Julius-Maximilians-
Universität
Würzburg

Fakultät für Chemie und Pharmazie

Quantifying Non-covalent Interactions – Rational *in-silico* Design of Guanidinium-based Carboxylate Receptors

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

vorgelegt von

Sebastian Schlund

aus Würzburg

Würzburg 2007

Eingereicht am: ______________________________

bei der Fakulät für Chemie und Pharmazie.

1. Gutachter: ______________________________

2. Gutachter: ______________________________

der Dissertation.

1. Prüfer: ______________________________

2. Prüfer: ______________________________

3. Prüfer: ______________________________

des öffentlichen Promotionskolloquiums.

Tag des öffentlichen Promotionskolloquium: ____________________

Doktorurkunde ausgehändigt am: ____________________

Die vorliegende Arbeit wurde unter Anleitung von Prof. Dr. Bernd Engels von Dezember 2003 bis Juli 2007 am Institut für Organische Chemie der Julius-Maximilians-Universität Würzburg angefertigt.

Teilergebnisse dieser Arbeit waren Gegenstand von Publikationen sowie von Postern und Kurzvorträgen.

Publikationen:

"Knock-out" Analogues as a Tool to Quantify Supramolecular Processes: A Theoretical Study of Molecular Interactions in Guanidiniocarbonyl Pyrrole Carboxylate Dimers" S. Schlund, C. Schmuck, B. Engels, *JACS* **2005**, *127*, 11115-11124.

"Geometry and Cooperativity Effects in Adenosine-carboxylic Acid Complexes" S. Schlund, M. Mladenovic, E. M. Basilio Janke, B. Engels, K. Weisz, *JACS* **2005**, *127*, 16151-16158.

"How important is molecular rigidity for the complex-stability of artificial host-guest systems? A theoretical study on Self-Assembly of Gas phase Arginine" S. Schlund, C. Schmuck, B. Engels, *Chem-Eur. J.* **2007**, *13 (23)*, 6644 – 6653.

"Conformational Analysis of Arginine in Gas Phase - A Strategy for Scanning the Potential Energy Surface effectively" S. Schlund, R. Müller, C. Graßmann, *J. Comp. Chem.* **2007**, DOI: 10.1002/jcc.20798

Poster und Vorträge:

S. Schlund, C. Schmuck, B. Engels; A Theoretical Study of Guanidiniocarbonyl Pyrrole Complexes as Model Systems for Carboxylate Receptors, Posterbeitrag, *Summer School International University of Bremen:: Advanced Modelling of Biological Function – Expression, Networks and Structure*, Bremen, August **2004**.

Schlund S., Engels B.; Theoretical Investigations of Non-covalent Interactions in Guanidiniocarbonyl Pyrrole Carboxylate Dimers, Vortrag, *Joint PhD Students Meeting of the SFBs 544 and 630*, Würzburg, November **2004**.

Schlund S., Engels B.; Theoretical Investigations of Non-covalent Interactions in Guanidiniocarbonyl Pyrrole Carboxylate Dimers, Posterbeitrag, *Joint PhD Students Meeting of the SFBs 544 and 630*, Würzburg, November **2004**.

S. Schlund, C. Schmuck, B. Engels; “Knock-out” Analogues as a Tool to Quantify Supramolecular Processes - Theoretical Investigations of Non-covalent Interactions in Guanidiniocarbonyl Pyrrole Carboxylate Dimers, Posterbeitrag, *41. Symposium Theoretische Chemie*, Innsbruck, September **2005**.

S. Schlund, C. Schmuck, B. Engels; “Knock-out” Analogues as a Tool to Quantify Supramolecular Processes - Theoretical Investigations of Non-covalent Interactions in Guanidiniocarbonyl Pyrrole Carboxylate Dimers, Posterbeitrag, *1st International Symposium the SFB* 630 – *Novel Agents against Infectious Deseases*, Würzburg, Februar **2006**.

Schlund S., Engels B.; How to Quantify Weak Molecular Interactions: Conformational Analysis of Monomeric Arginine and the Zwitterionic Dimer in Gas Phase, Posterbeitrag, *42. Symposium Theoretische Chemie*, Berlin, September **2006**.

Schlund S., Engels B.; How to Quantify Weak Molecular Interactions: Conformational Analysis of Monomeric Arginine and the Zwitterionic Dimer in Gas Phase, Posterbeitrag, *Second Joint PhD Students Meeting of the SFBs 544 and 630*, Heidelberg, November **2006**.

Table of Contents

Chapter 1 Introduction – Binding of anions in artificial receptor systems 1

Chapter 2 How to Quantify Intermolecular Interactions? 9

2.1 Experimentally 9
  2.1.1 Determination of association constants 9
  2.1.2 Determination of entropy and enthalpy contributions 12

2.2 Theoretically 14
  2.2.1 Calculation of dimerization energies 14
  2.2.2 Calculation of entropy and enthalpy contributions 15

Chapter 3 Theoretical Methods 19

3.1 Electronic structure methods 20
  3.1.1 A short review on the self-consistent field (SCF) theory 20
  3.1.2 Electron correlation methods 28
    3.1.2.1 Møller-Plesset perturbation theory 29
    3.1.2.2 Coupled-cluster theory 37
  3.1.3 Density functional theory 39
  3.1.4 Drawbacks and benefits of the electronic structure methods 47

3.2 Conformational analysis 49
  3.2.1 Brief introduction to molecular mechanics 49
  3.2.2 Conformational search algorithms 51

3.3 From gas-phase to solution 55
  3.3.1 Continuum solvation 56
    3.3.1.1 The Apparent Surface Charge Method 58
    3.3.1.2 The Conductor-like Screening Model 59
  3.3.2 Explicit solvation 62

3.4 Molecular simulations – Accounting for the entropy 63
  3.4.1 Molecular Dynamics 63
  3.4.2 Thermodynamic Integration 69

3.5 The QM/MM Method 71

Chapter 4 Results and Discussion 73

4.1 Quantifying Supramolecular Processes by Knock-out Analogues 73
  4.1.1 Introduction 73
  4.1.2 Computational Details 77
  4.1.3 Geometries 78
  4.1.4 Energies 82
  4.1.5 Conclusions 94

4.2 Arginine as Model System for Guanidinium-Carboxylate Interactions 95
  4.2.1 Conformational Analysis in Gas Phase – The Quest for the Global Minimum 96
    4.2.1.1 Introduction 96
    4.2.1.2 Computational Details 97
    4.2.1.3 Force-Field Validation 99
    4.2.1.4 Conformational Search Algorithms 101
    4.2.1.5 Electronic Structure Optimizations 104
    4.2.1.6 Calculated Spectra 113
    4.2.1.7 Conclusions 120
  4.2.2 Self-Assembly of Gas Phase Arginine 122
    4.2.2.1 Introduction 122
    4.2.2.2 Computational Details 124
    4.2.2.3 Geometries and Energies 125
    4.2.2.4 Importance of molecular rigidity for the stability of the dimer 130
    4.2.2.5 Conclusions 131

4.3 Gas-Phase Assemblies with Novel Structure Motifs 132
  4.3.1 Introduction 132
  4.3.2 Theoretical Details 133
  4.3.3 Novel Structure Motifs 134
    4.3.3.1 6-Membered Ring Structures 134
    4.3.3.2 Enlarging the π-System: Anellated Aromatic Rings 137
    4.3.3.3 Decoupled π –Systems: Biphenylene Derivatives 139
    4.3.3.4 Importance of Preorganisation and Fixation: Indole-Derivatives 141
    4.3.3.5 Benzene vs. Pyrrole Derivative: Molecular Solvation Effects 146
    4.3.3.6 Conclusion 155

4.4 Cooperativity Effects in Supramolecular Assemblies – NMR Shift Studies in Adenosine-Carboxylic Acid Complexes 157
  4.4.1 Introduction 157
  4.4.2 Experimental Findings 160
  4.4.3 Computational Details 163
  4.4.4 Calculation of averaged NMR – shifts 166

4.4.5 Discussion ... 172
4.4.6 Conclusion ... 176

Chapter 5 Summary ... 179

Chapter 6 Zusammenfassung ... 187

Chapter 7 Outlook ... 195

References and Notes ... 201

> *"Quantum mechanics is certainly imposing. But an inner voice tells me that it is not yet the real thing. The theory says a lot, but does not really bring us closer to the secret of the 'Old One.' I, at any rate, am convinced that He is not playing at dice."*
>
> -- ***Albert Einstein***

Symbols and Abbreviations

Ψ	total wave function
ψ	electronic wave function
Φ	nuclear wave function / electric field potential (chapter 3.3.1)
ψ^S	Slater determinant
φ	spin orbital
ϕ	spatial orbital
Ω	spin function
χ	basis function
$\mathbf{R}$	nuclear coordinates
$\mathbf{r}$	spatial electron coordinates
$\boldsymbol{\omega}$	spin coordinates
$\mathbf{x}$	spatial electron coordinates with spin coordinates
$\mathbf{H}$	Hamilton operator
$\mathbf{h}$	one electron operator
$\mathbf{g}$	two electron operator
$\mathbf{J}$	Coulomb operator
$\mathbf{K}$	exchange operator

HF	Hartree Fock
SCF	Self-Consistent Field
AO	Atomic Orbital
LCAO	Linear Combination of Atomic Orbitals
MO	Molecular orbital
CSF	Configuration state function
DFT	Density Functional Theory
MP2	Møller-Plesset perturbation theory 2nd order
CCSD(T)	Coupled Cluster theory with double excitations including triple excitations by perturbation theory
GGA	Generalized Gradient Approximation
B-LYP	Becke exchange functional and Lee, Yang and Parr’s correlation functional
B3-LYP	Becke’s 3 parameter functional and Lee, Yang and Parr’s correlation functional
RI	Resolution of Identity

COSMO	COnductor-like Screening Model
PCM	Polarizable Continuum Model
AM1	Austin Model 1
PM3	Parameterized Model 3
MNDO	Modified Neglect of Differential Overlap
MMFF94	Merck Molecular Force Field
OPLS-AA	Optimized Potentials for Liquid Solvents (All Atoms model)

Chapter 1 Introduction – Binding of anions in artificial receptor systems

Supramolecular chemistry as an independent research field has been developed in the beginning of the 1960s with the synthesis of macrocyclic ligands for metal cations like crown ethers (Pederson *et al.*) and three-dimensional cryptands (Lehn *et al.*) or spherands (Cram *et al.*).[1] The pioneering work of Cram, Lehn and Pederson in this new discipline was awarded in 1987 with the Alfred Nobel prize and nowadays supramolecular chemistry has become a fast-growing research field showing high interdisciplinarity between chemistry, biology and physics. It is defined as the chemistry of molecular assemblies which are formed by non-covalent interactions between two compounds, often denoted as host and guest. In the field of enzymology this concept has been known long before as receptor-substrate binding or molecular recognition. Already in 1894 Emil Fischer described the discrimination of various substrates towards an enzyme as a “key-lock” principle depending on the complementarity of the molecular shape with the binding pocket.

Today’s supramolecular chemistry comprises a large scope of research interests starting from biological relevant macrocyles, transport systems or neurotransmitters over artificial hosts that bind either cationic, anionic or neutral compounds to liquid crystals and other self-assembling systems. The common ground on which all of these structures are based is the non-covalent character of the intermolecular interactions. The nature of these supramolecular interactions originates from electrostatic interactions[2] which can be classified according to **Table 1**. The bond energies range from 350 kJ mol^{-1} similarly to covalent bonds to very weak van-der-Waals interactions.

The non-covalent binding of anions as guest compounds presents a field in supramolecular chemistry that, in contrast to cationic or neutral substrate binding, has to deal with difficulties in constructing artificial receptors arising from the intrinsic properties of anions.[3] Even simple inorganic anions and above all biological important anions show a non-spherical shape and have a much larger system size compared to common cations. Moreover, anion hosts have to compete with the surrounding aqueous medium resulting in large free energies of solvation. Solvation stabilizes anions to a larger extent than cations of equal charge density by the formation of multiple and strong hydrogen bonds. Further aspect which have to be considered when designing artificial anion receptor systems are

- the pH range dependency of anionic guests and cationic host molecules, especially those with protonated amino groups
- the complementarity and preorganisation of the host-guest complex
- the thermodynamic or kinetic selectivity of a host system towards several guest molecules

Since the engineering of a kinetic selectivity as it occurs in many biochemical enzymes is still quite challenging, a thermodynamic selectivity by intelligent application of the complementarity and preorganisation concepts in conjunction with optimized host-guest interactions is easier to achieve. The problem occurring during a kinetic controlled process is that the guest molecule must not bind too strong and the receptor molecule should on one hand be perfectly complementary to the substrate, but on the other hand should be also flexible enough to adapt to the conformational change and modification in the charge distribution of the guest molecule along the complexation pathway.

Table 1 Classification of the various non-covalent interactions

Interaction		Bond energy (kJ mol-1)	Example
Ionic interaction (Ion pairing)		100-350	NaCl
Ion-dipole interaction		50-200	Na^+-18-crown-6 complex
Hydrogen bonding	strong	60-120	HF
	medium	16-60	nucleobase pairing
	weak	<12	C—H···O hydrogen bond
Cation-π-interaction		5-80	K^+-benzene complex
π –π-stacking		0-50	DNA
Dipole-dipole interaction		5-50	Carbonyls
Van-der-Waals interaction		<5	non-polar solvents, *e.g.* n-octane

Arginine as carboxylate binding site

Anions play an important role in biochemical processes as enzyme substrates and cofactors.[4] For instance, phosphate residues occurring in ADP or ATP can bind non-covalently to ATPase enzymes that are part of the energy transport system in the respiratory chain. One particularly important amino acid residue found within many anionic binding pockets of proteic hosts is arginine which can bind effectively phosphates, sulfates and carboxylates. Arginine itself represents an amphipathic amino acid with a hydrophobic long side chain and a hydrophilic terminal guanidine group. The guanidine group has a high proton affinity and is thus protonated under physiological conditions (guanidinium: pK_a = 13.5). The cationic guanidinium rest in arginine can therefore form strong salt bridges with anionic substrates. One example for a naturally occurring arginine based anionic receptor is the carboxypeptidase A (CPA) that is responsible for the hydrolytic cleavage of ester bonds of C-terminal peptidic substrates possessing an aromatic rest in the β position. The X-ray structure of CPA with L-phenyllactate as enzyme inhibitor is given in **Figure 1** showing the carboxylate group of the substrate binding via two charge-assisted hydrogen bonds to the guanidinium moiety of an arginine residue.[5]

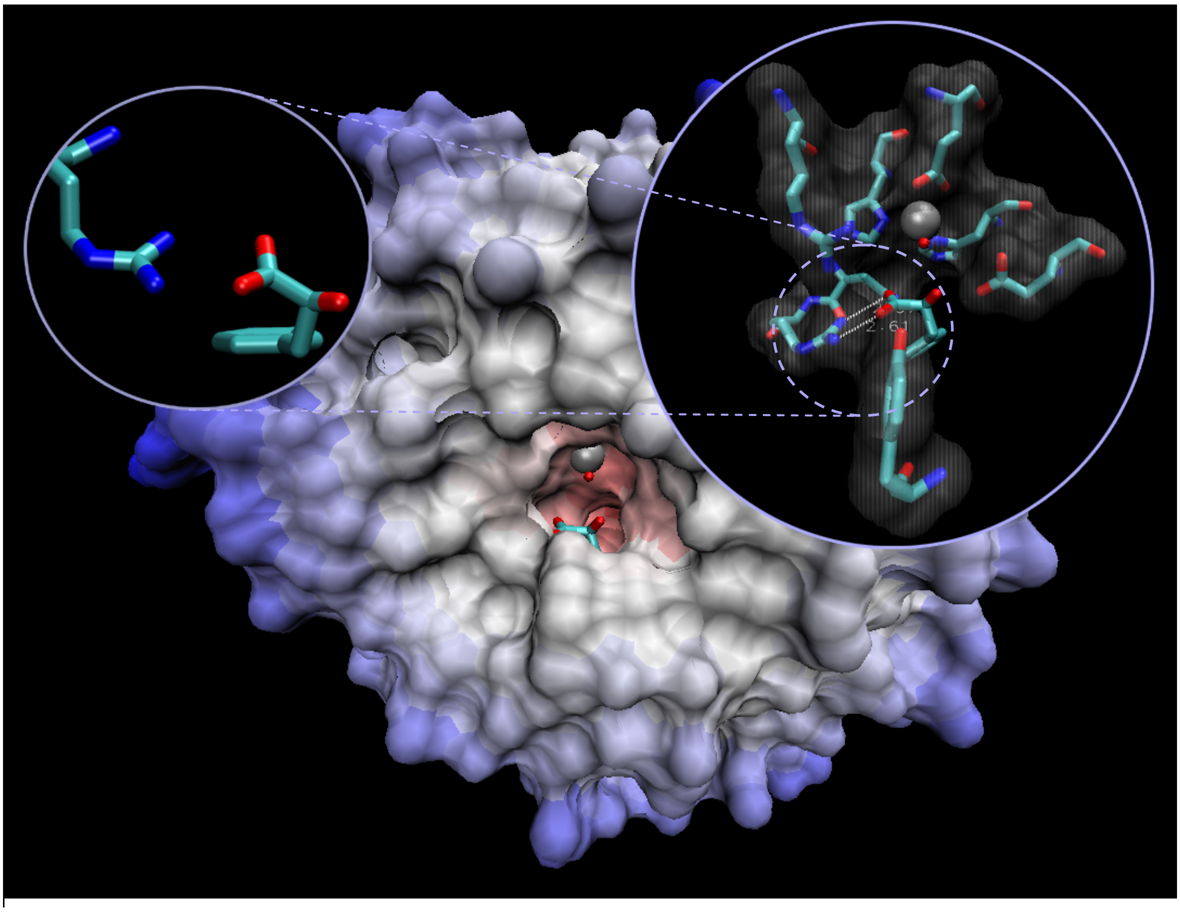
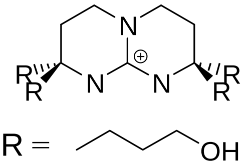


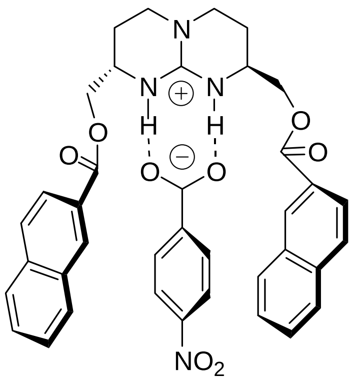
Figure 1 Crystal structure of L-phenyllactate binding to carboxypeptidase A.

Artificial guanidinium-based receptors

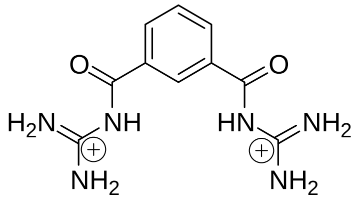
Already in the late 1970s Lehn and co-workers synthesized the first guanidinium-based anion receptors that were macrocycles containing up to three guanidinium moieties separated by various spacer groups. [3,6,7] However, the poor complexation behaviour based mainly on unspecific ionic interactions as well as macrocyclic and chelat effects made this type of receptors unfavourable. A decade later, in 1988, Schmidtchen and coworkers were able to present a bicyclic guanidinium system that showed improved binding characteristics due to the reduced hydration of the charged moiety and the introduction of anchor groups for stabilizing interactions with the backbone of the substrate.[8] The butanol substituted derivative of **A** was able to complex *p*-nitrobenzoat in acetonitril with an estimated lower limit for the dissociation constant K_{diss} of 10^{-4} M^{-1}. De Mendoza extended this scaffold to design chiral derivatives of **A** which were able to discriminate between *D*/*L* enantiomers of aromatic carboxylate anions. With *p*-nitrobenzoat as substrate the receptor **B** showed an association constant in $CDCl_3$ of $K = 1609$ M^{-1} and for N-acetyltryptophan the association for the *L*-isomer was twice as high as for the *D*-isomer (K_{ass} = 1051 vs. 534 M^{-1}).[9]



A



B



C

In 1992, Hamilton and coworkers developed a bisguanidinium-based receptor **C** that exhibited an excellent complexation behaviour in acetonitril towards phosphodiesters ($K_{ass} = 5 \times 10^4$ SM^{-1}).[10] It was shown that this receptor accelerates the phosphordiester cleavage in RNA by mimicking staphylococcal nuclease (SNase) by a factor of 300 compared to the uncatalysed reaction. Similar bisguanidinium systems were synthesized by Göbel or Anslyn that were able to complex multiple charged anions in rather polar solvents like DMSO or methanol with association constants up to 10^5 M^{-1}.[3] Unfortunately, all receptors showed poor association in increasingly competitive solvent media (from DSMO to water) and are therefore not capable to bind monoanionic substrates like protein carboxylate structures under physiological conditions effectively.

Guanidiniocarbonylpyrroles as carboxylate receptors

In 1999, Schmuck reported for the first time on a new class of receptor molecules for the binding of carboxylates in aqueous media.[11] These 2-(guanidiniocarbonyl)-1*H*-pyrroles improve the ion pairing of simple guanidinium cations with oxo anions through a combination of ion pairing and multiple hydrogen bonds (see **Figure 2**). Due to the increased acidity of the acyl guanidinium moiety and the additional H-bonds, these complexes are much stronger than those of simple guanidinium cations allowing the complexation of carboxylates even in highly polar solvents like aqueous DMSO.[3,4,6,12] An experimental comparative thermodynamic study with a series of structurally related guanidiniocarbonyl pyrroles demonstrated that the energetic contributions of the individual non-covalent interactions within this binding motif (the individual hydrogen bonds and the ion pair) are significantly different. Besides the ion pairing mainly the amide NH in position 5 of the pyrrole ring seemed to be important for the effective binding of the carboxylate substrate. Further studies showed that also the size and electronic structure of the aromatic ring is important.[13] Pyrrole systems are superior to the analogous benzene derivatives which in turn show a higher binding affinity than pyridine derivatives, in which the nitrogen lone pair exerts additional repulsive effects on the bound carboxylates.

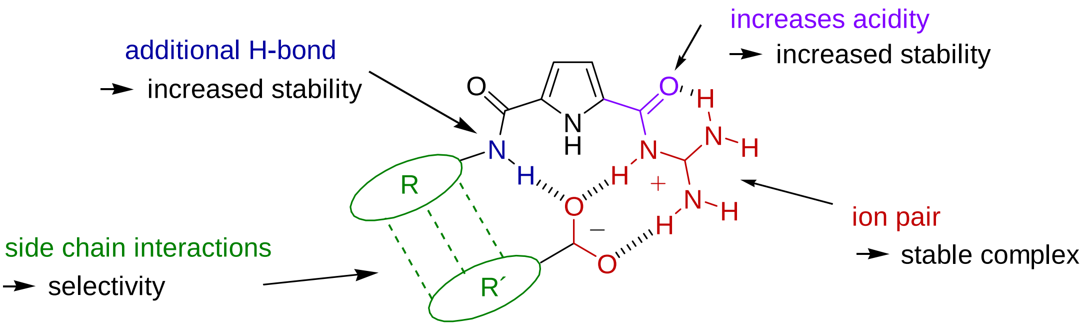


Figure 2 Guanidiniocarbonyl pyrroles efficiently bind carboxylates even in aqueous solvents due to a combination of ion pair formation and additional H-bonds.

A zwitterionic derivative of this new structure motif with substitution of a carboxylate group in position 5 of the pyrrole ring system shows strong self-assembly to discrete dimers with an estimated association constant of 170 M^{-1} in water.[14] The formation of discrete nanometresized cyclic dimers or vesicular structures could be observed in DMSO by separation of two zwitterionic monomeric units with different spacers (see **Figure 3**).[15]

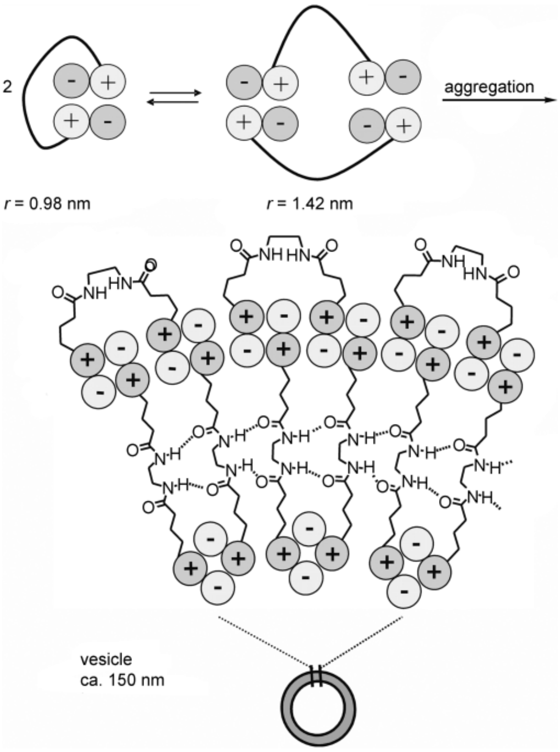


Figure 3 Formation of cyclic dimers consisting of bis-zwitterions, which then self-assemble into large vesicles [taken from ref. 15b].

The quantification of such supramolecular interactions has therefore been of great interest and one incremental approach, for example, is based on extrapolation from experimental thermodynamical data.[16] However, this ansatz is only able to provide a semi-quantitative thermodynamic ranking of intermolecular effects. A real quantitative discrimination between all the various contributions that sum up to the overall stability was not yet possible and the questions concerning the binding mode of the guanidinium based receptor motif that could not be answered satisfactorily are in particular:

- What are the contributions of each hydrogen bond to the dimerization energy and how important are cooperative effects?
- What role does the salt bridge play upon complexation?
- What influence do size and electronic structure of the ring system have on the anion binding?
- How important are molecular solvent effects?

To summarize, one question arises inevitably during any rational design process which can be formulated as follows: "*What makes a certain structure motif more favourable than others with respect to the property of interest and how can one further improve the structure motif in order to maximize the property in a given media?*"

On this account the present study has been performed in order to elucidate the various effects that affect the association process. The advantages of theoretical investigations are that one is able to quantify non-covalent interactions by means of fictional systems that may not be experimentally accessible. The newly gained knowledge on the binding modes can then be used for the development of a novel lead structure like for the rational design of artificial peptide receptors.[17]

Chapter 2 How to Quantify Intermolecular Interactions?

2.1 Experimentally

The quantification of non-covalent interactions by evaluation of association free energies is of central importance for the investigation and understanding of host-guest systems.[18] In the following chapter the discussion of physical-chemical approaches to determine the energetics of supramolecular complexes will be restricted to the most popular experimental methods used in modern research labs and gives only a rough sketch of the various theoretical backgrounds of the spectroscopic methods available.[19] Not regarded are for instance electrochemical, solubility-based or kinetic methods as well as calorimetry. Furthermore, it is assumed that the complexes are in a thermal equilibrium with 1:1 stoichiometry.

2.1.1 Determination of association constants

Considering a simple equilibrium reaction of a receptor *R* and a substrate *S* the association and dissociation constants are given by the law of mass action as

$$K_{Ass} = \frac{1}{K_{Diss}} = \frac{[RS]}{[R][S]} \qquad \text{eq. 2.1.1}$$

The equilibrium concentrations of the free receptors and substrates can be calculated from the mass balance equations and the initial concentrations.

$$\begin{aligned} [R] &= [R]_0 - [RS] \\ [S] &= [S]_0 - [RS] \end{aligned} \qquad \text{eq. 2.1.2}$$

By substituting eq. 2.1.2 into eq. 2.1.1 one obtains a quadratic equation

$$[RS]^2 - \{[R]_0 + [S]_0 + K_{Diss}\}[RS] + [R]_0[S]_0 = 0 \qquad \text{eq. 2.1.3}$$

that can be solved easily to

$$[RS] = \frac{([R]_0 + [S]_0 + K_{Diss}) - \sqrt{([R]_0 + [S]_0 + K_{Diss})^2 - 4[R]_0[S]_0}}{2} \qquad \text{eq. 2.1.4}$$

Regarding now any spectroscopic methods the property of interest (*e.g.* absorbance, fluorescence, NMR shifts, reaction rates, conductance etc.) can be expressed as sum of contributions from all components consisting of intrinsic molar properties x times the equilibrium concentrations. With the mass balance equations one can write for the observed property

$$X_{Obs} = x_R[R] + x_S[S] + x_{RS}[RS] = x_R[R]_0 + x_S[S]_0 + \Delta x[RS] \qquad \text{eq. 2.1.5}$$

with

$$\Delta x = x_{RS} - x_R - x_S \qquad \text{eq. 2.1.6}$$

The equilibrium concentration of the complex is given by eq. 2.1.4 and therefore eq. 2.1.5 represents a hyper surface determined by the variables $[R]_0$, $[S]_0$ and Δx. In general, the initial concentration of the receptor is held constant and the molar property difference Δx is recorded while the substrate is titrated to the solution. By a computer-aided non-linear curve regression of the observed titration curve one can eventually determine the dissociation constant K_{Diss} from eq. 2.1.5. This analysis holds for all spectroscopic methods describing 1:1 association processes. Reactions with different stoichiometry require other formulations of the law of mass action and hence the here derived equations for the equilibrium concentrations of the complex are no longer valid.

NMR studies

NMR titration experiments are popular for studying complexation reactions since besides the association constants one also obtains information about the structural properties of the receptor-substrate complex.[19] The “complexation-induced shifts” (CIS) are stronger for atoms directly involved in the association process by forming hydrogen bonds or undergoing other intermolecular interactions. Hence, these atoms show the largest modifications of the chemical (electronic) environment being reflected by the largest CIS’s. The major drawbacks of NMR studies of non-covalent interactions are on one hand the need for high concentrations in order to get a reasonable resolution of the NMR signals. However, this often entails solubility difficulties in polar solvents. In nonpolar media in contrast, one would observe too large association constants that make a non-linear regression analysis unfeasible. On the other hand, restrictions in the applicability can also derive from protonation reactions that give similar curve progressions or changes of the pH value during the titration experiment.

UV /Fluorescence spectroscopy

Optical methods like the UV and the fluorescence spectroscopy are able to circumvent the difficulties of NMR titration experiments by the usage of buffer solutions that suppress protonation reactions as well as by their sensitivity which allows working with low concentrations ($c \geq 10^{-7}$ M). However, optical methods require chromophores that absorb the radiated light in the specific wave length region. In case of UV spectroscopy (200-700 nm) aromatic compounds like pyrroles usually fulfil this criterion. The intrinsic molar property being detected is the molar extinction coefficient difference $\Delta\varepsilon$ at the wave length of the maximum absorption analogous to eq. 2.1.5. If no chromophore showing UV absorption or fluorescence is present in the receptor or in the substrate, an "indicator displacement assay" (IDA) can be used.[20] Hereby, the receptor is complexed by an indicator molecule called reporter which itself is then displaced by the addition of the actual substrate. The change in the absorption of the reporter can be used to evaluate the association constant of the receptor-substrate complex if the affinity of the reporter to the host is known.

Mass spectrometry (MS)

The thermodynamics and kinetics of host-guest complexation reactions are studied with increasing popularity by MS techniques since they allow a direct comparison of supramolecular complexes in solution and gas phase. It is thus possible to study the intrinsic molecular properties in the absence of solvent molecules.[21] However, soft ionization techniques are required which should preserve the host-guest complex and prevent fragmentation reactions. Besides various desorption methods (*e.g.* field desorption (FD), fast atom bombardment (FAB), matrix assisted laser desorption ionization (MALDI)) showing only limited scope, the electrospray ionization technique (ESI) has been proven to be a powerful tool for highly sensitive detection of specific association reactions (*e.g.* ligand-protein complexes).[22] Under certain assumptions competition experiments can be performed to determine relative binding constants to the receptor from the relative intensities of the complex peaks.

The soft ionization process occurring during an electrospray experiment can be separated into three stages.[23]

1. In the first stage small droplets are formed at the tip of a capillary. Due to the high voltage (positive pole) impressed on the capillary charge separation takes place and the enrichment of positive charge at the liquid surface causes the formation of a Taylor cone that results from the balance of the electric field and the surface tension. If the distance from the capillary end becomes too large, small droplets with positive excess charge are emitted in a constant spray (liquid filament) with $I \sim 10^{-7}$ A.
2. In the next step the droplets are reduced by solvent evaporation increasing the charge-to-volume ratio until the Raleigh limit is reached (electrostatic repulsion > surface tension). Elastic surface vibrations cause the fission of the droplets into micro-droplets and this procedure is repeated several times until only highly charged nano-droplets remain.
3. The formation of gaseous ions from the micro-droplets has been explained by two theories. The ‘charged residue model’ of Dole assumes that the nano-droplets contain only a single analyte ion that is brought into gas phase by further solvent evaporation. In contrast, the ‘ion emission model’ by Iribane and Thomson describes the formation of gaseous ions by emission from highly charged micro-droplets containing ~ 70 elementary charges and representing an alternative to a further Raleigh-decay of the droplet.

The principal function of the electrospray ionization mechanism is the disruption of the non-covalent interactions between analyte and solvent-molecules. In order to analyze non-covalently bound species the ESI-MS technique is thus reduced to ionisable associates that form considerably more stable complexes with each other than with the solvent molecules.

2.1.2 Determination of entropy and enthalpy contributions

The free energy of a complexation process composes of enthalpy and entropy contributions which can be determined either by calorimetry or by the temperature dependency of K.[18] The van’t Hoff isochore is given as

$$\frac{d \ln K}{dT} = \frac{\Delta H}{RT^2} \qquad \text{eq. 2.1.7}$$

and can be integrated under the assumption that the enthalpy difference ΔH is constant within a small temperature interval to

$$R\ln K = \int \Delta H \frac{1}{T^2} dT = \Delta H \int \frac{1}{T^2} dT = -\Delta H \frac{1}{T} + \Delta S \qquad \text{eq. 2.1.8}$$

The common approach is now to calculate ΔH as the slope of 1/T plotted against $R\ln K$ and the entropy difference is then determined from the free energy difference at standard temperature.

$$\Delta G = \Delta H - T\Delta S \qquad \text{eq. 2.1.9}$$

However, regarding common host-guest complexes the variation of ΔH often adds up to 20 kJ mol^{-1} for a temperature interval of 50 K. The temperature dependence of ΔH and ΔG is given by

$$\left(\frac{d\Delta H}{dT}\right)_p = T\left(\frac{d\Delta S}{dT}\right)_p = \Delta C_p \qquad \text{eq. 2.1.10}$$

Although the heat capacity itself also depends on temperature, an adequate assumption is to consider the change in heat capacity within a small temperature interval ΔC_p as more or less constant. Therefore, the enthalpy difference can be calculated as sum of a constant (intrinsic) value ΔH_i and a temperature correction.

$$\Delta H = \int d\Delta H = \int \Delta C_p dT = \Delta C_p T + \Delta H_i \qquad \text{eq. 2.1.11}$$

Analogously to eq. 2.1.10 and eq. 2.1.11 one write can the entropy difference as a sum of an intrinsic entropy ΔS_i and a temperature correction.

$$\Delta S = \int d\Delta S = \int \Delta C_p \frac{1}{T} dT = \Delta C_p \ln T + \Delta S_i \qquad \text{eq. 2.1.12}$$

By substituting the corrected expressions for the enthalpy and entropy differences (eq. 2.1.11 and eq. 2.1.12) into eq. 2.1.8 one gets a more sophisticated model for evaluation of enthalpy and entropy contributions to the Gibbs free energy.

$$R\ln K = -\Delta H_i \frac{1}{T} + \Delta S_i + \Delta C_p (\ln T - 1) \qquad \text{eq. 2.1.13}$$

It should be noted that enthalpies and heat capacities determined by the temperature dependence of K are less precise than those obtained by calorimetry. Especially ΔC_p requires accurate association constants whereas ΔH values are less sensitive towards the error bars of the determined K values and the considered temperature intervals. Moreover, inconsistent reaction conditions over large temperature ranges can also deteriorate the results making a reliable determination of ΔC_p unfeasible by this approach.

2.2 Theoretically

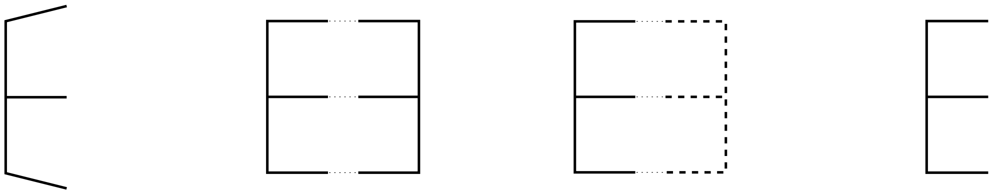
In principle, all theoretical models allow to determine directly the association energies of any given compound by the calculation of the potential energies of its components (see chapter 2.2.1). However, a sound description of complexation reactions as they occur in nature is only given by the Gibbs free energy difference of all Boltzmann weighted conformers composed of enthalpic and entropic terms. These contributions can be estimated either by analytical formulas for the gas phase (see chapter 2.2.2) or Molecular Dynamics simulations (see chapter 3.4) that are able to calculate free energy differences of reactions taking place in explicit solvation.

2.2.1 Calculation of dimerization energies

In the following only the aggregation of two components respectively monomeric units shall be considered. In general, the binding energy can be simply calculated by the energy differences of the individual monomers and the aggregate. However, by using electronic structure methods with basis set expansions for the electronic wave functions one has to encounter a basis set superposition error (BSSE). The binding energy of a non-covalently bound complex is overestimated since it is described by a much larger basis set than the monomers. A monomeric unit within a dimer can use additionally the basis functions of the counterpart to describe its own charge distribution. Metaphorically speaking, the electron density of one monomer is spread more over the whole dimer and, as a consequence, the complex system is described as too stable.

A larger basis set now decreases the basis set superposition error and for an infinite basis no error would occur. However, hard- and software restrictions limit the basis set size and a good estimate for the magnitude is necessary. The counterpoise (CP) correction after Boys and Bernardi gives an approximation of the BSSE by calculating the energy difference between the monomers in dimer geometry with and without additional basis functions on the atom positions of the monomeric counterparts in the dimer complex.[24] These orbitals are often denoted as "ghost orbitals" since they are not centred on any nucleus. To calculate the binding energy correctly, the difference in energy between the dimer and the monomers has to be reduced by the CP-correction.

$$\Delta E = \left\{ \left(E_A + E_B \right) - E_{A+B} \right\} - \left\{ \left(E_A^{Ghost} + E_B^{Ghost} \right) - \left(E_A^{Dimer} + E_B^{Dimer} \right) \right\} \qquad \text{eq. 2.2.1}$$



Monomer

Dimer

Monomer in dimer geometry with ghost orbitals

Monomer in dimer geometry without ghost orbitals

Figure 4 Various components of the counterpoise correction.

The magnitude of the counterpoise correction for medium-sized dimer complexes showing ionic interactions as considered within this work ranges from at least 25 kJ mol^{-1} for a single valence basis set to 5-8 kJ mol^{-1} for a triple ξ basis set with additional diffuse functions.

2.2.2 Calculation of entropy and enthalpy contributions

Thermodynamical entities like entropy and enthalpy of a given molecule in gas phase are accessible via the molecular partition function that can be dissected approximatively into a translational, rotational, vibrational and electronical part.[25] Since the latter component is described mainly by electronic structure calculations of the ground state, only the corrections for the translation, rotation and vibration have to be evaluated. The partition functions are approximated by the expression for an ideal gas in gas phase, *i.e.* no coupling between the degrees of freedom is assumed. The various partition functions can therefore be written (in a.u.) as

$$q_{trans} = v \cdot \left(\frac{2\pi \cdot m \cdot k \cdot T}{h^2} \right)^{\frac{3}{2}} \qquad \text{eq. 2.2.2}$$

$$q_{rot} = \frac{\sqrt{(2\pi \cdot kT)^3 \cdot I_A \cdot I_B \cdot I_C}}{\sigma \cdot \pi} \qquad \text{eq. 2.2.3}$$

$$q_{vib} = \prod_i \frac{1}{\left(1 - \exp\left(-\varepsilon_i / kT\right)\right)} \qquad \text{eq. 2.2.4}$$

Hereby, *v* denotes the volume per molecule of an ideal gas, *I* denotes the respective moment of inertia and *σ* is the symmetry number of the molecule. The zero point vibrational energy (ZVPE) in a.u. is calculated using a harmonic approximation to

$$ZPVE = f \cdot \frac{1}{2} \sum_i \varepsilon_i \qquad \text{eq. 2.2.5}$$

with *f* representing a scaling factor in the magnitude of 0.9 in order to adjust the overestimation of observed fundamentals by SCF calculated wavenumbers. The statistical thermodynamic gives the zero point energy corrected chemical potential as

$$\mu = ZPVE - RT \cdot \ln\left(q_{trans} \cdot q_{rot} \cdot q_{vib}\right) \qquad \text{eq. 2.2.6}$$

The inner energy can be written as the sum of the contributions of the translatory, rotatory and vibrational partition function, of which the first two are given by the simplified expressions for a rigid rotator by which every degree of freedom is set to 3/2 RT.[26] Only the inner energy of the vibrational partition function must be calculated as a sum over all frequencies.

$$\begin{aligned} U &= ZPVE + 3RT + u_{vib} = ZPVE + 3RT + RT^2 \frac{\partial \ln q_{vib}}{\partial T} = \\ &= ZPVE + 3RT + \sum_i \frac{\varepsilon_i \cdot \left(1 + \exp(-\varepsilon_i / kT)\right)}{2 \cdot \left(1 - \exp(-\varepsilon_i / kT)\right)} \end{aligned} \qquad \text{eq. 2.2.7}$$

By making use of the ideal gas law the enthalpy is simply given as

$$H = U + RT \qquad \text{eq. 2.2.8}$$

Finally, the entropy can be calculated by applying the following expression.

$$S = \frac{(H - \mu)}{T} \qquad \text{eq. 2.2.9}$$

These equations are based on approximations that are only applicable if the rotation can be treated classically, the molecules form an ideal gas or solution and the harmonic approximation regarding the vibrational modes are approvable. Other formulations in quantum chemistry software have been used that calculate the various contributions by inserting eq. 2.2.2 - eq. 2.2.4 into the statistical expression of the entropy.

$$S = kT \cdot \left(\frac{\partial \ln q}{dT} \right) + k \ln q \qquad \text{eq. 2.2.10}$$

In case of the translatory partition function one obtains the well-known Sackur-Tetrode equation for a monoatomic gas.

Chapter 3 Theoretical Methods

"Anyone who is not shocked by quantum theory has not understood it."

--***Niels Henrik David Bohr***

In this chapter the main features of the theoretical methods used throughout this work are briefly described. A detailed description of the various approaches is hardly feasible and therefore the reader is referred to the references for further information. The main focus is rather to give an overview over the shortcoming and the advantage of each method than to provide a full description. The aim is to show the problems modern theoretical chemistry faces when treating mid-range bioorganic molecules in a realistic, *i.e.* aqueous surrounding. An appropriate treatment therefore requires the usage of both quantum mechanics and classical mechanics.

3.1 Electronic structure methods

The explicit treatment of the electronic structure of a molecule is mandatory if one wants to describe inherent properties like ground state and excitation energies, charge distributions and interactions with external fields (electric or magnetic). In the recent past two approaches are commonly used, namely the electron correlation methods footed on the single determinant Hartree-Fock (HF) approach[27] (chapter 3.1.1 and 3.1.2), and the density functional theory (DFT)[28], which solves the Kohn-Sham equations (chapter 3.1.3) that are closely related to the HF formalism. Both methods can not be applied in a "black-box" manner since at least a basic knowledge of their theoretical fundament is needed in order to produce reasonable results (chapter 3.1.4).

3.1.1 A short review on the self-consistent field (SCF) theory

The following equations are all written in atomic units, *i.e.* the Planck's constant $\hbar$, the electron mass m_e and the permittivity of the vacuum $4\pi\varepsilon_0$ are all set to unity, whereas M_A, the mass of nucleus A, is given as a multiple of the electron mass. The time-independent, non-relativistic Schrödinger equation is given as

$$\mathbf{H}\Psi(\mathbf{R},\mathbf{r}) = E\Psi(\mathbf{R},\mathbf{r}) \qquad \text{eq. 3.1.1}$$

with **r** and **R** denoting the spatial electron and nuclear coordinates, respectively. The Hamilton operator is defined as a sum of kinetic and potential operators, the latter represent the electron-electron, electron-nucleus and nucleus-nucleus Coulomb interactions.

$$\begin{aligned}\mathbf{H} &= \mathbf{T}_e + \mathbf{T}_n + \mathbf{V}_{ee} + \mathbf{V}_{ne} + \mathbf{V}_{nn} \\ &= \sum_{i=1}^{N}\left(-\frac{1}{2}\nabla_i^2\right) + \sum_{A=1}^{M}\left(-\frac{1}{2M_A}\nabla_A^2\right) + \sum_{i}^{N}\sum_{j>i}^{N}\frac{1}{\left|\mathbf{r}_i - \mathbf{r}_j\right|} + \sum_{i}^{N}\sum_{A}^{M}\left(-\frac{Z_A}{\left|\mathbf{r}_i - \mathbf{R}_A\right|}\right) + \sum_{A}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{\left|\mathbf{R}_A - \mathbf{R}_B\right|}\end{aligned}$$

eq. 3.1.2

Since the mass of a nucleus is much greater than those of an electron, the heavy nuclei move more slowly than the light electrons. Their movement can therefore be considered in a reasonable approximation as being decoupled from the electron movement.

The Born-Oppenheimer approximation

In the Born-Oppenheimer approximation the nuclear kinetic energy operator $\mathbf{T}_n$ is neglected and the Schrödinger equation is solved only for the electronic part of the wave function for a fixed geometry of the nuclei ("clamped nuclei assumption"). The total wave function can be written in terms of a Taylor series with the expansion coefficients as functions of the nuclear coordinates.

$$\Psi(\mathbf{r},\mathbf{R}) = \sum_{i=1}^{\infty} \Phi_i(\mathbf{R})\psi_i(\mathbf{r},\mathbf{R}) \qquad \text{eq. 3.1.3}$$

The electronic Schrödinger equation therefore is given by

$$\mathbf{H}_{el}\psi_i(\mathbf{R},\mathbf{r}) = (\mathbf{T_e} + \mathbf{V_{ee}} + \mathbf{V_{ne}} + \mathbf{V_{nn}})\psi_i(\mathbf{R},\mathbf{r}) = \varepsilon_i^{el}\psi_i(\mathbf{R},\mathbf{r}) \qquad \text{eq. 3.1.4}$$

with ε_i^{el} denoting the eigenvalues of the electronic Schrödinger equation that give the potential energy surface (PES) for different positions of the nuclei. The dynamics of the nuclei is described by the Schrödinger equation of nuclear motion that gives as result the vibrations and rotations of a molecular system.

$$(\mathbf{T}_n + \varepsilon_i^{el})\Phi_i(\mathbf{R}) = E\Phi_i(\mathbf{R}) \qquad \text{eq. 3.1.5}$$

The exact solution of the Schrödinger equation also includes coupling elements C_{ij} between different electronic states which basically are contributions that derive from the nuclear kinetic operator acting on the electronic wave function.

$$(\mathbf{T}_n + \varepsilon_i^{el})\Phi_i(\mathbf{R}) + \sum_j \mathbf{C}_{ij}\chi_j(\mathbf{R}) = E\Phi_i(\mathbf{R}) \qquad \text{eq. 3.1.6}$$

In the Born-Oppenheimer approximation these coupling elements are now set to zero assuming an adiabatic behaviour of the electronic wave function. However, if the electronic wave function changes drastically along a nuclear coordinate ("avoided crossing"), the Born-Oppenheimer approximation fails and one has to take into account the coupling between different electronic states.

The Slater determinant

For the solution of the electronic Schrödinger equation one uses a wave function that incorporates the electron spin as a quantum effect by a product ansatz for the one-electron functions.

$$\varphi_i(\mathbf{x}_i) = \phi(\mathbf{r}_i) * \Omega(\boldsymbol{\omega}_i) \qquad \text{eq. 3.1.7}$$

The spin orbitals $\varphi_i(\mathbf{x}_i)$ consist of a spatial orbital part and a spin function part $\Omega(\boldsymbol{\omega})$, that can have either α or β spin, and which obeys the orthonormality condition $\delta_{\alpha\beta}$.

Regarding now a many-electron system, the Pauli principle requires an antisymmetric wave function with respect to the interchange of two electron coordinates. This is equal to the condition that electrons have to differ in at least one quantum number. A further aspect that one has to allow for is the indistinguishability of two electrons. As a result, a simple product ansatz consisting of one-electron functions (Hartree product) fails and the electronic wave function is therefore constituted by a Slater determinant.

$$\psi^S(\mathbf{x}_1 \cdots \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(\mathbf{x}_1) & \dots & \varphi_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \varphi_1(\mathbf{x}_N) & \dots & \varphi_N(\mathbf{x}_N) \end{vmatrix} = \left| \varphi_1(\mathbf{x}_1)\varphi_2(\mathbf{x}_2) \cdots \varphi_N(\mathbf{x}_N) \right| \qquad \text{eq. 3.1.8}$$

The Slater determinant fulfils all conditions mentioned above as it can be proved by regarding the spatial pair density $\left|\psi^S\right|^2 d\mathbf{r}_1 d\mathbf{r}_2$ of two arbitrary electrons. This quantity can be obtained by integrating the pair density over the electron spins which itself can be either alike or unlike.

The energy of a Slater determinant – The Hartree-Fock equations

The energy of an approximate and normalized wave function like the Slater determinant can be evaluated by calculating the expectation value of the Hamilton operator.

$$E_{el} = \frac{\langle \psi | \mathbf{H}_{el} | \psi \rangle}{\langle \psi | \psi \rangle} = \langle \psi | \mathbf{H}_{el} | \psi \rangle \qquad \text{eq. 3.1.9}$$

The variational principle now states that the exact energy of the ground state E_0 is a lower bound to the electronic energy of the approximate wave function.

$$E_0 \leq E_{el} \qquad \text{eq. 3.1.10}$$

In principle, by skilful choice of the trial wave function one is able to calculate the exact energy of any system. The parameters of the approximate wave function which can be varied are

1. the expansion coefficients of the atomic orbitals when expanding the Slater determinant into a basis (see "*Introduction of a basis – The Roothaan-Hall equations*") and
2. the expansion coefficients of a generic multi-determinant trial wave function (see Chapter 3.1.2).

The more flexible the trial wave function is constructed, the closer the expectation value will be to the exact ground state energy. The energy of a single determinant ansatz is described by the Hartree-Fock energy expression, in which the operators have been sorted according to the number of electron indices. The one and two electron operators are introduced as $\mathbf{h}_i = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|}$ and $\mathbf{g}_{ij} = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$, so that the electronic Hamilton operator gives

$$\mathbf{H}_{el} = \sum_{i=1}^{N} \mathbf{h}_i + \sum_{i=1}^{N}\sum_{j>i}^{N} \mathbf{g}_{ij} + \mathbf{V}_{nn} \qquad \text{eq. 3.1.11}$$

By representing the Slater determinant through a permutation operator acting on the Hartree product, it is possible to evaluate all terms which contribute to the electronic energy and the expectation value can be written as

$$E_{HF} = \sum_{i=1}^{N} \langle \varphi_i(\mathbf{x}_1) | \mathbf{h}_i | \varphi_i(\mathbf{x}_1) \rangle + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \left(\langle \varphi_j(\mathbf{x}_2) | \mathbf{J}_i | \varphi_j(\mathbf{x}_2) \rangle - \langle \varphi_j(\mathbf{x}_2) | \mathbf{K}_i | \varphi_j(\mathbf{x}_2) \rangle \right) + V_{nn}$$

with

$$\mathbf{J}_i \left| \varphi_j(\mathbf{x}_2) \right\rangle = \langle \varphi_i(\mathbf{x}_1) | \mathbf{g}_{12} | \varphi_i(\mathbf{x}_1) \varphi_j(\mathbf{x}_2) \rangle \qquad \text{eq. 3.1.12}$$

$$\mathbf{K}_i \left| \varphi_j(\mathbf{x}_2) \right\rangle = \langle \varphi_i(\mathbf{x}_1) | \mathbf{g}_{12} | \varphi_i(\mathbf{x}_2) \varphi_j(\mathbf{x}_1) \rangle$$

The **J**$_i$ operator is denoted as Coulomb operator and it describes the classical coulomb repulsion between two charges, whereas the **K**$_i$ operator is called the exchange operator that has no classical analogy. It interchanges the coordinates of two functions on the right hand side of the integral and results from the claim for an antisymmetric wave function. In other words, it accounts for the correlated movement of electrons having the same spin and that try to avoid each other due to the Pauli principle (“Fermi hole”).

The basic concept of Hartree-Fock is now to minimize the energy value by a variation of the spin orbitals under the orthonormality condition.

$$\frac{\partial\langle\psi|\mathbf{H}_{el}|\psi\rangle}{\partial\varphi_i}=0 \qquad \text{eq. 3.1.13}$$

with $\langle\varphi_i|\varphi_j\rangle=\delta_{ij}$

The constrained optimization problem can be solved by the introduction of Lagrange multipliers ε_{ij} into the linear system of equations.

$$L=E-\sum_i^N\sum_j^N\varepsilon_{ij}\left(\langle\varphi_i|\varphi_j\rangle-\delta_{ij}\right) \qquad \text{eq. 3.1.14}$$

The Lagrange function is stationary with respect to an orbital change.

$$\partial L=\partial E-\sum_i^N\sum_j^N\varepsilon_{ij}\left(\langle\partial\varphi_i|\varphi_j\rangle-\langle\varphi_i|\partial\varphi_j\rangle\right)=0 \qquad \text{eq. 3.1.15}$$

Inserting and differentiating the energy expression of eq. 3.1.12 leads to

$$\partial L=\sum_i^N\left(\langle\partial\varphi_i|\mathbf{F}_i|\varphi_i\rangle-\langle\varphi_i|\mathbf{F}_i|\partial\varphi_i\rangle\right)-\sum_i^N\sum_j^N\varepsilon_{ij}\left(\langle\partial\varphi_i|\varphi_j\rangle-\langle\varphi_i|\partial\varphi_j\rangle\right)=0 \qquad \text{eq. 3.1.16}$$

with

$$\mathbf{F}_i=\mathbf{h}_i+\sum_j^N(\mathbf{J}_j-\mathbf{K}_j)=\mathbf{h}_i+\boldsymbol{\upsilon}^{HF}(i) \qquad \text{eq. 3.1.17}$$

The Fock operator $\mathbf{F}_i$ is an effective one-electron energy operator describing the kinetic energy and the electron-nuclei attraction of an electron ($\mathbf{h}_i$ operator), as well as the energy within the mean field of all other electrons (Hartree-Fock potential $\boldsymbol{\upsilon}^{HF}$). With $\langle\varphi|\mathbf{O}|\partial\varphi\rangle = \langle\partial\varphi|\mathbf{O}|\varphi\rangle^*$ eq. 3.1.1 can be reformulated to

$$\partial L = \sum_i^N \langle\partial\varphi_i \| \left[\mathbf{F}_i|\varphi_i\rangle - \sum_j^N \varepsilon_{ij}|\varphi_j\rangle \right] + \text{complex conjugated} = 0 \qquad \text{eq. 3.1.18}$$

The expression within the brackets must equal zero which leads to the well-known Hartree-Fock equations.

$$\mathbf{F}_i|\varphi_i\rangle - \sum_j^N \varepsilon_{ij}|\varphi_j\rangle = 0 \quad \Rightarrow \quad \mathbf{F}_i|\varphi_i\rangle = \sum_j^N \varepsilon_{ij}|\varphi_j\rangle \qquad \text{eq. 3.1.19}$$

A unitary transformation diagonalizes the Lagrange multipliers matrix which leads to a set of pseudo-eigenvalue equations of canonical orbitals.

$$\mathbf{F}_i|\varphi_i'\rangle = \varepsilon_i|\varphi_i'\rangle \qquad \text{eq. 3.1.20}$$

The Lagrange multipliers are the expectation values of the Fock operator $\varepsilon_i = \langle\varphi_i'|\mathbf{F}_i|\varphi_i'\rangle$ and give the respective MO energies. Since the Fock operator depends on all occupied orbitals, these equations have to be solved iteratively until self-consistency is reached. The resulting orbitals are therefore called either *Hartree-Fock* (HF) or *Self-Consistent-Field* (SCF) orbitals. The total energy is the nuclear potential plus the sum over all SCF orbitals minus a correction term since the Fock operator counts the **J** and **K** operator twice.

$$E_{HF} = V_{nn} + \sum_{i=1}^N \varepsilon_i - \frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N \Big(\langle\varphi_j(\mathbf{x}_2)|\mathbf{J}_j|\varphi_j(\mathbf{x}_2)\rangle - \langle\varphi_j(\mathbf{x}_2)|\mathbf{K}_j|\varphi_j(\mathbf{x}_2)\rangle \Big) \qquad \text{eq. 3.1.21}$$

The Fock equations for a closed shell system can be derived by the separation of the spatial and spin functions and integrating from the left with $\langle\alpha(\omega_1)|$. The final equations can then be written as

$$\mathbf{f}_i|\phi_i\rangle = \varepsilon_i|\phi_i\rangle \qquad \text{eq. 3.1.22}$$

$$\mathbf{f}_i = \mathbf{h}_i + \sum_{j}^{N/2} (2\mathbf{J}_j - \mathbf{K}_j) \qquad \text{eq. 3.1.23}$$

with $\mathbf{f}_i$ as the closed-shell Fock operator

Introduction of a basis – The Roothaan-Hall equations

So far no further specifications about the nature of the molecular orbitals have been made. For finite sized molecular systems it is convenient to expand each MO in terms of basis functions that represent atomic orbitals.

LCAO ansatz: $\phi_i = \sum_{\nu=1}^{M} c_{\nu i} \chi_\nu$ eq. 3.1.24

Inserting eq. 3.1.24 into eq. 3.1.22 and integrating from the left with $\langle \chi_\mu |$ yields

$$\sum_{\nu=1}^{M} c_{\nu i} \langle \chi_\mu | \mathbf{f}_i | \chi_\nu \rangle = \varepsilon_i \sum_{\nu=1}^{M} c_{\nu i} \langle \chi_\mu | \chi_\nu \rangle \qquad \text{eq. 3.1.25}$$

This equation can be rewritten in short hand form as

$$\sum_{\nu=1}^{M} F_{\mu\nu} c_{\nu i} = \varepsilon_i \sum_{\nu=1}^{M} S_{\mu\nu} c_{\nu i} \qquad \text{, or as} \qquad \text{eq. 3.1.26}$$

$$\mathbf{FC} = \mathbf{SC\varepsilon} \qquad \text{eq. 3.1.27}$$

in matrix notation. $F_{\mu\nu}$ and $S_{\mu\nu}$ denote the Fock matrix and overlap matrix elements, respectively. The Fock matrix elements contain two-electron operators which are also expanded into a basis.[29]

$$\begin{aligned} F_{\mu\nu} &= \langle \chi_\mu(1) | \mathbf{h} | \chi_\nu(1) \rangle + \sum_{j}^{N/2} \langle \chi_\mu(1) | 2\mathbf{J}_i - \mathbf{K}_i | \chi_\nu(1) \rangle \\ &= h_{\mu\nu} + \sum_{j}^{N/2} \sum_{\lambda\sigma}^{AO} c_{\lambda j} c_{\sigma j} \cdot 2 \langle \chi_\mu(1) \chi_\lambda(2) \| \chi_\nu(1) \chi_\sigma(2) \rangle - \sum_{j}^{N/2} \sum_{\lambda\sigma}^{AO} c_{\lambda j} c_{\sigma j} \cdot \langle \chi_\mu(1) \chi_\lambda(2) \| \chi_\nu(2) \chi_\sigma(1) \rangle \end{aligned}$$

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\lambda\sigma} D_{\lambda\sigma} \cdot \left[(\mu\nu|\lambda\sigma) - \frac{1}{2} (\mu\sigma|\lambda\nu) \right] = h_{\mu\nu} + \sum_{\lambda\sigma} D_{\lambda\sigma} \cdot G_{\mu\nu\lambda\sigma} \qquad \text{eq. 3.1.28}$$

The AO coefficients are summarized to a density matrix $D_{\lambda\sigma}$ and the electron density can then be expressed in terms of the basis functions.

$$D_{\lambda\sigma} = 2\sum_{j}^{N/2} c_{\lambda j} c_{\sigma j} \tag{eq. 3.1.29}$$

$$\rho(\mathbf{r}) = \sum_{\lambda\sigma} D_{\lambda\sigma} \chi_{\lambda}^{*}(\mathbf{r}) \chi_{\sigma}(\mathbf{r}) \tag{eq. 3.1.30}$$

The SCF procedure

Since the Fock matrix depends on its own solution, the Roothaan-Hall equations must be solved iteratively as illustrated below.

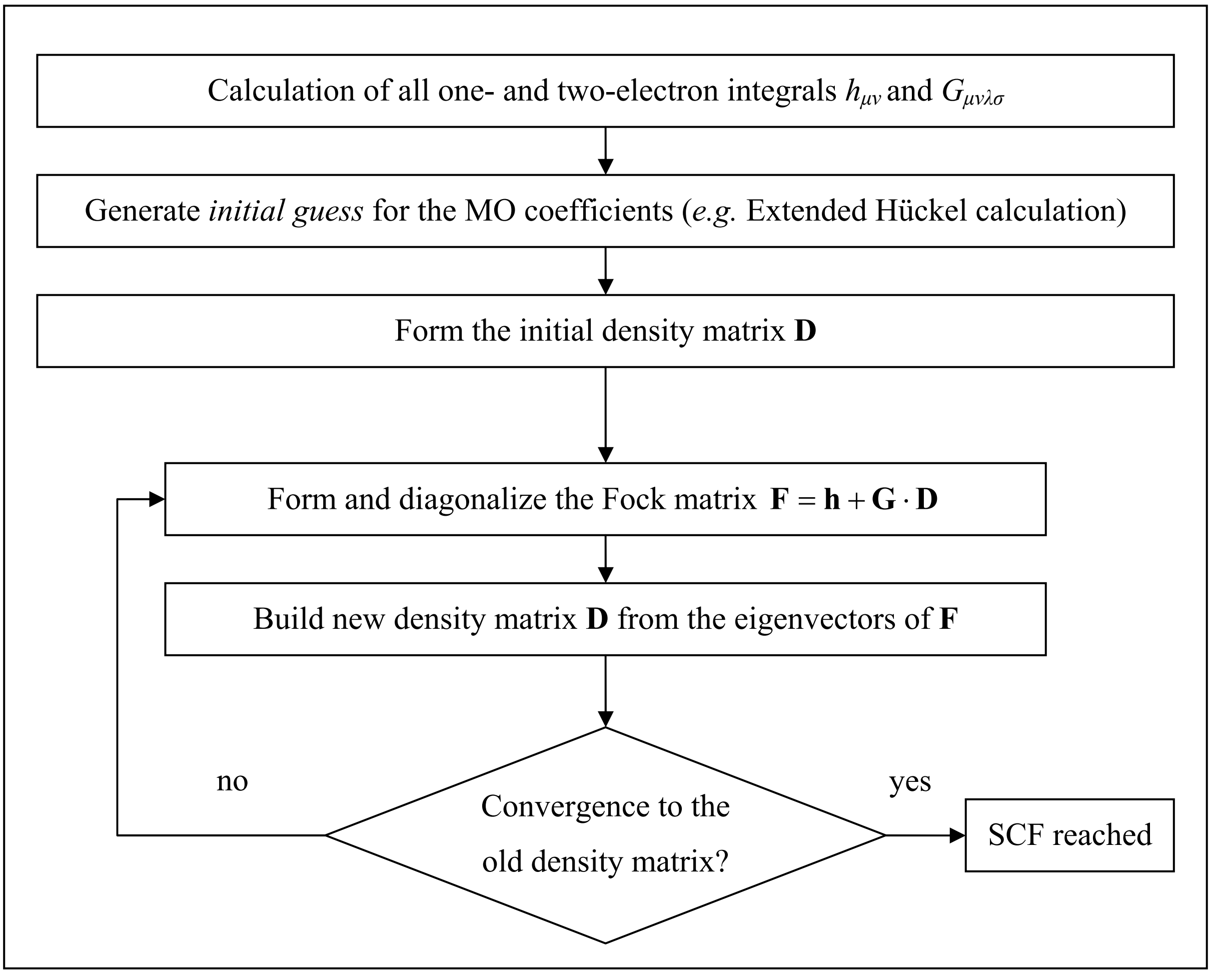

Calculation of all one- and two-electron integrals $h_{\mu\nu}$ and $G_{\mu\nu\lambda\sigma}$
Generate *initial guess* for the MO coefficients (*e.g.* Extended Hückel calculation)
Form the initial density matrix **D**
Form and diagonalize the Fock matrix $\mathbf{F} = \mathbf{h} + \mathbf{G} \cdot \mathbf{D}$
Build new density matrix **D** from the eigenvectors of **F**
Convergence to the old density matrix?
no
yes
SCF reached

Figure 5 Flowchart of the SCF procedure

3.1.2 Electron correlation methods

In the basis set limit the Hartree-Fock theory gives the exact energy of a Slater determinant wave function. The wave function fully incorporates the exchange correlation by the antisymmetric nature of the Slater determinant. However, since the Fock operator is only an effective one-electron operator which describes the interaction of an electron within the averaged field of the surrounding electrons, the Coulomb correlation is neglected in a restricted HF ansatz. The correlation energy is thus defined as the Coulomb correlation given by the energy difference between the Hartree-Fock limit and the exact (non-relativistic) energy.

Static and dynamic electron correlation

The electron correlation can be further decomposed into a static (near-degeneracy) correlation and a dynamic correlation. The static correlation can be seen by the wrong dissociation behaviour of a diatomic homonuclear covalent bond if only a single configuration of the system is taken into account. The Hartree-Fock ansatz gives in the dissociation limit an unphysical wave function with too large ionic contributions. By using a multi-determinant ansatz with already a few other configurations this bond dissociation failure can be circumvented. The HF wave function alone is not flexible enough to describe such kind of systems in a correct manner.

The dynamic correlation comprises the instantaneous electrostatic interaction between electrons and therefore takes into account the explicit correlated movement of two electrons due to their electronic repulsion. Hence, the electronic wave function exhibits a hole at small distances of two electrons (“Coulomb cusp”) that is not described by the Hartree-Fock theory.

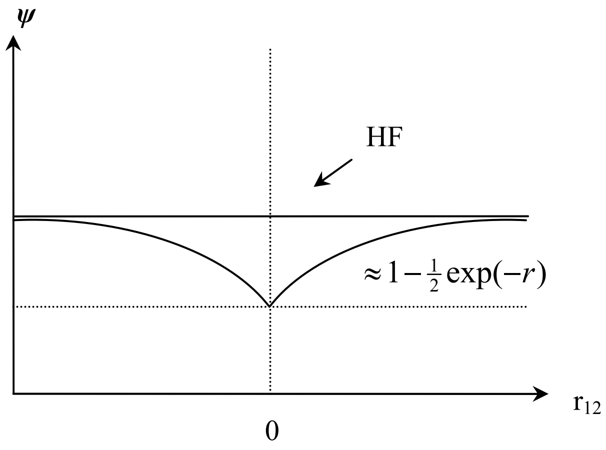


Figure 6 Graphical representation of the Coulomb cusp

Available electron correlation methods

The most straightforward method to include electron correlation is the configuration interaction (CI) of multiply excited determinants (singles (S), doubles (D), triples (T) etc.). The CI wave function is set up by a linear combination of Slater determinants weighted by expansion coefficients c_i, which are determined by solving the matrix eigenvalue problem $\mathbf{Hc} = E_{CI}\mathbf{c}$ either numerically or iteratively. The diagonalization of the CI-matrix **H** yields the CI energy. In general, CI is capable to calculate the exact energy within a given basis set if all possible excitations are considered ("full CI"). However, the fast growing size of the CI matrix makes this an unfeasible task already for small organic molecules so that different truncations have to be made. Variations of the configuration interaction scheme are the multi-configurational SCF (MCSCF) methods like CASSCF or RASSCF as well as the multi-reference CI ansatz. A detailed description of these methods is waived - chapter 3.1.4 will give a classification of the CI method within other electron correlation methods.

The electron correlation approaches that have been used throughout this work were the Møller-Plesset perturbation method 2nd order (MP2) based on the general many-body perturbation theory (chapter 3.1.2.1) and the coupled-cluster method (chapter 3.1.2.2). The latter has been fully applied for single and double excitations – triple excitations were approximated by perturbation methods. The so-called CCSD(T) approach is regarded as the "gold"-standard for single-point calculations in theoretical organic chemistry since it presents in conjunction with a reasonable basis set the best compromise of high accuracy and feasibility for small organic compounds.

3.1.2.1 Møller-Plesset perturbation theory

The basic concept behind perturbation theory is that a multi-dimensional problem can be approximated by the solution of a similar, but resolvable problem. The difference between the both problems can be regarded as a perturbation of the already known problem. The choice of the partitioning is in principle arbitrary, however, an important condition is that the perturbation must only be small in comparison to the reference system. The application field of this approach ranges from astrophysics (*e.g.* the calculation of planetary movement) to the description of matter in electric or magnetic fields. A general theoretical framework for describing a quantum mechanical perturbation is the many-body perturbation theory (MBPT) by Raleigh and Schrödinger. Starting from the MBPT equations, Møller and Plesset have developed a method to compute the electron correlation energy, in which the Coulomb correlation is seen as a perturbation of the uncorrelated Fock equations.[27]

Many-body perturbation theory

The many-body perturbation theory defines a Hamilton operator to consist of a reference Hamiltonian $\mathbf{H}^{(0)}$ and a perturbation operator $\mathbf{H}'$. The strength of the perturbation can be varied by the parameter λ.

$$\mathbf{H} = \mathbf{H}_0 + \lambda \mathbf{H}' \qquad \text{eq. 3.1.31}$$

with $0 < \lambda < 1$

The known solutions for the reference Hamiltonian can be chosen to be orthonormal eigenfunctions of the unperturbed system.

$$\mathbf{H}_0 \left| \psi_i^{(0)} \right\rangle = E_i^{(0)} \left| \psi_i^{(0)} \right\rangle$$

$$\text{with } \left\langle \psi_i^{(0)} \middle| \psi_j^{(0)} \right\rangle = \delta_{ij} \qquad \text{eq. 3.1.32}$$

For the perturbed system in the ground state the Schrödinger equation can be written as

$$\mathbf{H} \left| \Psi \right\rangle = (\mathbf{H}_0 + \lambda \mathbf{H}') \left| \Psi \right\rangle = E \left| \Psi \right\rangle \qquad \text{eq. 3.1.33}$$

If no perturbation takes place (λ = 0), eq. 3.1.33 will become equal eq. 3.1.32 to for the lowest energy state. The parameter λ can have any arbitrary value in the range from zero to one and the perturbed wave function and energy can be expressed in Taylor series in powers of the perturbation parameter.

$$\left| \Psi \right\rangle = \left| \psi^{(0)} \right\rangle + \lambda \left| \psi^{(1)} \right\rangle + \lambda^2 \left| \psi^{(2)} \right\rangle + \lambda^3 \left| \psi^{(3)} \right\rangle + \ldots \qquad \text{eq. 3.1.34}$$

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \ldots \qquad \text{eq. 3.1.35}$$

The perturbed wave function $\left| \Psi \right\rangle$ is intermediately normalized to the reference wave function (eq. 3.1.36). As a consequence from the orthonormality condition of eq. 3.1.32 one can see that all correction terms become orthogonal to the reference wave function.

$$\begin{aligned}\left\langle \Psi \middle| \psi^{(0)} \right\rangle &= 1 \\ &= \left\langle \psi^{(0)} + \lambda \psi^{(1)} + \lambda^2 \psi^{(2)} + ... \middle| \psi^{(0)} \right\rangle \\ &= \left\langle \psi^{(0)} \middle| \psi^{(0)} \right\rangle + \lambda \left\langle \psi^{(1)} \middle| \psi^{(0)} \right\rangle + \lambda^2 \left\langle \psi^{(2)} \middle| \psi^{(0)} \right\rangle + ..\end{aligned} \qquad \text{eq. 3.1.36}$$

$$\left\langle \psi^{(n)} \middle| \psi^{(0)} \right\rangle = 0 \qquad \text{eq. 3.1.37}$$

Inserting the Taylor expansions of eq. 3.1.34 and eq. 3.1.35 into eq. 3.1.33 yields

$$\begin{aligned}&(\mathbf{H}_0 + \lambda \mathbf{H}')\left(\left|\psi^{(0)}\right\rangle + \lambda\left|\psi^{(1)}\right\rangle + \lambda^2\left|\psi^{(2)}\right\rangle + \lambda^3\left|\psi^{(3)}\right\rangle + ...\right) \\ &= \left(E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + ...\right)\left(\left|\psi^{(0)}\right\rangle + \lambda\left|\psi^{(1)}\right\rangle + \lambda^2\left|\psi^{(2)}\right\rangle + \lambda^3\left|\psi^{(3)}\right\rangle + ...\right)\end{aligned} \qquad \text{eq. 3.1.38}$$

The brackets can now be expanded and the terms with the same power of λ are collected.

$$\begin{aligned}&\left\{\mathbf{H}_0\left|\psi^{(0)}\right\rangle - E^{(0)}\left|\psi^{(0)}\right\rangle\right\} + \\ &\lambda\left\{\mathbf{H}_0\left|\psi^{(1)}\right\rangle + \mathbf{H}'\left|\psi^{(0)}\right\rangle - E^{(0)}\left|\psi^{(1)}\right\rangle - E^{(1)}\left|\psi^{(0)}\right\rangle\right\} + \\ &\lambda^2\left\{\mathbf{H}_0\left|\psi^{(2)}\right\rangle + \mathbf{H}'\left|\psi^{(1)}\right\rangle - E^{(0)}\left|\psi^{(2)}\right\rangle - E^{(1)}\left|\psi^{(1)}\right\rangle - E^{(2)}\left|\psi^{(0)}\right\rangle\right\} + ... = 0\end{aligned} \qquad \text{eq. 3.1.39}$$

The choice of perturbation parameter is arbitrary and every bracket must therefore be equal zero giving the perturbation equations of the zero, first, second, n$^{\text{th}}$-order.

$$\begin{aligned}&\lambda^0 : \mathbf{H}_0\left|\psi^{(0)}\right\rangle = E^{(0)}\left|\psi^{(0)}\right\rangle \\ &\lambda^1 : \mathbf{H}_0\left|\psi^{(1)}\right\rangle + \mathbf{H}'\left|\psi^{(0)}\right\rangle = E^{(0)}\left|\psi^{(1)}\right\rangle + E^{(1)}\left|\psi^{(0)}\right\rangle \\ &\lambda^2 : \mathbf{H}_0\left|\psi^{(2)}\right\rangle + \mathbf{H}'\left|\psi^{(1)}\right\rangle = E^{(0)}\left|\psi^{(2)}\right\rangle + E^{(1)}\left|\psi^{(1)}\right\rangle + E^{(2)}\left|\psi^{(0)}\right\rangle \\ &... \\ &\lambda^n : \mathbf{H}_0\left|\psi^{(n)}\right\rangle + \mathbf{H}'\left|\psi^{(n-1)}\right\rangle = \sum_{i=0}^{n} E^{(i)}\left|\psi^{(n-i)}\right\rangle\end{aligned} \qquad \text{eq. 3.1.40}$$

These equations can now be integrated from the left with the reference wave function $\left\langle \psi^{(0)} \right|$ by making use of the intermediate normalization (eq. 3.1.37) and the hermiticity of the unperturbed Hamiltonian.

$$\left\langle \psi^{(0)} \middle| \mathbf{H}_0 \middle| \psi^{(n)} \right\rangle = \left\langle \psi^{(n)} \middle| \mathbf{H}_0 \middle| \psi^{(0)} \right\rangle^* = \left\langle \psi^{(n)} \middle| E^{(0)} \middle| \psi^{(0)} \right\rangle = E^{(0)} \left\langle \psi^{(n)} \middle| \psi^{(0)} \right\rangle = 0 \qquad \text{eq. 3.1.41}$$

As a result all correction energies of the nth-order can be expressed by matrix elements of $\mathbf{H}'$ of the unperturbed reference wave function and the (n-1)th-order wave function.

$$\begin{aligned}
\lambda^0 &: \left\langle \psi^{(0)} \middle| \mathbf{H}_0 \middle| \psi^{(0)} \right\rangle = \left\langle \psi^{(0)} \middle| E^{(0)} \middle| \psi^{(0)} \right\rangle = E^{(0)} \left\langle \psi^{(0)} \middle| \psi^{(0)} \right\rangle \\
&: E^{(0)} = \left\langle \psi^{(0)} \middle| \mathbf{H}_0 \middle| \psi^{(0)} \right\rangle \\
\lambda^1 &: \left\langle \psi^{(0)} \middle| \mathbf{H}_0 \middle| \psi^{(1)} \right\rangle + \left\langle \psi^{(0)} \middle| \mathbf{H}' \middle| \psi^{(0)} \right\rangle = E^{(1)} \left\langle \psi^{(0)} \middle| \psi^{(0)} \right\rangle - E^{(0)} \left\langle \psi^{(0)} \middle| \psi^{(1)} \right\rangle \\
&: E^{(1)} = \left\langle \psi^{(0)} \middle| \mathbf{H}' \middle| \psi^{(0)} \right\rangle \\
\lambda^2 &: \left\langle \psi^{(0)} \middle| \mathbf{H}_0 \middle| \psi^{(2)} \right\rangle + \left\langle \psi^{(0)} \middle| \mathbf{H}' \middle| \psi^{(1)} \right\rangle = E^{(2)} \left\langle \psi^{(0)} \middle| \psi^{(0)} \right\rangle - E^{(1)} \left\langle \psi^{(0)} \middle| \psi^{(1)} \right\rangle - E^{(0)} \left\langle \psi^{(0)} \middle| \psi^{(2)} \right\rangle \\
&: E^{(2)} = \left\langle \psi^{(0)} \middle| \mathbf{H}' \middle| \psi^{(1)} \right\rangle \\
&\ldots \\
\lambda^n &: E^{(n)} = \left\langle \psi^{(0)} \middle| \mathbf{H}' \middle| \psi^{(n-1)} \right\rangle
\end{aligned} \qquad \text{eq. 3.1.42}$$

Until now the wave function corrections $\left| \psi^{(n)} \right\rangle$ have not been specified. As the unperturbed Schrödinger equation results in a complete set of wave functions (eq. 3.1.32), it is possible to expand the unknown wave function corrections in terms of this basis.

$$\left| \psi^{(n)} \right\rangle = \sum_{k \neq 0} c_k^{(n)} \psi_k^{(0)} \qquad \text{eq. 3.1.43}$$

For the energy correction second-order, the first-order wave function correction thus can be written as

$$\left| \psi^{(1)} \right\rangle = c_1 \psi_1^{(0)} + c_2 \psi_2^{(0)} + c_3 \psi_3^{(0)} \ldots \qquad \text{eq. 3.1.44}$$

The expansion coefficients can be calculated by means of the excited unperturbed problem. Starting from the first-order perturbation equation (eq. 3.1.40), the terms are sorted by the perturbation index and are integrated from the left with the reference wave function in the kth excited state.

$$\begin{aligned}
\mathbf{H}_0 \left| \psi^{(1)} \right\rangle + \mathbf{H}' \left| \psi^{(0)} \right\rangle &= E^{(0)} \left| \psi^{(1)} \right\rangle - E^{(1)} \left| \psi^{(0)} \right\rangle \\
\mathbf{H}_0 \left| \psi^{(1)} \right\rangle - E^{(0)} \left| \psi^{(1)} \right\rangle &= -\mathbf{H}' \left| \psi^{(0)} \right\rangle - E^{(1)} \left| \psi^{(0)} \right\rangle \\
\left\langle \psi_k^{(0)} \middle| \mathbf{H}_0 \middle| \psi^{(1)} \right\rangle - \left\langle \psi_k^{(0)} \middle| E^{(0)} \middle| \psi^{(1)} \right\rangle &= -\left\langle \psi_k^{(0)} \middle| \mathbf{H}' \middle| \psi^{(0)} \right\rangle - E^{(1)} \left\langle \psi_k^{(0)} \middle| \psi^{(0)} \right\rangle \\
&= -\left\langle \psi_k^{(0)} \middle| \mathbf{H}' \middle| \psi^{(0)} \right\rangle
\end{aligned} \qquad \text{eq. 3.1.45}$$

Substituting the first order wave function correction with the Taylor expansion (eq. 3.1.44) and considering the intermediate normalization gives the expression for the kth expansion coefficient.

$$\begin{aligned}\left\langle \psi_k^{(0)} \middle| \mathbf{H}_0 \middle| \sum_{l\neq 0} c_l^{(1)} \psi_l^{(0)} \right\rangle - \left\langle \psi_k^{(0)} \middle| E^{(0)} \middle| \sum_{l\neq 0} c_l^{(1)} \psi_l^{(0)} \right\rangle &= -\left\langle \psi_k^{(0)} \middle| \mathbf{H}' \middle| \psi^{(0)} \right\rangle \\ c_l^{(1)} E_l^{(0)} \delta_{kl} - c_l^{(1)} E^{(0)} \delta_{kl} &= -\left\langle \psi_k^{(0)} \middle| \mathbf{H}' \middle| \psi^{(0)} \right\rangle \\ c_k^{(1)} \left(E^{(0)} - E_k^{(0)} \right) &= \left\langle \psi_k^{(0)} \middle| \mathbf{H}' \middle| \psi^{(0)} \right\rangle \\ c_k^{(1)} &= \frac{\left\langle \psi_k^{(0)} \middle| \mathbf{H}' \middle| \psi^{(0)} \right\rangle}{E^{(0)} - E_k^{(0)}}\end{aligned} \qquad \text{eq. 3.1.46}$$

The second-order energy correction can be written in the basis of the unperturbed excited wave functions.

$$E^{(2)} = \left\langle \psi^{(0)} \middle| \mathbf{H}' \middle| \sum_{k\neq 0} \psi_k^{(0)} \right\rangle c_k^{(1)} \qquad \text{eq. 3.1.47}$$

Finally, by inserting the expression for the expansion coefficients, the energy correction is written only in the basis of the unperturbed system.

$$E^{(2)} = \sum_{k\neq 0} \frac{\left\langle \psi^{(0)} \middle| \mathbf{H}' \middle| \psi_k^{(0)} \right\rangle \left\langle \psi_k^{(0)} \middle| \mathbf{H}' \middle| \psi^{(0)} \right\rangle}{E^{(0)} - E_k^{(0)}} \qquad \text{eq. 3.1.48}$$

The energy correction terms of higher orders can be derived analogously. However, when using perturbation theory for determining the electron correlation the second-order energy correction already accounts for ~80-90% of the correlation energy and higher order correction only play often a little role.

Møller-Plesset perturbation theory 2nd order

In order to calculate the correlation energy by means of perturbation theory one has to define a reasonable separation of the full many-body Hamiltonian into a reference Hamilton operator and a perturbation operator. Generally, this partition is arbitrary, but the perturbation should be as small as possible to assure good convergence. C. Møller and M. S. Plesset chose as zero-order reference Hamilton operator the sum of the one-electron Fock operators as defined in eq. 3.1.17 and as zero-order wave function the Hartree-Fock determinant.

$$\mathbf{H}^{(0)} = \sum_{i}^{occ} \mathbf{F}_i = \sum_{i}^{occ} \left(\mathbf{h}_i + \boldsymbol{\upsilon}^{HF}(i)\right) \qquad \text{eq. 3.1.49}$$

The perturbation operator is thus defined as the difference between the electronic Hamiltonian (eq. 3.1.11) and the reference Hamilton operator.[30]

$$\mathbf{H}' = \mathbf{H} - \mathbf{H}^{(0)} = \sum_{i<j}^{N} \mathbf{g}_{ij} - \sum_{i}^{occ} \boldsymbol{\upsilon}^{HF}(i) \qquad \text{eq. 3.1.50}$$

The Hartree-Fock potential operator is described as the sum of all Coulomb and exchange operators.

$$\begin{aligned} &\boldsymbol{\upsilon}^{HF}(i)\left|\varphi_k(\mathbf{x}_{2)}\right\rangle = \sum_{i}^{occ} (\mathbf{J}_i - \mathbf{K}_i)\left|\varphi_k(\mathbf{x}_{2)}\right\rangle \\ &= \sum_{i=1}^{occ} \left(\langle \varphi_i(\mathbf{x}_1)|\mathbf{g}_{12}|\varphi_k(\mathbf{x}_2)\varphi_i(\mathbf{x}_1)\rangle - \langle \varphi_i(\mathbf{x}_1)|\mathbf{g}_{12}|\varphi_i(\mathbf{x}_2)\varphi_k(\mathbf{x}_1)\rangle\right) \qquad \text{eq. 3.1.51} \\ &= \sum_{i=1}^{occ} \left(\langle i|\mathbf{g}_{12}|ki\rangle - \langle i|\mathbf{g}_{12}|ik\rangle\right) \end{aligned}$$

The zero-order energy is now just the sum over all occupied MO energies. As already shown by eq. 3.1.21, this is not the Hartree-Fock energy which must be corrected by a coulomb and an exchange term.

$$E^{(0)} = \left\langle \psi^{(0)} \middle| \mathbf{H}_0 \middle| \psi^{(0)} \right\rangle = \langle \psi_{HF} | \sum_{i}^{occ} \mathbf{F}_i | \psi_{HF} \rangle = \sum_{i}^{occ} \varepsilon_i \langle \psi_{HF} | \psi_{HF} \rangle = \sum_{i}^{occ} \varepsilon_i \qquad \text{eq. 3.1.52}$$

$$E_{HF} = E^{(0)} - \frac{1}{2} \sum_{i=1}^{occ} \sum_{j=1}^{occ} \left(\langle ij|\mathbf{g}_{ij}|ij\rangle - \langle ij|\mathbf{g}_{ij}|ji\rangle\right) \qquad \text{eq. 3.1.53}$$

The first order energy correction is given by

$$\begin{aligned} E^{(1)} &= \left\langle \psi^{(0)} \middle| \mathbf{H}' \middle| \psi^{(0)} \right\rangle = \left\langle \psi^{(0)} \middle| \sum_{i<j}^{N} \mathbf{g}_{ij} \middle| \psi^{(0)} \right\rangle - \left\langle \psi^{(0)} \middle| \sum_{i}^{occ} \boldsymbol{\upsilon}^{HF}(i) \middle| \psi^{(0)} \right\rangle \\ &= \frac{1}{2} \sum_{i=1}^{occ} \sum_{j=1}^{occ} \left(\langle ij | \mathbf{g}_{ij} | ij \rangle - \langle ij | \mathbf{g}_{ij} | ji \rangle \right) - \sum_{i}^{occ} \left\langle \psi^{(0)} \middle| \boldsymbol{\upsilon}^{HF}(i) \middle| \psi^{(0)} \right\rangle \\ &= -\frac{1}{2} \sum_{i=1}^{occ} \sum_{j=1}^{occ} \left(\langle ij | \mathbf{g}_{ij} | ij \rangle - \langle ij | \mathbf{g}_{ij} | ji \rangle \right) \end{aligned} \qquad \text{eq. 3.1.54}$$

By adding eq. 3.1.54 to eq. 3.1.53 one can see that the Hartree-Fock energy is the sum of the zero-order and first-order energy expression. The electron correlation energy is therefore taken into account beginning from the second-order energy correction. The energy correction term consists of matrix elements of $\mathbf{H}'$ between the Hartree-Fock determinant and all exited states (see eq. 3.1.48). All triple and higher excitations do not contribute to $E^{(0)}$ since the perturbation is a two-electron operator and thus all matrix elements vanish. For the singly excited states the matrix elements become also zero due to the orthonormality of the canonical HF orbitals and the Brillouin's theorem.

$$\begin{aligned} \left\langle \psi^{(0)} \middle| \mathbf{H}' \middle| \psi_a^r \right\rangle &= \left\langle \psi^{(0)} \middle| \mathbf{H} - \mathbf{H}^{(0)} \middle| \psi_a^r \right\rangle \\ &= \left\langle \psi^{(0)} \middle| \mathbf{H} \middle| \psi_a^r \right\rangle - \left\langle \psi^{(0)} \middle| \mathbf{H}^{(0)} \middle| \psi_a^r \right\rangle = \left\langle \psi^{(0)} \middle| \mathbf{H} \middle| \psi_a^r \right\rangle - \left\langle \psi_a^r \middle| \mathbf{H}^{(0)} \middle| \psi^{(0)} \right\rangle^* \\ &= \left\langle \psi^{(0)} \middle| \mathbf{H} \middle| \psi_a^r \right\rangle - E^{(0)} \left\langle \psi_a^r \middle| \psi^{(0)} \right\rangle^* = 0 \end{aligned} \qquad \text{eq. 3.1.55}$$

The Brillouin's theorem states that all CI matrix elements between the reference HF and singly excited Slater determinants vanish.

$$\begin{aligned} \left\langle \psi^{(0)} \middle| \mathbf{H} \middle| \psi_a^r \right\rangle &= \langle \chi_1 \cdots \chi_a \cdots \chi_N | \mathbf{H} | \chi_1 \cdots \chi_r \cdots \chi_N \rangle \\ &= \langle a | \mathbf{h} | r \rangle + \sum_{b}^{N} \langle a | \mathbf{J}_b | r \rangle - \langle a | \mathbf{K}_b | r \rangle = \langle a | \mathbf{F} | r \rangle = \varepsilon_r \langle a | r \rangle = 0 \end{aligned} \qquad \text{eq. 3.1.56}$$

Hence, only the doubly excited determinants contribute to the second-order energy correction which can therefore be written as

$$E^{(2)}=\sum_{a}\sum_{b<a}\sum_{r}\sum_{s<r}\frac{\left\langle\psi^{(0)}\middle|\mathbf{H}'\middle|\psi_{ab}^{rs}\right\rangle\left\langle\psi_{ab}^{rs}\middle|\mathbf{H}'\middle|\psi^{(0)}\right\rangle}{E^{(0)}-E_{ab}^{rs}}$$

$$=\sum_{a}\sum_{b<a}\sum_{r}\sum_{s<r}\frac{\left(\left\langle\psi^{(0)}\middle|\sum_{ij}\mathbf{g}_{12}\middle|\psi_{ab}^{rs}\right\rangle-\left\langle\psi^{(0)}\middle|\sum_{i}\boldsymbol{\upsilon}^{HF}(i)\middle|\psi_{ab}^{rs}\right\rangle\right)\left(\left\langle\psi_{ab}^{rs}\middle|\sum_{ij}\mathbf{g}_{12}\middle|\psi^{(0)}\right\rangle-\left\langle\psi_{ab}^{rs}\middle|\sum_{i}\boldsymbol{\upsilon}^{HF}(i)\middle|\psi^{(0)}\right\rangle\right)}{\sum_{i}^{occ}\varepsilon_{i}^{(0)}-\sum_{i}^{occ}\varepsilon_{i}^{ab\rightarrow rs}}$$

eq. 3.1.57

As $\boldsymbol{\upsilon}^{HF}(i)$ is an effective one-electron operator, these matrix elements become zero according to the Slater-Condon rules. By applying these rules to the remaining matrix elements, the final equation for the energy correction is obtained (in short-hand notation).

$$E^{(2)}=\sum_{a}\sum_{b<a}\sum_{r}\sum_{s<r}\frac{\left(\langle ab|rs\rangle-\langle ab|sr\rangle\right)^{2}}{\left(\varepsilon_{a}+\varepsilon_{b}\right)-\left(\varepsilon_{r}+\varepsilon_{s}\right)} \qquad \text{eq. 3.1.58}$$

Slater-Condon:

$$\begin{aligned}\left\langle\psi^{(0)}\middle|\mathbf{g}_{12}\middle|\psi_{ab}^{rs}\right\rangle&=\left\langle\varphi_{a}(\mathbf{x}_{1})\varphi_{b}(\mathbf{x}_{2})\middle|\mathbf{g}_{12}\middle|\varphi_{r}(\mathbf{x}_{1})\varphi_{s}(\mathbf{x}_{2})\right\rangle-\left\langle\varphi_{a}(\mathbf{x}_{1})\varphi_{b}(\mathbf{x}_{2})\middle|\mathbf{g}_{12}\middle|\varphi_{s}(\mathbf{x}_{1})\varphi_{r}(\mathbf{x}_{2})\right\rangle\\&=\langle ab|rs\rangle-\langle ab|sr\rangle\end{aligned}$$

The energy composed by the Hartree-Fock energy and its correction second-order is referred to as MP2 energy. A detailed discussion of the applicability and accuracy of the MP2 method is given in chapter 3.1.4.

3.1.2.2 Coupled-cluster theory

Introduced already in the 1950s by F. Coester and H. Kümmel and reformulated by J. Čížek in 1966 for quantum chemistry problems, the coupled-cluster theory (CC) has become one of the most popular single-reference approaches in last decades when determining accurate geometries and enthalpies of small compounds in gas phase.[27] Its major benefit arises from the strict multiplicativity of the chosen wave function resulting in a size consistent formulation of the electronic structure problem (see chapter 3.1.4). The basic concept in CC theory is to describe the exact wave function in a given basis set by an exponential ansatz of the HF wave function.

$$|\Psi\rangle = \exp(\mathbf{T})|\psi_{HF}\rangle \qquad \text{eq. 3.1.59}$$

The cluster operator **T** is defined as

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 + \ldots + \mathbf{T}_n \qquad \text{eq. 3.1.60}$$

with $\mathbf{T}_1$ denoting the excitation operator of all single excitations, $\mathbf{T}_2$ representing all double excitations and so on. In the creation/annihilation operator notation of the second quantization they are written as

$$\mathbf{T}_1 = \sum_{a,r} t_a^r \hat{a}_a^\dagger \hat{a}_r \qquad \text{eq. 3.1.61}$$

$$\mathbf{T}_2 = \sum_{a<b,r<s} t_{ab}^{rs} \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_r \hat{a}_s \qquad \text{eq. 3.1.62}$$

The expansion coefficients t are called amplitudes and have to be determined in order to find the approximate solution of $|\Psi\rangle$. The expansion of the cluster operator is restricted by the amount of electrons because no more than n excitations are possible. The exponential operator $\exp(\mathbf{T})$ can be expanded into a Taylor series given as

$$\exp(\mathbf{T}) = 1 + \mathbf{T} + \frac{\mathbf{T}^2}{2!} + \frac{\mathbf{T}^3}{3!} + \ldots = 1 + \mathbf{T}_1 + \mathbf{T}_2 + \frac{\mathbf{T}_1^2}{2!} + \mathbf{T}_1\mathbf{T}_2 + \frac{\mathbf{T}_2^2}{2!} + \ldots \qquad \text{eq. 3.1.63}$$

In practice, the expansion of **T** into excitation operators (see eq. 3.1.60) is terminated at second or third excitation class since contributions of higher excitation operators are very small. However, in contrast to CI methods also higher excitations are taken into account due to the exponential nature of the CC ansatz. Therefore, in case of the T_2 excitation operator also quadruple, sextuple etc. excitations are included.

The coupled-cluster energy is given as

$$E_{CC} = \langle \psi_{HF} | \mathbf{H} | \exp(\mathbf{T}) \psi_{HF} \rangle \qquad \text{eq. 3.1.64}$$

Expanding the exponential operator and making use of the fact that the Hamilton operator consists only of one and two electron operators one can write for the coupled-cluster energy (after applying Brillouins theorem and Slater-Condon rules)

$$E_{CC} = E_0 + \sum_{a<b,r<s} \left(t_{ab}^{rs} + t_a^r t_b^s - t_a^s t_b^r \right) \left(\langle \phi_a \phi_b | \phi_r \phi_s \rangle - \langle \phi_a \phi_b | \phi_s \phi_r \rangle \right) \qquad \text{eq. 3.1.65}$$

The coupled cluster energy is thus only determined by the singles and doubles amplitudes and the two-electron integrals. The determination of the cluster amplitudes t is done by left-multiplying the Schrödinger equation with excited determinants build from the HF reference wave function in order to generate as set of coupled, non-linear equations that have to be solved iteratively.

Inclusion of connected triple excitations arising from $\mathbf{T}_3$ is extremely expensive and only applicable to very small molecules. On this account various perturbation theoretical approaches to estimate the connected triples excitations have been developed. The most common method is denoted as CCSD(T) which fully includes the single and double excitation classes and uses a MP4-type correction for the triple excitations and terms describing the coupling between singles and triples.

3.1.3 Density functional theory

Density functional theory (DFT) in the framework of the Kohn-Sham theory was originally applied in the field of solid state physics already in the 1970s, but the models used by physicists were not accurate enough for quantum chemical calculations. The situation changed drastically in the beginning of the 1990s when new models emerged helping the DFT to its breakthrough and rapid ascent in computational chemistry. Nowadays, DFT has become the by far the most popular "tool" for investigating chemical questions and many program packages available require only little prior knowledge of the underlying theory. However, despite its great popularity one has to be aware of the conceptual deficiencies of DFT in order to produce reasonable results. Hence, a closer look at the theoretical basis is unavoidable and shall be given in this chapter, before chapter 3.1.4 will classify DFT in the context of the electron correlation methods.[28]

The exchange-correlation hole

In contrast to mere wave function based quantum mechanical methods, the key quantities of DFT are the electron density and the pair density, respectively, which are essential for describing the exchange and correlation effects. The electron density represents a probability density determining the probability of finding one electron with arbitrary spin within a defined volume element $d\mathbf{r}$ and is defined as the integral over all spins and all but one of the spatial coordinates.

$$\rho(\mathbf{r}_1) = N\int\left|\Psi(\mathbf{x}_1 \cdots \mathbf{x}_N)\right|^2 d\mathbf{s}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \qquad \text{eq. 3.1.66}$$

$\rho(\mathbf{r})$ is a non-negative function that vanishes at infinity and integrates to the total number of electrons of the molecular system.

$$\rho(\mathbf{r} \to \infty) = 0 \qquad \text{eq. 3.1.67}$$

$$\int \rho(\mathbf{r}_1) d\mathbf{r}_1 = N \qquad \text{eq. 3.1.68}$$

Important properties of the electron density are the discontinuity of the gradient at nuclear positions resulting in a cusp and the asymptotic exponential decay for large distances from the nuclei. The probability of finding two electrons with spin $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ at the same time within the volumes $d\mathbf{r}_1$ and $d\mathbf{r}_2$ is given by the pair density.

$$\rho_2(\mathbf{x}_1,\mathbf{x}_2) = N(N-1)\int \left|\Psi(\mathbf{x}_1 \cdots \mathbf{x}_N)\right|^2 d\mathbf{x}_3 \cdots d\mathbf{x}_N \qquad \text{eq. 3.1.69}$$

Note that the pair density is also dependent on the spin of the two electrons and thus contains all information about electron correlation. It is now convenient to express the dependency of the pair density on the exchange (Fermi) and coulomb correlation by decomposition into products of independent electron densities and introduction of a correlation factor.

$$\rho_2(\mathbf{x}_1,\mathbf{x}_2) = \rho(\mathbf{x}_1)\cdot\rho(\mathbf{x}_2)\cdot\left[1+f(\mathbf{x}_1,\mathbf{x}_2)\right] \qquad \text{eq. 3.1.70}$$

The conditional probability gives the probability of finding any electron at a certain position if the position of a second electron is determined.

$$\Omega(\mathbf{x}_1,\mathbf{x}_2) = \frac{\rho_2(\mathbf{x}_1,\mathbf{x}_2)}{\rho(\mathbf{x}_1)} \qquad \text{eq. 3.1.71}$$

As it can be seen easily by inserting eq. 3.1.66 and eq. 3.1.69 the conditional probability integrates to N-1 electrons excluding the reference electron at $\mathbf{x}_1$. The exchange-correlation hole is defined by the difference between the conditional probability and the uncorrelated electron density of an electron at $\mathbf{x}_2$ describing a depletion of the electron density in vicinity of the reference electron.

$$h_{XC}(\mathbf{x}_1,\mathbf{x}_2) = \frac{\rho_2(\mathbf{x}_1,\mathbf{x}_2)}{\rho(\mathbf{x}_1)} - \rho(\mathbf{x}_2) = \rho(\mathbf{x}_2)f(\mathbf{x}_1,\mathbf{x}_2) \qquad \text{eq. 3.1.72}$$

The integration of the exchange-correlation hole gives -1 being exactly the charge of one electron. With these concepts of densities and probabilities at hand one can now write the expectation value of the electron-electron repulsion term in the Hamiltonian by means of the pair density representing the interaction of two electrons weighted by the probability of the interelectronic distance or in terms of the exchange-correlation hole (cf. eq. 3.1.72).

$$\begin{aligned} E_{ee} &= \langle \Psi | \sum_{i<j} \frac{1}{r_{ij}} | \Psi \rangle = \frac{1}{2}\int \frac{\rho_2(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}} \cdot d\mathbf{r}_1 d\mathbf{r}_2 \\ &= \frac{1}{2}\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} \cdot d\mathbf{r}_1 d\mathbf{r}_2 + \frac{1}{2}\int \frac{\rho(\mathbf{r}_1)h_{XC}(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}} \cdot d\mathbf{r}_1 d\mathbf{r}_2 \end{aligned} \qquad \text{eq. 3.1.73}$$

The first term represents the classical electrostatic interaction energy of two charge distributions including the unphysical self-interaction, whereas the latter gives the interaction energy of a charge distribution with the exchange-correlation hole correcting for the self-interaction and correlation effects. The exchange-correlation hole can be separated into a exchange or Fermi part and a coulomb part which are associated with the Hartree-Fock definition of correlation, *i.e.* the Fermi hole accounts for the depletion in electron density due to the Pauli principle applying to electrons with the same spin, whereas the coulomb hole is due to electrostatic repulsion of two electrons (see **Figure 7**).

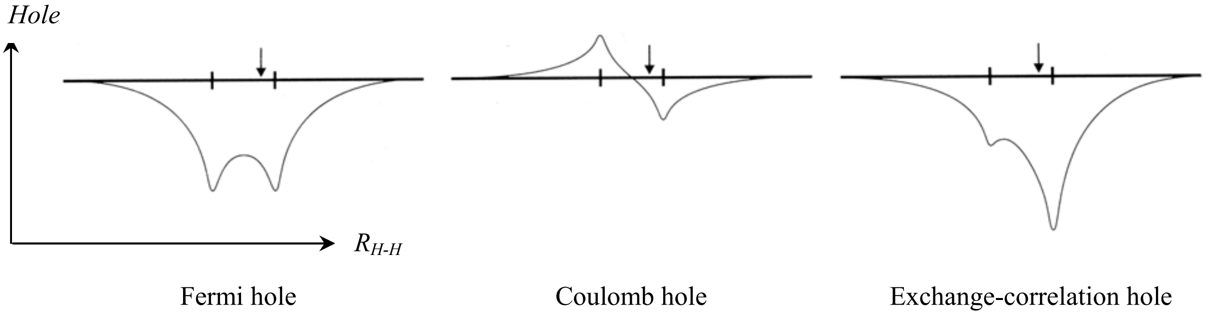


Figure 7 Fermi, coulomb and exchange-correlation hole for H_2 at the nuclear distance of 0.7 Å. The position of the reference electron is indicated by an arrow [taken from Ref. 28].

The Hohenberg-Kohn theorems

The basis for DFT was set by Hohenberg and Kohn in 1964 by proving that the ground-state energy and other properties of a system are uniquely defined by the electron density (first Hohenberg-Kohn theorem). [31] Therefore, *all* expectation values and as a consequence *all* observables are functionals of the ground state density of any molecular system. The electronic energy can thus be written as

$$\begin{aligned} E[\rho(\mathbf{r})] &= \langle \Psi[\rho(\mathbf{r})] | \mathbf{H} | \Psi[\rho(\mathbf{r})] \rangle \\ &= T[\rho(\mathbf{r})] + E_{ee}[\rho(\mathbf{r})] + \int \rho(\mathbf{r}) V_{ne} d\mathbf{r} \\ &= F_{HK}[\rho(\mathbf{r})] + \int \rho(\mathbf{r}) V_{ne} d\mathbf{r} \end{aligned} \qquad \text{eq. 3.1.74}$$

The kinetic energy of the electrons and the electron-electron interaction can be summarized to the Hohenberg-Kohn functional F_{HK} since they are universally valid for any molecular system in the sense that their form is independent of the positions and charges of the nuclei. The problem that has to be tackled is to find good approximations for $T[\rho(\mathbf{r})]$ and h_{XC} (see eq. 3.1.73) as no explicit forms for both of these functionals are known.

The second Hohenberg-Kohn theorem states that the energy of the true ground state electron density is a lower bound for any trial density. In other words, the variational principle ensures that the lowest total energy corresponds to the exact ground state electron density.

$$E_0[\rho] \leq E[\tilde{\rho}] \qquad \text{eq. 3.1.75}$$

The Kohn-Sham equations

The question that inevitably poses is how do we know the correct form for the Hohenberg-Kohn functional and how can we get the exact ground state electron density. In 1965, Kohn and Sham introduced their formulation and the foundation was laid for all modern DFT approaches.[32]

To understand the approach one has to get familiar with the concept of a "non-interacting reference system". Hereby, the Hamiltonian does not contain any electron-electron interaction and the electrons are moving within an effective, local potential $V_S(\mathbf{r})$.

$$\mathbf{H}_S = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N}V_S(\mathbf{r}) \qquad \text{eq. 3.1.76}$$

In complete analogy to HF theory (cf. eq. 3.1.22) the ground state wave function is represented by a Slater determinant consisting of Kohn-Sham (KS) spin orbitals that are determined by the Kohn-Sham equations of this auxiliary non-interacting system.

$$\mathbf{f}_i^{KS}\left|\phi_i^{KS}\right\rangle = \varepsilon_i^{KS}\left|\phi_i^{KS}\right\rangle \qquad \text{eq. 3.1.77}$$

The one-electron Kohn-Sham operator is defined as

$$\mathbf{f}_i^{KS} = -\frac{1}{2}\nabla_i^2 + V_S(\mathbf{r}) \qquad \text{eq. 3.1.78}$$

The link from this artificial system to the real system under investigation is to choose the effective potential V_S such that the electron density of the original many-body system is reproduced by summation over the squared moduli of the KS orbitals (the subscript S denotes the non-interacting reference system).

$$\rho(\mathbf{r}) \overset{def}{=} \rho_S(\mathbf{r}) = \sum_i^N \left|\phi_i^{KS}\right|^2 \qquad \text{eq. 3.1.79}$$

Early approximations to calculate the ground-state energy showed that the calculation of the kinetic energy of the electrons cannot be described by a simple functional form. The main idea of Kohn and Sham was to separate the functional F_{HK} into parts that can be well described and into a part that contains all of the unknown rest. The kinetic energy of the non-interacting reference system is given by

$$T_{\mathrm{S}} = -\frac{1}{2}\sum_{i=1}^{N}\left\langle \phi_i^{KS}\left|\nabla_i^2\right|\phi_i^{KS}\right\rangle \qquad \text{eq. 3.1.80}$$

The true kinetic energy of the electrons can be expressed as a sum of the kinetic energy of the non-interacting reference system and a correction term.

$$T = T_{\mathrm{S}} + T_C \qquad \text{eq. 3.1.81}$$

Furthermore, as eq. 3.1.73 already indicates, the electronic interaction energy consists of a term J representing the classical coulomb interaction of two electron densities and a part E_{ncl} that is due to the non-classical electrostatic contributions.

$$E_{ee} = J[\rho] + \frac{1}{2}\int \frac{\rho(\mathbf{r}_1)h_{XC}(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}} \cdot d\mathbf{r}_1 d\mathbf{r}_2 = J[\rho] + E_{ncl}[\rho] \qquad \text{eq. 3.1.82}$$

The total energy can now be written as

$$\begin{aligned} E[\rho] &= T_S[\rho] + J[\rho] + E_{ne}[\rho] + T_C[\rho] + E_{ncl}[\rho] \\ &= T_S[\rho] + J[\rho] + E_{ne}[\rho] + E_{XC}[\rho] \end{aligned} \qquad \text{eq. 3.1.83}$$

The first three terms are given by their well-known expressions whereas for the exchange-correlation energy term E_{XC} no explicit form is known. It is noteworthy that E_{XC} consists not only of an energy correction due to the exchange-correlation hole h_{XC}, but it also includes the correction for the true kinetic energy as well.

With these definitions at hand, the effective local potential can be derived under the orthonormality constraint δ_{ij} as

$$V_S(\mathbf{r}) = \left(\int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 - \sum_A \frac{Z_A}{r_{1A}} \right) + V_{XC}(\mathbf{r}_1) \qquad \text{eq. 3.1.84}$$

The potential V_{XC} is simply defined as the functional derivative of E_{XC} with respect to ρ.

$$V_{XC}(\mathbf{r}_1) \equiv \frac{\delta E_{XC}}{\delta \rho} \qquad \text{eq. 3.1.85}$$

Once a good approximation for the exchange-correlation energy is available, the KS orbitals, the ground state density and finally the ground state energy can be calculated iteratively by a self-consistent field procedure, since the effective local potential itself is dependent on the electron density.

From early attempts to gradient corrected approximations

Already in 1927 Thomas and Fermi developed a simple expression for the kinetic energy based on the uniform electron gas that gives in combination with the classical expression for the electron-electron and nuclear-electron potentials the Thomas-Fermi approximation for the total energy of an isolated atom.

$$T_{TF}[\rho] = \frac{3}{10}(3\pi^2)^{2/3} \int \rho^{5/3}(\mathbf{r}) d\mathbf{r} \qquad \text{eq. 3.1.86}$$

However, exchange and correlation effects are completely neglected and the kinetic energy is only very rough approximated. Therefore, the Thomas and Fermi model was only of very limited use. In 1951 Slater derived an explicit expression for the exchange energy by approximating the Fermi hole.[33] The approach also known as X_α or Hartree-Fock-Slater method became quite popular in physics but could not prevail in chemistry. Even more sophisticated models like the Thomas-Fermi-Dirac model describe the kinetic energy too poorly so that bonding in molecules does not occur. This problem was solved by the adiabatic connection of the non-interacting reference and the real system resulting in the “Local Density Approximation” (LDA) treating the electron density as a uniform electron gas. The exchange energy of a uniform electron gas is given by the Dirac formula and is identical to the X_α method if α is chosen as 2/3.

$$E_X^{LDA}[\rho] = -C_X \int \rho^{4/3}(\mathbf{r})d\mathbf{r} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \int \rho^{4/3}(\mathbf{r})d\mathbf{r} \qquad \text{eq. 3.1.87}$$

The correlation energy of a uniform electron density has been determined by Quantum Monte-Carlo simulations for various densities and was brought into analytical form *e.g.* by Vosko, Wilk and Nusair (VWN).[34] The LDA approach has been widely used in solid state physics, but the accuracy required for quantum chemical calculations was still unsatisfying. An improvement over the local density approximation was made by considering not only the electron density but also its derivatives in order to account for the non-uniform character of the electron density in molecules. The so-called “Generalized Gradient Approximations” (GGA) often achieve MP2-like accuracy for ground state energies and geometries in gas phase (see chapter 3.1.4). For the exchange expression the functionals by Perdew and Wang (PW86) and by Becke (B88) enjoy great popularity, whereas the gradient approximations for the correlation energy by Perdew (P86), Perdew and Wang (PW91) or Lee, Yang and Parr (LYP) are common.[35] The explicit forms of these functionals comprise complex formulas with various parameters that are fitted to experimental data or that are constructed by fundamental considerations of the properties a correct wave function should have. It is therefore a matter of opinion whether to consider GGA DFT an *ab initio* or semi-empirical approach. However, unlike for post-HF methods a systematic improvement of the level of theory is not possible since the exact form of the exchange and correlation functionals remains unknown and the real energy of a system does not consequently present a lower bound to the DFT calculated energies.

Beside the pure GGA based functionals especially the hybrid functionals have gained large influence in computational chemistry by mixing the exact exchange energy that can be calculated analogously to the HF scheme (eq. 3.1.88) with the GGA exchange expression.

$$E_X^{exact} = \frac{1}{2}\int \frac{\rho_0^{KS}(\mathbf{r}_1)h_X^{KS}(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \qquad \text{eq. 3.1.88}$$

However, a full replacement of the exchange energy expression would lead to a deterioration of the total energy since the definitions of both expressions are footed on different assumptions. The exact Fermi hole and the exact coulomb hole are both highly non-local quantities that add to result in the relative localized total exchange-correlation hole. However, in the LDA and GGA ansatz the energy expressions are derived for local densities assuming that both holes are localized since they depend solely on the density and its gradient at a particular point in space. A solution to this dilemma was provided by Becke in 1993 by expressing the exchange-correlation energy of the Kohn-Sham scheme through the coupling-strength integrated correlation hole as defined by the adiabatic connection method.

$$E_{XC} = \frac{1}{2}\int \frac{\rho(\mathbf{r}_1)\bar{h}_{XC}(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \qquad \text{eq. 3.1.89}$$

with the coupling-strength integrated correlation hole given as

$$\bar{h}_{XC}(\mathbf{r}_1,\mathbf{r}_2) \equiv \int_0^1 h_{XC}^{\lambda}(\mathbf{r}_1,\mathbf{r}_2) d\lambda \qquad \text{eq. 3.1.90}$$

A value of $\lambda = 0$ would indicate that all electrons are non-interacting as defined for the reference system and thus no correlation energy occurs. In this case, the exchange energy is given exactly by Hartree-Fock theory employing the KS-orbitals (cf. eq. 3.1.88). A value of $\lambda = 1$ includes also electron correlation and the energy can be approximated by an E_{XC} functional. The exact exchange-correlation energy would be given by a full integration of the integral given in eq. 3.1.89, however, since this turns out to be unfeasible, approximations have been generated of which the Becke-3-parameter functional in conjunction with the LYP correlation functional (B3LYP) soon became the most popular hybrid functional.[36]

$$E_{XC}^{B3LYP} = \left[(1-a)E_X^{LDA} + aE_X^{exact} + bE_X^{B88}\right] + \left[cE_C^{LYP} + (1-c)E_C^{LDA}\right] \qquad \text{eq. 3.1.91}$$

The empirical parameters were fitted to the G2 data set for atomization and ionization energies as well as for the proton affinities ($a = 0.20$, $b = 0.72$, $c = 0.81$).

3.1.4 Drawbacks and benefits of the electronic structure methods

The accuracy of the previously mentioned electronic structure methods is not only dependent on the basis set size or (in case of the wave function based electronic correlation methods) on the degree of truncation, but it is also restricted by the inherent methodology of the approach. Important issues concerning the applicability and performance are therefore discussed here briefly.

Size consistency and size extensity

Two of the issues that have to be addressed when discussing electronic structure theory are *size consistency* and *size extensivity*. On the one hand, size-consistency in the definition of Pople describes the additive separability of the wave function. Regarding for instance two molecules that are separated by an infinite distance, the energy of this non-interacting system should be in general the sum of the separately calculated energies of the two molecules. A stricter definition of size consistency states that the method should not only correctly describe the system in the non-interacting limit, but it should also be able to include the entire potential hyper surface of any dissociation process. Restricted Hartree-Fock or Kohn-Sham wave functions are size consistent for closed-shell fragmentation, however, when dissociating into open-shell fragments (*e.g.* H_2 to 2 H), the second definition of size-consistency is not fulfilled as the open-shell fragments can only be treated by unrestricted methods. Perturbative theory like MP2 and coupled cluster methods are both size consistent, whereas for configuration interaction this is only valid for full CI calculations. CISD as example for a truncated CI wave function neglects quadruply excited states at the non-interacting limit and thus does not describe the simultaneous double excitations of both separated systems.

Size extensivity on the other hand is a more mathematical formulation implying that the method scales linearly with the number of electrons. However, size extensivity does not ensure size consistency for dissociation processes, but at the non-interacting limit size extensivity also includes size consistency. All Hartree-Fock, DFT, perturbation theory and coupled cluster methods are size-extensive, whereas truncated CI methods are not.

Dispersion effects in DFT

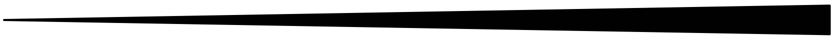
One major drawback of density functional theory is the poor description in current functionals of weak interactions due to dispersion effects. Unfortunately, these effects can play an important role in many organic biomolecules with π-π stacking interactions with the DNA being probably the most prominent example. It is therefore often necessary to validate DFT example calculations with respect to MP2 or coupled-cluster results in order to give estimates about the magnitude of error. As already mentioned in chapter 3.1.3 the popular B3LYP functional includes parameters that were fitted to the G2 test set. “Unusual” molecular structures that are not considered in this data base can cause significant error bars since the G2 test set consists of only 148 small and uncharged molecules.

Performance comparison

The formal scaling behaviour of important electronic structure methods in comparison to a rough quality ordering is given in **Table 2**. As one can see the effort increases non-linearly with respect to the basis set size for all methods and the very accurate CCSD(T) method already scales to the power of 7 with the number of basis functions making this approach only applicable to small compounds with a moderate basis set size. However, in order to reduce the formal scaling behaviour, techniques have been developed that are based on simplifying the calculation of the expensive two-electron-four-centre integrals. Popular methods are the resolution of identity (RI) formalism or the fast multipole moment method that are able to reach often linear scaling for HF, DFT or MP2 calculations.

The accuracy of DFT calculations strongly depends on the method and the application field. In general, GGA functionals outperform LDA functionals for non-periodic molecular systems. The hybrid functionals (*e.g.* B3LYP) usually offer some improvement over pure GGA functionals performing often better than MP2. However, absolute errors are always larger indicating that DFT methods sometimes fail for peculiar molecular systems and that one has to be aware of the interpretability of the results.

Table 2 The scaling behaviour as a function of basis functions N vs. accuracy. The average errors are given for the correlation energies compared to full CI applied to HB, H_2O and HF at equilibrium geometries (kcal mol^{-1}) and a polarized double ξ basis set.

Method	HF	MP2	CISD	CCSD	CCSD(T)	DFT
Scaling	N^4	N^5	N^6	N^6	N^7	N^4
Accuracy						n/a
Average. error	n/a	10.4	5.8	1.9	0.3	n/a

3.2 Conformational analysis

The determination of electronic binding energies of non-covalent assemblies requires the knowledge about the global minimum structures of the monomeric units as well as of their aggregates. Therefore, an extensive search strategy for finding the lowest lying minimum structure is indispensable. However, due to the increasing complexity associated with the system size a reasonable sampling of the conformational space is only feasible by computer-aided search algorithms. Moreover, these algorithms depend on fast computations of the potential energies making most quantum mechanical approaches unemployable (an exception could be semi-empirical methods). Hence, nearly all conformational searches use force-field potential energies to determine low energy conformers within a pre-defined energy window. Chapter 3.2.1 gives a very short overview over the calculation of potential energies via molecular mechanics. Afterwards, some of the most common conformational search algorithms are presented (chapter 3.2.2). An advanced discussion on the interpretability of force-field calculations can be found in most molecular modeling textbooks.[37]

3.2.1 Brief introduction to molecular mechanics

The molecular mechanics methods are based on a classical treatment of interaction potentials between atom types consisting of single atoms or atom groups ("united atoms"). These atom types are classified according to their characteristic binding patterns and each of them is described by a unique set of parameters. The force-field based interaction potentials characterize the geometry and ground-state energy of a molecule on the basis of analytical functions for the various potentials contributing to the potential energy. The electronic structure of the molecule is not taken into account explicitly and is only considered indirectly by the parameterization of the force field. The parameters are fitted either to experimental or theoretical data and the fitting method determines the quality and application field of each force field.

The force field potential energy is composed of multiple terms that describe bonded interactions, non-bonded interactions and terms that represent cross-terms of coupled movements. The analytical forms of the bonded interactions are represented by truncated Taylor-expansions around the equilibrium (stretch energy, bending energy) or by a truncated Fourier series (torsional energy). However, the computationally most demanding parts in a force field calculation are the non-bonded interactions consisting of the weak short-range interactions and the long-range electrostatic attractions/repulsions. The electrostatic interactions are given by the Coulomb potential for point-charges and have a large influence on the potential energy. The weak interaction is often spuriously equated with the van-der-Waals energy; however, it is actually composed of two contrary effects. The repulsive part dominates at very small distances and is due to the Pauli repulsion of two electrons with the same spin, whereas the attractive part characterizes interactions between transient multipoles in molecules without permanent multipole moments. These attractive forces are called either dispersion forces or van-der-Waals forces. The weak interaction potential combines both the attractive and the repulsive interaction and is represented in general by a 12-6- Lennard-Jones potential, but also other analytical expressions like the Buckingham potential are possible.

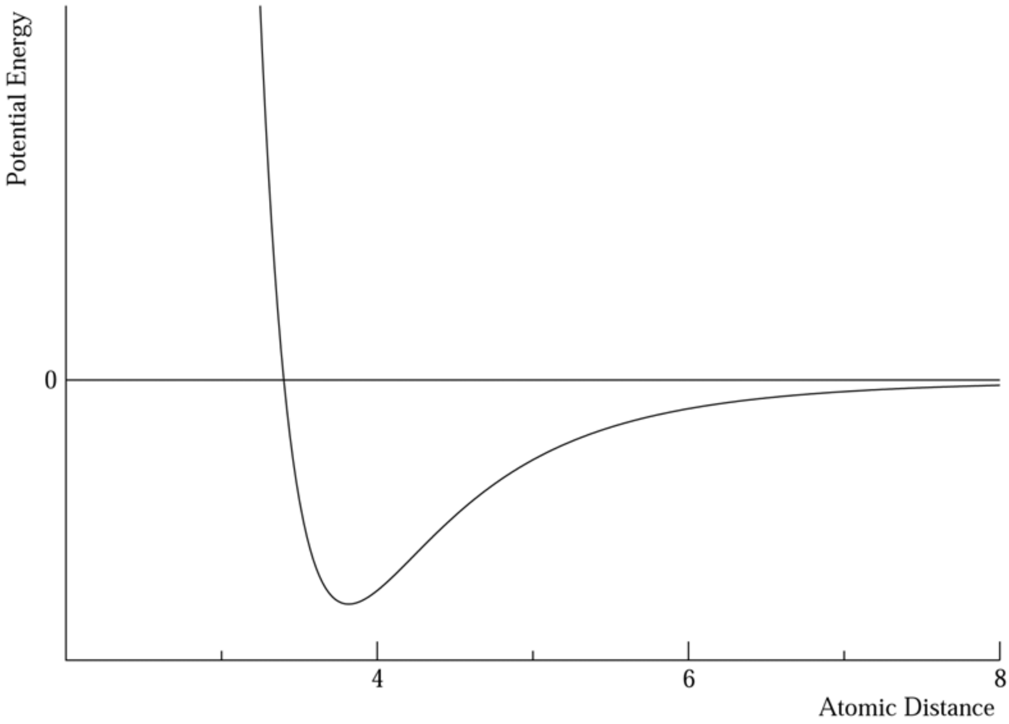


Figure 8 Lennard-Jones Potential: $U(r_{AB}) = 4 \cdot \varepsilon \cdot \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^{6} \right]$

3.2.2 Conformational search algorithms

For the conformational search problem a variety of different methods are available which are classified accordingly (see reference 37, Chapter 9):

- systematic search algorithms
- random search algorithms
- eigenvector following approaches
- molecular dynamics based conformational search
- distance geometry
- global optimization methods (genetic algorithms, simulated annealing, tabu search)

The global optimization methods can be applied on a wide range of multi-dimensional optimization tasks with the conformational search being only one of them. However, properties of real systems are usually averages from multiple low-energy structures and the consideration of a single structure is not sufficient for a quantitative modelling. The first three methods represent the most widespread approaches which are applied either purely or in a combined implementation. One problem that arises for all algorithms is the handling of cyclic molecules. Typically, ring structures are broken into acyclic structures that are subjected to ring closure constraints.

Systematic unbounded multiple minimum approach (SUMM)

The SUMM method represents a variation of a standard systematic search algorithm that overcomes two main difficulties of the depth-first strategy.[38] This search type starts at a terminal node (*e.g.* node 13 in **Figure 9**) of a search tree which represents a conformational state with all torsional angles set to a defined value. By variation of a single torsional angle the search algorithm explores the conformational space by backtracking the search path and moving to the next terminal node. The limitations associated with this strategy are that the generated conformers are very similar to their precursors and one has to scan the whole search tree to cover all relevant structures. Second, an increase of the resolution, *i.e.* the torsional angle increments, implies that all conformations that have already been scanned before are reproduced. The SUMM approach now uses a breadth-first strategy that begins at a coarse torsion angle resolution (120°) and uses low energy structures at each step that have been generated previously. A systematic variation of the torsional parameters generates highly diverse geometries and the search tree is scanned non-linearly. In case of the hexane example (see **Figure 9**), the SUMM algorithm would go from node 13 over node 22 to node 31 which covers the all branches of the root node. By recording all torsional variations that have been made on each minimum conformation, the search path will not be retraced. An increase in the torsional resolution thus results in previously undiscovered minimum conformers.

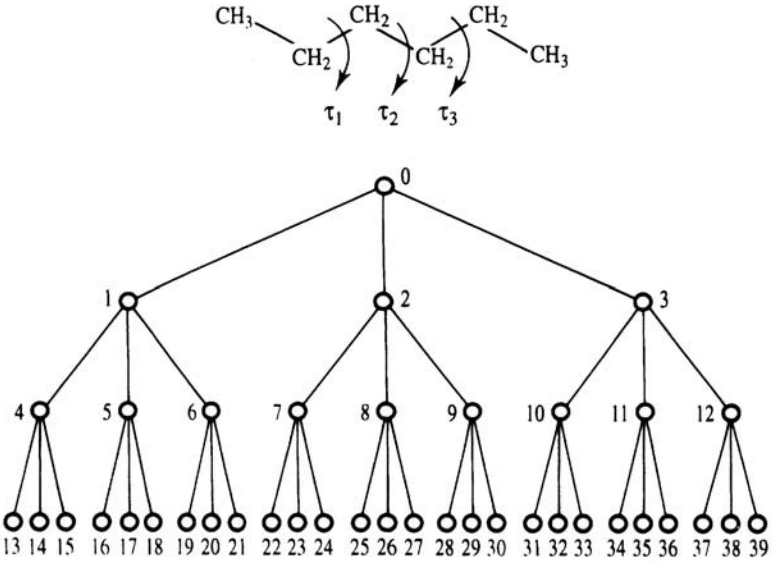


Figure 9 Search tree of the conformational search problem of hexane. Terminal nodes represent conformational states of the molecule whereas branches indicate a variation of the torsional angle (here: in 120° steps) of one single bond.

Monte-Carlo multiple minimum approach (MCMM)

In contrast to the systematic search strategy of the SUMM approach, the Monte-Carlo method as introduced by Chang et al. searches the conformational space on a random walk by making completely unpredictable modifications of the torsional angles.[39] The critical points in the algorithm are again the treatment of molecules containing rings and the proper choice of new input structures for the next search step. In order to steer the search into so far unvisited regions of the conformational space, the uniform usage protocol selects randomly starting structures from a pool of already generated conformers and weights the selection in favour of the least selected ones. The algorithm of the MCMM approach is given by the flow diagram in **Figure 10**.

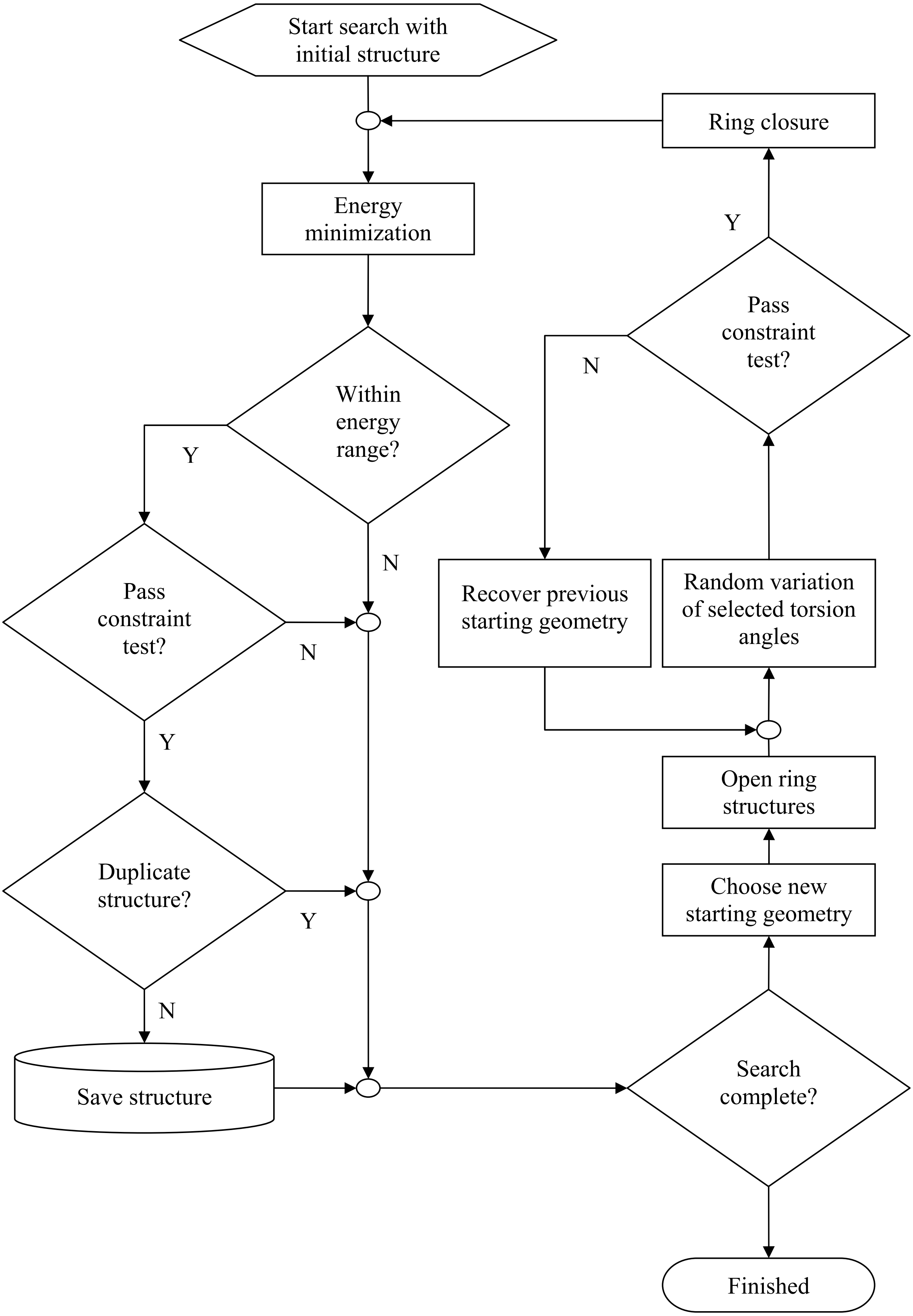

Start search with initial structure
Ring closure
Energy minimization
Y
Pass constraint test?
N
Within energy range?
Y
N
Pass constraint test?
N
Recover previous starting geometry
Random variation of selected torsion angles
Y
Open ring structures
Duplicate structure?
Y
Choose new starting geometry
N
Save structure
Search complete?
Finished

Figure 10 Flow diagram of the Monte-Carlo multiple minimum algorithm.

Low- mode approach

The low-mode approach is based on an eigenvector following concept that performs a normal mode analysis of the initial structure.[40] The low frequency modes (< 250 cm^{-1}) are stored in a non-mass-weighted Hessian matrix and the corresponding eigenvectors are searched systematically in both directions since they are considered to represent the torsional degrees of freedom. For this purpose the structure is perturbed in discrete steps along the eigenvectors until the potential energy exceeds a certain threshold during a single step. A subsequent energy minimization crosses a potential energy barrier in most cases and leads to new structures that are related to the initial structure. This conformational search approach therefore scans only the local neighbourhood of a minimum on the potential energy surface. The low-mode search never updates the eigenvector search direction which means that no re-evaluation of the Hessian matrix is performed.

In the case of the search directions being exhausted, the pure low-mode search can switch to a stochastic determination of the search direction by randomly mixing the eigenvectors. This procedure called mixed Monte-Carlo/low-mode approach guarantees the thorough scan of the potential hyper surface.

3.3 From gas-phase to solution

The goal in theoretical organic chemistry is the description of physicochemical properties (local or macroscopic) and the prediction of reaction rates and mechanisms. Unfortunately, almost every reaction takes place in a condensed phase implicating effects on the inherent molecular properties. Calculations in gas-phase are suitable to only a limited extent to model the real behaviour of any solute unless the solvation has only little effect on the particular property being studied.

A proper description of a solute in a condensed phase poses a nontrivial problem which lead to a variety of solvent models, each of them designed for different purposes. In the field of physics the emphasis was placed on creating various simplified expressions for the intermolecular interaction potentials for solvents footed solely on classical mechanics.[41] In theoretical chemistry, however, the focus is on the electronic structure of a solvated molecule which requires a quantum mechanical treatment. Besides the so-called "supermolecule" ansatz, which tries to mimic the solvation shell by considering explicit solvent molecules (see chapter 3.3.2), the electrostatic continuum methods have become a popular tool to estimate solvation effects in the recent past (chapter 3.3.1).

The process of bringing a solute from gas-phase to solution incorporates a variety of effects contributing to the free energy of solvation ΔG_S^0. In principle, it can be seen as interplay between the orientation and the polarization of the solvent molecules generating a reaction field on the one hand, and the electronic polarization of the solute itself on the other hand. The interaction energy between the solute and the solvent is in a balance with the energy costs of the solvation process. Contributions to the energy costs are from the self-interaction of the solvent molecules and the reaction field, the lost of configurational freedom, the costs to achieve polarization and the energy necessary to create a cavity in the condensed phase. It should be noted that the majority of the total cohesion energy (70-90%) of liquids derives from dispersion interactions. A reasonable approximation of the solvent effects on a solute thus should be based on an effective Hamiltonian method due to the non-classical origin of the dispersion energy. The free energy of solvation ΔG_S^0 consists formally of three components, whereas the van der Waals term can be split into a repulsive and dispersion part (see chapter 3.2.1).

$$\Delta G_S^O = \Delta G_{electric}^O + \Delta G_{vdW}^O + \Delta G_{cavity}^O = \Delta G_{electric}^O + \Delta G_{repulsive}^O + \Delta G_{dispersion}^O + \Delta G_{cavity}^O \qquad \text{eq. 3.3.1}$$

Experimentally it is not feasible to separate the various contributions to the free energy of solvation ΔG_S^0 and the continuum approaches being introduced in chapter 3.3.1 are only capable to make predictions for the electrostatic part of the solvation energy. Therefore, the results can only provide information about the relative energies of a particular system.

3.3.1 Continuum solvation

In the following part of this work the focus is on the quantum mechanical description of a solute M, whereas classical continuum models describe the solute only as a polarizable charge distribution.[42] The solvent molecules are supposed to be homogeneously distributed and can therefore be assumed in a good approximation to form a polarizable continuum as defined for dielelectric materials in electric fields.

By placing the solute into a cavity generated in the continuum solvent, the solute can now be described by an effective Hamiltonian consisting of the unperturbed molecular Hamiltonian (see eq. 3.1.2) and the electrostatic interaction potential of the solvent.

$$\mathbf{H}_M = \mathbf{H}_M^{(0)} + \mathbf{V}_\sigma \qquad \text{eq. 3.3.2}$$

The Schrödinger equation thus gives

$$\mathbf{H}_M \Psi = E\Psi \qquad \text{eq. 3.3.3}$$

and the solute-solvent interaction energy can be written by introducing the electrostatic field potential Φ_σ and the charge density of the solute M (nuclei and electrons) ρ_M.

$$W_{MS} = \int_{\text{all space}} \Psi^* \mathbf{V}_\sigma \Psi d\mathbf{r}_1 \ldots d\mathbf{r}_N, d\mathbf{R}_1 \cdots d\mathbf{R}_M = \int_{\text{all space}} \rho_M(\mathbf{r}) \Phi_\sigma(\mathbf{r}) d\mathbf{r}^3 \qquad \text{eq. 3.3.4}$$

The electrostatic field potential produced by the solvent is depending on the charge density of the solute and *vice versa*. The solution of this non-linear problem requires an iterative procedure until self-consistency is reached. Quantum approaches to determine the interaction energy between Φ_σ and ρ_M are amongst others multipole expansion (MPE) and apparent surface charge (ASC) methods of which the latter are of most importance.

The choice of the cavity shape can also have great influence on the quality of the employed method. Regular shapes like spheres or ellipsoids for the entire molecule are only very crude approximations and therefore molecular shapes consisting of overlapping atom-centred spheres or isodensity surfaces are mostly used. The radius of the spheres are slightly enlarged van-der-Waals radii with a factor of about 1.2 - 1.25. An inclusion of space within a solute that is not accessible by the solvent is considered as vacuum. For charged solutes (ions, zwitterions and ion pairs) the factor may need to be adopted due to non-linear effects in the polarization of the dielectric in high fields. It should be noted that dispersion, repulsion and cavitation terms require cavities of different size than the electrostatic terms.

The boundary conditions for the continuum solvation models are given in **Figure 11**. Inside the cavity the well-known Poisson's equation is fulfilled, whereas in the dielectric continuum no charge density of the solute occurs and thus changing the equation for the potential into Laplace's equation. At the boundary the electrostatic field potentials of inside and outside the cavity are equal, for the first derivative, however, the dielectric constant needs to be taken into account.

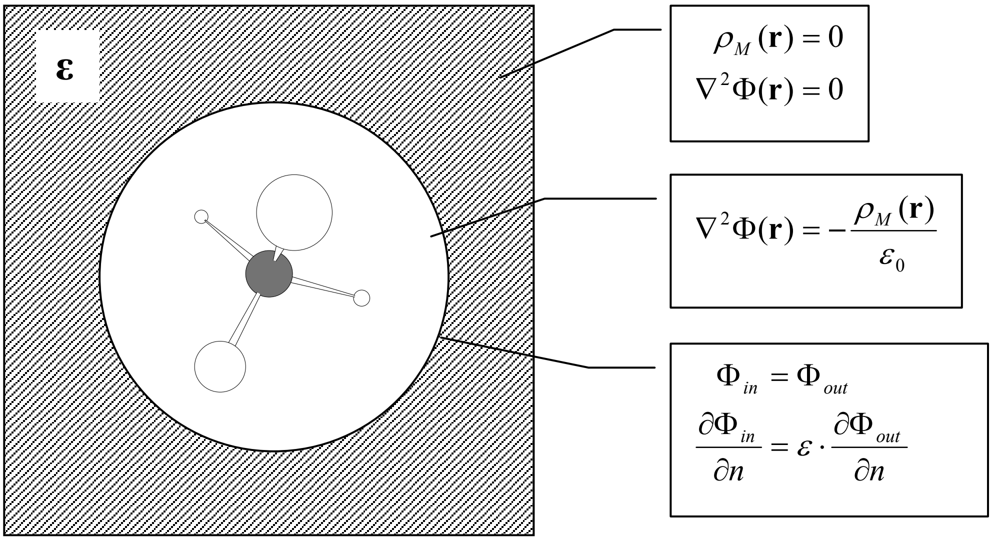

ε
$\rho_M(\mathbf{r}) = 0$
$\nabla^2\Phi(\mathbf{r}) = 0$
$\nabla^2\Phi(\mathbf{r}) = -\frac{\rho_M(\mathbf{r})}{\varepsilon_0}$
$\Phi_{in} = \Phi_{out}$
$\frac{\partial\Phi_{in}}{\partial n} = \varepsilon \cdot \frac{\partial\Phi_{out}}{\partial n}$

Figure 11 Boundary conditions of the continuum solvation model (in SI units)

3.3.1.1 The Apparent Surface Charge Method

The apparent surface charge method is based on the induction of small grid charges on the surface of the cavity. The surface polarization charge can be described in terms of an induced polarization vector projected on the outward normal of the surface segment.

$$\sigma = -\vec{\mathbf{P}} \cdot \vec{\mathbf{n}} \qquad \text{eq. 3.3.5}$$

The polarization vector itself is directly proportional to the electric field arising from the solute and acting on the continuum surface.

$$\vec{\mathbf{P}} = -\frac{\varepsilon - 1}{4\pi} \cdot \vec{\mathbf{E}} \cdot \vec{\mathbf{n}} = -\frac{\varepsilon - 1}{4\pi} \cdot \vec{\nabla}\Phi_{out} \cdot \vec{\mathbf{n}} \qquad \text{eq. 3.3.6}$$

Applying the boundary conditions and abbreviating the constant pre-factor one can write the polarization vector solely depending on the electric field potential inside of the cavity.

$$\vec{\mathbf{P}} = -\frac{\varepsilon - 1}{4\pi} \cdot \vec{\nabla}\Phi_{out} \cdot \vec{\mathbf{n}} = -\frac{\varepsilon - 1}{4\pi\varepsilon} \cdot \vec{\nabla}\Phi_{in} \cdot \vec{\mathbf{n}} = \kappa \cdot \vec{\nabla}\Phi_{in} \cdot \vec{\mathbf{n}} = \kappa \cdot \frac{\partial \Phi_{in}}{\partial \vec{\mathbf{n}}} \qquad \text{eq. 3.3.7}$$

The electrostatic potential can now be separated into one part arising from the solute and one part reducing the potential due to the induced screening charges (self-interaction).

$$\Phi_{in} = \Phi_M - \Phi_\sigma \qquad \text{eq. 3.3.8}$$

The electrostatic field potential can be derived from integrating the surface polarization charge over the cavity surface area *A* with respect to the vector **s** defining a point on that surface.

$$\Phi_\sigma(\mathbf{r}) = \int_A \frac{\sigma(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|} d^2\mathbf{s} \qquad \text{eq. 3.3.9}$$

The surface polarization charge is therefore dependent on its own electrostatic field potential.

$$\sigma = \kappa \cdot \frac{\partial\left(\Phi_{M,in} - \Phi_{\sigma,in}\right)}{\partial \vec{\mathbf{n}}} \qquad \text{eq. 3.3.10}$$

As the integration over a complex surface (see eq. 3.3.9) is very complicated the polarizable continuum models (PCM) use the boundary element method (BEM) that partitions the cavity surface into small elements A_k called tesserae. The electrostatic field potential can then be expressed by a sum over all apparent point charges q_k that are placed in the tesserae k at point $\mathbf{s}_k$.

$$\Phi_{\sigma}(\mathbf{r}) = \sum_k \frac{q_k}{\left|\mathbf{r} - \mathbf{s}_k\right|} \qquad \text{eq. 3.3.11}$$

with $q_k = A_k \cdot \sigma(\mathbf{s}_k)$ eq. 3.3.12

The solution of the electrostatic problem is done quantum-mechanically either by an iterative solution within a SCF cycle with a fixed potential of the solute, or by an implementation into the Fock equation.

3.3.1.2 The Conductor-like Screening Model

The conductor-like screening model (COSMO) is related to the PCM approach and is therefore sometimes denoted as conductor-like polarizable continuum model (CPCM).[43,44] The method is based on a simplification of the boundary conditions by using an infinitely strong dielectric like a conductor. As a result the total electrostatic potential vanishes on the cavity surface and by scaling of the ideal screening charge densities σ^* one can reintroduce finite dielectrics with the factor $f(\varepsilon)$.

$$\sigma = f(\varepsilon) \cdot \sigma^* \qquad \text{eq. 3.3.13}$$

with $f(\varepsilon) = \dfrac{\varepsilon - 1}{\varepsilon + x}$ eq. 3.3.14

The scaling factor $f(\varepsilon)$ includes a variable x that in general depends on the cavity and order of the lowest multipole moment of the solvent. However, in practical considerations it is set to a constant value (x = 0.5). For large dielectric constants the error is found to be very small and

at the lowest limit (ε = 2) one obtains a relative error that is within 10 %. The advantages, however, are lower computational costs and at the same time a greater numerical stability.
The implementation of the COSMO equations is straightforward into any SCF code, whether in HF or KS theory. The boundary conditions simplify the electrostatic problem and one can rewrite the electrostatic potential in vector notation (cf. eq. 3.3.8 and eq. 3.3.11) as

$$\mathbf{\Phi}(\mathbf{r}) = \mathbf{\Phi}_M(\mathbf{r}) + \mathbf{\Phi}_\sigma(\mathbf{r}) = \mathbf{\Phi}_M(\mathbf{r}) + \mathbf{A}\mathbf{q}^* = 0 \qquad \text{eq. 3.3.15}$$

The $m \times m$- matrix **A** represents a coulomb interaction matrix and the vector $\mathbf{q}^*$ includes all screening charges appearing on the m segments of the surface cavity. The ideal screening charges are therefore given by

$$\mathbf{q}^* = -\mathbf{A}^{-1} \cdot \mathbf{\Phi}_M \qquad \text{eq. 3.3.16}$$

resulting in a explicit expression for the dielectric screening charges depending now only on the electrostatic field potential of the solute.

$$\mathbf{q} = f(\varepsilon) \cdot \mathbf{q}^* = -f(\varepsilon) \cdot \mathbf{A}^{-1} \cdot \mathbf{\Phi}_M \qquad \text{eq. 3.3.17}$$

The reduction of the total energy due to the screening effect of the dielectric is only half of the interaction energy of the solute with the screening charges since, since according to the linear response theory the other half is needed for solvent polarization.

$$E_{\mathrm{diel}} = \tfrac{1}{2} E_{\mathrm{int}} = \tfrac{1}{2} \mathbf{q}\mathbf{\Phi}_{\mathrm{M}} = -\tfrac{1}{2} f(\varepsilon) \mathbf{\Phi}_{\mathrm{M}} \mathbf{A}^{-1} \mathbf{\Phi}_{\mathrm{M}} \qquad \text{eq. 3.3.18}$$

The electrostatic field potential of the solute can be decomposed into an n-dimensional vector **Q** representing the charge distribution and a matrix **B** defining the density-segment interactions.

$$\mathbf{\Phi}_{\mathrm{M}} = \mathbf{B}\mathbf{Q} \qquad \text{eq. 3.3.19}$$

In quantum approaches the charge distribution is equal to the elements of the density matrix. The energy reduction is thus given as

$$E_{\mathrm{diel}} = -\tfrac{1}{2} f(\varepsilon) \mathbf{Q}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}\mathbf{Q} = \tfrac{1}{2} \mathbf{Q}\mathbf{D}\mathbf{Q} \qquad \text{eq. 3.3.20}$$

with

$$\mathbf{D} = -\tfrac{1}{2} f(\varepsilon)\mathbf{B}\mathbf{A}^{-1}\mathbf{B} \qquad \text{eq. 3.3.21}$$

being the Green function of the cavity.

Finally, the total charge energy is given as

$$E_{\text{charge-charge}} = E_{\text{Coulomb}} - E_{\text{diel}} = \tfrac{1}{2}\mathbf{QCQ} - \tfrac{1}{2}\mathbf{QDQ} = \tfrac{1}{2}\mathbf{Q}(\mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B})\mathbf{Q} \qquad \text{eq. 3.3.22}$$

In a SCF scheme, the first step comprises the generation of all segments and the construction of the segment-segment interaction matrix **A**. Afterwards, the density-segment interaction matrix **B** is build and inverted, so that the Green function **D** can be calculated. This function can be used to modify all Coulomb interaction terms (nuclear energy, one-electron and two-electron terms) of the SCF calculation. The COSMO-SCF calculations require only little time to evaluate the dielectric energy reduction terms and due to the damping of density fluctuations the entire SCF calculation is often accelerated.[45]

3.3.2 Explicit solvation

All generic continuum models show one major disadvantage: they are simplified models which can only estimate the effect of solvation by the electrostatic interaction with the solute as a linear response to the electric field of the continuum solvent. Inhomogenities in the spatial distribution of the solvent molecules or the dynamical behaviour during geometrical changes of the solute are not taken into account, whereas dispersion and cavitation terms are often included semi-empirically by most continuum models. A quantum treatment of an explicitly solvated solute should therefore in general be able to compensate these drawbacks. However, in practice, an explicit solvation footed on a mere quantum mechanical description is not possible due to the following reasons:

1. The exact amount of solvent molecules that are necessary to solvate a solute can not be determined as the interactions between the various solvation shells must be considered too.
2. Edge effects of a small cluster of solvent molecules have a drastic impact on the proper description of the solute.
3. The spatial distribution of the solvent molecules is not known.

All of these problems make an explicit solvation of a target molecule for static quantum mechanical calculations unfeasible and can only be solved by simulation methods based on molecular mechanics, a QM/MM scheme or with the Car-Parrinello MD method.[46]

3.4 Molecular simulations – Accounting for the entropy

Molecular simulation methods like Molecular Dynamics (MD) or Monte-Carlo simulations are necessary when thermodynamic properties of a system (*e.g.* heat capacity, free energy) or the time-dependent behaviour are under investigation.[47] In some cases the global minimum structure determines the properties of a system or it is possible to detect all minimum configurations manually in order to set up the partition function properly. However, this procedure is only applicable for small systems in gas phase. In most cases a full quantitative description of the energy surface is unfeasible due to its enormous amount of closely separated minima and techniques were developed to circumvent this problem. In the following chapter only the basic concept of a MD simulation is described and the Thermodynamic Integration is presented as one approach to account for the entropy by approximating the free energy of a molecular process. Further important technical methods that are not considered here are, for example, periodic boundary conditions, stochastic boundary conditions, cuff-offs, constraints (*e.g.* SHAKE) and thermostats or heat baths (Nosé-Hoover/Berendsen).

3.4.1 Molecular Dynamics

Statistical mechanics

The determination of a thermodynamic property of a system is done experimentally by the measurement over a specified period τ of the property A depending on the momenta and positions of all N particles of the system. The time average is thus defined as

$$\langle A \rangle_{time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^{\tau} A(\mathbf{p}^N(t), \mathbf{r}^N(t))dt \qquad \text{eq. 3.4.1}$$

The ergodic hypothesis of statistical mechanics states that the time average of a given system is equal to an ensemble average. Ensembles in the sense of mathematical physics are collections of all possible systems which have different microscopic states, but which are in an identical macroscopic (thermodynamic) state. In molecular simulations one distinguishes between canonical (NVT), microcanonical (NVE) and isobaric-isothermal (NPT) ensembles, in which the particle number N, the volume V, the temperature T, the global potential energy

E or the pressure p are held constant. The ensemble average defines then the average value of the property A over all microscopic states of an ensemble during a simulation.

$$\langle A \rangle_{ensemble} = \iint A(\mathbf{p}^N, \mathbf{r}^N) \cdot \pi(\mathbf{p}^N, \mathbf{r}^N) d\mathbf{p}^N d\mathbf{r}^N = \int A(\mathbf{\Gamma}^N) \cdot \pi(\mathbf{\Gamma}^N) d\mathbf{\Gamma}^N \qquad \text{eq. 3.4.2}$$

The integration is done over the momenta and positions of all particles with $\pi(\mathbf{\Gamma}^N)$ denoting the probability density of finding the ensemble in a certain state or configuration in the phase space $\mathbf{\Gamma}^N$. For the canonical NVT ensemble the probability density is the familiar Boltzmann distribution

$$\pi(\mathbf{\Gamma}^N) = \frac{\exp\left(-\frac{E(\mathbf{\Gamma}^N)}{k_B T}\right)}{Q} \qquad \text{eq. 3.4.3}$$

with the partition function given as

$$Q = \frac{\int \exp\left(\frac{-E(\mathbf{\Gamma}^N)}{k_B T}\right) d\mathbf{\Gamma}^N}{h^{3N} N!} \qquad \text{eq. 3.4.4}$$

Newton's equation of motion

The movement of a particle (atom) is solved by integration of the differential equations obtained by Newton's second law.

$$\mathbf{F} = m \cdot \frac{d^2\mathbf{r}}{dt^2} = m \cdot \frac{d\mathbf{v}}{dt} \qquad \text{eq. 3.4.5}$$

The equations can be solved analytically if the particle moves constantly (no forces acting on it) or if it experiences only a constant potential. However, due to the coupled motions of the atoms in a molecular system, the molecular dynamics simulation faces a many-body problem and therefore one has to solve the equations of motions by integrating numerically employing finite difference methods.

The numerical integration yields a trajectory separated into small time steps that describes the positions, velocities and acceleration of the particles as they vary with time and from which the average values can be calculated. Initial coordinates for MD simulations of biomolecules are in general obtained by X-ray or NMR studies and the initial velocities are assigned randomly by a Maxwell-Boltzmann distribution for a given temperature. After calculation of the forces acting on each atom based on the force field energy function, the new positions and velocities after an adequate time step are determined by solving a numerical algorithm. Finally, the new forces are recalculated and the last step is repeated for a chosen period to generate the full trajectory (see **Figure 12**).

$$\begin{tikzcd}
\vec{r}_0, \vec{v}_0 \arrow[r] & \vec{r}_0, \vec{v}_0, \vec{F}_0 \arrow[r] & \vec{r}_{t_1}, \vec{v}_{t_1} \arrow[r] & \vec{r}_{t_1}, \vec{v}_{t_1}, \vec{F}_{t_1} \arrow[l, bend right]
\end{tikzcd}$$

Figure 12 Schematic representation of a Molecular Dynamics simulation

The key quantities of the equation of motion can be expressed by Taylor series expansions for infinitesimal time steps.

$$\begin{aligned}
\mathbf{r}(t+\delta t) &= \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 + \ldots \\
\mathbf{v}(t+\delta t) &= \mathbf{v}(t) + \mathbf{a}(t)\delta(t) + \frac{1}{2}\mathbf{b}(t)\delta t^2 + \ldots \\
\mathbf{a}(t+\delta t) &= \mathbf{a}(t) + \mathbf{b}(t)\delta t + \ldots
\end{aligned} \qquad \text{eq. 3.4.6}$$

The various algorithms in Molecular Dynamics use different truncated formulations for the positions and velocities at time $t+\delta t$. The most common and simplest algorithm is the Verlet algorithm which uses the accelerations and positions at a time t and the positions of the previous step ($t-\delta t$) to predict the new positions.

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \mathbf{a}(t)\delta t^2 \qquad \text{eq. 3.4.7}$$

Equation eq. 3.4.7 can be simply derived by adding the Taylor expansions in second order in time $t+\delta t$ and $t-\delta t$.

$$\begin{aligned}
\mathbf{r}(t+\delta t) &= \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 \\
\mathbf{r}(t-\delta t) &= \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2
\end{aligned} \qquad \text{eq. 3.4.8}$$

The velocities do not explicitly appear in the algorithm and can be calculated afterwards. Moreover, the Verlet algorithm needs a Taylor series for starting the simulation since at $t=0$ no set of positions for $t-\delta t$ exist. A further common algorithm that avoids these disadvantages is the Velocity-Verlet algorithm that gives positions, velocities and accelerations at the same time *t*.

$$\mathbf{r}(t+\delta t)=\mathbf{r}(t)+\mathbf{v}(t)\delta t+\frac{1}{2}\mathbf{a}(t)\delta t^2$$
$$\mathbf{v}(t+\delta t)=\mathbf{v}(t)+\frac{1}{2}\left[\mathbf{a}(t)+\mathbf{a}(t+\delta t)\right]\delta t$$

eq. 3.4.9

The implementation is somewhat more complicated since both the acceleration at time *t* and $t+\delta t$ is needed. However, the Velocity-Verlet algorithm is more precise than the simpler Verlet method implicating a higher storage requirement. Further algorithms that improve the deficiency of the Verlet algorithm are the leap-frog, the Beeman's and predictor-corrected algorithms which all make use of explicit velocities on the expense of higher computational costs.

Performing a MD simulation

A typical MD simulation consists of several consecutive procedures which should ensure energy conservation (see **Figure 12**). First, the initial structure is transformed into an internal coordinate structure and the force-field parameters are set for each atom type. The simulation is prepared by setting constraints, boundary conditions and cut-offs and an energy minimization is performed in order to avoid strong repulsive ("bad") contacts that can be artefacts from crystallographic or NMR data. The actual molecular dynamic simulation is composed of heating the system to a defined temperature using a heat bath, equilibrating until constant kinetic and potential energies are observed, performing the production run and in some cases cooling the system down. The data collected during the production run can be analysed by statistical methods to obtain the thermodynamic properties of interest.

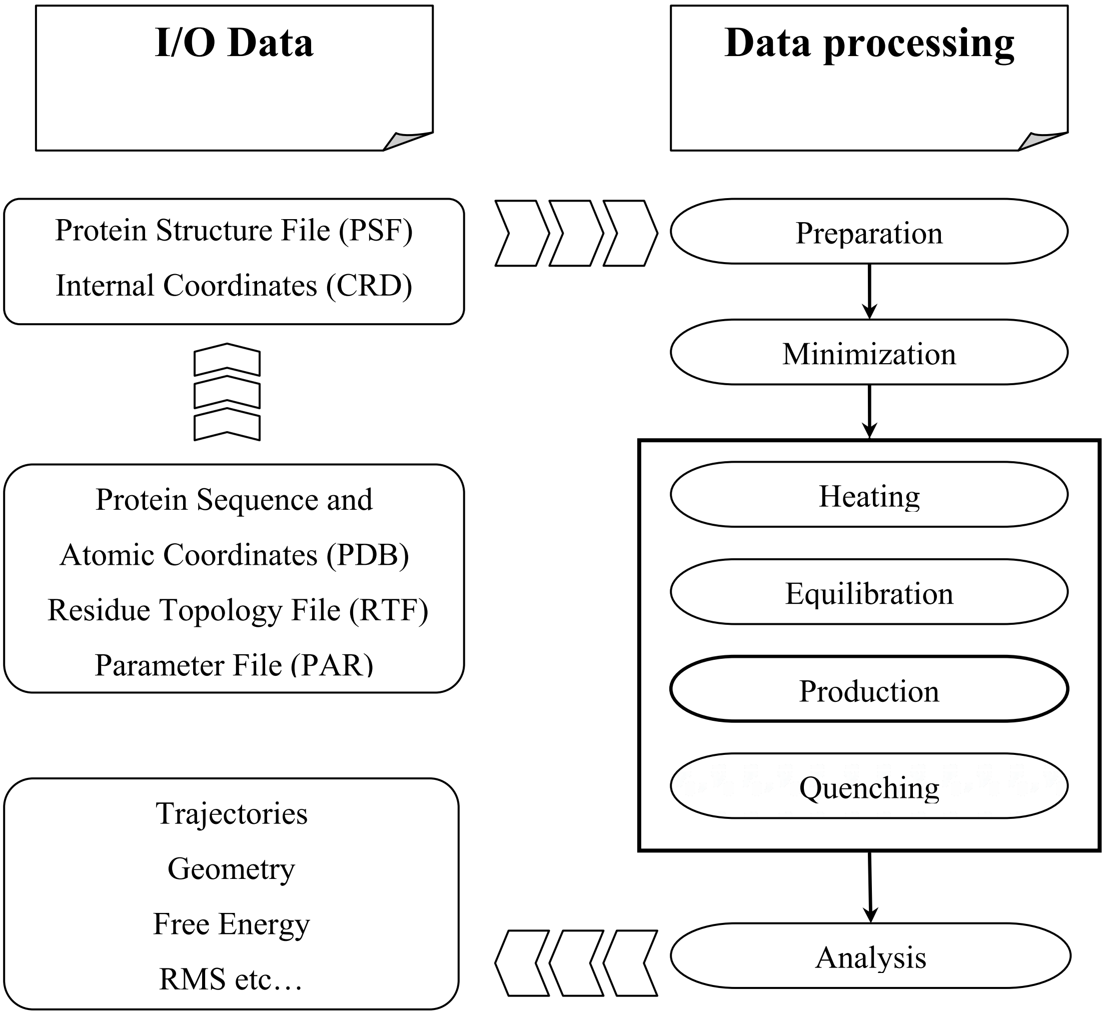

I/O Data
Data processing
Protein Structure File (PSF)
Internal Coordinates (CRD)
Preparation
Minimization
Protein Sequence and
Atomic Coordinates (PDB)
Residue Topology File (RTF)
Parameter File (PAR)
Heating
Equilibration
Production
Quenching
Trajectories
Geometry
Free Energy
RMS etc…
Analysis

Figure 13 Flow-diagram of a standard MD simulation by using a CHARMM like force field consisting of a parameter file and a residue topology file**.**

The Free Energy Problem

The calculation of mechanical properties like the internal energy *U*, the pressure *p* or the heat capacity C_V can be performed easily by means of standard MD or MC simulation methods whereas the determination of thermal (entropic) properties (free energies, chemical potentials, entropies) poses a problem. The reason for this inequality lies in the fact that mechanical properties are dependent on the derivative of the partition function. Thermal properties, however, are directly dependent on the partition function. Regarding the internal energy one can write it as the ensemble average of the Hamiltonian (neglecting normalization constants).

$$U = \frac{k_B T^2}{Q} \cdot \frac{\partial Q}{\partial T} = \frac{k_B T^2}{Q} \cdot \int \frac{H(\mathbf{\Gamma}^N)}{k_B T^2} \cdot \exp\left(\frac{-H(\mathbf{\Gamma}^N)}{k_B T} \right) d\mathbf{\Gamma}^N$$

$$= \int H(\mathbf{\Gamma}^N) \cdot \frac{\exp\left(\frac{-H(\mathbf{\Gamma}^N)}{k_B T} \right)}{Q} d\mathbf{\Gamma}^N = \int H(\mathbf{\Gamma}^N) \cdot \pi(\mathbf{\Gamma}^N) d\mathbf{\Gamma}^N = \left\langle H(\mathbf{\Gamma}^N) \right\rangle \qquad \text{eq. 3.4.10}$$

High values of the Hamiltonian have only very low probabilities and do not significantly contribute to the internal energy. A complete sampling of the phase space is therefore not necessary and the internal energy converges quite well during a MD or MC simulation. In contrast, high energy regions in the phase space make an important contribution on the Helmholtz free energy.[48]

$$A = -k_B T \ln Q = +k_B T \ln \left(\frac{1}{\int \exp\left(\frac{-H(\mathbf{\Gamma}^N)}{k_B T} \right) d\mathbf{\Gamma}^N} \right)$$

$$= k_B T \ln \left(\frac{\int \exp\left(\frac{+H(\mathbf{\Gamma}^N)}{k_B T} \right) \exp\left(\frac{-H(\mathbf{\Gamma}^N)}{k_B T} \right) d\mathbf{\Gamma}^N}{\int \exp\left(\frac{-H(\mathbf{\Gamma}^N)}{k_B T} \right) d\mathbf{\Gamma}^N} \right) \qquad \text{eq. 3.4.11}$$

$$= k_B T \ln \left(\int \exp\left(\frac{+H(\mathbf{\Gamma}^N)}{k_B T} \right) \frac{\exp\left(\frac{-H(\mathbf{\Gamma}^N)}{k_B T} \right)}{\int \exp\left(\frac{-H(\mathbf{\Gamma}^N)}{k_B T} \right) d\mathbf{\Gamma}^N} d\mathbf{\Gamma}^N \right)$$

$$= k_B T \ln \left(\int \exp\left(\frac{+H(\mathbf{\Gamma}^N)}{k_B T} \right) \pi(\mathbf{\Gamma}^N) d\mathbf{\Gamma}^N \right) = k_B T \ln \left\langle \exp\left(\frac{+H(\mathbf{\Gamma}^N)}{k_B T} \right) \right\rangle$$

As the Hamiltonian now appears in the exponent, the free energy is also dependent on high-energy regions of the phase space. However, sampling all points in phase space (= ergodic trajectory) is not applicable to real simulations leading to an erroneous determination of thermal properties. In contrast, the evaluation of the complete partition function is not necessary if one considers free energy differences between two states by sampling only the important parts extensively in which both states differ.

3.4.2 Thermodynamic Integration

For the calculation of free energy differences three methods are commonly used, namely the Slow Growth, the Thermodynamic Perturbation and the Thermodynamic Integration method.[49] These so-called coupling parameter approaches consider two states in the same phase space that are coupled via a parameter λ which determines a topographical transition coordinate. Besides structural transition coordinates (*e.g.* conformational change) and reaction coordinates (*e.g.* proton transfer) also creation/annihilation coordinates (*e.g.* interchange of a functional group or molecule) are possible.

The Thermodynamic Integration method applies a basic mathematical identity to the free energy function $A(\lambda)$ resulting in an exact expression for the free energy difference.

$$\Delta A = \int_0^1 \left(\frac{\partial A(\lambda)}{\partial \lambda} \right) d\lambda \qquad \text{eq. 3.4.12}$$

Deriving the expression for the Helmholtz free energy

$$A(\lambda) = -k_B T \ln Q = -k_B T \ln \left(\frac{1}{h^{3N} N!} \int \exp\left(\frac{-H(\mathbf{\Gamma}^N, \lambda)}{k_B T} \right) d\mathbf{\Gamma}^N \right) \qquad \text{eq. 3.4.13}$$

with respect to λ leads to

$$\frac{dA(\lambda)}{d\lambda} = -k_B T \left(\frac{\int \left(\frac{1}{k_B T} \cdot \frac{\partial H(\mathbf{\Gamma}^N, \lambda)}{\partial \lambda} \right) \cdot \exp\left(\frac{-H(\mathbf{\Gamma}^N, \lambda)}{k_B T} \right) \cdot h^{3N} N! d\mathbf{\Gamma}^N}{\int \exp\left(\frac{-H(\mathbf{\Gamma}^N, \lambda)}{k_B T} \right) d\mathbf{\Gamma}^N} \right)$$

$$= \int \left(\frac{\partial H(\mathbf{\Gamma}^N, \lambda)}{\partial \lambda} \right) \frac{\exp\left(\frac{-H(\mathbf{\Gamma}^N, \lambda)}{k_B T} \right) \cdot h^{3N} N!}{\int \exp\left(\frac{-H(\mathbf{\Gamma}^N, \lambda)}{k_B T} \right) d\mathbf{\Gamma}^N} d\mathbf{\Gamma}^N = \int \left(\frac{\partial H(\mathbf{\Gamma}^N, \lambda)}{\partial \lambda} \right) \pi(d\mathbf{\Gamma}^N) d\mathbf{\Gamma}^N \qquad \text{eq. 3.4.14}$$

$$= \left\langle \frac{\partial H(\mathbf{\Gamma}^N, \lambda)}{\partial \lambda} \right\rangle_\lambda$$

The free energy difference is in general calculated by a finite difference approximation using a step size $\Delta\lambda$ of 0.1. For each value of λ a MD simulation run is performed and the average of eq. 3.4.14 is determined.

Determining Entropy Differences

The estimation of entropy differences (see ref. 50) of two states *a* and *b* can be done directly either by using the free energy difference between the endpoints of a thermodynamic integration pathway of a canonical ensemble

$$\Delta S_{ab} = \frac{\Delta U_{ab} - \Delta A_{ab}}{T} \qquad \text{eq. 3.4.15}$$

or by employing a finite difference approximation on the temperature derivative of the free energy

$$S = -\left(\frac{\partial A}{\partial T}\right)_{N,V} \qquad \text{eq. 3.4.16}$$

leading to

$$\Delta S^{\Delta T} = -\frac{\Delta A_{ab}(T+\Delta T) - \Delta A_{ab}(T-\Delta T)}{2\Delta T} \qquad \text{eq. 3.4.17}$$

Latter approach assumes a constant difference in the heat capacity over the temperature range 2Δ*T*. A third method to obtain entropy differences can be derived from the thermodynamic integration formula expressing the entropy as a function of the coupling parameter.

$$\Delta S_{ab}^{TI} = \frac{1}{k_B T^2} \cdot \int_0^1 \left\{ \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda \langle H \rangle_\lambda - \left\langle \frac{\partial H)}{\partial \lambda} H \right\rangle_\lambda \right\} d\lambda \qquad \text{eq. 3.4.18}$$

3.5 The QM/MM Method

All studies of large molecular systems like proteins, nucleic acids, lipids, membranes or liquids require a classical treatment of the particles due to the enormous system size up to hundred thousands of atoms. Unfortunately, a mere force field based description does not account for bond formation or cleavage since changes in the covalence of an atom type cannot be parameterized. However, a quantum mechanical description of the entire system is not feasible so that in 1976 Warshel and Levitt proposed a hybrid QM/MM scheme that considers only a small part of the system in which the chemical reaction takes place by a QM calculation.[51] The surrounding of the active region and the boundary region are treated by molecular mechanics with an appropriate coupling scheme connecting the QM and the MM part of the full system.

Partitioning of the Hamiltonian

The energy of the system is described by the time-independent Schrödinger equation (cf. eq. 3.1.1) in which the quantal Hamiltonian is substituted by an effective Hamiltonian that splits into four parts.[44,52]

$$\mathbf{H}_{eff} = \mathbf{H}_{QM} + \mathbf{H}_{MM} + \mathbf{H}_{QM/MM} + \mathbf{H}_{boundary} \qquad \text{eq. 3.5.1}$$

The first term represents the quantum mechanical Hamiltonian in the Born-Oppenheimer approximation (cf. eq. 3.1.4), whereas the second term is simply replaced by the energy expression of a given force field (cf. chapter 3.2.1). The coupling between both regions is described by $\mathbf{H}_{QM/MM}$ and is commonly treated by taking the electrostatic and weak interactions into account.

$$\mathbf{H}_{QM/MM} = \underbrace{\left\{ -\sum_{iM} \frac{q_M}{\mathbf{r}_{iM}} + \sum_{\alpha M} \frac{Z_{\alpha M}}{\mathbf{r}_{\alpha M}} \right\}}_{\text{electrostatic ineraction}} + \underbrace{\left\{ \sum_{\alpha M} 4 \cdot \varepsilon \cdot \left[\left(\frac{\sigma_{\alpha M}}{\mathbf{r}_{\alpha M}} \right)^{12} - \left(\frac{\sigma_{\alpha M}}{\mathbf{r}_{\alpha M}} \right)^{6} \right] \right\}}_{\text{weak interaction}} \qquad \text{eq. 3.5.2}$$

The weak interaction is analytically described by a standard Lennard-Jones potential (cf. **Figure 8**) and the electrostatic part includes both the electrons (i) and nuclei (α) of the QM region interacting with the MM atoms (M).

Finally, the Hamiltonian for the boundary region is given by the boundary conditions of classical force field calculations (*e.g.* periodic boundary conditions or stochastic boundary conditions) and can be separated into one part that affects the QM region and one part for the MM region. The total energy of the system is then calculated by the expectation value of the effective Hamiltonian given by

$$E = \frac{\left\langle \Psi \middle| \mathbf{H}_{QM} + \mathbf{H}_{QM/MM} + \mathbf{H}_{boundary(QM)} \middle| \Psi \right\rangle}{\left\langle \Psi \middle| \Psi \right\rangle} + E_{MM} + E_{boundary(MM)} \qquad \text{eq. 3.5.3}$$

Fragmentation of large molecules and coupling methods

Considering a system consisting of well separated molecules like a solute surrounded by a solvent the determination of a QM and a MM region does not invoke severe problems. However, a suitable decomposition of a large biomolecule is much less obvious as one has to define QM and MM fragments that are connected via covalent bonds. In principle, one can distinguish two different approaches that deal with the problem how to describe the electronic structure of such a frontier bond (see '*Hybrid Quantum Mechanical/Molecular Mechanical (QM/MM) Methods*' in ref. 44) The link atom methods include the addition of an atom or pseudoatom, whereas the fragment orbital approach uses a transformation of the atomic orbitals of the frontier QM atom into hybrid orbitals that are collinear to the bond axis and that are excluded from the regular orbital basis in the QM calculations. The associated electron densities of the fragment orbitals are considered as external point charges acting on the cationic QM fragment.

The partitioning of the Hamiltonian and the definition of a QM/MM term as seen in equation eq. 3.5.2 is of course arbitrary and therefore various methods to describe the coupling between the QM and MM region have been developed.[53] They are classified according to the embedding of the QM part into the MM surrounding:

- The mechanical embedding scheme includes the interaction energy between both regions solely by a mechanic calculation of the system without any charge interactions between the QM and MM region. The QM part only "sees" the MM part in terms of a steric hindrance.
- The electrostatic embedding scheme includes the MM atoms as point charges into the QM Hamiltonian. The electrostatic part of the QM/MM interaction is therefore taken into account in each SCF cycle, whereas the van-der-Waals and covalent interactions are still treated by molecular mechanics.
- The polarized embedding scheme also includes the polarization of the MM region in order to improve the description of non-bonded QM/MM interactions.

Chapter 4 Results and Discussion

"If the facts don't fit the theory, change the facts."

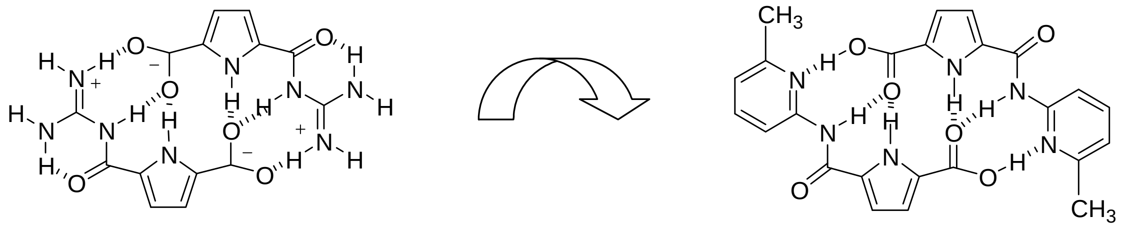
-- ***Albert Einstein***

4.1 Quantifying Supramolecular Processes by Knock-out Analogues

4.1.1 Introduction

The development of novel building blocks which are capable to self-assemble in polar solutions is one main goal in today's supramolecular chemistry[1,54] as molecular recognition-directed self-assembly and self-organization can lead to the formation of highly complex and fascinating structures with new and interesting properties.[55] However, so far only very few systems show strong self-assembly in polar, especially aqueous solution. For example, purely hydrogen bonded assemblies possess considerable association energies only in aprotic solvents of low polarity and are not stable in water due to the competitive solvation of donor and acceptor sites in water.[56] Therefore, to achieve strong self-assembly hydrogen bonds have to be combined with additional noncovalent interactions such as metal coordination,[57] salt bridges,[58] hydrophobic[59] or $\pi-\pi-$interactions. [60,61]

Based on the new recognition motif of Schmuck (cf. **Figure 2**) a self-complementary zwitterion **1** (see **Figure 14**) has been developed that forms extremely stable dimers as could be shown by X-ray, ESI-MS, and NMR solution studies.[14] The association constant is approximately $K > 10^{10}$ M^{-1} in DMSO and still surprisingly high ($K = 170$ M^{-1}, $\Delta G \approx -15$ kJ mol^{-1}) in water. Therefore compound **1** is one of the most efficient self-assembling systems relying solely on electrostatic interactions reported so far. Hence, an interesting question is which of the multiple binding interactions present in this dimer is mainly responsible for its unique binding properties? It could already be shown experimentally by comparison with a neutral amidopyridine pyrrole analogue that the charge interaction within the ion pair is crucial for its high stability. The neutral binding motif in this "knock-out" analogue **2** has the same H-bond pattern like dimer **1** as could be proved by X-ray analysis. Nevertheless, the dimerization is several orders of magnitude less efficient. Whereas **2** dimerizes in chloroform with $K > 10^4\ M^{-1}$ already the addition of > 5 % DMSO completely disrupts these dimers due to the competitive solvation of the H-bond donor by the polar solvent.



1 **2**

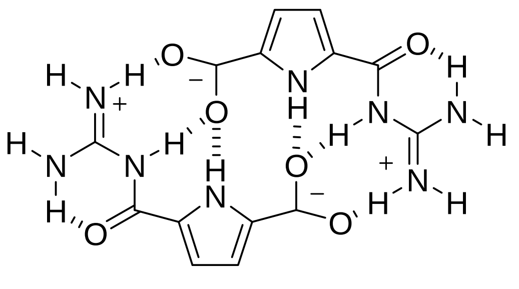
Figure 14 Amidopyridine pyrrole carboxylic acids as neutral “knock-out” analogues of zwitterionic guanidiniocarbonyl pyrrole carboxylates: Translating the zwitterionic dimer **1** into a neutral amidopyridine pyrrole carboxylic acid dimer **2** by “switching off” the ionic interactions while keeping the hydrogen bond network constant.

On the basis of these data one could assume that the main important factor being responsible for the high stability of zwitterion **1** is simply the charge interaction. However, already a single guanidiniocarbonyl pyrrole/carboxylate ion pair is much stronger than simple salt bridges between carboxylates and ammonium ions or even the parent guanidinium cation. Therefore, one has to account for the various H-bonds, their number and strength, the properties of the actual ion pair and further secondary electrostatic and cooperative effects. In order to design even better self-assembling systems for future applications (*e.g.* for supramolecular polymers) a detailed understanding of the importance of these various non-covalent interactions and their mutual interplay is needed. However, experimentally this is difficult to achieve as only the overall association energy can be determined. It is impossible to dissect this data into individual contributions of single interactions. The comparison of structurally closely related “knock-out” analogues is one way address this problem and to obtain at least semi-quantitative data (as shown above for zwitterion **1** and its neutral analogue **2**). However, very often the most interesting “knock-out” analogues cannot be made synthetically or might not even be stable molecules at all. Computational determination of their stabilities does not encounter any of these problems and indeed high level theoretical approaches have already proven quite useful to analyze supramolecular systems in general.[62]

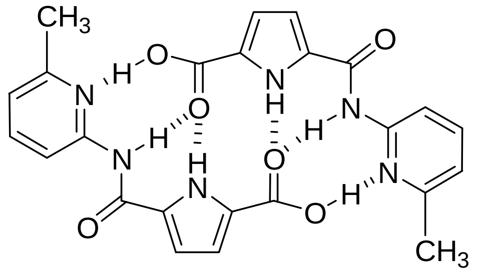
This approach is therefore used here to study in detail the various non-covalent interactions and factors that might be responsible for the high stability of zwitterion **1**.

In the present work the dissociation energies of a systematically varied series of “knock-out” analogues have been calculated by means of density functional approaches. This should give an insight into the strengths of the individual hydrogen bonds within these complex binding motifs which probably vary for every donor site. Additionally, it should elucidate the importance of cooperative effects (*e.g.* secondary interactions) which can be expected to be as important as already seen for the guanine cytosine pairing.[63]

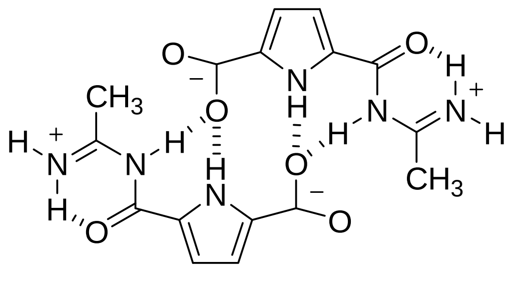
The compounds used in this study are shown in **Figure 15**. In each of these analogues one of the several non-covalent interactions present in **1** is switched-off. The amidopyridine dimer **2**, which was also already studied experimentally, has the same H-bond pattern but no charge interactions. The “knock-out” analogues **3a**, **4**, **5** and **6** are again zwitterionic but the individual N-H hydrogen bond donor sites are replaced by either methylene groups as in the methyl (**3a)**, amidine (**4**) and cyclopentadienyl (**5**) derivative or by an oxygen atom (in the furan derivative **6**). Dimer **3b** is obtained from **3a** by an internal rotation. It possesses the same hydrogen pattern than **1** and was included to study the influence of a methylation of the amidinium unit on the stability. For arginine such effects were found to stabilize the zwitterionic species with respect to the neutral one (*vide infra*).[64] For these latter “knock-out” analogues **3** - **6** no experimental data are available and at least for **3a** and **5** this is probably impossible to achieve due to their conformational (**3a**) and tautomeric (**5**) instability.



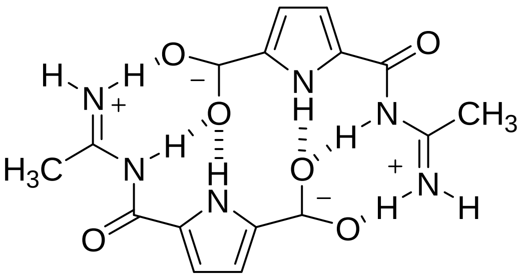
1



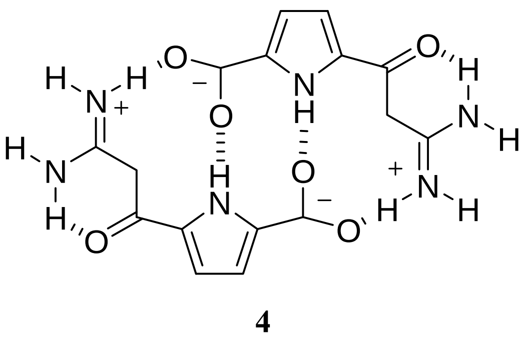
2



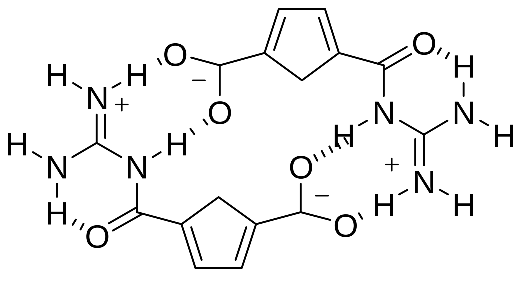
3a



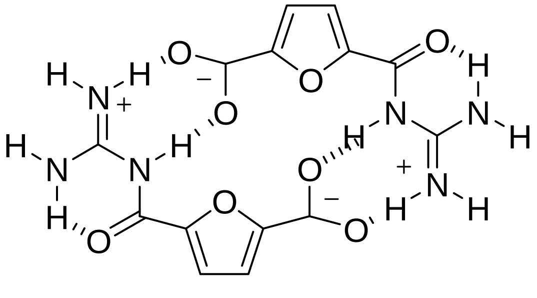
3b



4



5



6

Figure 15 Guanidiocarbonyl pyrrole carboxylate dimer (**1**), amidopyridine pyrrole carboxylic acid dimer (**2**), methyl derivative (**3a**, **3b**), amidine derivative (**4**), cyclopentadienyl derivative

4.1.2 Computational Details

The geometry optimizations of all compounds were performed with the TURBOMOLE program package[65] at the BLYP/TZVPP level of theory[35b-c,66] using the RI approximation.[67] For the zwitterionic species extra diffuse functions were added to the negative charged carboxylate oxygens in order to describe the diffuse shape of the valence orbitals properly. The TZVPP basis set was enlarged by 1s and 1p primitive uncontracted basis functions with an exponential coefficient of 0.068, whereas for the auxiliary basis sets the exponent was doubled (0.136). Dissociation energies were calculated including the counterpoise correction according to Boys and Bernardi.[24]

In most computations the influence of a solvent is dissected in several parts.[68] In the present paper the so-called electrostatic contributions (often also abbreviated as electrostatic component of solvation) were estimated using the COSMO[43] approach as implemented in TURBOMOLE[69] with a dielectric constant of $\varepsilon = 78$ to simulate a water like solvent. Since the COSMO implementation in TURBOMOLE only takes electrostatic contributions of the solvent into account, the non-electrostatic effects[68] were estimated by single-point calculations (BLYP/6-31++G(d,p))[70] on the optimized structures in water employing the Gaussian03 program package[71] implementation of the COSMO.

All optimized structures were characterized by harmonic frequency analysis employing either analytical derivatives (RI-DFT/BLYP/TZVP) for gas phase structures as implemented in TURBOMOLE or numerical derivatives (RIDFT/BLYP/TZVPP) for solvated structures using the SNF program, respectively.[72] Thermodynamic corrections for the gas phase were obtained with TURBOMOLE employing the standard approach.[65g-h] Thermodynamic corrections in solution were obtained by frequency calculations with SNF program of the TURBOMOLE-suite employing the COSMO approach with $\varepsilon = 78$.[72] For the computation of entropy effects resulting from the translation motion (ΔS_{trans}) this implementation uses the standard formula for gas phase.[65g-h] However, this approximation overestimates the absolute values. As a consequence the stability of dimer formation is underestimated as discussed recently.[73,74] Let us take **2** as an example: Employing the approximation of Williams and coworkers[73] to estimate ΔS_{trans} for a solvent the absolute value for $T\Delta S$ obtained with the standard formula decreases by about 20 kJ mol^{-1}. Despite this influence we refrained from considering this effect due to the following reasons: Within the approximation of Williams and coworkers differences between compounds arise only due to their masses, *i.e.* only a small fraction of the various effects are included. As a consequence mainly the absolute values change, but the differences between the various compounds studied here stay more or less constant.

As expected[75] test calculations employing various functionals and the MP2 approach[76] showed that the BLYP functional underestimates the dissociation energies. Nevertheless it gives geometrical parameters which are virtually identical to those obtained with the B3LYP functional. The latter predicted a stronger binding. Therefore we computed improved stabilities for solvent conditions employing the B3LYP functional based on previously optimized BLYP geometries. The thermodynamic corrections are also taken from BLYP calculations. Since we are more interested in solvent data the BLYP functional was employed for gas phase calculations throughout.

Coupled Cluster computations[75,77] indicate that also B3LYP often underestimates dissociation energies for hydrogen bonds. Consequently, its predictions may be looked upon as lower bounds for the dissociation energies. The computed differences between the various knock-out analogues, however, should possess a considerably higher accuracy since the binding situations are quite similar. To get a deeper insight into the variations appearing in our series of model compounds the electrostatic potentials of all compounds for both gas phase and solvent have been calculated to visualize variations in the electronic distributions and molecular interactions of guanidiniocarbonyl pyrrole carboxylate 1 and its knock-out analogues **2** - **6**. For these computations the Gaussian03 program package[71] was used.

4.1.3 Geometries

Table 3 summarizes selected computed geometrical parameters whereas compares computed and measured distances of the bonds described in **Figure 16**. **Table 4** contains data for compounds **1** and **2** for which X-ray data are available. For all covalent bonds, computed and measured structural parameters agree in the expected range (±0.02 Å). Measured and computed distances between the heavy centres of bonds 1-3 agree to about 0.1 Å. The larger deviations are expected due to the weakness of the bonds and crystal effects.

Table 3 Hydrogen bond lengths in (BLYP/TZVPP; solvent calculations performed with COSMO).[a]

bond	1		2		3	
compound	gas phase	solvent	gas phase	solvent	gas phase	solvent
1	1.68	1.86	1.58	1.69	1.79	1.81
2	1.77	1.72	1.85	1.88	1.82	1.86
2[b]	1.83	-	1.83	-	1.79	-
3a	-	-	1.58	1.76	1.76	1.78
3b	1.65	1.81	1.60	1.70	1.79	1.80
4	1.52	1.77	-	-	1.80	1.90
5	1.63	1.84	1.61	1.76	-	-
6	1.54	1.74	1.62	1.77	-	-

[a] Numbering of bonds according to **Figure 16**. [b] With fixed Cs symmetry

According to the available X-ray data compound **1** and **2** exhibit a planar structure. In contrast, geometry optimizations in gas phase or polar solvent give slightly bended geometries but the bending potential is extremely flat. For **1** the planar geometry, which represents a local minimum, lays only about 1 kJmol^{-1} higher than the bended structure. For **2** the energy difference is only 2 kJ mol^{-1}. The differences are so small that π-π stacking interaction within the crystal can explain the difference between experiment and theory. Additionally, already dynamic effects (large amplitude bending motion) are expected to lead to averaged planar geometries in X-ray experiments and in solution

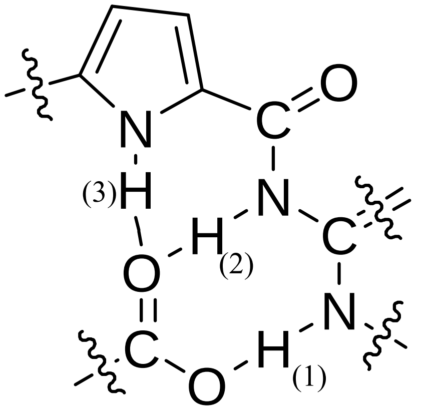


Figure 16 Numbering of the intermolecular bonds in the dimers.

Table 4 Heteroatomic C···N distances obtained from X-ray studies[a] and calculations (BLYP/TZVPP//gas phase).

bond	1		2		3	
compound	*X-ray*	*calc..*	*X-ray*	*calc.*	*X-ray*	*calc.*
1	2.85	2.75	2.68	2.65	2.73	2.77
2	2.62	2.80	2.72	2.88	2.73	2.81

[a] The amidopyridine pyrrole carboxylic acid dimer **2** was synthesized with hexyloxymethyl groups in position 3 and 4 of the pyrrole ring.[14]

Our calculations show that from the "knock-out" derivatives only the methyl derivatives **3a** and **3b** have a planar geometry. The geometries of the other compounds are more or less distorted due to steric or electronic effects. In and, which contain the electrostatic potentials, the distortions are best seen in the slight rotations of the carboxylate groups out of planarity. The optimization of the amidine derivative **4** revealed two conformers, differing only in the relative orientation of the methylene units within the dimer and resembling therefore in a side view a "boat" and a "chair" conformer, whereof the latter is ~ 5 kJ mol^{-1} more stable in gas-phase. In the cyclopentadienyl derivative **5** a hydrogen atom of the methylene group of the ring system points towards the carboxylate group, so that the cyclopentadienyl rings are forced into an up and down orientation. Also for the knock-out analogues **3a** and **4** the geometry optimizations lead to structures in which one of the hydrogen atoms of the methyl (**3**) or methylene (**4**) group is directed towards the carboxylate group. The distances are between 2.01 and 2.13 Å pointing to small attractive interactions. The distortions within the furan derivative **6** result from the electronic repulsion of the oxygen lone pairs of the furan oxygen and the carboxylate oxygen. This is expected to be a similar effect as observed experimentally for the pyridine derivatives.[13]

The calculated hydrogen bond lengths for **1** in gas phase and polar solvent show, that solvent effects influence the individual bonds differently (**Table 4**). As expected the influence decreases going from the outer (bond 1 in **Figure 2**) to the inner bond (bond 3). Bond 1 representing the second shortest one for the gas phase (1.68 Å) increases by 26 % and becomes the longest bond in a polar solvent (1.86 Å). Bond 2 is elongated by about 0.1 Å (7 %) but still remains the shortest bond. The influence on the inner bond is negligible (0.02 Å or 1%). Comparing the hydrogen bond lengths of **2** obtained for gas-phase with the values calculated using the COSMO approach the inner H-bond is only slightly longer in solvent than in vacuum. The largest change in a magnitude of about 0.05 Å can be observed for the outer bond, but in contrast to the zwitterionic dimer **1** the H-bond length now decreases a little upon solvation. This does not indicate increased bond strength but results from larger bending angles.

Table 5 Contributions to the total dissociation energies (all values given in kJ mol^{-1}). Thermodynamic corrections for T = 298 K).

	1	2	3a	3b	4	5	6
gas phase							
$^{1}\Delta E_{\text{elec}}$	+438/+464	--/--	+340/--	+443/--	+364/--	+364/--	+345/--
$^{2}\Delta E_{\text{elec}}$	+158/+170	+116/--					
$^{3}\Delta H^{\text{corr}}$	-9	-10	-11	-11	-21	-11	-6
$^{4}T\Delta S^{\text{corr}}$	+85	+61	+72	+74	+68	+80	+67
$^{6}\Delta G$	+344/+370	--/--	+256/--	+358/--	+275/--	+273/--	+272/--
$^{7}\Delta G$	+64/+76	+45/--					
solvent							
$^{8}\Delta E_{\text{elec}}$	+108	+48	+64	+111	+55	+64	+51
$^{9}\Delta E(\text{n.e.})$	+9	+3	+6	+9	+4	+5	+4
$^{3}\Delta H^{\text{corr}}$	+1	-6	-3	-4	-4	+1	+2
$^{4}T\Delta S^{\text{corr}}$	+62	+62	+66	+50	+59	+58	+56
$^{10}\Delta G$	+56	-17	+1	+66	-3	+13	-2

[1] Dissociation energy with respect to the zwitterionic monomers. The left value gives the BLYP result. For the right value the B3LYP functional was employed. [2] Dissociation energy with respect to the neutral monomers. The left value gives the BLYP functional results, for the right value the B3LYP functional was employed. [3] Correction to the free dissociation energies to obtain the enthalpy term (BLYP computations). [4] Correction to the free dissociation energies arising due to the entropy term (T=298 K, BLYP computations). [6] Free dissociation energies $\Delta G = \Delta E + \Delta H^{\text{corr}} - T\Delta S^{\text{corr}}$ with respect to the zwitterionic monomers. The left value gives the dissociation energy obtained with the BLYP functional, the right value gives the dissociation energy computed with the B3LYP functional. [7] Free dissociation energies $\Delta G = \Delta E + \Delta H^{\text{corr}} - T\Delta S^{\text{corr}}$ with respect to the neutral monomers assuming that the thermodynamic correction are similar to those computed for the dissociation into the zwitterionic monomers. The left value gives the dissociation energy obtained with the BLYP functional, the right value gives the results of the B3LYP functional. [8] Dissociation energies with respect to the lowest lying monomers. These represent the zwitterionic forms for **1**, **3**-**6**, while it is the neutral monomer for **2**. The B3LYP functional was employed in combination with BLYP geometries. [9] Corrections to the free dissociation energies arising due to the non-electrostatic interactions (free energy of cavity, dispersion-repulsion interaction between solute and solvent). The calculations were performed with GAUSSIAN03 (BLYP computations). [10] Free dissociation energies $\Delta G = \Delta E_{\text{elec}} + E(\text{n.e.}) + \Delta H^{\text{corr}} - T\Delta S^{\text{corr}}$. Thermodynamic corrections are obtained with the BLYP functional.

4.1.4 Energies

The calculated dissociation energies for gas-phase and solvent for all compounds **1-6** are given in Table 3. Table 3 also contains the computed thermodynamic corrections leading to the dimerization enthalpies and dimerization free energies. The electrostatic potentials mapped on isosurfaces of electron densities of all compounds in gas phase and polar solvent are given in **Figure 18** to **Figure 21**.

*The zwitterionic dimer (**1**)*

For the gas phase the dissociation energy (ΔE_{elec}) of zwitterion **1** with respect to the zwitterionic monomers is calculated to +464 kJ mol^{-1} (B3LYP/TZVPP). This value is surprisingly high compared to other guanidinium/carboxylate-based ion pairs, for example the arginine dimer.[64,77] Arginine possesses a high affinity to form an abundant number of clusters when electrosprayed into gas phase.[78] Theoretical studies[77] predict that zwitterionic dimers are formed which are stabilized by two guanidinium-carboxylate salt bridges. The dissociation energy of the zwitterionic structure was calculated to 199 kJ mol^{-1} by Goddard III and coworkers.[77] Hence, with respect to its zwitterionic monomers dimer **1** is more than twice as stable as the zwitterionic arginine dimer with respect to its zwitterionic monomers. A closer look at the dissociation channels reveals however that the possible reason for this extraordinary stability of dimer **1** lies more within the energy content of the corresponding monomers than the actual binding interactions within the dimers. **Figure 17** summarizes the computed values for **1** and for the arginine dimer.[64]

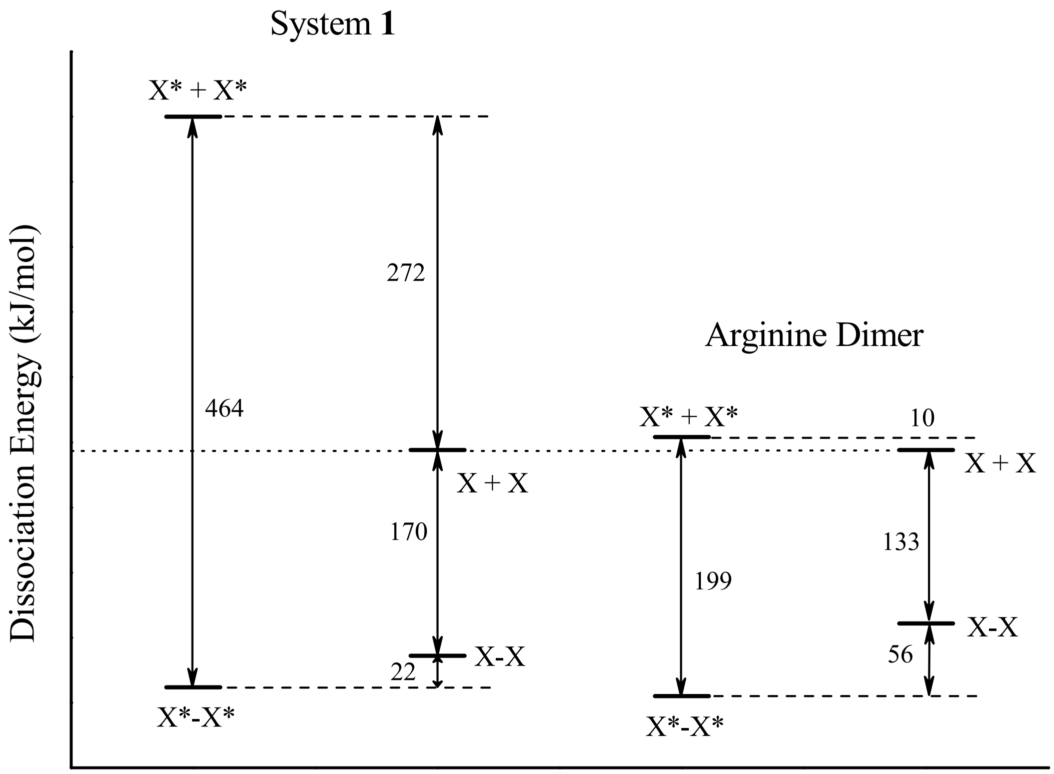


Figure 17 Left: reaction diagram of **1** in gas phase (B3LYP/TZVPP//BLYP/TZVPP). Right: reaction diagram of arginine in gas-phase (B3LYP/6-31G**).[34] The neutral form of a monomer is abbreviated as X, whereas X* denotes the zwitterionic analogue.

The dissociation energies mentioned above refer to the dissociation into two zwitterionic monomers. However, in the gas phase isolated zwitterions are normally energetically less stable than the corresponding neutral monomers. The stability of such zwitterionic monomers is significantly depending on the possibility of internal charge interactions. For example, for arginine the neutral monomer is still more stable than the zwitterion but the energy difference between both forms is rather small (~ 5 kJ mol^{-1}).[77,79] Due to the flexibility of the molecule an effective intramolecular charge interactions between the carboxylate and the guanidinium cation is possible, stabilizing the zwitterionic form. Methylation of the arginine[77] or the presence of an electric field[80] is already sufficient to make the zwitterionic form the absolute minimum. Similar effects were recently found for guanidiniocarbonyl pyrrole/carboxylate zwitterions with flexible linkers of varying chain length between both ionic groups. It was shown that the stability of the zwitterionic form depends on the length of the linker.[81] Only those zwitterions in which the linker is long enough to allow internal charge interactions are zwitterionic in the gas phase. For the smaller ones the neutral form is more stable.

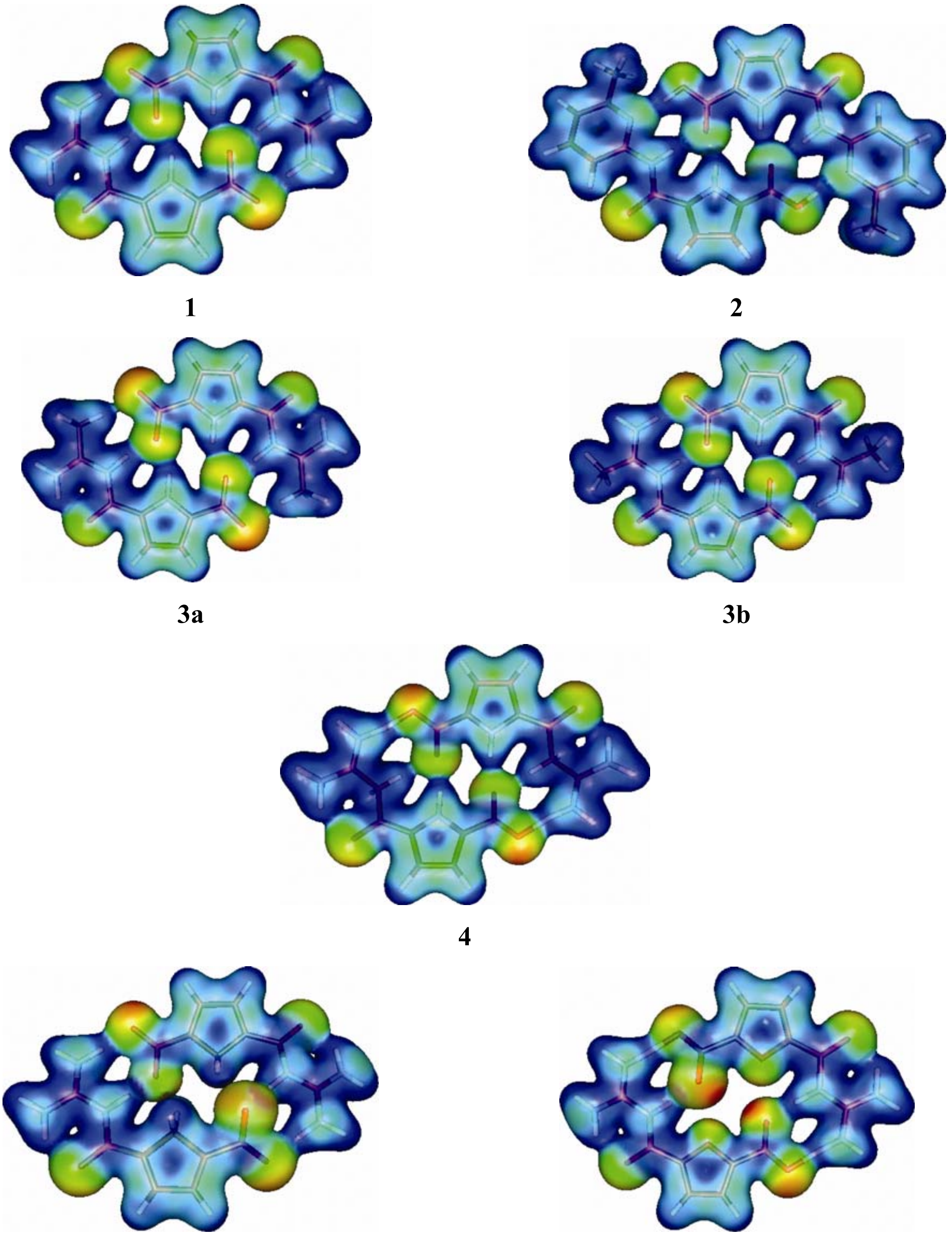
For zwitterion **1** no internal charge stabilization is possible due to the rigidity of the molecule. In accordance with this, we compute an energy difference of 136 kJ mol-1 between the zwitterionic and the neutral monomer of **1** (**Figure 17**, left hand side). Hence, the energy difference is one order of magnitude larger than for arginine (**Figure 17**, right hand side). For the dimers, however, even in gas phase the zwitterionic form represents the minimum structure. The neutral dimer structures obtained through a double proton transfer from the guanidinium to the carboxylate groups also represent local minima on the hyper surface but are less stable. For **1** B3LYP/TZVPP predicts the neutral structure to be 22 kJ mol* above the zwitterionic structure. For the arginine dimer the difference between the neutral and the zwitterionic structure is computed to 56 kJmol^{-1}.[77] As the guanidinium group in arginine (pKa = 13.5) is about six orders of magnitude less acidic than the acyl guanidinium group in **1** (pKa = 7-8), proton transfer in **1** is expected to be easier as reflected by these data.

The energetically most favourable dissociation channel should therefore lead from the zwitterionic dimers to the neutral monomers. If one considers this process, compound **1** and the arginine dimer become equally stable. For dimer **1** we compute a dissociation energy of 192 kJ mol^{-1} while Goddard III and coworkers[77] give a value of 189 kJ mol-1 for the arginine dimer (Figure 1). If one compares the dissociation of the less stable neutral dimers into its neutral monomers, the arginine dimer possesses a dissociation energy of about 133 kJ mol^{-1}, whereas for **1** we find a dissociation energy of 170 kJ mol^{-1}, respectively. The difference in the dissociation energies of both neutral structures is reasonable since **1** is stabilized by two additional hydrogen bonds between the pyrrole N-H unit and the carbonyl oxygen of the carboxylic acid.

A dimerization free energy of about ΔG = +76 kJ mol^{-1} is calculated for the energetically most favourable dissociation of the zwitterionic dimer **1** into the neutral monomers using the same thermodynamic corrections as calculated for the dissociation into two zwitterions (ΔG = +370 kJ mol-1). However, in an attempt to dissociate **1** in the gas phase using IRMPD-MS techniques only fragmentation due to covalent bond rupture was observed.[82] Assuming that such bond rupture needs energies in the range of a normal covalent bond (> 250 kJ mol^{-1}) this experimental outcome indicates that dissociation of zwitterionic dimer **1** requires more energy than expected based on the calculated stabilities of both the dimer and monomers. But this dissociation channel requires a double proton transfer. Obviously, this imposes a large energy barrier onto the dissociation.

As expected for electrostatic interactions, solvation by a polar solvent drastically affects the stability of dimer **1.** In general, the stability of salt bridges is influenced by the polarity of the solvent[83] or microsolvation.[84] In contrast to the situation in gas phase, for a polar solvent the zwitterionic form now represents the global minimum for both the monomer and the dimer.[14,85] Therefore, the dissociation channel that has to be considered now leads from zwitterionic dimer **1** directly to the zwitterionic monomers. The dissociation energy of **1** to the zwitterionic monomers is reduced to +108 kJ mol^{-1} in water ($\approx$ 23% of the gas-phase value) which translates into a dissociation free energy of ΔG = +56 kJ mol^{-1}. This is quite reasonable compared to the experimental value of $\Delta G \approx$ +15 kJ mol^{-1} measured from NMR dilution studies. In comparing these data, one has to take into account that our theoretical approach computes dissociation free energies for one single dimer of **1** in the solvent. The experimental values are however measured at millimolar concentrations. As the ionic strength (= salt concentration) of the solution has a tremendous destabilizing effect on the stability of salt bridges, it is not surprising that the experimental value is smaller than the calculated one. For example, Schneider assigns an upper limit of dissociation energies of $\Delta G \approx$ 8 kJ mol^{-1} to single organic ion pairs in an indefinite dilute solution,[86] but at millimolar concentrations the corresponding association constants of these ion pairs are more than a factor of 1000 smaller! Gallivan and Dougherty came to a similar conclusion on the basis of a theoretical characterisation of the methylammonium-acetate dimer.[62g] The computed and measured data underline that similar to the situation in gas phase also for water as solvent dimer **1** is much more stable than other organic zwitterionic dimers. For regular organic zwitterionic dimers a stability of $\Delta G \le 16$ kJ mol^{-1} would be expected based on Schneider's evaluation of literature data. Hence, our dimer **1** is at least three times more stable. Similar to the situation found in the gas phase part of this larger stability results probably again from the higher energy content of the rigid monomers of **1** compared to more flexible zwitterions. The instability of the zwitterionic monomer caused by its lack of intramolecular charge stabilization seems to emerge as an interesting principle for the realization of highly stable electrostatically driven self-assembly.

With respect to the gas phase (**Figure 18**) the electrostatic potential computed for a polar solvent (**Figure 20**) shows a considerably higher polarization. The electrostatic potentials reflect nicely the strong binding interaction between both monomers.



1

2

3a

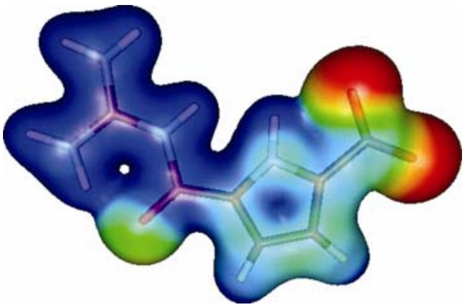
3b

4

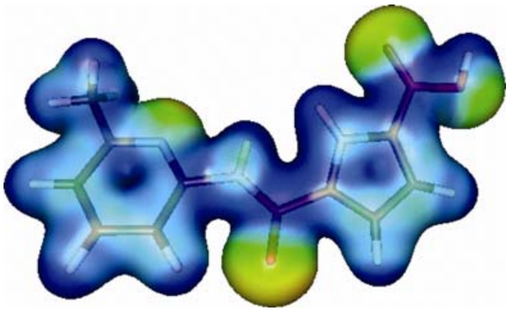
5

6

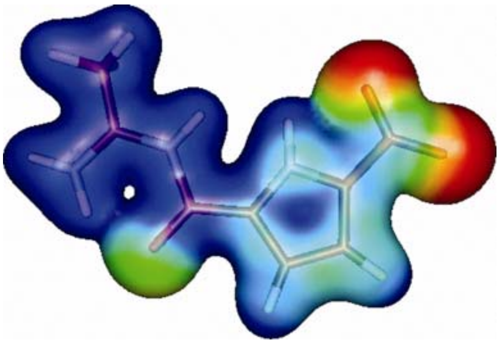
Figure 18 Electrostatic potential (contour value = 0.02) mapped on the electron density (contour value = 0.015) of the dimers **1-6** in gas phase.



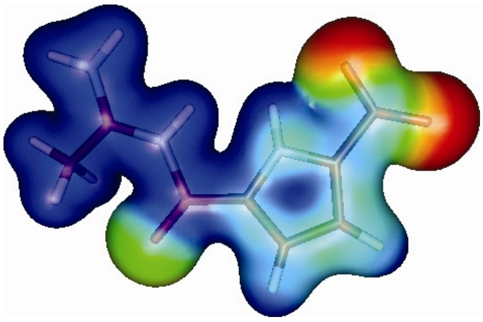
1



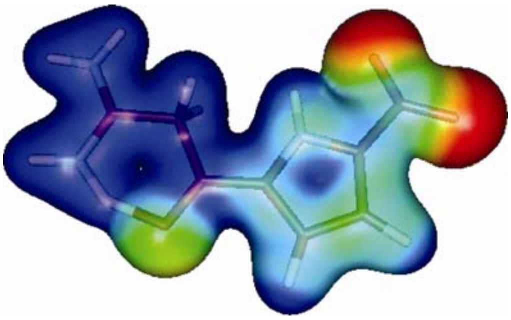
2



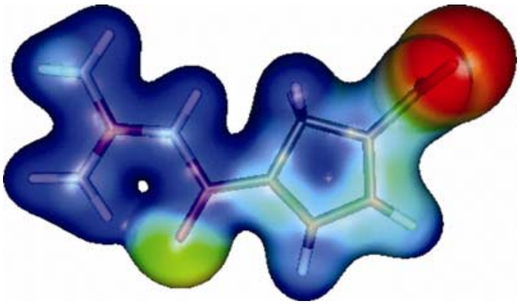
3a



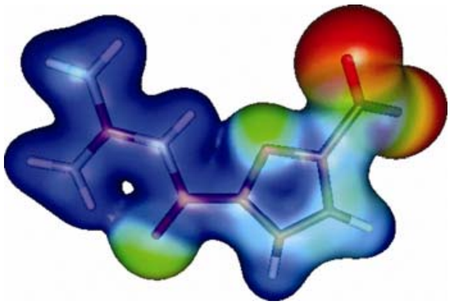
3b



4

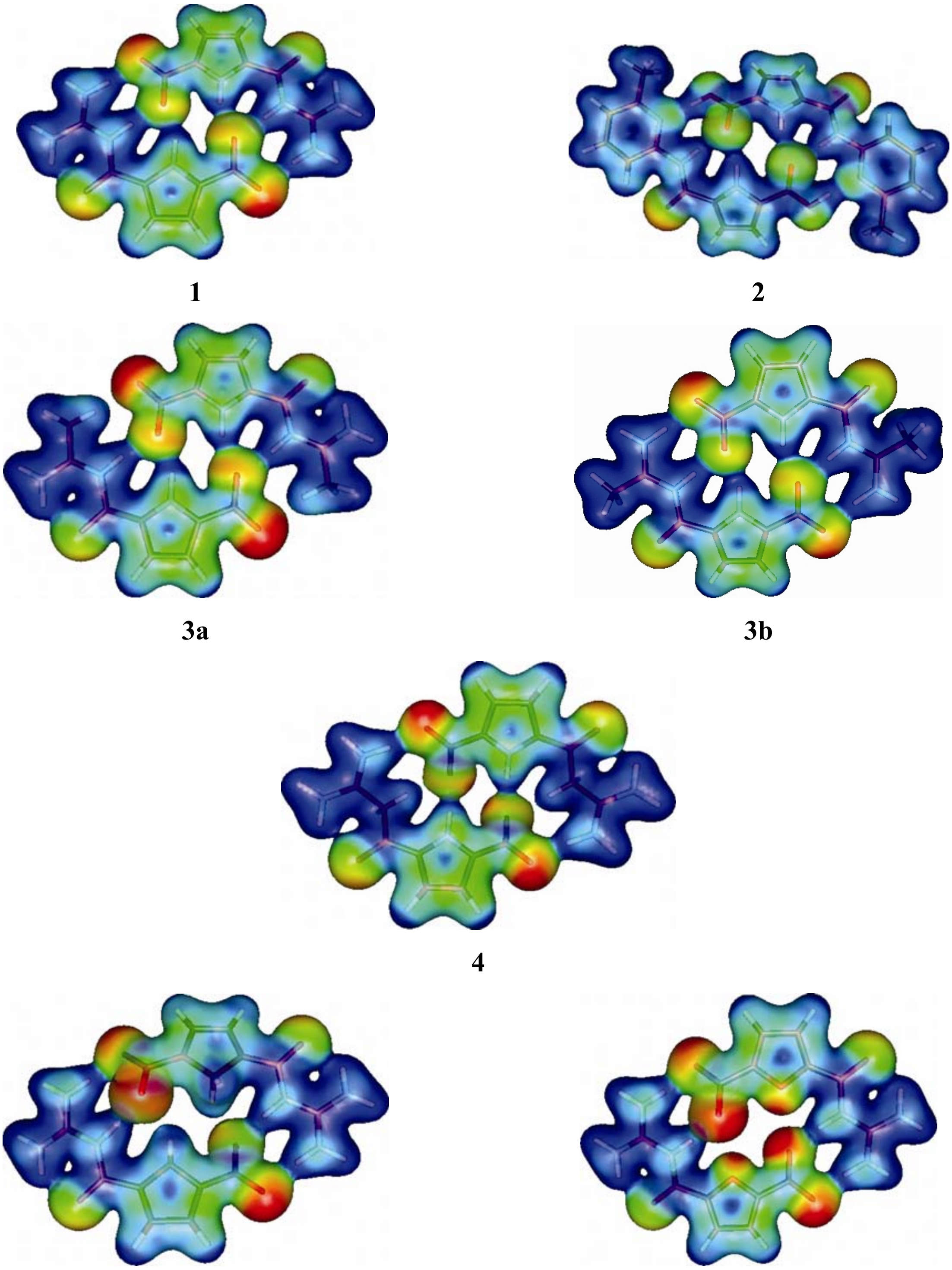


5



6

Figure 19 Electrostatic potential (contour value = 0.02) mapped on the electron density (contour value = 0.015) of the monomers **1**-**6** in gas phase.



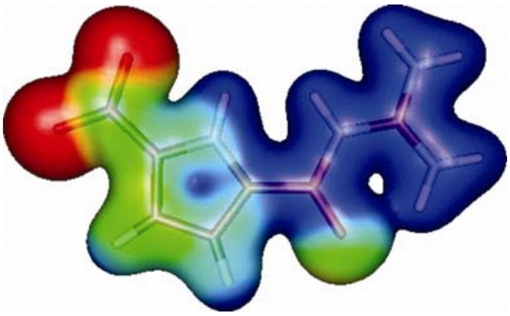
1 **2**

3a **3b**

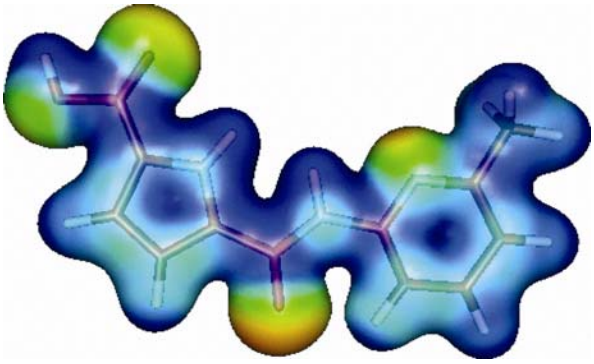
4

5 **6**

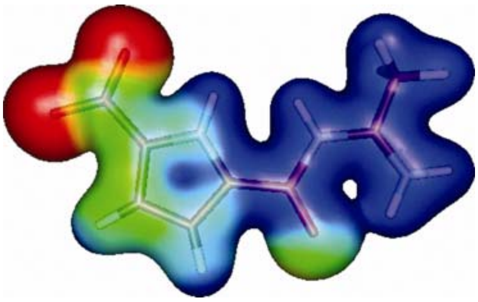
Figure 20 Electrostatic potential (contour value = 0.02) mapped on the electron density (contour value = 0.015) of the dimers **1-6** in solvent.



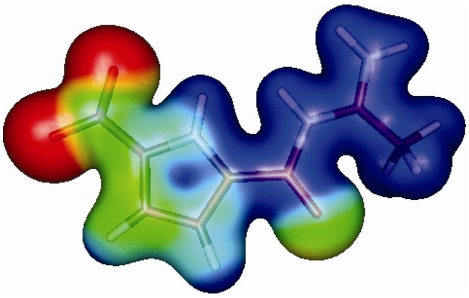
1



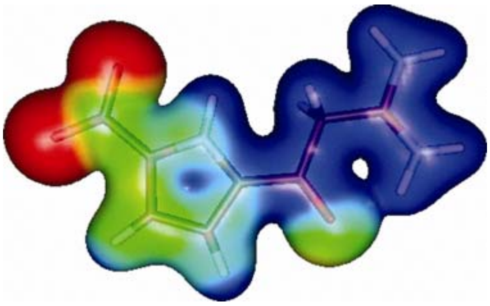
2



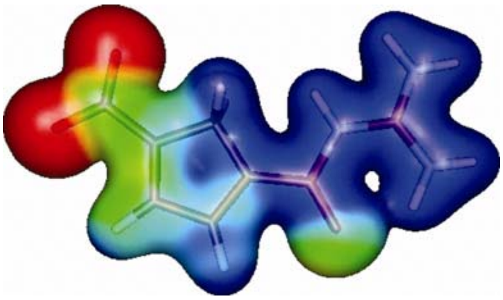
3a



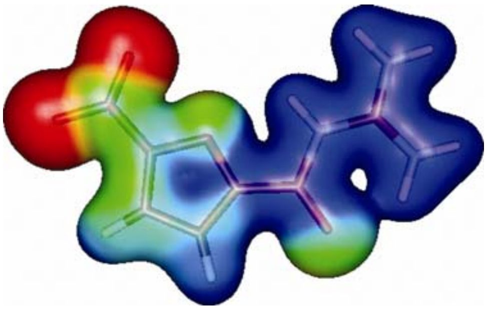
3b



4



5



6

Figure 21 Electrostatic potential (contour value = 0.02) mapped on the electron density (contour value = 0.015) of the monomers **1**-**6** in solvent.

*The neutral analogue (**2**)*

The calculated dimer dissociation energies (gas phase ΔE_{elec} = +116 kJ mol^{-1}, ΔG= +45 kJ mol^{-1}; polar solvent ΔE_{elec} = +48 kJ mol^{-1}, ΔG = -17 kJ mol^{-1}) are much smaller compared to the zwitterionic dimer **1** reflecting the great importance of charge interactions within this kind of dimers. The influence of the solvent on **2** (reduction by about 60 %) is weaker than on the zwitterion **1** for which a reduction by about 80 % is calculated. Such an effect is generally found if salt bridges are compared to neutral hydrogen bonds. Even though the electronic dissociation energy is still negative, a positive free energy of dimerization ΔG is computed in water showing that the hydrogen binding interactions within the dimer are not sufficient to compete with solvation. In polar solvents the neutral analogue **2** is therefore predicted to exist only in form of monomers which was indeed experimentally observed.[14] For the neutral dimer **2** the polarization upon solvation reflected by the electrostatic potential is less pronounced than for **1** (**Figure 18** to **Figure 21**).

*Knock-out analogues **3**-**6**: "switching off" single hydrogen bonds*

From the comparison of the stabilities of **1** and its neutral analogue **2** one could conclude that the main and most important factor responsible for the different stabilities is the zwitterionic nature of **1** and hence the resulting coulomb interaction between the monomers. However, that this interpretation is premature can be seen by taking a look at the knock-out analogues **3a**, **4**, **5** and **6**. These are all zwitterionic species with extensive charge interactions between the monomers (see **Figure 18** and **Figure 20** for the electrostatic potentials) but different hydrogen binding schemes compared to **1**. Despite their zwitterionic nature the calculated stabilities are much lower than for the parent zwitterion **1**. In gas phase the dissociation energies with respect to the zwitterionic monomers of the dimers **3a**, **4**, **5** and **6** lie between +340 kJ mol^{-1} and +364 kJ mol^{-1} corresponding to about 80 % of the value for the zwitterion **1** (BLYP computations).[87] Solvation reduces their dissociation energies to about +48 to +64 kJ mol^{-1} which are only about half of the corresponding dissociation energy of **1** (B3LYP calculations). This clearly demonstrates that the mere charge interaction is not enough to explain the stability of dimer **1**. The strength of the ion pair must be also influenced by the exact nature of hydrogen bond network.

But the computed dissociation energies of **3** – **6** indicate that a second simple model based on just counting the number of formal hydrogen bonds within the binding motifs is not sufficient either. If one neglects the weaker C-H···O^- bonds the knock-out analogues except **3b** all possess 4 instead of 6 formal hydrogen bonds in dimer **1**. Based on the number of H-bonds, one would expect dissociation energies of about +300 kJ mol^{-1} for the gas phase and about +70 kJ mol^{-1} for a polar solvent (≈ 66 % of stability of dimer **1**). On the one hand, the computed dissociation energies between +340 und +364 kJ mol^{-1} for the gas phase show that here the missing of two H-bonds relative to **1** is somehow compensated, *i.e.* the dimers are more stable than expected on the basis of this simple model. On the other hand, for a polar environment the dissociation energy decreases above average with respect to the number of H-bonds. With dissociation energies of about +48 to +64 kJ mol^{-1} the dimers are less stable than expected. Part of this finding can be probably explained by the effect, that in gas phase especially anionic groups have an extremely high energy content and benefit from any kind of molecular interaction, which allows a larger polarization of the negative charge.[88] This stabilizing effect is more or less independent from the exact chemical nature of the monomer and its binding motif. Therefore, the relative importance of any other non-covalent interaction (such as H-bonds or even ion pairs) for the stability of the dimers is reduced in the gas phase. In a polar solution, however, the anion is already stabilized by the solvent molecules. Hence, the relative importance of additional or missing H-bonds increases.

As both simple models (number of coulomb interactions and number of H-bonds) fail to predict the stability of these zwitterionic dimers, a more sophisticated insight into the various non-covalent interactions and their mutual interplay is needed. An estimate of the importance of the individual hydrogen bonds and of cooperative effects can be obtained by comparing the dissociation energies of **3a**, **3b**, **4**, **5** and **6**. Within this series compound **3b** possesses the same hydrogen bond pattern as **1** and the same kind of charge interactions. Therefore, its dissociation energies in gas phase (+443 kJ mol^{-1}) and polar solvent (+111 kJ mol^{-1}) are more or less identical to those of **1**.

As mentioned before all the other „knock-out“ analogues **3a** – **6** have a different H-binding pattern and all lead to a drastically reduced stability of the dimers compared to **1** (and **3b**). But even though their number of formal H-bonds is identical their stabilities differ significantly. This is most likely due to the different nature of the H-bonds and additional secondary electrostatic effects. For example, in dimers **3a** and **4** each carboxylate is bound by one neutral H-bond (from the pyrrole NH) and one ionic H-bond (from the amidinium or guanidinium moiety, respectively), whereas in dimers **5** and **6** both H-bonds are ionic. Furthermore, dimer **3a** exhibits bidentated hydrogen bonds to the inner carboxylate oxygen and the outer oxygen is not bound at all, whereas in dimers **4** – **6** both oxygens are hydrogen bonded.

Let us first compare dimers **3a** and **4**. In the gas phase the dissociation energy of **4** is 29 kJ mol^{-1} higher than the dissociation energy of **3**. Binding of both oxygen atoms by one H-bond each is obviously more efficient than two H-bonds to the same oxygen atom. However, for a polar solvent this trend is reversed. Upon solvation the stability of dimer **4** drops to 13 % of the gas phase value whereas the one of **3a** decreases to about 17 %. As a consequence, in a polar environment dimer **3a** is predicted to possess a higher dissociation energy than dimer **4** (+64 vs. +55 kJ mol^{-1}). This reflects the stronger impact of solvation on the solvent exposed hydrogen bond (bond 1) which is present in dimer **4** but not **3a**. This effect could already be seen in the variation of the bond distances of the parent zwitterion **1** (**Table 3**). For **3a** and **4** additional $C\text{-}H\cdots{}^{-}O_2C$ interactions have to be considered. Based on computations for $CH_4\cdots{}^{-}Cl$ (≈ 10 kJ mol^{-1})[89] we estimate these effects to about 10 kJ mol^{-1} for the gas phase and about 2-3 kJ mol-1 in a polar solvent.[90,91] Higher values than for $CH_4\cdots{}^{-}Cl$ could be assumed since the neighboured guanidinium group increases the acidity of the CH_3 or CH_2 group. A smaller value could be estimated since the charge of the carboxylate group is smeared over the whole unit. This effect will be enhanced by the interactions between the guanidinium group and the carboxylate group. In all respect the $C\text{-}H\cdots{}^{-}O_2C$ interactions can be considered to be much smaller than the effects discussed above.

Surprisingly, for gas phase the calculations for the cyclopentadienyl derivative **5** predict a dissociation energy of +364 kJ mol^{-1} which is equal to the amidine derivative **4**. One would expect a higher dissociation energy for **5** than for **4** since the H-bond pattern of **5** contains two ionic H-bonds instead of one neutral and one ionic one for **4**, and furthermore, the binding motif of **5** allows attractive secondary interactions. Additionally, **5** could be stabilized by an attractive interaction between the CH_2 group of the cyclopentadienyl unit and the carboxylate group. Obviously, this possible advantage is probably cancelled out to some extend by other factors. One possibility could be geometric strain in **5**. Additionally the $C\text{-}H\cdots{}^{-}O_2C$ interactions could be decreased since the charge of the carboxylate group is smeared out as discussed for **4**. For a polar environment **5** (decrease to 16 % of the dissociation energy in gas phase) becomes more stable than knock-out analogues **4** and **6** as expected for its binding motif with two ionic H-bonds and no further destabilizing secondary interactions. “Knock-out” analogue **3a** has a similar stability in water as **5**, despite its less efficient binding motif. This again probably reflects the fact that the influence of the solvent on the stability of the various H-bonds depends on their accessibility.

The furan derivative **6** exhibits the same H-bond pattern with two ionic H-bonds as the cyclopentadienyl derivative **5** and could have been expected to be equally stable. However, although the outer hydrogen bond is even shorter than in **5**, repulsive secondary electrostatic effects connected with oxygen lone pairs of the furan oxygen and the bound carboxylate reduces the dissociation energies about 19 kJ mol^{-1} in gas phase and 13 kJ mol^{-1} in a polar solvent, respectively. This repulsive interaction, which is also nicely reflected from the electrostatic potentials of **6** (**Figure 18** and **Figure 20**), makes dimer **6** even slightly less stable than the neutral analogue **2**. 33

The discussion so far was restricted to the mere electronic dissociation energies to analyze the intrinsic stabilities of the various binding motifs. Thermodynamic contributions leading from dissociation energies to the corresponding enthalpies and free energies considerably reduce the stability of all dimers (**Table 5**) with respect to their monomers. In gas phase the reduction is about 25 %. For a polar solvent the relative importance of the corrections is considerably stronger due to smaller absolute dissociation energies.[91] As shown in **Table 5** within a polar solvent the absolute values of the $T\Delta S$ term vary between 50 and 66 kJ mol^{-1}. The variations in ΔS among the series arrive mainly from the contributions of vibration (ΔS_{vib}) while the corrections due to translation and rotation are very similar (see supplementary material). One could expect that the variations mainly correlate with the magnitude of the binding interaction between the monomers since as a result of this binding various low lying bending vibrations of the monomers are hindered considerably. However, such a correlation is not found as most prominently shown by a comparison between **1** and **3b**. The dissociation energies of both compounds are very similar but their $T\Delta S$ terms differ by 11 kJ mol^{-1} ($\approx$ 20 %). This may result from the rigidity of the molecules studied here. The size of $T\Delta S$ is therefore probably determined by the reorganization of the whole electronic structure upon dimerization. The sum of the resulting subtle changes in all monomer bonds leading to various slight changes in many vibrations then determines the changes in ΔS_{vib}. It is important to note that even the thermodynamic corrections change the trend in the predicted stabilities to some small extent. However, considering the theoretical approximations differences smaller than 5 kJ mol^{-1} are too small for sound predictions.

On the basis of the computed ΔG values we see that besides **1** and **3b** only dimer **5** is expected to form stable dimers in water. For all other analogues the dimerization in water is endergonic. And even for dimer **5** the dissociation free energy is rather small (ΔG = +13 kJ mol^{-1} for a hypothetical infinite dilute solution), probably not allowing its experimental detection due to the salt effect mentioned above, which will further decrease the stability in macroscopic samples. Apart from the fact that **5** due to its tautomeric instability can never be studied experimentally.

4.1.5 Conclusions

The present study investigated the molecular interactions in 5-(guanidiniocarbonyl)-1*H*-pyrrole-2-carboxylate by computing various “knock-out” analogues in which single hydrogen bonds are switched off. The influence of a polar solvent is tested and the computations clearly show that simple models fail to predict the stability of the knock-out analogues.

The analysis of the “knock-out” analogues indicates that the following interactions seem to be important: a) charge interactions within ionic hydrogen binding networks are significantly more stable than simple point charge interactions, b) additional neutral H-bonds further stabilize the dimer but less efficiently than the ionic ones, c) solvation affects H-bonds differently depending on their accessibility and d) secondary electrostatic interactions further modulate the stability.

The comparison of 5-(guanidiniocarbonyl)-1*H*-pyrrole-2-carboxylate dimer with the arginine dimer in gas phase revealed a final important effect: The zwitterionic monomer of 5-(guanidiniocarbonyl)-1*H*-pyrrole-2-carboxylate has a considerably higher energy content than the zwitterionic form of arginine. The strong stabilization of the latter arises from the interaction of the charged ends which is prevented in 5-(guanidiniocarbonyl)-1*H*-pyrrole-2-carboxylate due to its stiffness. Transferring this knowledge to the situation in a polar medium the high stability of the 5-(guanidiniocarbonyl)-1*H*-pyrrole-2-carboxylate dimer (*e.g.* in comparison to arginine) seems to result also from the monomers which are less stabilized. This finding suggests a new approach for the optimization of supramolecular self-assembly. To have a strong dimerization affinity the monomers should be as rich in energy as possible, *i.e.* this principle does not only focus on the number and strengths of the bonds in the dimers but tries to enforce this effect by thermodynamically high lying monomers.

4.2 Arginine as Model System for Guanidinium-Carboxylate Interactions

Amongst all naturally occurring amino acids arginine is in the very focus of interest for various reasons. Due to the high basicity of the guanidine group situated in the side chain, its protonated form plays an important role in protein chemistry allowing the formation of strong salt bridge interactions with carboxylates or phosphates.[92] Arginine is therefore present in many reactive centres of enzymes and plays also an important role in secondary and tertiary structure formation.[93]

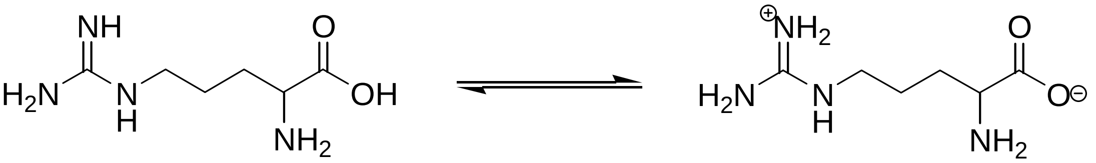


Figure 22 Tautomeric forms of neutral arginine. Left: canonical monomer, right: zwitterionic monomer.

Amino acids form stable zwitterions in aqueous solution whereas their canonical tautomers are strongly favoured in gas phase. An exception could be again arginine since the strong proton affinity of the guanidine group could outweigh the energy necessary for charge separation (see **Figure 22**). The question for the tautomeric form of the global minimum of arginine in gas phase has therefore been discussed widely by theoretical studies.[64,77c-d,94] An experimental study trying to shed light on this problem was performed by Chapo et al. employing infrared cavity ringdown laser absorption spectroscopy (IR-CRLAS). They identified two peaks at 1666 cm^{-1} and 1693 cm^{-1} which were assigned to carbonyl stretches of the carboxylic acid group present in the canonical form.[79] Since symmetric and asymmetric stretches of the carboxylate group in the zwitterionic arginine could not be found in the expected range, they concluded that in gas phase the canonical form of arginine is mainly populated. However, Rak et al. doubted the interpretation of the measured spectrum on the basis of new theoretical findings which predicted the carbonyl bands to occur in a region that has not be recorded.[77c] Nevertheless, on the basis of the computed relative energies also their study predicts that the canonical form should dominate in the gas phase by 7 kJ mol^{-1}.

Due to its ability to form strong salt bridges, arginine is also an ideal model system to study guanidinium-carboxylate interactions.[77a-b,78,95] Moreover, the strong non-covalent binding interactions of the guanidinium moiety with anionic groups is the basic concept of a research field trying to mimic biological receptor systems in order to improve ligand-receptor interactions and to understand molecular recognition processes.

4.2.1 Conformational Analysis in Gas Phase – The Quest for the Global Minimum

4.2.1.1 Introduction

Computational investigations are often hampered by the multiple-minimum problem appearing for all molecules which possess many rotatable single bonds. Due to the small amount of energy necessary for the internal rotation around such bonds a large number of local minima occur which are comparable in energy. The resulting difficulties arise since reliable theoretical studies require at least the knowledge of the global minimum. One example is the computation of aggregation energies which are of ample interest for many different fields of research *e.g.* supramolecular chemistry, protein-protein interaction etc. Exact predictions of aggregation energies for example presuppose the exact energies of the global minima of both the underlying monomers and dimers. The determination of free energy differences even needs information about all low lying conformers being populated for a given temperature.

Therefore, one has to be aware of such problems when studying non-covalent interactions between arginine-like structural motifs occurring in biological and artificial host guest complexes.[3,4,6,12,96] Chapter 4.1 disclosed the effects causing the large variations in the dissociation energies of different complexes which are presumably mainly bonded through the same interaction, namely ionic interactions between guanidinium and carboxylate moieties. This study uses arginine as reference system and thus a highly accurate computation of its dimerization energy is necessary.

Besides the conformational state also the tautomeric state of arginine dominating the gas phase is still under dispute.[64,77c-d,79,94] Like all other naturally occurring amino acids arginine forms stable zwitterions in aqueous solutions – in that case by an intramolecular protonation of the side chain guanidine. However, going from solution to gas phase conditions the canonical conformers of amino acids are strongly favoured. As already mentioned, a possible exception could be arginine which strong proton affinity could outweigh the energy necessary for charge separation. An addition of single water molecules or counterions can already strongly stabilize the zwitterionic structure as theoretical studies could prove.[80,83b] In gas phase arginine is also capable to form clusters consisting of zwitterionic monomeric units which form discrete dimers or higher aggregates.[77a,97] As already noted, a detailed understanding of the non-covalent bonding pattern of these aggregates is of special interest since they reveal various contributions (salt bridge, hydrogen bridge, cooperative effects…) which are similar to those occurring in guanidinium based carboxylate receptors (see review by Schug and Lindner).[6]

Starting from arginine monomer and dimer structures given in the literature extensive conformational searches were performed in order to ensure that no minima are missed. These computations revealed a new global minimum which possesses a completely different structural arrangement than the minimum given in the literature. While all already known structures are stabilised only by directed hydrogen bonds, the new arrangements also allow additional stacking interactions. This structural motif is also found in many other low lying conformers which underlines its importance. Also these local minima were overlooked so far. Beside the new minima the various pitfalls in the conformational search which prevented the previous studies from finding these new minima are revealed.

4.2.1.2 Computational Details

The success of a conformational search depends on the trustiness of all methods employed in the various steps since a failure in a previous step (normally performed with a less accurate approach) can not always be corrected by the subsequent steps for which normally better methods are used. If for example the force field based generation of conformers only provides structures far away from the global minimum the subsequent refining procedures (geometry optimizations on higher theoretical levels) are not able to straighten out this error as the used optimization routines only lead to the next local minimum. Since conformation algorithms and force fields are often biased, in most cases the first step is performed with various combinations. For a validation of conformational search algorithms the stochastic Monte-Carlo Multiple Minimum (MCMM)[39] approach, the Pure Low Mode (LM) approach,[40] the Mixed MCMM/Low Mode approach and the Systematic Unbound Multiple Minimum (SUMM)[38] approach as implemented in the MacroModel8.0 program package[98] have been used for scanning the conformational space. All conformational searches took between 2000 to 5000 steps and were repeated from different starting structures if necessary. For the canonical monomer the number of generated conformers within an energy range of 50 kJ/mol exceeded several hundreds conformers so that the different structures were clustered by the XCluster program[99] based on atomic RMSD (root-mean-square distance) differences of all atoms. The choice of an appropriate force field was made dependent on the protonation state of the respective isomer. For the zwitterionic species the OPLS-AA force field[100] as well as the MMFF94 force field[101] yielded reasonable structures, however, for the canonical structure only the MMFF94 force field was found to give the most promising results (see Chapter 4.2.1.3). The lowest lying energy structures were then pre-optimized on a B3LYP/TZVPP level using the TURBOMOLE program package.[35b,e,65,102] Within this screening step DFT was used since it is well known to describe many properties with an excellent cost-benefit value.[103] Afterwards the most promising monomer structures were fully optimized on RI-BLYP, B3LYP and RI-MP2 level of theory employing either a TZVPP basis set or with additional diffuse functions on the carbonyl atoms in order to describe the diffuse shape of electron of the carboxylate atoms in the zwitterionic conformers properly (denoted as TZVPP+ in the following). Therefore, the basis set was enlarged by 1s and 1p primitive uncontracted basis functions with an exponential coefficient of 0.068, whereas for the auxiliary basis sets the exponent was doubled (0.136).[104] On the MP2 optimized monomer conformers CCSD(T) calculations using the MOLPRO program package[105] were performed employing a cc-pVDZ basis and an aug-cc-pVDZ basis for the oxygen atoms.[106] For CCSD(T) the employment of larger basis sets were not feasible due to hardware and software restrictions.

All optimized structures were characterized by harmonic frequency analysis and thermodynamic corrections which were obtained with TURBOMOLE on a RI-MP2/TZVP level. The free energies were calculated with a scaling factor for the wavenumbers of 0.937.[107] The IR spectra were simulated employing a Gauss fit for the line spectra. The contributions from all excitations were added according to the following equation:

$$\varepsilon(E) = \frac{\sum_i I_i \cdot \exp\left\{-\left[-\frac{(E-\Delta E_i}{2\sigma}\right]^2\right\}}{\sqrt{2\pi\sigma}} \qquad \text{eq. 4.2.1}$$

with σ as the full width of at half maximum (here σ = 0.001) and ΔE_i (in eV) and I_i as the calculated excitation energies and intensities, respectively.[108]

4.2.1.3 Force-Field Validation

The initial generation of conformers is mostly performed on a force field level as this step requires very high numbers of computations. Therefore, reliable force fields are essential since poor starting geometries for subsequent optimizations on higher levels may not lead to the global minimum. For the zwitterionic arginine the conformational search (see below) employing the OPLS-AA force field resulted directly the thitherto known lowest zwitterionic conformer Z3. It is depicted in Figure 1 which gives the geometrical arrangements of the most important conformers obtained from RI-MP2/TZVPP+ optimizations. It should be noted that force field and MP2 geometries do not differ considerably. A second search employing the MMFF94 force field often gave the same or similar structures. Comparing both force fields it can be seen that the OPLS-AA force field mostly favours structures with directed hydrogen bonds. This effect results since the OPLS-AA force field, which was primarily designed to represent conformational energies and nonbonded parameters like hydrogen bonding of peptides and proteins properly[100], lays more stress on the ionic interactions of the charged ends than the MMFF94 force field. Nevertheless, also stacked structures, *e.g.* the ZW19 structure are obtained. The MMFF94 force field yields both structure types, however, the stacked conformers are favoured in the energy listing (see below). The quality of the MMFF94 force field was already discussed by Boyd and coworkers.[109] Based on X-ray geometries as reference quantities they found that the MMFF94 force field yields often more accurate results than MP2-computations. That a carefully parameterized force field can give more accurate results than high level *ab-initio* procedures is known, especially for weak dispersion interactions which are very difficult to obtain by *ab-initio* procedures but can be parameterized quite well. However, using X-ray structures to determine the accuracy seems to be very problematic. The determination of X-ray structures includes severe optimization cycles which rely on force-field approaches. As a consequence of these fitting procedures the X-ray structures should be more similar to the force field than to *ab-initio* computations.

It should be noted that the conformational search for the zwitterionic species is less delicate due to the strong ionic interaction which reduces the number of low lying conformers drastically. The preference of directed hydrogen bonds by the OPLS-AA force field leads to severe problems in the description of the aggregation behaviour of arginine-like zwitterionic monomers.

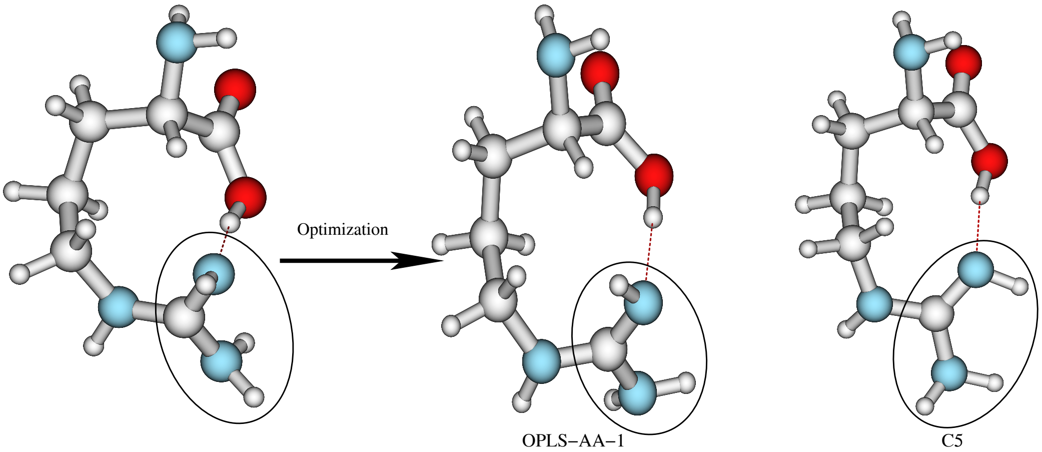


Figure 23 The lowest energy canonical conformer OPLS-AA-1 generated by the Mixed MCMM/LowMode algorithm in conjunction with the OPLS-AA force field (left). Geometry optimization on DFT level (B-LYP/TZVPP+) gives a conformer which differs mainly by the geometry of the guanidine group (middle). It deviates from the C5 conformers (right) by the orientation of the terminal amino moieties.

Such a dominating interaction does not exist for the canonical arginine conformers so that the conformational search generates a lot more conformers. A comparison between the two force fields MMFF94 and OPLS-AA showed that the latter indeed gives as first hit a structure resembling the minimum structure C5 in regard of the alkyl backbone but differing in the guanidine geometry. Here, the hydrogen of the proton acceptor nitrogen is displaced nearly perpendicular to the molecular plane of the guanidine group on force field level giving after *ab initio* geometry optimization the unfavourable tautomer OPLS-AA-1 shown in **Figure 23**, which lies 23 kJ/mol higher than the minimum structure C5 (RI-BLYP/TZVPP+).

The lowest conformers of the MMFF94 force field are structurally different to the literature minimum conformer C5, which itself could not be found since it does not represent a local minimum in any of the force fields. A similar conformer, N20, was found on rank 20 differing only by +2 kJ/mol (RI-MP2/TZVPP+) in energy from C5 (**Figure 24**). Both deviate by the orientation of the amine hydrogen atoms of the guanidine group. An identical backbone including the correct representation of the guanidine group was found for a conformer listed on rank 392. However, a small rotation of the α-amino group results in a relative potential energy of +5 kJ/mol with regard to the lowest energy conformer C5 (RI-MP2/TZVPP+).

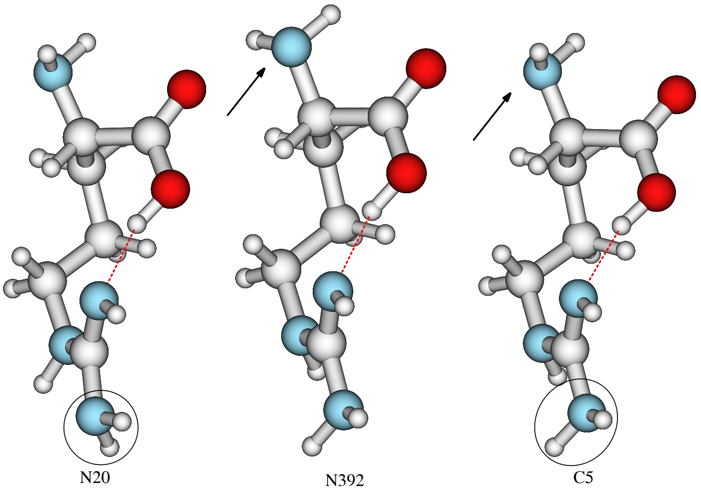


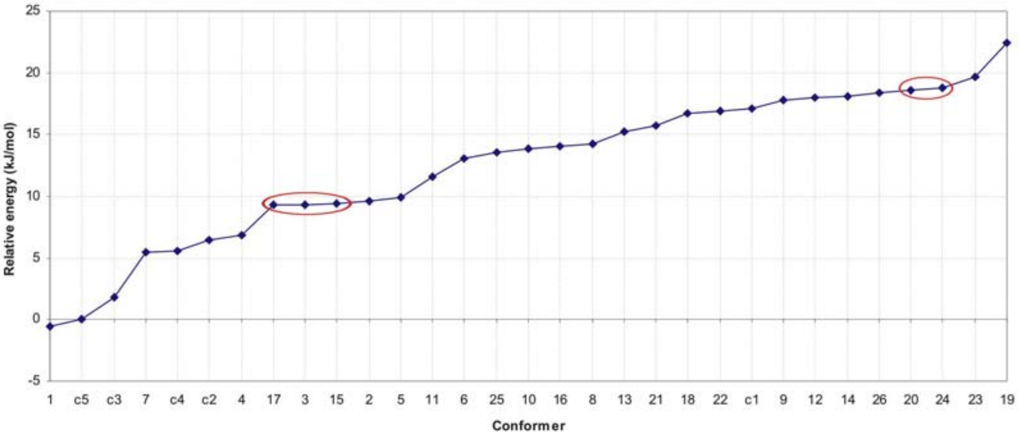
Figure 24 RI-MP2/TZVPP+ optimized geometries of the canonical conformers N20, N392 and C5. For N20 the structural difference to C5 is restricted to a flip of the amine hydrogens of the guanidine group, whereas in case of N392 a rotation of the α-amino group is found.

4.2.1.4 Conformational Search Algorithms

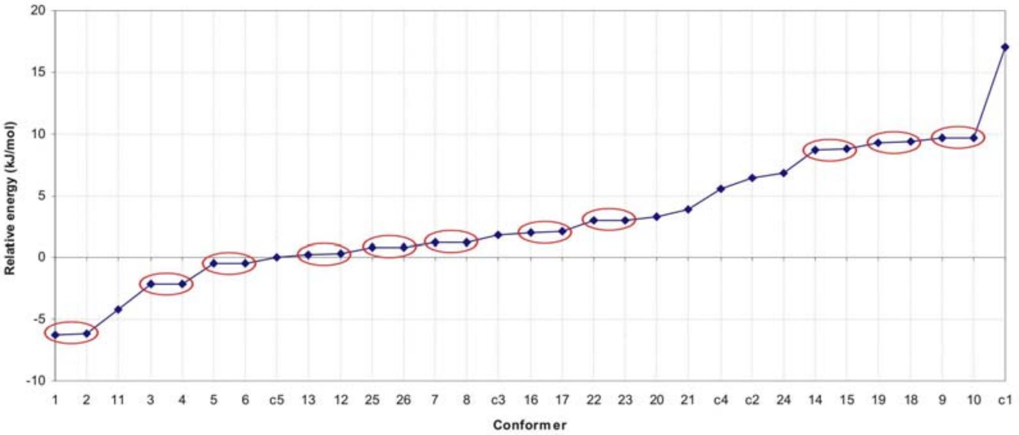
The conformational search algorithms used throughout this work follow different strategies. The Systematic Unbound Multiple Minimum (SUMM) approach[38] obtains new conformers by a systematic variation of the torsion angles with an increasing resolution. The Pure Low Mode (LM)[40] approach uses the second derivates from a given geometry and follows the lowest mode to escape the minimum. This approach possesses a local character, *i.e.* it only searches the near surrounding of a given minimum. The Monte-Carlo Multiple Minimum (MCMM) approach[39] represents a stochastic variation of torsion and therefore possesses a more global character. A combination of MCMM and LM should comprise both local and global characters.

The strong ionic interactions between the guanidinium moiety and the carboxylate group do not only reduce the number of possible low lying conformers of the zwitterionic structure. They also increase the well depth of the respective minima. The method of choice for searching this PES effectively was found to be the Systematic Unbound Multiple Minimum (SUMM) method. Its benefit results from the fast algorithm which varies the dihedrals systematically. Unlike stochastic methods, it is not retracing its search path resulting in a more rapid convergence in comparison to random search algorithms.

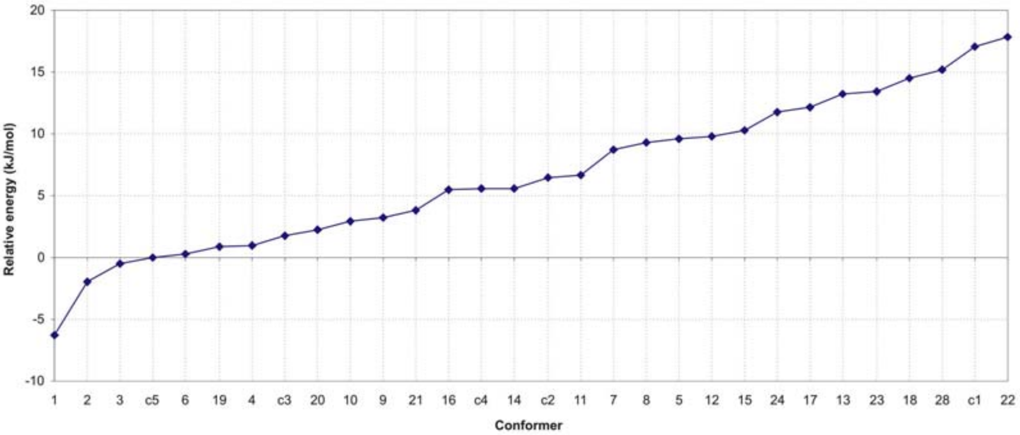
Since the potential energy surface (PES) of canonical arginine is considerably smoother, sensitive search strategies are required in order to detect the numerous local minima. Due to the coarse grid in its first step, the SUMM approach turned out to be inefficient for such a situation. The success of the remaining algorithms is compared in **Figure 25** which gives the relative RI-MP2/TZVPP//B3LYP/TZVPP energies of the generated conformers with respect to structure (C5) given by Rak et al. For a better comparison only the MMFF94 conformers are given. As one can see in **Figure 25**the stochastic MCMM algorithm produces only one conformer that is lower in energy than C5. All others are clearly higher in energy. Taking into account that the various low lying minima possess a quite similar structure (see **Figure 26**), this indicates that the global character of this search impedes the finding of local minima lying close to each other on the PES. The conformational search in using the LM approach gave many conformers which possessed the same electronic energy after DFT geometry optimization (see **Figure 25**). A closer inspection showed that they represent the L and D configurations of the respective conformers, *i.e.* the LM algorithm overcomes the high energy barrier separating both enantiomers. This shows the strong local character of the LM approach since two enantiomers are more similar with respect to the internal coordinates than two conformers. However, this behaviour is unfavourable since in most conformational searches only one enantiomer is of interest. The mixed MCMM/LM algorithm combining local and global elements yields new low lying energy conformers which represent different minima also on the DFT PES. In summary, the mixed MCMM/LM approach in conjunction with the MMFF94 force field turned out to be the best choice for generating new and unique low-energy conformers that can be handed over to more sophisticated *ab initio* methods.



MCMM



LowMode



mixed MCMM/LM

Figure 25 Comparison of MP2/TZVPP//B3LYP/TZVPP energies (in kJ mol-1) of conformers generated by different conformational search algorithms using the MMFF94 force field with the previously published structures by Rak et al. The new conformers are termed according to their rank in the force field energy listing. Conformers which collapsed to a single minimum during the DFT optimization are encircled. Structures being identical to those given by Rak et al[77c] are indicated by a C (C1, C2, C3, C4).

4.2.1.5 Electronic Structure Optimizations

A strong uncertainty results from the differences found in the energy ordering predicted by the force fields and by the *ab-initio* pre-optimisation. To minimize the unreliability of our approach a great amount of conformers were taken into account in order to ensure that all possible lowest level conformers are included. The selection of new candidates was done manually for the purpose to reduce the amount of conformers which where treated in the pre-optimization to 30 per conformational search. The pre-optimization consists of a geometry optimization on the B3LYP/TZVPP level followed by single-point MP2/TZVPP computation. The single-point computations were necessary since DFT is well known for an excellent description of various properties[110] and especially the interactions between charged species[111], but often DFT does not encounter all important interactions properly (see Chapter 3.1.4). For the most promising conformers additional geometry optimizations employing DFT and MP2 methods in conjunction with the TZVPP+ basis set were performed. The RI-MP2/TZVPP+ optimized geometries for both the canonical and zwitterionic conformers of arginine are given in **Figure 26**. The corresponding relative energies and bond length and angles are shown in **Table 6** and **Table 7**, respectively. The two zwitterionic conformers have been selected from conformational searches employing the OPLS-AA (ZW19) as well as the MMFF94 force field (ZW1), whereas all new canonical structures resulted from a mixed MCMM/LM conformational search based on the MMFF94 force field. As noted before the structures C5 and Z3 were already described in the literature.[77c]

Furthermore, the multiple hydrogen binding patterns within the stacked structures comprehend hydrogen bonds of the type N–H···O from the guanidine to the protonated hydroxyl oxygen (N1, N3, N4). Such hydrogen bond types have not been mentioned for the arginine monomer yet. Nearly all new canonical structures also posses N–H···N bonds from the α-amino group to the guanidine rest (N1, N3, N4, N6, N19) having a back-donating-like character. N6 and N19 additionally show weak C–H···C hydrogen bonds (2.4 - 2.5 Å) stabilizing the backbone.

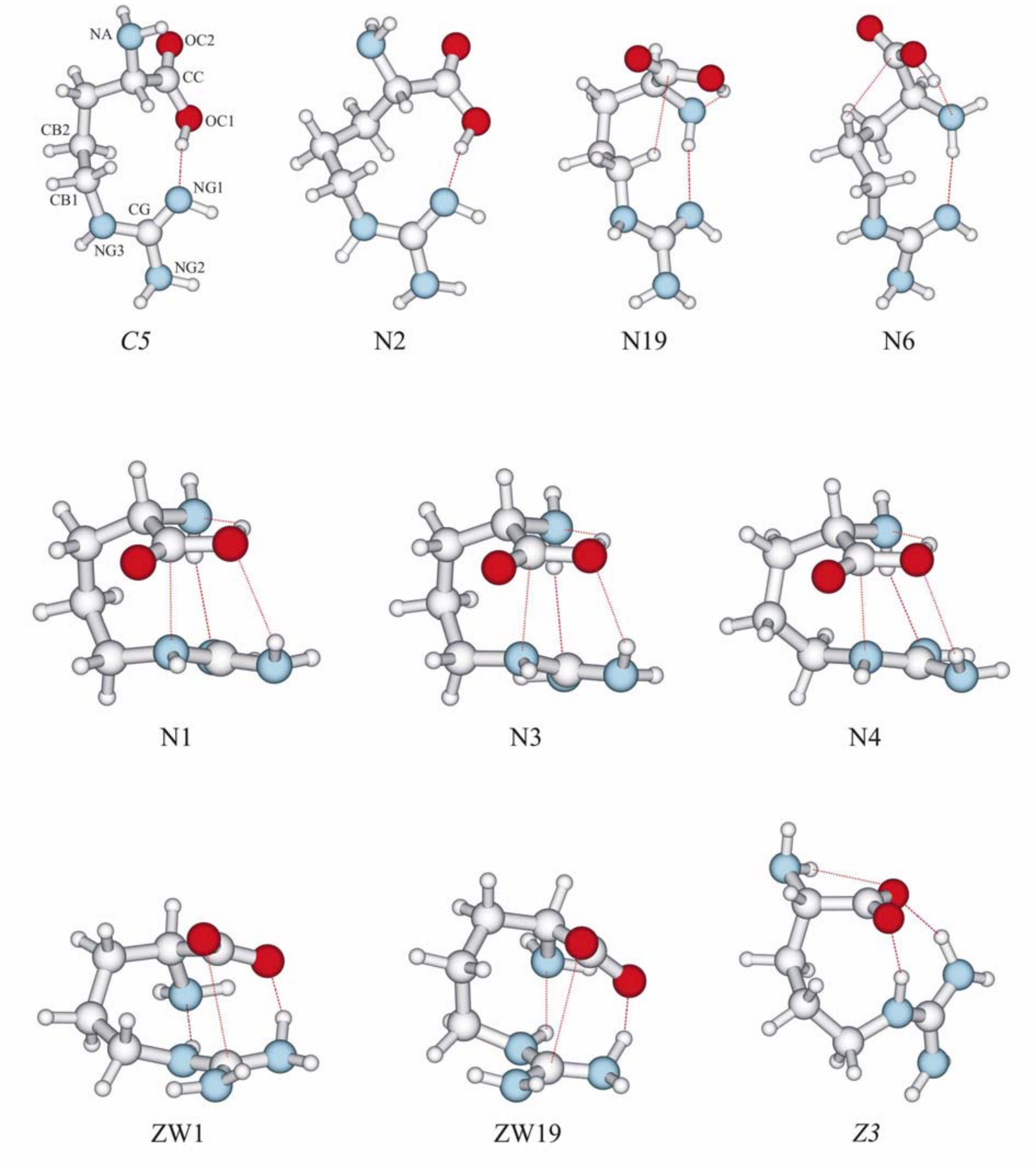

NA
OC2
CC
OC1
CB2
NG1
CB1
CG
NG3
NG2
C5
N2
N19
N6
N1
N3
N4
ZW1
ZW19
Z3

Figure 26 RI-MP2/TZVPP+ optimized geometries for canonical and zwitterionic conformers of arginine monomer. The nomenclature of the new conformers is based on the ranking after the conformational search and in case of the lowest reference conformations (C5, Z3) the names are adopted from the literature.

Table 6 Relative electronic energies in kJ/mol of the fully optimized conformer geometries in dependence of the level of theory

Method	N1	N2	N3	N4	N6	N19	*C5*	ZW1	ZW19	*Z3*
B3LYP/TZVPP	0.0	-8.6	+4.5	+4.9	-3.6	-1.1	-6.8	-	-	-
MP2/TZVPP[1]	0.0	+4.3	+5.8	+7.3	+6.6	+7.2	+6.3	-	-	-
B-LYP/TZVPP+	0.0	-10.6	+4.2	+4.4	-6.3	-2.6	-9.4	+1.8	-0.8	-0.8
B3-LYP/TZVPP+	0.0	-9.5	+4.6	+5.0	-4.0	-1.2	-7.4	+4.3	+3.5	+2.8
RI-MP2/TZVPP+	0.0	+5.5	+5.6	+7.1	+8.5	+8.7	+8.3	+7.1	+6.7	+15.7
MP2/cc-VDZ[2]	0.0	+8.7	+8.4	+8.4	+15.4	+15.5	+11.2	-2.8	-2.5	+11.4
CCSD(T)/cc-VDZ[2]	0.0	+7.4	+8.2	+7.7	+14.9	+14.2	+9.2	+1.6	+1.6	+15.0
CCSD(T)/ extrapol.[3]	0.0	+4.2	+5.4	+6.4	+8.0	+7.4	+6.3	+11.5	+10.8	+19.3

---

[1] Relative single point energies on B3LYP/TZVPP optimized geometries

[2] Relative single point energies on RI-MP2/TZVPP+ optimized geometries

[3] Δ E(CCSD(T)/extrapol.)=ΔE(CCSD(T)/cc-VDZ)+{Δ E(MP2/TZVPP)-ΔE(MP2/cc-VDZ)

Table 7 Hydrogen bond lengths and selected intermolecular distances of canonical and zwitterionic tautomers of arginine (in Å) calculated on RI-MP2/TZVPP+ level of theory. For the labelling of the atoms see **Figure 26**.

Structure	OC1–H···NG1	CB–H···CC	NA–H···NG1	NG2–H···OC1	OC1–H···NA	NG3···CC
C5	1.79	-	-	-	-	-
N2	1.76	-	-	-	-	-
N1	-	-	2.32	2.48	1.82	2.70
N3	-	-	2.53	2.36	1.83	2.73
N4	-	-	2.32	2.48	1.83	2.83
N6	-	2.53	1.98	-	1.81	-
N19	-	2.42	2.12	-	1.83	-

Structure	NG1–H···OC1	NG3–H···OC2	NG3–H···NA	NA–H···OC2		OC2···CG
Z3	1.63	1.63	-	2.32		-
ZW1	1.55	-	1.91	-		3.03
ZW19	1.55	-	2.01	-		3.10

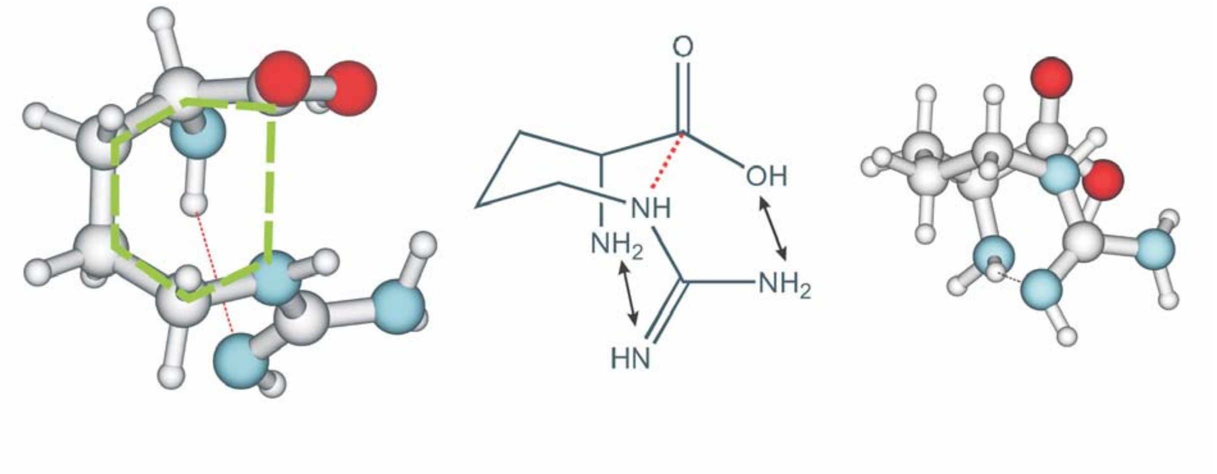


Figure 27 Chair-like conformation of N1

All structures show strong interactions between the guanidine and the carboxyl moiety. However, while the already known structure type (*e.g.* N2, C5, Z3) is solely stabilized by directed hydrogen bonds, the new structure motif shows strong stacking interactions between the terminal groups leading to a parallel orientation of the carboxyl and the guanidine moiety. It is noteworthy that such stacked arrangements appear in canonical as well as zwitterionic conformers.

One reason why these structures were overlooked before is shown by **Table 6** revealing a strong dependency of the relative energies on the level of theory, *i.e.* on basis set size and method. In case of the canonical conformers the relative energies of DFT optimized conformers are remarkably higher (10-15 kJ mol^{-1}) if stacked structures are considered, whereas structures showing classical O–H···N directed hydrogen bonds like N2 or C5 are predicted to be the most stable. This energy order is inverted if MP2/TZVPP single-point computations are performed on top of the B3LYP/TZVPP geometries. Such computations already predict the stacked N1 structure to be the most favourable conformer. If geometry optimizations are performed on the MP2/TZVPP+ level the relative energies change to some extend but the main trends remain. In order to estimate the accuracy of the MP2 results single-point CCSD(T) calculations have been performed on top of the optimized MP2 geometries. Unfortunately, these calculations were only feasible with a correlation-consistent double-ξ basis set. Therefore the energies for a larger basis set have only been extrapolated using the basis set dependency of the MP2 calculations. Especially the N6 and the N19 conformer both showing in principle the same binding pattern are affected by an enlargement of the basis by about 7 kJ mol^{-1}. On this level of theory all directed bonded structures (C5, N2, N19, N6) are between 6-9 kJ mol^{-1} higher in energy than the new global minimum structure N1.

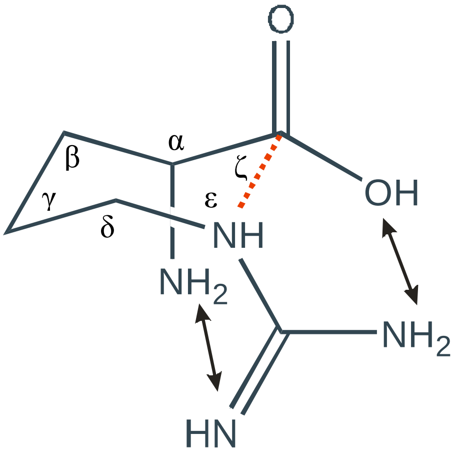
Regarding the zwitterionic conformers it has to be noted that for some conformers (*e.g.* ZW19) DFT is even not able to predict the structure correctly. BLYP and B3LYP calculate the zwitterionic structures to lie about 10-12 kJ mol^{-1} above the lowest canonical structure predicted by DFT being N2. Additionally, DFT wrongly favours the directed hydrogen bonding pattern of Z3 with respect to the stacked structures of ZW1 and ZW19. Geometry optimizations on MP2/TZVPP+ level yield that the lowest zwitterionic conformer is only 7 kJ mol^{-1} less stable than the lowest canonical structure on this level of theory, which is N1. Compared to the N2 conformer this difference decreases even to 1 kJ mol^{-1}. Furthermore, MP2/TZVPP+ computations predict the stacked structures to be energetically more favourable than the structure Z3 by about 8 kJ mol^{-1}. CCSD(T)/cc-DVZ calculations amplify this unbalance making the zwitterionic conformers ZW1 and ZW19 now almost as stable as the canonical structures (+1.6 kJ mol^{-1}), whereas the relative energy of Z3 remains nearly unchanged. However, the stacked structures show a very strong basis set dependency and thus these structures are proportionally stronger destabilized by the basis set extrapolation than Z3 and are therefore estimated to lie +11 kJ mol^{-1} higher than the lowest canonical structure. Since CCSD(T) should depend more strongly on the basis set size than MP2 we expect that the correct energy difference is slightly higher.

The reason for the underestimation of stacked structures by DFT methods must be discussed separately for the canonical and zwitterionic structures since both discrepancies seem to result from different underlying effects. Figure 5 shows the chair-like structure of the canonical N1 conformer that forms an open six-membered ring. The structure is only slightly strained as can be seen in the angles (see **Table 8**). The main stabilization of the structure results from the OC1–H···NA bond. It is only slightly longer than the OC1–H···NG1 bond which stabilizes the structures N2 and C5. Additional stabilizations arise from weak dispersive interactions between the guanidine nitrogen and the carboxylate carbon atom. With a distance of 2.70 Å the hydrogen bond is about 0.6 Å shorter than the sum of the van der Waals radii.[112] Moreover, as a consequence of the stacked orientation the arginine is able to form three additional hydrogen bonds of which two are real hydrogen bonds between the carboxylic acid and the guanidine moieties. Although each bond is remarkably larger than the hydrogen bond in N2 or C5 they also contribute slightly to the stabilization. From the distances of the hydrogen bonds (**Table 7**) also N6 could be expected to be lower in energy than N1. However, the missing stacking effects between the guanidinium and the carboxylate moiety present in N1 but not in N6 seem to outweigh the advantage of stronger hydrogen bonds. The multiple bonding pattern makes the stacked structure N1 the global minimum structure if MP2 or CCSD(T) methods are employed. In contrast, DFT fails for these types of conformers since it is well known to overestimate hydrogen bonding and the common functionals like B3LYP are not able to describe dispersion interactions like π-π-stacking.

The N3 conformer is almost identical to the global minimum structure N1 differing only in the orientation of the hydrogen situated at the sp^2-hybridized guanidine nitrogen. The resulting repulsive interactions with the backbone hydrogen atoms destabilize this conformation of about 5 kJ mol^{-1}. While N1 possesses a chair-like conformation the N4 structure represents the twist conformer lying 7 kJ mol^{-1} higher in energy.

Table 8 Angles in degree of the chair-like conformers N1, N3 and N4 (RI-MP2/TZVPP+).

	N1	N3	N4
α	113.0	112.8	113.0
β	116.4	116.6	118.7
γ	115.8	115.7	115.7
δ	113.9	113.7	110.0
ε	103.4	101.4	103.8
ζ	87.5	87.7	82.3



For the zwitterionic structures an analogous picture is found. Here, DFT (*e.g.* B3-LYP/TZVPP+) predicts both structure types to be more or less similar in energy whereas MP2 and CCSD(T)/extrapol. calculations compute the new stacked structures ZW19 and ZW1 to be lower in energy by about 9 kJmol^{-1}.

An analysis of the situation in ZW19 is provided by **Figure 28**. Geometry optimization with RI-BLYP/TZVPP+ results in the local minimum structure given on the left hand side of **Figure 28** which lies ~10 kJ mol^{-1} above the global minimum N2 on this level of theory. If this structure is used as a starting point for a MP2 geometry optimization (RI-MP2/TZVPP+) the structure ZW19 shown on the right hand side of **Figure 28** is obtained. RI-MP2/TZVPP+ predicts ZW19 to be 7 kJ mol^{-1} lower in energy than the DFT optimized structure and about 7 kJmol^{-1} less stable than the global minimum N1. A single point RI-BLYP/TZVPP+ calculation on the MP2 optimized structures predicts it to be 19 kJ mol^{-1} higher in energy than the BLYP optimized structure.

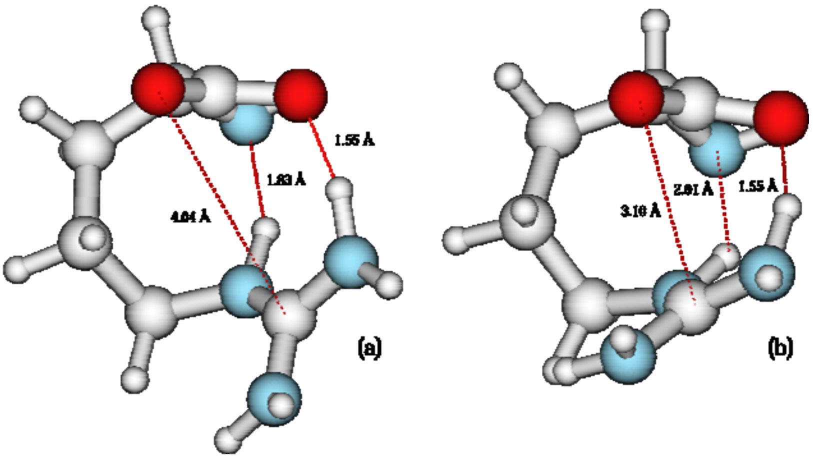


Figure 28 Geometries of the zwitterionic conformer ZW19 in dependence of the method. The a) B-LYP/TZVPP+ optimized geometry and b) MP2/TZVPP+ geometry differ in energy of 7.4 kJ mol^{-1} (MP2/TZVPP+)

The reasons for the variations can be seen from an analysis of the ESP fit charges. Both approaches agree with that most of the positive charge is localized on the hydrogen atoms but they slightly differ in the charge delocalization between the heavier centres. Using MP2/TZVPP+ positive and negative charge is a bit more localized on the heavier centres in contrast to DFT which is in line with findings showing that DFT overestimates delocalizing effects.[113] As a consequence of the more localized charges for MP2, the Coulomb interaction between the carboxylate oxygen and the central carbon atom of the guanidinium moiety leads to a decreased oxygen-carbon distance (**Figure 28**). A complete parallel orientation of both groups may be impeded by repulsive interactions between the oxygen and the nitrogen centres. It is noteworthy that the strong hydrogen bond between one of the oxygen atoms and one of the NH2 groups of the guanidinium remains nearly unchanged. Comparing the two low lying zwitterionic structures Z3 and ZW19 (**Figure 26**) it becomes clear that the increased electrostatic attractions outweigh at least one strong hydrogen bond. Besides the differences in the localization, one can also expect that the inability of DFT to account for dispersion effects is a second reason for the discrepancies. That dispersion effects are very important for guanidinium cations was recently shown by Brady and coworkers.[114]

In accordance to Saykally and coworkers, Rak et al. predicted a canonical conformer of arginine to represent the global minimum in the gas phase.[77,79] Using the MP2/6-31++G** approach the lowest lying zwitterionic structure was computed to lie only 7 kJmol^{-1} higher but the energy difference increased to about 17 kJmol^{-1} if CCSD was employed. Since the stacked zwitterionic structure ZW19 is found to lie considerably lower in energy than the already known zwitterionic structure Z3, the question arises if it becomes the global minimum. This is not the case as can be seen from **Table 6**. Using the MP2/TZVPP+ approach the structure ZW19 is found to lie about 7 kJmol^{-1} above the lowest lying neutral structure N1 (**Figure 26**). CCSD(T) computations were performed to check the MP2 predictions. After the basis set extrapolation the zwitterionic structure ZW19 lies 11 kJ mol^{-1} higher in energy than the global minimum N1. However, since CCSD(T) should be more influenced by basis set effects than MP2 the 11 kJ mol^{-1} represents a lower limit.[115]

Table 6 reveals that the computed energy differences between the various low lying isomers are so small that zero-point vibrational energy contributions and thermal effects are not negligible. The various contributions are summarized in with respect to structure N1. **Table 9** shows that the structures exhibiting directed hydrogen bonds (N2, N6, N19) slightly benefit from the enthalpy and entropy corrections. Due to the thermodynamic corrections the energy difference between the global minimum (N1) and the now lowest lying zwitterionic structure ZW1 decreases to 9 kJmol^{-1} (extrapolated CCSD(T) values). In comparison to previous works the present approach could identify new global minimum structures for both the canonical as well as the zwitterionic tautomers and finds that the expected energy gap between these structures is only about half as large as predicted by earlier calculations which employed less accurate methods.[77c]

Table 9 Thermodynamic corrections in kJmol^{-1} for RI-MP2/TZVPP+ optimized geometries calculated on RI-MP2/TZVP level (T=298.15 K).

	N1	N2	N3	N4	N6	N19	C5	ZW1	ZW19	Z3
ΔH_{corr}	0.0	-0.7	+0.2	-0.3	-0.2	0.0	-0.8	-4.0	-3.7	-5.4
$-T\Delta S_{\mathrm{corr}}$	0.0	-2.9	+0.2	-0.5	-2.6	-2.1	-2.5	-2.0	+2.2	-0.8
ΔG_{MP2}[1]	0.0	+2.0	+5.9	+6.3	+5.7	+6.6	+5.0	+5.0	+5.2	+9.5
$\Delta G_{\mathrm{CCSD(T)}}$[2]	0.0	+0.7	+5.4	+5.6	+5.2	+5.3	+3.0	+9.4	+9.3	+13.1

---

[1] $\Delta G_{\mathrm{MP2}} = \Delta E_{elec}(\text{RI-MP2/TZVPP+}) + \Delta H^{\mathrm{corr}} - T\Delta S^{\mathrm{corr}}$

[2] $\Delta G_{\mathrm{CCSD(T)}} = \Delta E_{elec}(\text{CCSD(T)/extrapol.}) + \Delta H^{\mathrm{corr}} - T\Delta S^{\mathrm{corr}}$

4.2.1.6 Calculated Spectra

According to our study arginine possesses conformers with large structural differences already within an energy range of less than 10 kJmol^{-1}. Taking into account remaining uncertainties resulting from the conformational search and from the notoriously difficult estimate of entropy effects only experiment can provide an unambiguous answer which structure type represents the global minimum. Due to the differences in the intramolecular interactions vibrational spectroscopy should at least be able to distinguish between the various types of conformers (stacked arrangements vs. directed hydrogen bonds) and tautomers (canonical vs. zwitterionic form). This approach was utilized for the first time by Saykally and coworkers who concentrated on the region between 1500 and 1600 cm^{-1} in order to determine experimentally whether the canonical or the zwitterionic form represents the global minimum. However, as already pointed out by Rak et al. this energy range is not sufficient for a definite answer.

To estimate which interval of the spectra is best suited for an unambiguous identification the RI-MP2 method was used to compute the various IR spectra. Figure 5 shows the calculated, unscaled line spectra of the various low lying energy conformers of canonical and zwitterionic arginine with the superimpositions of Gauss fitted curves. Characteristic peaks were assigned by vibrational mode analysis. The resulting labelling is given if peaks could be related to more or less uncoupled vibrations.

All spectra can be divided into three major regions: (a) the fingerprint region below 1500 cm^{-1} wavenumbers showing coupled scaffold and bending vibrations, (b) the region between 1600 cm^{-1} and 1900 cm^{-1} consisting mainly of C=O and C=N stretch vibrations and (c) the hydrogen stretch vibrations between 2600 cm^{-1} and 4000 cm^{-1}. Chapo et al. as well as Rak et al. both concentrated on region (b) in order to determine the tautomeric state of arginine. The spectra of the canonical conformers show mainly a C=O stretch band around 1860 cm^{-1} and a clear peak for one C=N stretch vibration in the guanidine part at ~1710 cm^{-1}. Some other peaks occur representing COH bending or coupled vibrations which are quite similar for both forms of geometrical arrangements (linear H-bond vs. stacking orientation). Regarding the spectra of the zwitterionic arginine the symmetric O=C=O stretch band is rather weak and coupled with other vibrations. The large C=N stretch peak is shifted to larger wavenumbers (1830-1860 cm^{-1}) in comparison to the canonical conformers and it is also coupled with NH stretch vibrations. However, whereas the assignments for the various conformers are different, the experimentally accessible properties such as frequencies and intensities are too similar for a definitive determination of the structure of the global minimum.

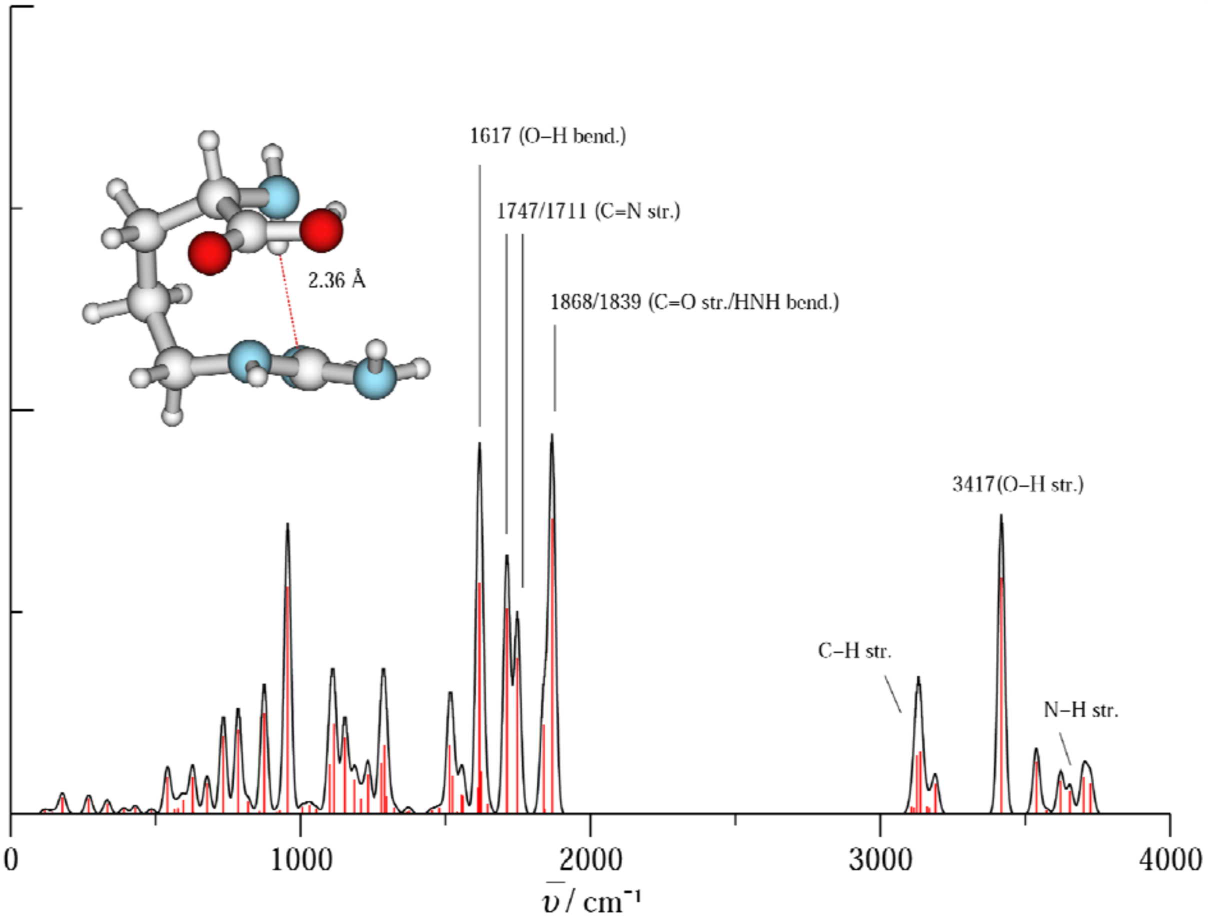
As expected region (c) containing the N-H and O-H vibrations would allow an unambiguous identification which type of conformer (stacked arrangements vs. directed hydrogen bonds) and tautomers (canonical *vs.* zwitterionic form) predominates in the gas phase. The conformers C5 and N2 possess a very intense peak at about 3300 cm^{-1} assigned to the stretch vibration of the O-H group. The high intensities result from their involvement in the directed hydrogen bond to the guanidinium group. The spectra of N2 and C5 show a slightly different energy gap between the OH-stretch and the CH-stretch vibrations which may be used to differentiate between both conformers. In any case this slight difference could be used to determine if both conformers were present in gas phase.

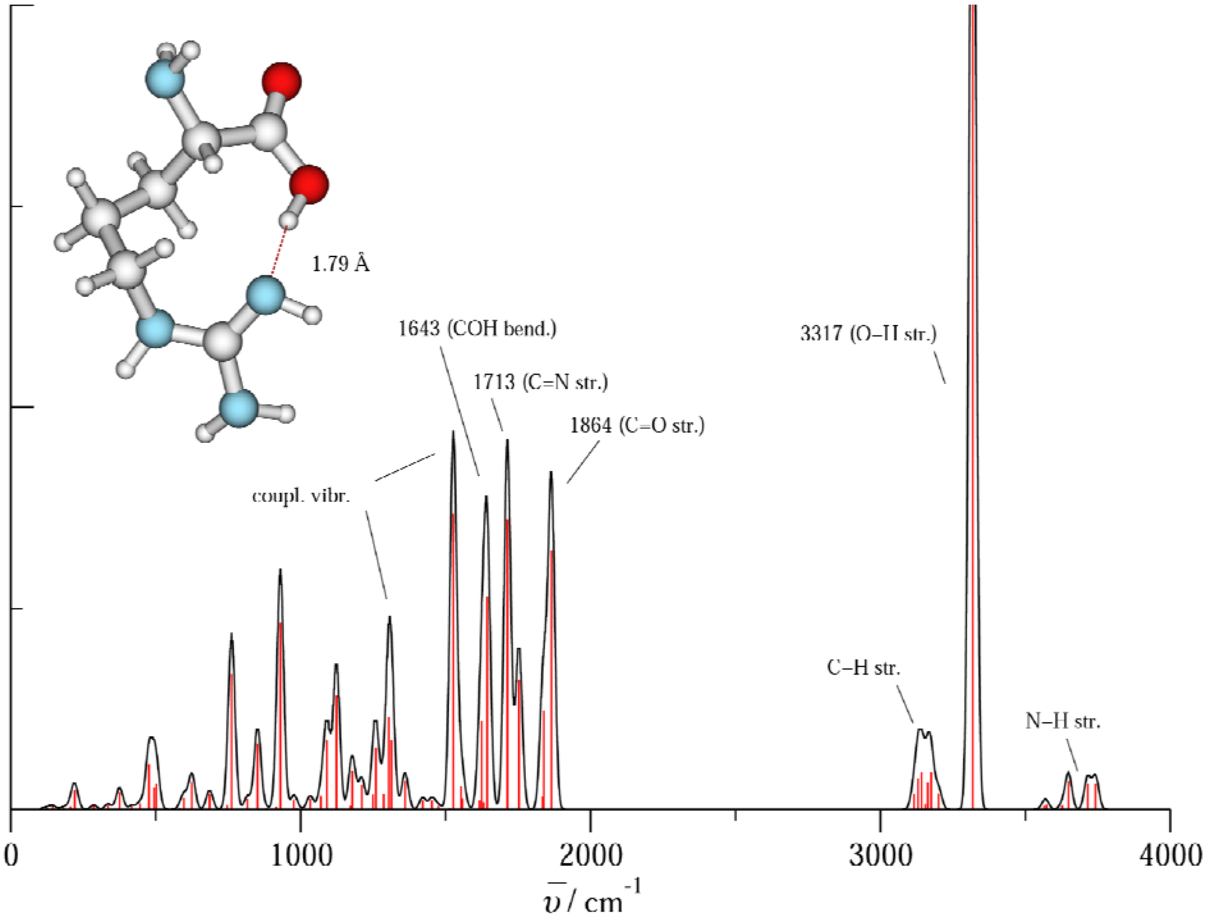
The intensity of this vibration is drastically reduced when the arginine shows a stacked conformation (N1, N3, N4) and also a small shift towards ~3400 cm^{-1} is predicted. Therefore, this peak allows a differentiation between directed and stacked canonical conformers, whereas it is hardly possible to distinguish between N3 and N4 or N6 and N19, respectively.

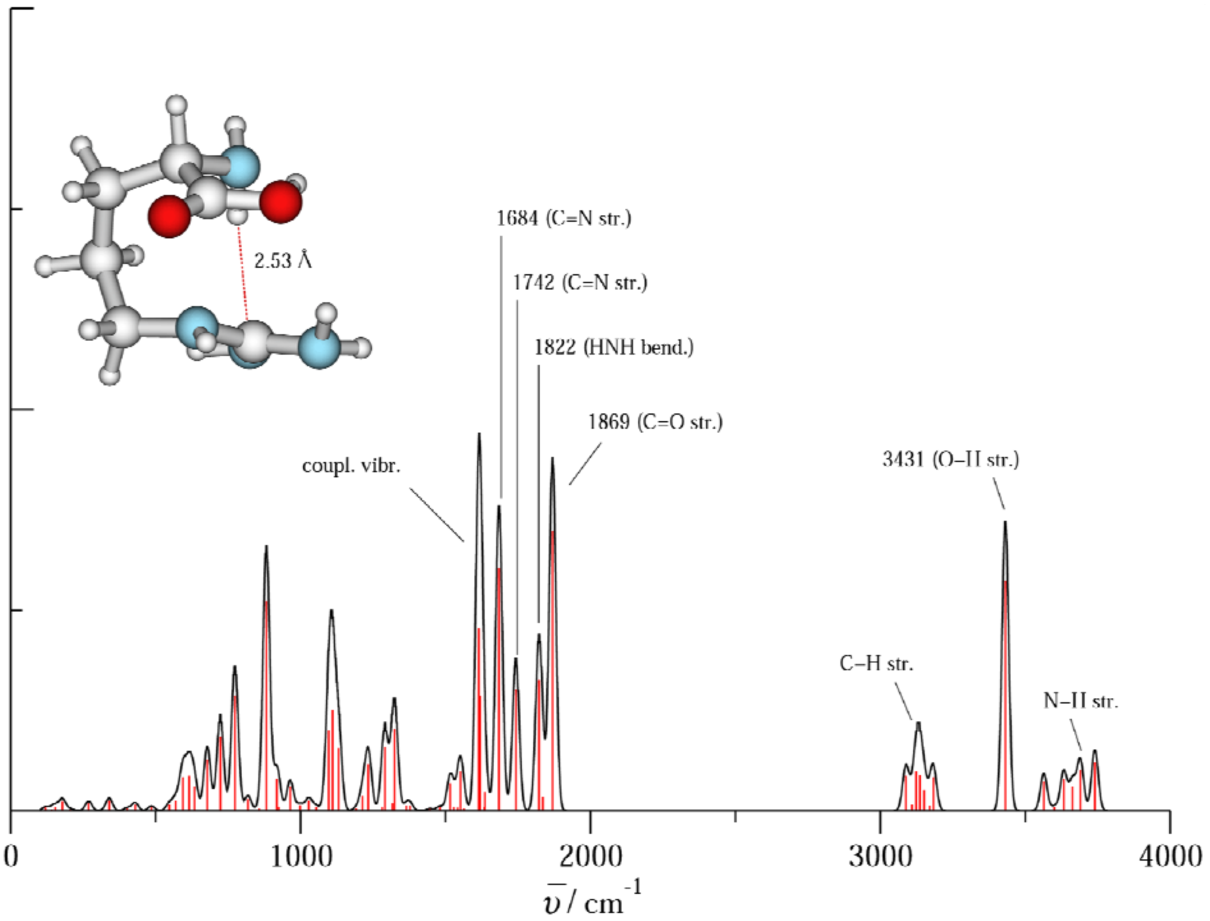
For the zwitterionic conformers Z3, ZW1 and ZW19 the N-H stretch vibrations of the zwitterionic hydrogen bond (NH···O) appear in this energy range. These vibrations all occur at wavenumbers well below 3300 cm^{-1} and should therefore be a characteristic evidence for the existence of zwitterionic conformers in the gas phase. A differentiation between stacked or directed types of conformers is also easily possible since the latter possesses two strong absorptions within a small energy range. In contrast, the stacked conformer ZW1 and ZW19 shows strong peaks at about 3200 and 3400 cm^{-1}, respectively. They are assigned to the NH stretch vibration of the NH···N hydrogen bond between the guanidinium group and the α-amino nitrogen centre. Our calculations strongly suggest that the region around 3000 cm^{-1} can be used to experimentally differentiate between the various structures. Hence, this way allows an unambiguous determination of the global minimum of the arginine monomer.

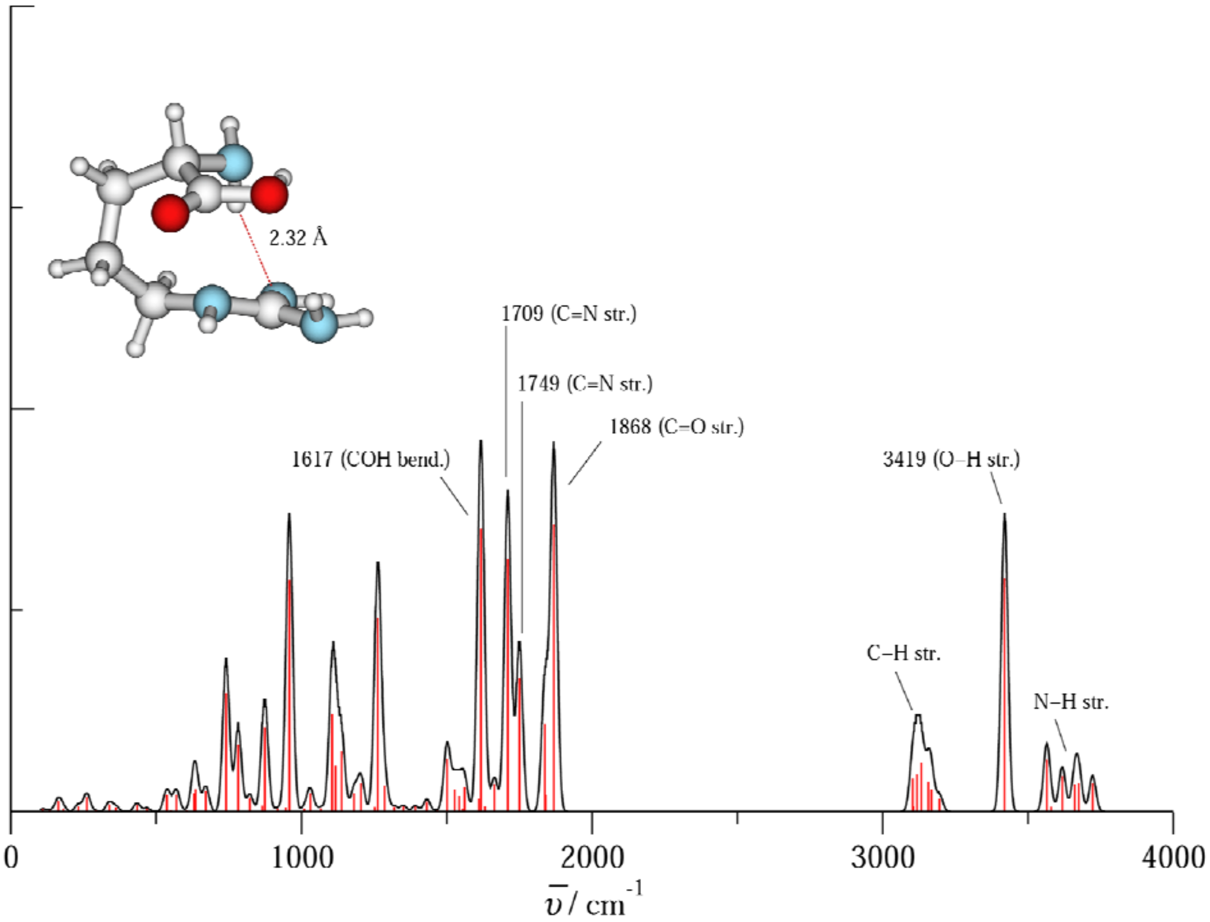
Figure 29 Pages 115 to 119: Gauss fitted curves of calculated vibration line spectra of neutral arginine conformers (RI-MP2/TZVP). Some relevant unscaled vibration modes are given. The peaks of high intensity were cut for the sake of clarity.

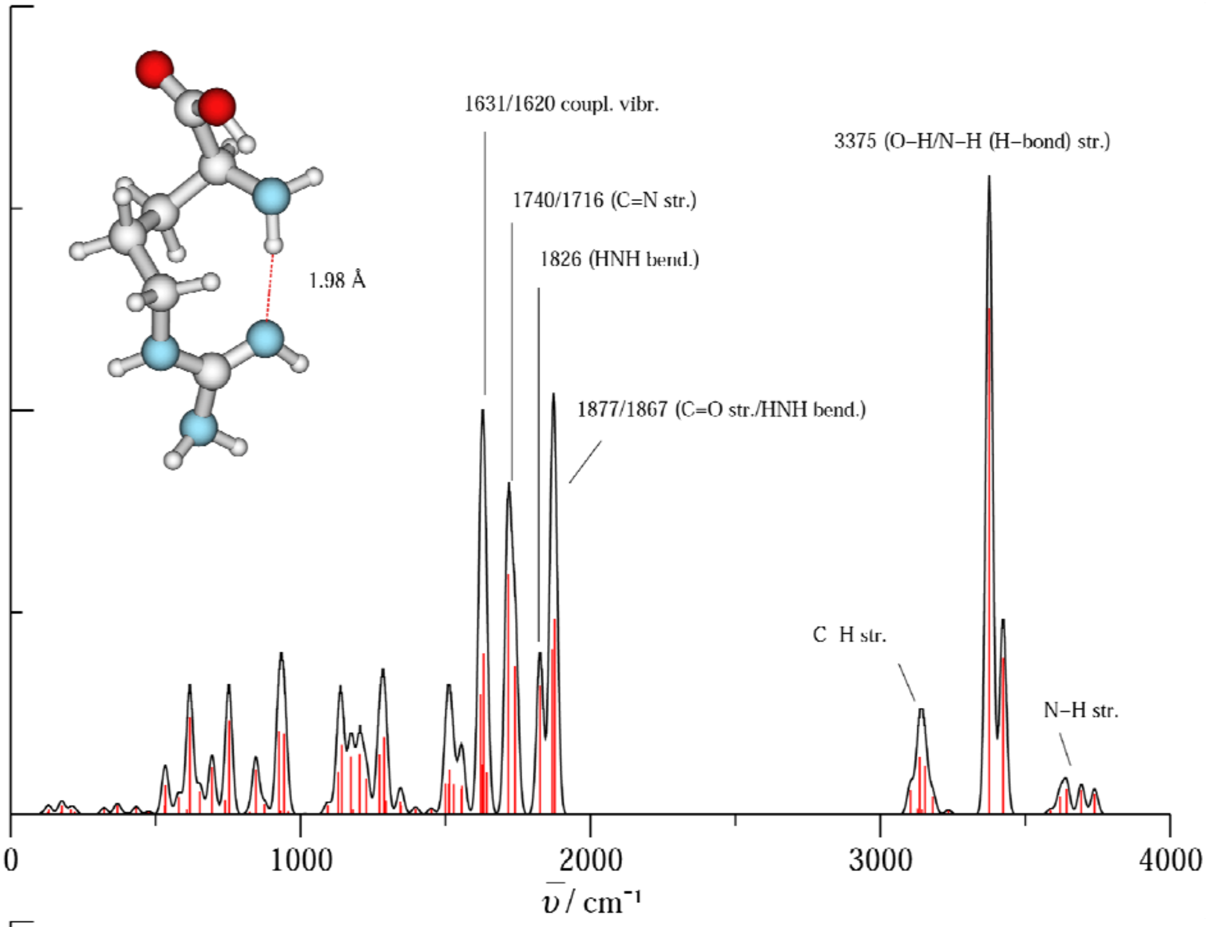
p. 115: N1 (top), N2 (bottom); p. 116: N3 (top), N4 (bottom); p. 117: N6 (top), N19 (bottom); p. 118: C5 (top); Z3 (bottom); p. 119: ZW1 (top), ZW19 (bottom);

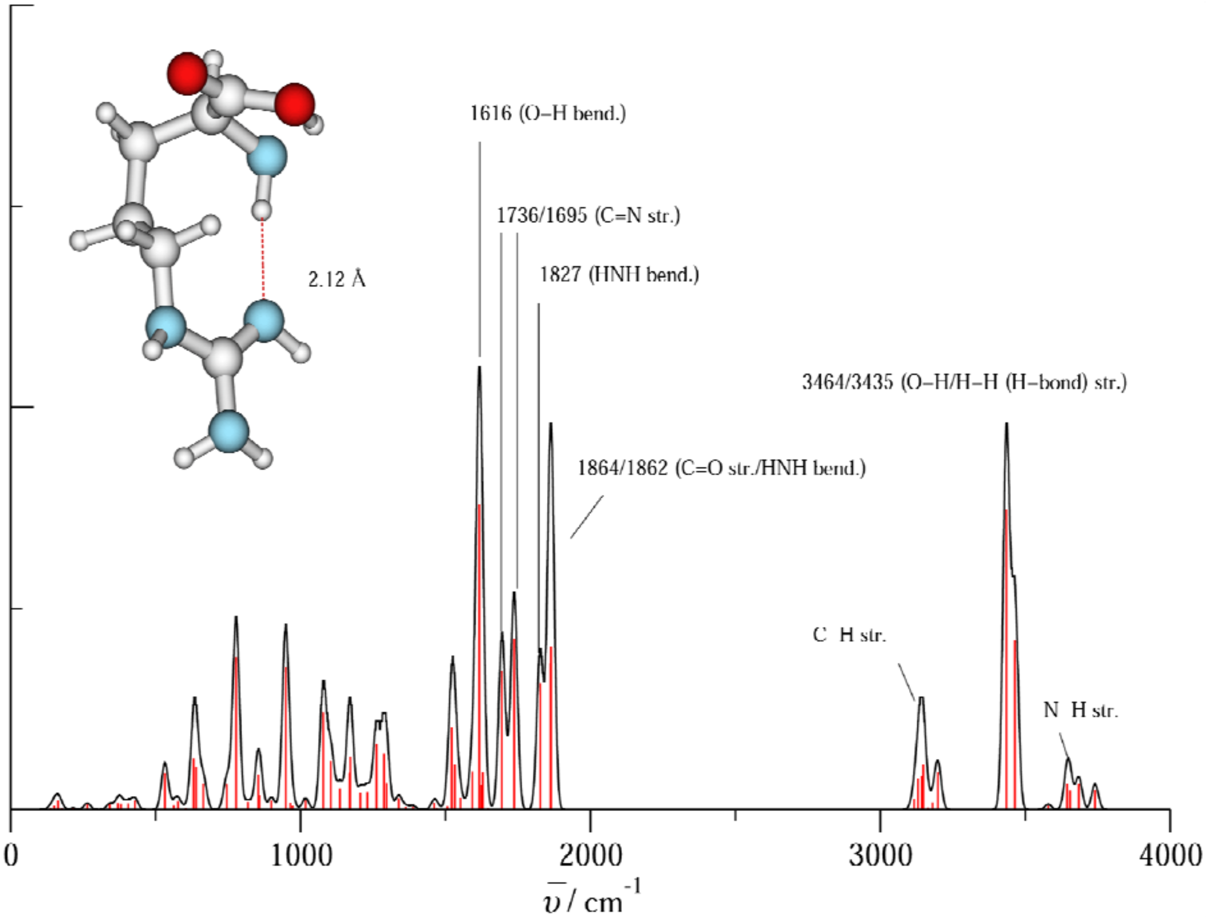

1617 (O–H bend.)
1747/1711 (C=N str.)
1868/1839 (C=O str./HNH bend.)
2.36 Å
3417(O–H str.)
C–H str.
N–H str.
0
1000
2000
3000
4000
$\bar{\upsilon}$ / cm^{-1}

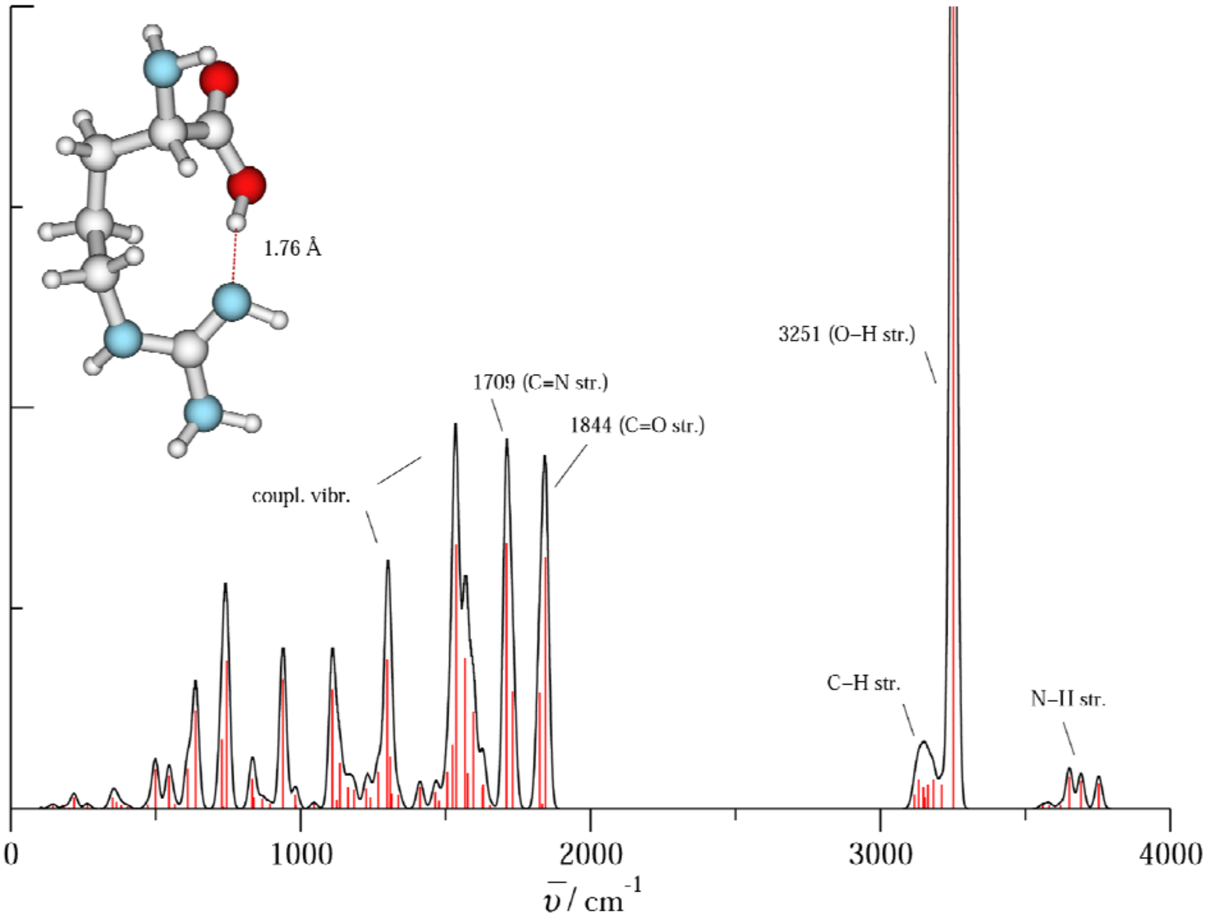

1.79 Å
1643 (COH bend.)
1713 (C=N str.)
1864 (C=O str.)
coupl. vibr.
3317 (O–H str.)
C–H str.
N–H str.
0
1000
2000
3000
4000
$\bar{\upsilon}$ / cm^{-1}

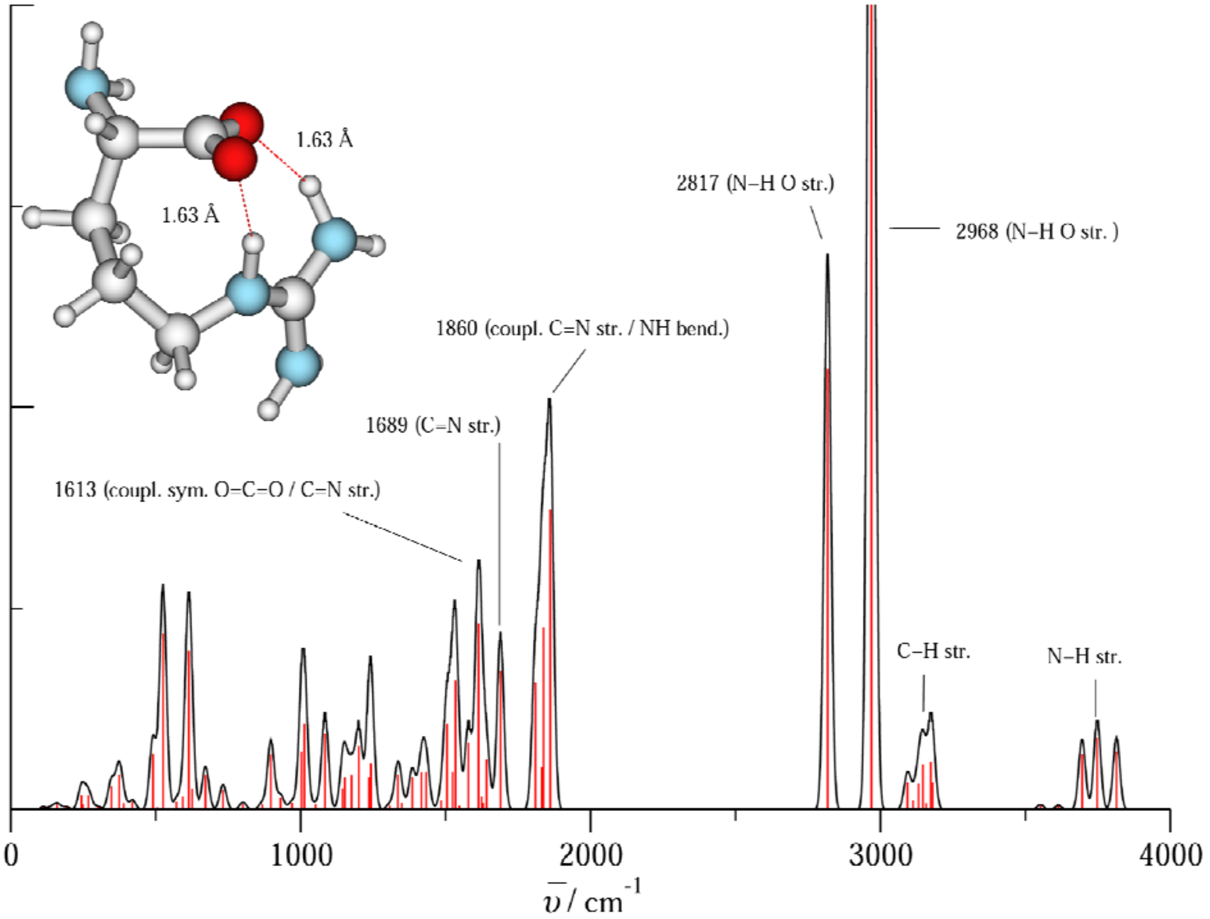

2.53 Å
1684 (C=N str.)
1742 (C=N str.)
1822 (HNH bend.)
1869 (C=O str.)
coupl. vibr.
3431 (O–H str.)
C–H str.
N–H str.
0
1000
2000
3000
4000
ῡ / cm-1

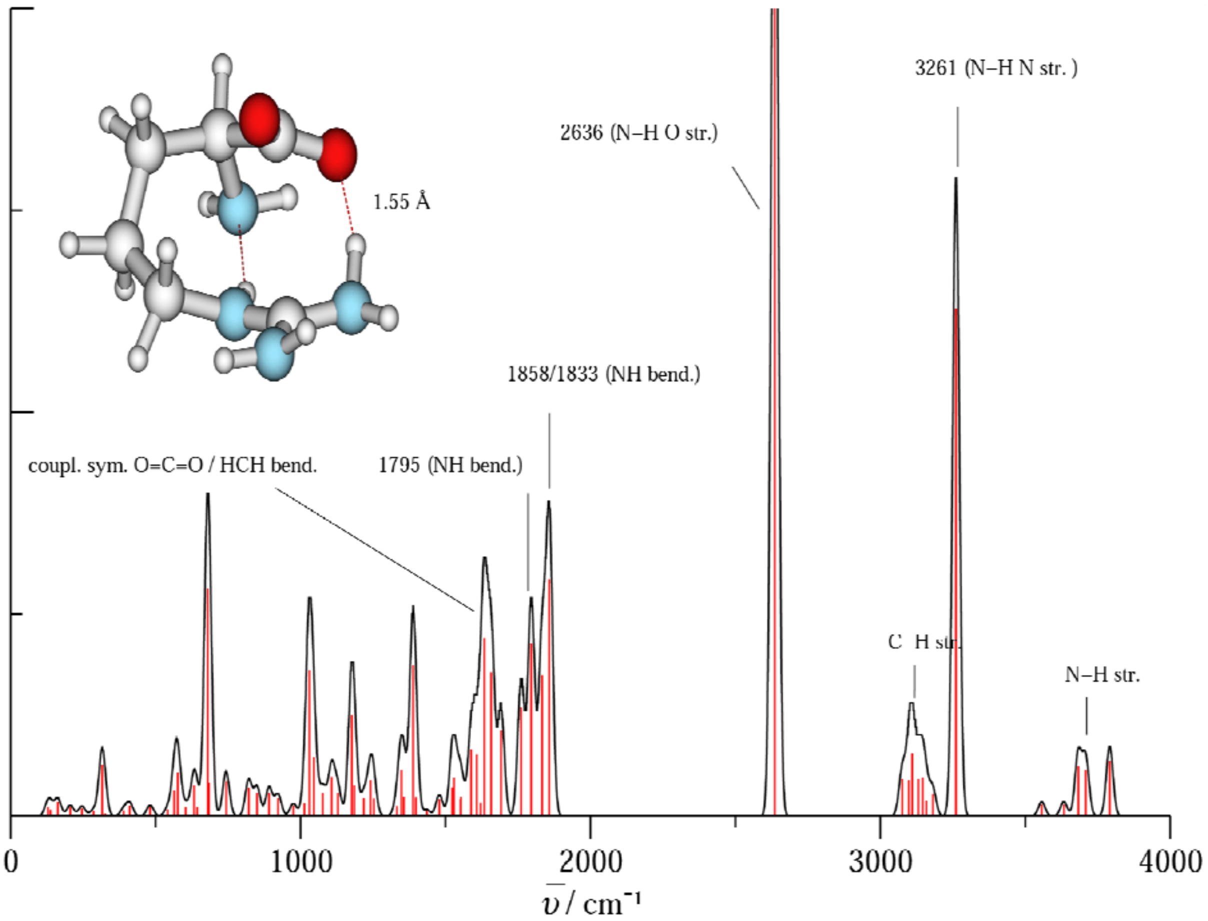

2.32 Å
1709 (C=N str.)
1749 (C=N str.)
1868 (C=O str.)
1617 (COH bend.)
3419 (O–H str.)
C–H str.
N–H str.
0
1000
2000
3000
4000
ῡ / cm-1

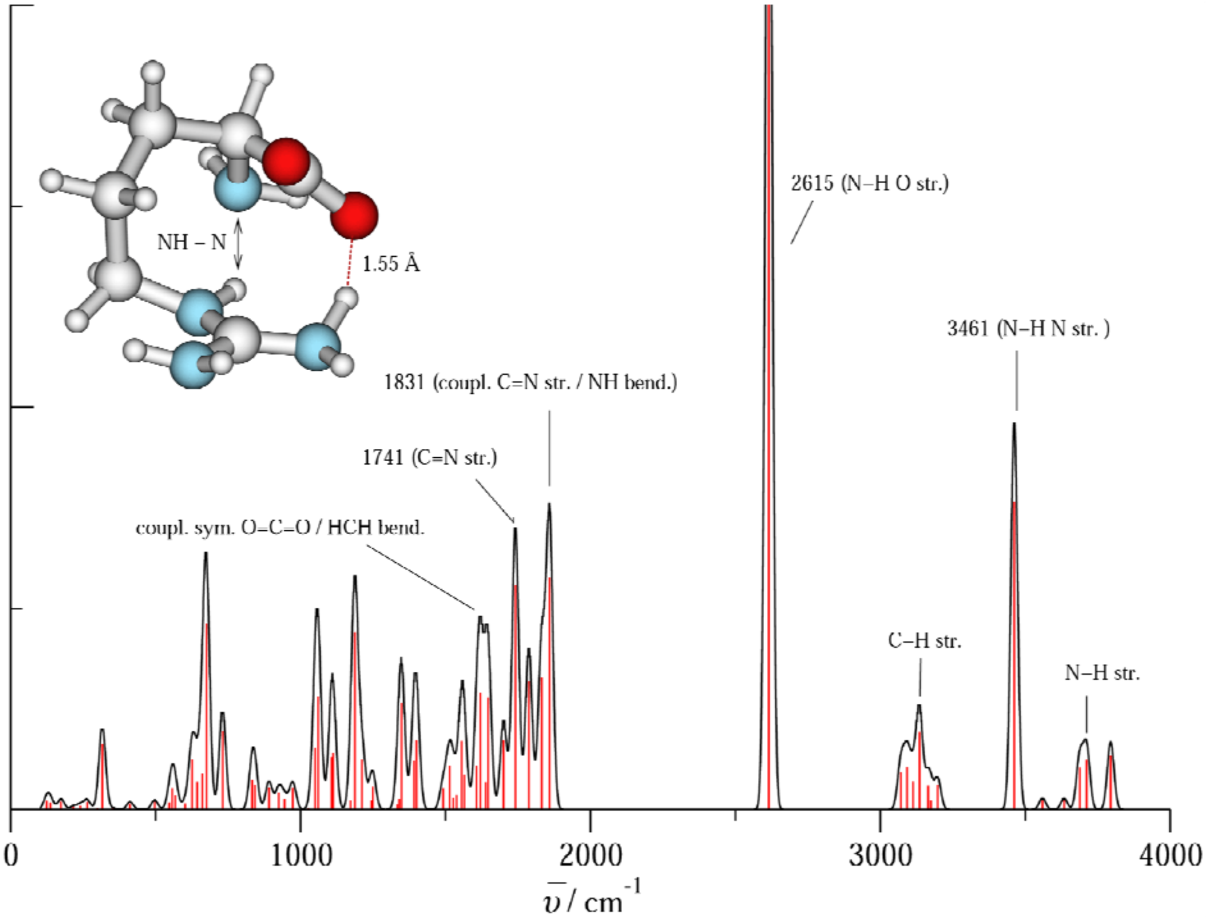

1.98 Å
1631/1620 coupl. vibr.
1740/1716 (C=N str.)
1826 (HNH bend.)
1877/1867 (C=O str./HNH bend.)
3375 (O–H/N–H (H–bond) str.)
C H str.
N–H str.
0
1000
2000
3000
4000
υ / cm⁻¹


2.12 Å
1616 (O–H bend.)
1736/1695 (C=N str.)
1827 (HNH bend.)
1864/1862 (C=O str./HNH bend.)
3464/3435 (O–H/H–H (H–bond) str.)
C H str.
N H str.
0
1000
2000
3000
4000
υ / cm⁻¹


1.76 Å
3251 (O–H str.)
1709 (C=N str.)
1844 (C=O str.)
coupl. vibr.
C–H str.
N–H str.
0
1000
2000
3000
4000
$\bar{\upsilon}$ / cm^{-1}


1.63 Å
1.63 Å
2817 (N–H O str.)
2968 (N–H O str.)
1860 (coupl. C=N str. / NH bend.)
1689 (C=N str.)
1613 (coupl. sym. O=C=O / C=N str.)
C–H str.
N–H str.
0
1000
2000
3000
4000
$\bar{\upsilon}$ / cm^{-1}


1.55 Å
2636 (N–H O str.)
3261 (N–H N str.)
1858/1833 (NH bend.)
coupl. sym. O=C=O / HCH bend.
1795 (NH bend.)
C H str.
N–H str.
0
1000
2000
3000
4000
$\bar{\upsilon}$ / cm^{-1}


NH – N
1.55 Å
2615 (N–H O str.)
3461 (N–H N str.)
1831 (coupl. C=N str. / NH bend.)
1741 (C=N str.)
coupl. sym. O=C=O / HCH bend.
C–H str.
N–H str.
0
1000
2000
3000
4000
$\bar{\upsilon}$ / cm^{-1}

4.2.1.7 Conclusions

In this chapter a strategy was presented that enables to identify minimum energy conformers in gas phase of the arginine monomer serving as a prototype for medium sized organic molecules that show strong intramolecular interactions. The methods used for various stages are shown and various pitfalls are discussed. Starting from an exhaustive force field based conformational search using different conformational search algorithms for the zwitterionic and the canonical tautomeric state of arginine, the lowest energy structures have been selected based on clustering and visual inspection. The subsequent optimizations on DFT and MP2 level of theory employing a large basis set revealed the shortcoming of density functional approaches when describing stacking-like interactions. In this case, new geometrical arrangements were found that have not been reported so far comprising a stacked orientation of the terminal groups. For the arginine the explanation for this finding is twofold. In the case of the canonical structure a geometry optimization solely based on density functional theory would neglect dispersive interactions between the guanidine group and the carboxylic acid which can occur due to the high flexibility of the arginine. For the zwitterionic monomer it was shown that in addition to the lack of dispersion terms also the overestimation of hydrogen bonding and delocalization is the main reason for the deficiency of B3LYP or BLYP functionals. As a consequence, the electrostatic interactions between the carbonyl and guanidine moieties are underestimated and the stacked conformers are described poorly. Therefore, as a thumb rule, if treating non-covalently interacting systems that can undergo a π-π like stacking interaction, at least a simple electron correlation method like MP2 is mandatory in order to get appropriate relative energies of all possible conformations. This assumption should also hold for biochemical systems which are only hardly accessible to solvent molecules like many active sites in enzymes are. Single point energy calculations on CCSD(T)/cc-VDZ level finally confirmed the previously found trends for the relative energies of the conformers and a hitherto unknown global minimum conformer of arginine in gas phase (N1) was detected that is more than 8 kJ mol^{-1} lower in energy than the published structures.

It was also shown that the new zwitterionic conformer ZW1 is energetically near the canonical global minimum (7 kJ mol^{-1}) implicating that a rigorous exclusion of a zwitterionic state in gas phase as it was proposed by several studies before is not tenable. An unambiguous proof can therefore only be given by experiment. For this purpose we computed the gauss fitted vibrational spectra for the lowest monomer structures and we were able to show that a comparison of the hydrogen stretch vibrations between 2600 cm^{-1} and 4000 cm^{-1} would make it possible to assign which tautomer (zwitterionic *vs.* canonical) and which type of conformer (directed hydrogen bonds vs. stacked orientation) dominates in gas phase.

4.2.2 Self-Assembly of Gas Phase Arginine

4.2.2.1 Introduction

A successful rational design of new artificial receptor systems requires a profound knowledge of all inter- and intramolecular interactions as well as the ability to distinguish between molecular inherent and solvent effects. On this account theoretical studies are often used to investigate the gas phase properties of such systems which best reflect the molecular inherent effects. Moreover, theory also allows a differentiation between various interactions. In order to study the gas phase binding properties of guanidinium based artificial receptor systems we started to research on the dimerization of arginine monomers. A similar approach has already been performed by Goddard III and co-workers who calculated bonding energies of arginine dimers and trimers with respect to the global minimum of arginine published by Rak and coworkers.[77a,c]

However, our conformational search in conjunction with accurate electron correlation computations yielded yet unknown conformers of arginine dimer systems which exhibit completely new types of geometrical arrangements not reported before. Additionally, new global minimum structures could be identified. Therefore, the first aim is to discuss the interactions which stabilize the new conformers and to give a brief outline why these structures could not be identified in the previous works.

Second, it is studied to what extent a stiffening of a system can enhance its ability to self-aggregate, *e.g.* to form stable dimers. The question arises from the previous study of the binding properties of the efficient carboxylate receptors developed by Schmuck and coworkers (see chapter 4.1). Within these dimer systems the non-covalent complexation includes a variety of effects like ionic interactions, hydrogen bonding and cooperativity which all contribute to the stabilization energy. However, recent theoretical studies of the dissociation processes of the dimer indicate that not only the additional hydrogen bonds compared to the parent guanidinium cation, but also the rigidity of the monomer is of utmost decisive importance for the stability of the dimer. Due to its strong rigidity the 2-(guanidiniocarbonyl)-1H-pyrrole-5-carboxylate monomer cannot be stabilized by intramolecular interactions between the charged terminal groups. As a result, the formation of assemblies leads to high dimerization energies. In contrast, arginine has a significant flexibility because of the large amount of rotatable bonds in the side chain so that the oppositely charged groups strongly interact already in the monomer. Arginine is expected to show drastically reduced dimerization energies in comparison to the artificial systems introduced by Schmuck et al.. This difference can not only be explained simply by the difference in the binding motifs (*e.g.* number of H-bonds or acidity). To estimate how an artificial stiffening of arginine would enhance its complexation ability we computed the dimerization energy of a linear conformer of arginine. These calculations indicate that the dimerization energy of an artificial rigid arginine is about twice as large as for the flexible arginine. Therefore, at least for self-complementary systems, the stiffening of a molecule seems to be a suitable instrument to steer its complexation properties. A comparison to the completely rigid the 2-(guanidiniocarbonyl)-1H-pyrrole-5-carboxylate shows that its inherent rigidity accounts to about 50% to the dimer stability.

4.2.2.2 Computational Details

An efficient although exhaustive search of the conformational space as well as the choice of an appropriate force field is the crux in determining low-energy conformers of canonical and zwitterionic arginine monomers and their non-covalent assemblies. An extensive validation revealed that for the arginine monomer and dimer the Mixed Monte-Carlo Multiple Minimum/Low Mode (MCMM/LowMode) approach[39,40] and the Systematic Unbound Multiple Minimum (SUMM) algorithm[38] as implemented in the MacroModel8.0 program package[98] are the most effective tools for scanning the conformational space, whereas the OPLS-AA[100] and the MMFF94[101] force fields gave the best structures and energy order of the conformers.

All conformational searches took between 2000 to 5000 steps and were repeated from different starting structures. In the next step the most promising structures were optimized on the RI-BLYP/SV(P) level of theory employing the TURBOMOLE program package, respectively. [35b,e,65,102,65] In these calculations a great number of conformers should be taken into account since the energy ordering resulting from the force field based conformational searches and from these DFT computations differ largely (see chapter 4.2.1). To keep the efforts manageable, the conformers treated in this step were manually selected after visual inspection.

To determine the lowest lying structures of the dimer system the same strategy like for the monomer was used (see chapter 4.2.1). Due to software and hardware restrictions the RI-MP2 optimizations were only feasible with a TZVP basis. The final electronic energies were then calculated by single-point calculations on RI-MP2 level employing the TZVPP+ basis in which the oxygen centres was augmented by one s and one p function with low exponents (ζ=0.068) in order to describe the diffuse shape of electrons of the carboxylate atoms in the zwitterionic conformers properly.[104] For the auxiliary basis sets the exponent was doubled (ζ=0.136). CCSD(T) computations with reasonable basis sets were not possible. Dissociation energies of the dimer species were calculated including the counterpoise correction according to Boys and Bernardi.[24]

4.2.2.3 Geometries and Energies

The conformational search for low lying conformers of the arginine dimer was performed with the same strategy as used for the monomer. All detected low lying structures consist of two zwitterionic species interacting through several charged H-bonds between the guanidinium and carboxylate moieties. Dimers consisting of canonical monomers are unfavourable since the zwitterionic forms are stabilized through the electric field of its counterpart. Hence, the formation of the strong bonded salt bridges outweighs the energy necessary to form the zwitterionic monomers. The deep minimum wells resulting from these strong salt bridges also explain the low number of conformers.

The three lowest lying conformers detected within our conformational search (MMFF1, MMFF2 and MMFF4), which were all predicted by the MMFF94 force field, are depicted in **Figure 30**. **Figure 30** also contains the lowest conformer given by Goddard III and co-workers (DZ1) and the lowest structure predicted by the OPLS-AA force field (OPLS-AA1).[77a] **Table 10** summarizes the most important geometrical parameters. The geometrical differences between the MMFF structures and both other conformers (OPLS-AA1 and DZ1) are striking. In all structures the zwitterionic monomers form strong salt bridges between the guanidinium moiety and the carboxylate group of the counterpart through a network of directed hydrogen bonds. The striking differences result from the intramolecular interactions of the carboxylate and the guanidinium group of a given monomer. For the OPLS-AA1 and the DZ1 conformer these units also interact through one directed hydrogen bond necessitating a planar structure for the dimer. In contrast, in the MMFF structures the guanidinium and the carboxylate moiety of one monomer adopt a more parallel orientation leading to pocket-like structures.

Within **Figure 30** the opening of the pocket is in the foreground (R(O(2)···N(8)) = 4-5 Å). The bond lengths between O(1) and HN(7) (R = 1.98 - 2.12 Å) indicate moderate hydrogen bond strengths. The structural arrangement points to interactions resembling those found in the ZW19 conformer. With dimerization energies of about 200 kJ mol^{-1} (MP2/TZVPP) these pocket-like structures are about 60 kJ mol^{-1} more stable than the planar structures DZ1 and OPLS-AA1 (see **Table 11**).

The stronger stabilization of the MMFF structures does not only result from the differences in the bonding network between the guanidinium and carboxylate moieties, but also from an additional hydrogen bond between the α-amino-nitrogen and the guanidinium moiety which cannot be formed in DZ1 or OPLS-AA1 due to their planarity.

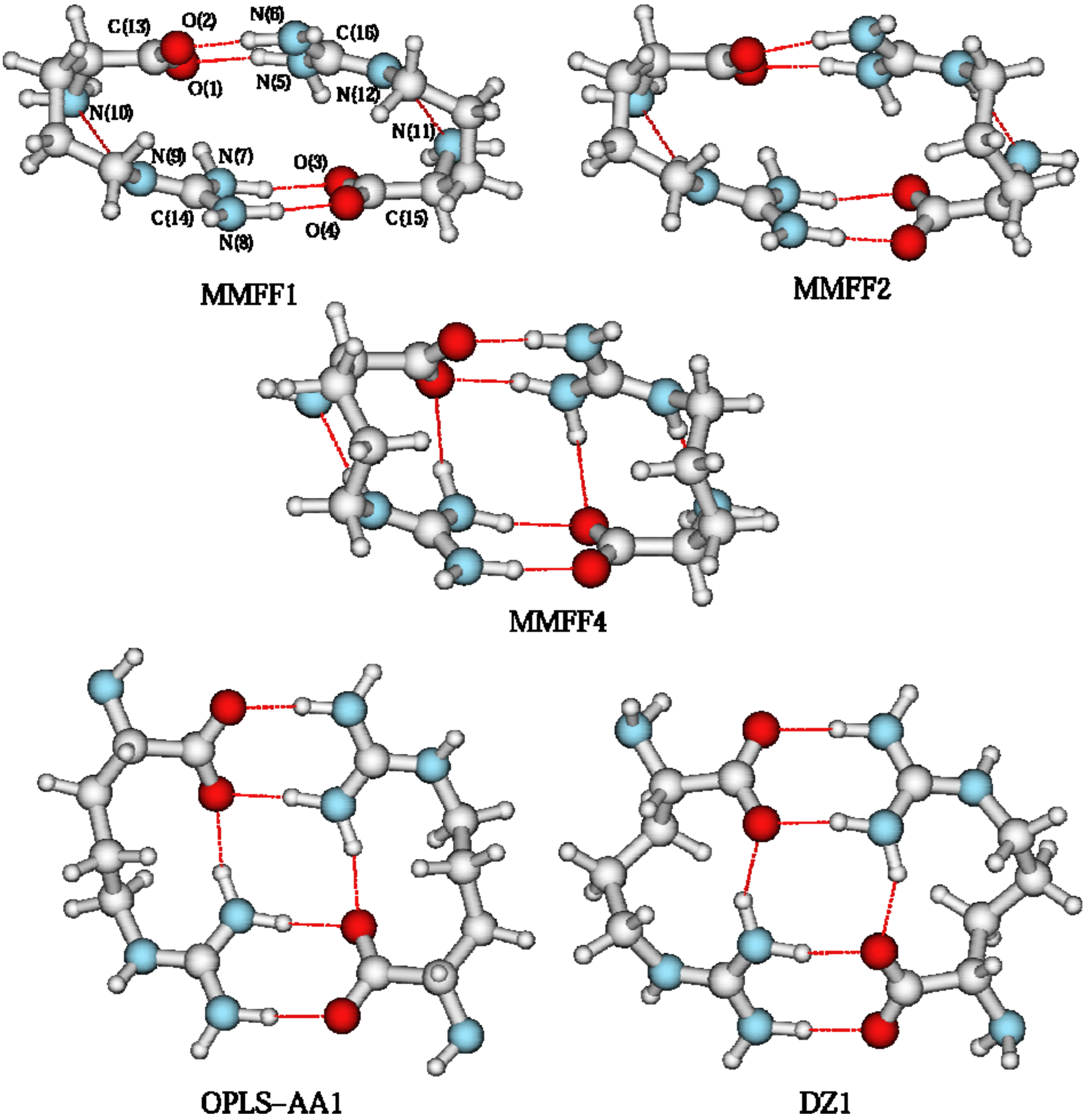
The energy difference between MMFF1 and MMFF4 results from the interplay of the various bonding effects. In MMFF4 the intramolecular hydrogen bonds are shorter but the distances between the carboxylate and the guanidinium moieties (R(C(13)···C(14)) = 3.56 Å) are enlarged with respect to the MMFF1 dimer (R(C(13)···C(14)) = 3.22 Å). This indicates stronger hydrogen bonds but smaller electrostatic interactions between the negatively and positively charged terminal groups. The structural differences are caused by a flip of the alkyl backbone. MMFF2 represents a mixture between MMFF1 and MMFF4.

The interactions stabilizing the pocket structures seem to comprise electrostatic and dispersions contributions. This is indicated by the differences between the dimerization energies computed with DFT and with MP2 as shown in **Table 12** and underlines again the importance of stacking effects as already discussed for the monomers. For the MMFF structures DFT always yields considerably smaller stabilization energies (16-28 kJ mol^{-1}) since it cannot account for the dispersion part. For the DZ1 and OPLS-AA1 DFT predicts slightly higher dimerization energies ($\Delta E \approx 7$ kJ mol^{-1}) since the strengths of the directed hydrogen bonds seem to be overestimated in comparison to MP2.

In conclusion, our calculations reveal new structures for the arginine dimer which are twice as stable relative to the monomer as the previously predicted structures. This shows that careful conformational searches are necessary since the lowest lying structures can be counter-intuitive even for such well-known species as arginine.

Table 10 Selected atomic distances in zwitterionic arginine dimer conformers optimized on RI-MP2/TZVP level of theory. All values are given in Ångstrøm.

Dimer	Atoms	Distance (Å)
MMFF1	O(1)···HN(5), O(3)···HN(7)	1.79
	O(2)···HN(6), O(4)···HN(8)	1.67
	O(1)···HN(7), O(3)···HN(5)	2.12
	N(10)···HN(9), N(11)···HN(12)	1.92
	C(13)···C(14), C(15)···C(16)	3.22
MMFF2	O(1)···HN(5)	1.77
	O(2)···HN(6)	1.68
	O(1)···HN(7)	2.10
	N(10)···HN(9)	1.91
	C(13)···C(14)	3.26
	O(3)···HN(7)	1.86
	O(4)···HN(8)	1.63
	O(3)···HN(5)	1.97
	N(11)···HN(12)	1.86
	C(15)···C(16)	3.57
MMFF4	O(1)···HN(5), O(3)···HN(7)	1.82
	O(2)···HN(6), O(4)···HN(8)	1.66
	O(1)···HN(7), O(3)···HN(5)	1.98
	N(10)···HN(9), N(11)···HN(12)	1.86
	C(13)···C(14), C(15)···C(16)	3.56
OPLS-AA1	O(1)···HN(5), O(3)···HN(7)	1.71
	O(2)···HN(6), O(4)···HN(8)	1.67
	O(1)···HN(7), O(3)···HN(5)	1.80
DZ1	O(1)···HN(5), O(3)···HN(7)	1.69
	O(2)···HN(6), O(4)···HN(8)	1.71
	O(1)···HN(7), O(3)···HN(5)	1.77


C(13)
O(2)
N(6)
C(16)
O(1)
N(5)
N(12)
N(10)
N(11)
N(9)
N(7)
O(3)
C(14)
N(8)
O(4)
C(15)

MMFF1 MMFF2

MMFF4

OPLS–AA1 DZ1

Figure 30 RI-MP2/TZVP optimized structures of zwitterionic arginine dimers.

Table 11 Dimerization energies of zwitterionic arginine conformers calculated on a RI-MP2/TZVP//RI-MP2/TZVPP+ level of theory with the respective thermodynamic corrections (T=298.15 K) determined on a RI-MP2/TZVP level of theory. All values are given in kJmol^{-1}.

Dimer	MMFF1	MMFF2	MMFF4	OPLS-AA1	DZ1
ΔE	+224.7	+220.6	+218.6	+151.3	+156.5
BSSE	-16.3	-16.2	-15.8	-16.8	-16.9
$\Delta E_{\text{(BSSE corr.)}}$	+208.4	+204.4	+202.8	134.5	+139.6
ΔH_{corr}	-5.8	-5.3	-5.0	-3.8	-2.7
$-T\Delta S_{\text{corr}}$	-67.2	-66.0	-65.1	-60.2	-60.1
ΔG_{corr}	+135.4	+133.1	+132.7	+70.5	+76.8

Table 12 Comparison between counterpoise corrected electronic dimerization energies calculated for optimized structures on DFT and MP2 level of theory (TZVPP+ basis). All energies are given in kJ mol^{-1}.

Dimer	MMFF1	MMFF2	MMFF4	OPLS-AA1	DZ1
ΔE_{corr}(B3LYP)	+180.7	+183.0	+187.1	+141.4	+147.2
ΔE_{corr} (MP2)	+208.4	+204.4	+202.8	+134.5	+139.6
$\Delta\Delta E$	-27.7	-21.4	-15.7	+6.9	+7.6

4.2.2.4 Importance of molecular rigidity for the stability of the dimer

Compared to the arginine dimer (see **Figure 30**) the high stability of the 2-(guanidiniocarbonyl)-1H-pyrrole-5-carboxylate dimer (**Figure 31**) can be traced back to an improved hydrogen bonding network, the higher acidic strength of the NH's as well as the energy contents of the monomers. The energy contents of the arginine and the 2-(guanidiniocarbonyl)-1H-pyrrole-5-carboxylate monomer differ strongly since a stabilizing interaction between the oppositely charged ends can only take place in the flexible zwitterionic arginine. A comparison between arginine, an artificially stiffened arginine and the 2-(guanidiniocarbonyl)-1H-pyrrole-5-carboxylate allows an estimate of the importance of the various effects. As model system for the artificially stiffened arginine a conformer was chosen in which the methylene groups are arranged in an all-trans orientation (**Figure 31**). The dimerization energy of 410 kJmol^{-1} (RI-MP2/TZVPP+//B3LYP/6-311++G**) is about 180 kJmol^{-1} higher than for the regular arginine dimer (224 kJmol^{-1}), although, due to geometrical constraints, only one hydrogen bond can be formed within one salt bridge.[43] The 2-(guanidiniocarbonyl)-1H-pyrrole-5-carboxylate dimer possesses a dimerization energy of 581 kJmol^{-1} (RI-MP2/TZVPP+) which is once more about 170 kJmol^{-1} higher than for the stiffened arginine dimer system. This comparison shows that the distinct higher stability of the 2-(guanidiniocarbonyl)-1H-pyrrole-5-carboxylate dimer results to about 50% from the rigidity of the monomeric units. The rest is due to the improved H-bonding network and the increased acidity of the acyl guanidinium moiety.

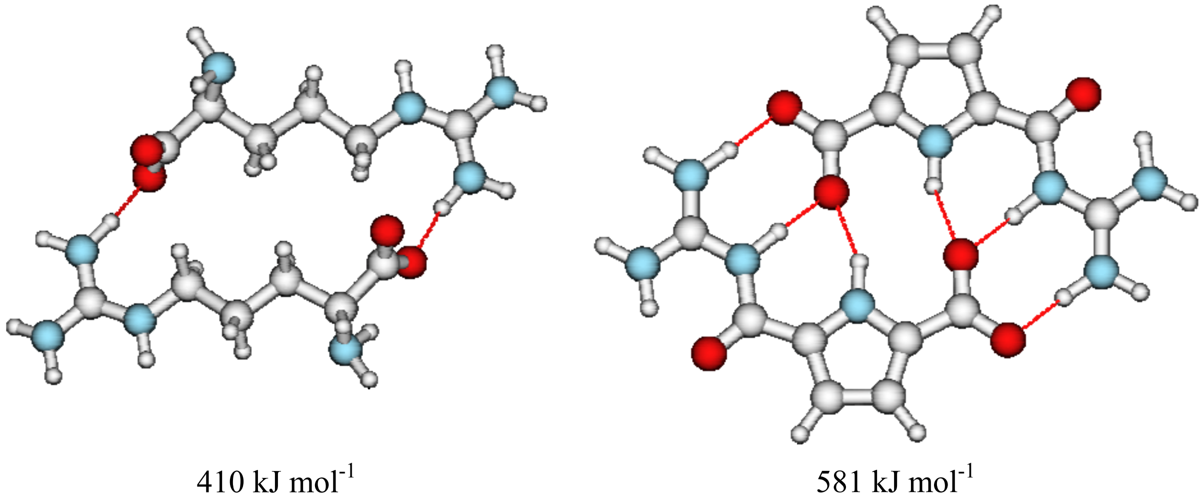

410 kJ mol^{-1} 581 kJ mol^{-1}

Figure 31 MP2/TZVPP+ calculated dimerization energies of an artificially linear arginine dimer (left) and 2-(guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate (right).

4.2.2.5 Conclusions

The present study shows that the rigidity of a molecule significantly influences its self-assembling properties. The prevention of stabilizing intra-molecular interactions within the monomers due to geometrical constraints strongly enhances the corresponding dimer stability. This was shown by comparing the self-assembly of arginine with that of 2-(guanidiniocarbonyl)-1H-pyrrole-5-carboxylate.

The importance of rigidity was analyzed by calculating the dimerization energy of an artificially stiffened arginine dimer system and comparing it with the dimerization energy of the 2-(guanidiniocarbonyl)-1H-pyrrole-5-carboxylate dimer.

The analysis shows that the high binding affinity of the 2-(guanidiniocarbonyl)-1H-pyrrole-5-carboxylate results to about 50% from the rigidity of the monomers which cannot be stabilized by intramolecular interactions and are therefore high in energy. As a result dimerization is more favourable for rigid monomers stabilizing the terminal charges. This effect should therefore be strongly considered when optimizing the complexation ability of artificial self-complementary systems. A similar effect might be expected also for flexible non-complementary systems in which other stabilizing interactions can take place intramolecularly.

4.3 Gas-Phase Assemblies with Novel Structure Motifs

4.3.1 Introduction

From early days until now chemical intuition and experience has guided experimentalists through the rational design of novel structure motifs. In the last years it was shown by several studies that theoretical chemistry can have an important share to the process of developing novel lead structures since possible candidates can be identified already in early stages before experimental studies have to be performed.[116] Besides the prediction of new substrates that show improved binding affinities towards certain receptors also the investigation and explanation of the binding modes make theoretical studies interesting for experimentalists.
Lately, it has become apparent that it is of utmost interest to find new lead structures for an efficient anion binding in polar media.[3,4] By now only very few synthetic receptors are capable to form stable non-covalent interactions with an anionic substrate in highly competitive media like aqueous solutions.[6-10,12] In the end of the 1990's Schmuck developed the 2-(guanidiniocarbonyl)-1*H*-pyrroles as novel lead structures for carboxylate receptor systems that improve the ion pairing of simple guanidinium cations with oxo anions through a combination of multiple hydrogen bond patterns and increased acidity of the acyl guanidinium moiety (see **Figure 2**).[11,13] It could be shown that 5-subsituted carboxylate derivates are able to form stable dimer systems in DMSO or even water with association constants of approximately $K > 10^8$ M^{-1} and $K = 170$ M^{-1} ($\Delta G \approx -15$ kJ mol^{-1}), respectively.[14] Therefore compound **1** is one of the most efficient self-assembling systems relying solely on electrostatic interactions reported so far. In chapter 4.1 we investigated the various contributions of the hydrogen bonding network as well as the importance of the salt bridge by a comparative study of knock-out analogues of compound **1**. That the rigidity of a zwitterionic monomeric unit is of decisive importance for an effective self-assembly has been shown by investigations on the structurally related arginine dimer systems (chapter 4.2). A rigid scaffold prevents the charged ends of the zwitterionic monomers to interact intramolecularly which stabilizes the monomers energetically and thus results in a smaller energy gain upon intermolecular complexation. With these concepts at hand we are now able to build up new guanidinium-based zwitterionic carboxylate receptors with variations in the rigid scaffold in order to enhance the willingness for self-assembly. The questions being subject of the present study are:

1. How important is the preorganisation or geometrical orientation of the carboxylate and guanidinium moiety towards each other?
2. What influence does the carbonyl group in vicinity of the guanidinium group have on its acidity?
3. How important is the dipole moment for the aggregation strength and to what extent does delocalization over the entire molecule play a role?
4. How large are molecular solvents effects and is there a way to treat them approximately with quantum chemical calculations?

In order to answer these questions a great number of novel structures has been generated and the respective dimerization energies were calculated on a density functional level of theory. The emphasis hereby was less on creating synthetically available compounds than on trying to understand the principles of the binding modes of an effective anion receptor. Moreover, it should be stressed that only the potential energy values of the compounds relative to each other are of interest as only a qualitative description of their dimerization properties can be given. A quantitative interpretation is very delicate and should be avoided since solvation is taken into account only by a continuum ansatz. Therefore it is impossible to get a realistic description of the binding situation present in an explicit solvation shell. Also dynamic effects and thus entropic contributions play an important role in the association process and are only treated approximately by analytic formulas originally derived for ideal gases.

4.3.2 Theoretical Details

The novel structure motifs have been pre-optimized on a force-field level (MMFF94/PRCG minimization) using the Macromodel V8.0 suite.[98,101] In some cases a conformational search has been necessary in order to detect all low-lying minimum conformers. This step was performed employing the mixed MCMM/LowMode algorithm.[39,40] Due to the high rigidity of most of the designed structures a total step number of 1000 has been proven to be sufficient.

The DFT-based structure optimizations of all compounds were performed with the TURBOMOLE program package[65] at the BLYP/TZVPP+ level of theory[35b,e,66] using the RI approximation.[67] The standard TZVPP basis set was hereby enlarged by 1s and 1p primitive uncontracted basis functions with an exponential coefficient of 0.068, whereas for the auxiliary basis sets the exponent was doubled (0.136). The influence of a solvent was estimated for some compounds using the COSMO[43] approach as implemented in

TURBOMOLE[69b] with appropriate dielectric constants for DMSO ($\varepsilon = 46.7$), methanol ($\varepsilon = 32.63$) and water ($\varepsilon = 78.0$).

On order to estimate the molecular effects of explicit solvent molecules some calculations have been performed with either a DMSO or a methanol molecule surrounded by gas phase or a continuum. Hereby, the positioning and structures of the non-covalently bound solvent molecules were obtained from force-field based conformational searches (mixed MCMM/LowMode/5000 steps). The structures for quantum mechanical calculations were selected by visual inspection.

Thermodynamic corrections of solvated structures at $T = 298.15$ K were obtained by numerical differentiation employing the NumForce module as implemented in the Turbomole V5.8 suite. Previous studies have shown that for the given basis set size and level of theory the basis set superposition error is small by only 5 to 10 kJ mol^{-1} and thus has only little influence on the accuracy of the DFT calculated energies. This holds in particular for dimerization energies in solution which were obtained by the continuum ansatz which limits the interpretability of these values to a qualitative extent. We therefore refrained from evaluating the counterpoise-corrected dimerization energies and hence only the uncorrected values are given.

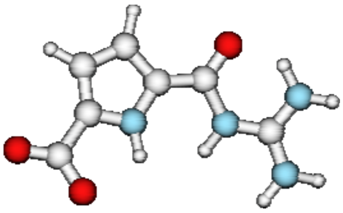
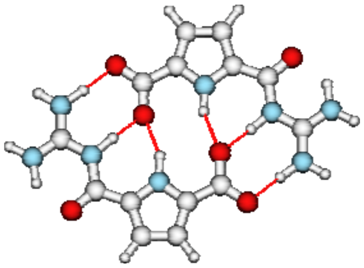
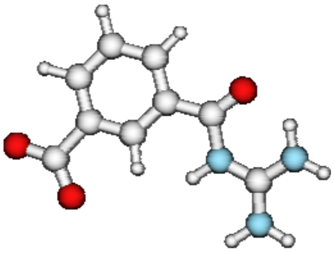
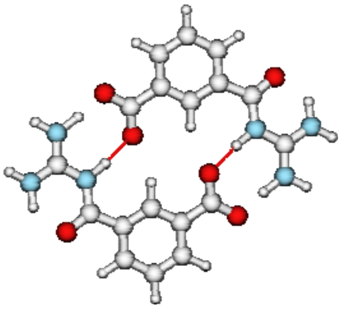
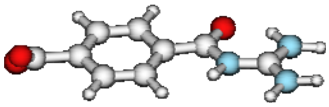
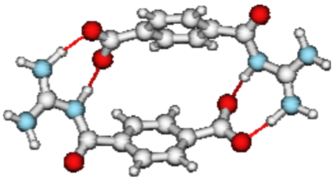
4.3.3 Novel Structure Motifs

Starting from simple six-membered aromatic ring systems with substitutions either in para other meta position the scaffold was extended to condensed aromatic systems and decoupled π-systems represented by the biphenylene derivatives. To investigate the importance of molecular solvent effects a more detailed study on *m*-carbonyl-guandiniobenzoate and the original pyrrole structure 2-guanidinio-carbonylpyrrole-6-carboxylate was performed elucidating the apparently contradictory experimental results.

4.3.3.1 6-Membered Ring Structures

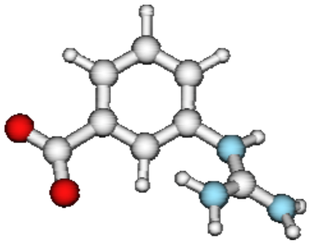
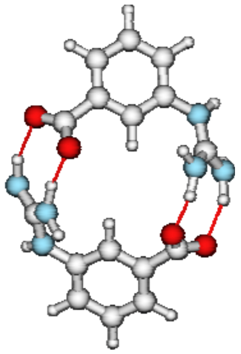
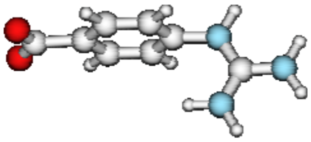
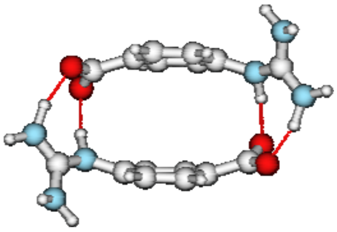
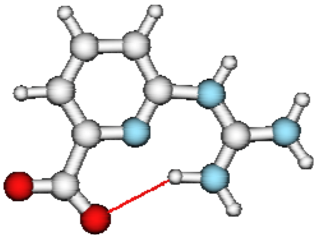
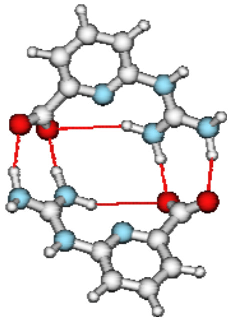
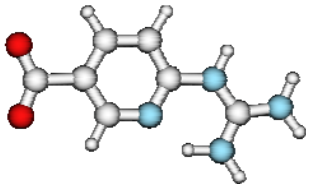
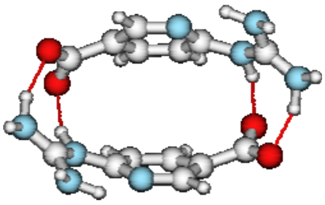
The simplest interchange of the aromatic pyrrole ring system is the transformation to a benzene derivative as indicated by the graphical representation in **Table 13**. The dimerization energies given are electronic energy differences between monomers and dimer in gas phase without correcting for the basis set superposition error. Surprisingly, the removal of the pyrrole hydrogen bond donor affects the dimerization energy only by less than 5 % being in disagreement to experimental observations predicting the structure **7a** to form much less stable aggregates in DMSO or DMSO/methanol mixtures (see chapter 4.3.3.5). The para substituted benzene derivative **7b** shows already a diminution of dimerization energy to about -380 kJ mol^{-1}. This clearly demonstrates the importance of the formation of unstrained and linearized hydrogen bonds which cannot be compensated (at least under gas phase conditions) by increased dipole moments of the monomeric building blocks (19.9 vs. 23.9 D).

Table 13 Gas phase dimerization energies (in kJmol^{-1}) and optimized structures of guanidiniocarbonyl-carboxylate ring systems calculated on a RI-BLYP/TZVPP+ level of theory (no CP).

Name	Energy	Structure monomer	Structure dimer
2-Guanidinio-carbonylpyrrole-6-carboxylate **1**	-443		
m-Guanidinio-carbonylbenzoate **7a**	-424		
p-Guanidinio-carbonylbenzoate **7b**	-378		

The removal of the carbonyl function separating the ring system from the hydrogen donating guanidinium group results in even more destabilized dimer structures, although the difference between meta and para substitution vanishes. As depicted in **Table 14** both structures (**7c**, **7d**) show more strained geometries that can be seen by the twisted guanidinium groups in the monomers leading to less directed hydrogen bonds in the dimer.

Table 14 Gas phase dimerization energies (in kJ mol^{-1}) and optimized structures of guanidiniocarboxylate ring systems calculated on a RI-BLYP/TZVPP+ level of theory (no CP).

Name	Energy	Structure monomer	Structure dimer
m-Guanidinio- benzoate **7c**	-331		
p-Guanidinio- benzoate **7d**	-330		
1-Guanidinio- pyridine-6- carboxylate **7e**	-278		
1-Guanidinio- pyridine-5- carboxylate **7f**	-283		

An undistorted planar geometry should therefore result in less deformed dimer geometries and thus give more stabilized aggregates. However, the pyridine based structure **7e** clearly shows the dilemma of forcing a planar monomeric unit. In order to avoid the repulsive interaction between the benzene hydrogen and the guanidinium group one has to substitute this atom group by a single sp^2-hybridized nitrogen atom. This atom can now serve as further hydrogen bond acceptor and thus stabilizes the monomer intramolecularly. This stabilizing effect of is in competition with the intermolecular interactions during the aggregation processes and therefore reduces the dimerization energy to less than -280 kJmol^{-1} in gas phase. The para substituted pyridine derivative **7f** again gives similar energies which can be explained by the slightly favourable situation in the monomer which is, in contrast to **7e,** not additionally stabilized by the proximate carboxylate group.

4.3.3.2 Enlarging the π-System: Anellated Aromatic Rings

The influence of the dipole moment plays an important role as structures **7d** and **7f** in **Table 14** could already show. This effect should be amplified in polar solvents which stabilize and increase partial charges in molecules. A simple proof is to enlarge the aromatic scaffold by further rigid anellated aromatic rings as shown in **Table 15** and **Table 16**. A nice trend is observed rising from -424 kJ mol^{-1} for the benzene derivative **7a** to -521 kJ mol^{-1} for the anthracene compound **8b**. The non-linear increase in dimerization energies can be explained by the stronger stabilization of the benzene monomer in which the charged moieties are less separated in space and are thus stabilized stronger intramolecularly via direct interactions of the charged ends. Already in the naphthalene monomer the opposite charges are separated too far in order to experience a significant stabilization. A further elongation of the spacer like in anthracene has much less impact on the dimerization energy as observed for the naphthalene derivative **8a**. In summary, the charged moieties of a monomeric unit are stabilized via coulomb interactions which decay as the square of the distance. As seen for the 6-membered ring structures the dimerization energy is decreased by about 80 kJ mol^{-1} after removal of the carbonyl group (structure **8c)** which is a result of geometrical stress causing weakened hydrogen bonding. The planarization of the monomeric unit in the quinoline derivative **8d** again stabilizes the monomer and therefore makes a dimerization less favourable.

Table 15 Gas phase dimerization energies (in kJ mol^{-1}) and optimized structures of anellated aromatic ring systems calculated on a RI-BLYP/TZVPP+ level of theory (no CP).

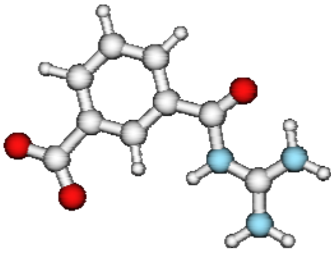
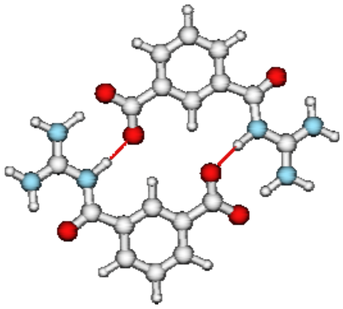
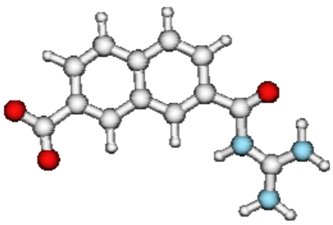
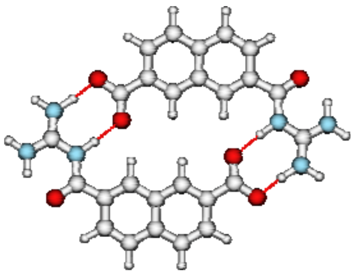
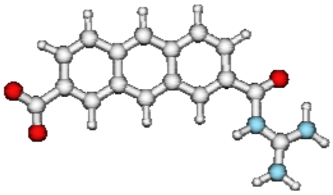
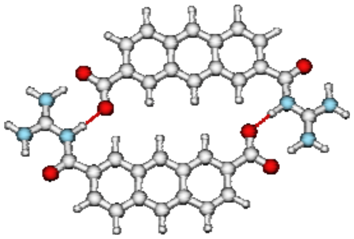
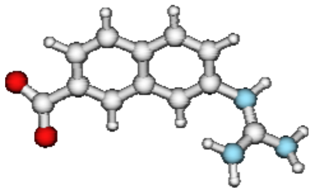
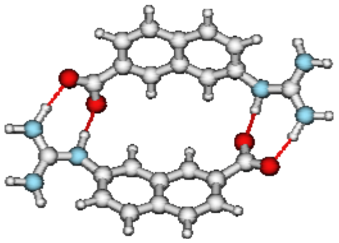
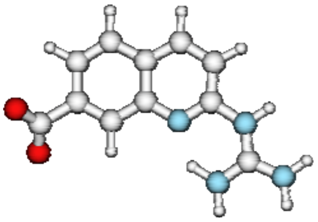
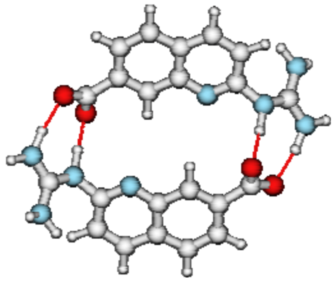
Name	Energy	Structure monomer	Structure dimer
m-Guanidiniocarbonylbenzoate **7a**	-424		
2-Guanidiniocarbonylnaphthalene-7-carboxylate **8a**	-498		
2-Guanidiniocarbonylanthracene-7-carboxylate **8b**	-521		
2-Guanidinionaphthalene-7-carboxylate **8c**	-416		
2-Guanidinioquinoline-7-carboxylate **8d**	-312		

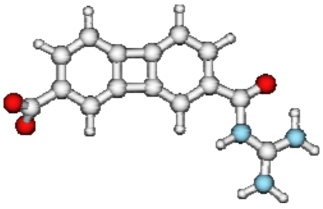
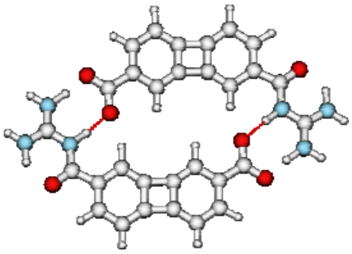
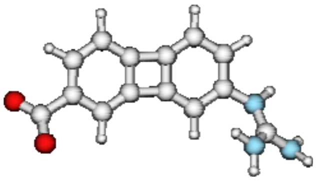
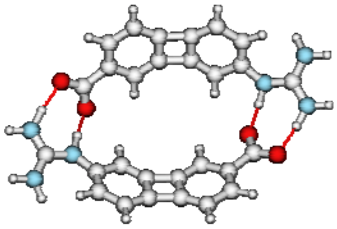
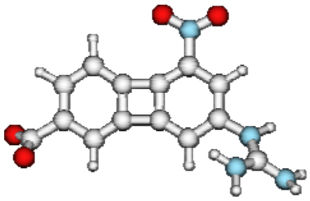
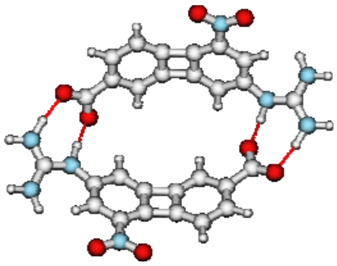
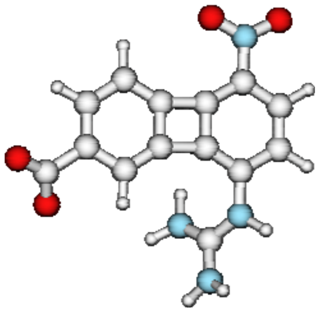
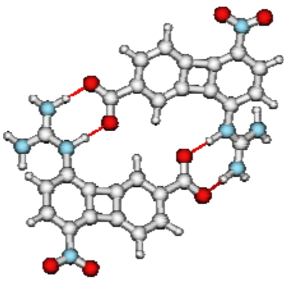
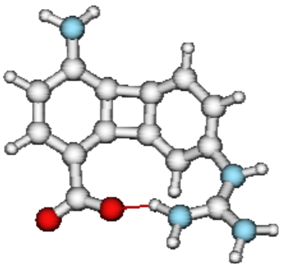
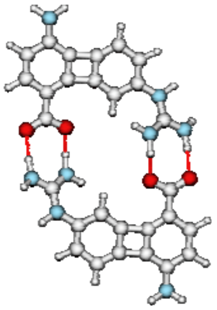
Table 16 Dipole moments (in Debye) and dimerization energies (in kJ mol^{-1}) of selected anellated ring structures.

Monomer structure		ΔE_{Dim}	Dipole moment
m-Guanidiniocarbonylbenzoate	**7a**	-424	19.9
2-Guanidiniocarbonylnaphthalene-7-carboxylate	**8a**	-498	25.2
2-Guanidiniocarbonylanthracene-7-carboxylate	**8b**	-529	29.7

4.3.3.3 Decoupled π –Systems: Biphenylene Derivatives

In order to investigate the influence of the conjugated π-system biphenylenes can be used to set up monomeric units with electronically decoupled aromatic acceptor and donor moieties. **Table 17** shows the dimer of the guanidiniocarbonyl derivative **9a** which is 17 kJ mol^{-1} less stable as the structurally related naphthalene compound **8a**, although the latter possesses a slightly smaller dipole moment (26.0 vs. 25.2 D). Like before, the direct attachment of the guanidinium group to the ring system (**9b**) reduces the gas phase energy of aggregation by further 40 kJ mol^{-1}. In principle, an electron withdrawing group at the aromatic ring in the donor site should increase the acidity of the guanidinium group. However, substitution of a nitro group in meta position to the guanidinium group (**9c**) has no considerable influence on the dimerization process which is even slightly decreased contrariwise (-426 vs. -430 kJ mol^{-1}). The para substitution of either a nitro/guanidinium group (**9d**) or an amino/carboxylate group (**9e**) both show no improvement over the 2-guanidinio-biphenylen-7-carboxylate **9b**, although the dimer structures feature both less bended geometries with structurally complementary hydrogen donor and acceptor sites. However, the close intramolecular vicinity of the carboxylate and guanidinium group again stabilizes the monomeric unit too much so that the dimerization process is less favourable.

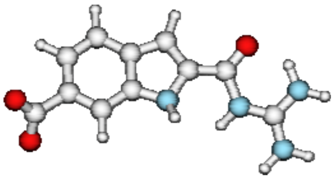
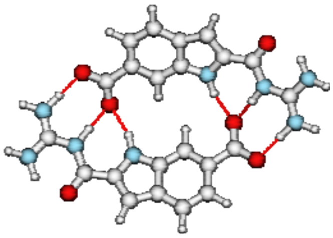
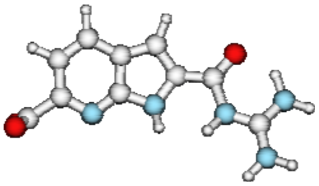
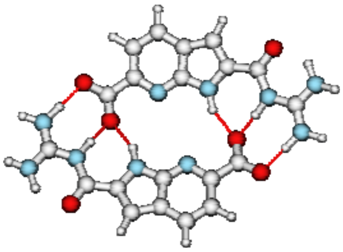
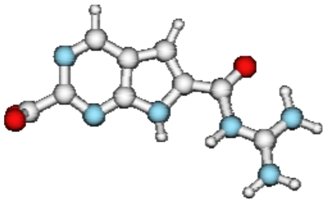
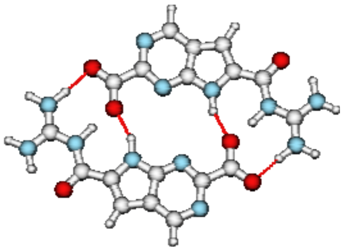
Table 17 Gas phase dimerization energies (in kJ mol^{-1}) and optimized structures of biphenylene derivatives calculated on a RI-BLYP/TZVPP+ level of theory (no CP).

Name	Energy	Structure monomer	Structure dimer
2-Guanidinio-carbonyl-biphenylen-7-carboxylate **9a**	-481		
2-Guanidinio-biphenylen-7-carboxylate **9b**	-438		
3-Guanidinio-1-nitroso-biphenylen-6-carboxylate **9c**	-430		
4-Guanidinio-1-nitroso-biphenylen-6-carboxylate **9d**	-426		
1-Amino-6-guanidinio-biphenylen-4-carboxylate **9e**	-410		

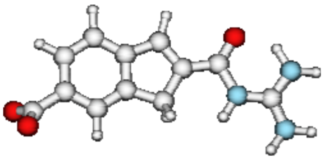
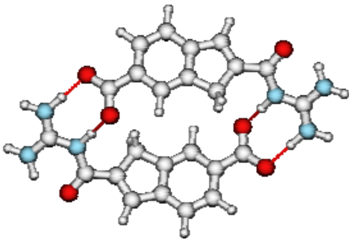
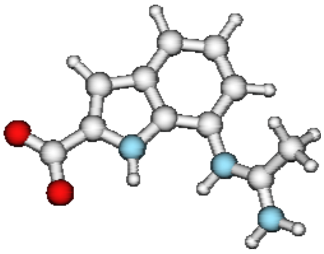
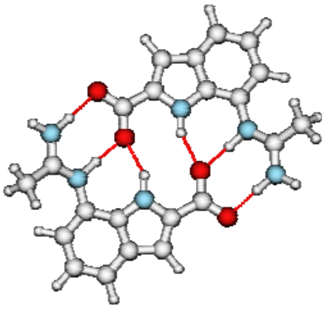
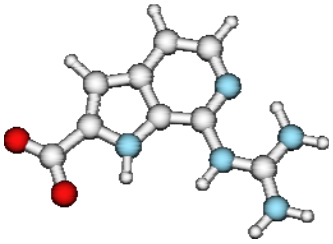
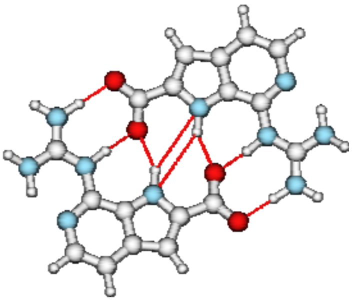
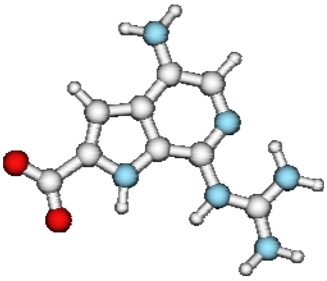
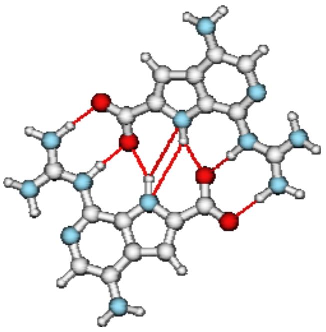
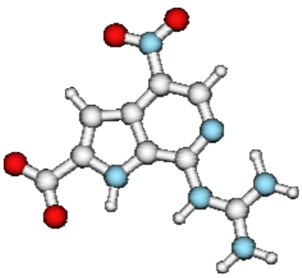
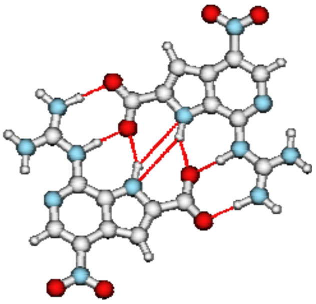
4.3.3.4 Importance of Preorganisation and Fixation: Indole-Derivatives

The previous structure motifs already showed that the carbonyl function separating the ring system and the guanidinium group is of central importance for an effective carboxylate binding. The reason therefore seems to be primarily a geometrical advantage as in the dimer system both monomers show a good complementarity leading to more or less planar structures with directed hydrogen bonds (see structures **1**, **7a**, **8a**, **8b**, **9a**). Hence, the good preorganisation in conjunction with an increased dipole moment should improve the binding affinity of the initial structure **1**. The interchange of the pyrrole ring by the enlarged indole ring system leads to structure **4a** (see **Table 18**) that indeed improves the binding energy by about 90 kJmol^{-1} in comparison to structure **1**.

Table 18 Gas phase dimerization energies (in kJ mol^{-1}) and optimized structures of indole derivatives calculated on a RI-BLYP/TZVPP+ level of theory (no CP).

Name	Energy	Structure monomer	Structure dimer
2-Guanidinio-carbonylindole-6-carboxylate **10a**	-530		
7-Aza-2-guanidinio-carbonylindole-6-carboxylate **10b**	-527		
5,7-Diaza-2-guanidinio-carbonylindole-6-carboxylate **10c**	-529		

Continuation of ***Table 18***

2-Guanidinio-carbonylindene-6-carboxylate **10d**	-422		
7-Amidinio-indole-2-carboxylate **11a**	-442		
6-Aza-7-guanidinioindole-2-carboxylate **11b**	-431		
4-Amino-6-aza-7-guanidinioindole-2-carboxylate **11c**	-420		
6-Aza-7-guanidinio-4-nitrosoindole-2-carboxylate **11d**	-453		

Attempts to increase the binding affinity by destabilizing the monomer with the insertion of one or two nitrogen atoms in α – position of the carboxylate carbon atom (see structure **10b** and **10c**) failed. In contrast to the regular indole derivative **10a** both azaindole monomers show twisted carboxylate groups perpendicular to the ring systems improving the electron density overlap with the lone pair of the nitrogen atom. This stabilization is lost by dimerization during which the carboxylate groups are rotating back into the molecular plane. Moreover, the two oppositely lying lone pairs of the nitrogen atoms of the monomers lead to additional repulsive interactions making these structures less favourable.

The importance of the pyrrole N-H donor group can be seen by structure **10d** which is about 100 kJ mol^{-1} less stable than the dimer structure **10a**. Here, the indole scaffold has been substituted by an indene scaffold showing no interaction with the carboxylate oxygen atoms. Besides the knock-out of two hydrogen bonds which have an additional entropic share in the high dimerization energy of structure **10a** (chelate effect) also the aromaticity is distorted in the indene dimer. However, this distortion has only a slight influence on the dipole moment of the monomer **10d** (25.1 D) which is decreased by only 0.3 D in comparison to monomer **10a** (25.4 D).

In summary, the present results indicate that the strong stability of the dimer **10a** arises from the cooperative, salt bridged hydrogen bonding network, the rigid backbone which separates the opposite charges in the zwitterionic monomer effectively and thus induces a higher dipole, as well as the perfect complementarity and steric orientation between the guanidinium and carboxylate groups.

The insertion of the carbonyl function separating the ring system from the guanidinium group seems to be very important for the steric orientation of the hydrogen donor and acceptor site, whereas the strong acidity which is expected to highly contribute to the dimer stability has not yet been determined. To check the effect of the carbonyl function on the acidity of the guanidinium group the structures **11a** to **11d** have been calculated. The structures are characterized by a substitution of the carbonyl function with an aromatic ring system consisting either of a benzene or pyridine ring attached to the pyrrole ring. Although they therefore represent indoles or azaindoles, their dimerization energies must be compared to structure **1** (ΔE_{Dim} = -443 kJ mol^{-1}) instead of structure **10a**. Structure **11a** possesses the same binding energy as the guanidiniocarbonylpyrrole **1** showing that the increase of acidity of the hydrogen donor ability by the adjacent carbonyl function is of minor importance for the strong dimerization. However, it should be noted that structure **11a** differs from structure **1** in the hydrogen donor site as the guanidinium group has been replaced by an amidinium group.

The modification of monomer **11a** with a guanidinium group leads to strong repulsive interactions of the NH_2 rest with the hydrogen atoms of the 6-membered ring. This implicates a twist of the entire guanidinium group with respect to the molecular plane leading to an intramolecular hydrogen transfer to the carboxylate group. Of course, this effect is strongly advantaged by the long range charge interactions in gas phase and should not occur in solvation. In order to prevent such an undesirable proton transfer the relevant amino group is substituted by a methyl group (see structure **11a**) and therefore strongly resembles the “knock-out” analogue **3b** (see chapter 4.1). This analogue of structure **1** is the rotamer of structure **3a** in which one of the amide groups of the guanidinium moiety has been replaced by a methyl group. In comparison to the guanidiniocarbonylpyrrole **1** one sees that the dimerization of monomer **11a** benefits from a stronger destabilized monomer structure which additionally possesses a more localized positive charge distribution on the amidinium nitrogen atoms than for the strongly delocalized guanidinium group. This effect could be observed for the “knock-out” structure **3b** which showed a slightly higher dimerization energy even in solution than the reference structure **1** (-66 vs. -56 kJ mol^{-1}).

An unfavourable twisted structure is avoided in monomer **11b**. It consists of an azaindole in which the interfering CH unit is substituted with a nitrogen atom. **Table 18** clearly shows that in comparison to structure **11a** the association is only very slightly affected (ΔE = -11 kJ mol^{-1}). Although the monomeric unit is stabilized by a strong intramolecular hydrogen bond, the dimerization energy should not be affected since this stabilization is also present in the dimer. Actually one would expect a slight increase in the binding energy as the electron negative nitrogen atom raises the acidity of the guanidinium group. By the positioning of electron pushing or electron pulling groups at the electron poor aromatic ring system one can either reduce (structure **11c**) or increase (structure **11d**) the dimerization affinity. The nitro substituted structure **11d** even possesses a higher dimerization energy than structure **1.** It is important to note that the fixation of the rotation around the pyrrole – carbonyl carbon bond as possible in structure **1** is prevented in all structures **11a** to **11d** which should increase in general the affinity for dimerization as the stiffening of the scaffold leads to a smaller entropic lost upon complexation.

In summary, on the basis of gas-phase calculated electronic dimerization energies structure **10a** seems to be the most promising candidate for improving the dimerization affinity of guanidinium based carboxylate receptors. As shown by structures **11a** to **11d** the carbonyl function separating the guanidinium group from the ring system seems to be mainly important for a good preorganisation leading to an effective carboxylate binding and less for increasing the acidity. Calculations of azaindole derivatives which substitute the more electron negative oxygen atom by an aromatic nitrogen atom (structures **11b** to **11d**) indicate that the electronic influence is only of minor importance on the aggregation energy. The calculation of a 7-guanidinioindole-2-carboxylate monomer in gas phase failed due to an intramolecular proton shift. However, solvation effects should stabilize this structure and thus an experimental determination of the dimerization energy would give a clear answer about the importance of the carbonyl function with respect to the hydrogen donor activity of the guanidinium moiety.

4.3.3.5 Benzene vs. Pyrrole Derivative: Molecular Solvation Effects

The association constants of the pyrrole derivative **7a** in DMSO/methanol mixtures obtained experimentally by isothermal titration calorimetry experiments (ITC) are given in **Table 19**. In contrast to the gas phase calculations (see **Table 13**) the experimental findings show that the benzene derivative *m*-guanidiniocarbonyl-benzoate (**7a**) forms much less stable aggregates in DMSO/methanol mixtures than structure **1** in DMSO. With an estimated dimerization constant K_{dim} of at least 10^8 M^{-1} the free enthalpy difference ΔG_{298} of the pyrrole derivative is calculated to -45 kJ mol^{-1}, so that the energy difference between **1** and **7a** can be assumed in a rough guess to be in the range of 30-40 kJ mol^{-1} (see **Table 19**). Surprisingly, with an increasing amount of methanol the measured free enthalpy difference of **7a** even rises, although methanol is known for weakening non-covalently bound aggregates stronger. It formally possesses a smaller dielectric than DMSO, but due to the protic character of the hydroxyl group it is able to form competitive hydrogen bonds to the solute. The enthalpic contribution to the overall Gibbs free energy declines for the benzene derivative **7a** in DMSO/methanol mixtures with increasing percentage of methanol, whereas the free enthalpy itself arises pointing to an entropy steered reaction. Regarding solely the relative gas phase dimerization energies between **1** and **7a**, the experimentally observed energy difference is clearly underestimated ($\Delta E \approx 20$ kJ mol^{-1}) although in the right order of magnitude.

Table 19 Experimental K_{diss} and ΔH values obtained by ITC measurements of association constants and the calculated enthalpy contributions to the Gibbs free energy (in kJ mol^{-1}) at T = 295.15 K.

	K_{diss} [mM]	ΔH	K_{dim} [M^{-1}]	ΔG
25 % DMSO in methanol	0.23	-26.3	4580	-20.7
50 % DMSO in methanol	0.54	-27.1	1840	-18.4
100 % DMSO[†]	*2.8 ± 1.4*	*-2.5 ± 0.2*	*513*	*-15.3*

[†] The K_{diss} values measured in pure DMSO are not really significant due to their large variance. NMR-titration experiments gave an dimerization constant K_{Dim} = 4

Table 20 Calculated dimerization energies (B-LYP/TZVPP+) of structures **1** and **7a** in gas phase and solution. The solvent is simulated by either a pure continuum model or with an additional explicit solvent molecule. All values are given in kJ mol^{-1}.

Continuum dimerization energy

	ε	2-Guanidiniocarbonyl-pyrrole-6-carboxylate	*m*-Guanidiniocarbonyl-benzoate
Gas phase	0	-443	-424
DMSO	46.7	-112	-79
Methanol	32.63	-119	-85

Explicit solvation energy of monomers

	ε	2-Guanidiniocarbonyl-pyrrole-6-carboxylate	*m*-Guanidiniocarbonyl-benzoate
+ 1 DMSO	0	-133	-135
	46.7	-22	-18
+ 1 Methanol	0	-96	-116
	32.63	-16	-22

Dimerization energy of explicitly solvated monomers

	ε	2-Guanidiniocarbonyl-pyrrole-6-carboxylate	*m*-Guanidiniocarbonyl-benzoate
DMSO	0	-276	-241
	46.7	-95	-71
Methanol	0	-338	-274
	32.63	-111	-68

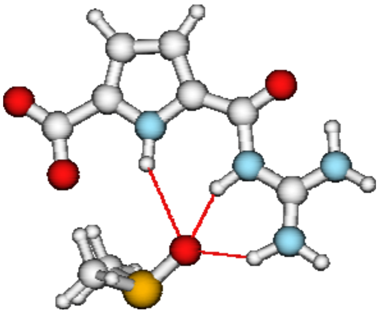
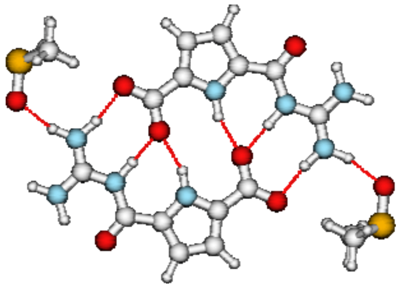
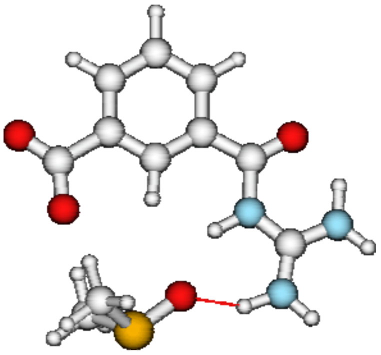
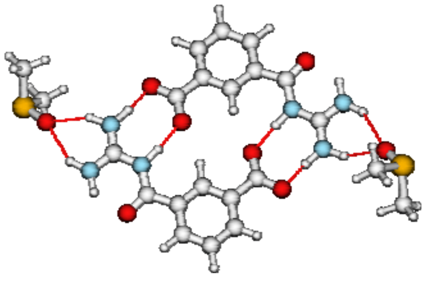
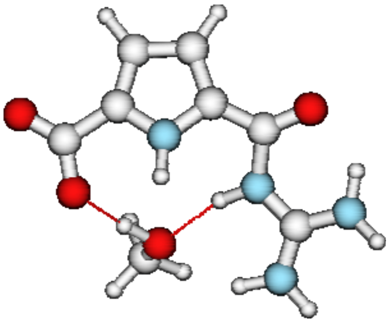
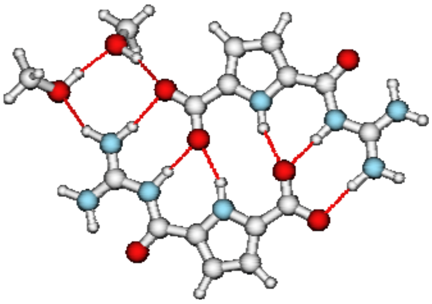
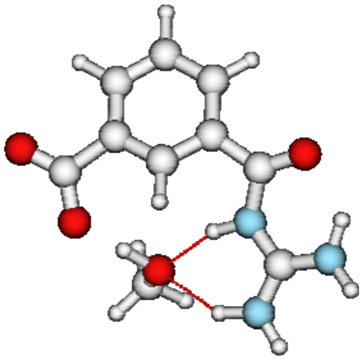
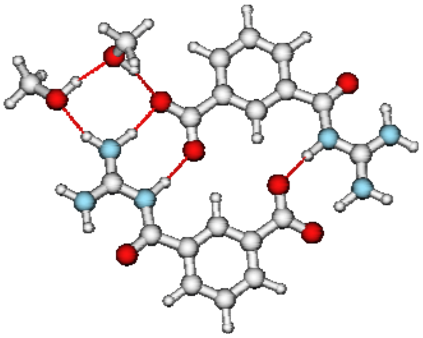
	Structure monomer	**Structure dimer**
1		
7a		
1		
7a		

Figure 32 Gas-phase optimized (B-LYP/TZVPP+) structures of monomers and dimers of **1** and **7a** explicitly solvated by one DMSO (top) or one methanol (bottom) per monomeric unit.

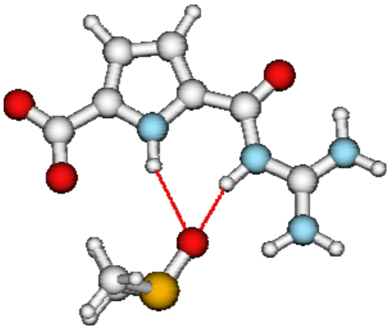
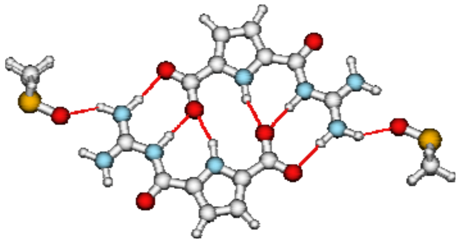
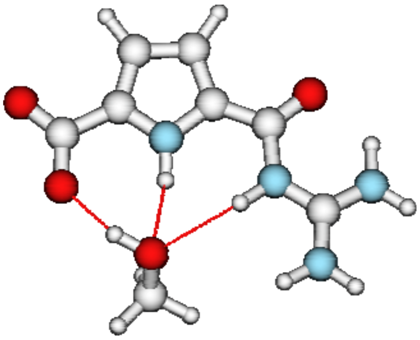
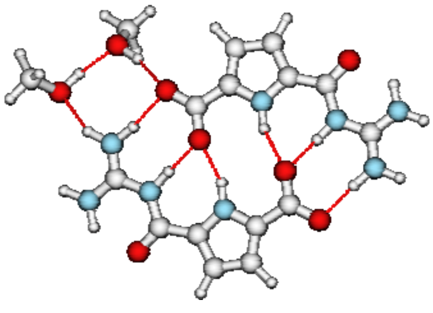
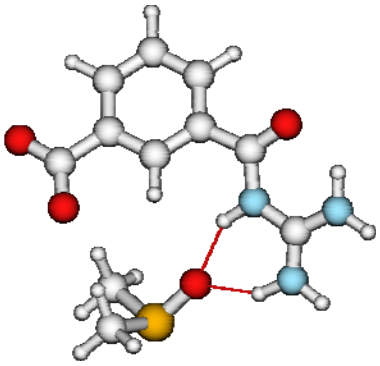
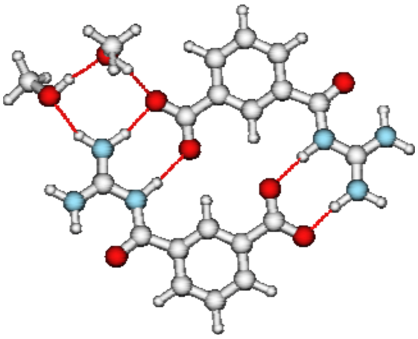
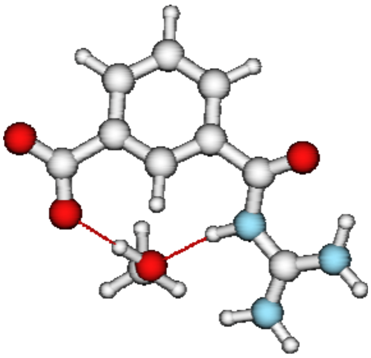
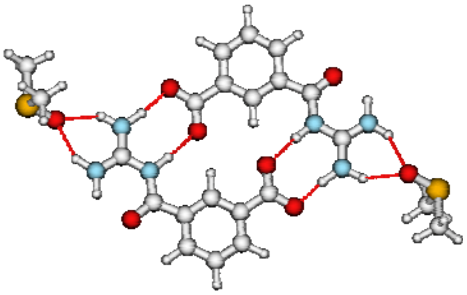
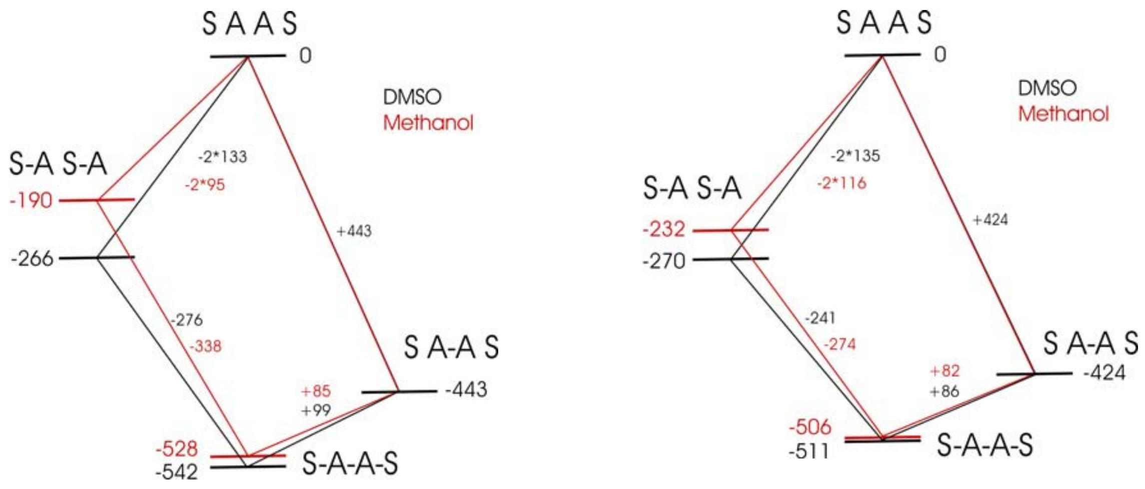
	Structure monomer	Structure dimer
1		
7a		
1		
7a		

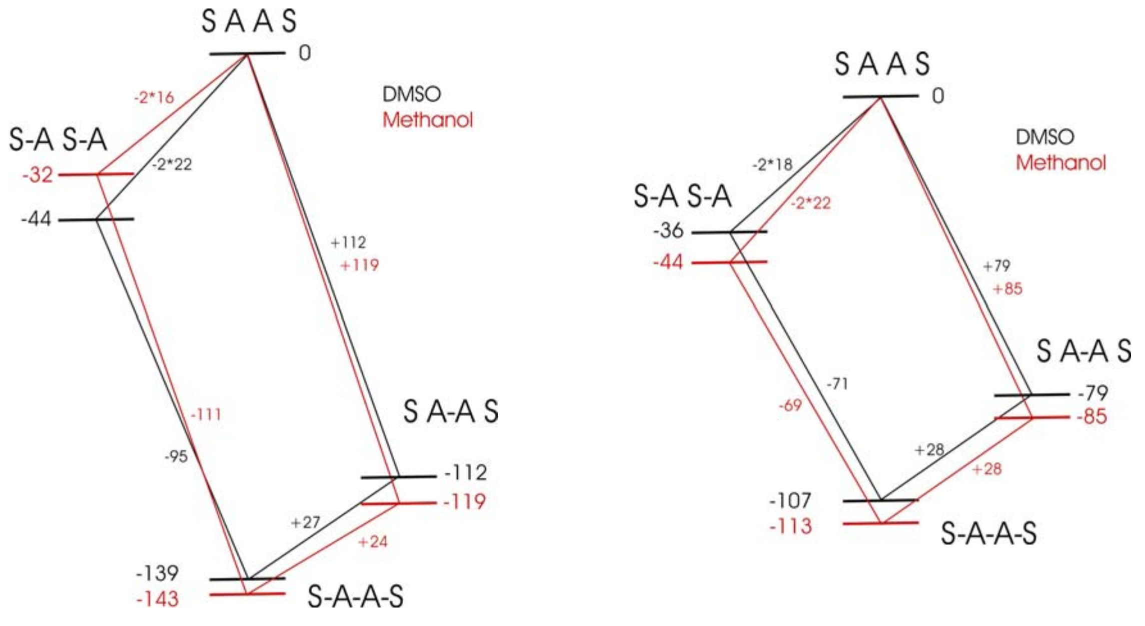
Figure 33 B-LYP/TZVPP+ optimized monomers and dimers of **1** and **7a** explicitly solvated by one DMSO (top) or one methanol (bottom) per monomeric unit and further solvated by a continuum model (COSMO) of the solvent with ε = 46.7 and 32.63 for DMSO and methanol.

To estimate the effect of solvation on the stability of the dimers the electronic energies of structures **1** and **7a** have been calculated on a B-LYP/TZVPP+ level of theory (**Table 20**) employing a continuum model (COSMO) for the solvent with the respective dielectrics for DMSO ($\varepsilon = 47.6$) and methanol ($\varepsilon = 32.63$). As one can see from **Table 20** the continuum solvation drastically reduces the dimerization energies of both compounds by several hundreds kJ mol^{-1}. However, the difference between the dimerization energies of **1** and **7a** remains nearly unchanged. This causes a larger relative difference between the pyrrole and the benzene derivative for solvation, whereas in gas phase the difference of both dimerization energies is small in relation to the overall magnitude. Moreover, in solution the benzene derivative **7a** is slightly more destabilized resulting in a relative energy difference between both derivatives of about 30 kJ mol^{-1} in both solvents. Although the relative electronic energies agree quite well with experimental observations, the unexpected inversion of the dimer stability of the benzene derivative by the increasing addition of methanol to DSMO can not be explained by this simple continuum model. Obviously molecular solvent effects have to be taken into account which seem to be important for the dimer stability. For this reason calculations were performed which include one explicit solvent molecule, *i.e.* one DMSO or one methanol, respectively. **Table 20** shows the binding energies of the solvent molecules to the monomers as well as the dimerization energies of the explicitly solvated monomers in gas phase or solvation. The DFT optimized structures are given in **Figure 32** for gas phase and in **Figure 33** for continuum solvation. The DMSO molecule binds significantly stronger to both monomers in gas phase than a single methanol due to its larger dipole moment (~ 4 D) that counteracts the large dipole of the zwitterionic monomers. However, in a dielectric environment this effect is negated and the binding of a methanol to a benzene derivative becomes even more favourable than for DMSO (-22 vs. -18 kJ mol^{-1}). Regarding the explicitly solvated dimerization energies of *m*-guanidiniocarbonylbenzoate (**7a**) the energies obtained for methanol and DSMO seem to be equally high, whereas for the pyrrole derivative (**1**) solvation in methanol should yield much higher stabilization upon aggregation. This conclusion seems to be contradictory to experimental experience and a solution to this dilemma is only provided by regarding the full thermodynamic cycles of the dimerization processes (see **Figure 34**). In gas phase, the benzene derivative binds a methanol molecule significantly stronger than the pyrrole derivative (-232 vs. -190 kJ mol^{-1}) making the methanol solvated dimer nearly as stable as the DMSO solvated one (-506 vs. -511 kJ/mol). This trend can be observed also for the continuum solvated aggregates, which now even favour the methanol complexed monomer by about 8 kJ mol^{-1} and the dimer by about 6 kJ mol^{-1}. In a sufficiently diluted solution of a 50% DMSO/methanol mixture this energy differences correspond to Boltzmann distributions of about 1:25 and 1:11, respectively. However, reducing the percentage of DMSO in the solvent mixture entropically benefits the formation of methanol solvated *m*-guanidiniocarbonylbenzoate monomers as the possibility for the formation of the solvent complex increases statistically. Although the dimer formation of the explicit solvated monomers shows in case of methanol a slightly smaller binding energy than for DMSO (-69 vs. -71 kJ mol^{-1}), the entropy gain makes the dimer formation more favourable in mixtures with a larger percentage of methanol than DMSO.

Pyrrole derivative (1) **Benzene derivative (7a)**


S A A S
0
DMSO
Methanol
S-A S-A
-190
-266
-2*133
-2*95
+443
-276
-338
S A-A S
+85
+99
-443
-528
-542
S-A-A-S
S A A S
0
DMSO
Methanol
S-A S-A
-2*135
-2*116
-232
-270
+424
-241
-274
S A-A S
+82
+86
-424
-506
-511
S-A-A-S

Gas-phase Gas-phase

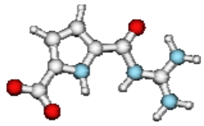
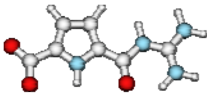
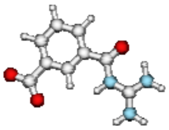
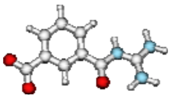

S A A S
0
DMSO
Methanol
-2*16
S-A S-A
-32
-2*22
-44
+112
+119
-111
-95
S A-A S
-112
-119
+27
+24
-139
-143
S-A-A-S
S A A S
0
DMSO
Methanol
-2*18
S-A S-A
-2*22
-36
-44
+79
+85
S A-A S
-79
-85
-71
-69
+28
+28
-107
-113
S-A-A-S

with COSMO with COSMO

Figure 34 Thermodynamic cycles of the dimerization process of structures **1** and **7a** (abbreviated as A) in the presence of one explicit DMSO or methanol solvent molecule (S) per monomeric unit. The electronic energies (in kJ mol^{-1}) given in the graphs are calculated on a B-LYP/TZVPP+ level of theory either in gas-phase or with a continuum model (COSMO) of the solvent with ε = 46.7 for DMSO and ε = 32.63 for methanol, respectively.

In order to obtain the Gibbs free energies of the dimerization processes the enthalpic and entropic contributions to the electronic energies have been calculated numerically on a B-LYP/TZVPP+ level of theory (see **Table 21**). The influence of the solvent on the thermodynamic corrections has been estimated by a continuum model with a dielectric of ε = 78 corresponding to an aqueous surrounding. This solvent has been chosen as for the 2-guanidiniocarbonylpyrrole-6-carboxylate (**1**) the dimerization constant K_{Dim} has been extrapolated from NMR titration experiments to 170 M^{-1} which corresponds to a Gibbs free energy of about 13 kJ mol^{-1}. In principle, two conformations in the zwitterionic monomer structure are possible: (1) the curved orientation of the guanidiniocarbonyl group that exhibits the same conformation as present in the dimer and (2) the linearized conformation that possesses a larger dipole moment. While in gas phase a high dipole moment destabilizes such conformers, solvation generally favours the opposite. This principle is best reflected by regarding the dimerization energies of the pyrrole derivative **1**. The continuum model stabilizes the linearized conformation stronger leading to a reduced dimerization energy. However, the benzene derivative **2** is more destabilized in the linear conformation although the dipole moment of this conformation is with 39 D much larger than in the curved geometry (29 D). The reason for this alleged anomaly lies in the destabilization of the monomer by a slight rotation of the guanidinium group due to the repulsive interaction between the backward amide hydrogen and the benzene hydrogen in para position. This close contact is not present in the linearized pyrrole derivative **1**. One might now assume that the linearized conformation of the monomeric pyrrole derivative **1** should be present in a realistic (*i.e.* explicit) solvation and the calculated Gibbs free energy of 15 kJ mol^{-1} seems to fit also perfectly to the experimental value (13 kJ mol^{-1}). However, test calculations with two explicit water molecules have shown a good fit of the solvent molecules into the binding pocket forming a hydrogen bonding network from the guanidinium moiety to the carboxylate oxygen atoms. In protic solvents it is therefore reasonable to presume a curved conformation which is stabilized mainly by molecular solvation effects. The large discrepancy between the experimental and calculated Gibbs free energy for the dimerization lies in the deficient description of thermodynamic corrections by continuum models. In contrast to absolute values the relative Gibbs free energies between two derivatives should be more meaningful as for both systems similar errors are made for the same solvent. In case of structure **1** and **7a** the free energies differ about 23 kJ mol^{-1} which would make the benzene dimer **7a** unstable in aqueous solution.

Table 21 Electronic energies and thermodynamic corrections for structures **1** and **7a** calculated on a B-LYP/TZVPP+ level of theory and employing a continuum model (COSMO) for water as solvent ($\varepsilon = 78.0$). All values are given in kJ mol^{-1}.

	Monomer conformation	ΔE	ΔH	-TΔS	ΔG
2-Guanidiniocarbonyl-pyrrole-6-carboxylate (**1**)		104	-2	-64	+39
		83	-3	-66	+15
m-Guandinio-carbonyl-benzoate (**7a**)		73	+1	-58	+16
		76	+0	-59	+18

4.3.3.6 Conclusion

In this chapter novel structure motifs that are able to bind carboxylate anions effectively have been examined by DFT gas-phase studies of dimer assemblies consisting of zwitterionic monomers. The 2-guanidiniocarbonylpyrrole-6-carboxylate (**1**) which shows a strong dimer stability with an association constants of $K \approx 170\ M^{-1}$ in water has been used as origin structure for improving the binding modes by substituting the rigid scaffold with various ring structures. In the previous chapters the effect of salt bridges, multiple hydrogen bonding patterns and rigid scaffolds on the dimer stability of guanidinium-based carboxylate receptors has already been reported. The focus of this comparative study was now to elucidate the influence of dipole moment, preorganisation, complementarity, delocalization and solvent effects on the dimer stability and to make predictions for an improved structure motif.

The influence of the dipole moment was clearly shown by the dimerization energies of various anellated ring structures. An increased separation of the carboxyl and guanidinium moiety results in larger dipoles which are more likely to aggregate. The decoupled biphenylene structures with additional electron pushing or withdrawing substituents could not achieve further improvement on the dimerization affinity. The importance of monomeric complementarity and unstrained dimer structures has been underlined by the benzene derivatives. As seen already for the dimerization of monomeric arginine (chapter 4.2.2) the intramolecular interactions present in both pyridine derivatives strongly reduce the dimerization energies. The indole derivative **10a** combines the advantages of a pyrrole based receptor with an increased dipole moment which makes this structure motif more stable in gas phase than the pyrrole derivative **1**. The carbonyl function separating the ring system from the hydrogen donor site is hereby responsible for the good preorganisation whereas the effect on the acidity of the guanidinium group is only of minor importance as seen for the azaindole derivatives. Furthermore, the high dimerization energies of the pyrrole and indole derivatives underline the importance of the pyrrole N-H hydrogen bond leading to an additional fixation of the carboxyl group in the binding pocket ("Gulliver effect").[17]

Finally, by comparison with experimental dimerization constants in varying DSMO/methanol mixtures showed the great influence of molecular solvation effects on the dimer stability in different solvents. A pure continuum model like COSMO is often not accurate enough to include all important solvent effects, but we could show that the inclusion of already one explicit solvent molecule can be sufficient to explain trends in the binding constants of *m*-guanidiniocarbonylbenzoate diluted in such solvent mixtures. The calculation of Gibbs free enthalpies is still an unsolved problem since the analytic formulas used here have been derived originally for ideal gases and are thus in principle inapplicable for condensed phases. However, the comparison of ITC measured dimerization constants in DMSO of the pyrrole derivative **1** and the benzene derivative **7a** with calculated dimerization energies showed that the relative energy differences are estimated quite well with a continuum model. The calculations of thermodynamic corrections can then give a hint about the stability of a given dimer system if only the relative Gibbs free enthalpy with regard to 2-guanidiniocarbonylpyrrole-6-carboxylate (**1**) is taken into account. Therefore, the 2-guanidiniocarbonylindole-6-carboxylate (**10a**) should be a promising candidate for an effective dimerization even in highly competitive media.

4.4 Cooperativity Effects in Supramolecular Assemblies – NMR Shift Studies in Adenosine-Carboxylic Acid Complexes

4.4.1 Introduction

Hydrogen bond interactions constitute a major driving force in the formation of specific molecular and complex geometries. Thus, protein and nucleic acid secondary and tertiary structural elements as well as many natural and artificial host-guest complexes are partly based on the directive power of intra- and intermolecular hydrogen bond formation. [117,118] In the past, NMR spectroscopic techniques have been established as a powerful tool to study the strength and geometry of hydrogen bonds in both the solid and the liquid state.[119] For the latter, however, NMR signals normally correspond to an average over fast exchanging hydrogen bonded species at ambient temperatures, thus restricting the detailed characterization of hydrogen bonds of individual hydrogen bonded associates.[120] However, employing deuterated freonic mixtures as NMR solvents allows high resolution NMR spectra to be acquired in the liquid state down to 100 K where the regime of slow hydrogen bond exchange within the NMR time scale is reached for even weakly hydrogen bonded systems.[121]

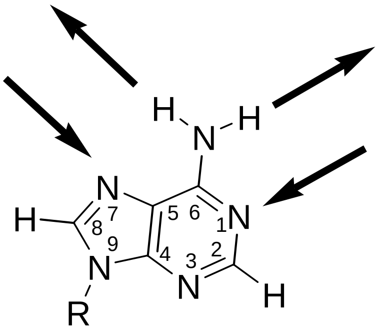


Figure 35 Adenine nucleobase with hydrogen bond donor and acceptor sites at the Hoogsteen (left) and Watson-Crick face (right).

Weisz et al. have recently presented low-temperature NMR studies on the binding of an acetic acid ligand to adenosine.[122] Because of its multiple functionalities, this nucleobase can engage in cyclic hydrogen bonds with a carboxylic acid at either its Hoogsteen or Watson-Crick site (**Figure 35**). In fact, only trimolecular complexes A·HAc$_2$ (**12a**, **Figure 36**) and A$_2$·HAc (**12b**,) with both Watson-Crick and Hoogsteen sites of the central adenine base occupied were observed at low temperatures. In all complexes the acetic acid OH proton is still localized closer to the oxygen of acetic acid in a neutral, non-ion pairing complex.

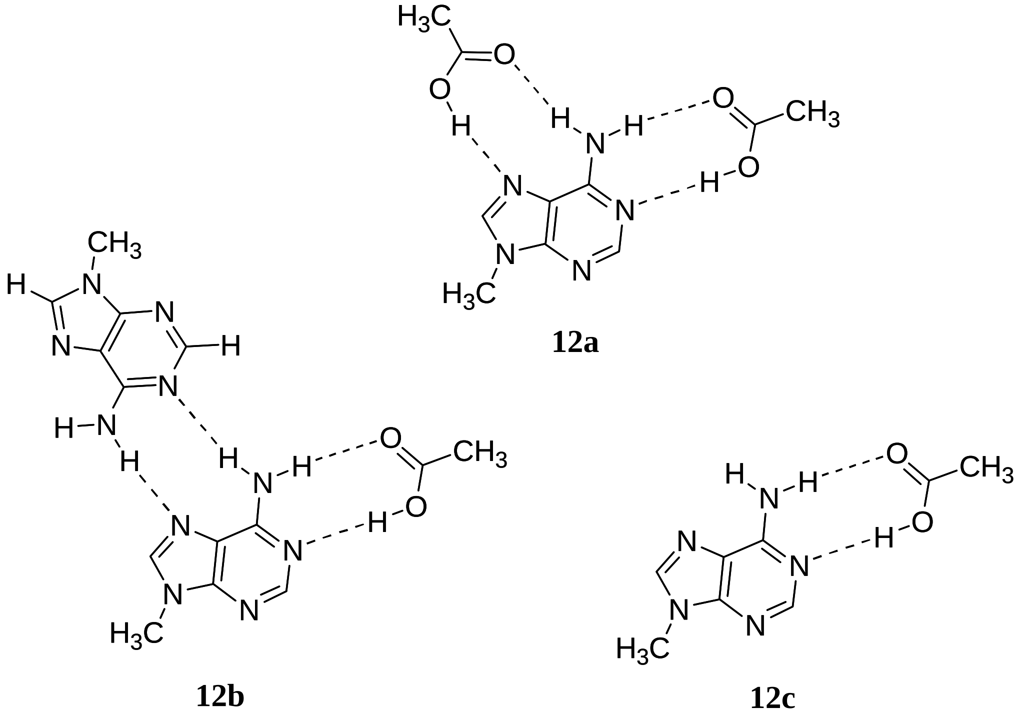


Figure 36 Computed complexes of 9-methyladenine and acetic acid.

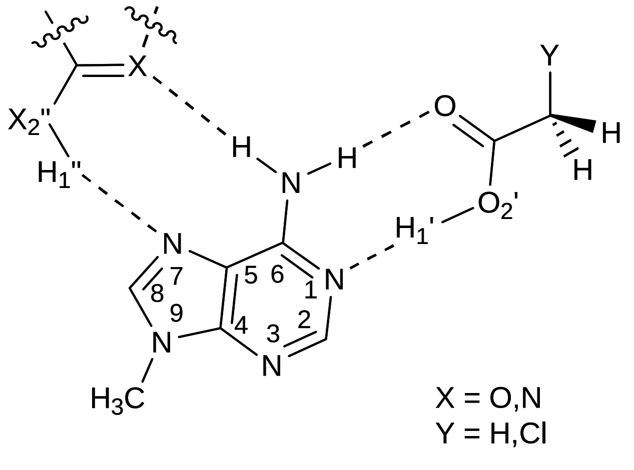


Figure 37 Atom numbering used for the adenine complexes

Comparing the A·HAc$_2$ (**12a**) and A$_2$·HAc complex (**12b**), the NMR experiment finds a more deshielded proton for the latter in the O-H---N1 hydrogen bridge at the Watson Crick site (see **Figure 37** for atom numbering). For **12a** the measured value of $\delta(H_{1'})$ is 17.1 ppm while 17.76 ppm is obtained for **12b**. Obviously, some cooperativity effects arise between Hoogsteen and Watson-Crick binding since the proton chemical shift constitutes a sensitive indicator for the relative hydrogen bond strength.[3,121,123,124,125,126,127] Anticipated cooperativity effects in these higher-order complexes[124,125] and their dependence on the type of ligand[124] are little understood, yet are not only important for the development of adenine receptor molecules that are often based on carboxylic acids[128] but also for a better understanding of interactions within nucleic acids and nucleic acid - protein complexes.[129]

The downfield shift observed in $\delta(H_{1'})$ when going from **12a** to **12b** may result from a displacement of the $H_{1'}$ proton towards the N1 acceptor atom pointing to a strengthening of the hydrogen bond. Such a strengthening of the hydrogen bond is often associated with an elongation of the covalent bond between the hydrogen and the donor. A corresponding example was recently reported by Kar and Scheiner[124] who investigated cooperative effects in chains consisting of up to five water molecules. For the gas phase their computations predict an elongation of the covalent OH bond by about 0.003 Å for the donor water molecule positioned at the end of the chain when going from the water dimer to the water trimer. This small change in the geometry was accompanied by a variation of 0.7 ppm in the chemical shift of the bridging hydrogen computed at the equilibrium geometry. Going from the water monomer to the dimer the variations were 0.006 Å and 2.8 ppm, respectively.

A similar finding was reported by Dingley et al.[125] who studied cooperative effects in T·A-T and C·G-C triplets and found a strong correlation between the ^{1}H chemical shift of the imino proton and the size of the two-bond scalar coupling across a hydrogen bond $^{2h}J_{NN}$. A similar but inverse correlation is found between the proton chemical shift and the (absolute) size of the covalent $^1J_{HN}$ scalar coupling. Based on DFT computations for a model system they concluded that most of the experimentally observed variations in the parameter are likely caused by changes in the donor acceptor distances. In this study and in the investigation of Kar and Scheiner[124] the computed NMR chemical shifts did not include effects resulting from vibrational motions.

The downfield shift could also result from vibrational effects. These effects have been shown to be important for the ^{1}H chemical shift[126,130,131,132] and were considered in extensive investigations on primary and secondary geometric H/D isotope effects on low barrier hydrogen bonds performed by Limbach and coworkers. In a series of papers[126] correlations between ^{1}H chemical shifts and hydrogen bond geometries were studied with the inclusion of vibrationally averaged bond distances allowing for the transfer of chemical shift data into hydrogen bond geometries. Employing empirical corrections for anharmonic zero-point vibrations, these correlations are able to describe isotope effects on hydrogen bond geometries from the weak to the strong hydrogen bond regime, taking into account single and double-well situations.[126] Isotope effects for intramolecular hydrogen bonds were also studied.[133] The influence of vibration anharmonicity was studies by J. Del Bene and coworkers.[134]

Important in this respect are also the extensive investigations of Steiner et al.[135] who used data of low-temperature neutron diffraction measurements to establish geometric hydrogen bond correlations.

The present work continues the investigations of such cooperative effects in complexes of substituted acetic acid and adenosine as model systems. In order to extend the series in a most appropriate way, new experimental data for additional compounds is briefly presented. Insights into the importance of various effects are obtained by high-level computations which include solvent effects and account for vibrational effects. Both are found to be extremely important. In addition to providing detailed insights into the contribution of various effects, computations are also used to extend the known series to yet experimentally inaccessible complexes. A critical evaluation of the theoretical approach is enabled by the new experimental data and the combination of theory and experiment provides for a more detailed and comprehensive understanding of the effects in multiple binding.

4.4.2 Experimental Findings

In order to investigate adenosine complex geometries and cooperativity effects on ligand binding, Weisz and coworkers studied the association of adenosine with chloroacetic acid through low temperature NMR experiments using a freonic solvent in order to reach the slow hydrogen bond exchange regime below 133 K.[1] ^{1}H NMR spectra acquired at 123 K for mixtures of chloroacetic acid (HAcCl) and 2',3',5'-tri-*O*-(*tert*-butyldimethylsilyl)-1-^{15}N-adenosine in various stoichiometric ratios are shown in **Figure 38**. With adenosine in excess, two doublets at 18.67 ppm and 17.64 ppm are observed at low field (**Figure 38** top). Employing additional ^{1}H{^{15}N}decoupling experiments, both of these OH protons are unambiguously identified as bound to the adenine Watson-Crick site through their scalar coupling of J_{NH} ~ 79 Hz and J_{NH} ~ 85 Hz with the labelled endocyclic N1 nitrogen. Such large couplings to the nitrogen acceptor atom demonstrate that these protons are mostly transferred to adenine N1 adopting ion pairing species. This is also confirmed by the increase in the J_{NH} scalar coupling for the more upfield shifted resonance and by the corresponding ^{15}N chemical shifts measured with a 2D ^1H-^{15}N heteronuclear multiple-quantum coherence (HMQC) experiment at 125 K. Thus, more upfield shifted Watson-Crick proton resonances scalar coupled to ^{15}N are correlated with more upfield shifted $^{15}N1$ signals as expected for a proton location closer to nitrogen.

---

[1] NMR experiments were performed on a Bruker AMX500 spectrometer. Temperatures were adjusted by a Eurotherm Variable Temperature Unit to an accuracy of ±1.0 °C. ^{1}H chemical shifts in a Freon mixture were referenced relative to $CHClF_2$ (δ_H = 7.13 ppm).

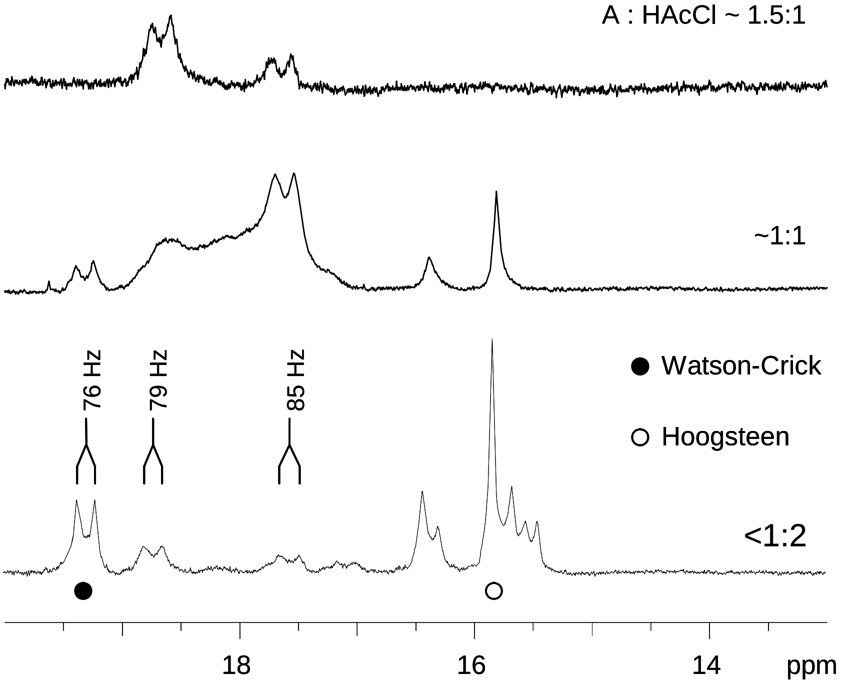
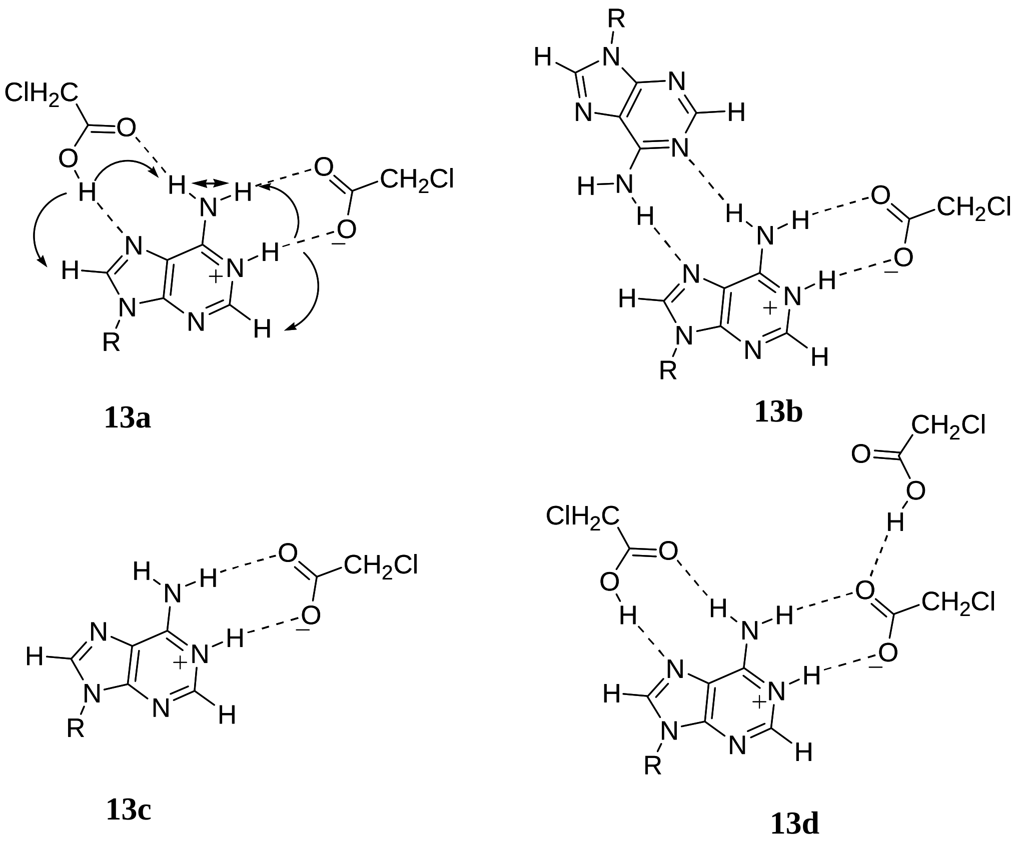


Figure 38 Carboxylic acid OH proton resonances for mixtures of 2’,3’,5’-tri-*O*-(*tert*-butyldimethylsilyl)-1-^{15}N-adenosine and chloroacetic acid with different molar ratios in Freon at 123 K. Solid and open circles denote the Watson-Crick and Hoogsteen bound proton in a 1:2 A:HacCl complex.

With all of the acid exclusively bound at the adenine Watson-Crick site, the two resonances at low concentrations of chloroacetic acid with stoichiometric ratios of A:HAcCl ~ 1.5:1 can be assigned to 2:1 or 1:1 complexes **13b** and **13c** (**Figure 39**). Adding more acid, the situation gets significantly more complex and an increasing number of signals appear, some of them being significantly exchange broadened at a 1:1 stoichiometric ratio (**Figure 38** centre). Obviously, with the acid in excess several different species coexist in solution making a detailed structural characterization of all the complexes a difficult task. However, protons hydrogen bonded to adenine N1 are easily recognized by their ^1H-^{15}N scalar coupling. Information on the more upfield shifted singlet signals comes from 2D NOE contacts observed at 123 K. As expected from their participation in a Watson-Crick hydrogen bond, the ^{15}N scalar coupled signals exhibit NOE crosspeaks to adenine H2. Correspondingly, NOE crosspeaks to adenine H8 identify protons bound to the Hoogsteen face of adenosine. Note, that in addition to the absence of a corresponding $^{2}J_{NH}$ scalar coupling, H8 protons of adenosine are easily distinguished from H2 protons by their NOE crosspeaks to sugar protons.[136] In addition, a third type of OH resonances at 15.45 and 15.55 ppm exhibiting no connectivities to any adenine base protons must be assigned to protons in a OHO hydrogen bond between two carboxylic acid molecules.

The two resonances at 19.30 ppm and 15.83 ppm having the same intensity on integration display NOE contacts to the same amino protons at 11.39 and 9.17 ppm and must be assigned to a 1:2 A-$(HAcCl)_2$ complex **13a** (**Figure 39**). As is apparent from **Figure 38**, they are increasingly formed with increasing acid concentration as expected. No attempt was made to assign additional, higher-order complexes that coexist at low temperatures. Interestingly, however, no chloroacetic acid dimers are formed even with an excess of acid as is evident from comparison with a chloroacetic acid solution at 123K (δ_{dimer} = 13.15 ppm). Rather, binding of a third acid molecule to form linear aggregates like **13d** (**Figure 39**) may account for the observed most upfield shifted OHO proton resonances at 15.45 and 15.55 ppm (**Figure 38** bottom).



13a **13b** **13c** **13d**

Figure 39 Geometries of complexes formed between adenosine and chloroacetic acid at low temperatures in a Freon solvent; arrows in complex **13a** indicate experimentally observed NOE contacts.

4.4.3 Computational Details

^{1}H NMR chemical shifts of inter- or intramolecular hydrogen bonds are considerably influenced by vibrational motions.[126,130,131,132] Additionally, temperature[137] and solvent effects[127] have to be taken into account. The latter may quite strongly affect the proton dynamics.[126,138] To account for the vibrational effects, a two-dimensional treatment including the high- and low-frequency stretching of the hydrogen bond would be desirable[131] but requires a two dimensional potential energy surface. Since we are primarily interested in the trends along a series of molecules, such a treatment would be overly expensive. Moreover, many studies indicate that an effective one-dimensional treatment covers the major effects.[126d,130,132,139] The present work therefore employs an effective one dimensional treatment to investigate its suitability in explaining the experimental data. To largely include the coupling between the high- and low-frequency stretching of the hydrogen bond, the hydrogen bond stretching potential was computed by relaxing the N-O distance at each point.

Within this treatment one moves along the minimum energy path of the two-dimensional surface[131] thus effectively accounting for the coupling between both motions.

The effective one-dimensional potential for the shuttling motion of $H_{1'}$ between $O_{2'}$ and N1 (for an atom numbering see **Figure 37**) was computed employing the B3LYP functional[36,35e] in combination with a TZVP basis set.[66] As described elsewhere the B3LYP functional in combination with a triple zeta split valence basis should provide very accurate results for geometries and energies of hydrogen bonded structures.[140] In these computations all internal coordinates except for the vibrational coordinate were optimized. All computations were performed with the TURBOMOLE program package.[65] Solvent effects are included by using the *Conductor-like Screening Model* (COSMO).[43,69] Because the dielectric constant of Freon solutions varies to some extent depending on temperature and composition, we choose ε=40 as a compromise.[141]

For the present problem both the O-H [$R(H_{1'}$-$O_{2'})$] and the N-H [$R(H_{1'}$-N1)] distance of the Watson-Crick hydrogen bond were considered as vibrational coordinates with calculated potentials exhibiting two minima under solvent conditions. In the following, $R_0(H_{1'}$-$O_{2'})$ and $R_0(H_{1'}$-N1) represent the equilibrium bond length of the proton covalently attached to oxygen (neutral structure) and to nitrogen (ion pairing structure), respectively. For $R(H_{1'}$-$O_{2'}) > R_0(H_{1'}$-$O_{2'})$, the $R(H_{1'}$-$O_{2'})$ coordinate can serve as the vibrational coordinate, *i.e.*, the effective one-dimensional potential can be computed for optimized geometries with fixed $R(H_{1'}$-$O_{2'})$ values. However, for $R(H_{1'}$-$O_{2'}) < R_0(H_{1'}$-$O_{2'})$ a steep repulsive potential and therefore an erroneous behaviour of the effective potential curve arises because the carboxylic acid molecule would rather dissociate from the adenine base (breaking the O–H···N hydrogen bond) than compressing the OH-bond. The dissociation energy amounts to about 30 kJmol^{-1} (2500 cm^{-1}). This branch of the potential can only be obtained if the NH distance is used as vibrational coordinate. Since a corresponding problem exists for small NH distances, both potentials have to be merged in order to obtain the full effective potential. In the present paper the potential with fixed N-H distances is used for the values of the vibrational coordinate larger than 1.25 Å. For those smaller than 1.25 Å, the O-H bond is employed. The correlation given in **Figure 40** allows for a better understanding of the relationship between the vibrational coordinate and the actual bond distances. Also, the distances for the computed minima are given in the respective tables.

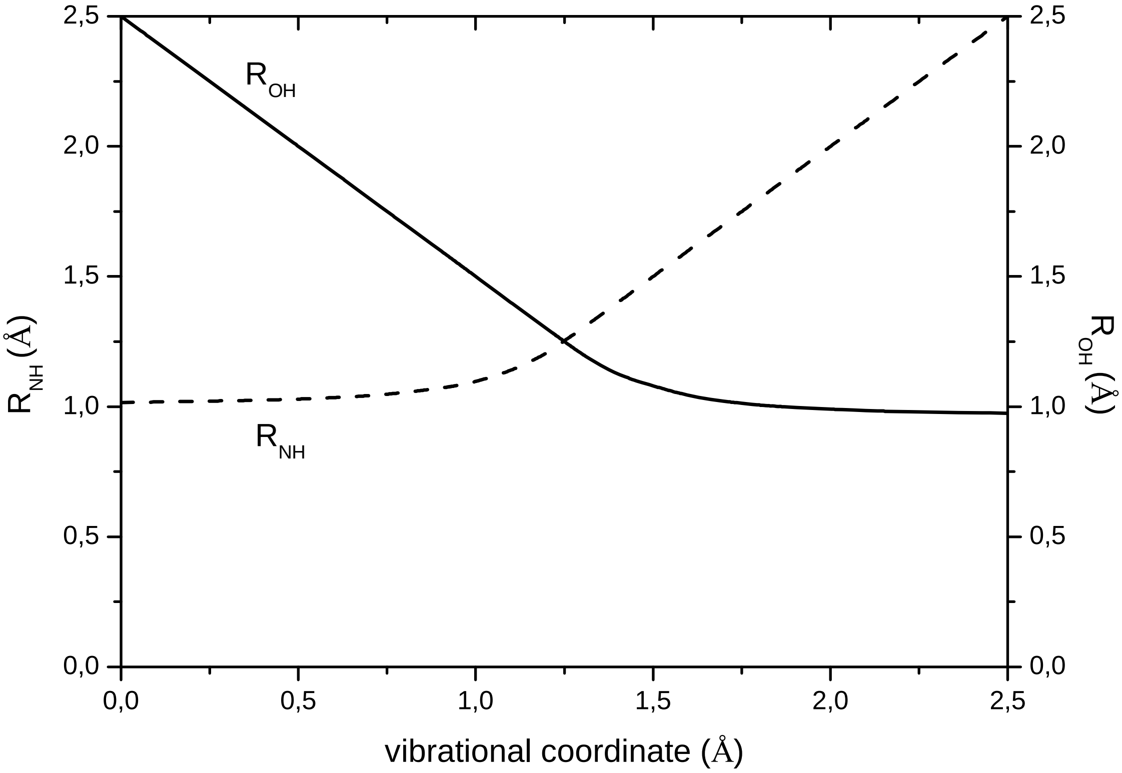


Figure 40 Correlation between the vibrational coordinate and the bond distances R_{OH} (solid line) and R_{NH} (broken line). All values are given in Ångstrøm.

NMR shielding tensors were calculated on a RHF/TZVP level using the GIAO approach for the gas phase. The relative ^{1}H chemical shifts are calculated with the ^{1}H shielding constant of TMS ($\sigma(H_{TMS})$ = 32.257). This level was used in various computations of relative ^{1}H shielding constants[127,142] which showed that the RHF/TZVP approach possesses an accuracy of about 0.2 ppm for relative ^{1}H chemical shifts in most cases. In a recent review[142b] it was stated that in contrast to the success of DFT in the calculation of molecular structures and energies DFT does not provide a systematic improvement over RHF in the calculation of magnetic shielding constants. Hence, we use DFT only to compute the potential energy surfaces.

In order to obtain the vibrationally averaged NMR shielding constants we used the approach developed by Peric et al.[143] For the computation of the reduced mass μ the whole motion of the supermolecule has to be considered. Between the minima [$R(H_{1'}\text{-}N1) > R_0(H_{1'}\text{-}N1)$ and $R(H_{1'}\text{-}O_{2'}) > R_0(H_{1'}\text{-}O_{2'})$], the reduced mass can be approximated by 1 a.u. because mainly the proton moves. As described above for $R(H_{1'}\text{-}O_{2'}) < R_0(H_{1'}\text{-}O_{2'})$, the whole carboxylic acid moves away leading to a flatter potential compared to the OH bond compression. Consequently, for this part of the potential curve a higher value for the reduced mass had to be taken. A similar problem arises for the part of the potential surface with $R(H_{1'}\text{-}N1) < R_0(H_{1'}\text{-}N1)$. Test computations employing different reduced masses show, however, that these effects are small. Therefore, a value of μ=1 a.u. was taken for the whole surface. Temperature effects were taken into account assuming a Boltzmann distribution with v = 0 and v = 1 vibrational levels. Test calculations show that the influence of the v = 2 level is negligible.

4.4.4 Calculation of averaged NMR – shifts

Results of gas phase computations on the Watson-Crick hydrogen bond geometry and proton chemical shift $\delta(H_{1'})$ of complexes **12a-c** (**Figure 36**) and **13a** (**Figure 39**) are summarized in **Table 22**. In all systems only the structure with the proton $H_{1'}$ attached to the carboxylate group (neutral structure) represents a minimum. This situation is in agreement with the experimental data for **12a-c** but in disagreement for **13a** with its experimentally observed ion pairing structure. The $\delta(H_{1'})$ values computed for the equilibrium structures of **12a-c** deviate by about 1.0 ppm from the measured data, *i.e.*, despite the agreement in the overall geometrical structure the deviation between experiment and theory for $\delta(H_{1'})$ is much larger than the expected error bars.[127,142] Nevertheless, while the computed absolute values for $\delta(H_{1'})$ disagree, the difference between **12a** and **b** is nicely reproduced on this level of theory.

Table 22 Summary of the gas phase computations. Structural parameters d are given in Å, NMR chemical shifts δ in ppm

System	[a]d(N··H··O)	vibrational coordinate	$\delta(H_{1'})$
12a	1.707/1.016	1.707	16.20
12b	1.691/1.020	1.691	16.76
12c	1.693/1.019	1.693	16.59
13a	1.668/1.023	1.668	17.41

[a] Left and right numbers refer to the proton-nitrogen and proton-oxygen distance, respectively.

Table 23 Summary of the computed energy differences ΔE (kJmol^{-1}), structural parameters d (Å) and NMR chemical shifts δ (ppm) of the minima for ε = 40. The respective potential curves are given in **Figure 41** and **Figure 42**.

System	$^{a}\Delta E$	$^{b}\Delta E_1^{\ddagger}$	^{c}d(N··H··O)	vibrational coordinate	^{d}d(N··H··O)	vibrational coordinate	$^{e}\delta(H_{1'})$	$^{f}\delta(H_{1'})$
12a	5.2	9.5	1.681/1.022	1.681	1.082/1.551	0.949	16.81	20.53
12b	2.8	8.2	1.686/1.022	1.686	1.077/1.572	0.928	16.95	20.08
12c	3.6	8.6	1.683/1.022	1.683	1.075/1.578	0.922	16.87	19.86
13a	7.4	1.9	1.603/1.041	1.603	1.066/1.611	0.889	18.75	19.04

[a]Energy difference between both minima. [b]Barrier height from the minimum in which the proton is attached to the carboxylate group. [c]Structural parameters of the minimum for which the proton is attached to the oxygen. Left and right numbers refer to the proton-nitrogen and proton-oxygen distance, respectively. [d]Structural parameters of the minimum for which the proton is attached to the nitrogen. Left and right numbers refer to the proton-nitrogen and proton-oxygen distance, respectively. [e]NMR chemical shift for the proton attached to oxygen (neutral structure). [f]NMR chemical shift for the proton attached to nitrogen (ion pairing structure).

If the influence of the Freon matrix is taken into account within the framework of the COSMO approach (ε = 40), ionic structures in which the proton $H_{1'}$ is bound to N1 rather than to $O_{2'}$ are sufficiently stabilized to become minima (**Table 23**). For the systems **12a-c** these minima are only 3-5 kJ/mol above the global minimum, which corresponds to the equilibrium structure found for the gas phase computations (neutral structure). The OH distances increase by only 0.002 – 0.006 Å in the Freon matrix with the $\delta(H_{1'})$ values computed for the global minima moving towards the measured data. For compound **12a** it only differs by 0.3 ppm from the experimental value which is within the range of the expected error bars. For **12b** the deviation still amounts to 0.8 ppm being larger than the expected inaccuracy. The computed $\delta(H_{1'})$ values for **12a** and **12b** differ by only 0.1 ppm, *i.e.*, at this level theory cannot explain the difference of nearly 0.7 ppm observed experimentally.

If the influence of the Freon matrix is taken into account, theory and experiment agree with respect to the hydrogen bond geometry of compound **13a** since the ion pairing structure becomes the global minimum lying more than 7 kJmol^{-1} below the minimum of a neutral structure. For this ion pairing species the computed $\delta(H_{1'})$ value is 19.0 ppm and thus only 0.3 ppm lower than the experimental value. This nice agreement supports the expectation that differences between the $\delta(H_{1'})$ values of systems **12a** and **12b** on the one hand and **13a** on the other hand result from a proton migration.

The computations also predict two potential minima for the Hoogsteen hydrogen bond to adenine N7 in a Freon matrix (not shown). However, with energy differences of 17 kJmol^{-1} (**12a**) and 13 kJmol^{-1} (**13a**), respectively, the influence of these minima are negligible.

For an accurate calculation of proton chemical shifts an average over the vibrational motions is often necessary.[130] This is also expected for the present systems since $\delta(H_{1'})$ considerably depends on the vibrational coordinate (**Figure 41**; for a correlation between the vibrational coordinate and the bond distances see **Figure 40**). This dependence is also indicated by the shape of the vibrational wave functions which are exemplarily shown for complexes **1c** and **13a** in **Figure 42**. A comparison of the various potentials is given in **Figure 43**. The vibrationally averaged $\delta(H_{1'})$ values are summarized in **Table 24**. A comparison with the experimentally accessible chemical shifts (**Table 25**) clearly shows that at this level of theory the values predicted for compounds **12a** and **13a** (T = 128 K) show an excellent agreement with the measured data. To some extent this excellent agreement is surely fortuitous since the absolute experimental values depend to some extent on the explicit Freon composition of the particular sample which is not accounted for in our theoretical approach. Note, however, that the relative shifts do not depend on the actual composition. For **12b** the deviation is about 0.3 ppm, *i.e.* all computed $\delta(H_{1'})$ values agree with their experimental counterparts within error bars. This indicates that this computational approach accounts for all important effects (see below).

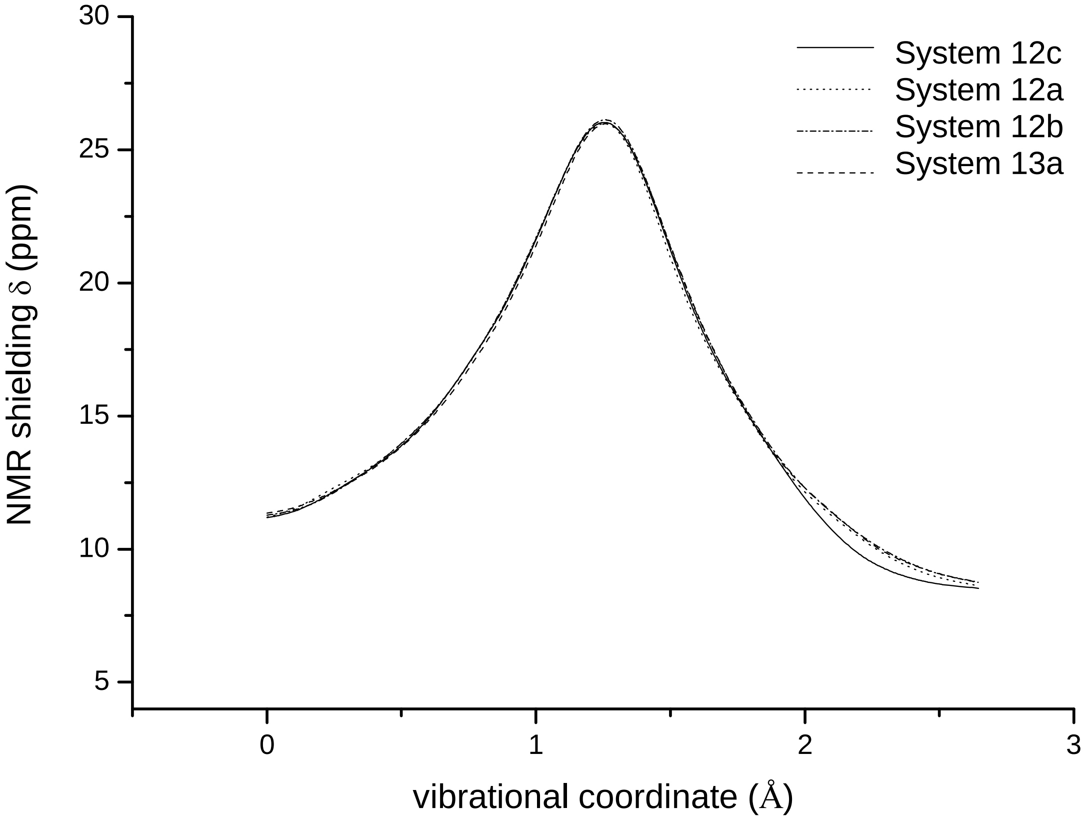


Figure 41 Comparison of the computed relative ^{1}H chemical shifts $\delta(H_{1'})$ along the potential curves for systems **12a-c** and **13a**. The computations were performed with the RHF/TZVP ansatz.

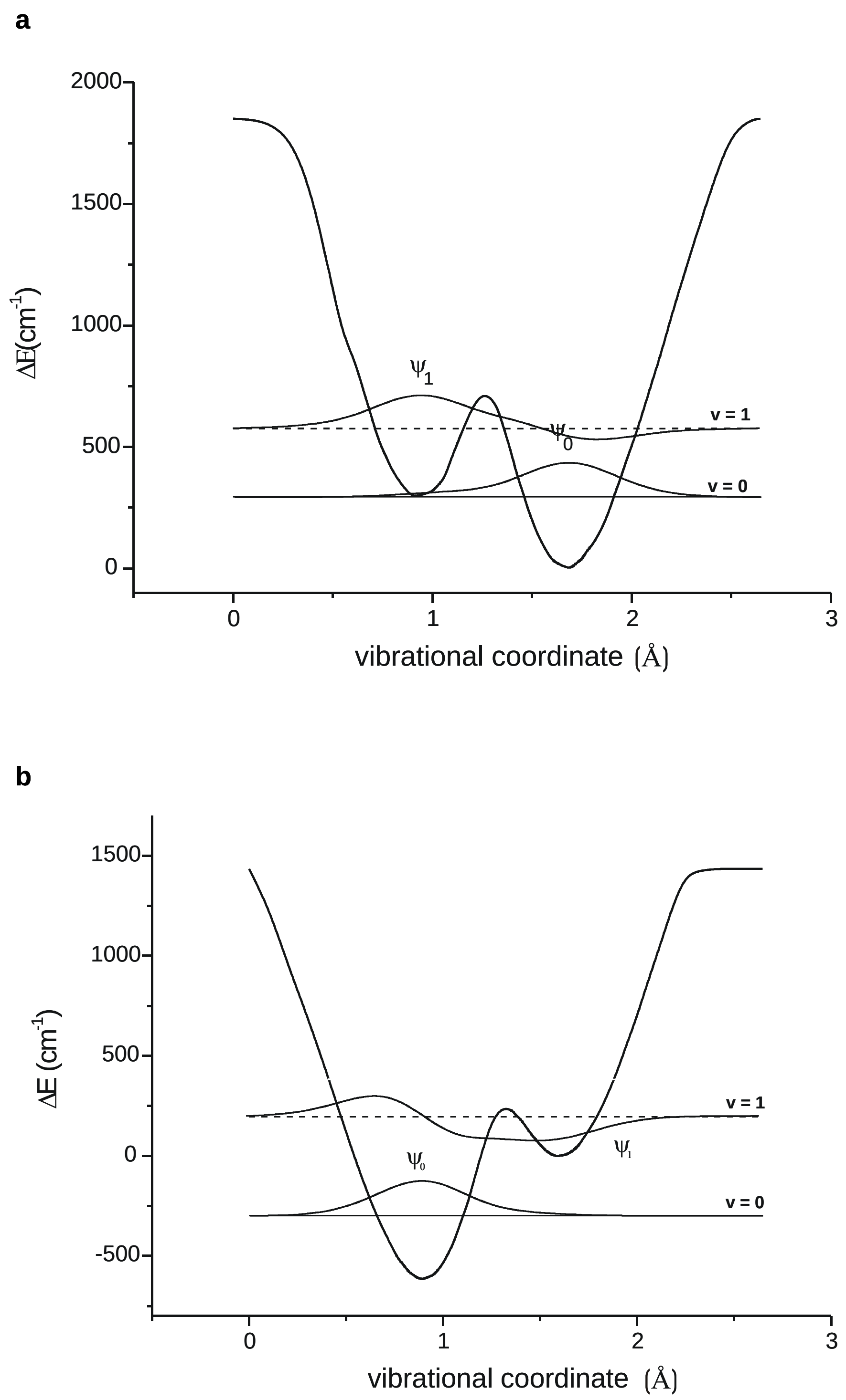


Figure 42 Computed potential curves, energy positions of the vibrational states and the corresponding wave functions for system **12c** (a) and **13a** (b). The computations were performed with the B3LYP/TZVP ansatz. The minimum on the right hand side corresponds to a proton close to the carboxyl oxygen.

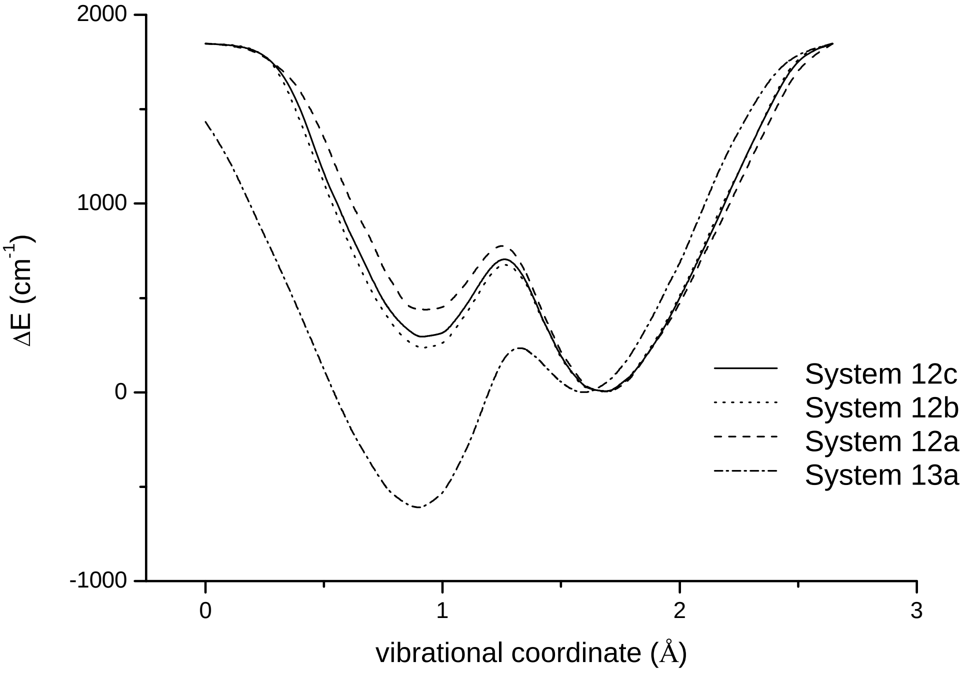


Figure 43 Comparison of the computed potential energy curves. All computations were performed with the B3LYP/TZVP ansatz.

Table 24 Summary of averaged NMR chemical shifts δ (ppm)

System	v = 0		v = 1			
	[a]energy	$\delta(H_{1'})$	[a]energy	$\delta(H_{1'})$	[b]Boltz.	$\delta(H_{1'})$128 K
12a	301	17.1	691	20.5	0.013	17.2
12b	295	17.9	532	20.1	0.07	18.1
12c	296	17.6	578	20.2	0.042	17.7
13a	316	19.3	809	20.1	0.004	19.3

[a]With respect to the global minimum in cm^{-1}. [b]Boltzmann distribution taking only v=0 and v=1 into account. Test calculations showed a negligible influence of the v = 2 levels.

Table 25 Experimental ^1H chemical shifts δ (ppm) and J_{NH} coupling constants (Hz) of $H_{1'}$ and $H_{1''}$ protons in Watson-Crick and Hoogsteen hydrogen bonds of adenosine-acid complexes

complex	pK_a	N1-H-O		N7-H-O	
	(25 °C)	$\delta(H_{1'})$	J_{NH}	$\delta(H_{1''})$	J_{NH}
12a[a]	4.75	17.10	0	15.11	0
12b[a]	4.75	17.76	0	--	--
13a	2.85	19.30	76	15.83	0

[a] Reference 122

4.4.5 Discussion

Low temperature NMR experiments on adenosine-chloroacetic acid mixtures in aprotic solvents indicate various coexisting higher-order complexes whose formation depends on the acid to adenine base molar ratio. In contrast, previous measurements on various adenine-acid mixtures under fast exchange conditions at ambient temperatures have been interpreted in terms of a simple 1:1 association model and higher-order complexes have mostly been excluded.[144,145,146] This might be attributed to the different enthalpic as well as entropic factors that will increasingly favour aggregation to higher-order aggregates at lower temperatures.

The nature of hydrogen bonds can be assessed by NMR spectral parameters like ^1H NMR chemical shifts or ^1H-^{15}N scalar couplings of the proton in the hydrogen bridge. Based on our computations (**Figure 41**) a weakening of the proton-donor bond accompanied by a strengthening of the proton-acceptor bond will give rise to a maximum in the proton chemical shift at about 26 ppm for a centralized proton (neglecting vibrational effects). If the proton is further transferred towards the acceptor atom, the proton chemical shift is predicted to decrease again. Such a proton shift is nicely seen for both Hoogsteen and Watson-Crick hydrogen bonds when going from the 1:2 to the 2:1 adenosine-acetic acid complex and finally to the 1:2 adenosine- chloroacetic acid complex, all of which are experimentally accessible. J_{NH} scalar couplings as measured in specifically 1-^{15}N labelled adenine nucleoside complexes unambiguously point to the formation of neutral complexes with acetic acid (**Table 25**).[6] In contrast, for chloroacetic acid the Watson-Crick bound proton is transferred to the nucleobase forming ion pairing species. Correspondingly, no J_{NH} scalar coupling is observed for Watson-

Crick adenosine - acetic acid complexes whereas a coupling of 76 Hz for the chloroacetic acid 1:2 complex, close to the value expected for a covalent N-H bond with bond order one, indicates an almost complete proton transfer to the N1 acceptor. Interestingly, excess chloroacetic acid does not form additional cyclic carboxylic acid dimers but seems to hydrogen-bond to the negatively charged carboxylic acid oxygen of the ion pairing associates in a linear fashion. This behaviour corresponds to the formation of charge relay chains previously observed between acetic acid (HAc) and pyridine at low temperatures.[126f,18]

Disregarding the N3 endocyclic nitrogen, the adenine base possesses two proton acceptor sites which can engage in a cyclic hydrogen bond with carboxylic acids, namely N1 and N7. In water the pK_a for N1-protonation of adenosine is 3.5 whereas N7 protonation does not take place above pH 1-2. Correspondingly, the hydrogen bond proton is more shifted toward the more basic N1 nitrogen in a Watson-Crick geometry when compared to the N7 in a Hoogsteen geometry. In no cases a resolved scalar coupling is found for N7 bound protons when using a 7-^{15}N labelled adenine nucleoside, *i.e.*, neutral complexes are always formed at the Hoogsteen face.

The differences found in the hydrogen bond geometries of **12a** and **13a** allow for a critical evaluation of the theoretical approach. Indeed, with respect to the geometries of the various complexes, experiment and theory only agree if the influence of the Freon matrix is taken into account. By including solvent effects, theory predicts a neutral structure for **12a** and **b** and a ion pairing structure for **13a** (see **Figure 43**). Also, in line with the experiment, theory always finds hydrogen atoms located closer to the oxygen donor within the hydrogen bonds at the Hoogsteen site. In addition to the hydrogen bond geometries, the computed proton chemical shifts also agree well with their experimental counterparts, however, vibrational effects have to be included (**Table 23** and **Table 24**).

The good agreement indicates that the employed approach seems to include all important effects. Consequently, theory can also provide an insight into the origin responsible for the trends. Clearly, the differences between **12a** and **b** on the one hand and **13a** on the other hand result from a gradual proton transfer induced by the stronger acidity of the chloroacetic acid. For the differences between **12a** and **12b**, however, structural effects seem to be less important. The variations in the equilibrium bond distances when going from **12a** to **12b** ($\Delta R_{O\text{-}H}$ < 0.001 Å; $\Delta R_{N\text{---}H} \approx$ 0.005 Å) are small and the chemical shifts computed for the equilibrium structures (16.81 *vs.* 16.95 ppm) theory strongly underestimates the experimentally determined difference in $\delta(H_{1'})$ (17.1 *vs.* 17.76 ppm)

Table 26 Summary of the various contributions to the averaged NMR chemical shifts δ(ppm). All computations were performed with ε = 40. For more information see text.

Potential curve from system	NMR values from system	$\delta(H_{1'})$ v = 0
12a	**12a**	17.12
[a]**12a**	**12a**	16.92
12a	**12b**	17.37
12b	**12b**	17.90
12b	**12a**	17.64
[a]**12b**	**12b**	17.46

[a]The average was obtained using a potential curve in which only the minimum with the proton at the carboxylate group was taken into account.

An analysis of the various effects is given in **Table 26**. By combining the potential curve of system **12a** with the NMR shielding curve of system **12b**, a vibrationally averaged $\delta(H_{1'})$ value of 17.37 ppm is obtained. It deviates by 0.25 ppm from the “pure” $\delta(H_{1'})$ value of **12a**, *i.e.*, the slight difference in the NMR shielding curve (**Figure 41**) accounts for one third of the overall effect. The strong influence of the potential curve can be seen from the calculation which combines the potential curve of **12b** with the NMR shielding curve of **12a** increasing the value of $\delta(H_{1'})$ to 17.64 ppm. This is about 2/3 of the overall effect for T = 0.0. The influence of the double-well nature of the potentials of **12a** and **12b** can be seen from the second and the last row of **Table 26**. If the second minima are omitted, averaged $\delta(H_{1'})$ values of 16.92 ppm (**12a**) and 17.46 ppm (**12b**) are computed. They deviate by 0.2 ppm (**12a**) and 0.4 ppm (**12b**) from the values obtained with the double-well potentials. Assuming a Boltzmann distribution for T = 128 K the value of $\delta(H_{1'})$ increases by about 0.1 ppm for **12a** and 0.2 ppm for **12b** (**Table 24**).

With vibrationally averaged bond distances $\langle R(N{\cdots}H)\rangle_{v=0}$ the trends can be explained by changes in the averaged geometries. For **12a** and **12b** the values amount to 1.6855 Å and 1.6514 Å, respectively, in line with an effective proton displacement towards the nitrogen acceptor. If the $\delta(H_{1'})$ values for these distances are taken from **Figure 41**, one obtains $\delta(H_{1'})$ = 16.71 ppm for **12a** and $\delta(H_{1'})$ = 17.56 ppm for **12b**. Using this picture our model is in line with the findings of Limbach and coworkers.[126f,147] Due to the importance of solvent effects for the present problem, the question arises whether the used theoretical approach correctly mirrors the physical situation.[148] The adaptation of the environment to changes in the electronic structure of the solute can be divided into polarization effects of the electron shells of the solvent molecules and effects resulting from molecular reorientation (rotation, translation, etc.). While the former will be much faster than the proton transfer process described by the computed double-minimum potentials, the latter (and probably more important ones) will be slower. Thus, the solvent will mainly be adapted to the situation at the global minimum. Going from the global to the local minimum, an only partial adaptation will result in an imperfect orientation of the environment. In contrast, continuum approaches assume a complete reorientation of the surroundings. As a consequence, the computed energy differences between global and local minima are smaller than in the real situation of a shuttling proton. Correcting for this shortcoming, the importance of the second minima diminishes in **12a** and **12b** and their influence may even vanish completely (this is even more true for **13a**). This situation is shown by the second and the last row of **Table 26** where the second minima have not been taken into account. The computed values for $\delta(H_{1'})$ are 16.92 ppm (**12a**) and 17.46 ppm (**12b**), *i.e.*, even in this case the trend is nicely reproduced and vibrational effects are still the major contributors. Although absolute values deviate more from their experimental counterparts, the agreement is still good considering the uncertainties in experiment and theory.

Clearly, a 1:1 adenosine-acetic acid complex **12c** serves as an important reference for the evaluation of cooperativity effects in binding a second ligand. Unfortunately, such a 1:1 complex was not found experimentally at low temperatures, however, our computations also allow a structural characterization for this bimolecular complex. As shown in Figure 9, binding a second ligand at the adenine Hoogsteen site primarily affects energy differences between minima for the potential curve of the Watson Crick OH····N hydrogen bond. These effects manifest themselves in a weakening of the H-bond and in a shielding of the $H_{1'}$ proton upon binding a second acetic acid at the Hoogsteen site in **12a**. In contrast, a strengthening of the H-bond and a deshielding of the $H_{1'}$ proton is calculated upon binding a second adenine base in **12b** (see **Table 24**). Likewise, an averaged proton position closer to N1 is suggested by a more upfield chemical shift measured in the ion pairing complex **13c** compared to **13a** with a second Hoogsteen bound acid.

These (anti)cooperativity effects can be rationalized in terms of electron-withdrawing and electron-releasing effects exerted by bound ligands. Electron density is expected to gradually decrease in the whole purine ring system upon binding ligands with increasing electron withdrawing properties. This in turn should reduce the acceptor capability of the endocyclic nitrogen with a concomitant shift of the averaged proton location closer to the donor atom. However, when substituting the electron-releasing adenine nucleobase for a Hoogsteen bound second carboxylic acid, the proton is further shifted towards the more potent N1 acceptor resulting in a downfield or upfield shift in a neutral or ion pairing complex, respectively. Also, binding of a second adenine being essentially basic may even benefit from protonation at N1 and from the concomitant increase in acidity of the exocyclic amino donor.

In the present study the cooperative effects between N···H-O hydrogen bonds are mediated by the purine ring system. This can be compared to the results of Kar and Scheiner[124] who investigated the cooperativity in OH···O hydrogen bonds employing chains of water molecules $(H_2O)_n$ up to n=5. Going from n=2 (no additional second ligand) to n=3 (one additional second ligand) correspond to the changes found between **12c** and **12a**. In their model system the overall effect is smaller due to the lower acidity of the second ligand. Nevertheless, having a smaller mediating system the overall effects could be similar. Indeed, based on gas phase results this seems to be the case. For the water chain the covalent O-H bond elongates by about 0.003 Å, *i.e.*, the effect is similar to our system with a change in the equilibrium geometry of also about 0.003 Å (**Table 22**). For the water chain the computed chemical shift of the bridging hydrogen changes by about 0.7 ppm and is thus comparable to the change of about 0.4 ppm in our system. A comparison of the effects in a solvent is not possible since Kar and Scheiner only discussed the influence of a solvent on binding energies but not on geometries and chemical shifts. Nevertheless, they concluded that the cooperative effects will diminish since the computed binding energies decrease dramatically. As for variations in equilibrium distances this is indeed found in this study, however, due to vibrational effects variations in chemical shifts could still be more pronounced.

4.4.6 Conclusion

Chemical shift calculations applying vibrational averaging faithfully reproduce experimental proton chemical shifts in adenosine-carboxylic acid complexes obtained at very low temperatures in the slow hydrogen bond exchange regime. Therefore, geometries on complexes not accessible under the solution conditions employed can be calculated complementing the available experimental data. On the basis of this combination of experimental and theoretical data, reliable information can be obtained on the strength of individual hydrogen bonds and on the cooperativity of multiple binding to the adenine nucleobase. Clearly, such knowledge is not only important for the design of artificial adenine receptors but also for a better understanding of the many biological interactions involving adenine, *e.g.*, as a component in polymeric nucleic acids or in adenine cofactors.

Chapter 5 Summary

The effective binding of anions like carboxylates and phosphates in aqueous solutions is of particular interest for various reasons. On the one hand anions play an important role in many biochemical processes as enzyme substrates or co-factors making a profound understanding of the binding modes under physiological conditions essential. On the other hand there is a vast number of possible uses for artificial anion receptor systems. It ranges from rational drug design of active agents against *methicillin-resistant staphylococcus aureus* (MRSA) germs or other infectious diseases threatening immuno-deficient patients to the development of self-assembling functional materials consisting of zwitterionic building blocks. However, it is challenging to develop such artificial structure motifs that are able to complex anions efficiently, as a highly competitive media like an aqueous surrounding strongly weakens non-covalently formed complexes.

The natural archetypes of effective anion receptors are enzymes that contain often arginine as relevant amino acid in the binding pocket. For this reason, one class of artificial anion receptors that emerged more than two decades ago mimics the anion binding with the guanidinium group present in the amino acid side chain. In 1999, Schmuck and coworkers developed a new class of guanidinium-based oxo anion receptor that binds carboxylates even in aqueous media. The binding modes of the 2-(guanidiniocarbonyl)-1*H*-pyrroles are based on individually weak non-covalent interaction between artificial host and substrate like ion pairing and multiple hydrogen bonds. The combination of these effects leads to unprecedented high association constants (“Gulliver effect”). The zwitterionic derivative with substitution of a carboxylate group in position 5 of the pyrrole ring system (structure **1, Figure 44**) shows a strong self-assembly to discrete dimers with an estimated association constant of 170 M^{-1} even in water. In order to further improve the structure motif for an effective oxo anion binding it is therefore of great interest to quantify the different intermolecular interactions between two monomeric units of **1**.

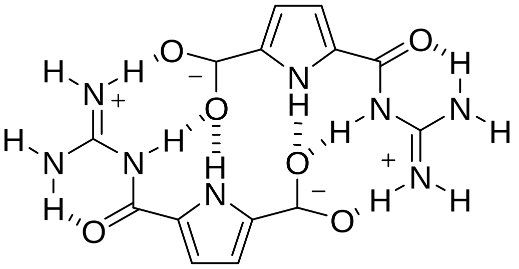


Figure 44 2-(Guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate dimer (**1**) as prototype for an effective guanidinium-based carboxylate receptor.

Against this background several theoretical *ab initio* studies were conducted in order to elucidate the influences of intrinsic properties as well as solvent effects on the stability of self-assembled dimers. In chapter 4.1 the molecular interactions in the 2-(guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate dimer (**1**) were investigated by comparison to various “knock-out” analogues. In these analogues single hydrogen bonds were switched off by substitution of hydrogen donor atoms with either methylene groups or ether bridges. The calculations were done for vacuum and solvation, as represented by a conductor-like polarizable continuum. It could be shown that the application of a simple continuum solvent model fails to predict the absolute energies of the knock-out analogues in strongly polar solvents. However, the calculated trends can explain the relative stabilities. The analysis of the stabilities of the various “knock-out” analogues thus revealed four issues that seem to be important for an efficient binding:

1. The salt-bridged hydrogen bonds are essential for the overall stability and are much more effective than simple point charge interactions without hydrogen bonds.
2. Additional neutral hydrogen bonds from the pyrrole ring further stabilize the dimer system but their influence is much smaller than that of ionic hydrogen bonds.
3. Solvation effects are strongly dependent on the accessibility, *i.e.* solvent-exposed hydrogen bonds are more weakened than inner hydrogen bonds.
4. A multiple hydrogen bonding pattern comprises secondary interactions that contribute significantly to the dimer stability.

Finally, the comparison of the 2-(guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate dimer (**1**) with the arginine dimer showed that the extraordinary stabilization upon dimerization in gas phase is ascribed to the rigidity of the zwitterionic monomers which have an considerably higher energy content than arginine. The intramolecular interactions of the charged moieties which are due to the flexible backbone are responsible for the strong stabilization of monomeric arginine. As this *intra*molecular stabilization is prevented in structure **1** the dimers are stronger stabilized by the *inter*molecular complexation.

In chapter 4.2 the structural similarity of arginine with 2-(guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate (**1**) was used in order to examine the dependence of self-assembly from the flexibility of the molecular structure. Arginine as model system has several advantages when studying intermolecular guanidinium-carboxylate interactions, namely (1.) it is a natural occurring amino acid of biological relevance showing multiple hydrogen bonded guanidinium-carboxylate interactions, (2.) it possesses a flexible backbone which allows strong conformational changes in order to form best possible hydrogen bonds and (3.) many experimental and theoretical studies are already at hand. The calculation of dimerization energies requires the determination of the global (and sometimes also local) minimum structures which poses a non-trivial problem even for medium-sized molecules with a flexible scaffold like arginine. It is therefore essential to develop a strategy to identify all relevant minimum energy conformers with reliable relative energy ranking.

In chapter 4.2.1 new global minimum structures of the canonical and zwitterionic arginine in gas phase were found by means of exhaustive force field based conformational searches in conjunction with *ab initio* structure optimizations of the lowest energy conformers on a MP2 level of theory with a large basis set comprising additional diffuse functions. Moreover, most of the newly identified minimum conformers of both the zwitterionic and canonical tautomer revealed geometrical arrangements with hitherto unreported stacked orientations of the terminal groups. These unusual geometrical arrangements were overlooked so far by other theoretical studies due to the deficiencies of density functional approaches. Therefore, it could be shown that for non-covalently bound systems possessing π-π like stacking interactions only *post*-HF electron correlation methods are able to give the appropriate relative energies of all possible conformers. All trends found on a MP2 level of theory could be confirmed by CCSD(T) calculations on the MP2 optimized geometries and finally a novel global minimum structure (see **Figure 45**) was detected that is more than 8 kJ mol^{-1} lower in energy than the previously published conformers.

There was a vivid discussion in the literature whether arginine occurs in gas phase as zwitterion or in the canonical state.[64,77c-d,79,94] The strong proton affinity of the guanidine group could outweigh the energy required for charge separation and it is well known that already one cation strongly stabilizes zwitterionic species. We could show that the lowest zwitterionic conformer ZW1 (see **Figure 45**) is only 7 kJ mol^{-1} (MP2/TZVPP+) higher in energy than the canonical global minimum N1. Therefore, the rigorous exclusion of the zwitterionic state in the gas phase that was proposed in several theoretical and experimental studies is no longer valid and an unambiguous proof could only be given by experiments. One possibility to distinguish between both tautomers would be the comparison of the hydrogen stretch vibrations between 2600 and 4000 cm^{-1} as we could demonstrate by computed vibrational spectra of the lowest monomer structures.

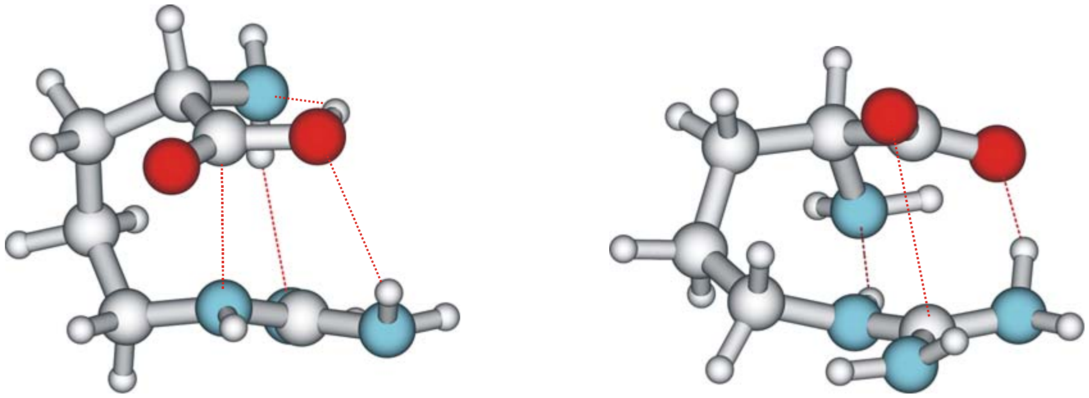

N1 ZW1

Figure 45 Minimum structures of arginine: global minimum (N1) and the lowest zwitterionic minimum (ZW1) optimized on a RI-MP2/TZVPP+ level of theory

The same strategy for finding minimum energy conformers of the arginine monomer has also been employed for the arginine dimer structures. While previous theoretical studies favoured directed hydrogen bonds (see dimer structure DZ1 in **Figure 46**) the new global minimum structure MMFF1 is about 60 kJ mol^{-1} (RI-MP2/TZVPP+//RI-MP2/TZVP) more stable and exhibits a stacked orientation of the guanidinium and carboxylate groups.

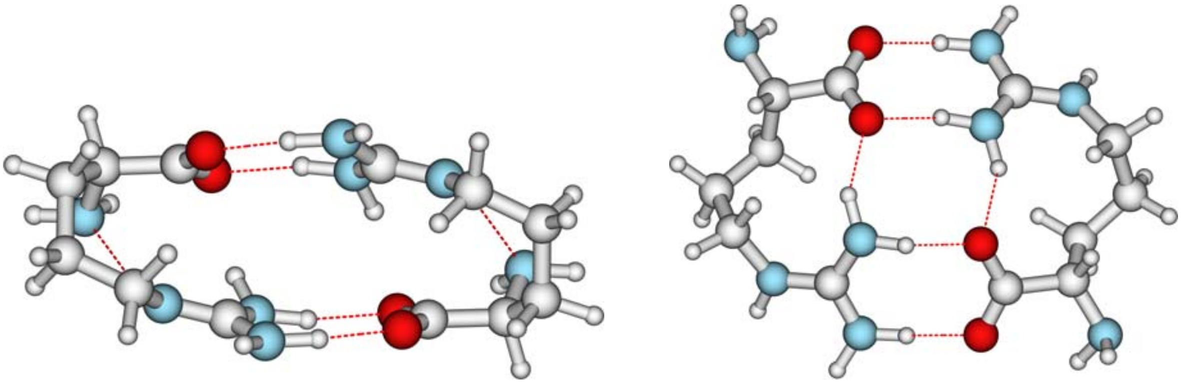

MMFF1 DZ1

Figure 46 Minimum structures of arginine dimer: new global minimum structures MMFF1 (left) and the local minimum structure DZ1 published by Goddard III et al.[77a] (right) both optimized on a RI-MP2/TZVP level of theory.

The importance of rigidity on the dimer stability was proven by calculations of an artificially stiffened arginine dimer system. Although the hydrogen bonding pattern of the artificial arginine is less distinct, the comparison with the 2-(guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate dimer (see **Figure 47**) revealed that it shows a dimerization energy that is nearly of the same magnitude as for the pyrrole derivative and twice as large as for the global minimum structure MMFF1. Thus the high binding affinity of the 2-(guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate dimer (**1**) results by about 50% from the rigidity of the monomers which prevents any intramolecular stabilization.

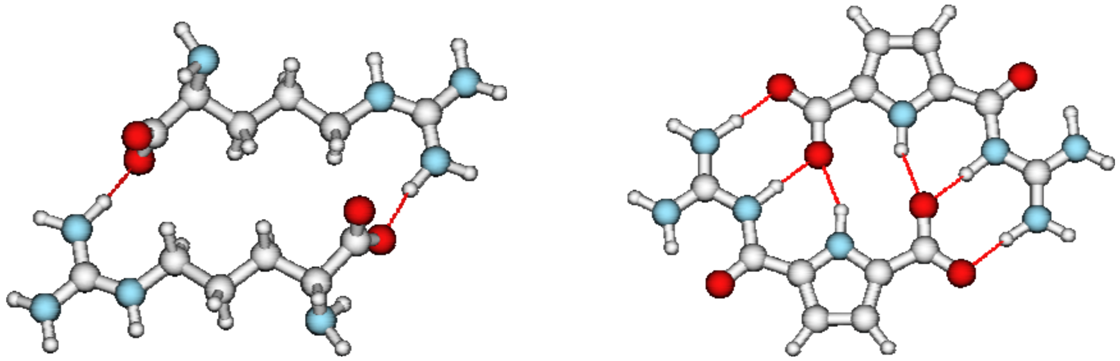

410 kJ mol^{-1} **1:** 581 kJ mol^{-1}

Figure 47 MP2/TZVPP+ calculated dimers of an artificially linear arginine (left) and 2-(guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate (right).

In chapter 4.3 novel structure motifs with varying ring systems have been examined on a DFT level of theory in order to make proposals for an improved carboxylate binding motif. The study of benzene and pyridine derivatives showed the importance of complementarity of the monomers which minimizes possible geometrical strains in the dimers. Again many structures clearly indicated that any intramolecular stabilization within the monomers strongly reduces the dimerization energy. The direct dependency of the dimerization energy on an increasing dipole moment was demonstrated by various anellated ring structures. The influence of the delocalization in the monomer on the dimerization energy was examined by variation of the electronic structure of electronically decoupled biphenylenes with electron pushing or withdrawing substituents. However, this approach did not allow to improve the dimerization affinity. A significant increase in the dimerization energy of about 90 kJ mol^{-1} with respect to the 2-(guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate dimer was observed for the indole derivative **10a** which benefits from a larger dipole moment in conjunction with the good binding behaviour of the guanidiniocarbonylpyrrole moiety (see **Figure 48**). In both the pyrrole (**1**) and the indole derivative (**10a**) the additional pyrrole hydrogen bond stabilizes the dimer by a stronger fixation of the carboxyl group in the binding pocket ("Gulliver effect").

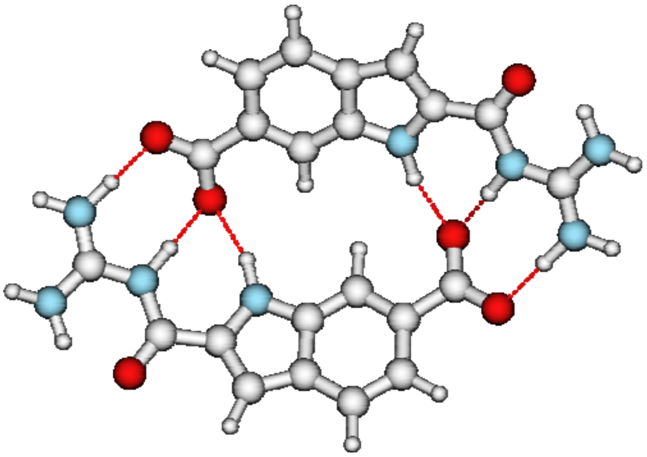


Figure 48 Optimized structure of 2-guanidinio-carbonylindole-6-carboxylate dimer (**10a**) in gas phase (B-LYP/TZVPP+).

Schmuck et al. proposed that the carbonyl function in vicinity of the guanidinium group implies an increase of the acidity of the hydrogen donor site which is essential for the strong binding affinity. However, with the aid of various substituted 7-guanidinioindole-2-carboxylate derivatives we could show that the carbonyl function is mainly responsible for the advantageous preorganisation, whereas the effect on the acidity seems to be only of minor importance.

The influence of molecular solvent effects on potential energies was investigated by calculated dimerization energies in continuum solvation and experimental dimerization constants of *m*-guandiniocarbonylbenzoate **7a** obtained in varying DMSO/methanol solvent compositions. We could show that only by inclusion of explicit solvent molecules it is possible to explain the experimentally observed trends. Furthermore, the calculation of Gibbs free enthalpies for association reactions is still problematic since thermodynamic contributions for the condensed phase were calculated by means of analytical formulas of an ideal gas. However, ITC measured dimerization constants indicate that the computed relative energy differences are already in good accordance with experimental findings. As a consequence, the calculated Gibbs free enthalpies can then be used to estimate the dimer stability in a given solvent if the experimental association constant of a reference system that is structurally related is available. Based on these results the 2-guanidiniocarbonylindole-6-carboxylate (**10a**) seems to be a very promising candidate that should form discrete dimers even in highly polar solvents.

In the last chapter cooperativity effects in supramolecular assemblies have been investigated. This was achieved by NMR shift calculations of adenosine-carboxylic acid complexes as model systems and comparison to experimental low-temperature NMR studies. We could demonstrate that only by applying vibrational averaged NMR shifts the experimental proton shifts obtained at very low temperatures in the hydrogen bond exchange regime could be reproduced. As a result, by combining experimental and theoretical data one can obtain reliable information on the strength of individual hydrogen bonds and of the cooperativity within multiple hydrogen binding patterns of self-assembling systems.

Chapter 6 Zusammenfassung

Eine effektive Bindung von Anionen wie Carboxylaten und Phosphaten in wässrigen Lösungen ist aus verschiedenen Gründen von großem Interesse. Auf der einen Seite spielen Anionen in vielen biologischen Prozessen eine wichtige Rolle als Enzymsubstrate oder Co-Faktoren. Deshalb ist ein tiefgreifendes Verständnis der Bindungsmodi unter physiologischen Bedingungen besonders wichtig. Auf der anderen Seite können künstliche Anionenrezeptorsysteme vielseitig eingesetzt werden. Dies reicht vom rationalen Drug Design von Wirkstoffen gegen *Methicillin-resistente Staphylococcus Aureus* (MRSA) Keime und andere Infektionskrankheiten, die besonders für immunschwache Patienten eine Gefahr darstellen, bis hin zur Entwicklung von Funktionswerkstoffen, welche sich aus zwitterionischen Bausteinen selbst-organisierend zusammensetzen. Die Entwicklung von solchen künstlichen Strukturmotiven, welche Anionen effizient komplexieren können, ist jedoch sehr anspruchsvoll, da die nicht-kovalent gebundenen Komplexe durch hochkompetitive Medien wie wässrige Lösungen stark geschwächt werden.

Die natürlichen Vorbilder effektiver Anionenrezeptoren sind Enzyme, welche oftmals Arginin als entscheidende Aminosäure in der Bindungstasche tragen. Aus diesem Grund wurde vor mehr als zwei Jahrzehnten eine Klasse von künstlichen Anionenrezeptoren entwickelt, welche versucht die natürliche Bindung von Anionen in Arginin zu imitieren. Die positiv geladenene Guanidiniumgruppe, wie sie in der Seitenkette von Arginin vorkommt, ist daher das zentrale Strukturmerkmal für viele Anionenrezeptoren. Im Jahre 1999 gelang es Schmuck und Mitarbeitern eine neue Klasse von Guanidinium-basierten Oxoanionenrezeptoren zu entwickeln, die Carboxylate sogar in wässrigen Medien binden können. Die Bindungsmodi der 2-(Guanidiniocarbonyl)-1*H*-pyrrole basieren auf einer Kombination von einzeln betrachtet schwachen nicht-kovalenten Wechselwirkungen wie Ionenpaarbildung und multiplen Wasserstoffbrückenbindungen zwischen künstlichem Rezeptor und Substrat. Jedoch erst die Kombination dieser Effekte führt zu den hohen Assoziationskonstanten („Gulliver-Effekt“). Durch Substitution einer Carboxylatgruppe in Position 5 des Pyrrolringes erhält man ein zwitterionisches Derivat (Struktur **1**, siehe **Figure 49**), welches sich in Wasser mit einer

Assoziationskonstante von schätzungsweise 170 M^{-1} zu einzelnen Dimeren zusammenlagert. Um das Strukturmotiv hinsichtlich einer noch effektiveren Anionenbindung weiter verbessern zu können, ist es daher von großem Interesse, die verschiedenartigen intermolekularen Wechselwirkungen zwischen den beiden monomeren Einheiten von Dimer **1** zu quantifizieren.

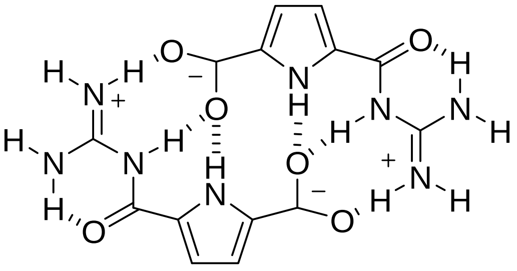


Figure 49 2-(Guanidiniocarbonyl)-1*H*-pyrrol-5-carboxylat Dimer (**1**) als Prototyp für einen effektiven Guanidinium-basierten Carboxylat Rezeptor.

Vor diesem Hintergrund wurden verschiedene theoretische *ab initio* Studien durchgeführt, um die Einflüsse von intrinsischen Eigenschaften sowie von Solvenseffekten auf die Stabilität sich selbst zusammenlagernden Dimeren aufzuklären. In Kapitel 4.1 wurden die molekularen Wechselwirkungen im 2-(Guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylat-Dimer (**1**) durch Vergleich mit verschiedenen „Knock-out“ Analoga untersucht. In diesen Analoga wurden einzelne Wasserstoffbrückenbindungen durch Substitution von Wasserstoffdonoren mit Methylengruppen oder Etherbrücken ausgeschaltet. Die Berechnungen wurden für Vakuum oder Solvensumgebung durchgeführt, wobei letztere durch ein polarisierbares, so genanntes „conductor-like“ Kontinuummodell dargestellt wurde. Es konnte gezeigt werden, dass die Anwendung eines vereinfachten Kontinuum-Solvensmodells nicht ausreicht, die absoluten Energien der „Knock-out“ Analoga in stark polaren Lösungsmitteln vorherzusagen, jedoch können die berechneten Trends Auskunft über die relativen Stabilitäten geben. Die Analyse der Stabilitäten der verschiedenen „Knock-out“ Analoga ergab, dass folgende vier Punkte für eine effiziente Bindung wichtig zu sein scheinen.

1. Die durch die Salzbrücke verstärkten Wasserstoffbrückenbindungen sind essentiell für die Gesamtstabilität und zudem deutlich effektiver als einfache Punkladungswechselwirkungen ohne Wasserstoffbrückenbindungen.

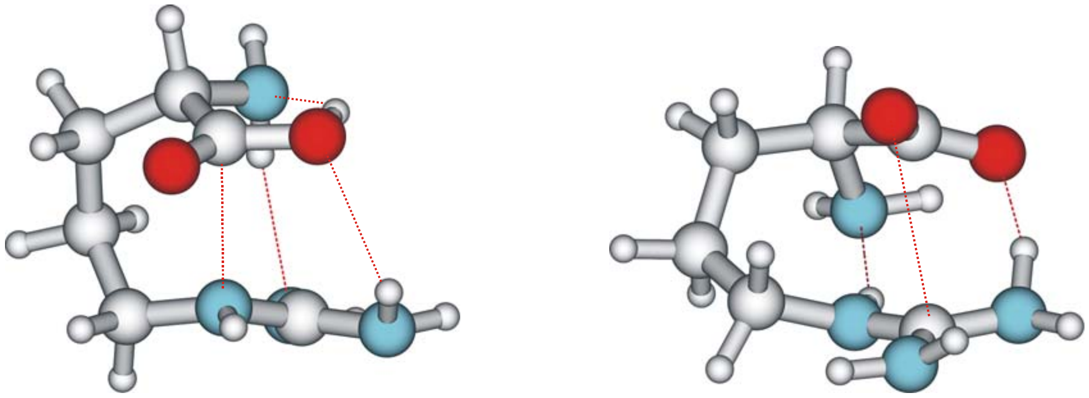
2. Weitere neutrale Wasserstoffbrückenbindungen ausgehend vom Pyrrolring stabilisieren zusätzlich das Dimersystem, jedoch ist ihr Einfluss deutlich geringer als der von ionischen Wasserstoffbrückenbindungen.
3. Lösungsmitteleffekte sind stark abhängig davon, wie leicht eine Bindungsstelle für Solvensmoleküle zugänglich ist. Lösungsmittelexponierte Wasserstoffbrückenbindungen werden stärker geschwächt als innere Wasserstoffbrückenbindungen.
4. Ein multiples Wasserstoffbrückenmuster weist sekundäre Wechselwirkungen auf, welche entscheidend zur Dimerstabilität beitragen.

Letztendlich zeigte der direkte Vergleich des 2-(Guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylat-Dimers (**1**) mit dem Arginin Dimer in Gasphase, dass die außergewöhnliche Stabilisierung bei der Dimerisierung auch der Rigidität der zwitterionischen Monomere zugeschrieben werden muss. Die starke Stabilisierung des monomeren Arginins ergibt sich durch intramolekulare Wechselwirkungen der geladenen Enden, welche sich aufgrund des flexiblen Grundgerüsts ausbilden können. Da diese *intra*molekulare Stabilisierung in Struktur **1** verhindert wird, sind die einzelnen Monomere deutlich energiereicher, sodass durch die *inter*molekulare Komplexierung stärker stabilisierte Dimere resultieren.

In Kapitel 4.2 wurde die strukturelle Ähnlichkeit von Arginin mit 2-(Guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylat (**1**) ausgenutzt, um die Abhängigkeit der Stärke der Dimerisierung von der Flexibilität der molekularen Struktur eingehender zu untersuchen. Die Wahl von Arginin als Modellsystem hat mehrere Vorteile bei der Untersuchung intermolekularer Guanidinium-Carboxylat Wechselwirkungen, nämlich (1.) handelt es sich hierbei um eine natürlich vorkommende Aminosäure mit biologischer Relevanz, welche multiple wasserstoffverbrückte Guanidinium-Carboxylat Wechselwirkungen aufweist, (2.) besitzt Arginine ein flexibles Grundgerüst, welches große konformative Veränderungen erlaubt, um bestmögliche Wasserstoffbrückenbindungen ausbilden zu können und (3.) sind viele experimentelle und theoretische Studien bereits vorhanden. Die Berechnung von Dimerisierungsenergien erfordert die Bestimmung der globalen (und manchmal auch lokalen) Minimumsstrukturen, was sogar für mittelgroße Moleküle mit flexiblem Gerüst ein nicht-triviales Problem darstellt. Es ist daher wichtig eine Strategie zu entwickeln, die es einem erlaubt, alle relevanten energetischen Minimumskonformere zu identifizieren und dabei die korrekte Rangfolge der relativen Energien mit hoher Zuverlässigkeit wiederzugeben.

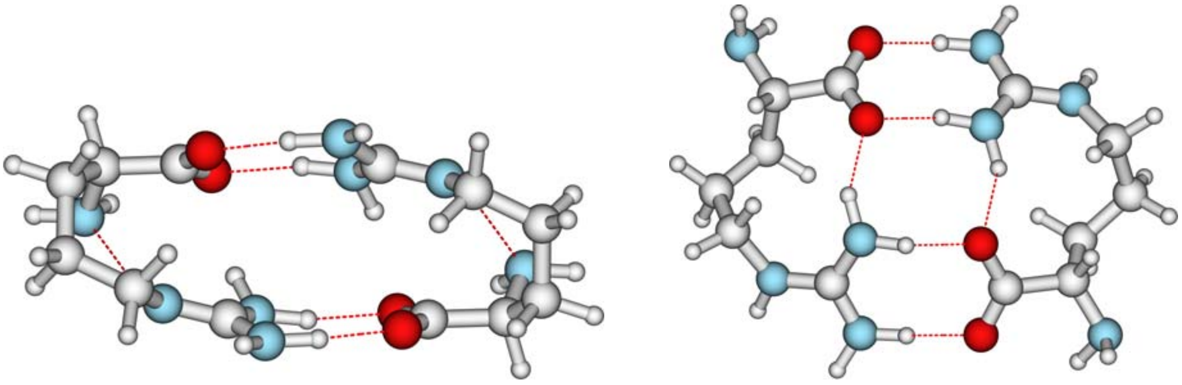
In Kapitel 4.2.1 wurden neue globale Minimumsstrukturen des kanonischen und zwitterionischen Arginins in der Gasphase bestimmt. Dies geschah mit Hilfe von umfangreichen kraftfeldbasierten Konformationssuchen in Verbindung mit *ab initio* Strukturoptimierungen der energetisch niedrigsten Konformere auf MP2 Niveau und großem Basissatz mit zusätzlichen diffusen Funktionen. Zudem zeigten die meisten der neu identifizierten Minimumskonformere sowohl des zwitterionischen als auch des kanonischen Tautomers geometrische Anordnungen mit bis dahin unbekannten gestapelten Orientierungen der endständigen Gruppen. Diese ungewöhnlichen geometrischen Strukturen wurden bislang von anderen theoretischen Studien aufgrund von Unzulänglichkeiten der dichtefunktionalbasierten Verfahren übersehen. Es konnte daher gezeigt werden, dass nur *post*-HF Elektronenkorrelationsmethoden in der Lage sind, für nicht-kovalent gebundene Systeme mit π-π-artigen Stapelwechselwirkungen die richtigen relativen Energien aller möglichen Konformere wiederzugeben. Durch CCSD(T) Rechnungen auf den MP2 optimierten Geometrien konnten alle auf MP2 Niveau gefundenen Trends bestätigt werden und es wurde letztendlich eine neuartige globale Minimumsstruktur (N1) gefunden, welche eine um mehr als 8 kJ mol^{-1} niedrigere Energie besitzt als die bislang veröffentlichten Konformere (siehe **Figure 50**).

Es wurde in der Literatur intensiv darüber diskutiert, ob Arginin in der Gasphase im zwitterionischen oder kanonischen Zustand vorliegt. [64,77c-d,79,94] Die starke Protonenaffinität der Guanidingruppe könnte hierbei die Energie aufwiegen, welche für eine Ladungstrennung nötig wäre. Es ist außerdem bekannt, dass ein schon zusätzliches Kation das zwitterionische Tautomer in Gasphase stabilisieren kann. Wir konnten zeigen, dass das niedrigste zwitterionische Konformer ZW1 (siehe **Figure 50**) um nur 7 kJ mol^{-1} energetisch höher liegt als das kanonische globale Minimum N1. Daher ist es nicht möglich, einen zwitterionischen Zustand in der Gasphase rigoros auszuschließen und ein eindeutiger Beweis könnte nur durch das Experiment erbracht werden. Eine Möglichkeit, um zwischen beiden Tautomeren zu unterscheiden, wäre der Vergleich der Wasserstoffstreckschwingungen zwischen 2600 und 4000 cm^{-1}, wie wir mit berechneten Vibrationsspektren der niedrigsten Konformere zeigen konnten.



N1 ZW1

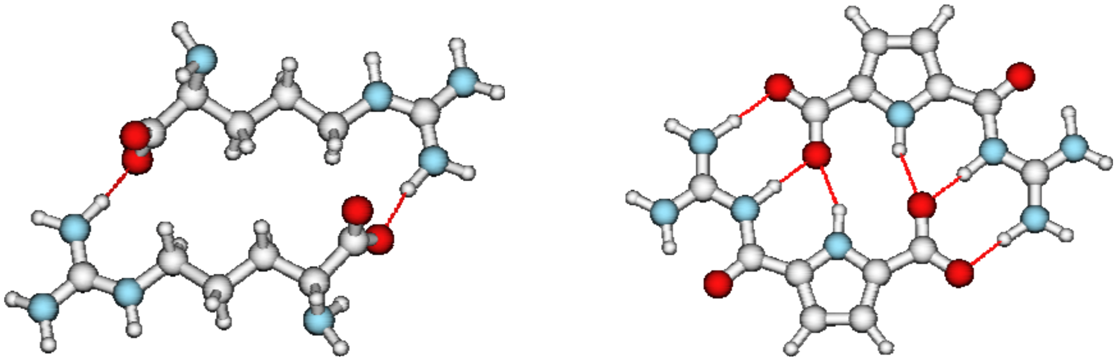
Figure 50 Minimumsstrukturen von Arginin: globales Minimum (N1) und das niedrigste zwitterionische Minimum (ZW1). Beide Strukturen wurden auf RI-MP2/TZVPP+ Niveau optimiert.



MMFF1 DZ1

Figure 51 Minimumsstrukturen des Arginin Dimers: globales Minimum (MMFF1) und das von Goddard III et al. veröffentlichte lokale Minimum (DZ1)[77a]. Beide Strukturen wurden auf RI-MP2/TZVPP+ Niveau optimiert.

Die gleiche Strategie für das Auffinden von energetischen Minimumskonformeren, wie sie bereits für das Arginin Monomer benutzt wurde, wurde auch im Falle der Dimere von Arginin verwendet. Im Gegensatz zu vorhergehenden theoretischen Untersuchungen, welche ausschließlich gerichtete Wasserstoffbrückenbindungen bevorzugten, ist die neue globale Minimumsstruktur ungefähr 60 kJ mol^{-1} (RI-MP2/TZVPP+//RI-MP2/TZVP) stabiler und weist ebenfalls eine gestapelte Orientierung der Guanidinium- und Carboxylatgruppen auf (siehe **Figure 51**).



410 kJ mol^{-1} **1:** 581 kJ mol^{-1}

Figure 52 MP2/TZVPP+ berechnete Dimere eines artifiziellen linearen Arginins (links) und 2-(Guanidiniocarbonyl)-1*H*-pyrrol-5-carboxylat (rechts).

Der Einfluss der Rigidität auf die Dimerstabilität wurde durch Berechnungen eines künstlich versteiften Arginin Dimersystems bewiesen. Der Vergleich mit dem 2-(Guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylat Dimer (siehe **Figure 52**) offenbarte, dass das artifiziell versteifte Arginin eine Dimerisierungsenergie besitzt, die, (a) doppelt so groß wie die der globalen Minimumsstruktur MMFF1 ist und (b) in einer ähnlicher Größenordnung wie für das Pyrrolderivat liegt, obwohl das Wasserstoffbrückenbindungsmuster weniger ausgeprägt ist. Die hohe Bindungsaffinität des 2-(Guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylat Dimers (**1**) ergibt sich daher zu etwa 50% aus der Rigidität der Monomere, welche jegliche intramolekulare Stabilisierung verhindert.

Um Vorschläge für ein verbessertes Carboxylatbindungsmotiv machen zu können, wurden in Kapitel 4.3 neuartige Strukturmotive mit veränderten Ringsystemen auf DFT Niveau untersucht. Die Studie von Benzol- und Pyridinderivaten zeigte die Wichtigkeit der Komplementarität der Monomere zueinander, welche mögliche geometrische Spannungen in den Dimerstrukturen minimiert. Es konnte erneut anhand vieler Strukturen klar gezeigt werden, dass jede intramolekulare Stabilisierung innerhalb der Monomere die Dimerisierungsenergie erheblich verringert. Die direkte Abhängigkeit der Dimerisierungsenergie von einem zunehmenden Dipolmoment wurde durch verschiedene anellierte Ringstrukturen bewiesen. Der Einfluss der Delokalisierung in den Monomeren auf die Dimerisierungsenergie wurde durch Veränderung der Elektronenstruktur von elektronisch entkoppelten Biphenylenen mit elektronenziehenden und elektronenschiebenden Substituenten untersucht. Diese Vorgehensweise führte jedoch zu keiner Verbesserung der

Dimerisierungsaffinität. Eine erhebliche Zunahme der Dimerisierungsenergie von etwa 90 kJ mol^{-1} bezogen auf das 2-(Guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylat Dimer wurde für das Indolderivat **10a** beobachtet, welches von einem größerem Dipolmoment in Verbindung mit dem guten Bindungsverhalten der Guanidiniumcarbonylpyrrol-Gruppe profitiert (siehe **Figure 53**). Die zusätzlich vorhandenen Wasserstoffbrückenbindungen des Pyrrolrings im Pyrrolderivat (**1**) als auch im Indolderivat (**10a**) verstärken die Stabilisierung des Dimers durch stärkere Fixierung des Carboxylats in der Bindungstasche („Gulliver-Effekt").

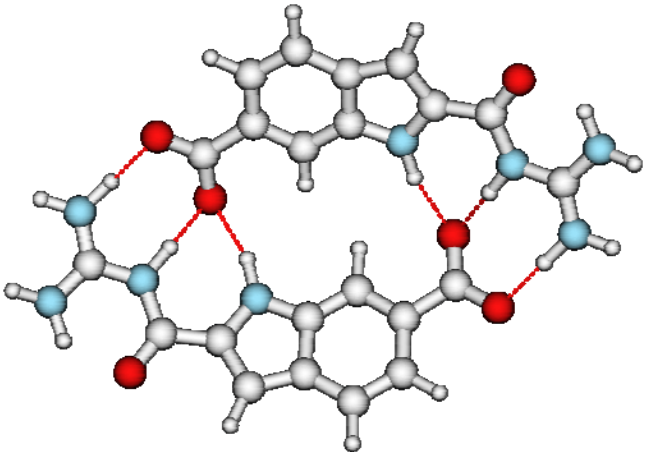


Figure 53 Optimierte Dimerstruktur von 2-Guanidiniocarbonylindole-6-carboxylat (**10a**) in Gasphase (B-LYP/TZVPP+).

Es wurde von Schmuck et al. vermutet, dass die Carbonylfunktion, welche an die Guanidiniumgruppe angrenzt, eine Erhöhung der Azidität der Wasserstoffdonorseite verursacht, die entscheidend für die starke Bindungsaffinität sei. Mit Hilfe von verschiedenen substituierten 7-Guanidinioindole-2-carboxylaten konnte gezeigt werden, dass die Carbonylfunktion hauptsächlich für eine gute Präorganisation verantwortlich ist, wohingegen der Effekt auf die Azidität eine geringere Bedeutung besitzt.

Der Einfluss von molekularen Solvenseffekten auf potentielle Energien wurde durch den Vergleich von im Kontinuum-Solvensmodell berechneten Energien von *m*-Guanidiniocarbonylbenzoat **7a** mit experimentellen Dimerisierungskonstanten untersucht, welche in unterschiedlichen DMSO/Methanol Mischungen gemessen wurden. Wir konnten beweisen, dass es nur mit Berücksichtigung von expliziten Lösungsmittelmolekülen möglich ist, die experimentell beobachteten Trends zu erklären. Zudem konnte gezeigt werden, dass die Berechnung von Gibbs freien Enthalpien für Assoziationsreaktionen weiterhin problematisch ist. Dieser Umstand ist darauf zurückzuführen, dass die thermodynamischen Beiträge für die kondensierte Phase mit Hilfe von analytischen Formeln für ideale Gase berechnet wurden. Allerdings zeigen die ITC gemessenen Dimerisierungskonstanten, dass die berechneten relativen Energieunterschiede bereits in guter Übereinstimmung mit den experimentellen Werten sind. Dies hat zur Folge, dass man die berechneten Gibbs freien Enthalpien dazu verwenden kann, die Dimerstabilität in einem gegebenen Lösungsmittel abzuschätzen, falls die experimentelle Assoziationskonstante eines strukturell verwandten Referenzsystem in demselben Solvens bekannt ist. Aufgrund dieser Ergebnisse scheint das 2-Guanidiniocarbonylindol-6-carboxylat (**10a**) ein sehr viel versprechender Kandidat zu sein, Dimere sogar in hochpolaren Lösungsmitteln ausbilden zu können.

Im letzten Kapitel wurden Kooperativitätseffekte in supramolekularen Systemen untersucht. Als Modellsysteme dienten hierbei Adenosin-Carbonsäure-Komplexe, deren berechnete NMR Verschiebungen mit experimentellen Niedrigtemperatur-NMR-Studien verglichen wurden. Wir konnten zeigen, dass nur durch die Verwendung von schwingungsgemittelten NMR Verschiebungen die experimentelle Protonenverschiebung reproduziert werden kann, welche unter Tieftemperaturbedingungen im Austauschregime von Wasserstoffbrückenbindungen erhalten wurde. Als Ergebnis dieser Studie zeigt sich, das durch die Verknüpfung von experimentellen und theoretischen Daten verlässliche Informationen über die individuellen Wasserstoffbrückenbindungsstärken und die Kooperativität innerhalb des multiplen Wasserstoffbrückenbindungsnetzwerkes von nicht-kovalent gebundenen Systemen erhalten werden können.

Chapter 7 Outlook

> *"An expert is a man who has made all the mistakes which can be made in a very narrow field."*
>
> --***Niels Henrik David Bohr***

In the course of this work an approach was presented which enables to quantify non-covalent interactions of supramolecular systems by using force-field conformational searches and subsequent *ab initio* computations. Hereby it has become apparent that the proper description of solvated assemblies is still unsatisfying. As already mentioned the polarizable continuum models like COSMO lack two important issues that are necessary for a sound description of any assembling process. On the one hand the contributions arising from solvent inhomogenities, dispersion terms or cavitation are neglected as only the electrostatic interactions between solvent and solute are considered. On the other hand, and probably the more critical point, the proper description of thermodynamic effects is crucial for a correct determination of free energies of association. However, in contrast to gas-phase systems no explicit analytical functions for the enthalpy and entropy contributions exist. Especially the contributions arising from translational degrees of freedom of the solution have a great influence on the stability of the complex under investigation. Most quantum mechanical program packages simply use the same formulas derived for gas-phase systems which results in erroneous free energies for the complexation process. On this account Williams and co-workers have proposed an approximation scheme which corrects for the overestimation of the translation motion. However, a correct description of the association in solution can only be given by simulation methods as no analytical functions without extensive approximations are known for the thermodynamic corrections for solutes. Hardware and software limitations make a quantum chemical description of the full system, *i.e.* solute surrounded by solvent molecules with boundary conditions applied, for the moment and in presumably the near future unfeasible and therefore at least parts of the system have to be approximated by force field methods. Hereby, two principally different proceedings are possible. On the one hand one can simulate the whole association process with a well-parameterized force-field being able to account for changes in the electronic structure during the aggregation. On the other hand, it is possible to employ a QM/MM ansatz (see Chapter 3.5) which incorporates the electrostatic effect of the solvent molecules as point charges acting on the quantum system.

Polarizable Force Fields

The performance of classical force fields is strongly dependent on the application field. This can be seen for CHARMM that was developed for large-scale bio-systems like proteins and nucleic acids or by MMFF94 which was designed primarily for smaller organic compounds. In case of the macromolecular force fields fast computations are mandatory whereas the experimental accuracy to which geometries can be determined is low compared to the possibilities available for small systems. Hence, experiments and computations match quite well for these types of force fields applied on macromolecular systems. In order to achieve a high efficiency most of the terms are truncated after the quadratic term affecting especially the calculation of non-bonded interactions that represent the bottle-neck of any force field. Moreover, the proteins und nucleic acids are dissected into residues (amino acids or nucleobases) which are parameterized to experimental data. As a result, the parameters employed in these force fields are excellent only for macromolecules containing the standard residues. A variation in the molecular structure yields either less accurate results or the force field is not even able to recognize the system as interactions between some atom types have not been parameterized. The more general force fields like MMFF94 are mainly developed to reproduce the potential energy surfaces of small to mid-size organic compounds in gas phase with a good accuracy. However, using an explicit solvation ansatz with several thousands of solvent molecules makes these force field types unfavourable for molecular simulations since the computations of non-bonded interactions become too time-consuming in comparison to macromolecular force fields. Moreover, all force fields represented so far show one major drawback: in all static force fields the electron distribution of the molecule is simply reduced to point charges located at the atom centres. Therefore, these force fields do not allow for systems in which the polarizability of the electron sheath has a large impact in the potential energy. This is the case when studying highly polar fluids like water or assembling processes in which strong intermolecular forces influence the charge distribution of the monomeric units. A solution to this dilemma is provided by the development of so-called “polarizable force-fields” which use either inducible multipoles or fluctuating charges in order to account for changes in the electron distribution. While the former method solves the interactions of the formally introduced multipoles self-consistently, the latter considers the partial charges as dynamic variables. The energy necessary for transferring a charge between two atom centres is usually approximated by a quadratic function. Currently, many research groups are developing new polarizable force fields which seem to be very promising for simulating association processes and non-covalent enzyme-ligand interactions. However, beside the inclusion of polarizability the application of such a force field is still highly dependent on the parameterization to experimentally or quantum chemically obtained data.

QM/MM – MD Calculations

A second approach to circumvent the inaccuracy of continuum solvation models for the calculation of supramolecular processes is to employ a QM/MM ansatz. The molecular assemblies are treated by semi-empirical or *ab initio* methods, whereas the surrounding solvent is approximated by force field methods. Regarding reaction studies of enzymes confined mainly to proton transfers in the active site, QM/MM calculations have been successfully applied in the recent past. In contrast to these well-defined chemical systems the investigation of assembling processes faces more fundamental difficulties which should be addressed briefly here. The main issues are

(1) Reaction coordinate
(2) System size and boundary conditions
(3) Constraints
(4) Thermodynamic correction factors
(5) QM level of theory

In the following the solvated dimer of 2-(guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate is taken as example to disclose the technical and principal hindrances that have to be overcome when performing a QM/MM-MD simulation of any supramolecular process in explicit solvent.

Re (1):

The right choice of a reaction coordinate is of decisive importance for a meaningful simulation since it is not possible to determine unambiguously the association pathway of the assembling process. Therefore, we have chosen a coordinate system in which the dimer is located in the z-plane and the origin is placed in the middle of the connecting line of both pyrrole hydrogen atoms being the x-coordinate (see **Figure 54** left). A displacement of each monomer in +x and –x direction, respectively, represents the reaction coordinate along which the phase space will be sampled stepwise in order to estimate the free energy of association.

The coupling parameter λ indicates the step length of the gradual separation of both monomers. A reasonable step size should be between 0.2 and 0.5 Å in order to achieve a distance of 5 Å in total within 10 to 20 steps. Since each step requires a full MD run including heating and sampling of several picoseconds, the simulations should be on one hand long enough to reach energy conservation, but on the other hand they should be as short as possible due to long computation times (one MD run of 60 ps needs 7 to 10 days on a 2.6 GHz Xeon-Workstation).

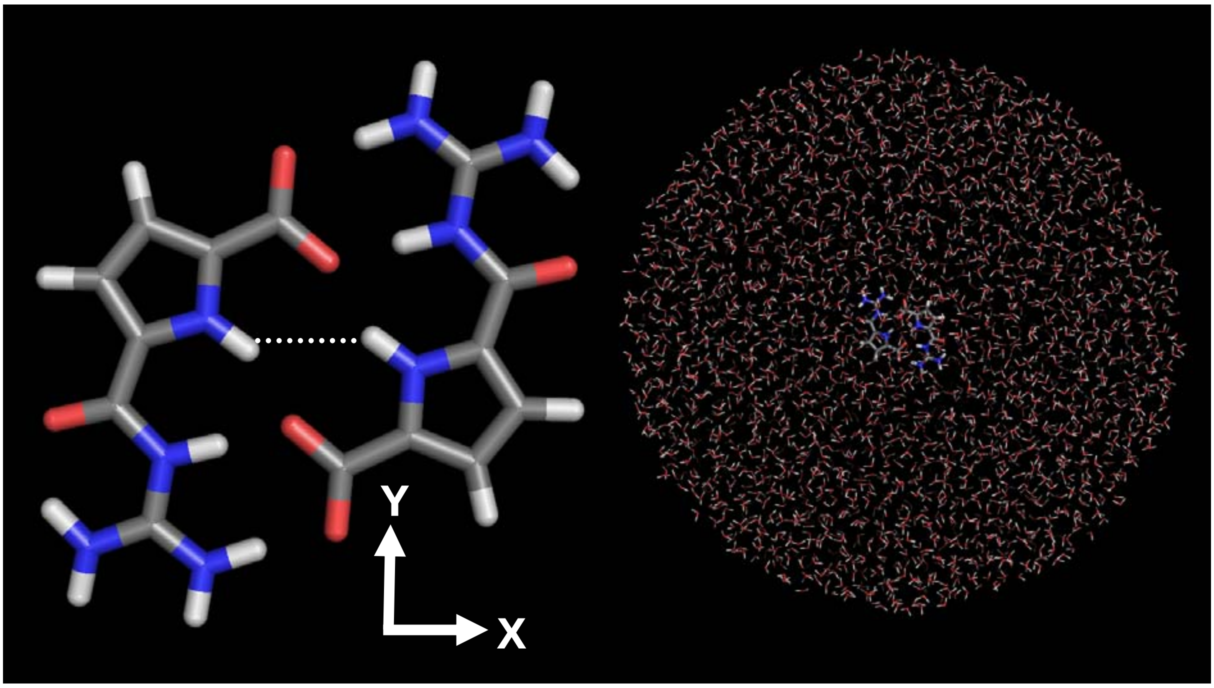


Figure 54 Left: The isolated 2-(guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate dimer system in the orientation of the reaction coordinate represented by the abscissa. Right: The solvated structure with stochastic boundary conditions

Re (2):

The system size of course determines the computational costs, but it should also be chosen large enough to minimize boundary effects. Since it is by now not possible to apply a periodic boundary condition in the ChemShell environment, one has to simulate a stochastic boundary potential by freezing an outer layer of solvent molecules of 15 Å. Adding the total size of the assembly and the maximum length of separation this yields a solvation sphere of 35 Å radius including 6 325 waters (see **Figure 54** right).

Re (3):

Besides the constraint of the coupling parameter λ that determines the distance between the two monomers other constraints can be useful in order to avoid unwanted geometrical arrangements. In this example a gradual displacement of the monomeric units without any further restriction would lead to a linear head-to-tail arrangement of the solutes. Since one is interested in the free energy of the non-covalent interaction of the entire system, both binding sites in the dimer should be separated simultaneously. This implies further constraints that should conserve in average the distance for all parts of the monomeric units by setting a harmonic potential between another atom pair.

Re (4):

The use of constraints results in unphysical approximations which should be corrected after the free energy simulation. Although all solvent molecules in the inner shell can freely move and thus adapt to the changing situation during the displacement of the monomers, the degrees of freedom of the solutes are still restricted. Therefore a correction factor for the translational and rotational motion of the solute should be added which could be determined by simple test systems.

Re (5):

Finally, the level of theory of the quantum mechanical part dominates the quality of the QM/MM calculation. Since for every time step one SCF cycle is performed, the calculations should be very fast and the theory should also exhibit robustness and reliability. The performance criterion restricts the application to semi-empirical methods or density functional theory in conjunction with density fitting methods (*e.g.* resolution of identity (RI) approach) and small basis sets. In the case of 2-(guanidiniocarbonyl)-1*H*-pyrrole-5-carboxylate DFT is already to time-consuming to perform MD simulations and therefore the semi-empirical AM1, PM3 and MNDO have been taken into consideration. Gas-phase optimization and comparison to higher level DFT calculations revealed that PM3 seems to be best suitable for describing the non-covalent interactions as it can be seen from **Table 27**.

Method	H-bond 1	H-bond 2	H-bond 3	ΔE (kJ mol^{-1})
B3LYP/TZVPP+	*1.68*	*1.58*	*1.78*	*464*
PM3	1.71	1.69	1.78	464
AM1	1.91	1.94	2.08	401
MNDO	-	-	2.87	290

Table 27 Gas-phase hydrogen bond length (in Å) and dimerization energies (in kJ mol^{-1}) of various semi-empirical methods in comparison to ab-initio values (see **Table 5**).

Coincidentally, the B3-LYP/TZVPP+ optimization with BSSE correction gives the exactly the same dimerization energy as PM3. It is known that PM3 calculates the hydrogen bond lengths and energies quite good, whereas AM1 gives distorted hydrogen bond geometries, low rotational barriers and no ideal flat π-systems. AM1 is underestimating the hydrogen bond network whereas MNDO is not even able to predict the planar geometry of the monomers correctly as the guanidinium groups are optimized perpendicular to the pyrrole ring system. Furthermore, it is also well known for its large deficiencies in describing weak interactions like hydrogen bonds.

In conclusion, a physically grounded and sufficiently accurate access to determine thermodynamically corrected aggregation energies of supramolecular processes is still a challenging task for future researches. However, the first steps towards this objective are made by preliminary QM/MM-MD simulations. Despite the above mentioned difficulties it still seems to be a feasible approach to calculate free energy differences of such non-covalent interactions. If it will be possible to perform such calculations, it will be also of great interest to include ions in order to simulate solutions with varying salt concentrations.

References and Notes

---

[1] For general references on supramolecular chemistry see: (a) Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry*, Wiley: Chichester, **2000**. (b) Lehn, J.-M. *Supramolecular Chemistry*; *Concepts and Perspectives*, VCH: Weinheim, **1995**. (c) Vögtle, F. *Supramolecular Chemistry*, J. Wiley & Sons: Chichester, **1991**.

[2] Hydrophobic effects between two compounds do not derive from additional forces but are due to the entropic gain upon complexation of the released polar solvent molecules. This lowering of the overall free energy is thus a pure entropic effect and the underlying interactions are footed on electrostatics.

3 For a review on artificial anion receptors see: Schmidtchen, F.P.; Berger, M. *Chem. Rev.* **1997**, *97*, 1609-1646.

[4] For an overview on biological anion receptors see: Mangani, S.; Ferraroni, M. “Natural Anion Receptors: Anion Recognition by Proteins”, in *Supramolecular Chemistry of Anions*, Bianchi, A.; Bowman-James, K.; Garcia-Espana, E., Wiley: New York, **1997**, 63-78.

[5] Teplyakov, A.; Wilson, K. S.; Orioli, P.; Mangani, S. *Acta Crystallogr. Sect.D* **1993**, *49*, 534-540.

[6] Schug, K. A.; Lindner, W. *Chem. Rev.* **2005**, *105*, 67-113.

[7] Dietrich, B.; Fyles, T. M.; Lehn, J.M.; Pease, L.G.; Fyles, D. L. *J. Chem. Soc., Chem. Commun.* **1978**, 934-936.

[8] Müller, G.; Riede, J.; Schmidtchen, F. P. *Angew. Chem.* **1988**, *100*, 1574-1775.

[9] Echavarren, A.; Galan, A.; Lehn, J. M.; de Mendoza, J. *J. Am. Chem. Soc.* **1989**, *111*, 4994-4995.

[10] Dixon, R. D.; Geib, S. J.; Hamilton, A. D. *J. Am. Chem. Soc.* **1992**, *114*, 365-366.

[11] (a) Schmuck, C.; Geiger, L. *Curr. Org. Chem.* **2003**, *7*, 1485-1502. (b) Schmuck, C. *Chem. Eur. J.* **2000**, *6*, 709-718. (c) Schmuck, C. *Chem. Commun.* **1999**, 843-844.

[12] (a) Best, M. D.; Tobey, S. L.; Anslyn, E. V. *Coord. Chem. Rev.* **2003**, *240*, 3-15. (b) Gale, P. A. *Coord. Chem. Rev.* **2003**, *240*, 191-221. (c) Fitzmaurice, R. J.; Kyne, G. M.; Douheret, D.; Kilburn, J. D. *J. Chem. Soc., Perkin Trans 1* **2002**, 841-864. (d) Snowden, T. S.; Anslyn,

---

E. V. *Curr. Opin. Chem. Biol.* **1999**, *3*, 740-746. (e) Beer, P. D.; Schmitt, P. *Curr. Opin. Chem. Biol.* **1997**, *1*, 475-482. (f) Seel, C.; Galán, A.; deMendoza, J. *Top. Curr. Chem.* **1995**, *175*, 101-132.

[13] Schmuck, C.; Machon, U. *Chem. Eur. J.* **2005**, *11*, 1109-1118.

[14] (a) Schmuck, C.; Wienand, W. *J. Am. Chem. Soc.* **2003**, *125*, 452-459. (b) Schmuck, C. *Eur. J. Org. Chem.* **1999**, 2397-2403.

[15] (a) Schmuck, C.; Rehm, T.; Gröhn, F.; Klein, K.; Reinhold, F. *J. Am. Chem. Soc.* **2006**, *128*, 1430-1431. (b) Schmuck, C.; Rehm, T.; Klein, K.; Gröhn, F. *Angew. Chem. Int. Ed.* **2007**, *46*, 1-6.

[16] Hunter, C. A. *Angew. Chem.. Int. Ed.* **2004**, *43*, 5310-5324.

[17] Schmuck, C. *Coord. Chem. Rev.* **2006**, *250*, 3053-3057.

[18] For supramolecular thermodynamics and experimental methods see: Schneider, H. J.; Yatsimirsky, A. *Principles and Methods in Supramolecular Chemistry*, Wiley-VCH: Weinheim, **2000**.

[19] For practical guidelines to determine binding constants see: Hirose, K. *J. Incl. Phen. Macrocyc. Chem.* **2001**, *39*, 193-209.

[20] Connors, K. A. *Binding Constants*, Wiley: New York, **1987**.

[21] (a) Schalley, C. A. *Mass Spec. Rev.* **2001**, *20*, 253-309. (b) Baytekin, B.; Baytekin, H. T.; Schalley, C. A. *Org. Biomol. Chem.* **2006**, *4*, 2825-2841.

[22] Gaskell, S. J. *J. Mass Spec.* **1997**, *32*, 677-688.

[23] (a) Bruins, A. P. *J. Chromatogr. A*. **1998**, *794*, 345-357. (b) Kerbarle, P. *Anal. Chem.* **1993**, *65*, 972A-986A.

[24] Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

[25] Cramer, C. J. 'Thermodynamic Properties' in *Essentials of Computational Chemistry*, Wiley: Chichester, **2002**.

[26] The expressions for the enthalpy and entropy corrections are taken from the output explanations of the module 'FREEH' of the Turbomole program package.[65]

---

[27] For general textbooks on quantum chemistry see: (a) Szabo, A.; Ostlund, N. S. *Modern quantum chemistry*, Dover: Mineaola, **1989**. (b) Jensen, F. *Introduction to computational chemisty*, Wiley: Chichester, **2007**. An overview on coupled cluster theory is given by: Crawford, T. D.; Schaefer III, H. F. ‘An Introduction to Coupled Cluster Theory for Computational Chemists’ in *Reviews in Computational Chemistry Vol. 14*, Lipkowitz, K. B.; Boyd, D. B. Wiley: Chichester, **1999**.

[28] A general textbook on DFT theory is given by: Koch, W.; Holthausen, M. C. *A Chemist’s Guide to Density Functional Theory*, Wiley-VCH: Weinheim, **2000**.

[29] The two-electron integrals are written for the sake of clarity in the chemist’s notation.

[30] The V_{nn} potential can be neglected since it represents only a constant within in the framework of the Born-Oppenheimer approximation.

[31] Hohenberg, P.; Kohn; W. *Phys. Rev.* **1964**, *136*, B364-B871.

[32] Kohn, W.; Sham, L. J. *Phys. Rev*. **1965**, *140*, A1133-A1138.

[33] Slater, J. C. *Phys. Rev.* **1951**, *81*, 385-390.

[34] Vosko, S. J.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200-1211.

[35] The numbers appearing in the acronyms indicate the year of publication: (a) PW91: Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1986**, *33*, 8800-8802. (b) B88: Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098-3100. (c) P86: Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822-8824. (d) PW91: Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671-6687. (e) LYP: Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789.

[36] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5684-5651.

[37] For a comprehensive textbook on force fields and modeling methods see: Leach, A. R. *Molecular Modelling – Principles and Applications*, Prentice Hall: Harlow, **2001**.

[38] Goodman, J. M.; Still, W. C. *J. Comp. Chem.*, **1991**, *12*, 1110-1117.

[39] Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc***.**, **1989**, *111*, 4379-4386.

[40] Kolossvary, I.; Guida, W. C. *J. Comp. Chem.*, **1999**, *20*, 1671-1684.

---

[41] An overview over various solvation models is given by ref. 37 in chapters 11.9 – 11.12.

[42] Reviews on continuum solvation are given by: (a) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999-3093. (b) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2017-2094.

[43] Klamt, A.; Schüürmann, G. J. *J. Chem. Soc. Perkin Trans*. 2 **1993**, *2*, 799-805.

[44] Schleyer, P. v. R. (ed.) *Encyclopedia of Computational Chemistry*, Wiley: Chichester, **1998**.

[45] Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1*, 70-77.

[46] The Car-Parrinello MD (CPMD) simulation method is often mentioned as a QM-MD simulation, but strictly speaking the CPMD method is not a full QM-MD simulation method in terms of solving the electronic Schrödinger equation (Born-Oppenheimer MD).

[47] General textbooks on molecular dynamics are: (a) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, Academic Press: San Diego, **2003**. (b) ref. 37, chapter 6-7.

[48] The factor $N!h^{3N}$ appearing in the partition function is dropped for the sake of clarity.

[49] Mezei, M.; Beveridge, D. L. 'Free energy simulations' in *Computer Simulation of Chemical and Biomolecular System, Vol. 482*, Annals of the NY Academy of Science: **1986**.

[50] Peter, C.; Oostenbrink, C.; van Dorp, A.; van Gunsteren, W. *J. Chem. Phys.* **2004**, *120 (6)*, 2652-2661.

[51] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227-249.

[52] Cramer, C. J. 'Hybrid Quantal/Classical Models' in *Essentials of Computational Chemistry*, Wiley: Chichester, **2002**.

[53] Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580-10594.

[54] (a) Reinhoudt, D. N.; Crego-Calama, M. *Science* **2002**, *295*, 2403-2407. (b) Menger, F. M. *Proc. Nat. Acad. Sci.* **2002**, *99*, 4818-4822. (c) Stoddart, F. J.; Tseng, H.-R. *Proc. Nat. Acad. Sci.* **2002**, *99*, 4797-4800. (d) Lehn, J.-M. *Proc. Nat. Acad. Sci.* **2002**, *99*, 4763-4768. (e) Whitesides, G. M.; Simanek, E. E.; Mathias, J. P.; Seto, C. T.; Chin, D.; Mammen, M.; Gordon, D. M. *Acc. Chem. Res.* **1995**, *28*, 37-44.

---

[55] (a) Whitesides, G. M.; Boncheva, M. *Proc. Nat. Acad. Sci.* **2002**, *99*, 4769-4774. (b) Philp, D.; Stoddart, J. F. *Angew. Chem.* **1996**, *108*, 1242-1286. (c) Lawrence, D. S.; Jiang, T.; Levitt, M. *Chem. Rev.* **1995**, *95*, 2229-2260.

[56] (a) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*, Oxford University Press, New York 1997. (b) Israelachvili, J. *Intermolecular & Surface Forces*, 2nd ed., Academic Press, London, 1992.

[57] For review articles on metal-templated self-assembly see: (a) Fujita, M. *Chem. Soc. Rev.* **1998**, *27*, 417-426. (b) Stang, P. J. *Chem. Eur. J.* **1998**, *4*, 19-27. (c) Linton, B.; Hamilton, A. D. *Chem. Rev.* **1997**, *97*, 1669-1681. (d) Stang, P. J.; Olenyuk, B. *Acc. Chem. Res.* **1997**, *30*, 502-518.

[58] For work on capsule formation in polar solvents based on the hetero association of oppositely charged ions see *e.g.*: (a) Grawe, T.; Schrader, T.; Zadmard, R.; Kraft, A. *J. Org. Chem.* **2002**, *67*, 3755-3763. (b) Corbellini, F.; Flammengo, R.; Timmerman, P.; Crego-Calama, M.; Veslius, K.; Heck, A. J. R.; Luyten, I.; Reinhoudt, D. N. *J. Am. Chem. Soc.* **2002**, *124*, 65969-6575. (c) Fiammengo, R.; Timmerman, P.; de Jong, F.; Reinhoudt, D. N. *Chem. Commun.* **2000**, 2313-2314. (d) Hamilin, B.; Jullien, L.; Derouet, C.; Hervé du Penhoat, C.; Berthault, P. *J. Am. Chem. Soc.* **1998**, *120*, 8438-8447. (e) Bok Lee, S.; Hong, J.-I. *Tetrahedron Lett.* **1996**, *37*, 8501-8504.

[59] For reviews on hydrophobic interactions see: (a) Widom, B.; Bhimalapuram, P.; Koga, Kenichiro. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3085-3093. (b) Pratt, L. R.; Pohorille, A. *Chem. Rev.* **2002**, *102*, 2671-2691. (c) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. *J. Phys. Chem. B* **2002**, *106*, 521-533.

[60] For examples of recent work on the importance of π-stacking in supramolecular aggregates see *e.g.*: (a) Lahiri, S.; Thompson, J. L.; Moore, J. S. *J. Am. Chem. Soc.* **2000**, *122*, 11315-11319. (b) Sirish, M.; Schneider, H. J. *J. Am. Chem. Soc.* **2000**, *122*, 5881-5882. (c) Guckian, K. M.; Schweitzer, B. A.; Ren, R. X.-F.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 2213-2222. (d) Isaacs, L.; Witt, D.; Fettinger, J. C. *Chem. Commun.* **1999**, 2549-2550.

[61] For a recent review on aromatic interactions see: Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. *J. Chem. Soc., Perkin Trans. 2* **2001**, 651-669.

---

[62] Some selected examples can be found in: (a) Cannizzaro, C. E.; Houk, K. N. *J. Am. Chem. Soc.* **2002**, *124*, 7163-7169. (b) Raymo, F. M.; Bartberger, M. D.; Houk, K. N.; Stoddart, J. F. *J. Am. Chem. Soc.* **2001**, *123*, 9264-9267. (c) Houk, K. N.; Menzer, S.; Newton, S. P.; Raymo, F. M.; Stoddart, J. F.; Williams, D. J. *J. Am. Chem. Soc. 1999*, *121*, 1479-1487. (d) Brown, S. P.; Schaller, T.; Seelbach, U. P.; Koziol, F.; Ochsenfeld, C.; Klärner, F.-G.; Spiess, H. W. *Angew. Chem.*, *Int. Ed.* **2001**, *40*, 717-720. (e) Kamieth, M.; Klarner, F.-G.; Diederich, F. *Angew. Chem.*, *Int. Ed.* **1998**, *37*, 3303-3306. (f) Ma, J. C.; Dougherty, D. A. Chem. Rev. **1997**, *97*, 1303-1324. (g) Gallivan, J. P.; Dougherty, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 870-874.

[63] (a) Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008-2010 (b) Pranata, J.; Wierschke, S.G.; Jorgensen W.L. J. *Am. Chem. Soc.***1991**, *113*, 28102819. (c) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 209-216; (d) Lukin, O.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 67756782; (e) Lukin, O.; Leszczynski, J. *J. Phys. Chem. A* **2003**, *107*, 9251-9252.

[64] Julian, R. R., Jarrold, M. F. *J. Phys. Chem. A*, **2004**, *108*, 10861-10864.

[65] TURBOMOLE V5.6 and V.8, since 1988 (a) Ahlrichs, R.; Bär, M.; Häser, M.; Horn H.; Kölmel, C. *Chem. Phys. Letters* **1998**, *162*, 165-169. (b) Häser M,; Ahlrichs, R. *J. Comput. Chem.* **1989**, *10*, 104-111. (c) Haase, F.; Ahlrichs, R. *J. Comp. Chem.* **1993**, *14*, 907-912. (d) Treutler, O; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346-354. (e) Deglmann, P.; May, K,; Furche, F.; Ahlrichs, R. *Chem. Phys. Letters* **2004**, *384*, 103-107. (f) Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331-340. (g) Deglmann, P.; Furche, F.; Ahlrichs. R. *Chem. Phys. Letters* **2002**, *362*, 511-518. (h) Deglmann, P.; Furche, F. *J. Chem. Phys.* **2002**, *117*, 9535-9538.

[66] Schäfer, A.; Huber, C; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829-5835.

[67] (a) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514 – 518. (b) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *242*, 652-660.

[68] (a) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187-4225. (b) Luque, J. F. ; Curutchet, C.; Munoz-Muriedas, J.; Bidon-Chanal, A.; Soteras, I..; Morreale, A.; Gelpi, J. L.; Orozco, M. *PCCP*, **2003**, *5*, 3827-3836. (c) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161-2200. (d) Orozco, M.; Colominas, C.; Luque, F. *J. Chem. Phys.* **1996**, *209*, 19-29. (e)

Curutchet, C.; Cramer, C. J.; Truhlar, D. G.; Ruiz-Lopez, M. F.; Rinaldi, D.; Orozco, M.; Luque, F. J. *J. Comp. Chem.* **2003**, *24*, 284-297.

[69] Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187-2193.

[70] Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys*. **1984**, *80*, 3265-3269.

[71] (a) Gaussian 03, Revision C.02, Pople, J. A. and coworkers; Gaussian, Inc., Wallingford CT, 2004. (b) NBO Version 3.1, E. D. Glendening, A. E. Reed, J. E. Carpenter, and F. Weinhold.

[72] (a) Kind C.; Reiher, M.; Neugebauer, J.; Hess, B. A. *SNF – a program for quantum chemical calculations of vibrational spectra. University of Erlangen-Nürnberg*, 1999-2001. (b) Neugebauer, J.; Reiher, M.; Kind, C.; Hess, B. A. *J. Comp. Chem.* **2002**, *23*, 895-910.

[73] (a) Doig, A. J.; Williams, D. H. *J. Am. Chem. Soc.* **1992**, *114*, 338-343. (b) Searle, M. S.; Williams, D. H. *J. Am. Chem. Soc.* **1992**, *114*, 10690-10697.

[74] Hupp, T.; Sturm, Ch.; Basílio Janke, E.M.; Pérez Cabre, M.; Weisz, K.; Engels, B. *J. Phys. Chem. A* **2005**, *109*, 1703-1712.

[75] (a) Xu, X.; Goddard III, W. A. *J. Phys. Chem. A* **2004**, *108*, 2305-2313. (b) Xu, X.; Zhang, Q.; Muller, R. P.; Goddard III, W. A. *J. Chem. Phys.* **2005**, *122*, 014105.

[76] MP2 computations employing the COSMO approach were not possible.

[77] (a) Julian, R.R.; Beauchamp, J. L.; Goddard III, W. A. *J. Phys. Chem. A* **2002**, *106*, 32-34. (b) Melo, A.; Ramos, M. J.; Floriano, W. B.; Gomes, J. A. N. F.; Leao, J. F. R.; Magalhaes, A. L.; Maigret, B.; Nascimento, M.C.; Reuter, N. *J. Molec. Struct. (Theochem)* **1999**, *463*, 81-90. (c) Rak, J.; Skurski, P.; Simons, J.; Gutowski, M. *J. Am. Chem. Soc.* **2001**, *123*, 11695-11707. (d) Z. B. Maksić, B. Kovačević, *J. Chem. Soc., Perkin Trans. 2* **1999**, 2623-2629.

[78] Julian, R. R.; Hodyss, R., Beauchamp, J. L. *J. Am. Chem. Soc*. **2001**, *123*, 3577-3583.

[79] Chapo, C. J.; Paul, J. B.; Provencal, R. A.; Roth, K.; Saykally, R. J. *J. Am. Chem. Soc.* **1998**, *120*, 12956-12957.

[80] (a) Jockusch, R. A.; Price, W. D.; Williams, E. R. *J. Phys. Chem. A* **1999**, *103*, 9266-9274. (b) Wyttenbach, T.; Witt, M.; Bowers, M. T. *J. Am. Chem. Soc*. **2000**, *122*, 3458.

---

[81] Schäfer, M.; Schmuck, C.; Geiger, L.; Chalmers, M. J.; Hendrickson, C. L.; Marshall, A. G. *Int. J. Mass. Spec* **2004**, *237*, 33-45.

[82] Schmuck, C.; Schäfer, M. unpublished results.

[83] (a) Barril, X. ; Alemán, C. ; Orozco, M.; Luque, F. J. *PROTEINS* **1998**, *32*, 67-79. (b) Zheng, Y.-J.; Ornstein, R. L. *J. Am. Chem. Soc.* **1996**, *118*, 11237-11243. (c) Mason, P. E.; Neilson, G.W.; Enderby, J. E.; Saboungi, M.-L.; Dempsey, C. E.; MacKerell Jr., A. D.; Brady, J. W. *J. Am. Chem. Soc*. **2004**, *126*, 11462-11470.

[84] (a) Ahn, D.-S.; Park, S.-W., Jeon, I..-S.; Lee, M.-K.; Kim, N.-H.; Han, Y.-H.; Lee, S. *J. Phys. Chem. B* **2003**, *107*, 14109-14118. (b) Jeon, L.-S.; Ahn, D.-S.; Park, S.-W.; Lee, S.; Kim, B. *Int. J. Quant. Chem*. **2005**, *101*, 55-66.

[85] Attempts to find a local minimum for the neutral structure of dimer **1** in a polar solvent (COSMO calculation) failed. All optimizations starting from neutral geometries (proton attached to the carboxylate groups) transformed into zwitterionic structures without any barrier. Geometrical structures with fixed O-H distances (≈1.044 Å) lie about 150 kJ/mol above the zwitterionic structure.

[86] Hossain, M. A.; Schneider, H.-J. *Chem. Eur. J.* **1999**, *5*, 1284-1290.

[87] As for **1** the lowest dissociation channel of these knock-out analogues in the gas phase would lead to the neutral monomers. We are, however, mainly interested in a comparison of the various binding schemes and in the results obtained for a polar solvent. Therefore, we will only compare the dissociation energies with respect to the zwitterionic monomers both in gas phase and in water.

[88] (a) Lau, E. Y.; Newby, Z. E.; Bruice, T. C.; *J. Amer. Chem. Soc.* **2001**, *123*, 3350-3357. (b) Laitinen, T.; Rouvinen, J.; Peräkylä, M.; *J. Org. Chem.* **1998**, *63*, 8157-8162. (c) Helten, H.; Schirmeister, T.; Engels, B. *J. Phys. Chem. A*, **2004**, *108*, 7691-7701. (d) Garau, C.; Frontera, A.; Quinonero, D.; Ballester, P.; Costa, A.; Deya, P. M, *J. Phys. Chem. A* **2004**, *108*, 9423-9427.

[89] Hay, B. P.; Gutowski, M.; Dixon, D. A.; Garza, J.; Vargas, R.; Moyer, B. A. *J. Amer. Chem. Soc.* **2004**, *126*, 7925-2934.

---

[90] The decrease assumes that the $CH_4^{\cdots -}$ Cl interaction similarly behaves upon solvation than all other effects.

[91] The major contribution to ΔS is the changes in translation and rotation resulting from the formation of one single dimer from two independent monomers. This part of ΔS, however, is not influenced by the solvent. Its effect on the already smaller absolute dissociation energies in solvent is therefore more pronounced.

[92] (a) Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.*1**983**, *168*, 867-885. (b) Singh, J.; Thornton, J. M.; Snary, M.; Campbell, S.F. *FEBS Letters* **1987**, *224*, 161-171.

[93] (a) Riordan, J. F.; McElvany, R. D.; Borders, C. L. *Science* **1977**, *195*, 884-886. (b) Gandini, D., Gogioso, L., Bolognesi, M., Bordo, D. *PROTEINS* **1996**, *24*, 439-449. (c) Sundaralingam, S., Sekhadrudu, Y. C., Yathindra, N., Ravichadran, V. *PROTEINS* **1987**, *2*, 64-71. (d) Lyu, P. C.; Gans, P. J.; Kallenbach, N. R. *J. Mol. Biol.* **1992**, *223*: 343-350.

[94] Skurski, P.; Gutowski, M.; Barrios, R.; Simons, J. *Chem. Phys. Lett.* **2001**, *337*, 143-150.

[95] (a) Julian, R. R.; Beauchamp, J. L. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 616-624. (b) Kinsel, G. R.; Zhao, Q.; Narayanasamy, J.; Yassin, F.; Dias, H. V. R.; Niesner, B.; Prater, K.; Marie, C. St.; Ly, L.; Marynick, D. S. *J. Phys. Chem. A* **2004**, *108*, 3153-3161.

[96] Schrader, T.; Hamilton A., Eds. *Functional Synthetic Receptors*, Wiley-VCH: Weinheim, **2005**, ISBN 3527306552.

[97] Strittmatter, E. F.; Williams, E. R. *J. Phys. Chem. A* **2000**, *104*, 6069-6076.

[98] Schrödinger, Inc., 101 SW Main Street, Suite 1300, Portland, OR, 97204

[99] Shenkin, P. S.; McDonald, D. Q. *J. Comp. Chem.*, **1994**, *15*, 899-916.

[100] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11235.

[101] Halgren, T. A. *J. Comp. Chem.* **1996**, *17*, 490-512; 520-552; 553-586; 587-615; 616-641.

[102] (a) Becke, A. D. *J. Chem. Phys.* **1997**, *107*, 8554-8560 .(b) Binkley, J. S. ; Pople, J. A. *J. Am. Chem. Soc.* **1980**, *102*, 939-947 .

[103] (a) Engels, B.; Schoeneboom, J. C.; Munster, A. F.; Groetsch, S.; Christl, M. *J. Am. Chem. Soc.* **2002**, *124*, 287-297. (b) Adam, W.; Bottke, N.; Engels, B.; Krebs, O. *J. Am. Chem. Soc.*

2001, *123*, 5542-5548. (c) Schlund, S.; Mladenovic, M.; Basilio Janke, E. M.; Engels, B. Weisz, K. *J. Am. Chem. Soc.* **2005**, *127*, 16151-16158. (d) Hupp. T.; Sturm, C.; Basilio Janke, E. M.; Perez Cabre, M.; Weisz, K.; Engels, B. *J. Phys. Chem.*. **2005**, *109*, 1703-1712.

[104] Poirier, R.; Kari, R.; Csizmadia, I. G. *Handbook of Gaussian Basis Sets (Physical sciences data; 24)*, Elsevier: Amsterdam, 1985.

[105] MOLPRO, version 2006.1, a package of *ab initio* programs, H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, P. Celani, T. Korona, G. Rauhut, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, C. Hampel and G. Hetzer, A. W. Lloyd, S. J. McNicholas, W. Meyer and M. E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni and T. Thorsteinsson.

[106] (a) Dunning Jr, T. H. *J. Chem. Phys.* **1989**, *90*, 1007-1023. (b) Wilson, A. K.; van Mourik, T.; Dunning Jr, T. H. *J. Mol. Struct.* **1996**, *388*, 339-349. (c) Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796-6806.

[107] Scott, A. P., Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502-16513.

[108] Grimme, S. *Rev. Comp. Chem.* **2004**, *20* , 153-218.

[109] Ercanli, T.; Boyd, B. *J. Chem. Inf. Model.* **2005**, *45*, 591-601.

[110] (a) Helten, H.; Schirmeister, T.; Engels, B. *J. Org. Chem.* **2005**, *70*, 233-237. (b) Helten, H.; Schirmeister, T.; Engels, B. *J. Phys. Chem. A.* **2004**, *108*, 7691-7701.

[111] Mladenovic, M.; Schirmeister, T.; Thiel, S.; Thiel, W.; Engels, B. *Chem. Med. Chem.* **2007**, *2*, 120-128.

[112] Huheey, J.E.; Keiter, E.A.and Keiter, R.L. in *Inorganic Chemistry : Principles of Structure and Reactivity*, 4th edition, HarperCollins, New York, USA, 1993.W.W. Porterfield in *Inorganic chemistry, a unified approach*, Addison Wesley Publishing Co., Reading Massachusetts, USA, 1984

[113] King, R. A.; Crawford, T. D.; Stanton, J. F.; Schaefer III, H. F. *J. Am. Chem. Soc.* **1999**, *121*, 10788-10793.

[114] Mason, P. E.; Neilson, G. W.; Enderby, J. E.; Saboungi, M.-L.; Dempsey, C. E.; MacKerell, A. D., Jr.; Brady, J. W. *J. Am. Chem. Soc.*, **2004**, 126, 11462-11470.

---

[115] Helgekar, T.; Jørgensen, P.; Olsen, Jeppe Molecular Electronic-Structure Theory, Wiley & Sons: Chichester, 2000.

[116] (a) Kirchner, B.; Reiher M. *J. Am. Chem. Soc.* **2005**, *127*, 8748-8756. (b) Kirchner, B.; Reiher, M. "Theoretical Methods in Supramolecular Chemistry" in *Analytical Methods in Supramolecular Chemistry*, Schalley, C. A., Wiley-VCH: Weinheim, **2007**.

[117] Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer: Berlin, 1994.

[118] Lonergan D. G.; Deslongchamps G.; Tomás S. *Tetrahedron Lett*. **1998**, *39*, 7861-7864.

[119] Hibbert, F.; Emsley, J. *Adv. Phys. Org. Chem*. **1990**, *26*, 255-279.

[120] Iwahashi, H.; Kyogoku, Y. *J. Am. Chem. Soc*. **1977**, *99*, 7761-7765.

[121] (a) Siegel, J. S.; Anet, F. A. L. *J. Org. Chem*. **1988**, *53*, 2629-2630. (b) Golubev, N. S.; Denisov, G. S. *J. Mol. Struct*. **1992**, *270*, 263-276. (c) Weisz, K.; Jähnchen, J.; Limbach, H.-H. *J. Am. Chem. Soc*. **1997**, *119*, 6436-6437. (d) Dunger, A.; Limbach, H.-H.; Weisz, K. *Chem. Eur. J.* **1998**, *4*, 621-628. (e) Smirnov, S. N.; Benedict, H.; Golubev, N. S.; Denisov, G. S.; Kreevoy, M. M.; Schowen; R. L.; Limbach, H.-H. *Can. J. Chem*. **1999**, *77*, 943-949. (f) Dunger, A.; Limbach, H.-H.; Weisz, K. *J. Am. Chem. Soc*. **2000**, *122*, 10109-10114. (g) Basílio Janke, E. M.; Dunger, A.; Limbach, H.-H.; Weisz, K. *Magn. Reson. Chem*. **2001**, *39*, 177-182.

[122] Basílio Janke, E. M.; Limbach, H.-H.; Weisz, K. *J. Am. Chem. Soc*. **2004**, *126*, 2135-2141.

[123] Kumar, G. A.; Allister, M. A. *J. Org. Chem.* **1998**, *63*, 6968-6972.

[124] Kar, T.; Scheiner, S. *J. Phys. Chem A* **2004**, *108*, 9161-9168.

[125] Dingley, A. J.; Masse, J. E.; Peterson, R. D.; Barfield, J. F.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 6019-6027.

[126] (a) Tolstoy, P. M.; Schah-Mohammedi, P.; Smirnov, S. N.; Golubev, N. S.; Denisov, G. S.; Limbach, H.-H. *J. Am. Chem. Soc*. **2004**, *126*, 5621-5634. (b) Limbach, H.-H.; Pietrzak, M.; Benedict, H.; Tolstoy, P. M.; Golubev, N. S.; Denisov, G. S. *J. Mol. Struct.* **2004**, *706*, 115-119. (c) Smirnov, S. N.; Benedict, H.; Golubev, N. S.; Denisov, G. S.; Kreevoy, M. M.; Schowen, R. L.; Limbach, H.-H. *Can. J. Chem*. **1999**, *77*, 943-949. (d) Benedict, H.; Limbach H.-H.; Wehlan, M.; Fehlhammer, W.-P.; Golubev, N. S.; Janoschek, R. *J. Am. Chem. Soc.*

1998, *120*, 2939-2950. (e) Ramos, M.; Alkorta, I.; Elguero, J.; Golubev, N. S.; Denisov, G. S.; Benedict, H.; Limbach H.-H. *J. Phys Chem. A* **1997**, *101*, 9791-9800. (f) Smirnov, S. N.; Golubev, N. S.; Denisov, G. S.; Benedict, H.; Schah-Mohammedi, P.; Limbach, H.-H. *J. Am. Chem. Soc.* **1996**, *118*, 4094-4101. (g) Benedict, H.; Hoelger, C.; Aguilar-Parrilla, F.; Fehlhammer, W. P.; Wehlan, M.; Janoschek, R.; Limbach, H.-H. *J. Mol. Struct.* **1996**, *378*, 11-16.

[127] (a) Brown, S. P.; Schaller, T.; Seelbach, U. P.; Koziol, F.; Ochsenfeld, C.; Klärner, F.-G.; Spiess, H. W. *Angew. Chem. Int. Edit.* **2001**, *40*, 717-720. (b) Ochsenfeld, C.; Kussmann, J.; Koziol, F. *Angew. Chem. Int. Edit.* **2004**, *43*, 4485-4489. (c) Fokkens, M.; Jasper, C.; Schrader, T.; Koziol, F.; Ochsenfeld, C.; Polkowska, J.; Lobert, M.; Kahlert, B.; Klärner, F.-G. *Chem. Eur. J.* **2005**, *11*, 477-494. (d) Ochsenfeld, Ch. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2164-2159. (e) Ochsenfeld, Ch.: Brown, S. P.; Schnell, I.; Gauss, J.; Spiess, H. W. *J. Am. Chem. Soc.* **2001**, *123*, 2597-2606.

[128] (a) Goswami, S.; Hamilton, A. D. *J. Am. Chem. Soc.* **1989**, *111*, 3425-3426. (b) Zimmermann, S. C.; Wu, W. *J. Am. Chem. Soc.* **1989**, *111*, 8054-8055. (c) Adrian, J. C.; Wilcox, C. S. *J. Am. Chem. Soc.* **1989**, *111*, 8055-8057. (d) Conn, M. M.; Deslongchamp, G.; De Mendoza, J.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1993**, *115*, 3548-3557. (e) Alkorta, I.; Elguero, J.; Goswami, S.; Mukherjee, R. *J. Chem. Soc., Perkin Trans. 2* **2002**, 894-898. (f) Castellano, R. K.; Gramlich, V.; Diederich, F. *Chem. Eur. J.* **2002**, *8*, 118-129.

[129] Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1992**, *61*, 1053-1095.

[130] (a) Benzi, C.; Crescenzi, O.; Pavone, M.; Barone, V. *Magn. Reson. Chem.* **2004**, *42*, 57-67. (b) Golubev, N. S.; Melikova, S. M.; Shchepkin, D. N.; Shenderovich, I. G.; Tolstoy, P. M.; Denisov, G. S. *Z. Phys. Chem.* **2003**, *217*, 1549-1563. (c) Ruden, T. A., Ruud, K. *Ro-Vibrational Corrections to NMR Parameters* in *Calculation of NMR and EPR Parameters*, Ed. Kaupp, M.; Bühl, M.; Malkin, V. G., Wiley-VCH: Weinheim, 2004; pp 153-173. (d) Ruud, K.; Astrand, P.-O.; Taylor, P. R. *J. Am. Chem. Soc.* **2001**, *123*, 4826-4833. (e) Sundholm, D.; Gauss, J.; Schäfer, A. *J. Chem. Phys.* **1996**, *105*, 11051-11059.

[131] Stare, J.; Jezierska, A.; Ambrozic, G.; Kosir, I. J.; Kidric, J.; Koll, A.; Mavri, J.; Hadzi, D. *J. Am. Chem. Soc.* **2004**, *126*, 4437-4443.

---

[132] (a) Munch, M.; Hansen, A. E.; Hansen, P. E.; Bouman, T. D. *Acta. Chem. Scand.* **1992**, *46*, 1065-1071. (b) Abildgaard, J.; Bolvig, S.; Hansen, P. E. *J. Am. Chem. Soc.* **1998***, 120*, 9063-9069.

[133] Dziembowska, T; Rozwadowski, Z.; Hansen, P. E. *J. Mol. Struct.* **1997***, 437*, 189-199. (b) Perrin, C. L; *Science* **1994**, *266*, 1665-1668.

[134] (a) Alkorta, I.; Elguero, J.; Mo, O.; Yanez, M.; Del Bene, J. E. *J. Phys. Chem. A*, **2002**, *106*, 9325-9330. (b) Del Bene, J. E.; Elguero, J.; Alkorta, I.; Mo, O.; Yanez, M. *J. Phys. Chem. A*, **2005**, *109*, 2350-2355.

[135] (a) Steiner, T.; Saenger, W. *J. Am. Chem. Soc*. **1992**, *114*, 7123-7126. (b) Steiner, T.; Saenger, W. *Acta Cryst. B* **1994**, *50*, 348-357. (c) Steiner, T. *J. Chem. Soc. Chem. Commun.* **1995**, 1331-1332. (d) Steiner, T. *J. Phys. Chem. A* **1998**, *102*, 7041-7052.

[136] Wüthrich, K. *NMR of Proteins and Nucleic Acids*, Wiley-Interscience: New York, 1986, pp. 205-214.

[137] Garcia-Viloca, M.; Gelabert, R.; Gonzalez-Lafont, A.; Moreno, M.; Lluch, J. *J. Am. Chem. Soc.* **1998***, 120*, 10203-10209.

[138] Stare, J.; Mavri, J.; Ambrozic, G.; Hadzi, D. THEOCHEM **2000**, *500*, 429-440.

[139] (a) Somorjai, R. L.; Hornig, D. F. *J. Chem.Phys*. **1962**, *36*, 1980-1987. (b) Marechal, Y.; Witkowski, A. *J. Chem.Phys*. **1968**, *48*, 3697-3701. (c) Singh, T. R.; Wood, J. L. *J. Chem.Phys*. **1969**, *50*, 3572-3576. (d) Janoschek, R.; Weidemann, E. G.; Pfeiffer, H.; Zundel, G. *J. Am. Chem. Soc*. **1972**, *94*, 2387-2396.

[140] (a) Sahu, P. K. ; Mishra, R. K.; Lee, S.-L. *J. Chem. Phys. A* **2005**, *109*, 2887-2893. (b) Rabuck, A. D.; Scuseria, G. E. *Theor. Chem. Acc.* **2000**, *104*, 439-448.

[141] (a) Golubev, N. S.; Denisov, G. S.; Smirnov, S. N.; Shchepkin, D. N.; Limbach, H.-H. *Z. Phys. Chem.* **1996***, 196*, 73-84. (b) Shenderovich, I. G.; Burtsev, A. P.; Denisov, G. S.; Golubev, N. S.; Limbach, H.-H. *Magn. Reson. Chem*. **2001**, *39*, S91-S99.

[142] (a) Gauss, J.; Stanton, J. F. *Electron-Correlated Methods for the Calculation of NMR Chemical Shifts* in *Calculation of NMR and EPR Parameters*, Ed. Kaupp, M.; Bühl, M.; Malkin, V. G., Wiley-VCH: Weinheim, 2004. (b) van Wüllen, C. *Chemical Shifts with Hartree-Fock and Density Functional Methods* in *Calculation of NMR and EPR Parameters*,

---

Ed. Kaupp, M.; Bühl, M.; Malkin, V. G., Wiley-VCH: Weinheim, 2004. (c) Wang, B.; Fleischer, U.; Hinton, J. F.; Pulay, P. *J. Comp. Chem.* **2001**, *22*, 1887-1895. (d) Wang, B.; Hinton, J. F.; Pulay, P. *J. Comp. Chem.* **2003**, *23*, 492-497. (e) Wang, B.; Hinton, J. F.; Pulay, P. *J. Phys. Chem A* **2003**, *107*, 4683-4687.

[143] (a) Peric, M.; Hachney, M. R. J.; Grein, F. *J. Chem. Phys.* **2000**, *113*, 9011-9021. (b) Peric, M.; Runau, R.; Roemelt, J.; Peyerimhoff, S. D.; Buenker, R. J. *J. Mol. Spectrosc.* **1979**, *78*, 309-332.

[144] Lancelot, G. *J. Am. Chem. Soc*. **1977**, *99*, 7037-7042.

[145] Pistolis, G.; Paleos C. M.; Malliaris, A. *J. Phys. Chem*. **1995**, *99*, 8896-8902.

[146] Rao, P.; Ghosh, S.; Maitra, U. *J. Phys. Chem. B* **1999**, *103*, 4528-4533.

[147] Golubev, N. S.; Smirnov, S. N.; Gindin, V. A.; Denisov, G. S.; Benedict, H.; Limbach, H.-H. *J. Am. Chem. Soc.* **1994***, 116*, 12055-12056.

[148] (a) Searles, D. J.; Huber, H. *Molecular Dynamics and NMR Parameter Calculations* in *Calculation of NMR and EPR Parameters*, Ed. Kaupp, M.; Bühl, M.; Malkin, V. G., Wiley-VCH: Weinheim, 2004; a good description of the problems can be taken from ref. 126e.

LEBENSLAUF

Persönliche Daten

Name	Sebastian Schlund
Geburtsdatum	24. Februar 1978
Geburtsort	Würzburg
Adresse	Josefplatz 13 97080 Würzburg

Schulbildung

1984-1988	Grundschule Würzburg-Lengfeld
1988-1997	Friedrich-Koenig-Gymnasium, Würzburg Abschluss Abitur

Zivildienst

09/1997-09/1998	Jugendgästehaus Venusberg, Bonn

Studium

10/1998-02/2003	Studium der Chemie an der Julius-Maximilians-Universität Würzburg
02/2003-11/2003	Diplomarbeit im Arbeitskreis von Prof. Dr. Bernd Engels mit dem Titel: *"Theoretical Studies of Guanidiniocarbonyl Pyrrole Complexes"*
11/2003	Diplom
12/2003-09/2007	Promotionsarbeit im Arbeitskreis von Prof. Dr. Bernd Engels mit dem Titel: *"Quantifying Non-covalent Interactions -* *Rational in silico Design of Guanidinium-based Carboxylate Receptors"*

DANKSAGUNG

An dieser Stelle möchte ich allen danken, die es mir ermöglicht haben, diese Arbeit anzufertigen und mir bei Fragen oder Problemen mit Rat und Tat zur Seite standen.

Besonderer Dank gilt hierbei allen Mitarbeitern, F-Praktikanten und Ehemaligen des Arbeitskreises Engels, mit denen ich auch abseits der Arbeit viel Spaß hatte und die mich die letzten Jahre begleitet und unterstützt haben. Hervorzuheben ist hier vor allem Milena Mladenovic, mit der ich das schwierige Gebiet der QMMM Simulationen gemeinsam durchkämpft habe.

Weiterhin möchte ich Herrn Priv.-Doz. Dr. Reinhold Fink dafür danken, dass er sich immer die Zeit genommen hat, mir bei den Schwierigkeiten der Quantenchemie zu helfen und meine Fragen zu allerlei theoretischen Problemen zu beantworten.

Auch unserer guten Seele des Arbeitskreises, Uschi Rüppel, gilt mein besonderer Dank dafür, dass sie uns Doktoranden stets hilfreich beim alltäglichen bürokratischen Wahnsinn zur Seite stand.

Außerdem möchte ich meinen Kooperationspartnern, Herrn Prof. Dr. Carsten Schmuck (Uni Würzburg) und Herrn Prof. Dr. Klaus Weisz (Uni Greifswald), für die interessanten Fragestellungen und die angenehmen und fruchtbaren Diskussionen danken.

Zuletzt gilt mein besonderer Dank meinem Doktorvater Herrn Prof. Dr. Bernd Engels für die Möglichkeit, diese Arbeit anfertigen zu können. Er hatte stets ein offenes Ohr für Probleme wissenschaftlicher, aber auch privater Natur und die angenehme Arbeitsatmosphäre trug entscheidend zum Gelingen dieser Arbeit bei. Zudem möchte ich ihm danken, dass immer eine Finanzierung der Doktorandenstelle möglich war und ich mir darüber gerade in der Endphase der Dissertation keine Gedanken machen musste.

Hiermit erkläre ich an Eides statt, dass ich die Dissertation

„Quantifying Non-covalent Interactions –
Rational *in-silico* Design of
Guanidinium-based Carboxylate Receptors“

selbständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe.

Ich erkläre weiterhin, dass diese Dissertation weder in gleicher oder anderer Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ich habe früher außer den mit dem Zulassungsgesuch urkundlich vorgelegten Graden keine weiteren akademischen Grade erworben oder zu erwerben versucht.

Würzburg, 27.07.2007

Sebastian Schlund