity against IN-mediated strand transfer (values for 3'-processing are five- to tenfold higher). In comparison to the parent acids, the esters are less active (or not active at all). Ligand **L**¹ and its complexes (IC₅₀ values ranging from 5-69 μM) are generally less active than ligand **L**³ and its complexes (IC₅₀ values ranging from 0.26-0.99 μM). This finding points out the major influence of the aromatic moiety on the ligand's potency. In the **9a-f** series, the metal complexes show an increase in potency, compared to their parent free ligand **L**¹: For example, the cobalt complex **9d** displays an almost 14-fold increase in inhibitory activity. Generally, in the **9a-f** series, a metal-dependency is observed: Shifts in potency are related to different divalent metal ions, exemplified by the case of **9a** (Mg²⁺, IC₅₀ = 44 μM) in comparison to **9d** (Co²⁺, IC₅₀ = 5 μM). In the **11a-f** series, the metal complexes generally show a decrease in potency, compared to their parent free ligand. Exceptions are **11a** and **11c**: These complexes show an increased potency. Concerning the esters **HL**² and **HL**⁴, only one complex per ester has been synthesized (**10**, **12**). In the case of **HL**², which itself shows no inhibitory activity at 333 μM, the respective complex **10** also has no effect on IN-mediated strand transfer. Complex **12** shows a sixfold increase in potency, compared to its free ester **HL**⁴.

To address the primary question whether diketo acid derivatives act as complexes, data of the potentiometric measurements and the biological testing need to be compared. The central point is to compare inhibitory activity data of free ligands and complexes. It is reasonable to assume that if the inhibitory activity of a complex is comparable to the activity of the free ligand, this very complex resembles the species formed by the free ligand and metal ions (with the respective metal-to-ligand ratio) under physiological conditions. Looking at Table 5, the free ligand **H₂L**¹ has an IC₅₀ of 69 μM. Potentiometric measurements suggest that this ligand preferably forms [M₂L]²⁺ complexes under physiological conditions. Indeed, complex **13** (with this very metal-to-ligand ratio of 2:1) has an inhibitory activity of 88 μM (IC₅₀ error margins of ~10 %), which is comparable to the IC₅₀ value of the free ligand **H₂L**¹. For the ester **HL**², no preferred complex form could be deduced from the activity data (same for ester **HL**⁴). Potentiometric measurements indicate that the free ligand **H₂L**³ preferably forms [M₂L₂] complexes, which is further supported by the comparable inhibitory activities of its complexes **11a-f** (metal-to-ligand ratio of 1:1, the stoichiometry does not distinguish between a 1:1 and a 2:2 ratio).

The hypothesis of DKAs acting as complexes has been supported by experimental data. Now, further issues and questions, especially on the structural level, need to be addressed. First and foremost, it is important to understand how DKA complexes bind to the IN active site (which normally possesses a metal cofactor itself): Do the metal complexes bind to the IN active site with or without its metal cofactor? Does an exchange of metal ions between complexes and the IN active site happen? Of course, the latter issue is virtually impossible to be addressed by conventional modelling techniques, since an exchange is a dynamic process. However, modelling studies could shed light upon further important issues related to DKA complexes: The coordination bifunctionality of the diketo acid moiety allows different speciation models for the complexes, meaning that the ligands can chelate the metal ions by either their diketo function or keto-carboxylate function (see Figure 14). Up to now, there is no evidence which form is prevalent. Additionally, different metal-to-ligand ratios for a given ligand need to be investigated, i.e. ratios of 1:1 versus 2:2 (complexes with ligand L^3) or 1:1 versus 1:2 (ligands L^2 , L^4). Consequently, the DKA mode of action on a structural level still has to be revealed. Extensive docking studies addressing this issue are the subject of the following chapter.

2.4.2 Comparative docking of free ligands and complexes

The hypothesis that inhibitors HL^1 - HL^4 act as complexes was tested on a structural level by performing comparative docking studies. Here, the free ligands were docked to the IN active site containing a Mg^{2+} ion and the respective ligand complexes were docked both to a metal-free IN active site and to the protein including its metal ion. Furthermore, to address the aforementioned question of different speciation models, various chelation patterns were tested for Mg^{2+} complexes. Finally, the results from the docking studies were compared with respect to favorable binding energies and binding geometries, indicating whether HL^1 - HL^4 act preferably as complexes or free ligands.

Figure 17 gives an overview of the docked species: The free ligands were docked in their keto-enol (HL^1 - HL^4) and their keto-enolate form (L^1 - L^4). For ligand L^1 , three complexes were created: The metal-to-ligand ratio 1:1 comprised two chelation patterns, i.e. the chelation via acetyl-acetonate (**9a_aa**) and via hydroxy-carboxylate (**9a_hc**); for the metal-to-ligand ratio 2:1 (which was found to be the dominant species in potentiometric measurements), complex **13a** was created. Two species were created for ligand L^2 : Metal-to-ligand ratio 1:1 and 1:2 (see Table 4), represented by complexes **10a_mono** and **10a_bis**, respectively. Since a carboxylate group is missing in this ligand, only the acetyl-acetonate chelation is possible. For ligand L^3 , one complex

form was created (**11a**), having a metal-to-ligand ratio of 2:2. Complexes of ester **L**⁴ were created in the same way as for ester **L**² (**12a_mono** and **12a_bis**).

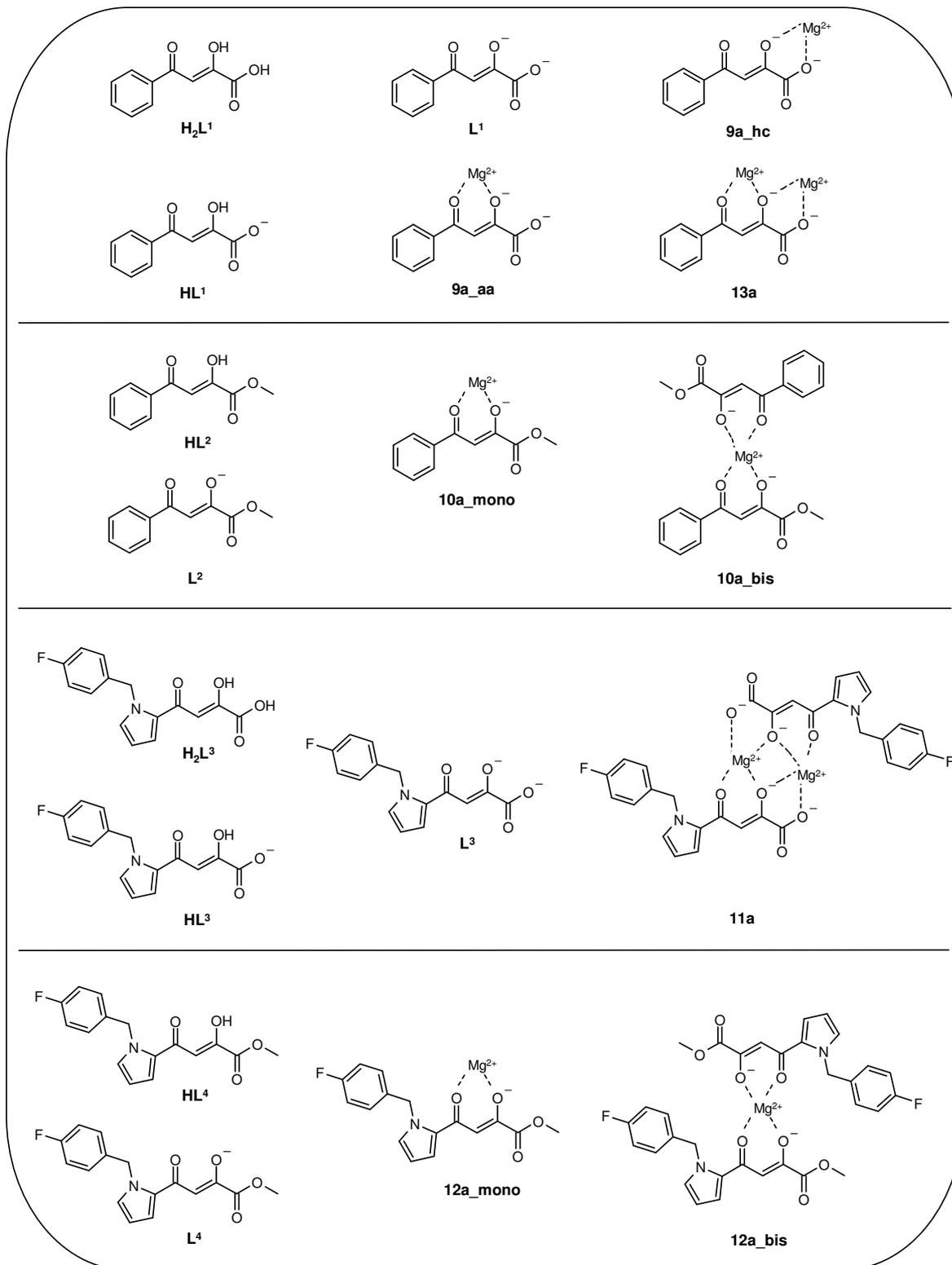


Figure 17: Free ligands and their possible complex species analyzed in the comparative docking study: Free ligands were docked to the IN active site with its cofactor Mg²⁺, complexes were docked to the active site with and without its metal cofactor. The compound codes follow the naming convention of the original publication⁵⁹.

Methods

For a general introduction to the methodologies applied in the thesis, see Chapter 5.1. Free ligand structures were generated by converting SMILES to mol2 format using CORINA3.40⁷⁷. Subsequently, hydrogens were added and Gasteiger-Marsili charges⁷⁸ were assigned using Sybyl7.1⁷⁹. Complex structures were created by placing the metal ions manually in a 2 Å distance to the chelating oxygen atoms of the respective ligand, again using Sybyl7.1. Charges of +2.000 were assigned to the metal ions. The *autotors* module was employed to create the ligand docking input (pdbq) files. Bonds within the conjugated framework were manually set as rigid.

AutoDock3.0⁸⁰ was used for the docking experiments. The protein input file was the modified 1QS4 structure (generated by Sotriffer *et al.*⁸¹), which includes the active site loop. Polar hydrogens were added using the adapted AMBER *protonate* tool⁸². Kollman united-atom partial charges^{83,84} and solvation parameters (derived from the method of Stouten *et al.*⁸⁵) were assigned by the inhouse-tool *add_chrgsol*. The docking site was defined by a cubic box with 60 grid points in each dimension, centered on the IN active site Mg²⁺, and the grid spacing was set to 0.375 Å. Interaction maps were calculated with AutoDock's *autogrid* module. The docking protocol used an initial population size of 50, a maximum number of 3.0·10⁶ energy evaluations, a mutation rate of 0.02, a crossover rate of 0.80, and an elitism value of 1. The Solis and Wets local search parameters were set as follows: The probability of performing a local search on an individual was set to 0.06, a maximum number of 300 iterations was performed, and the number of consecutive successes or failures before changing the size of the local search space was 4. 50 GA runs per ligand were carried out. The resulting binding poses were clustered, with a cluster tolerance of 1.0 Å.

Results – Free ligands docked to protein with metal cofactor

Docking of free ligands shows no complexation of the active site Mg²⁺ ion. The carboxy (ester) groups of **HL**¹, **HL**³, and **HL**⁴ are placed close to Lys156 and Lys159, respectively. Comparison of the keto-enol with the keto-enolate form shows almost identical binding orientations, with deviations in the case of **HL**¹. The keto-enolate form (**L**¹, **L**², **L**³, and **L**⁴) yields more favorable energies (see Figure 18, Table 6).

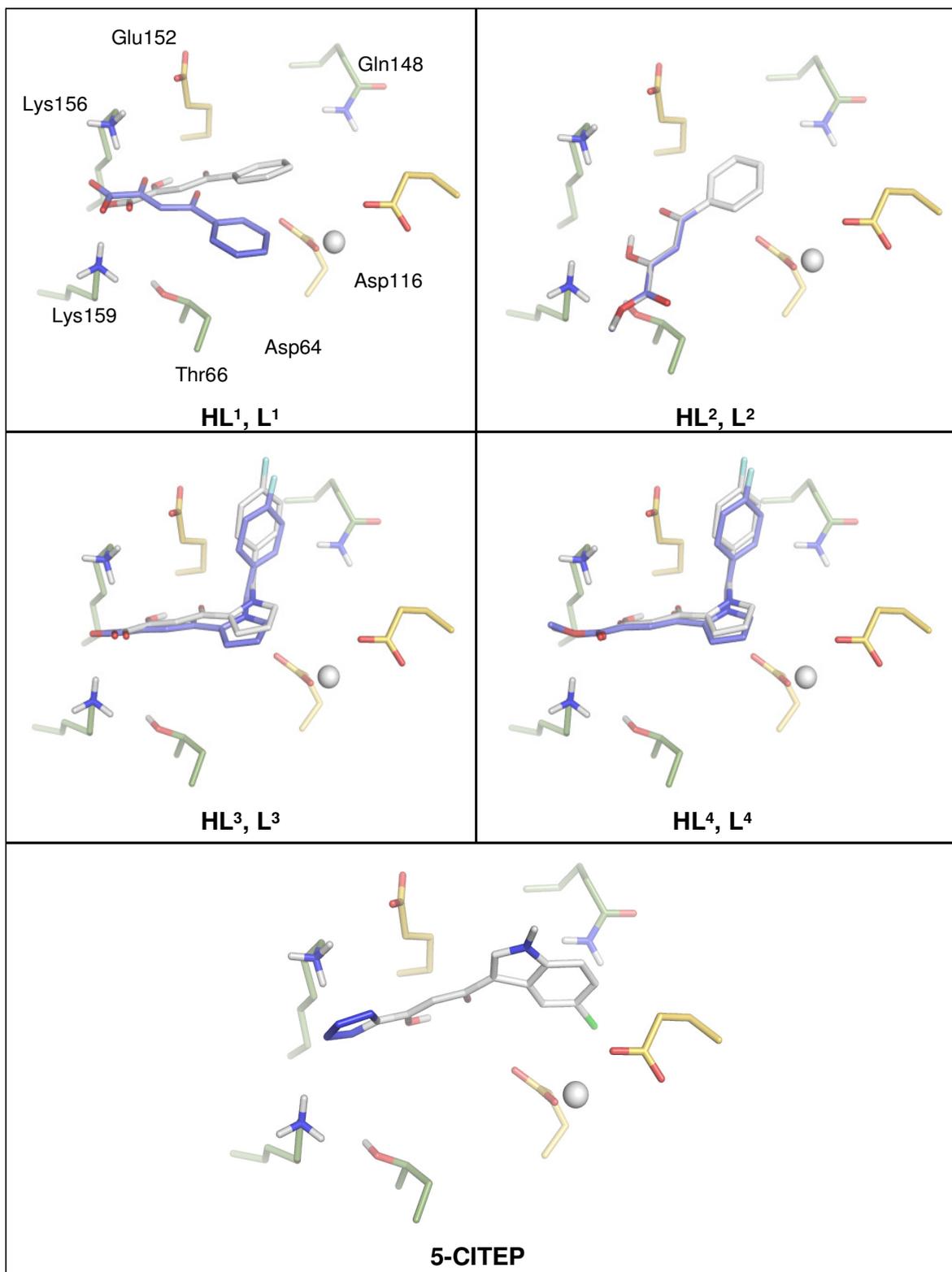


Figure 18: Docked binding modes of the free ligands, in both enol form (HL¹ - HL⁴, gray) and enolate form (L¹ - L⁴, blue). The catalytic triad Asp64, Asp116, and Glu152 is shown in yellow sticks, the metal cofactor is shown in gray spheres. For reference purposes, the docked binding mode of 5-CITEP is also shown.

Good clustering was observed in nearly all cases, only the docking runs of **HL**⁴ did not cluster very well. As a reference, 5-CITEP was docked and showed a similar orientation of the keto-enol group. A per-atom energy (i.e. the intermolecular energy divided by the number of heavy atoms and polar hydrogens) was calculated to allow for a better comparison between monomeric ligands and dimeric complexes. To get an impression of the active site metal ion's contribution to the interaction energies, these contributions were subtracted from the total energies: The active site metal ion contributes 5-10 % to the overall interaction energy (see Table 6).

	Number of runs in first cluster	Calc. free energy [kcal/mol]	Intermolecular energy [kcal/mol]	Per-atom energy [kcal/mol]
HL ¹	50	-6.06 (-5.89)	-6.06 (-5.89)	-0.40 (-0.39)
HL ²	50	-5.25 (-5.23)	-5.56 (-5.54)	-0.35 (-0.35)
HL ³	50	-7.32 (-7.18)	-7.94 (-7.81)	-0.36 (-0.36)
HL ⁴	12	-5.97 (-5.80)	-6.91 (-6.74)	-0.30 (-0.29)
L ¹	47	-6.52 (-6.15)	-6.52 (-6.15)	-0.46 (-0.44)
L ²	50	-6.27 (-5.63)	-6.27 (-5.94)	-0.41 (-0.40)
L ³	50	-7.85 (-7.50)	-8.47 (-8.12)	-0.40 (-0.39)
L ⁴	31	-6.40 (-6.17)	-7.33 (-7.10)	-0.33 (-0.32)

Table 6: Docking results: Clustering and energy data of free ligands. Intermolecular energy refers to the sum of electrostatic and van-der-Waals contributions of protein-ligand interactions. The calculated free energy refers to the sum of the intermolecular energy plus a rotatable bond penalty. For the values in parentheses, the active site metal ion's contribution to the overall interaction energies has been subtracted.

Results – Complexes docked to protein without metal cofactor

The docking runs clustered mainly well, exceptions were **11a** and **12a_bis**, which have more rotatable bonds than the other ligands. See Table 7 for energy and clustering data.

To dissect the interaction energies of the complex metal ions, their energetic contributions were subtracted from the overall interaction energies: Clearly, the metal ions are essential for the protein-complex interactions; they contribute between 30 and 90 % (the latter in complex **13**) to the overall interaction energy.

	Number of runs in first cluster	Calc. free energy [kcal/mol]	Intermolecular energy [kcal/mol]	Per-atom energy [kcal/mol]
9a_aa	36	-6.27 (-1.66)	-6.27 (-1.66)	-0.45 (-0.12)
9a_hc	50	-7.45 (-4.67)	-7.45 (-4.67)	-0.50 (-0.33)
13a	45	-9.34 (-0.85)	-9.34 (-0.85)	-0.58 (-0.06)
10a_mono	19	-7.11 (-3.92)	-7.42 (-4.23)	-0.46 (-0.28)
10a_bis	37	-6.20 (-4.56)	-6.82 (-5.18)	-0.22 (-0.17)
11a	10	-8.34 (-4.15)	-9.59 (-5.40)	-0.22 (-0.13)
12a_mono	29	-8.04 (-4.36)	-8.98 (-5.29)	-0.39 (-0.24)
12a_bis	12	-7.37 (-4.69)	-9.23 (-6.83)	-0.21 (-0.16)

Table 7: Docking results: Clustering and energy data of the Mg²⁺ complexes docked to the protein without its metal cofactor. For the values in parentheses, the energetic contributions from the complex metal ions to the overall interaction energies have been omitted.

Figure 19 shows binding orientations of complexes **9a_hc**, **13a**, and **11a**, where the ion is coordinated by the catalytic dyad Asp64 and Asp116. The binding mode of **9a_aa** shows favorable electrostatic interactions between the carboxy group and Lys156. **12a_mono** and **12a_bis** act as hydrogen bond acceptor with Gln148. Due to their size, the dimeric complexes **10a_bis**, **11a** and **12a_bis** do not fit very well in the active site. The binding modes of the complexes and their parent free ligands differ in all cases.

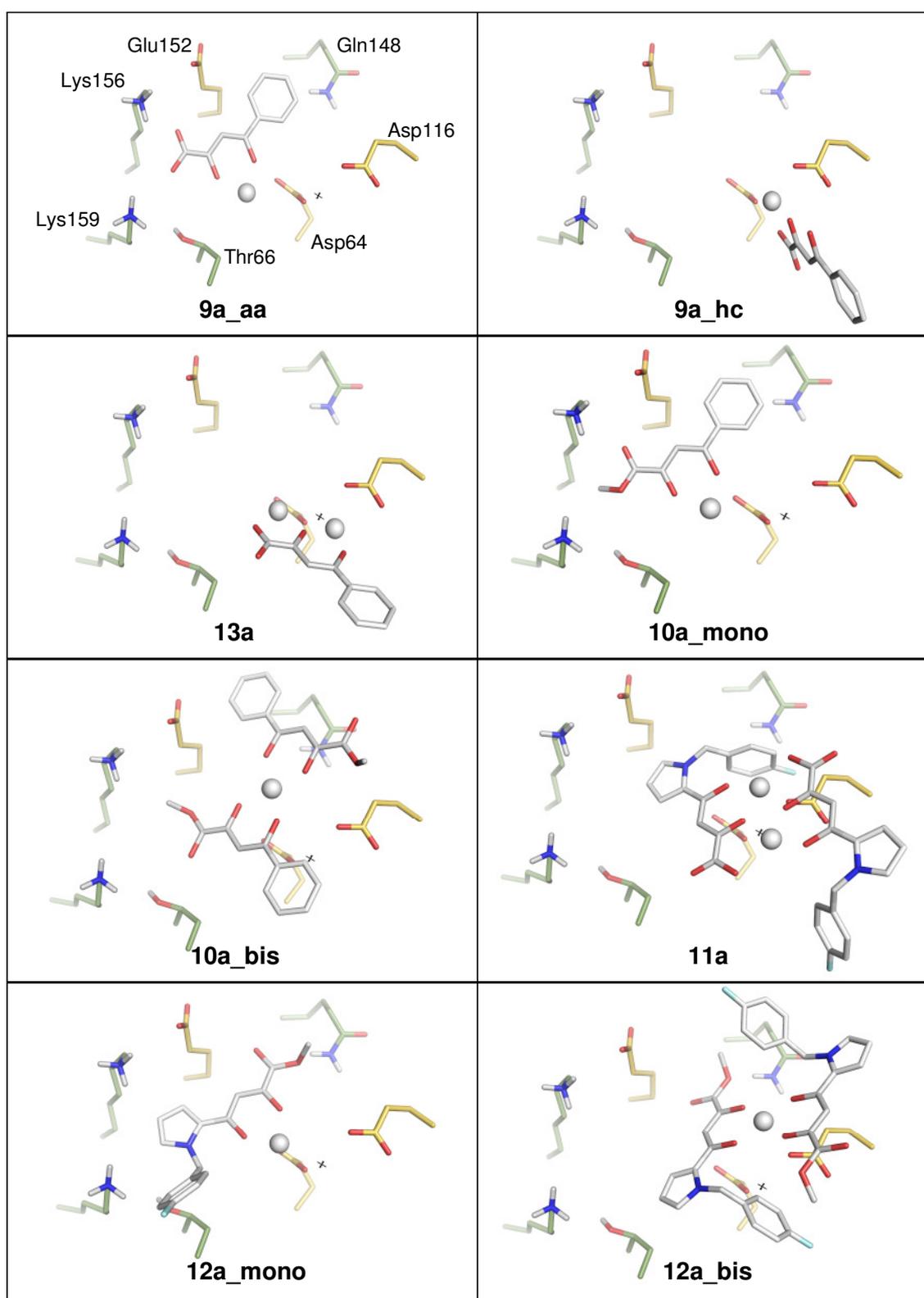


Figure 19: Binding modes of Mg^{2+} complexes, docked to IN active site where the metal cofactor had been removed (original position marked by the black cross).

Results – Complexes docked to protein with metal cofactor

The docked binding modes of the complexes (see Figure 20) show that the carboxylate group of **9a_hc** coordinates the active site metal ion, **13a** is likewise. The carboxylate group of **9a_aa** is placed close to Lys156 and Lys159, the ligand orientation resembles the binding mode obtained when docking **9a_aa** to the active site without its metal cofactor (see Figure 19). Docking of **10a_mono** yields a binding mode with favorable interactions between the active site metal ion and the ester and enolate oxygen atoms; the magnesium ion of the ligand is coordinated by Asp64. The distance between these two magnesium ions is 3.2 Å, which is close to the 3.6 Å, observed between the two active site metal ions in the ASV integrase PBD structure 1VSH. The position of **10a_mono** would grant space for the third catalytic (and flexible) residue E152 to fold towards the active site and interact with the ligand. **11a** chelates the active site magnesium ion by one carboxylate group; the second carboxylate group is in vicinity to Lys159. Docking of **12a_mono** shows a quite remarkable and favorable orientation in the binding site: Striking is the fact that the fluorophenyl ring is well buried; the ester group is close to Lys156. The Mg²⁺ of the complex is coordinated by the catalytic residue Asp64. The distance between this ion and the active site magnesium ion is 4.0 Å (cf. binding mode of **10a_mono**). Due to its size, the dimeric complex **12a_bis** can not be well accommodated by the active site.

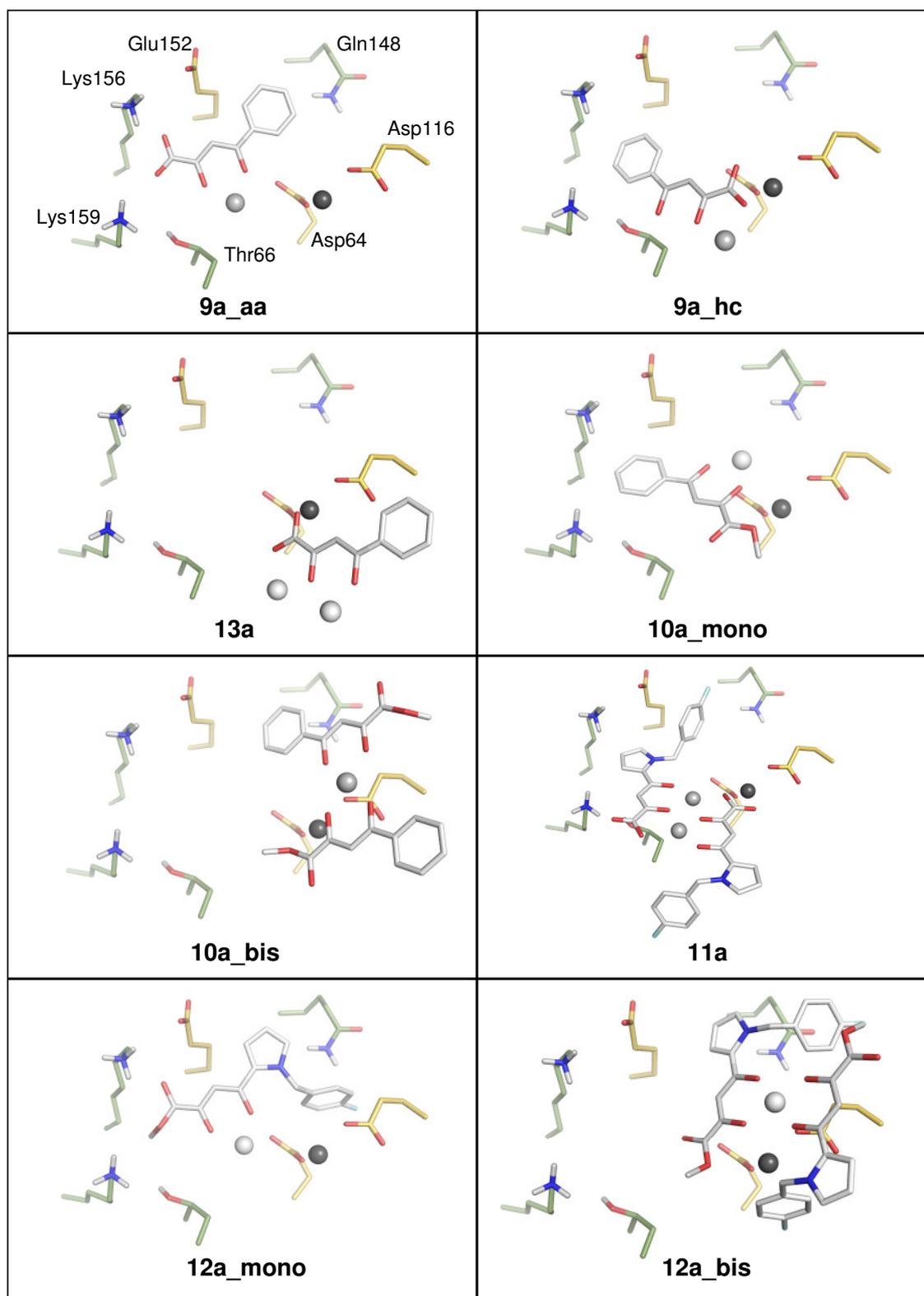


Figure 20: Binding modes of Mg^{2+} complexes, docked to IN active site with its metal cofactor (black sphere).

Poor clustering is observed in the docking runs of compounds with the pyrrole-fluorophenyl functionality (i.e. **11a**, **12a_mono**, and **12a_bis**). In the mentioned cases, there are clusters with more members than the top ranked cluster, but they have significantly less favorable energies and less meaningful binding geometries. Docking energies and clustering data are listed in Table 8. Subtraction of the complex metal ion's energetic contribution shows a significant, yet not dramatic decrease of the intermolecular energy. In the case of dimeric complexes **10a_bis**, **11a**, and **12a_bis**, there is even an increase in intermolecular energy, due to repulsive interactions of the complex metal ions with the protein.

	Number of runs in first cluster	Calc. free energy [kcal/mol]	Intermolecular energy [kcal/mol]	Per-atom energy [kcal/mol]
9a_aa	50	-6.48 (-5.62)	-6.48 (-5.62)	-0.43 (-0.40)
9a_hc	50	-6.68 (-5.76)	-6.68 (-5.76)	-0.45 (-0.41)
13a	50	-6.69 (-3.10)	-6.69 (-3.10)	-0.42 (-0.21)
10a_mono	38	-5.79 (-5.59)	-6.10 (-5.90)	-0.38 (-0.37)
10a_bis	17	-6.46 (-6.75)	-7.09 (-7.38)	-0.23 (-0.25)
11a	6	-8.46 (-8.48)	-9.70 (-9.72)	-0.22 (-0.23)
12a_mono	2	-6.47 (-5.42)	-7.41 (-6.36)	-0.32 (-0.29)
12a_bis	3	-6.86 (-7.53)	-8.72 (-9.39)	-0.19 (-0.21)

Table 8: Docking results: Clustering and energy data of the Mg^{2+} complexes docked to the active site with its metal cofactor. For the values in parentheses, the contributions of the complex metal ion have been subtracted.

Discussion of the obtained results

When modelling the free ligands, the question is whether to keep the enol oxygen protonated or not: On the one hand, the intramolecular hydrogen bond in the keto-enol system is strong, on the other hand deprotonation upon metal complexation is conceivable. Thus, both cases were modelled. As expected, the deprotonated form yielded more favorable energies due to the negative charge, in contrast to the protonated form, where the hydrogen bond was kept fixed and thus did not have much interaction possibilities.

In the first scenario, free ligands were docked to the IN active site containing its metal cofactor, and the respective complexes were docked to the active site missing the cofactor. It was discovered that free ligands and their corresponding complexes in most of the cases do not have similar orientations, as expected, due to their different interaction sites (keto-enol function versus coordinated metal ion). In Table 9, a first comparison of the per-atom energies of the free ligands (both enol and enolate form) and their respective complexes shows an energetic preference of the complex form in the case of **9a_hc**, **13a**, **10a_mono**, and **12a_mono**. **9a_aa** shows a similar per-atom energy, compared to the free ligand (-0.45 kcal/mol and -0.46 kcal/mol, respectively). The dimers **10a_bis**, **11a**, and **12a_bis** have less negative energies, though. This is due to their size and planarity, which prevent a good accommodation by the protein.

	Per-atom energy [kcal/mol]			
	Complex (no metal cofactor)	Complex (metal cofactor)	Free ligand (enol) (metal cofactor)	Free ligand (enolate) (metal cofactor)
9a_aa	-0.45 (-0.12)	-0.43 (-0.40)		
9a_hc	-0.50 (-0.33)	-0.45 (-0.41)	-0.40 (-0.39)	-0.46 (-0.44)
13a	-0.58 (-0.06)	-0.42 (-0.21)		
10a_mono	-0.46 (-0.28)	-0.38 (-0.37)		
10a_bis	-0.22 (-0.17)	-0.23 (-0.25)	-0.35 (-0.35)	-0.41 (-0.40)
11a	-0.22 (-0.13)	-0.22 (-0.23)	-0.36 (-0.36)	-0.40 (-0.39)
12a_mono	-0.39 (-0.24)	-0.32 (-0.29)		
12a_bis	-0.21 (-0.16)	-0.19 (-0.21)	-0.30 (-0.29)	-0.33 (-0.32)

Table 9: Per-atom energies of Mg²⁺ complexes docked to the active site with and without its metal cofactor (values in parentheses: Metal contributions have been subtracted). Also, the per-atom energies of the free ligands (enol and enolate form) docked to the active site with metal cofactor are given.

In the second scenario, complexes were docked to the active site containing its metal cofactor. This docking study resulted in generally less negative energies, when compared to the first docking study. This is due to the occupation (by the active site metal cofactor) of the interaction site near the catalytic residues D64 and D116. Compared to the free ligands' enol forms, complexes **9a_aa**, **9a_hc**, **13a**, **10a_mono**, and **12a_mono** show more favorable energies.

Overall, comparing the per-atom energies of complexes with the free ligands (enol and enolate form), the complexes perform better in case of **9a_hc**, **13a**, **10a_mono**, and **12a_mono**. The dimeric complexes **11a**, **10a_bis**, and **12a_bis** have the least favorable per-atom energy. This is not surprising, given the fact that such bulky flat and rigid molecules can not be accommodated well by the IN active site. It is, therefore, less probable that the parent ligands will bind as dimeric complexes. However, the monomer complexes energetically performed better than the free ligands: From this point of view, the hypothesis of DKA derivatives acting as complexes could be supported.

Visual inspection of the binding modes of **9a_aa**, **10a_mono**, and **12a_mono** is interesting: The positioning of the metal ions resembles the respective ASV integrase (PDB code 1VSH) structural features. ASV integrase is a structure homologue to HIV-1 IN, with the same catalytic triad. The 1VSH structure shows a second metal ion which is not observed in IN structures, but has been reported to be necessary for IN activity. In addition to the favorable positioning of the metal ions, the **10a_mono** binding mode would grant space for the third catalytic (and flexible) residue E152 to fold towards the active site and interact with the ligand.

When discussing the results of the study, it should be mentioned that docking of metal complexes is not a standard procedure for current docking tools. The position of the metal ion within the complex had to be edited manually, also, a simple formal charge was assigned to the metal ion. This procedure might not fully reflect the delicate charge distribution within this system. The approach of quantum mechanics charge calculations was deferred, since AutoDock is parametrized for the use of charges derived from the Gasteiger-Marsili method.

DKA-derived compounds act selectively as inhibitors of the strand transfer reaction. The second metal ion (which is more mobile and therefore not observed in crystallographic structures) has been reported to play an important role during the strand transfer. The absence of this second metal ion in crystal structures led to the hypothesis that the metal ion is bound to the substrate (i.e. DNA). Considering DKAs as competitive inhibitors of this substrate, the hypothesis of DKA derivatives acting as metal complexes, as in the case **9a_aa**, **10a_mono** and **12a_mono**, is in good agreement with this idea.

2.5 Dimerization interface binders – Design of a novel inhibitor class

The last chapter has outlined new insights into the mechanism of action of DKA derivatives. These chemical entities belong to a well-known class of IN inhibitors, which target the enzyme's active site and compete with the DNA substrate for binding. An alternative strategy for inhibition can be derived from the fact that IN exerts its function as a multimer: For the concerted integration of the viral DNA into the host DNA, an IN tetramer is required, whereas IN dimers are sufficient for the 3'-processing reaction. Higher-order oligomers are formed in the pre-integration complexes. Biasing the equilibrium between the different states of multimerization has been shown to result in inhibition⁶⁶. Accordingly, it is conceivable to develop inhibitors which do not address the active site, but rather bias IN multimerization.

2.5.1 Protein-protein interfaces: An emerging class of therapeutic targets

In biological systems, processes are often carried out by multiprotein complexes. Therefore, protein-protein interactions are an emerging valuable class of potential therapeutic targets^{86,87}. Several modulators of protein-protein interactions are known: Therapeutic antibodies have the distinction of having a high specificity to their target, peptides can interrupt protein multimerization. However, in both cases, the poor metabolic stability as well as the low bioavailability are major drawbacks. Far better drug candidates are small-molecule modulators of protein-protein interactions. At a first glance it seems virtually impossible to block protein-protein interactions with small-molecule inhibitors, given the large interaction surfaces of protein interfaces. In addition, such interfaces are generally flat and do not have distinct pockets observed in proteins binding to small molecules. A more optimistic view arises from studies of protein-protein versus protein-peptide complexes: Energetic data suggest that both types of complexes show similar thermodynamic properties and stability, despite significant differences in the interface sizes⁸⁸. Based on these findings, it has been hypothesized that the contributions to binding affinity are not evenly distributed over the entire buried surface of interacting protein-protein interfaces. More likely, the major contributions to binding affinity originate from a smaller interaction interface (so-called functional epitope), which has a similar size among protein-protein and protein-peptide complexes. Additionally, alanine scanning studies of these macromolecular complexes have revealed that only a small subset of amino acids within the overall interface shows significant contributions to the binding affinity^{89,90}. For the design of small-molecule inhibitors of protein-protein interactions, these facts imply that it might not be

necessary for the small molecule to cover the entire protein-protein interface (which is not feasible, given its size), but to address the much smaller functional epitopes⁹¹. The validity of these findings has been proven by recent studies, which have demonstrated the feasibility of blocking protein-protein interactions with small-molecule compounds (see Table 10).

	Molecular mass [Da]	Function	PDB code	Affinity [nM]
Bcl-X_L				
peptidic binding partner	3,110	Regulator of apoptosis	2BZW ⁹²	0.6
small-molecule disruptor	813		2YXJ ⁹³	0.6
HDM2				
peptidic binding partner	1,808	Degradates tumor-suppressant p53	1YCR ⁹⁴	600
small-molecule disruptor	581		1RV1 ⁹⁵	90
HPV E2				
peptidic binding partner	24,630	Transcription factor of Human Papilloma Virus	1TUE ⁹⁶	60
small-molecule disruptor	684		1R6N ⁹⁷	6
IL-2				
peptidic binding partner	24,790	Cytokine, key role in graft rejection	1Z92 ⁹⁸	11
small-molecule disruptor	663		1PY2 ⁹⁹	60
TNF				
peptidic binding partner	17,381	Key factor in inflammation processes	1TNF ¹⁰⁰	n.a.
small-molecule disruptor	584		2AZ5 ¹⁰¹	13,000

Table 10: Examples of small-molecule protein-protein interaction disruptors. Beneath the target, the peptidic binding partner is given, followed by the small-molecule disruptor.

Despite the major challenges in addressing protein-protein interfaces, the successful development of the small-molecule binders listed in Table 10 is an incentive to develop IN interface binders, which block the formation of the catalytically active multimers.

In the following chapter, an approach towards the design of such multimerization inhibitors is introduced: First, the lack of structural data is overcome by molecular dynamics simulations; the obtained protein structures then serve as input for structure-based pharmacophore model design. Finally, a virtual screening is performed. These studies can serve as groundwork for the development of a novel IN inhibitor class.

2.5.2 The dimerization interface: Molecular dynamics simulation and docking studies

The main problem for the rational design of IN dimerization inhibitors is the lack of relevant structural data about the targeted monomeric form. All available IN structures are dimeric, i.e., the dimerization interface of the first monomer is covered by the second monomer's interface. To overcome this lack of data, molecular dynamics (MD) simulations of the IN dimerization interface were performed. In contrast to the recent study of Hu *et al.*¹⁰², who used MD simulations to calculate binding energies and to propose binding locations of a large inhibitory peptide at the interface, the focus of the present work is to reveal detailed interaction sites on the IN dimerization interface that can be addressed by small-molecule dimerization disruptors. Hence, the goal of this study is to elucidate the structural features of the interface by means of MD simulations in order to derive valuable information for subsequent structure-based design efforts.

The dimerization interface is formed primarily by the β 3-sheet (Tyr83-Ile89), the α 1-helix (Thr93-Arg107), and the α 5-helix (His171-Lys186). Most important for the formation of the interaction surface are the residues which point towards the surface with their side chains. These are: Tyr83, Glu85, and Glu87 of the β 3-sheet; Glu96, Tyr99, Phe100, Lys103, and Arg107 of the α 1-helix; Lys173, Gln177, Val180, and Asn184 of the α 5-helix (see Figure 21).

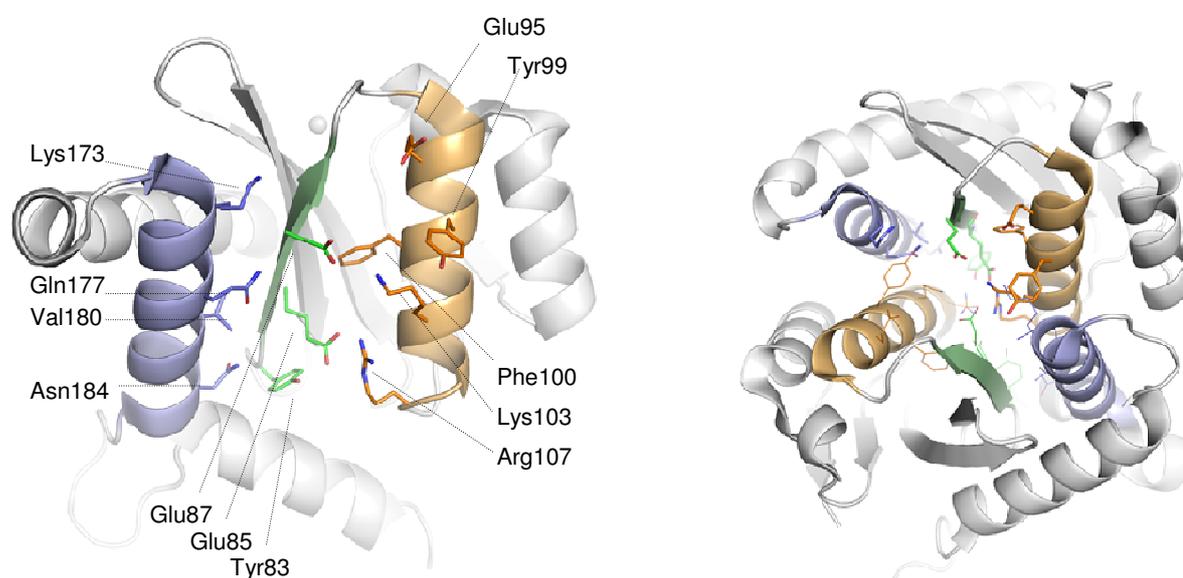


Figure 21: IN dimerisation interface, comprising the α 5-helix (blue), the β 3-sheet (green), and the α 1-helix (wheat). Left: Secondary structure elements and residues forming the dimer interface, which is located on the opposite site of the catalytic center (metal ion shown as gray sphere). Right: Symmetrical IN dimer, where the α 1-helix of one monomer (top) interacts with the α 5-helix of the other monomer (bottom) and vice versa.

The MD simulation study follows the rationale that the interface conformation as seen in the (static) IN dimer crystal structure does not necessarily correspond to the conformations shown by the solvated monomer. A small molecule binding to the monomer will primarily “see” these monomer conformations and tend to stabilize them, whereas it is rather unlikely that it may be able to stabilize the same conformation as the large protein binding partner in the dimer (unless the interface is intrinsically very rigid). Accordingly, structural dynamics of the interface region in the monomer by means of MD simulations of the monomer (as well as of the dimer, for reference purposes) were performed. Subsequently, trajectory snapshots were selected as input structures for hot-spot analyses and docking studies of a hexameric inhibitory peptide binding to the dimerization interface. The hot-spots and the suggested binding mode of the peptide indicate crucial protein-ligand interactions, which serve to derive guidelines for the design of small-molecule dimerization inhibitors as a new class of IN inhibitors.

MD simulation study – Methods

The structure of the IN catalytic-core-domain monomer used for the MD simulations corresponds to subunit A of PDB file 1QS4, which was set up as in previous docking studies. Five crystallographic water molecules (1035, 1229, 1234, 1263, and 1340) were retained due to their important role in stabilizing a cavity at the dimer interface (further outlined in the results and discussion section). Hydrogens were added to the structure using the *tleap* module of AMBER9¹⁰³ after removing the ligand 5CITEP from the structure but retaining the Mg²⁺ ion observed in the active site. Despite its role in the mechanism of the catalytic reaction, a second Mg²⁺ ion was not used in the simulation. As pointed out by Bacchi *et al.*^{Fehler! Textmarke nicht definiert.}, only one Mg²⁺ ion has been detected in the active site in crystallographic studies, and the presence of DNA substrate is believed to be critical for the binding of the second metal ion. Accordingly, substrate-free IN (as simulated in the present study) should be modeled with only one metal ion bound to its active site.

Using the *tleap* module, parameters of the Amber 2003 force field¹⁰⁴ were assigned. A short energy minimization of 200 steps was performed using a generalized Born implicit solvent model^{105,106}. The minimized protein structure was again loaded into *tleap* and solvated in a rectangular water box with a minimum distance of 8 Å between solute and box wall using pre-equilibrated TIP3P water boxes¹⁰⁷. This gave a 65·65·50 Å³ simulation box with 6580 water molecules. To neutralize the protein charge, two chloride counterions were added using the *addions* command of *tleap*.

MD simulations were performed using the SANDER module of AMBER9 with the Amber 2003 force field. The minimized and solvated protein structure was gradually heated to the simulation temperature of 300 K in a two-step procedure. The first step was to disorder the solvent molecules in a constant-volume simulation, keeping the protein position fixed by applying the *ibly* option. The temperature was increased from 100 K to 300 K over 20 ps and then decreased from 300 K back to 100 K over 5 ps, employing the Berendsen weak coupling algorithm with a time constant of 0.5 ps. In the second step, the whole system (i.e. solute and solvent) was allowed to move and was heated to 300 K over 25 ps. To allow a time step of 0.002 ps, bonds involving hydrogen atoms were constrained using the SHAKE algorithm throughout the MD simulation. An 8 Å cutoff was used for the van der Waals interactions, while the electrostatics were treated by the particle mesh Ewald method¹⁰⁸. After the heatup procedure, the constant-volume periodic boundary conditions were switched to constant-pressure; isotropic position scaling was employed, with a default reference pressure of 1 bar and a pressure relaxation time of 1.0 ps. The time constant of the heat bath coupling was switched to 1.0 ps. After a 300 ps equilibration phase, MD trajectory data were collected over 13 ns for the monomer simulation. The dimer simulation carried out for comparative purposes was sampled over 6 ns. The coordinates were saved at intervals of 1 ps. Postprocessing of the trajectory was done using the *ptraj* module.

MD simulation study – Results and Discussion

To probe the conformation space accessible to the dimer interface (see Figure 21), an MD simulation of the IN catalytic-core domain monomer was carried out. As a measure of overall stability, the root-mean-square deviation (RMSD) of the C α atoms was followed. Figure 22 shows that the protein fold was fairly stable during the simulation, with a mean C α RMSD of 1.80 ± 0.29 Å (averaged over the entire trajectory). Nevertheless, significant structural changes are apparent from this diagram. The highest per-residue RMSD values (averaged over the trajectory, see Figure 23) were observed for the flexible loop close to the active site and the terminal loop near the dimer interface, as well as for the C-terminal residues. Excluding the residues 140-150 and 186-196 of these exposed loop regions from the RMSD calculation, a trajectory average of 1.42 ± 0.29 Å was obtained (see Figure 22). The interface residues (as defined above) show low to moderate average RMSD values, and the comparison between the all-atom data (top diagram) and the values for the C α atoms (bottom diagram) in Figure 23 suggests that the structural changes are mostly restricted to the side-chain level.

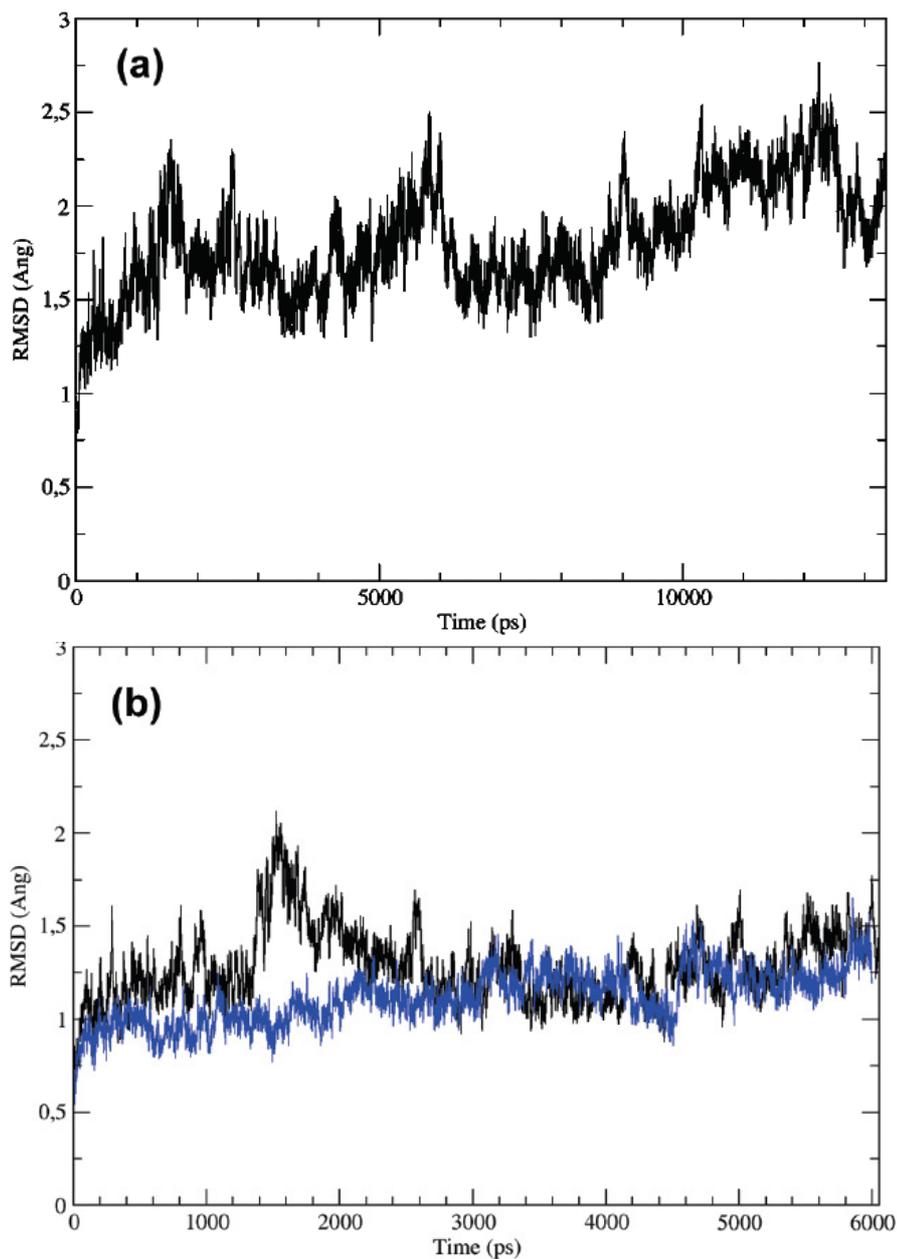


Figure 22: (a) Average RMSD of the C α atoms from the starting structure in the IN catalytic-core monomer MD as a function of simulation time. (b) Comparison of the average RMSD from the starting structure in the MD simulations of the isolated monomer (black line) and the intact dimer (blue line). In both cases, the RMSD values were calculated for the C α atoms of one monomer and excluding the residues of the two exposed loop regions 140-150 and 186-196.

Because tracing possible conformational changes at the dimer interface was of pivotal interest, a 2D-RMSD plot was recorded, covering solely the dimer interface region as defined by the surface-exposed residues given above. For the 2D-RMSD calculation, every twentieth frame (snapshot) of the trajectory was used, corresponding to an interval of 20 ps. The pairwise RMSD value between all 650 frames was calculated after superimposing the aforementioned dimer

interface residues. The resulting plot (Figure 24) shows RMS deviations of up to 3.08 Å. One large cluster covering major parts of the simulation and two smaller clusters at the start and the end of the simulation are immediately apparent. For further studies, representatives of distinct clusters and subclusters were chosen, covering especially the beginning of the simulation to follow the most immediate structural changes upon relaxation of the interface in aqueous solution. After detailed visual inspection of the clusters and subclusters within the 2D-RMSD plot, the following eleven frames were selected: 25, 78, 100, 113, 132, 200, 330, 460, 530, 560 and 610, corresponding to simulation times of 500, 1560, 2000, 2260, 2640, 4000, 6600, 9200, 10600, 11200, and 12200 ps, respectively.

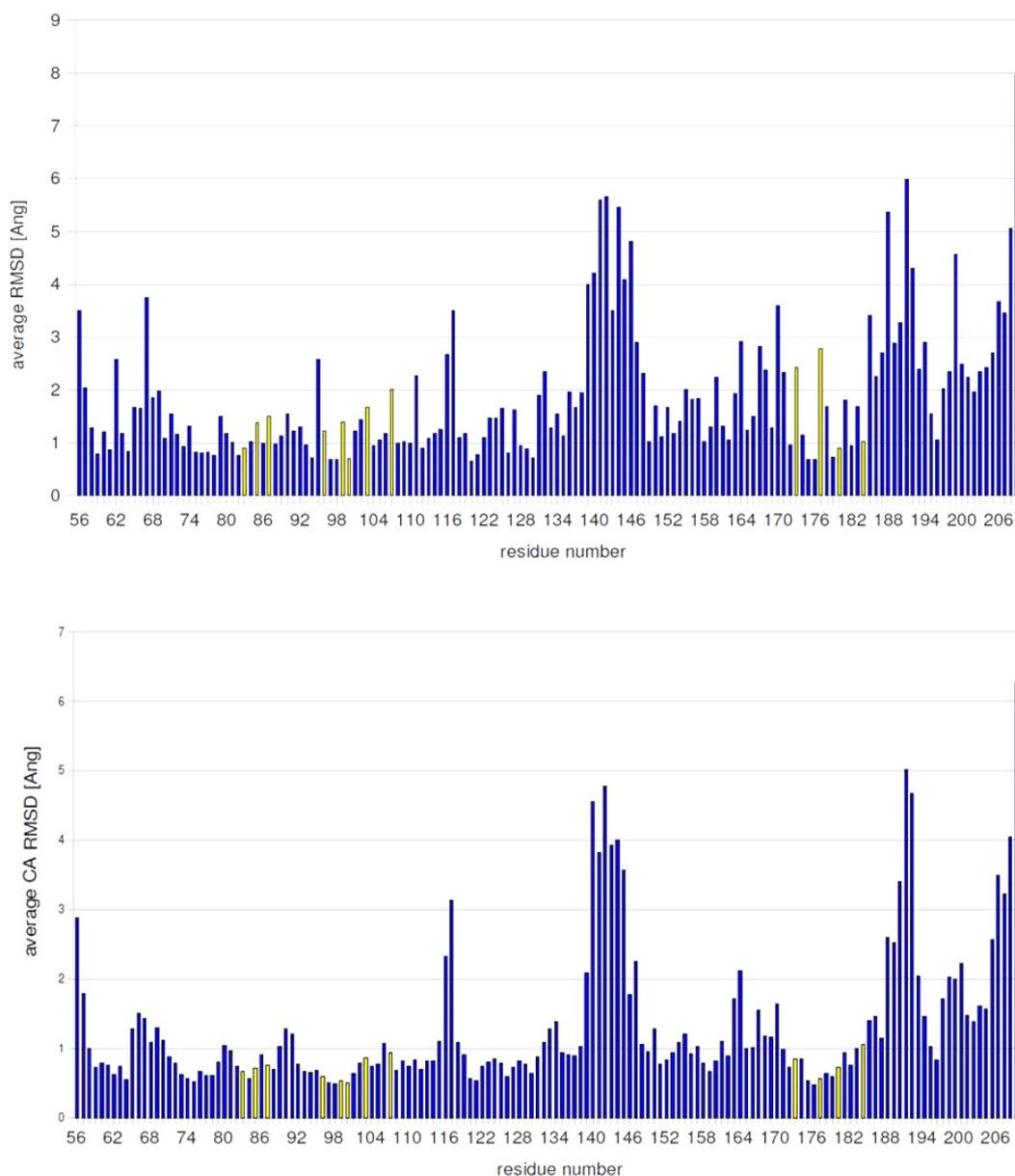


Figure 23: Per-residue RMSD values in the IN catalytic-core monomer MD. Top: Fit and RMSD calculation based on all atoms. Bottom: Fit and RMSD calculation based on C α atoms. The interface residues are highlighted in yellow.

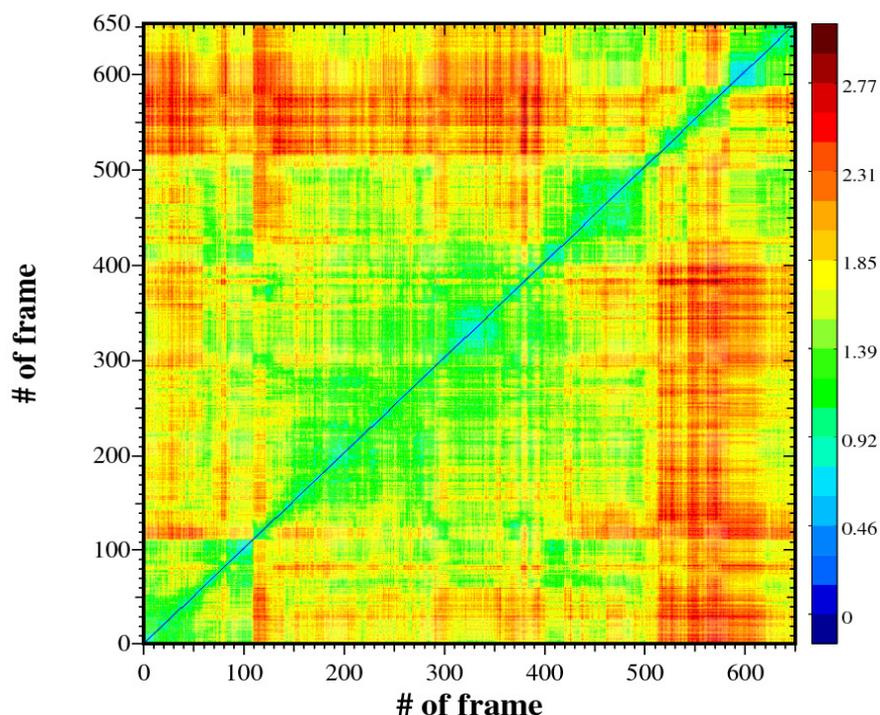


Figure 24: 2D-RMSD plot for the 13 ns monomer simulation, based on a superposition of all dimer interface atoms (as defined in the text) of snapshots taken every 20 ps. Each square in the plot holds the mutual RMSD value (in Å) between two snapshots, represented with the color code shown in the right column. The maximum value observed in the plot corresponds to 3.08 Å.

First, the secondary structure at the dimer interface was analyzed. Figure 25 shows a structural comparison of the eleven selected snapshots. While no dramatic changes in the arrangement of the helices are observed, the β 3-sheet shows a slight drift towards the α 1-helix. The latter shows fluctuations in its helical content (cf. states at 9200 ps and 12200 ps), as does the α 6-helix to a greater extent (the α 6-helix corresponds to the uncolored helix at the southern end of the interface). Interestingly, an alteration of secondary structure elements at the dimer interface has also been observed by Circular Dichroism measurements. Of course, the time range of feasible MD simulations does not allow one to follow major structural changes (which happen on a microsecond time scale). Nevertheless, correct tendencies can be observed. In this case, the fluctuations in the helical content of the two helices show that they are most likely liable to major structural changes (as observed in the Circular Dichroism spectra).

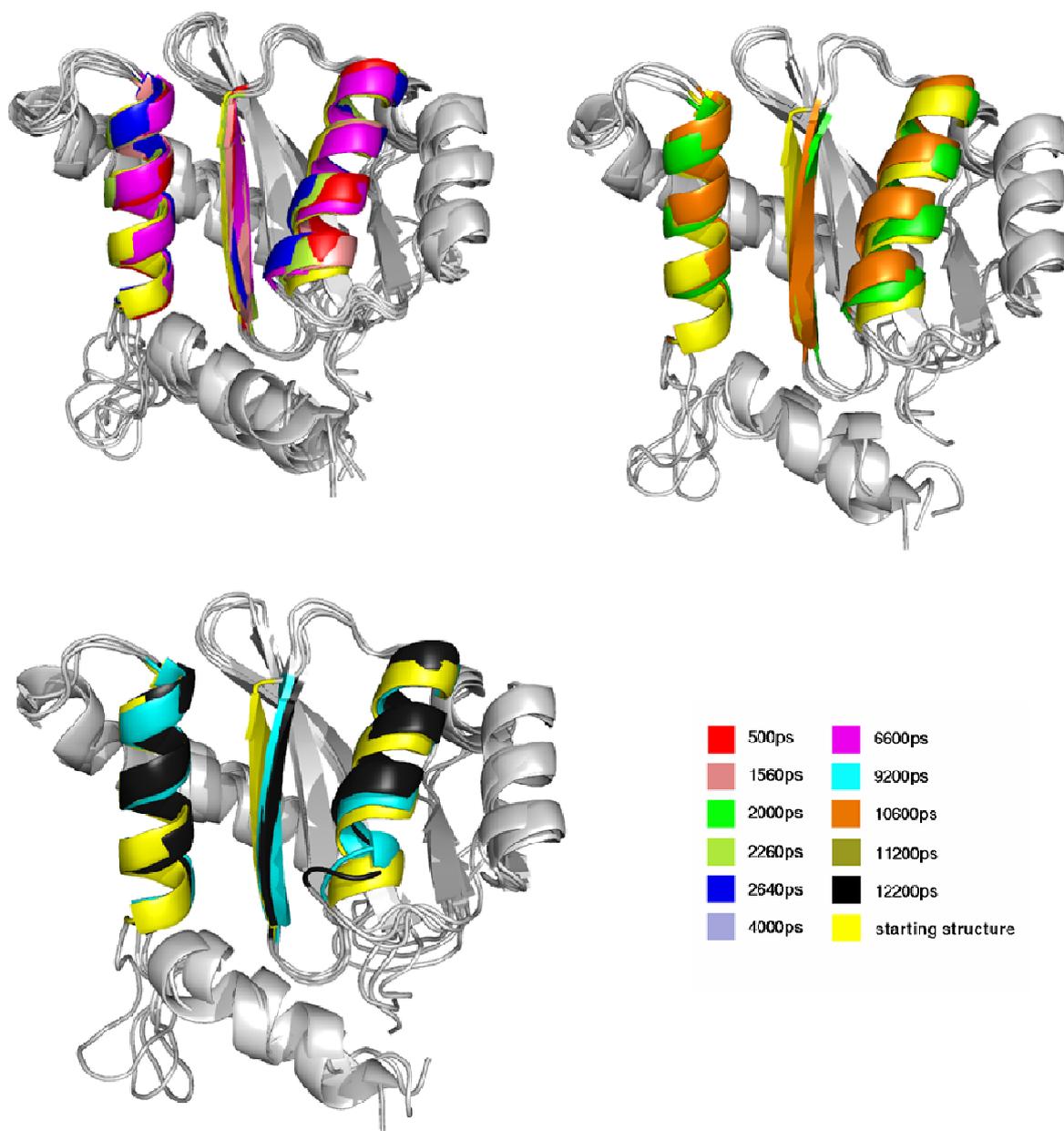


Figure 25: Secondary structure overview of the dimer interface. The colored areas correspond to α 1-helix (right), α 5-helix (left) and β 3-sheet (middle). Upper left: Snapshots at $t = 500, 1560, 2260, 2640, 4000, 6600,$ and 11200 ps; upper right: Snapshots at $t = 2000$ and 10600 ps; lower left: Snapshots at $t = 9200$ and 12200 ps. The starting structure has been superposed for reference purposes.

In the next step, the dynamics of the dimerization interface residues (i.e. Tyr83, Glu85, Glu87, Glu96, Tyr99, Phe100, Lys103, Arg107, Lys173, Gln177, Val180, and Asn184) were analyzed.

Torsion-angle dialplots were recorded for the interface residues. The χ_1 and χ_2 angles of the side chains were measured throughout the trajectory using *ptraj*. Postprocessing of this data to generate the dialplots was done using an in-house script. The results are shown in Figure 26, where the dihedral angle of the corresponding residue in the starting structure is indicated by a red line. Tyr83 keeps its χ_1 and χ_2 starting dihedrals with only minor changes and small fluctuations. For Glu85, χ_1 first oscillates around the initial value, whereas in the last third of the simulation a 90° shift is observed. This χ_1 change coincides with a shift in χ_2 of about 140° . The striking differences between the mean and the initial value for the χ_2 angle of Glu85, Glu87, and Gln177 indicate that the (minimized) starting conformations of these residues are not representative for the monomeric state in aqueous solution. The χ_1 dihedral of Glu96 smoothly oscillates around the starting value, whereas the χ_2 dihedral shows a second distinct side-chain orientation in the last quarter of the simulation, characterized by a 160° shift. A similar behavior is observed for the the Tyr99 χ_1 dihedral. Phe100 shows relatively weak changes. Lys103 has χ_1 and χ_2 values close to the starting value, intermitted by temporary shifts. In Arg107, Lys173, and Gln177, χ_1 shows interesting oscillations of $\sim 100^\circ$ during the simulation. A similar observation is made for the Lys173 χ_2 dihedral. Val180 and Asn184 do not show any significant changes with respect to the starting orientation of their side chains.

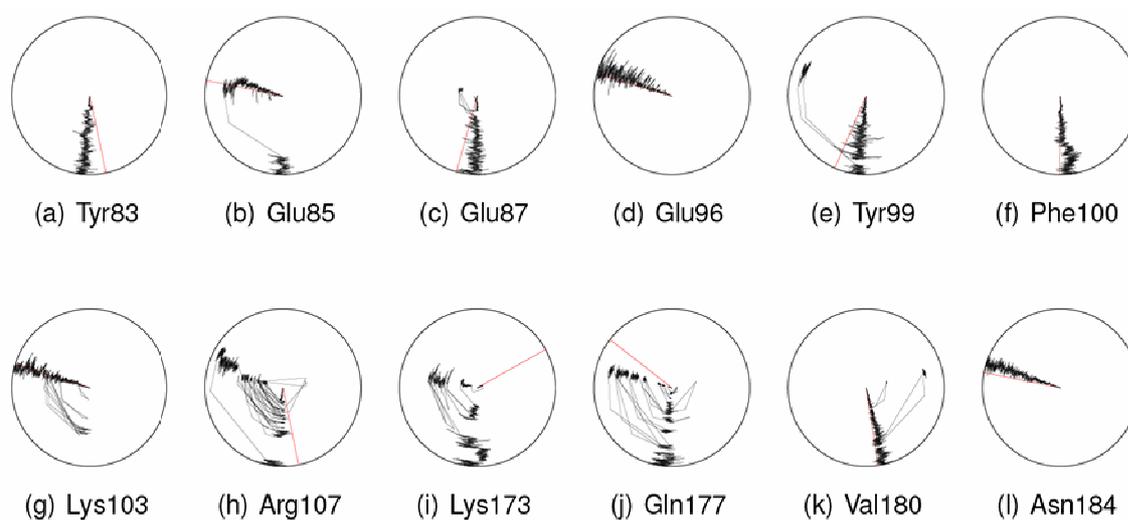


Figure 26a: Dialsplots of the interface-residue torsion angles. The side-chain angle χ_1 was measured. Values in the starting structure are indicated by red lines. The plots show the time-dependent change of the dihedrals, with the center of the circle corresponding to $t = 0$ ps.

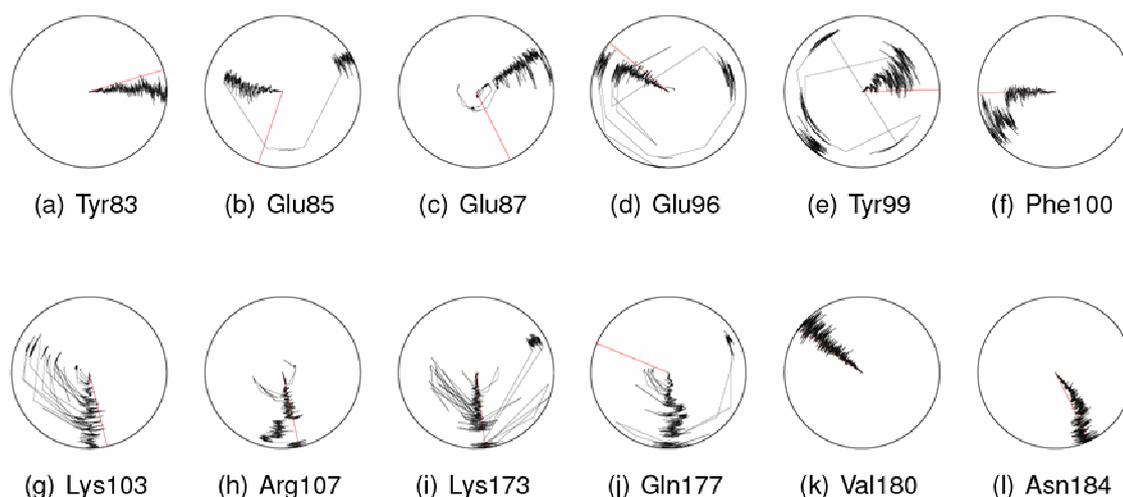


Figure 26b: Dialsplots of the interface-residue torsion angles. The side-chain angle χ_2 was measured. Values in the starting structure are indicated by red lines. The plots show the time-dependent change of the dihedrals, with the center of the circle corresponding to $t = 0$ ps.

Figure 27 (A) shows the side-chain orientations of the residues found to be less flexible during the MD simulation, in particular Tyr83 (located on the $\beta 3$ -sheet), Phe100 (located on the $\alpha 1$ -helix), and Val180 as well as Asn184 (both located on the $\alpha 5$ -helix). These residues are rather buried and less solvent exposed, which explains the lower degree of flexibility. Figure 27 (B) shows the side-chain orientations of the residues found to be moderately flexible during the MD simulation, i.e. Glu85 and Glu87 (both located on the $\beta 3$ -sheet), as well as Glu96 and Tyr99 (both located on the $\alpha 1$ -helix). Tyr99 oscillates weakly at the beginning of the simulation, whereas a higher degree of mobility is observed in the second half of the trajectory. Glu87 shows moderate flexibility, but changes in the orientation of the terminal carboxy group have a dramatic effect on a cavity formed with Lys103. The higher flexibility of Glu85, Glu87, Glu96, and Tyr99 compared to the previously mentioned interface residues is due to their more direct contact to the surrounding solvent.

Figure 27 (C) shows the side-chain orientations of the residues found to be highly flexible during the MD simulation: Lys173 and Gln177 (both located on the $\alpha 5$ -helix), as well as Lys103 and Arg107 (both located on the $\alpha 1$ -helix). All four residues are highly solvent exposed.

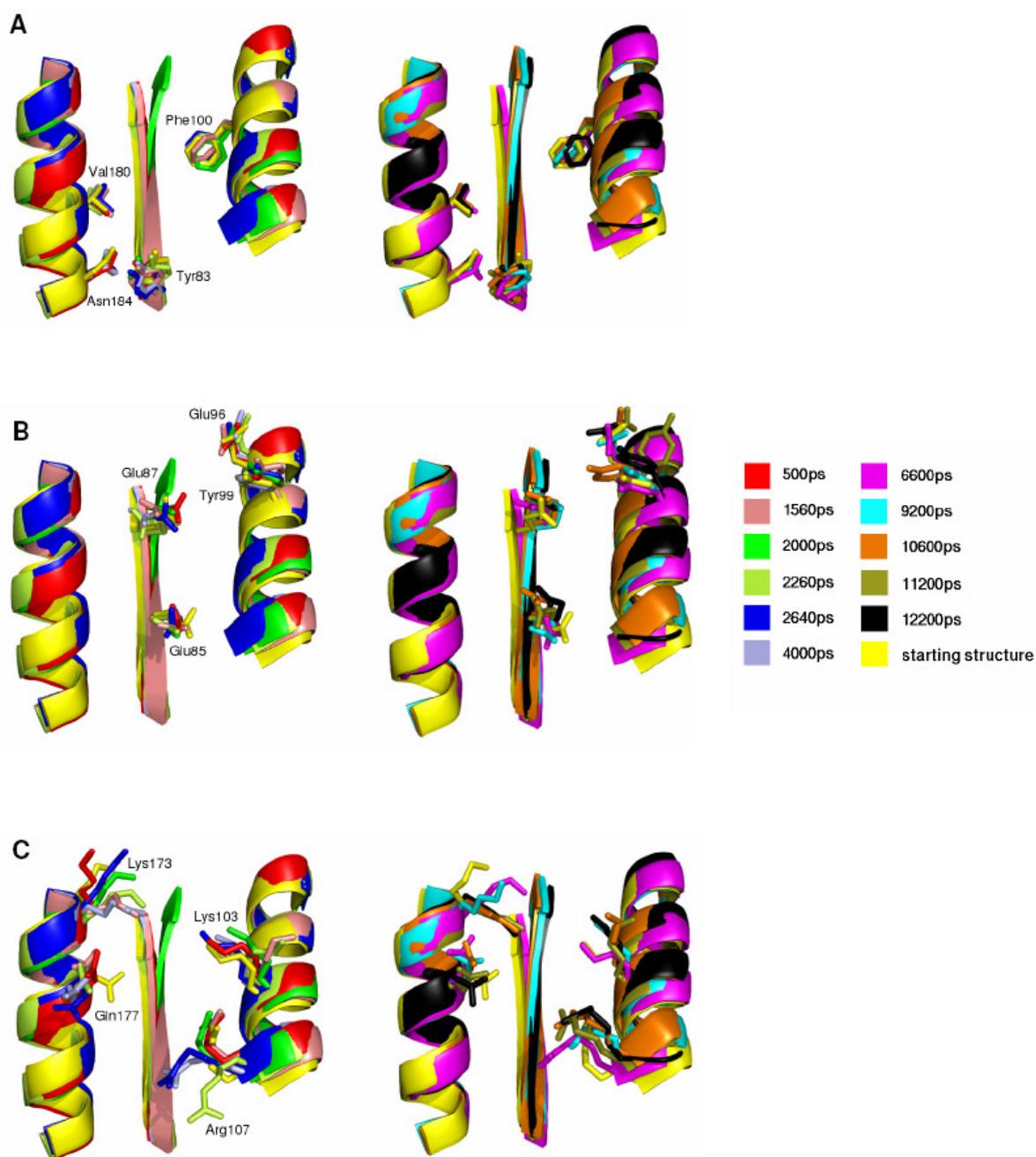


Figure 27: Left: Different structural states observed from the first half of the trajectory (at $t = 500, 1560, 2000, 2260, 2640,$ and 4000 ps, respectively) superimposed with the starting structure. Right: Different structural states from the second half of the trajectory (at $t = 6600, 9200, 10600, 11200,$ and 12200 ps, respectively) superimposed with the starting structure. A (top): Less flexible side chains; B (middle): Moderately flexible side chains; C (bottom): Highly flexible side chains.

The flexibility of Lys103 has dramatic effects on a cavity formed with Glu87. During the simulation, Lys173 displays a tendency to bend towards the groove between the two helices, with the sheet forming the "bottom" of this groove. Bending of Lys173 blocks this groove, best seen in the surface representation of the dimer interface (cf. Figure 28, where the essential residues Glu87, Phe100, Lys103, and Lys173 are colored). Structural changes in this groove region are of pivotal importance, bearing in mind that this is the putative binding site for inhibitors of dimerization (based on analogy with the binding site of the $\alpha 1'$ -helix of the second IN monomer).

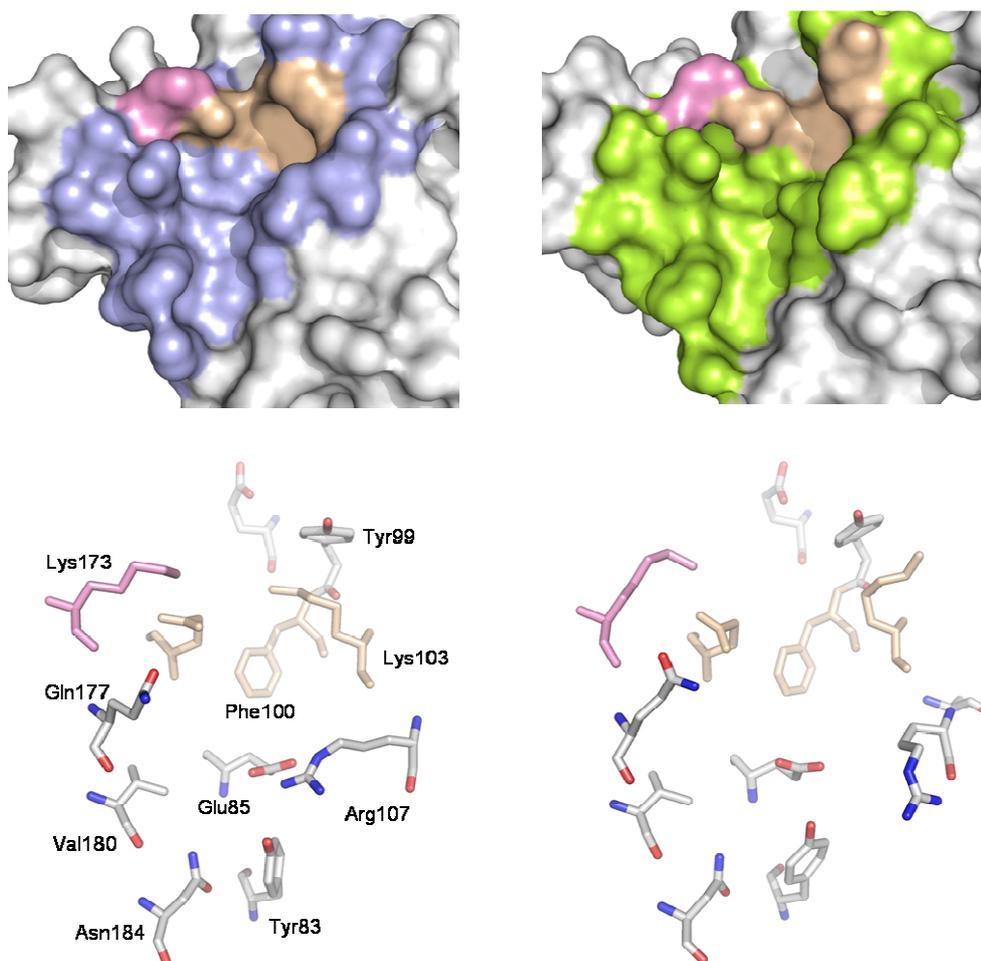


Figure 28: Surface and stick representations of the dimer interface. Colored areas indicate surfaces of the $\alpha 5$ -helix, the $\alpha 1$ -helix, and the $\beta 3$ -sheet. The left column represents the snapshot at 4000 ps with a blocked groove and a formed cavity, whereas the right column represents the snapshot at 2260 ps with a blocked groove and a collapsed cavity.

The surface representation of the different snapshots illustrates the changes in the interface region, especially with respect to the groove between the two helices and the cavity formed by

Glu87 and Lys103. A crucial factor influencing the groove is Lys173. In the starting structure, this residue is directed towards the aqueous surrounding (referred to as "unblocked" state), therefore contributing only marginally to the dimer interface. During the simulation, Lys173 bends towards the groove and narrows it (referred to as "blocked" state). In the trajectory, an oscillation between unblocked (500 ps, 2000 ps, 2640 ps and 9200 ps) and blocked state (1560 ps, 2260 ps, 4000 ps, 6600 ps, 10600 ps, 11200 ps and 12200 ps) can be observed, with the blocked state prevailing at the end of the simulation.

The cavity formed by Glu87 and Lys103 is distinct in the starting structure; during the MD simulation, fluctuations in the shape of the cavity are observed, best seen in the surface representation of the interface region. As a quantitative measure of the cavity structure fluctuations, the distance between the terminal amino group of Lys103 and the terminal carboxy group of Glu87 was recorded and plotted versus simulation time (Figure 29). From this graph, the oscillation in the distance leading to a collapse and recurring formation of the cavity is clearly visible. A distinct cavity (i.e. a relatively close contact between Glu87 and Lys103) is observed in the snapshots at $t = 500$ ps, 2640 ps, 4000ps, 6600 ps, 9200 ps, and 12200 ps, but not in the snapshots at $t = 1560$ ps, 2000 ps, 2260 ps, 10600 ps, and 11200 ps. The distance curve in Figure 29 indicates that formation and collapse of the cavity is an oscillating process, without clear tendency towards stabilization or collapse.

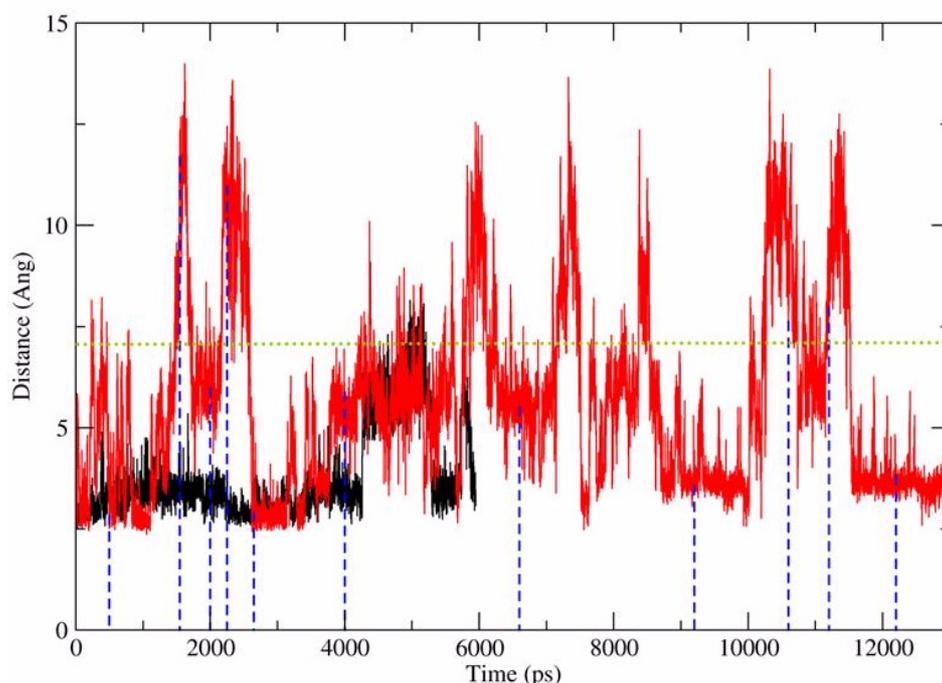


Figure 29: Glu87-Lys103 distance versus simulation time. Red: Monomer MD simulation; black: Dimer MD simulation; dotted blue lines: Selected snapshots. The dotted green line indicates the formation and collapse (below and above, respectively) of the cavity.

For reference purposes, a 6 ns MD simulation of the dimeric IN structure was performed. The protein fold did not show significant structural deviations from the crystal structure, proving that the crystallographic structure is stable in a noncrystalline environment. In line with the expectation, structural deviations and fluctuations were observed mainly in the exposed loop regions mentioned above. Excluding these loop regions, average C α atom RMSD values of 1.23 ± 0.14 Å and 1.12 ± 0.14 Å were observed for the two monomers in the 6 ns dimer simulation (cf. Figure 22). Both monomeric units displayed similar structural behaviour over the trajectory. The first monomer was extracted, analyzed, and used as a reference for the monomer MD simulation. Looking at the Glu87-Lys103 distances obtained in the dimer MD simulation (see Figure 29, black line), the fluctuations are rather slight and that the largest distance (at $t = 5000$ ps) is only roughly one-half of the largest distance obtained in the monomer MD. In contrast to the free monomer, no collapse of the cavity formed by Glu87 and Lys103 is observed in the dimer.

In most of the IN structures, crystallographic water molecules are observed at the dimer interface. Five such molecules are found in the vicinity of Glu87 and Lys103, building a water network between Glu87 and Glu96. In a first attempt, a solute-solvent system was prepared, where all crystallographic water molecules had been removed prior to the solvation process (no direct water mediation between Glu87 and Lys103 via hydrogen bonds is observed in the crystal structure). However, the *tleap* routine of AMBER9 did not place water molecules at the corresponding positions. MD simulations based on this system yielded a collapse of the cavity formed by Glu87 and Lys103 (data not shown). In the actual study, the aforementioned five water molecules have been retained. In contrast to the former MD simulation, a collapsed state of the cavity is not maintained: The water network between Glu87 and Glu96 seems to reduce the mobility of Glu87. Additionally, the water molecules placed close to Glu87 and Lys103 begin to bridge both residues in the course of the MD simulation. Both features lead to the assumption that these water molecules play a crucial role in limiting the movements of Glu87 and Lys103, thereby preventing the cavity from a permanent collapse.

In summary, the most interesting features of the dimerization interface with respect to ligand accommodation appear to be the groove and the cavity. To roughly estimate the ability of the different snapshots to accommodate a ligand, the number of potential contact atoms in the protein capable of interacting with a ligand was counted using the *Site Finder* methodology implemented in MOE: Originally, the purpose of the MOE¹⁰⁹ Site Finder utility is to identify potential binding sites in a receptor by means of a geometric method based on Alpha Shapes. Apart from graphical output for visual analysis, the algorithm provides the number of receptor contact atoms in the putative binding sites and partitions them into hydrophobic or hydrophilic

as well as backbone or side-chain atoms. Using the MD-derived snapshots as input structures, Site Finder was applied to retrieve the number of contact atoms within the dimer interface binding site.

Here, the whole protein structure was scanned, and in each snapshot the dimer interface was revealed as the site with the highest contact-atom count. In Figure 30, the total number of contact atoms in the dimer interface is reported for each snapshot. The lowest number of contact atoms (52) is found for the starting structure. All simulation snapshots show higher values, with the maximum reached at snapshot 4000 ps, corresponding to a total of 137 contact atoms. Accordingly, the protein conformation corresponding to this snapshot would appear to be most suited for accommodating potential ligands.

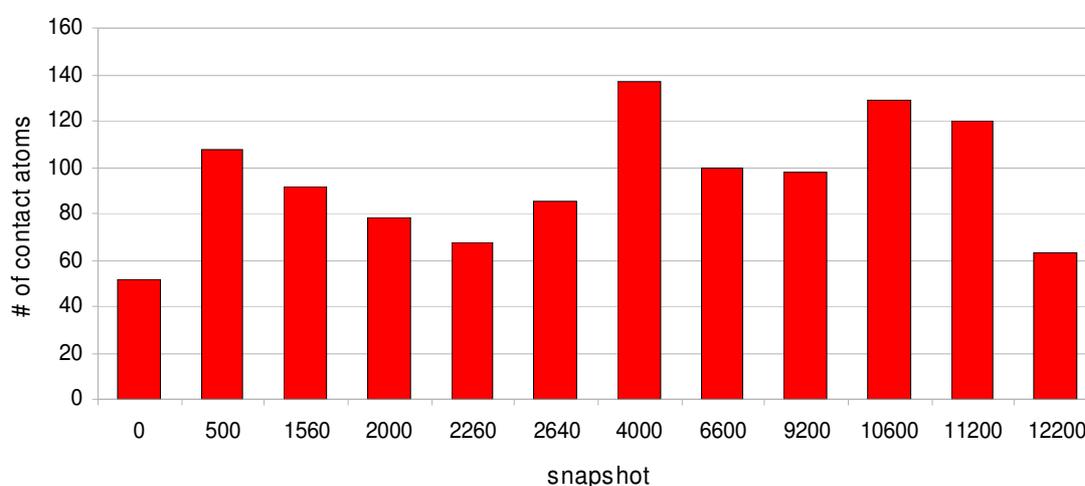


Figure 30: Total number contact atoms at the protein dimer interface, according to the *Site Finder* analysis (see text for further explanations). The snapshot at 4000 ps shows the highest count.

To further confirm this assumption, all extracted snapshots were subjected to a comparative docking analysis using the peptide YFLLKL as ligand: This hexapeptide was derived from a parent inhibitory dodecamer (TAYFLLKLAGRW). This dodecamer has the same amino acid sequence as the IN α 1'-helix (residues 97'-108'), which binds (as part of the second monomer) to the dimer interface of the first IN monomer. It was shown⁶⁶ that this synthetic dodecamer competes for the same binding site, preventing dimerization of IN. Interestingly, mutation studies⁶⁸ showed that the lysine residue can be replaced by an isoleucine residue without affecting the affinity. This seems surprising, given the role of this lysine (Lys103') in salt-bridge formation with Glu87 across the interface. However, it has been hypothesized that this salt bridge contributes only to a minor extent to the overall dimer stability¹⁰. Truncation studies of the dodecamer showed that the minimal active structure is the hexameric peptide YFLLKL⁶⁸. In

analogy to the dodecamer, one can expect that this hexamer competes for the same binding site in the dimer interface region. However, in the dimeric crystal structure, this hexameric portion shows only very few contacts to the first monomer. This leads to the assumption that the inhibitory peptide YFLLKL may still bind to the same site, yet with a different binding mode which allows more favorable interactions. Consequently, docking studies of this peptide to the different MD-derived snapshots were performed to obtain potential binding modes.

Docking of the inhibitory peptide YFLLKL – Methods

For the docking study, GOLD3.2¹¹¹ was used. The helical peptide YFLLKL was constructed using MOE. The backbone was kept rigid in the standard-helix conformation throughout the docking runs, while the side chains were allowed to rotate. All eleven snapshots selected from the MD simulation after the 2D-RMSD analysis were used as protein input structures for docking to identify the most favored conformer for binding the inhibitory peptide. For each protein conformer, 50 independent docking runs were performed. The binding site was defined by a sphere of a 10 Å radius centered at the amide hydrogen of Ala86. The genetic algorithm parameters were set as follows: Population size 100, number of islands 5, selection pressure 1.1, number of operations 150000, niche size 2, migrate 10, mutate 95, and crossover 95. Clustering of the resulting docking poses was done using the *rms_analysis* utility, using a threshold of 1.5 Å. For the determination of preferred interaction sites (hot-spots) within the peptide binding site, the HotSpots function of DrugScore^{CSD} 112,113,114 was employed. Sybyl atom type C.ar served as aromatic probe, C.3 as hydrophobic probe, and N.3 as hydrophilic nitrogen probe. The grid size was set to 97·97·64 points around the dimerization interface center, with a grid spacing of 0.5 Å. The contour plots were normalized for a more intuitive score ranging from 0 to 1 for each probe.

Docking of the inhibitory peptide YFLLKL – Results and Discussion

The results were analyzed in terms of the scoring values and by visual inspection. Comparing the GOLDscore values from docking to different protein conformations, it is striking that the snapshot taken from the central part of the trajectory (i.e. 4000 ps) yielded the most favorable score; moreover, good clustering of the top-ranked results was observed (see Table 11:). This protein conformer is part of the largest cluster in the 2D-RMSD plot and, therefore, most likely corresponds to a relatively stable conformation of the IN monomer. The scores indicate that

YFLLKL will preferably bind to this protein conformation, presumably stabilizing it further. The favorable accommodation of the ligand by the snapshots at 4000 ps is in fair agreement with the expectations based on the *Site Finder* analysis described above.

	number of members in top-ranked cluster	total GOLDScore	h-bond score	hydrophobic score	internal score
start structure	24	13.74	0.60	39.54	-26.40
500 ps	49	-23.53	5.15	-1.71	-26.97
1560 ps	4	11.37	0.00	38.10	-26.73
2000 ps	6	12.95	0.16	13.31	-26.42
2260 ps	43	24.20	5.72	45.18	-26.70
2640 ps	1	16.19	0.00	42.98	-26.79
4000 ps	47	28.34	3.40	53.10	-28.16
6600 ps	8	13.28	3.12	36.30	-26.14
9200 ps	46	-4.63	0.63	23.57	-28.83
10600 ps	5	-4.46	0.00	22.88	-27.34
11200 ps	1	-10.50	0.00	16.44	-26.94
12200 ps	4	6.38	0.00	32.58	-26.20

Table 11: Clustering and scoring data of the docking study analyzing the interaction of YFLLKL with the different MD snapshots. The best total score as well as very good clustering is achieved when taking the snapshot at 4000 ps as input structure (bold).

Figure 31 shows the binding mode of YFLLKL to snapshot 4000 ps. Residue Y1 is well accommodated by the cavity formed by Glu87 and Lys103. The aromatic system displays interactions with the phenyl ring of Phe100. Furthermore, the Tyr-OH group donates a hydrogen bond to the backbone oxygen of Phe100. The aromatic system of F2 is placed in the groove. The cationic residue K5 binds to the smaller cavity formed by Glu87 and Lys173, where it forms a salt bridge with Glu87.

When using AutoDock3.0 as docking tool (data not shown), the aromatic residues of the hexamer were placed towards the aqueous surrounding, which is not quite likely. Furthermore, subjecting the different binding modes obtained from GOLD and AutoDock3.0 to a rescoring

with SFCscore¹¹⁵ (employing both the SFC_290p and SFC_290m scoring schemes) favored the binding modes from GOLD.

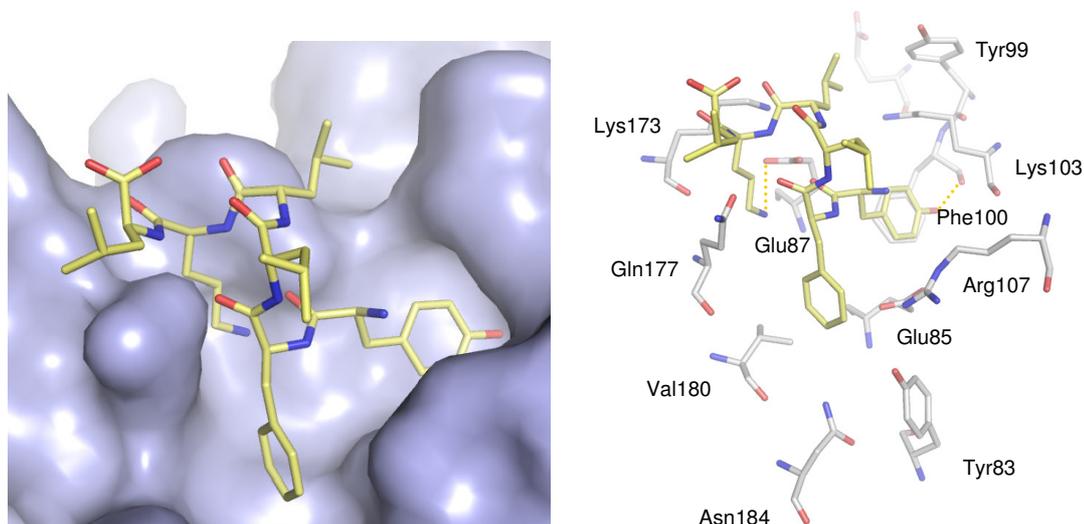


Figure 31: Top-ranked binding mode obtained by docking of the hexameric peptide YFLLKL to the dimerization interface (snapshot at $t = 4000$ ps). Left: Surface representation of the protein. Right: Stick representation of the protein; the hydrogen-bond interactions described in the text are highlighted as dotted lines.

This was further supported by the hot-spot analysis (Figure 32): In the GOLD binding mode, the favorable positions for aromatic probes in the cavity are occupied by Y1 and F2, respectively. Their prominent positions indicate a major contribution to the overall affinity, which is also in accordance with experimental data: Alanine scanning studies of the parent dodecameric peptide TAYFLLKLAGRW show that tyrosine and phenylalanine are essential for inhibitory activity; furthermore, hexameric peptides derived from this dodecamer are inactive when they did not include these two residues.

The hot-spot analysis (see Figure 32) yielded favorable interaction spots within all of the three interface features: The larger cavity formed by Glu87/Lys103 with Phe100 at the bottom displays aromatic and hydrophobic hot-spots, as does the groove. The smaller cavity formed by Glu87/Lys173 appears to have amphiphilic character, as favorable hot-spots for a variety of probes are observed in this area. This is likely due to the presence of hydrophobic “floor” residues in this superficially polar cavity, namely Val88 and Val176, which explain the presence of favorable hydrophobic interaction sites. Hot-spots for a hydrophilic probe (aliphatic amine N)

were also observed in this cavity, showing a similar prevalence as the hydrophobic probe. These polar hot-spots indicate possible interaction sites with the amide oxygen of Gln177 and the carboxy function of Glu87, both located at the rim of the cavity.

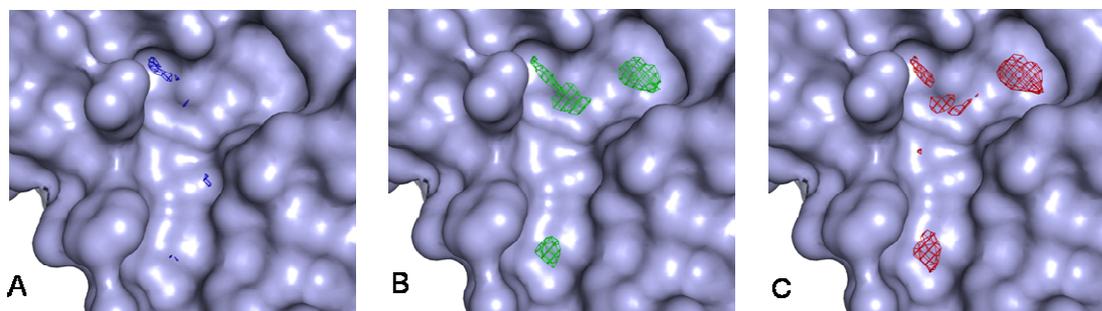


Figure 32: Important subsites of the dimer binding site (snapshot at $t = 4000$ ps): Drugscore^{CSD} contourplots (contour level = 0.70) are shown, with a nitrogen probe (A, blue), an aliphatic probe (B, green) and an aromatic probe (C, red), illustrating the Y-shaped dimerization interface.

Conclusion

An inspection of the dimer interface has so far mostly been ignored in MD simulation studies of HIV-1 IN. In this study, important structural features at the dimer interface could be revealed: A groove was identified which very likely accommodates the inhibitory peptide YFLLKL. This groove became more distinct during the MD simulation; a highly flexible residue, Lys173, was observed to be able to obstruct this groove. Additionally, a cavity adjacent to the groove was identified and analyzed. This cavity, formed by Glu87 and Lys103, was found to be of transient nature, oscillating between formation and collapse. Furthermore, a second, less distinct cavity formed by Glu87 and Lys173 was observed. Groove and cavities appear to be suitable regions for accommodating inhibitors at the dimer interface. Snapshots derived from the MD simulation were used as protein input structures in a docking study with the peptide YFLLKL to reveal its potential binding mode. The docking procedure showed that the peptidic ligand binds to a protein conformation where the groove is blocked by Lys173 and the Glu87/Lys103 cavity is formed. Altogether, the groove and the two cavities form a Y-shaped site. Hot-spot analyses gave information about ligand atom types which favorably interact with the respective subsite. Therefore, this study yielded valuable information for the design of IN dimerization inhibitors.

2.5.3 Dimerization inhibitor design I

In the last chapter, an alternative strategy towards IN inhibition was introduced: IN dimerization interface binders could bias the delicate multimerization equilibrium, thereby preventing the formation of IN tetramers, which are required for catalytic activity. The lack of structural data was overcome by extensive MD simulation studies of the dimer interface. The protein conformation 4000ps was identified as a representative structure of the dimer interface in solution: It is characterized by a Y-shaped site at the dimer interface, which can be exploited for the design of interface binders. Based on these findings, pharmacophore models can be created and employed in a virtual screening for IN dimerization inhibitors, which eventually should be tested in an IN inhibition assay.

Pharmacophore model I – Model creation

For the pharmacophore model creation, MOE was used. For an introduction to MOE pharmacophore modelling, see Chapter 5.1. Based on the DrugScore^{CSD} hot-spot analysis of the interface (see Figure 33), distinct pharmacophore features were set.

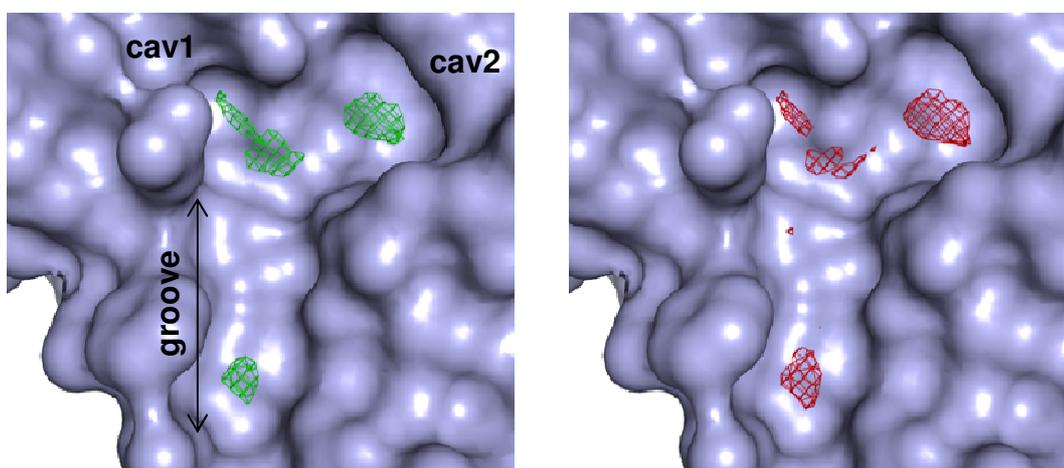


Figure 33: Interface regions cav1 („left“ cavity), cav2 („right“ cavity), and groove. Left: Aliphatic hotspots (atom type C.3, contour level = 0.70). Right: Aromatic hot-spots (atom type C.ar, contour level 0.70).

The aromatic hotspots within cav2 and the groove are in good agreement with the docked binding mode of the YFLLKL, where the aromatic ring systems are placed on the corresponding positions in cav2 and the groove, respectively. Accordingly, aromatic features were set on these

hot-spots. The third position within cav1 was addressed by a hydrophobic feature; smaller aliphatic groups are expected to fit this narrower subpocket better than aromatic ring systems. To maximize hydrophobic interactions between ligand and protein, a good shape-complementarity is required: The Y-shaped binding site should be addressed by a Y-shape ligand. Accordingly, the three-point pharmacophore model was extended by a volume constraint: Along the binding site, exterior volumes were set to dismiss compounds demanding space beyond the Y-shaped binding site, whereas interior volumes were defined (at the same positions) to retrieve only those molecules which are able to fill up the Y-shaped binding site completely. Combining both features lead to a volume constraint which ensured that only Y-shape compounds for an optimal protein-ligand shape complementarity were retrieved. The resulting pharmacophore model I is shown in Figure 34.

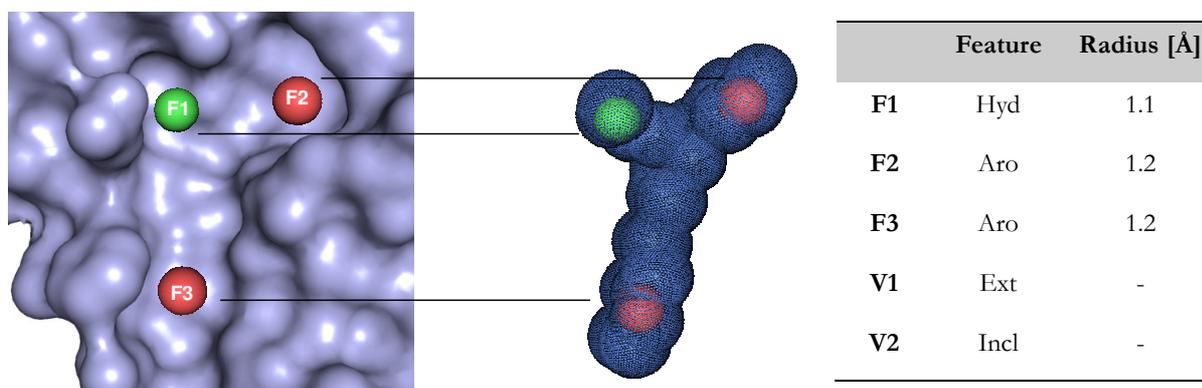


Figure 34: Structure-based pharmacophore model I design. Left: Feature placement within the dimerization interface. Middle: Three-point pharmacophore model with additional Y-shaped volume constraints. Right: Feature and volume description, according to the MOE annotation scheme.

Pharmacophore model I – Virtual screening

Using the created pharmacophore model I as query, the ZINC druglike database¹¹⁶, containing ~1.8 M compounds, was screened in MOE. The initial hit collection was subjected to further filtering steps: First, amide and ester containing compounds were rejected, due to their liability for hydrolysis. Then, only those molecules with an optimal pharmacophore model fit (RMSD < 0.5 Å) were retained. To get rid of undesirably flexible molecules, a rotatable bond count filter ($N_{rotb} \leq 7$) was introduced. Since known peptidic interface binders are not negatively charged, a formal charge filter was introduced to dismiss compounds with negative

formal charge ($F_{\text{charge}} \geq 0$). In the last step, remaining hit molecules were clustered (fingerprint-based clustering method: MACCS structural keys, Tanimoto coefficient metric, similarity = 0.7, overlap = 0.7). Out of every cluster, one representative was retrieved, finally leading to a total of 100 hits (see Figure 35).

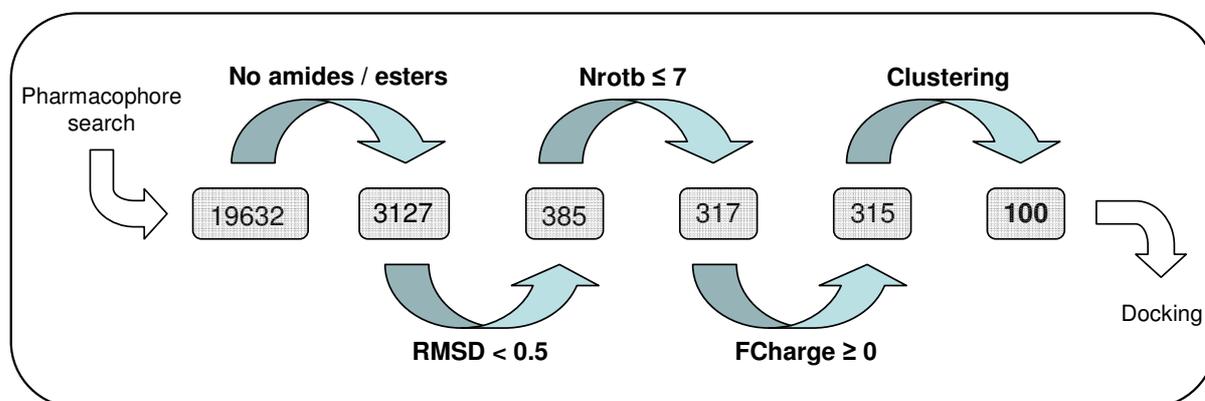
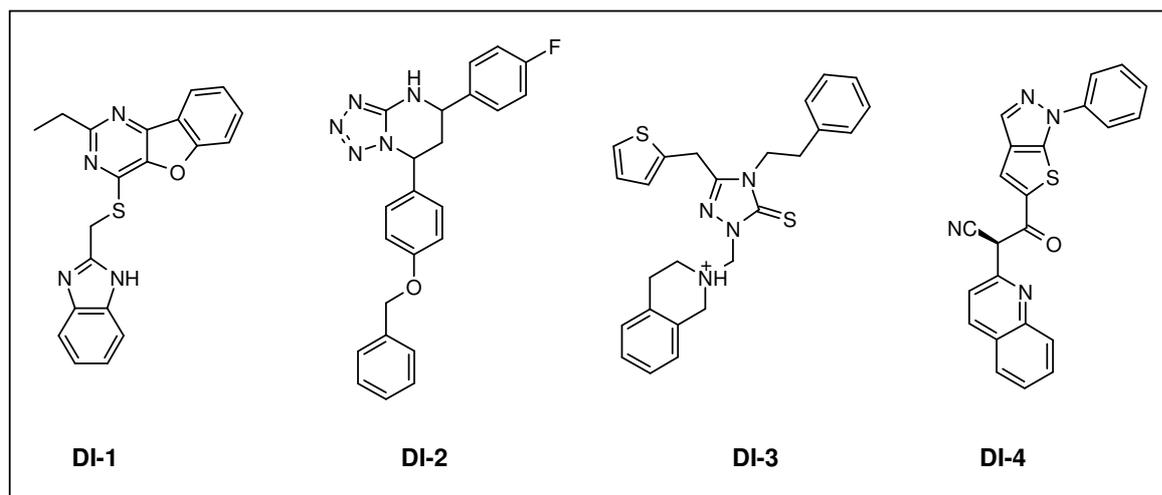


Figure 35: Virtual screening filters. The resulting 100 compounds were subjected to docking.

Pharmacophore model I – Docking

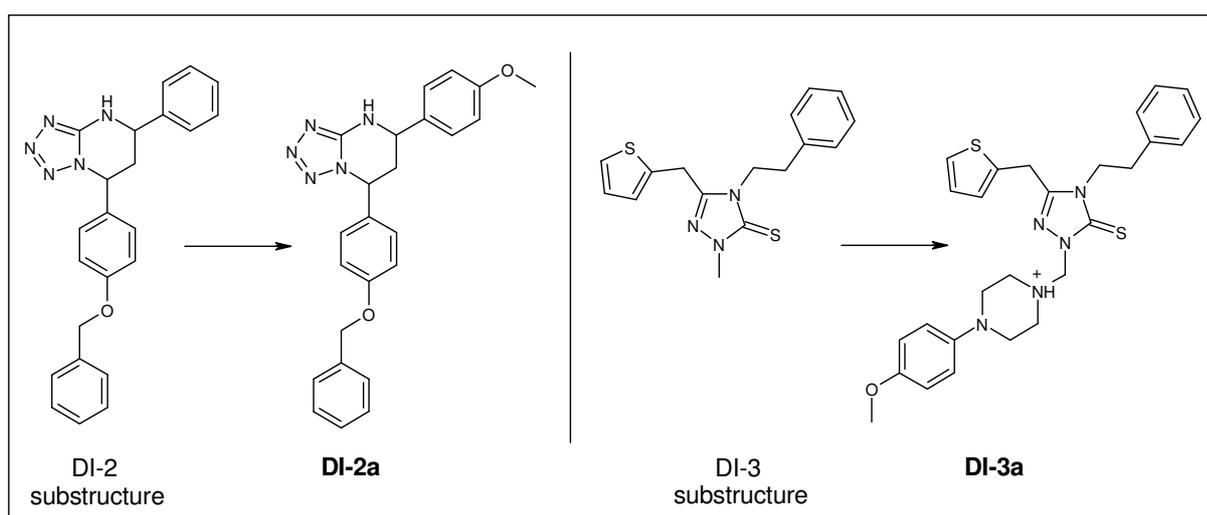
The 100 hit compounds were docked to the dimer interface, using the MD-derived snapshot 4000 ps as protein input structure. The binding site was defined by a sphere of 10 Å radius centered at the amide hydrogen of Ala86. The search efficiency was set to 100 % and genetic algorithm parameters were automatically assigned by GOLD3.2, depending on the molecule's size. For each compound, a total of 50 GA runs were performed. The docking results were then analyzed by means of visual inspection. Here, the docking poses were critically assessed, in particular with regard to shape complementarity: Candidate compounds were required to occupy the dimerization interface subsites (cav1, cav2, groove); in addition to that, docking poses with highly solvent-exposed parts of the molecule were rejected. In the last step, it was checked whether the compounds were available for purchase.

Finally, four purchasable molecules showed a favorable fit to the binding site, along with good scoring. These four molecules were selected as candidate compounds (see Scheme 1).



Scheme 1: Selected hit compounds from virtual screening using pharmacophore model I.

For a detailed view of the binding modes, see Figure 36a/b. To find more analogues of these promising scaffolds, a substructure search of the emolecules¹¹⁷ online database was performed: Derivatives were found for **DI-2** and **DI-3** (see Scheme 2). The derivatives were also docked and showed a good fit to the binding site (see Figure 36a/b). Accordingly, six candidate compounds (**DI-1**, **DI-2**, **DI-2a**, **DI-3**, **DI-3a**, and **DI-4**) were obtained from this first virtual screening.



Scheme 2: Substructure search queries, yielding derivatives of virtual screening hit compounds.

Looking at the binding mode of **DI-1**, one can see that the tricyclic moiety nicely occupies cav2, forming a hydrogen bond to the guanidinium group of Arg107. The ethyl group addresses cav1, whereas the benzimidazole moiety is accommodated by the groove. Overall, a good shape complementarity is visible, confirming the Y-shaped ligand approach. **DI-2** also shows this Y-shaped binding motif: The fluorophenyl ring is harbored by cav2, the tetrazolopyrimidine system points towards cav1. The phenoxy group is placed right above Val180, leading to favorable hydrophobic interactions, whereas the terminal phenyl ring is in a staggered position to Phe181. **DI-2a** has a similar orientation, as expected. **DI-3** and its derivative **DI-3a** share a similar binding mode: The thiophene ring occupies cav1, having its plane in a favorable orientation towards the aliphatic side chain carbon atoms of Gln177, thereby exploiting hydrophobic interactions. The second cavity, cav2, accommodates the ligand's phenyl ring, which in the case of **DI-3a** is almost in edge-to-face orientation to Phe100. This edge-to-face orientation is also observed in the docked binding mode of **DI-4**: Here, the quinoline ring occupies cav2 in this very orientation. The aromatic nitrogen acts as hydrogen bond acceptor for the guanidinium group of Arg107. The sulfur atom of the thienopyrazole moiety points towards the hydrophobic "floor" of the groove (Figure 36b). Overall, the molecules show a good shape-complementarity to the dimerization interface.

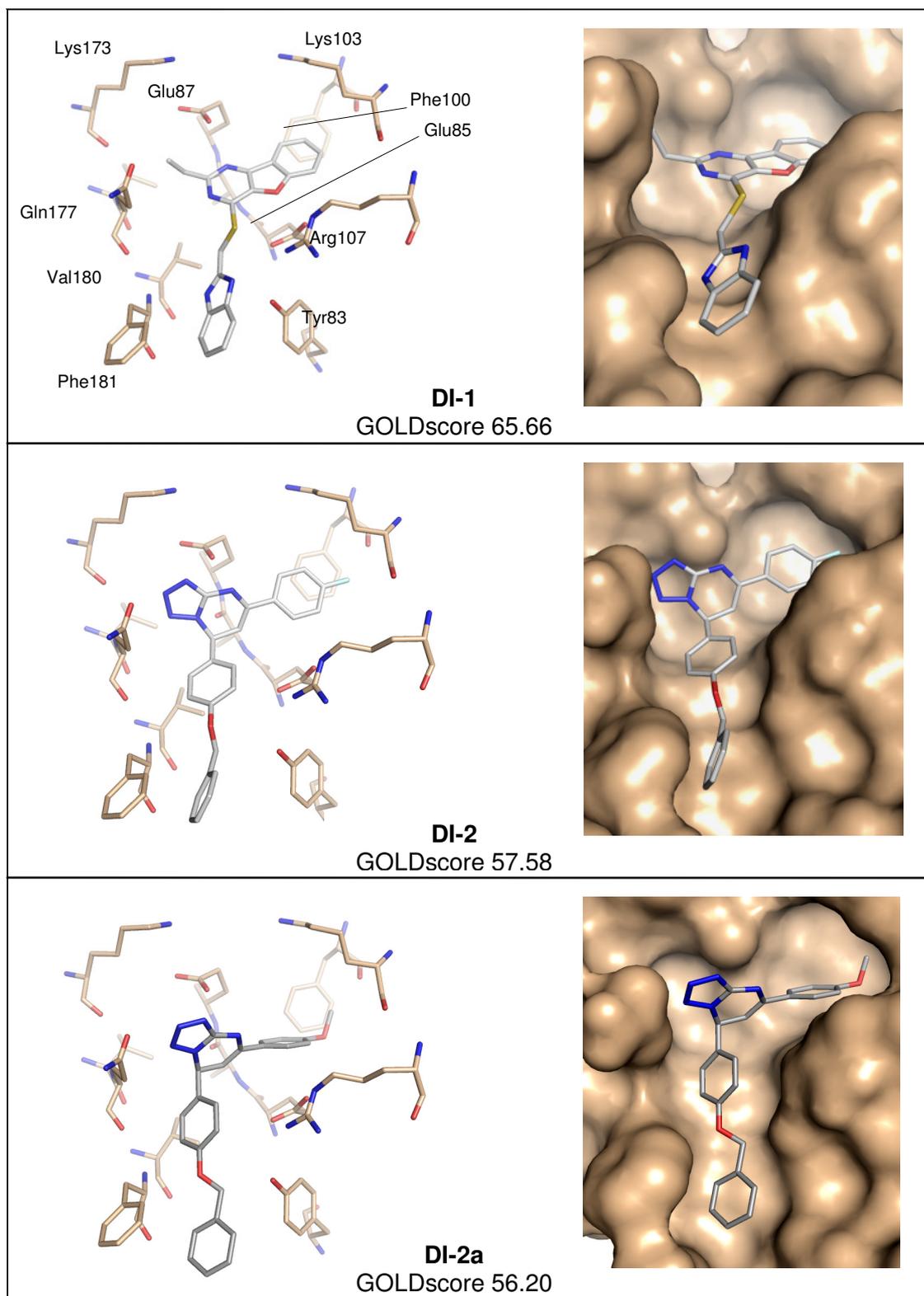


Figure 36a: Docked binding poses of hit compounds DI-1 – DI-2a. On the left, the protein is shown as sticks, on the right it is displayed in surface representation.

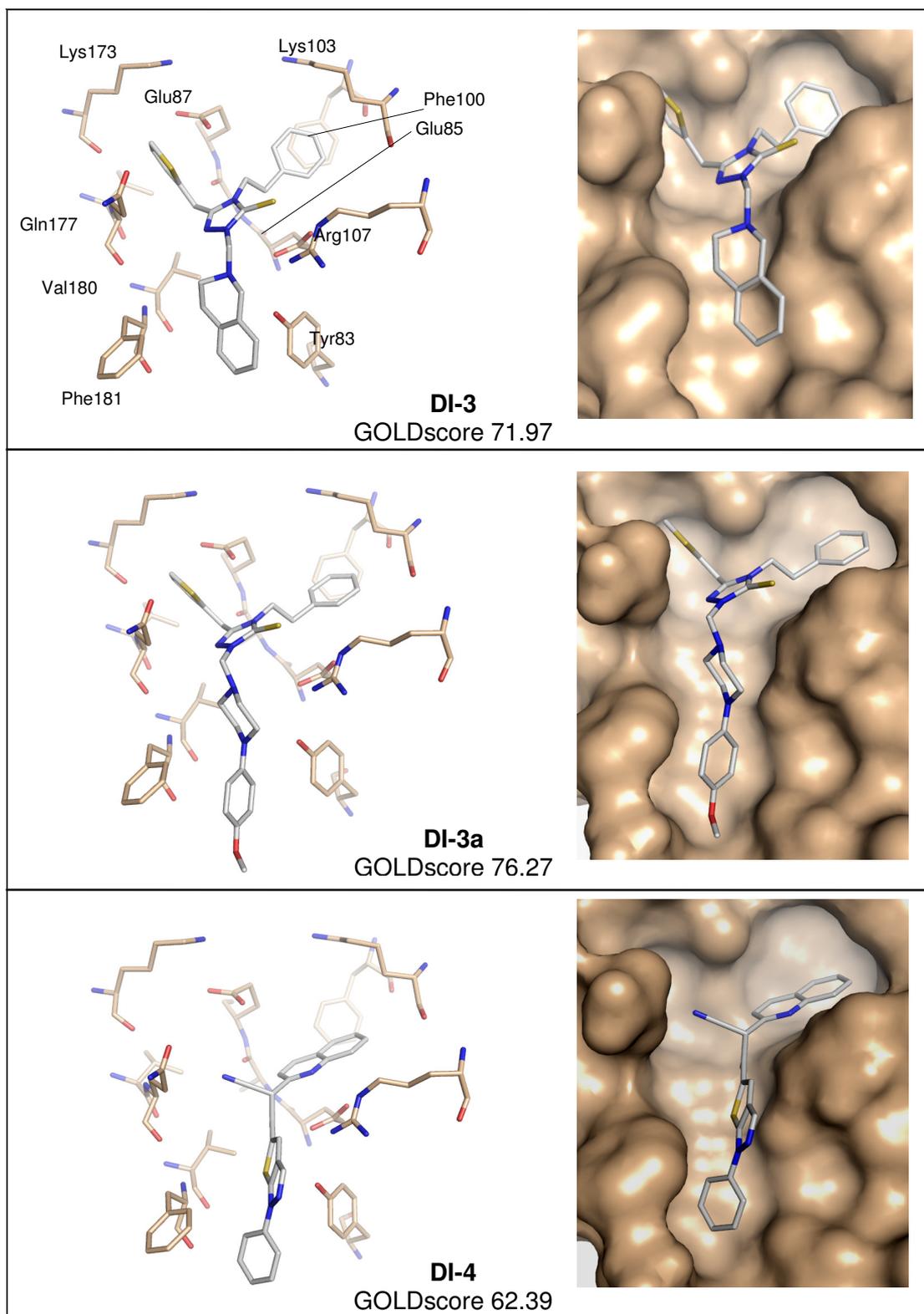
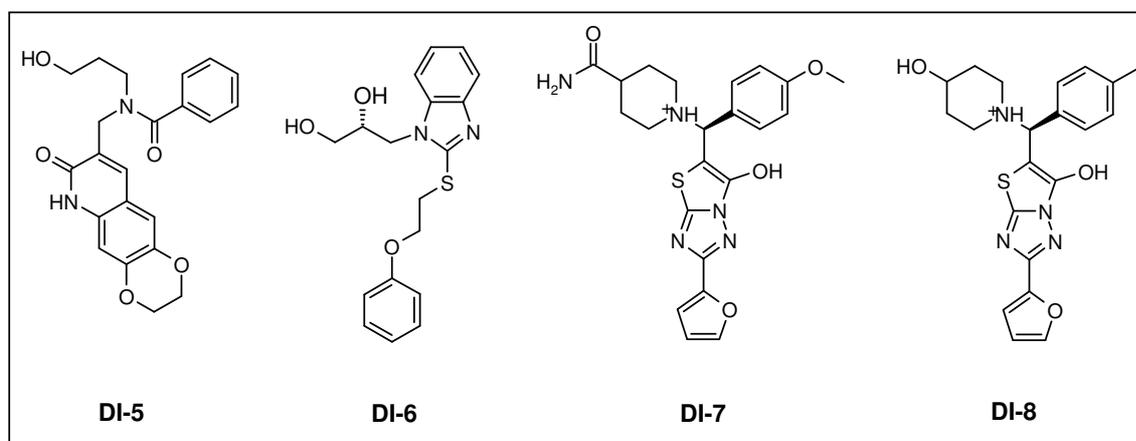


Figure 36b: Docked binding poses of hit compounds DI-3 – DI-4. On the left, the protein is shown as sticks, on the right it is displayed in surface representation.

Pharmacophore model II – Model creation, virtual screening, and docking

The hot-spot analysis of the dimer interface also showed a donor hotspot in cav1. This information was used to modify pharmacophore model I: Feature F1 was changed to a donor feature, while the remaining features and volumes were kept unchanged. The resulting pharmacophore model II was employed for screening. Since only ten matching molecules could be retrieved from the ZINC database, a second database was screened: The MOE conformational database comprises ~660,000 molecules in different conformations. Using pharmacophore model II as query, 54 hits were found. Without further filtering, these compounds were subjected to docking: The same protocol (see above) was used to dock the 54 compounds. After visual inspection of the poses, four purchasable compounds were selected (see Scheme 3).



Scheme 3: Selected hit compounds from virtual screening using pharmacophore model II.

As shown in Figure 37, **DI-5** places its tricyclic system within the groove, cav2 is occupied by a phenyl ring. Congruent with the pharmacophore model, cav1 is addressed by the terminal hydroxy group, which acts as a hydrogen bond donor with Gln177. Also in the binding mode of **DI-6**, cav1 is occupied by hydrogen bond donating groups: The dihydroxy moiety interacts with the backbone oxygen of Lys173 and with the carboxyl group of Glu87. The compound's phenyl ring is in a favorable edge-to-face orientation to Phe100 in cav2. Arg107 donates a hydrogen bond to the benzimidazole nitrogen. Interestingly, in the case of **DI-7**, cav1 is not addressed by a hydrogen bond donor, but cav2: The terminal amide group, which according to the pharmacophore model was supposed to render the hydrogen bond donor for cav1 rather binds to cav2 by forming a hydrogen bond with the backbone oxygen of Phe100. The positively charged nitrogen of the piperidine ring is placed right between the carboxylate groups of Glu85 and Glu87, respectively. The sulfur atom of the triazolothiazol moiety points towards the hydrophobic "floor" of the groove. A similar binding mode is observed for **DI-8**: Again, the positively charged nitrogen is placed between the Glu85 and Glu87.

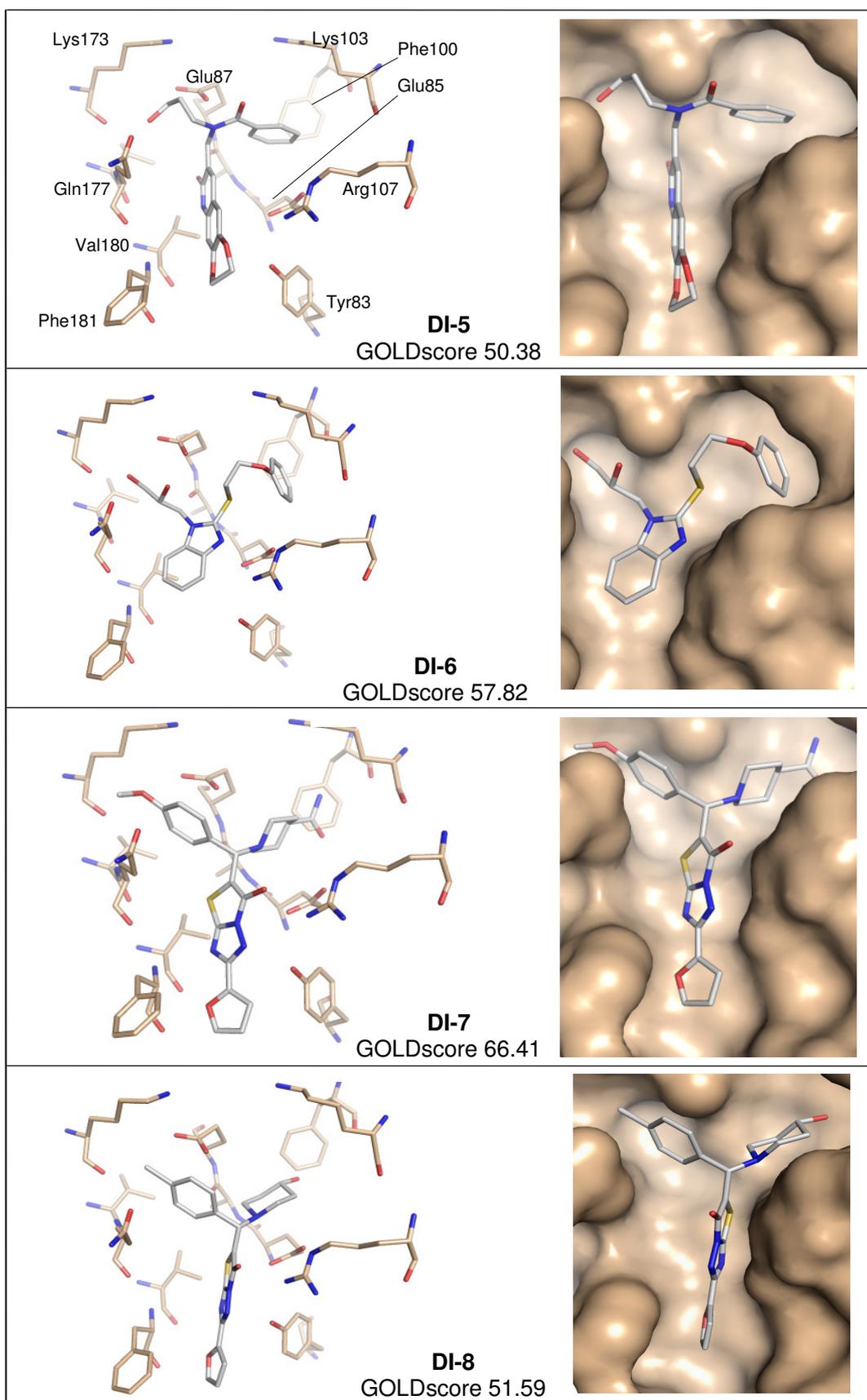
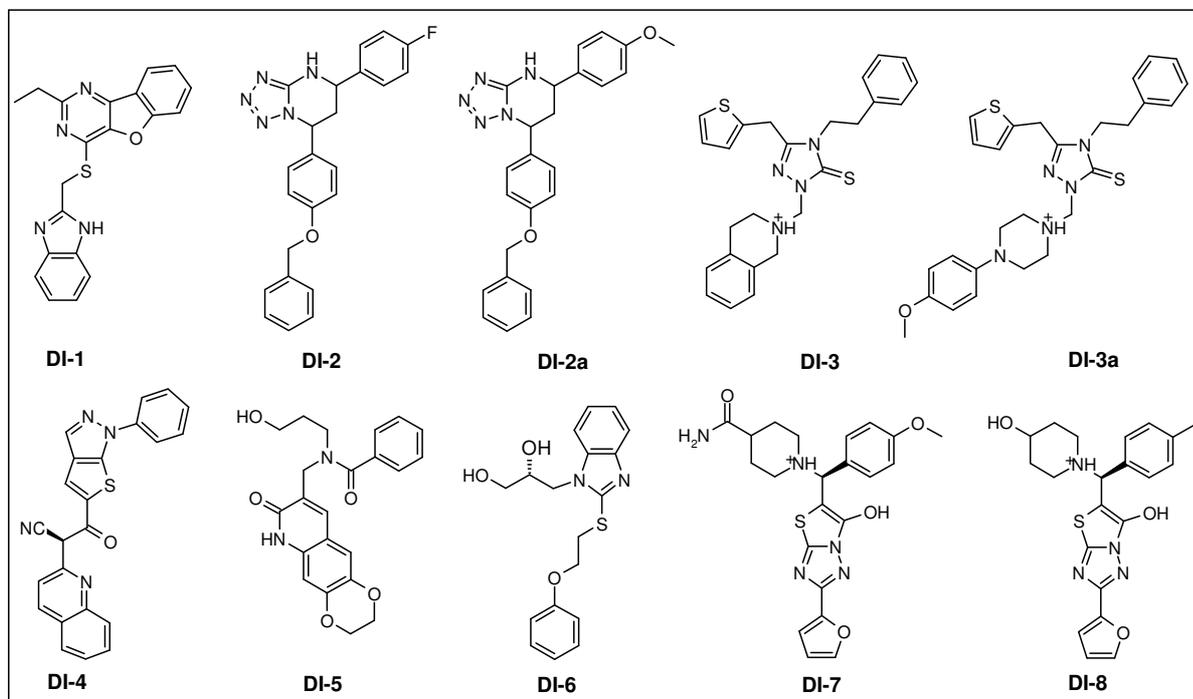


Figure 37: Docked binding poses of hit compounds DI-5 – DI-8. On the left, the protein is shown as sticks, on the right it is displayed in surface representation.

Dimerization inhibitor candidates for biological testing

The virtual screenings performed so far yielded a total of 10 purchasable dimerization inhibitor candidates, which were submitted to biological testing in an IN integration assay (Scheme 4).



Scheme 4: Final selection of candidate compounds for biological testing.

All compounds have the distinction of a good shape complementarity with the dimerization interface, which is reflected by high hydrophobic contributions to the overall score (Table 12).

	GOLDScore	S(hbond)	S(hydrophobic)	S(internal strain)
DI-1	65.66	3.84	63.81	-1.99
DI-2	57.58	0.00	63.09	-5.51
DI-2a	56.20	0.00	63.96	-7.76
DI-3	71.97	0.19	80.71	-8.93
DI-3a	76.27	0.00	84.64	-8.37
DI-4	62.39	2.78	63.95	-4.34
DI-5	50.38	6.05	57.66	-13.33
DI-6	57.82	4.84	60.88	-7.90
DI-7	66.41	1.48	70.43	-5.50
DI-8	51.59	2.24	56.84	-7.49

Table 12: Contributions to the overall GOLDScore for the ten selected candidate compounds.

2.5.4 Biological testing of dimerization inhibitor candidates

The IN inhibition assays were performed in the lab of Dr. Nouri Neamati at the University of Southern California, Los Angeles. In this assay, radioactive-labelled DNA (21 base-pair length) is processed by IN, which leads to a mixture of DNA strands of different length: 3'-processing (3'P) yields a 19-mer, whereas the strand transfer (ST) reaction results in a DNA adduct, i.e., the DNA substrate undergoes a recursive integration, leading to a polymer. This mixture of DNA strands is separated via polyacrylamide gel electrophoresis (PAGE) and detected by a radiosensitive device. Inhibition can be deduced from absence of characteristic bands (19-mer for 3'P, polymer for ST).

Methods

Preparation of stock solutions: For each candidate compound, 10 mM stock solutions were prepared, which were then further diluted to 1000 μ M, 333 μ M, 111 μ M, 37 μ M and 12 μ M solutions, respectively.

Preparation of radioactive-labelled DNA: With this procedure, two unlabelled single DNA strands are transformed to radioactive-labelled double-stranded DNA: In a kinase-catalyzed reaction, radioactive-labelled ATP (32 P-ATP) is introduced. Preparation: 10 μ l H₂O, 8 μ l PNK buffer, 5 μ l 21-mer (single-stranded DNA, referred to as “top oligo”), 2 μ l T4-Kinase, and 45 μ l 32 P-ATP were incubated 45 min at 37 °C, followed by another 15 min at 85 °C. Then, 10 μ l 21-mer (complementary single-stranded DNA, referred to as “bottom oligo”) and 150 μ l H₂O were added. The mixture was incubated at 85 °C for 2 min. The heater was turned off. After 4 h, the mixture was incubated at 37 °C for 30 min. The columns were spun twice to remove buffer, before spinning at 10,000 rpm for 1 min to collect fractions for storage.

Preparation of DNA-cocktail: 100 μ l MOPS (125 mM, pH = 7.4), 55 μ l glycerol, 52 μ l H₂O, and 20 μ l radioactive-labelled DNA were mixed.

Preparation IN-cocktail: 36 μ l MnCl₂ (100 mM), 48 μ l BSA, 48 μ l BME (2 %), 64 μ l H₂O, and 12 μ l IN (200 nM) were mixed.

Preparation of acrylamide gel: 55 ml acrylamide were polymerized using 390 μ l APS (10 %) and 39 μ l TEMED.

Procedure: In each test tube, 7 μ l IN-cocktail and 1.6 μ l test compound were preincubated for 30 min at 30 °C (allowing the test compound to interact with its target enzyme). After adding 7.4 μ l DNA-cocktail, the reaction mixture was incubated for 1 h at 30 °C (addition of the IN substrate (DNA) leads to an interplay between protein, substrate, and test compound). The

reactions were quenched with loading dye (98 % formamide, 0.025 % xylene cyanol, 0.025 % bromophenol blue) and transferred onto the polyacrylamide gel (the loading dye's formamide denatures the DNA substrate and therefore stops the reaction). The electrophoresis was performed at P = 55 W in TBE buffer. After a sufficient separation of the bands, the gel was dried, exposed in a PhosphorImager cassette for 12 h and analyzed on a Typhoon 8610 variable-mode imager (Amersham Biosciences). In some cases, the assay had to be repeated, due to unsteady separation and exposure.

Results and discussion

Gel images are shown in Figure 38. For each compound, five dilutions were tested to determine IC_{50} values: Concentrations of 100 μ M, 33 μ M, 11 μ M, 3.7 μ M, and 1.2 μ M were used.

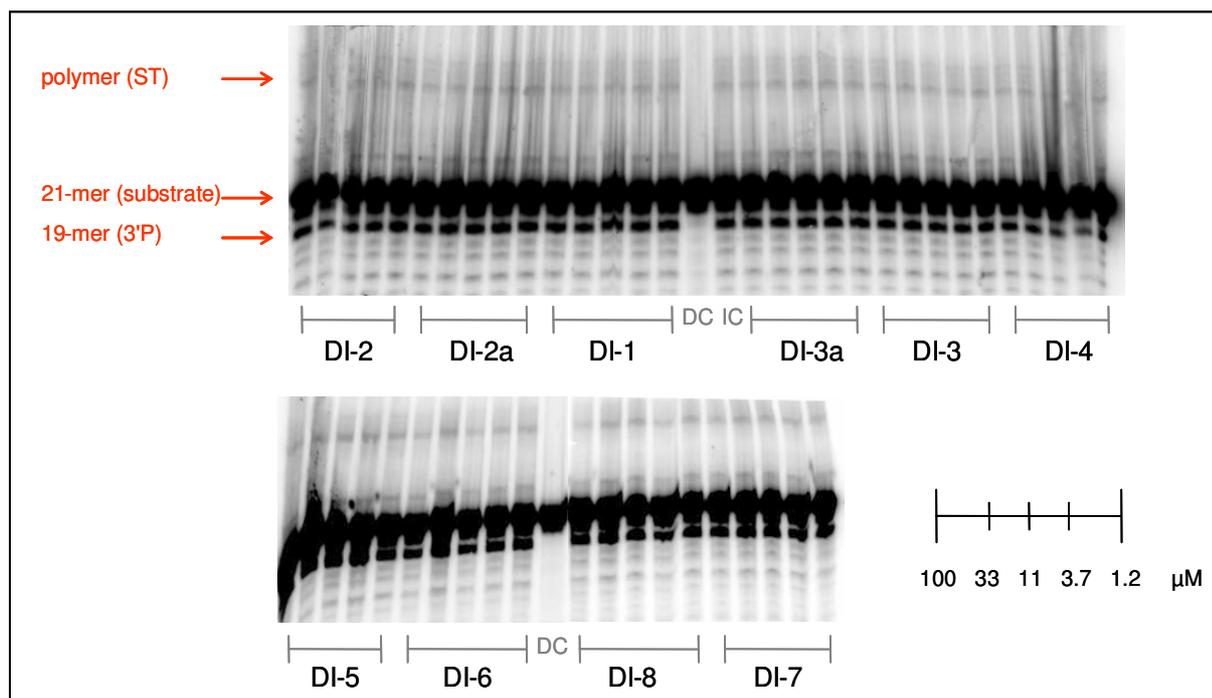


Figure 38: Gel images of compound testings. Every compound was tested in five dilutions (see legend). For reference purposes, a DNA control (DC, IN- and compound-free sample) and an IN control (IC, compound-free sample) are included.

Compared to the IN control, the gel images do not show a significant decrease in intensity of the strand transfer or the 3'-processing band. In several cases (e.g. **DI-4**), the bands appear somehow brightened, but this is most probably due to uneven loading of the reaction mixtures. The assay was repeated, also with a longer incubation time (4 h instead of 1 h), but none of the compounds showed reproducible inhibitory activity.

The assay results indicate that the compounds tested so far do not inhibit IN activity and therefore are unlikely to act as dimerization inhibitors. However, still some doubts remain: The use of BSA (bovine serum albumine, used to block protein aggregation) is critical, because it is not improbable that BSA interfaces with the hydrophobic compounds and therefore prevents them from binding to the IN dimerization interface.

2.5.5 Dimerization inhibitor design II

The IN inhibition assay results could not confirm the hypothesis of compounds **DI-1 – DI-8** acting as dimerization inhibitors, although still some doubts remain (use of BSA). However, the strategy for designing dimerization interface binders may have to be reconsidered: Originally, the rationale was to find compounds which are complementary to the Y-shaped dimer interface groove. The known peptidic binders (YFLLKL, TAYFLLKLAGRW) have a prevailing hydrophobic character: Therefore, it is reasonable to assume that binding interactions are driven by rather hydrophobic interactions. Accordingly, a good shape-complementarity should be a sound basis for good protein-ligand interactions. However, a tradeoff for these favorable hydrophobic interactions is unavoidable: Hydrogen bond acceptors or donors (HBAs/HBDs) of the protein might not be sufficiently satisfied by corresponding ligand counterparts upon ligand binding. These considerations finally lead to the development of further pharmacophore models. This time, the satisfaction of interface hydrogen bonding groups by ligands received major attention.

GRID analysis of the dimer interface site

The program GRID (version 22a)^{118,119,120,121} was used to scan the dimer interface for interaction hot-spots. Here, special attention was turned on polar probes, capable of satisfying the hydrogen bonding groups within the binding site. Input structure for the dimer interface scan was the MD-derived snapshot 4000 ps, which was found to be a representative protein conformation. The GRID computation site was defined by a box with the dimensions 19.9 · 24.6 · 16.9 Å³, to cover

the entire IN dimer interface. Default settings were used for calculations with the following probes: O1 (hydroxy group), N3+ (cationic amine), N1: (N-H with lone pair), N: (nitrogen with lone pair), and O (carbonyl oxygen). These probes were selected as representatives of H-bond-donating (O1, N3+, N1:) and accepting (N:, O:) ligand atoms. An MOE SVL script, provided by the CCG upon request, was used to convert the resulting GRID .kont files. The readable .grid files were imported in MOE for subsequent pharmacophore model creation. See Figure 39 for contour plots of the different probes. The contour levels were adjusted in a way that distinct hotspots for the different probes were obtained.

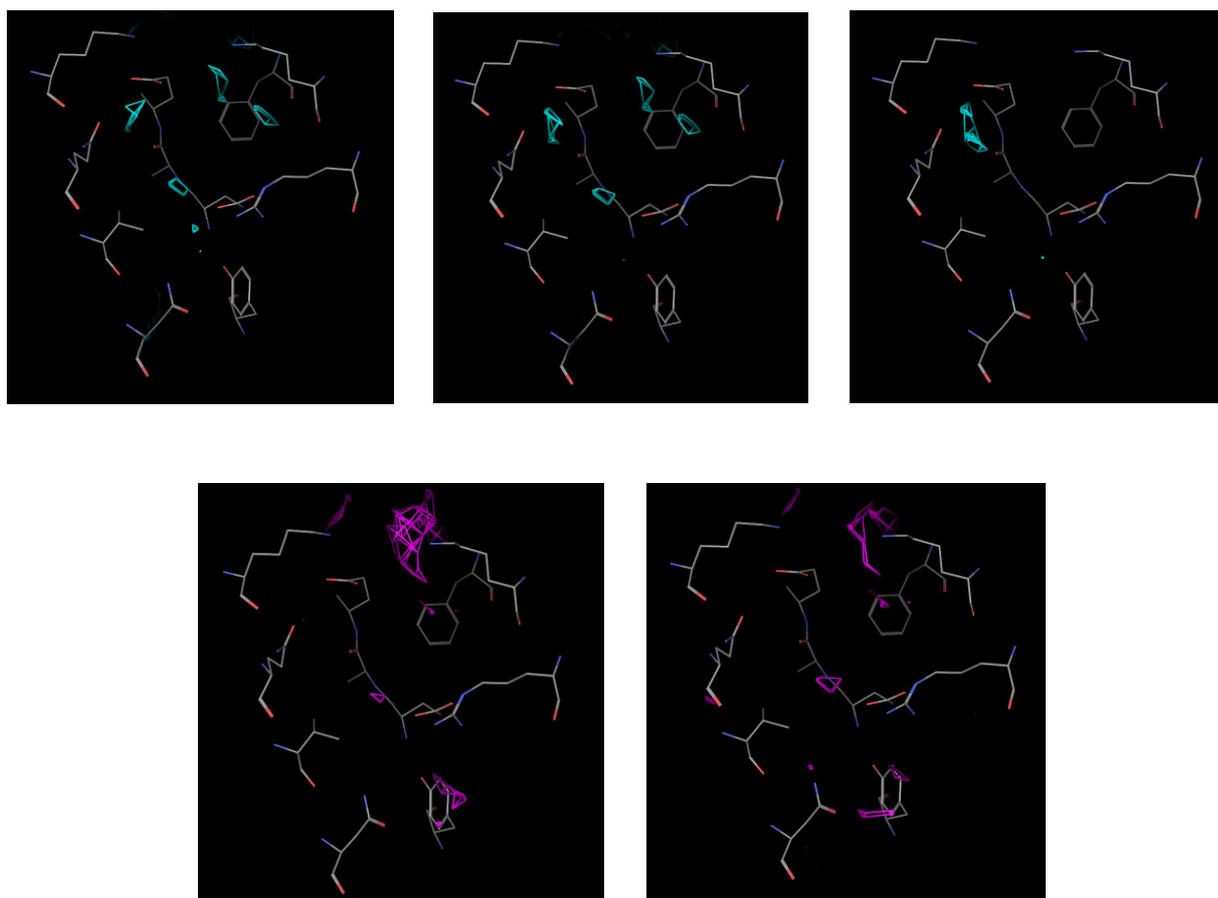


Figure 39: Top: GRID donor hotspots. Left: O1 probe (contour level = -7.4 kcal/mol), middle: N1 probe (contour level = -8.1 kcal/mol), right: N3+ probe (contour level = -10.0 kcal/mol). Bottom: GRID acceptor hotspots. Left: O: probe (contour level = -5.0 kcal/mol), right: N: probe (contour level = -5.4 kcal/mol).

The contour plots show that there is a favorable interaction point for probe N3+, which is located in cav1, close to Glu87. This is in accordance with the lysine residue position of the inhibitory peptide YFLLKL. Also, similar hot-spots for the donating probes O1 and N1 could be detected, however at less negative contour levels. These hot-spots are located in cav1 with its

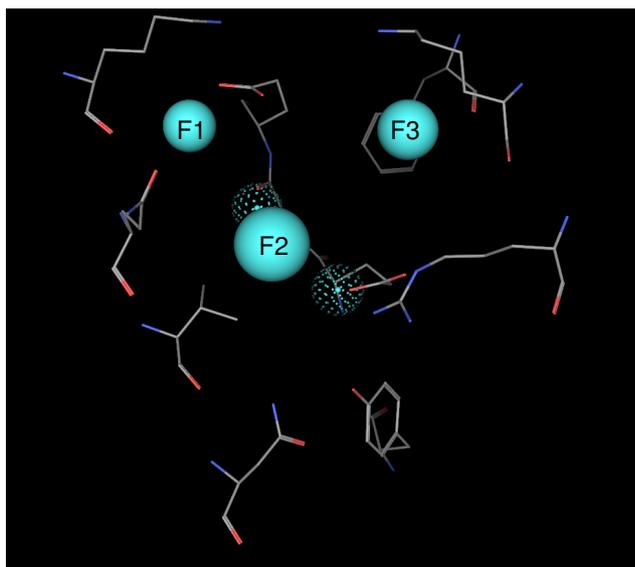
three acceptor groups (terminal amide oxygen of Gln177, carboxyl group of Glu87, and the backbone oxygen atom of Lys173); in addition, the backbone oxygen of Ala86 is close. Furthermore, two distinct donor hot-spots were detected within cav2: Apparently, these hot-spots represent interaction points with the backbone oxygen of Phe100 and the carboxyl group of Glu87. A fourth hot-spot indicates a possible donation to the backbone oxygen of Ala86 (see Figure 42 for Ala86 position) and the carboxyl group of Glu85, which are located within the groove. In contrast to that, the accepting probes N: and O show only few interaction points at significantly less negative contour levels: Hot-spots appear close to the amide hydrogen of Ala86 and the terminal amine group of Lys103.

The results of the GRID interface scan indicate that the main interaction points are clearly within cav1 and the northern part of the groove, i.e. close to Ala86. Furthermore, a donating group within cav2 is also favored (interaction with the backbone oxygen of Phe100). Based on these findings, a pharmacophore model was created: Donor features were placed on the respective hot-spots to find molecules with the desired interaction points.

Pharmacophore model III – Model creation and virtual screening

For the pharmacophore model creation, the MOE pharmacophore methodology was employed. First, the GRID hotspots and the protein structure 4000 ps were imported, before the respective pharmacophore feature spheres were manually placed on the hot-spot positions. In addition to the donor features, excluded volume spheres were set along the interface residue atoms: Thereby, potential clashes between ligand and protein were avoided. Since a nitrogen probe (N3+) showed the most favorable interaction energies within cav1, the respective donor feature was required to have a nitrogen atom by setting the adequate SMARTS expression. Projected features of donor F2 (see Figure 40) address the backbone oxygen of Ala86 and the carboxy group of Glu85, respectively. The sphere radii were set to find the best possible tradeoff between a permissive and a restrictive pharmacophore model. Partial matching was allowed: Here, only one of the projected features needs to be matched (the donor features were set to “essential”, meaning that these must be matched; volume features are always required).

Virtual screening of the ZINC database yielded a total of 1103 unique compounds, all of which were subjected to docking.



	Feature	Radius [Å]
F1 (ess.)	Don&“N“	1.0
F2 (ess.)	Don	1.5
F3 (ess.)	Don	1.2
F4	Don2	1.0
F5	Don2	1.0
V1	Excl.	1.5
Partial match	at least 4	

Figure 40: Pharmacophore model III. Here, satisfaction of hydrogen bonding groups was pivotal, the criterion for molecules to have a Y-shape was secondary. See the table on the right for information about features and sphere radii.

Pharmacophore model III – Docking of virtual screening hits

In a first docking study, the compounds were docked with GOLD4.0. Inspection of the docked binding poses revealed that often bad contacts (hydrophobic ligand groups close to hydrophilic protein residues and vice versa) are formed. Such binding modes are rather improbable. Accordingly, GOLD was discarded as appropriate docking tool for this study, which focuses on the best possible match between ligand and protein hydrogen bonding groups. Thus, AutoDock3.0 was used for compound docking. The AutoDock scoring function has an explicit term for electrostatic interactions and furthermore a penalty term for desolvation of polar protein atoms.

The compounds were docked to the dimer interface of snapshot 4000 ps. Gasteiger-Marsili charges were added to the fully protonated compounds by Sybyl8.0, before *autotors* was employed to create the required ligand input files. Polar hydrogens were added to the protein input structure using the *protonate* tool. Kollman united-atom partial charges and solvation parameters (derived from the method of Stouten *et al.*) were assigned by the inhouse-tool *add_chrgsol*. The binding site was defined by a cubic box with 65 grid points in each dimension, centered on the Ala86 nitrogen, and the grid spacing was set to 0.375 Å. Interaction maps were calculated with AutoDock's *antogrid* module. The docking protocol used an initial population of 50, a maximum

number of $1.0 \cdot 10^6$ energy evaluations, a mutation rate of 0.02, a crossover rate of 0.80, and an elitism value of 1. The Solis and Wets local search parameters were set as follows: The probability of performing a local search on an individual was set to 0.06, a maximum number of 300 iterations was performed, and the number of consecutive successes or failures before changing the size of the local search space was 4. 10 GA runs per compound were carried out. To allow a better comparison between molecules of different size, per-atom energies were calculated: The docked energy was divided by the number of heavy atoms of the respective molecule.

A random visual inspection of the docking results showed that in many cases hydrogen bonding groups of the protein and the ligand, respectively, remain unsatisfied in the docked ligand binding pose. To retrieve those binding poses with the least number of unsatisfied polar groups, a descriptor was applied to the poses, which detects unsatisfied buried HBAs/HBDs of both protein and ligand. Such a descriptor is implemented in the HERMES suite¹²². Accordingly, the binding poses generated by AutoDock were postprocessed with the HERMES descriptor, which counts the number of unsatisfied buried HBAs/HBDs of both protein and ligand within each docking pose (see Figure 41).

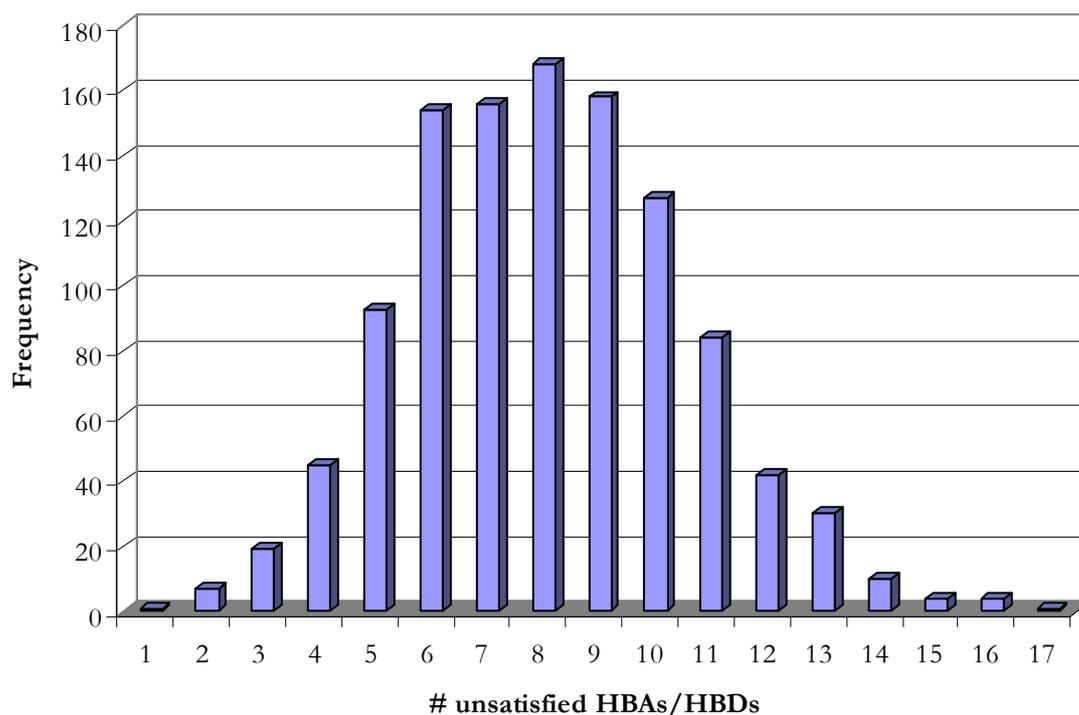


Figure 41: Count of unsatisfied buried HBAs/HBDs of both protein and ligand within each of the top-ranked docking results of all 1103 compounds.

None of the docking poses displays a perfect complementarity of protein and ligand in terms of hydrogen bonding. Surprisingly, 93 % of all docking poses have ≥ 5 unsatisfied HBAs/HBDs. For a detailed consideration of this aspect, see Chapter 2.5.7. The threshold for an acceptable unsatisfied HBAs/HBDs count was set to 4 (see Chapter 2.5.7), yielding 72 compounds for further analysis.

Pharmacophore model III – Selection of compounds

The remaining 72 molecules were subjected to visual inspection: Compounds with solvent-exposed hydrophobic groups were rejected. The hydrogen bonding pattern of the individual compounds was further analyzed with respect to proper geometries (distance, angle), since the unsatisfied HBAs/HBDs count used a rather permissive hydrogen bonding definition (see Chapter 2.5.7). Furthermore, the docked binding modes were assessed with respect to further favorable features like attractive electrostatic interactions. Finally, the visual inspection yielded six compounds (Scheme 5), which are planned to be tested in cell-based HIV infection assays. For the latter, molecules are required to have cell-permeability properties: A rough estimation of the compounds' permeabilities can be deduced by calculating logP and polar surface area descriptors: Egan and coworkers¹²³ demonstrated that well-absorbed compounds have logP values ranging from -1 to 5.9 and a polar surface area between 0 and 132 Å². Accordingly, the hit molecules are further characterized by these descriptors: The MOE database descriptor calculator was used to calculate logP and polar surface area (PSA) values. As shown in Table 13, the hit molecules fulfil the criteria for cell permeability. For a size-independent measure of the intermolecular energies, a per-atom energy (docked energy divided by the number of heavy atoms) was introduced. Here, all of the six hit compounds are characterized by very favorable per-atom energies (for example **DI-9** with -0.45 kcal/mol).

The docked binding modes are shown in Figure 42. A first overview revealed that none of the poses correlate with the pharmacophore model: Placement of donating groups does not coincide with the position of the pharmacophore features (which were derived from the hot-spot analysis).

Most probably, the mismatches between docked and pharmacophore model positions are due to the large number of donors / acceptors within the interface binding site and their peculiar arrangement: Within a relatively small region, there is a frequent alternation of donating and accepting groups. In other words, there is a vast number of different interaction patterns between protein and ligand; so the ligand's donor groups can easily be "distracted" from the original pharmacophore feature position. The peculiarity of the IN dimerization interface is discussed in more detail below. However, despite the fact that the pharmacophore motif was not matched, the docked binding poses show binding motifs with interesting and meaningful protein-ligand interactions.

DI-9 is favorably anchored in cav2 with its purinone system, which interacts with the phenyl ring of Phe100 via stacking. Furthermore, the backbone oxygen of Phe100 accepts a hydrogen bond from the ligand. Glu87 is addressed by the donating N-H within the five-membered ring, whose second nitrogen accepts a hydrogen bond donated by Arg107. The ligand's hydroxy group is placed close to the backbone oxygen of Ala86, the terminal methyl group occupies cav1. The groove largely remains unoccupied. Although the positions of donating groups after docking do not coincide with the original hot-spots (from which the pharmacophore model had been derived), the docked binding mode of **DI-9** shows favorable protein-ligand interactions. The binding mode of **DI-10** also does not coincide with the pharmacophore model features, but also shows interesting aspects. The cavities are not occupied, but the groove is completely blocked: The hydroxypyrazole substructure nicely addresses the acceptor-donor-acceptor sequence (backbone O, N-H of Ala86, carboxy group of Glu85) within the groove. Also, favorable hydrophobic interactions between the heteroaromatic ring and Val180 are observed. The hydroxy group of the naphthene moiety is close to the backbone oxygen of Phe181. **DI-11** has its positive charge placed right between Glu85 and Glu87. The latter forms a strong hydrogen bond with the hydroxypropylamine N-H group. The hydroxy group donates its hydrogen to the amide oxygen of Gln177. The aminopyrazole substructure acts as donor simultaneously to Glu85 and the backbone oxygen of Gln177. **DI-12** nicely stretches from cav2 throughout the groove. Favorable stacking interactions are observed between the pyrrole ring and Phe100. Hydrogen bonds are formed with Ala86 (via its backbone N-H group) and Tyr83. **DI-13** also interacts with the Ala86 N-H and the hydroxy group of Tyr83. As a further interesting feature, the positively charged quaternary nitrogen is placed close to Glu85 and right above the phenyl ring of Phe100, enabling cation- π interactions. Similar to **DI-11**, also **DI-14** has its positive charge placed between the negatively charged dyad Glu85/Glu87. In addition to that, the ligand donates two hydrogen bonds to the backbone oxygens of Ala86 and Lys173.

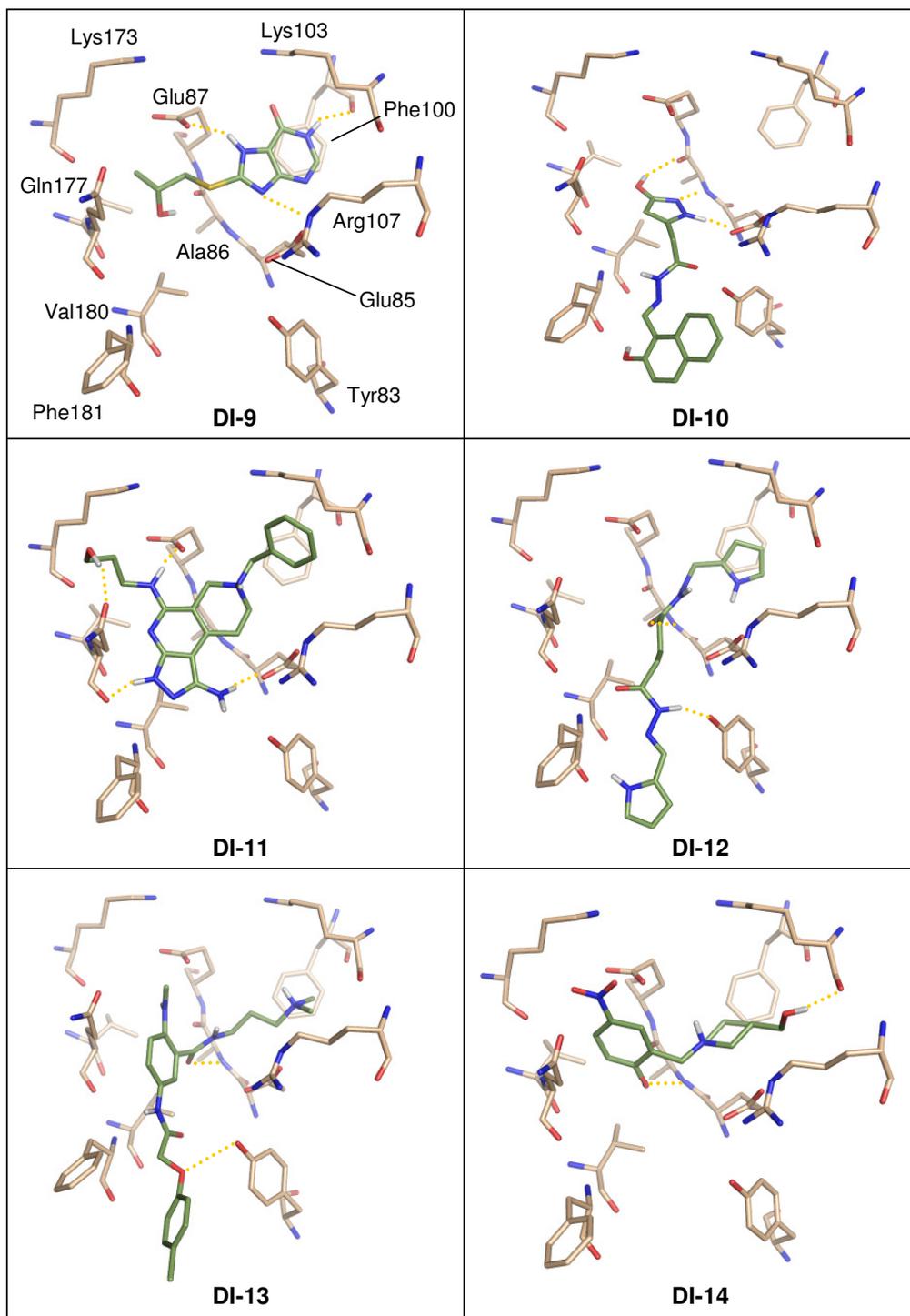


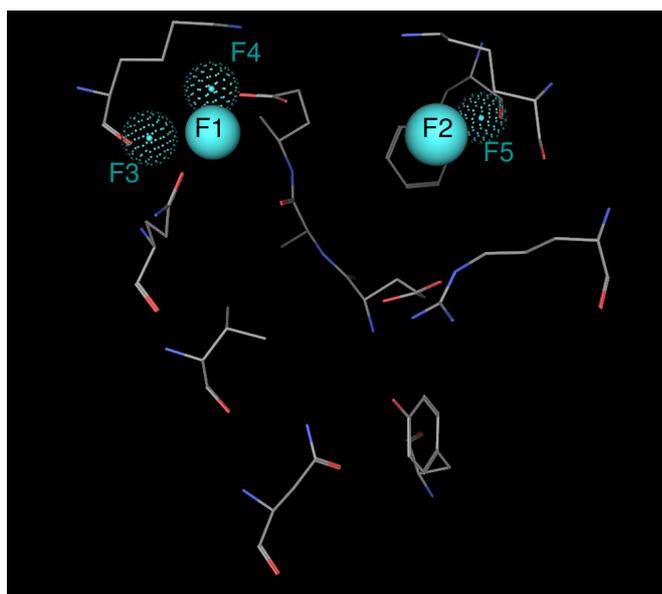
Figure 42: Docked binding modes of pharmacophore model III screening hit compounds.

Biological testing of compound **DI-14** is in progress. Compounds **DI-9 – DI-13** are purchasable, but currently out of stock.

Pharmacophore model IV – Model creation and virtual screening

In an additional pharmacophore model, special attention was directed to cav1 and cav2: The donor features in both cavities were extended by two (cav1) and one (cav2) directed donor features, respectively. Thereby, only those molecules are retrieved, which optimally address both cavities, i.e. hydrogen bonding with Glu87 / Lys173 (cav1) and Phe100 (cav2). Accordingly, the projected donor features were set close to the carboxy oxygen of Glu87, the backbone oxygen of Lys173, and the backbone oxygen of Phe100. Constraint matching was employed: For cav1, only one projected donor feature was required (either for Glu87 or Lys173). The groove was not longer addressed: Setting an additional feature led to a model which was too restrictive to retrieve a sufficient number of compounds (data not shown).

With this pharmacophore model IV, it was assumed that addressing both cavities in the best possible way could overcompensate a missing interaction within the groove (see Figure 43).



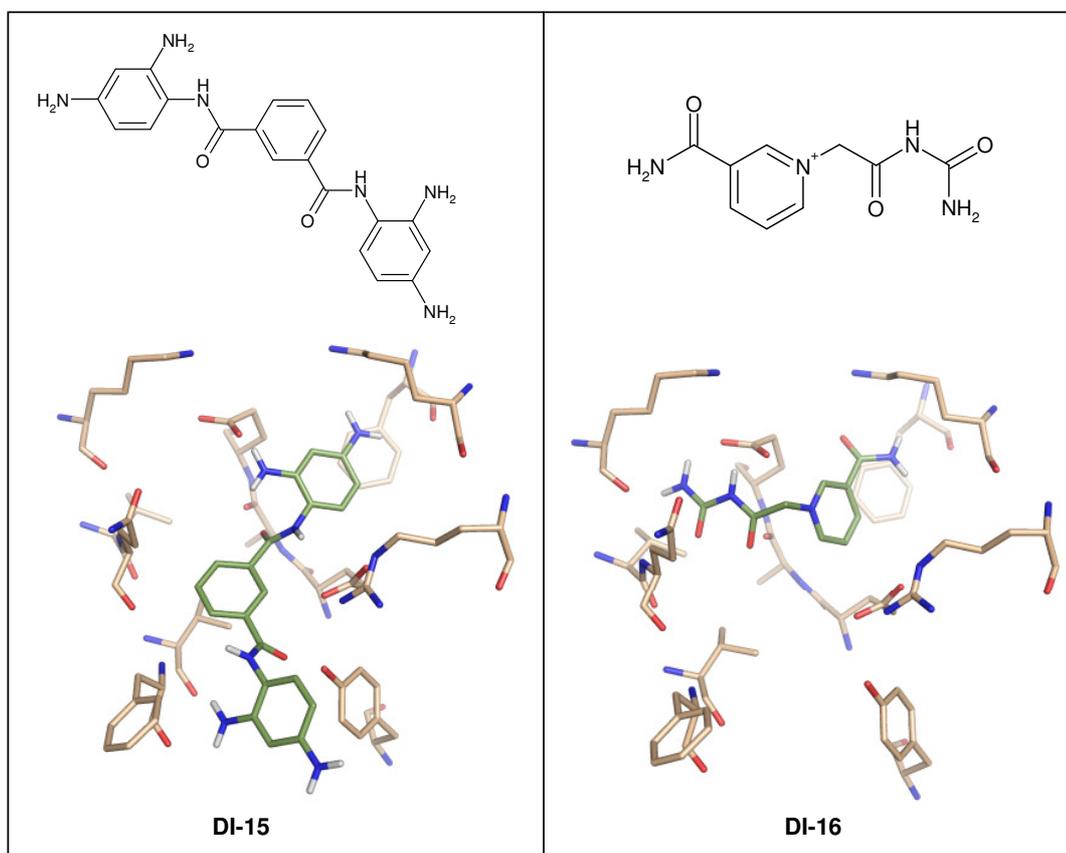
	Feature	Radius [Å]
F1	Don&“N“	1.0
F2	Don	1.2
F3	Don2	1.0
F4	Don2	1.0
F5	Don2	1.0
V1	Excl.	1.5
C1	at least one [F3,F4]	

Figure 43: Pharmacophore model IV. Here, donor features and their corresponding projections were placed solely within cav1 and cav2, the groove was left blank.

Virtual screening with this pharmacophore query yielded a total of 746 unique compounds, all of which were subjected to docking. The same protocol (cf. pharmacophore model III) was employed: After docking with AutoDock3.0, the resulting binding modes were postprocessed with the hydrogen bond descriptor. Visual inspection of the remaining 65 docking poses followed the same criteria (cf. pharmacophore model III).

Pharmacophore model IV – Selection of compounds

Finally, two commercially available compounds were selected (Figure 44). **DI-15** has a triaminobenzene substructure which occupies cav2 and donates hydrogen bonds to Glu87 and Phe100. The rest of the ligand nicely stretches throughout the groove. The Ala86 N-H donor group is close to the internal amide oxygen. The binding motif of **DI-16** bears several favorable interaction features: The positively charged nitrogen is placed between Glu85 and Glu87, whereas Phe100 is addressed by the terminal amide group, which donates a hydrogen bond. The urea moiety occupies cav1 and favorably interacts with acceptor atoms of Glu87 and Lys173. **DI-16** fulfils the pharmacophore model IV features; **DI-15** does not match the pharmacophore features, yet has an interesting binding mode. Biological testing is pending.



	$E_{\text{intermolecular}}$ [kcal/mol]	Per-atom energy [kcal/mol]	Molecular weight [Da]	Calc. logP	PSA [Å ²]
DI-15	-10.70	-0.28	376.4	1.2	109.5
DI-16	-8.09	-0.39	223.2	-1.1	81.9

Figure 44: Docking poses and properties of pharmacophore model IV screening hit compounds.

Virtual screening – Summary

Out of several hundred pharmacophore search hit molecules, which were subjected to docking, only a small percentage (72 out of 1103, ~7 %; 65 out of 746, ~9 %) passed the unsatisfied HBAs/HBDs analysis filter. This underlines the difficulty in properly addressing the alternating basic/acidic/hydrophobic residues within the IN dimerization interface. Subsequent visual inspection of the docking results again diminished the number of promising compounds: Unfavorable interaction motifs, for example aromatic/hydrophobic groups covering Glu85 or Glu87, were frequently observed. This example illustrates the delicate vicinity of hydrophobic and hydrophilic residues: Placement of an aromatic/hydrophobic group close to Phe100 yields favorable interactions, whereas a minor shift in the position of such a group leads to unfavorable contacts with Glu85 or Glu87.

Looking at the docking results, it is striking that the docked orientation of a compound does not match the orientation of donor features intended by the pharmacophore models: The compound seem to be “distracted” by other hydrogen bonding groups of the dimerization interface. Apart from that, of course, the actual number of hydrogen bonding groups within a given compound is in most cases higher than just the number of acceptor / donor features within a pharmacophore model: Various hydrogen bonding patterns between protein and ligand can be realized, which may differ from the pattern which was originally intended by the pharmacophore model.

After the virtual screening (pharmacophore model search – docking – unsatisfied HBAs/HBDs analysis – visual inspection) had been performed, only eight compounds could be selected which showed a sufficient number of favorable interaction motifs. In addition to pharmacophore models III and IV, further alternative models were created and employed (not shown), yet without an improvement in the number of hits. This underlines how delicate the task to find binders to the IN interface really is.

The fact that only very few precast compounds from vendor databases were retrieved suggests the design of custom compounds, tailored specifically for the IN dimerization interface. The compound structures shown here could serve as guidelines, for example in the case of **DI-16**: This urea derivative may be extended by custom synthesis to address also the groove, thus leading to stronger binding.

2.5.6 The peculiarity of the dimerization interface

Compared to other protein-protein interfaces, the IN dimerization interface is characterised by frequently alternating basic, acidic, and hydrophobic residues. It is not a well-ordered binding site with continuous hydrophobic areas and distinct hydrogen bond donors/acceptors. This is illustrated in Figure 45: Compared to the IN interface, the Bcl-X_L interface (regulator of apoptosis, cf. Table 10) has a “well-arranged” binding site, with a continuous hydrophobic area, formed by phenylalanine and valine residues. Polar residues are located at the outer rim of the pocket.

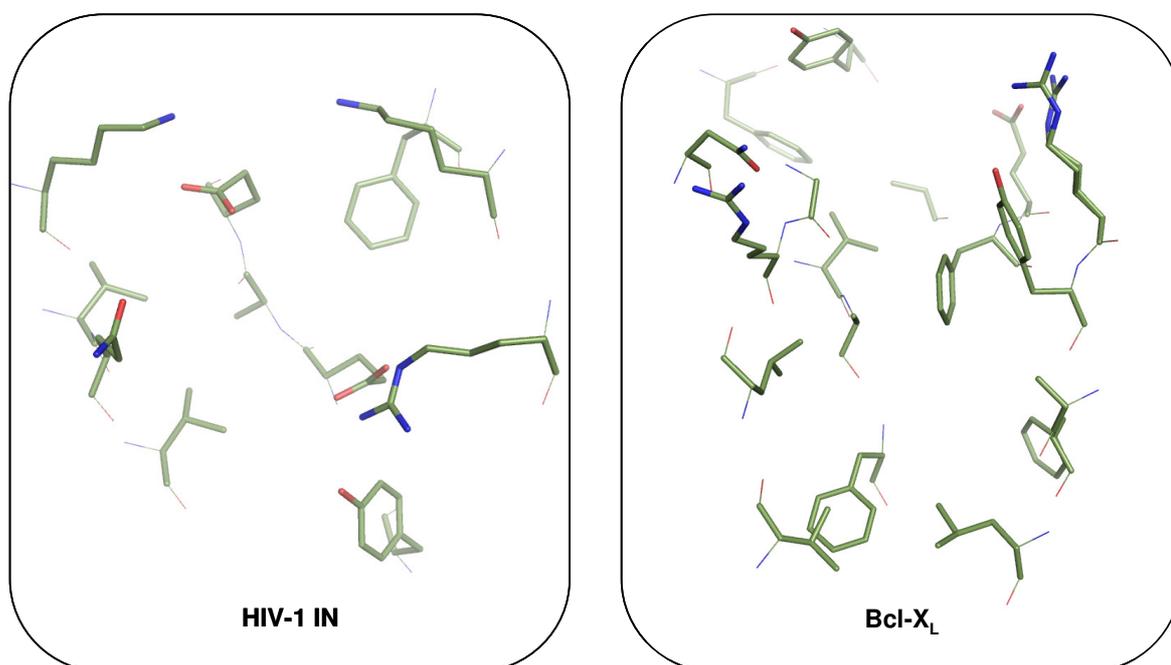


Figure 45: Comparison of the IN dimerization interface with the Bcl-X_L interface. The latter shows a distinct polar rim at the upper part and a large hydrophobic surface in the middle and lower part of the pocket. The IN interface is less clearly ordered.

One can argue that a peculiar binding site like the IN dimerization interface requires tailor-made ligands, which are not expected to be found within a compound collection, but rather would have to be synthesized. This means that even though the hit compounds from the virtual screening performed so far may bind to the IN dimer interface, however, a perfect protein-ligand fit can be achieved solely by custom-made compounds. Besides the implications for IN interface binder design, this virtual screening / docking study stresses the difficulties for docking tools to address the question of polar group desolvation: Many of the docked binding modes obtained in this

study (more in GOLD, fewer in AutoDock) show unsatisfied HBAs/HBDs of both protein and ligand. On the one hand, this is unfavorable for ligand binding; on the other hand, such unfavorable interactions could be compensated by an overall better fit (hydrophobic interactions). The question arises to which extent binding modes with unsatisfied HBAs/HBDs appear in crystallographic protein-ligand complexes: Do native ligand orientations perfectly match the hydrogen bonding pattern of the respective target protein? If not, is there only one or maybe more unsatisfied HBAs/HBDs? A separate study was dedicated to these questions, which is described in the following chapter.

2.5.7 Satisfaction of hydrogen bonding groups in protein-ligand complexes

To address the aforementioned question of potential unsatisfied HBAs/HBDs in native protein-ligand complexes, a collection of high-quality crystallographic structures, the Cheng dataset¹²⁴, comprising 195 protein-ligand complexes, was subjected to analysis.

Methods

The Cheng dataset provides separate protein and ligand structure files, which are fully protonated. However, the hydrogens are not necessarily in optimal positions, particularly with regard to rotatable -OH or -NH groups. Accordingly, protein and ligand structures had to be reprotonated in their respective complex environment, using the sophisticated MOE protonate3D methodology: Here, proper rotamers for -OH, -NH, and -SH groups are calculated. To speed up the calculations, the ligand binding sites were cut out of the protein structures with the tool *fconv* (kindly provided by Gerd Neudert, University of Marburg), which automatically detects the binding site within an 8 Å sphere around the ligand. Then, the ligand mol2 files were converted into pdb format using CORINA, before they were merged with their respective protein pdb file counterpart. The resulting complex pdb files were then imported into an MOE database, which was subjected to protonate3D calculations. The reprotonated complexes were again dissected in protein and ligand structures, which then served as input for the descriptor calculations using the HERMES descriptor calculator. This tool extracts information about non-saturated, buried HBAs/HBDs of both ligand and receptor upon complex formation (i.e., unsatisfied, but exposed hydrogen bonding groups will not be considered). The HERMES

descriptor definition of a hydrogen bond sets the DH..A angle to 90°, and the maximum distance to 3.0 Å. This rather permissive definition is explained in the discussion section (see below).

Results

The descriptor calculator output was parsed for the frequency of unsatisfied hydrogen bonding groups of ligand, protein, and complex. Results are given in Table 14.

Number of unsatisfied HBAs/HBDs	Frequency (ligand)	Frequency (protein)	Frequency (complex)
0	45	50	17
1	49	61	29
2	42	38	38
3	26	18	28
4	18	15	18
5	4	9	29
6	7	4	10
7	1	-	6
8	1	-	6
9	1	-	6
10	0	-	3
11	0	-	1
12	1	-	2
13	-	-	2
	Σ 195	Σ 195	Σ 195

Table 14: Unsatisfied HBAs/HBDs analysis results of 195 complexes. The frequency of unsatisfied hydrogen bonding groups is given for ligand, protein, and the complex.

It is clearly visible that in most complexes, only few hydrogen bonding groups remain unsatisfied: In 92 % of the complexes, ligands have less or equal than four unsatisfied HBAs/HBDs. Regarding the protein, this percentage goes up to 93 %. When accounting for both ligand and

protein, 67 % have four or less unsatisfied hydrogen bonding groups. Some outliers, which have up to 13 unsatisfied HBAs/HBDs, are also observed (see Figure 46).

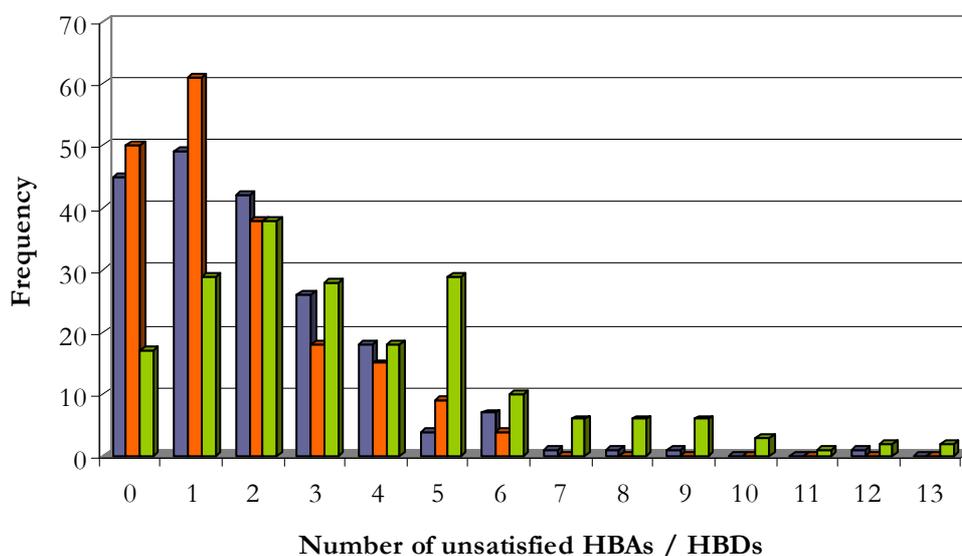


Figure 46: Illustration of the results. The number of unsatisfied HBAs/HBDs is given for ligand (blue), protein (orange), and complex (green).

Figure 47 shows an example structure (PDB code 1ELB), which has six unsatisfied HBAs/HBDs of both ligand and protein: The backbone oxygens have no donating ligand counterpart, the amide hydrogen lacks an accepting ligand group. On the other side, the ligand's donor/acceptor pair remains unsatisfied by the protein.

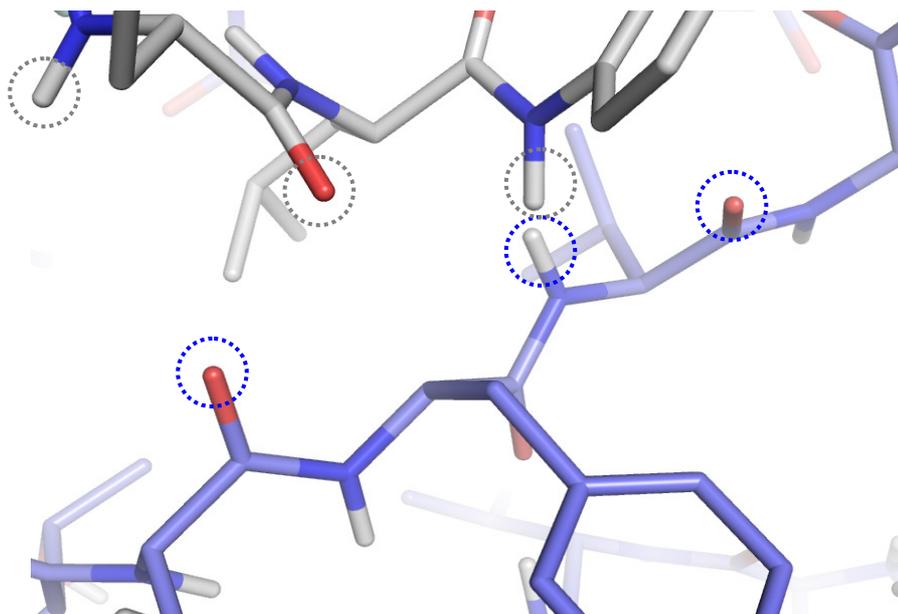


Figure 47: Example structure 1ELB with six unsatisfied HBAs/HBDs of both protein (blue circles) and ligand (grey circles).

Discussion

It could be shown that complexes with few unsatisfied hydrogen bonding groups are prevailing among the Cheng dataset. Although several outliers with a high count are present, a clear tendency towards a low frequency of unsatisfied HBAs/HBDs is observed. Water molecules were not present during this study. This was done intentionally: Most docking scenarios skip water molecules, due to the difficult handling; to draw a comparison between docking results and native binding modes, it is therefore reasonable to skip water molecules also in this crystal structure analysis. Of course, one could argue that water molecules may satisfy HBAs/HBDs, which otherwise would remain unsatisfied, but on the other hand, water molecules themselves are potential unsatisfied hydrogen bonding groups.

The HERMES descriptor uses a rather permissive definition of a hydrogen bond (see above). More common definitions use stricter parameters (D-H..A angle = 150°, maximum distance = 2.5 Å). The latter definition of course would lead to an increase in the number of unsatisfied HBAs/HBDs. The fact that protein-ligand complexes show donor-acceptor contacts which deviate from this strict definition implies that even though strong hydrogen bonding may not be observed, the donor/acceptor dipoles at least are in a proper mutual orientation. Such donor – acceptor contacts are regarded in the more permissive HERMES definition. This aspect can be extended also to ionic interactions: Visual inspection of complexes with a high count of unsatisfied HBAs/HBDs showed that often phosphate or carboxylate groups of the ligand are placed close to lysine or arginine residues, albeit in an angular orientation where even the more permissive HERMES descriptor regards the ligand acceptors / protein donors as unsatisfied. Nevertheless, interactions between negatively charged ligand atoms and positively charged protein residues, are of course favorable protein-ligand contacts. For such electrostatic interactions, also the HERMES descriptor fails.

These considerations urge to extend the question of unsatisfied HBAs/HBDs: If a complex has unsatisfied hydrogen bonding groups, are the donors/acceptors still in a polar environment (as described above) or are they surrounded by apolar groups? This is a decisive difference: In the former scenario, favorable polar interactions apart from “classical” hydrogen bonding are present, whereas the latter scenario represents generally unfavorable protein-ligand contacts. This question needs to be investigated in an affiliated study, which requires a sophisticated discrimination between (un-)favorable polar and apolar protein-ligand contacts. It is reasonable to expect that many of the complexes with a high count of unsatisfied HBAs/HBDs (see Table 14) show favorable polar donor – acceptor contacts, which are not reflected in a hydrogen bond descriptor. Still, in some cases, unfavorable contacts (donor – donor / acceptor – acceptor, see

Figure 47, or polar – hydrophobic) are observed: Apparently, the overall favorable protein-ligand interactions within such complexes are able to overcompensate “bad” contacts.

The HERMES hydrogen bond descriptor was successful in identifying a characteristic of the Cheng dataset: 92 % of all ligands and 93 % of all binding sites have equal or less than four unsatisfied HBAs/HBDs. As already done in the virtual screening described above, these descriptor values are valuable for the assessment of docking results: Binding modes or compounds not matching these values can be rejected, because they are rather unlikely to be true binders.

2.6 Summary and conclusion

The computational studies of HIV-1 IN have provided valuable information for inhibitor design. Docking experiments supported the hypothesis that the well-known diketo acid inhibitors enter the IN active site not as free ligands, but rather as metal complexes. Different complex species were docked, however, visual inspection of the binding modes as well as the docked energies suggest that monomeric complexes with a metal-to-ligand ratio of 1 are the preferred species. These results help to reveal the mechanism of action of this important class of IN inhibitors.

As a major part of the project, a new strategy towards IN inhibition was introduced: An alternative binding site, the dimerization interface of an IN catalytic core domain monomer, was explored for inhibitor design. The lack of structural data of the free monomer was overcome by extensive MD studies. An inspection of the dimerization interface had so far mostly been ignored. The starting structure for the simulation carried out here was an IN monomer extracted from the dimeric crystal structure. This means that the interface was removed from its protein binding partner and exposed to an aqueous environment instead of its protein binding partner. Such changes could lead to major structural rearrangements, which may not be fully detectable on accessible MD simulation timescales. However, the aim of the MD simulation studies was not a complete characterization of large structural rearrangements, but to explore conformational flexibility of the interface residues in the vicinity of the dimer state, as this may provide the expected target structure of inhibitory peptides (like YFLLKL) derived from the binding partner across the interface in the dimer. These inhibitory peptides form the basis for the IN dimerization inhibitor development.

In the MD simulation studies presented here, a groove and two cavities were revealed as important features at the IN dimerization interface. Groove and cavities are of transient nature: During the MD simulation, the groove shows a blocked and an unblocked state (depending on

the Lys173 movement, which also affects the smaller Glu87/Lys173 cavity1), whereas the Glu87/Lys103 cavity2 oscillates between formation and collapse. Snapshots derived from the MD simulation were used as protein input structures in a docking study with the known inhibitory peptide YFLLKL to reveal its potential binding mode, which had not been reported in the literature so far. The docking procedure showed that the peptidic ligand preferably binds to a protein conformation where the groove is blocked by Lys173 and the Glu87/Lys103 cavity2 is formed. The groove – cavity1 – cavity2 interface features a Y-shaped binding site which accommodates the inhibitory peptide well. The docked binding mode of YFLLKL matches the favorable positions for aliphatic/aromatic residues found by a hot-spot analysis, which further supports the validity of the binding mode. The next step was to address this protein conformation with small, non-peptidic molecules. Based on the fact that YFLLKL has a prevailing hydrophobic character and that aliphatic/aromatic hotspots were detected within the dimerization interface, the first strategy towards finding small-molecule interface binders was to create a pharmacophore model with hydrophobic features and shape constraints, aiming to find molecules with a good complementarity to the Y-shaped dimerization interface. Virtual screening yielded a total of 10 compounds, which all displayed good shape complementarity and favorable hydrophobic interactions. These compounds were tested in IN inhibition assays. Unfortunately, none of the compounds showed a reproducible inhibitory activity. Some doubts remain about the validity of the assay results: The use of BSA (which had to be added according to the standard protocol) was critical, since it is not unlikely that BSA interfered with the hydrophobic candidate compounds. Additionally, it has to be critically mentioned that virtual screening campaigns sometimes have very poor hit rates (1 - 2 %) ¹²⁵. This would mean that of about hundred compounds suggested by virtual screening, only one shows activity. Of course, testing of hundreds of compounds is hardly feasible in a non-industrial environment. Various reasons could lead to a failure of a virtual screening campaign, e.g. inexact structural information. Also, the computational methods like docking may reach their limits in virtual screenings where a large, diverse set of compounds needs to be assessed; given the great diversity of molecules in a screening database, it is a delicate task to differentiate between binders and nonbinders for a given target.

Although polar interactions were also observed in the docked binding modes, the hydrophobic character of the candidate compounds was prevailing, in analogy to the inhibitory hexameric peptide YFLLKL. Interestingly, the polar lysine residue was found to be non-essential in the parent dodecameric peptide TAYFLLKLAGRW, as a K7I mutation even lead to an increase in potency. Therefore, it is difficult to elucidate the crucial features of this peptide-protein interaction. YFLLKL was docked to the MD-derived snapshot; however, docking of the peptide

was performed using a rigid helical backbone, only the sidechains were allowed to move: YFLLKL is expected to have a helical shape, but deviations from the standard helix conformation are possible. Correct modelling of backbone flexibility during docking is hardly feasible, due to the delicate interplay of internal backbone hydrogen bonding and sidechain interactions and the high count of degrees of freedom. It is therefore conceivable that the hydrophobic character of the peptide is not the only factor for protein binding: Small deviations from the standard helical backbone orientation could allow also polar contacts between protein and ligand.

Accordingly, in the second virtual screening approach, the satisfaction of hydrogen bonding residues at the dimerization interface, which had not sufficiently been addressed by the overall hydrophobic compounds tested so far, was of major interest. Two pharmacophore models were employed, which retrieved several hundred hit molecules. However, docking of these molecules showed that still many hydrogen bonding groups of the protein remained unaddressed by the ligands. A systematic study of the Cheng dataset, comprising 195 protein-ligand complexes, showed that a high number of unsatisfied hydrogen bonding groups in both protein and ligand are rarely observed. Instead, in over 90 % of all cases, ligands and proteins have equal or less than four unsatisfied hydrogen bonding groups. Accordingly, for docking pose analysis of the pharmacophore hits, a threshold of four unsatisfied HBAs/HBDs was introduced, which led to rejection of over 90 % of the docked binding modes. Eventually, after visual inspection, only eight molecules were selected as candidate compounds for further testing (results pending). This small “yield” underlines the difficulties in finding interface binders: The IN dimerization interface is a peculiar target with frequently alternating basic, acidic, and hydrophobic residues. It is not a well-ordered binding site with continuous hydrophobic areas and distinct hydrogen bond donors or acceptors. Other protein-protein interfaces show such well-ordered binding sites. Accordingly, the peculiarity of the IN dimerization interface, in addition to the delicate task of disrupting protein-protein interactions at all, makes the development of IN dimerization inhibitors very challenging. Still, it is a worthwhile goal: Small-molecule dimerization disruptors would pose a whole new class of IN inhibitors.

Due to the lack of synthesis capacities, the aim of the project was to find purchasable compounds from common vendor databases. Given the peculiarity of the binding site, it could be argued that instead of precast compounds, tailor-made molecules are the better option. Also for this option, the studies shown here yield valuable information: Fragments like triaminobenzene (as shown in compound **DI-15**) or urea-pyrimidine (**DI-16**) can be used as a starting point for the synthesis of dimerization inhibitor candidates.

3. *Legionella pneumophila* Macrophage Infectivity Potentiator

3.1 Legionnaires' Disease and MIP

The aim of this project was to design small-molecule inhibitors of MIP, which could eventually lead to novel therapeutics against Legionnaires' Disease. In the present chapter, the characteristics of this infectious disease are outlined, followed by an introduction to its major pathogenic factor, the Macrophage Infectivity Potentiator, MIP.

The Legionnaires' Disease (named after an outbreak of pneumonia during a convention of the American Legion in Philadelphia in July 1976) is an infectious disease causing pneumonia. The causative agent of Legionnaire's Disease is the ubiquitous aquatic *Legionella pneumophila*. This gram-negative bacterium can also cause the so-called Pontiac Fever, which is less severe and resembles influenza. Both Legionnaires' Disease and Pontiac Fever are summarized by the term Legionellosis¹²⁶. Legionellosis infection is transmitted by inhaling *Legionella*-containing aerosols. The origins of such aerosols include air-conditioning systems, room-air humidifiers, whirlpools or even showers, i.e. systems with stagnant, warm water. 20 % of all Legionellosis cases in Europe are considered to be travel-related (hotels, ships)¹²⁷. However, *Legionella* infections are not solely originated in destinations with a warm climate (e.g. southern Europe), but also in central Europe: Outbreaks have been reported in the Netherlands (1999; 300 infections, 32 fatalities) and the United Kingdom (2002; 172 infections, 7 fatalities). Overall, in Europe there are an estimated 4,600 infections per year, for the U.S. the number of cases is estimated between 8,000 and 18,000. Due to underreporting and underutilization of diagnostic testing, accurate information is not available¹²⁸. Data of Legionnaires' Disease outbreaks in developing countries are rare. Accordingly, it is acknowledged that the actual morbidity and mortality of Legionnaires' Disease is vastly underrated. For the future, the impact of *Legionella* infections is expected to grow, since the climate change may foster *Legionella* spread. These facts urge to develop novel and effective drugs against Legionnaires' Disease and Legionellosis in general.

Current treatment of Legionnaires' Disease comprises quinolones like levofloxacin or macrolides like clarithromycin, targeting the bacterial gyrase and ribosomes, respectively. However, new targets need to be exploited for an effective antibacterial therapy. Once *Legionella* has entered the lung, it is phagocytosed by alveolar macrophages, but remains intact. The bacterium can multiply within the macrophage by inhibiting the phagosome-lysosome fusion. Eventually, this leads to the death of the macrophage and the release of new bacteria¹²⁹. Among the pathogenic factors of

Legionella pneumophila, the extracellular macrophage infectivity potentiator (MIP) is of major importance. Its exact role in infection has been discovered only recently: MIP enables the bacterium to transmigrate through lung epithelial cells to access its macrophage host cell. Within the extracellular matrix of the lung tissue, MIP binds to collagen IV. It has been hypothesized that MIP interconverts proteolytically insensitive *cis* prolyl bonds within the collagen protein to proteolytically sensitive *trans* prolyl bonds. Then, in a concerted action of MIP and a serine protease, the extracellular matrix is degraded, yielding access for the bacterium to its macrophage host cell (see Figure 48)¹³⁰.

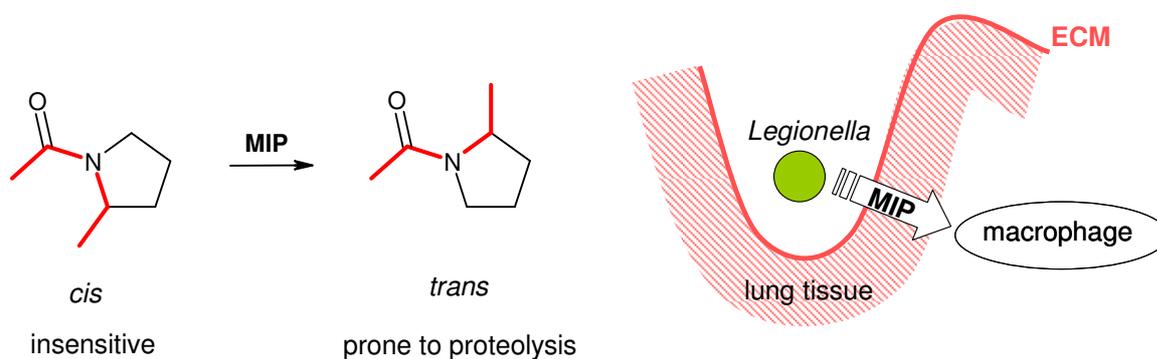


Figure 48: MIP-catalyzed interconversion of prolyl bonds (left) and MIP-assisted transmigration of *Legionella* through the extracellular matrix (ECM), finally leading to macrophage infection.

Accordingly, MIP is a very attractive target for anti-*Legionella* treatment: Inhibition of this pathogenic factor would significantly lower macrophage infection. Up to date, MIP is not targeted in Legionellosis treatment and only one single macrocyclic inhibitor is known, which itself is not an appropriate drug candidate (see below). The goal of the present work is to develop a small-molecule inhibitor of MIP, which eventually could serve as a lead structure for the development of a MIP-tailored Legionellosis treatment.

3.2 MIP and FKBP12: Peptidyl prolyl *cis/trans* isomerases

MIP belongs to the family of peptidyl prolyl *cis/trans* isomerases (PPIases), which catalyze the isomerization of *cis*-proline to *trans*-proline. PPIases are also common in mammals, where they play an important role in immunological processes. Hence, they are termed immunophilins. One of the most prominent immunophilins is FKBP12, named after its ligand, the macrolide FK506 (better known as tacrolimus). Tacrolimus and other macrolidic FKBP12 binders like rapamycin (sirolimus, see Figure 49) are immunosuppressants, which are indicated in post-transplantational therapy or autoimmune diseases, to name a few^{131,132}.

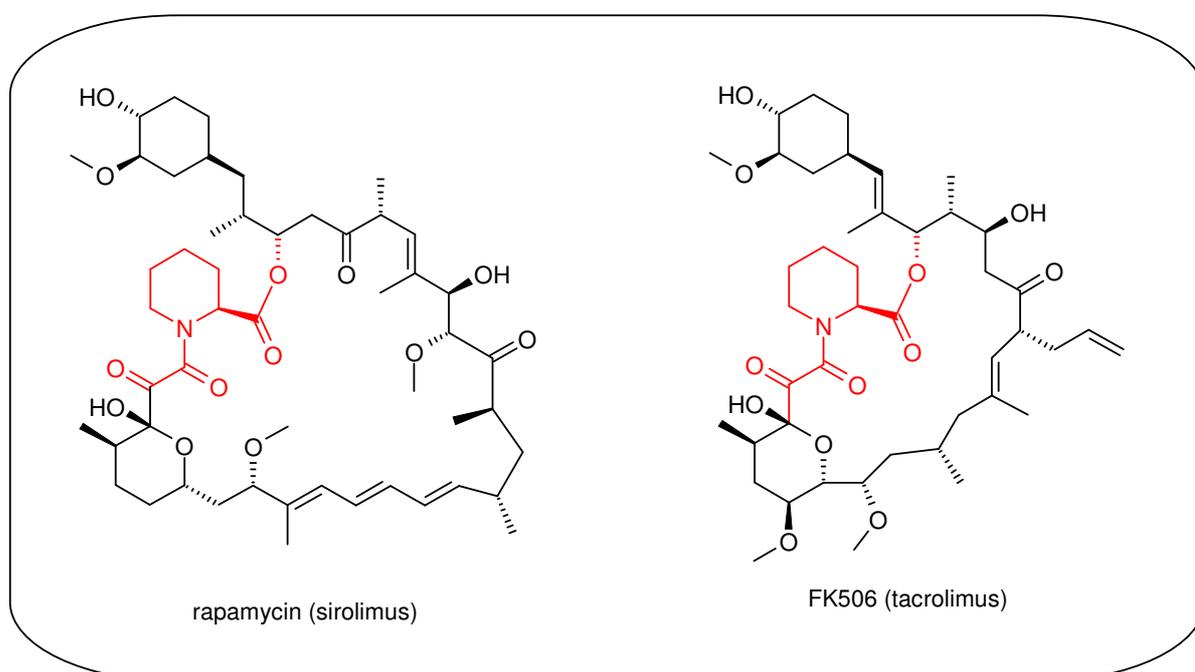


Figure 49: FKBP12 ligands rapamycin and FK506. The common pipecoline substructure (red) binds to FKBP12's active site, whereas the major portion of the macrocycle binds to a second cellular protein, yielding a ternary complex, which exerts an immunosuppressive action.

Both tacrolimus and rapamycin share a common substructure, the FKBP binding domain. This pipecoline moiety is anchored in a hydrophobic pocket within the FKBP12 active site. The second domain is unique to the particular macrolides and determines the second binding protein. In the case of tacrolimus, this second protein is calcineurin, leading to a ternary complex FKBP12-tacrolimus-calcineurin. For rapamycin, it is the mammalian target of rapamycin (mTOR), yielding the FKBP12-rapamycin-mTOR complex¹³³. In both cases, these ternary complexes enter (separate) signalling cascades, eventually leading to immunosuppression.

For the binding of macrolides like rapamycin to FKBP12, only the FKBP12 binding domain is relevant. This was proven by systematic studies on rapamycin derivatives, lacking the mTOR binding domain¹³⁴. Formation of the immunosuppressive ternary complex is thereby averted, so these derivatives are termed non-immunosuppressive agents (NIAs, see Figure 50). NIAs are promising candidates for the treatment of neurodegenerative diseases like Alzheimer's or Parkinson's Disease. By inhibiting FKBP12, they exert neuroprotective effects. Interestingly, increased FKBP12 concentrations have been found in brain tissue. Still, the underlying mechanism of the neuroprotective activity needs to be revealed^{135,136,137}.

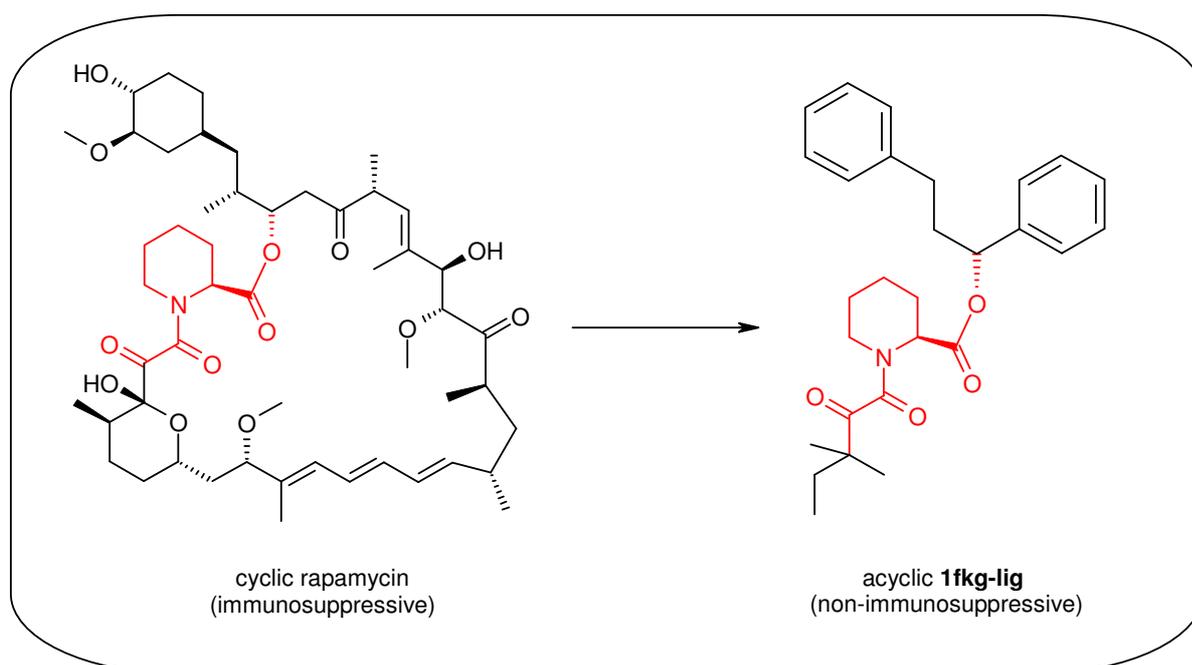


Figure 50: Development of non-immunosuppressive agents (NIAs) against FKBP12: The rapamycin-derived 1fk-g-lig (ligand in PDB structure 1FKG)¹³⁸ still bears the FKBP12-binding pipercoline domain, but lacks the mTOR binding domain; formation of the ternary complex with subsequent immunosuppression is thereby averted.

Binding of NIAs to FKBP12 has been studied well and crystallographic data of protein-ligand complexes are available via the PDB. As mentioned earlier, both FKBP12 and MIP are PPIases; they belong to the same family of enzymes. Indeed, rapamycin binds not only to FKBP12, but also to MIP. However, rapamycin is not an option for Legionnaires' Disease treatment, since it would also trigger an immunosuppressive response, which of course is highly undesired in bacterial infections.

3.2.1 MIP structures

MIP is a homodimeric surface protein localized at the membrane of *Legionella pneumophila*. Each of the 22.8 kDa monomers consists of a C-terminal domain (which resembles FKBP12) and an N-terminal domain, with a long connecting helix in between. The N-terminal domain mediates homodimerization. The C-terminal domain consists of six β -strands, which form an antiparallel sheet. A short helix is located across this sheet (Figure 51). Right between this helix and the interior side of the sheet, the hydrophobic active site of MIP is located.

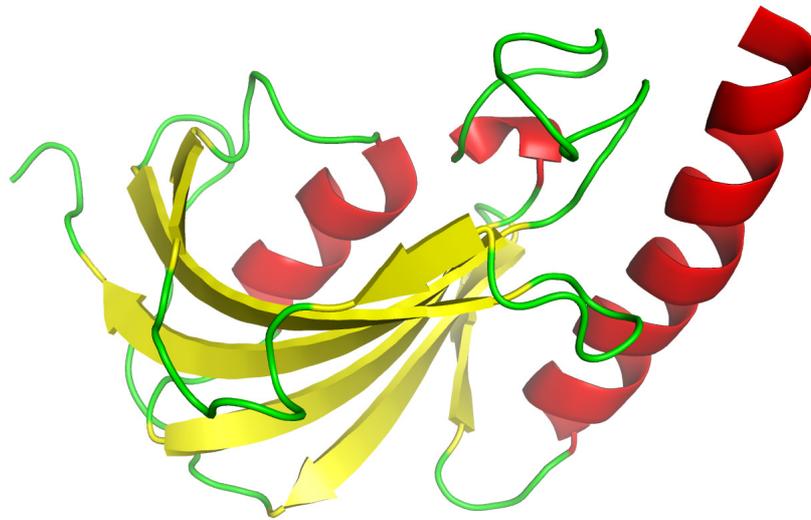


Figure 51: Secondary structure elements of the MIP C-terminal domain (PDB code 2VCD). The active site is located between the sheet and the short helix. The long helix on the right is the junction to the N-terminal domain (not present).

The active site is characterized by its high hydrophobicity. It is formed by Tyr55, Phe65, Asp66, Ala75, Phe77, Val82, Ile83, Trp86, Tyr109, and Phe126. Trp86 forms the ground of a pocket, which is limited by Tyr55, Ala75, Phe77, Val82, Ile83, and Phe126. At the outer rim of this pocket, Phe65, Asp66, and Tyr109 are located. The latter residue is part of a loop region adjacent to the active site (see Figure 52). Currently, three MIP structures are available via the PDB: One X-ray structure (**1FD9**¹³⁹), an NMR-derived solution structure (**2UZ5**¹³²), and an NMR structure with rapamycin bound to the active site (**2VCD**¹³²).

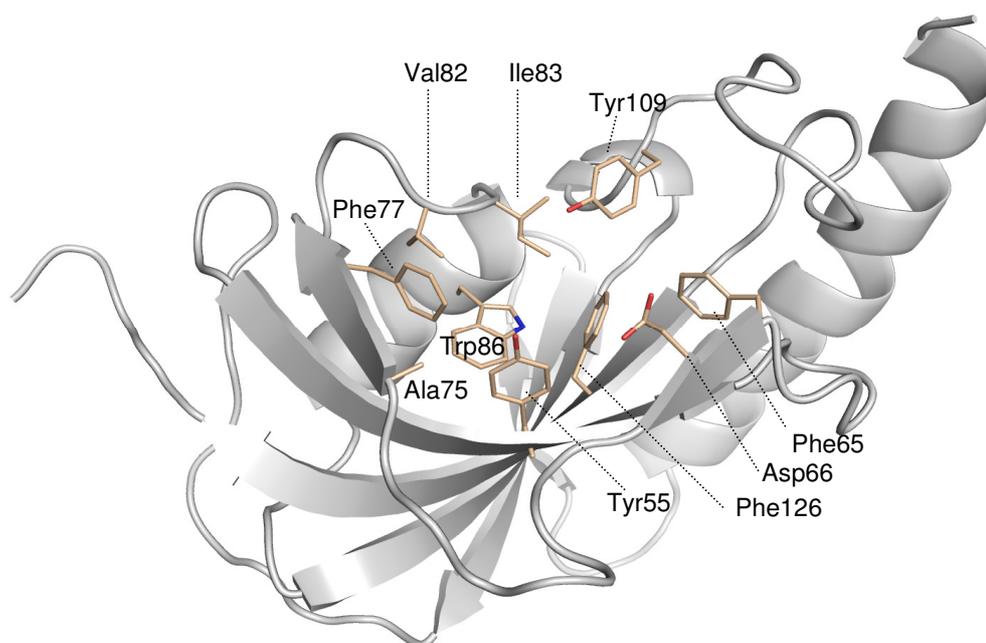


Figure 52: MIP active site residues and secondary structure elements. Val82 and Ile83 are part of a short connecting loop, whereas Tyr109 is located on a larger loop adjacent to the active site.

Mutual superposition of the different MIP structures reveals interesting differences: Superposition of the *apo* structures **1FD9** and **2UZ5** shows obvious displacements of the active site residue positions (see Figure 53), with the most striking difference occurring at Tyr109, located on the loop adjacent to the binding site. In **2UZ5**, Tyr109 can hardly be regarded as an active site residue anymore.

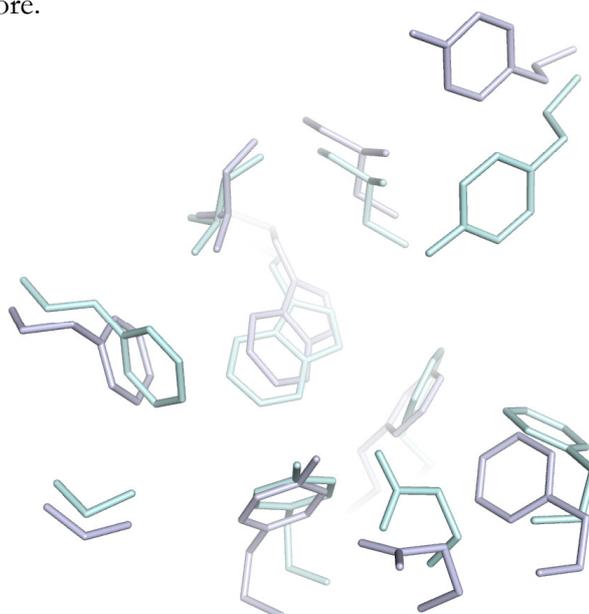


Figure 53: Superposition of the MIP *apo* structures **1FD9** (X-ray, cyan) and **2UZ5** (first out of ten NMR ensemble structures, blue).

The third MIP structure, **2VCD**, is the NMR-derived structural ensemble of the MIP-rapamycin complex. Superposition of **2VCD** with both **2UZ5** and **1FD9** revealed that the complex solution structure is more similar to the crystallographic *apo*-structure **1FD9** than to NMR-derived **2UZ5** (see Figure 54). The complex structure shows a tighter active site: Ile83, Tyr109, and Phe126 are tilted towards the active site center.

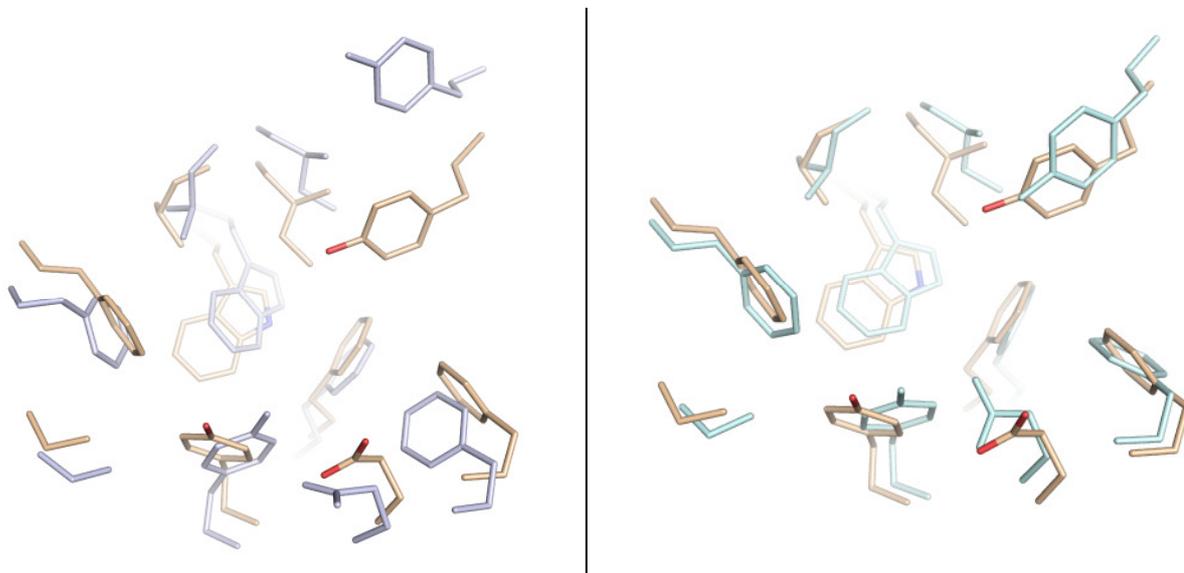


Figure 54: Superposition of complex structure **2VCD** (wheat) with uncomplexed structures **2UZ5** (blue, left) and **1FD9** (cyan, right). Tyr109 shows a distinct incongruence among the three structures.

In general, the available structural data of MIP show diverse protein conformations, especially in the case of Tyr109. The question remains which of these MIP structures should serve as protein input for structure-based ligand design. Since a complex structure is expected to render a ligand-bound active site conformation best, **2VCD** will be used for further studies.

3.2.2 FKBP12 structures

Several FKBP12 structures are available from the PDB repository, comprising crystallographic and solution structures with and without ligands. Examples for uncomplexed structures are **2PPN**¹⁴⁰(X-ray) and **1FKT**¹⁴¹ (NMR), whereas **1FKB**¹⁴² represents a crystallographic complex. Superposition of crystal structure **2PPN** with the solution structure **1FKT** shows differences in the orientation of Asp37: In the latter structure, this residue points towards the catalytic center. All other residue positions match well (Figure 55).

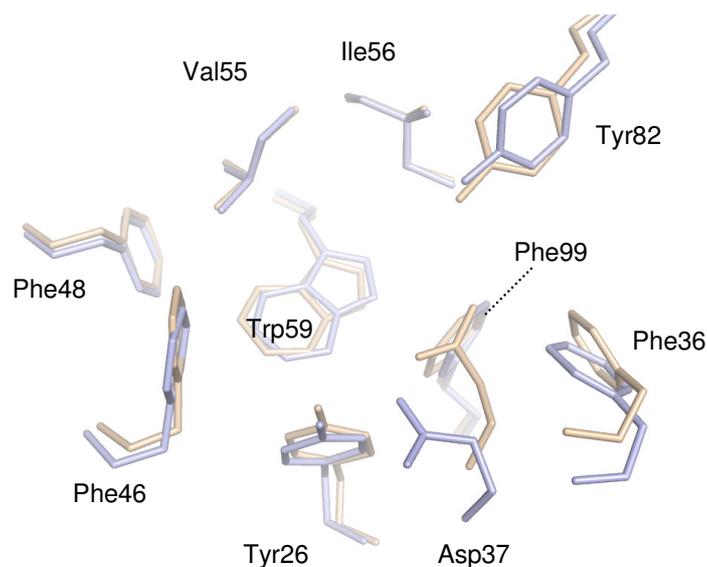


Figure 55: Superposition of the FKBP12 *apo* structures 2PPN (X-ray, blue) and 1FKT (NMR, wheat).

Superposition of the crystallographic *apo* structure **2PPN** with the rapamycin complex structure **1FKB** shows that the overall arrangement of the active site residues is similar. In the case of **2PPN**, Asp37 as well as Val55/Ile56 are somewhat more tilted towards the active site center, but overall, the residues match well (see Figure 55).

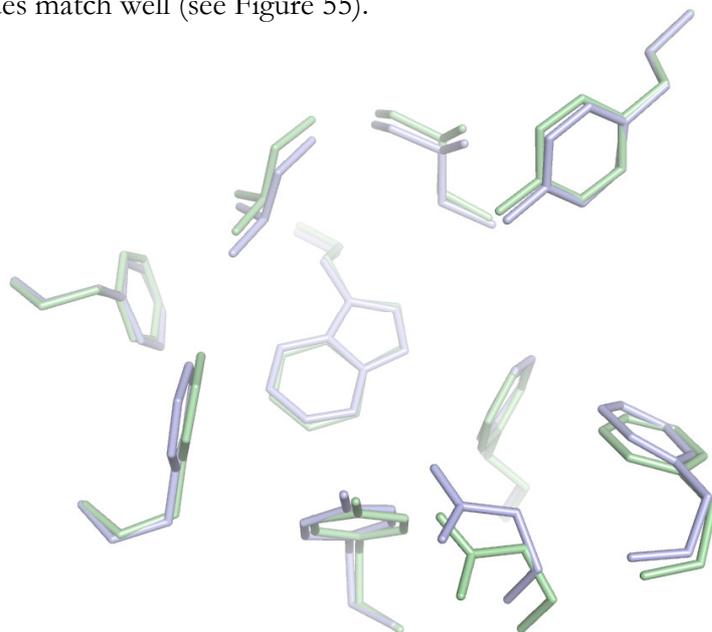


Figure 56: Superposition of 2PPN (X-ray, blue) and the complex structure 1FKB (X-ray, green).

To sum up, NMR and crystallographic data of both complexed and uncomplexed FKBP12 show consistent structural arrangements within the catalytic center. Apart from slight differences, the active site residues display a good congruency (in contrast to MIP structures).

3.2.3 Comparison of MIP and FKBP12

The homologous PPIases MIP and FKBP12 share common binding site residues, which can be revealed by superposition of the respective PDB structures. In Table 15, the active site residues of MIP and their respective FKBP12 counterparts are listed.

MIP	FKBP12
Tyr55	Tyr26
Phe65	Phe36
Asp66	Asp37
<i>Ala75</i>	<i>Phe46</i>
Phe77	Phe48
Val82	Val55
Ile83	Ile56
Trp86	Trp59
Tyr109	Tyr82
Phe126	Phe99

Table 15: MIP active site residues and their respective FKBP12 counterparts.

Conservation is observed in all cases, except for Ala75 / Phe46.

Superposition of MIP and FKBP12 shows conserved active site residues, except for Ala75 (MIP) and Phe46 (FKBP12, see Figure 57). Nevertheless, the orientation of Phe46 and Phe48 in FKBP12 creates a similar hydrophobic environment within the binding pocket as does Phe77 in MIP. Interestingly, superposition of **1FD9** and **2PPN** shows a higher similarity of the active site shape between MIP and FKBP12 than the superposition of **1FD9** and **2UZ5**, both being exclusively MIP structures. Superposition of the rapamycin complex structures **2VCD** and **1FKB** reveals a clear distinction between both PPIases: Tyr109 (MIP) protrudes much more into the hydrophobic pocket than its FKBP12 counterpart. This difference has a major impact on ligand binding. To sum up, MIP and FKBP12 have common binding site features and an overall shape similarity (even more similar than between the two free MIP structures). Differences are observed in the position of Tyr109/Tyr82. However, Tyr109 has been found very flexible among MIP structures.

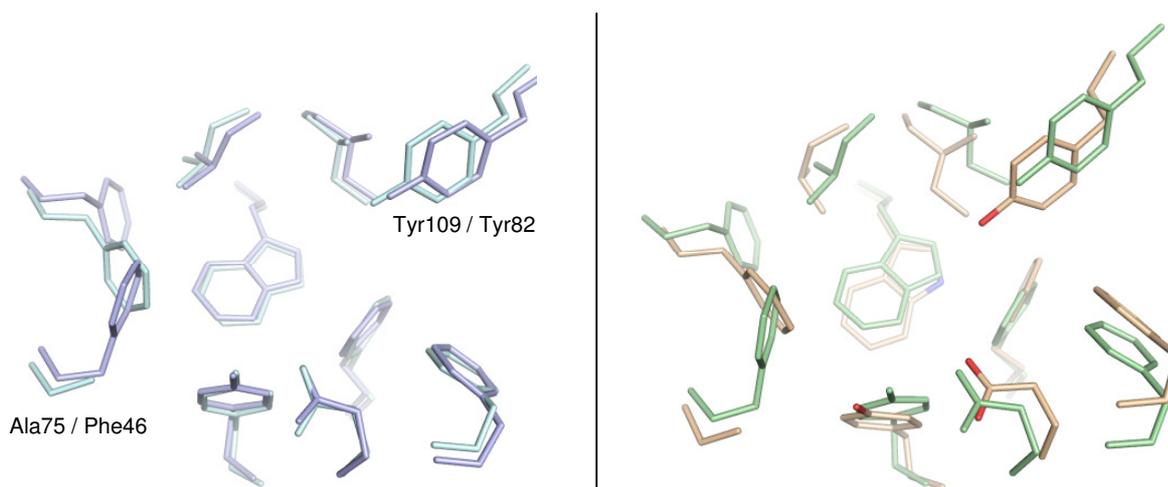


Figure 57: MIP - FKBP12 superposition. Left: *Apo* structures 1FD9 (MIP, cyan) and 2PPN (FKBP12, blue); Ala75 of MIP has Phe46 as FKBP12 counterpart. Right: Rapamycin complexes 2VCD (MIP, wheat) and 1FKB (FKBP12, green); Tyr109 of MIP protrudes significantly further into the active site pocket.

Rapamycin is a common ligand for both PPIases. In both cases, the pipecoline moiety is anchored in the active site pocket. The macrocyclic portion is mainly exposed, ready to bind a second cellular protein like mTOR. Here, the focus was put on the pipecoline anchor group; possible differences were revealed by comparing the FKBP12-rapamycin structure **1FKB** with the MIP-rapamycin complex **2VCD**. Figure 58 shows how the pipecoline moiety is accommodated by the FKBP12 pocket: Favorable hydrophobic interactions are complemented by hydrogen bonding between Ile56 and the ester C=O as well as Tyr82 and the amide C=O (however in a less favorable out-of-plane geometry).

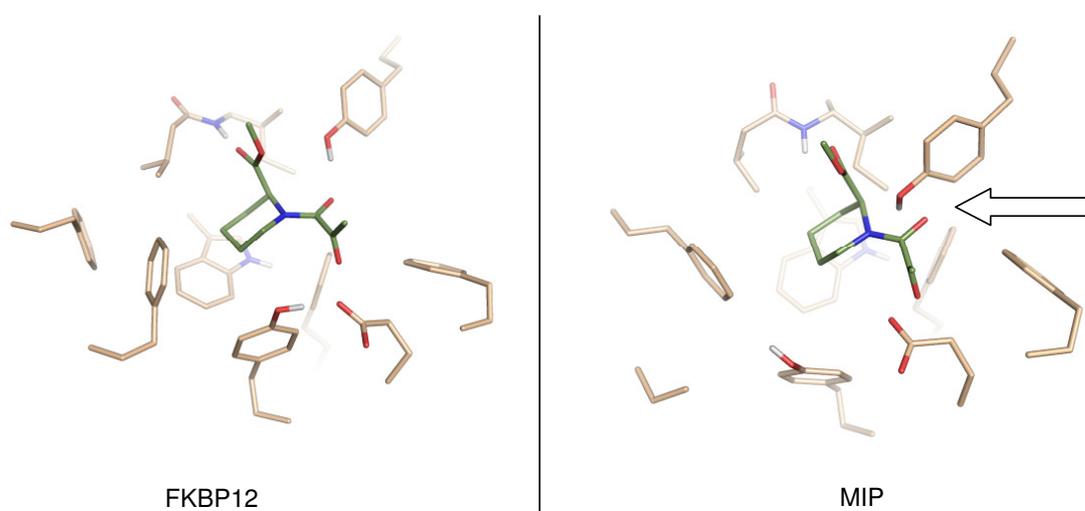


Figure 58: Left: Rapamycin binding mode to FKBP12: The pipecoline moiety is anchored in the active site pocket and addresses Ile56 and Tyr82. Right: Superposition of this binding mode with the MIP active site: Clashes between Tyr109 and the amide C=O impede a similar binding motif in MIP.

Transferring this binding mode to MIP, clashes between the amide group of the pipercoline anchor and Tyr109 would occur (Figure 58). Hence, a similar binding motif is impeded. Indeed, the MIP-rapamycin complex shows a different orientation within the active site pocket. **2VCD** is a structural ensemble comprising 16 single structures. In Figure 59, two structures which show discrepancies in the pipercoline binding mode are depicted.

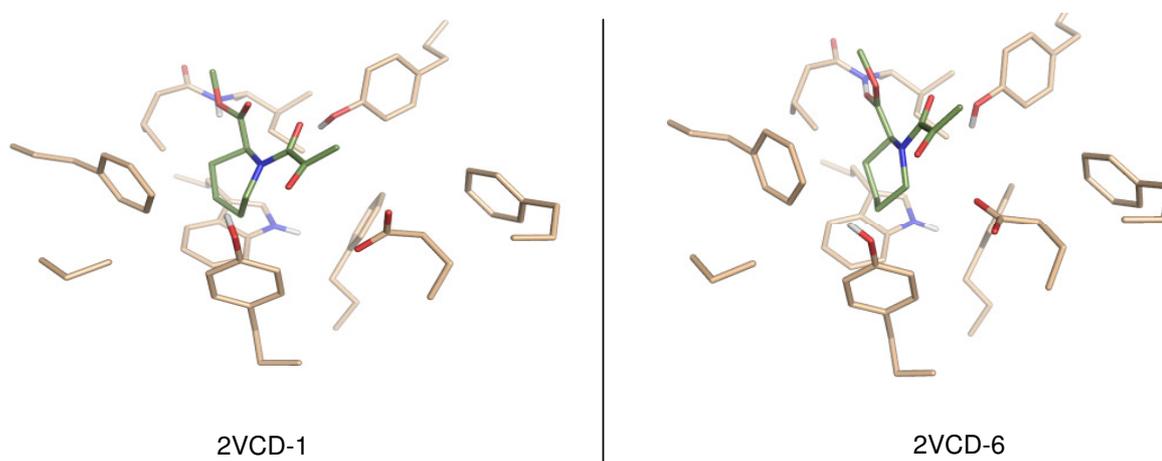


Figure 59: Rapamycin binding mode in MIP: Two single structures from the 2VCD ensemble are depicted, showing discrepancies in the pipercoline binding mode with respect to the ester group.

In contrast to the FKBP12 binding mode, the pipercoline ring in MIP is in a kind of upright position, due to the protruding Tyr109, which does not act as hydrogen bond donor anymore. According to the NMR-derived, fully protonated structure, the Tyr-OH of Tyr55 points away from the diketo function, although hydrogen bonding would be feasible. Planarity of the diketo function is maintained (cf. FKBP12), whereas the amide function has lost its planarity in **2VCD-6**. The discrepancy between the two single structures lies in the orientation of the ester: In **2VCD-1**, the C=O group points towards the surrounding, whereas in **2VCD-6**, it is directed to Ile83 (similar motif as in FKBP12).

In contrast to X-ray crystallography, NMR experiments reflect the structural behaviour of proteins in their physiological environment (i.e. in solution). The protein solution structure ensemble is elucidated by force field calculations using NMR-derived restraints; force fields are parameterized for proteins, nucleic acids, or small molecules (see Chapter 5.1): A macrocyclic ligand like rapamycin poses a delicate task for the force field calculations, since it is neither a small molecule nor a protein / nucleic acid. In other words, the force field may fail to assign proper parameters, which would lead to limitations in the accurate elucidation of the MIP-rapamycin complex structure.

3.2.4 Analysis of the MIP structural ensemble 2VCD

As mentioned earlier, the rapamycin complex **2VCD** is a structural ensemble of 16 single structures. In general, the active site residues show a similar arrangement throughout all structures. However, there are some differences which have influence on the active site shape. For better feasibility of the structure-based ligand design, only one out of the 16 structures is considered: This structure should be rather permissive in terms of ligand accommodation, i.e., the binding site should not be too narrow. Accordingly, the structural ensemble **2VCD** was searched for single structures which have a rather dilated active site pocket. For this purpose, a triangle spanned by residues Asp66 (carboxylate oxygen adjacent to the active site pocket), Phe77 (distant ring carbon), and Tyr109 (hydroxyl oxygen, see Figure 60) was defined, and the area was trigonometrically computed; the area sizes were taken as a criterion for pocket dilation.

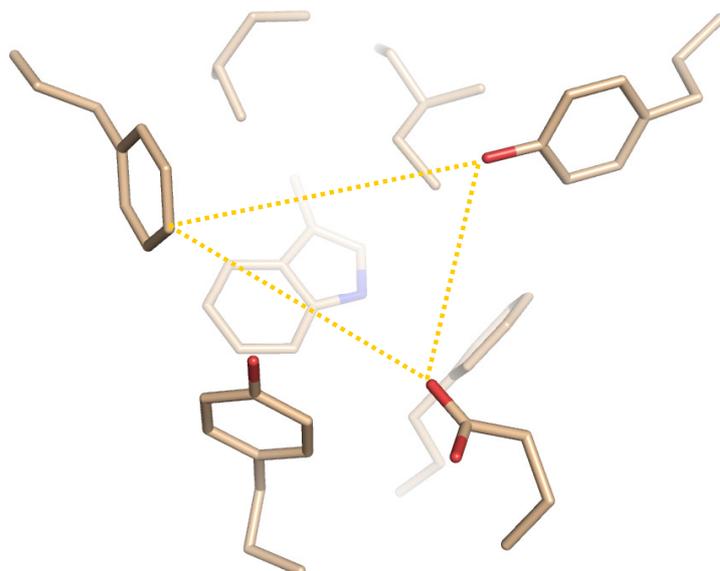


Figure 60: Triangle spanned by Asp66, Phe77, and Tyr109 to measure the active site areas of the 2VCD ensemble structures.

Table 16 shows the broad range of distances between the respective residues: Asp66 and Phe77 have distances ranging from 6.44 Å (**2VCD-6**) up to 9.03 Å (**2VCD-7**), Asp66 and Tyr109 show a similar behaviour with distances between 4.65 Å (**2VCD-10**) and 8.54 Å (**2VCD-4**), and the distances between Phe77 and Tyr109 go from 7.07 Å (**2VCD-8**) up to 8.15 Å (**2VCD-12**). These variations are reflected by the different triangle areas, which serve as a measure of dilatation of

the active site pocket. **2VCD-4** has the largest area of about 29 Å², followed by **2VCD-15** and **2VCD-16**. The smallest area is observed in case of **2VCD-6** (16 Å²).

	d(Asp66–Phe77) [Å]	d(Asp66–Tyr109) [Å]	d(Phe77–Tyr109) [Å]	A(triangle) [Å ²]
2VCD-1	7.77	7.38	7.82	25.34
2VCD-2	8.51	5.85	7.95	22.40
2VCD-3	7.13	6.25	7.58	20.75
2VCD-4	8.17	8.54	7.85	28.92
2VCD-5	7.00	5.71	7.35	18.76
2VCD-6	6.44	4.96	7.62	15.85
2VCD-7	9.03	5.20	7.23	18.79
2VCD-8	7.26	6.07	7.07	19.69
2VCD-9	7.08	6.25	7.36	20.32
2VCD-10	8.49	4.65	7.43	17.23
2VCD-11	7.09	5.54	7.26	18.33
2VCD-12	7.29	5.61	8.15	19.94
2VCD-13	7.27	7.33	7.91	24.27
2VCD-14	8.77	6.56	7.24	23.29
2VCD-15	8.64	7.38	8.10	27.65
2VCD-16	8.99	7.16	7.94	27.16

Table 16: Triangle areas and distances between the triangle-spanning residues Asp66, Phe77, and Tyr109.

Accordingly, **2VCD-4** represents a protein conformation within the structural ensemble which is not too narrow, but permits accommodation of larger ligands also. **2VCD-4** will be employed as protein input for the forthcoming studies.

3.3 FKBP12 ligands as MIP inhibitors

As mentioned above, rapamycin is not a suitable MIP inhibitor, due to its immunosuppressive action. However, non-immunosuppressive rapamycin derivatives (NIAs) are available as FKBP12 ligands, sharing the pipercoline moiety as binding site anchor. Given the similarity of MIP and FKBP12, and the fact that they have rapamycin as common ligand, it is reasonable to assume that rapamycin-derived NIAs could also bind to MIP.

The FKBP12 complex structure **1FKG**¹³⁸ shows the binding mode of a non-immunosuppressive rapamycin derivative (**1fkg-lig**, $K_i = 10$ nM), depicted in Figure 61. It is obvious that the pipercoline substructure of this ligand is almost congruent with the pipercoline moiety of rapamycin. Again, it serves as an anchor for the whole ligand. The phenyl rings are rather solvent-exposed and hydrophobic interactions are hardly visible. Re-docking of **1fkg-lig** using AutoDock3.0 was successful: A well-clustered, near-native binding mode (RMSD = 1.05 Å) could be obtained as top-ranked result.

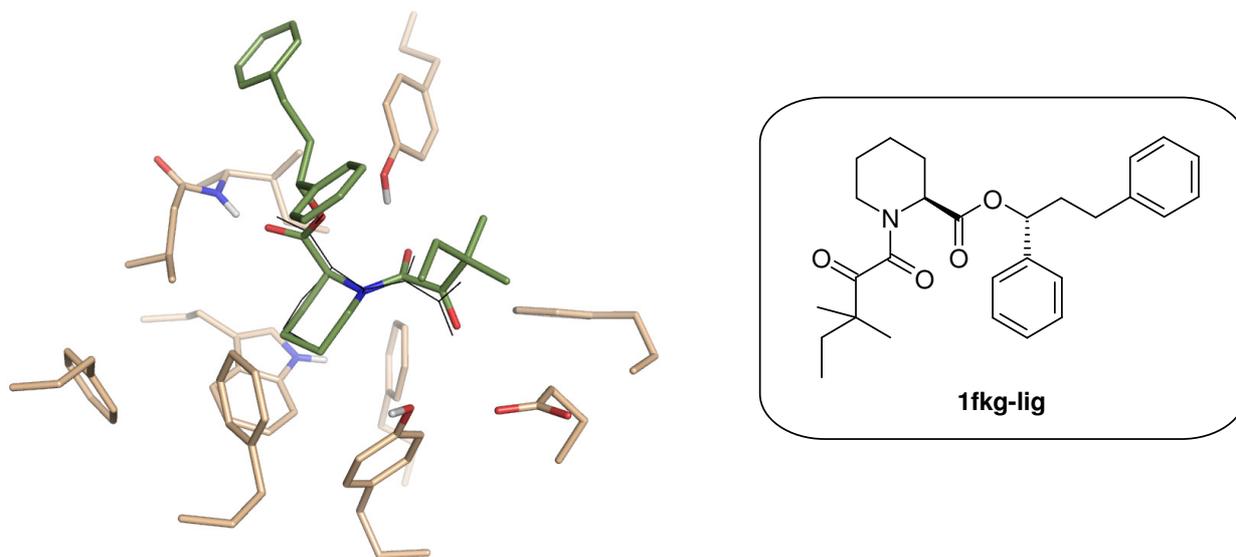


Figure 61: Binding mode of **1fkg-lig**, a FKBP12 ligand. The pipercoline moiety of rapamycin is depicted in black lines, the congruency of both substructures is apparent.

Also in MIP, rapamycin places its pipercoline system into the active site pocket. To reveal whether **1fkg-lig** is able to bind in a similar fashion, docking studies with MIP as target structure were performed.

Docking of 1fkg-lig to MIP

Docking studies were performed using AutoDock3.0. **2VCD-4** was used as protein input; rapamycin and water molecules as well as all hydrogens were removed, before polar hydrogens were re-added with the *protonate* tool. Finally, Kollman united-atom partial charges and solvation parameters (derived from the method of Stouten *et al.*) were assigned by the inhouse-tool *add_chrgsol*. The active site was defined by a box with 60·60·60 grid points, centered between Phe77 and Tyr109. The grid spacing was set to 0.375 Å. Subsequently, the interaction maps were calculated with the *autogrid* module. The ligand **1fkg-lig** was protonated and saved in mol2 format. Gasteiger-Marsili charges were added using Sybyl8.0. The conversion to the pdbq format required as input for AutoDock3.0 was done with the *autotors* module. The docking protocol used an initial population of 50 individuals, a maximum number of $3.0 \cdot 10^6$ energy evaluations, a mutation rate of 0.02, a crossover rate of 0.80, and an elitism value of 1. The Solis and Wets local search parameters were set as follows: The probability of performing a local search on an individual was set to 0.06, a maximum number of 300 iterations was performed, and the number of consecutive successes or failures before changing the size of the local search space was 4. 100 GA runs were carried out. The resulting binding poses are shown in Figure 62.



Figure 62: Docked binding modes of 1fkg-lig in the active site of MIP. Left: Top-ranked, but rarely observed binding mode. The pipercoline moiety is harbored by the binding pocket; the rapamycin structure is shown in grey lines. Right: Frequently observed docking result, where the pocket is occupied by a phenyl ring.

Docking yielded a top-ranked binding mode, where the pipercoline ring is harbored by the active site pocket. Compared to the rapamycin binding mode, the ring position is translated, but has a similar orientation. The ester carbonyl oxygen accepts a hydrogen bond donated by Tyr55. The phenyl rings are rather exposed, but match the hydrophobic chain positions in the rapamycin binding mode. A certain similarity to the binding mode of **1fkg-lig** within FKBP12 is

recognizable. However, this top-ranked binding mode ($E_{\text{docked}} = -10.03$ kcal/mol, cf. with $E_{\text{docked}} = -11.80$ kcal/mol in the case of FKBP12; the docked energy E_{docked} is the sum of the intermolecular energy and the internal energy of the ligand) is observed only in 2 out of 100 runs. A more frequently found docked binding mode (10 out of 100 runs) shows occupation of the active site pocket by the phenyl ring, whereas the pipercoline moiety is placed outside the pocket. Here, no hydrogen bonding is observed. Placing the phenyl ring as anchor group apparently prevents close contacts between protein and ligand, which are unavoidable when the pipercoline ring is placed within the active site pocket. The reason for that is the protruding residue Tyr109 (see Figure 58). On the one hand, the poor clustering of the docking results indicates that **1fkg-lig** with its *N*- α -ketoacyl pipercoline anchor fragment does not readily find a favorable binding mode in the MIP active site; furthermore, the docked energy of the top-ranked, but poorly clustered result is significantly less favorable than in the case of FKBP12. On the other hand, the *N*- α -ketoacyl pipercoline moiety is apparently able to occupy the MIP active site, as seen in the top-ranked binding mode and in the MIP-rapamycin structure **2VCD**. Also, the structural diversity (especially concerning Tyr109), as seen in **2UZ5**, **1FD9**, and **2VCD**, permits the consideration that the active site residues could adapt to the ligand and thus allow binding of **1fkg-lig**. Accordingly, it is of importance to test **1fkg-lig** against MIP, in order to gain final information whether the *N*- α -ketoacyl pipercoline moiety is a suitable anchor fragment for non-immunosuppressive ligands or not.

Besides ligands like **1fkg-lig**, further NIAs are described in the literature. Compound **MI-1**¹⁴⁵ (FKBP12: $K_i = 230$ nM) bears also a pipercoline system, albeit with a sulfonamide instead of an *N*- α -ketoacyl group, leading to a difference in the steric demands (see Figure 63). Again, docking studies were performed to assess potential MIP binding of **MI-1**.

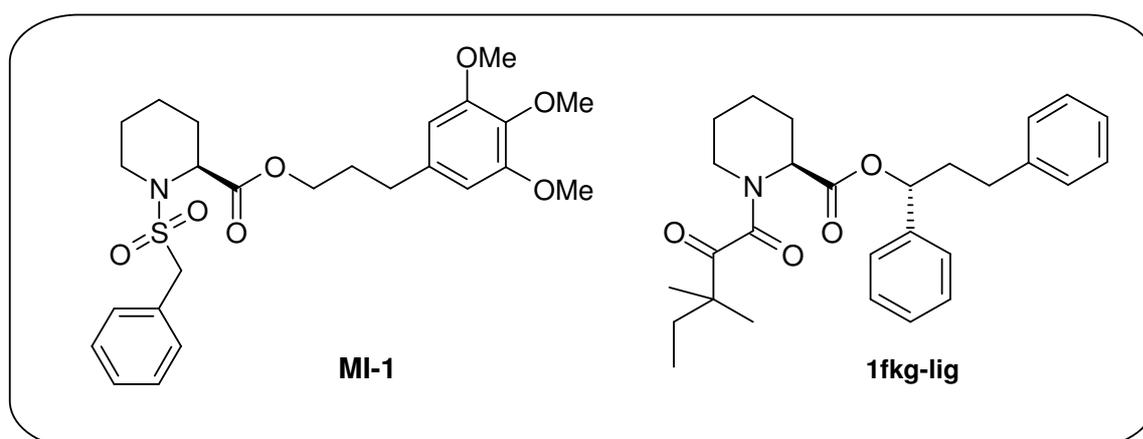
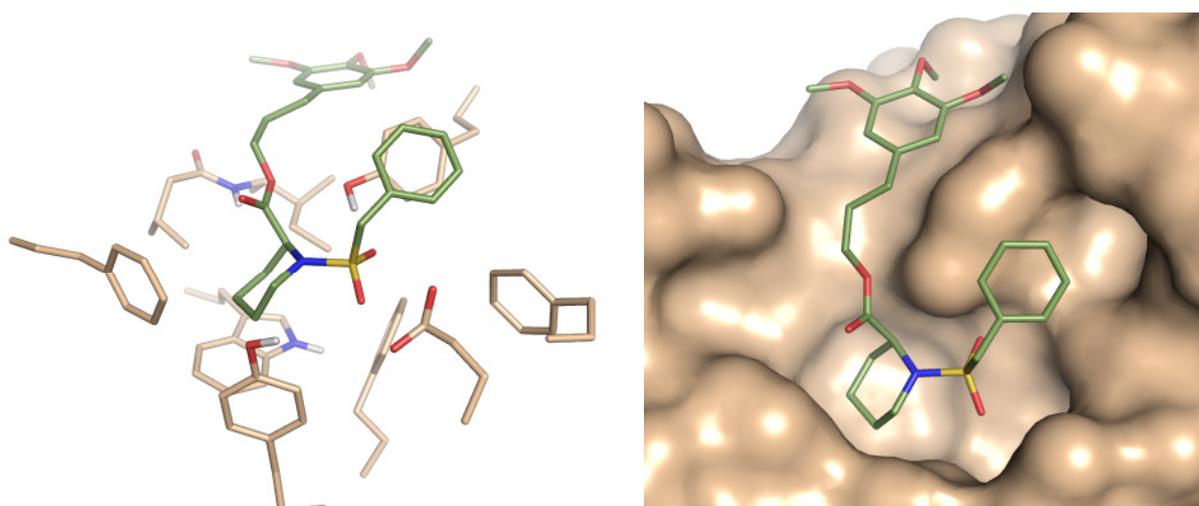


Figure 63: Non-immunosuppressive ligands of FKBP12. Ligand MI-1 also has a pipercoline substructure, yet with a sulfone substituent. The substituents have different steric demands, which is a decisive factor for binding to the narrow MIP pocket.

Docking of MI-1

MI-1 was built using the MOE molecule builder, minimized using the MMFF94x force field (modification of the MMFF94s force field¹⁴⁴, minimization is terminated if the RMSD gradient is ≤ 0.05 Å), and saved in mol2 format. This time, for practical reasons, GOLD4.0 with GOLDScore was taken for docking (GOLD4.0 was also employed to dock **1fkg-lig**, yielding similar binding modes as described above). 50 GA runs were performed. The binding site was defined by a sphere of 8 Å radius, centered between Tyr55 and Tyr109 of the fully protonated structure **2VCD-4**. Genetic algorithm parameters were set as follows: Population size 100, number of islands 5, selection pressure 1.1, number of operations 89500, niche size 2, migrate 10, mutate 95, and crossover 95. Clustering of the docking poses was performed using a threshold of 2 Å (this rather high value should account for the flexibility of the trimethoxy-phenyl side chain). The resulting binding modes and scoring data are shown in Figure 64.



	GOLDScore	S(hbond)	S(hydrophobic)	S(internal strain)	Clustering
MI-1	69.73	0.00	82.66	-12.93	30/50

Figure 64: Top-ranked docking result of MI-1 and scoring data. The pipercoline ring is placed in the active site pocket, showing a good shape complementarity (best seen in the surface representation on the right). The phenyl ring of the sulfonamide moiety occupies a subpocket and contributes to the high hydrophobic score. “Clustering” refers to the number of members in the top-ranked cluster.

Looking at the scoring table, the high total GOLDScore and especially the high hydrophobic contributions are striking. In a comparative docking study, **MI-1** was also docked to its original target, FKBP12. Here, an overall similar binding mode was obtained, however with a lower score (GOLDScore = 58.14, S(hbond) = 1.87, S(hydrophobic) = 62.05, S(internal strain) = -5.78). The top-ranked and well-clustered docked binding mode of **MI-1** shows a good fit of the pipercoline-sulfonamide anchor group, which is reflected in the high hydrophobic score. In contrast to **1fkg-lig**, the amide substituent (i.e. the sulfone), slides underneath the hydroxyl group of Tyr109, thereby allowing a more favorable orientation of the pipercoline moiety, having its wedge parallel to Trp86. This points out the essential influence of the amide substituent: The ketoacyl substituent forms a planar amide, which is too bulky for the narrow MIP binding pocket and clashes with Tyr109; in contrast to that, the more flexible non-planar sulfonamide fits underneath the tyrosine. Interestingly, the hydrogen bond score is insignificant, although the hydroxyl group of Tyr109 is close to the sulfone oxygen. An unsuitable angle may prevent the formation of a strong hydrogen bond. However, given the flexibility of this residue, it is conceivable that Tyr109 moves slightly backwards, thereby enabling hydrogen bond formation. As a matter of fact, the orientation of the pipercoline-sulfonamide substructure within the MIP active site pocket resembles the orientation of a close sulfonamide analogue within the FKBP12 active site pocket, as observed in crystal structure 1J4H¹⁴⁵ (not shown). The trimethoxy-phenyl ring displays solvent exposure, but at the same time nicely lies on the protein surface, again yielding favorable hydrophobic interactions. The exposure of aromatic rings has already been observed in the case of **1fkg-lig** binding to FKBP12. The phenyl ring of the sulfonamide moiety is accommodated by a subpocket close to Phe65. Modifications on the *para*-position of this phenyl ring could lead to the development of MIP-selective ligands: The MIP subpocket occupied by the phenyl ring grants space for a substituent, whereas the corresponding FKBP12 subpocket is narrowed by a protruding histidine residue. Overall, **MI-1** is a promising candidate for synthesis and subsequent testing in biological assays.

Synthesis was performed by Christina Juli, University of Wuerzburg. **MI-1** (and potential derivatives) is readily accessible via a four-step synthesis (see Figure 66). Here, the pipercolinic acid educt is not enantiomerically pure (due to the high cost); the product is a racemic mixture of *S*-**MI-1** and *R*-**MI-1**. For the docking studies shown here, the *S*-enantiomer (which is observed in the structural data of FKBP12 complexes and the MIP-rapamycin complex) is regarded. However, a docking study with *R*-**MI-1** was also performed: Compared to *S*-**MI-1**, the pipercoline anchor fragment of *R*-**MI-1** is twisted and the sulfone oxygens are exposed (see Figure 65). These

differences are reflected by the scores: The hydrophobic contributions are significantly lower (75.76, compared to 82.66 for *S*-**MI-1**). Like for *S*-**MI-1**, the hydrogen bond score is zero.

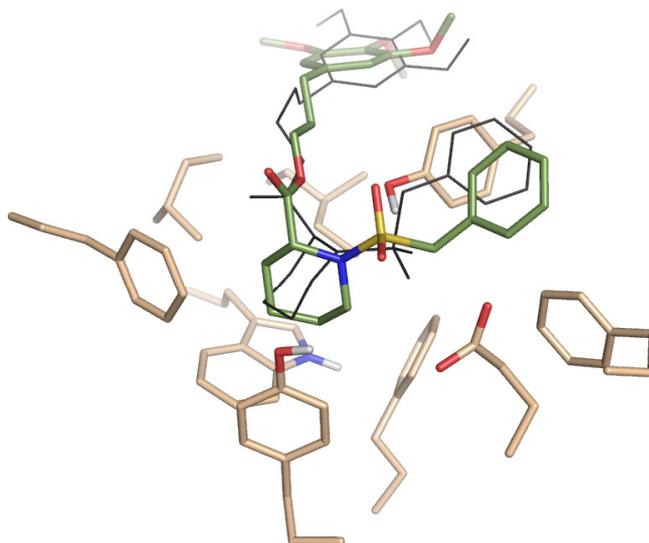


Figure 65: Docked binding mode of *R*-**MI-1** to MIP. The pipecoline ring has a different orientation, the sulfone oxygens point towards the surrounding, compared to *S*-**MI-1** (black lines).

Docking of *R*-**MI-1** yielded a well-clustered (26 out of 50 runs in the top-ranked cluster) binding mode, but the less favorable hydrophobic score suggests that this enantiomer may bind weaker to MIP than *S*-**MI-1**. Therefore, once the inhibitory activity of the **MI-1** scaffold is confirmed, synthesis with enantiomerically pure pipecolinic acid is planned.

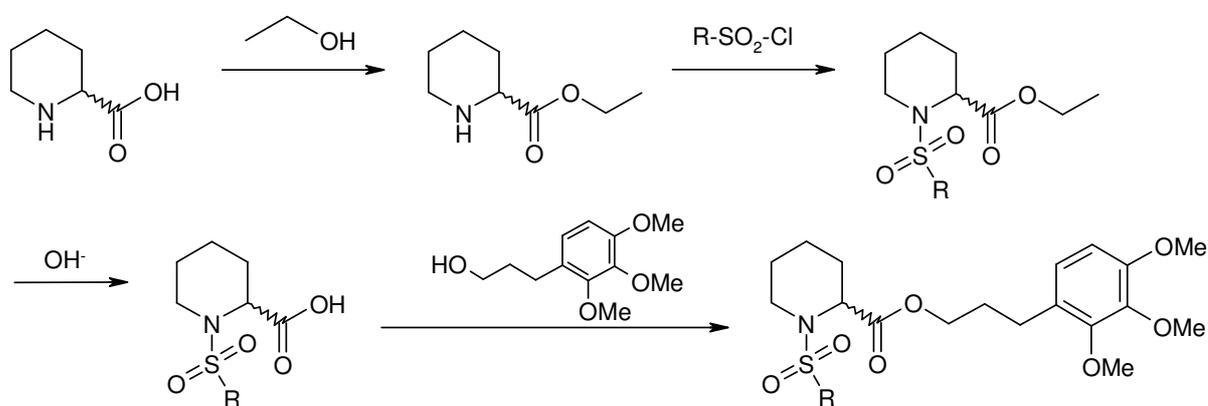


Figure 66: Preparation of the pipecoline-sulfonamide scaffold. R stands for benzyl or phenyl. 1fkg-lig is amenable by using the keto-acid chloride instead of the sulfone chloride.

Synthetic modifications on **MI-1** lead to the close analogue **MI-2** (again as a racemic mixture), which has a phenyl instead of a benzyl group attached to the sulfone. Based again on docking with GOLD4.0 (performed as described above, again with the *S*-enantiomer), also this compound shows a convincing binding mode, which is close to the one of **MI-1**, but with a significant loss in hydrophobic interactions (as reflected by the score). However, the shorter sulfonamide “tail” grants more possibilities for aromatic substitution (preferably in *meta*-position).

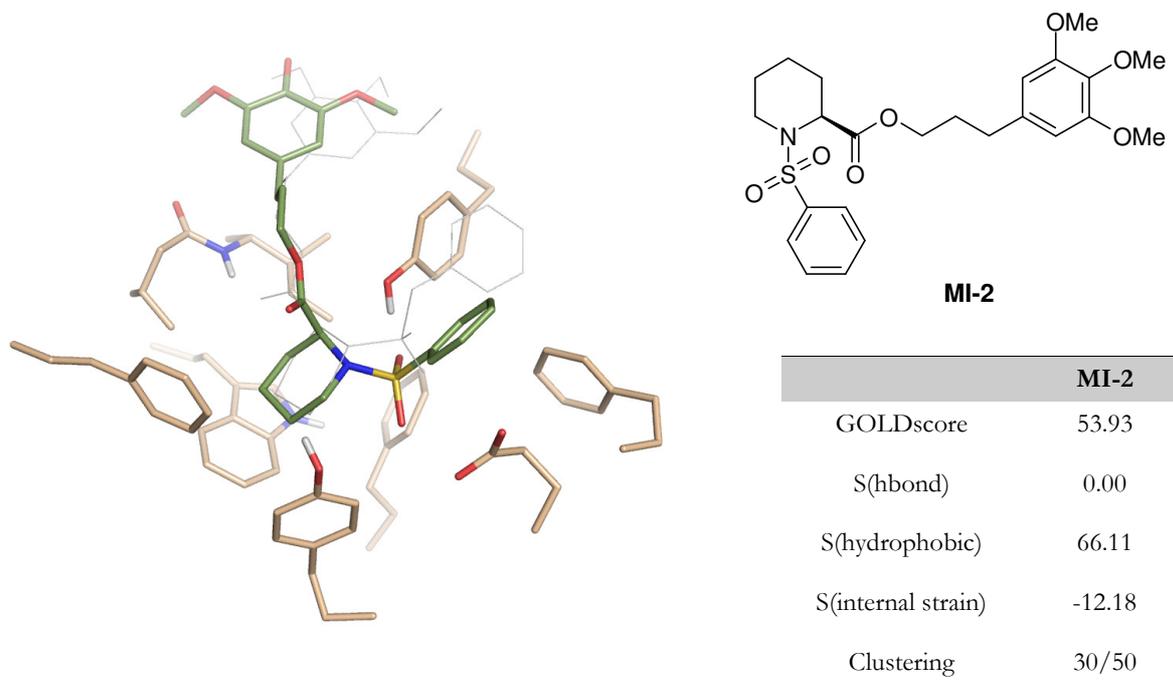


Figure 67: Docked binding mode of MI-2 and scoring data. The binding mode resembles the one of MI-1, yet with a lower score.

3.4 Biological testing of compounds

Testings were performed by Alexandra Thiele, Max-Planck-Institute for Enzymology, Halle. The biological testing is based on a chymotrypsin-coupled fluorometric PPIase assay. Here, the proline residue of a pentapeptide bearing the *para*-nitroaniline fluorophore isomerizes from its *cis* to its *trans* conformation. Only the latter is prone to proteolysis by chymotrypsin, leading to release of the fluorophore, which itself can be detected by fluorometry at 390 nm. The proline *cis-trans* isomerization is catalyzed by a PPIase like FKBP12 or MIP. Accordingly, inhibition of these enzymes can be monitored by fluorometric measurements.

To dissolve lipophilic test compounds, usually DMSO is added. As a matter of fact, usage of this solvent has major influence on the enzymatic activity (see next chapter). Only very recently, the assay has been tuned for the use of ethanol as solvent, which is expected to have less influence.

Assay results are shown in Table 17. Here, the decrease in enzymatic activity depending on compound concentration is listed.

c [μM]	MIP activity [%] 1fkg-lig (diast.)	MIP activity [%] MI-1 (rac.)	MIP activity [%] MI-2 (rac.)
1	84	105 (107)	74 (99)
5	73	70 (75)	60 (68)
10	126	70 (50)	60 (67)
20	90	60 (33)	56 (64)

Table 17: Activity data of 1fkg-lig (diastereomeric mixture), MI-1 (racemic mixture), and MI-2 (racemic mixture) against MIP. The decrease of catalytic activity upon compound addition was monitored. Values were derived from the assay containing DMSO (values in parentheses were obtained with the ethanol protocol).

The activity data shows that for both protocols (DMSO/ethanol), a significant decrease in enzymatic activity is obtained upon addition of **MI-1** and **MI-2**. In the case of **MI-1**, the values of both protocols differ significantly, whereas for **MI-2**, the values are similar. A continuous decrease in activity is observed only for the ethanol protocol of **MI-1**; in other cases, doubling of the inhibitor concentration does not lead to further significant decrease in enzymatic activity (see values for **MI-2**). Still, at a concentration of 20 μM , both **MI-1** and **MI-2** show an explicit inhibitory activity. These results are even more encouraging, given the fact that both compounds were tested as racemic mixtures: Docking studies suggested that *S*-**MI-1** (*S*-**MI-2**) is the preferred enantiomer; synthesis of enantiomerically pure **MI-1** and **MI-2** is currently performed and expected to show an increased inhibitory activity.

Although the values for **1fkg-lig** show an initial decrease in enzymatic activity, no significant inhibitory activity is observed for the highest concentration ($c = 20 \mu\text{M}$). Values of the ethanol protocol have not been recorded yet. Furthermore, the diastereomeric mixture may have a low concentration of the potentially active diastereomer (see Figure 63). However, higher concentrations of the diastereoisomeric mixture are problematic, due to compound precipitation.

The results obtained so far suggest that **MI-1** and **MI-2** are MIP inhibitors, with the potential to improve their affinity by preparing enantiomerically pure compounds. **MI-1** has been subjected to testing in a so-called transmigration assay (performed by Michael Steinert, University of Braunschweig): Here, *Legionella* penetration of a model ECM is monitored; successful MIP inhibitors should block the transmigration.

3.5 DMSO as new substructure?

DMSO addition leads to a significant decrease in MIP catalytic activity. The question arises whether this decrease can be traced back to a partial denaturation of the protein in presence of the solvent or rather to a specific binding of DMSO to the MIP active site. To address this question, HSQC-NMR studies were performed by Christian Schweimer, University of Bayreuth. In this experiment, chemical shift perturbations of protein residues upon DMSO addition were observed. After assigning the chemical shift perturbation signals to the respective protein residues, one can see that especially residues within or close to the active site are affected. Figure 68 shows residues with high (red) and medium (orange) perturbations.

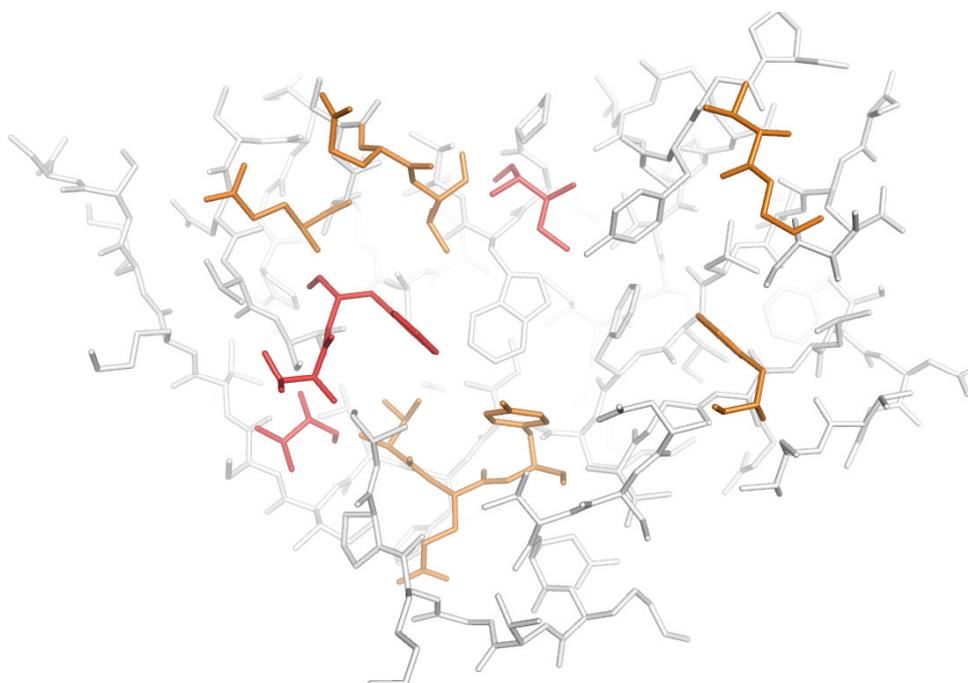


Figure 68: Chemical shift perturbation data from HSQC-NMR experiments performed by Christian Schweimer, mapped onto the MIP structure. Highly affected residues are shown in red, medium perturbations are indicated by orange color.

The largest perturbations were detected for residues Thr52, Thr76, Phe77, and Ile83. Also for residues Val53, Glu54, Tyr55, Phe65, Gln78, Ser80, Gln81, Val82, Val114, and Gly115, perturbations were observed. These residues are located within or at least nearby the catalytic center. Hence, results from the HSQC-NMR experiments strongly suggest that DMSO specifically binds to the MIP active site, thereby decreasing its catalytic activity.

DMSO binding mode in MIP – Docking

To get an impression of the DMSO binding mode within the MIP active site, docking experiments were carried out: GOLD4.0 with GOLDScore was employed, the docking site was defined by an 8 Å sphere centered between Tyr55 and Tyr109. The ligand was built and fully protonated using MOE. Default parameters were taken to perform 10 GA runs, the docking poses were subsequently clustered with an RMSD tolerance of 1.5 Å. Results are shown in Figure 69.

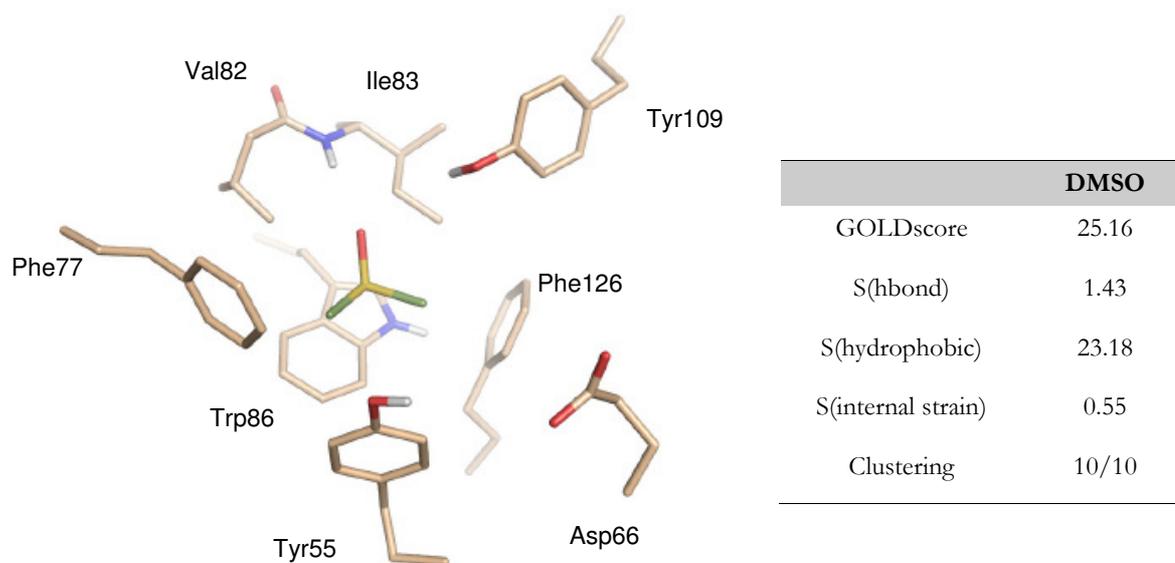


Figure 69: Binding mode and scoring data of DMSO docked to MIP. The ligand occupies the active site, acting as hydrogen bond acceptor with Ile83. The polarizable sulfur atom as well as the methyl groups are favorably placed within the hydrophobic pocket.

Docking yielded a well-clustered binding mode which shows favorable interactions such as hydrogen bonding between the sulfoxide oxygen and the backbone N-H group of Ile83. Furthermore, both methyl groups and the polarizable sulfur atom are buried within the hydrophobic active site pocket. Also a subtle interaction between the polarized S-methyl group and the π -system of Phe77 is conceivable (similar to methionine-phenylalanine interactions within proteins), however, such an interaction is not considered by the docking algorithm. The docked binding mode is in agreement with the NMR shift perturbation data and explains the enzymatic activity loss upon DMSO addition.

Interestingly, DMSO binding was also observed in the homologous FKBP12 ($K_d = 20$ mM): The crystallographic structure 1D7H¹⁴⁶ includes DMSO, which occupies the active site. Also in this case, DMSO forms a hydrogen bond with the backbone N-H group of Ile56 (corresponding to Ile83 in MIP) and the dimethyl-sulfur system is buried within the hydrophobic active site pocket.

The binding mode was reproducible in docking with GOLD4.0 (RMSD = 0.65 Å, data not shown).

3.5.1 DMSO derivatives – cyclic sulfoxides

The docking study further validated the assumption of specific DMSO binding on a structural level. In the next step, the question of how to exploit this finding needs to be addressed: Can DMSO serve as substructure for the design of sulfoxide fragments in general? The binding mode of DMSO in MIP shows that the S=O group favorably interacts with Ile86 via hydrogen bonding. The hydrophobic active site pocket is addressed by both methyl groups. Still, the pocket grants more space for hydrophobic groups. According to the docking results, **MI-1** (see above) places its pipercoline moiety in the active site pocket. Similar motifs are also frequently found in FKBP12 ligands, with both five- or six-membered rings. It seems therefore advisable to place an aliphatic ring system in the pocket to fully exploit potential hydrophobic interactions. Consequently, the DMSO sulfoxide substructure and an aliphatic ring system could be combined to yield a new anchor moiety: A cyclic internal sulfoxide, tetrahydrothiophenoxide (**CF-1**, Figure 70), which was also reported as FKBP12 inhibitor ($K_d = 200 \mu\text{M}$)¹⁴⁶.

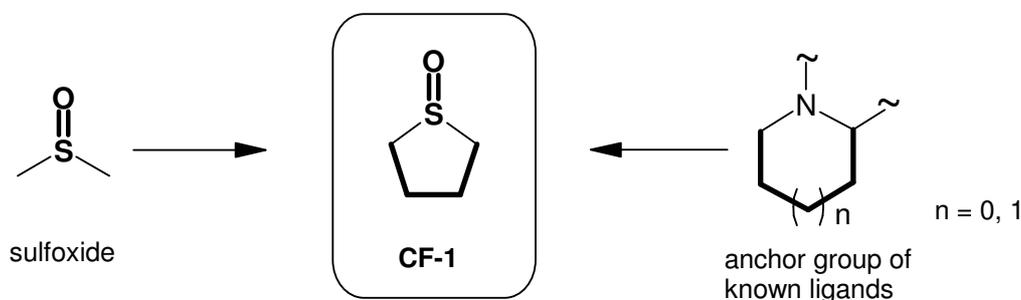


Figure 70: Combination of the DMSO sulfoxide with the aliphatic ring substructure, yielding a cyclic sulfoxide, which potentially enables both hydrogen bonding and hydrophobic interactions.

Thereby, the hydrogen bonding interactions with Ile83 via the sulfoxide substructure (see DMSO binding mode) should be maintained, while the hydrophobic pocket of the MIP active site is exploited by the aliphatic ring. Accordingly, **CF-1** is a suitable anchor group. To test this hypothesis, docking experiments had to be carried out: **CF-1** was docked to the MIP binding site with GOLD4.0 (GOLDscore). The ligand was built by converting the SMILES code into mol2 format using CORINA. Default docking parameters were employed for 10 GA runs, the top-

ranked binding mode was rescored with the SFCscore fragment scoring function¹¹⁵. Results are shown in Figure 71.

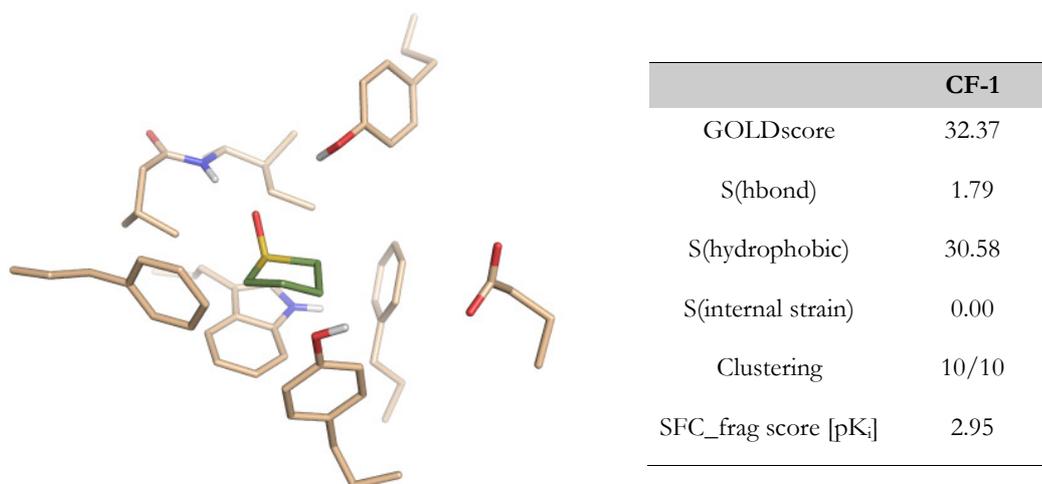


Figure 71: Binding mode and scoring data of CF-1 docked to MIP. Like DMSO, the sulfoxide accepts a hydrogen bond from Ile83. The carbon ring atoms are buried within the pocket.

The resulting docked binding mode confirmed the expectations that **CF-1** should fit nicely: Like DMSO, the sulfoxide group acts as hydrogen bond acceptor with Ile83. Also, the ring system fills the hydrophobic pocket, in analogy to the piperidine ring in **MI-1** (see above). **CF-1** therefore unites the favorable features of known MIP binders, yielding an interesting anchor group for further development. NMR experiments (in analogy to the studies performed for DMSO) are planned to confirm binding of **CF-1** to MIP. Interestingly, tetrahydrothiophenoxide is an inhibitor of liver alcohol dehydrogenase, which itself plays an important role in ethanol metabolism¹⁴⁷. In this context, synthetic pathways to extend this lead structure had been established, e.g. for 3-substituted alkyl derivatives. To study the impact of fragment extension on the binding mode, an example structure, 3-propyl-tetrahydrothiophenoxide (**CF-2**), was docked to the MIP active site (Figure 72): As expected, the tetrahydrothiophenoxide substructure takes a similar orientation as **CF-1**, but in the case of **CF-2**, the sulfoxide is directed in a more favorable angle towards the backbone N-H of Ile83. To compare both compounds, their binding modes were assessed by the SFC_frag scoring function: **CF-2** is ranked higher than **CF-1** (calculated pK_i = 4.09 and 2.95, respectively), i.e., **CF-2** is expected to have a higher affinity. Confirmation by experimental testing is pending.

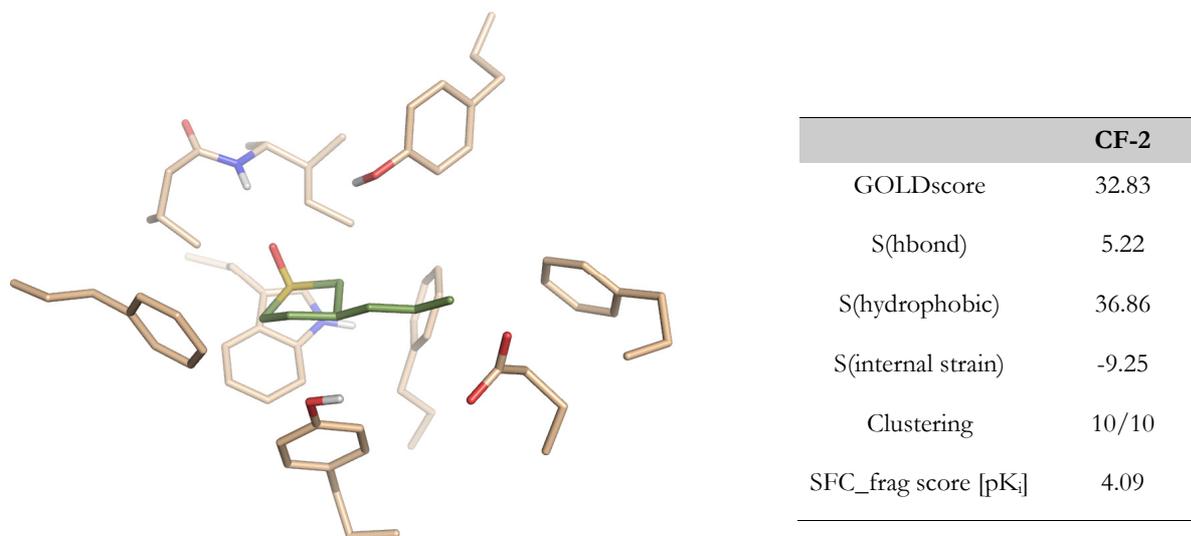


Figure 72: Binding mode and scoring data of CF-2 docked to MIP. Again, Ile83 is addressed by the sulfoxide. The alkyl chain points towards a subpocket of the MIP active site.

The propyl side chain of **CF-2** is directed towards a subpocket within the MIP active site. This subpocket could be addressed by further extending the **CF-2** anchor fragment. Generally, **CF-2** shows a high potential for being a suitable anchor group for MIP inhibitors. Of course, this compound is too small for sufficiently strong binding. It has to be extended into the already mentioned subpocket. At this point, fragment-based design comes into play.

3.6 Extension of the anchor: Fragment-based design

3.6.1 Fragment-based drug design

Screening of large compound libraries (typically containing several million molecules) is one of the standard techniques in the drug discovery process. This common approach routinely identifies binders to a given target. However, there are also drawbacks, for example the frequently low hit rates: The molecular complexity of molecules from large compound collections lowers the probability of a good complementarity with the target protein¹⁴⁸. In other words, a library of less complex molecules should yield a higher probability of a suitable match, i.e., the hit rates are increased¹⁴⁹. This is the idea of the so-called fragment-based drug design: Instead of searching for complex molecules matching the complete active site of a target protein, molecular fragments are

searched which address distinct interaction subsites. In the last step, these subsite-binding fragments are connected to a single molecule, which optimally addresses the entire pocket. Fragments are defined as compounds with a low molecular weight ($\sim 120\text{-}250\text{ Da}$)¹⁵⁰. A library of such small compounds explores the available chemical space much better than a library of complex molecules: Combinatorial assembly of the matching fragments leads to an exponential growth of the number of possible molecules. Due to the higher hit rates, the fragment-based approach allows screening of smaller libraries (typically containing several thousand compounds). This "low-throughput" screening allows the use of techniques like NMR or X-ray for identification of binding fragments¹⁵¹. Due to their low molecular weight, fragments mostly show low affinity values: Thus, fragment-based screening techniques must have a high sensitivity to detect compounds with high micromolar or even millimolar affinity. Besides the experimental techniques, also a virtual screening of a fragment database can be employed to retrieve small compounds for a given target binding site.

3.6.2 MIP active site subpocket

According to the docking results, the alkyl side chain of **CF-2** occupies the "entrance" of the subpocket formed by Val114, Gly115, Gly116, Ile118, and Leu124. The outer rim of the subpocket is characterized by the backbone C=O and N-H groups of Val114 and Gly116, respectively (see Figure 73). This subpocket is interesting, because the corresponding FKBP12 pocket (formed by residues His87, Pro88, Gly89, Ile90, and Ile91) has a significantly different arrangement in terms of hydrogen bonding groups and shows protruding residues (His87, Ile90) which narrow the subpocket. Hence, it is conceivable to obtain selective MIP ligands through subpocket fragments which fit to MIP but not to FKBP12. Accordingly, the MIP subpocket was analyzed for ligand interaction sites, and fragments addressing this subpocket were identified by means of a pharmacophore-based virtual screening of a fragment database. Ultimately, fitting fragments could be linked with the core fragment **CF-2** to yield a ligand with potentially high anti-MIP activity.

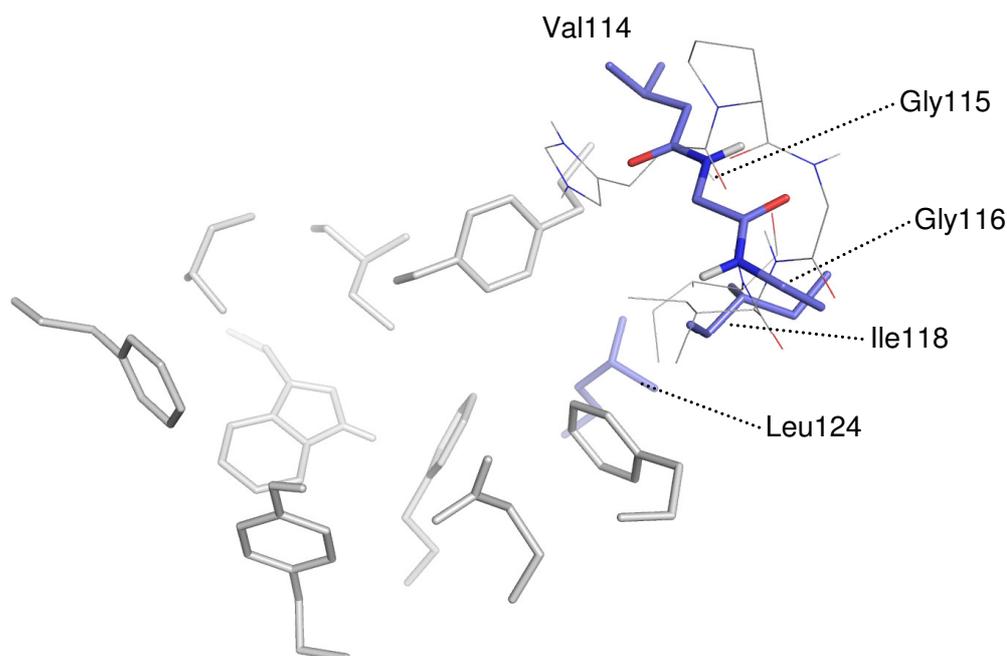


Figure 73: MIP Subpocket (blue sticks) adjacent to the active site pocket (grey sticks). The FKBP12 subpocket residues (thin lines) have a significantly different arrangement (in particular, the histidine restraints access to this subpocket).

Methods

Apparently, the backbone N-H of Gly116 and C=O of Val114 are the central interaction points within the subpocket. Therefore, an MOE pharmacophore model was created (see Figure 74), having a donor feature to address Val114 and an acceptor feature for Gly116; additionally, projected features for donor and acceptor were set. Constraints were imposed which required the acceptor / projected acceptor and donor / projected donor pairs to be fulfilled by the same atom, respectively. Due to the small size of the subpocket and of the fragments, no hydrophobic feature was set. The MOE fragment database served as input for the search. To limit the size of the candidates, all fragments having a molecular weight greater than 120 Da were skipped. Matching molecules were docked to the subpocket, defined by a 6 Å radius around Ile118, again using GOLD4.0 with GOLDScore. To select only those binding modes which fulfill the desired hydrogen bonding pattern to Val114 and Gly116, a hydrogen bond descriptor within the HERMES program was defined to filter out non-matching molecules after docking. Finally, the remaining fragments were analyzed by visual inspection.

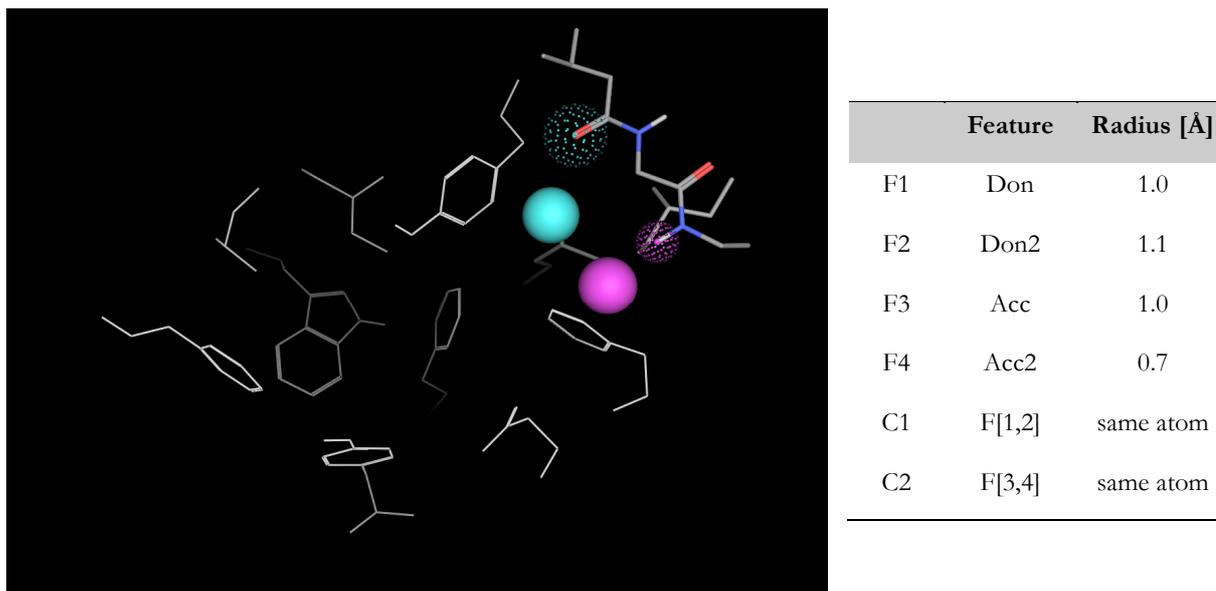


Figure 74: Four-point pharmacophore model addressing the subpocket: Donor (solid, cyan) and acceptor (solid, violet) features were completed by projected donor and acceptor features (dots). Constraints were set, which require that the feature / projected feature pair is fulfilled by the same atom.

Results

8600 compounds with a molecular weight below 120 Da remained for the pharmacophore-based screening: 2372 fragments passed the query and an additional filter, removing groups with negative formal charge (to avoid electrostatic repulsion with Asp66) and fragments with more than one chiral center (for better synthetic accessibility). After docking, the hydrogen bond descriptor retrieved 184 fragments which fulfilled the criterion to address both Val114 and Gly116. To limit the flexibility, fragments with more than two rotatable bonds were rejected and the remaining 61 compounds were analyzed by visual inspection: Here, special attention was paid to the placement of atoms near the subpocket entrance (i.e. close to the alkyl side chain of **CF-2**, to allow an easier linkage of the core with the subpocket fragment); favorable hydrophobic interactions, e.g. with the phenyl ring of Tyr109, were an additional criterion; furthermore, appropriate compounds must not display solvent exposure. Compounds with a convincing binding mode were subsequently assessed for their synthetic feasibility. Since the MOE fragment database contains artificial (i.e. non-purchasable) fragments, vendor databases were searched for precursors or derivatives, which were required to have functional groups for easier linkage with the core fragment. Finally, two cyclic fragments, **F1** and **F2** (see Figure 75), were selected. Both display hydrogen bonding with the backbone of Val114 and Gly116 via their amide function. **F2**

shows a further hydrogen bond to Asp66. Additionally, the aliphatic/aromatic ring systems fit well in the hydrophobic environment created by Tyr109, Ile118, and Phe65. Also, the rings reach to the “entrance” of the subpocket, i.e. in vicinity to the alkyl side chain of **CF-2**, which is important for fragment linkage. As a further favorable feature, the fragments are rather rigid and therefore have a low entropic loss upon binding.

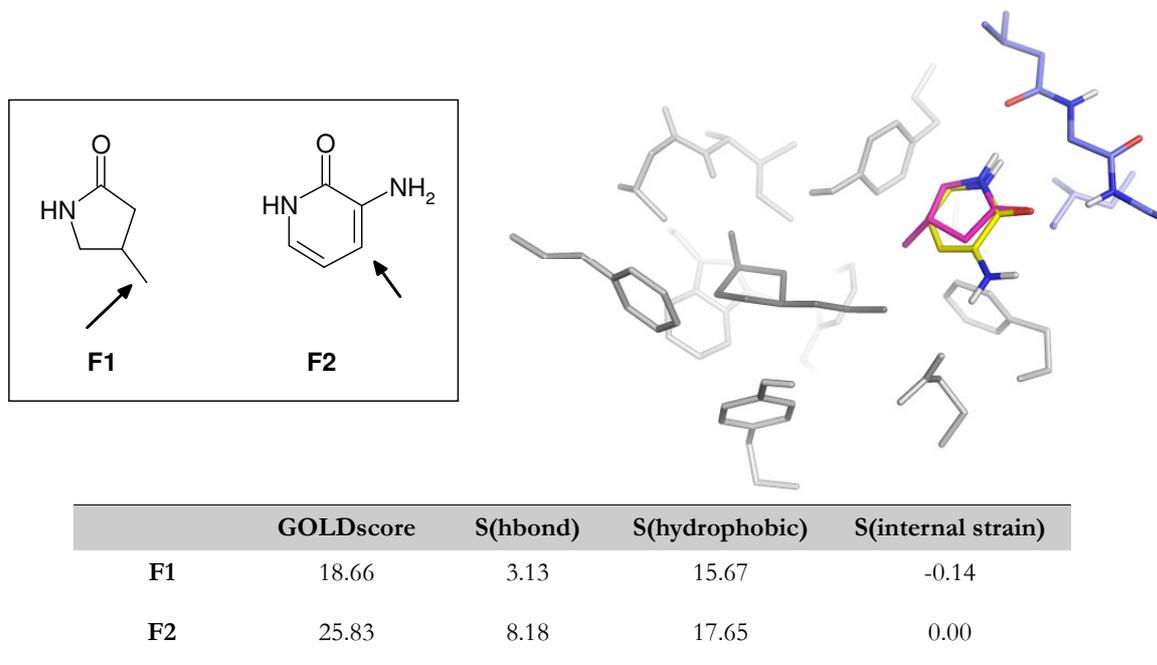


Figure 75: Fragments for the subpocket. The arrows indicate the position for the following linkage with the core fragment **CF-2**. Color coding: F1 (magenta), F2 (yellow), **CF-2** (gray).

Fragment linkage

In the next step, the core fragment **CF-2** needs to be connected to the subpocket fragments **F1** and **F2**, respectively, yielding two single molecular entities. Here, the pivotal question is how to link the core fragment with the subpocket fragment: This linkage, of course, needs to be synthetically feasible. Therefore, the subpocket fragments were assessed in terms of synthetic accessibility: 2-pyrrolidinone (**F1**) is a common compound, which is commercially available with functional groups (e.g. ester) at the required 4-position (see Figure 76). For the 3-amino-2-pyridone **F2**, precursors are commercially available: A functional group is required in the 4-position. As mentioned before, synthesis of the core fragment **CF-2** has already been described in the literature. This synthesis needs to be slightly modified to introduce a terminal functional

group in the alkyl side chain. Thereby, the core fragment can be connected with the subpocket fragments.

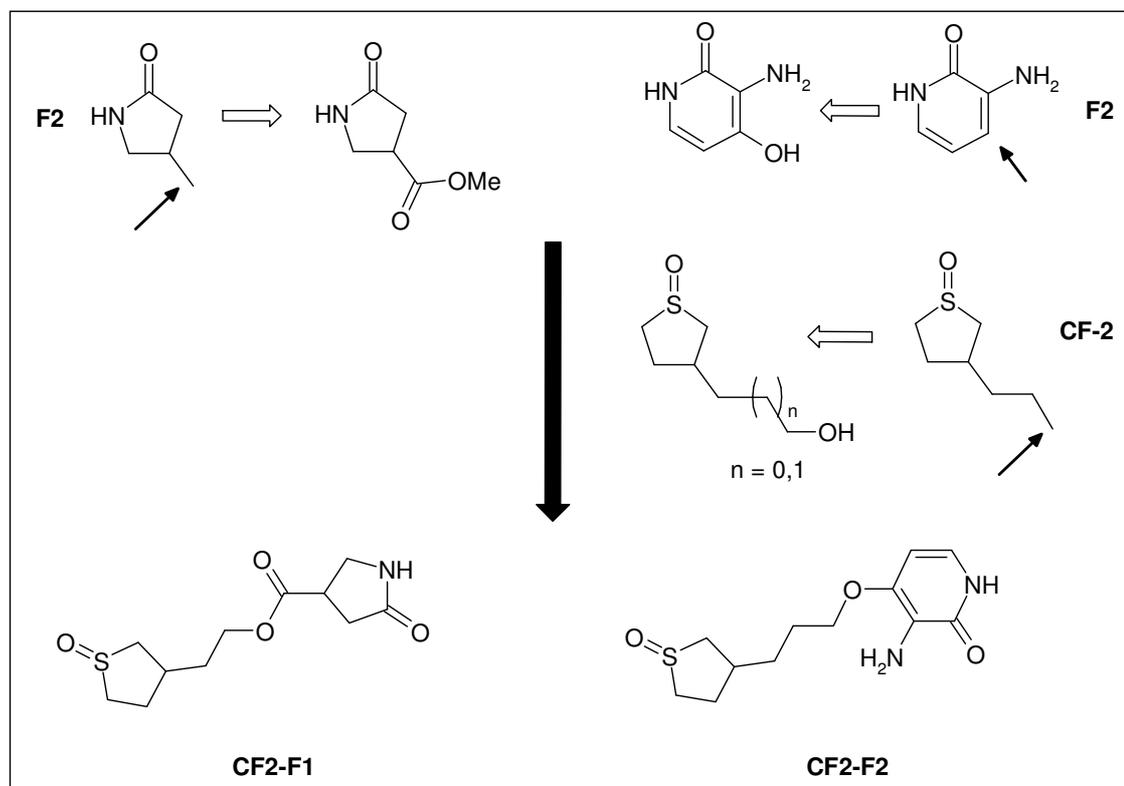
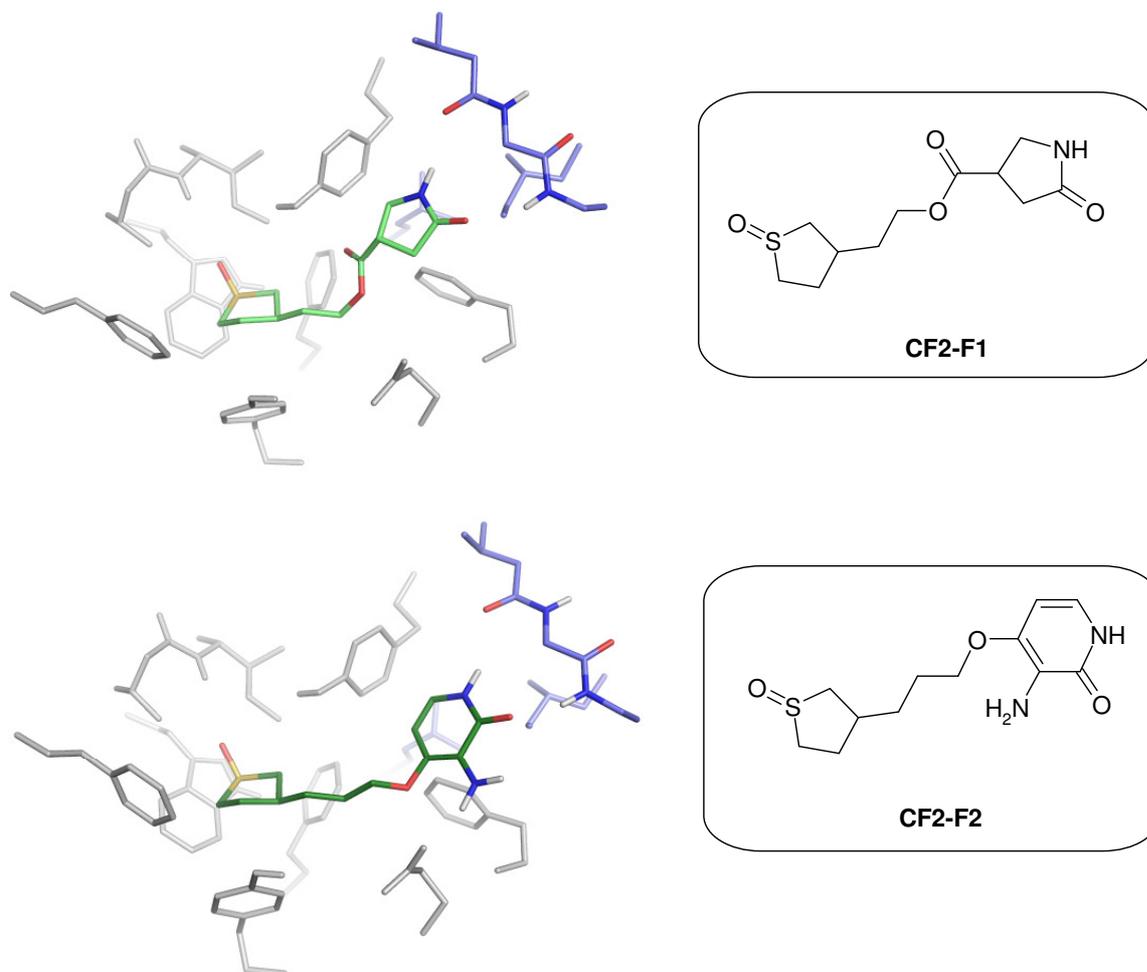


Figure 76: Schematic modifications of the fragments prior to linkage. The F2 ester is commercially available, also a precursor of the hydroxylated F2 fragment. CF2-F1 and CF2-F2 are the resulting molecules, where the core and the subpocket fragments have been linked.

The resulting molecules were docked to the complete MIP active site, i.e. active site pocket plus adjacent subpocket, considering the energetically favored pseudo-equatorial position of the five-membered ring substituents. Results are shown in Figure 77. Here, the positioning of the core and subpocket groups yields the same interaction pattern, compared to the docked binding mode of the single fragments: The sulfoxide moiety still occupies the core pocket, whereas the pyrrolidinone / pyridone moieties are placed within the subpocket, addressing the hydrogen bonding residues Val114, Gly116, and Asp66, respectively. Looking at the scoring data, it is striking that the strength of the hydrogen bonding contribution of the molecule is weaker than the hydrogen bonding contribution of the respective fragments (see scoring data). Although the positioning of the molecules' core / subpocket moieties displays an overall good match with the single fragments, slightly shifted positions result in weakening of the hydrogen bonds. Obviously, when fragments are connected to a molecule, the translational freedom of the moieties is restrained and therefore an optimal hydrogen bonding geometry maybe hindered. Nevertheless,

the required residues (Val114, Gly116, Asp66 for the subpocket fragments, Ile83 for the core fragment, respectively) are still addressed. Inspection of the MIP solution structure ensemble 2VCD has shown an inherent slight flexibility of these residues, which could lead to an improved hydrogen bonding geometry for **CF1-F1** and **CF1-F2**.



	GOLDScore	S(hbond)	S(hydrophobic)	S(internal strain)
CF2-F1	47.00	3.14	48.36	-4.50
CF2-F2	41.42	6.08	51.02	-15.68
<i>CF-2</i>	<i>32.83</i>	<i>5.22</i>	<i>36.86</i>	<i>-9.25</i>
<i>F1</i>	<i>18.66</i>	<i>3.13</i>	<i>15.67</i>	<i>-0.14</i>
<i>F2</i>	<i>25.83</i>	<i>8.18</i>	<i>17.65</i>	<i>0.00</i>

Figure 77: Docked binding modes and scoring data of the single molecular entities CF2-F1 and CF2-F2, The orientation of the fragmental substructures matches the original fragment orientations. For reference purposes, also the scoring data of the fragments are given (in italics).

3.7 Conclusion

The studies done so far have revealed two experimentally validated MIP inhibitors: **MI-1** and **MI-2** significantly reduce MIP enzymatic activity. Although **MI-1** is a known inhibitor of FKBP12, it still is an attractive lead structure, since it is expected to be non-immunosuppressive (in contrast to the known MIP inhibitor rapamycin). **MI-1** has entered the next stage and is currently tested in a cell-based assay for anti-MIP activity. **MI-2** is a close analogue, for which structural modifications were suggested which can increase its affinity.

A detailed analysis of the available structural data of MIP and a comparison to the human PPIase counterpart, FKBP12, pointed out a conformational diversity among the MIP structures and a crucial difference between both PPIases, which could be traced to mainly one residue (Tyr109). The detailed comparison of FKBP12 and MIP complex structures made it possible to give an explanation, why a ketoacyl-substituted pipercoline derivative most probably does not bind to MIP, but a sulfone-substituted pipercoline derivative does bind to MIP.

Experimental validation was done via a MIP inhibition assay, which was found to be very sensitive to DMSO addition. Currently, the protocol is changed for the use of ethanol, which is expected to have less influence on the catalytic activity of MIP. The sulfone-substituted pipercoline derivatives **MI-1** and **MI-2** were shown to significantly decrease MIP activity in both protocols (i.e. DMSO and ethanol), however with different activity values. It is expected that enantiomerically pure compounds (for the tests, racemic mixtures were taken) will lead to an increase in inhibitory activity of **MI-1** and **MI-2**: Docking studies revealed that the *S*-enantiomer is preferred. The ketoacyl-substituted pipercoline derivative **1fkg-lig** did not show a significant decrease in catalytic activity using the DMSO protocol. However, for consistent data, the assay testings need to be repeated using the ethanol protocol, to get a definite result about its inactivity against MIP.

The question was addressed whether the decrease in catalytic activity of MIP upon DMSO addition originated from a partial denaturation of the enzyme or rather a specific binding to the active site. Docking studies of DMSO to MIP were performed and a binding mode within the active site was suggested which is in good agreement with NMR shift perturbation data. The sulfoxide substructure was subsequently extended to yield cyclic sulfoxides, **CF-1** and **CF-2**, which show favorable and consistent binding modes. Interestingly, tetrahydrothiophenoxide (**CF-1**) and its derivative **CF-2** are described in the literature as promising agents against alcohol intoxication, by inhibiting the liver alcohol dehydrogenase. The cyclic sulfoxides pose an interesting alternative to the pipercoline-based anchor groups of **MI-1** and **MI-2**. In a fragment-based design approach, a subpocket which is close to the binding domain of **CF-2**, has been

addressed by means of pharmacophore-based screening for fitting fragments. Taking care of the synthetic feasibility, two fragments (**F1**, **F2**) with a molecular weight below 120 Da were discovered, which are able to build a strong hydrogen bonding network within the subpocket. In the case of FKBP12, this subpocket is obstructed by a histidine residue; therefore, finding fragments which fill the MIP subpocket is a first step towards the development of MIP-selective inhibitors. Finally, the core fragment **CF-2** and the subpocket fragments were connected, yielding two molecules **CF2-F1** and **CF2-F2**, which show an overall good match between their molecules' substructures and the parent fragments. However, the scoring data suggested that the hydrogen bonding was weaker in the molecules compared to the single fragments, due to slight shifts in the orientations. It was suggested that this decrease in hydrogen bond strength could be compensated in a slightly different protein conformation, as observed in the structural ensemble solution structure of MIP. This points to the deficiencies of working with only one protein conformation: Proteins are not static, but show inherent flexibility. Thus, a designed ligand which may not fit optimally to one protein conformation, may fit perfectly to another protein conformation. Therefore, all available conformational data should be taken into consideration. However, this would mean a much greater effort, which is not always feasible.

The synthesis of the fragments and the connected molecules is currently elaborated. The core fragment **CF-2** is planned to be analyzed by means of NMR spectroscopy. Due to its low molecular weight, it is very unlikely to compete with the substrate in the biological assay. Nevertheless, extension of the core fragment with the subpocket fragments is expected to yield compounds which display activity in MIP inhibition assays.

4. Summary / Zusammenfassung

4.1 Summary

In this thesis, computational structure-based design approaches were employed to target the HIV-1 integrase and the macrophage infectivity potentiator (MIP) of *Legionella pneumophila*. Both enzymes are currently not fully exploited for the treatment of HIV infection and Legionellosis, respectively. Therefore, the development of novel agents against these targets is of major importance. The thesis yields valuable information about the mechanism of action of a known class of integrase inhibitors and a novel approach towards enzyme inhibition, which still is mainly unaddressed in current integrase research. For the MIP enzyme, the available structural data were critically assessed and differences to a homologous human enzyme were shown. Two small-molecule MIP inhibitors were discovered and approaches towards a fragment-based design of further inhibitory compounds were outlined.

The computational studies of HIV-1 integrase have provided valuable information for IN inhibitor design. Docking experiments supported the hypothesis that the well-known diketo acid inhibitors enter the IN active site not as free ligands, but rather as metal complexes. Different complex species were docked, however, visual inspection of the binding modes as well as the docked energies suggest that monomeric complexes with a metal-to-ligand ratio of 1:1 are the preferred species. These results help to reveal the mechanism of action of this important class of IN inhibitors.

Diketo acids and other integrase active site binders have been the focus of intensive research in industry and academia since more than a decade. To give an impulse for the development of a novel class of inhibitors, a new strategy towards IN inhibition was introduced: An alternative binding site, the dimerization interface of an IN catalytic core domain monomer, was explored for inhibitor design. The lack of structural data of the free monomer was overcome by extensive MD studies. An inspection of the dimerization interface had so far mostly been ignored. Snapshots derived from the MD simulation were used as protein input structures in a docking study with the known inhibitory peptide YFLLKL to reveal its potential binding mode, which had not been reported in the literature so far. The docking procedure showed that the peptidic ligand binds to a dimerization interface conformation which shows a Y-shaped binding site, formed by two cavities and a groove. The next step was to address this protein conformation with small, non-peptidic molecules. Based on the fact that YFLLKL has a prevailing hydrophobic

character and that aliphatic/aromatic hot-spots were detected within the dimerization interface, the first strategy towards finding small-molecule interface binders was to create a pharmacophore model with hydrophobic features and shape constraints, aiming to find molecules with a good complementarity to the Y-shaped dimerization interface. Virtual screening yielded a total of 10 compounds, which all displayed good shape complementarity and favorable hydrophobic interactions. Unfortunately, none of the compounds showed a reproducible inhibitory activity in biological assays. Some doubts remain about the validity of the assay results: The use of BSA (which had to be added according to the standard protocol) was critical, since it is not unlikely that BSA “intercepted” the hydrophobic candidate compounds. Additionally, it has to be critically mentioned that virtual screening campaigns sometimes have very poor hit rates (1 - 2 %). This would mean that of about hundred compounds suggested by virtual screening, only one shows activity. Of course, testing of hundreds of compounds is hardly feasible in a non-industrial environment. Various reasons could lead to a failure of a virtual screening campaign, e.g. inexact structural information. Also, the computational methods like docking may reach their limits in virtual screenings where a large, diverse set of compounds needs to be assessed; given the great diversity of molecules in a screening database, it is a delicate task to differentiate between binders and nonbinders for a given target.

The first strategy towards finding small-molecule dimerization inhibitors was reconsidered: In the second approach, the satisfaction of hydrogen bonding residues at the dimerization interface, which had not sufficiently been addressed by the overall hydrophobic compounds tested so far, was of major interest. Two pharmacophore models were employed, which retrieved several hundred hit molecules. However, docking of these molecules showed that still many hydrogen bonding groups of the protein remained unaddressed by the ligands. A systematic study of the Cheng dataset, comprising 195 protein-ligand complexes, showed that a high number of unsatisfied hydrogen bonding groups in both protein and ligand are rarely observed. Instead, in over 90 % of all cases, ligands and proteins have equal or less than four unsatisfied hydrogen bonding groups. Systematic studies of unsatisfied HBAs/HBDs in protein-ligand complexes have not been reported so far: The studies described here are currently extended in the group of Prof. Sotriffer by analyzing polar protein-ligand contacts which have an improper geometry for hydrogen bonding. The goal is to develop a descriptor which can be employed to reject improper binding modes generated by docking algorithms.

For the pharmacophore-based search, a threshold of four unsatisfied HBAs/HBDs was introduced, which led to rejection of over 90 % of the docked binding modes. Eventually, after visual inspection, only eight molecules were selected as candidate compounds for further testing (results pending). This small “yield” underlines the difficulties in finding interface binders: The

IN dimerization interface is a peculiar target with frequently alternating basic, acidic, and hydrophobic residues. It is not a well-ordered binding site with continuous hydrophobic areas and distinct hydrogen bond donors / acceptors. Other protein-protein interfaces show such well-ordered binding sites. Accordingly, the peculiarity of the IN dimerization interface, in addition to the delicate task of disrupting protein-protein interactions at all, makes the development of IN dimerization inhibitors very challenging. Still, it is a worthwhile goal: Small-molecule dimerization disruptors would pose a whole new class of IN inhibitors. Due to the lack of synthesis capacities, the aim of the project was to find purchasable compounds from common vendor databases. Given the peculiarity of the binding site, it could be argued that instead of precast compounds, tailor-made molecules are the better option. Also for this option, the studies shown here provide valuable information: Fragments like triaminobenzene or urea-pyrimidine (included in the hit compounds), can be used as a starting point for the synthesis of dimerization inhibitor candidates.

For MIP, the studies revealed two experimentally validated MIP inhibitors: The small-molecule compounds **MI-1** and **MI-2** significantly reduce MIP enzymatic activity. To our knowledge, no small-molecule MIP inhibitor has been reported in the literature so far. **MI-1** has entered the next stage and is currently tested in a cell-based assay for anti-MIP activity. **MI-2** is a close analogue, for which structural modifications were suggested which can increase its affinity.

A detailed analysis of the available structural data of MIP and a comparison to the human PPIase counterpart, FKBP12, pointed out a conformational diversity among the MIP structures and a crucial difference between the two PPIases, which could be traced to mainly one residue (Tyr109). The detailed comparison of FKBP12 and MIP complex structures made it possible to give an explanation, why a ketoacyl-substituted pipercoline derivative most probably does not bind to MIP, but a sulfone-substituted pipercoline derivative does bind to MIP.

Experimental validation was done via a MIP inhibition assay, which was found to be very sensitive to DMSO addition. The question was addressed whether the decrease in catalytic activity of MIP upon DMSO addition originated from a partial denaturation of the enzyme or rather a specific binding to the active site. Docking studies of DMSO to MIP were performed and a binding mode within the active site was suggested which is in good agreement with NMR shift perturbation data. The sulfoxide substructure was subsequently extended to yield cyclic sulfoxides, **CF-1** and **CF-2**, which show favorable and consistent binding modes. Interestingly, tetrahydrothiophenoxide (**CF-1**) and its derivative **CF-2** are described in the literature as promising agents against alcohol intoxication, by inhibiting the liver alcohol dehydrogenase. In a fragment-based design approach, a subpocket which is close to the binding domain of **CF-2** has

been addressed by means of a pharmacophore-based virtual screening for fitting fragments. Taking care of the synthetic feasibility, two fragments (**F1**, **F2**) with a molecular weight below 120 Da were discovered, which are able to build a strong hydrogen bonding network within the subpocket. In the case of FKBP12, this subpocket is obstructed by a histidine residue; therefore, finding fragments which fill the MIP subpocket is a first step towards the development of MIP-selective inhibitors. Finally, the core fragment **CF-2** and the subpocket fragments were connected, yielding two molecules **CF2-F1** and **CF2-F2**, which show an overall good match between their molecules' substructures and the parent fragments. The synthesis of the fragments and the connected molecules is currently elaborated. Knowledge of *Legionella* MIP inhibitors could be transferred also to other organisms (e.g. trypanosomes), where homologous MIP proteins are also pathological factors.

4.2 Zusammenfassung

Die vorliegende Arbeit thematisiert das computergestützte strukturbasierte Design auf dem Gebiet der HIV-1-Integrase und des Macrophage Infectivity Potentiator (MIP) von *Legionella pneumophila*. In der gängigen Behandlung der HIV-Infektion bzw. der Legionellose spielen beide Enzyme bisher eine untergeordnete Rolle. Daher ist die Entwicklung neuartiger Inhibitoren dieser Enzyme von entscheidender Bedeutung. Die vorliegende Arbeit gibt wertvolle Aufschlüsse über den Wirkmechanismus einer bekannten Integrase-Inhibitorenklasse and zeigt darüber hinaus einen neuartigen Ansatz zur Integrase-Inhibition auf. Im Falle des MIP-Enzyms konnten zwei niedermolekulare Inhibitoren ermittelt werden; ferner wurde mit Hilfe eines fragmentbasierten Ansatzes das Design weiterer Inhibitoren vorangetrieben.

Die Integrase-Studien ergaben wertvolle Informationen im Hinblick auf das Design neuer Inhibitoren. Docking-Experimente konnten die Hypothese weiter untermauern, nach der die Klasse der Diketosäure-Inhibitoren nicht als freie Liganden, sondern als Metallion-Komplexe an das aktive Zentrum der Integrase binden. Bei den Docking-Experimenten wurden verschiedene Komplexierungs-Muster analysiert; sowohl die eingehende Untersuchung der Bindemoden als auch die Betrachtung der Docking-Energien legen jedoch nahe, dass monomere Komplexe mit einem Metall-Ligand-Verhältnis von 1:1 die bevorzugte Spezies sind. Die Ergebnisse dieser Studie helfen dabei, das Verständnis über den Wirkmechanismus dieser wichtigen Klasse von Integrase-Inhibitoren weiter zu vertiefen.

Inhibitoren des aktiven Zentrums (wie beispielsweise die Klasse der Diketosäuren) standen und stehen seit langem im Zentrum der Integrase-Forschung. Um der Entwicklung von Integrase-Inhibitoren einen neuen Impuls zu geben, wurde eine neue Strategie zur Inhibition dargelegt: Anstatt an das aktive Zentrum soll eine neue Inhibitor-Klasse an das Dimerisierungs-Interface eines Integrase-Monomers binden, die katalytisch notwendige Dimerisierung verhindern und somit die enzymatische Aktivität stören. Das Hauptproblem hierbei bestand in den fehlenden Strukturdaten des freien Monomers. Hierzu wurden Molekulardynamik-Simulationen durchgeführt, um nähere strukturelle Informationen zu erhalten. Aus der MD-Simulation abgeleitete Momentaufnahmen unterschiedlicher Konformationen („Snapshots“) dienten als Input-Strukturen für eine Docking-Studie mit dem bekannten peptidischen Inhibitor YFLLKL, um dessen Bindemodus aufzuklären. Hierbei zeigte sich, dass dieser peptidische Ligand an eine Interface-Konformation bindet, die durch eine Y-förmige Bindestelle aus zwei Kavitäten und einem Graben charakterisiert ist. Im nächsten Schritt sollte diese Protein-Konformation mit kleinen, nicht-peptidischen Molekülen adressiert werden. Da das Peptid YFLLKL einen überwiegend hydrophoben Charakter hat und hydrophobe HotSpots innerhalb des Dimerisierungs-Interface detektiert wurden, bestand die erste Strategie darin, ein Pharmakophor-Modell zu erstellen, das zur Suche nach Molekülen mit einer guten Komplementarität zur Y-förmigen Bindetasche geeignet ist. Das folgende virtuelle Screening ergab zehn Verbindungen, die eine gute Komplementarität und günstige hydrophobe Wechselwirkungen aufwiesen. Leider zeigte keine der Verbindungen eine reproduzierbare Aktivität im Integrase-Assay. Hierbei verbleiben jedoch gewisse Zweifel, da in dem Assay die Zugabe von BSA vorgeschrieben war, das möglicherweise die hydrophoben Inhibitor-Kandidaten gebunden und somit deren Wechselwirkung mit der Integrase verhindert hat.

Außerdem muss angemerkt werden, dass nicht selten die Hit-Raten von virtuellen Screenings bei 1 - 2 % liegen, was bedeuten würde, dass von hundert getesteten Verbindungen gerade einmal ein bis zwei aktiv sind. Hierbei können mehrere Faktoren eine Rolle spielen, z. B. unzureichend genaue Strukturinformationen über das Zielprotein. Die verwendeten Methoden wie beispielsweise das Docking können beim virtuellen Screening insbesondere von großen, diversen Moleküldatenbanken an ihre Grenzen stoßen, wenn es darum geht, die Verbindungen als bindend / nicht-bindend zu klassifizieren.

Die erwähnte erste Strategie wurde überdacht: In einem zweiten Ansatz galt die Hauptaufmerksamkeit der Absättigung von wasserstoffbrückenbildenden Resten. Diese waren zuvor von den eher hydrophoben Verbindungen nicht optimal abgesättigt worden. Zwei Pharmakophor-Modelle wurden erstellt und in einem virtuellen Screening eingesetzt: Hieraus ergaben sich mehrere hundert Hits. Docking-Studien dieser Hits zeigten jedoch, dass nach wie

vor viele wasserstoffbrückenbildende Reste des Proteins nicht vom Liganden abgesättigt wurden. Daraufhin wurde eine systematische Studie des Cheng-Datensatzes (umfasst 195 Protein-Ligand-Komplexe) durchgeführt um zu klären, wie oft solche Nicht-Absättigungen in Protein-Ligand-Komplexen überhaupt zu finden sind. Tatsächlich zeigte sich, dass in 90 % aller Fälle sowohl auf Seiten des Proteins als auch des Liganden weniger als vier nicht-abgesättigte wasserstoffbrückenbildende Gruppen vorkamen. Momentan wird diese Studie im Arbeitskreis von Prof. Sotriffer weiter ausgebaut und auf die Untersuchung polarer Protein-Ligand-Wechselwirkungen, die keine Wasserstoffbrücke ergeben, ausgedehnt. Ziel ist es, daraus einen Deskriptor abzuleiten, der ungünstige Docking-Bindemoden detektiert.

Für die durchgeführte Pharmakophor-Suche wurde der Grenzwert von vier nicht-abgesättigten Wasserstoffbrücken-Akzeptoren / -Donoren auf die Docking-Posen angewendet: 90 % aller Hits mussten so verworfen werden. Nach abschließender eingehender Betrachtung der Bindemoden der verbliebenen Moleküle konnten nur acht für weitere Testungen ausgewählt werden (Ergebnisse der experimentellen Testung durch Kooperationspartner stehen noch aus). Diese geringe „Ausbeute“ an geeigneten Verbindungen für das Integrase-Dimerisierungsinterface zeigt, wie schwer dieses Target zu adressieren ist: Das Interface weist eine schnell wechselnde Abfolge von basischen, sauren und hydrophoben Resten auf. Im Gegensatz zu anderen Protein-Protein-Interfaces zeigt das Integrase-Interface keine „aufgeräumte“ Bindetasche mit klar voneinander getrennten hydrophoben und hydrophilen Bereichen, sondern eine unregelmäßige Anordnung verschiedenartiger Reste. Neben den generellen Schwierigkeiten bei der Adressierung von Protein-Protein-Interfaces macht diese „Unordnung“ die Entwicklung von Dimerisierungs-Inhibitoren der Integrase zu einer großen Herausforderung. Dennoch lohnt sich der Aufwand, wenn man bedenkt, dass hieraus eine völlig neue Klasse von Integrase-Inhibitoren hervorgehen kann. Das Integrase-Projekt war auf das Auffinden von vorgefertigten, käuflichen Verbindungen ausgerichtet. In Anbetracht der Besonderheiten des Integrase-Interfaces lohnen sich Überlegungen, nach denen solch ein Target eher maßgeschneiderte Verbindungen verlangt, die diesen Besonderheiten besser gerecht werden. Auch für diese Möglichkeit wurden durch die hier durchgeführten Studien wertvolle Informationen erhalten: Fragmente wie beispielsweise Triaminobenzol oder Pyrimidin-Harnstoffderivate, die als Substruktur der Hit-Moleküle aus dem virtuellen Screening interessante Bindemoden zeigten, können als Startstrukturen für die zielgerichtete Synthese dienen.

Für das zweite hier betrachtete Enzym, MIP, konnten mit Hilfe des computergestützten strukturbasierten Designs zwei niedermolekulare Inhibitoren, **MI-1** und **MI-2**, gefunden werden. Beide Verbindungen führten zu einer deutlichen Abnahme der katalytischen Aktivität im Assay.

Soweit bekannt, sind bisher keinerlei niedermolekulare MIP-Inhibitoren veröffentlicht worden. Die vielversprechende Verbindung **MI-1** wird nun im nächsten Schritt auch in einem zellbasierten Assay auf ihre anti-MIP-Aktivität getestet.

Die detaillierte Analyse der verfügbaren MIP-Strukturen und der Vergleich mit einer humanen PPIase ergaben, dass es zwischen den verschiedenen MIP-Strukturen (NMR, Röntgenstruktur) eine teils große konformationelle Vielfalt gibt. Der Vergleich von MIP mit der humanen PPIase FKBP12 zeigte eine größtenteils ähnliche Tasche, die jedoch einen entscheidenden Unterschied aufweist, nämlich in der Orientierung des Restes Tyr109. Hierdurch wird die MIP-Bindetasche im Vergleich zur FKBP12-Tasche wesentlich verengt. Die detaillierte Betrachtung der Strukturdaten beider Enzyme konnte schließlich eine Erklärung liefern, warum ein ketoacyl-substituiertes Pipecolinderivat mit hoher Wahrscheinlichkeit nicht an MIP bindet, ein sulfonsubstituiertes Pipecolinderivat hingegen das Enzym inhibiert.

Die experimentellen Validierungen wurden mit Hilfe eines MIP-Inhibitionsassays durchgeführt. Hierbei zeigte sich, dass das Enzym sehr empfindlich auf die Zugabe von DMSO reagiert. Die Frage stellte sich nun, ob die beobachtete Abnahme der katalytischen Aktivität bei Zugabe des Lösungsmittels auf eine teilweise Denaturierung oder auf eine spezifische Bindung des DMSOs zurückzuführen ist. Docking-Studien wurden durchgeführt, die stark darauf hindeuten, dass DMSO tatsächlich an das aktive Zentrum von MIP bindet und das Enzym dadurch inhibiert. Der Bindemodus ist im Einklang mit Daten aus NMR-Experimenten. Die Sulfoxid-Substruktur wurde zu zyklischen Sulfoxiden (**CF-1** und **CF-2**) ausgebaut: In beiden Fällen zeigten sich konsistente, vielversprechende Bindemoden. Ein weiterer interessanter Aspekt an diesen Tetrahydrothiophen-Derivaten ist, dass beide bereits als vielversprechende Inhibitoren der Alkoholdehydrogenase erwähnt wurden.

Ein fragmentbasierter Ansatz diente im nächsten Schritt zur Erweiterung dieser Sulfoxid-Kernfragmente: Für eine Subtasche, die nahe der Bindestelle von **CF-2** liegt, wurden mit Hilfe eines pharmakophor-basierten virtuellen Screenings passende Fragmente gesucht. Hierbei wurde Wert auf eine spätere Realisierbarkeit in der Synthese gelegt. Zwei Fragmente mit einem Molekulargewicht unter 120 Da wurden ausgewählt, die sich durch günstige Wasserstoffbrücken-Bindungen innerhalb der Subtasche auszeichneten. Die Adressierung dieser Subtasche ist gerade deshalb interessant, weil das FKBP12-Pendant durch einen Histidin-Rest verschlossen ist: Hieraus lässt sich eine Strategie zur Entwicklung selektiver Inhibitoren ableiten. Im letzten Schritt wurde das Kernfragment mit den Subtaschen-Fragmenten verbunden; die resultierenden Moleküle wurden abermals einer Docking-Studie unterzogen und zeigten eine gute Übereinstimmung in der Platzierung ihrer Substrukturen mit denen der jeweiligen Fragmente. Die Synthese dieser Verbindungen wird zur Zeit von Kooperationspartnern ausgearbeitet. Die

Erkenntnisse über das Inhibitoren-Design für *Legionella*-MIP können auch auf andere Organismen (z.B. Trypanosomen) übertragen werden, bei denen ebenfalls das (homologe) MIP ein wesentlicher Pathogenitätsfaktor ist.

5. Appendix

5.1 Methodologies

Molecular representations – SMILES and SMARTS codes

SMILES (simplified molecular input line entry system) is a simple chemical language which describes molecules using ASCII characters representing atoms and bonds. The SMILES notation is a fast and space-saving method to encode molecules. Atoms are represented by the respective letter (C, O, N, etc.); single bonds do not have to be specified, whereas double and triple bonds have their own character (“=” and “#”, respectively). For example, the SMILES code for ethane is simply “CC” (hydrogens are omitted), the code for formaldehyde is “C=O”. These linear molecule codes can be transferred to actual three-dimensional structures with programs like CORINA (“*coordinates*”), which automatically builds low-energy conformations. The SMARTS (“SMILES arbitrary target specification”) language is a SMILES-related language, which was designed especially to find substructures within molecules. A simple expression is “[Cl]”, which stands for any chlorine atom within a given compound. SMARTS-based programs can thus search molecular databases for pre-defined substructures¹⁵².

Fingerprints

Molecular fingerprints encode information about the presence or absence of distinct features or fragments (e.g. aromatic rings, conjugated systems, halogens, etc.) within a molecular structure. Depending on the fingerprint scheme, different features of a molecule are emphasized. The goal is to convert a molecular structure into a simple descriptor set, which allows straightforward, computationally inexpensive comparison of molecules. One of the most prominent fingerprint schemes is the MACCS keys scheme. In the publicly available version, which is used in the present work, it comprises 166 keys. To compare fingerprints (and thereby the parent molecules), a similarity metric has to be introduced: A common metric is the Tanimoto coefficient, ranging from 0 (maximum dissimilarity) to 1 (maximum similarity). Fingerprint-based similarity measurements within a given set of molecules allow clustering: Here, compounds being highly similar to each other are pooled in one cluster. If in the next step one representative per cluster is extracted, then only those molecules from the original dataset are retrieved which are structurally diverse¹⁵³.

Hot-spot analysis

Given the fact that a protein's active site can be regarded as a negative imprint of the binding ligand, a pharmacophore model can be created based on the knowledge of crucial active site interaction points, the so-called hot-spots. Such energetically favorable interaction sites can be found with programs like GRID or DrugScore's HotSpot analysis. GRID puts the protein structure onto a 3D grid and calculates interaction energies of a probe (e.g., sp^3 carbon or hydroxy group) at each grid point. Using different probes, favorable positions (hot-spots) for the most important pharmacophore features within the protein's active site can be deduced and subsequently used to create a pharmacophore model. The DrugScore HotSpot analysis takes a similar approach: The protein is also placed on a 3D grid and the interactions of a given probe are evaluated on each grid point. However, this HotSpot analysis relies on the DrugScore knowledge-based pseudo-potentials derived from crystallographic data (cf. DrugScore as a scoring function). Nevertheless, this method also gives information about favorable positions of pharmacophoric features, which can be exploited in the creation of a protein-based pharmacophore model^{112-114, 118-121}.

Pharmacophore modelling using MOE

The MOE pharmacophore annotation scheme automatically derives pharmacophore annotation points (hydrogen bond donor / acceptor, hydrophobic, etc.) of a given molecule. Based on these annotation points, pharmacophore features can be set which finally constitute the pharmacophore model (pharmacophore query). MOE has numerous annotation points, the most common being hydrogen bond donor ("Don") and acceptor ("Acc"), hydrophobic ("Hyd") and aromatic ("Aro"), or cationic ("Cat") and anionic ("Ani"). The hydrogen bonding annotation points are completed by so-called projected annotations, which imply the location of their donating / accepting partners, i.e., they are located along lone pairs ("Acc2", "Don2"). Thereby, the proper orientation of hydrogen bonding groups can be incorporated into the pharmacophore model. Hydrophobic annotation points are assigned as either atom-centered ("HydA") or as centroids ("Hyd") of a group of hydrophobic atoms.

Using the pharmacophore query editor, the annotation points of interest can be selected and transformed to pharmacophore features, which finally constitute the pharmacophore model. The different features are represented by spheres, centered on the corresponding annotation point; the radii are freely adjustable, yielding a more or less restrictive pharmacophore model. The

features can be further refined by attributing SMARTS expressions; e.g., if a donor feature should be matched only by a nitrogen atom, one would extend this feature to "Don&[N]". Logical operators "|", "&", "!" (meaning "or", "and", "not", respectively) can also be employed. In addition to the annotation point features, spatial constraints (volumes) with individual radii can be set. These volumes are of three types: An included volume constraint ("V incl.") requires at least one atom to be located within this volume, an excluded volume ("V excl.") requires that no atom lies within this volume, and an exterior volume ("V ext.") is satisfied if no atom is located outside this volume. With these definitions, the pharmacophore query can be further refined to yield ligands with a specific shape. Volumes can be defined as single spheres or as a union of individual spheres (the latter allows easier handling of the volume constraints). The pharmacophore query can be further modified by setting constraints. Groups of features can be constrained ("C") by different expressions: For example, a group of three features ("F[1,2,3]") can be attributed with the "at least one" ("at least two", etc.) constraint type, meaning that at least one out of these three features needs to be matched. Further constraint types are "exact one" ("exact two", etc.), "all or none", and "same atoms" (meaning that the features must be derived from the same atom, e.g. donor feature / projected donor feature). MOE pharmacophore modelling enables partial matching, i.e., not all query features need to be matched. A threshold can be set which specifies the minimum number of features that must be matched. Features that must be matched in any case, even in partial matching, can be attributed with the "essential" ("ess.") keyword. Once the pharmacophore model has been finished, it can serve as query to screen a database: Annotation points of the database molecules are generated on-the-fly, and compounds matching the pharmacophore query are finally written to an output database¹⁰⁹.

Docking

Docking is a two-step procedure: First, different orientations and conformations of a ligand binding to a given protein are generated (sampling). Second, these orientations/conformations, the binding modes, are assessed (scoring). For both sampling and scoring, a wide range of methods is available.

Sampling methods: Given the huge conformational space of a molecule and the myriad of possible binding orientations within the protein active site, robust methods have to be applied which search the conformational / orientational space sufficiently within a reasonable amount of time. One example of such search methods is the so-called genetic algorithm (GA), which is

implemented in the docking tools AutoDock and GOLD. GAs are based on Darwin's theory of evolution. The information of the binding mode, i.e., the ligand's conformational (intramolecular) as well as rotational and translational (with respect to the protein) state is encoded in a "chromosome" by means of the respective coordinates, angles, etc. Each binding mode is represented by an individual with its own chromosome. Just like evolution, these individuals undergo reproduction, while genetic operations like crossover or mutation change the genome of the individuals' progeny (i.e. representing new binding modes). In analogy to Darwin's "survival of the fittest", only the lowest-energy binding mode is retained and transferred to the next generation. After evolving a distinct number of generations, the "fittest" binding mode, i.e., the binding mode with most favorable properties is retained.

Scoring methods: The assessment of the docked binding modes is done with scoring functions. The ultimate scoring function should be capable of 1) rank a near-native pose of a given ligand best and 2) rank different ligands according to their predicted binding affinity correctly, thereby retrieving active compounds from a set of decoys. This goal still remains to be achieved, even if in several cases the available scoring functions show good results. While most scoring functions are implemented in the parent software packages (e.g. AutoDock scoring function, GOLDscore), there are also standalone scoring functions available (DrugScore^{CSD}, SFCscore). These standalone scoring functions are useful in re-scoring, i.e. the re-assessment of docking poses generated by arbitrary docking tools or for consensus scoring, where two or more scoring functions are used to score a given docking solution. DrugScore^{CSD} belongs to the class of the so-called knowledge-based scoring functions, which have been developed by statistical analysis of crystal structures. The analysis yields information about the distribution of distances between atom-type pairs. From this distribution, an artificial score is derived for each pair of atom-types. The final score is the sum over all atom-type pair contributions. SFCscore is trained on experimental affinities and thus belongs to the class of empirical scoring functions. Here, the total score is the sum of individual, scalable contributions. For this, descriptors are used, reflecting interaction properties of the ligand pose to be assessed, for example the number of hydrogen bonds, and buried polar surface, to name a few.

AutoDock: Before the ligand is actually docked, AutoDock precalculates the interaction energies of each ligand atom type: The *autogrid* program embeds the protein in a three-dimensional grid, places a probe atom at each grid point, and calculates the interaction energies. This energy grid is computed for each ligand atom type (carbon, nitrogen, oxygen, hydrogen, etc.). Additionally, an electrostatic potential grid is computed by using a point charge of +1. The affinity of a given ligand orientation within the protein can then be evaluated by interpolation of the eight grid

points surrounding each ligand atom. This grid-based procedure allows a speed-up of the docking algorithm, since the interaction energies do not have to be calculated from a sum over all protein-ligand atom pairs. The AutoDock scoring function is an empirical free-energy function including the following terms:

$$\Delta G = \Delta G_{\text{vdW}} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \Delta G_{\text{hbond}} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{\text{hbond}} \right) \\ + \Delta G_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + \Delta G_{\text{tor}} N_{\text{tor}} + \Delta G_{\text{sol}} \sum_{i_c, j} S_i V_j e^{(-r_{ij}^2/2\sigma^2)}$$

This free energy function is the sum over all pairwise protein-ligand contacts with the distance r_{ij} , it comprises terms for van-der-Waals interactions (Lennard-Jones 12-6 potential), a hydrogen bonding term (directional 12-10 potential, where $E(t)$ is a directional angle-dependent weight and E_{hbond} is the estimated average energy of hydrogen bonding of water with a polar atom, i.e. desolvation of polar ligand atoms), an electrostatic term (distant-dependent dielectric constant $\epsilon(r_{ij})$), a measure of entropy loss upon ligand binding (proportional to the number of rotatable bonds within the ligand, N_{tor}), and a fifth term modelling desolvation and the hydrophobic effect (applicable to carbon ligand atoms, solvation parameter S_p and fragmental volume of surrounding protein V_j). The ΔG coefficients were empirically determined by linear regression analysis of a protein-ligand complex dataset with known affinities.

AutoDock requires a protein structure with polar hydrogens. A fully protonated ligand bearing Gasteiger-Marsili charges is transformed to a ligand structure where non-polar hydrogens are merged with their heavy atom partners: This is done by the *autotors* tool, which also specifies rotatable bonds within the ligand. For the docking stage (pose generation), AutoDock3.0 employs an extended genetic algorithm: The GA-driven global search for a favorable binding pose is supplemented by the Solis and Wets (SW) local search method, which optimizes global search binding poses. In terms of the genetic algorithm, this means a modification of the phenotype, which itself can change the genotype; this notion refers to Lamarckian genetics, therefore this modified GA is termed Lamarckian genetic algorithm (LGA). After a user-defined number of LGA runs, AutoDock outputs the results and clusters them according to their mutual RMSD values.

GOLD: For pose generation, GOLD uses so-called fitting points to initially place the ligand within the protein active site. First, it adds fitting points to hydrogen bonding groups on protein and ligand. Then, acceptor points on the ligand are mapped onto protein donor points and vice

versa. Additionally, GOLD creates hydrophobic fitting points within the protein active site, onto which hydrophobic ligand atoms are mapped. GOLD requires fully protonated protein and ligand structures, yet without any charges. Formal charges are automatically derived from the protonation states of the respective atoms. The GOLD scoring function has the form:

$$GOLDscore = S(hbond) + S(van-der-Waals_ext) + S(van-der-Waals_int)$$

In contrast to the AutoDock scoring function, the GOLD scoring function does not yield calculated free energies, but artificial scores: The better the binding pose, the higher the (positive) score. The hydrogen bond term $S(hbond)$ is calculated as the sum of all protein-ligand hydrogen bonds, including a penalty for desolvation. Hydrophobic interactions between protein and ligand are rendered by the $S(van-der-Waals_ext)$ term, whereas the ligand internal strain is included in the $S(van-der-Waals_int)$ expression. In contrast to the AutoDock scoring function, GOLDScore waives an electrostatic contribution^{80, 111-115}.

Molecular dynamics simulations

Molecular dynamics (MD) simulations are employed to explore the dynamics and thereby the conformational space of a molecular system (e.g. a protein). MD simulations are based on solving Newton's equations of motion for every atom of the system, the associated potential function is rendered by force fields. Force fields include terms for both bonded interactions (bond, bond angle, dihedral contributions) and non-bonded interactions (van-der-Waals, electrostatic contributions). The general form of a force field with its individual terms is given by:

$$E_{total} = \sum_{bonds} K_b(b - b_{eq})^2 + \sum_{angle} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

E_{total} is the sum over all individual contributions. K_b and K_θ are the force constants of the bond and bond angle, respectively; the equilibrium bonds length and bond angle is given by b_{eq} and θ_{eq} ; dihedrals are characterized by the force constant V_n and the angle ϕ (γ is the phase angle with values of either 0° or 180°). Non-bonded interactions between atoms i and j are represented by a Lennard-Jones potential (with parameters A_{ij} and B_{ij}) and an electrostatic term with atomic charges q_i and q_j and the dielectric constant (usually 1.0 in systems with explicit solvent). All parameters are derived from experimental and theoretical data. Depending on the system to be simulated (protein, small molecule, nucleic acid, etc.), different parameters (i.e. different force

fields) are available. An appropriate force field for simulating proteins is the AMBER 2003 force field, which uses an all-atom representation, i.e., both heavy atoms and hydrogens are considered (in contrast to united-atom force fields, where non-polar hydrogens are merged with their heavy atoms).

Practically, starting from an existing structure (for example a PDB crystal structure), the protein is solvated and counterions are added to simulate the aqueous environment. The solvent can be explicitly modelled by adding water molecules or, otherwise, the solvent effect can be implicitly modelled by using implicit solvent models. The first method is computationally more expensive, but at the same time more accurate. Either way, the solvated system is subsequently minimized to remove close contacts or improper geometries, which otherwise would lead to unnaturally high energies during the actual MD simulation. After randomly assigning an initial velocity to every atom (according to the Boltzmann distribution at a given temperature, in general 300 K), the dynamics of the system evolve for a user-defined amount of time. Pressure and temperature of the system are continuously monitored and maintained at a nearly constant level by implementing appropriate algorithms. After the MD simulation has finished, AMBER's *ptraj* module offers a plethora of analysis tools to retrieve valuable information about the system's dynamics, for example time-dependent atom-atom distances or atomic fluctuations, to name a few^{103,104}.

5.2 Abbreviations

³² P-ATP	: ³² P-labelled adenosine triphosphate
2D-RMSD	: two-dimensional root mean square deviation
3'-P	: 3'-processing
5-CITEP	: 1-(5-chloroindol-3-yl)-3-(tetrazolyl)-1,3-propadione
aa	: acetyl-acetate
AIDS	: acquired immunodeficiency syndrome
AMBER	: assisted model building with energy refinement
APS	: ammonium persulfate
ASV	: avian sarcoma virus
BAF	: barrier-to-autointegration factor
Bcl	: B-cell lymphoma
BME	: beta-mercaptoethanol
BSA	: bovine serum albumine
CA	: cytidine-adenosine
CAPE	: caffeic acid phenethyl ester
cav1/2	: cavity 1 / 2
CCD	: catalytic core domain
CCG	: Chemical Computing Group
CCR5	: C-C chemokine receptor type 5
CD4	: cluster of differentiation 4
CORINA	: coordinates
CSD	: Cambridge structural database
CTD	: C-terminal domain
DC	: DNA control
DI	: dimerization inhibitor
DKA	: diketo acid
DNA	: deoxyribonucleic acid
ECM	: extracellular matrix
e.g.	: <i>exempli gratia</i>
et al.	: <i>et alii</i>
etc.	: <i>et cetera</i>
FIV	: feline immunodeficiency virus
FKBP12	: FK506-binding protein

GA	: genetic algorithm
GOLD	: genetic optimization of ligand docking
gp41/120	: glycoprotein 41 / 120
GT	: guanosine-thymidine
HAART	: highly active antiretroviral therapy
HBAs/HBDs	: hydrogen bond acceptors / hydrogen bond donors
hc	: hydroxy-carboxylate
HDM2	: human MDM2 (human mouse double minute gene 2)
hDNA	: host desoxyribonucleic acid
HIV	: human immunodeficiency virus
HMGA	: high mobility group chromosomal protein A1
HPV	: human papilloma virus
HSQC	: hetero single-quantum coherence
IC	: integrase control
IC ₅₀	: half-maximal inhibitory concentration
i.e.	: id est
IL-2	: interleukine-2
IN	: integrase
INH1/5	: integrase helix 1 / 5
INI1	: integrase interactor 1
INSTI	: integrase strand transfer inhibitor
LCA	: L-chicoric acid
LEDGF	: lens epithelium growth factor
LTR	: long terminal repeats
M:L	: metal-to-ligand ratio
MACCS	: molecular access system
MD	: molecular dynamics
MIP	: macrophage infectivity potentiator
MOE	: molecular operating environment
MOPS	: 3-(N-morpholino)propanesulfonic acid
mTOR	: mammalian target of rapamycin
n.a.	: not available
NIA	: non-immunosuppressive agents
NMR	: nuclear magnetic resonance
Nrotb	: number of rotatable bonds

NTD	: N-terminal domain
PAGE	: polyacrylamide gel electrophoresis
PDB	: protein data bank
PIC	: pre-integration complex
PNK	: polynucleotide kinase
PPIase	: peptidyl-prolyl <i>cis/trans</i> isomerase
PR	: protease
PSA	: polar surface area
QLZ	: quinalizarin
RMSD	: root mean square deviation
RNA	: ribonucleic acid
RSV	: rous sarcoma virus
RT	: reverse transcriptase
SANDER	: simulated annealing using NMR-derived energy restraints
SFC	: scoring function consortium
SMILES	: simplified molecular input line entry specification
SMARTS	: SMILES arbitrary target specification
SQL	: styryl quinoline
ST	: strand transfer
SVL	: scientific vector language
TBE (buffer)	: tris-borate-EDTA (tris-borate ethylenediaminetetraacetic acid)
TEMED	: tetramethylenediamine
TMS	: tetramethyl spiro-bisindan
TNF	: tumor necrosis factor
USC	: University of Southern California
vDNA	: viral desoxyribonucleic acid
WHO	: World Health Organisation
ZINC	: zinc is not commercial

5.3 Publications

5.3.1 Articles

Bacchi, A.; Biemmi, M.; Carcelli, M.; Carta, F.; Carlotta, C.; Fiscaro, E.; Rogolino, D.; Sechi, M.; Sippel, M.; Sotriffer, C.; Sanchez, T. W.; Neamati, N. From Ligand to Complexes. Part 2. Remarks on Human Immunodeficiency Virus Type 1 Integrase inhibition by Diketo Acid Metal Complexes. *J. Med. Chem.* **2008**, *51*, 7253–7264.

Sippel, M.; Sotriffer, C. Molecular Dynamics Simulations of the HIV-1 Integrase Dimerization Interface: Guidelines for the Design of a Novel Class of Integrase Inhibitors. *J. Chem. Inf. Model.* **2010**, *50*, 604-614.

5.3.2 Poster presentations

Sippel, M.; Sotriffer, C. Legionnaires' Disease and *Legionella* MIP: Fragment-based design and synthesis of the first small-molecule MIP inhibitor. International SFB symposium, Würzburg, **2009**.

Sippel, M.; Sotriffer, C. Dimerisation Disruptors: A Novel Class of HIV-1 Integrase Inhibitors. DECHEMA 5th Status Seminar Chemical Biology, Frankfurt, **2008**.

Sippel, M.; Sotriffer, C. HIV-Integrase als Zielenzym – Computergestütztes Design neuer Inhibitoren. GDCh Jungchemiker-Forum Chem-SyStM, Würzburg, **2007**.

Sippel, M.; Sotriffer, C. Targeting HIV-1 Integrase: Discovering peptidomimetics as a novel class of inhibitors. Darmstädter Molecular Modelling Workshop, Erlangen, **2007**.

Sippel, M.; Klebe, G.; Sotriffer, C. Targeting HIV-1 Integrase: Investigating a new binding site for inhibitor design. Summer School Medicinal Chemistry, Regensburg, **2006**.

5.3.3 Oral presentations

Sippel, M.; Sotriffer, C. Legionnaires' Disease and *Legionella* MIP: Fragment-based design and synthesis of the first small-molecule MIP inhibitor. International SFB symposium, Würzburg, **2009**.

Sippel, M.; Sotriffer, C. Dimerisation Disruptors: A Novel Class of HIV-1 Integrase Inhibitors. DECHEMA 5th Status Seminar Chemical Biology, Frankfurt, **2008**.

Sippel, M.; Sotriffer, C. HIV-Integrase als Zielenzym – Computergestütztes Design neuer Inhibitoren. GDCh Jungchemiker-Forum Chem-SyStM, Würzburg, **2007**.

Sippel, M.; Sotriffer, C. Targeting HIV-1 Integrase: Discovering peptidomimetics as a novel class of inhibitors. Darmstädter Molecular Modelling Workshop, Erlangen, **2007**.

Sippel, M.; Klebe, G.; Sotriffer, C. Targeting HIV-1 Integrase: Investigating a new binding site for inhibitor design. Summer School Medicinal Chemistry, Regensburg, **2006**.

6. Bibliography

- 1 Talele, T. T.; Khedkar, S. A.; Rigby, A. C. Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. *Curr. Top. Med. Chem.* **2010**, *15*, 127-141.
- 2 UNAIDS report on the global HIV/AIDS epidemic: executive summary **2008**.
- 3 Marx, J. L. New Disease Baffles Medical Community. *Science* **1982**, *217*, 618-621.
- 4 De Silva, T. I.; Cotten, M.; Rowland-Jones, S. L. HIV-2: the forgotten AIDS virus. *Trends Microbiol.* **2008**, *12*, 588-595.
- 5 Coffin, J.; Haase, A.; Levy, J. A.; Montagnier, L.; Oroszlan, S.; Teich, N.; Temin, H.; Toyoshima, K.; Varmus, H.; Vogt, P. What to call the AIDS virus? *Nature* **1986**, *321*, 10.
- 6 Campbell, J. Biologie. *Spektrum Verlag* **2006**.
- 7 Flexner, C. HIV drug development: the next 25 years. *Nat. Rev. Drug Discov.* **2007**, *6*, 959-966.
- 8 Matthews, T.; Salgo, M. Enfuvirtide: The First Therapy to Inhibit the Entry of HIV-1 into Host CD4 Lymphocytes. *Nat. Rev. Drug Discov.* **2004**, *3*, 215-225.
- 9 Fischl, M. A.; Richman, D. D.; Grieco, M. H.; Gottlieb, M. S.; Volberding, P. A.; Laskin, O. L.; Leedom, J. M.; Groopman, J. E.; Mildvan, D.; Schooley, R. T. The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial. *N. Engl. J. Med.* **1987**, *317*, 185-191.
- 10 Cahn, P.; Sued, O. Raltegravir: a new antiretroviral class for salvage therapy. *Lancet* **2007**, *369*, 1235-1236.
- 11 Martin, J. A.; Redshaw, S.; Thomas, G. J. Inhibitors of HIV Proteinase. *Prog. Med. Chem.* **1995**, *32*, 239-288.
- 12 May, M. T.; Sterne, J. A.; Costagliola, D.; Sabin, C. A.; Phillips, A. N.; Justice, A. C.; Dabis, F.; Gill, J.; Lundgren, J.; Hogg, R. S.; de Wolf, F.; Fätkenheuer, G.; Staszewski, S.; d'Arminio Monforte, A.; Egger, M.; HIV treatment response and prognosis in Europe and North America in the first decade of highly active antiretroviral therapy: a collaborative analysis. *Lancet* **2006**, *368*, 451-458.
- 13 Pommier, Y.; Johnson, A. A.; Marchand, C. Integrase Inhibitors to treat HIV/AIDS. *Nat. Rev. Drug Discov.* **2005**, *4*, 236-248.
- 14 Rice, P.; Craigie, R.; Davies, D. R. Retroviral integrase and their cousins. *Curr. Opin. Struct. Biol.* **1996**, *6*, 76-83.
- 15 Chiu, T. K.; Davies, D. R. Structure and Function of HIV-1 Integrase: An Update. *Frontiers Med. Chem.* **2006**, *3*, 3-22.

- 16 Vink, C.; van Gent, D. C.; Elgersma, Y.; Plasterk, R. H. Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. *J. Virol.* **1991**, *65*, 4636-4644.
- 17 Turlure, F.; Devroe, E.; Silver, P. A.; Engelman, A. Human cell proteins and human immunodeficiency virus DNA integration. *Front. Biosci.* **2004**, *9*, 3187-3208.
- 18 Vink, C.; Groenink, M.; Elgersma, Y.; Fouchier, R. A.; Tersmette, M.; Plasterk, R. H. Analysis of the junctions between human immunodeficiency virus type 1 proviral DNA and human DNA. *J. Virol.* **1990**, *64*, 5626-5627.
- 19 Chow, S. A.; Vincent, K. A.; Ellison, V.; Brown, P. O. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* **1992**, *255*, 723-726.
- 20 Chiu, T. K.; Davies, D. R. Structure and function of HIV-1 integrase. *Curr. Top. Med. Chem.* **2004**, *4*, 965-977.
- 21 Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **1994**, *266*, 1981-1986.
- 22 Engelman, A.; Craigie, R. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J. Virol.* **1992**, *66*, 6361-6369.
- 23 Greenwald, J.; Le, V.; Butler, S. L.; Bushman, F. D.; Choe, S. The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry* **1999**, *38*, 8892-8898.
- 24 Yang, W.; Steitz, T. A. Recombining the structures of HIV integrase, RuvC and RNaseH. *Structure* **1995**, *3*, 131-134.
- 25 Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davies, D. R. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9150-9154.
- 26 Maignan, S.; Guilloteau, J. P.; Zhou-Liu, Q.; Clement-Mella, C.; Mikol, V. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *J. Mol. Biol.* **1998**, *282*, 359-368.
- 27 Goldgur, Y.; Craigie, R.; Cohen, G.H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D.R. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13040-13043.
- 28 Bujacz, G.; Alexandratos, J.; Wlodawer, A.; Merkel, G.; Andralke, M.; Katz, R. A.; Skalka, A. M. Binding of different cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J. Biol. Chem.* **1997**, *272*, 18161-18168.

- 29 Lins, R. D.; Straatsma, T. P.; Briggs, J. M. Similarities in the HIV-1 and ASV Integrase Active sites Upon Metal Cofactor Binding. *Biopolymers* **2000**, *53*, 308-315.
- 30 Jenkins, T. M.; Hickman, A. B.; Dyda, F.; Ghirlando, R.; Davies, D. R.; Craigie, R. Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6057-6061.
- 31 Zheng, R.; Jenkins, T. M.; Craigie, R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13659-13664.
- 32 Cai, M.; Zheng, R.; Caffrey, M.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* **1997**, *4*, 567-577.
- 33 Cai, M.; Huang, Y.; Caffrey, M.; Zheng, R.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. Solution structure of the His12 --> Cys mutant of the N-terminal zinc binding domain of HIV-1 integrase complexed to cadmium. *Protein Sci.* **1998**, *7*, 2669-2674.
- 34 Wang, J.-Y.; Ling, H.; Yang, W.; Craigie, R. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.* **2001**, *20*, 7333-7343.
- 35 Chen, J. C. H.; Krucinski, J.; Miercke, L. J. W.; Finer-Moore, J. S.; Tang, A. H.; Leavitt, A. D.; Stroud, R. M. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8233-8238.
- 36 Lodi, P. J.; Ernst, J. A.; Kuszewski, J.; Hickman, A. B.; Engelman, A.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* **1995**, *34*, 9826-9833.
- 37 Eijkelenboom, A. P. A. M.; Sprangers, R.; Hard, K.; Puras Lutzke, R. A.; Plasterk, R. H. A., Boelens, R.; Kaptein, R. Refined Solution Structure of the C-Terminal DNA-Binding Domain of Human Immunovirus-1 Integrase. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 556-564.
- 38 Wielens, J.; Crosby, I. T.; Chalmers, D. K. A three-dimensional model of the human immunodeficiency virus type 1 integration complex. *J. Comput. Aid. Mol. Des.* **2005**, *19*, 301-317.
- 39 Faure, A.; Calmels, C.; Desjobert, C.; Castroviejo, M.; Caumont-Sarcos, A.; Tarrago-Litvak, L.; Litvak, S.; Parissi, V. HIV-1 integrase crosslinked oligomers are active in vitro. *Nuc. Acids Res.* **2005**, *33*, 977-986.

- 40 Davies, D. R.; Goryshin, I. Y.; Reznikoff, W. S.; Rayment, I. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **2000**, *289*, 77-85.
- 41 Barreca, M. L.; De Luca, L.; Iraci, N.; Chimirri, A. Binding Mode Prediction of Strand Transfer HIV-1 Integrase Inhibitors Using Tn5 Transposase as a Plausible Surrogate Model for HIV-1 Integrase. *J. Med. Chem.* **2006**, *49*, 3994-3997.
- 42 Ason, B.; Knauss, D. J.; Balke, A. M.; Merkel, G.; Skalka, A. M.; Reznikoff, W. S. Targeting Tn5 Transposase Identifies Human Immunodeficiency Virus Type 1 Inhibitors. *Antimicrob. Agents Chemother.* **2005**, *49*, 2035-2043.
- 43 Ferro, S.; De Luca, L.; Barreca, M. L.; Iraci, N.; De Grazia, S.; Christ, F.; Witvrouw, M.; Debysers, Z.; Chimirri, A. Docking Studies on a New Human Immunodeficiency Virus Integrase-Mg-DNA Complex: Phenyl Ring Exploration and Synthesis of 1 H-Benzylindole Derivatives through Fluorine Substitutions. *J. Med. Chem.* **2009**, *52*, 569-573.
- 44 Liao, C.; Karki, R. G.; Marchand, C.; Pommier, Y.; Nicklaus, M. C. Virtual screening application of a model of full-length HIV-1 integrase complexed with viral DNA. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5361-5365.
- 45 Chen, X.; Tsiang, M.; Yu, F.; Hung, M.; Jones, G. S.; Zeynalzadegan, A.; Qi, X.; Jin, H.; Kim, C. U.; Lee, W. A.; Swaminathan, S.; Chen, J. M. Modeling, Analysis, and Validation of a Novel HIV Integrase Structure Provide Insights into the Binding Modes of Potent Integrase Inhibitors. *J. Mol. Biol.* **2008**, *380*, 504-519.
- 46 Molteni, V.; Greenwald, J.; Rhodes, D.; Hwang, Y.; Kwiatkowski, W.; Bushman, F. D.; Siegel, J. S.; Choe, S. Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain. *Acta Crystallogr. Sect. D* **2001**, *57*, 536-544.
- 47 Markowitz, M.; Morales-Ramirez, J. O.; Nguyen, B. Y.; Kovacs, C. M.; Steigbigel, R. T.; Cooper, D. A.; Liporace, R.; Schwartz, R.; Isaacs, R.; Gilde, L. R.; Wenning, L.; Zhao, J. Antiretroviral activity, pharmacokinetics, and tolerability of MK-0518, a novel inhibitor of HIV-1 integrase, dosed as monotherapy for 10 days in treatment-naive HIV-1 infected individuals. *J. Acquir. Immune Defic. Syndr.* **2006**, *43*, 509-515.
- 48 Summa, V.; Petrocchi, A.; Bonelli, F.; Crescenzi, B.; Donghi, M.; Ferrara, M.; Fiore, F.; Gardelli, C.; Gonzalez Paz, O.; Hazuda, D. J.; Jones, P.; Kinzel, O.; Laufer, R.; Monteagudo, E.; Muraglie, E.; Nizi, E.; Orvieto, F.; Pace, P.; Pescatore, G.; Scarpelli, R.; Stillmock, K.; Witmer, M. V.; Rowley, M. Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J. Med. Chem.* **2008**, *51*, 5843-5855.

- 49 Hazuda, D. J. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **2000**, *287*, 646-650.
- 50 Billich, A. S-1360 Shionogi-GlaxoSmithKline. *Curr. Opin. Investig. Drugs* **2003**, *4*, 206-209.
- 51 Hazuda, D. J.; Anthony, N. J.; Gomez, R. P.; Jolly, S. M.; Wai, J. S.; Zhuang, L.; Fischer, T. E.; Embrey, M.; Guare, J. P.; Egbertson, M. S.; Vacca, J. P.; Huff, J. R. A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11233-11238.
- 52 Sato, M.; Motomura, T.; Aramaki, H.; Matsuda, T.; Yamashita, M.; Ito, Y.; Kawakami, H.; Matsuzaki, Y.; Watanabe, W.; Yamataka, K.; Ikeda, S.; Kodama, E.; Matsuoka, M.; Shinkai, H. Novel HIV-1 Integrase Inhibitors Derived from Quinolone Antibiotics. *J. Med. Chem.* **2006**, *49*, 1506-1508.
- 53 Serrao, E.; Odde, S.; Ramkumar, K.; Neamati, N. Raltegravir, elvitegravir, and metoogravir: the birth of „me-too“ HIV-1 integrase inhibitors. *Retrovirology* **2009**, *6*.
- 54 Vacca, J.; Wai, J.; Fisher, T. Discovery of MK-2048: subtle changes confer unique resistance properties to a series of tricyclic hydroxypyrrrole integrase strand transfer inhibitors. *4th International AIDS Society (IAS) Conference Jul 22-25, 2007. Sydney, Australia.*
- 55 Shimura, K.; Kodama, E.; Sakagami, Y.; Matsuzaki, Y.; Watanabe, W.; Yamataka, K.; Watanabe, Y.; Ohata, Y.; Doi, S.; Sato, M.; Kano, M.; Ikeda, S.; Matsuoka, M. Broad anti-retroviral activity and resistance profile of a novel human immunodeficiency virus integrase inhibitor, elvitegravir (JTK-303/GS-9137). *J. Virol.* **2007**, *82*, 747-764.
- 56 Garvey, E. P.; Johns, B. A.; Gartland, M. J.; Foster, S. A.; Miller, W. H.; Ferris, R. G.; Hazen, R. J.; Underwood, M. R.; Boros, E. E.; Thompson, J. B.; Weatherhead, J. G.; Koble, C. S.; Allen, S. H.; Schaller, L. T.; Sherrill, R. G.; Yoshinaga, T.; Kobayashi, M.; Wakasa-Morimoto, C.; Miki, S.; Nakahara, K.; Noshi, T.; Sato, A.; Fujiwara, T. The naphthyridinone GSK364735 is a novel, potent human immunodeficiency virus type 1 integrase inhibitor and antiretroviral. *Antimicrob Agents Chemother.* **2008**, *3*, 901-908.
- 57 Marchand, C. Metal-dependent inhibition of HIV-1 integrase by diketo acids and resistance of the soluble double-mutant (F185K/C280S). *Mol. Pharmacol.* **2003**, *64*, 600-609.
- 58 Sotriffer, C. A.; Ni, H.; McCammon, J. A. HIV-1 integrase interactions at the active site: prediction of binding modes unaffected by crystal packing. *J. Am. Chem. Soc.* **2000**, *43*, 4109-4117.
- 59 Sechi, M.; Bacchi, A.; Carcelli, M.; Compari, C.; Duce, E.; Fisicaro, E.; Rogolino, D.; Gates, P.; Derudas, M.; Al-Mawsawi, L.-Q.; Neamati, N. From Ligand to Complexes: Inhibition of

- Human Immunodeficiency Virus Type 1 Integrase by Diketo Acid Metal Complexes. *J. Med. Chem.* **2006**, *49*, 4248-4260.
- 60 Fesen, M. R.; Kohn, K. W.; Leteurtre, F.; Pommier, Y. Inhibitors of human immunodeficiency virus integrase. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2399-2403.
- 61 Parrill, A. L. HIV-1 Integrase Inhibitors: Binding Sites, Structure Activity Relationships and Future Perspectives. *Curr. Med. Chem.* **2003**, *10*, 1811-1824.
- 62 Farnet, C. M.; Wang, B.; Hansen, M.; Lipford, J. R.; Zalkow, L.; Robinson, W. E.; Siegel, J.; Bushman, F. Human Immunodeficiency Virus Type 1 cDNA Integration: New Aromatic Hydroxylated Inhibitors and Studies of the Inhibition Mechanism. *Antimicrob. Agents. Chemother.* **1998**, *42*, 2245-2253.
- 63 Ryabinin, V. A.; Sinyakov, A. N.; de Soultrait, V. R.; Caumont, A.; Parissi, V.; Zakharova, O. D.; Vasyutina, E. L.; Yurchenko, E.; Bayandin, R. B.; Litvak, S.; Tarrago-Litvak, L.; Nevinsky, G. A. Inhibition of HIV-1 integrase-catalyzed reaction by new DNA minor groove ligands: the oligo-1,3-thiazolecarboxamide derivatives. *Eur. J. Med. Chem.* **2000**, *35*, 989-1000.
- 64 Mazumder, A.; Neamati, N.; Ojwang, J. O.; Sunder, S.; Rando, R. F.; Pommier, Y. Inhibition of the human immunodeficiency virus type 1 integrase by guanosine quartet structures. *Biochemistry* **1996**, *35*, 13762-13771.
- 65 Lutzke, R. A. P.; Eppens, N. A.; Weber, P. A.; Houghton, R. A.; Plasterk, R. H. A.; Identification of a hexapeptide inhibitor of the human immunodeficiency virus integrase protein by using a combinatorial chemical library. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 11456-11460.
- 66 Maroun, R. G.; Gayet, S.; Benleulmi, M. S.; Porumb, H.; Zargarian, L.; Merad, H.; Leh, H.; Mouscadet, J.-F.; Troalen, F.; Femandjian, S. Peptide Inhibitors of HIV-1 Integrase Dissociate the Enzyme Oligomers. *Biochemistry*, **2001**, *40*, 13840-13848.
- 67 Zhao, L.; O'Reilly, M. K.; Shultz, M. D.; Chmielewski, J. Interfacial peptide inhibitors of HIV-1 integrase activity and dimerization. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1175-1177.
- 68 Li, H.-Y.; Zawahir, T.; Song, L.-D.; Long, Y.-Q.; Neamati, N. Sequence-Based Design and Discovery of Peptide Inhibitors of HIV-1 Integrase: Insight into the Binding Mode of the Enzyme. *J. Med. Chem.* **2006**, *49*, 4477-4486.
- 69 Cherepanov, P.; Ambrosio, A. L.; Rahman, S.; Ellenberger, T.; Engelman, A. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17308-17313.

- 70 Shkriabai, N.; Patil, S. S.; Hess, S.; Budihias, S. R.; Craigie, R.; Burke, T. R.; Le Grice, S. F. J.; Kvaratskhelia, M. Identification of an inhibitor-binding site to HIV-1 integrase with affinity acetylation and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6894-6899.
- 71 Al-Mawsawi, L. Q.; Fikkert, V.; Dayam, R.; Witvrouw, M.; Burke, T. R.; Borchers, C. H.; Neamati, N. Discovery of a small-molecule HIV-1 integrase inhibitor-binding site. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10080-10085.
- 72 Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. Discovery of a Novel Binding Trench in HIV Integrase. *J. Med. Chem.* **2004**, *47*, 1879-1881.
- 73 Kessl, J. J.; Eidahl, J. O.; Shkriabai, N.; Zhao, Z.; McKee, C. J.; Hess, S.; Burke, T. R.; Kvaratskhelia, M. An allosteric mechanism for inhibiting HIV-1 integrase with a small molecule. *Mol. Pharmacol.* **2009**, *76*, 824-832.
- 74 Long, Y. Q.; Jiang, X. H.; Dayam, R.; Sanchez, T.; Shoemaker, R.; Sei, S.; Neamati, N. Rational design and synthesis of novel dimeric diketoacid-containing inhibitors of HIV-1 integrase: implication for binding to two metal ions on the active site of integrase. *J. Med. Chem.* **2004**, *47*, 2561-2573.
- 75 Grobler, J. A.; Stillmock, K.; Hu, B.; Witmer, M.; Felock, P.; Espeseth, A. S.; Wolfe, A.; Egbertson, M.; Burgeois, M.; Melamed, J.; Wai, J. S.; Young, S.; Vacca, J.; Hazuda, D. J. Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6661-6666.
- 76 Espeseth, A. S.; Felock, P.; Wolle, A.; Witmer, M.; Grobler, J.; Anthony, N.; Egbertson, M.; Melamed, J. Y.; Young, S.; Hamill, T.; Cole, J. L.; Hazuda, D. J. HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11244-11249.
- 77 CORINA 3.40. Molecular Networks GmbH **2006**.
- 78 Gasteiger, J.; Marsili, M. A New Model for Calculating Atomic Charges in Molecules. *Tetrahedron Lett.* **1978**, 3181-3184.
- 79 SYBYL 7.1. Tripos, Inc. **2005**.
- 80 Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.* **1998**, *19*, 1639-1662.
- 81 Sotriffer, C. A.; Ni, H.; McCammon, J. A. Active site binding modes of HIV-1 integrase inhibitors. *J. Med. Chem.* **2000**, *43*, 4109-4117.
- 82 Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, III, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. A.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.;

- Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 5, University of California, San Francisco, 1997.
- 83 Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies and free energies of solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620-9631.
- 84 Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-Behaved Electrostatic Potential Based Method Using Charge Restraints For Determining Atom-Centered Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269.
- 85 Stouten, P. F. W.; Frömmel, C.; Nakamura, H.; Sander, C. An effective solvation term based on atomic occupancies for use in protein simulations. *Mol. Simulat.* **1993**, *10*, 97-120.
- 86 Gonzalez-Ruiz, D.; Gohlke, H. Targeting Protein-Protein Interactions with Small Molecules: Challenges and Perspectives for Computational Binding Epitope Detection and Ligand Finding. *Curr. Med. Chem.* **2006**, *13*, 2607-2625.
- 87 Wells, J. A.; McClendon, C. L. Reaching for the high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007**, *450*, 1001-1009.
- 88 Stites, W. E. Protein-Protein Interactions: Interface Structure, Binding Thermodynamics, and Mutational Analysis. *Chem. Rev.* **1997**, *97*, 1233-1250.
- 89 Clackson, T.; Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **1995**, *267*, 383-386.
- 90 Hu, Z.; Ma, B.; Wolfson, H.; Nussinov, R. Conservation of polar residues as hot spots at protein interfaces. *Proteins* **2000**, *39*, 331-342.
- 91 Arkin, M. R.; Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.* **2004**, *4*, 301-317.
- 92 Kelekar, A.; Chang, B. S.; Harlan, J. E.; Fesik, S. W.; Thompson, C. B. Bad is a BH3 domain-containing protein that forms an inactivating dimer with Bcl-XL. *Mol. Cell. Biol.* **1997**, *17*, 7040-7046.
- 93 Bruncko, M. et al. Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-XL. *J. Med. Chem.* **2007**, *50*, 641-662.
- 94 Kussie, P. H. et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **1996**, *274*, 948-953.
- 95 Vassilev, L. T. et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **2004**, *303*, 844-848.
- 96 Abbate, E. A.; Berger, J. M.; Botchan, M. R. The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2. *Genes Dev.* **2004**, *18*, 1981-1996.

- 97 Wang, Y. et al. Crystal structure of the E2 transactivation domain of human papillomavirus type 11 bound to a protein interaction inhibitor. *J. Biol. Chem.* **2004**, *279*, 6976-6985.
- 98 Rickert, M.; Wang, X.; Boulanger, M. J.; Goriatcheva, N.; Garcia, K. C. The structure of interleukin-2 complexed with its receptor. *Science* **2005**, *308*, 1477-1480.
- 99 Raimundo, B. C. et al. Integrating fragment assembly and biophysical methods in the chemical advancement of small-molecule antagonists of IL-2: an approach for inhibiting protein-protein interactions. *J. Med. Chem.* **2004**, *47*, 3111-3130.
- 100 Eck, M. J.; Sprang, S. R. The structure of tumor necrosis factor α at 2.6 Å resolution. Implications for receptor binding. *J. Biol. Chem.* **1989**, *264*, 17595-17605.
- 101 He, M. M. et al. Small-molecule inhibition of TNF- α . *Science* **2005**, *310*, 1022-1025.
- 102 Hu, J. P.; Gong, X. Q.; Su, J. G.; Chen, W. Z.; Wang, C. X. Study on the molecular mechanism of inhibiting HIV-1 integrase by EBR28 peptide via molecular modeling approach. *Biophys. Chem.* **2008**, *132*, 69-80.
- 103 Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. AMBER 9, University of California, San Francisco, 2006.
- 104 Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Cieplak, P.; Luo, R.; Lee, T. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999-2012.
- 105 Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.
- 106 Srinivasan, J.; Trevathan, M.; Beroza, P.; Case, D. A. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **1999**, *101*, 426-434.
- 107 Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.
- 108 Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald-an Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
- 109 MOE, version 2007_09; Chemical Computing Group, Montreal, Canada, 2007.

- 110 Al-Mawsawi, L. Q.; Hombrouck, A.; Dayam, R.; Debyser, Z.; Neamati, N. Four-tiered interaction at the dimeric interface of HIV-1 integrase critical for DNA integration and viral infectivity. *Virology* **2008**, *377*, 355-363.
- 111 Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727-748.
- 112 Velec, H. F. G.; Gohlke, H.; Klebe, G. DrugScoreCSD - Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction. *J. Med. Chem.* **2005**, *48*, 6296-6303.
- 113 Gohlke, H.; Hendlich, M.; Klebe, G. Predicting binding modes, binding affinities and "hot spots" for protein-ligand complexes using a knowledge-based scoring function. *Persp. Drug Discov. Des.* **2000**, *20*, 115-144.
- 114 Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337-356.
- 115 Sotriffer, C. A.; Sanschagrin, P.; Matter, H.; Klebe, G. SFCscore: Scoring functions for affinity prediction of protein-ligand complexes. *Proteins* **2008**, *73*, 395-419.
- 116 Irwin, J. J. Shoichet, B. K. ZINC-a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177-182.
- 117 <http://www.emolecules.com>
- 118 Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important molecules. *J. Med. Chem.* **1985**, *28*, 849-857.
- 119 Wade, R. C.; Clark, K. J.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 140-147.
- 120 Wade, R. C.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 148-156.
- 121 Carosati, E.; Sciabola, S.; Cruciani, G. Hydrogen Bonding Interactions of Covalently Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the GRID Force Field. *J. Med. Chem.* **2004**, *47*, 5114-5125.
- 122 HERMES 1.3.1. CCDC Software Limited. 2009
- 123 Egan, W. J.; Merz, K. M.; Baldwin, J. J. Prediction of drug absorption using multivariate statistics. *J. Med. Chem.* **2000**, *43*, 3867-3877

- 124 Cheng, T.; Xun, L.; Yan, L.; Zhihai, L.; Wang, R. Comparative Assessment of Scoring Functions on a Diverse Test Set. *J. Chem. Inf. Model.* **2009**, *49*, 1079-1093.
- 125 Forino, M. Virtual docking approaches to protein kinase B inhibition. *J. Med. Chem.* **2005**, *48*, 2278-2281.
- 126 Cianciotto, N. P. Pathogenicity of *Legionella pneumophila*. *Int. J. Med. Microbiol.* **2001**, *5*, 331-343.
- 127 Steinert, M.; Hentschel, U.; Hacker, J. *Legionella pneumophila*: An aquatic microbe goes astray. *FEMS Microbiol. Rev.* **2002**, *26*, 149-162.
- 128 *Legionella and the prevention of legionellosis. World Health Organization* **2007**.
- 129 Banga, S.; Gao, P.; Shen, X.; Fiscus, V.; Zong, W. X.; Chen, L.; Luo, Z. Q. *Legionella pneumophila* inhibits macrophage apoptosis by targeting pro-death members of the Bcl2 protein family. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 5121-5126.
- 130 Wagner, C.; Khan, A. S.; Kamphausen, T.; Schmausser, B.; Ünal, C.; Lorenz, U.; Fischer, G.; Hacker, J.; Steinert, M. Collagen binding protein Mip enables *Legionella pneumophila* to transmigrate through a barrier of NCI-H292 lung epithelial cells and extracellular matrix. *Cell. Microbiol.* **2007**, *9*, 450-462.
- 131 Dumont, F. J.; Su, Q. Mechanism of Action of the Immunosuppressant Rapamycin. *Life Sci.* **1996**, *58*, 373-395.
- 132 Ceymann, A.; Horstmann, M.; Ehse, P.; Schweimer, K.; Paschke, A.-K.; Steinert, M.; Faber, C. Solution structure of the *Legionella pneumophila* Mip-rapamycin complex. *BMC Structural Biology* **2008**, *8*, 17.
- 133 Brown, E. J.; Albers, M. W.; Shin, T. B.; Ichikawa, K.; Keith, C. T.; Lane, W. S.; Schreiber, S. L. A mammalian protein targeted by G1-arresting rapamycin-receptor complexes. *Nature* **1994**, *369*, 756-758.
- 134 Steiner, J. P.; Hamilton, G. S.; Ross, D. T.; Valentine, H. L.; Guo, H.; Connolly, M. A.; Liang, S.; Ramsey, C.; Li, J. H.; Huang, W.; Howorth, P.; Soni, R.; Fuller, M.; Sauer, H.; Nowotnik, A. C.; Suzdak, P. D. Neurotrophic immunophilin ligands stimulate structural and functional recovery in neurodegenerative animal models. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2019-2024.
- 135 Zhao, L.; Huang, W.; Liu, H.; Wang, L.; Zhong, W.; Xiao, J.; Hu, Y.; Li, S. FK506-Binding Protein Ligands: Structure-Based Design, Synthesis, and Neurotrophic/Neuroprotective Properties of Substituted 5,5-Dimethyl-2-(4-thiazolidine)carboxylates. *J. Med. Chem.* **2006**, *14*, 4059-4071.

- 136 Stebbins, J. L.; Zhang, Z.; Chen, J.; Wu, B.; Emdadi, A.; Williams, M. E.; Cashman, J.; Pellechia, M. Nuclear Magnetic Resonance Fragment-Based Identification of Novel FKBP12 Inhibitors. *J. Med. Chem.* **2007**, *50*, 6607-6617.
- 137 Steiner, J. P.; Connolly, M. A.; Valentine, H. L.; Hamilton, G. S.; Dawson, T. M.; Hester, L.; Snyder, S. H. Neurotrophic actions of nonimmunosuppressive analogues of immunosuppressive drugs FK506, rapamycin and cyclosporin A. *Nat. Med.* **1997**, *3*, 421-428.
- 138 Holt, D. A.; Luengo, J. I.; Yamashita, D. S.; Oh, H.-J.; Konialian, A. L.; Yen, H.-K.; Rozamus, L. W.; Brandt, M.; Bossard, M. J.; Levy, M. A.; Eggleston, D. S.; Liang, J.; Schultz, L. W.; Stout, T. J.; Clardy, J. Design, Synthesis, and Kinetic Evaluation of High-Affinity FKBP Ligands and the X-ray Crystal Structures of Their Complexes with FKBP12. *J. Am. Chem. Soc.* **1993**, *115*, 9925-9938.
- 139 Riboldi-Tunncliffe, A.; König, B.; Jessen, S.; Weiss, M. S.; Rahfeld, J.; Hacker, J.; Fischer, G.; Hilgenfeld, R. Crystal structure of MIP, a prolyl isomerase from *Legionella pneumophila*. *Nat. Struct. Biol.* **2001**, *8*, 779-773.
- 140 Szep, S.; Park, S.; Boder, E. T.; Van Duyne, G. D.; Saven, J. G. Structural coupling between FKBP12 and buried water. *Proteins* **2009**, *74*, 603-611.
- 141 Michnick, S. W.; Rosen, M. K.; Wandless, T. J.; Karplus, M.; Schreiber, S. L. Solution structure of FKBP, a rotamase enzyme and receptor for FK506. *Science* **1991**, *252*, 836-839.
- 142 Van Duyne, G. D.; Standaert, R. F.; Schreiber, S. L.; Clardy, J. Atomic structure of the Rapamycin Human Immunophilin Fkbp-12 Complex. *J. Am. Chem. Soc.* **1991**, *113*, 7433-7440.
- 143 Holt, D. A.; Konialian, A. L.; Oh, H. J.; Yen, H. K.; Rozamus, L. W.; Krog, A. J.; Erhard, K. F.; Ortiz, E.; Levy, M. A.; Brandt, M.; Bossard, M. J.; Luengo, J. I. Structure-Activity Studies of Synthetic FKBP Ligands as Peptidyl-Prolyl Isomerase Inhibitors. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 315-320.
- 144 Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720-729.
- 145 Sun, F.; Li, P.; Ding, Y.; Wang, L.; Bartlam, M.; Shu, C.; Shen, B.; Jiang, H.; Li, S.; Rao, Z. Design and Structure-Based Study of New Potential FKBP12 Inhibitors. *Biophys. J.* **2003**, *85*, 3194-3201.
- 146 Burkhard, P.; Taylor, P.; Walkinshaw, M. X-ray structures of small ligand-FKBP complexes provide an estimate for hydrophobic interaction energies. *J. Mol. Biol.* **2000**, *295*, 953-962.

- 147 Chadha, V. K.; Leidal, K. G.; Plapp, B. V. Inhibition by carboxamides and sulfoxides of liver alcohol dehydrogenase and ethanol metabolism. *J. Med. Chem.* **1983**, *26*, 916-922.
- 148 Hann, M. M. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856-864.
- 149 Hajduk, P. J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* **2007**, *6*, 211-219.
- 150 Carr, R. A. E.; Congreve, M.; Murray, C. W.; Rees, D. C. Fragment-based lead discovery: leads by design. *Drug Disc. Today* **2005**, *10*, 987-992.
- 151 Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531-1534.
- 152 Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31-36.
- 153 Durant, J. L. et al. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273-1280.