Evaluation and Development of Quantum Chemical Methodologies for Noncovalent Interactions and Supramolecular Thermochemistry

Rebecca Sure

Dissertation

Evaluation and Development of Quantum Chemical Methodologies for Noncovalent Interactions and Supramolecular Thermochemistry

Dissertation

zur Erlangung des Doktorgrades
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Rebecca Sure
aus Iserlohn

– Bonn 2015 –

Dekan:	Prof. Dr. Ulf-G. Meißner
Erster Gutachter:	Prof. Dr. Stefan Grimme
Zweiter Gutachter:	Prof. Dr. Thomas Bredow
Tag der Promotion:	29.01.2016
Erscheinungsjahr:	2016

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

Publications

Parts of this thesis have already been published in international, peer-reviewed journals:

1. R. Sure and S. Grimme, "Corrected small basis set Hartree-Fock method for large systems.", *J. Comput. Chem.* **2013**, *34*, 1672–1685.

2. A. Gansäuer, C. Kube, K. Daasbjerg, R. Sure, S. Grimme, G. D. Fianu, D. V. Sadasivam and R. A. Flowers, "Substituent effects and supramolecular interactions of titanocene(III) chloride: Implications for catalysis in single electron steps." *J. Am. Chem. Soc.* **2014**, *136*, 1663–1671.

3. R. Sure, J. Antony and S. Grimme, "Blind prediction of binding affinities for charged supramolecular host–guest systems: achievements and shortcomings of DFT-D3.", *J. Phys. Chem. B* **2014**, *118*, 3431–3440.

4. G. Meyer-Eppler, R. Sure, A. Schneider, G. Schnakenburg, S. Grimme and A. Lützen, "Synthesis, chiral resolution, and absolute configuration of dissymmetric 4,15-difunctionalized [2.2]paracyclophanes." *J. Org. Chem.* **2014**, *79*, 6679–6687.

5. J. Antony, R. Sure and S. Grimme, "Using dispersion-corrected density functional theory to understand supramolecular binding thermodynamics.", *Chem. Commun.* **2015**, *51*, 1764–1774.

6. R. Sure and S. Grimme, "Comprehensive benchmark of sssociation (free) energies for realistic host–guest complexes", *J. Chem. Theory Comput.* **2015**, *11*, 3785–3801.

7. R. Sure, J. G. Brandenburg and S. Grimme, "Small basis set first-principles quantum chemical methods for large molecular and periodic systems", *ChemistryOpen* **2015**, accepted.

Further Publications:

1. A. Gansäuer, M. Seddiqzai, T. Dahmen, R. Sure and S. Grimme, "Computational study of the rate constants and free energies of intramolecular radical addition to substituted anilines" *Beilstein J. Org. Chem.* **2013**, *9*, 1620–1629.

2. A. Gansäuer, D. von Laufenberg, C. Kube, T. Dahmen, A. Michelmann, M. Behlendorf, R. Sure, M. Seddiqzai, S. Grimme, D. V. Sadasivam, G. D. Fianua and R. A. Flowers "Mechanistic study of the titanocene(III)-catalyzed radical arylation of epoxides." *Chem. Eur. J.* **2015**, *21*, 280–289.

Conference presentations:

1. Poster: "Dispersion Corrected Hartree-Fock/Small Basis Set Calculations for Large Supramolecular and Biochemical Systems", *48th Symposium on Theoretical Chemistry*, **2012**, Erlangen, Germany.

2. Poster: "Thermochemistry of Supramolecular Complexes Using DFT-D3 and HF/Minimal Basis Set Approaches", *Symposium on Weak Molecular Interactions*, **2013**, Pécs, Hungary.

3. Poster: "Supramolecular Host-Guest Binding Affinities: Can DFT-D3 Win a Blind Test Challenge?", *European Summerschool in Quantum Chemistry*, **2013**, Palermo, Italy.

4. Poster: "Supramolecular Host-Guest Binding Affinities: Can DFT-D3 Win a Blind Test Challenge?", *49th Symposium on Theoretical Chemistry*, **2013**, Karlsruhe, Germany.

5. Poster: "The S30L Test Set for Supramolecular Host-Guest Association Free Energies", *Sostrup Summer School - Quantum Chemistry and Molecular Properties*, **2014**, Ry, Denmark.

6. Poster: "The S30L Test Set for Supramolecular Host-Guest Association Free Energies", *50th Symposium on Theoretical Chemistry*, **2014**, Vienna, Austria.

7. Talk: "Supramolecular Association Free Energies by Dispersion Corrected Density Functional Theory", *Symposium on Supramolecular Chemistry*, **2015**, Berlin, Germany.

8. Talk: "Calculating Association Free Energies of Realistic Host-Guest Complexes", *Women in Science@Spin Centers*, **2015**, Eisenach, Germany.

9. Poster: "1812 Fullerenes: DFT-D3 Based Investigation of the Stability of the C_{60} Isomers", *51th Symposium on Theoretical Chemistry*, **2015**, Potsdam, Germany.

Abstract

This thesis focuses on the application and development of electronic structure methods for noncovalent interactions in general and the evaluation of multilevel methodologies for an accurate description of supramolecular thermochemistry in particular. Noncovalent interactions are omnipresent in systems of various domains of science, such as supramolecular chemistry, structural biology, and surface science. Within supramolecular chemistry, host-guest complexes are of particular importance due to their diverse applicability in various fields like molecular recognition or self-assembly.

The binding situation in a supramolecular complex is often unknown and sampling many different conformations is desired. Therefore, the first part of this thesis is concerned with cost-efficient density functional theory (DFT) and Hartree–Fock (HF) based electronic structure methods for noncovalent interactions, which are about a factor of 50 to 100 faster than calculations in a large basis set. The main errors in a DFT or HF calculation with small atomic orbital basis sets are the missing London dispersion and the basis set superposition error (BSSE). An exemplary benchmark study shows that modern correction strategies clearly outperform plain DFT or HF for energies and geometries of small dimers, large supramolecular complexes, and molecular crystals. Further, the development and evaluation of a minimal basis set Hartree–Fock method with three atom-pairwise corrections for London dispersion, BSSE, and basis set incompleteness (HF-3c) is presented. With nine global parameters, the empiricism of HF-3c is moderate, the method is self-interaction error free, and noiseless analytical frequencies can be obtained. HF-3c provides accurate geometries of organic supramolecular systems and small proteins, and good noncovalent interaction energies. The mean absolute deviations (MADs) for the S22 set of small noncovalently bound dimers and the S12L set of supramolecular host-guest association energies are 0.6 and 4.4 kcal mol^{-1}, respectively. This is excellent compared to dispersion corrected DFT methods whose MADs are in the range of 0.3–0.5 and 2–5 kcal mol^{-1}, respectively.

The second part focuses on the application and evaluation of multilevel methodologies for an accurate description of Gibbs free energies of association (ΔG_a) for supramolecular host-guest complexes in solution. First, state-of-the-art dispersion corrected DFT (DFT-D3^{ATM}) is used together with a large quadruple-zeta (QZ) basis set to obtain association energies in the gas phase. A semiempirical method is utilized to compute the thermostatistical corrections from energy to free energy and last, a continuum solvation

model is employed. The general procedure is illustrated with a case study on eight typical complexes. The SAMPL4 blind test challenge provides a unique opportunity to test this methodology in a realistic setting. Relative ΔG_a in water are predicted for a cucurbit[7]uril host and 14 guest molecules containing ammonia groups. The HF-3c method was employed to sample possible binding conformations and the final ΔG_a were calculated on the PW6B95-D3$^{\mathrm{ATM}}$/QZ level with HF-3c thermal corrections and COSMO-RS solvation contributions. Compared to other methods theses predictions rank in the top three of all statistical measurements. The MAD and RMSD are only 2.0 and 2.6 kcal mol^{-1}, respectively. Further, the S30L benchmark set is proposed as an extension of the S12L set for association (free) energies of host-guest complexes. Larger systems with up to 200 atoms, more divers interaction motifs, and higher charges are represented by experimentally measured complexes with ΔG_a values in the range from -0.7 to -24.7 kcal mol^{-1}. In order to obtain a theoretical best estimate for ΔG_a different dispersion corrected density functionals, semiempirical methods, and continuum solvation models are tested. The best method combination is similar to the one used for the SAMPL4 bind test and yields an MAD with respect to experiment of only 2.4 kcal mol^{-1}. Inclusion of counterions for the charged systems (S30L-CI) were found to improve the results overall.

Synergy between theory and experiment is demonstrated in the last part of this thesis with the application of quantum chemical methods to two specific chemical problems related to supramolecular chemistry. Experimentally, it was found that titanocene(III) catalysts can be stabilized by chloride additives and the calculations reveal that the stabilities of these adducts are determined by the extent of hydrogen bonding between the catalyst and the ammonium cation. 1,1'-Binaphthol based ligands can be used to obtain enantiomerically pure double- and triple-stranded helicates with transition-metal ions in a completely diastereoselective self-assembly process. Electronic circular dichroism spectra of precursors for paracyclophane based ligands have been investigated computationally in order to identify their absolute configuration.

Zusammenfassung

Diese Dissertation befasst sich mit der Anwendung und Entwicklung von Elektronenstrukturmethoden zur Beschreibung nichtkovalenter Wechselwirkungen im Allgemeinen und der Evaluierung von Multilevelmethodiken für die Thermochemie supramolekularer Systeme im Speziellen. Nichtkovalente Wechselwirkungen sind allgegenwärtig in Systemen unterschiedlichster Fachgebiete, wie supramolekulare Chemie, Strukturbiologie und Oberflächenforschung. Auf Grund ihrer vielfältigen Einsatzmöglichkeiten, u.a. in der molekularen Erkennung und der Selbstassemblierung, spielen Wirt-Gast Komplexe in der supramolekularen Chemie eine besondere Rolle.

Da die Bindungssituation in einem solchen supramolekularen Komplex oft nicht bekannt ist, ist ein Testen verschiedener Konformationen wünschenswert. Daher beschäftigt sich der erste Teil der Arbeit mit kostengünstigen Dichtefunktionaltheorie (DFT) und Hartree–Fock (HF) basierten Methoden, die nichtkovalente Wechselwirkungen gut beschreiben und 50 bis 100 mal schneller sind als Rechnungen in einem großen Basissatz. Die beiden gravierendsten Fehler in einer DFT oder HF Rechnung mit kleinem Atomorbital-Basissatz sind die fehlende London Dispersion und der Basissatzsuperpositionsfehler (BSSE). Eine exemplarische Studie zeigt, dass moderne Korrekturstrategien reine DFT und HF hinsichtlich ihrer Genauigkeit für Energien und Geometrien kleiner Dimere, großer Supramoleküle und molekularer Kristalle deutlich übertreffen. Weiterhin wird die Entwicklung und Evaluierung einer HF Methode mit minimalen Basissatz gezeigt (HF-3c). Diese nutzt drei Korrekturen zur Berücksichtug der London Dispersion, des BSSE und des Basissatzunvollständigkeitsfehler. Der Empirismus von HF-3c ist mit neun globalen Parametern moderat, die Methode ist selbstwechselwirkungsfehlerfrei und rauschfreie analytische Frequenzen sind zugänglich. HF-3c liefert genaue Geometrien für organische Supramoleküle und kleine Proteine sowie gute Energien nichtkovalenter Wechselwirkungen. Die gemittelte absolute Abweichung (MAD) für den S22 Testsatz kleiner nichtkovalent gebundener Dimere und den S12L Testsatz supramolekularer Wirt-Gast Komplexe sind 0.6 und 4.4 kcal mol^{-1}. Dieses Ergebnis ist ausgezeichnet, verglichen mit modernen dispersionskorrigierten DFT Methoden, welche MADs von 0.3–0.5 und 2–5 kcal mol^{-1} liefern.

Der zweite Teil der Dissertation befasst sich mit der Anwendung und Evaluierung von Multilevelmethodiken zur genauen Beschreibung der freien Enthalpie (Gibbs-Energie) der Assoziation (ΔG_a) supramolekularer Wirt-Gast Komplexe in Lösung. Die Assoziationsenergie in der Gasphase wird über moderne dispersionskorrigierte DFT (DFT-D3$^{\text{ATM}}$)

Rechnungen in einem quadrupel-zeta (QZ) Basissatz erhalten, semiempirische Methoden liefern die thermostatistischen Korrekturen zur freien Enthalpie und Solvatationseffekte werden mit Hilfe von Kontinuumsmodellen berücksichtigt. Das Protokoll wird anhand einer Fallstudie von acht typischen Systemen dargelegt. Die Teilnahme am SAMPL4 Blindtest bietet die ideale Möglichkeit die Methodik unter realen Bedingungen zu prüfen. Für einen Cucurbit[7]uril Wirt und 14 Gast Moleküle mit Ammoniumgruppen werden relative ΔG_a Werte vorhergesagt. HF-3c wurde zum Testen verschiedener Bindungskonformationen genutzt und die Methodenkombination aus PW6B65-D3$^{\mathrm{ATM}}$/QZ Energien, HF-3c thermostatistischen Korrekturen und COSMO-RS freien Solvatationsenergien liefert ΔG_a, welche unter den Top drei aller eingereichten Methoden für alle Statistikmaße rangiert. Der MAD and der RMSD sind nur 2.0 and 2.6 kcal mol^{-1}. Weiterhin wurde der S12L Testsatz auf 30 Komplexe erweitert (S30L), sodass größere Systeme mit bis zu 200 Atomen, ein breiteres Spektrum an Wechselwirkungsmotiven und höhere Ladungen (-1 bis +4) enthalten sind. Zu allen Komplexen liegen experimentelle Daten vor und die ΔG_a Werte liegen im Bereich von -0.7 bis -24.7 kcal mol^{-1}. Um die beste Methodenkombination zu finden, wurden verschiedene dispersionskorrigierte Dichtefunktionale, mehrere semiempirische Methoden und unterschiedliche Solvatationsmodelle getestet. Die beste Kombination ist ähnlich zu der bereits im SMAPL4 Blindtest verwendeten und ergibt einen MAD von 2.4 kcal mol^{-1} verglichen mit dem Experiment. Berücksichtigung von Gegenionen für die geladenen Systeme verbessert die Ergebnisse im Mittel leicht.

Das Zusammenwirken von Theorie und Experiment wird im letzten Teil der Arbeit anhand zweier Kooperationsprojekte mit Bezug zur supramolekularen Chemie dargestellt. Experimente zeigen, dass Titanocen(III)-Katalysatoren durch Chlorid-Additive stabilisiert werden können. Die Berechnungen erklären die Stärke dieser Stabilisierung mit dem Ausmaß der Wasserstoffbrückenbindungen. Chirale Liganden basierend auf 1,1'-Binaphtyl können über einen Selbstassemblierungsprozess mit Übergangsmetallionen zwei-oder dreisträngige enantiomerenreine Helicate bilden. Die absolute Konfiguration von Vorläufern Paracyclophan-basierter Liganden wurde mithilfe von theoretischen Circulardichroismus-Spektren aufgeklärt.

Contents

Part I.

Introduction and theoretical background

1. Introduction

Traditional chemistry focuses on covalent bonds and new substances are obtained via synthetic routes that involve cleavage and formation of those bonds. Supramolecular chemistry as defined by Lehn is the *"chemistry beyond the molecule"*,[1,2] which implies that not only one single molecule but a number of assembled molecules is regarded. To emphasize the difference to the interactions responsible for forming or breaking covalent bonds within a molecule, the intermolecular forces governing the assembly or disassembly of supramolecules are termed noncovalent interactions.[3,4]

The fundaments of supramolecular chemistry date back to the late 19$^{\text{th}}$ century, when some of the most basic concepts were already introduced. In 1891, Villiers and Hebd discovered the cyclodextrins, the first host molecules.[5] Werner formulated the idea of coordination chemistry in 1893,[6] and in 1894 Fischer introduced the lock-and-key concept,[7] which already assumes molecular recognition and host-guest chemistry. The static picture of the lock-and-key concept is reasonable as long as the binding molecules remain in their initial pre-binding conformation. In the beginning of the 20$^{\text{th}}$ century, noncovalent bonds were gradually understood in more detail, e.g. through the description of hydrogen bonding by Latimer and Rodebush in 1920.[8] 1958, Koshland refined the lock-and-key idea and formulated the induced fit concept for biomolecules which undergo a conformational change during the binding event.[9] This provided a more dynamic view of the often significant structural reorganization. An important breakthrough was the elucidation of the double helix structure of DNA, which is stabilized by hydrogen bonds connecting the base pairs (see Figure 1.1 (a)) and stacking interactions between those pairs.[10] Watson, Crick, and Wilkins were honored for their discovery with the Nobel Prize in Physiology or Medicine in 1962.[11] Gradually, chemists were able to transfer the introduced concepts to synthetic systems, which finally lead to the discovery of crown ethers by Pedersen in 1967 (see Figure 1.1 (b)).[12,13] Inspired by this work, Lehn and Cram focused their own research on supramolecular chemistry. Together, the three pioneers of this field were awarded with the Nobel Prize in Chemistry in 1987 *"for their development and use of molecules with structure-specific interactions of high selectivity"*.[14]

Today, supramolecular chemistry is a highly interdisciplinary field that bridges areas like biochemistry, nanochemistry, and material sciences.[16,17] Nevertheless, it is still tightly connected to host-guest chemistry, which is also the focus of this thesis. A host molecule favorably binds a guest molecule and thus, the two molecules form a stable complex

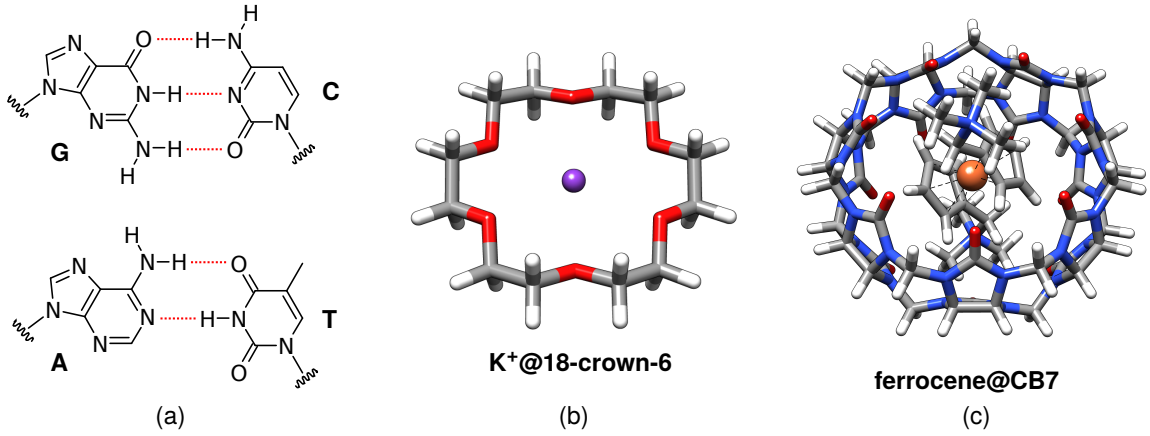


Figure 1.1.: (a) Hydrogen bonded DNA base pairs: Guanine-cytosine (**G-C**) and adenine-thymine (**A-T**). (b) Structure of a crown ether alkali metal ion complex.[12] (c) Structure of a host-guest complex.[15]

(see Figure 1.1 (c) for a typical example). If the binding of a certain guest molecule is specific, the process is called molecular recognition.[16–18] Many different hosts with functional shapes, e.g. bowls, cages, pincers, tweezers, container, and capsules have been designed in the past years.[16,17,19–21] Besides molecular recognition, host-guest complexes are used in the fields of template-directed synthesis, biomimetics, and even as reaction containers.[1,2,13,16,17,22]

The formation and stability of supramolecular architectures in general and host-guest complexes in particular is determined by the intermolecular interactions. These non-covalent interactions are based on electrostatics, induction and charge transfer effects, and London dispersion.[3,4,23,24] For their understanding the chemical concepts of ion-ion and ion-dipole interactions, hydrogen and halogen bonding, $\pi - \pi$ stacking, anion- and cation-π interactions, and van der Waals interactions can be employed. Within this thesis, the terms of the chemical concepts and the underlying physical principles are not used separately but rather combined to complement each other. Note, that all these types of noncovalent interactions can of course also occur within one single molecule of certain size. Thus, noncovalent interactions are omnipresent and control the structures of DNA and proteins, antigen-antibody recognition, host-guest and enzyme-substrate binding, self-assembly processes, and the orientation of molecules on surfaces or in molecular crystals.[25–29]

Although supramolecular chemistry is mainly an experimental discipline, computational methods have emerged as invaluable tools for analyzing, characterizing, and understanding of supramolecular systems.[30] Theoretical methods with reliable predictive power could partly replace tedious and costly experimental work and help in the design of new systems. Therefore, a central aspect of this thesis is the evaluation of computational methodologies

to accurately describe supramolecular host-guest thermochemistry of realistic systems. As typical complexes consist of a few hundred atoms or more, the method of choice has to have a low computational cost but be accurate enough at the same time. Kohn-Sham density functional theory (DFT) provides a good ratio of cost and accuracy, for a more detailed description of DFT see Chapter 2.2. It is able to treat chemically relevant systems and yields quantitative relative energies of many reactions and accurate geometries of various molecules. The success of DFT can partly be attributed to the fact that electron correlation at short and intermediate distances is taken into account in a very efficient manner. However, a fundamental problem of common local, semi-local and hybrid density functional approximations is their incapability to describe long-range correlation, namely London dispersion.[23,24,31] Since the common functionals are based on the exponentially decaying electron density, they fail to provide the asymptotically correct $1/R^6$ dependence of the London dispersion energy on the inter-atomic distance R. Nevertheless, for all noncovalently bound systems the dispersion energy is crucial and cannot be neglected.[3,4] In the past years, various concepts to correct this error have been developed.[32–35] The work presented in this thesis mainly uses the dispersion-corrected DFT scheme that was developed in the Grimme group over the last years (DFT-D3).[36,37] Within the D3 scheme the dispersion energy is calculated as a sum of atom-pairwise contributions $-C_6/R^6$, where the C_6 are the interatomic dispersion coefficients. A more detailed description of this semi-classical correction scheme can be found in Chapter 2.4. Nevertheless, even with highly accurate association energies, DFT calculations are still limited to the gas phase.

Solvation is a crucial aspect in supramolecular chemistry and thus, cannot be ignored.[17] Solvent molecules can be competitors to the guest molecules, and their polarity and size are adjusted in order to tune a binding constant.[17] Within a calculation, a solvent can be included either explicitly or implicitly. As the regarded systems within this thesis are usually too large to be treated routinely in an environment of hundreds of explicit solvent molecules, simpler continuum solvation models will be used. Common continuum solvation models like the polarizable continuum model (PCM)[38,39] or the conductor like screening model (COSMO)[40] cannot describe the statistical entropy effects of the liquid phase. The conductor like screening model for real solvents (COSMO-RS)[41,42] is capable of including those effects by using statistical mechanics. One part of this work is concerned with the evaluation of the accuracy of COSMO-RS for host-guest systems. More on the theoretical background of COSMO and COSMO-RS can be found in Chapter 2.6.

In order to compute the association free energy ΔG_a of a host-guest system in solution, the thermodynamic cycle shown in Figure 1.2 is used. First, the geometries of all compounds are optimized on a feasible level of theory. Usually, TPSS-D3 with a medium sized atomic orbital basis set is employed. As only one single conformation is taken into account, this approach is static. The association energy ΔE in the gas phase is

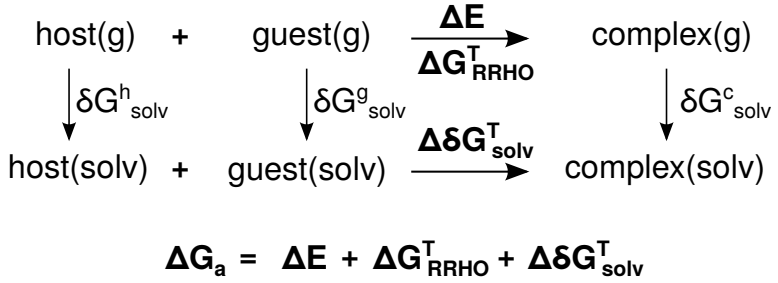


Figure 1.2.: Thermodynamic cycle used to compute the association free energy ΔG_a of a host-guest complex in solution

computed with DFT-D3$^{\mathrm{ATM}}$ using a hybrid functional including the three-body contributions to the D3 dispersion energy and a large basis set in a supermolecular approach, i.e. $\Delta E = E(\mathrm{complex}) - E(\mathrm{host}) - E(\mathrm{guest})$. The thermostatistical corrections from energy to free energy ΔG_{RRHO} in the gas phase (see Chapter 2.5 for details) are obtained mostly with semiempirical methods due the greatly reduced computational cost. For an overview of the semiempirical methods employed, see Chapter 2.3. And finally, the solvation free energy for each gas phase species δG_{solv} is calculated with COSMO-RS. This protocol has been used before for supramolecules in general[43–45] and host-guest systems in particular. 12 typical textbook examples of host-guest complexes (S12L) have been compiled by Grimme and treated with this multilevel methodology.[46] The ΔG_a values on the PW6B95-D3$^{\mathrm{ATM}}$/QZ level of theory including DFTB3-D3 thermostatistical corrections and COSMO-RS 2012 solvation free energies, have an overall mean absolute deviation (MAD) of 2.1 kcal mol^{-1} compared to experiment.[46] This thesis further evaluates and applies the multilevel approach to realistic host-guest systems with the aim to show its predictive power.

After this introduction and the theoretical background, the thesis is divided into three major parts. Part II focuses on the evaluation and development of cost-efficient DFT and Hartree–Fock (HF) based electronic structure methods for noncovalent interactions. As the binding situation in a supramolecular complex is often unknown, sampling many possible conformations is desired. This cannot be afforded with a high level method like DFT-D3$^{\mathrm{ATM}}$/QZ for calculating the gas phase association energies. Therefore, computationally less demanding approaches with still reasonable accuracies are needed. Reducing the number of atomic orbital basis functions to a double-zeta or minimal basis set yields an increase in speed by a factor of 50 to 100 compared to a quadruple-zeta basis set calculation. Unfortunately, two major errors arise in a plain semi-local DFT or HF calculation with a small basis set. These are the missing London dispersion and the basis set superposition error (BSSE) which are analyzed in Chapter 3. Modern correction strategies to

circumvent them are presented and reviewed. An exemplary benchmark study is carried out to compare the accuracy of those correction schemes and to evaluate the improvement over plain DFT and HF. The performance is tested for interaction energies and geometries of noncovalently bound dimers, host-guest complexes, and organic molecular crystals.
The development and evaluation of a minimal basis set Hartree–Fock method with three atom-pairwise corrections (HF-3c) is presented in Chapter 4. The D3 scheme is employed to include London dispersion, a geometric counter-poise correction (gCP)[47] is used to account for BSSE, and a third short-range term is applied to correct for basis sets incompleteness effects. This results in nine global parameters and thus, moderate empiricism. Compared to a DFT approach with a small basis set, HF has the advantages of being self-interaction error free, providing noiseless analytical frequencies and it does not require the construction of a grid for integrating the exchange-correlation functional. The accuracy of HF-3c is evaluated for the geometries of small covalently bound organic molecules, geometries and interaction energies of small noncovalent dimers and large supramolecular host-guest systems, geometries of small proteins, and thermostatistical corrections. In terms of cost and accuracy HF-3c might fill the gap between conventional semiempirical and DFT methods.

Part III focuses on the application and evaluation of the multilevel methodology described above, to obtain free energies of association for supramolecular host-guest complexes in solution. In Chapter 5 the general procedure is reviewed, and an overview of recent calculations on supramolecular systems that were carried out before in our group and by others is provided. To illustrate the methodology and its general applicability, a case study on eight typical host-guest complexes is carried out. Two tweezer complexes, two pseudorotaxanes, two cucurbit[7]uril complexes, and two complexes of the fullerene C_{70} in cycloparaphenylenes are chosen.
Chapter 6 focuses on the participation in a blind test challenge. Blind tests provide the unique opportunity to test a computational method in a realistic setting without a bias towards known answers. The Statistical Assessment of Modeling of Proteins and Ligands (SAMPL) aims at testing protein, supramolecular, and small molecule modeling.[48] In the 4^{th} SAMPL blind test (SAMPL4), (relative) association free energies of two different host and several guest molecules in water are to be predicted.[49] The first host is the rigid cucurbit[7]uril and the 14 guest molecules contain either one or two ammonia groups. The cucurbit[n]uril (CBn, $n = 5, 6, 7, 8$) family of molecular containers is one of the major tools to study molecular recognition in water.[50–52] Usually, guest molecules with a rigid hydrophobic core such as ferrocene[53] or adamantane[15,50] in combination with cationic ammonium groups yield very high binding affinities. The participation in the blind test challenge is carried out in cooperation with Dr. Jens Antony, who focuses on the second host system which is a more flexible, eightfold charged basket shaped octa-acid molecule.

The HF-3c method is applied to sample possible binding conformations. Counterions are introduced and the geometries are re-optimized on the TPSS-D3/TZ level of theory. The final ΔG_a values are computed on the PW6B95-D3$^{\mathrm{ATM}}$/QZ level with HF-3c thermal corrections and COSMO-RS 2012 solvation contributions. The results are submitted to the SAMPL4 organizing committee, prior to publication of the experimental results, and compared to the experimentally obtained values and to those computed by other participants.
In the case study as well as in the SAMPL4 blind test participation outliers for ΔG_a values compared to experiment are present. Often, but not always these systems are charged. Thus, it is evident that further testing of the multilevel approach is necessary and the extension of the S12L test set to 30 host-guest systems is described in Chapter 7. An extensive literature search is carried out in order to find interesting systems for which measured free energies of association exist. The aim is to include larger and slightly more flexible systems, higher charges, more diverse interaction motifs, and different solvents. In order to obtain a theoretical best estimate for ΔG_a, different dispersion-corrected density functionals (ΔE) are used in combination with several semiempirical methods (ΔG_{RRHO}) and various COSMO-RS parametrizations, as well as a density dependent continuum solvent model (SMD)[54] ($\Delta\delta G_{solv}$). The influence of counterions, i.e. chloride for cations and sodium for anions, is investigated.

Part IV contains two projects that are carried out in collaboration with experimentally working groups. Experiment and theory are often complementary and thus, should ideally be combined. Theory can give insight into experimental findings and make suggestions for new compounds, and in return, experimental results and newly synthesized systems can trigger the development of novel theoretical methods. Chapter 8 presents a computational study on supramolecular interactions of titanocene(III)-chloride that is realized in a collaboration with the group of Prof. Gansäuer. Titanocene(III) catalysts are employed in atom-economical catalytic radical reactions to synthesize tetrahydrofuran[55–58] and arylate epoxides.[59] It was found that thermal instability of the most efficient electron-deficient catalysts could be circumvented by the addition of a chloride source. Thus, the stability and geometry of the formed hydrochloride-titanocene adducts is investigated computationally in order to explain this finding.
The focus of Chapter 9 is the computational study of 4,15-difunctionalized [2.2]paracyclophanes in collaboration with the group of Prof. Lützen. It was shown, that chiral 1,1'-binaphthol based ligands can be used to obtain enantiomerically pure double- and triple-stranded helicates with transition-metal ions[60–63] or even larger polyhedral structures.[64] The underlying process is a completely diastereoselective self-assembly. Similar to binaphtyl, paracyclophanes are chiral and can be used as precursors for analogous ligands.[65] In order to identify the absolute configuration of the difunctionalized

[2.2]paracyclophanes, electronic circular dichroism spectra are calculated and compared to experiment.

2. Theoretical background

The following chapters provide an overview of the theoretical background of the methods applied within this thesis. The main focus lies on electronic ground state theory. Starting with the electronic Hamiltonian, the Hartree–Fock approximation is explained in Chapter 2.1 and the concept of electron correlation energy is defined. Chapter 2.2 gives an introduction to density functional theory and an overview of the classes of density functionals. Simplifications to the Hartree–Fock and DFT methods resulting in several semiempirical methods are reviewed in Chapter 2.3. A semi-classical scheme to account for the London dispersion interactions is introduced in Chapter 2.4. The necessary steps from energy to free energy are discussed in Chapter 2.5 and models to take solvation effects into account are introduced in Chapter 2.6.

2.1. The Hartree–Fock approximation

The time-independent, non-relativistic Schrödinger equation $\hat{H}\Psi = E\Psi$ is the starting point in a wave function based, quantum-mechanical treatment of atoms and molecules.[66] Within the Born–Oppenheimer (BO) approximation, the comparably slow motion of the nuclei is separated from that of the fast moving electrons.[67] This leads to the electronic Schrödinger equation for a fixed nuclear conformation in atomic units:

$$\hat{H}_e\Psi_e = \left[\sum_i^N \hat{h}_i + \hat{V}_{ee}\right] = E_e\Psi_e \tag{2.1}$$

$$\hat{h}_i = -\frac{1}{2}\hat{\nabla}_i^2 - \sum_A^M \frac{Z_a}{r_{iA}}\,, \quad \hat{V}_{ee} = \sum_i^{N-1}\sum_{j>i}^{N}\frac{1}{r_{ij}} \tag{2.2}$$

The electronic Hamiltonian $\hat{H}_e$ consists of two parts: The one-electron part $\hat{h}_i$ describes the kinetic energy of the ith electron and its Coulomb interaction with the fixed nuclei, and the two-electron part $\hat{V}_{ee}$ is the Coulomb repulsion between electron pairs ij. The operator $\hat{\nabla}_i^2$ yields the second derivative of the electronic wave function Ψ_e with respect to the coordinates of electron i (kinetic energy), r_{iA} and r_{ij} are the distances between electron i and nucleus A or another electron j, and Z_A is the charge of nucleus A.
The electronic energy E_e is the expectation value of $\hat{H}_e$ with respect to the normalized

electronic wave function Ψ_e:

$$E_e = \langle \Psi_e | \hat{H}_e | \Psi_e \rangle \tag{2.3}$$

In order to obtain the total energy E_{tot} of a given system, the potential energy of the nuclei V_{nn}, which is a constant for the fixed, classically treated nuclei in the BO approximation, has to be added to E_e.

$$E_{tot} = E_e + V_{nn} = E_e + \sum_{A}^{M-1} \sum_{B>A}^{M} \frac{Z_a Z_B}{r_{AB}} \tag{2.4}$$

If not mentioned otherwise, electronic energies will be considered within this chapter and the index e is dropped from here on.

The wave function of an electronic ground state of a system with N electrons can often be approximated with good accuracy by a single Slater Determinant Φ_0.[68]

$$\Phi_0 = \frac{1}{\sqrt{N!}} \sum_{n}^{N!} (-1)^{p_n} \hat{P}_n \prod_{i}^{n} \phi_i \tag{2.5}$$

Φ_0 consists of products of independent one-electron wave functions ϕ_i, so called molecular orbitals (MOs). The permutation operator $\hat{P}_n$ exchanges the electrons between orbitals, p_n is the number of transpositions necessary to achieve the nth permutation, and $1/\sqrt{N!}$ is the normalization factor. By construction, the Slater determinant fulfills the Pauli principle, that is the anti-symmetry of the wave function with respect to the interchange of two electrons. Within this thesis, only those systems will be investigated for which a single Slater determinant is a good approximation for the electronic ground state.

The variational principle states that the energy expectation value of any trial wave function $\tilde{\Phi}$ cannot be lower than the energy of the true wave function Ψ.

$$\tilde{E} = \langle \tilde{\Phi} | \hat{H} | \tilde{\Phi} \rangle \geq \langle \Psi | \hat{H} | \Psi \rangle = E \tag{2.6}$$

The most important procedure to minimize the energy expectation value of a single Slater determinant Φ_0 is probably the Hartree–Fock (HF) approximation.[69–71] It forms the basis for a large number of more accurate and sophisticated wave function based methods and thus, is fundamental for the whole of quantum chemistry. In HF theory, Ansatz 2.5 is used for the wave function and the energy is minimized with respect to the orbitals ϕ_i. The method of Lagrange multiplier with the constraint of orthonormal orbitals finally

leads to the HF eigenvalue problem.

$$\hat{f}_i \phi_i = \varepsilon_i \phi_i \tag{2.7}$$

$$\hat{f}_i = \hat{h}_i + \hat{v}_i^{HF} \ , \quad \hat{v}_i^{HF} = \sum_j^N \hat{J}_{ij} - \hat{K}_{ij} \tag{2.8}$$

Hence, the MOs ϕ_i are eigenfunctions of an effective one-electron operator, i.e. the Fock-operator $\hat{f}_i$. In the HF method, the exact electron-electron Coulomb potential is approximated by the interaction of an electron i with the averaged field $\hat{v}_i^{HF}$ of all other electrons. Therefore, Hartree–Fock is a mean-field theory. $\hat{v}_i^{HF}$ consists of the Coulomb operator $\hat{J}_{ij}$ and the exchange operator $\hat{K}_{ij}$.

$$\hat{J}_{ij}|\phi_i\rangle = \langle\phi_j|\frac{1}{r_{ij}}|\phi_j\rangle|\phi_i\rangle \tag{2.9}$$

$$\hat{K}_{ij}|\phi_i\rangle = \langle\phi_j|\frac{1}{r_{ij}}|\phi_i\rangle|\phi_j\rangle \tag{2.10}$$

$\hat{J}_{ij}$ incorporates the Coulomb repulsion between all electrons and $\hat{K}_{ij}$ originates from the anti-symmetry of the wave function leading to an additional effective interaction between electrons with the same spin. As $\langle\phi_i|\hat{J}_{ii}|\phi_i\rangle$ equals $\langle\phi_i|\hat{K}_{ii}|\phi_i\rangle$, there is no unphysical interaction of an electron with itself. Thus, the HF method is free of the self-interaction error (SIE).

The MOs can be expanded in a linear combination of known basis functions, usually atomic orbitals ψ (AOs, LCAO ansatz),[72,73]

$$\phi_i = \sum_\mu C_{\mu i} \psi_\mu \tag{2.11}$$

where the $C_{\mu i}$ are the LCAO-MO coefficients. This approach is exact if the atomic orbitals form a complete basis. For a finite basis set, it leads to the so called basis set incompleteness error (BSIE). Within the LCAO ansatz, the eigenvalue problem 2.7 can be reformulated as a linear matrix equation (Roothan–Hall equation).[72,73]

$$\mathbf{FC} = \mathbf{SC}\varepsilon \tag{2.12}$$

The Fock matrix $\mathbf{F}$ contains elements of the form $\langle\phi_i|\hat{f}_i|\phi_j\rangle$, and $\mathbf{C}$ consists of the LCAO-MO coefficients. $\mathbf{S}$ is the overlap matrix with the elements $S_{ij} = \langle\phi_i|\phi_j\rangle$, as the AOs are usually not orthogonal in the molecular system. Solving Equation 2.12 yields the energies (eigenvalues) ε of the orbitals. The Fock operator depends on the orbitals and thus on the MO coefficients (c.f. Eqs. 2.8, 2.9, 2.10). Consequently, the solution has to be carried out

iteratively which is done by a self-consistent field (SCF) procedure via linear variation of the MO coefficients.

Once the self-consistent orbitals are obtained, the HF energy can be calculated.

$$E^{HF} = \sum_{i}^{N} \langle \phi_i | \hat{h}_i | \phi_i \rangle + \sum_{i}^{N-1} \sum_{j>i}^{N} \left(\langle \phi_i | \hat{J}_{ij} | \phi_i \rangle - \langle \phi_i | \hat{K}_{ij} | \phi_i \rangle \right) \tag{2.13}$$

The computational effort of HF formally scales with $\mathcal{O}(N^4)$, where N is the number of AOs used. Speed-ups can be achieved e.g. through the resolution of the identity (RI) approximation, in which the four-center two-electron integrals are approximated by three-center ones.[74]

The HF approximation includes the correlated motion of two electrons with parallel spin that arises due to the Pauli principle (Fermi correlation), but the correlated motion of two electrons that originates from their mutual Coulomb repulsion (Coulomb correlation) is neglected. The reason for the missing Coulomb correlation is the approximation that each electron only experiences an averaged potential of the other electrons. Thus, even in the limit of a complete basis set, HF cannot yield the exact energy of the non-relativistic Schrödinger equation within the BO approximation for a system with more than one electron. The difference between the exact energy E and the HF limit E^{HF} is defined as electron correlation energy E_{corr}.

$$E_{corr} = E - E^{HF} \tag{2.14}$$

E_{corr} can be computed by electron correlated methods, e.g. by perturbative methods like the Møller–Plesset (MP) theory, configuration interaction (CI) and coupled cluster (CC) methods, which are all based on the single-determinant wave function obtained from HF. Having a formal scaling of $\mathcal{O}(N^5)$ with the system size or higher, these methods are computationally much more demanding. However, for an accurate treatment of chemical systems, correlation cannot be neglected. Coupled cluster theory that includes single and double excitations and treats the triple excitations in an approximated, perturbative way (CCSD(T)) has emerged as 'the gold standard' of quantum chemistry and is often used to calculate reference values for benchmarking simpler methods. Recent developments in local coupled-cluster methods allow the treatment of molecules in the regime of about 100 atoms.[75–78] Nevertheless, the convergence of the energy with the basis set size is inverse cubic, which is rather slow compared to HF that converges exponentially. As the systems regarded in this thesis are generally too large to be naturally treated with these kinds of correlated methods, they will not be discussed further.

2.2. Density functional theory

The idea to use the electron density ρ instead of a wave function dates back to Thomas and Fermi. In 1927 they independently assumed that the ground state of a system can be obtained through an energy functional of the density.[79,80] The existence of such an energy functional was proven by Hohenberg and Kohn in 1964, who showed that the energy of the electronic ground state is completely determined by its density.[81] Thus, according to the first Hohenberg–Kohn theorem, there has to be a density functional (DF) $E[\rho]$ that directly connects the ground state electron density to the exact energy. The second Hohenberg–Kohn theorem is the analogue to the variation principle in wave function theory (WFT), and shows that for any valid trial density $\tilde{\rho}$, $E[\tilde{\rho}]$ yields the upper bound to the true ground state energy.[82,83] However, the Hohenberg–Kohn theorems do not provide a construction formalism for the shape of this functional.

As the exact density functional is not known, finding a good approximation to $E[\rho]$ is the goal of density functional theory (DFT) methods. $E[\rho]$ can conveniently be divided into separate functionals that include different contributions.

$$E[\rho] = T_e[\rho] + V_{en}[\rho] + V_{ee}[\rho] \tag{2.15}$$

$$V_{ee}[\rho] = J[\rho] + K[\rho] \tag{2.16}$$

$T_e[\rho]$ gives the kinetic energy, $V_{en}[\rho]$ yields the Coulomb attraction between electrons and nuclei and $V_{ee}[\rho]$ describes the electron-electron interaction, that can be split further into a Coulomb part $J[\rho]$ and an exchange part $K[\rho]$. Up to this point, the formulation of DFT is orbital-free. The advantage compared to wave function based theories is the dependence of the density on only three variables, i.e. the three Cartesian coordinates, instead of three variables per electron. The functionals $V_{en}[\rho]$ and $J[\rho]$ can be described by their classical expressions.

$$V_{en}[\rho] = -\sum_{A}^{M} \int \frac{Z_A \rho(\mathbf{r})}{|\mathbf{R}_A - \mathbf{r}|} \mathrm{d}\mathbf{r} \tag{2.17}$$

$$J[\rho] = \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \tag{2.18}$$

Expressions for $T_e[\rho]$ and $K[\rho]$ were first approximated by Thomas and Fermi and Dirac,

respectively, based on the uniform electron gas (UEG).

$$T_e^{UEG}[\rho] = \frac{3}{10}(3\pi)^{\frac{2}{3}} \int \rho(\mathbf{r})^{\frac{5}{3}} \mathrm{d}\mathbf{r} \tag{2.19}$$

$$K^D[\rho] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho(\mathbf{r})^{\frac{4}{3}} \mathrm{d}\mathbf{r} \tag{2.20}$$

These early attempts to approximate the true functional do not yield useful results for chemical problems, as chemical bonding cannot be described. The main reason for this is the inaccurate expression of the kinetic energy in the Thomas–Fermi model. Kohn and Sham introduced the calculation of the kinetic energy via a fictitious reference system of non-interacting quasi-particles which is supposed to have the same density as the true system.[84] This Kohn–Sham (KS) approach to DFT (KS-DFT) is nowadays the most common one and therefore, the KS prefix will be dropped. The drawback is that orbitals (KS-orbitals) have to be introduced in order to evaluate the kinetic energy T_{KS}. Thus, the number of variables grows to $3N$ as in wave function theory. Usually, this approach yields 98 to 99 % of the true kinetic energy. The missing difference in the kinetic energy for the independent compared to the correlated electrons, as well as the overall correlation and the exchange effects are described by the exchange-correlation functional E_{XC}. E_{XC} is usually divided into an exchange E_X and a correlation part E_C. The sum of all theses contributions is the total DFT energy.

$$E^{DFT} = T_{KS}[\phi] + V_{en}[\rho] + J[\rho] + E_{XC}[\rho] \tag{2.21}$$

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho] \tag{2.22}$$

$$T_{KS}[\phi] = -\frac{1}{2}\sum_i^N \langle \phi_i | \hat{\nabla}_i^2 | \phi_i \rangle \tag{2.23}$$

The electronic energy and the respective set of KS-orbitals are obtained iteratively by solving the KS-equations.

$$\hat{f}_i^{KS}[\rho]\phi_i = \left[\hat{h}_i[\rho] + \sum_j \left(\hat{J}_{ij}[\rho] + v_{XC}[\rho]\right)\right]\phi_i = \varepsilon_i \phi_i \tag{2.24}$$

Analogous to the Fock-operator in HF, the Kohn–Sham operator $\hat{f}_i^{KS}[\rho]$ is an effective one-electron operator. But instead of the exchange operator in HF, DFT uses an exchange-correlation potential $v_{XC}[\rho]$ that is the derivative of the $E_{XC}[\rho]$ functional with respect to the density.

The similarity of DFT and HF is also reflected in the same computational effort. Reducing the formal scaling of $\mathcal{O}(N^4)$ to $\mathcal{O}(N^3)$ can be achieved by using the efficient resolution

of the identity approximation for the Coulomb integrals (RI-J),[74] The major advantage of DFT over HF is the inclusion of correlation effects when a sufficiently accurate approximation for $E_{XC}[\rho]$ is applied. While DFT is formally exact, the commonly used density functional approximations (DFAs) are not. A drawback of DFT compared to WFT is, that it cannot be systematically improved, e.g. by increasing the number of excitation configurations as in coupled cluster theory. Therefore, the development of DFAs is often based on a trial and error approach, i.e. the accuracy of the functional is evaluated by testing it on typical, exemplary systems. As only a small part of the chemical space can be covered, uncertainties of the accuracy remain.

2.2.1. Hierarchy of density functional approximations

Although the systematic improvement of density functionals is difficult, the accuracy and computational effort of a density functional can be classified. Perdew introduced the picture of 'Jacob's ladder' which ascends from the 'Hartree-hell' to the 'heaven of chemical accuracy'.[85] Density functionals can be categorized into rungs of that ladder. The higher the rung, the more information of the systems is used in the functional and the more expensive it is. The increase in accuracy with each rung cannot be guaranteed, but that the general picture holds has been verified statistically on a large scale.[86]

The local density approximation

The first rung of Jacobs ladder includes functionals that only take into account the local electron density. This local density approximation (LDA) is based on the assumption that the electron density varies slowly, and that it behaves like the uniform electron gas (UEG). The LDA exchange functional is a modification of the Dirac functional (Eq. 2.20) for which a pre-factor was introduced by Slater (also called X_α method).[87] An LDA description of the correlation energy was derived by Vosko, Wilk and Nusair (VWN)[88] via analytic interpolation formulae based on accurate Monte-Carlo calculations.
The LDA is widely and successfully used for the description of metallic solids, as their electronic structures are similar to the UEG. LDA functionals yield reasonable molecular structures but unfortunately, tend to overbind most molecular systems.

The general gradient approximation

The major reason for the inaccuracies of LDA methods for molecular systems is the strongly varying electron density of a molecule, which cannot be described by an approximation solely based on he UEG. Therefore, in addition to the local density, functionals on the second rung of the ladder take the gradient of the density $\nabla\rho$ into account. This

general gradient approximation (GGA) is based on an LDA description modified by an enhancement factor F_{XC}^{GGA}, which depends on both the electron density and its gradient.

$$E_{XC}^{GGA} = \sum_{\sigma=\alpha,\beta} \int \rho \, E_{XC}^{LDA}[\rho] \, F_{XC}^{GGA}[\rho, \nabla\rho] \, \mathrm{d}\mathbf{r} \tag{2.25}$$

σ is the spin variable for α or β spin, and the LDA exchange and correlation functionals are usually modified separately. Common GGA functionals are the PBE exchange and correlation functional by Perdew, Burke, and Ernzerhof,[89,90] Beckes's B88 exchange functional[91] and the LYP correlation functional by Lee, Yang, and Parr.[92]

meta-GGA functionals

The third rung of Jacobs ladder comprises functionals, which additionally include higher order derivatives of the electron density, such as the electron density Laplacian $\nabla^2\rho$. The Laplacian was found to be numerically unstable and thus, meta-GGA functional are based on enhancement factors that include the related orbital kinetic energy density instead. Meta-GGA functionals are often more accurate than GGA functionals, although this cannot be generalized and the improvement is by far not as large as going from LDA to GGA. The probably most popular meta-GGA functional is the TPSS functional developed by Tao, Perdew, Staroverov, and Scuseria,[93] which was also extensively used in this thesis for the optimization of geometries. GGA and meta-GGA functionals are called semi-local functionals, since they are not only evaluated based on local electron density, but also take information about its close proximity into account.

Hybrid functionals

The forth rung of the ladder contains functionals that use additional non-local information based on the occupied KS-orbitals. This is achieved by substituting a part of the DFT exchange with non-local Fock-exchange evaluated with the KS-orbitals (E_X^{HF}), also called 'exact' exchange. This approach can be motivated based on the adiabatic connection.[94]

$$E_{XC}^{hybrid} = E_C^{(meta-)GGA} + (1 - a_X) E_X^{(meta-)GGA} + a_X E_X^{HF} \tag{2.26}$$

Functionals of this class are named hybrid functionals and its most prominent example is the B3LYP functional. It contains 20 % of Fock-exchange ($a_X = 0.2$), 0.08 % of Slater's LDA and 0.72 % B88 exchange, and 0.19 % VWN-LDA and 0.81 % LYP correlation.[95,96] Further examples of typical hybrid functionals are BHLYP[97] ($a_X = 0.5$) and PBE0[98] ($a_X = 0.25$). Popular hybrid functional based on meta-GGA functionals are Zhao and Truhlar's M05[99] and M06[100] classes of functionals, also frequently called Minnesota

functionals, which are highly parametrized with up to 40 parameters. The PW6B95[101] hybrid functional ($a_X = 0.28$) is also based on a meta-GGA and used in this thesis to a large extent. All the so far mentioned hybrids have in common, that they employ the same amount of Fock-exchange over the whole space. Therefore, these hybrid functionals are also called global hybrids.

The principal problem with global hybrid functionals is that the underlying (meta-)GGA potential decays exponentially. Thus, in the asymptotic limit the hybrid exchange potential decays with a_x/r instead of the correct $1/r$ behavior. This behavior can be corrected by introducing a range separation that retains the short-range error cancellation between exchange and correlation of a global hybrid and yields the correct potential of 100 % Fock exchange in the asymptotic limit. In these so called range-separated hybrid functionals, the two-electron operator $\frac{1}{r_{12}}$ is partitioned into a short-range and a long-range component using the error function (erf).[102,103]

$$\frac{1}{r_{12}} = \frac{1 - \mathrm{erf}(\mu r_{12})}{r_{12}} + \frac{\mathrm{erf}(\mu r_{12})}{r_{12}} \tag{2.27}$$

The short-range part is then treated by an exchange functional, and the long-range part by HF exchange. Examples of this type of functionals are ωB97 and ωB97X developed by Chai and Head-Gordon,[104] and their advancements ωB97X-D[105] and ωB97X-D3.[106] Further examples are LC-BLYP and CAM-B3LYP by Handy,[107] though the latter one is not asymptotically correct as it contains a finite amount of GGA exchange in the long-range limit. Nevertheless, CAM-B3LYP has been used successfully for the computation of electronic circular dichroism spectra within this thesis. In general, range-separated functionals perform good for the calculations of excited states by means of time-dependent DFT, due to the much more appropriate orbital energies.

Virtual-orbital dependent functionals

Functionals on the fifth and last rung of the ladder take the virtual KS-orbitals into account when calculating the correlation energy. Several approaches to accomplish this have been published, including perturbation methods as done by Görling and Levy,[108,109] and random phase approximation (RPA) methods.[110] The probably most widely used approach is that of a double-hybrid density functional (DHDF) proposed by Grimme.[111] A part of the correlation energy is computed by second order Møller–Plesset perturbation theory (MP2) from the KS-orbitals of a preceding hybrid functional SCF calculation.

$$E_{XC}^{DHDF} = (1 - a_X)E_X^{(meta-)GGA} + a_X E_X^{HF} + (1 - a_C)E_C^{(meta-)GGA} + a_C E_C^{MP2} \tag{2.28}$$

The most accurate DHDFs are those containing rather high amounts of Fock exchange and much smaller amounts of non-local correlation. E.g. the B2PYLP [111] functional employs 53 % exact exchange ($a_X = 0.53$) and 27 % MP2 correlation ($a_C = 0.27$). Double-hybrid functionals are very accurate for reactions energies and basic molecular properties. [86] However, as the MP2 part with its formal scaling of $\mathcal{O}(N^5)$ is rather expensive for large molecules and MP2 in general cannot accurately treat $\pi - \pi$ stacked systems, DHDFs were not employed in the work of this thesis.

2.3. Semiempirical methods

In the early days of quantum chemistry the computational resources were very limited and carrying out a Hartree–Fock or DFT calculation was challenging even for small systems using a small basis set. Nowadays, even with the present computational power there are systems, e.g. supra- or biomolecules, which exceed the size of those practically accessible by tHF or DFT methods. Originally, the so called semiempirical methods (SE-MO) were introduced as simplifications to *ab initio* MO treatments, usually HF. Drastic integral approximations are introduced to gain speed-up and empirical parameters are used to partly recover the lost accuracy. In this thesis modified neglect of differential overlap (MNDO) type methods have been used and will be discussed in the first part of this chapter. Conceptually different are the semiempirical tight-binding (TB) versions of DFT methods. The initial DFTB approach [112] as well as the self-consistent charge (SCC) DFTB method [113,114] are employed in this thesis will be introduced in the second part of this chapter.

2.3.1. MNDO type methods

MNDO type methods are typically restricted to the valence electrons only and use a minimal basis set of Slater type orbitals (STOs), i.e. one basis function per valence orbital reduces the number of integrals. They are based on the neglect of diatomic differential overlap (NDDO) approximation, in which all products of basis functions that depend on the same electron coordinates but are located on different atoms A and B are zet to zero. [115,116] The first consequence of the NDDO approximation is, that the overlap matrix $\mathbf{S}$ is reduced to a unity matrix.

$$S_{\mu\nu} = \langle \psi_\nu | \psi_\mu \rangle \equiv \langle \mu | \nu \rangle = \delta_{\mu\nu}\delta_{AB} \tag{2.29}$$

Thus, a special eigenvalue problem has to be solved. As this is not the correct HF eigenvalue problem, the exchange is usually described badly. Further, one-electron integrals

that involve three centers, i.e. two from the basis functions and one from the operator, are set to zero. The ones that remain are the following,

$$\langle\mu_A|\hat{h}|\nu_A\rangle = \langle\mu_A| - \frac{1}{2}\nabla^2 - \frac{Z'_A}{|R_A - r|}|\nu_A\rangle - \sum_{a\neq A}^{nuclei}\langle\mu_A|\frac{Z'_a}{|R_a - r|}|\nu_A\rangle \tag{2.30}$$

$$\langle\mu_A|\hat{h}|\nu_B\rangle = \langle\mu_A| - \frac{1}{2}\nabla^2 - \frac{Z'_A}{|R_A - r|} - \frac{Z'_B}{|R_B - r|}|\nu_B\rangle \tag{2.31}$$

where $\hat{h}$ is the one-electron operator. Z'_a is the nuclear charge reduced by the number of core electrons which assumes complete shielding.
And finally, all three- and four-center two-electron integrals are neglected.

$$\langle\mu_A\nu_B|\lambda_C\sigma_D\rangle = \delta_{AC}\delta_{BD}\langle\mu_A\nu_B|\lambda_A\sigma_B\rangle \tag{2.32}$$

For a basis set containing s and p functions only, there are 27 different one- and two-center two-electron integrals, whereas this number rises over 500 if d functions are added. In case of the sp basis, only five one-center two-electron integrals remain:

$$\begin{aligned} &\langle ss|ss\rangle = G_{ss} \quad \langle sp|sp\rangle = G_{sp} \quad \langle ss|pp\rangle = H_{pp} \\ &\langle pp|pp\rangle = G_{pp} \quad \langle pp'|pp'\rangle = G_{p2} \end{aligned} \tag{2.33}$$

The G-type parameters are Coulomb terms and the H parameter is an exchange integral. The G_{p2} integral involves two different types of p-functions.

Besides NDDO, two other approximations exist which differ mainly in the treatment of the two-electron integrals. The intermediate neglect of differential overlap approximation (INDO) neglects all two-center two-electron integrals which are not of Coulomb type and in the complete neglect of differential overlap (CNDO) only Coulomb one-center and two-center two-electron integrals survive.

The difference in the various NDDO based methods lies in treatment of the remaining integrals. They are either calculated from the functional form of the atomic orbitals, estimated from parameters based on a few atomic experimental data, or obtained via parameters which are fitted to molecular (experimental) data. Older modified NDDO methods, i.e. the MNDO model,[117] Austin Model 1 (AM1),[118] and MNDO Parametric Method number 3 (PM3),[119] use only s and p functions and calculate the overlap integrals $S_{\mu\nu}$ explicitly. They differ in the treatment of the core-core repulsion and how the parameters for the integrals are obtained. MNDO and AM1 use atomic experimental data that was assigned by hand, whereas the PM3 parameters were fitted on a large set of experimental training data. All of these methods are available only for a small number of elements, usually H, C, N, O, F, Si, P, Cl, Br and I.

An older common INDO type method is SINDO1,[120] which is available for the elements H, C, N, O, F, Li, Be, and B. For its modification, MSINDO, parameters exist also for main group elements and transition metals of the third row.[121–123]

Within this thesis the more recent MNDO type methods PM6, and[124] PM7[125] and OM2[126] have been applied. PM6 is a successor of PM3, which uses improved core-core interactions, a *spd* basis set, and a more general parametrization for over 70 elements. PM7 can be considered as the most general purpose semiempirical method. It aims at correcting many faults of the PM6 method, and partially includes dispersion and hydrogen bond corrections. The OM2 method introduces two one-electron orthogonalization corrections. The first one takes valence orthogonalization of the resonance integrals into account and the second one includes the dominant core-valence repulsions through an effective core potential.

2.3.2. Density functional tight binding methods

In a similar way the so far discussed conventional semiempirical methods are approximations to HF theory, the density functional tight binding (DFTB) methods are approximations to DFT. The DFT energy of a system is obtained through solving the KS equations (Eq. 2.24) and adding the nuclear-nuclear repulsion energy E_{nn}. Within the DFTB approach the electron density is expressed as the sum of a reference density ρ_0 and its deviation $\Delta\rho$ from the DFT ground state density ρ. The exchange-correlation energy E_{XC} is expanded in a Taylor series around ρ_0 with respect to $\Delta\rho$ up to second-order.

$$\begin{aligned} E_{XC}[\rho(\mathbf{r}) + \delta\rho(\mathbf{r})] =& E_{XC}[\rho_0(\mathbf{r})] + \int \left[\frac{\delta E_{XC}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})}\right]_{\rho_0} \Delta\rho(\mathbf{r})\,\mathrm{d}\mathbf{r} \\ &+ \frac{1}{2}\int \left[\frac{\delta^2 E_{XC}[\rho(\mathbf{r})]}{\delta^2\rho(\mathbf{r}')}\right]_{\rho_0,\rho_0'} \Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}')\,\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \end{aligned} \tag{2.34}$$

The reference density ρ_0 is build from a superposition of atomic densities and in order to yield a reasonable molecular density, the atomic ones are compressed by an external

potential. The total energy is given as

$$\begin{aligned} E[\rho(\mathbf{r})+\Delta\rho(\mathbf{r})] = & \underbrace{\sum_i \langle\phi_i| -\frac{1}{2}\nabla^2 + V_{en} + \int \frac{\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}' + v_{XC}|\phi_i\rangle}_{E^{H0}} \\ & \underbrace{-\frac{1}{2}\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' - \int v_{XC}[\rho_0(\mathbf{r})]\rho(\mathbf{r})\,\mathrm{d}\mathbf{r} + E_{XC}[\rho_0(\mathbf{r})] + E_{nn}}_{E^{rep}} \\ & \underbrace{+\frac{1}{2}\int \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \left.\frac{\delta^2 E_{XC}}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right|_{\rho_0,\rho_0'} \right) \Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}')\, d\mathbf{r}d\mathbf{r}'}_{E^{\gamma}} \end{aligned} \tag{2.35}$$

The first term, E^{H0}, is the energy contribution depending on the reference density ρ_0 only. Since ρ_0 is build from a superposition of atomic densities, the matrix elements of the Hamiltonian $H^0_{\mu\nu}$ can be computed in an atomic orbital basis set in advance.

$$E^{H0} = \sum_i \sum_{\mu\in a} \sum_{\nu\in b} n_i c_{\mu i} c_{\nu_i} H^0_{\mu\nu} \tag{2.36}$$

The second contribution, E^{rep}, is an approximation to the electron Coulomb interaction, the core repulsion, and the exchange-correlation contributions. In a tight-binding approach, these are approximated as a sum of one-center terms and short-ranged two-center potentials V^{rep}_{ab}. As the atomic one-center contributions result in a constant shift which cancels when considering energy differences, they are neglected within DFTB.

$$E^{rep} = \frac{1}{2}\sum_{ab} V^{rep}_{ab}[\rho_{0,a}, \rho_{0,b}, r_{ab}] \tag{2.37}$$

The second order term is neglected in the original DFTB approach and in this case, the energy can be obtained non-self-consistently. Within the self-consistent charge DFTB (SCC-DFTB) the second order contribution is approximated with E^{γ}, which needs to be evaluated self-consistently.

$$E^{\gamma} = \frac{1}{2}\sum_{ab} \Delta q_a \Delta q_b \gamma_{ab} \tag{2.38}$$

$\Delta q_a = q_a - q_{0,a}$ is the net charge of atom a and γ is a function that takes the electron-electron interaction into account. γ is given by the integral over the product of two normalized Slater-type spherical charge densities. For large distances r_{ab}, γ_{ab} reduces to $1/r_{ab}$ and for $a = b$ it describes the self-repulsion $\gamma_{aa} = U_a$ introducing the Hubbard parameter U_a.

The total energy within the SCC-DFTB method is given as a sum of these three con-

tributions.

$$E^{SCC-DFTB} = E^{H0} + E^{rep} + E^{\gamma} \tag{2.39}$$

$$= \sum_i \sum_{\mu \in a} \sum_{\nu \in b} n_i c_{\mu i} c_{\nu_i} H^0_{\mu\nu} + \frac{1}{2} \sum_{ab} V^{rep}_{ab}[\rho_{0,a}, \rho_{0,b}, r_{ab}] + \frac{1}{2} \sum_{ab} \Delta q_a \Delta q_b \gamma_{ab} \tag{2.40}$$

Within this thesis only the SCC-DFTB approach is used, and for simplicity, the prefix SCC is dropped in the following. Further, the latest third order inclusive version with empirical damping of hydrogen containing pair-potentials and self-consistent charge redistribution is employed,[127] and the 3OB Slater–Koster files constructed by Elstner and coworkers are used.[128,129]

2.4. London dispersion corrections to DFT and HF

One problem that all mean-field methods, i.e. semi-local functionals as well as Hartree–Fock and its semiempirical variants share, is that they cannot provide the correct $-C_6/R^6$ dependence of the long-range correlation, so called London dispersion energy in the asymptotic limit. Here, R is the the inter-atomic distance.[3,4,23,24,31] Therefore, the description of noncovalently bound systems, which are in the focus of this thesis, is usually bad. Over the recent years, many different approaches to treat London dispersion were proposed. For reviews and overviews of state-of-the-art dispersion corrections see Refs. [32–34]. In the following, only the DFT-D scheme introduced by Grimme and coworkers will be discussed in more detail as it was extensively used in this thesis. The DFT-D scheme provides a semi-classical dispersion energy E^D_{disp} that can simply be added to any converged DFT, HF or SE-MO calculation.

$$E^{DFT\text{-}D} = E^{DFT} + E^D_{disp} \tag{2.41}$$

The correction is based on a perturbative treatment of the interaction and the multipole expansion of the intermolecular potential $V(R)$ at large distances R. The application of perturbation theory leads to the potential for the dispersion energy $V_{disp}(R_{AB})$ of two systems A and B.[3]

$$V_{disp}(R_{AB}) = -\sum_{n=6,7,8,9,..} \frac{C^{AB}_n}{R^n_{AB}} \tag{2.42}$$

The C^{AB}_n are the system dependent n^{th} order dispersion coefficients. There is no convention for the sign of the C^{AB}_n in the literature. Within this thesis, positive C^{AB}_n are used, which gives rise to the minus sign in the definition of the inherently negative dispersion energy and provides a stronger interaction for larger coefficients. For the approximation

that the interacting systems are spherical the odd terms vanish and the resulting potential

$$E_{disp} = -\frac{1}{2}\sum_{A\neq B}\sum_{n=6,8,10,..}\frac{C_n^{AB}}{R_{AB}^n} \tag{2.43}$$

can be used to create an isotropic atom-pairwise dispersion correction to DFT. DFT accounts for some correlation energy in the short- to medium-range regime, and the sum in Equation 2.43 diverges for small R_{AB}. Thus, the dispersion energy needs to be damped for small distances.

The first version of the DFT-D dispersion correcting scheme (DFT-D1), was published in 2004.[130] It takes only the leading order (dipole-dipole, C_6^{AB}) term of the multipole expansion (Eq. 2.43) into account and introduces a damping function that lets the dispersion energy vanish for short distances. The C_6 dispersion coefficients are element-specific averages over different hybridization states based on experimental dipole oscillator strength distributions. DFT-D1 is only available for a few elements (H,C–Ne) and was therefore extended in 2006.[131] The main change in the second version (DFT-D2) is that the C_6 are calculated using PBE0/QZVP values of the atomic static polarizabilities and ionization potentials for all elements up to xenon via the original London formula for the dispersion interaction.[31]

The major revision in 2010, named DFT-D3,[36] is less empirical than the previous versions, and is the one that is extensively used in this thesis. It also includes a higher order (dipole-quadrupole, C_8^{AB}) term to account for medium-range dispersion effects.

$$E_{disp}^{D3} = -\frac{1}{2}\sum_{A\neq B}\sum_{n=6,8} s_n \frac{C_n^{AB}}{R_{AB}^n + f_{damp}^{BJ}(R_0^{AB})^n} \tag{2.44}$$

The C_n^{AB} denote the averaged isotropic n^{th} order dispersion coefficient for each atom pair AB which depend on the coordination number. The dispersion coefficients C_6^{AB} are computed starting from the Casimir–Polder formula.[132]

$$C_6^{AB} = \frac{3}{4}\int_0^\infty \alpha^A(i\omega)\alpha^B(i\omega)\mathrm{d}\omega \tag{2.45}$$

$\alpha(i\omega)$ is the averaged dipole polarizability at imaginary frequency ω calculated *ab initio* by time-dependent (TD) DFT for appropriate reference systems. The higher-order coefficients are calculated from the C_6^{AB} with a recursion formula.
The s_n in Equation 2.44 are the global scaling factors. For common density functionals s_6 is usually set to unity to ensure the correct asymptotic behavior, whereas s_8 is optimized for each functional. Note that the s_8 was set to zero for most of the Minnesota class functionals except M05 due to double counting of the medium-ranged dispersion that is

fitted into the functionals themselves.

The initially proposed damping function is very close to the damping function used in DFT-D1/2. As the dispersion energy vanishes for short distances, this damping is called zero-damping and thus, the correction scheme is dubbed D3(0). Problematic is that the zero-damping may result in unphysical repulsion forces. Becke and Johnson (BJ) introduced a rational damping function f_{damp}^{BJ} that leads to a constant contribution of the dispersion energy to the total correlation energy from spatially close, i.e. directly bonded, pair of atoms.[133,134]

$$f_{damp}^{BJ}(R_0^{AB}) = a_1 R_0^{AB} + a_2 \tag{2.46}$$

a_1 and a_2 are fitting parameters and $R_0^{AB} = \sqrt{C_8^{AB}/C_6^{AB}}$ is the cut-off parameter. This BJ-damping function is the default in the DFT-D3 scheme and thus, we will refer to it as DFT-D3 instead of DFT-D3(BJ) in the following.[37]

The dispersion energy is mainly additive but non-additive many-body contributions can play a role in large systems that are in the focus of this thesis. The importance of many-body dispersion interactions has been recently analyzed by various groups.[135–137] Within the D3 dispersion correction scheme an Axilrod-Teller-Muto (ATM) type three-body (dipole-dipole-dipole) term is available.[138,139]

$$E_{disp}^{(3)} = -\frac{1}{6} \sum_{A \neq B \neq C} \frac{C_9^{ABC}(3 \cos\theta_a \cos\theta_b \cos\theta_c + 1)}{(R_{AB} R_{BC} R_{AC})^3} f_{damp}^0(R_0^{AB}) \tag{2.47}$$

θ_a, θ_b and θ_c are the three angles between the three atoms A, B and C, and R_{AB}, R_{BC} and R_{AC} are the respective interatomic distances. The C_9^{ABC} dispersion coefficients are approximated from the C_6 coefficients.

$$C_9^{ABC} \approx -\sqrt{C_6^{AB} C_6^{BC} C_6^{AC}} \tag{2.48}$$

The damping function is the zero-damping variant similar to the one used in DFT-D1/2. It employs geometrically averaged distances and averaged cut-off radii as well as modified, functional independent parameters. Note that the three-body term is usually repulsive for dense systems.

A drawback of the D3 scheme is its semiempirical character and the need of a parameter fit for every functional. Further, the D3 correction does not depend on the density and thus, the electronic structure is not directly affected. Although, an indirect effect due to the altered geometry (coordination number) is present. The independence on the density can be advantageous for semiempirical methods, since their density is usually inaccurate. Another flaw is that the C_6 coefficients are not reasonable if no appropriate reference systems exist. Therefore, the D3 scheme can be problematic for anions, cations, organo-

metallic compounds, and systems with a small or vanishing gap, e.g. metals.
One advantage of the D3 scheme over other dispersion corrections is its availability for almost all elements in the periodic table and its possible combination with all relevant HF, DFT and semiempirical methods. Further, the computation is extremely fast and analytical nuclear gradients exist. Even for thousands of atoms the scaling behavior with system size is low and the computational pre-factor is small. Another benefit is that the corrections can simply be added on top of any DFT, HF or semiempirical calculation without the necessity for a specific implementation into a program package.

2.5. Gas phase thermodynamic properties

Up to now, potential and kinetic energies of electrons and nuclei in a single atom or molecule and the total electronic energy of that atom or molecule were regarded. As in a chemical reaction usually an ensemble of molecules is present, this chapter deals with the necessary steps to make the transition from single-molecule energies to ensemble thermodynamic variables.

Within the BO approximation the electronic energy is calculated for a fixed nuclear conformation that is treated classically. Even if nuclear motion on the BO potential energy surface (PES) is allowed, no quantum effects for the nuclei are considered. Nevertheless, there exists a lowest non-zero vibrational energy level for any bound molecule. In order to obtain the full energy at zero temperature ($T = 0$), the zero-point vibrational energy (ZPVE) hat to be taken into account. The usual approach to describe molecular vibration is the quantum harmonic oscillator (HO) approximation. The nuclear Schrödinger equation for a quadratic potential needs to be solved in order to obtain the energy eigenvalues for the nuclear vibrations. With a transformation into a unique set of mass-dependent, vibrational normal coordinates $\mathbf{q}$ it is possible to separate the initially $3N$ dimensional equation into $3N$ one-dimensional Schrödinger equations.

$$\left[-\frac{1}{2\mu_i}\frac{d^2}{dq_i^2} + \frac{1}{2}k_i(q_i - q_{eq,i})^2\right]\Xi(q_i) = E\Xi(q_i) \tag{2.49}$$

There, μ_i is the reduced mass, k_i is the bond force constant, i.e. the second derivative of electronic energy with respect to q_i at $q_{eq,i}$. These equations can be solved analytically, their eigenfunctions are products of Hermite polynomials and Gaussian functions, and

their energy eigenvalues $E_{i,n}$ depend on the force constant and the reduced mass.

$$E_{i,n} = \left(n + \frac{1}{2}\right) \hbar\omega_i \tag{2.50}$$

$$\omega_i = \frac{1}{2\pi}\sqrt{\frac{k_i}{\mu_i}} \tag{2.51}$$

ω_i is the vibrational frequency of the ith vibrational mode, and $\hbar$ is the reduced Planck constant. The lowest energy level for each normal vibration is then given as $\hbar\omega_i/2$ and the sum over all these energies defines the zero-point vibrational energy. Adding the ZPVE to the electronic energy yields the the internal energy U_0 for a molecule at 0 K.

$$U_0 = E_{elec} + \sum_{i}^{modes} \frac{1}{2}\hbar\omega_i \tag{2.52}$$

As the least harmonic modes are soft modes, e.g. weakly hindered torsions, which have small vibrational frequencies, their contribution to the ZPVE is also small. Thus, the harmonic oscillator approximation provides good results for the ZPVE if the frequencies themselves are accurate enough. Since the second derivatives of the energy with respect to the nuclear coordinates gets very expensive for large systems, a common approach is to scale the frequencies obtained with simpler and less accurate electronic structure methods. E.g. for HF a scaling factor of about 0.9 is necessary to bring the computed frequencies in good agreement with experiment.

A number of molecules at $T > 0$ are described with statistical mechanics, which requires certain external conditions to be constant. For a canonical ensemble these constants are the total number of identical molecules N, the volume V and the temperature T. The canonical ensemble is described by its partition function $Q(N, V, T)$.

$$Q(N, V, T) = \sum_{i} e^{-E_i(N,V)/k_B T} \tag{2.53}$$

The index i runs over all possible energy states E_i of the system and k_B is Boltzmann's constant. Within the canonical ensemble the internal energy U, the enthalpy H, the

entropy S, and the Gibbs free energy G can be calculated from its partition function.

$$U = k_B T^2 \left(\frac{\partial lnQ}{\partial T} \right)_{N,V} \tag{2.54}$$

$$H = U + PV \tag{2.55}$$

$$S = k_B \, lnQ + k_B T \left(\frac{\partial lnQ}{\partial T} \right)_{N,V} \tag{2.56}$$

$$G = H - TS \tag{2.57}$$

The problem is that finding an explicit expression for $Q(N, V, T)$ for the real system is non-trivial, and therefore, simplifications are made. First, the ensemble is approximated as an ideal gas. Thus, the indistinguishable molecules do not interact with each other, which reduces the problem of finding the partition function of the ensemble $Q(N, V, T)$ to finding a molecular partition function $q(V, T)$. Further, the PV term in the enthalpy equation (Eq. 2.55) can be replaced with Nk_BT. Second, the molecular energy is assumed to be a separable sum of electronic, translational, rotational, and vibrational terms, which leads to a molecular partition function which is a product of these individual components.

$$Q(N, V, T) = \frac{[q(V, T)]^N}{N!} \tag{2.58}$$

$$q(V, T) = q_{elec}(T) \, q_{trans}(V, T) \, q_{rot}(T) \, q_{vib}(T) \tag{2.59}$$

The electronic and translational partition functions will not be discussed further, as their resulting contribution to U and S is either zero (q_{elec}) or very small (q_{trans}) for common and within this thesis considered closed-shell molecules. To find an expression for the molecular rotational partition function, the molecule is assumed to be a quantum rigid rotor (RR). For a diatomic molecule, the rigid rotor Schrödinger equation can be solved and the analytical expression for the energy eigenvalues E_J, which depends on the moment of inertia I, is:

$$E_J = \frac{J(J+1)h^2}{8\pi^2 I} \tag{2.60}$$

The solution for the diatomic case is general for any linear molecule, as long as I is computed in the appropriate way for the whole molecule. The solution for the general rigid rotor with three unique axes and associated moments of inertia is non-trivial. Therefore, the generalization of the classical rigid rotor problem is transferred to the quantum problem and a simple expression for the rotational partition function is possible.

$$q_{rot}(T) = \frac{\sqrt{\pi I_A I_B I_C}}{\sigma} \left(\frac{8\pi^2 k_B T}{h^2} \right)^{3/2} \tag{2.61}$$

I_A, I_B, and I_C are the principal moments of inertia, and σ is a symmetry number, that gives the number of pure rotations that carry the molecule into itself. With $q_{rot}(T)$ the rotational contributions to the internal energy U_{rot} and the entropy S_{rot} can be derived.

$$U_{rot} = \frac{3}{2}RT \tag{2.62}$$

$$S_{rot} = R\left[\frac{3}{2} + \ln\left(\frac{\sqrt{\pi I_A I_B I_C}}{\sigma}\left(\frac{8\pi^2 k_B T}{h^2}\right)^{3/2}\right)\right] \tag{2.63}$$

R is the gas constant and k_B is Boltzmann's constant. Any methodology which yields accurate geometries and thus, accurate principal moments of inertia is useful for constructing the rotational partition function and the thermodynamic variables computed therefrom.

To construct an expression for the molecular vibrational partition function $q_{vib}(T)$, it is assumed that the vibrational energy can be expressed as a sum of individual energies associated with each mode. Each of theses $3N-6$ vibrational modes is described within the quantum harmonic oscillator approximation (Eq. 2.49, Eq. 2.50). If the ZPVE is included in the zero of energy U_0 (Eq. 2.52), the zeroth vibrational energy level is zero for every mode. The vibrational partition function $q_{vib}(T)$ is a product of all partition functions of the individual modes.

$$q_{vib}(T) = \prod_i^{3N-6}\left(\sum_k^{\infty} e^{-kh\omega_i/k_BT}\right) = \prod_i^{3N-6}\left(\frac{1}{1-e^{-h\omega_i/k_BT}}\right) \tag{2.64}$$

As the vibrational modes are treated as independent variables, a full geometry optimization and subsequent evaluation of the frequencies via the second derivatives is necessary to compute the vibrational contributions to the internal energy U_{vib} and the entropy S_{vib}.

$$U_{vib} = R\sum_i^{3N-6}\frac{h\omega_i}{k_B(e^{h\omega_i/k_BT}-1)} \tag{2.65}$$

$$S_{vib} = R\sum_i^{3N-6}\left[\frac{h\omega_i}{k_BT(e^{h\omega_i/k_BT}-1)} - \ln\left(1-e^{-h\omega_i/k_BT}\right)\right] \tag{2.66}$$

Low-lying vibrational frequencies are inaccurate in the harmonic treatment and the numerical noise in standard quantum chemical calculations causes further errors. Unfortunately, they have a great impact on the vibrational entropy since the Bose-Einstein factor diverges for small frequencies (Eq. 2.66). Thus, an interpolated rigid-rotor-harmonic-oscillator (RRHO) approach proposed by Grimme will be used. This partially replaces the contribution of the low-lying modes to the entropy by a corresponding rotational entropy.[46] For any low-lying normal vibrational mode the moment of inertia μ of a rigid

rotor with the same energy is calculated from the frequency. To restrict this quantity to a reasonable value, the average molecular moment of inertia I_{av} is used a as limiting value for very small ω and thus, very large μ.

$$\mu = \frac{h}{8\pi^2\omega} \tag{2.67}$$

$$\mu' = \frac{\mu I_{av}}{\mu + I_{av}} \tag{2.68}$$

Via the effective moment of inertia μ' the entropy of a low-lying mode is evaluated as a rotational entropy given in Equation 2.63.
In order to smoothly interpolate between the rotational and harmonic treatment of the vibrations, S_{rot} and S_{vib} are combined by using the Head-Gordon weighting function $w(\omega)$.[105]

$$S = w(\omega)S_{vib} + [1 - w(\omega)]\, S_{rot} \tag{2.69}$$

$$w(\omega) = \frac{1}{1 + (\omega_0/\omega)^4} \tag{2.70}$$

This approach ensures interpolation between the harmonic vibrational entropy for $\omega \gg \omega_0$ and a pure and finite rotational entropy for small ω. The default value for ω_0 is 100 cm^{-1}. A similar cut-off value has recently been used by Truhlar and co-workers.[140,141]

2.6. Continuum solvation models

So far, all energies and thermodynamic properties have been regarded in the gas phase. However, most of chemistry including all of biochemistry takes place in solution and solvation effects are often crucial. Especially for supramolecular host-guest systems which are in the focus of this thesis it is common to tune the binding constant of a specific host-guest complex by changing the polarity and size of the solvent molecules.[16,17] From a theoretical point of view, describing a chemical reaction in a solution is a very complex problem, and the involved possibilities of an explicit treatment of solvation via molecular mechanics will not be discussed. Within this thesis the simpler approach of an implicit treatment of the solution environment via continuum solvation models[142,143] has been employed and these methods are introduced briefly in this chapter.

All implicit solvation models are based on the Poisson equation of classical electrostatics, which expresses the electrostatic potential $\phi(\mathbf{r})$ as a functions of the charge density $\rho(\mathbf{r})$ and the dielectric constant ε of the medium.

$$\nabla^2\phi(\mathbf{r}) = -\frac{4\pi\rho(\mathbf{r})}{\varepsilon} \tag{2.71}$$

The explicitly treated solute forms a cavity inside the homogeneous dielectric medium and thus, there are two regions, one inside and one outside the cavity, for which the Poisson equation can be formulated as:

$$\nabla\varepsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{2.72}$$

The cavity for a given solute, also called the solvent accessible surface (SAS), can be constructed e.g. by superposition of atomic spheres with respective Van-der-Waals radii. For arbitrarily shaped cavities, the analytical solution of the Poisson equation is not possible. In order to solve the equation numerically, the SAS is tessellated into i surface segments with an area s_i.

Instead of solving the resulting equations for the exact dielectric boundary conditions the conductor-like screening model (COSMO) approximates the dielectric medium as a perfect conductor, i.e. ε is infinite. Thus, the electrostatic potential vanishes at each of the surface segments and the resulting equations are much easier to solve.

$$\mathbf{Aq} + f(\varepsilon)\,\mathbf{BQ} = 0 \tag{2.73}$$

The vector $\mathbf{q}$ contains the outside surface charges q_i and $\mathbf{Q}$ contains the given inside screening charges Q_i of the solute cavity segments. The matrices $\mathbf{A}$ and $\mathbf{B}$ generate the electrostatic potential on the surface segments and the solute cavity segments, respectively. $f(\varepsilon)$ is a scaling function for the screening charges in order to achieve an approximation for a finite dielectric.

$$f(\varepsilon) = \frac{\varepsilon - 1}{\varepsilon + 0.5} \tag{2.74}$$

The surface charges q_i can be used as external charges within the SCF procedure of a HF or DFT calculation. Due to the altered electron density and subsequent change in the surface charges, the solute electron density and the polarization charges are iterated to self-consistency. COSMO is a purely electrostatic model that influences the electronic energy only. Explicit solute-solvent interactions like hydrogen bonding or dispersion interactions are neglected and a solvation correction to the free energy is not accessible.

The conductor-like screening model for real solvents (COSMO-RS) overcomes these deficiencies by using the screening charge densities σ from a precedent COSMO calculation to calculate the chemical potential μ of a solute in solution.[41,42] Based on the resulting chemical potentials other thermodynamic equilibrium properties such as the free energy of solvation, activity coefficients, partition coefficients, solubility, and vapor pressure are available.

Within the COSMO-RS model, the liquid phase is approximated to be incompressible.

Further, it is assumed that all parts of the solute and solvent molecule surface can be in contact with each other, but only pairwise interactions of the surface segments are allowed. For the solute as well as the solvent molecule a histogram of the charge densities $p(\sigma)$, the so called σ profiles are constructed and are the only information used. This way, all geometrical constrains vanish and the statistical thermodynamics of the liquid ensemble reduces to the much simpler statistical thermodynamics of the corresponding ensemble of pairwise interacting surface segments. For an ensemble S of solvent molecules the chemical potential μ_S is determined as

$$\mu_S(\sigma) = -\frac{k_B T}{a_{eff}} \ln \int p_S(\sigma') \exp\left(-\frac{a_{eff}\left(E_{int}(\sigma,\sigma') - \mu_S(\sigma')\right)}{k_B T}\right) \mathrm{d}\sigma' \tag{2.75}$$

where a_{eff} is the effective area of the surface segment. The interaction energy E_{int} is the sum of the electrostatic E_{misfit} and the hydrogen bonding energy E_{hb}.

$$E_{misfit}(\sigma) = \frac{\alpha}{2}(\sigma + \sigma')^2 \tag{2.76}$$

$$E_{hb}(\sigma) = c_{hb}(T) \min[0, \sigma\sigma' - \sigma_{hb}^2] \tag{2.77}$$

α, C_{hb}, and σ_{hb}^2 are fitting parameters, obtained from a large training set of data. The electrostatic interaction has been dubbed "misfit" energy because it results from the mismatch of the interacting charged surface segments compared to the perfect conductor.

Equation 2.75 needs to be solved iteratively, usually starting with $\mu_S = 0$. After convergence, μ_S describes the affinity of a solvent S for a molecular surface of polarity σ. The chemical potential of any solute molecule X in this solvent S is calculated by integrating μ_S over the surface of X.

$$\mu_S^X = \int p^X(\sigma)\mu_S(\sigma)\mathrm{d}\sigma + \mu_{comb,S}^X + \mu_{disp} \tag{2.78}$$

$\mu_{comb,S}^X$ is a combinatorial correction term and μ_{disp} is the dispersion contribution, which is originally a simple surface proportional term but has been replaced by the D3 scheme (see section 2.4) in the newest version.

One of the main advantages of COSMO-RS is its capability to treat all kinds of solvents and solutes. For a solvent mixture, the respective σ-profiles are simply averaged according to their mole fractions. Within this thesis, the COSMO-RS model has been successfully employed for the computation of free energies of solvation.

Part II.

Evaluation and development of cost-efficient methods for noncovalent interactions

Part II contains two chapters on the evaluation and development of cost-efficient DFT and HF based methods for the description of noncovalent interactions. It has been shown that dispersion corrected hybrid density functional approximations together with a large atomic-orbital basis set yield accurate interaction energies for noncovalently bound systems.[35,46,86,144] However, if the systems are larger than a few hundred atoms, like small proteins, or if there are many of them, e.g. whenever sampling various binding conformation for a given host-guest complexes is necessary, these calculations are too expensive. Several semiempirical methods exist (see Chapter 2.3) which are two to three orders of magnitude faster than HF or DFT calculations in a large basis set. But even if a dispersion correction is used, the interaction energies are typically not accurate enough. Therefore, the following two chapters aim at filling the gap between the semiempirical methods and the DFT/large basis calculations in terms of cost and accuracy (See Figure 2.1).

The first idea that comes to mind when the computational cost has to be decreased, is the reduction of the number of basis functions. Thus, Chapter 3 evaluates the accuracy of DFT and HF calculations with double-zeta basis sets. B3LYP/6-31G* has been in extensive use for years and which can yield surprisingly good results for various systems and reactions. As mentioned before, a semi-local density functional cannot describe the asymptotically correct $1/R^6$ dependence of the London dispersion energy on the interatomic distance R. However, if a dispersion correction is applied, the results are often worse. This is due to the basis set superposition error (BSSE), which is the second major error source in a DFT or HF calculation with a small basis set and thus, a second correction is needed. Both errors are analyzed further in Chapter 3 and modern correction schemes to include dispersion and prevent BSSE are presented. Among other methods, the combination of the D3 dispersion correction scheme[36,37] with a recently published

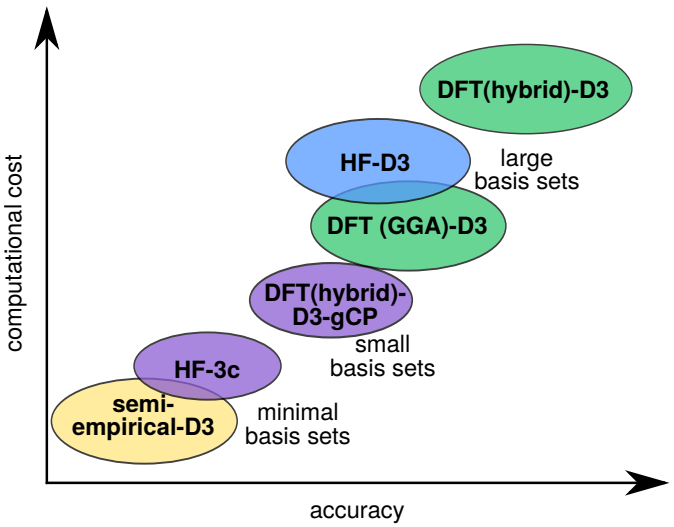


Figure 2.1.: Sketch of the cost-accuracy ratio of several quantum chemical methods. A dispersion correction is assumed in all cases.

very efficient geometric counterpoise correction (gCP)[47] to correct for the BSSE is used which has shown promising results before.[145,146] An exemplary benchmark study on interaction energies and geometries of noncovalently bound dimers, host-guest complexes and organic molecular crystals is presented in order to compare the accuracy of those correction schemes and to evaluate the improvement over plain DFT and HF.

When the basis set is reduced even further to a minimal basis set, the basis set incompleteness error (BSIE) becomes problematic. A novel HF based method that includes dispersion, and corrects for BSSE and BSIE (HF-3c) is introduced in Chapter 4 Three atom-pairwise corrections are added on top of a HF calculation in a minimal basis set. These are the D3 dispersion correction scheme,[36,37] the gCP correction to efficiently correct for BSSE,[47] and a short-ranged correction to compensate for the BSIE effects. HF-3c has nine global parameters and is less empiric than conventional semiempirical methods, but it is about a factor of 50 slower. Nevertheless, in terms of cost and accuracy HF-3c might fill the gap between existing semiempirical methods and density functionals with a large basis set. The accuracy of HF-3c is evaluated for the geometries of small covalently bound organic molecules, geometries and interaction energies of small noncovalent dimers and large supramolecular host-guest systems, geometries of small proteins, and thermostatistical corrections.

3. Small basis set first-principles quantum chemical methods for large molecular and periodic systems

Rebecca Sure*, Jan Gerit Brandenburg*, and Stefan Grimme*

Received 5th of October 2015

Reprinted (adapted) with permission from
R. Sure, J. G. Brandenburg, and S. Grimme, *ChemistryOpen* **2015**, *accepted*.

Own manuscript contribution:

- Performance of most calculations for the molecular systems
- Interpretation of data
- Writing major parts of the manuscript

*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany

Abstract

In quantum chemical computations the combination of Hartree–Fock or a density functional approximation with relatively small atomic orbital basis sets of double-zeta quality is still widely used, e.g., in the popular B3LYP/6-31G* approach. We critically analyze the two main sources of error in such computations, that is the basis set superposition error on the one hand and the missing London dispersion interactions on the other. We review various strategies to correct those errors and present exemplary calculations on mainly noncovalently bound systems of widely varying size. Energies and geometries of small dimers, large supramolecular complexes, and molecular crystals are covered. We conclude that it is not justified to rely on fortunate error compensation, as the main inconsistencies can be cured by modern correction schemes which clearly outperform the plain mean-field methods.

3.1. Introduction

Kohn-Sham density functional theory (KS-DFT, or simply DFT in the following)[147,148] has evolved to be today's most widely used electronic structure method and has emerged as the theory of choice for application to various problems in the chemical and physical science. Due to its good cost-accuracy ratio this especially holds for large molecular systems and solids. The number of collaborative experimental and theoretical studies grew tremendously in the last decade. The usage of complementary theoretical and experimental information can generate valuable new insights and it is nowadays possible to explain and describe various phenomena in a detailed mechanistic way based on routine quantum chemical calculations.

DFT is considered as the natural theory for extended systems but its current, partially semiempirical character requires extensive benchmarking on theoretical or experimental reference values. Over the past years, such benchmark studies have been carried out with diligence, mainly focusing on energetic properties[86,149,150] and more recently also regarding structures of small to medium sized molecules[150–153]. However, the number of proposed density functionals is already too huge to be covered comprehensively. Consequently, the task to select an appropriate and efficient level of theory for a specific problem is highly non-trivial. Thus, it comes as no surprise that non-experts often choose methods purely because of their popularity and those are not necessarily the best options for their application. This eventually results in a waste of computational as well as human resources.

One prominent example is the combination of the B3LYP functional[88,92,96,154] with the 6-31G* double-ζ one-particle AO basis set[155] in particular, or similar functionals with a

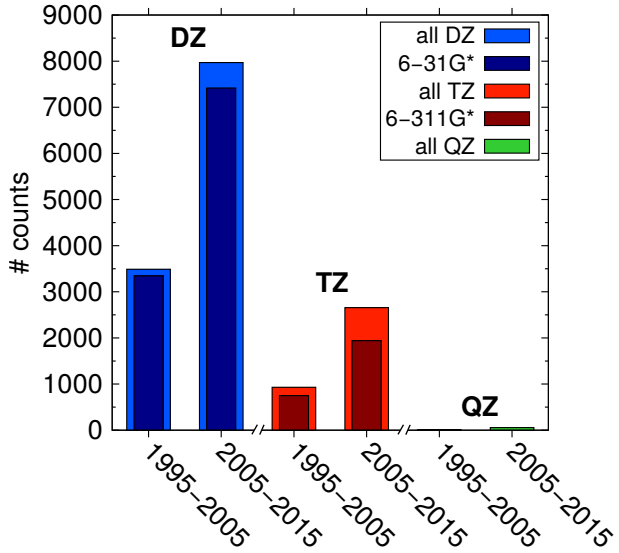


Figure 3.1.: SciFinder[156] hits (dated May 2015) for journal articles containing the B3LYP functional in combination with DZ (6-31G*, def2-SVP, cc-pVDZ, aug-cc-pVDZ), TZ (6-311G*, def2-TZVP, cc-pVTZ, aug-cc-pVTZ) and QZ (def2-QZVP, cc-pVQZ, aug-cc-pVQZ) basis sets from the periods 1995-2005 and 2005-2015.

small double-ζ (DZ) basis set in general. Over the last decade the computational power has increased immensely and using a well converged basis set (BS) is feasible in many cases. A search with SciFinder[156] for the exemplary B3LYP functional reveals that the ratio of journal articles using it in combination with a DZ basis and those using it with a triple-ζ (TZ) basis is roughly the same compared to the previous decade. In the years 1995 to 2005 this ratio was about 3.5:1 and it only dropped slightly to about 3:1 during the last decade (Figure 3.1).

In 2005 Ahlrichs *et. al* published the efficient def2-SVP (DZ) and def2-TZVP (TZ) BSs which were specifically designed for SCF calculations. However, the 6-31G* (DZ type) and 6-311G*[157] (TZ type) published by Pople *et. al* in 1972 and 1980, respectively, are still widely used in DFT calculations.

Compared to the citations of DZ and TZ basis sets, the number of articles employing B3LYP in combination with quadruple-ζ (QZ) type expansions is tiny. Utilizing QZ basis sets in HF or DFT calculations leads to results which are chemically very close to the complete basis set (CBS) limit and this is our general recommendation if this level is affordable. Because of the faster BS convergence compared to correlated post-Hartree–Fock methods, normally no further BS extrapolation scheme is needed. However, these calculations are routinely possible on standard workstations only for medium-sized systems with about 100 atoms or less.

If the system size increases, or one has to perform very many calculations, and one

is rather limited in the computational resources, as most mainly experimentally working groups are, a DZ basis is sometimes the only choice. Even with modern computational equipment, a sufficiently fast, and at the same time, reasonably accurate and interaction consistent electronic structure method is mandatory for the screening of a large conformational space, for instance in the fast growing field of organic crystal structure prediction.[158–160] Therefore, in this short review article we want to emphasize the problems that arise from using a small DZ or related BSs, give an overview of methods to circumvent these problems, and discuss some exemplary calculations to provide a survey on the accuracy of the selected methods. This work extends our previous activities in the field which were focused specifically on B3LYP/6-31G* thermochemistry[145]. For related papers concerning noncovalent interactions see Refs. [32–34,161–165]

3.2. Problems of double-ζ basis sets

There are two major shortcomings of small BS Hartree–Fock (HF) or DFT calculations. The first one is the BS error. This error can be split further into the basis set superposition error (BSSE) and the basis set incompleteness error (BSIE).

Almost all quantum chemical simulations rely on systematic error compensations between the initial (reactant) and final state (product) calculation. The BSSE is caused by the fact that with a small BS the monomers and the complex in a reaction are not treated on equal footing which destroys the error compensation. Typically, this is discussed in the context of noncovalently bound complexes but the same phenomenon also appears for covalent bond-forming chemical reactions as well as in intramolecular transformations. In a dimer complex AB the BS is larger than the individual ones of the monomers A and B because the unoccupied orbitals from A can be used by B and vice versa. This variational 'borrowing' of basis functions leads to an artificial energy lowering of the complex.

The most common approach to circumvent an intermolecular BSSE is the counterpoise (CP) correction scheme proposed by Boys and Bernadi (BB-CP)[166]. The BB-CP counterpoise correction ΔE^{CP} for a dimer complex AB is defined as

$$\Delta E^{CP} = E(A)_a - E(A)_{ab} + E(B)_b - E(B)_{ab} \tag{3.1}$$

where a and b are the BSs belonging to the monomers A and B in their frozen AB complex geometries. This approach is also termed molecular CP correction as only two fragments (the former monomers) are taken into account. Although the BB-CP approach is not free of criticism[167–169], it is widely used and found to be a robust approximation for the self-consistent field (SCF) methods HF and DFT when applied to molecular aggregates.

The BSSE depends on the number of virtual functions that are supplied by the addi-

tional fragment in the complex and on their respective overlap. Because the HF/DFT total energies converge exponentially with respect to the BS size, the initial increase of BSSE with BS size eventually decreases as the CBS limit is approached. The electron density decays exponentially with the distance and the corresponding exponent is determined by the ionization potential of the fragment.[170,171] Because the inclusion of Fock exchange in a hybrid functional increases the ionization potential, this leads to a more compact density, a smaller density overlap of neighboring atoms, and a maller BSSE. This can be qualitatively described as

$$E_{\mathrm{BSSE}} \propto N_{bf} \times \exp\left(-N_{bf}\right) \times \exp\left(-\sqrt{2\,I}\;r\right) , \tag{3.2}$$

with the number of virtual basis functions N_{bf}, ionization potential I, and electron-molecule distance r. We have adjusted this function with variable pre-factors to the Boys-Bernadi CP energy of the S66[172] noncovalent dimers (see below) for functionals with varying amount of Fock exchange (PBE: 0%, B3LYP: 20%, PBEh-3c: 42%, HF: 100%) and increasing BS size (MINIX, def2-SV(P), def2-TZVP, def2-QZVP). The corresponding contour plot is shown in Figure 3.2.

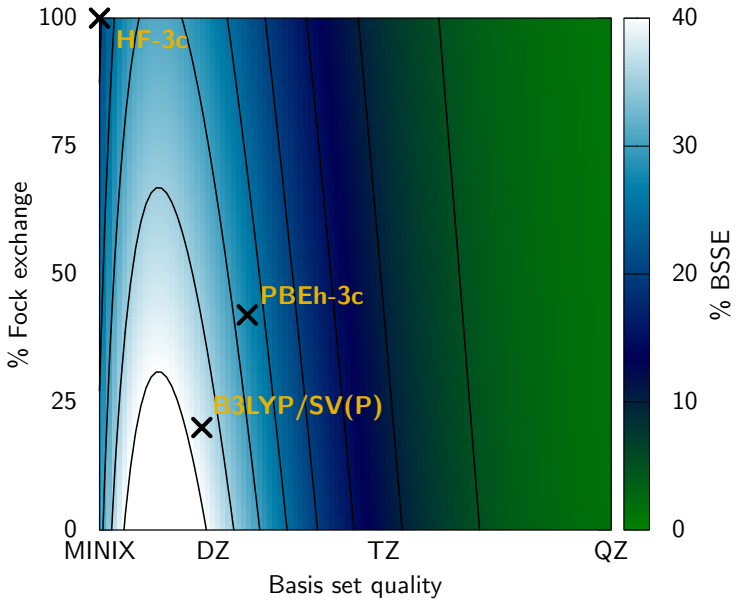


Figure 3.2.: Contour plot of the relative BSSE as a function of the basis set size (minimal, DZ, TZ, QZ) and the amount of Fock-exchange. An interpolating function (equation 3.2) was fitted based on the S66 Boys-Bernardi counterpoise energies calculated for the functionals PBE, B3LYP, PBEh-3c, and HF.

The BSSE is most pronounced for medium sized BSs of double-ζ quality and can be more than 40% of the binding energy. In a minimal BS, the neighboring fragment has only few (even zero for rare gas atoms) virtual orbitals and the extension of the variational freedom is minor (small BSSE). In a CBS, the virtual space is huge, but the energy gain is

zero because it is already converged in the single fragment basis (no BSSE). For medium sized BSs, the increase in the number of virtual functions and the corresponding lowering of the energy can be substantial. In general, BSSE leads to overestimated binding energies and underestimated interatomic distances.

Similar to the formation of a complex out of monomers, one part of a molecule such as a functional group can borrow basis function from another nearby part. This leads to the concept of intramolecular BSSE (IBSSE).[173–175] A uniform and clear definition of the IBSSE is missing, but its influence on energetics and structures of molecules has been recognized. The BSIE is an inherent problem of any finite BS expansion. It leads to insufficient descriptions of physical effects such as Pauli repulsion, electrostatics and polarization and thus, often to a systematic lengthening of bonds[176]. In practice, BSSE and BSIE are not strictly distinguishable but we will focus on the effects due to BSSE in the following.

Though one should try to minimize the BSSE (corresponding to the 'green areas' in Figure 3.2), a small BSSE is not a sufficient criterion for a good basis set. A minimal BS for instance has a relatively small BSSE, but cannot describe certain physical effects like polarization well. Furthermore, additional basis functions do not automatically lead to a more complete basis. They need to have the proper shape, which is a non-trivial requirement and basis set optimizations have been carried out for decades. We typically find the Ahlrichs sets optimal for molecular SCF type calculations and they have been only slightly adjusted and optimized for composite methods like PBEh-3c.[177] Similarly, the amount of HF exchange should not be increased too much to lower the BSSE because this would eliminate the account of important (static) electron correlation effects. The correct electron density can be best reproduced with a medium amount of HF exchange (about 20% to 50%), but other options (GGA or plain HF) can have advantages, too.

The second major shortcoming of common HF and (semi-local) DFT approximations is the inherent lack of a correct description of the London dispersion energy. For large interatomic distances > 4.5 Å the interaction between atoms or comparably unpolar molecules is dominated by long-range correlation effects, called London dispersion. This type of interaction has a $-C_6/R^6$ distance dependence and are not included in any semi-local exchange-correlation functional. Modern density functionals exist[100,178], which include correlation effects in the medium distance regime (2.5 to 4.5 Å) to a strongly varying degree but they do not provide the correct asymptotic behavior. Density functionals with a non-local correlation kernel exist and have been shown to yield reasonably good geometries and reasonably accurate binding energies[179–181]. While these special functionals can in principle also be evaluated in small basis sets, this combination is rarely applied and in the present review we focus on inherently more efficient methods.

The reason why small BS DFT (or HF) calculations like B3LYP/6-31G* can perform

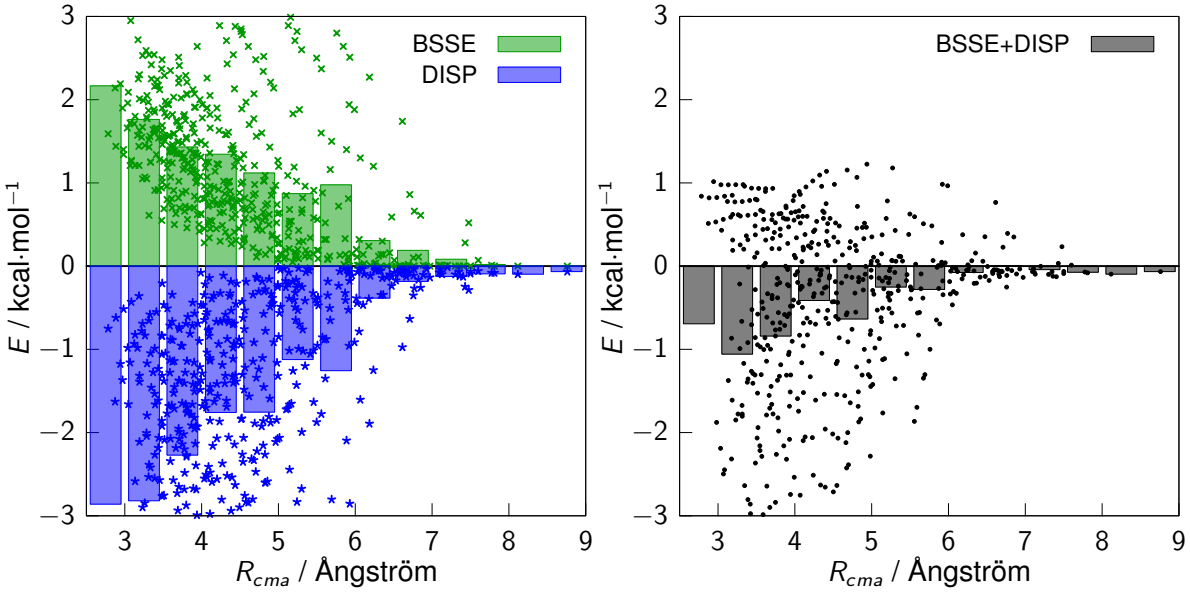


Figure 3.3.: Distance behavior of the dispersion energy and BSSE as calculated with second-order DFT-SAPT and the Boys-Bernadi counterpoise scheme, respectively, for the PBE0/SV(P) method on the S66x8 molecular dimers. The individual energies with integrated contributions in 0.5 Å bins (bars, left) and the possible error compensation if both contributions are neglected (bars, right) are shown. The crosses refer to the individual values for the 528 complex geometries. Note that while the BSSE is a negative quantity which has the same sign as the stabilizing dispersion energy, the plotted CP correction is positive (repulsive).

surprisingly well is immediately recognized when looking at the two largest error sources and their (partial) compensation. The first one is the BSSE which leads to too strongly bound complexes while the second flaw is the missing London dispersion energy resulting in too weak interactions. The prerequisite for a favorable error compensation is that dispersion and BSSE are of similar magnitude in a sufficiently large distance regime. However. this does not hold in general due to the fundamentally different functional dependence of BSSE and dispersion with respect to the distance separation, (exponential vs. R^{-6}) which is highlighted in Figure 3.3. We have calculated the dispersion and BSSE contribution for the S66x8 set (66 molecular dimers at eight different center of mass distances) with DFT-SAPT[182,183] and the Boys-Bernadi method, respectively, at the PBE0/SV(P) level. While the two contributions roughly cancel each other on average (bars, right plot), the individual values for the complexes have a significant scatter showing that either dispersion or BSSE can dominate. It is clear that systematically accurate results can not be obtained if both contributions are not properly included.

If we assume that BSSE and London dispersion effects cancel precisely at the equilibrium distance of the stacked benzene dimer, this can not hold for non-equilibrium distances (compare with Figure 3.4). Thus, for reliable results one needs to correct for

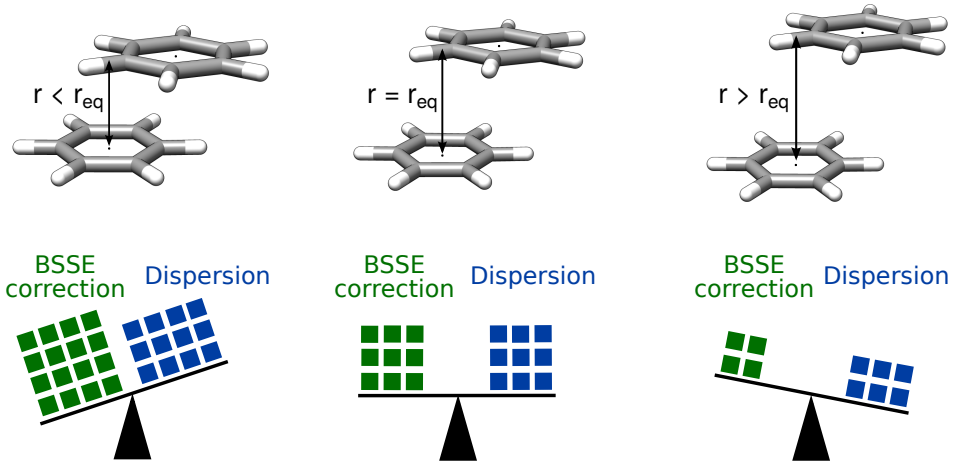


Figure 3.4.: Sketch of the error compensation between BSSE and dispersion. In this example for the noncovalently bound benzene dimer we assume that they are balanced at the intermolecular equilibrium distance. This is not the case for smaller or larger distances due to the different distance dependence of BSSE and dispersion. The BSSE decays exponentially whereas dispersion decreases slower with R^{-6}.

both dispersion and BSSE at the same time. In the following section we will present an overview over existing methods that were designed for that purpose and are used in combination with HF or DFT.

3.3. Methods treating dispersion and BSSE

One possibility to include London dispersion effects in DFT calculations is to correct the long-range interaction by atom-centered potentials, so called dispersion correcting potentials (DCPs)[184–186]. Though the correct physical terms leading to the London dispersion interaction (zero point energy of coupled frequency dependent polarizabilities) cannot be described by DCPs, their mathematical form together with parameter adjustment can empirically capture attractive dispersion-like forces to a rather high degree. The DCPs designed by DiLabio *et al.* resemble traditional effective core potentials (ECPs) and are similar to the earlier plane-wave approach proposed for periodic DFT by Lilienfeld *et al.*[187,188]. The general idea is to use a set of reference data and fit the interaction which is not covered by the density functional into additional atom-centered potentials. These potentials need a high degree of flexibility and should distinguish atoms in different hybridization states.

Typical potentials $U_l(r)$ are composed of atom-centered Gaussian-type functions and

have the following form:

$$U_l(r) = r^{-2} \sum_{i=1}^{N_l} c_{li}\ r^{n_{li}}\ e^{-\xi_{li} r^2} \tag{3.3}$$

where l is the angular momentum, N_l corresponds the number of Gaussian functions, n_{li} is the power of r (electron-nucleus distance), c_{li} is the coefficient of the Gaussian function, and ξ_{li} is its exponent.

For B3LYP-DCP[185] the n_{li} are fixed to the value of two and the exponents and coefficients were optimized for a set of 16 noncovalently bonded dimers. Two sets of DCPs have been developed: one intended for use with a CP correction and one for use without. The latter also compensate for the BSSE to a significant extent. Thus, BSSE and dispersion are treated simultaneously but the correct R^{-6} asymptotic behavior of the London dispersion is not met (though this is in principle possible by a complete expansion with all orders in N_l).

The advantage of this approach is its simple implementation and that DCPs can be used with any computational chemistry program package that can handle ECPs. The corresponding exponents and coefficients for each element are provided in the input files and no modifications of the program themselves are necessary. Further, with the use of DCPs or related approaches, dispersion effects are included on the electronic structure level and the electron density can adjust to these effects. A disadvantage in particular for large systems is that the incorporation of DCPs into the self consistent field (SCF) procedure often increases the number of cycles needed for convergence and hence the overall computational time. Another drawback is that a DCP has to be fitted for each element in combination with a specific density functional and a given BS. For each element one has to gather enough reliable reference data, which can be difficult when aiming at an extension to heavier elements. Currently, for small DZ basis sets, which are the focus of this review article, DCPs are available for the elements H, C, N and O and the B3LYP functional[185,186]. They are suitable for the 6-31+G(d,p) basis set or larger, but the use of 6-31+G(2d,2p) is recommended. The first DCP for the carbon atom needed to be revised due to the too large exponents that hampered the correct description of C-C bond breaking or bond formation and with the revised DCP, noncovalent interactions and covalent chemical reactions are described with similar accuracy[186]. We will refer to and use these improved DCPs throughout this article.

In the original publication the performance of B3LYP-DCP/6-31+G(2d,2p) for noncovalent interactions was tested on several benchmark sets and we will give here only some examples. Its accuracy for the S66[172] test set of small noncovalently bound dimers is excellent. The mean absolute deviation (MAD) of the binding energies compared to the reference is only 0.19 kcal mol^{-1}. For comparison, B3LYP-D3(BJ) with the quadruple-ζ basis set def2-QZVP yields an MAD of 0.28 kcal mol^{-1} [189]. For the HSG[190] set of 21

dimers and trimers which are present in the complex of the inhibitor indinavir and HIV-II protease the performance is also encouraging. The MAD compared to the revised reference values (HSG-A[191]) is 0.16 kcal mol^{-1}. Further, B3LYP-DCP/6-31+G(2d,2p) was applied to the S12L[46] set of supramolecular complexes. One out of the 12 complexes contains a Cl atom and another one involves Fe. As no DCPs exist for Cl and Fe these atoms were left uncorrected. The final MAD for the (reduced) S12L set is 2.6 kcal mol^{-1}. This result is similar to those obtained with PBE-D3 or PBE-NL in combination with def2-QZVP (2.1 and 2.3 kcal mol^{-1}, respectively)[192]. Overall, these examples show, that B3LYP-DCP/6-31+G(2d,2p) as a method on the double-ζ level can provide results of quadruple-ζ quality.

Its good performance was confirmed by Goerigk who compared B3LYP-DCP, B3LYP-NL and B3LYP-D3(BJ) in combination with the 6-31+G(2d,2p) basis set for noncovalent complexes, relative energies of conformers, basic properties and reaction energies[34]. An overall comparison revealed B3LYP-NL as the most robust and accurate approach, closely followed by B3LYP-D3. However, for these two methods the influence of BSSE effects on the binding energies of noncovalently bound complexes can be larger than it is the case for B3LYP-DCP. Further, it was verified that the revised DCP for carbon actually does improve the overall performance, though the change for electron affinities and ionization potentials is negligible.

Recently, the DCP scheme was coupled to the atom-pairwise D3 dispersion correction (*vide infra*) for the BLYP functional and the 6-31+G(2d,2p) basis set[193]. In this BLYP-D3-DCP approach the exponents of the DCP tend to be larger than those for the ones developed previously. Thus, they mostly have an impact on the electron density close to the nuclei and mainly influence the covalently bonded parts. This is reflected in the large improvement for barrier heights compared to BLYP-D3 but only small enhancements for noncovalent interaction energies. This result indicates that also typical GGA problems like the self-interaction-error (SIE) can be corrected with DCP (see also[194,195]).

A different approach is the combination of a dispersion correction and a CP correction which are developed independently from each other, but which are simultaneously employed in a calculation. For the treatment of London dispersion we use our efficient atom-pairwise D3(BJ) correction[36,37] that can simply be added on top of a converged standard DFT or HF calculation. For reviews and overviews of other state-of-the-art dispersion corrections see Refs. [32–35]. Within the D3(BJ) scheme the energy contribution is calculated as a sum over all atom pairs AB

$$E_{disp}^{D3} = -\frac{1}{2}\sum_{A,B}\sum_{n=6,8} s_n \frac{C_{AB}^n}{R_{AB}^n + f(R_{AB}^0)^n} \tag{3.4}$$

where, C_n^{AB} denotes the averaged coordination-number dependent (isotropic) n^{th} order dispersion coefficient for each atom pair AB. The order n equals 6 and 8, introducing a R_{AB}^{-6} long-range and a R_{AB}^{-8} medium-range term. The s_n are the global scaling factors. For common density functionals s_6 is usually set to unity to ensure the correct asymptotic behavior, whereas s_8 is optimized for each functional. $f(R_{AB}^0)$ is the damping function as introduced by Becke and Johnson[133,134]

$$f(R_{AB}^0) = a_1 R_{AB}^0 + a_2 \tag{3.5}$$

with the fitting parameters a_1 and a_2, and the cut-off radii $R_{AB}^0 = \sqrt{C_8^{AB}/C_6^{AB}}$. For simplicity we will refer to D3(BJ) (which is the current default for the method) as D3 in the following. An Axilrod-Teller-Muto (ATM) type three-body (dipole-dipole-dipole) term is also available in the D3 code including its analytical derivatives.[138,139] The importance of many-body dispersion interactions has been recently analyzed by various groups,[135–137] but is not in the focus of this review.

The D3 dispersion correction can in principle be combined with any BSSE correction, e.g., with the standard BB-CP procedure. In this scheme, however, the computational cost quickly increases for larger complexes because full BS calculations for the fragments have to be conducted. If each atom is considered as an individual fragment, one can define an atomic counterpoise correction (ACP)[196] as done by Jensen. The ACP(x) correction $\Delta E^{ACP(x)}$ is expressed as a sum over all atoms A

$$\Delta E^{ACP(x)} = \sum_A E(A)_a - E(A)_{as} \tag{3.6}$$

where a denotes the regular basis set and as is a subset of a which always includes the regular basis function on A. For the intramolecular case this subset further includes all basis functions from atoms x bonds apart, and for the intermolecular case it contains all basis function of the other monomer. When all basis functions of the whole system are included in the subset, the ACP(1) correction equals the CP^{aa} correction published earlier by Galano and Alvarez-Idaboy.[197] These BSSE corrections have a highly reduced computational cost and the advantage to treat inter- and intra-molecular effects conceptually on the same level. Unfortunately, these approaches lack the availability of nuclear gradients. Therefore, we recently developed a geometrical counterpoise correction (gCP)[47] that solely depends on the molecular geometry. It provides a fast, conceptually simple but physically reasonable energy and gradient correction for the BSSE in large molecules and condensed phase systems.

Within the gCP scheme the difference in atomic energy E_A^{miss} between a large, nearly complete BS and the target basis (here DZ) is calculated (and tabulated) for each atom

at the HF or B3LYP level and used as a measure for the BS incompleteness. The E_A^{miss} are then multiplied with a decay function depending on the interatomic distance R_{AB} and summed up over all atom pairs AB

$$E_{BSSE}^{gCP} = \sigma \sum_{A} \sum_{A \neq B} E_A^{miss} \frac{exp\left(-\alpha (R_{AB})^{\beta}\right)}{\sqrt{S_{AB} N_B^{virt}}} \tag{3.7}$$

where α, β and σ are functional and BS specific fitting parameters. As the density has an exponential tail, the decay function is exponential. Due to the strong dependence of the BSSE on the charge density overlap in SCF methods, this function is normalized by the square-root of the Slater-overlap S_{AB} times the number of virtual orbitals N_B^{virt} on atom B. The overlap integrals S_{AB} are evaluated over single s-type orbitals centered on each atom using optimized Slater exponents and weighted by the last fitting parameter η. The fit was performed for HF and B3LYP together with the target basis set on the S66x8 test set[172]. Standard BB-CP corrected interaction energies for the respective method were employed as reference values. The accuracy gained by a re-fit for different density functionals is negligible and thus, the use of the B3LYP parameters is recommended for common GGA or hybrid functionals.

One advantage of the D3-gCP combination is its availability for almost all elements in the periodic table, and the existence of analytical nuclear gradients. Further, the scaling behavior with system size is low and the computational pre-factor is small. This results in very fast computations even for thousands of atoms. Another benefit is that the corrections can simply be added on top of any DFT or HF calculation without need for a specific implementation into a program package. A drawback is the semiempirical character of both corrections and thus, the need for a parameter fit for every functional in case of D3 and each functional/BS combination in case of gCP. However, as mentioned before, the gCP dependence on the functional was found to be negligible and hence only adjustments for each basis set and for HF or DFT have to be made. Further, the corrections do not depend on the density and thus, the electronic structure is not directly affected, though, an indirect effect due to the altered geometry is present.

Note, that in general the gCP scheme can be combined with any dispersion correction. One example is a recent publication by Yoshida *et. al*, who used gCP for HF/6-31G(d) together with their own dispersion correction to describe the HIV-1 protease and its potent inhibitor KNI-10033[198]. The good performance of DFT-D3-gCP/DZ and HF-D3-gCP/DZ for non-ncovalent interactions was already noted in the original gCP publication. The gCP correction is able to provide a reasonable estimate for the intermolecular BSSE with an error of 10-30%. For the S22 benchmark set[199] e.g. PW6B95-D3-gCP/def2-SVP yields an MAD of 0.84 kcal mol^{-1} for interaction energies. In case of B3LYP/6-31G*

the MAD can be reduced from 2.67 to 0.88 kcal mol^{-1} upon application of both the D3 and gCP correction. Geometry optimizations of the S22 complexes showed that B3LYP-D3-gCP/6-31G* as well as HF-D3-gCP/SV reproduce the reference structures well. In case of 9-helicene the non-bonded C-C distances can be accurately computed with HF-D3-gCP/SV within a few pm[47]. Somewhat unexpectedly, of all various method/basis set combinations tested, HF-D3-gCP/MINIS performs particularly well for noncovalent interactions.

In a recent publication the shortcomings of the B3LYP/6-31G* model chemistry, as explained in the previous section, were analyzed and it was shown that D3-gCP can account for the major deficiencies and that B3LYP-D3-gCP/6-31G* yields reasonably accurate thermochemical results[145]. Benchmark calculations on the general main group thermochemistry, kinetics and non-ncovalent interactions meta-database GMTKN30[86] showed a statistical improvement when both corrections are used. The weighted MAD decreased from 8.8 (B3LYP/6-31G*) to 6.9 kcal mol^{-1} (B3LYP-D3-gCP/6-31G*). It was statistically confirmed that the partial error compensation of missing dispersion and BSSE in plain B3LYP/6-31G* is unsystematic and depends on the chemical nature of the system at hand. The improvement gained with the D3-gCP scheme is largest for systems that exhibit noncovalent interactions but reaction energies and barrier heights are also improved.

Goerigk and Reimers used DFT-D3-gCP/DZ and HF-D3-gCP/DZ for geometry optimizations of several test sets which aim at describing important interactions in protein structures[200]. Various functionals as well as HF in combination with different DZ basis sets were employed for the P26 test set[201], in order to investigate their performance for conformers of five tri-peptides containing aromatic side chains. For the 6-31G* basis without any correction as an example, structural RMSDs around 0.5 Å are observed. When only gCP is employed the RMSDs rise, and with solely the D3 correction the RMSDs drop significantly. When the combined D3-gCP scheme is used the RMSDs decrease to values of about 0.15 Å, which are slightly higher than those with the D3 correction only. It seems that in this specific case without gCP a fortunate error compensation occurs which, however, does not hold in general as discussed above.

Martinez *et. al* showed that uncorrected DFT or HF with DZ basis sets can yield good geometries for small proteins[202]. They compiled a set of 58 proteins with up to 35 residues (up to 600 atoms) and compared their results to experimental X-ray or nuclear magnetic resonance derived structures. The ab initio methods HF and ωPBEh are able to provide geometries of the same quality as highly parametrized force fields and are consistently better at reproducing experimental structures for proteins with disordered regions, judged by standard health metrics.

Reimers *et. al* optimized a portion of an ensemble of conformationally flexible lysosome

structures by a divide-and-conquer approach and compared their results to X-ray crystallography data[146]. The functionals BP86 and B3LYP as well as HF were employed together with the 6-31G* basis set and in combination with the D3-gCP scheme. Regarding amd all atom RMSD and the R-factor, the best and most consistent structures are obtained when both the D3 and the gCP correction are used. Compared to the uncorrected methods, employing only D3 gives similar results and only gCP yields worse values. This observations resemble the ones made for small peptides[200] and again show that one cannot rely on error compensation effects.

Extension of the gCP correction to periodic HF/DFT calculations enables the use of the D3-gCP scheme for molecular crystals[203]. The corrections were applied to PBE and B3LYP for the X23 molecular crystal test set[204] and reduce the MAD of the sublimation energies significantly by more than 70 % and 80 %, respectively, to small residual MADs of about 2 kcal mol^{-1} (corresponding to 13 % of the average sublimation energy). Further, variation of the interlayer distances for graphite yielded a potential energy surface that is very close to the converged basis set reference and agrees very well with experimental stacking distances.

Up to now, D3 and gCP were fitted independently of each other, but applied at the same time in a calculation. We introduced two composite methods, that make also use of these corrections but which were fitted or adjusted in the presence of each other and thus are suggested as one composite approach with a fixed basis set. As we noticed the good performance of HF-D3-gCP/MINIS for noncovalent interactions during the development of the gCP correction, we proposed HF-3c, a minimal basis set Hartree–Fock method with three atom-pairwise corrections: D3, gCP, and an additional term, which corrects for short-range basis (SRB) set incompleteness effects[205]. The six parameters of the gCP and D3 correction terms were fitted together on the S66 test set and were kept constant in the subsequent fitting procedure of the third SRB term. This composite method corrects for both dispersion and BSSE and is suggested as an alternative to semiempirical methods or DFT, in particular when SIE is acute. HF-3c yields reasonable noncovalent interaction energies and good geometries of small organic molecules, as well as supramolecular complexes and small proteins[200,205,206]. As this review focuses on DZ basis sets, we will not discuss this method further.

A related composite approach is our recently developed PBEh-3c method, a global hybrid functional with a DZ basis set, that is meant to fill the gap between existing semiempirical methods or HF-3c and large basis set DFT with respect to the cost-accuracy ratio[177]. The term '3c' indicates its relation to HF-3c, and the corrections are a slightly modified gCP, D3, and minor modifications to the def2-SV(P) BS (dubbed def2-mSVP) for boron to neon in order to ensure consistent bond lengths for all elements. PBEh-3c yields accurate geometries which was verified for small molecules as well as medium

sized molecules, noncovalently bound complexes, and molecular crystals. The overall deviations from reference structures are tiny and practically of MP2/def2-TZVPP quality while the geometries are obtained at a much lower computational cost (speedup of about 50-100). All other DFT/small BS methods tested yielded larger deviations. For the S22 set of noncovalent complexes PBEh-3c agrees well with the MP2 reference geometries, the mean deviation (MD) for intermolecular center-of-mass distances is only 3 pm. For molecular crystals, the PBEh-3c accuracy for geometries in the X23 and ICE10[207] sets approaches TPSS-D3/'large BS' results. The mean absolute deviations in the computed unit cell volume are 2.7 % and 5.0 %, respectively, for X23 and ICE10. Although PBEh-3c was mainly designed for the computation of structures it yields reasonable results for thermochemistry, barrier heights and general noncovalent interactions. Clearly, due the small BS the accuracy of dispersion-corrected hybrid DFT in a QZ basis set can not be reached.

3.4. Comparison of methods for noncovalently bound systems

In the following we will compare the performance of the various discussed methods for some exemplary noncovalently bound systems. We chose HF, HF-D3-gCP, B3LYP, B3LYP-D3-gCP, B3LYP-DCP, M06-2x, and PBEh-3c. The def2-SV(P) basis set will be applied in all cases (modified for PBEh-3c) except for B3LYP-DCP where the 6-31+G(2d,2p) basis will be used. An overview of the capability of these method to treat dispersion and BSSE is provided in Table 3.1.

Table 3.1.: Overview of the applied methods and their capability to treat BSSE and dispersion. B3LYP-DCP will be used with the 6-31+G(2d,2p) basis set, all other methods with the def2-SV(P) basis set (modified in case of PBEh-3c).

method	**BSSE correction**	**dispersion correction**
HF	no	no
HF-D3-gCP	yes	yes
B3LYP	no	no
B3LYP-D3-gCP	yes	yes
B3LYP-DCP	yes	(yes)[a]
M06-2X	no	(yes)[b]
PBEh-3c	yes	yes

[a] BSSE and dispersion are treated together in one ECP leading to the wrong asymptotic behavior for the dispersion interaction. [b] The dispersion interaction has the wrong asymptotic behavior.

3.4.1. Noncovalent interaction energies

The accuracy for noncovalent interaction energies of the aforementioned methods is tested on several benchmark sets. We chose three sets for the interaction of small to medium sized systems (WATER27[208,209], S22[191,199], S66[172]), two sets for large and supramolecular systems (L7[144], S30L[206]), and two test sets for molecular crystals (ICE10[207], X23[204,210]). For each test set one exemplary system is depicted in Figure 3.5.

The WATER27 test set contains 27 neutral and charged water clusters with up to 20 water molecules. The S22 set consists of 22 noncovalently bound model complexes that show hydrogen bonding, dispersion interactions and mixed electrostatic-dispersion binding motifs. The S66x8 test set is similar to S22 but with less emphasis on nucleobases. Further, reference geometries and energies are provided at eight different distances of the monomers, which allows the extraction of the minimum of the intermolecular potential energy surface (PES) of a given method via an interpolation procedure. The reference energies for these three sets refer to the estimated CCSD(T)/CBS level of theory. For the S22 we use the revised values by Sherill *et al.*[191]. For the $(H_2O)_{20}$ complex contained in the WATER27 set we use the reference values computed on the incremental CCSD(T)(F12*)|MP2-F12+ΔMP2 level by Friedrich[209]. The L7 test set comprises seven larger, mostly dispersion-stabilized complexes of organic molecules. We use the revised reference values on the estimated DLPNO-CCSD(T)/CBS* level of theory[211]. The S30L set is an extension of the S12L set[46,192], which was the first test set for large host-guest complexes. It contains 30 realistic host-guest complexes with charges from -1 to $+4$ and up to 200 atoms, featuring various typical noncovalent binding motifs like hydrogen and halogen bonding, $\pi-\pi$ stacking, nonpolar dispersion, CH$-\pi$, and cation-dipolar interactions. The reference association energies are back-corrected values from experimentally measured association free energies. ICE10 includes ten ice polymorphs and X23 compiles molecular crystals that show mainly van-der-Waals or hydrogen bonding or a mixture of these two interaction motifs. For these two sets the reference lattice energies were derived from experimental values which are further corrected for zero-point vibrational and thermal effects.

As the absolute interaction energies differ by almost three orders in magnitude, we give mean absolute relative deviations (MARDs in %) from the reference energies for all test sets and methods in Figure 3.6. Because of SCF convergence problems for some molecular crystals, the HF results for the periodic benchmarks were omitted. The values are color-coded as suggested by Martin[212] in order to provide an easy overview and the best two methods for each test set are highlighted.

As expected, the plain B3LYP functional or HF without any corrections cannot properly describe noncovalent interactions. Already for small systems contained in the WATER27,

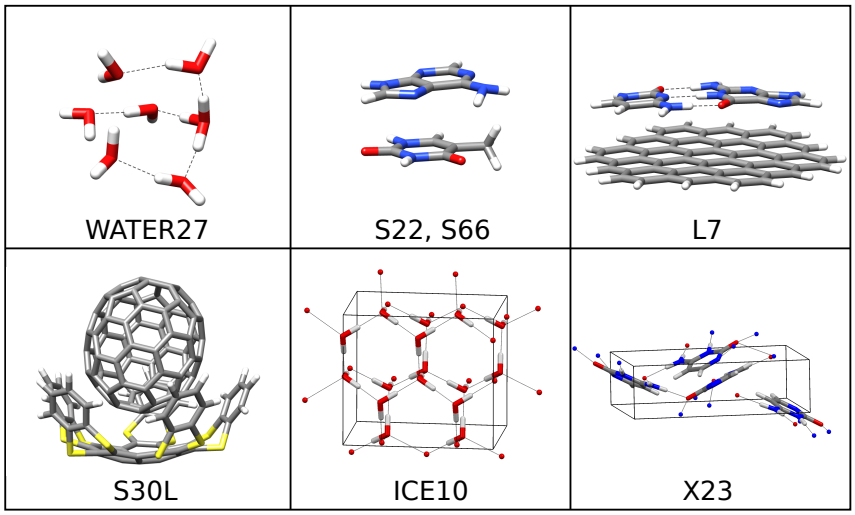


Figure 3.5.: Example systems for each test set. Hydrogen bonds are indicated by dotted lines.

S22, or S66x8 sets huge MARDs of 50 to 80% are obtained. In many cases these methods yield unbound complex states. The same is observed for the supramolecular test sets and the MARDs for L7 and S30L are even larger (80 to 160%). The performance of B3LYP for the ice polymorphs is similar to the molecular WATER27 set. For both the MARD is about 60 %. Surprisingly, the MARD for the X23 set of molecular crystals is with 34 % much smaller than the corresponding values for the S22 and S66x8 sets (about 60 %) and the performance is actually similar to M06-2X and B3LYP-DCP. For the mixed hydrogen bonded crystals, the error compensation between missing dispersion and neglected BSSE in plain B3LYP is rather good around the corresponding equilibrium geometry. While this explains the slightly smaller error compared to the other test sets, this compensation does not hold for stronger hydrogen bonding (significant overbinding of the various ICE10 polymorphs due to dominant BSSE) nor for purely London dispersion bonded X23 systems (significant underbinding due to missing dispersion interaction).

When the D3 and gCP corrections are added, the plain HF and B3LYP results can be improved tremendously. In case of HF, no improvement is observed for the WATER27 set, but for all others the MARD for HF-D3-gCP drops to 15 to 25 % which is very reasonable for such a simple method. For B3LYP-D3-gCP the enhancement for WATER27 and ICE10 is much smaller than for the other sets, but still, the MARD is reduced from 60 % to about 35 %. Very good results are obtained for the S22, S66x8, L7, S30L and X23 test sets which have MARDs of 10 to 14 %. Compared to B3LYP-D3 with the large def2-QZVP basis set, the MARD for B3LYP-D3-gCP/DZ on the S22 set is doubled (6.2 %[86] vs. 13 %). For the large supramolecular complexes B3LYP-D3-gCP/DZ yields equally good values or even better results than B3LYP-D3/QZ. The MARD for the S30L set is 13.8 % for B3LYP-D3-gCP/DZ and very similar for B3LYP-D3/QZ (13.2 %[206]). For the L7 set

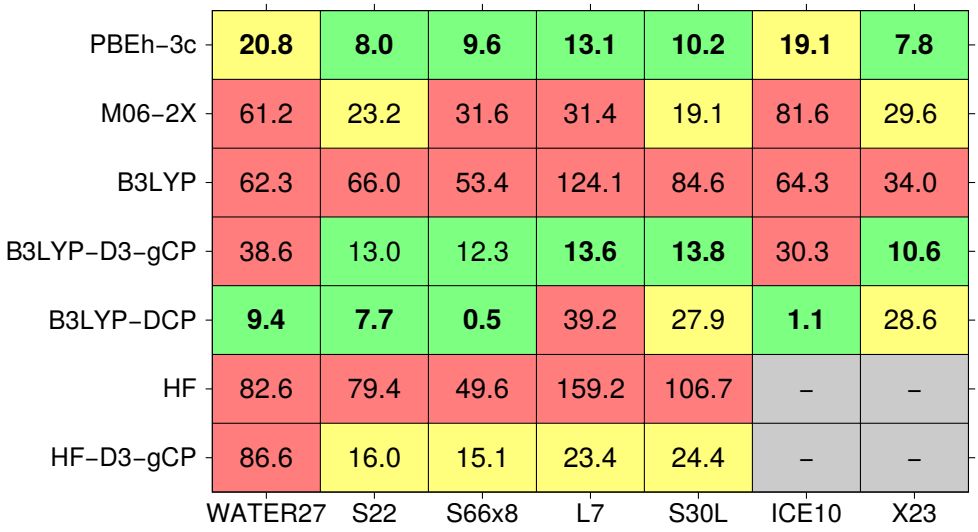


Figure 3.6.: Mean absolute relative deviations (MARDs, in %) for different methods compared to the reference values for several test sets. MARDs below 15 % are color-coded in green, those below 30 % in yellow and those higher than 30 % in red. For each set the two best performing methods are highlighted. PBEh-3c includes the ATM three-body dispersion term by default, for B3LYP-D3-gCP and HF-D3-gCP it was included for the large systems (L7 and S30L test sets). In case of B3LYP-DCP two systems of the S30L were omitted due to missing functions of the 6-31+G(2d,2p) basis set for iodine and for X23, eight systems had to be disregard due to SCF convergence problems.

B3LYP-D3-gCP/DZ yields an MARD of 13.6 %, which is less than half of the value for B3LYP-D3/QZ (32.5 %[144,211]).

If DCPs are used for B3LYP instead of the D3-gCP correction the behavior is very different. First we note the extraordinary good performance for the water containing systems. The MARD of B3LYP-DCP for WATER27 is 9.4 % and for ICE10 1.1 %. Even with large basis set dispersion-corrected DFT calculations it is difficult to reach this accuracy. This can partially be attributed to the basis set (6-31+G(2d,2p)) which contains two sets of additional polarization functions as well as a diffuse set of *sp*-functions on non-hydrogen atoms which is known to be important for these systems[86,208,209]. Therefore, the number of basis functions per atom is more comparable to a TZ basis and much larger than in def2-SV(P). The B3LYP-DCP results for S22 and S66x8 are also very good and the MARDs of 7.7 and 0.5 %, respectively, are the lowest ones reported here. The MARD for S22 is very close to the already mentioned B3LYP-D3/QZ result (6.2 %[86]). For large systems, however, the performance of B3LYP-DCP deteriorates significantly. The MARDs for the S30L and L7 sets are 39.2 and 27.9 %, respectively. Although these values are about three times smaller than those for plain B3LYP, they are still about three times larger than for B3LYP-D3-gCP. For the X23 set there is no improvement compared to plain B3LYP. Both MARDs are about 35 %, which is again three times as large compared

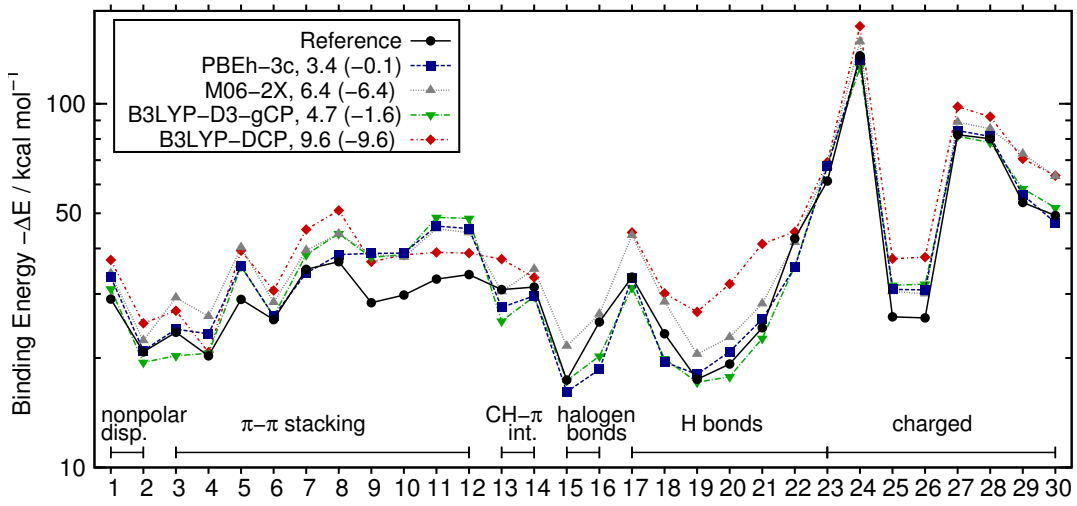


Figure 3.7.: Comparison of the PBEh3c, M06-2X, B3LYP-D3-gCP, and B3LYP-DCP binding energies with reference values for the S30L test set. The complexes are sorted according to the most prominent type of interaction. In case of B3LYP-DCP the complexes **15** and **16** were omitted due to missing functions of the 6-31+G(2d,2p) basis set for iodine. The MADs and MDs in parentheses are provided for each method in kcal mol^{-1}.

to B3LYP-D3-gCP.

Finally, we discuss the recently published composite method PBEh-3c[177] Its overall performance for all test sets is very good and consistent accuracy for small as well as large complexes is evident. The MARDs for the S22, S66x8, L7, S30L and X23 sets are found to be in the range of 8 to 13 %. The performance for WATER27 and ICE10 is worse, the MARDs are about 20 %. This indicates that the applied corrections cannot repair the higher basis set requirements in condensed hydrogen bonded systems compared to only medium polar dimers. But nevertheless, the similar MARDs for WATER27 and ICE10 as well as S22/S66x8, L7/S30L, and X23 show that PBEh-3c treats the noncovalent interactions in small, large and periodic systems with the same accuracy. PBEh-3c is always one of the two best performing methods on any test set. The others are either B3LYP-DCP (WATER27, S22, S66x8, ICE10) or B3LYP-D3-gCP (L7, S30L, X23).

As this article mainly focuses on large systems, a closer look to the supramolecular complexes of the S30L set is appropriate. The association energies ΔE range from -17.4 kcal mol^{-1} for the halogen bonded complex **15** up to -135.5 kcal mol^{-1} for the doubly positive charged complex **24**. Figure 3.7 shows a comparison of the ΔE values for PBEh-3c, M06-2X, B3LYP-D3-gCP and B3LYP-DCP with the reference values.

As one can easily see, B3LYP-DCP and M06-2X exhibit large systematic overbinding as indicated by MDs of -9.6 and -6.4 kcal mol^{-1}, respectively. The largest errors for B3LYP-DCP are observed for the charged systems **23** to **30** and range from -11 to -27 kcal mol^{-1}. Somewhat surprisingly, the errors are also large (>-10 kcal mol^{-1}) for most of the hydrogen

bonded systems (**17**, **19** to **21**). As seen before B3LYP-DCP performs exceptionally well for the WATER27 and ICE10 sets, the hydrogen bonded dimers in S22/S66x8, but seems to fail for hydrogen bonds in these supramolecular complexes. Further, the errors are large (-8 to -14 kcal mol^{-1}) for some of the $\pi-\pi$-stacked systems (**5**, **7** to **10**). The complexes **25** and **26** which also exhibit $\pi-\pi$-stacking as major interaction show a similar error. Obviously, for the small systems (on which the DCPs are fitted) the description of dispersion and the compensation for BSSE is reasonable and accurate results can be obtained. For these large supramolecular complexes the balancing of dispersion effects and BSSE is different which is difficult to describe by a correction potential lacking the correct physics. As explained above, missing dispersion results in too weak and BSSE in too strong bound complexes. B3LYP-DCP overestimates the binding energy for all host-guest complexes and thus, the remaining BSSE seems to be the major error source. Whether the diffuse functions in the 6-31+G(2d,2p) basis play an additional negative role in the larger systems due to a more long-ranged BSSE is currently not clear.

The largest errors for plain M06-2X (-10 to -19 kcal mol^{-1}) are obtained for the complexes **5**, **9**, **11**, **12**, **17**, **29** and **30**. Thus, M06-2X seems to have less trouble to accurately describe hydrogen bonded complexes than to reproduce the reference values for the $\pi-\pi$-stacked systems. As the MARD for S30L is similar to that of S22 and even better than for S66x8 (Figure 3.6), the incorrect asymptotic treatment of the London dispersion seems not to be a major error source. Much more problematic is the unaccounted BSSE and therefore, binding energies are overestimated. When a TZ basis is used, the MAD drops to 2.5 kcal mol^{-1} and the MD is just 1.4 kcal mol^{-1}, indicating underbinding due to the missing long-range dispersion contribution[206].

When dispersion and BSSE are both independently accounted for, as in B3LYP-D3-gCP and PBEh-3c, the errors decrease substantially. For B3LYP-D3-gCP the largest errors of -12 to -21 kcal mol^{-1} are observed for complexes **9** to **12**. For PBEh-3c the complexes **11** to **13**, **22** and **24** show the largest errors of 6 to 9 kcal mol^{-1}. B3LYP-D3-gCP and PBEh-3c reach MDs of -1.6 to -0.1 kcal mol^{-1}, respectively, indicating small to almost no systematic overbinding. Compared to M06-2X and B3LYP-DCP, the MAD values for B3LYP-D3-gCP and PBEh-3c are with 4.7 kcal mol^{-1} and 3.4 kcal mol^{-1}, respectively, much lower. For comparison, the previously best results for S30L were obtained with PW6B95-D3/def2-QZVP, which yields an MAD of 2.4 and an MD of -0.1 kcal mol^{-1}[206]. B3LYP-D3/def2-QZVP is one of the worse performers at the large BS level and has an MAD of 4.1 and an MD -2.7 kcal mol^{-1}[206]. Thus, B3LYP-D3-gCP is able to provide close to QZ quality results but at a small fraction of computational cost, and PBEh-3c almost approaches the accuracy of PW6B95-D3/QZ. These examples show clearly how important it is in large systems to properly and consistently treat both, dispersion and BSSE.

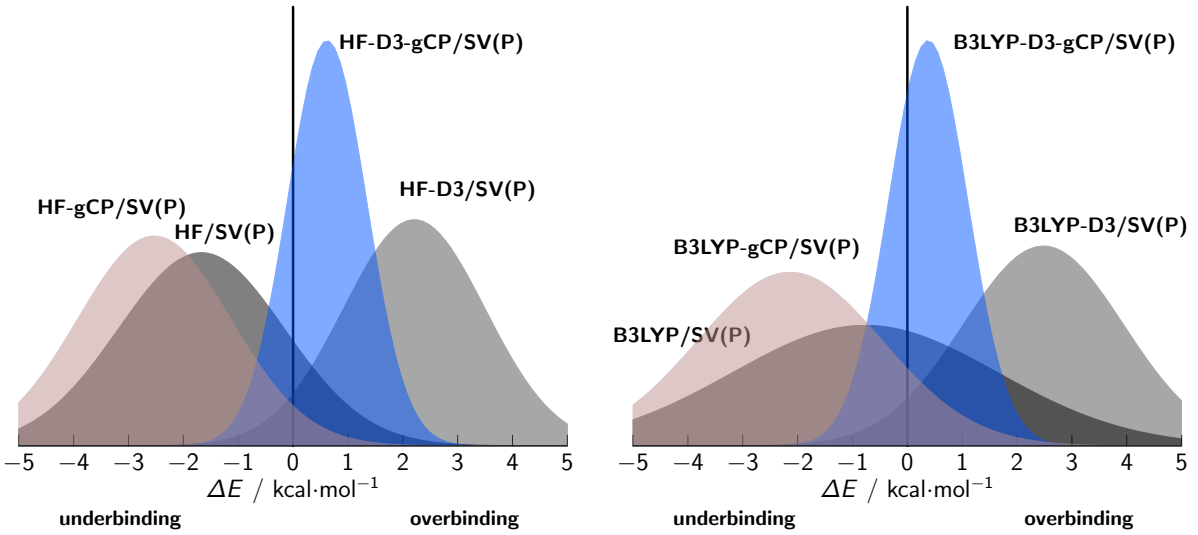


Figure 3.8.: Error statistics of S66 equilibrium binding energies for corrected and uncorrected HF and B3LYP in a SV(P) basis set converted into normal error distributions.

In Figure 3.8, we summarize the different contributions to the binding energy of the S66 dimers (minimum extracted from S66x8 potentials) for the HF and B3LYP methods evaluated in a def2-SV(P) basis set with and without correction schemes. We show the statistics of the deviations to CCSD(T) references as normal error distributions. The behavior of HF and B3LYP mean-field methods is very similar, which is typical for purely noncovalent interactions. Without any corrections, the error spread is large (broad distribution) with a slight systematic underbinding. When only the gCP correction is applied, the error spread decreases, but the underbinding is increased. The sole application of the D3 correction leads similarly to a smaller error spread and a systematic overbinding. Only the combination of both schemes leads to an excellent agreement with the reference data with MAD of 0.8 kcal mol^{-1} and 0.6 kcal mol^{-1} for HF-D3-gCP and B3LYP-D3-gCP, respectively.

3.4.2. Structures of noncovalently bound systems

In the following, the accuracy of the methods for optimized structures of noncovalent complexes is evaluated. As examples for small systems we chose the S22 and S66x8 sets that were already employed for the interaction energies as well as the P26[201] set, which contains different conformers of four peptides. For S22 and P26 the reference geometries were calculated on the MP2/TZ level of theory and the root mean square deviation (RMSD) of the heavy atom positions as well as the deviation of the intramolecular center-of-mass distance of the monomers are used as performance measures. For the S66x8 the PES is used to determine the optimal intramolecular center-of-mass distance of the

monomers and therefore the CCSD(T)/CBS level of theory is the reference as conducted similarly before.[177,213] For the supramolecular systems we face the problem that there are no reference geometries available. The L7 and S30L systems were optimized on the TPSS-D3/TZ level of theory, which is certainly a good choice but not accurate enough to serve as a reference. In order to show exemplary the influence of London dispersion and BSSE in large supramolecular complexes, system **5** from S30L and the **phe** complex from L7 were re-optimized on the TPSS-D3/def2-QZVP(-g/f) level and these structures are used for comparison.

For the molecular crystal test sets ICE10 and X23 the reference geometries refer to experimental X-ray data which are isotropically corrected for zero-point vibrational and thermal effects[177,207]. Here, we use the deviation of the unit cell volume as measure to evaluate the accuracy of the tested methods.

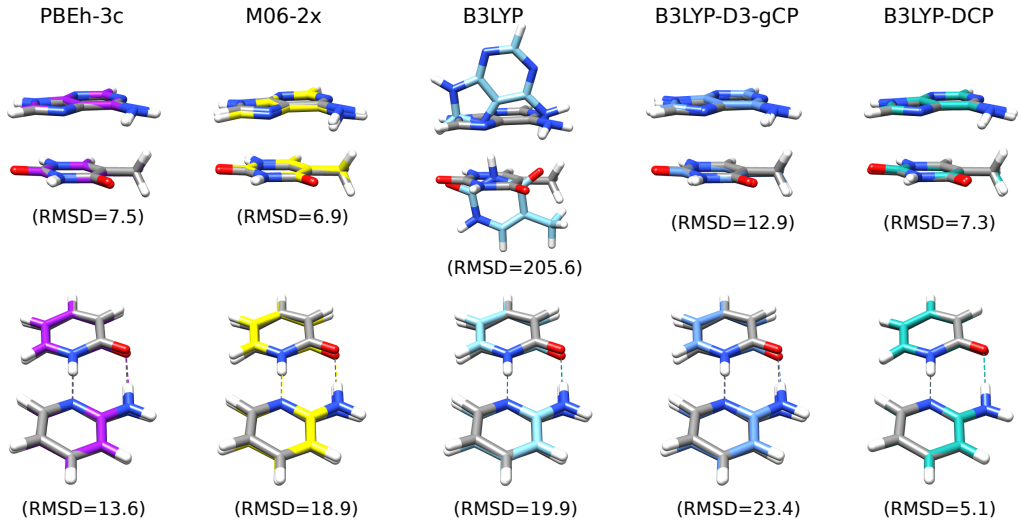


Figure 3.9.: Comparison of the PBEh-3c, M06-2X, B3LYP, B3LYP-D3, B3LYP-D3-gCP, and B3LYP-DCP geometries (colored) with the reference MP2/TZ structure (in gray) for the $\pi-\pi$ stacked cytosine-uracil base pair (top) and the hydrogen bonded cytosine-uracil base pair (bottom) from the S22 test set. Hydrogen bonds are indicated by dotted lines. The RMSDs for the heavy atom positions are given in pm.

Figure 3.9 depicts the geometries of two exemplary systems, the $\pi-\pi$ stacked and hydrogen bonded cytosine-uracil base pair, as obtained with the DFT methods in comparison with the reference structures. The results for HF and HF-D3-gCP are similar to B3LYP and B3LYP-D3-gCP and therefore, these geometries are not shown.

One can see immediately, that plain B3LYP can not even qualitatively correctly reproduce the $\pi-\pi$ stacked structure (Figure 3.9, top). As the London dispersion is completely missing the $\pi-\pi$ stacked dimer is not a minimum on the PES and the optimization leads to the hydrogen bonded structure. In contrast, all other four methods which include

Table 3.2.: Statistical data for the deviations of the intramolecular center-of-mass distances R_{CMA} from the reference values for the S22 and S66x8 test set and the mean RMSDs ($\overline{\text{RMSD}}$) for the heavy atom positions in case of S22 and P26 compared to the reference. An MD > 0 denotes too large intermolecular distances. All values are given in pm.

	S22			**S66x8**			**S22**	**P26**
	MD	MAD	MAX	MD	MAD	MAX	$\overline{\text{RMSD}}$	$\overline{\text{RMSD}}$
PBEh-3c	7.6	12.7	98.4	3.2	**5.7**	16.7	9.9	16.1
M06-2X	-7.7	**8.4**	-44.7	-9.5	9.6	-24.6	**6.5**	**8.4**
B3LYP	78.7	82.6	389.9	40.2	43.0	362.5	55.1	48.8
B3LYP-D3-gCP	4.6	11.2	43.8	10.1	10.2	28.0	8.2	17.5
B3LYP-DCP	-4.4	**5.7**	33.8	1.3	20.5	59.6	**3.9**	**9.4**
HF	81.8	82.2	263.2	54.2	54.2	362.5	53.2	49.6
HF-D3-gCP	8.6	15.6	135.4	1.7	**2.8**	f 18.4	11.7	12.7

dispersion effects can describe the $\pi - \pi$ stacking. B3LYP-DCP and PBEh-3c yield the most and B3LYP-D3-gCP the least accurate structure. The missing dispersion terms in plain B3LYP are less problematic for the hydrogen bonded dimer because hydrogen bonds are mainly caused by electrostatic and induction interactions which most density functionals cover rather accurately. Thus, for the hydrogen bonded dimer plain B3LYP yields a geometry which is as good as with M06-2X and even slightly better than the B3LYP-D3-gCP structure. Again, B3LYP-DCP and PBEh-3c have the smallest RMSD compared to the reference.

Table 3.2 presents the statistical data of the S22, S66x8 and P26 test sets for small systems. The two best performing methods for each set are highlighted and the data are converted to normal error distributions in Figure 3.10.

In general, plain B3LYP and HF cannot correctly reproduce the structures. Especially, when $\pi - \pi$ stacking or nonpolar dispersion interactions are involved these methods yield a practically unbound geometry or a different conformation, like the hydrogen bonded dimer in the example shown above. Thus, the mean (absolute) deviations for the intermolecular center-of-mass distances R_{CMA} for the S22 and S66x8 sets as well as the heavy atom positions in the S22 and P26 sets are unacceptably large. When the D3-gCP corrections are added, the MDs and MADs for the R_{CMA} of about 80 pm for S22 and about 40-50 pm for S66x8 drop significantly. B3LYP-D3-gCP yields an MD of only 5 pm and an MAD of 11 pm in case of S22 and the same values for both measures of 10 pm for the S66x8 set. HF-D3-gCP gives an MD of 9 pm and and MAD of 16 pm for the S22 set. For S66x8, HF-D3-gCP in fact is one of the two best methods and yields an MD of only 2 pm and and MAD of only 3 pm. The $\overline{\text{RMSD}}$ for the heavy atom positions are similar for B3LYP-D3-gCP and HF-D3-gCP, about 10 pm for the S22 and about 15 pm for the P26.

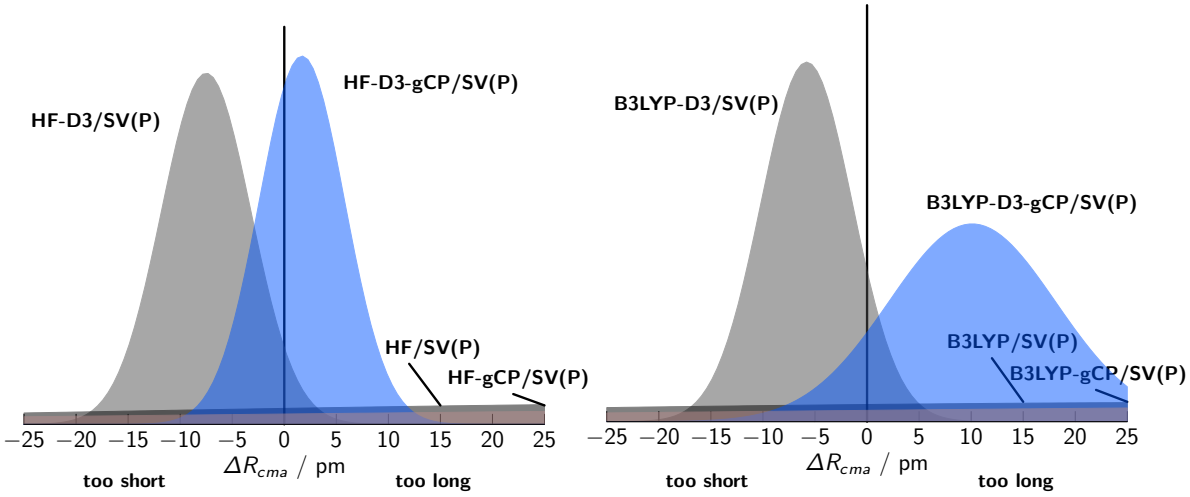


Figure 3.10.: Error statistics of S66 equilibrium center-of-mass distances for corrected and uncorrected HF and B3LYP in a SV(P) basis set converted into normal error distributions.

B3LYP-DCP performs somewhat better than B3LYP-D3-gCP for S22 and gives the best results for this set with an MAD of -4 pm, an MAD 6 pm, and an $\overline{\text{RMSD}}$ value for the heavy atom positions of only 4 pm. In case of the S66x8 set the MD of 1 pm is even lower and the best value obtained but the MAD of 21 pm is much larger than those for all other dispersion-corrected methods. M06-2X yields one of the two best values for the heavy atom $\overline{\text{RMSD}}$s of 7 and 8 pm for the S22 and P26 set, respectively. It is the only method which consistently gives too small intermolecular center-of-mass distances R_{CMA} for both, the S22 and S66x8 set. The MDs are -8 and -10 pm, respectively. The MADs are very similar, indicating that this error is systematic. PBEh-3c yields results comparable to HF-D3-gCP and slightly worse than M06-2x, with an MD of 8 pm, an MAD of 13 pm and an $\overline{\text{RMSD}}$ value of 10 pm for the S22 set. In case of the S66x8 the performance is better. The MD is 3 pm and the MAD is with 6 pm the second best value obtained.

Note that the S66 set consists of the most reliable reference data, while the S22 and P26 systems are only optimized at the MP2 level. A comparison of the S66 equilibrium CMA distances calculated at MP2 level reveals an MAD of 3.3 pm compared to the coupled cluster reference. Therefore, MAD values of a few pm on the S22 and P26 sets do not indicate significant deviations and are within the MP2 error. Even more important in this context are systematic errors in the reference structures. For instance some of the largest PBEh-3c outliers for S22 occur for the π-stacked benzene dimer, for which MP2 is known to overbind significantly[214]. Presumably, in this case the MP2 reference is in fact off (too short distance) as indicated by a distance underestimation by 3.4% compared to the CCSD(T) reference for the benezene dimer in the S66x8 set.

The influence of the dispersion and counterpoise correction schemes for the S66 equilib-

rium distances is summarized in Figure 3.10. In analogy to the binding energy analysis in the previous paragraph, we show HF and B3LYP deviations with and without correction converted into normal error distributions. These distributions mainly confirm the analysis given above. While the uncorrected methods yield rather bad structures, the most consistent methods are the dispersion and counterpoise corrected ones, though B3LYP-D3-gCP yields slightly too large intermolecular distances.

Among the dispersion-corrected methods B3LYP-DCP seems to be the best performer for all three test sets but we could not identify a method which is clearly superior to others. It is important to note, that due to the larger 6-31+G(2d,2p) basis set and the dispersion correcting potentials themselves, the geometry optimizations with B3LYP-DCP are an order of magnitude slower than with all other methods employed.

In order to show the influence and the interplay of dispersion and BSSE for supermolecular systems we optimized complex **5** of the S30L and the **phe** complex with plain B3LYP, B3LYP-D3, B3LYP-gCP and B3LYP-D3-gCP. The overlays of these geometries with the reference structure are presented in Figure 3.11.

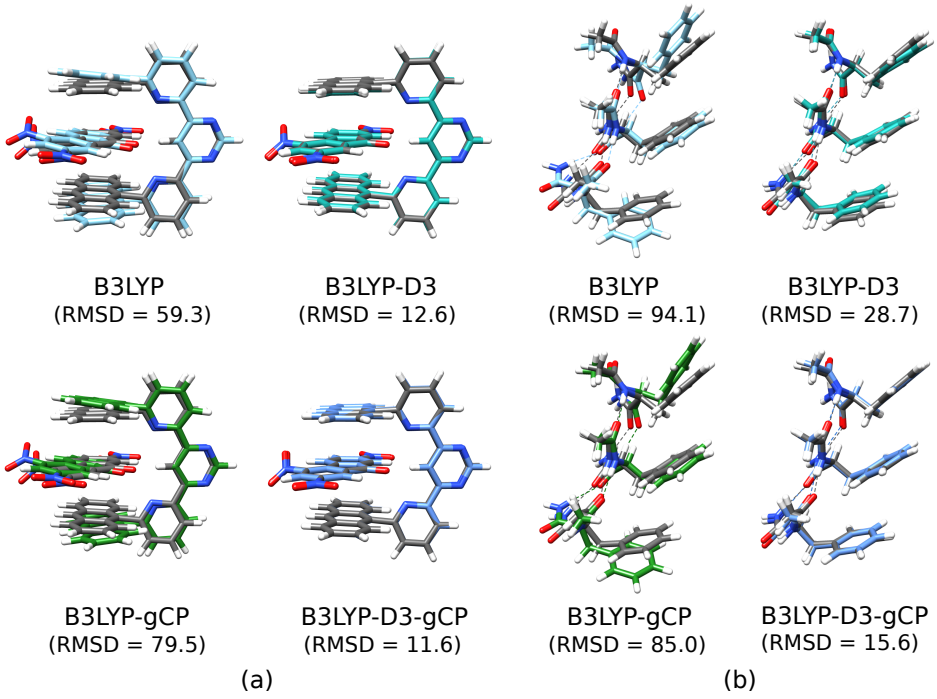


Figure 3.11.: Comparison of the B3LYP, B3LYP-D3, B3LYP-gCP, and B3LYP-D3-gCP geometries (colored) with the reference TPSS-D3/def2-QZVP(-gf) structure (in gray) for (a) complex **5** of the S30L set and (b) the **phe** complex of the L7 set. The RMSDs for the heavy atoms are given in pm.

As already observed for the small complexes the plain B3LYP functional gives too large distances for the $\pi - \pi$ stacked systems due to the dominant effect of missing dispersion. The RMSD of the heavy atom positions is 59 pm for **5** and 94 pm for **phe**. If the D3 scheme

is employed, the distances are slightly too small due to the BSSE. The RMSDs drop to 13 and 29 pm, respectively. When we only correct for the BSSE by gCP, the distances are again far too large and the RMSD values are similar to those of plain B3LYP. Only if dispersion and BSSE corrections are employed together (B3LYP-D3-gCP level), accurate geometries are obtained. Visually, the agreement with the reference structure is very good and the heavy atom RMSDs are only 12 pm for **5** and 16 pm for **phe**. These two examples clearly show, that not only for interaction energies, but also for geometries of large systems it is important to treat London dispersion and BSSE on the same footing.

Finally, we investigate the structures of the previously introduced X23 and ICE10 solid state benchmark sets. As noted before, we use the experimental crystal densities (or crystal volumes) from X-ray measurements. These mass densities have been back-corrected for zero-point and thermal effects, which is important as it can alter the mass density by 1 to 5% with a typically decreased density (enlarged unit cell volume). In Table 3.3, we give the statistical deviations from the reference unit cell volumes for the methods PBEh-3c, M06-2X, B3LYP, B3LYP-D3-gCP, and B3LYP-DCP all with the same basis sets as in the molecular calculations.

The general picture that emerged from the molecular complexes is confirmed for the crystals. However, because of the larger long-range contributions to the interaction the differences between the tested methods are more pronounced. Again, M06-2X and B3LYP-DCP are numerically problematic and suffer from SCF convergence problems. As already seen for the molecular dimers, M06-2X suffers from BSSE, which leads to systematically too small unit cells by 13% and 15% for the X23 and ICE10 set, respectively. For plain B3LYP inconsistent behavior for the two test sets is found. For the more dispersion dominated X23 systems, the unit cells are substantially too large by more than 20% though for some systems the error compensation leads to better results than B3LYP is

Table 3.3.: Statistics (MD, MAD, SD, MAX)[a] for the relative deviations of the cell volume for the X23 and ICE10 sets. All values are given in %. The two best performing methods with smallest MAD are highlighted.

	X23				**ICE10**			
	MD	MAD	SD	MAX	MD	MAD	SD	MAX
PBEh-3c	1.8	**2.7**	3.2	10.2	2.5	**5.0**	7.7	16.6
M06-2X[b]	-12.5	12.5	4.4	23.3	-14.9	14.9	2.2	17.3
B3LYP	22.1	22.1	15.5	57.3	-5.7	5.7	1.4	8.3
B3LYP-D3-gCP	5.5	7.6	6.4	14.7	7.6	8.3	5.5	15.8
B3LYP-DCP[b]	-3.5	**3.6**	2.1	7.5	-4.4	**4.4**	1.9	7.0

[a] An MD > 0 denotes a too large cell volume.
[b] only 70% of the systems could be converged.

inherently capable. For instance, the geometries of the oxalic acid polymorphs are very reasonable with only 4% deviation from the reference density. The ice polymorphs are more strongly dominated by electrostatic and induction effects with only small dispersion contribution. Here, the BSSE is even larger compared to the missing dispersion leading to too small unit cells. Applying both correction schemes (D3 and gCP) results in a more consistent performance. At this level, for both test sets a reasonable MAD of about 8% is obtained. A combined optimization of the gCP and D3 parameters would probably lead to even better geometries. B3LYP-DCP is based on a larger basis set with rather diffuse functions, which explains some of the convergence problems. However, this also minimizes the BSSE and the results are good with MADs for the X23 and ICE10 reference unit cell slightly below and slightly above 4%, respectively. Again, especially the ice polymorphs are described to a high accuracy consistent with the excellent lattice energies. On the X23 set, B3LYP-DCP is only outperformed by the new PBEh-3c composite method. The geometries are competitive to more expensive calculations based on converged PAW basis sets with typical unit cell errors of about 3%.[177,215]

As prototypical example for London dispersion dominated crystal, we investigate the benzene crystal in more detail. It has various energetically close-lying polymorphs[216,217] and it was used extensively to test and judge electronic structure methods (including wave-function expansions[218–222], dispersion corrected DFT[204,210,215,223–225], and semiempirical MO methods[226–228]). We show a potential energy surface (PES) scan of the benzene crystal in Figure 3.12. Each structure corresponds to a constrained volume optimization at the TPSS-D3 level in a converged PAW[229,230] basis set and we additionally highlight the equilibrium point. The reference point refers to the back-corrected experimental unit cell volume combined with a highly accurate CCSD(T) computed lattice energy.[231]

Because of SCF convergence problems, M06-2X results are not included. Concerning the other methods, substantial differences in the computational speed are observed. With identical numerical setups, the relative timing for one single-point energy calculation of PBEh-3c, B3LYP/SV(P), and B3LYP-DCP/6-31+G* are 1.0 : 1.2 : 8.2 with PBEh-3c being the fastest and B3LYP-DCP/6-31+G* the slowest method. The higher computational cost is mainly due to the larger and more diffuse basis set, which leads (especially in periodic boundaries) to a substantially higher number of computed integrals.

The benzene crystal nicely reflects the basic properties of the described methods regarding the treatment of dispersion dominated systems. Plain B3LYP just shows a shallow BSSE related minimum. This correctly disappears when the gCP correction is applied. Only in combination with both correction schemes (B3LYP-D3-gCP), a very reasonable PES is obtained with nearly perfect lattice energy (-13.16 (B3LYP-D3-gCP) vs. -13.22 kcal mol^{-1} (reference)) but slightly to large unit cell (by 2.6% too low mass density). The B3LYP-DCP approach shows a clear minimum which is somewhat too low

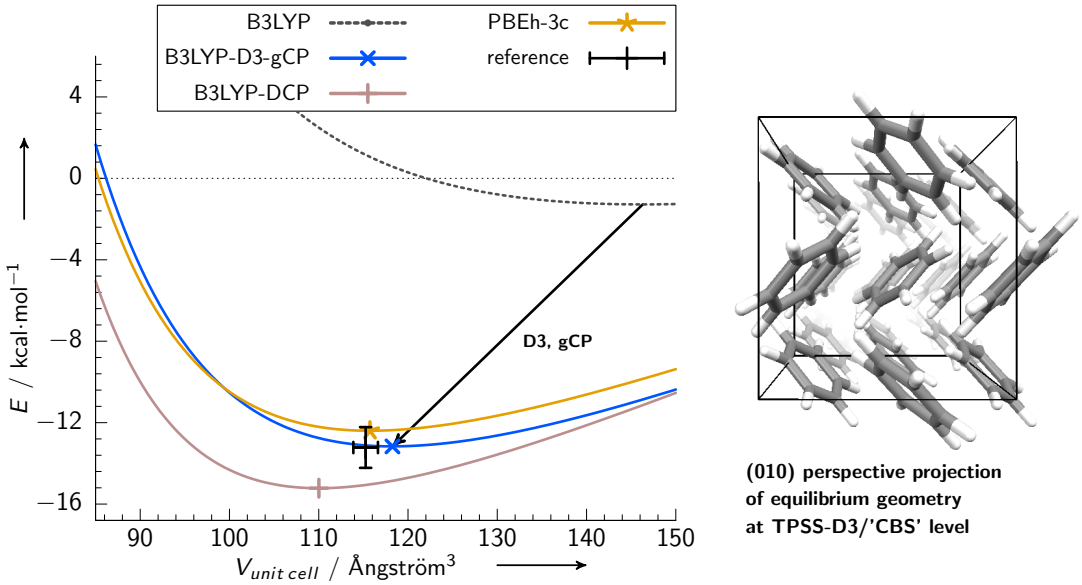


Figure 3.12.: Lattice energy of the benzene crystal along a PES based on contained volume optimizations (TPSS-D3/'CBS' level). The experimental geometry is back-corrected for zero-point and thermal effects as described in Refs. [177,207] and the reference lattice energy corresponds to a CCSD(T) estimate.[231]

(lattice energy of -15.22 kcal mol^{-1}) and too small unit cell (by 4.5% too large mass density). The potential of PBEh-3c agrees excellently with the reference and both the unit cell volume and the lattice energy are within 1.2% and 0.8 kcal mol^{-1}, respectively.

3.5. Conclusions

In this short topical review, we have critically analyzed widely used quantum chemical HF and DFT computations employing relatively small single-particle basis sets of double-zeta quality. As indicated by the tremendous number of publications which are based on this or similar theoretical levels, these methodologies are practically very relevant. We highlighted the two main error sources in standard applications, namely the BSSE and the missing London dispersion interaction. Different strategies to treat and correct the errors were reviewed and tested on mainly noncovalently bound systems with varying size. We analyzed both energetic and geometric properties. Due to the efficiency of the methods, their main applications are large supramolecular or periodic systems, which were also the focus of our analysis.

As main result of our investigations, it is nowadays not justified to rely on fortuitous error compensation as e.g. in the popular B3LYP/6-31G* approach. Without additional computational overhead, the main error sources can be treated with semi-classical potentials and the composite method B3LYP-D3-gCP/DZ outperforms the plain functional clearly. Further improved results are obtained, when the BSSE and London dispersion are

directly included in the method design as recently done in the PBEh-3c functional with good to excellent results on all tested geometries and reference energies. Only for some systems with particular high requirements on the basis set (e.g. very strong hydrogen bonds or anions) the performance is slightly worse compared to e.g. a 'hybrid'/QZ level.

Using ECP type potentials to simultaneously cure the functional and basis set errors works very well for small complexes similar to those used in the training sets for the method. However, with the tested B3LYP-DCP scheme, the computational costs are closer to that of a triple-zeta basis set. More importantly, the quality of the results for larger systems deteriorates and the performance is unsatisfying. We attribute this inconsistency to a wrong distance behavior of the correction potential as the finite ECP expansion can not recover the correct R^{-6} limit of the dispersion interaction.

In summary, it is indeed possible to effectively use quantum chemical methods with small basis set expansions when all arising errors are treated properly. The good results for both, energetic and geometric properties of large and periodic systems is encouraging and we expect this to translate into globally accurate potential energy surfaces, which is important for thermodynamic properties and *ab initio* molecular dynamics.

In this context we would like to mention the problem of solvation effects that was not discussed in this review. Dispersion effects are omnipresent and also occur for any molecule when it is solvated as in most chemical applications. Molecular dispersion (or BSSE) effects are then partly quenched, i.e., intramolecular contributions are replaced by intermolecular ones with the solvent. An accurate account of these effects requires sophisticated solvation models with the same high accuracy as the quantum chemical treatments which is difficult to obtain at present. Whenever comparisons of computed molecular to experimental liquid phase data are made we recommend to inlcude consistent continuum solvation models like COSMO-RS[41,232] or DCOSMO-RS[41] (for geometry optimizations). Only such treatments eventually will lead to the 'right answer for the right reason'. In any case due to the broad area of possible applications of e.g. the new PBEh-3c composite scheme in describing host-guest binding enthalpies, lattice enthalpies of organic crystals, and structures of larger biologically relevant molecules, the future for quantum chemical modeling of these systems seems bright.

3.6. Computational details

For the single-point energy calculations on the benchmark sets S22[199,214], S66x8[172], WATER27[208], L7[144], S30L[206], ICE10[207], and X23[204,210] the geometries were taken as provided in the corresponding references. The computations for the molecular systems were carried out with either the current development version of ORCA 3.0[233,234] in case of B3LYP[88,92,96,154]-DCP[185,186]/6-31+G(2d2p)[155] or TURBOMOLE 7.0[235,236] for all

other methods (HF, B3LYP, M06-2X[100], PBE0[98] and PBEh-3c[177]) in combination with the double-ζ basis set def2-SV(P)[237] (modified in case of PBEh-3c for boron to neon, for details see Ref.[177]). The resolution-of-identity (RI) approximation for the Coulomb integrals[74] was applied in all cases except B3LYP-DCP using matching default auxiliary basis sets[238]. For the integration of the exchange-correlation contribution the numerical quadrature grids *m4* (*m5* in case of M06-2X)[239] and grid 6 were employed in TURBOMOLE and ORCA, respectively. The energy convergence criteria were set to 10^{-7} E_h in all cases.

The periodic calculations were conducted with a developer version of the CRYSTAL14 program[240]. It is the ideal choice for cost effective DFT calculations in small basis sets as it can exploit full point and space group symmetry. The Brillouin zone is sampled with a Γ-centered k-mesh with grid density of approximately 0.025 $Å^{-1}$ (for details see references[204,207]). Standard integral thresholds and large DFT integration grids were used.

The calculation of the D3(BJ)[36,37] dispersion correction and the gCP BSSE correction[47] were carried out with our own programs `dftd3` and `gcp`, respectively. These programs are freely available from our website[241].

For the structure optimizations of the benchmark sets S22 and P26[201] the geometries as provided in the corresponding references were taken as start coordinates. Again, for B3LYP-DCP/6-31+G(2d,2p) ORCA 3.0 was employed and TURBOMOLE 6.6 was used in for HF, B3LYP, M06-2X, PBE0, and PBEh-3c in combination with the def2-SV(P) basis set (modified in case of PBEh-3c). The D3(BJ) and gCP corrections to the gradients were again calculated with our own programs `dftd3` and `gcp`. The convergence criteria were set to 10^{-7} E_h for energies and 10^{-5} E_h/Bohr for gradients.

4. Corrected small basis set Hartree–Fock method for large systems

Rebecca Sure* and Stefan Grimme*

Received 28th of February 2013, Published online 14th of June 2013

Reprinted (adapted) with permission from
R. Sure and S. Grimme, *J. Comput. Chem.* **2013**, *19*, 1672–1685.
 DOI: 10.1002/jcc.23317

Own manuscript contribution:

- Implementation of the method into ORCA
- Performance of the calculations
- Interpretation of data
- Writing the manuscript

*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany

Abstract

A quantum chemical method based on a Hartree–Fock calculation with a small Gaussian AO basis set is presented. Its main area of application is the computation of structures, vibrational frequencies and noncovalent interaction energies in huge molecular systems. The method is suggested as a partial replacement of semiempirical approaches or DFT in particular when self-interaction errors are acute. In order to get accurate results three physically plausible atom pair-wise correction terms are applied for London dispersion interactions (D3 scheme), basis set superposition error (gCP scheme), and short-ranged basis set incompleteness effects. In total nine global empirical parameters are used. This so called Hartree–Fock-3c (HF-3c) method is tested for geometries of small organic molecules, interaction energies and geometries of noncovalently bound complexes, for supramolecular systems, and protein structures. In the majority of realistic test cases good results approaching large basis set DFT quality are obtained at a tiny fraction of computational cost.

4.1. Introduction

Noncovalent interactions such as van der Waals interactions or H-bonding play a crucial role in the chemistry of supra-molecular and bio-molecular systems as well as for nano-structured materials.[16,17] They control host-guest and enzyme-substrate binding, structures of proteins and DNA, antigen-antibody recognition or the orientation of molecules on a surface. Theoretical methods based on first principles to complement experimental studies which often can provide only limited information about these complex soft-matter systems seem indispensable.

Many of these systems or at least reasonable models thereof can nowadays be computed routinely with quite good accuracy by (dispersion corrected) density functional theory (DFT) together with relatively large basis sets (triple-ζ quality or better). For recent reviews how to treat the important long-range London dispersion interactions in DFT see Refs. [32,33]. One perspective of such treatments is to provide accurate input data to parametrize simpler force-field or even coarse-grained theoretical models although full protein structures can be treated[242]. But despite of the good cost-accuracy ratio of DFT for large systems, these calculations are often prohibitive in terms of the necessary computational efforts. Furthermore the quadrature of the exchange- correlation energy in DFT causes numerical noise in geometry optimizations or frequency calculations which is a particular problem in these often flexible systems. Accurate harmonic frequencies are an important ingredient for the computation of thermodynamic properties as for example free energies of association of supra-molecules[46]. Another issue in DFT are charged

systems (e.g. proteins with charged residues) where the self-interaction error (SIE[243,244]) can lead to artificial charge-transfer and convergence problems of the self consistent field (SCF)[242,245,246] at least when 'cheap' semi-local functionals of general gradient approximation (GGS) type are used. Modern semiempirical methods like DFTB3[247], OM2[126] or PM6[124] (for an overview see Ref. [248]) represent an alternative in principle but suffer from missing parametrization for important elements or robustness in certain situations (e.g. charged complexes[192]).

As will be shown in this work, most of the above mentioned problems can be alleviated by applying Hartree–Fock (HF) theory together with small AO basis sets. The basic idea is to fill the gap between existing semiempirical methods and DFT in terms of the cost-accuracy ratio with a physically sound approach. Using HF has the following advantages: First, in contrast to DFT, HF does not suffer from SIE and extended charged systems even when treated un-screened (in vacuo) are unproblematic. Second, a HF calculation is performed completely analytical, including the computation of gradients and Hessians so that no problems with numerical noise in geometry optimizations or frequency calculations occur. Third, contrary to standard semiempirical approaches HF is inherently able to treat the important hydrogen bonding so that there is no need for atom-type dependent H-bond corrections which are normally applied for neglect of differential overlap (NDDO) type methods.[249]. Furthermore, the proposed HF method can be applied without any parametrization to almost any element of the periodic table and includes important physical effects like Pauli-exchange repulsion correctly. The accurate description of these steric interactions was always a problem in semiempirical methods[248] and even current density functionals are not free of inaccuracies for short inter-atomic distances[172,250]. For density functionals which try to mimic the HF short-range repulsive behavior see e.g. Ref.[251]

It is clear, however, that the Coulomb correlation energy is entirely missing in HF and a small basis set can introduce further severe errors. The suggested approach is hence not meant to be generally applicable or as a replacement of DFT. Rather, it should yield reasonable results for simple molecular properties like equilibrium structures or vibrational frequencies or for noncovalent interactions, i.e., when changes in the basic electronic structure during a chemical process is small. The accurate computation of chemical reaction energies requires the account of various short-ranged polarization and correlation effects and is not of concern here (and likely not computable with a minimal or small AO basis set).

Several years ago Pople noted that HF/STO-3G optimized geometries for small molecules are excellent, better than HF is inherently capable of yielding.[252,253] Similar observations were made by Kołos already in 1979, who obtained good interaction energies for a HF/minimal-basis method together with a counterpoise-correction as well as a correction to account for the London dispersion energy.[254] It seems that part of this valuable knowl-

edge has been forgotten during the recent 'triumphal procession' of DFT in chemistry. The true consequences of these intriguing observations could not be explored fully at that time due to missing computational resources but are the main topic of the present work.

We recently noted the good performance of HF/large-basis in combination with our latest dispersion correction scheme D3[36,37] for noncovalent interactions and we will use this well-established dispersion correction (see Ref.[212,255–257] for recent D3 applications) also in the present work. Work along similar lines (i.e. using HF-D3/STO-3G) has been done by the group of T. Martinez recently[202]. The basis set superposition error (BSSE) is significant for a small or minimal basis set and will be treated with our recently developed geometrical counterpoise correction (gCP).[47] Importantly, this approach also accounts for intramolecular BSSE which is difficult to correct efficiently otherwise. Both schemes are used essentially in unmodified form here. Additionally, a new short-ranged basis (SRB) incompleteness correction term is applied. This corrects for systematically overestimated bond lengths for electronegative elements (e.g. N,O,F) when employing small basis sets. According to common practice basis set effects are separated into BSSE and BSIE (basis set incompleteness error). In this sense the SRB term corresponds to the BSIE and the gCP scheme accounts for the atom pair- wise part of the BSSE (for related BSSE correction schemes see Refs. [196,197]).

The basis set used here is of minimal quality for the often occurring ('organic') elements H,C,N,O and mostly of split-valence (SV) or polarized SV (SVP) quality for the other elements. It is dubbed 'MINIX' from now on and an inherent (fixed) ingredient of the method. For simplicity this HF-D3-gCP-SRB/MINIX method will be abbreviated HF-3c in the following where the term '3c' stands for the three applied corrections and the mentioned compound basis set is always implied. It should also indicate that the method accounts for the important dispersion contributions by the relatively accurate D3 scheme[36,37].

We present HF-3c results in comparison to those obtained with the semiempirical PM6[124] method and to standard DFT. The PM6 method is used because it is parametrized for very many elements so that the same systems can be calculated for comparison. We investigate geometries of small organic molecules as well as interaction energies and geometries of small noncovalent complexes. As more realistic tests, geometries and association free energies of supramolecular complexes will be considered. This also includes a test of the quality of the harmonic vibrational frequencies. Finally, HF-3c results for protein structures will be presented and compared to experimental X-ray and solution NMR data.

Table 4.1.: Composition of the MINIX basis set.

element	basis
H-He, B-Ne	MINIS
Li-Be	MINIS+1(p)
Na-Mg	MINIS+1(p)
Al-Ar	MINIS+1(d)
K-Zn	SV
Ga-Kr	SVP
Rb-Xe	def2-SV(P) with ECP

4.2. Theoretical and computational Methods

4.2.1. The HF-3c method

The starting point for calculating the electronic energy is a standard HF treatment with a small Gaussian AO basis set. The herein used so called MINIX basis set consists of different sets of basis functions for different groups of atoms (Table 4.1). The valence scaled minimal basis set MINIS[258] and the split valence double-ζ basis sets SV, SVP[259] and def2-SV(P)[237] (the latter together with effective core potentials (ECP)[260] for heavier elements) are employed. Many other possibilities have been considered but the chosen one not only represents a very good compromise between accuracy and speed but furthermore this basis seems to be balanced and easily to correct for deficiencies (see below).

The HF calculations are conducted in conventional mode, i.e., the two-electron integrals are computed once and stored on disk or in memory if possible. This option is a further advantage of the small basis set approach and leads to large computational savings. Only huge systems are treated in direct mode by re-calculating integrals in every SCF iteration. The so called resolution of the identity (RI) approximations are not applied because the savings are negligible for small basis sets and this approach can even slow-down the computations due to overhead from the necessary linear algebra parts.

Three terms are added to correct the HF energy $E_{tot}^{HF/MINIX}$ in order to include London dispersion interactions, to account for the BSSE and to correct for overestimated bond lengths. The corrected total energy is calculated as

$$E_{tot}^{HF-3c} = E_{tot}^{HF/MINIX} + E_{disp}^{D3(BJ)} + E_{BSSE}^{gCP} + E_{SRB}. \tag{4.1}$$

The first correction term $E_{disp}^{D3(BJ)}$ is the atom-pair wise London dispersion energy from

the D3 correction scheme[36] and applying Becke-Johnson (BJ) damping[37,133,134]

$$E_{disp}^{D3(BJ)} = -\frac{1}{2}\sum_{A\neq B}^{atoms}\left(s_6\frac{C_6^{AB}}{R_{AB}^6 + \left(a_1 R_0^{AB} + a_2\right)^6} + s_8\frac{C_8^{AB}}{R_{AB}^8 + \left(a_1 R_0^{AB} + a_2\right)^8}\right) \tag{4.2}$$

Here, C_n^{AB} denotes the nth-order dispersion coefficient (orders = 6, 8) for each atom pair AB, R_{AB} is their internuclear distances and s_n are the order-dependent scaling factors. The cutoff radii $R_0^{AB} = \sqrt{C_8^{AB}/C_6^{AB}}$ and the fitting parameters a_1 and a_2 are used as introduced in the original works[133,134]. For the present method, the three usual parameters s_8, a_1 and a_2 were re-fitted using reference interaction energies of the the S66 test set complexes[172]. This results in $s_8 = 0.8777$, $a_1 = 0.4171$ and $a_2 = 2.9149$. The parameter s_6 was set to unity as usual to enforce the correct asymptotic limit and the gCP correction (see below) was already applied in this fitting step.

The second term E_{BSSE}^{gCP} denotes our recently published geometrical counterpoise (gCP) correction[47] for BSSE, which depends only on the atomic coordinates of a given molecule. The difference in atomic energy E_A^{miss} between a large (nearly complete) basis set and the target basis set (MINIX in our case) for each free atom A is calculated for the HF Hamiltonian. The E_A^{miss} term is multiplied with a decay function depending on the interatomic distances R_{AB}. The sum over all atom pairs reads

$$E_{BSSE}^{gCP} = \sigma\sum_{A}^{atoms}\sum_{A\neq B}^{atoms} E_A^{miss}\frac{exp\left(-\alpha\left(R_{AB}\right)^{\beta}\right)}{\sqrt{S_{AB}N_B^{virt}}}, \tag{4.3}$$

where α, β and σ are fitting parameters, S_{AB} is a Slater-type overlap integral and N_B^{virt} is the number of virtual orbitals on atom B in the target basis. The S_{AB} is evaluated over a single s-type orbital centered on each atom and using optimized Slater exponents weighted by the fourth fitting parameter η. The gCP parameters were fitted in a least-squares sense against counterpoise correction data obtained by the scheme of Boys and Bernadi[166] as described in the original publication[47]. This way, for each combination of a Hamiltonian (HF or DFT) and a basis set, a specific set of parameters α, β, σ and η was created. We found that this gCP correction performs particularly well for HF in combination with a small basis set. For further details and recent applications see[47,145].

The last term E_{SRB} is a short-ranged correction to deal with basis set deficiencies which occur when using small or minimal basis sets. It corrects for systematically overestimated covalent bond lengths for electronegative elements and is again calculated as a sum over all atom pairs:

$$E_{SRB} = -s\sum_{B\neq B}^{atoms}\left(Z_A Z_B\right)^{3/2}\exp(-\gamma(R_{AB}^{0,D3})^{3/4}R_{AB}) \tag{4.4}$$

Here, $R_{AB}^{0,D3}$ are the default cut-off radii as determined *ab initio* for the D3 dispersion correction scheme[36] and Z_A, Z_B are the nuclear charges. The correction is applied for all elements up to argon. The empirical fitting parameters $s = 0.03$ and $\gamma = 0.7$ were determined to produce vanishing HF-3c total atomic forces for the B3LYP-D3(BJ)/def2-TZVPP equilibrium structures of 107 small organic molecules. The other two correction terms were included in the fitting procedure of E_{SRB}, which was carried out by minimizing the HF-3c RMS gradient for the reference geometries. The D3 and gCP parameters were kept constant at their previously optimized values in this procedure. Because the SRB correction also effects covalent bond energies, the thermochemical properties of HF-3c are different from those of HF-D3-gcp/MINIX. Some cross-checking for standard reaction energies of organic molecules showed that HF-3c performs reasonably well but further tests which are out of the scope of the present work should be conducted to validate this finding.

In summary, the HF-3c method consists of only nine empirical parameters, three for the D3(BJ) dispersion, four in the gCP scheme and two for the SRB correction. Because the fits are done independently, this parametrization procedure was easy to perform and changes in the setup of the fit are not expected to have any major effect on the method. No element or pair-specific terms need to be determined, i.e., the nine parameters apply globally for all elements considered (i.e., currently up to iodine). Total energies and 3c-components for a few molecules are given in the SI.

4.2.2. Technical details

All HF/MINIX and B3LYP[91,92]-D3(BJ)/def2-TZVPP[237] calculations were performed using TURBOMOLE 6.4.[261] In case of B3LYP the resolution-of-identity (RI) approximation for the Coulomb integrals[74] was applied using matching default auxiliary basis sets.[238] The numerical quadrature grid *m4* was employed for integration of the exchange-correlation contribution. The 3c-terms to energy and analytical gradient were calculated by a new code which basically merges the freely available programs `dftd3` and `gCP`[241]. For both, HF and DFT, computations of the harmonic vibrational frequencies were performed analytically using the `aoforce` code of TURBOMOLE. The 3c-contributions to the Hessian are computed numerically by two-point finite differences of analytical gradients.

All PM6[124] and PM6-DH2[262] calculations were undertaken using MOPAC 2012[263] for the calculation of energies and gradients but the `relax` or `statpt` codes from TURBOMOLE 6.4 for executing the geometry relaxation steps. Vibrational frequencies were computed numerically using MOPAC 2012.

The COSMO-RS model[41,42] was used as implemented in COSMOtherm[264] to ob-

tain all solvation free energies. Single-point calculations on the default BP86[91,265]/def-TZVP[266] level of theory were performed on the optimized gas phase geometries.

All visualizations of molecules were done with USCF Chimera version 1.6.1.[267]

The root mean square deviation (RMSD) of two geometries was calculated using a quaternion algorithm[268] in order to get an all atom best-fit.

The HF-3c method has also been implemented into the upcoming version of the free ORCA software[234] where it is invoked simply by keyword.

4.2.3. Computation of free energies of association

Free energies of association for host and guest molecules in a solvent X at a temperature T are calculated as

$$\Delta G_a = \Delta E + \Delta G^T_{RRHO} + \Delta\delta G^T_{solv}(X). \tag{4.5}$$

Here, ΔE denotes the gas phase interaction energy of the fully optimized molecules and G^T_{RRHO} is the sum of thermal corrections from energy to free energy within a rigid-rotor-harmonic-oscillator approximation for each molecule in the gas phase at a given temperature T, including the zero-point vibrational energy. All harmonic frequencies are scaled with a factor of 0.86 for HF-3c. For obtaining the vibrational entropy, low-lying modes below $\approx$100 cm^{-1} are treated within a rigid-rotor model in order to reduce their error in the harmonic approximation, for details see Ref[46]. The solvation free energy $\delta G^T_{solv}(X)$ is calculated for each gas-phase species by employing the COSMO-RS model[41,42]. No further (empirical) corrections are applied and the so computed values can be directly compared to experimental data.

4.3. Results and Discussion

4.3.1. Geometries of small organic molecules

The fitting set for the SRB correction of basis set deficiencies consists of 107 small organic molecules (2 to 34 atoms) containing the elements H, B, C, N, O, F, Si, P, S, and Cl. All standard functional groups are represented within this test set (for a detailed list of molecules see SI). The B3LYP-D3(BJ)/def2-TZVPP geometries, which have been proven to be reliable for organic molecules, were used as reference structures in the fitting procedure. PM6 calculations were performed to compare the HF-3c results to those from a widely used semiempirical approach.

Geometry optimization of these organic molecules using the final 3c-parameters yield an average root mean square deviation (RMSD) between the HF-3c and B3LYP-D3 Cartesian coordinates of 0.033 Å. This is considered to be a very good result meaning that

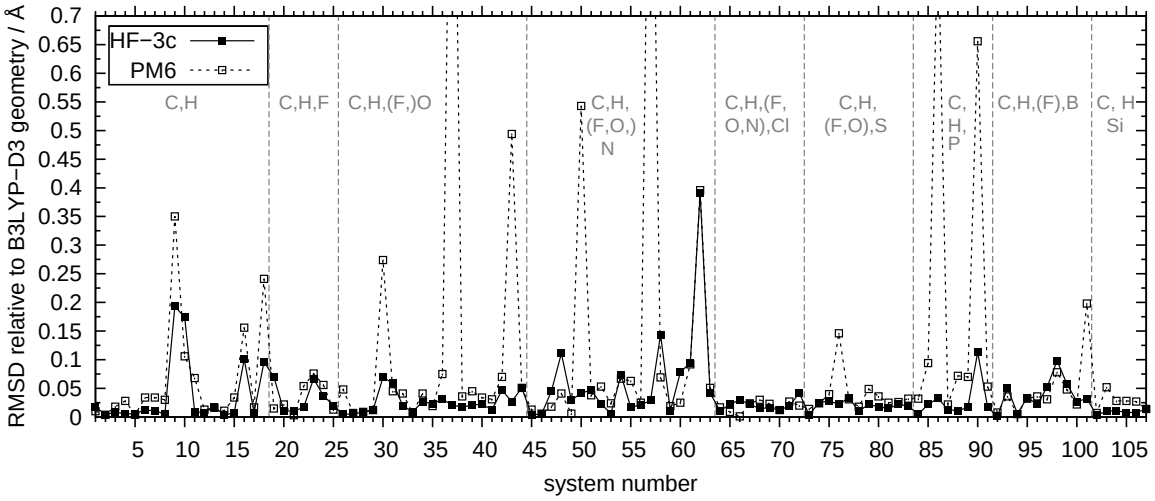


Figure 4.1.: Root mean square deviation (RMSD) between HF-3c or PM6 and B3LYP-D3/def2-TZVPP geometries for 107 small organic molecules. The molecules are sorted according to the type of atoms and hence to the functional groups they contain. The atoms given in brackets are only rarely represented in the corresponding group. The lines between the data points are drawn just to guide the eye.

at least for the fit set HF-3c yields structures of almost B3LYP/large-basis quality. The RMSD values for the individual molecules are shown in Figure 4.1. One of the rare 'outliers' with a notably higher RMSD (adenine, 63) merely shows a methyl group rotated by 180° compared with the reference structure. PM6 shows more 'outliers' than HF-3c and the average RMSD of 0.910 Å is much larger. Also the PM6 geometries of adenine as well as methyl acetate (43) exhibit a rotated methyl group. Further, hydrogen peroxide (30) is planar whereas glyoxal (37) and urea (58) are not as they should be. Hydrazine (50), diphosphane (87) and PH_2NH_2 (91) adopt the anti instead of the gauche conformation when optimized with PM6. These drastic conformational changes do not occur in optimizations with the HF-3c method.

Comparison of the lengths for the most frequent bonds (C-C, C=C, conjugated C-C/C=C, C-H, O-H N-H, P-H, B-H, C-F, C=O, C-O, C-N, conjugated C-N/C=N, C-S, C-Cl, C-B and C-Si) results in an overall mean deviation (MD) with respect to the reference structures of 0.012 Å for HF-3c and 0.005 Å in case of PM6. With a few exceptions (C=C, B-H and C-F) the HF-3c bond lengths tend to be slightly too long. The mean absolute deviation (MAD) for all considered bond lengths in HF-3c and PM6 structures is 0.015 Å and 0.016 Å respectively. Hence, the overall error for bond lengths is similar for both methods. Due to a better description of bond angles and dihedral angles, HF-3c geometries generally show smaller RMSD values than PM6 structures.

The accuracy as demonstrated above also results from the SRB correction. This is more clearly seen by comparing some critical bond lengths with and without this term in

Table 4.2.: Critical bond lengths for some exemplary molecules at the HF/MINIX, HF-D3-gCP/MINIX, HF-3c and B3LYP-D3/def2-TZVPP level. All distances are given in Å.

molecule	bond	R(HF/ MINIX)	R(HF-D3-gCP/ MINIX)	R(HF-3c)	R(B3LYP-D3)
acetone	C=O	1.264	1.268	1.206	1.209
urea	C=O	1.275	1.280	1.216	1.218
methaneimine	C=N	1.294	1.298	1.260	1.264
ethanol	C-O	1.478	1.486	1.428	1.428
urea	C-N	1.423	1.427	1.397	1.372
hexaflouroethane	C-F	1.413	1.429	1.343	1.334
H_2S_2	S-S	2.132	2.136	2.122	2.073

typical molecules. For example the C=O bond length in a ketone like acetone is 1.268 Å at the HF-D3-gCP/MINIX level (1.264 Å at HF/MINIX) which is too long by about 0.06 Å. This systematic deviation it corrected with HF-3c and the computed length of 1.206 Å is sufficiently close to the B3LYP reference value fo 1.209 Å. Another example is hexaflouroethane where the corresponding values for the C-F bond length are at 1.429 Å the HF-D3-gCP/MINIX level (1.413 Å at HF/MINIX) and 1.343 Å at the HF-3c level (1.334 Å at B3LYP). A few more comparisons are given in Table 4.2 where in general the strong influence is seen for several bonds in polar situations.

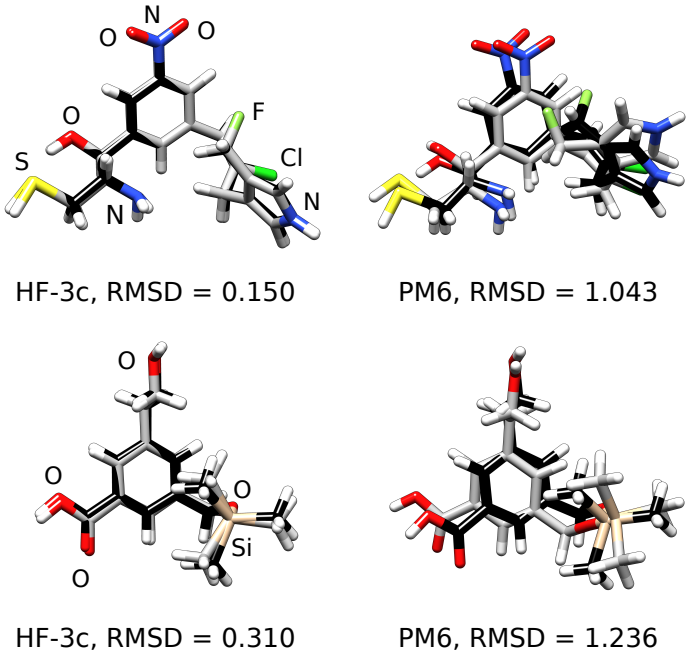


Figure 4.2.: Two artificially constructed organic molecules optimized with HF-3c (left grey structures) and PM6 (right grey structures). Black coloured B3LYP-D3/def2-TZVPP geometries serve as reference. All RMSDs are given in Å.

As a cross-validation, two artificial neutral organic molecules containing a few heteroatoms were constructed in a more or less arbitrary fashion and fully optimized with all three methods taking again B3LYP-D3(BJ)/def2-TZVPP as reference. The RMSD relative to the reference structure is 0.15 Å for HF-3c and 1.043 Å for PM6 in case of the first molecule and 0.310 Å for HF-3c and 1.236 Å for PM6 in case of the second molecule (see Figure 4.2). For both structures HF-3c performs significantly better than PM6. Additionally, PM6 is not able to correctly describe the bond angle at the oxygen-atom of the silyl ether group in the second molecule but instead yields an almost linear coordination geometry.

Additionally, we performed single-point calculations for 10 conformers of the tripeptide phenylalanyl-glycyl-glycine (PCONF set[269]), 15 conformers of the n-alkanes butane, pentane and hexane (ACONF set[270]), 15 conformers of the sugar 3,6-anhydro-4-O-methyl-D-galactitol (part of the SCONF set[271]) and 10 conformers of cystein (CYCONF set[272]) as included in the GMTKN30 benchmark set[86]. The reference energies were taken from the original publications. For PCONF, SCONF and CYCONF they were calculated on the estimated coupled cluster with singles an doubles excitations and perturbative triples at the estimated complete basis set limit (CCSD(T)/CBS) level of theory and the ones for ACONF on the W1h-val level. The mean absolute deviation (MAD) for all conformational energies is 1.4 kcal/mol for HF-3c, which is an reasonable result in particular because this property is quite sensitive to the quality of the AO basis set. PM6-DH2 yields a much higher MAD of 2.8 kcal/mol while B3LYP-D3/def2-QZVP gives a much smaller MAD of 0.3 kcal/mol. The D3-correction contributes significantly to this good result, since plain B3LYP/def2-QZVP yields an MAD of 1.5 kcal/mol (i.e., is worse than HF-3c).

Further cross-validation studies for structures are performed on noncovalent complexes and their fragments as discussed in the next sections.

4.3.2. Geometries and interaction energies for S22 and S66 sets

In order to test the capability of the HF-3c method to describe noncovalent interactions, single-point calculations as well as geometry optimizations for the S22[199] and S66[172] test sets were carried out. Due to under representation of some interaction motifs, the S66 set was published by the Hobza group as a revised and extended version of the S22 set.[172] We also used their recently published X40 test set, which was designed to cover different halogen bonding interactions.[273] Reference values for interaction energies and geometries were taken from the original publications. The interaction energies refer to the estimated CCSD(T)/CBS level and the geometries were optimized on the MP2/cc-pVTZ(CP) or CCSD(T)/cc-pVTZ(noCP) level of theory.

Table 4.3.: Mean deviation (MD) and mean absolute deviation (MAD) for the single-point interaction energies of the S22, S66 and X40 test sets for the three methods HF-3c, PM6 and PM6-DH2. All energies are given in kcal/mol.

	HF-3c		PM6		PM6-DH2	
	MD	MAD	MD	MAD	MD	MAD
S22	-0.01	0.55	3.39	3.39	0.13	0.39
S66	-0.09	0.38	2.68	2.68	0.35	0.65
X40	-0.80	1.44	1.19	1.73	0.35	1.46

Again, PM6 optimized geometries and interaction energies are used for comparison. Additionally, the DH2 correction[262] to PM6 for hydrogen-bonding and dispersion was employed which is mandatory for this kind of benchmark. Due to known problems with this correction for geometry optimizations, the scheme of calculating PM6-DH2 energies on PM6 geometries proposed by Hobza *et al.* was applied.[262,274]

For the S22 and S66 sets, the single-point HF-3c interaction energies are rather accurate with MADs of 0.55 kcal/mol and 0.39 kcal/mol, respectively. These values are considerably lower than the previously published ones (0.64 and 0.51 kcal/mol) for HF/mini calculations applying just the D3 and gCP correction.[47] Thus, the modified basis set together with the SRB correction term and re-parametrization gives a further significant improvement. This accuracy is comparable or even better than obtained for some density functionals at the DFT-D3/large-basis level[250].

The MD values of -0.01 kcal/mol in case of S22 and -0.09 kcal/mol in case of S66 are almost insignificant. In case of the X40 test set both the MAD of 1.44 kcal/mol and the MD of -0.80 kcal/mol are much higher than for S22 (MAD of 0.55 kcal/mol) and S66 (MAD of 0.38 kcal/mol) but they are still reasonable for the applied theoretical level. In conclusion it is clear that HF-3c is able to provide a qualitatively correct and quantitatively reasonable description of general noncovalent interactions. For a detailed analysis of responsible systematic error compensations see Ref.[47]

In contrast, PM6 single-point calculations result in equal values for the MD and MAD of 3.39 kcal/mol for the S22 and 2.68 kcal/mol for the S66 test set which indicates a systematical underbinding. This error can be reduced by applying the DH2 correction which accounts for dispersion and H-bonding. PM6-DH2 yields an MD of 0.13 kcal/mol and an MAD of 0.39 kcal/mol in case of the S22 and an MD of 0.35 kcal/mol and an MAD of 0.65 kcal/mol for the S66 set. Again, for the X40 set the deviations are much higher (MAD of 1.46 kcal/mol, MD of 0.35 kcal/mol). Altogether, the HF-3c method performs slightly better than PM6-DH2 in reproducing the interaction energies.

For the S22 set HF-3c geometry optimizations lead to an MD of 0.42 kcal/mol and an MAD of 0.94 kcal/mol for the interaction energies. Optimizations on the PM6 level of

theory results in much higher values of 3.11 kcal/mol for both MD and MAD. Except for complex 10, which shows an imaginary vibrational mode for methyl rotation on the HF-3c level of theory, all optimized complexes are minima on the corresponding potential energy surface (PES) for both methods when started straightforwardly from the reference coordinates. In various cases the convergence criteria for energy and gradient and the step size for the numerical PM6 frequency calculations had to be adjusted in order to remove small artificial imaginary frequencies. Similar numerical problems do not occur in HF-3c calculations. PM6-DH2 single-point calculations on PM6 geometries yield an MD of 0.1 kcal/mol and an MAD of 0.76 kcal/mol which are slightly lower than the corresponding values for HF-3c although the inconsistencies in the PM6 optimizations should be kept in mind.

Comparison of the resulting geometries with the reference structures yields an average RMSD of 0.21 Å in case of HF-3c and 0.45 Å for PM6. As shown in Figure 4.3(a), there are more outliers for PM6 than for HF-3c geometries. The HF-3c geometries of both, the T-shaped benzene dimer (20) and the T-shaped benzene$\cdots$indole complex (21) show structures in between a T-shaped and parallel-stacked one. The rings of two parallel stacked systems, namely the benzene dimer (11) and the benzene$\cdots$indole complex (14), are rotated towards each other compared with the reference structures. Altogether the general structural motifs of the S22 complexes can be reproduced well with HF-3c keeping in mind the flatness of the corresponding PES. In contrast, PM6 seems to systematically disfavor parallel stacked geometries. Instead of a parallel stacking the benzene dimer (11) shows a T-shaped stacking, the uracil dimer (14) an H-bonded geometry and the benzene$\cdots$indole complex (14) a structure between parallel-stacked and T-shaped. Further, the orientation of the monomers in PM6 optimized geometry of the methane dimer (8) differs from the one in the reference structure. Overall, the HF-3c geometries in the S22 set match the reference structures better than the PM6 ones.

The results for the S66 set reveal a similar picture. Geometry optimizations of the complexes yield an MD of 0.08 kcal/mol and an MAD of 0.59 kcal/mol for the interaction energy in case of HF-3c and again the same value for the MD and MAD of 2.33 kcal/mol for PM6. The PM6-DH2 single-point calculations on PM6 geometries result in an MD of 0.33 kcal/mol and an MAD of 0.81 kcal/mol which are slightly higher than the values for HF-3c. Similar to the S22 set there are more outliers for PM6 than for HF-3c geometries (Figure 4.3b) compared to the reference. The average structural RMSD is 0.20 Å in case of HF-3c and 0.68 Å for PM6. All structures were proven to be minima on the corresponding PES though PM6 again shows problems with numerical noise. In general, HF-3c geometries reproduce the reference structures very well. The acetamide dimer (21) shows a rotated methyl group and the rings of the parallel stacked benzene$\cdots$uracil complex (28) are differently rotated towards each other compared to the reference structures. In

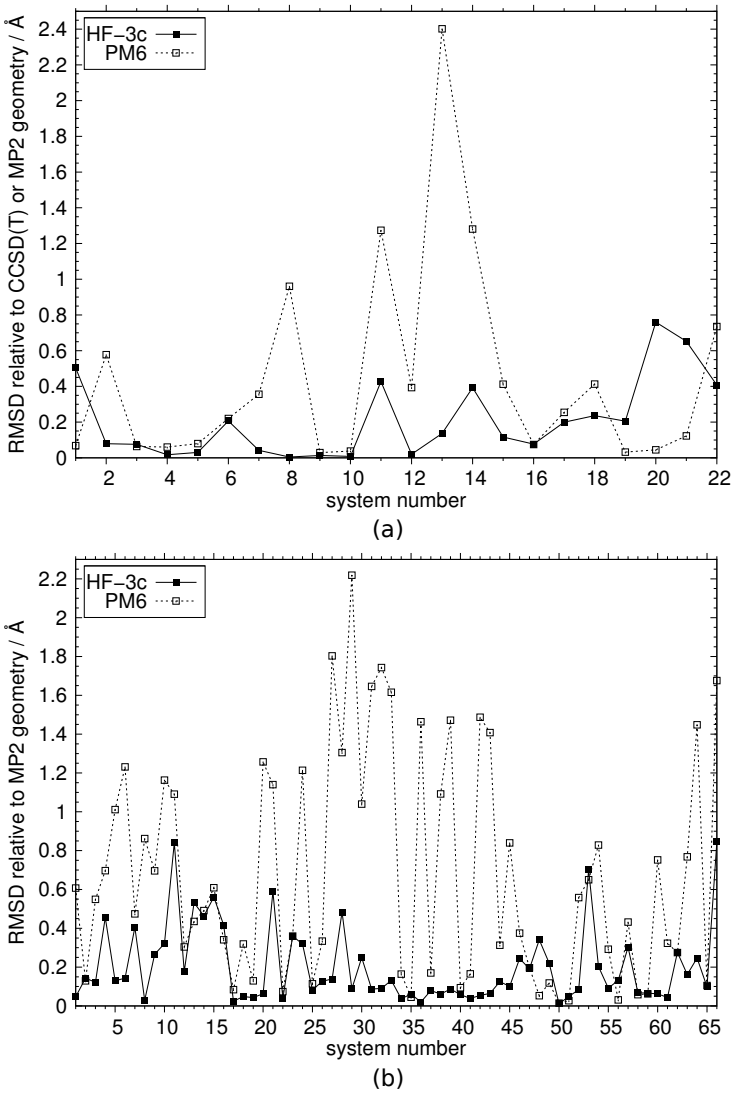


Figure 4.3.: RMSD between HF-3c or PM6 and CCSD(T)/cc-pVTZ(noCP) or MP2/cc-pVTZ(CP) reference geometries for S22 (a) and S66 (b). The lines between the data points are drawn just to guide the eye.

all cases, the basic interaction motifs are preserved in the HF-3c geometries which is a very important result.

PM6 geometries of the acetic acid dimer (20), acetamide dimer (21) and the ethyne$\cdots$acetic acid complex (60) feature a rotated methyl group. As already observed for the S22 set PM6 prefers T-stacked geometries over parallel stacked ones. Almost every parallel stacked reference geometry shows T-shaped binding when optimized with PM6. Furthermore, the pyridine$\cdots$uracil complex (29) shows an H-bonded geometry instead of parallel stacking and the H-bonded pyridine$\cdots$methylamine complex (66) does not exhibit an H-bond at all.

Overall the HF-3c method reproduces the reference geometries of the S22 and S66 sets better than PM6. The RMSD is smaller and the general interaction motives are preserved

in all cases indicating robustness in practical applications. The MDs and MADs for HF-3c interaction energies derived from optimized structures are similar to single-point values indicating that the HF-3c and reference PES are reasonably parallel to each other. The accuracy for HF-3c computed noncovalent interaction energies approaches that of dispersion corrected DFT but is less than the best DFT-D3/large-basis variants.

4.3.3. Thermal corrections to Gibbs free energies for small organic molecules and noncovalent complexes

Vibrational frequency calculations and the corresponding zero-point energy and thermal corrections to Gibbs free energies are supposed to be a main area of application of HF-3c. We randomly chose ten molecules out of 107 from the geometry fitting set, four complexes from S22, and six from the S66 test set. For these 20 molecules the $E \rightarrow G(298)$ corrections were calculated using HF-3c, PM6, and B3LYP-D3/def2-TZVPP as reference. The scaling factors for the harmonic vibrational frequencies were set to 0.86 for HF-3c, 1.0 for PM6 and 0.97 for B3LYP. Low-lying modes below $\approx$100 cm^{-1} were treated within a rigid-rotor model[46] in order to reduce their error in the harmonic approximation when obtaining the vibrational entropy. The final thermal corrections for all 20 molecules are listed in the SI (table 4).

Comparison of HF-3c with the B3LYP reference values shows a good agreement with an MD of 0.8 kcal/mol and an MAD of 1.9 kcal/mol (corresponding to about 3% relative error). For most molecules the deviations range from only -1.3 to 2.7 kcal/mol. The four molecules with the highest deviations are tetramethylsilane, the ethane-pentane complex and the cyclopentane-neopentane complex where the HF-3c thermal corrections are 4.2 to 7.4 kcal/mol too large and the T-shaped benzene dimer for which the HF-3c value is 7 kcal/mol too small. The large error for the benzene dimer can be attributed to the very shallow potential energy surface. In case of PM6 the thermal corrections for all regarded molecules except ammoniaborane are too small. The MD with respect to the B3LYP-D3/def2-TZPP values is -7.0 and the MAD is 7.2 kcal/mol, i.e., significantly worse than for HF-3c.

4.3.4. Geometries and association free energies of supramolecular complexes

Recently, we compiled a set of 12 supramolecular complexes (S12L set) and compared calculated free energies of association with experimental data.[46] This set was very recently used to benchmark various dispersion corrections to DFT[192] and will be taken in this work for cross-validation of the HF-3c method on large realistic systems.

The investigated complexes are two "tweezer" complexes with tetracyanoquinone (TCNQ) and 1,4-dicyanobenzene (DCB) (1a and 1b measured in $CHCl_3$)[275], two "pincer" complexes of organic π-systems (2a and 2b in CH_2Cl_2)[276], the fullerenes C_{60} and C_{70} in a "buckycatcher" (3a and 3b in toluene)[277], complexes of an amide macrocycle (mcycle) with glycine anhydride (GLH) and bezoquinone (BQ) (4a and 4b in $CHCl_3$),[278] complexes of cucurbit[6]uril (CB6) with butylammonium ($BuNH_3$) and propylammonium ($PrNH_3$) (5a and 5b in a 1:1 mixture of formic acid and water)[279] and complexes of cucurbit[7]uril (CB7) with a di-cationic ferrocene derivative (FECP) and 1-hydroxyadamantane (ADOH) (6a and 6b in water).[280]

Computations at the PW6B95-D3(BJ)/def2-QZVP'//TPSS-D3(BJ)/def2-TZVP level for gas phase interaction energies ΔE together with a rigid rotor harmonic oscillator model for thermodynamical corrections ΔG_{RRHO} and the COSMO-RS model for solvation free energies $\Delta\delta G_{solv}$ are able to reproduce the experimental values for association free enthalpies for these complexes with good accuracy. The MAD from experimental data was about 2 kcal/mol.[46] These results were used as a reference to test the performance of HF-3c for geometries and free enthalpies of association of the S12L set of supramolecular complexes. Again, PM6-DH2//PM6 calculations are performed for comparison.

Figure 4.9 (a) shows the magnitudes of the contributions to the association free energy (ΔE, ΔG_{RRHO} and $\Delta\delta G_{solv}$) for HF-3c, PW6B95-D3//TPSS-D3 as reference and PM6 or PM6-DH2//PM6, respectively. The HF-3c gas phase interaction energy tends to be lower than the PW6B95-D3 energy, the deviation for the complexes 1a, 1b, 2a, 2b, 4a, 4b and 6b is 0.5 to -2 kcal/mol. For C_{60}@Catcher (3a) and C_{70}@Catcher (3b) HF-3c is overbinding by 5 to 6 kcal/mol, for $BuNH_3$@CB6 (5a) and $PrNH_3$@CB6 (5b) by 10 kcal/mol and for FECP@CB7 (6a) by 12.6 kcal/mol. The result for FECP@CB7 is not surprising since HF is known to describe transition metal complexes in general badly. Additionally, the complex has a double positive charge, which is challenging for a small basis set method due to large polarization effects. Consistent with this, the two complexes 5a and 5b with a larger error also carry a positive charge. These errors demonstrate that HF-3c is well-behaved and performs as expected.

Overall, the HF-3c gas phase interaction energies have an MD of -4.2 and an MAD of 4.4 kcal/mol compared with the PW6B95-D3//TPSS-D3 reference values. The MD indicates a small systematical overbinding and the MAD is similar to various dispersion corrected DFT methods employing large AO basis sets[46].

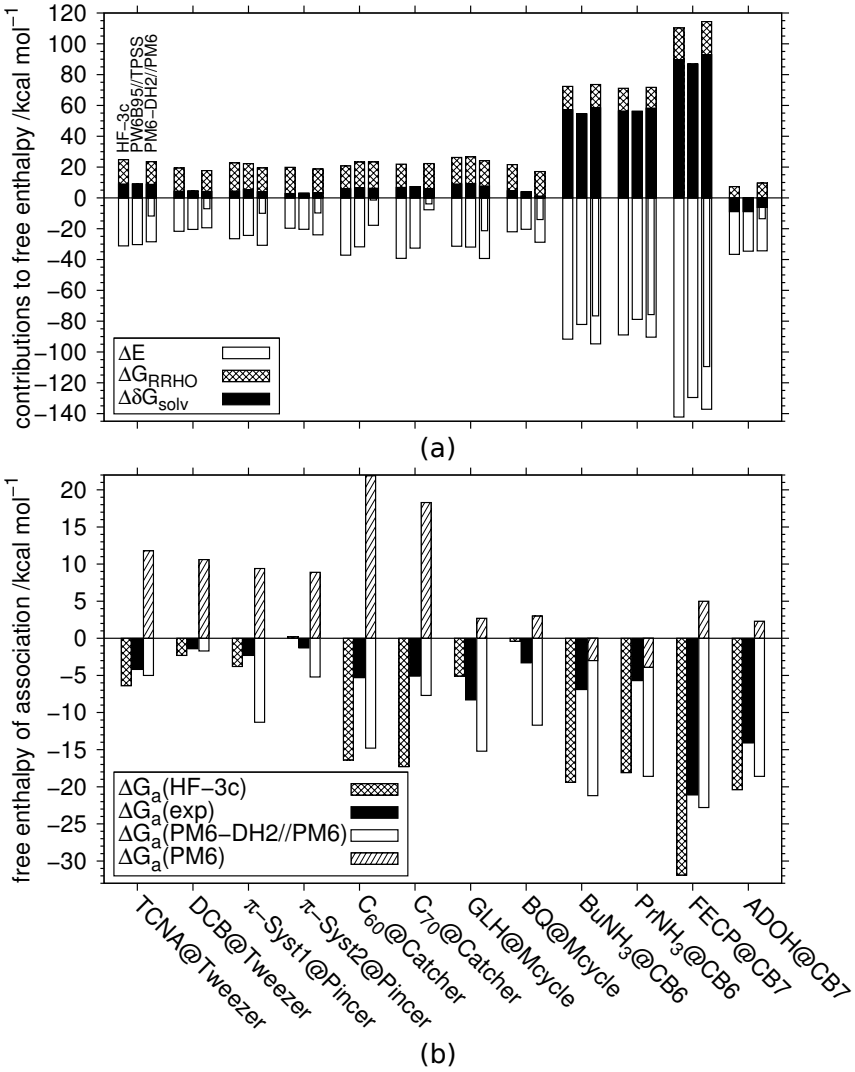


Figure 4.4.: (a) Contributions to free energy of association (interaction energy ΔE, RRHO free energy correction ΔG_{RRHO} and solvation free energy $\Delta\delta G_{solv}$). PW6B95-D3/def2-QZVP'//TPSS-D3/def2-TZVP values are taken from Ref.[46] and are shown for comparison. The left bar for each complex always presents the HF-3c values, the bar in the middle the PW6B95-D3//TPSS-D3 values and the right bar the PM6-DH2//PM6 (pure PM6 results for ΔE are shown with narrower bars) values. Not all ΔG_{RRHO} have been computed at the DFT level. (b) Total free energy of association ΔG_a for all supramolecular complexes on the HF-3c, PM6 and PM6-DH2//PM6 levels of theory. Experimental values are taken from Refs. [275,276,278–281] and are shown for comparison.

All PM6 interaction energies are much higher than the reference values, the deviation ranges from 3 up to 30 kcal/mol. Applying the PM6-DH2//PM6 approach, the deviations decrease but remain larger than for HF-3c (6.1 kcal/mol compared to 4.4 kcal/mol). Exceptions are C_{70}@Catcher (3b) and FECP@CB7 (6a) with an error of -3.6 and -7.6 kcal/mol, respectively. Except for complexes 1a and 1b, PM6-DH2 overbinds and the MD (-5.6 kcal/mol) is absolutely larger than for HF-3c.

Comparison of the HF-3c geometries with the TPSS-D3 reference structures yield a minimal RMSD of 0.04 Å for the complex C_{60}@Catcher (3a) and a maximal RMSD of 0.48 Å for π-Syst1@Pincer (2a). The average RMSD is 0.19 Å. The corresponding values for PM6 are 0.11 Å, 0.97 Å and 0.45 Å. For both methods the complexes $BuNH_3$@CB6 (5a) and $PrNH_3$@CB6 (5b) show a slightly different coordination of the guest molecule compared with the reference geometries. Similar to the small noncovalent complexes, the HF-3c method reproduces the reference structures better than PM6.

Since the geometry enters the COSMO-RS calculation, the better performance of HF-3c is also reflected in the solvation free energies $\Delta\delta G_{solv}$ of the complexes. The $\Delta\delta G_{solv}$ values based on the HF-3c geometries deviate from the reference values in the range from only -0.5 to +2.6 kcal/mol whereas the deviation based on PM6 geometries ranges from -2.7 to +6.1 kcal/mol.

Because of the high computational cost, the thermodynamic correction ΔG_{RRHO} on the TPSS-D3/def2-TZVP level of theory has been computed only for three complexes (2a, 3a and 4a).[46] Both simpler methods match the three reference values relatively well. The highest deviation is 1.5 kcal/mol in case of HF-3c and 1.3 kcal/mol for PM6 corresponding to about 5-10% of ΔG_{RRHO}. Because the number of comparisons is very small we can only guess that both methods might perform equally well.

The sum of all these contributions, the association free energy ΔG_a, is shown in Figure 4.9 (b) in comparison to the experimental values. Since the gas phase interaction energy is the largest contribution and also most sensitive to the quality of the underlying electronic structure method, the error in ΔG_a mainly reflects the error in ΔE. Therefore, HF-3c yields ΔG_a values which are too low (overbinding). Nevertheless, the calculated ΔG_a values from HF-3c are surprisingly good regarding the simplicity of the method and an MD of -5.2 and an MAD of 6.2 kcal/mol seems to be very respectable. The PM6-DH2//PM6 values are even lower and hence, the overbinding is even stronger than for HF-3c in most cases. The only significant exception is the complex FECP@CB7, whose ΔG_a(PM6-DH2//PM6) matches the reference value much better than the HF-3c one. Since the HF-3c geometries are quite accurate and the derived values for $\Delta\delta G_{solv}$ and ΔG_{RRHO} in particular are reasonable, a single-point DFT-D3/large-basis calculation on the HF-3c geometries is suggested for improved performance. For screening applications or scanning of supra-molecular potential energy surfaces, however, HF-3c seems to be sufficiently

accurate.

4.3.5. Geometries of small proteins

Recently, Martinez *et al.* composed a set of 58 small proteins with 5 to 35 residues in length and total charges ranging from -2 to +2.[202] To test the performance of HF-3c, these proteins were fully optimized starting from the experimental geometries, which were taken from the Protein Databank (PDB).[282] Eight structures were excluded due to problems with the original PDB file (residues were missing or charges could not be assigned according to Ref.[202]). In case of multiple protein structures in one PDB file, the first one was always used. Again, PM6 optimizations were performed for comparison.

During the HF-3c geometry optimization procedure of almost all proteins, the charged termini of the protein backbone neutralize via proton transfer from the protonated amino group to the carboxylate, if they are in close proximity or close to a lysine and aspartic or glutamic acid. This was also observed when two of those amino acids are too close. The protonation states and final charges were determined with USCF Chimera, which uses an empirical procedure for adding hydrogen atoms to the protein structure and AMBER ff99SB parameters[283] to assign the overall charge. Hence, it is not completely sure weather this is the same protonation state the protein would adopt in its natural environment. Six final HF-3c geometries (1T2Y, 2I9M, 2NX6, 2NX7, 2RLJ, 2RMW) exhibit a very small imaginary vibrational frequency below -22 cm^{-1}, all other structures are true minima on the PES. In case of PM6 this hydrogen transfer is observed for only a few proteins. Contrary to the unproblematic HF-3c calculations, the PM6 optimization of ten proteins showed convergence problems which could not be solved. Additionally, 13 optimized structures exhibit persistent imaginary frequencies. Nevertheless, all structures also with imaginary frequencies are included in the geometry analysis.

As a first examination, the backbone RMSD between the calculated and the starting experimental geometries was evaluated using USCF Chimera[267]. The results are shown in Figure 4.5. All C^{α} atom pairs were included, even if the calculated secondary structure strongly deviates from the reference one. In this way, the RMSD value gives a hint how good the computed secondary structure is. The minimal RMSD for the HF-3c geometries is 0.45 Å for 3NJW, the maximal value is 5.21 Å for 2PJV and the average RMSD is 2.02 Å. The average RMSD between different models of solution NMR structures in the whole set of 58 proteins is 1.73 Å.[202] Hence, the average RMSD for the HF-3c geometries is acceptable. In most cases the general secondary structure is preserved. Figure 4.6 shows four protein geometries with a very small RMSD in comparison to the experimental structures. We consider 13 protein structures which exhibit a backbone RMSD higher than 2.5 Å (arbitrarily chosen threshold) as some kind of outliers and these are now discussed

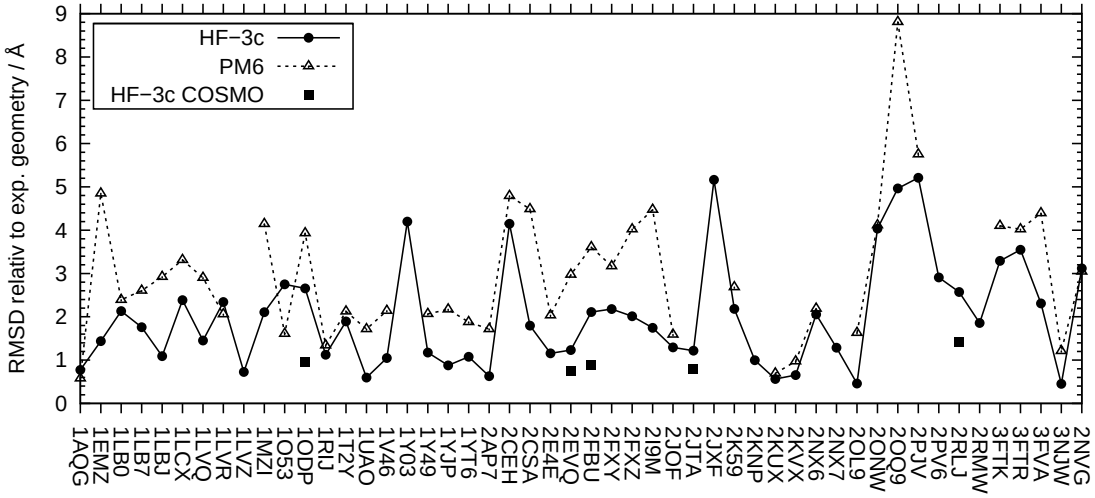


Figure 4.5.: Backbone RMSD for all optimized protein structures on the HF-3c and PM6 level of theory relative to the experimental starting structure. The lines between the data points are drawn just to guide the eye.

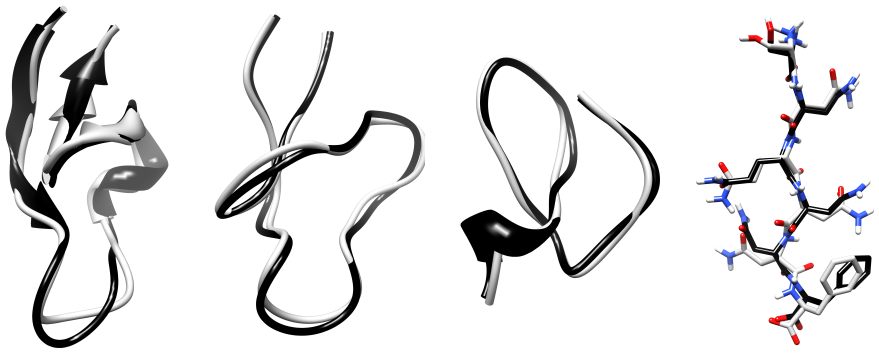


Figure 4.6.: HF-3c structures (grey) for four proteins with a small backbone RMSD in comparison to experimental ones (black). The RMSDs are given in Å. Hydrogens at carbon atoms in structure (d) are omitted for clarity.

in more detail.

Figure 4.7 shows the HF-3c geometries of four proteins with a high RMSD and the experimental structure in comparison. The experimentally determined α-helix of 1Y03 is bent but straight in the HF-3c calculation (Figure 4.7a). The opposite applies for 2JXF (Figure 4.7b) and 2OQ9, where the experimental structure exhibits a straight helix and the calculated geometry a bent one. In case of 2PJV (Figure 4.7c), 2PV6, 1ODP and 1O53 the α-helix is strongly distorted compared with the experimental geometry. For 2ONW (Figure 4.7d), 3FTK, 3FTR and 3NVG the backbone of the experimental structures is more or less linear whereas it is folded in HF-3c optimized geometries. Protein 2CEH neither has a α-helix nor a β-sheet structure and the HF-3c geometry is disordered in a different way than the experimental one. 2RLJ exhibits a larger helix part when optimized

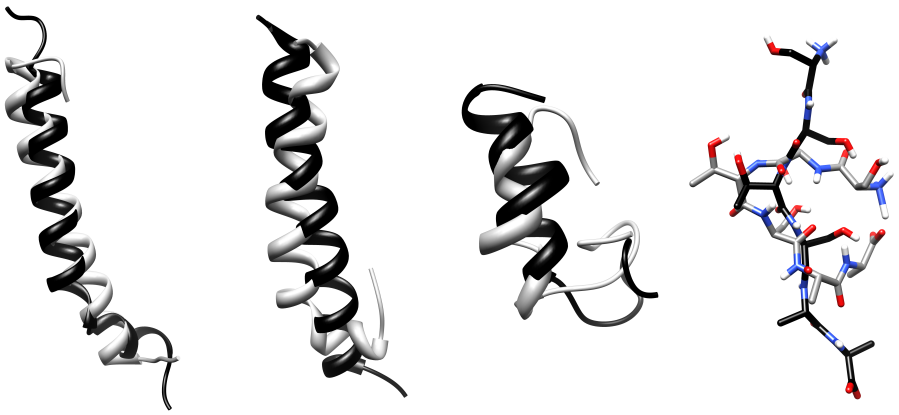


Figure 4.7.: HF-3c structures (grey) for four proteins with a high backbone RMSD in comparison to experimental ones (black). The RMSDs are given in Å. Hydrogens at carbon atoms in structure (d) are omitted for clarity.

with HF-3c compared to the experimentally obtained geometry.

In case of PM6 the minimal backbone RMSD is 0.58 Å for 1AQG and the maximal value is 8.81 Å for 2OQ9. The average backbone RMSD of 2.96 Å is much higher than for the HF-3c optimized geometries. For more than half of the investigated proteins the PM6 structure yields an RMSD larger than 2.5 Å and in most cases PM6 is not able to reproduce the general secondary structure.

Standard health checks to characterize the protein structures were used as described in Refs. [284–286]: (1) clashcores or steric overlaps greater than 0.4 Å per 1000 atoms, (2) percentage of bad side-chain dihedrals or rotamers, (3) number of β-carbon deviations greater than 0.25 Å from the expected position based on the backbone coordinates, (4) percentage of backbone dihedrals that fall into a favored region on a Ramachandran plot and (5) percentage of those, which are Ramachandran outliers, (6) percentage of bad bonds and (7) percentage of bad angles. These health checks were performed for the calculated as well as the starting experimental structures. No structural improvements, e. g. allowing Asn/Gln/His flips, were made. To provide one single number that represents the quality of a protein structure, the MolProbity score was defined as a logarithmic-weighted combination of clashores, percentage of Ramachandran outliers and percentage of bad side-chain rotamers.[284] The averaged results are shown in Table 4.4, the individual values for each protein are provided in the supporting information.

The health check data for the HF-3c structures match the values obtained for the experimental geometries very well. The values for clashcores and bad angles are only slightly higher. The most defective health criterion is the percentage of bond outliers. Compared to the values published by Martinez *et al.*[202] for HF-D3/mini the application of the geometrical counterpoise correction and the additional short-range term in the HF-

Table 4.4.: Averaged health criteria for the HF-3c (50 proteins) and PM6 (41 proteins) optimized structures as well as the experimental starting geometries (50 proteins). Values for HF-D/mini and HF/6-31G were taken from Ref.[202] for comparison (all 58 proteins).

	exp.	HF-3c	PM6	HF-D3/mini	HF/6-31G
clashcore / 1000 atoms	29	34	54	43	8
bad side-chain rotamers	19 %	13 %	21 %	18 %	10 %
C^{β} deviations	0.2	0.2	0.0	0.5	0.3
Ramachandran outliers	5 %	6 %	8 %	7 %	3 %
Ramachandran favored	81 %	81 %	71 %	77 %	86 %
bad bonds	0.5 %	8 %	3 %	79 %	1 %
bad angles	1 %	4 %	1 %	10 %	1 %
MolProbity score	2.7	3.3	3.9	3.1	1.9

3c method gives an improvement for all health criteria. This is particularly obvious for the percentage of bond outliers, which is much smaller for the HF-3c geometries than for the ones obtained with HF-D3/mini. Compared to the results from the original publication for HF and the 6-31G basis set, the HF-3c health criteria are almost compatible. The highest deviation is found again in the percentage of bond outliers. Additionally, the number of clashcores is substantially smaller for HF/6-31G than for both HF-3c and experiment. Overall, we conclude that HF-3c is able to yield good geometries for the tested proteins. Because the method includes only minor empiricism and was not parametrized specifically for protein structures, we think that this conclusion holds in general and suggest it as a tool in structural bio-chemistry.

The health checks for PM6 geometries give worse results than those for HF-3c for most criteria. The number of clashcores and the percentage of poor rotamers is higher and the percentage of favored Ramachandran dihedrals is much smaller. The results for bond and angle outliers are slightly better than for HF-3c but overall the PM6 structures are not as good as the HF-3c ones. Additionally, in many cases the positively charged guanidinium group of the amino acid arginine is not planar when optimized with PM6.

In general HF-3c seems to predict too many hydrogen bonds (Figure 4.8). On average, the calculation yields six hydrogen bonds too much compared to the corresponding experimental structures. PM6 shows on average four excessive hydrogen bonds. The hydrogen bond search was done with USCF Chimera[267] applying default criteria.

To test the influence of the solvent (i.e. artificially neglected water molecules) on the observed hydrogen transfer and the excess of hydrogen bonds, five proteins (1ODP, 2EVQ, 2FBU, 2JTA and 2RLJ) were optimized with HF-3c using the COSMO model[40] for continuum solvation. The dielectric constant ϵ was set to 78 for pure water. For all optimizations including COSMO, considerably less hydrogen transfers are observed. 1ODP

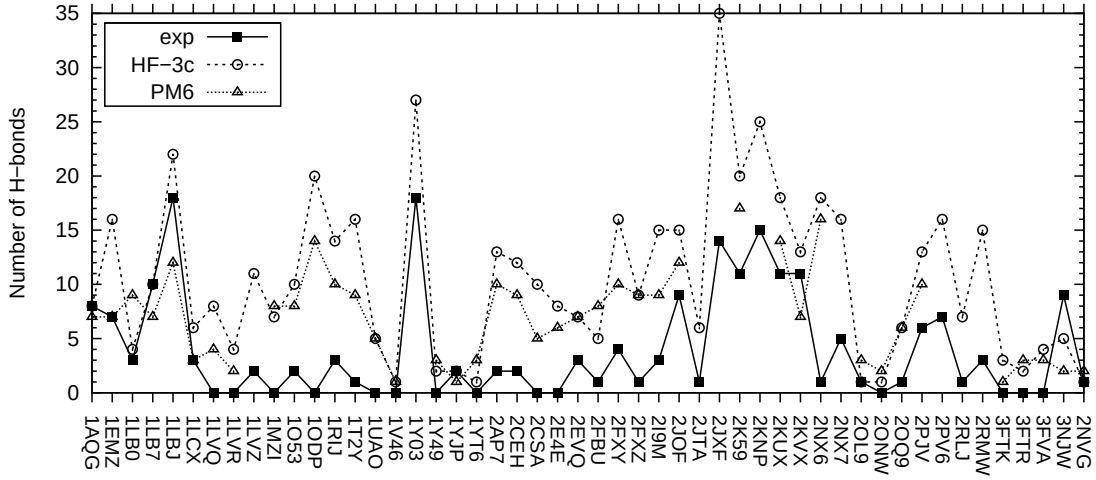


Figure 4.8.: Number of hydrogen bonds for the experimental, HF-3c and PM6 protein structures. The lines between the data points are drawn just to guide the eye.

and 2RLJ do not show a hydrogen transfer at all. For the other three proteins, the number of transferred hydrogens is reduced from two in case of 2EVQ and 2JTA and four in case of 2FBU to just one. Regarding the hydrogen bonds, only the 2FBU structure exhibits more H-bonds in the HF-3c-COSMO optimization than with plain HF-3c. The other four proteins exhibit two or three H-bonds less when optimized with COSMO. Nevertheless the number of computed hydrogen bonds is still higher compared to the experiment. Because HF-3c performs very well for the structures and energies of all hydrogen bonded systems in S22 and S66, it is not clear in how far this conclusion is based on biased experimental data instead of errors of the theoretical model.

The geometries of all five proteins improve regarding all health checks when using COSMO in the optimization (for explicit values see SI). In particular, the number of clashcores is reduced and the percentage of Ramachandran favored dihedrals is increased. Also the backbone RMSD relative to the experimental geometry is much smaller, i.e., it decreases by a factor of about two. The largest improvement was observed for 1ODP, its RMSD is reduced from 2.656 Å to only 0.956 Å . Thus, inclusion of the COSMO model in the optimization yields a further improvement to already good HF-3c protein "gas phase" structures.

4.4. Conclusion

A fast method based on a Hartree–Fock calculation with a small (in part minimal) basis set is presented (dubbed HF-3c from now on). Three corrections, namely the D3 scheme to include include long-range London dispersion and medium-range correlation effects, a geometrical counterpoise to handle intra- and inter-molecular BSSE and a short-range

term to correct basis set deficiencies for bond lengths are added to improve the plain HF energy. Detailed benchmarks for a variety of molecular properties were presented.

The method is able to yield good geometries for small covalently bound organic molecules, small noncovalent complexes as included in the S22 and S66 test sets as well as large supramolecular complexes. Fully optimized geometries of small proteins with up to 550 atoms yield good results in standard protein structure health checks and reasonable RMSD agreement compared to experimental structures.

By construction the method gives a physically sound description of noncovalent interactions which is reflected in accurate interaction energies for a variety of systems. The mean absolute deviation of the interaction energies compared with theoretical reference values is only 0.55 kcal/mol for the S22 and 0.38 kcal/mol for the S66 test set. For 12 supramolecular complexes, the fully *ab initio* computed association free energy has an MAD of 6.2 kcal/mol with respect to experimentally obtained values. The MAD for the corresponding gas phase interaction energies is 4.4 kcal/mol. To put this into perspective, dispersion corrected DFT methods yield MADs in the range of 2-5 kcal/mol while MP2/CBS yields an MAD of 16 kcal/mol[46] for the same set of realistic complexes. For the S66 set the MAD for the best DFT-D3/large basis variants and MP2/CBS are 0.2-0.3 and 0.45 kcal/mol, respectively[250].

Compared to widely used semiempirical approaches (PM6 and PM6-DH2 used here as typical examples), the presented Hartree–Fock based method is slower but generally more accurate, robust and numerically stable. It is easier to handle in large-scale geometry optimizations as shown by the protein studies. The method can be used routinely even on small desktop computers to optimize systems with hundreds of atoms and in parallel it can be applied to those with a few thousand atoms. Analytical vibrational frequency calculations are straightforward and the derived statistical thermodynamic corrections seem to be reasonable. Thus, the HF-3c method might be able to fill the gap between semiempirical and DFT methods in terms of cost and accuracy and is recommended as a standard quantum chemical tool in bio-molecular or supra-molecular simulations. Current work in our laboratory investigates its applicability for the computation of molecular crystals.

Acknowledgement

We thank Dr. Holger Kruse and Dr. Andreas Hansen for their help with the implementation of HF-3c into the ORCA programm suit. This work was supported by the Fonds der Chemischen Industrie.

Part III.

Evaluation of methodologies for supramolecular thermochemistry

Part III focuses on the application and evaluation of a multilevel methodology to obtain free energies of association for supramolecular host-guest complexes in solution. This approach was first published for a set of 12 host-guest systems, termed S12L set. This set contains six different typical host molecules, a tweezer, a pincer, a catcher, a macrocycle and two cucurbiturils, which are relatively rigid. Each host forms a complex with two different guest molecules (see Figure 4.9),[46] and for every complex only one binding conformation is taken into account. In order to obtain the association free energy ΔG_a of the host-guest systems in solution, three contributions are added. First, the association energy ΔE is computed for optimized structures on the DFT(hybrid)-D3/QZ level of theory in the gas phase. Second, a semiempirical method is used to calculate the thermostatistical corrections from energy to free energy ΔG_{RRHO}. And third, a continuum solvation model is employed to include solvation effects ($\Delta\delta G_{solv}$).

The methods of choice for the S12L set were PW6B95-D3/QZ for ΔE, DFTB3-D3 for ΔG_{RRHO} and COSMO-RS with the 2012 parametrization for $\Delta\delta G_{solv}$. With this combination a mean absolute deviation (MAD) of only 2.1 kcal mol^{-1} compared to the experiment was achieved.[46] It was demonstrated that using only ΔE in the gas phase is certainly not enough to explain experimentally observed binding affinities, and solvation effects are crucial to reproduce the measured results. This thesis further explores the multilevel approach with applications to realistic host-guest systems in order to evaluate its predictive power.

In Chapter 5 the general procedure is reviewed, and put into perspective by giving an overview of other approaches for various examples of supramolecular systems. To illustrate the multilevel methodology and its general applicability, eight typical host-guest complexes are studied. The chosen examples for the case study are two tweezer complexes, two pseudorotaxanes with either one or two crownethers as wheels, two cucurbit[7]uril complexes, and two complexes of the fullerene C_{70} in cycloparaphenylenes. In Chapter 4

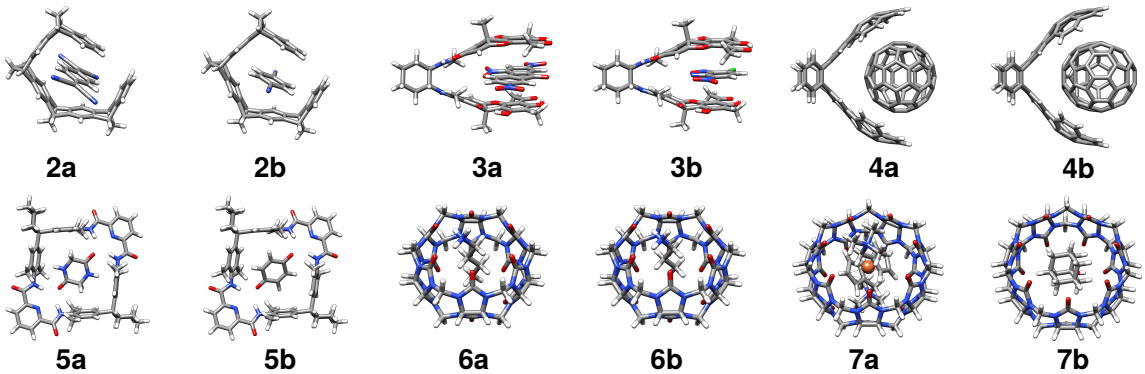


Figure 4.9.: Structures of the host-guest complexes contained in the S12L set.[46] **6a** and **6b** carry a positive, and **7a** has a two-fold positive charge.

it was shown, that HF-3c yields accurate thermostatistical contributions and thus, HF-3c is mainly used to compute ΔG_{RRHO} from here on.

With the participation in a blind test challenge this approach is used in a hard and realistic examination. This blind test challenge is the focus of Chapter 6. Without a bias towards known answers, achievements of diverse computational methods can be evaluated. In the 4$^{\text{th}}$ Statistical Assessment of Modeling of Proteins and Ligands (SAMPL4) relative association free energies of the rigid cucurbit[7]uril and 14 guest molecules that contain either one or two ammonia groups are to be predicted.[49] The HF-3c method is employed to sample possible binding conformations, and the ones with the highest binding affinities are then treated with higher level methodologies. Counterions are introduced and the geometries are re-optimized on the TPSS-D3/TZ level of theory. The PW6B95-D3$^{\text{ATM}}$/QZ level of theory is used to obtain the gas phase association energies, HF-3c is employed for thermal corrections, and COSMO-RS 2012 is applied to compute the solvation free energies for each compound. The relative ΔG_a values are submitted to the organizing committee of the SAMPL4 blind test, and compared to the results obtained by other participants by means of a statistical analysis. During the previous investigations on the S12L set, the case study presented in Chapter 5 and in the SAMPL4 blind test participation a few ΔG_a values were computed that deviate more than 4 kcal mol^{-1} from the experimental results. Frequently, these outliers are charged complexes and thus, further testing of the multilevel approach is necessary especially for charged systems. Chapter 7 presents the extension of the S12L test set to 30 host-guest systems (S30L), with the aim to include larger and slightly more flexible systems, higher charges (cations and anions) and a broader variety of interaction motifs. The ferrocene@cucurbit[7]uril complex of the S12L set (**7a**) was disregarded in order to provide a set that contains only organic systems since many semiempirical methods and force fields are not parametrized for treating metals. The influence of counterions (CI) is investigated by adding chloride for cations and sodium ions for anions to the structures of the charged systems and re-optimizing them including the COSMO model (S30L-CI). In order to obtain a theoretical best estimate for ΔG_a, different dispersion-corrected density functionals are tested for ΔE and combined with several semiempirical methods to calculate ΔG_{RRHO} and COSMO-RS with different parametrizations as well as SMD to compute $\Delta\delta G_{solv}$.

5. Using dispersion-corrected density functional theory to understand supramolecular binding thermodynamics

Jens Antony*, Rebecca Sure*, and Stefan Grimme *

Received 26th of August 2014, Published online 14th of November 2014

Reprinted (adapted) with permission from
J. Antony, R. Sure, and S. Grimme, *Chem. Commun.* **2015**, *52*, 1764–1774.
 DOI: 10.1039/C4CC06722C

Own manuscript contribution:

- Performance of the calculations
- Interpretation of data
- Writing parts of the manuscript

*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany

5. Using dispersion-corrected density functional theory to understand supramolecular binding thermodynamics

Abstract

A recently published theoretical approach employing a nondynamic structure model using dispersion-corrected density functional theory (DFT-D3) to calculate equilibrium free energies of association (*Chem. Eur. J.*, 2012, **18**, 9955–9964) is illustrated by its application to eight new supramolecular complexes. We compare with experimentally known binding constants which span the range from -3.3 to -20.3 kcal mol^{-1}. The mean deviation of calculated from measured ΔG_a results in 0.4 kcal mol^{-1}, the mean absolute deviation in 1.8 kcal mol^{-1} excluding two outliers for which the computed solvation free energies are identified as the largest error source. A survey of previous applications of the theoretical approach and related computational studies is given underlining its good accuracy. It is concluded that structures and gas phase interaction energies can be computed routinely with good to high accuracy (relative errors for interaction energies of 5-10%) for complexes with about 200-300 atoms.

5.1. Introduction

Noncovalent interactions are of crucial importance for the binding of small molecules to enzymes and receptors, the folding of proteins and DNA to their three-dimensional structures, and the orientation of substrates on surfaces and of molecules in molecular crystals.[28,29] Therefore, hydrogen bonding, cation-π and ion pair interactions, and London dispersion forces[3,24] have a central role in providing matter with unique functions and properties. Consequently, the understanding and control of noncovalent interactions has a prominent place in chemistry, biology, and materials science. A particular example to this is the area of host-guest and supramolecular chemistry, mimicking concepts like self-assembly and molecular recognition, that are operative in many biological processes, on a manageable scale in a well defined environment.[16,17]

Although supramolecular chemistry is essentially an experimental discipline, theoretical methods have a firm standing in the arsenal of tools for analyzing host-guest systems.[30] The methods of our choice for modeling supramolecular complexation are based on density functional theory (DFT),[82,83] as it combines the predictive power of a first-principles approach with computational efficiency allowing the routine treatment of molecules containing hundreds of atoms. Unfortunately, practically all common density functional approximations have well known deficiencies, among which their inability to account for long-range correlation effects is particularly obstructive when treating intermolecular interactions.[287–289] Numerous ways to fix this error were presented.[32,33]

In the mean time the dispersion problem of density functional theory is considered as being more or less solved in the sense that applications to realistic questions of chemical

interest like supramolecular systems that are the focus of this feature article are treated repeatedly and with reasonable accuracy.[290] However, the development of dispersion corrections[291–293] and of dispersion-including density functionals[294] is still a highly active area in order to properly treat effects like polarizability anisotropy, polarizability of delocalized π states, many-body effects beyond three atoms, resonant structures, dispersion interactions in polarizable solids, etc.

Our contribution to the field of dispersion corrections denoted as DFT-D3[36,37] has proven to provide accurate intermolecular interaction energies with a typical error below 5%.[86,250] Molecular dispersion coefficients are on average accurate to better than 5 % as well[32]. The error denotes the average performance of better functionals in the vicinity of the equilibrium distance, which is the relevant case for this work (see Fig. 4 in Ref. [250]Q). For less good performing functionals and for non-equlibrium distances the relative error is larger, and of course outliers can occur. Concerning the issue of scalability we refer to Ref. [192] where mean absolute percentage deviations of DFT-D3 from empirical reference energies for the S12L complexes of down to 5% are reported. Consequently, DFT-D3 has been applied recently to calculate host-guest binding affinities in solution with an unprecedented accuracy.[46] The key role of London dispersion interactions for the binding enthalpies of cucurbituril host-guest systems in water has very recently been emphasized also by Fenley *et al.*[295] using microsecond time scale classical molecular dynamics simulations.

After a short description of the theoretical approach, a survey of recent and related applications is given and then its performance is illustrated by new calculations on host-guest complexes. Thereby, both parts of this article are considered as exemplary rather than exhaustive and as not all relevant classes of supramolecular structures are covered by the eight complexes treated, one or the other important contribution might be omitted in the survey.

5.2. Computation of free energies of association with DFT-D3

Equilibrium association free energies ΔG_a for typical medium-sized supramolecular host-guest systems can be computed with good accuracy by DFT-D3 together with a relatively large atomic-orbital basis set in a nondynamic single-structure approach without any system-dependent adjustments or empirical corrections.[46,49,192,296] The target ΔG_a for a host-guest system in a given solvent X at a certain temperature T is calculated as the sum of three contributions, the electronic gas phase association energy ΔE, the difference in thermal corrections from energy to free energy ΔG^T_{RRHO}, and the difference in solvation

free energies $\Delta\delta G^T_{solv}(X)$:

$$\Delta G_a = \Delta E + \Delta G^T_{RRHO} + \Delta\delta G^T_{solv}(X) \tag{5.1}$$

Here, the Δ terms refer to differences in the process host + guest $\rightarrow$ complex while the δ indicates that the quantity is already a difference between two states. All molecules, host, guest and the complex, are fully optimized in the gas phase on an affordable DFT level together with the D3(BJ) (D3 with Becke-Johnson damping[133,134]) dispersion correction (e.g. TPSS-D3/def2-TZVP). The gas phase association energy ΔE of the optimized system is then calculated by employing a hybrid density functional with the D3(BJ) dispersion correction in combination with an extended quadruple-ζ basis set (e.g., PW6B95-D3/def2-QZVP') in the supermolecular approach, i.e. $\Delta E = E(complex) - E(host) - E(guest)$, with E being the total electronic energy. In addition to the two-body dispersion interaction energy $\Delta E^{(2)}_{disp}$ being operative upon geometry optimization the three-body contribution $\Delta E^{(3)}_{disp}$ is also taken into account in the single-point energy calculations. $\Delta E^{(3)}_{disp}$ was found to destabilize the complex by typically 1-3 kcal mol^{-1} and therefore cannot be neglected.[46,49] For a recent discussion of many-body dispersion effects see Refs. [297–299] The overall association energy ΔE is thus a sum of the pure electronic DFT energy ΔE^{DFT}_{el}, two-body $\Delta E^{(2)}_{disp}$ and three-body dispersion $\Delta E^{(3)}_{disp}$:

$$\Delta E = \Delta E^{DFT}_{el} + \Delta E^{(2)}_{disp} + \Delta E^{(3)}_{disp}. \tag{5.2}$$

The sum of thermal corrections from energy to free energy G^T_{RRHO} including the zero-point vibrational energy is obtained for each species in the gas phase at the given temperature T and a pressure of 1 atm by using a combined rigid-rotor-harmonic-oscillator approach. To calculate the vibrational entropy, low lying modes below 100 cm^{-1} are treated within a rigid-rotor model to reduce their error in the harmonic approximation, for details, see Ref. [46]. It has already been discussed in the original publication[46] that the applied rotor approximation for the entropy works much better than a standard treatment, since anharmonic calculations are currently absolutely impossible for such large complexes. The harmonic vibrational frequencies themselves are calculated with dispersion-corrected semiempirical methods like PM6,[124] DFTB[114,300,301] or our recently developed HF-3c method[205] (minimal basis set Hartree–Fock with three corrections). DFT-D3 and semiempirical frequencies provide very similar ΔG^T_{RRHO} values as shown in the original work.[46]

The solvation free energy $\delta G^T_{solv}(X)$ for each species is calculated by employing the DFT based COSMO-RS solvation model (conductor-like screening model for real solvents)[41,42] which is simply used as a black box with default parametrization. As an alternative, SMD

(universal solvent model based on the solute electron density)[54] is tested. COSMO-RS as well as SMD approximately include all non-electrostatic effects of solvation. In particular the standard state correction for the free energy is implicitly included in COSMO-RS. Whether the continuum solvation models are of sufficient accuracy for large complexes is a topic of actual research and this feature article. For small organic molecules the computed solvation free energies have an accuracy of 0.5 kcal mol^{-1} on average.[232]

Alternatives to COSMO-RS/SMD continuum solvation models like MM/MD or QM/MM/MD require the statistical sampling of the distributions of solvent and counterions. For a recent survey of force-field based free energy perturbation calculations of solvation free energies see Ref. [302]. Another route to solvation modeling is provided by couplings of QM with integral equation theory of liquids like 3D-RISM.[303,304]

5.3. Survey of previous calculations

5.3.1. Scope of the survey

The following survey of literature is on the application of quantum chemical methods to help understanding structure and energetics of noncovalent supramolecular association in recent years. The focus lies on density functional theory and wave-function based as well as semiempirical methods but not on approaches based on classical force fields. The size of the systems under investigation lies between those treated in common benchmark data sets for intramolecular interaction energies[199] (up to about 20 atoms) on the one hand and macromolecular systems such as protein-ligand complexes on the other (more than 500-1000 atoms). Consequently, no highly accurate theoretical reference data are available routinely and comparison is mainly made to experiment. Finally, neither metal-containing systems nor chemical reactions are considered below.

5.3.2. Recent studies in our group

The nondynamic structure approach as described above[46] has been applied successfully to several different host-guest systems of similar size as in the original work.[43,49,192,205,296,305] The calculated free energy of dimerization of the [5,15-diphenyl-10,20-bis[4-(*N*-methyl)pyridinum]porphyrin] dication in water yielded an excellent agreement with measurements (ΔG_{calc}=-6.9 vs. ΔG_{exptl}=-8.2 kcal mol^{-1}) when counterions (chlorides) are included in the calculations leading to overall charge neutrality.[43] The computed ΔG_a value of the bicyclophane-*t*Bu_6-HBC (hexa-*peri*-hexabenzocoronene) complex in THF solution of -7 kcal mol^{-1} agreed reasonably well with the experimental value of -2.5 kcal mol^{-1} obtained for two related somewhat larger bicyclic oligophenylene structures.[296] The inclu-

sion complex of the biscation $[Cp^*Ir(H_2O)_3]^{2+}$ with cucurbit[7]uril (CB7) being neutralized by association of two chloride anions was studied at the (ZORA)BLYP-D3(0)/AE(all electron)-TZP level suggesting that noncovalent concealment is way more favorable thermodynamically than the chelation of the Ir center by a pair of carbonyl oxygen atoms of CB7.[305]

Back-corrected intermolecular interaction energies for the S12L test set of supramolecular complexes were used to evaluate the performance of a range of modern dispersion-corrected density functional methods combined with quadruple-ζ basis sets when possible or with 1/2 counterpoise correction for triple-ζ and the three-body dispersion correction.[192] Most dispersion-corrected DFT methods as well as the M06 functional performed well, whereas standard MP2 theory was less accurate for these complexes. Semiempirical (or minimal basis HF) methods are useful for e.g., conformer screening.

The applicability of the HF-3c method, which is based on a Hartree–Fock(HF)/small basis set calculation, for free energies of association and rotamer sampling was tested for the twelve supramolecular complexes contained in the S12L) set.[205] The HF-3c method was able to yield good geometries and reasonable analytical frequencies for thermochemical corrections of supramolecular complexes. It gave good free energies of association (mean absolute deviation (MAD) of 6.2 kcal mol^{-1} w.r.t. experiment on the S12L test set), and provided a possibility for efficient rotamer sampling.

In the 2013 version of the SAMPL aiming at testing protein and small molecule modeling,[49] the supramolecular host-guest binding affinities for the cucurbit[7]uril host and 14 different amine guest molecules as well as for the octa-acid host (OA) and nine different carboxylic acid guest molecules were predicted.[306] For the CB7 host our submission was one of the top three among 20 submissions in all statistical analyses, whereas for the OA host it ranked in the second half among twelve submissions. In both cases the inclusion of counterions systematically improved the agreement with experiment.

5.3.3. Related studies by other groups

The interaction of nine small organic molecules and amolecular tweezer featuring two (+)-usnic acid moieties as tethered pincers and (1*R*,2*R*)-1,2-diaminocyclohexane as spacer was modeled using dispersion-corrected density functional theory approaches at the B97-D/def2-QZVP//B97-D/6-31G(d) level of theory including solvent effects through single-point polarizable continuum model (PCM) calculations.[307,308] The theoretical best estimates for the free complexation energies of two of the guests where X-ray structures have been obtained (-5.7 kcal mol^{-1} for 2,4,7-trinitrofluorenone and -2.2 kcal mol^{-1} for 7-chloro-4-nitrobenzoxadiazole), which were quite close to the experimental values of -2.3 and -1.3 kcal mol^{-1}, respectively.[307] Extension to a series of molecules complexed in

the same tweezer for which the experimental geometry was unknown and whose Gibbs energies fell in the narrow range from -2 to -1 kcal mol^{-1} was additionally complicated by uncertainties concerning the experimental conditions. However, it turned out that the best statistics (regression coefficient and standard deviation) in comparison to experiment was found with the complexation energies rather than with the Gibbs energies including solvent effects and entropic corrections, tempting the authors to suggest to predict experimental Gibbs energy results from complexation energies only.[308] Our results based on large basis set DFT-D3 calculations do not support such a treatment.

The complexes of C_{60} with the buckycatchers $C_{60}H_{28}$ and $C_{60}H_{32}S_8$ and with the respective pincers $C_{20}H_{10}$ and $C_{20}H_{10}S_4$ were studied by Zhao and Truhlar at the M06-2X/6-31+G(d,p)//M06-L/MIDI! level of theory.[309] The free energy of association for $C_{60}@C_{60}H_{28}$ in vacuum was calculated to be -8.6 kcal mol^{-1} and only 3.2 kcal mol^{-1} more negative than the experimental value measured in toluene. Consequently, the difference was ascribed to desolvation missing in the theoretical treatment. The other three structures were estimated to be unbound in vacuum which for $C_{60}@C_{20}H_{10}$ is consistent with the difficulty to detect this supramolecule experimentally. Due to flexibility of the host and uncertainties concerning the structure of the inclusion complexes, no computed free energies of association were reported in addition to those determined by ^{1}H NMR experiments in toluene-d_8 in Ref. [310] for the complexes of a tridental molecular clip with cyclotriveratrylene tether and three corannulene pincers and fullerenes C_{60} and C_{70}.

The encapsulation of C_{60} by three cycloparaphenylenes (9CPP, 10CPP, and 11CPP) was studied using density functional theory calculations at the M06-2X/6-31G(d) level of theory.[311] The encapsulation of C_{70} and a functionalised C_{70} by the cycloparaphenylenes 10CPP and 11CPP was treated with the same approach.[312] While the structural information obtained from energy-minimisation was in agreement with NMR results and the X-ray analysis, free energies of binding were overestimated almost by an order of magnitude (ΔG_{calc}= -236 vs. ΔG_{exptl}=-28 kJ mol^{-1} for C_{70}@10CPP and ΔG_{calc}= -211 vs. ΔG_{exptl}=-30 kJ mol^{-1} for C_{70}@11CPP), a discrepancy which most probably originated only partly from the solvation effect by toluene missing in the theoretical description. A possible explanation is the basis set superposition error (BSSE)at this level to which M0X functionals were shown to be very sensible.[313]

The complexation of five molecular tweezers based on oligo-Tröger's base derivatives with tetracyanobenzene was investigated by DFT calculations combined with a dielectric continuum solvent model on the one hand and by the potential of mean force approach using umbrella sampling and the weighted histogram analysis method with molecular dynamics simulations on the other.[314] Compared to experimental binding free energies determined by NMR titration, the DFT calculations correctly provided the observed trends in complex stability but overestimate the magnitudes of complexation energies. The over-

estimation might in part be ascribed to the relatively small atomic orbital basis sets used not fully exploring the basis set limit. The semiempirical PM6-DH2X method yielded better magnitudes of the binding energies but not the proper order. Although not in the focus of the current presentation it is indicated that the molecular dynamics simulations provided the most realistic Gibbs binding energies of the study.[314] This finding is in line with the computational results given in Ref. [315] for the free energies of binding of nine cyclic carboxylate guest molecules to the octa-acid host in the SAMPL4 blind test challenge[306] employing four different approaches. While the absolute binding affinities calculated using minimized DFT structures gave intermediate-quality results with MADs of 5-9 kJ mol^{-1} and R^2=0.6-0.8, depending on how the structures where obtained, standard free-energy perturbation calculations of relative binding affinities, performed at the molecular mechanics level, amounted to good results providing the best prediction (of twelve in total) submitted to SAMPL4.

The structure, the energy, and vibrational, electronic, and NMR spectra of the bucky-onions C_{60}@C_{180} and C_{60}@C_{240} were investigated using dispersion-corrected DFT methods in Ref. [316] Although, not being in the very focus of this survey, the work is included as another relevant example to illustrate the necessity of including dispersive interactions in the DFT treatment to obtain a meaningful description of the structure and energetics of the two systems investigated. This finding is far from self-evident, as the existence of e.g. the following studies demonstrates: an ab initio investigation of electronic structure, molecular electrostatic potential, and NMR chemical shifts in cucurbit[n]urils (n=5-8), ferrocene, and their complexes,[317] an investigation of binding patterns, NMR, and vibrational spectra of SF_6 in complex with CB6,[318] both at the B3LYP/6-31G(d) level of theory modeling effects of solvation through the self-consistent reaction field theory calculations incorporating the polarizable continuum model, and a recent density functional based investigation on structure, binding energy, and vibrational spectra of complexes between hexa- and penta-valent actinyls and functionalised cucurbit[5]uril on the B3LYP/TZVP//BP86/def2-SV(P) level of theory (for the valence orbitals of the actinide ions, a def-SV(P) is taken for geometry optimizations and def-TZVP for energetics, while the core orbitals are modeled via the def-ECP pseudo potential).[319]

Sundararajan investigated the binding of methane, ethane, butane, isobutane, n-pentane, cyclopentane, neopentane, and n-hexane to cucurbit[6]uril with dispersion-corrected DFT calculations.[320] Calculated binding free energies at the B97-D/TZVP//BP86/def2-SV(P) level forethane, butane, isobutane, and n-pentane in vacuum deviated from the experimental free energies by 1.23 kcal mol^{-1} (mean absolute error). Though incorporation of solvent effects did not alter the overall binding affinity trends it might have been instructive to report the effect of COSMO on the absolute calculated values as well. Using the dispersion-corrected PM7 semiempirical method, binding of long chain alkanes (up to

diammonium-C_{12}-alkane) to several supramolecular hosts (cucurbit[6]uril, cucurbit[8]uril, and a dimeric capsule of a tetraimide derived from resorcinarene) was also studied. Along similar lines, binding of oxazine-1 to two cyclodextrin hosts, namely β-cyclodextrin and γ-cyclodextrin, was examined theoretically in support of photophysical studies.[321] This time the COSMO continuum solvation model for water was invoked in estimation of the binding energies, but no calculated binding free energies were reported for direct comparison to the measured binding constant values.

The barriers for expulsion of atoms and small molecules (N_2, CO, H_2, Ar, Kr, Xe, H_2O) from an open C_{60} fullerenes cage (I20) and related molecular containers ($C_{40}H_{20}$, [5]beltene, cucurbit[5]uril) determined on the M05-2X/6-311+G(d,p)//M05-2X/6-31(d) level of theory were reported in Ref. [322] The SMD method with standard parameters was used to calculate free t energies of solvation when comparison was made to measurements in solution. Results of N_2@I20 and CO@I20 were compared with experimental kinetic data, yielding a better agreement for CO@I20 than for N_2@I20. The calculation for Xe@CB[5] in water provided qualitative agreement with the experiment.

DiLabio and coworkers applied dispersion-correcting potentials being developed for use with B3LYP/6-31+G(2*d*,2*p*) amongst others to determine complexation energies of the complexes in the S12L test set with a mean absolute error of 2.6 kcal mol^{-1} but did not extend the approach to the calculation of free energies.[186] Bachrach reported structures and binding energies for the complex of the so called ExBox^{4+} host with the aromatic hydrocarbons benzene, naphthalene, anthracene, and tetracene guests using the ωB97X-D functional and the 6-31G(d) and 6-311G(d,p) basis sets in both the gas and the solution phases, the latter using the COSMO continuum model with acetonitrile as solvent.[323] The formation of all four complexes was calculated to be exergonic, but the binding was not as strong in solution as in the gas phase. Comparison of the computed solution-phase binding free energy of the complex with anthracene to the experimental value yielded a large discrepancy (ΔG_{calc}=-18.8 vs. ΔG_{exptl}=-4.01 $\pm$ 0.073 kcal mol^{-1}).

The tetracationic cyclobis(paraquat-*p*-phenylene) ring in complex with the [2]pseudorotaxane without and with four chloride counterions was studied considering B3LYP, PBE, X3LYP, M06-L, M06, M06-2X, and M06-HF density functionals and the Hartree–Fock method, using the 6-31G(d,p) basis set for geometry optimization and the 6-311++G(d,p) basis set for subsequent single-point energy calculations.[324] Solvent corrections were based on single-point self-consistent Poisson-Boltzmann continuum solvation calculations for acetonitrile. Comparison to the experimental enthalpy of formation showed that only the M06-class of functionals predicted structural and binding properties qualitatively correctly. DFT calculations at the level of M06-2X/6-311G and the Poisson-Boltzmann model for MeCN were carried out for the synthetic receptor Ex^{2}Box^{4+} in complex with anthracene, 9,10-anthraquinone, and 1,4-anthraquinone in order to understand their modes

of binding.[325] For the complex with anthracene, the co-conformation observed in the solid state was a local minimum, while the computed global minimum was 1.25 kcal mol^{-1} lower in energy. The position and orientation of 1,4-anthraquinone inside Ex^2Box^{4+} observed experimentally was in close agreement with the DFT prediction, while in the case of 9,10-anthraquinone the guest resided in one of the DFT local energy minima. The size of the relative energies compared to the complex of Ex^2Box^{4+} with four MeCN molecules determined by DFT calculations correlated with the measured binding parameters, but no calculated free energies of binding for direct comparison were given.

DFT calculations on the dispersion-corrected BP86-D/def2-TZVP//BP86-D/def2-SV(P) level of theory which were performed on two open-cage fullerenes and their water encapsulated complexes yielded barrier heigths for water entering the cages whose difference of 16 kJ mol^{-1} compared well to that observed in ^{1}H NMR experiments (13 kJ mol^{-1}) when the COSMO model with $\epsilon=\infty$ was used in the calculations.[326] However,a second transition state lying about 53 kJ mol^{-1} below the one considered for one of the cages was not used for the comparison. Employing the same level of theory, dispersion-corrected density functional calculations of binding and activation energies for the anion of one of the two open-cage C_{60} fullerene derivatives encapsulating one water molecule provided a barrier height of 110 kJ mol^{-1} which was in excellent agreement with the experimental value of 104±4 kJ mol^{-1} determined from the temperature dependence of the ion kinetics of the water evaporation reaction. The computational result was hardly affected by the charge of the complex.[327]

Dieckmann and Houk reported a benchmark of DFT calculations performed on the dispersion-corrected M06-2X/def2-TZVPP//B97-D/6-31G(d,p) level of theory to estimate free energies of complex formation by comparison to experimental values for six charge transfer complexes and a hydrogen bonded model system.[328] Subsequently, the authors studied the energetics of complex formation in a number of artificial self-replicating systems by this procedure.[329] While the stability of the termolecular complexes relative to the bimolecular complexes was represented well by the computed energetics of complex formation, inclusion of entropic effects reversed the balance in five of eight cases considered.

The binding of the trihalomethanes chloroform, bromoform, and fluoroform as well as tetrachloroethane to a C_3-symmetric imidazole-containing cyclopeptide cavitand and its building blocks was studied on the M05-2X, B3LYP, and B3LYP-D3 levels of theory with a mixed 6-31G(d)/cc-pVTZ basis set for structure optimization and 6-311++G(d,p), aug-cc-pVTZ for B3LYP single-point calculations.[330] The calculations revealed that a very high dispersion energy is responsible for the measured complex stability, supported by DF-DFT-SAPT for host-guest model complexes, but no computed free energies of binding were given. The same computational procedure was applied to the

cryptophane-E and cryptophane-A host molecules proving their ability to enclose chloroform, dichloromethane, tetrachloroethane, chlorobenzene, bromobenzene, and dichlorobenzene due to energetic stabilisation by dispersion interactions.[331]

By using diffusion Monte Carlo calculations, Tkatchenko and coworkers reported reference binding energies for two supramolecular complexes of the S12L benchmark set: the buckyball catcher complex ($C_{60}@C_{60}H_{28}$) and glycine anhydride interacting with an amide macrocycle.[332] The results were used to assess the performance of their dispersion-corrected DFT methods and were compared to the back corrected binding energies derived from the experimental free energies of association by removing enthalpic and entropic contributions, with which they agree within 2.2 and 1.4 kcal mol^{-1}. The approach was extended to four more complexes of S12L, the tetracyanoquinone-tweezer, 1,4-dicyanobenzene-tweezer, butylammonium-cucurbit[6]uril cation, and 1-hydroxyadamantane-cucurbit[7]uril, with differences from the back corrected binding energies in the range from 1.4 to 3.6 kcal mol^{-1}.[333] The results were compared to binding energies for the PBE functional including long-range correlation energy calculated from coupled atomic response functions. The binding energies obtained by the latter method for the entire S12L database were reported to deviate from the back-corrected experimental association free energies by 1.6 kcal mol^{-1} on average (mean absolute error), but the approach was not incorporated into a scheme for directly determining association free energies.[334] Reliable estimates for the binding energies of the S12L complexes are tried to approach towards by accompanying the DFT-SAPT results with those from the nonlocal DFT (NLDFT) and MP2 coupled (MP2C) methods in Ref. [335]

Continuing an investigation on the supramolecular homodimers, mixed dimers, and complexes with C_{60} and C_{70} formed with corannulene, sumanene, and pentaindenocorannulene,[336] Denis reported gas phase complexation energies and free energies for C_{60} and C_{70} interacting with chrysaorole and pentaindenocorannulene and the buckycatchers that employ the latter as pincers at the M06-2X/6-31G level of theory[337] and with 26 receptors for C_{60}, 15 had been synthesized and 11 were proposed by the author, and 22 for C_{70}, four of them designed theoretically, at the M06-2X/6-31G and M06-2X/6-311G levels.[338] The effect of solvation in toluene on the formation of the supramolecular complexes was evaluated employing the solvation model developed by Tomasi, Barone and coworkers.[339–344] Results on relative complexation strengths were utilized to make suggestions regarding the synthesis of these receptors, but lacking a suitable reference in most cases, no final judgement of the accuracy of the reported data was done.

Joseph and Masson studied complexes of three biphenyls and cucurbit[7]- and cucurbit[8]uril by DFT calculations at the TPSS-D3(BJ)/def2-TZVP level.[345] Instead of calculating binding affinities in comparison to their measured data, the authors focused on torsional barriers by using dispersion-corrected B3LYP and triple-ζ doubly polarized def2-

TZVPP basis sets, with enthalpic and entropic corrections at nonzero temperatures. The deviation of the experimental torsional barriers inside CB7 and CB8]from the calculated barrier of the free guest in gas phase and in aqueous solution did not exceed 1.4 kcal mol^{-1}.

5.4. Results of illustrative calculations on eight example complexes

In this section we present new results for eight supramolecular complexes (see Figure 5.1) to illustrate the performance of our method for obtaining association free energies as described above. The chosen examples, which have not been treated with our approach yet, contain neutral as well as charged molecules, show the typical supramolecular interactions such as hydrogen bonding, π-stacking, nonpolar dispersion dominated binding and cation-dipolar interactions, and cover a wide range of measured ΔG_a values from -3.3 to -20.3 kcal mol^{-1}.

We investigate the following systems: Two tweezer complexes with 1,2,4,5-tetracyanobenzene (TCNB) and 1-fluor-2,4-dinitrobenzol (FDNB) as guest molecules (**1** and **2** that were measured in $CDCl_3$ and are calculated in $CHCl_3$),[346] two pseudorotaxanes with N,N'-(anthracene-9,10-diylbis(methylene))bis(1-phenylmethanaminium) (ADMPA) as axle and one or two dibenzo[24]crown-8 (DB24C8) wheels (**3** and **4** measured in a 1:1 mixture of CD_3CN and CD_3NO_2 and calculated in a 1:1 mixture of CH_3CN and CH_3NO_2),[347] two cucurbit[7]uril (CB7) complexes with adamantan-1-aminium ($AdNH_3^+$) and adamantan-1-ylmethanaminium ($AdCH_2NH_3^+$) as guests (**5** and **6** in H_2O),[348] and two complexes of the fullerene C_{70} in [10]cycloparaphenylene (10CPP) and [11]cycloparaphenylene (11CPP) (**7** and **8** in toluene).[312]
The tweezer complexes **1** and **2** and the fullerene complexes **7** and **8** are mainly dominated by π-stacking and nonpolar dispersion interactions. The pseudorotaxanes **3** and **4** and the CB7 complexes **5** and **6** mostly feature hydrogen bonding and cation-dipole interactions.

As molecular symmetry influences the rotational part of the entropy and the effect of symmetry number σ is significant for highly symmetric molecules,[349] we treat the symmetric molecules in their respective point group: the tweezer host is C_2 symmetric, TCNB is D_{2h} and FDNB C_s symmetric, ADMPA has C_2 and the fullerene C_{70} D_{5h} symmetry. The other host and guest molecules and all complexes do not show any symmetry.

The tweezer complexes **1** and **2** and the fullerene complexes **7** and **8** are neutral. The CB7 complexes **5** and **6** are positively charged and the pseudorotaxanes **3** and **4** are twice positively charged. These systems are investigated with and without counter ions. For simplicity, chloride ions are chosen for all complexes although for **3** and **4** hexafluorophosphate was used in the experiment.[347] From calculations for CB7 complexes with

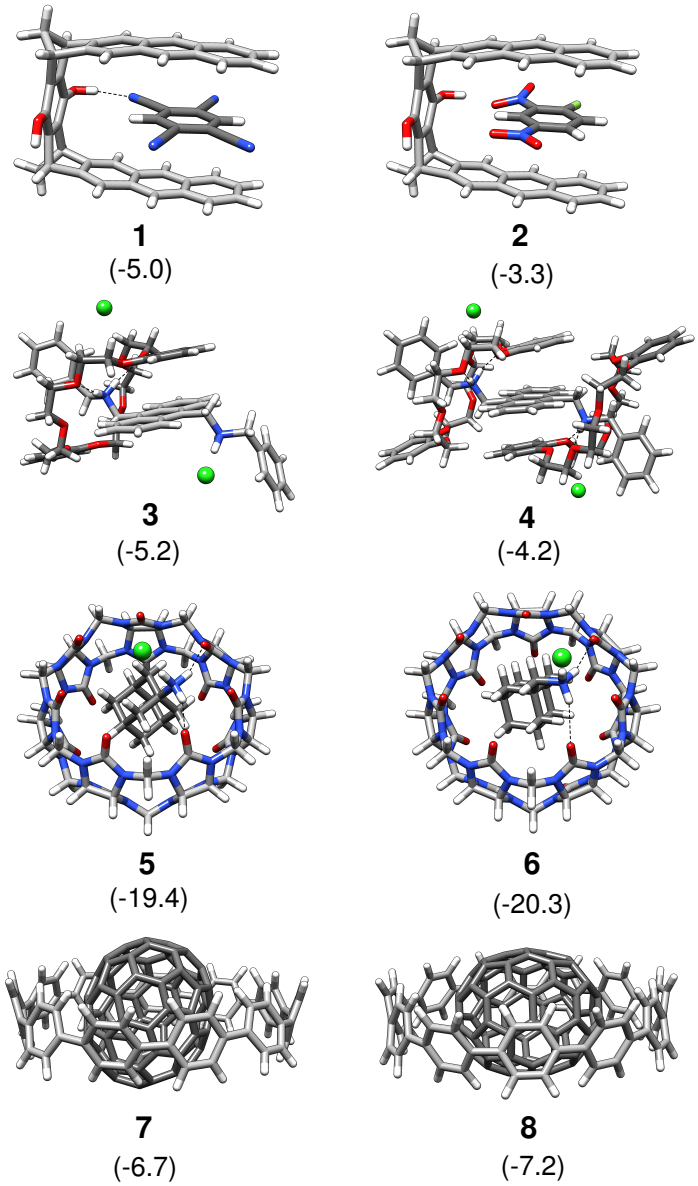


Figure 5.1.: Structures of the eight investigated complexes. Host molecules are shown in light grey, guest molecules in dark grey and H-bonding interactions are indicated by dotted lines. The atoms in green represent the chloride counter ions included in the calculations Experimental free energies of binding are reported in kcal mol^{-1} in parentheses.

ammonium guests similar to **5** and **6** we concluded that the inclusion of chloride counterions significantly improves the results.[49] Also, counterions were found to be necessary in calculations of multiply charged species when using the COSMO-RS solvation model.[43]

On the PW6B95-D3/def2-QZVP' level the calculated ΔG_a values for the CB7 complexes **5** and **6** show no improvement when chloride counterions are included. Without counterions both complexes are slightly overbound, **5** by 3.6 kcal mol^{-1} and **6** by 0.6 kcal mol^{-1}. These results for the charged species are already very accurate. Chloride

ions shift the ΔG_a values so that the complexes now become slightly underbound, **5** by 3.8 kcal mol^{-1} and **6** by 1.9 kcal mol^{-1}, but the error compared to the experiment is similar. In case of the complexes **3** and **4** the improvement is tremendous. Without counterions ΔG_a for both complexes is highly overestimated by -12.3 kcal mol^{-1} and -14.0 kcal mol^{-1}, respectively. When chloride ions are included, **3** is overbound by 3.8 kcal mol^{-1} and **4** is now underbound by 1.3 kcal mol^{-1} and the error is drastically reduced. In the following, we will only discuss results including counterions for these four complexes, although no improvement is observed for **5** and **6**.

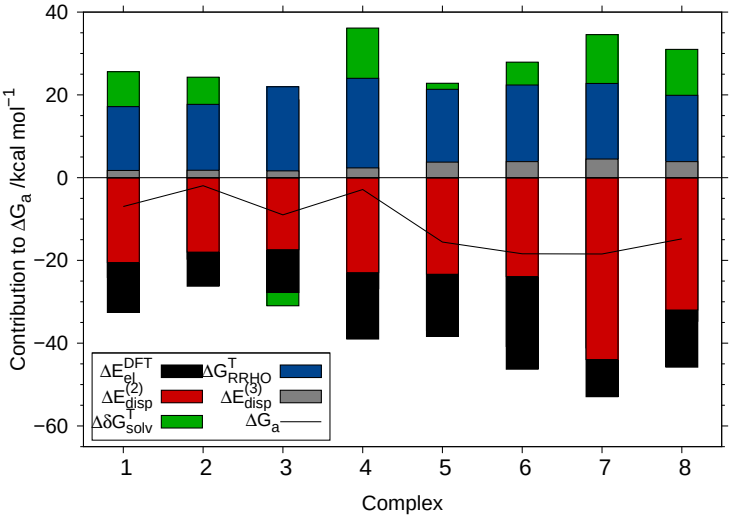


Figure 5.2.: Contributions to ΔG_a for all complexes (pure electronic energy (ΔE_{el}^{DFT}), two-body ($\Delta E_{disp}^{(2)}$) and three-body ($\Delta E_{disp}^{(3)}$) dispersion energy, thermal corrections from energy to free energy (ΔG_{RRHO}^{T}) and the solvation free energy ($\Delta \delta G_{solv}^{T}$) calculated by using the COSMO-RS solvent model) and total ΔG_a values on the PW6B95-D3/def2-QZVP'//TPSS-D3/def2-TZVP level of theory. For **3**, **4**, **5** and **6** chloride counterions are included.

The individual contributions to the total free energy of association on the PW6B95-D3/def2-QZVP' level of theory for the eight investigated complexes are shown in Figure 5.2. As all these values are large in magnitude, they have to have a small intrinsic error, otherwise adding them up would result in a highly defective ΔG_a. Most components in our partitioning are very important to achieve reasonable accuracy. Solvation alone would explain not even qualitative trends.

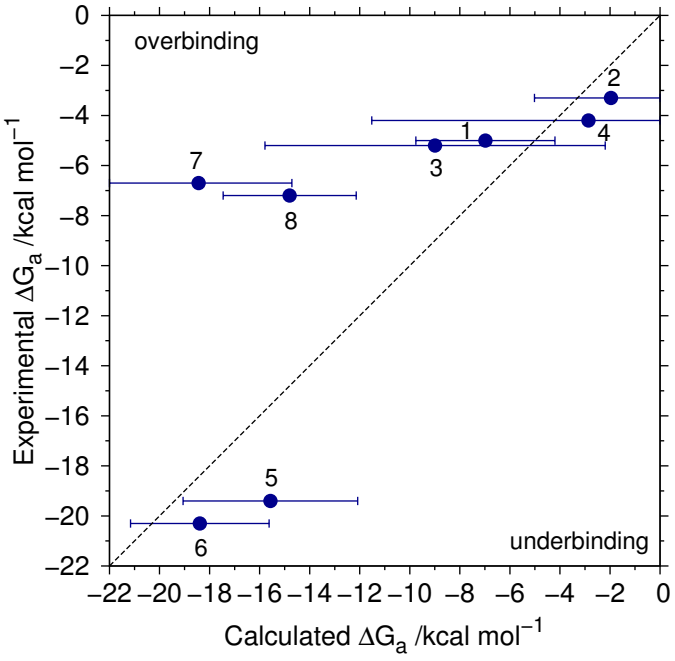
The total interaction energy ΔE for our set of complexes ranges from -26.2 kcal mol^{-1} for **2** to -48.5 kcal mol^{-1} for **7**. As expected, the contribution of the two-body dispersion energy $\Delta E_{disp}^{(2)}$ is large, at least as big as the pure electronic DFT energy ΔE_{el}^{DFT}, but for most complexes it is even higher. The outstanding example is **7**, whose two-body dispersion contribution to binding is about five times as large as the pure DFT interaction energy. In general, $\Delta E_{disp}^{(2)}$ ranges from -17.4 kcal mol^{-1} (**3**) to -44.0 kcal mol^{-1} (**7**),

whereas ΔE_{el}^{DFT} only lies between -8.2 kcal mol^{-1} (**2**) and -22.4 kcal mol^{-1} (**6**). The three-body contribution to the dispersion is always repulsive and lies between 1.6 for **3** to 4.5 kcal mol^{-1} for **7**. As typical supramolecular ΔG_a values are in the range 0 to -10 kcal mol^{-1},[17] this term is significant and cannot be neglected, especially, since it has the opposite sign than $\Delta E_{disp}^{(2)}$ and therefore shifts all ΔG_a values towards the less bound region. Altogether, this clearly shows that an accurate account of the dispersion energy is mandatory for obtaining best possible association energies.

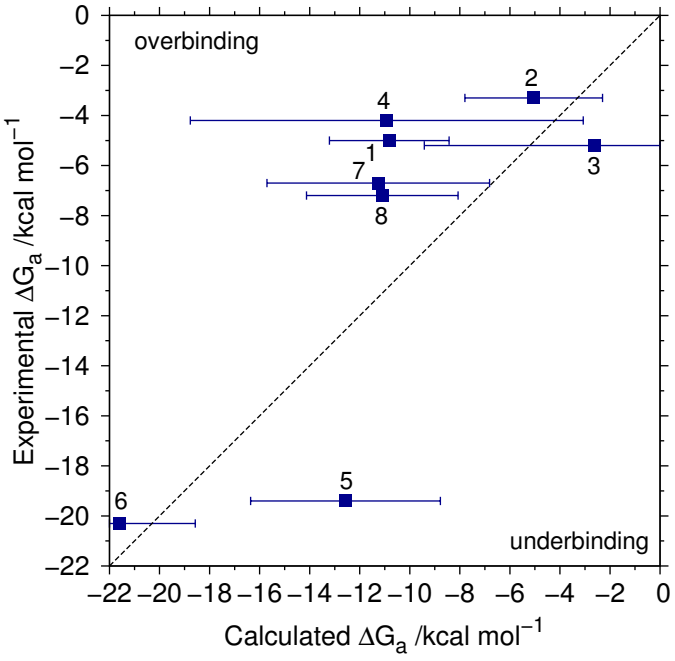
The thermal corrections from energy to free energy ΔG_{RRHO}^{T} on the HF-3c level lie between 15.4 and 21.6 kcal mol^{-1}. This is in the usual range that was observed before for complexes of similar size.[46,49] We also showed previously that HF-3c is sufficiently accurate for calculating the harmonic frequencies of such complexes.[205] Because it avoids any quadrature schemes, it is practically free of numerical noise which is of particular importance for low-lying vibrational frequencies.

The solvation contribution $\Delta\delta G_{solv}^{T}$ is positive for all complexes except **3** and ranges from 1.5 kcal mol^{-1} (**5**) to 12.1 kcal mol^{-1} (**4**). **3** shows a small negative $\Delta\delta G_{solv}^{T}$ amounting to -3.2 kcal mol^{-1}. We assume, that this term has the largest error in our calculation protocol for ΔG_a, since it depends on the chosen COSMO-RS parametrization and the COSMO-RS model itself breaks down for large molecular surface charges which appear for bare ions with localized electronic structure as in **3**, **4**, **5** and **6** (see also Ref. [115]). Further, it has been found before that SMD performs worse for ions than e.g. SM6 or SM8.[350]

The agreement between the total association free energies on the PW6B95-D3/def2-QZVP'//TPSS-D3/def2-TZVP level with the experiment is very good for the complexes **1** to **6**, see Figure 5.3 (a) for a direct comparison, but the complexes **7** and **8** are overbound. The error is -11.6 kcal mol^{-1} for **7** and -7.6 kcal mol^{-1} for **8** and will be discussed in the next paragraph. The next largest errors are found for complexes **3** and **5**, for which ΔG_a is overestimated by 3.8 kcal mol^{-1} and underestimated by 3.8 kcal mol^{-1}, respectively. Compared to the other rigid complexes, **3** and **4** contain one and two very flexible DB24C8 molecules, but even for those our nondynamic single-structure approach works fine. The mean deviation (MD) of calculated from experimental ΔG_a results in -2.1 kcal mol^{-1} (0.4 kcal mol^{-1} excluding **7** and **8**) and the mean absolute deviation (MAD) in 4.2 kcal mol^{-1} (1.8 kcal mol^{-1} excluding **7** and **8**)
In all our investigations conducted so far, the errors for **7** and **8** are the largest ones we observed.



(a) COSMO-RS



(b) SMD

Figure 5.3.: Comparison of experimental and computed total free association energies ΔG_a on the PW6B95-D3/def2-QZVP'//TPSS-D3/def2-TZVP level of theory with COSMO-RS (2012) (a) and SMD (b) as solvation models. Theoretical error bars were obtained via the sum of the following three estimates (absolute values): difference between the PW6B95/QZ' and TPSS/QZ' interaction energies, 5% of the ΔG^T_{RRHO} and 10% of the $\Delta\delta G^T_{solv}$ term. In the experimental work, very small errors were reported for **5**, **6**, **7**, and **8** (about 0.1 kcal mol^{-1}), while no uncertainties were given for complexes **1**, **2**, **3**, and **4**.

Among the eight complexes studied in this work, **7** and **8** contain the largest π-systems. A similar but less pronounced overbinding was already observed for the two buckycatcher-fullerene complexes in Ref. [46]
In order to shed light on this, we first calculate the association energy ΔE with two other density functionals, TPSS and B3LYP, and the same basis set to investigate, whether PW6B95 might give an exceptionally bad result in this case. As the deviations in ΔE lie in the range of only $\pm$(1-2) kcal mol^{-1} we can rule out any significant effect of the chosen density functional. However, as the three-body term is large for **7** and **8**, ΔE might be overestimated due to underestimation of further many-body effects that are not captured with the D3 correction. The importance of many-body effects beyond the three-body term has been pointed out before.[135,297–299] Nevertheless, as the D3 and MBD (many body dispersion) approaches are different, the conclusions drawn there cannot be transferred one-to-one to D3.

Next, we calculate DFTB3-D3 frequencies in addition to the HF-3c frequencies, but the deviations in ΔG^T_{RRHO} are only within 1 kcal mol^{-1}. Hence, as we assumed before, the COSMO-RS model to calculate $\Delta\delta G^T_{solv}$ seems to be the largest error source in our calculation protocol. We therefore employ six different COSMO-RS parameter sets (based on BP86/def-TZVP and BP86/def2-TZVPD from the years 2012, 2013 and 2014) and obtain values for $\Delta\delta G^T_{solv}$ in the range of 10.3 to 14.4 kcal mol^{-1} for **7** and 9.5 to 14.2 kcal mol^{-1} for **8**. These deviations are large compared to the total ΔG_a value and by far more than the 10% error that were assumed for estimating the theoretical error bars. For the other complexes the deviation with the parametrization is less pronounced but still about $\pm$2 kcal mol^{-1}. The largest values are obtained by the 2014 parametrization based on BP86/def2-TZVPD including a dispersion term similar to D3, which is quite important for a solvent like toluene and gives better results in this case. For comparison we employ SMD[54] as an alternative DFT based solvation model for calculating $\Delta\delta G^T_{solv}$. The solvation contribution for **7** and **8** is 18.9 and 14.8 kcal mol^{-1}, respectively, an thereby even higher than the result obtained by the 2014 COSMO-RS parametrization. Therefore the total ΔG_a is with -11.3 kcal mol^{-1} for **7** and -11.1 kcal mol^{-1} for **8** closer to the experimental values than using COSMO-RS and the 2014 BP86/def2-TZVPD parameterization, but still not very accurate. For complex **6** the ΔG_a value is also slightly better when SMD is used, but for the other three charged complexes it provides higher absolute deviations of up to 6.8 kcal mol^{-1} compared to experiment and hence, performs worse (see Figure 5.3 (b)). When using SMD, the MD for the calculated ΔG_a with respect to experiment is similar (-2.4 kcal mol^{-1}) and the MAD is larger (5.6 kcal mol^{-1}) compared to COSMO-RS. No general recommendation regarding the choice of the solvation model can be given, but for the presented examples COSMO-RS seems to be the better choice. Note, however, that whatever solvation model is used, the results are much more accurate

than they would be when solvation effects are ignored completely. This example indicates, that the most inaccurate part of our approach is the solvation free energy. Hence, one has to be cautious when using implicit solvation models like COSMO-RS or SMD and maybe should consult potentially more accurate methods like QM/MM in the future.

The association energy ΔE for all complexes was calculated also on the B3LYP-D3/def2-QZVP' and TPSS-D3/def2-QZVP' levels to show the influence of the density functional and on the TPSS-D3/def2-TZVP level to illustrate the effect of the basis set. The resulting ΔG_a for all applied methods are presented in Figure 5.4. For all complexes except **2**, **3** and **7** PW6B95/def2-QZVP' yields the most accurate ΔG_a values compared to experiment. For **2** and **7** TPSS-D3/def2-TZVP has a slightly smaller error and in case of **3** all other methods provide a better ΔG_a than PW6B95/def2-QZVP'.

Compared to PW6B95/def2-QZVP' all other tested methods result in a slightly smaller MD (-1.1 kcal mol^{-1} for TPSS/TZ, -1.4 kcal mol^{-1} for TPSS-QZ' and -1.6 kcal mol^{-1} for B3LYP/QZ') but a much larger MAD (5.1 kcal mol^{-1} for TPSS/TZ, 6.8 kcal mol^{-1} for TPSS-QZ' and 6.6 kcal mol^{-1} for B3LYP/QZ'). If **7** and **8** are excluded, PW6B95 has the smallest MD and MAD, while for the other methods the improvement is much smaller. The MDs for TPSS/TZ and TPSS/QZ' get even larger (1.8 and 3.1 kcal mol^{-1}). Hence, it is concluded, that similar to the S12L set the PW6B95 functional again gives the best results for ΔE.

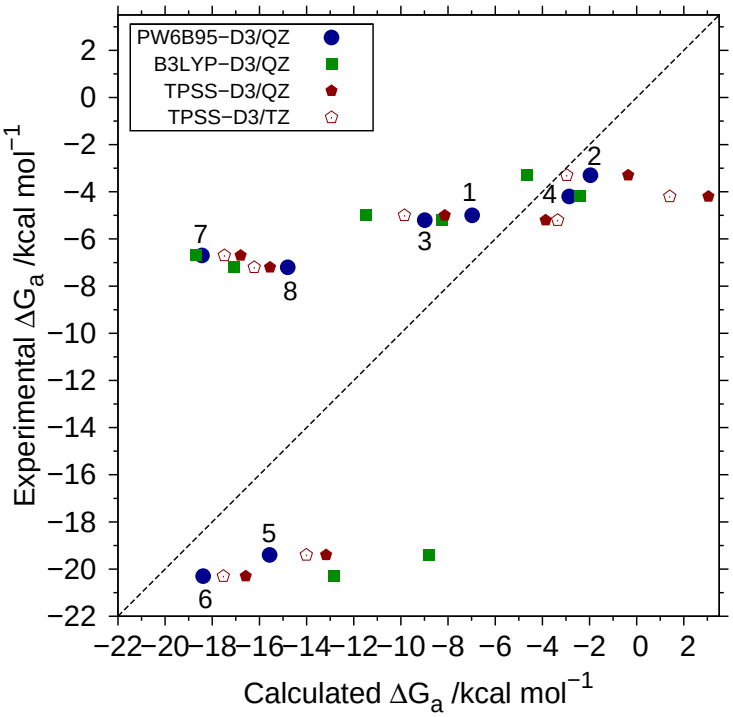


Figure 5.4.: Comparison of experimental and computed total free association energies for four levels of theory to calculate ΔE.

When comparing TPSS/TZ and TPSS/QZ' one clearly sees the effect of the basis set superposition error (BSSE). For all complexes except **3** the ΔE value for the quadruple-ζ basis is less negative by 1-2 kcal mol^{-1} than for the triple-ζ basis set. As the basis set incompleteness error (BSIE) has the opposite sign compared to the BSSE, ΔE is not necessarily less negative for the larger basis set calculation (see complex **3**). Thus, for a correct description without relying on error compensation a single-point quadruple-ζ basis set treatment is mandatory.
In order to check whether the quadruple-ζ results are converged we estimated the basis set limit with a two point exponential extrapolation[351–353] using the triple and quadruple-ζ energies and the coefficient fitted for the def2 basis sets as proposed by Neese et. al.[354] The interaction energies calculated from the extrapolated energies are at most 0.35 kcal mol^{-1} lower than the quadruple-ζ results, so we are are confident that a ΔE on the quadruple-ζ level is converged.

5.5. Computational details

HF-3c[205] was used for pre-optimizations and calculations of harmonic frequencies. Re-optimizations were done on the TPSS[93]-D3(BJ)[36,37] level employing the def2-TZVP basis set.[237] When chloride counter ions were included, the COSMO model[40] was used in the optimization. Single-point calculations for the energies in gas phase were obtained on the PW6B95[101]-D3(BJ), B3LYP[88,92,96,154]-D3(BJ), and TPSS-D3(BJ) levels together with the def2-QZVP' basis set (def2-QZVP with g- and f-functions discarded on the non-hydrogen and hydrogen atoms, respectively).[237]
All HF and DFT calculations were performed using the TURBOMOLE 6.4 program package.[261] The 3c-corrections to the pure HF energy and analytical gradient were calculated by our freely available standalone program.[241] Computations of harmonic vibrational frequencies were performed analytically using the *aoforce* code from TURBOMOLE. The 3c-contributions to the Hessian are computed numerically by two-point finite differences of analytical gradients.[241] For calculating ΔG^{T}_{RRHO} the HF-3c frequencies were scaled with a factor of 0.86.[205]

Additional frequency calculations were performed with DFTB-D3 using the full third-order correction and self consistent charges (SCC)[114,300,301] together with the most recent Slater–Koster files provided by the group of M. Elstner.[355] D3 parameters for DFTB were fitted on the S66 test set.[227] The DFTB-D3 frequencies were used unscaled.

In case of the DFT calculations the resolution-of-identity (RI) approximation for the Coulomb integrals[74] was applied using matching default auxiliary basis sets.[238] For the integration of the exchange-correlation contribution the numerical quadrature grid *m4*[239] was employed. For the geometry optimizations as well as the single-point calculations the

default convergence criteria were used. The three-body contributions to the dispersion energy were calculated using the dftd3 program.[241]

The COSMO-RS solvation model[41,42] was used as implemented in COSMOtherm employing six different parameter sets (BP86/def-TZVP and BP86/def2-TZVPD from the years 2012, 2013 and 2014).[356] To obtain the solvation free energies the standard procedure with two single-point calculations (one in the gas phase and one in an ideal conductor with $\epsilon = \infty$) on the default BP86[91,265]/def-TZVP[266] or the BP86/def2-TZVPD[237] levels of theory were performed on the optimized geometries and then used as input for COSMOtherm. The results presented in Figure 5.3 (a) refer to the 2012 BP86/def-TZVP parametrization. SMD[54] calculations based on COSMO charges were performed with the implementation in the current development version of ORCA 3.0[234]. For the solvent mixture of **3** and **4** the solvation free energies were averaged.

All visualizations of molecules were done with USCF Chimera version 1.8.1[267] and all graphs were plotted with Gnuplot 4.4.[357]

5.6. Conclusions

In this feature article we give an overview of the authors' and other groups research on using dispersion-corrected density functional theory to understand supramolecular binding thermodynamics. Although almost being a truism to date, we reiterate here as most important finding that proper modeling of the London dispersion energy is crucial when applying density functional theory in the study of intermolecular interactions in a quantitative way. The contribution of dispersion to the complexation energy is significant sometimes reaching or even exceeding 100% of the total gas phase interaction energy. A corollary to this is the necessity of using large basis sets as close to the complete basis set limit as affordable. Commonly applied double-ζ type basis sets are therefore not recommended. This conclusion is even more relevant for potential users of wave-function theory methods (e.g. MP2 or CCSD(T)) because an accurate account of dispersion by these methods requires even more extensive basis sets including multiple sets of diffuse functions. The relatively fast convergence of the dispersion-corrected DFT interaction energy with basis set size to a limit of very reasonable accuracy is probably its best 'selling argument'.

A second major prerequisite for a successful *ab initio* based modeling of binding thermodynamics is that the structure of the supramolecular complex is represented well by a single conformer or maximally a few most important conformations being either known from experimental evidence or unambiguously identified through conformation search. Generally, structures are better reproduced than energies if not too many too closely spaced conformational alternatives are accessible to the host-guest complex. Furthermore,

semi-local density functionals which are used for optimization perform best when not too large, localized charges are present. This is related to the so called self-interaction error of approximate density functionals and this problem is largely avoided by using hybrid- or range-separated hybrid functionals.

Finally, it is of utmost importance that the conditions under which the experimental free energies of binding are recorded and the assumptions of the modeling study match. For instance, it must be excluded that auto-association of the guest disturbs the measured host-guest association constants[308] or that complexation is hindered by a too high barrier of activation.[320] Inclusion of counterions generally improves the results as charge neutrality is obtained although the flexibility of the host-guest complex is enhanced.[43,49] And last but not least, solvation effects have to be properly taken care of. Implicit solvation models have their drawbacks, but without any solvation model the calculation of reasonable association free energies is not possible in most cases.

Acknowledgement

Financial support by the German research foundation by grant AN 793/1-2 is gratefully acknowledged.

6. Blind prediction of binding affinities for charged supramolecular host–guest systems: achievements and shortcomings of DFT-D3

Rebecca Sure*, Jens Antony*, and Stefan Grimme *

Received 26th of November 2013, Published online 3rd of March 2014

Reprinted (adapted) with permission from
R. Sure, J. Antony, and S. Grimme, *J. Phys. Chem. B* **2014**, *118*,3431–3440 .
 DOI: 10.1021/jp411616b

Own manuscript contribution:

- Performance of the calculations for the CB7 host
- Interpretation of data for the CB7 host
- Writing major parts of the manuscript

*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany

6. Blind prediction of binding affinities for charged supramolecular host–guest systems: achievements and shortcomings of DFT-D3

Abstract

Association free energies ΔG_a are calculated for two different types of host-guest systems, the rigid cucurbit[7]uril (**CB7**) and the basket shaped octa-acid (**OA**), and a number of charged guest molecules each by quantum chemical methods from first principles in the context of a recent blind test challenge (SAMPL4). For **CB7** the overall agreement between theory and experiment is excellent. In comparison with all other submitted calculated relative $\Delta G_{a,rel}$ values for this part of the blind test, our results ranked on top. Modeling the binding free energy in the case of the **OA** host mainly suffers from the problem that the binding situation is undefined with respect to the charge state and due to its intrinsic flexibility the host-guest complex is not represented well by a single configuration, but qualitative features of the binding process such as the proper binding orientation and the order of magnitude of ΔG_a are represented in accord with the experimental expectations even though an accurate ranking is not possible.

6.1. Introduction

Noncovalent interactions, such as dispersion interactions, hydrogen and halogen bonding or $\pi \cdots \pi$ stacking play an important role in structural biology and supramolecular chemistry.[25–27] They control the structures of proteins and DNA, host-guest and enzyme-substrate binding, antigen-antibody recognition, or the orientation of molecules on surfaces or in molecular crystals.[16,17] Association free energies ΔG_a for typical supramolecular host-guest systems can be computed with good accuracy by dispersion corrected density functional theory (DFT-D3) together with a relatively large basis set as demonstrated recently[46,192]. Blind tests provide a unique opportunity to test a method in a realistic setting without a bias towards known answers. One of those blind test challenges is the Statistical Assessment of Modeling of Proteins and Ligands (SAMPL).[48] The latest challenge, SAMPL4, consisted of three different parts: (1) the prediction of hydration free energies, (2) the prediction of protein-ligand binding constants and (3) the prediction of association free energies of supramolecular host-guest systems. The results presented in this article were obtained while participating in the latter part on host-guest binding affinities and submitted to the SAMPL committee before publication of the experimental results. After the submission deadline and the statistical assessment of all results, the so far disclosed experimental binding affinities were made available to the participants. An overview of all the results, calculations as well as experiments, will be given by the SAMPL4 committee in an upcoming article.[49]

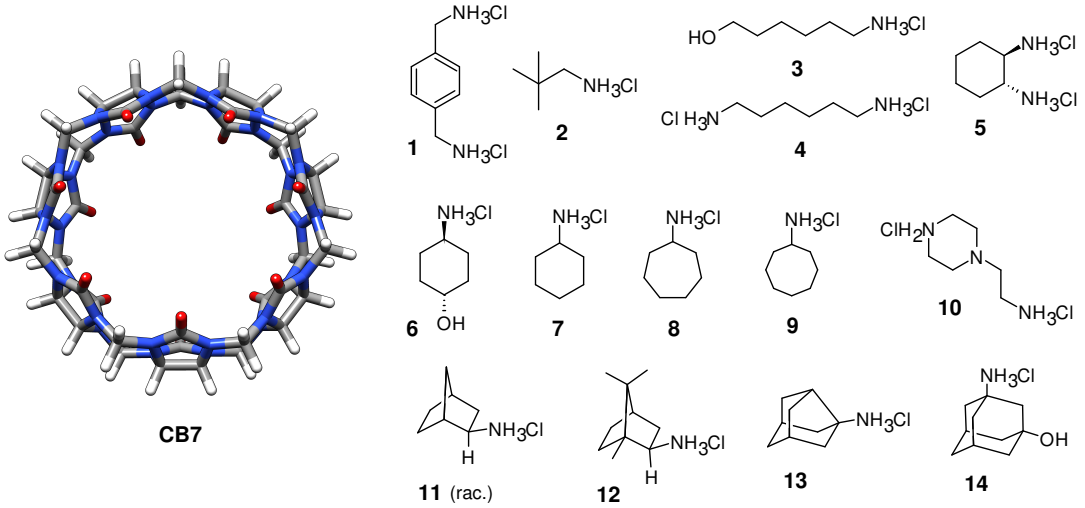


Figure 6.1.: Host and guest molecules for the **CB7** host system. Guest **1** is the reference compound.

Association free energies ΔG_a had to be predicted for two different types of host systems and a number of guest molecules each. The first host is the rigid cucurbit[7]uril (**CB7**), which is known for its ability to bind guest molecules in aqueous solution with a very high binding affinity (ΔG_a as large in magnitude as -20 kcal mol^{-1}).[50,53,358] The highest binding constants have been observed for guests with a rigid hydrophobic core, such as ferrocene[53], adamantane[50,358] or [2.2.2]bicyclooctane[358] in combination with cationic ammonium groups, that can bind to the carbonyls at the rims of the host. The guest molecules (**1**-**14**) for this host also have at least one ammonia group in combination with various cores, see Figure 6.1. The experimental binding affinities for this part of SAMPL4 were obtained via competition experiments with respect to guest **1** monitored by ^{1}H-NMR spectroscopy at pH 7.4.[359] Hence, it is expected that all ammonia groups are likely to be protonated in solution and that we have to treat positively charged guest molecules.

The second host is a basket shaped octa-acid (**OA**)[360,361], which has four flexible propionate side chains bearing two rotatable single bonds each and, hence, is not as rigid as a cucurbituril. The guest molecules for this host are several carboxylic acids (benzoic acid **15**, 4-methyl-benzoic acid **16**, 4-ethyl-benzoic acid **17**, 4-chloro-benzoic acid **18**, 3-chloro-benzoic acid **19**, cyclohexanecarboxylic acid **20**, *trans*-4-methyl-cyclohexylcarboxylic acid **21**, cyclopentylcarboxylic acid **22**, and cycloheptanecarboxylic acid **23**), see Figure 6.2. Since dimers of **OA** are known to encapsulate steroids up to the size of estradiol and estriol[361], all of the guests **15-23** are expected to be easily accomodated by this host in various orientations. The experimental binding affinities determined by ITC and NMR were measured at high pH (9.2),[362] so the octa-acid host might be fully deprotonated having a maximum charge of -8. However, also the charge of -6 with two of the carboxylic acids at the bottom of the basket taken to be diagonally across from each other

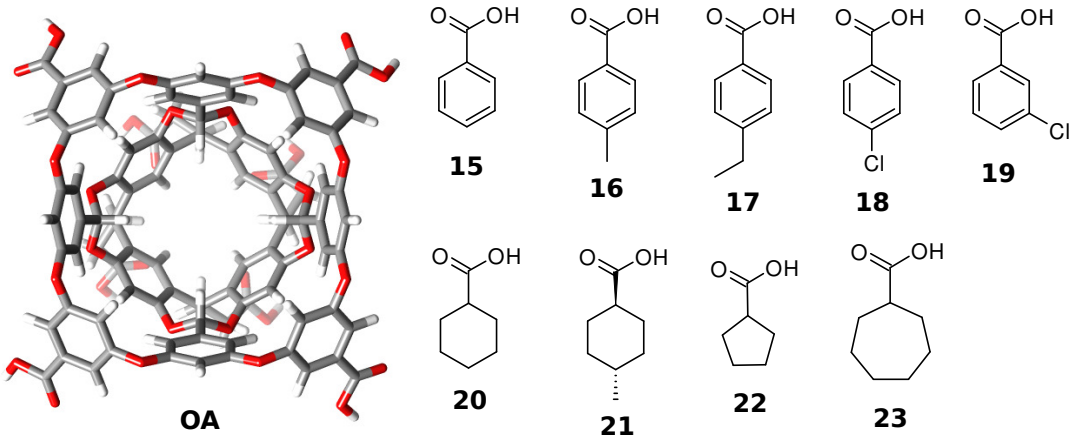


Figure 6.2.: Host and guest molecules for the **OA** host system.

being protonated was suggested[363]. Thus, the charge of the octa-acid is not completely clear. As the guests may be fully deprotonated as well under the experimental conditions, one is faced with the unusual case of anion-anion binding, which represents a huge challenge for quantum chemical methods due to the inherent electrostatic repulsion. Solvent screening is expected to play an important role in the process.

6.2. General approach to calculate association free energies

The equilibrium association free energies ΔG_a for the SAMPL4 host-guest systems in solution are calculated by quantum chemical methods. The theoretical approach employs a (nondynamic) single-structure model and has been successfully applied before to several different host-guest systems of similar size.[43,45,46]

The free energies of association for a host and guest molecule in a given solvent X at a certain temperature T is computed as:

$$\Delta G_a = \Delta E + \Delta G^T_{RRHO} + \Delta \delta G^T_{solv}(X) \tag{6.1}$$

The solvent in our cases is always water and the temperature was assumed to be 298.15 K. ΔE denotes the electronic gas phase association energy of the fully optimized host and guest molecules and is calculated by dispersion corrected density functional theory (DFT-D3(BJ), D3 with Becke-Johnson damping)[36,37,133] with extended basis sets (triple-zeta and quadruple-zeta quality). The solvation free energy $\delta G^T_{solv}(X)$ is calculated for each gas-phase species by employing the COSMO-RS solvation model.[41,42] G^T_{RRHO} is the sum of thermal corrections from energy to free energy within a rigid-rotor-harmonic-oscillator approximation for each molecule in the gas phase at a given temperature T and 1 atm,

including the zero-point vibrational energy. For obtaining the vibrational entropy from harmonic frequencies, low-lying modes below 100 cm^{-1} are treated within a rigid-rotor model in order to reduce their error in the harmonic approximation, for additional details see Ref. [46]. All harmonic frequencies were calculated with our recently developed HF-3c method.[205] HF-3c is a fast quantum chemical procedure based on a minimal(small) basis set Hartree–Fock (HF) calculation with three atom-pairwise correction terms (D3(BJ), geometrical counterpoise correction (gCP)[47] and an additional short-range term to correct for basis set deficiencies). To determine ΔG^{T}_{RRHO} the HF-3c vibrational frequencies were scaled with a factor of 0.86.

HF-3c was also used to pre-optimize several complexes with possible binding situations. The complexes with the lowest association free energy were then re-optimized on the TPSS[93]-D3(BJ)[36,37,133]/def2-TZVP[237] level of theory. Final single-point energies were calculated on the PW6B95[101]-D3(BJ)/def2-QZVP[237] level as in the original work.[46] We also included three-body-dispersion[36] which amounts to 2-3 kcal mol^{-1} for this type of systems as discussed recently[192]. No empirical modifications to the published first-principles procedure[46] were made.

6.3. Computational details

All HF/MINIX, TPSS-D3/def2-TZVP and PW6B95-D3/def2-QZVP' calculations (def2-QZVP with discarded g- and f-functions on the non-hydrogen and hydrogen atoms, respectively) were performed using the TURBOMOLE 6.4 program package.[261] In case of the DFT calculations the resolution-of-identity (RI) approximation for the Coulomb integrals[74] was applied using matching default auxiliary basis sets[238]. The numerical quadrature grid *m4* (grid 2 for the SCF iterations and grid 4 for the final energy)[239] was employed for integration of the exchange-correlation contribution. During the geometry optimizations the convergence criteria were set to 10^{-7} a.u. for the energy change and 10^{-5} a.u. for the maximum gradient norm. For the single-point energy a convergence criterion of 10^{-8} a.u. for DFT and 10^{-7} a.u. for HF calculations was used.

In the case of HF-3c, the 3c-terms to energy and analytical gradient were calculated by a freely available code which basically merges the two programs dftd3 and gCP.[241] Computations of harmonic vibrational frequencies were performed analytically using the *aoforce* code from TURBOMOLE. The 3c-contributions to the Hessian are computed numerically by two-point finite differences of analytical gradients. The three-body contributions to the dispersion energy were calculated using the dftd3 program.[241] The COSMO-RS solvation model[41,42] was used as implemented in COSMOtherm.[356] To obtain the solvation free energies the standard procedure with two single-point calculations, one in the gas phase and one with an infinite dielectric constant, on the default BP86[91,265]/def-TZVP[266] level

of theory were performed on the TPSS optimized geometries and then used as input for COSMOtherm.

6.3.1. Specifics for the CB7 host

We first used the protonated and hence positively charged species without any counter ions. Checking for several binding modes with HF-3c was done by chemical intuition and not by employing a special algorithm to search for conformers. For all guest molecules except **10@CB7** all ammonia groups are stabilized by hydrogen bonds. ΔG_a for **10@CB7** with guest **10** in its triply protonated state is unreliably high (-103.4 kcal mol^{-1} on the TPSS-D3/def2-TZVP level). It seemed reasonable to us to use the doubly protonated state for guest **10** (see below for further discussion).

We then included chloride counter ions by simply adding them to the structures at hand and re-optimizing the neutral molecules. Including counter ions was already found to be necessary in calculations of multiply charged species by COSMO-RS[43]. These optimizations are not reasonable in the gas phase and hence we used the COSMO solvation model[40] with the dielectric constant for water ($\varepsilon = 78$). The binding mode as well as the overall geometry did not change significantly when including counter ions. Since analytical frequencies cannot be computed when using COSMO and the numerical ones often show numerical noise we therefore used the frequencies obtained for the positively charged compounds in gas phase also for calculating ΔG^T_{RRHO} of the neutral complexes with counter ions in solution. For calculating the rotational part of the entropy C_2 symmetry was assumed for guests **1** and **4**, and C_1 symmetry for all others. Single-point calculations on the PW6B95-D3/def2-QZVP' as well as the TPSS-D3/def2-TZVP level were performed in the gas phase to avoid double counting of electrostatic solvation effects when applying COSMO-RS.

The absolute ΔG_a for the racemic mixture of compound **11@CB7** was calculated from the ΔG_a values from **11a@CB7** and **11b@CB7** via the Boltzmann average:

$$\Delta G_a(11) = -RTln\left(exp\left(-\frac{\Delta G_a(11a)}{RT}\right) + exp\left(-\frac{\Delta G_a(11b)}{RT}\right)\right) \tag{6.2}$$

6.3.2. Specifics for the octa-acid host

The octa-acid host is used in its fully protonated and thus neutral form while all guests are treated as singly negatively charged. Various binding modes are sampled with HF-3c in the following way: The guest is rotated from 0 to 180 degrees (polar angle) in seven steps and from 0 to 45 degrees (azimuth angle) in four steps inside the host with respect to its symmetry axis to generate 28 starting structures per guest for full geometry optimization with HF-3c. Harmonic frequency calculations on the optimized structures provide internal

(gas-phase) entropies. For determination of the rotational part of the entropy, a symmetry number according to four-fold symmetry of the host and a symmetry number according to two-fold symmetry of guests **15** to **18** is assumed. The symmetry number accounts for equivalent molecular conformations in the partition function. It is not necessarily given by the symmetry of the lowest energy conformation which for the empty fully protonated host is C_4 and for the deprotonated guests **15** and **18** is C_2, but the symmetry of guests **16** and **17** themselves is C_s at best. Guests **19** to **23** and due to a tilt of the guest with respect to the host all complexes have no symmetry and their symmetry numbers are 1.

The octa-acid host bears four propionate groups whose inner two C-C-C-C dihedral angles span a sizable conformational space. A complete sampling of all low energy structures is not yet practical on the level of theory used in this work. One possible strategy to deal with this flexibility is to truncate the host by substituting each $(CH_2)_2COO^-$ with one hydrogen atom. Alternatively and more closely preserving the electronic structure of the host in this work the reference structure of the uncomplexed host is obtained by reoptimization starting from its geometry in the complex thus eliminating the influence of intramolecular hydrogen bonds at the outside of the host on the estimated free energies of binding ΔG_a within the single-structure model. Preliminary estimates of ΔG_a are obtained by the COSMO-RS continuum solvation model based on DFT (BP86/def-TZVP) single-point energy calculations and TPSS-D3(BJ)/def2-TZVP single-point association energies. Host-guest structures with most negative ΔG_a are then fully optimized with TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) to obtain structures that account for desolvation more properly than in vacuo.

The final ΔG_a values obtained by single-point calculations with the PW6B95 meta-hybrid functional and the large def2-QZVP' basis set, the three-body dispersion contribution to the association energy, and redone COSMO-RS calculations on the TPSS-D3(BJ) structures, are finally supplemented by a neutral to ionic host correction. This correction is done by TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) single-point energy calculations on a host where the protons of the four benzoic acid carboxylate groups at the rim of the host are replaced by Na^+ ions in a relative geometry derived from a fully HF-3c optimized anionic structure of the host with twelve additional explicit water molecules. The relative Na geometries are the same for the nine different guests and are specified in the supporting information. The four propionate groups are not ionized because a fully deprotonated host of charge -8 is unlikely to change the correction term much. The difference between the interaction energies in the anionic and the neutral hosts $\Delta\Delta E$=$\Delta E(4Na^+)$-$\Delta E(4H)$ ranges from 0.8 to 2.1 kcal mol^{-1} with an average value of 1.5 kcal mol^{-1} (see Tables S6 and S7 of the supporting information). It is added to ΔE for the results submitted to the SAMPL4 blind test. Subsequent to the evaluation of the results by the SAMPL4 committee several alternative approaches to include ionic effects have been tested consid-

ering also the case of the fully deprotonated host. They are described below in the results section.

6.4. Results and discussion

6.4.1. The CB7 host complexes

We first considered the protonated and hence positively charged guest species without any counter ions and optimized the geometries and calculated the association free energy at the HF-3c level. Checking for several binding situations with the **CB7** host was done by trying to achieve the maximum number of hydrogen bonds for all complexes.

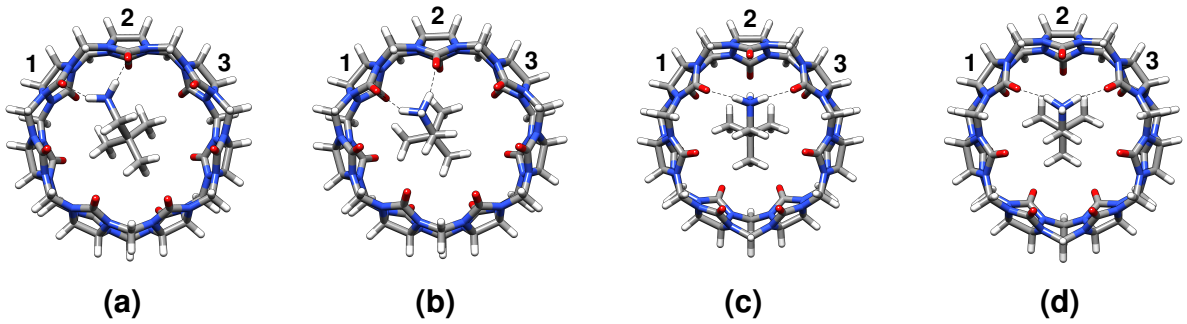


Figure 6.3.: Four different possible binding modes for the ammonia group of guest **2** to the host **CB7**. (a) 1,2-Binding with third H-atom pointing inwards the cavity. (b) 1,2-Binding with third H-atom pointing outwards the cavity. (c) 1,3-Binding with third H-atom pointing inwards the cavity. (d) 1,3-Binding with third H-atom pointing outwards the cavity.

For each ammonia group of the guest molecules there are four possible binding modes, shown in Figure 6.3. The first hydrogen forms a hydrogen bond with an arbitrary oxygen atom at the rim of the host. The second hydrogen can then bind to the neighbouring oxygen (1,2-binding, Figure 6.3 (a) and (b)) or to the next but one oxygen (1,3-binding, Figure 6.3 (c) and (d)). The third hydrogen is either pointing inwards (Figure 6.3 (a) and (c)) or outwards (Figure 6.3 (b) and (d)) the host cavity.

For those guests with two ammonia groups (**1**, **4** and **5**) we additionally checked whether both groups bind to the same rim or one to the upper and one to the lower one. **3**, **6** and **14** have a hydroxyl group in addition to the ammonia group. Hence, we had to test if hydrogen bonding with the host or hydrogen bonding with the solvent water molecules is favored and the OH group points outwards the cavity.

It was not possible to obtain all these different binding situations for all complexes because sometimes two different starting points lead to the same final geometry. When both geometries with the third non-bonding hydrogen atom of the ammonia group pointing inwards and outwards the host cavity could be optimized, the one with the hydrogen atom

pointing outwards is always the one with the lower free energy of association. Also in general the 1,2-binding mode of the guest molecules with the oxygens is favored over the 1,3-binding mode. From the guests with two ammonia groups, **1** and **4** bind to both, the upper and the lower rim of the host, whereas **5** forms hydrogen bonds with one rim only. From the guests with a hydroxyl group, only **14** shows a lower ΔG_a when the OH groups points outwards into the solvent. **3** and **6** form an additional hydrogen bond with an oxygen atom at the opposite rim.
Guest **10** is the only one with three ammonia groups and also the only one for which not all ammonia groups can be stabilized via hydrogen bonds. ΔG_a for **10@CB7** with **10** in its triply protonated state is unreliably high (-103.4 kcal mol^{-1} on the TPSS-D3/def2-TZVP level). **10** can only form two hydrogen bonds via the primary and secondary ammonia group. It is known that in aqueous solution tertiary amine groups are normally not protonated due to hydration effects.[364] So it seemed reasonable to us to use the doubly protonated state for guest **10**, which then yielded more reliable ΔG_a values.

The complexes with the lowest ΔG_a obtained at the HF-3c level were then re-optimized with TPSS-D3/def2-TZVP. If two conformers of one complex had similar ΔG_a values, we re-optimized both to be sure to find the one with the lowest ΔG_a on the TPSS-D3 level. The final geometries are shown in Figure 6.4. For three complexes, **2**, **6** and **11b**, the favored binding mode differs between TPSS-D3 and HF-3c. The hydrogen bonds in the HF-3c geometries show a 1,2 binding mode, whereas the 1,3 binding mode is present in the TPSS-D3 structures. For the other 12 complexes the average root mean square deviation (RMSD) between the HF-3c and TPSS-D3 geometries is only 0.206 Å(max. RMSD = 0.542 Å; min. RMSD = 0.073 Å). In general, in most complexes the host molecule shows an elliptical deformation compared to the free state without a guest molecule. The average RMSD between the geometries of the free host and the host in the complex is 0.567 Å (max. RMSD = 1.917 Å; min. RMSD = 0.117 Å).

We then included chloride counter ions by simply adding them to the gas phase structures above the third non-bonding hydrogen atoms of the ammonia groups. We re-optimized the now neutral complexes with TPSS-D3/def2-TZVP including the COSMO continuum solvation model ($\varepsilon = 78$ for water). The binding modes did not change and the overall geometries do not differ significantly compared to the charged complexes optimized in the gas phase. The average RMSD between the gas phase geometries and those in solution is 0.16 Å on the HF-3c and 0.19 Å on the TPSS-D3 level of theory.

The contributions to ΔG_a, namely the gas phase electronic energy ΔE_{el}, the two-body $\Delta E^{(2)}_{disp}$ and three-body $\Delta E^{(3)}_{disp}$ dispersion energy, the thermal correction from energy to free energy ΔG^{T}_{RRHO} and the solvation free energy $\Delta\delta G^{T}_{solv}$ are show in Figure 6.5. The electronic association energy varies in a large range from -7.5 to -37.0 kcal mol^{-1}.

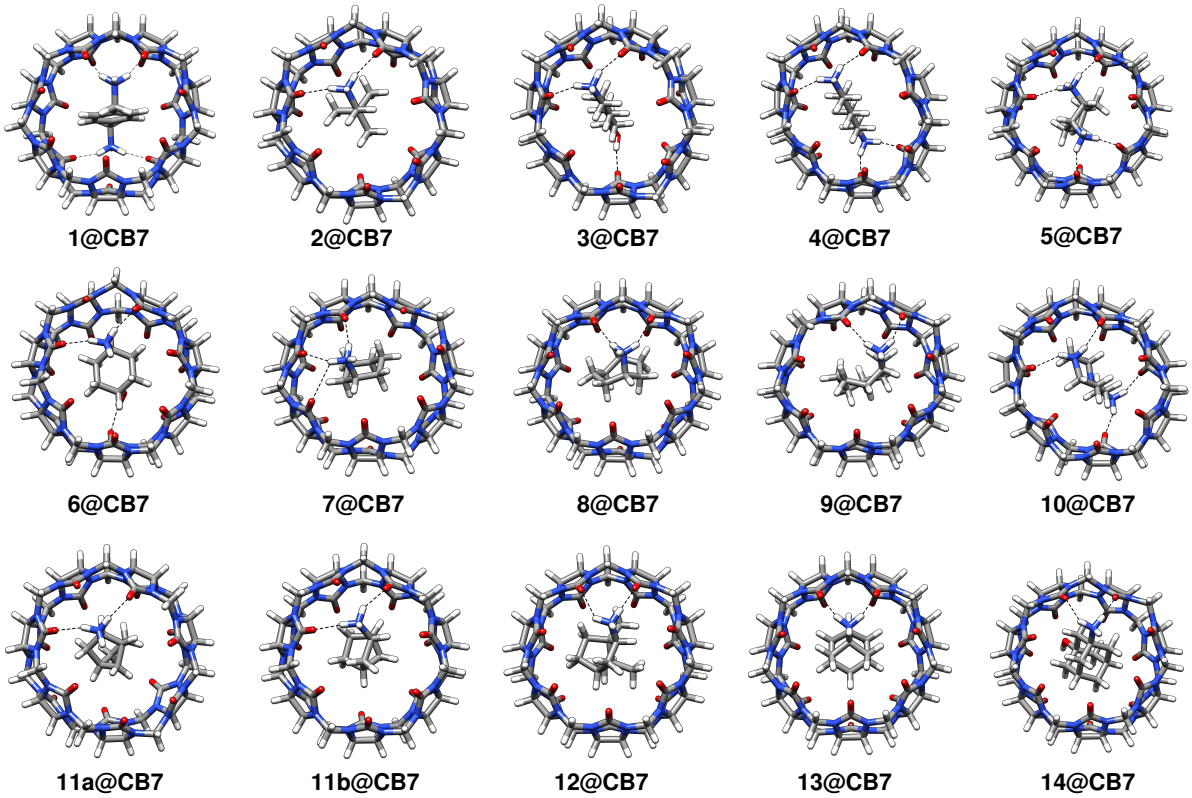


Figure 6.4.: Geometries of all 15 **CB7** complexes on the TPSS-D3/def2-TZVP level in the gas phase optimized without chloride counter ions.

The guests with two ammonia groups (**1**, **4**, **5** and **10**) exhibit a higher ΔE_{el} compared to those with just one ammonia group, mainly due to twice the number of hydrogen bonds. The higher ΔE_{el} of those complexes is compensated by a larger solvation free energy with opposite sign. For **3@CB7**, **6@CB7** and **14@CB7** with an additional hydroxy group the absolute value of $\Delta\delta G^{T}_{solv}$ is much smaller and for the guests with only one ammonia group it is almost zero. Complexes **7@CB7**, **8@CB7**, **9@CB7**, **12@CB7** and **13@CB7** show a negative $\Delta\delta G^{T}_{solv}$, whereas it is positive for all others. As expected, the two-body dispersion energy is large and for most complexes even higher than the pure DFT contribution. $\Delta E^{(2)}_{disp}$ varies between -14.0 and -24.0 kcal mol^{-1}, depending on the size and nature of the guest molecules. The three-body dispersion energy is similar for all complexes (2-3.5 kcal mol^{-1}) and therefore constantly shifts all ΔG_a values. The thermal correction ΔG^{T}_{RRHO} on the HF-3c level lies between 15.6 and 19.2 kcal mol^{-1}. Note that this term also contains changes of the zero-point vibrational energy and other thermal contributions (H(0 K) to H(298 K)) and not only (but mostly) accounts for entropic changes. Previously, we showed that HF-3c is sufficiently accurate to calculate the harmonic frequencies of supramolecular complexes of similar size.[205] The final absolute ΔG_a values span a large range from -3.4 kcal mol^{-1} for complexes **6@CB7** to -17.8 kcal mol^{-1} for **12@CB7**. ΔG_a of the racemic mixture of **11@CB7**

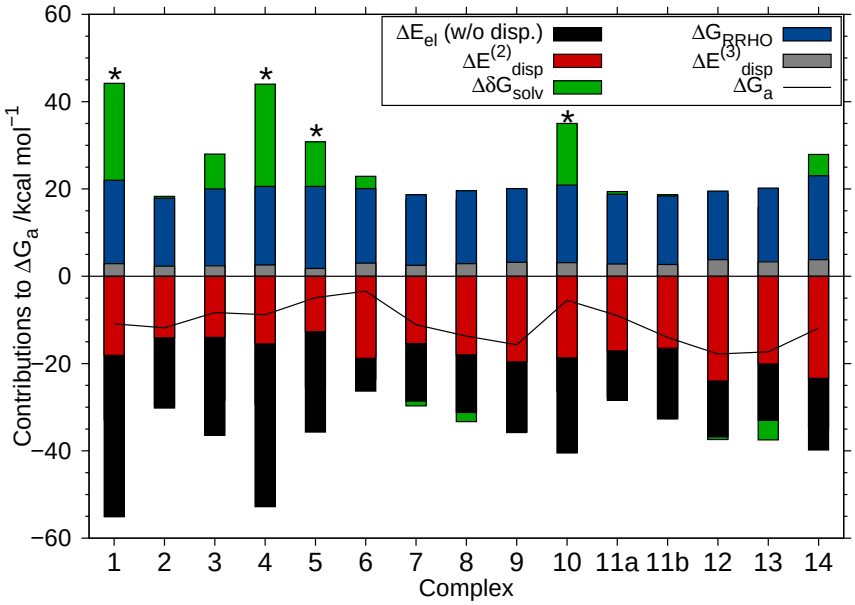


Figure 6.5.: Contributions to ΔG_a and final ΔG_a values on the PW6B95-D3/def2-QZVP(-g,-f)/COSMO-RS//TPSS-D3-cosmo/def2-TZVP/HF-3c(freq.) level for all 14 **CB7** complexes including Cl^- counter ions. Complexes marked by $\star$ have guests with two ammonia groups.

was calculated as a Boltzmann average (Equation 6.2) of the values of **11a@CB7** and **11b@CB7**, resulting in -14.0 kcal mol^{-1}. The reference compound **1@CB7** has a ΔG_a of -10.9 kcal mol^{-1}. In order to get the relative $\Delta G_{a,rel}$, which had to be submitted to the SAMPL4 committee according to the guidelines, this value was subtracted from all the others (see Table S2 in the supporting information). The absolute experimental ΔG_a of **1@CB7** (-9.9 kcal mol^{-1}) was then added by the committee to both the relative calculated and the relative measured $\Delta G_{a,rel}$ of the other complexes in order to ensure a fair evaluation of the results. Compared to the values obtained this way, all our calculated absolute ΔG_a values are shifted by the error in ΔG_a(**1@CB7**) of 1.0 kcal mol^{-1}.

The comparison of our calculated and the experimental results is shown in Figure 6.6. We included estimated theoretical error bars, which have been obtained by taking half of the absolute difference in interaction energy on the TPSS-D3/def2-TZVP and PW6B95-D3/def2-QZVP' levels of theory and adding 3% of the computed absolute values of the ΔG^T_{RRHO} and $\Delta\delta G^T_{solv}$ terms.

The overall agreement between theory and experiment is very good. The best results are obtained for complexes **3@CB7**, **4@CB7**, **7@CB7** and **8@CB7** for which the error is less than 1 kcal mol^{-1}. As it can be seen from Figure 6.6, the calculations yield overbinding for strongly bound complexes and underbinding for weakly bound complexes, respectively. The main outlier of the overbound complexes with a deviation from the experimental value of -3.5 kcal mol^{-1} is **12@CB7**, which has the most bulky guest. The

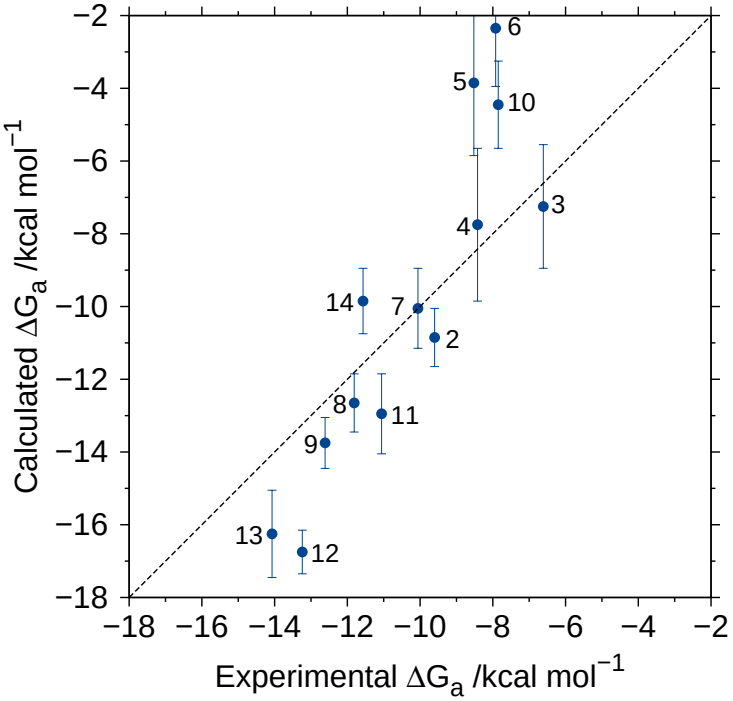


Figure 6.6.: Comparison of calculated and experimentally obtained association free energies ΔG_a for all **CB7** complexes with respect to the reference compound **1@CB7**. The absolute experimental ΔG_a of **1@CB7** was added to both the relative calculated and the relative measured $\Delta G_{a,rel}$ to get absolute values ΔG_a.

main outliers of the underbound complexes with an error larger than +3 kcal mol^{-1} are the complexes **5@CB7** and **6@CB7** with two functional groups and complex **10@CB7** with the not clearly specified charge state and at least two ammonia groups. Presumably, we were not able to find the conformer with the lowest ΔG_a and therefore, the single-conformer approach may be not best suited for these systems.

In comparison with all other submitted calculated relative $\Delta G_{a,rel}$ values for this part of the SAMPL4 blind test, our results achieved the best Pearson correlation coefficient (R= 0.90 ± 0.05), the second best value for the mean deviation (MD= 0.27 ± 0.72 kcal mol^{-1}), the second best value for the mean absolute deviation (MAD= 2.02 ± 0.46 kcal mol^{-1}), the second best value for the root mean square deviation (RMSD= 2.62 ± 0.51 kcal mol^{-1}) and the third best Kendall's τ measure ($\tau = 0.74 \pm 0.10$) with respect to the experimental values. However, for the SAMPL4 overview article the RMSD was calculated after subtracting the MD. The committee justified, they did not want to put participants at a disadvantage when their result for the reference compound **1@CB7** had a large error.[49] With respect to this measure our results only ranked eighth.

To conclude this chapter on the **CB7** systems, we want to mention the importance of counter ions. As described before, ΔG_a values were calculated for all complexes with a charge and with chloride counter ions and the changes in the corresponding geometries

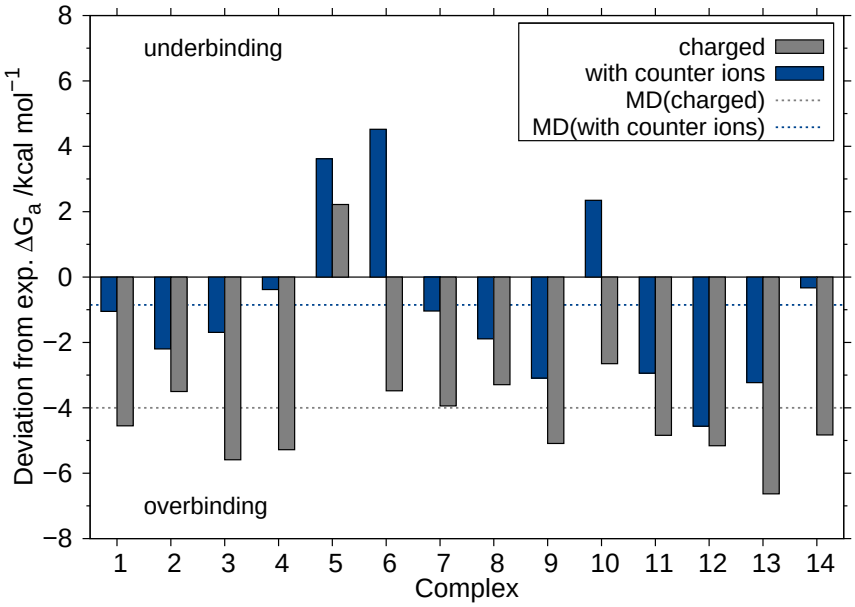


Figure 6.7.: Deviations from experimental $\Delta G_a, rel$ values for all complexes with and without chloride counter ions.

were small. We decided to submit the $\Delta G_{a,rel}$ values for the complexes including chloride ions, because the inclusion of counter ions seems physically more reasonable to us. As it can be seen from Figure 6.7, with only two exceptions the deviation from the experimental association free energy is always smaller for the calculations with counter ions. Hence, counter ions systematically improve the agreement with experiment. Tentatively, this can be attributed to a breakdown of the COSMO-RS solvation model for large surface charges which appear for bare ions with localized electronic structure as considered here.

6.4.2. The octa-acid host complexes

After showing the success of the combined DFT-D3/HF-3c/COSMO-RS approach in predicting the free energy of binding for a rigid, neutral host we now turn to the very challenging situation of the highly charged and considerably flexible octa-acid host. Two types of binding modes of the anionic guests inside the neutral octa-acid host were obtained by the HF-3c based conformational search, in both of them the guest is roughly aligned to the axis of the host. In one type, the carboxylate group points out of, in the other into the binding pocket. Structures with an orientation of the guest perpendicular to the octa-acid axis are possible only at the cost of a strong deformation of the host preventing a rotation of the guest inside the pocket between up and down. On the other hand for each of the two bound configurations various orientations of the guest inside the host with similar energies are obtained indicating considerable intrinsic flexibility of the host-guest complex.

In terms of the gas phase association energy ΔE, orientations with the carboxylate pointing into the pocket are favored over those with the carboxylate oriented towards the outside by several kcal mol^{-1} mainly due to weak hydrogen bonds with the CH groups inside the octa-acid. However, this energy difference is overcompensated by the $\Delta\delta G^{T}_{solv}$ term being over 10 kcal mol^{-1} more positive when the carboxylate is inside the pocket, thereby resulting in positive (i.e. not binding) ΔG_a values. All in all, in the most favorable complexes in terms of ΔG_a, the polar head group of the guest is located at the portal of the hydrophobic pocket, and the host is minimally deformed with the only exception of guest **22@OA** which has a positive computed ΔG_a (see below). In order to reduce artifacts also in structures with the proper orientation of the guest inside the host by overestimating interactions between the carboxylate and the host, geometries were derived by using the COSMO continuum solvation model instead of optimizations in vacuum. Figure 6.8 shows two representative structures with minimal free energy of binding ΔG_aon the PW6B95-D3/def2-QZVP(-g,-f)/COSMO-RS//TPSS-D3(BJ)/def2-TZVP+COSMO(ϵ=78)/def2-TZVP/HF-3c(freq.) level of theory.

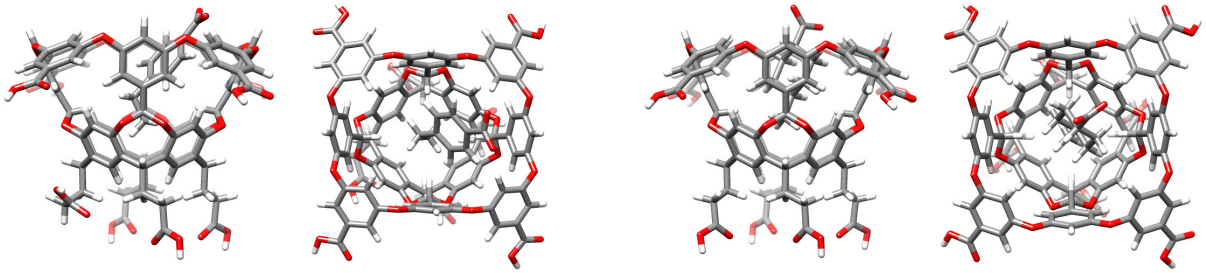


Figure 6.8.: Geometries of two **OA** complexes on the TPSS-D3/def2-TZVP level optimized with the COSMO continuum solvation model (ϵ=78): side-view and plan-view of guests **17** (left) and **21** (right).

The neutral model of the host binds the anionic guest according to the ΔG_a values at the TPSS-D3(BJ)/def2-TZVP level with one exception: no negative ΔG_a was obtained for guest **22**. After adding single-point PW6B95-D3(BJ)/def2-QZVP' association energy corrections, the three-body dispersion contribution, and the neutral to ionic host correction, the final ΔG_a values are also positive for complexes **15@OA** and **20@OA**. Note that ΔG_a is a sum of the intramolecular, noncovalent association energy ΔE with competing, usually positive solvent and entropic corrections (Equation 6.1). Since these compensation terms are generally much larger than the resulting free energy of binding, we already consider it as a success to be in the right order of magnitude[46] for such a difficult system. Comparison of the final ΔG_a values for octa-acid to the experimental affinities reveals that the calculated result is not completely wrong, but underestimates the experiment due to an inappropriate modelling of the charge state (Figure C.2, top).

Consequently the results rank between seventh and eleventh among the twelve submissions to SAMPL4 in eight different statistical measures (note that the results presented here are not identical to those that were available at the SAMPL4 deadline, see Table S7 of the supporting information for a description of the difference between the submitted data and those shown in Figure C.2, top). In particular the results performed worse than two null models, like all other submissions based on quantum mechanical energy models[49] indicating a common shortcoming in the case of octa-acid. More importantly our submitted results are in general slightly worse than a submission by the group of Ryde also based on DFT-D3 calculations on one single configuration per host-guest complex[365]. Among others the DFT-D3 based free energies of binding reported in ref.[365] are derived from TPSS-D3/def2-QZVP intermolecular interaction energies for a fully charged host. Without a final judgement with respect to the most important source of the discrepancy between between the performances of the two sets of similarly derived results we conclude that the full potential of the approach is not yet exhausted with our contribution to the SAMPL4 octa-acid challenge.

The importance of a proper modeling of the charge state is demonstrated by single-point energy calculations after adding a Na^+ counterion and two explicit water molecules to the guest in a relative orientation obtained by TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) geometry optimization of guest **15** + Na^+ + $2H_2O$ (Figure C.2, middle). While ΔE of guest **15** is smaller by about 11 kcal mol^{-1}, $\Delta\delta G^T_{solv}$ is reduced as well by more than 18 kcal mol^{-1}, resulting in an overbinding ΔG_a of -7.6 kcal mol^{-1}. Thus, adding a counterion corrects the systematic deviation of values except for the complex of octa-acid with guest **22**, but the scattering of the data is increased. This is reflected by the smaller correlation coefficient of 0.18 with (Figure C.2, middle) compared to 0.66 without (Figure C.2, top) Na^+ counterion (both after removing the outlier complex **22@OA**). That the stabilizing effect of adding a Na^+ counter ion and two explicit water molecules on ΔG_a is not also visible for the **22@OA** complex is due to the repulsion between the additional water molecules and the host overriding the former. Even for a structure of **22@OA** in which the carboxylate group of cyclopentanecarboxylic acid roughly points out of the host, the Na^+ counter ion and two explicit water molecules can not be added in the plane of the guests' carboxylate group without a clash with the host as for the other eight cases but have to be tilted by 60 degrees out of this plane. The extra energy is removed by the restricted geometry optimization described below in the next paragraph which results in a binding ΔG_a also for guest **22** which was not obtained without counter ion due to artificial contacts between the carboxylate group of cyclopentanecarboxylic acid and the host.

In order to improve the computational model of the octa-acid host accordingly, an anionic C_4 symmetrical structure with eight Na^+ counterions plus three additional explicit

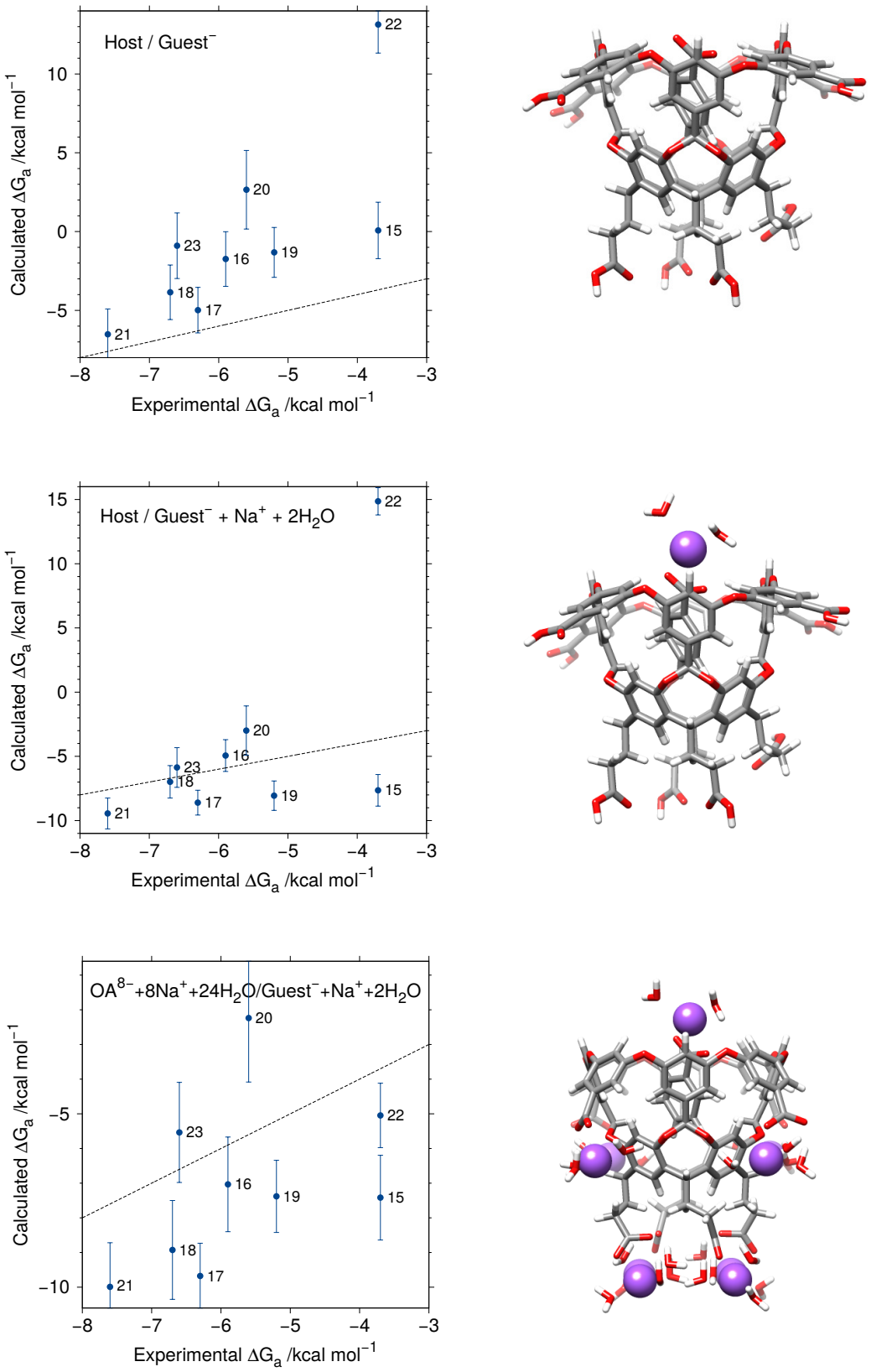


Figure 6.9.: Computed vs. experimental binding free energy (kcal mol^{-1}) for the anionic guests in the neutral octa-acid host without (top) and with (middle) one Na^+ counterion and in the anionic octa-acid host with totally nine Na^+ counterions (bottom). Geometries of the **15@OA** complex are shown on the right.

water molecules (totally 24 extra water molecules) was obtained by HF-3c optimization in vacuum. The ionic model replaces the neutral host in the best TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) structure (i.e. the one with most negative ΔG_a except for complex **22@OA** where instead of the structure with lowest positive ΔG_a a less favorable structure is taken in which the carboxylate group of cyclopentanecarboxylic acid points out of instead of into the binding pocket) by matching the positions of all carbon atoms (minimal RMSD). Finally, the guest together with its Na^+ counterion and two explicit water molecules is optimized at the TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) level in a fixed host, i.e. by setting the forces on the host atoms, eight Na^+ counterions, and 24 water molecules zero. The resulting ΔG_a values are composed of the TPSS-D3(BJ)/def2-TZVP intermolecular interaction energy ΔE between the guest with its Na^+ counterion plus two explicit water molecules and the C_4 symmetric host including the three-body dispersion contribution, the $\Delta\delta G^T_{solv}(X)$ term for the same structure, a TPSS-D3(BJ)/def2-TZVP to PW6B95-D3(BJ)/def2-QZVP' correction for the fully TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) optimized structure of the anionic guest with the neutral host, and the ΔG^T_{RRHO} term derived from the harmonic frequencies of the HF-3c optimized the neutral host/anionic guest structure (Figure C.2, bottom).

Comparing the results for ΔG_a in the anionic (Figure C.2, bottom) to those in the neutral (Figure C.2, middle and top) host, a considerable improvement is obtained. Instead of underbinding all guests are overbound now except for complexes **20@OA** and **23@OA**. In particular complex **22@OA** is bound, too, as interaction between the carboxylate group of the smallest and thereby most mobile guest in the series and the host is prohibited by the presence of the Na^+ counterion. However, the deviation of the calculated ΔG_a from the reference values still ranges from -3.7 to 3.4 kcal mol^{-1} with a mean absolute deviation (MAD) of 2.3 kcal mol^{-1}. Also the value of the correlation coefficient is only slightly increased to 0.43, but now also including the previous outlier, guest **22**. A similar picture results for an anionic C_2 symmetrical host structure with six Na^+ counterions plus three additional explicit water molecules (totally 18 extra water molecules) and an anionic C_4 symmetrical host structure with four Na^+ counterions plus three additional explicit water molecules (totally 12 extra water molecules). The error range and MAD are -4.5 to 2.6 (-3.5 to 2.5) kcal mol^{-1} and 2.0 (1.6) kcal mol^{-1} for seven (five) Na^+ counterions and the correlation coefficient is 0.45 (0.47). The corresponding graphs are provided as supporting information where also the raw data and their components are tabulated. Thus, the assumed charge state of the host appears to be not as influential for the resulting ΔG_a (Figure C.2, bottom) as is that of the guest whose modification results in qualitatively correct (i.e. binding) estimates (Figure C.2, top and middle). Unfortunately neither of the three anionic host structures gives a proper ranking.

It is very interesting to know whether the improved results with counter ions are caused

by the counter ions and not because the host was kept rigid with the same geometry for all complexes, avoiding some of the conformational problems. This is checked by calculating single-point energies also without the counter ions, i.e. re-replacing eight Na^+ by H and deleting the 24 coordinating water molecules. The resulting ΔG_a values differ from those in the fully deprotonated host from 0.3 to -0.8 kcal mol^{-1} (see Table S3 of the supporting information). The differences in the individual components (ΔE and $\Delta\delta G^T_{solv}$ are slightly larger but opposite in sign and therefore mutually compensating. Consequently the error range is slightly shifted (-4.2 to 2.8 kcal mol^{-1}), while MAD (2.3 kcal mol^{-1}) and correlation coefficient (0.49) are hardly affected compared to the fully protonated host supporting the statement at the end of the preceding paragraph.

The overbinding of an anionic guests by an anionic model of the host is surprising at first sight. Most likely this is related to the extension of the guest by the Na^+ counterion plus its two coordinating water molecules leading to the inclusion of additional interaction terms in ΔE. Therefore the two water molecules were omitted from interaction energy and free energy of desolvation calculation.[365] The resulting ΔG_a values are consistently reduced by 1.8 to 4.5 kcal mol^{-1}, so that overbinding is essentially removed for guests **15**, **17**, **19**, and **21**. Since this is true also for the initially underbinding guests **20** and **23**, all remaining complexes (**16@OA**, **18@OA**, **20@OA**, **22@OA**, and **23@OA**) end up underbinding without consideration of the two explicit water molcules. The error ranges from -0.9 to 5.8 with an MAD of 2.1 kcal mol^{-1} and correlation coefficient of 0.42. That is the results are more or less shifted from over- to underbinding. The same holds for the other two incompletely deprotonated hosts: MAD 2.0 kcal mol^{-1} (2.0 kcal mol^{-1}) and correlation coefficient 0.59 (0.54) without the two explicit water molecules for seven (five) Na^+ counter ions are similar to the case with the two explicit water molecules, while the error range -1.2 to 5.1 kcal mol^{-1} (-1.2 to 4.9 kcal mol^{-1}) is shifted upwards. However the underbinding of the remaining five complexes can be easily attributed to the underlying potentially not optimal binding configuration. All in all modeling the binding free energy in the case of the octa-acid host mainly suffers from the problem that the binding situation is insufficiently defined with respect to the charge state and due to its intrinsic flexibility the host-guest complex is probably not represented well by a single configuration in our approach.

6.5. Conclusion

The presented DFT-D3 based blind prediction of binding affinities for charged supramolecular host-guest systems within the SAMPL4 framework shows that the success of the approach strongly depends on the validity of one (or few) minimal energy structures to represent the complex properly. Extensive sampling of several hundred thousand or even

million of configurations is not yet compatible with the high level of electronic structure theory applied here. Furthermore, neutral host-guest complexes are treated more accurately than charged ones and cationic ones are easier to handle than anionic systems in our DFT treatment.

Hence, in case of the rigid and neutral **CB7** host the single conformer approach is successful to calculate the binding free energies of cationic guest molecules. The inclusion of chloride counter ions to neutralize the ammonia groups improves the results even further. Compared to all other submissions to the SAMPL4 blind test our calculated ΔG_a ranked top and our result is one out of the best three for all statistical analyses applied. With an MAD of only 2 kcal mol^{-1} w.r.t. experiment we were able to predict the binding affinity with a high accuracy.

In case of the **OA** host the addition of counter ions and explicit water molecules, while neutralizing the structures, increases the flexibility and furthermore introduces an ambiguity into the dissection between host and guest which is necessary for computing ΔG_a. It is reassuring that the qualitative features of the binding process are represented in accord with the experimental expectations also in the case when the quantitative agreement of the computed free energy of binding with the measured one is not fully satisfactory. It is interesting to note that whereas the present method performs worse on **OA** than on **CB7**, the opposite is true fore nearly all the other submissions to SAMPL4.[49] So having tested the limits of our approach to ΔG_a determination we feel even more confidence than before in the predictive power of dispersion corrected density functional theory in determining binding affinities for a wide range of supramolecular systems.

Acknowledgement

The authors thank the members of the SAMPL4 committee David Mobley, John Chodera, Tom Peat, Terry Stouch, Vijay Pande and Mike Gilson for organizing this blind test.
Financial support by the German research foundation by grant AN 793/1-2 is gratefully acknowledged.

7. Comprehensive benchmark of association (free) energies of realistic host–guest complexes

Rebecca Sure* and Stefan Grimme*

Received 30th of March 2015, Published online 19th of June 2015

Reprinted (adapted) with permission from
R. Sure and S. Grimme, *J. Chem. Theory Comput.* **2015**, *11*, 3785–3801.
 DOI: 10.1021/acs.jctc.5b00296

Own manuscript contribution:

- Performance of the calculations
- Interpretation of data
- Writing the manuscript

*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany

Abstract

The S12L test set for supramolecular Gibbs free energies of association ΔG_a (Chem. Eur. J. **2012**,*18*, 9955–9964) is extended to 30 complexes (S30L), featuring more diverse interaction motifs, anions and higher charges (-1 up to +4) as well as larger systems with up to 200 atoms. Various typical noncovalent interactions like hydrogen and halogen bonding, $\pi - \pi$ stacking, non-polar dispersion, CH$-\pi$, and cation-dipolar interactions are represented by "real" complexes. The experimental Gibbs free energies of association (ΔG_a^{exp}) cover a wide range from -0.7 to -24.7 kcal mol^{-1}. In order to obtain a theoretical best estimate for ΔG_a, we test various dispersion corrected density functionals in combination with quadruple-ζ basis sets for calculating the association energies in the gas phase. Further, modern semiempirical methods are employed to obtain the thermostatistical corrections from energy to Gibbs free energy, and the COSMO-RS model with several parametrizations as well as the SMD model are used to include solvation contributions. We investigate the effect of including counterions for the charged systems (S30L-CI), which is found to overall improve the results. Our best method combination consists of PW6B95-D3 (for neutral and charged systems) or ωB97X-D3 (for systems with counterions) energies, HF-3c thermostatistical corrections, and Gibbs free energies of solvation obtained with the COSMO-RS 2012 parameters for non-polar solvents and 2013-fine for water. This combination gives a mean absolute deviation for ΔG_a of only 2.4 kcal mol^{-1} (S30L) and 2.1 kcal mol^{-1} (S30L-CI), respectively, with a mean deviation of almost zero compared to experiment. Regarding the relative Gibbs free energies of association for the 13 pairs of complexes which share the same host, the correct trend in binding affinities could be reproduced except for two cases. The MAD compared to experiment amounts to 1.2 kcal mol^{-1} and the MD is almost zero.

The best-estimate theoretical corrections are used to back-correct the experimental ΔG_a values in order to get an empirical estimate for the "experimental", zero-point vibrational energy exclusive, gas phase binding energies. These are then utilized to benchmark the performance of various "low-cost" quantum chemical methods for noncovalent interactions in large systems. The performance of other common DFT methods as well as the use of semiempirical methods for structure optimizations is discussed.

7.1. Introduction

Noncovalent interactions between atoms and molecules, such as dispersion interactions, $\pi - \pi$ stacking, or hydrogen and halogen bonding play an important role in structural biology and supramolecular chemistry[25–27]. They control the structures of DNA and proteins, antigen-antibody recognition, host-guest and enzyme-substrate binding, or the orientation of molecules on surfaces or in molecular crystals.[16,17] Due to their omnipresence in diverse fields of science, the investigation and understanding of these noncovalent interactions (NCIs) has advanced to a major topic in modern chemistry. Although they are frequently termed "weak" interactions, especially the London dispersion part can account for a large percentage of the total interaction energy and often outranks electrostatic or hydrogen bonding contributions[3].

Noncovalently bound host-guest complexes are of particular importance in supramolecular chemistry. They are utilized in the fields of molecular recognition, template-direct synthesis, biomimetics, self-assembly, and even as reaction containers.[1,2,16,17,22] Therefore, the characterization and subsequent tuning of the different stabilizing interactions are of particular interest. Quantitative descriptions and predictions of the binding thermodynamics of these supramolecular complexes by means of computational electronic structure methods are still a challenge. This can be attributed to the facts that even the smallest experimentally synthesized systems are built up from a hundred atoms or more, and solvation and entropic effects play an important role in the binding process. Hence, the computational costs in a theoretical, preferably highly accurate quantum chemical treatment, are large and often unaffordable.

Density functional theory (DFT) has been proven to provide a good ratio in terms of cost and accuracy for many applications in modern quantum chemistry. However, standard exchange-correlation functional approximations inherently lack the correct description of London (long-range) dispersion interactions.[287,288,366] Several different ways of treating dispersion forces within the Kohn-Sham DFT framework have been presented in the past years and have emerged as a standard in the field, for recent reviews see Refs. [32,33], and for comparison of their performance see e.g. Refs. [34,35]. As demonstrated recently, Gibbs free energies of association ΔG_a (in the following referred to as free energies) for typical medium-sized supramolecular host-guest systems can be computed with good accuracy by dispersion corrected density functional theory (e.g. DFT-D3[36]) together with a relatively large basis set in a non-dynamic single-structure approach without any system-specific adjustments.[46,192,296,367]. For the so called S12L set of complexes,[46] which was the first benchmark set for NCI in large systems, the DFT-D3 gas phase interaction energies were confirmed by independent DFT-SAPT calculations[368] and high level electronic Quantum Monte-Carlo simulations.[299] For the related L7 benchmark set for NCI in large model

complexes see Ref. [144].

Our procedure to obtain the target ΔG_a value involves three steps: First, the electronic interaction energy in the gas phase ΔE is computed for equilibrium structures. The molecules of interest (host, guest and complex) are optimized in the gas phase on an affordable level including a dispersion correction, e.g. the TPSS meta-GGA density functional[93] together with the triple-ζ basis set def2-TZVP[237] and the D3 dispersion correction with Becke-Johnson damping[133,134] (D3(BJ)).[37] Single-point energies are obtained by employing a hybrid-functional like PW6B95[101]-D3 together with a large quadruple-ζ basis set, e.g. def2-QZVP[237]. The gas phase association energy ΔE is calculated in the supermolecular approach

$$\Delta E = E(\text{complex}) - E(\text{host}) - E(\text{guest}), \tag{7.1}$$

where E is the total electronic energy of the species involved. At a QZ basis set level, basis set superposition errors (BSSE) diminish to typically less than 2% of ΔE and can be ignored[192] for (hybrid)GGA functionals. In addition to the two-body dispersion interaction $\Delta E^{(2)}_{disp}$, the Axilrod-Teller-Muto-type three-body dispersion (ATM) energy $\Delta E^{(3)}_{disp}$ is also included. The three-body dispersion contribution to ΔE was found to be always positive and to contribute significantly with 2-3 kcal mol^{-1} for typical supramolecular systems[46]. We disregard many-body dispersion effects beyond the three-body term, see Refs. [135,137,299,332,369] for examples and further discussion. Note that $\Delta E^{(3)}_{disp}$ varies in methodologically different approaches,[136,299] and we decided to use the efficient D3 method for its computation without any empirical adjustments. However, missing higher-order many-body dispersion effects are maybe on the order of 1-2 kcal mol^{-1} which can be estimated from a comparison of D3+ATM data and the MBD approach[297] for molecular crystals.[207,370] This is relevant when aiming at absolute ΔG_a values of about 5-10 kcal mol^{-1} and further investigation of their accurate computation seems mandatory. The total electronic interaction energy ΔE thus comprises the pure electronic DFT energy ΔE^{DFT}_{el}, two-body $\Delta E^{(2)}_{disp}$, and three-body dispersion energy $\Delta E^{(3)}_{disp}$:

$$\Delta E = \Delta E^{DFT}_{el} + \Delta E^{(2)}_{disp} + \Delta E^{(3)}_{disp} \tag{7.2}$$

Second, the sum of thermostatistical corrections from energy to free energy G^T_{RRHO} are obtained for each molecule in the gas phase at a given temperature T and normal pressure of 1 atm, including the zero-point vibrational energy. Since low-lying vibrational frequencies are inaccurate in the harmonic approximation, a modified rigid-rotor-harmonic-oscillator scheme is used[46]. In this approach vibrational modes below 100 cm^{-1} are treated within a rigid-rotor model with smooth interpolation to the standard harmonic

regime, for additional details, see Ref. [46] Its good performance has been verified repeatedly also for computing reaction thermochemistry, see e.g. Refs[371–373]. The computation of the harmonic frequencies is carried out with semiempirical methods like PM6[124]-D3, DFTB[113]-D3 or our recently developed minimal basis set Hartree–Fock (HF-3c) method[205]. HF-3c is a fast quantum chemical procedure with three atom-pairwise correction terms (D3(BJ), geometrical counter-poise correction (gCP)[47], and an additional short-range term to correct for basis set deficiencies). It yields very reasonable structures and interaction energies for typical supramolecular systems and can therefore be used for pre-optimizations or screening applications.

Third, a continuum solvation model like COSMO-RS[41,42] or SMD[54] is used in a black-box manner to calculate the solvation free energy $\delta G^T_{solv}(X)$ of each gas phase species at the temperature T in solvent X. The resulting values implicitly contain the conversion to standard state conditions. An alternative approach to the solvation problem is for example the coupling of quantum mechanics to an integral equation theory of liquids, like 3D-RISM.[303,304] As the current implementation unfortunately does not support all needed solvents, we did not apply this method here. Another route is represented by classical molecular dynamics (MD) or hybrid quantum mechanics/molecular mechanics (QM/MM) MD simulations, which require a statistical sampling of the distributions of solvent molecules, for recent examples see Refs. [295,374]

The association free energy ΔG_a is the sum of the three contributions:

$$\Delta G_a = \Delta E + \Delta G^T_{RRHO} + \Delta \delta G^T_{solv}(X) \tag{7.3}$$

In the original publication,[46] this approach was applied to 12 rigid organic complexes (S12L) with up to 160 atoms and yielded accurate ΔG_a values with a mean absolute deviation (MAD) of only 2.1 kcal mol^{-1} with respect to experiment for PW6B95-D3/def2-QZVP' energies, DFTB-D3 vibrational frequencies and solvation effects treated with COSMO-RS (2012 parametrization). Moreover, it has recently been employed in the SAMPL4 blind test challenge. Binding affinities for 14 cucurbit[7]uril complexes with cationic guest molecules were to be predicted and our results ranked in the top three of all submissions in all statistical measures.[49,306,375]

The present work continues and extends the application of this approach to 30 complexes with up to 200 atoms and charges from -1 up to +4. In addition to 11 of the original complexes (excluding the ferrocene@CB7 complex because many semiempirical methods or force-fields cannot handle transition metals), 19 new chemically interesting complexes were chosen, which feature slightly less rigid hosts like crown ethers and cyclodextrins and host molecules with flexible alkyl side chains. The experimentally obtained ΔG_a values cover a wide range from -0.7 to -24.7 kcal mol^{-1} and the complexes cover the most im-

portant typical supramolecular interactions such as hydrogen and halogen bonding, $\pi-\pi$ stacking, non-polar dispersion, CH$-\pi$, and cation-dipolar interactions. We investigate the influence of counterions for charged complexes since we found in previous studies that the inclusion of chloride (model) ions significantly improved the results.[49,367] Also, counterions were found to be necessary in calculations of multiply charged species when using the COSMO-RS solvation model.[43]

Furthermore, we test several density functionals to compute the binding energies, different semiempirical methods to calculate vibrational frequencies, and various COSMO-RS parametrizations as well as SMD to describe solvation effects in order to find the combination yielding the most accurate ΔG_a values. No empirical adjustments to any of the applied methods were made. With the presumably best approaches for ΔG^T_{RRHO} and $\Delta\delta G^T_{solv}$, we subsequently back-correct the experimental binding free energies to obtain empirical ("experimental") gas phase binding energies ΔE^{emp}. These values are then used to benchmark the performance of low-cost simplified quantum chemical methods and are suggested together with the best DFT values as reference data in future methodological studies.

7.1.1. The test set complexes

In the following, we briefly describe the investigated complexes. Figure 7.1 shows the optimized equilibrium structures of the complexes contained in the S30L test set and Table 7.1 summarizes the experimental conditions as well as the net charges and the experimental ΔG^{exp}_a values. The complexes are sorted according to the most prominent type of interaction.

The first two complexes mainly feature non-polar dispersion interactions and were already part of the S12L set. **1** and **2** are tweezer complexes with tetracyanoquinodimethane (TCNQ) and para-dicyanobenzene (DCB) (measured in $CHCl_3$ at 298 K).[275]

The next group consists of ten complexes which mainly feature $\pi-\pi$ stacking interactions. **3** and **4** were taken from the S12L and are two pincer complexes with 2,4,7-trinitro-9-fluorenone (TNF) and 4-chloro-7-nitrobenzofurazan (NDB) as guests (in CH_2Cl_2 at 298 K).[276] Further, two tweezer complexes with TNF and TCNQ as guest molecules (**5** and **6** in $CHCl_3$ at 298 K)[376] were chosen. These two complexes are able to undergo shape switching and are therefore interesting as a nanomechanical devices.[16,385] Also, two ring-in-ring complexes were considered, namely [5]cycloparaphenyleneacetylene (5CPPA) in 8CPPA and 6CPPA in 9CPPA (**7** and **8** in $CHCl_3$ at 328 K).[377] These complexes show concave-convex $\pi-\pi$ interactions, which are important in the formation of e.g. so called bucky onions[386,387] and fullerene peapods.[388] **7** has a ΔG_a value comparable to C_{60}@6CPPA giving evidence that concave-convex $\pi-\pi$ interactions are not limited to

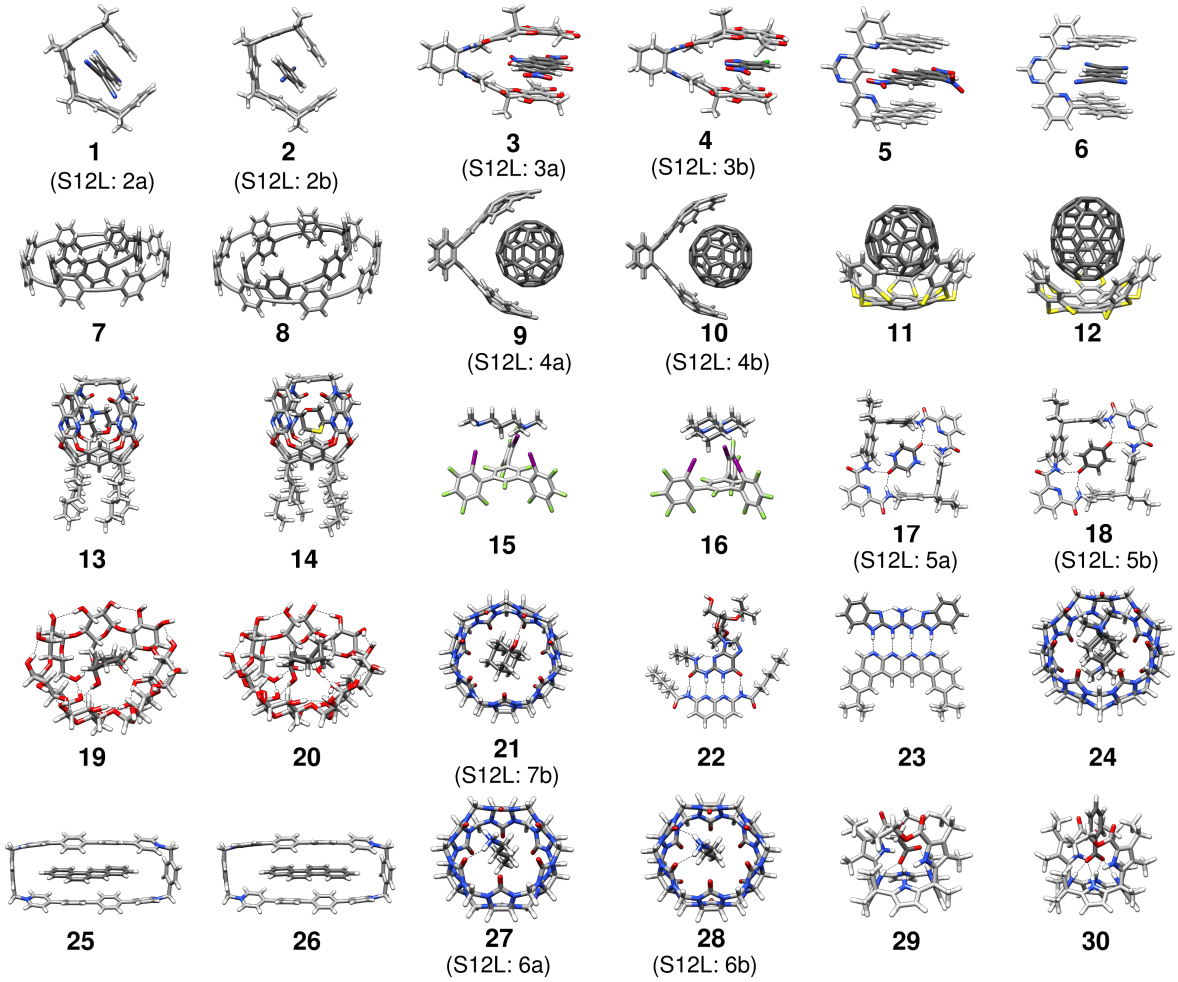


Figure 7.1.: Structures of the 30 supramolecular complexes contained in the S30L test set. C-atoms of the host molecules are shown in light grey, those of the guest molecules in dark grey, and H-bonding interactions are indicated by dotted lines. For convenience, the old S12L numbering is given if appropriate.

fullerenes only.[378,389] **9** and **10** were already part of the S12L and are two buckycatcher complexes that capture C_{60} and C_{70} (in toluene at 293 K).[281] Next, two other fullerene catcher complexes witch pentakis(1,4-benzodithiino)corannulene (CA10) as host and also C_{60} and C_{70} as guests (**11** and **12** in CS_2 at 298 K)[378] were chosen. In addition to the already mentioned concave-convex $\pi - \pi$ interactions the CA10 host contains sulfur atoms, which provide somewhat enhanced dispersion interactions compared to second-row atoms.

The third group consists of two complexes which mainly show C-H$\cdots\pi$ interactions. Two resorcin[4]arene-based container (RA4) complexes with morpholine and tioxane as guests (**13** and **14** in mesitylene at 303 K) were selected.[379] These complexes feature non-polar dispersion in addition to C-H$\cdots\pi$ interactions, and due to the latter, heterocyclic guests show a much larger binding affinity than cyclic hydrocarbons, which were also

Table 7.1.: Overview of the investigated complexes, their charges (in case of S30L), the experimental measurement conditions (solvent and temperature T in K) and the experimental association free energies ΔG_a^{exp} given in kcal mol^{-1}.

	complex	charge	solvent	T	ΔG_a^{exp}
1	TCNQ@tweezer[275]	0	$CHCl_3$	298	-4.2
2	DCB@tweezer[275]	0	$CHCl_3$	298	-1.4
3	TCNB@pincer[276]	0	CH_2Cl_2	298	-1.5
4	NBD@pincer[276]	0	CH_2Cl_2	298	-1.8
5	TNF@tweezer2[376]	0	$CHCl_3$	298	-5.2
6	TCNQ@tweezer2[376]	0	$CHCl_3$	298	-4.6
7	5CPPA@8CPPA[377]	0	$CHCl_3$	328	-5.5
8	6CPPA@9CPPA[377]	0	$CHCl_3$	328	-2.2
9	C_{60}@catcher[281]	0	toluene	293	-5.3
10	C_{70}@catcher[281]	0	toluene	293	-5.1
11	C_{60}@CA10[378]	0	CS_2	298	-4.4
12	C_{70}@CA10[378]	0	CS_2	298	-4.2
13	morpholine@RA4[379]	0	mesitylene	303	-10.0
14	tioxane@RA4[379]	0	mesitylene	303	-9.0
15	TMPDA@XB-donor[380]	0	cyclohexane	298	-0.7
15	HHTAP@XB-donor[380]	0	cyclohexane	298	-5.1
17	BQ@mcycle[278]	0	$CHCl_3$	298	-8.3
18	GLH@mcycle[278]	0	$CHCl_3$	298	-3.3
19	C_5H_9OH@β-CD[381]	0	H_2O	298	-3.0
20	$C_8H_{15}OH$@β-CD[381]	0	H_2O	298	-4.9
21	AdOH@CB7[280]	0	H_2O	298	-14.1
22	DAAD@ADDA[15]	0	$CHCl_3$	298	-11.7
23	AAAA@DDDD$^+$[358]	+1	CH_2Cl_2	298	-17.3
24	$Ad_2(NMe_3)_2$@CB7[382]	+2	H_2O	298	-24.7
25	tetraphene@Ex^{2}Box[383]	+4	CH_3CN	298	-4.4
26	chrysene@Ex^{2}Box[383]	+4	CH_3CN	298	-5.3
27	$BuNH_4^+$@CB6[279]	+1	formic acid:H_2O 1:1	298	-6.9
28	$PrNH_4^+$@CB6[279]	+1	formic acid:H_2O 1:1	298	-5.7
29	acetate@CP4[384]	-1	CH_3CN	298	-8.2
30	benzoate@CP4[384]	-1	CH_3CN	298	-7.7

investigated experimentally. The RA4 container has four flexible hexyl side chains that were fully included in the calculations. Thus, **13** and **14** are the largest complexes with about 200 atoms.

The forth group consists of two complexes which exhibit halogen bonding (**15** and **16** in cyclohexane at 298 K).[380] The host molecule (XB-donor) features three polyfluoroiodoarenes that are orientated perpendicular to the central benzene core and thus form a tridentate halogen bond donor motif. It binds various diamines and triamines such as the chosen guests N,N,N',N'-tetramethylpropane-1,3-diamine (TMPDA) and hexahydro-

1H,4H,7H-3a,6a,9a-triazaphenalene (HHTAP). Halogen bonds are more directional than hydrogen bonds and thus, describing these multi-point interactions is much more challenging.[390–392]

The fifth group of eight complexes feature hydrogen bonds as the major type of interaction. The first two complexes are two amine macroycyle (mcycle) complexes with benzoquinone (BQ) and glycine anhydride (GLH) from the S12L (**17** and **18** in $CHCl_3$ at 298 K).[278] Note, that the electronic structure of BQ is somewhat non-trivial in the sense that some quantum chemical codes produce a wrongly occupied orbital guess in the D_{2h} symmetry so that the SCF converges to an incorrect excited state. Next, two β-cyclodextrin (β-CD) complexes with cyclopentanol and cyclooctanol as guests (**19** and **20** in H_2O at 298 K) were investigated.[381] Cyclodextrins posses a hydrophobic central cavity and a hydrophilic outer surface and can therefore be used to improve the delivery for poorly soluble drugs.[393] Therefore, in addition to hydrogen bonds, non-polar dispersion is also important. **21** and **24** are two cucurbit[7]uril (CB7) complexes with 1-hydroxyadamantane (AdOH) (taken from S12L)[280] and a double positively charged diamantane diammonium ($Ad_2(NMe_3)_2$)[358] as guests (in H_2O at 298 K). **23** shows an especially high ΔG_a due to the perfect alignment of the guest along the axis of the host and hence, maximal dispersion interaction of the diamantane core with the hydrophobic region of the inner circumference of the CB7 and seven ion-dipole interactions of the NMe_3^+ groups with each carbonyl of the ureidyl portal. Over the past decade, the cucurbit[n]uril (CBn, $n = 5, 6, 7, 8$) family of molecular containers has advanced to a major tool for studying molecular recognition in water.[50–52] Guest molecules with a rigid hydrophobic core, such as ferrocene[53] or adamantane,[50,358] in combination with cationic ammonium groups, have been found to yield very large binding affinities. Further, two quadruple hydrogen bond arrays were chosen. The first one is an ADDA-DAAD type array where both host and guest are a double donor (D) and a double acceptor (A) (**22** in $CHCl_3$ at 298 K).[15] The second one is an AAAA-DDDD$^+$ complex, where the host molecule is a quadruple acceptor and the guest a positively charged quadruple donor (**23** in CH_2Cl_2 at 298 K).[382] This array exhibits exceptional stability for such a small system even in hydrogen bond disrupting solvents.

The last group contains eight charged complexes (**23** and **24** also belong to this group). We investigated two complexes of the recently synthesized Ex^2Box^{4+} macrocycle with tetraphene and chrysene (**25** and **26** in CH_3CN at 298 K).[383] This macrocycle, composed of two biphenyl-bridged bipyridinium units and thus a charge of +4, has the unusual capability of binding π-electron rich as well as π-electron-poor guests, either two small molecules at the same time or one large guest like the chosen tetraphene and chrysene. In this case, the main interaction occurs through π-π-stacking. **27** and **28** were already included in the S12L and are two CB6 complexes with butylammonium ($BuNH_3$) and

propylammonium ($PrNH_3$) guests (in a 1:1 mixture of formic acid and H_2O at 298 K).[279] The last complexes that were investigated are two calix[4]pyrrole (CP4) complexes that bind various mono anions. Our chosen guests are acetate and benzoate (**29** and **30** in CH_3CN at 298 K).[384] Anions are usually more challenging for DFT methods than cations due to the fact that the energy of highest molecular orbital is often calculated to be positive leading to artifical charge-transfer.

7.2. Computational details

The geometries of the S30L complexes **1**, **2**, **3**, **4**, **9**, **10**, **17**, **18**, **21**, **27** and **28** as well as those of their host and guest molecules were taken from Ref[46]. Geometry optimizations of all other complexes, hosts and guests (applying the appropriate charge of the molecules) were done on the same level, that is employing the density functional TPSS[93] together with the triple-ζ basis set def2-TZVP[237]. Hence, opposed to other large complex benchmarks like the L7 set[144],treatment of S30L requires the computation of the so called relaxation energy, i.e., the effect of full optimization of all fragments. In all DFT calculations except for M06-2X the D3 dispersion correction[36] with Becke-Johnson (BJ) damping[37,133,134] was used. In case of the S30L-CI set, chloride counterions were added to the cationic structures and sodium ions to the anionic structures of the charged complexes **23** to **30** at appropriate positions and re-optimized on the TPSS-D3/def2-TZVP level using the COSMO continuum solvation model[40]. No significant effect to the host-guest binding by the specific choice of the counterions is expected and they should mainly compensate for the present charges and reduce the electrostatic contribution to ΔE and the COSMO-RS solvation free energies. This approach was successfully applied before.[43,394,395] Symmetric molecules were always treated in their respective point group (*vide infra*). Corrections for basis set superposition error (except for HF-3c which uses gCP) were not made.

Single-point calculations for the final gas phase interaction energies were conducted at the PW6B95[101]-D3, B3LYP[88,91,92,95,96]-D3, BLYP[91,92]-D3, PBE[89,90]-D3, TPSS-D3 and ωB97X-D3[106] levels together with the quadruple-ζ basis set def2-QZVP' (QZ) with discarded g- and f-functions on the non-hydrogen and hydrogen atoms, respectively.[237] For such a large basis set the basis set superposition error (BSSE) almost vanishes (remaining BSSE at this level is typically 1-2% of ΔE, see Ref. [192]) and hence, no special treatment, e.g. such as a computationally demanding counter-poise correction, is required. Basis set extrapolation (TZ,QZ) of the energies for similar complexes yielded at maximum a change of about 0.5 kcal mol^{-1} for ΔE.[367] For iodine the pseudopotentials from the Stuttgart/Cologne group were used.[260]

All DFT calculations were performed using the TURBOMOLE 6.4 program pack-

age[236,261], except for the ωB97X-D3 single-points which were carried out with ORCA 3.0.1.[233,234] and the M06-2X[100] single-points which were calculated with TURBOMOLE 6.6. The resolution-of-identity (RI) approximation for the Coulomb integrals[74] was applied in all cases using matching default auxiliary basis sets.[238] For the integration of the exchange-correlation contribution the numerical quadrature grids *m4* (*m5* in case of M06-2X)[239] and grid 5 (final-grid 6) were employed in TURBOMOLE and ORCA, respectively. In the geometry optimizations as well as for the single-point energies the default convergence criteria were used (10^{-7} E_h for energies and 10^{-5} E_h/Bohr for gradients). The three-body contribution to the dispersion energy were calculated using the `dftd3` program.[241] Computations of TPSS-D3 frequencies for comparison were carried out numerically and in parallel with SNF 2.2.1[396] and the values were used unscaled.

The HF-3c[205] method was used for calculations of harmonic vibrational frequencies, geometry optimizations and energies. The HF part was computed using TURBOMOLE 6.4 (with the same convergence criteria as given above) and the 3c-terms to the energy and the nuclear gradients were calculated with our own freely available code[241]. For the optimizations of complexes with counterions the COSMO model was used. Computations of the HF harmonic vibrational frequencies were performed analytically using the `aoforce` code of TURBOMOLE. The 3c-contributions to the Hessian are computed numerically by two-point finite differences of analytical gradients.[241] To determine G^T_{RRHO}, the HF-3c vibrational frequencies were scaled with a factor of 0.86. For the host molecules tweezer, CA10, and XB-donor and complex **23** small imaginary frequencies of about $< i50$ cm^{-1} were obtained and inverted to the corresponding positive value in order to minimize the error resulting from missing low-lying modes.

SCC-DFTB3-D3 energies, geometries and frequencies were computed with DFTB+[397] using the full third-order correction[127], self consistent charges (SCC) and the empirical damping for hydrogen-containing pair potentials[113,114,300] together with the most recent Slater–Koster files provided by the group of Elstner.[355] For simplicity the SCC-DFTB3-D3 method is referred to as DFTB-D3. The SCC tolerance was set to 10^{-7} E_h and refitted D3 parameters determined recently[227] were used. For the geometry optimizations the `statpt` code of TURBOMOLE 6.4 was used for executing the relaxation steps and again, for the complexes with counterions the COSMO model was employed. The DFTB-D3 frequencies were used unscaled for calculating G^T_{RRHO}. For some complexes (**4**, **22**, **23**, **27** **28**), host (XB-donor and CB6) and guest molecules (5CPPA, 6CPPA, TMPDA, DAAD, and acetate) small imaginary frequencies of about $< i50$ cm^{-1} were obtained and therefore inverted. Imaginary frequencies larger than $i100$ cm^{-1} were observed for TCNB and TNF, disregarded and not used in the calculation of G^T_{RRHO}.

PM6-D3[124], PM6-D3H2[262] and PM7[125] calculations were carried out with MOPAC 2012[263] but employing the TURBOMOLE modules for executing the geometry relax-

ation steps. As the parametrization of the hydrogen bonding corrections was done in combination with D3 employing the zero damping (termed D3(0)), we used this scheme consistently for all PM6-D3 calculations. The same holds for DFT with the M06-2X functional where only D3(0) is appropriate. For the H+ corrections[398] we used our own code, the H4 corrections[399] were computed using the standalone code provided by the Hobza group. Vibrational frequencies for PM6-D3 were calculated numerically using MOPAC 2012. To obtain G^T_{RRHO} the frequencies were used unscaled. For most host molecules and most complexes imaginary frequencies were obtained, and again those smaller than $i50$ cm^{-1} were inverted and larger ones disregarded.

OM2[126] energies were calculated with MNDO 2005[400]. The SCF convergence criterion was set to 10^{-6} eV. The dispersion contribution was calculated using the `dftd3` standalone code.

The COSMO-RS solvation model[41,42] was used as implemented in COSMOtherm[401] employing the 2012, 2013 and 2014 BP86/def-TZVP parametrization as well as 2013 and 2014 BP86/def2-TZVPD parameters(dubbed fine parametrizations in the following). To obtain the solvation free energies the standard procedure with two single-point calculations (one in the gas phase and one in an ideal conductor) on the default BP86[91,265]/def-TZVP[266] or BP86/def2-TZVPD[402] levels of theory were performed on the optimized gas-phase geometries and then used as input for COSMOtherm.

The SMD[54] calculations based on COSMO charges were performed on the BP86/def2-SVP level with the implementation in NWChem 6.4[403]. For the solvent mixture in case of **27** and **28** the solvation free energies were averaged.

All visualizations of molecules were done with USCF Chimera version 1.8.1[267] and all graphs were plotted with Gnuplot 4.6.[357].

7.3. Results and discussion

Molecular symmetry influences the rotational part of the entropy and the effect of the symmetry number σ is significant for highly symmetric molecules.[349] Therefore, we treated the symmetric guest and host molecules in their respective point group: C_{60} has I_h and C_{70} has D_{5h} symmetry; 5CPPA is D_{5h}, 6CPPA and CB6 D_{6h}, 8CPPA D_{8h} and 9CPPA D_{9h} symmetric; TCNQ, DCB, TCNB and BQ have D_{2h} symmetry; tetraphene has C_{2h} symmetry; benzoate and the tweezer are C_{2v} symmetric; XB-donor and HHTAP have C_{3v} symmetry; pincer and tweezer2 are C_2 symmetric; FDNB, ADDA, crysene, acetate, CA10, DAAD and AAAA are C_s symmetric; and $Ad_2(NMe_3)_2$ has C_i symmetry. Out of all complexes only four exhibit any symmetry. **9** is C_{2v}, **16** is C_{3v}, and **27** and **28** are C_s symmetric.

The complexes **23**, **27** and **28** are positively charged, **24** is doubly positive charged,

25 and **26** carry a fourfold positive charge and **29** and **30** are negatively charged. These systems were investigated with and without counterions. For simplicity, chloride ions were chosen for all cationic complexes and sodium ions for the anionic ones, although in the experiment iodide was present in the case of **24**, $[B(3,5\text{-}(CF_3)_2C_6H_3)_4]^-$ for **23** and hexafluorophosphate in case of **25** and **26**. For an easier assignment of the systems we will refer to the set without counterions as S30L and the one including them as S30L-CI. In general, we assume that the additional chloride or sodium ions do not effect the vibrations significantly. Hence, the same ΔG_{RRHO} values as for the charged systems without counterions are used to calculate the final ΔG_a.

In the following sections we evaluate the different possible method combinations for ΔG_a, investigate the influence of counterions, and find the presumably best method combination to reproduce the experimental values. With help of the most accurate ΔG^T_{RRHO} and $\Delta G^T_{solv}(X)$ empirical ("experimental") binding energies are created and used for benchmarking low-cost semiempirical methods. And finally, the performance of these semiempirical methods for the geometries of the complexes is discussed.

7.3.1. Evaluating the various method combinations

The three (free) energy components in Eq. 7.3 can independently be computed by different theoretical methods and very many combinations are possible. A priori it is not clear if certain method combinations can benefit from error compensation and are thus preferred over others to compute accurate ΔG_a values. Finding such cases and exploring the general sensitivity of the results to the applied theory levels is the topic of this section. Note that it is currently not clear which of the three components ultimately limits the overall accuracy and in which direction future theoretical work should be invested.

For calculations of the binding energies ΔE we tested the following functionals together with the D3(BJ) dispersion correction: PW6B95-D3 because it worked well for this purpose before[46,306] and yielded good results for the GMTKN30 general thermochemistry test set[86], B3LYP-D3 and PBE-D3 because they are widely used, TPSS-D3 as another (meta-)GGA and ωB97X-D3 as a long-range corrected range-separated hybrid which should be preferable for charged systems due to reduced self-interaction error. For all DFT calculations the three-body dispersion (ATM) contribution is included and D3(BJ)+ATM is abbreviated as just "D3" in the following, unless noted otherwise. The harmonic frequencies for evaluation of ΔG^T_{RRHO} were calculated with the semiempirical methods HF-3c, PM6-D3 and DFTB-D3 and $\Delta G^T_{solv}(X)$ was obtained with COSMO-RS and SMD. For COSMO-RS several parameter sets were used: the BP86/def-TZVP parameters from 2012, 2013 and 2014 and the BP86/def2-TZVPD parameters from 2013 and 2014 (called fine in the following).

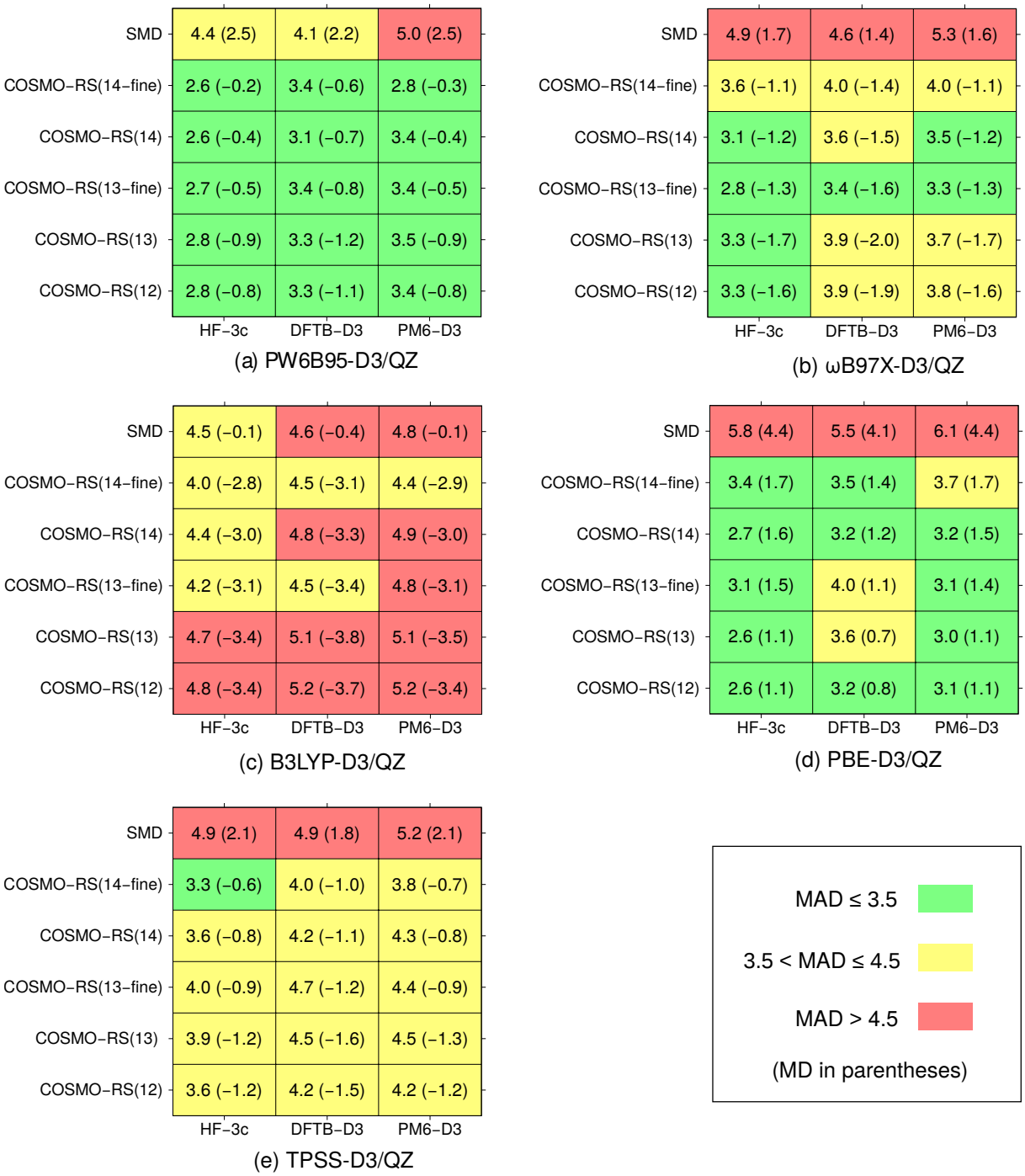


Figure 7.2.: MADs (and MDs) for S30L for several combinations of functionals (ΔE), semiempirical methods for vibrational frequencies (ΔG^{T}_{RRHO}), and continuum solvation models ($\Delta\delta G^{T}_{solv}$) with respect to the experimental ΔG^{exp}_{a} values. A negative MD corresponds to overbinding, a positive one to underbinding. All values are given in kcal mol^{-1}.

Figure 7.2 shows the mean absolute deviation (MAD) from experiment (and the mean deviation (MD) in parenthesis) for all possible method combinations. The values are color-coded in order to provide an easy overview as suggested by Martin and co-workers.[212]

The choice of the functional has the largest effect on the performance of the entire procedure. Within our of course limited test suite, PW6B95-D3 and PBE-D3 perform best yielding lowest MADs of about 3 kcal mol^{-1}. This confirms our previous finding that PW6B95-D3 yields accurate supramolecular binding energies. The good performance of PBE-D3 is somewhat surprising due to its known tendency to overbind hydrogen bonds[86,207]. However, we find that compared to e.g. PW6B95-D3 the error for hydrogen bonded systems is here not larger than for other complexes. ωB97X-D3 and TPSS-D3 give slightly worse results with MADs increased by about 0.5-1 kcal mol^{-1}. This conclusion is rather independent of the choice of the source of the thermostatistical as well as solvation corrections (see below). For B3LYP-D3 the MADs of about 5 kcal mol^{-1} are the largest obtained. The MD of B3LYP-D3 is also worst (about -3.5 kcal mol^{-1}) and indicates a strong tendency to overbind. This is also the reason why B3LYP-D3 is the only functional that does not yield worse results together with SMD as the SMD solvation contributions are usually larger than those from COSMO-RS. Since B3LYP-D3 performs very well for small noncovalently bound complexes[250], we tentatively attribute this to an inconsistent treatment of many-body effects (see below) in this very over-repulsive functional (PW6B95, PBE, and TPSS are inherently much less repulsive).

The conclusion for the performance of the functionals is not affected by the inclusion of the three-body dispersion term. Disregarding it worsens almost all MADs by at least 0.5 kcal mol^{-1} when COSMO-RS is used as a solvation model. When SMD is employed in most cases the MAD is better without the three-body dispersion due to the already mentioned larger values for $\Delta\delta G^{T}_{solv}$. For analysis purposes we tested if scaling the three-body dispersion (in a range from 0.5 to 2.1) improves the performance. In case of PW6B95-D3 and ωB97X-D3 the unscaled three-body contribution yields mean absolute deviations and standard deviations (SD) very close to the optimum. For PBE-D3 a slight improvement in both measures is observed when the three-body dispersion is scaled down, for TPSS-D3 a slight improvement is gained when it is scaled up. In the case of B3LYP-D3 scaling up the three-body term yields a significant improvement of 0.5 kcal mol^{-1} of both the MAD and SD. This is consistent with the finding that B3LYP-D3 shows a strong tendency to overbind. TPSS-D3 and B3LYP-D3 both have a relatively large scaling factor s_8 of about two for the higher order two- body dispersion term in the D3 scheme. To check whether this might be connected, we also tested the BLYP functional which has a similarly large s_8 value of about 2.7. The behavior is the same as seen for B3LYP and it is even more pronounced. Particularly the complexes **9**-**12** with C_{60} or C_{70} as guests where the ATM term is large are overbound with these functionals. These observations suggest that functionals with a smaller s_8 values in the D3 scheme yield in general better results for the binding energies of large systems.

When comparing the semiempirical methods tested for the frequencies it is evident

that HF-3c performs slightly better than DFTB-D3 and both yield better results than PM6-D3. HF-3c also seems to be generally more stable for this purpose, since HF avoids any quadrature schemes and is practically free of numerical noise. This is of particular importance for low-lying vibrational modes. Contrary to HF-3c, imaginary modes were obtained for a few complexes, host and guest molecules when DFTB-D3 was used (see computational details) and for almost all complexes when PM6-D3 was employed. For eight complexes (**1**, **3**, **7**, **9**, **15**, **17**, **22**, **25**), we also calculated (unscaled) TPSS-D3/def2-TZVP vibrational frequencies to obtain ΔG^{T}_{RRHO} values for comparison. The HF-3c results agree very well with the DFT ones, the absolute deviation is only 0.4 kcal mol^{-1} on average for theΔG^{T}_{RRHO} term (see SI for details, Table S5). If PM6-D3 and DFTB-D3 are used for the frequencies the MAD is larger, 1.1 and 1.6 kcal mol^{-1}, respectively. Note that anharmonic vibrational treatments for systems as large as the complexes discussed here are currently not possible.

For the solvation models the trend is not as clear as for the functionals and the thermostatistical contributions. In general, COSMO-RS outperforms SMD, but several parametrizations yield equally good results. It has been reported before that SMD performs worse for ions than other SMx methods, e.g. SM6 or SM8.[404] Excluding the charged systems (**23** to **30**) improves the results for both solvation models slightly (see supporting information, Figure S1). As the improvement is similar, the charged complexes cannot cause the worse performance of SMD compared the COSMO-RS. When COSMO-RS is used, the resulting MADs lie between 2.6 and 2.8 kcal mol^{-1} in combination with PW6B95-D3 energies and HF-3c thermostatistical corrections, and between 2.6 and 3.4 kcal mol^{-1} for PBE-D3 and ΔG^{T}_{RRHO}(HF-3c). We expected the 2014-fine parametrization to perform better than others because it uses the theoretically superior D3 scheme to describe solvation dispersion interactions instead of a much simpler surface-proportional approach employed for the other parameter sets, but this seems not to be the case. All COSMO-RS versions yield similar MADs and MDs, and for all of them outliers are observed. This will be discussed in more detail in the next sections.

Overall, there is no method combination that very clearly outperforms others and at this point we recommend to use the PW6B95-D3 functional (together with a quadruple-ζ basis set) for the gas phase association energies ΔE, HF-3c for the thermostatistical corrections ΔG^{T}_{RRHO} and COSMO-RS for the solvation contributions $\Delta\delta G^{T}_{solv}$.

7.3.2. Influence of counterions

The results discussed so far were obtained without including counterions for the charged systems **23** to **30**. In the S30L-CI test chloride ions were included for the cationic systems and sodium ions for the anionic ones meaning that in all quantum chemical calculations of

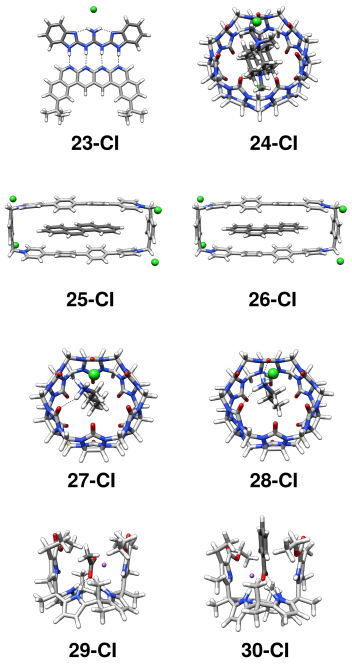


Figure 7.3.: Structures of the eight supramolecular complexes contained in the S30L-CI test set which have counterions (chloride ions drawn green, sodium ions are purple). C-atoms of the host molecules are shown in light grey, those of the guest molecules in dark grey, and H-bonding interactions are indicated by dotted lines.

energeis (including those required by the continuum solvation models) only overall neutral species appear. The structures of complexes **23-CI** to **30-CI** are shown in Figure 7.3. As mentioned before, the same thermostatistical corrections as for the S30L were used, so they will not be discussed again.

Figure 7.4 shows the color-coded MADs with respect to experiment (and the MDs in parentheses) for all possible method combinations for the S30L-CI in the same manner as before. Whether the inclusion of counterions improves or deteriorates the results strongly depends on the functional. The largest improvement is observed for ωB97X-D3, which now outperforms PW6B95-D3. B3LYP-D3 also yields better results and now performs similar to TPSS-D3. For PW6B95-D3 the MADs stay similar and for TPSS-D3 and PBE-D3 a slight deterioration is obtained compared to the results for S30L. The observations for the solvation contributions are the same as for S30L: COSMO-RS outperforms SMD for $\Delta\delta G^{T}_{solv}$ but the five versions give similar results.

The individual errors in ΔG_a for the charged complexes with respect to experimental data are shown for ωB97X-D3 in Figure 7.5. The largest improvement is observed for complex **24**. Without counterions it is overbound by -11.1 kcal mol^{-1}, and upon inclusion of chloride ions the error is reduced to only -0.3 kcal mol^{-1}. In case of **23**, **25**, **26**, **27** and **28**, the deviation compared to experiment decreases by about 2-4 kcal mol^{-1}. For

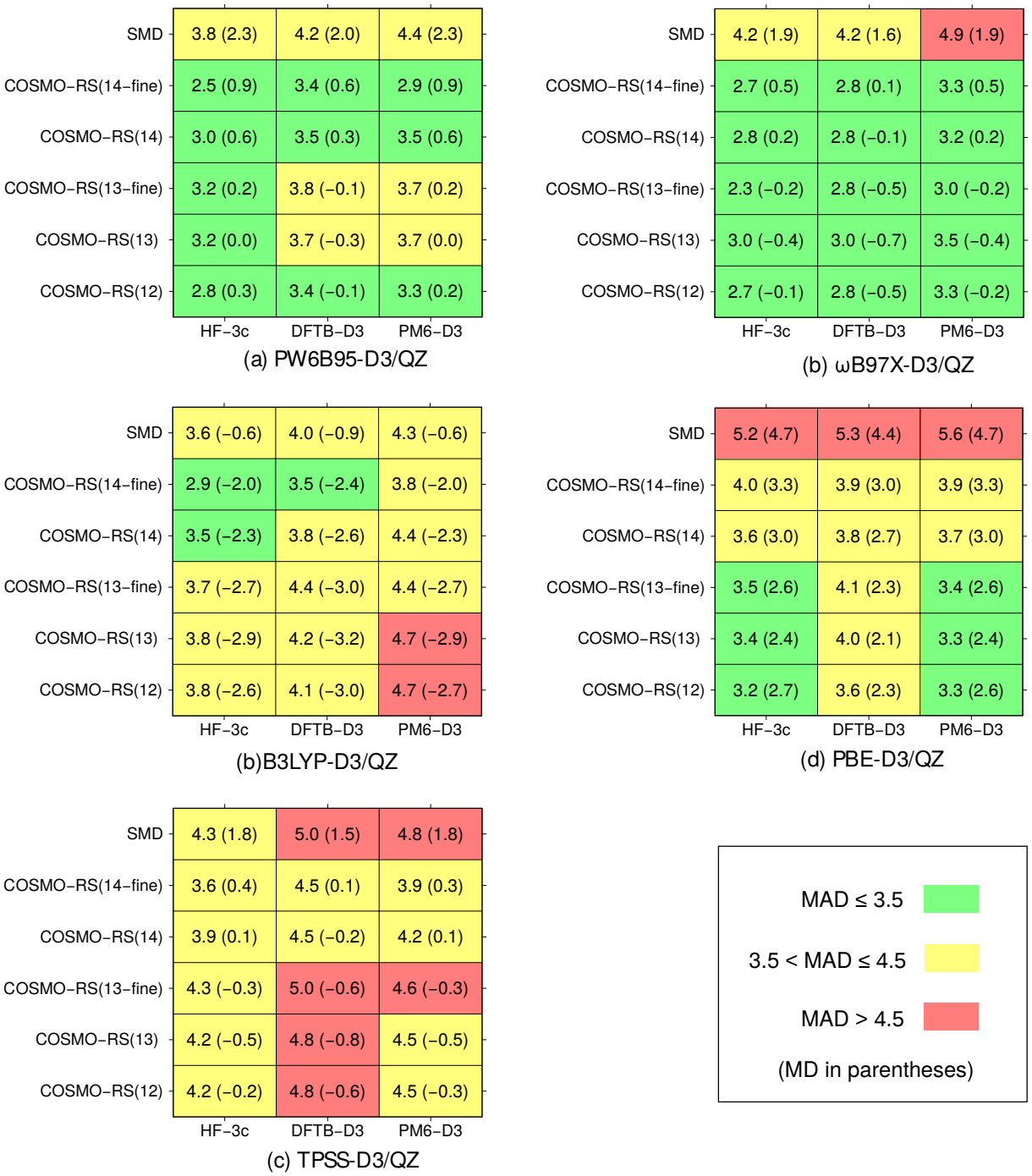


Figure 7.4.: MADs (and MDs) for S30L-CI for several combinations of functionals (ΔE), semiempirical methods for vibrational frequencies (ΔG^{T}_{RRHO}), and continuum solvation models ($\Delta\delta G^{T}_{solv}$) with respect to the experimental ΔG^{exp}_{a} values. The three-body dispersion contribution $\Delta E^{(3)}_{disp}$ is always included. A negative MD corresponds to overbinding, a positive one to underbinding. All values are given in kcal mol^{-1}.

25 and **26**, the complexes with the highest charge of +4, the improvement is not larger when counterions are added than for other complexes. Due to the fully conjugated π-

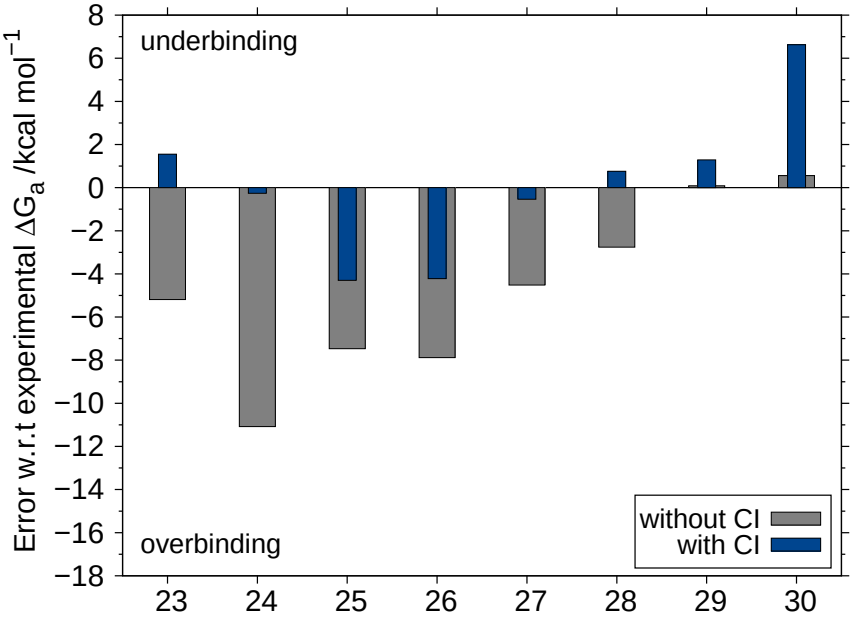


Figure 7.5.: Deviations for the method combination ωB97X-D3/QZ, HF-3c and COSMO-RS(13-fine) including and disregarding counterions for the charged systems **23** to **30** with respect to the experimental ΔG_a^{exp} in kcal mol^{-1}.

system, the charges are highly delocalized and thus seem to cause less problems for the continuum solvation models compared to the more localized charges in the other systems. Other COSMO-RS parametrizations gave similar results in combination with the ωB97X-D3 functional for the six cationic systems. For other functionals the improvement or deterioration upon inclusion of chloride ions can differ from the results shown for ωB97X-D3.

In contrast, for the two anionic systems ωB97X-D3 yields worse results when sodium counterions are taken into account. All other functional behave in the same way, though the extent of deterioration may differ. Again, these observations do not depend on the COSMO-RS parametrization employed. This result is surprising because the tested density functional approximations should have more problems in describing anions than cations and this might suggest the general omission of sodium counterions for anionic systems.

7.3.3. Finding the optimal method combination

As described above, PW6B95-D3 yields best gas phase interaction energies for S30L, and ωB97X-D3 performs best for the S30L-CI set. HF-3c consistently gives the best thermostatistical contributions. In terms of MAD and MD all COSMO-RS parametrizations perform similar, and for all of them outliers exist.

Figure 7.6(top) shows the comparison of the calculated ΔG_a with the experimental

ΔG_a^{exp} for the S30L set when the COSMO-RS(12) parameters are used. For non-polar solvents, like chloroform or dichloromethane, COSMO-RS(12) gives the most consistent results, which cannot be improved when any other parametrization is used. **22** clearly is an outlier for all COSMO-RS versions. According to our previous experience COSMO-RS(14-fine) yields better values for solvents like toluene, which can do $\pi - \pi$-stacking, in combinations with fullerene complexes due to the aforementioned better description of dispersion in COSMO-RS(14-fine).[367] For the two buckycatcher complexes **9** and **10** in toluene, the deviation of COSMO-RS(12) and COSMO-RS(14-fine) is small (up to 2 kcal mol^{-1}). Therefore, we decided to use the COSMO-RS(12) values. Moreover, the calculations are less time-consuming when the regular instead of the fine parametrizations are used.

For water as the solvent, the situation is different. With COSMO-RS(12) the complexes **21**, **24**, **27** and **28** show a large deviation from experiment, which can be reduced when a fine parametrization is used. For the six complexes in water the best and most consistent results were obtained with COSMO-RS(13-fine) (Figure 7.6(bottom)). An explanation for this may be, that the fine parametrizations employ the larger basis set def2-TZVPD instead of def-TZVP. It is well-known that for water a large basis set with diffuse functions is necessary for an accurate description[86]. The largest errors ($\geq$4 kcal mol^{-1}) for S30L are now observed for **6**, **11**, **13**, **22**, **25** and **26**. Complexes **6**, **11**, **25** and **26** are overbound by -4.0, -5.4, -4.5, and 4.6 kcal mol^{-1}, respectively, while **13** and **22** are underbound by 4.9, and 8.9 kcal mol^{-1}, respectively.

Complex **11** belongs to the ones which have a fullerene as guest molecule. For these type of large, electronically delocalized systems a partial breakdown of pairwise dispersion interaction schemes has been observed and analyzed before.[405,406] This is confirmed by our study as reflected in the overbinding observed for all complexes including fullerenes, though less pronounced for **9**, **10**, and **12**. Although within our D3 dispersion model three-body-effects are only very approximately included, the three-body dispersion for these four systems is qualitative correctly and found to be larger (3.3 to 5.1 kcal mol^{-1}) compared to other complexes (excluding **24**). Many-body-dispersion effects beyond the three-body term are disregarded in our model, but the observed overbinding provides evidence that neglecting the higher order terms introduces an error of about 1-2 kcal mol^{-1} for theses systems. Recent work on molecular crystals and comparison of D3 and MBD data indicates,[207,370] that many-body effects beyond three-body dispersion are probably rather small for saturated or not too unsaturated organic molecules. This conclusion is in agreement with previous results from wave function theory based analysis (for cooperativity in noncovalent interactions of biologically relevant molecules see Ref. [407] and for a study on crystalline benzene see Ref. [137]). For six of the complexes (**1**, **2**, **9**, **17**, **21**, and **27**) PBE-MBD* calculations exist.[299] Compared to our PBE-D3 values the PBE-MBD*

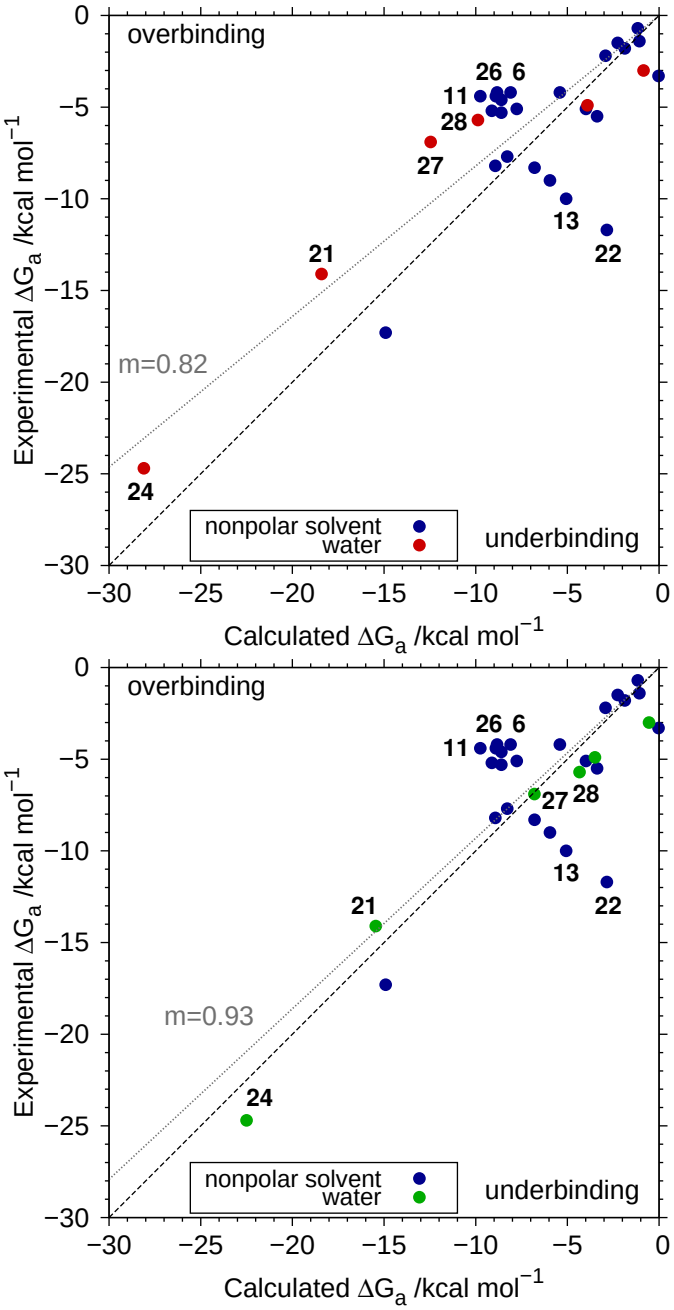


Figure 7.6.: Comparison of experimental ΔG_a^{exp} and computed total free association energies ΔG_a for S30L obtained with the method combination PW6B95-D3/HF-3c/COSMO-RS(12) throughout (top), and PW6B95-D3/HF-3c/COSMO-RS(12) for non-polar solvents and COSMO-RS(13-fine) for water (bottom). The dashed grey line shows the result of a linear regression with slope m.

results for ΔE are always more negative and the deviations lie in the range of -0.6 to -1.8 kcal mol^{-1} for five of the complexes. For **21**, the deviation is with -4 kcal mol^{-1} much larger, but compared to the DQMC values PBE-MBD* seems to be overshooting in this case.

For S30L-CI ωB97X-D3 performs better than PW6B95 in terms of MAD (the MD is

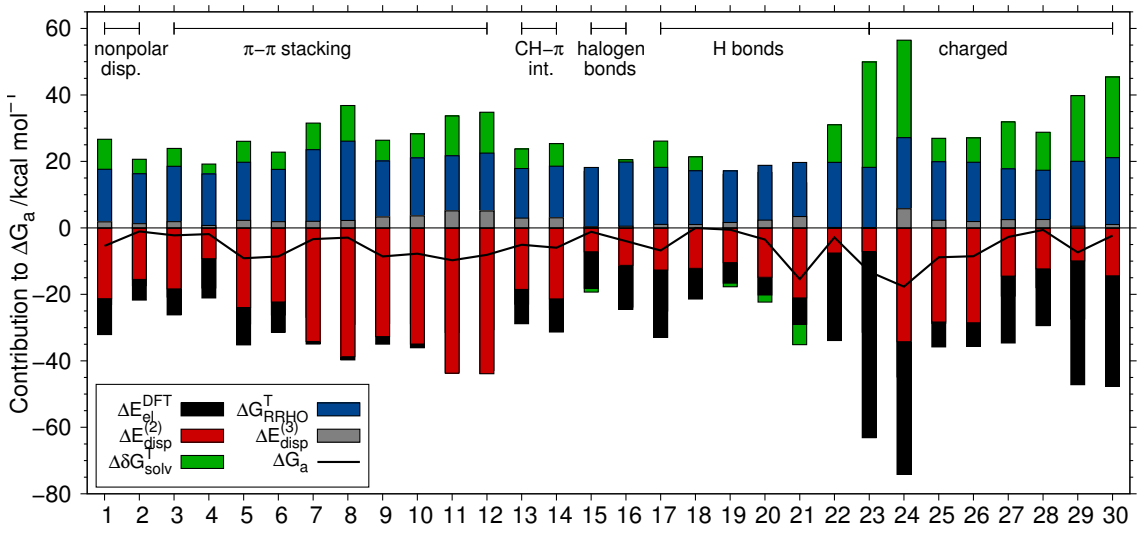


Figure 7.7.: Contributions to ΔG_a for S30L-CI (pure electronic energy (ΔE_{el}^{DFT}), two-body ($\Delta E_{disp}^{(2)}$) and three-body ($\Delta E_{disp}^{(3)}$) dispersion energy, thermal corrections from energy to free energy (ΔG_{RRHO}^{T}) and solvation free energy ($\Delta \delta G_{solv}^{T}$)) and total ΔG_a values for the method combination PW6B95-D3/HF-3c/COSMO-RS(12/13-fine). The lengths of the bars represent the size of the contributions. All values are given in kcal mol^{-1}.

similar and about zero). Complexes **1** to **22** are the same as in the S30L and the observations described for PW6B95-D3 and the various COSMO-RS parameter sets hold for ωB97X-D3. In case of the complexes with counterions the COSMO-RS(13-fine) parametrization also yields the most consistent results. It does not always give the smallest error, but no outlier is observed.

Our best method combination to compute ΔG_a now consists of PW6B95-D3 gas phase association energies including the three-body dispersion term or ωB97X-D3 energies when counterions are included, HF-3c frequencies to obtain the thermostatistical corrections from energy to free energy, and COSMO-RS with the 2012 parametrization to calculate the solvation free energies for non-polar solvents and the 2013-fine parametrization for water and when counterions are included.

This combination yields an MAD of only 2.4 kcal mol^{-1} for the S30L and PW6B95-D3 (2.9 kcal mol^{-1} for the S30L-CI) and 2.1 kcal mol^{-1} for the S30L-CI and ωB97X-D3 (2.6 kcal mol^{-1} for the S30L). When judging this deviation one should keep in mind that the small ΔG_a values result as a sum of individually large and oppositely signed contributions as discussed already in Ref[46] (see Figure 7.7). The MD for both approaches is almost zero (-0.5 to 0.5 kcal mol^{-1}) indicating the absence of systematic errors (or a very favorable systematic compensation). Because the final ΔG_a values for many (neutral) complexes are rather small (between -5 and -10 kcal mol^{-1}), the mean relative deviation for the whole set is with about 50% rather large. The overall correlation between experiment

and theory is reasonable, the Pearson correlation coefficient is $R = 0.80$ for PW6B95-D3 on S30L and $R = 0.89$ for ωB97X-D3 on the S30L-CI set. A linear regression gives a slope of 0.93 for S30L/PW6B95-D3 and 0.92 for S30L-CI/ωB97X-D3 indicating only minor systematic deviations.

Figure 7.7 shows the individual contributions to the total free energy of association exemplary for the method combination PW6B95-D3/HF-3c/COSMO-RS(12/13-fine) for S30L-CI. The pure electronic DFT energy ΔE_{el}^{DFT} ranges from -53.3 kcal mol^{-1} for **23-CI**, which shows nearly pure hydrogen bonding, to almost zero for **11** and **12**, which are mainly bound by dispersion interactions. The the two-body dispersion $\Delta E_{disp}^{(2)}$ contribution ranges from -5.2 to -43.9 kcal mol^{-1}. For **22** and **23-CI**, which geometrically have a small contact area between host and guest, the $\Delta E_{disp}^{(2)}$ part is with about -5 to -7 kcal mol^{-1} small compared most other complexes. As we have already shown previously for many supramolecular systems[46,49,367] the two-body dispersion contribution easily outranks the pure electronic DFT energy. The three-body dispersion term $\Delta E_{disp}^{(3)}$ ranges from almost zero for the nearly planar complexes **22** and **23-CI** up to 5.9 kcal mol^{-1} for **24-CI**. Again, we find that $\Delta E_{disp}^{(3)}$ contributes repulsively with 2 to 3 kcal mol^{-1} on average. The thermal corrections from energy to free energy ΔG_{RRHO}^{T} vary less, between 15 and 20 kcal mol^{-1}. The solvation contribution $\Delta\delta G_{solv}^{T}$ is positive in most cases except for complexes **19**, **20** and **21**, and ranges from -6.2 for **21** to 31.7 kcal mol^{-1} for **23**. As Figure 7.7 also shows the total ΔG_a values, one can easily see that a large gas phase association energy does not necessarily result in a large association free energy in solution if the solvation contribution is big. Therefore, all parts are significant, and solely ΔE cannot be used to explain binding affinity trends in solution.

In 13 cases two complexes (deliberately) share the same host molecule. For these systems relative binding affinities $\Delta\Delta G_a$ were calculated and compared to experiment in order to evaluate if our best method combination is able to correctly reproduce those. The results are shown for PW6B95-D3/HF-3c/COSMO-RS(12/13-fine) in Figure 7.8. For two pairs of complexes, **3/4** and **13/14** the wrong trend is observed. For all others it is correct but the differences in binding tend to be overestimated. Compared to experiment the MAD for $\Delta\Delta G_a$ is 1.2 kcal mol^{-1} and the MD is almost zero. This improved accuracy compared to the results for ΔG_a indicates favorable error compensation.

7.3.4. Binding energy reference values for benchmarking purposes

In order to provide convenient gas phase binding energies for benchmarking, we back-correct the experimental binding free energies ΔG_a^{exp} to obtain empirical ("experimental") binding energies ΔE^{emp} as reference values in analogy to the original work for S12L.[46,192] Therefore, we subtract the best thermostatistical corrections calculated with HF-3c and

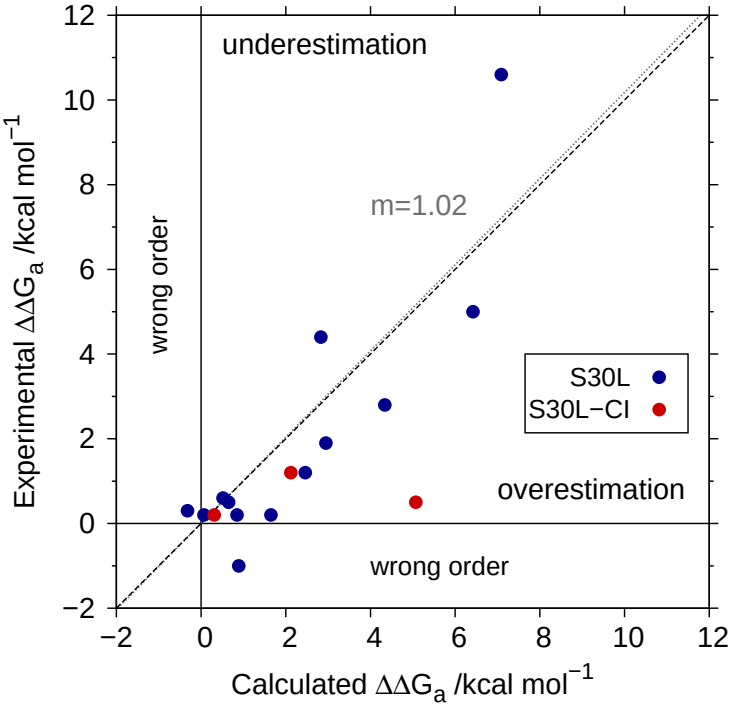


Figure 7.8.: Comparison of experimental $\Delta\Delta G_a^{exp}$ and computed $\Delta\Delta G_a$ for complexes with the same host and different guest molecules on the PW6B95-D3/HF-3c/COSMO-RS(12/13-fine) level. The dashed grey line shows the result of a linear regression for S30L with slope 1.02.

the solvation contributions obtained with COSMO-RS(12) for non-polar solvents and COSMO-RS(13-fine) for water from the experimental ΔG_a^{exp}:

$$\Delta E^{emp} = \Delta G_a^{exp} - \Delta G_{RRHO}^{T}(\text{HF-3c}) - \Delta\delta G_{solv}^{T}(X)(\text{COSMO-RS(12/13-fine)}) \qquad (7.4)$$

These values are collected in Table 7.2. Since the $\Delta\delta G_{solv}$ values vary for different COSMO-RS parametrizations and we think that the solvation contribution is the least accurate theoretical component, we decided to define our error in ΔE^{emp} as 10% of the chosen $\Delta\delta G_{solv}$ COSMO-RS results. For comparison the calculated ΔE values on the presumably most accurate DFT level (PW6B95-D3/def2-QZVP') are provided. Further, a comparison of the two best functionals PW6B95-D3 and ωB97X-D3 with the reference values ΔE^{emp} is shown in Figure 7.9. Except for the aforementioned outlier **22** the PW6B95-D3 energies agree well with the back corrected ΔE^{emp}. As noted above, for S30L both data sets have an MAD of 2.4 and 2.8 kcal mol^{-1}, respectively, and for the S30L-CI 2.6 and 2.1 kcal mol^{-1}, respectively. These values should be kept in mind when other theoretical methods are benchmarked against ΔE^{emp}. At present it seems difficult to guess if the residual errors in the DFT-D3 treatment or the inherent errors in the experimental data and the back-correction scheme are larger. As will be shown below, however, the agreement achieved is sufficient to benchmark lower-level quantum chemical methods

Table 7.2.: Empirical binding energies ΔE^{emp} for the S30L and S30L-CI sets obtained via back correcting the experimental ΔG_a^{exp} values and our best calculated binding energies ΔE^{calc} on the PW6B95-D3/def2-QZVP' level for comparison. Complexes of the S30L-CI including counterions are indicated with a "CI". All values are given in kcal mol^{-1}.

	ΔE^{emp}	ΔE^{calc}		ΔE^{emp}	ΔE^{calc}		ΔE^{emp}	ΔE^{calc}
1	-29.0 ± 0.9	-30.2	**14**	-31.3 ± 0.7	-28.28	**27**	-82.2 ± 6.0	-82.1
2	-20.8 ± 0.4	-20.4	**15**	-17.4 ± -0.1	-17.84	**28**	-80.1 ± 6.0	-78.8
3	-23.5 ± 0.5	-24.2	**16**	-25.1 ± 0.1	-24.02	**29**	-53.5 ± 2.6	-54.3
4	-20.3 ± 0.3	-20.3	**17**	-33.4 ± 0.8	-31.87	**30**	-49.3 ± 2.1	-49.9
5	-29.0 ± 0.6	-32.9	**18**	-23.3 ± 0.4	-20.38	**23(CI)**	-67.3 ± 3.2	-63.2
6	-25.5 ± 0.5	-29.5	**19**	-17.5 ± -0.1	-15.01	**24(CI)**	-75.4 ± 2.9	-68.4
7	-35.1 ± 0.8	-32.9	**20**	-19.2 ± -0.2	-17.85	**25(CI)**	-29.1 ± 0.7	-33.5
8	-36.8 ± 1.1	-37.5	**21**	-24.2 ± -0.6	-25.56	**26(CI)**	-29.4 ± 0.7	-33.7
9	-28.4 ± 0.6	-31.7	**22**	-42.6 ± 1.1	-33.79	**27(CI)**	-36.3 ± 1.4	-32.1
10	-29.8 ± 0.7	-32.4	**23**	-61.3 ± 2.6	-58.9	**28(CI)**	-32.0 ± 1.1	-26.9
11	-33.0 ± 1.2	-38.3	**24**	-135.5 ± 8.9	-133.4	**29(CI)**	-47.5 ± 2.0	-46.7
12	-33.9 ± 1.2	-37.8	**25**	-26.0 ± 0.4	-30.5	**30(CI)**	-52.1 ± 2.4	-46.7
13	-30.8 ± 0.6	-25.9	**26**	-25.8 ± 0.4	-30.4			

which are mostly less accurate than the mentioned 3 kcal mol^{-1} uncertainty. This holds even more for common force-field approaches, which are expected to have problems to reach average errors below 10 kcal mol^{-1} for S30L according to some preliminary test calculations.

For some complexes already contained in the old S12L set diffusion quantum Monte Carlo (DQMC) calculations were performed,[299] and for all of them symmetry adapted perturbation theory (DFT-SAPT) computations exist.[368] The binding energies obtained with these methods are in good agreement with our new ΔE^{emp}. A table with a direct comparison is provided in the SI (table S6). The maximal deviation between ΔE^{emp} and the DQMC and DFT-SAPT results amounts to 2.5 kcal mol^{-1} which is more than reasonable. This gives us confidence in the back correcting scheme and the choice of $\Delta\delta G_{solv}$ contributions.

Error distributions for the relative deviation from ΔE^{emp} are depicted in Figure 7.10 for the S30L and all tested functionals. The statistical data (MD, MAD, mean absolute relative deviation (MARD),mean relative deviation (MRD), and standard relative deviation (SRD)) are given in Table 7.3. Note that all methods including the semiempirical ones are benchmarked as usual by single-point energy computations. The widths (SRD) of the error distributions for PW6B95-D3/QZ, ωB97X-D3/QZ, and PBE-D3/QZ are small with about 10%. B3LYP-D3/QZ, BLYP-D3/QZ, and TPSS-D3/QZ have a broader distribution with widths of about 14-16%. PW6B95-D3/QZ and ωB97X-D3/QZ show no systematic

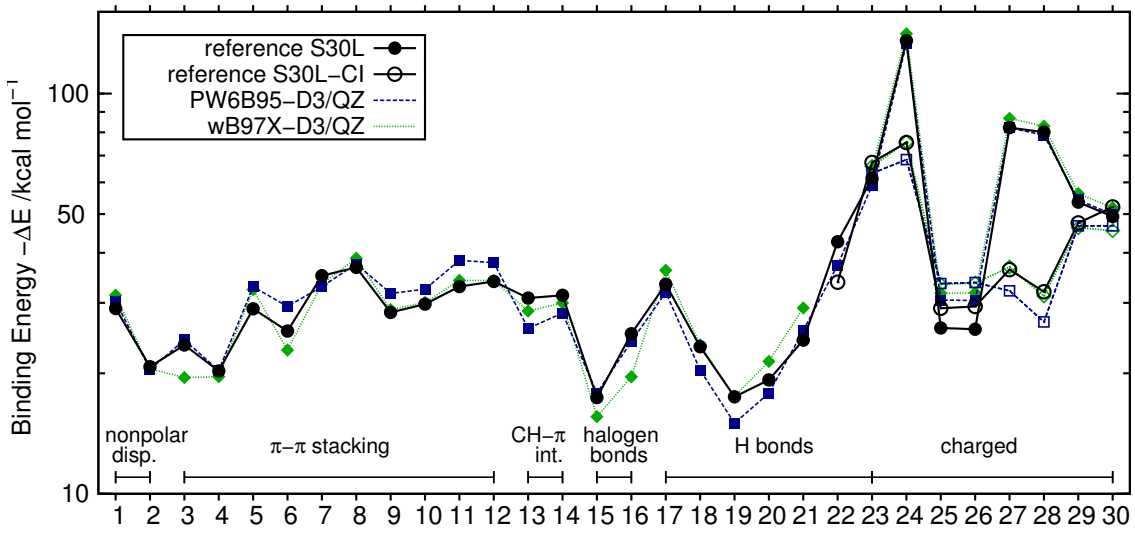


Figure 7.9.: Empirical binding energies ΔE^{emp} in comparison to binding energies for the best functionals PW6B95-D3 and ωB97X-D3 together with a QZ basis set.

error, whereas PBE-D3/QZ tends to underbind by about 7% on average. TPSS-D3/QZ, B3LYP-D3/QZ as well as BLYP-D3/QZ exhibit a systematic overbinding by about 5, 10 and 14%, respectively. TPSS-D3-/TZ performs similar to TPSS-D3/QZ with a bit worse MD and MRD and slightly better MAD, MARD and SRD.

Because the Minnesota functionals[100] include medium-range dispersion effects and are widely used for noncovalently bound systems, we tested the prototypical M06-2X functional also at the TZ basis set level so that the results are directly comparable to TPSS-D3/TZ. M06-2X/TZ yields a smaller standard deviation than TPSS-D3/TZ of about 8% and tends to underbind by about 5% similar to TPSS-D3/TZ. Adding the D3(0)+ATM corrections yields overbinding by about 14 % on average indicating the appearance of the residual BSSE.

Table 7.3.: MAD and MD in kcal mol^{-1}, MARD, MRD and SRD in % for S30L for various functionals compared to ΔE^{emp}. A negative MD means overbinding and a positive MD underbinding.

	MAD	MD	MARD	MRD	SRD
PW6B95-D3/QZ	2.4	-0.1	7.9	0.9	10.1
ωB97X-D3/QZ	2.6	-0.9	8.0	1.8	10.5
B3LYP-D3/QZ	4.1	-2.7	13.2	9.6	13.8
BLYP-D3/QZ	4.8	-4.1	16.0	13.8	16.0
PBE-D3/QZ	2.8	1.8	8.4	-5.5	9.7
TPSS-D3/QZ	3.6	-0.5	11.8	3.1	14.7
TPSS-D3/TZ	3.5	-1.6	11.1	5.8	13.2
M06-2X/TZ	2.5	1.4	8.1	-5.2	8.4
M06-2X-D3/TZ	4.8	-4.4	15.0	13.5	10.0

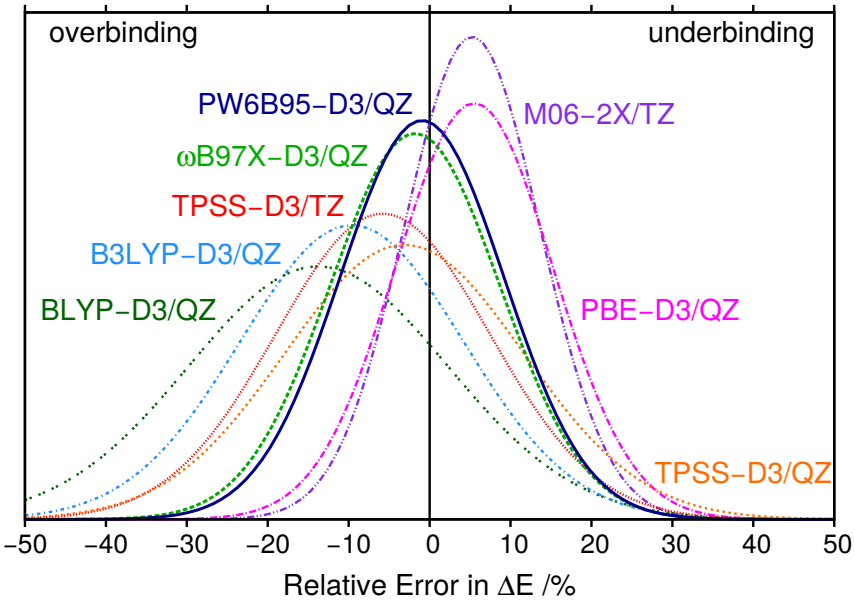


Figure 7.10.: Visualization of the statistical distribution of the relative errors in ΔE with respect to ΔE_{emp} for several functionals for S30L. The half widths of the Gaussians represent the SD and their shift from the origin corresponds to the MD.

7.3.5. Performance of semiempirical methods

The empirical binding energies ΔE^{emp} were used to benchmark the performance of various semiempirical (minimal basis set) methods which are typically about two to three orders of magnitude faster than DFT/"large basis" calculations. Thus, these methods can be used for e.g. pre-screening of different possible conformers or binding modes in supramolecular complexes if they yield sufficiently accurate binding energies.

We tested HF-3c, DFTB-D3, PM6-D3, PM6-D3 in combination with hydrogen bonding corrections of second generation H2[262] (PM6-D3H2), third generation H+[398] (PM6-D3H+), and fourth generation H4[399] (PM6-D3H4), as well as PM7[125] and OM2-D3[126]. For the semiempirical methods the abbreviation D3 does not include the three-body dispersion. If the ATM term is included it is stated explicitly. In the case of PM6, the zero damping for the D3 correction was used throughout, as the hydrogen bonding corrections were parametrized in this way. Of all the methods tested, HF-3c is the most expensive one because all integrals are computed (about 50 times slower than the "truly" semiempirical PMx or OMx methods). Nevertheless, HF-3c calculations can be carried out routinely on a standard workstation for hundreds to about one thousand atoms.

The MADs (and MDs) with respect to ΔE^{emp} are given in Figure 7.11 for the S30L and S30L-CI sets and with and without inclusion of the three-body dispersion. OM2 has no parameters for sodium, sulfur, chlorine, and iodine and hence, no results could be

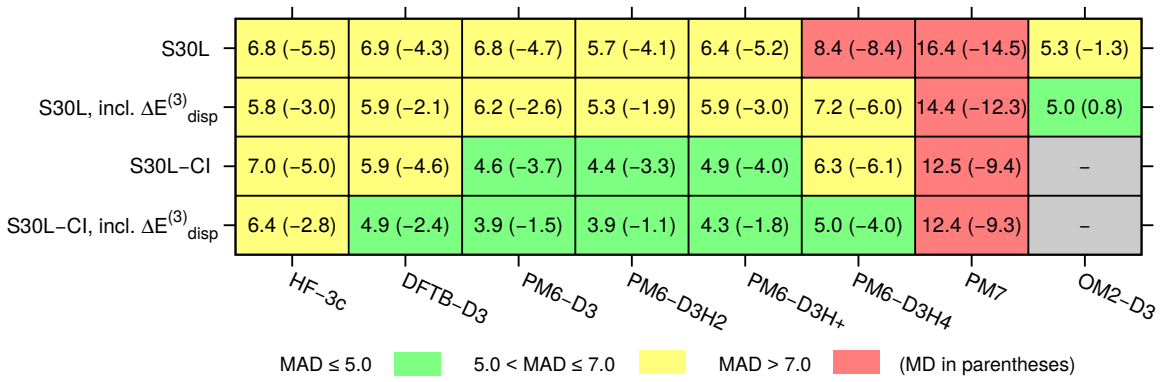


Figure 7.11.: MADs (and MDs) for calculated ΔE of S30L and S30L-CI for several semiempirical methods w.r.t. ΔE^{emp} in kcal mol^{-1}. The values are given with and without inclusion of the three-body dispersion term $\Delta E^{(3)}_{disp}$. For OM2-D3 six complexes of the S30L had to be disregarded due to missing parameters.

obtained for the complexes **4**, **11**, **12**, **14**, **15**, **16** and all complexes including counterions (**23-CI** to **30-CI**). Thus, for S30L-CI half of the complexes could not be treated and we disregarded the statistical analysis for S30L-CI with OM2-D3.

With the sole exception of PM7 (which shows a huge MAD of about 15 kcal mol^{-1}), all tested semiempirical methods can be recommended and in particular for neutral systems. Methods like PM6-D3H2 treat the dispersion and hydrogen bonding corrections independently of PM6, whereas for PM7 explicit terms for dispersion and hydrogen bonding are included in the method itself and optimized during the parameter fit. PM7 performs well for small organic complexes like in the S22[199] and S66[172] test sets but still slightly worse than e.g. PM6-DH2.[125,408,409] This result clearly shows that the PM7 parameters obtained for small systems cannot easily be used to describe the interactions in large supramolecular complexes.

For the S30L set, OM2-D3 and PM6-D3H2 yields the lowest MADs of 5.3 and 5.7 kcal mol^{-1}, respectively. PM6-D3H+, PM6-D3, HF-3c and DFTB-D3 show similar results ranging from 6.4 to 6.9 kcal mol^{-1}. PM6-D3H4 performs worse than PM6-D3 in combination with the other two hydrogen bond corrections tested, and yields a larger MAD of 8.4 kcal mol^{-1}.

As discussed for the DFT methods above, one can see a general improvement of the interaction energies when chloride counterions are used, except for HF-3c for which the MAD and MD almost stay the same. Furthermore, all methods benefit from the inclusion of the three-body dispersion. For DFTB-D3 and all PM6 based methods the results obtained with counterions and inclusion of the three-body dispersion contribution yield the smallest MADs compared to ΔE^{emp}. The MADs for these six methods range from 3.9 kcal mol^{-1} (PM6-D3) to 5.0 kcal mol^{-1} (PM6-D3H4). PM6-D3H2 and PM6-D3H+

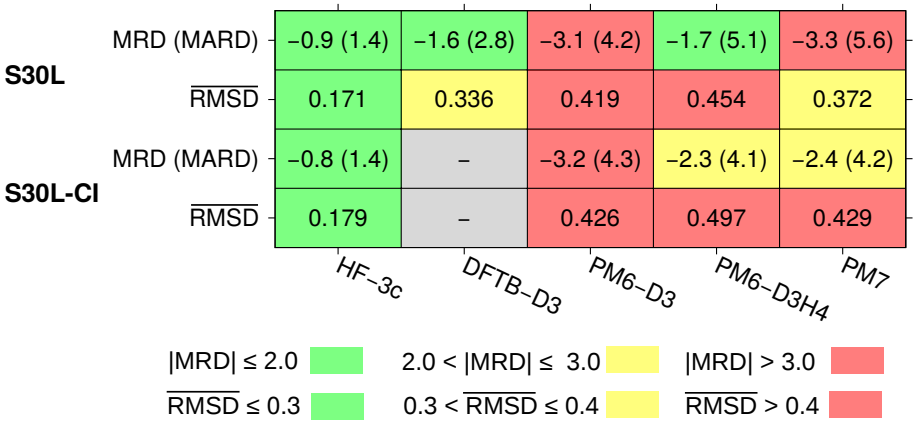


Figure 7.12.: The averaged RMSD ($\overline{\text{RMSD}}$) in Å for the heavy atoms and the MRD (and MARD) of the rotational constants in % w.r.t. TPSS-D3/def2-TZVP geometries for S30L and S30L-CI. For the rotational constants a negative MRD corresponds to a structure with a too small spatial extent.

perform similar to plain PM6-D3, and PM6-D3H4 performs worse. HF-3c has a larger MAD of 6.2 kcal mol^{-1}, which is mainly caused by the larger error for the halogen bonded systems **15** and **16** and the anionic ones **29/29-CI** and **30/30-CI**. For all methods the MDs are small and negative and lie between -1.1 (PM6-D3H2) and -4.0 kcal mol^{-1} (PM6-D3H4), showing a slight to small systematic overestimation of the binding energies.

The improvement gained with the various hydrogen bonding corrections compared to plain PM6-D3 for the whole set of complexes is small and in most cases negligible. PM6-D3H4 seems to perform generally a bit worse than plain PM6-D3. For the complexes whose interactions are dominated by hydrogen bonding (**17** to **23**) association energies obtained with PM6-D3 and PM6-D3 in combination with any hydrogen bonding correction differ by 3 kcal mol^{-1} at most. We could not identify a version that clearly outperforms the others or is better than PM6-D3 for all of these seven complexes. This was observed before for smaller model complexes as contained in e.g. the S66[172] or JSCH[199] test sets.[410]

7.3.6. Semiempirical methods for structure optimization

Finally, we compare the geometries of the complexes obtained with HF-3c, DFTB-D3, PM6-D3, PM6-D3H4 and PM7 with the ones obtained from TPSS-D3/def2-TZVP optimizations taken as reference. We left out OM2 because of convergence problems and PM6-D3H2 due to known small errors in the H2 gradient.[262] DFTB-D3 can not be applied to S30L-CI set due to the missing COSMO implementation in the DFTB+ code. As quality measure the average root mean square deviation ($\overline{\text{RMSD}}$) of the heavy atom coordinates and the mean relative deviation (MRD) and the mean absolute relative deviation (MARD) of the rotational constants were calculated (see Figure 7.12). For the rotational

constants, we define the error as $B_{semiemp.} - B_{TPSS-D3/def2-TZVP}$ so that a negative MRD denotes a structure with too small spatial extent. One has to keep in mind that TPSS-D3 structures have small errors themselves, but better hybrid DFT (or MP2 for saturated complexes) cannot be obtained routinely for such large systems which require typically hundreds of structure optimization steps. In general, for small noncovalent complexes DFT-D3 yields reasonable geometries.[411] For small covalently bound organic molecules TPSS-D3 yields geometries with slightly too large spatial extent[152].

All tested semiempirical methods except PM7 yield very similar $\overline{\text{RMSD}}$s for the complexes with counterions included than for those without. For PM7 the $\overline{\text{RMSD}}$ increases by about 0.06 Å. HF-3c yields by far the smallest $\overline{\text{RMSD}}$ of 0.171 Å for S30L, which is about half the magnitude of the values observed for all other tested semiempirical methods. These give values in the range of 0.336 Å (DFTB-D3) up to 0.454 Å (PM6-D3H4). The good agreement of HF-3c and DFT-D3/TZ structures has been observed before.[200,205] Note, that PM7 yielded the worst association energies but the geometries are as good as DFTB-D3 and (hydrogen bonding corrected) PM6-D3 results.

Regarding the rotational constants, HF-3c yields the best results with the smallest MRD of -0.9 % and the smallest MARD of 1.4 %. DFTB-D3 and PM6-D3H4 perform a bit worse for the MRD (-1.6 to -1.7 %) and PM6-D3 and PM7 yield much larger MRDs above 3 %. The MARDs are at least twice as large as for HF-3c and range from 2.8 % for DFTB-D3 to 5.6 % for PM7. For HF-3c and PM6-D3 the mean relative deviation of the rotational constants does not change much when counterions are included. In case of PM6-D3H4 a deterioration of 0.5 % and in case of PM7 an improvement of 0.9 % is observed. Overall, we find that HF-3c yields the most accurate geometries of all tested semiempirical methods, which agrees well with our previous finding that HF-3c yields the best thermostatistical corrections.

7.4. Conclusion

The S12L test set for supramolecular association free energies ΔG_a was extended to 30 complexes (S30L). It features complexes with higher charges (up to +4) and anions (-1), slightly less rigid hosts, more diverse types of noncovalent interactions, and larger system sizes (up to 200 atoms). The ΔG_a values were obtained in a non-empirical, static single structure approach by adding the computed gas phase binding energy, the thermostatistical corrections from energy to free energy and the solvation free energy.

Various dispersion corrected density functionals (PW6B95-D3, B3LYP-D3, TPSS-D3, PBE-D3, ωB97X-D3) in combination with a quadruple-ζ basis set were tested for calculating the association energies in the gas phase including our standard Axilrod-Teller-Muto type three-body dispersion correction. Various minimal basis set, semiempirical

methods (HF-3c, PM6-D3, DFTB-D3) were used to obtain the thermostatistical contributions. Several versions of the COSMO-RS model as well as SMD were employed to include solvation effects. In order to find the best procedure to predict theoretical ΔG_a values we investigated and statistically analyzed all possible combinations of these methods. The best performing density functional and the best method for thermostatistical corrections could be clearly identified, whereas for the solvation free energies many COSMO-RS parametrizations perform similarly. When looking closer, we found that the COSMO-RS(12) parameters perform best and most consistently for non-polar solvents, whereas for water COSMO-RS(13-fine) yields better results. Thus, the proposed method combination consists of PW6B95-D3/def2-QZVP' energies on TPSS-D3/def2-TZVP optimized geometries, HF-3c thermostatistical corrections, and COSMO-RS(12/13-fine) (for non-polar solvents/water) solvation free energies.

Further, we investigated the effect of counterions for the charged systems (S30L-CI) on the gas phase binding energy calculations as well as on the solvation term. The inclusion of counterions reduced the error in most cases for the cationic systems and is thus recommended as a default procedure. For the association energies ωB97X-D3 slightly outperforms PW6B95-D3 and thus is advised for charged systems. In case of the solvation term the COSMO-RS(13-fine) parametrization again works best. Our best method combinations give a mean absolute deviation of only 2.4 kcal mol^{-1} for S30L (PW6B95-D3) and 2.1 kcal mol^{-1} for S30L-CI (ωB97X-D3) and a mean deviation of almost zero compared to experimental ΔG_a^{exp}.

13 pairs of complexes (deliberatly) share the same host. For those relative association free energies were calculated in order to evaluate if the correct trend in binding affinities is observed. Except for two cases this always is the case. The MAD compared to experiment amounts to 1.2 kcal mol^{-1} and the MD is almost zero.

The thermostatistical and solvation data above were used to back-correct the experimental association free energies from solution measurements in order to get an empirical estimate for the "experimental" binding energies in the gas phase as done previously for the S12L set. These reference data are utilized to benchmark the performance of various semiempirical, minimal basis set methods for binding energies. HF-3c, DFTB-D3, OM2, PM6-D3 (with and without various hydrogen bonding corrections) and PM7 were tested. Apart from PM7 (whose errors are huge), all these methods perform rather similar although clearly worse than dispersion corrected DFT/"large basis set". They can be recommended in general for neutral complexes. For charged systems, the "simple" methods perform less well and their application in such cases requires careful testing on the specific problem under consideration. The choice of the hydrogen bonding correction for PM6-D3 has no significant impact. Again, inclusion of counterions improves the semiempirical results and so does the three-body dispersion term.

Finally, we investigated the geometries of the complexes obtained with the semiempirical methods. HF-3c yields the by far smallest averaged RMSD of 0.18 Å compared to TPSS-D3/def2-TZVP reference data. All other methods perform similar and give $\overline{\text{RMSD}}$s in the range of 0.34 Å (DFTB-D3) up to 0.45 Å (PM6-D3H4). This picture changes only slightly in favor of PM6-D3H4 when rotational constants are used to measure the quality of the structures.

In summary, the future for the prediction and understanding of supramolecular interactions by dispersion corrected DFT seems bright. If the structures are not too flexible and only a few conformers have to be considered, reasonably accurate absolute as well as relative binding affinities can be computed routinely. We found no indications in S30L that slightly more flexible complexes exhibit larger errors but note, that extended π-systems seem to be somewhat problematic in the gas phase interaction part. The small residual deviations of 2-3 (typically 5-10% of ΔE) are impressive for large complexes with 200 atoms from a theoretical point of view. However, achieving "chemical accuracy" on an 1 kcal mol^{-1} level for ΔG_a seems to be extremely difficult and likely requires improvement of gas phase interaction energies, inclusion of anharmonic and dynamic effects as well as a much more accurate solvation treatment.

Acknowledgement

The authors thank A. Hansen for helpful discussions, G. Brandenburg for technical support regarding the H+ hydrogen bonding corrections for the PM6-D3 method, and A. Goeller (Bayer AG) for comparing our SMD values to those he obtained with Gaussian.

Part IV.

Synergy between theory and experiment: Application of quantum chemical methods to two specific problems

In Part II and III of this thesis the performance of electronic structure methods for noncovalent interactions and the achievements of a multilevel methodology for describing the thermochemistry of host-guest complexes have been presented. Part IV focuses on the application of quantum chemical methods to two specific problems in order to demonstrate the possible synergy between theory and experiment.

The first project is a collaboration with Prof. Gansäuer and his group who work on titanocene(III) catalyzed atom-economical reactions. One example of the investigated reactions is the intramolecular arylation of epoxides.[59] In this reaction, the titanocene(III) catalyst opens the epoxide, an intramolecular radical addition to the arene takes place, the electron is transferred back to the Ti(IV), and a proton transfer yields the desired product as well as the regenerated catalyst. The proposed catalytic cycle of the reaction, as well as mechanistic alternatives were studied computationally.[412] In agreement with synthetic and kinetic studies, the calculations confirmed the electron transfer as the turnover limiting step and several possible side reactions could be ruled out. During the experimental work on this and other reactions, it is found that the most efficient electron-deficient catalysts are thermally unstable. This can be circumvented by the addition of chloride sources. Thus, Chapter 8 focuses on the computational investigation of the supramolecular interaction of the titanocene with the hydrochloride additives, and the stability of the formed adducts.

The second project is a collaboration with Prof. Lützen and his group who work on the synthesis of functional supramolecular systems. Self-assembly processes are used to design metallo supramolecular coordination compounds. For example these are enantiomerically pure double- and triple-stranded helicates which can be obtained by combining chiral 1,1'-binaphthol based ligands and transition metal ions.[60–63] Similarly, paracyclophane based ligands can be used to create helicates.[65] The determination of the absolute configuration of 4,15-difunctionalized [2.2]paracyclophanes is the topic of Chapter 9. If a crystal structure cannot be obtained, one way to identify an enantiomer is by measuring an electronic circular dichrosim (ECD) spectrum and comparing it to a calculated one. As the standard time dependent (TD) DFT approach is computationally demanding, a recently published simplified TD-DFT (sTD-DFT) version is used.[413,414] ECD spectra are calculated for different difunctionalized paracyclophanes employing the hybrid functional BHLYP[97] and the range-separated functional CAM-B3LYP.[107]

As mentioned above, the functionalized paracyclophanes can be used to synthesize ligands which form double- and triple-stranded helicates with transition metal ions. One example, a triple-stranded helicate with two zink ions is shown in Figure 7.13. In general, three diastereomers are possible: (Δ,Δ)-, (Δ,Λ)-, or (Λ,Λ)-$[Zn_2L_3]^{4+}$. In ongoing work, the experimental spectrum is compared to the calculated ones of all three diastereomers to identify the diastereomer that was obtained in the self-assembly process. The

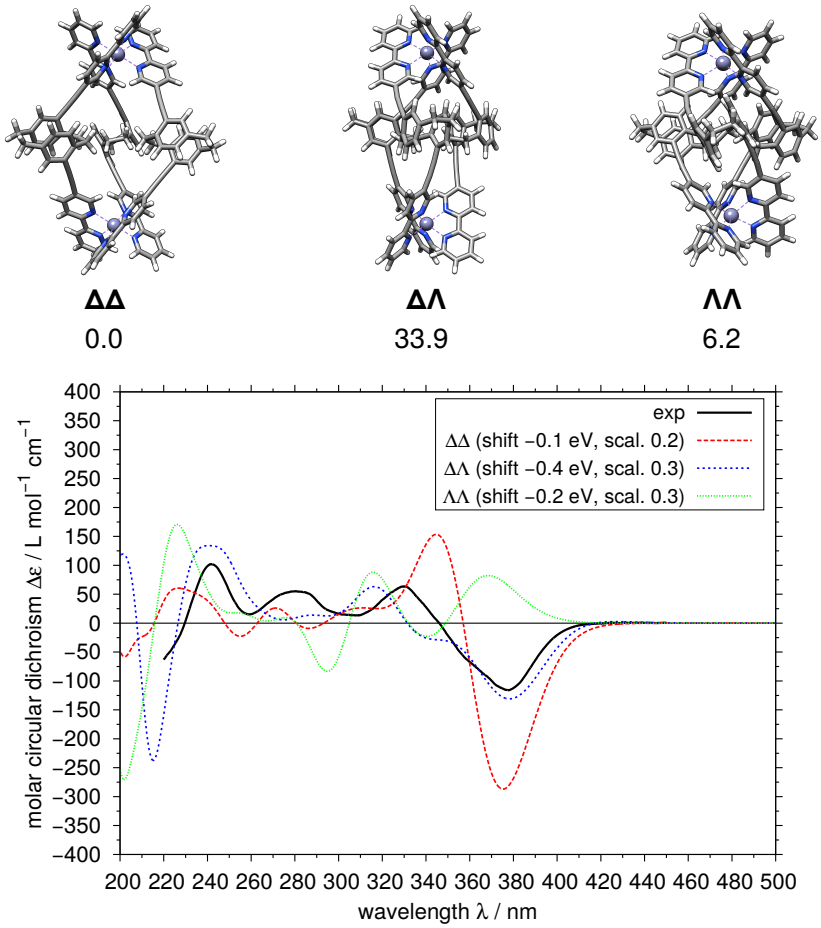


Figure 7.13.: Geometries (PBE-D3/def2-TZVP, cosmo=35.7) of three possible $[Zn_2L_3]^{4+}$ helicates and their corresponding calculated circular dichroism spectra (sTD-DFT CAM-B3LYP/def2-TZVP, cosmo=35.7) compared to experiment. Relative association free energies $\Delta\Delta G_a$ are given in kcal mol^{-1} on the PW6B95-D3/def2-QZVP' level of theory with PBE/def2-SVP thermal corrections and COSMO-RS 2012 solvation effects.

(Λ, Λ)-diastereomer can be ruled out but the general features of the calculated spectra for the (Δ, Δ)- and (Δ, Λ)-diastereomers are similar and the agreement with the experimental spectrum is equally good. Therefore, the relative association free energies are calculated on the PW6B95-D3/QZ level of theory with PBE/def2-SVP thermostatistical corrections and COSMO-RS 2012 solvation effects in order to identify the most stable helicate. (Δ, Δ)-$[Zn_2L_3]^{4+}$ is the most stable diastereomer and thus, most likely the helicate obtained in the self-assembly process.

8. Substituent Effects and Supramolecular Interactions of Titanocene(III) Chloride: Implications for Catalysis in Single Electron Steps

Andreas Gansäuer*, Christian Kube*, Kim Daasbjerg†, Rebecca Sure‡, Stefan Grimme‡, Godfred D. Fianu§, Dhandapani V. Sadasivam§, Robert A. Flowers§

Received 28th of November 2013, Published online 3rd of January 2014

Reprinted (adapted) with permission from
A. Gansäuer, C. Kube, K. Daasbjerg, R. Sure, S. Grimme, G. D. Fianu, D. V. Sadasivam, and R. A. Flowers *J. Am. Chem. Soc.* **2014**, *136*, 1663–1671.
 DOI: 10.1021/ja4121567

Own manuscript contribution:

- Performance of the calculations
- Interpretation of calculated data
- Co-writing the manuscript (theoretical part)

*Kekulé-Institute of Organic Chemistry and Biochemistry, University of Bonn, Gerhard-Domagk-Straße 1, 53121 Bonn, Germany
†Department of Chemistry, Aarhus University and Interdisciplinary Nanoscience Center, Langelandsgade 140, 8000 Aarhus C, Denmark
‡Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany
§Department of Chemistry, Lehigh University, Bethlehem, Pennsylvania 18015, United States

Abstract

The electrochemical properties of titanocene(III) complexes and their stability in THF in the presence and absence of chloride additives were studied by cyclic voltammetry (CV) and computational methods. The anodic peak potentials of the titanocenes can be decreased by as much as 0.47 V through the addition of an electron-withdrawing substituent (CO_2Me or CN) to the cyclopentadienyl ring when compared with Cp_2TiCl. For the first time, it is demonstrated that under the conditions of catalytic applications low-valent titanocenes can decompose by loss of the substituted ligand. The recently discovered effect of stabilizing titanocene(III) catalysts by chloride additives was analyzed by CV, kinetic, and computational studies. An unprecedented supramolecular interaction between $[(C_5H_4R)_2TiCl_2]$ and hydrochloride cations through reversible hydrogen bonding is proposed as a mechanism for the action of the additives. This study provides the critical information required for the rational design of titanocene-catalyzed reactions in single electron steps.

8.1. Introduction

The design of novel and efficient catalytic transformations is at the heart of chemistry.[415–417] Radical-based transformations offer attractive features such as ease of generation, high functional group tolerance, and ability to add to unsaturated functional groups.[418–420] In view of these advantages, it is surprising that their potential as key intermediates in catalytic atom-economical C–C bond formation is largely untapped.[421–423] Among the rare examples of such reactions are atom-transfer radical addition reactions, especially those catalyzed by Ru complexes,[424–426] the highly important Cu-catalyzed atom-transfer radical polymerizations,[427–429] and H_2-mediated Cr- or Co-catalyzed reductive cyclization of dienes.[430–432] We have introduced reagent-controlled examples[433–435] of atom-economical catalytic radical reactions with our titanocene-catalyzed tetrahydrofuran synthesis[55–58] and radical arylations of epoxides.[59] In all of the aforementioned processes, radical generation and trapping can be regarded as oxidative additions and reductive eliminations in single electron steps.[436] Thus, catalysis of atom-economical radical chemistry should be considered as catalysis in single electron steps and is therefore part of the framework of classical organometallic catalysis.[437,438] In line with this notion, the titanocene-catalyzed tetrahydrofuran synthesis and radical arylation are critically influenced by the electronic properties of the ligands. Moreover, in reactions with the most efficient electron-deficient catalysts, thermal stability of the catalyst was a serious issue. Addition of chloride sources resolved this issue and allowed a significant reduction of catalyst loading. Thus, the success of these transformations is due to an intricate interplay

of thermodynamic properties of the titanocene(III) reagents and their kinetic stability under highly specific reaction conditions. As a consequence, it is essential to unravel the interplay between catalyst stability and reactivity in order to provide a toolbox for the design of efficient reactions. Here we describe our efforts to achieve this goal with the aid of cyclic voltammetry (CV), kinetic, and computational studies. CV is ideally suited for the identification of the components of mixtures of redox-active compounds and the study of their properties and the kinetics of their reactions.[439–441] CV has been applied to resolve the composition and reactivity of parent Cp_2TiCl-derived reagents in a number of solvents and transformations.[442–446] Computational studies provide the relative stabilities of the complexes between Ti(III) catalysts and amine hydrochlorides, and kinetic studies show the impact of additives on catalyst reactivity.

8.2. Results and discussion

The CV investigations were divided into three parts. First, the electrochemical reduction of the titanocenes was studied. In this manner, the properties of $[(C_5H_4R)_2TiCl_2]^-$ could be studied, providing benchmarks for examining the impact of chloride additives. Second, the influence of ligand substitution on solutions of Zn-reduced $(C_5H_4R)_2TiCl_2$ was investigated. These solutions contain $(C_5H_4R)_2TiCl$, the active species in titanocene(III)-catalyzed reactions, and its dimer. Finally, since chloride sources are often employed in catalytic applications of Ti(III) reductants, the effect of chloride additives to solutions of Zn–$(C_5H_4R)_2TiCl_2$ was probed. In addition to the CV experiments, kinetic and computational studies aimed at understanding the nature of the interactions of these additives and catalysts were examined.

8.2.1. Substituent effects for electrochemically reduced titanocene(IV) dichlorides in THF

In general, the electrochemical reduction of Cp_2TiCl_2 (Figure 8.1, R = R' = H) can be described according to an E_qC_r reaction scheme as originally proposed by Laviron and co-workers.[447,448] The quasi-reversible electrochemical reduction of Cp_2TiCl_2 (E_q) is followed by a chemically reversible cleavage of chloride from $[Cp_2TiCl_2]^-$ (C_r).[443,449] Figure 8.1 also includes a chemically irreversible step (C_i), namely, loss of a cyclopentadienyl anion ligand rather than chloride upon electrochemical reduction, which has never been observed until this point. Previously, it was shown that the most characteristic feature of the cyclic voltammogram of Cp_2TiCl_2 at a low sweep rate ($\nu < 1$ V s^{-1}), is the presence of the $Cp_2TiCl_2/Cp_2TiCl_2^-$ wave in tetrahydrofuran (THF) (*vide infra*).[443] This is due to the fact that the second-order back association reaction between Cp_2TiCl and Cl^- in

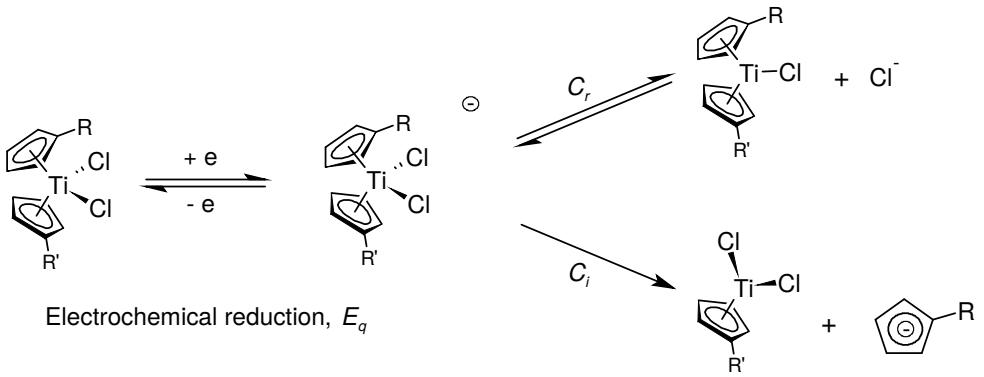


Figure 8.1.: E_qC_r and E_qC_i schemes for the electrochemical reduction of titanocene dichlorides.

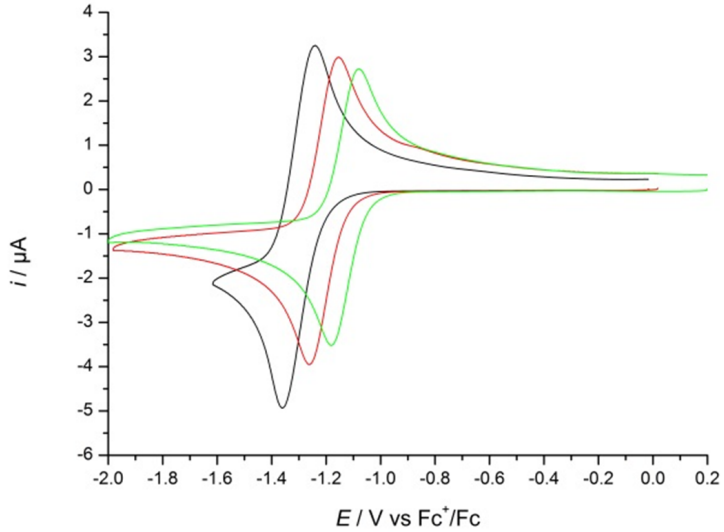


Figure 8.2.: Cyclic voltammograms of 2 mM Cp_2TiCl_2 (**4**) (—), $(C_5H_4Cl)CpTiCl_2$ (**5**) (—), and $(C_5H_4Cl)_2TiCl_2$ (**7**) (—) recorded at a glassy carbon disk electrode using $\nu = 0.1$ V s^{-1} in 0.2 M Bu_4NPF_6/THF.

the C_r step is fast, thus leaving $[Cp_2TiCl_2]^-$ as the only species detectable on the reverse sweep. In principle, at sufficiently high sweep rates a situation can be reached where the back association reaction is outrun before regeneration of $[Cp_2TiCl_2]^-$. Indeed, cyclic voltammograms recorded at $\nu = 20$ V s^{-1} show a small additional oxidation wave pertaining to the oxidation of Cp_2TiCl.[443] Ring-substituted titanocenes have only recently been employed in catalytic electron transfer applications.[59,436] It was assumed that ring substitution would strongly influence the redox behavior of the complexes. Until now only a few such examples have been investigated.[450,451] Therefore, in this study we decided to investigate various mono- and disubstituted titanocene complexes of the type $(C_5H_4R)(C_5H_4R')Ti(IV)Cl_2$ with R = *t*Bu, H, Cl, COOMe, or CN and R' = *t*Bu, H, Cl, or COOMe (see Figure 8.1).

Figure 8.2 shows cyclic voltammograms of Cp_2TiCl_2, $(C_5H_4Cl)CpTiCl_2$, and $(C_5H_4Cl)_2$-

$TiCl_2$ recorded at $\nu = 0.1$ V s^{-1} in 0.2 M Bu_4NPF_6/THF. As can be seen, the overall electrochemical behavior of the two substituted complexes follows that of Cp_2TiCl_2 in which the characteristic redox wave of $(C_5H_4R)(C_5H_4R')TiCl_2/[(C_5H_4R)(C_5H_4R')TiCl_2]^-$ is the only one observable at low ν. However, the position of the wave is shifted in a positive direction when chlorine is introduced as a substituent. Hence, CV constitutes a good method of choice for studying the properties of the anionic complexes $[(C_5H_4R)(C_5H_4R')TiCl_2]^-$.

Table 8.1 summarizes the effect of ligand substitution on the redox properties of the titanocene complexes. The potential of the reduction peak ($E_{p,c}$) varies from –1.44 V vs ferrocenium/ ferrocene (Fc^+/Fc) for the reduction of Kagan's complex (**1**) to –1.06 V vs Fc^+/Fc for the reduction of $(C_5H_4CN)CpTiCl_2$ in THF (recorded at $\nu = 0.1$ V s^{-1}); the pertinent anodic potentials ($E_{p,a1}$) were found to be 100–160 mV less negative. The standard potentials of the $(C_5H_4R)(C_5H_4R')TiCl_2/[(C_5H_4R)(C_5H_4R')TiCl_2]^-$ systems (E_1°) were determined by digital simulations and are included in the last column of Table 8.1.[452–454]

The order of the potentials is substituent-dependent, displaying a reasonable correlation with the Hammett substituent coefficient σ_p (σ_p = –0.20, 0, 0.22, 0.45, and 0.66 for the *t*Bu, H, Cl, COOMe, and CN substituents, respectively).[465] In addition, the substituent effect on the potential shift is seen to be almost additive for Cl (compounds **5** and **7**) with, on average, 75 mV per substituent and also for COOMe (compounds **6** and **8**) with, on average, 130 mV per substituent, as deduced from the E_1° values. In contrast, the *t*Bu substituent (complexes **2** and **3**) exerts much less than the expected effect ($\leq$ 20 mV) on the measured potentials. For Kagan's complex **1** (Figure 8.3),[455–457] a precatalyst in highly enantioselective electron transfer reactions,[466–471] the effect on E_1° is larger. This is due to destabilizing steric interactions between the two Ti-bound chlorides and the conformationally locked[456] and sterically congested cyclopentadienyl ligands.

In this respect, it is also interesting to note that for complexes **1**–**5** and **7** we found that $|i_{p,a1}/i_{p,c}| > 0.8$ at $\nu = 0.1$ V s^{-1}, where $i_{p,a1}$ is the anodic peak current of the first anodic wave and $i_{p,c}$ is the cathodic peak current. Such high values of $|i_{p,a1}/i_{p,c}|$ show that the chemical reactions present are quasi-reversible. As mentioned previously, by increasing ν

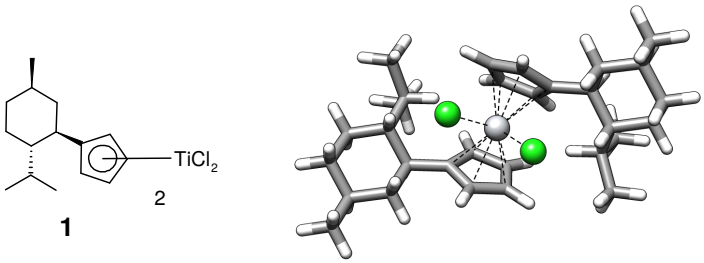


Figure 8.3.: Structure of Kagan's complex (**1**).

Table 8.1.: CV Data for the reduction of titanocene dichlorides **1**–**9** in THF in terms of peak potentials for the cathodic wave ($E_{p,c}$) and the two anodic waves ($E_{p,a1}$ and $E_{p,a2}$) along with the determined standard potential (E_1°)[a]

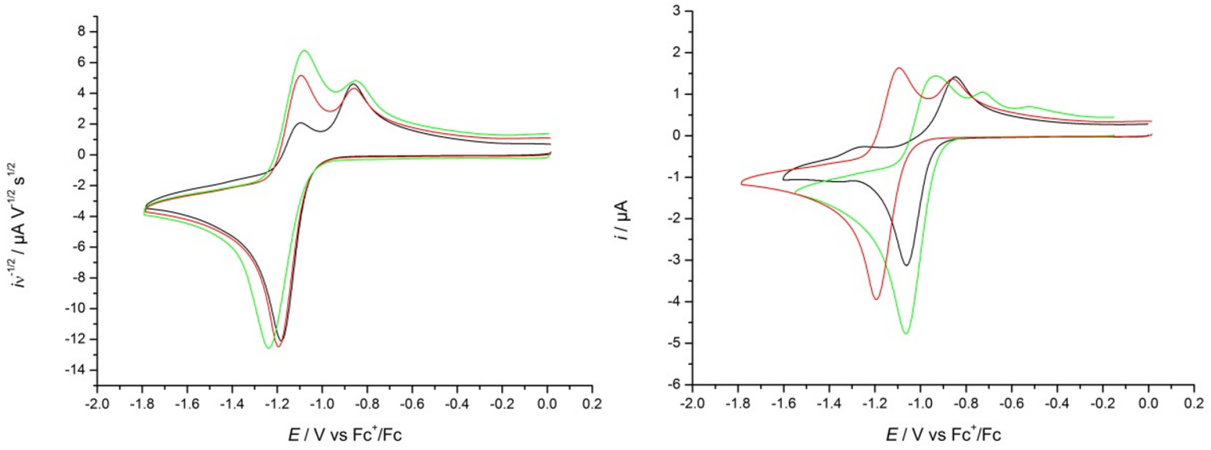
compound	$E_{p,c}$ [b]	$E_{p,a1}$ [b]	$E_{p,a2}$ [b]	E_1° [c]
Kagan's complex (**1**)[455–457]	-1.44	-1.32		-1.37
$(C_5H_4tBu)CpTiCl_2$ (**2**)[458]	-1.39	-1.27		-1.36
$(C_5H_4tBu)_2TiCl_2$ (**3**)[459]	-1.36	-1.25		-1.34
Cp_2TiCl_2 (**4**)	-1.36	-1.24		-1.27
$(C_5H_4Cl)CpTiCl_2$ (**5**)[460]	-1.26	-1.15		-1.20
$(C_5H_4COOMe)CpTiCl_2$ (**6**)[461]	-1.20	-1.09	-0.85	-1.15
$(C_5H_4Cl)_2TiCl_2$ (**7**)[460,462]	-1.18	-1.08		-1.12
$(C_5H_4COOMe)_2TiCl_2$ (**8**)[463]	-1.08	-0.92	-0.74	-1.01
$(C_5H_4CN)CpTiCl_2$ (**9**)[436]	-1.06		-0.85	-1.00

[a] All potentials are given in units of V vs Fc^+/Fc and can be converted to V vs SCE by adding 0.52 V.[443,464]
[b] Recorded at a glassy carbon disk electrode with $\nu = 0.1$ V s^{-1} in 0.2 M Bu_4NPF_6/THF.
[b] Determined by digital simulation (see Appendix E).

substantially it is possible to reach a situation where the back association reaction in the C_r step is outrun and can be disregarded. Experimentally, this is seen as a decrease in $|i_{p,a1}/i_{p,c}|$ and the appearance of a new peak at less negative potentials pertaining to the oxidation of $(C_5H_4R)(C_5H_4R')TiCl$. In a specific study of Cp_2TiCl itself, its oxidation wave recorded at $\nu = 0.1$ V s^{-1} appeared at about –0.8 V vs Fc^+/Fc.[443]

In contrast to the findings described above, the reduction of **6**, **8**, and textbf9 follows a different course as evidenced by CV. Figure 8.4(a) shows cyclic voltammograms recorded for $(C_5H_4COOMe)CpTiCl_2$ (**6**) at three different sweep rates. Besides an oxidation wave appearing at $E_{p,a1} = -1.09$ V vs Fc^+/Fc, a second wave at $E_{p,a2} = -0.85$ V is seen. Unambiguously, this shows that the generated anionic Ti(III)–species **6** is unstable and undergoes a chemical follow-up reaction.

Two observations preclude that the mechanism can be the "expected" E_qC_r mechanism. First, the addition of a chloride source ($Bu_4N^+Cl^-$) did not affect the voltammograms (see the Supporting Information), even though the C_r follow-up reaction should be greatly suppressed by the presence of Cl^-.[443] Second, the first anodic wave grows at the expense of the second wave (after background subtraction) as ν is enhanced from 0.05 to 0.5 V s^{-1}. This behavior is the opposite of what would be expected if the rate-controlling step had been the back association in a C_r step. Rather, this behavior would be in much better agreement with the existence of another rate-controlling chemical follow-up step in which the cyclopentadienyl anion, $[C_5H_4R]^-$, rather than chloride is lost irreversibly,[472] as



(a) (b)

Figure 8.4.: (a) Cyclic voltammograms of 2 mM $(C_5H_4COOMe)CpTiCl$ (**6**) recorded at a glassy carbon disk electrode using $\nu = 0.05$ (—), 0.1 (—), and 0.5 V s^{-1} (—) in 0.2 M Bu_4NPF_6/THF.
(b) Cyclic voltammograms of 2 mM $(C_5H_4CN)CpTiCl_2$ (**9**) (—), $(C_5H_4COOMe)CpTiCl_2$ (**6**) (—), and $(C_5H_4COOMe)_2TiCl_2$ (**8**) (—) recorded at a glassy carbon disk electrode using $\nu = 0.1$ V s^{-1} in 0.2 M Bu_4NPF_6/THF.

included in figure 8.1.

Figure 8.4 (b) further compares the CV behavior of **6**, **8**, and **9** for ν fixed at 0.1 V s^{-1}. Since the reductions of both **6** and **9** would be expected to proceed exclusively through loss of the cyclopentadienyl ligand possessing the electron-withdrawing group to achieve the highest possible stabilization of the released anion, this would in both cases result in the formation of $CpTiCl_2$ as the other product. This was indeed the case, as the peak potential of the second oxidation wave was the same for both complexes ($E_{p,a2} = -0.85$ V vs Fc$^+$/Fc) and identical to that measured for the oxidation peak of $CpTiCl_2$ generated by electrochemical reduction of $CpTiCl_3$ (see the Supporting Information).

Loss of $[C_5H_4CO_2Me]^-$ was also observed after reduction of **8**. In this case, the second anodic wave has a lower potential ($E_{p,a2} = 0.74$ V vs Fc$^+$/Fc) than for **6** because of the formation of $(C_5H_4CO_2Me)TiCl_2$, in agreement with the E_qC_i mechanism in Figure 8.1. A third anodic wave (–0.52 V vs Fc$^+$/Fc) was also observed. It seems reasonable to assume that this is due to the formation of $TiCl_3$ through ligand loss from $(C_5H_4CO_2Me)TiCl_2$, even though this was not further investigated.

Hence, for **6**, **8**, and **9** we propose the E_qC_i mechanism in Figure 8.1, where the quasi-reversible electrode process is followed by an irreversible loss of the cyclopentadienyl anion. The time frame for the follow-up reactions in the case of **6** and **8** can be easily monitored in CV by varying the sweep rate as illustrated in Figure 8.4 (a). From such studies of the voltammetric response recorded as a function of sweep rate, the dissociation rate constant

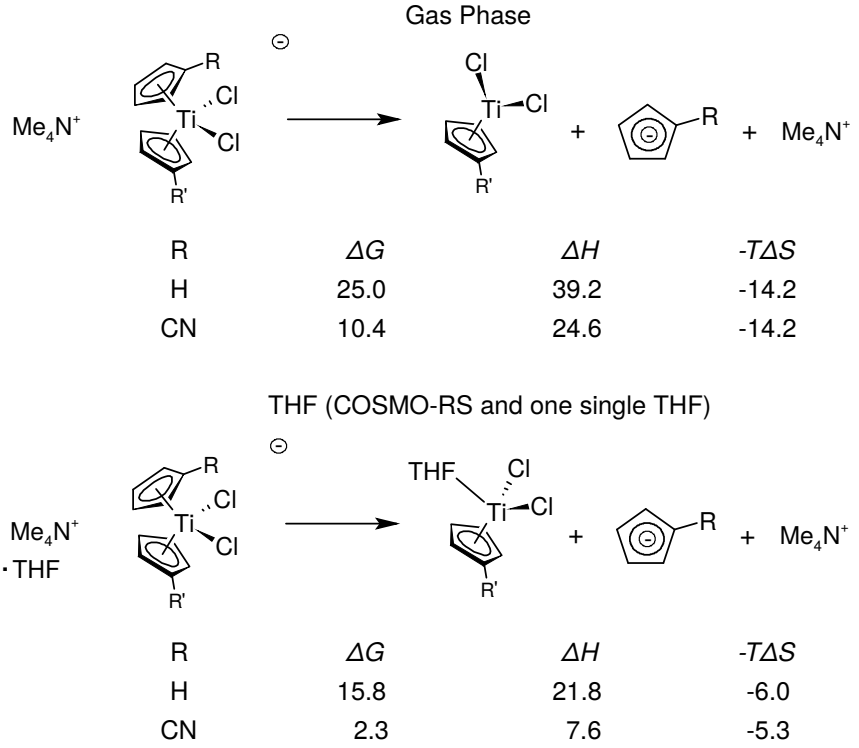
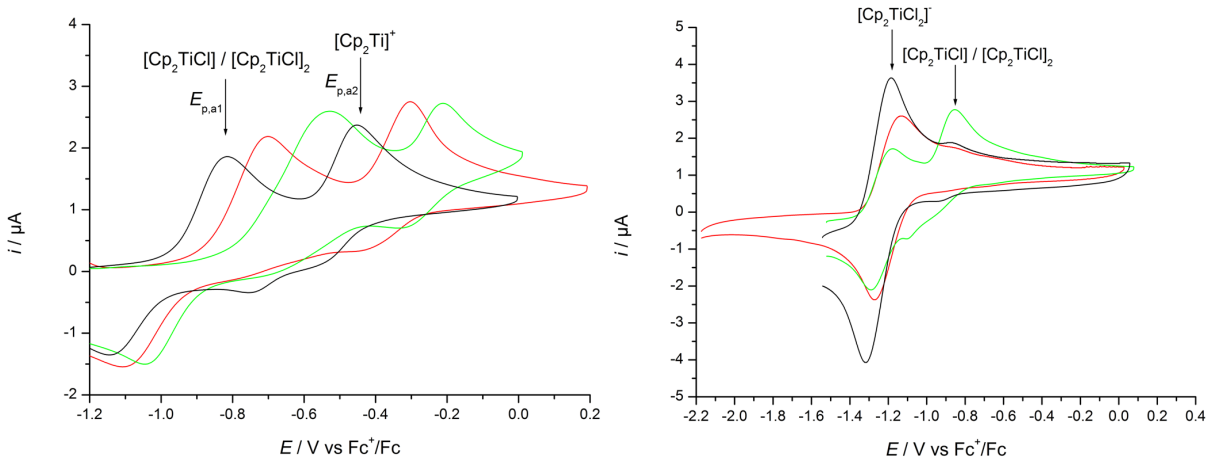


Figure 8.5.: Model systems for the thermodynamic features of cyclopentadienyl ligand loss (R = H, CN) at 298.15 K (All Energies in kcal mol^{-1}.)

(k_{dis}) may be extracted using digital simulations (see the Supporting Information).[452–454] Using this approach provided k_{dis} values of 0.11 and 10 s^{-1} for the anions of **6** and **9**, respectively. Thus, the more electron-deficient cyano-substituted cyclopentadienyl ligand[473] dissociates faster than the corresponding ester-substituted one as a consequence of the enhanced ability of the cyano group to stabilize the negative charge in the cleaved $[C_5H_4R]^-$ anion.

The thermodynamic features of the loss of the cyclopentadienyl ligands were studied computationally[261] for the model systems with Bu_4N^+ replaced by Me_4N^+ shown in Figure 8.5 in the gas phase (TPSS-D3/def2-TZVP)[36,37,93,133,134,237] and in solution with the COSMO continuum solvation model for real solvents (RS) (TPSS-D3-COSMO-RS[41,42]/def2-TZVP//TPSS-D3/def2-TZVP). In the gas phase, both dissociation reactions are unfavorable (R = H, ΔG = +25.0 kcal mol^{-1}; R = CN, ΔG = +10.4 kcal mol^{-1}). The difference between the ΔG values is caused by the difference in ΔH and is a reflection of the weaker binding of $[C_5H_4CN]^-$. The effect of solvation was studied by employing two modifications. First, COSMO-RS was included to simulate the effect of bulk solvent. Second, one molecule of THF was introduced in order to understand its interactions with NMe_4^+ of the substrate and the Lewis acidic product $CpTiCl_2$ on a molecular level. The results clearly show that dissociation of the cyclopentadienyl ligands is more advantageous in solution than in the gas phase. This is mainly due to a more favorable ΔH that is



(a) (b)

Figure 8.6.: (a) Cyclic voltammograms of 2 mM Zn−Cp_2TiCl_2 (**4**) (—), Zn−$(C_5H_4Cl)CpTiCl_2$ (**5**) (—), and Zn−$(C_5H_4Cl)_2TiCl_2$ (**7**) (—) recorded at a glassy carbon disk electrode using $\nu = 0.1$ V s^{-1} in 0.2 M Bu_4NPF_6/THF. (b) Cyclic voltammograms of 2 mM Zn−Cp_2TiCl_2 containing 2 eq $Hex_3N{\cdot}HCl$ (—), 2 eq (Coll · HCl (—), and 2 eq Py · HCl (—) as additive. Recordings were performed at a glassy carbon disk electrode using $\nu = 0.1$ V s^{-1} in 0.2 M Bu_4NPF_6/THF.

caused by coordination of THF to the Lewis acid $CpTiCl_2$. The differences between the ΔG values (14.6 and –113.5 kcal mol^{-1}, respectively) in the gas phase and in solution are almost identical. This suggests that binding of $[C_5H_5]^-$ to Ti is more favorable than binding of $[C_5H_4CN]^-$ by about 14 kcal mol^{-1}.

8.2.2. Zn–titanocene(IV) dichlorides in THF

In stoichiometric and catalytic applications of titanocene(III) complexes, the active reagent is practically always generated by reduction of the titanocene dichlorides with Mn or Zn dust.[433–435] The understanding of the influence of ligand substitution on the redox properties of the titanocene species is therefore of high practical relevance and was investigated next. Previous studies of solutions of metal–Cp_2TiCl_2 in THF by CV have shown that these solutions consist of the Cp_2TiCl monomer and the $(Cp_2TiCl)_2$ dimer.[442–446] The Cp_2Ti^+ cation is also observed in the voltammograms but is generated only in the diffusion layer during sweeping from the oxidized products of Cp_2TiCl and $(Cp_2TiCl)_2$.

Figure 8.6 (a) shows cyclic voltammograms recorded for Zn–**4**, Zn–**5**, and Zn–**7**. In line with the previous assessment, the first oxidation wave is assigned to the monomer/dimer couple and the second one to the cation. Furthermore, it was noted that the second wave diminishes with increasing ν, thus confirming that the cation indeed is formed in follow-up reactions induced by the sweeping and thus is not present in any substantial amount for

Table 8.2.: CV Data for the oxidations of Zn–**1** through Zn–**9** in terms of peak potentials for the anodic Waves, $E_{p,a1}$ and $E_{p,a2}$.[a]

compound	$E^b_{p,a1}$	$E^b_{p,a2}$
Kagan's complex (**1**)	-0.82	-0.47
$(C_5H_4tBu)CpTiCl_2$ (**2**)	-0.84	-0.44
$(C_5H_4tBu)_2TiCl_2$ (**3**)	-0.74	-0.45
Cp_2TiCl_2 (**4**)	-0.83	-0.42
$(C_5H_4Cl)CpTiCl_2$ (**5**)	-0.71	-0.32
$(C_5H_4COOMe)CpTiCl_2$ (**6**)	-0.58	-0.23
$(C_5H_4Cl)_2TiCl_2$ (**7**)	-0.54	-0.2
$(C_5H_4COOMe)_2TiCl_2$ (**8**)	-0.43	-0.1
$(C_5H_4CN)CpTiCl_2$ (**9**)	-0.35	-0.2

[a] Potentials were recorded at a glassy carbon disk electrode with $\nu = 0.1$ V s^{-1} in 0.2 M Bu_4NPF_6/THF; the values are given in V vs Fc^+/Fc and can be converted to V vs SCE by adding 0.52 V.[443,464]

any of the compounds **1**–**9**.

To determine the exact monomer/dimer composition, a detailed analysis of the first wave as a function of ν and concentration involving digital simulation would be required. Since this is not the purpose of the present investigation, the focus is rather addressed at describing the redox properties of the Ti(III) species by determining the peak potentials. At the same time, it would be pertinent to elucidate whether the cyclopentadienyl ligand is lost in the chemical reduction of **6**, **8**, and **9** as was seen for the electrochemical reduction.

Table 8.2 summarizes the CV data in terms of the anodic peak potentials $E_{p,a1}$ and $E_{p,a2}$ obtained for Zn–**1** through Zn–**9** in THF. Notably, the potentials of Zn–**1**[455–457] and Zn–**4** are very similar. The success of Zn–**1**[455–457] in enantioselective and regiodivergent epoxide opening reactions[466–471,474] is therefore due to steric and not electronic effects. This should also be the case in the recently described enantioselective cyclizations of ketyl radicals catalyzed by Brintzinger's complex.[475]

The introduction of electron-withdrawing substituents has a dramatic effect on $E_{p,a1}$, which changes from –0.84 to –0.35 V vs Fc^+/Fc. For the cationic titanocenes, the effect of ligand substitution is less dramatic, with the $E_{p,a2}$ values ranging from –0.47 to –0.20 V vs Fc^+/Fc.

A very important finding is that for the Zn-reduced solutions the loss of electron-deficient cyclopentadienyl ligands, which is a major decomposition pathway for the electrochemically reduced complexes, was never observed. This is likely due to the efficient abstraction of chloride from $[(C_5H_4R)CpTiCl_2]^-$ by Zn_2^+ ions in THF. The same is true for Mn_2^+ ions.

8.2.3. Zn–titanocene(IV) dichlorides in THF in the presence of chloride additives

The increased use of Cp_2TiCl as a reagent is a consequence of the development of conditions catalytic in the reagent. In most of these reactions, turnover is mediated by protonation or silylation of Ti–O bonds through the addition of stoichiometric amounts of either Coll · HCl or Coll · Me_3SiCl (Coll = 2,4,6-trimethylpyridine).[433–435] In order to understand the impact of added chloride ions on the coordination sphere and the redox properties of titanocene(III) chlorides, we studied the CV behavior of Zn–Cp_2TiCl_2 in the presence of various chloride donors, namely, $Bu_4N^+Cl^-$, Hex_3N · HCl, LiCl, Py · HCl (Py = pyridine), and Coll · HCl.

Figure 8.6 (b) shows the cyclic voltammograms recorded for the selected chloride donors at a glassy carbon disk electrode with $\nu = 0.1$ V s^{-1} in 0.2 M Bu_4NPF_6/THF. A common effect of all of these additives is that they make the oxidation wave of $[Cp_2TiCl_2]^-$ appear at low sweep rates. In fact, in the presence of $Bu_4N^+Cl^-$, LiCl, and to a large extent also Py · HCl, the voltammograms recorded are by and large identical to those of electrochemically reduced Cp_2TiCl_2 in 0.2 M $Bu_4N^+PF_6^-$/THF, as essentially no other wave but that for $[Cp_2TiCl_2]^-$ is detectable. This implies that the association of chloride to either Cp_2TiCl or $(Cp2\,TiCl)_2$ cannot be outrun in the presence of these additives, at least for the sweep rates employed (Figure 8.1).

The same is true to some extent for Hex_3N · HCl, although the $Cp_2TiCl_/(Cp_2TiCl)_2$ oxidation wave becomes dominant in the high sweep rate range. In general, Coll · HCl leads to the formation of a distinctly lower amount of $[Cp_2TiCl_2]^-$ in the CV compared with Py · HCl, and interestingly, the peak current ratio of the two oxidation waves was found to be constant over the range of sweep rates from 0.05 to 20 V s^{-1} .

This tuning of the kinetics of the formation of $[Cp_2TiCl_2]^-$ and its concentration through additives is to the best of our knowledge unprecedented. Since Coll · HCl is distinctly less soluble in THF than Hex3 N · HCl and Py · HCl, a straightforward explanation for this behavior is the smaller concentration of Coll · HCl in THF.

However, this analysis does not take into account the stability of the hydrochloride adducts, which should not be a function of the concentration of the hydrochloride. This issue was further investigated by computational means after the study of the kinetic effects of adduct formation.

In order to further understand the effect of ligand substitution on the interaction of the titanocenes with additives experimentally, we studied the voltammetry of Zn–$(C_5H_4tBu)_2TiCl_2$ and Zn–$(C_5H_4Cl)_2TiCl_2$ in the presence of Coll · HCl. Compared with Cp_2TiCl_2, the experiments at higher sweep rates show a more rapid decline in the intensity of the oxidation wave of $[(C_5H_4tBu)_2TiCl_2]^-$, and therefore, the complexation of

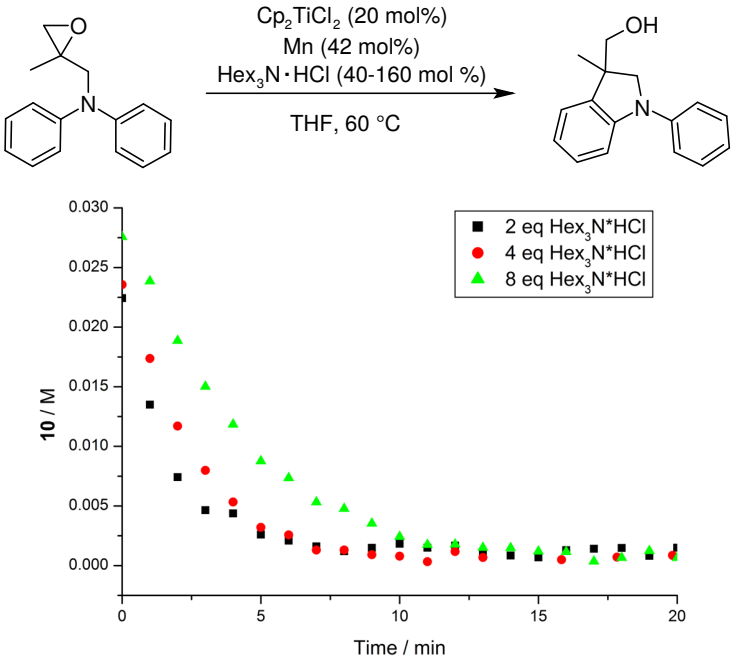


Figure 8.7.: Effect of $Hex_3N \cdot HCl$ on the rate of the radical arylation of **10**. For further experimental details see Appendix E.

Cl^- is slower for the bulky $(C_5H_4tBu)_2TiCl$ than for Cp_2TiCl. In agreement with intuition, a higher proportion of $[(C_5H_4Cl)_2TiCl_2]^-$ was observed for the more Lewis acidic $(C_5H_4Cl)_2TiCl$ over all sweep rates.

The formation of $[(C_5H_4R)_2TiCl_2]^-$ reduces the concentration of Cp_2TiCl, the active species in titanocene-catalyzed reactions. Therefore, there should be an inverse kinetic order of chloride concentration on rate caused by addition of the hydrochloride salts. This was probed for the Cp_2TiCl-catalyzed radical arylation of **10** (Figure 8.7).[59] In these experiments $Hex_3N \cdot HCl$ was chosen as the chloride source rather than $Coll \cdot HCl$ because of the limited solubility of the latter in THF.

The observed rate constants for the catalytic reaction were measured at constant initial substrate and catalyst concentrations and three concentrations of $Hex_3N \cdot HCl$ by following the decay of epoxide **10**. The decays (Figure 8.7) fit well to a single exponential. The k_{obs} values from the decays are given in Table 8.3.

With higher amounts of $Hex_3N \cdot HCl$, lower values of k_{obs} were observed. This finding is consistent with a reduction of the Cp_2TiCl concentration through the formation of $[(C_5H_4R)_2TiCl_2]^- \cdot Hex_3NH^+$. Moreover, the radical arylation of **10** proceeds faster in the presence of $Coll \cdot HCl$ than in the presence of $Hex_3N \cdot HCl$. This is consistent with the CV data showing a lower amount of adduct formation.

Table 8.3.: Observed rate constants for the arylation of **10** at different concentrations of $Hex_3N \cdot HCl$.

equiv of $Hex_3N \cdot HCl$[a]	k_{obs} / min^{-1}
2	0.57 ± 0.04
4	0.39 ± 0.01
8	0.24 ± 0.01

[a] With respect to catalyst concentration.

8.2.4. Computational study of the hydrochloride adducts

To evaluate our hypothesis of hydrogen bonding and to understand the nature of the interactions of the hydrochlorides with titanocene(III) complexes on a molecular level, a computational study was performed using the TURBOMOLE 6.4 program package.[261] All of the DFT structures were fully optimized at the TPSS-D3/def2-TZVP level including the COSMO[40] model. Final reaction free energies were obtained via single-point calculations on the PW6B9535[101]-D3/def2-QZVP[237] level in the gas phase and applying the COSMO-RS model to include solvation.

For all of the titanocenes, the $Et_3NH \cdot HCl$ adducts are the most stable ones (Table 8.4). The $Coll \cdot HCl$ adducts are enthalpically more favorable than the $Py \cdot HCl$ adducts and also thermodynamically more stable except for the case of $[(C_5H_4Cl)_2TiCl_2]^- \cdot PyH^+$. These results can be explained by a modulation of chloride binding through hydrogen bonding. The stronger acid $Py \cdot HCl$ will interact more strongly with the basic chloride ligands than the weaker acids $Coll \cdot HCl$ and $Et_3NH \cdot HCl$. This is in agreement with the calculated bond lengths for H–Cl(1) and Ti–Cl(1) shown in Table 8.4. The shortest of the "short" Ti–Cl bonds is found in $[Cp_2TiCl_2]^- \cdot Et3\,NH^+$ (2.50 Å) and the longest in $[Cp_2TiCl_2]^- \cdot PyH^+$ (2.53 Å). The notion that chloride binding makes the most important contribution to ΔH is also supported by the observation that the most Lewis acidic titanocene, $(C_5H_5Cl)_2TiCl$, forms the most stable adducts. The contributions of entropy are less relevant and are caused by differences in the entropy of solvation.

Analysis of the adduct structures (Table 8.5) also supports the idea that supramolecular interactions modulate the Ti–Cl bonding. In all of the structures, hydrogen bonding between the N–H and the chlorides is observed. The $Coll \cdot HCl$ adduct **12** and $Py \cdot HCl$ adduct **11** are distinguished by the orientations of the respective arenes. While in **11** the arene is almost in the plane containing both chlorides and Ti, in **12** the arene is nearly perpendicular to this plane (Figure 8.8). This is consistent with unfavorable steric interactions between the methyl groups at the 2- and 6-positions of collidine and the two chlorides that disfavor the "in-plane" binding. As a consequence, the hydrogen bonding

Table 8.4.: Computed free energies (PW6B95-D3-COSMO-RS/def2-QZVP//TPSS-D3-COSMO/def2-TZVP) of formation of hydrochloride adducts in THF at 298.15 K. All energies are given in kcal mol^{-1}.

[Ti]	additive	ΔH	$-T\Delta S$	ΔG
Cp_2TiCl	$Py \cdot HCl$	-4.9	4	-1.0
Cp_2TiCl	$Coll \cdot HCl$	-9.6	6.9	-2.6
Cp_2TiCl	$Et_3N \cdot HCl$	-15.2	6.6	-8.6
$(C_5H_4Cl)_2TiCl$	$Py \cdot HCl$	-8	3.8	-4.20
$(C_5H_4Cl)_2TiCl$	$Coll \cdot HCl$	-10.4	6.6	-3.7
$(C_5H_4Cl)_2TiCl$	$Et_3N \cdot HCl$	-17.2	6.9	-10.3
$(C_5H_4tBu)_2TiCl$	$Py \cdot HCl$	0	4.5	4.50
$(C_5H_4tBu)_2TiCl$	$Coll \cdot HCl$	-5.4	8.1	2.7
$(C_5H_4tBu)_2TiCl$	$Et_3N \cdot HCl$	-9.2	8.8	-0.4

in **11** is distinctly different from that in **12**. In **11** the two N–H–Cl hydrogen bonds have the same length, and the arrangement is symmetrical. In **12** the two N–H–Cl hydrogen bond lengths are significantly different. In **13**, the hydrogen bonding pattern is similar to that **12**. Therefore, Et_3NH^+ is best regarded as a cation with a steric bulk similar to $CollH^+$. This notion is further corroborated for the adducts of $(C_5H_4Cl)_2TiCl$ and $(C_5H_4tBu)_2TiCl$ (see Appendix E for details).

The slightly less favored formation of $[(C_5H_4Cl)_2TiCl_2]^- \cdot CollH^+$ compared with $[(C_5H_4Cl)_2TiCl_2]^- \cdot PyH^+$ is a consequence of "out-of-plane" binding, which results in an unfavorable interaction between one of the methyl groups attached to the arene and the Cl substituent of a Cp ligand in $[(C_5H_4Cl)_2TiCl_2]^- \cdot CollH^+$ (see Appendix E).

Thus, our CV, kinetic, and computational studies clearly highlight that the addition of hydrochloride additives has a profound and unprecedented influence on the composition of metal-reduced solutions of Cp_2TiCl_2 and its substituted derivatives. Adducts are formed that consist of hydrogen-bonded tight ion pairs of $[Cp_2TiCl_2]$ and the ammonium ion. The stability and rate of formation of the adduct can be finetuned by the steric bulk and

Table 8.5.: Selected Structural Data for **11**, **12**, and **13**. The dihedral angle is measured between the Cl–Ti–Cl and C(2)–N–C(6) planes

H–Cl(1) / Å	H–Cl(2) / Å	Ti–Cl(1) / Å	Ti–Cl(2) / Å	dihedral angle / °
2.34	2.34	2.53	2.53	-14.2
2.07	2.73	2.51	2.55	79.9
2.06	3.24	2.5	2.55	NA

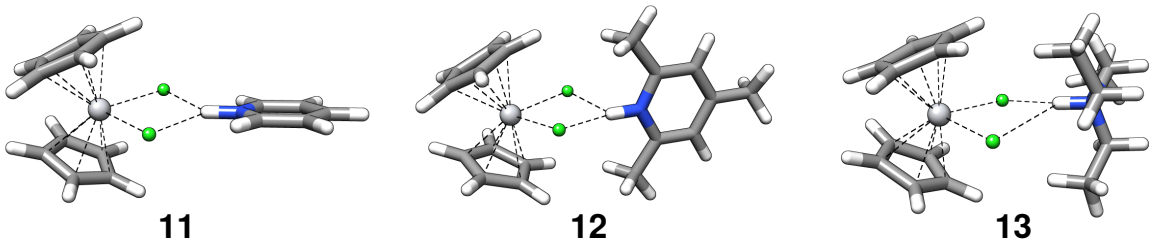


Figure 8.8.: TPSS-D3-COSMO/def2-TZVP-optimized structures of $[Cp_2TiCl_2]^- \cdot PyH^+$ (**11**), $[Cp_2TiCl_2]^- \cdot CollH^+$ (**12**), and $[Cp_2TiCl_2]^- \cdot Et_3NH^+$ (**13**).

acidity of the additive's cation. The consequences of adduct formation on the performance of these reagents in catalysis will be discussed next.

8.2.5. Implications of adduct formation for catalysis

The $Cp_2TiCl/(Cp_2TiCl)_2$ couple was introduced by Nugent and RajanBabu as a very mild and chemoselective stoichiometric reagent for the reductive opening of epoxides.[476–479] In seminal contributions it was demonstrated that the epoxide-derived radicals could be employed in classical radical reactions such as 5-exo cyclizations, additions to acrylates, and radical reduction via hydrogen atom transfer.

Later, catalytic conditions employing additives such as $Coll \cdot HCl$ to mediate turnover were developed.[480–482] Even though under these circumstances the concentration of the active species Cp_2TiCl is further depleted by chloride binding, the catalytic conditions are superior for the realization of kinetically difficult radical processes. Examples include epoxypolyene cyclizations featuring slow 6-endo, 7-endo, and transannular cyclizations,[483–489] 4-exo cyclizations,[490–494] tandem processes combining cyclizations and intermolecular additions,[495,496] and atom-economical tetrahydrofuran syntheses.[55–58]

In all of these processes, the trapping of intermediate radicals by Cp_2TiCl is an undesired intermolecular side reaction. Reduction of the Cp_2TiCl concentration through reversible $[Cp_2TiCl_2]^-$ formation efficiently suppresses radical trapping and therefore increases the radical lifetime. Even better results can be obtained with electron-deficient titanocenes,[59,436] which form the hydrochloride adducts more readily and constitute less efficient reductants. These effects can even be exploited under stoichiometric conditions, as demonstrated in a recent synthesis of (–)-maoecrystal Z.[497,498]

Another beneficial aspect of adduct formation has been described recently.[55–58] After addition of hydrochlorides, Cp_2TiCl- catalyzed reactions can be run at high temperatures without catalyst decomposition. Therefore, the hydrochloride adducts are thermally more stable than Cp_2TiCl and its dimer.

8.3. Conclusion

We have investigated the composition and properties of solutions of electrochemically reduced $(C_5H_4R)_2TiCl_2$, Zn–$(C_5H_4R)_2TiCl_2$, and Zn–$(C_5H_4R)_2TiCl_2$ in the presence of chloride additives by cyclic voltammetry, kinetic studies, and DFT calculations. Through this combined approach, the redox properties of representative $(C_5H_4R)_2TiCl$ complexes, their dimers, and anionic chloride adducts $[(C_5H_4R)_2TiCl_2]^-$ were determined. The stability of the electrochemically generated complexes depends on the substituents of the cyclopentadienyl ligands. With –CO_2Me- or –CN-containing ligands, $[Cp(C_5H_4R)TiCl_2]^-$ decomposes through loss of $[C_5H_4R]^-$.

In the presence of organoammonium chlorides, ammonium-adducts of $[(C_5H_4R)_2TiCl_2]$ are generated from Zn–$(C_5H_4R)_2TiCl_2$. The stabilities of these adducts and the rates of their formation are determined by the extent of hydrogen bonding between the catalyst and the ammonium cation. The degree of adduct formation can also be controlled by the solubility of the hydrochloride. The fine-tuning of the supramolecular interactions provides a novel platform for the design of more efficient and sustainable titanocene catalysts and titanocene-catalyzed processes.

8.4. Computational details

All quantum chemical calculations on the study of the loss of a cyclopentadienyl ligand as well as the formation of hydrochloride adducts have been performed with the TURBOMOLE 6.4 program package.[261] The geometry optimizations were performed on the DFT level using the TPSS density functional[93] together with the polarized triple-zeta Gaussian AO basis set def2-TZVP.[237] This choice avoids major BSSE effects without employing counter-poise corrections. Further, for the hydrochloride adduct formation reaction the cosmo continuum solvation model (COSMO)[40] was included during the optimizations. As the reactions were performed in THF the dielectric constant was set to 7.4. For all DFT calculations the resolution-of-identity (RI) approximation for the Coulomb integrals[74] with matching default auxiliary basis sets[238] was applied. The numerical quadrature grid *m4* was employed for integration of the exchange-correlation contribution.[239] For all geometry optimizations as well as single-point calculations, the D3 dispersion correction scheme[36] applying Becke-Johnson (BJ) damping[37,133,134] was used. The final level used for geometry optimization is dubbed TPSS-D3/def2-TZVP in the following. For a detailed description of the dispersion correction, that is of great importance in studies of large molecules, including many illustrative examples see Ref. [11], for recent chemical applications of this method see e.g. Ref. [12]. We discuss in the manuscript enthalpies H(298) and free energies G(298) that are obtained by a rigid-rotor,

harmonic vibrational statistical treatment. For the study of the loss of cyclopentadienyl ligands the vibrational frequencies are computed at the HF-3c[205] level (minimal basis set Hartree–Fock with three corrections for dispersion, basis set superposition error and short-range basis incompleteness) with a scaling factor of 0.86. Low-lying vibrational modes (below 100 cm^{-1}) are treated by a special rigid-rotor approximation in order to avoid numerical artifacts in the entropy calculations.[46] These gas phase calculations are also used to characterize the stationary points as minima or transition states, respectively, and do not include explicit solvent molecules. The corresponding thermo-statistical corrections are also used to correct the explicitly solvated results from energy to enthalpy or free energy. The methods employed here have been used recently to compute in an *ab initio* manner free energies of association for typical supramolecular complexes in solution (see below) with an unprecedented accuracy of about 1–2 kcal mol^{-1}.[46] For the recent computation of transition-metal thermochemistry in solution see Ref.[44]. For the hydrochloride adduct formation reactions the computations of the harmonic vibrational frequencies were performed numerically due to the inclusion of the COSMO model using the TURBOMOLE module `numforce`. Thermal corrections from energy to free energy were obtained as described for the other study before. The vibrational frequencies were used unscaled. Solvent effects on the thermochemical properties have been obtained by the COSMO-RS model[41,42] that was used as implemented in COSMOtherm[401] to obtain all solvation free energies. Single-point calculations employing the default BP86[91]/def-TZVP[266] level of theory were performed on the optimized cosmo geometries. Solvation contributions to free energies at 298 K for THF solution are computed from the TPSS-D3 gas phase structures and added to the TPSS-D3 gas phase free energies values. For the study of the loss of a cyclopentadienyl ligand test calculations for the functional dependence of the gas phase reaction energies have been performed with the PBE[89,90]-D3 (GGA) and PBE0[98]-D3 (hybrid) approximations but only small differences to the reported TPSS-D3 results of about 1 kcal mol^{-1} have been found. The employed def2-TZVP basis set usually yields results to within 2–3 kcal mol^{-1} of the basis set limit for such reactions. The COSMO-RS solvent correction for the here considered reaction is 8–9 kcal mol^{-1}. With a conservative error estimate of 10-20 % for this contribution one arrives at a final estimate for accuracy of the computed Δ value of $\pm$4–5 kcal mol^{-1}. For the formation of hydrochloride adducts, single-point energies for the hydrochloride adduct formation were obtained in the gas phase on the PW6B95[101] and B3LYP[91,92,96] levels together with the extended quadruple-zeta basis set def2-QZVP.[237] The difference in the gas phase reaction energy between those two functional amounts to 2–3 kcal mol^{-1}. The larger def2-QZVP basis has been used in this part because it yields an even smaller BSSE than the def2-TZVP basis which is important for these hydrogen binding associations.

Acknowledgement

We gratefully acknowledge generous support by SFB 813 ("Chemistry at Spin Centers") and the National Science Foundation (CHE-1123815).

9. Synthesis, Chiral Resolution, and Absolute Configuration of Dissymmetric 4,15-Difunctionalized [2.2]Paracyclophanes

Georg Meyer-Eppler*, Rebecca Sure†, Andreas Schneider‡, Gregor Schnakenburg‡, Stefan Grimme†, and Arne Lützen*

Received 30th of May 2014, Published online 27th of June 2014

Reprinted (adapted) with permission from
G. Meyer-Eppler, R. Sure, A. Schneider, G. Schnakenburg, S. Grimme, and A. Lützen *J. Org. Chem.* **2014**, *79*, 6679–6687.
 DOI: 10.1021/jo501212t

Own manuscript contribution:

- Performance of the calculations
- Interpretation of calculated spectra
- Co-writing the manuscript (theoretical part)

*Kekulé-Institute of Organic Chemistry and Biochemistry, University of Bonn, Gerhard-Domagk-Straße 1, 53121 Bonn, Germany
†Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany
‡Institute of Inorganic Chemistry, University of Bonn, Gerhard-Domagk-Straße 1, 53121 Bonn, Germany

Abstract

Despite the fact that functionalized planar chiral [2.2]- paracyclophanes have received a lot of attention, the chemistry of pseudo- meta 4,15-distubstituted [2.2]paracyclophanes is largely unexplored. This is mainly due to the fact that the 4,5-dibromo-functionalized [2.2]- paracyclophane is much less prone to halogen-metal exchange reactions than its constitutional pseudo-*ortho* or pseudo-*para* isomers. Here, we give an account of an efficient protocol to achieve this, which allows the synthesis of a broad variety of 4,15-disubstituted [2.2]paracyclophanes. Furthermore, we were able to resolve several of the racemic compounds via chiral HPLC and assign the absolute configurations of the isolated enantiomers by X-ray diffraction and/or by the comparison of calculated and measured CD-spectra.

9.1. Introduction

Although known for 65 years now [2.2]paracyclophane (**1**) is still far from its retirement age as it still offers lots of opportunities and challenges. The archetype of layered compounds is still fascinating chemists around the world due to its special physical and chemical properties.[499–502] In fact, it was only very recently, e.g., that the slightly twisted arrangement of the two layered aromatic rings could be proven experimentally.[503] This chiral D_2 symmetric structure, which represents a challenging test case for approximate density functional theory (DFT), has been predicted by high level quantum chemical calculations already 10 years ago.[504] In most cases, substitution of the aromatic rings results in the formation of planar chiral compounds even when the twist in the equilibrium structure is not considered, which could lead to possible diastereomers which influences their CD spectra.[505] For a discussion of possible diastereomers and their CD spectra, see ref[505]. Mono- and pseudo-*ortho* 4,12-disubstituted derivatives have found application as chiral building blocks in the synthesis of materials,[506–513] chiral catalysts,[514–519] or synthetic receptors.[520] Therefore, it is surprising that pseudo-*meta* 4,15-disubstituted [2.2]paracyclophanes are largely unexplored, so far, although the 4,15-dibromo-[2.2]paracyclophane (2) is easily accessible via bromination of [2.2]paracyclophane and has been known for a long time.[521,522] In fact only very few 4,15-disubstituted [2.2]paracyclophanes have been synthesized by D. J. Cram[523] in the early days of paracyclophane-chemistry to study their spectral properties. Despite the work of H. Hopf who brought this class of compounds back into focus in the early 2000s,[520,524] the number of pseudo-*meta* disubstituted derivatives is still very low. One reason for this might be that research has mainly been focused on the synthesis of the 4,12-derivatives as bidentate ligand or ligand precursors. However, another reason might be that the pseudo-meta dibromide was found

to be much less reactive compared to its pseudo-*ortho* substitutes isomer concerning a bromine-lithium exchange. This particular reaction is most often used to get access to functionalized derivatives.

Our group has been interested in rigid dissymmetric chiral scaffolds with uncommon stereogenic elements for quite some time now, and we have synthesized and resolved 9,9'-spirobifluorenes with stereogenic spirocenters,[525] Troöger's base derivatives with stereogenic nitrogen atoms,[526,527] and planar chiral pseudo-*ortho* 4,12-disubstituted [2.2]paracyclophanes[528] and used these for the synthesis of ditopic ligands for the stereoselective self-assembly of metallosupramolecular aggregates.[64,529–532] Hence, pseudo-*meta* 4,15-disubstituted [2.2]paracyclophanes also caught our attention because this substitution pattern brings functional groups in an angle of 120°, which makes them interesting building blocks for the formation of (metallo-)supramolecular aggregates.

Here, we report on the synthesis of various pseudo-*meta* 4,15-disubstituted [2.2]paracyclophanes and the chiral resolution of some complementarily substituted derivatives via semipreparative and preparative HPLC on chiral stationary phases. In doing so we were not only able to improve the syntheses of already known compounds such as 4,15-dihydroxy[2.2]paracyclophane,[522] [2.2]paracyclophane-4,15-di-carboxylic acid,[522,523], 4,15-diformyl[2.2]paracyclophane,[524] and 4,15-diamino[2.2]para-cyclophane,[523] but we could also synthesize the formerly unknown 4,15-diiodo[2.2]para-cyclophane, 4,15-di(4,4,5,5-tetramethyl-1,3,2-dioxoborolane)[2.2]paracylcophane, and [2.2]-paracyclophane-4,15-diazide, which can easily be converted into the corresponding diamine. Furthermore, we were able to separate the enantiomers of 4,15-dihydroxy[2.2]para-cyclophane, [2.2]paracyclophane-4,15-dicarboxylic acid (indirectly in form of its di(4-bromophenyl) ester and subsequent saponification), 4,15-diformyl[2.2]paracyclophane, 4,15-diamino[2.2]paracyclophane, and 4,15- di(4,4,5,5-tetramethyl-1,3,2-dioxoborolane)[2.2]para-cyclophane. Enantiomerically pure 4,15-diidodo[2.2]paracyclophane could be obtained via a Sandmeyer reaction from the corresponding enantiomerically pure diamine. The absolute configuration of the resolved compounds could be determined by single crystal X-ray diffraction and/or by comparison of quantum chemically calculated electronic circular dichroism (CD) spectra with experimentally obtained CD spectra.

9.2. Results and discussion

Our synthesis started from nonsubstituted [2.2]paracyclophane (**1**), which is commercially available and easy to functionalizevia well-known bromination.[521,524] To afford the desired (*rac*)-4,15-dibromo[2.2]paracyclophane ((*rac*)-**2**) we used the method developed by Hopf in 2008,[522] which also leads to achiral 4,16-dibromo[2.2]paracyclophane (**3**). Compound *(rac)*-**2** then served as our starting material for all further reactions (Figure 9.1).

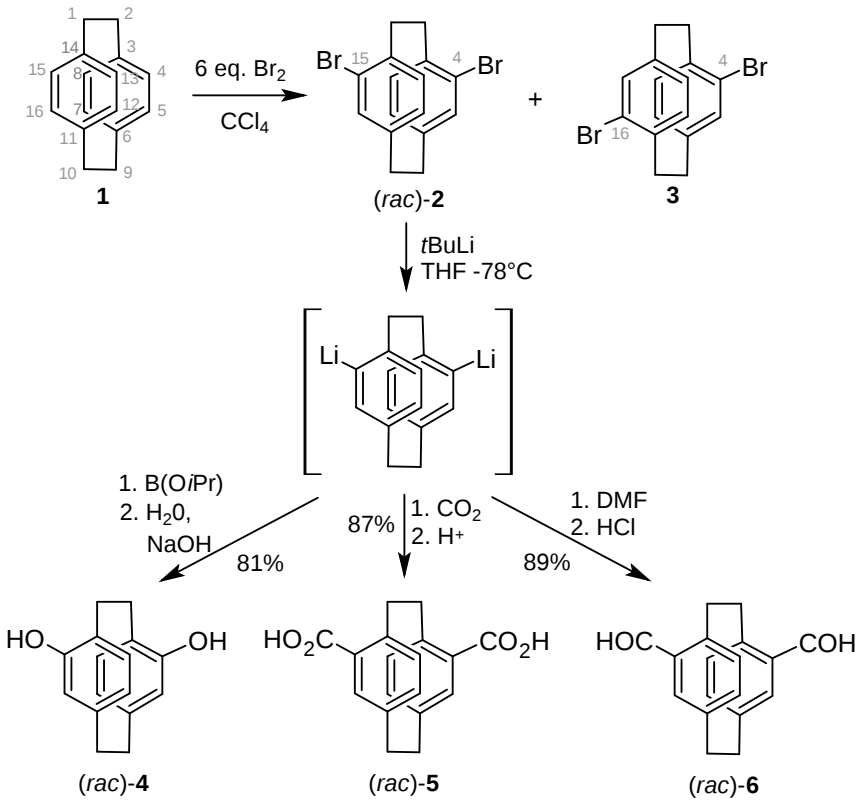


Figure 9.1.: Synthesis of 4,15-difunctionalized [2.2]paracyclophanes by bromine lithium exchange of (*rac*)-**2** and addition of electrophiles

The synthesis of compounds (*rac*)-**4**, (*rac*)-**5**, and (*rac*)-**6** has already been reported by H. Hopf,[522,524] however, only in rather low yields. We tried to follow these procedures but it soon became clear that the bromine-lithium exchange reaction must be the hindered step, which prevents higher yields. Hence, our first task was to improve this step. This was finally achieved by adding a solution of (rac)-**2** in dry THF to a solution of *tert*-butyllithium (*t*BuLi) in dry THF at –78 X°C. The color of the solution turns from flashy yellow to pale yellow, and after 1 h of stirring at –78 °C the bromine-lithium exchange is complete.

The synthetic procedure for the preparation of diol (*rac*)-**4** established by H. Hopf *et al.* allots the use of 2.4 equiv of *n*-butyllithium (*n*BuLi) in diethyl ether at room temperature followed by oxidation of the lithiated (*rac*)-**2** with nitrobenzene at –78 °C to obtain (*rac*)-**4** in 30% yield.[522] Interestingly, standard addition of $B(OMe)_3$ to the dilithiated intermediate and subsequent oxidative cleavage of the intermediate diboronate did not lead to the desired product. Alternatively, Hopf performed a stepwise synthesis involving monolithiation with 1.2 equiv of *n*BuLi in diethyl ether at room temperature followed by the addition of $B(OMe)_3$ with subsequent oxidation and saponification of the borate leading to the monohydroxy-monobromo compound. This compound wasthen etherified to protect the hydroxyl-function and subsequently subjected to the complete

sequence again to finally obtain the monomethoxy-monohydroxy compound in an overall yield of 65%. With our lithiation approach and the usage of $B(OiPr)_3$ instead of $B(OMe)_3$ we were able to increase the yield of (*rac*)-**4** dramatically to 81%.

Similarly, we were able to improve the synthesis of dicarboxylic acid (r*ac*)-**5** to 87% yield which H. Hopf *et al.* obtained via addition of CO_2 to dilithiated (*rac*)-**2** followed by acidification in 75% yield.[522]

In 2004 H. Hopf described the synthesis of (*rac*)-4,15- diformyl[2.2]paracyclophane ((*rac*)-**6**) in 62% from (*rac*)-**2** via a bromine-lithium exchange with s*ec*-butyllithium (*s*BuLi) in THF followed by the addition of *N*-formylpiperidine and subsequent treatment with aqueous HCl. Again, we were able to improve the synthesis by employing *t*BuLi and *N*,*N*-dimethylformamide (DMF) instead and we obtained (*rac*)-**6** after quenching with aqueous HCl in 89% yield.[524]

As we mentioned above our group is interested in enantiomerically pure compounds. Hence, the next challenge was the resolution of the racemic mixtures. As described earlier HPLC on a chiral stationary phase proved to be a versatile method to separate various 4,12-disubstituted [2.2]-paracyclophanes.[528] So we tried to apply this approach also for the resolution of the 4,15-disubstituted [2.2]paracyclophanes. In fact, both (*rac*)-**4** and (r*ac*)-**6** could be resolved on an analytical and a semipreparative scale in a very effective manner by using a CHIRALPAK IA as the stationary phase and different mixtures of n-hexane and ethanol as the eluent (see Supporting Information (SI)). This method allowed us to obtain both enantiomers in optically pure forms on a semipreparative scale.

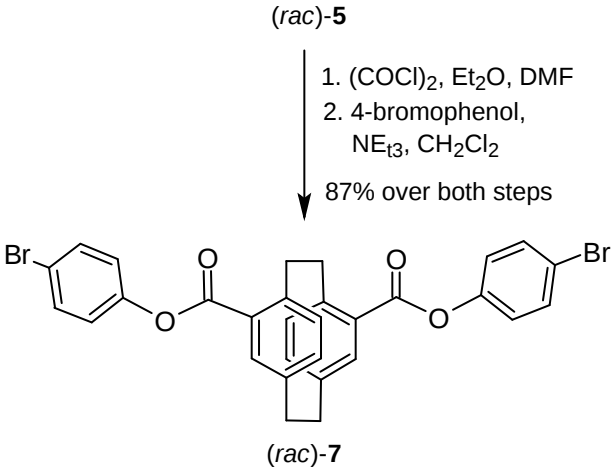


Figure 9.2.: Synthesis of di(bromophenyl) ester (*rac*)-**7**

To achieve the esterification the diacid was first transformed into the corresponding dicarboxylic acid chloride upon reaction with oxalyl chloride. This was reacted with 4-bromophenol to afford the desired diester (*rac*)-**7** in 87% yield (Figure 9.2). As hoped, we are also able to resolve the enantiomers in the same way on a semipreparative scale using *n*-hexane/ethanol (80:20 v/v) as the eluent. Saponification of the ester under alkaline

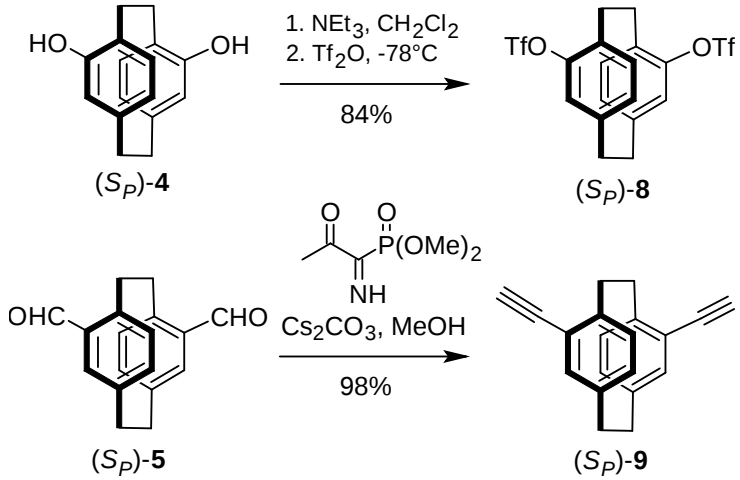


Figure 9.3.: Synthesis of enantiomerically pure ditriflate **8** and enantiomerically pure diethynyl-[2.2]paracyclophane **9**

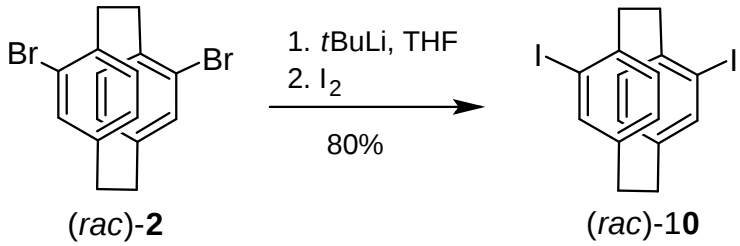


Figure 9.4.: Synthesis of racemic diiodo-[2.2]paracyclophane **10**

conditions then provided the enantiomerically pure acids.

The next step was to use the separated compounds as starting materials for the construction of further functionalized derivatives. Therefore, we decided to transform (*rac*)-**4** into the corresponding ditriflate **8**, which can be used as a coupling reagent in various types of cross-coupling reactions. The enantiomercially pure aldehyde **5** could be transformed into the corresponding dialkyne by treating it with the Bestmann-Ohira[533,534] reagent following an approach introduced by Hopf.[524] This dialkyne **9** is a promising starting material, e.g., for Sonogashira cross-coupling reactions (Figure 9.3).

Our next idea was to synthesize the 4,15-diiodo[2.2]-paracyclophane ((*rac*)-**10**), which is also a versatile starting material for various types of transformations including cross-coupling reactions. Again, we performed the bromine-lithium exchange in THF at – 78 °C by using *t*BuLi and quenched dilithiated **2** with iodine to get access to the desired diiodinated (*rac*)-**10** in 80% yield (Figure 9.4).

Unfortunately, we were not able to resolve (*rac)*-**10** directly into its enantiomers by HPLC on chiral stationary phases. Hence, we decided to develop a second strategy to synthesize **10** from the corresponding 4,15-diamino[2.2]paracyclophane ((*rac*)-**12**) via a Sandmeyer reaction hoping that this might by easier to resolve via chiral HPLC (Figure 9.5).

The first step is the lithiation of (rac)-2 with subsequent addition of *p*-toluensulfonyl azide leading to the diazide (*rac*)-**11**, which is quite stable and can be purified via column chromatography on silica gel to separate it from *p*-toluensulfonate and defunctionalized [2.2]paracyclophane. Unfortunately, however, we were not able to separate it from the monoazide byproduct at this stage. Hence, the resulting mixture was reduced to the corresponding amines with sodium borohydride, which could easily be separated via column chromatography on silica gel to afford the pure desired diamine (r*ac*)-**12** in a good overall yield of 68% starting from dibromide **2**. As hoped, this diamine could be resolved by HPLC on a CHIRALPAK IB column as the stationary phase and *n*-hexane and ethanol (70:30 v/v) as the eluent both on an analytical and a preparative scale. Enantiomerically pure 12 could then be converted into the enantiomerically pure diodide **10** by a Sandmeyer reaction, which was running smoothly without stereochemical leakage to afford the desired product in 74% yield.

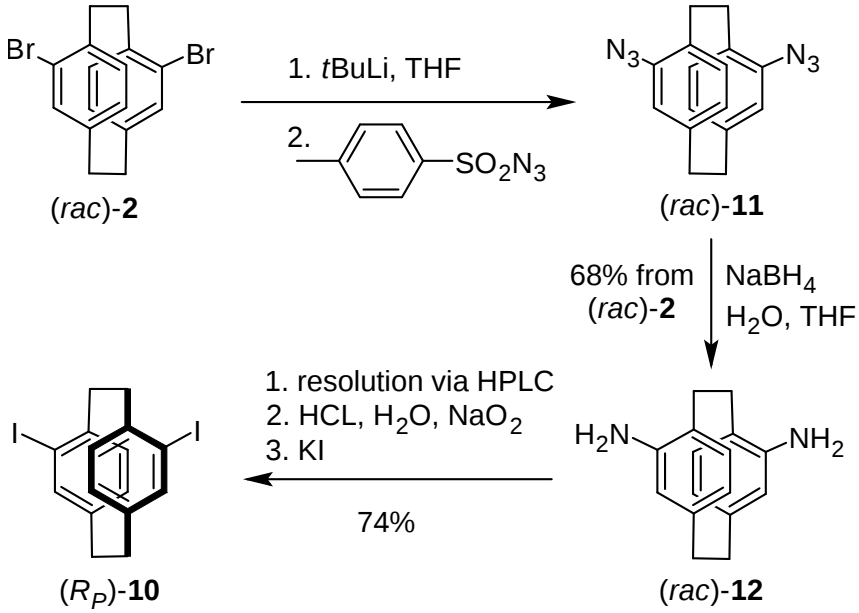


Figure 9.5.: Synthesis of enantiomerically pure 4,15-diiodo[2.2]paracyclophane **10** via racemic diazide **11** and racemic diamine **12** with subsequent chiral resolution via HPLC

Having achieved the syntheses of two enantiomerically pure starting materials for potential cross-coupling reactions (**8** and **10**) and the dialkyne (**9**), which might also serve as the starting material for a transmetalation agent [the reactive Cu-alkyne is formed *in situ* during a Sonogashira-type cross-coupling reaction], we wanted to broaden the spectrum of versatile precursors for the synthesis of more sophisticated molecular architectures based on the 4,15-difunctionalized [2.2]paracyclophane motif even further. Thus, we decided to synthesize a [2.2]paracyclophane-4,15-diboronic acid derivative next. Having access to enantiomerically pure starting materials our first approach was the direct

formation of the boronic pinacol ester via a Pd-catalyzed Miyaura cross-coupling reaction with bis(pinacolato)diboron,[535] but, unfortunately, this reaction did not lead to the desired product. Also, the boronic acid itself seems to be quite instable, and its synthesis via bromine-lithium exchange reaction and subsequent borylation with $B(OMe)_3$ or $B(O\mathit{i}Pr)_3$ followed by ester hydrolysis with aqueous ammonium chloride solution only led to an undefined mixture of products, which neither contained the desired boronic acid nor the nonsubstituted [2.2]paracyclophane. Even the addition of pinacol to the reaction mixture did not provide the boronic ester. This is in agreement with the fact that 4-substituted [2.2]paracyclophane boronicacid esters have been found to be unstable under similar conditions so far. Finally, we were able to solve the problem by using commercially available isopropoxyboronic acid pinacol ester, which is known to react directly with lithiated aromatic molecules to afford the usually rather stable boronic acid pinacol esters.[536,537]

This approach turned out to be very effective, and we were able to prepare the bis(boronic pinacol ester) ((*rac*)-**13**) in an excellent yield of 91%. Fortunately, racemic **13** could also be resolved via HPLC by using a CHIRALPAK IB column as the stationary phase and *n*-hexane/chloroform (98:2 v/v) as the eluent both in analytical and preparative scale (Figure 9.6).

Having achieved the successful resolution of racemic **4**, **6**, **7**, **12**, and **13** the final task was to determine the absolute configuration of the isolated enantiomers. A powerful method to determine the absolute configuration is X-ray diffraction analysis of suitable single crystals and the analysis of the Flack parameter. This requires the presence of a heavy atom such as bromine, iodine, or sulfur in the structure, which only applies to compound **7**. Conveniently, **4** can easily be transformed into the corresponding ditriflate and **8** and **12** into the corresponding diiodide **10**. To determine the absolute configuration of dialdehyde **6** it was transformed into the corresponding (4- bromophenyl)hydrazone **14** (Figure 9.7).

Fortunately, we succeeded in growing single crystals of compounds **7**, **10**, and **14** suitable for XRD-measurements (seeSI). Thus, we were able to unambiguously assign the (R_P)-configuration to the (–)-enantiomer of **7**, the (R_P)-configuration to the (–)-

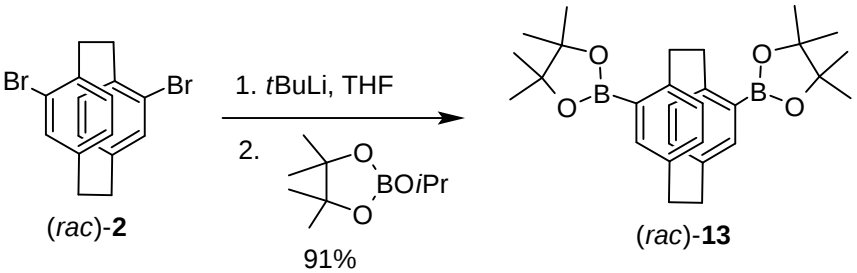


Figure 9.6.: Synthesis of racemic diboronic pinacolatoester **13**

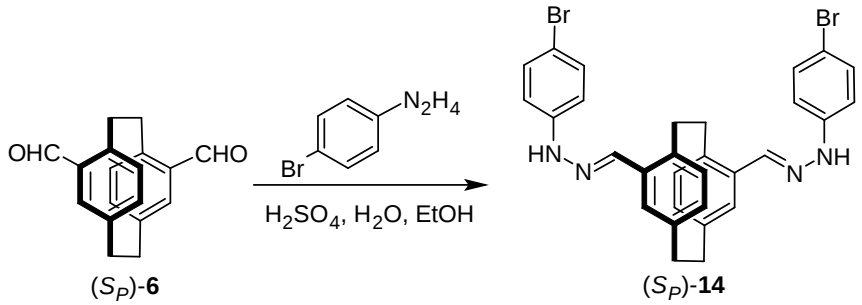


Figure 9.7.: Synthesis of (+)-(S_P)-**14** from enantiomerically pure (+)-(S_P)-**6**

enantiomer of **10**, and the (S_P)-configuration to the (+)-enantiomer of **14**. This enabled us also to conclude on the absolute configurations of the enantiomerically pure precursors **5**, **6** and **12**. Hence, (–)-**5** is (R_P)-configurated, (+)-6 is (S_P)-configurated, and (–)-12 is (R_P)-configurated.

Unfortunately, we were not successful in growing suitable single crystals of **8** as it turned out to be highly viscous oil in its enantiomerically pure form. Additionally, we could not elucidate the absolute configuration of **13** because of the lack of any heavy atom in the structure. Therefore, we turned our attention to another analytical method that is well established to assign the absolute stereochemistry of enantiomerically pure chiral molecules, the circular dichroism (CD) spectroscopy.

In order to allow an assignment from experimental CD spectra it is probably best to compare them with those obtained from quantum chemical calculations (Figure F.1). These calculations were done by employing the recently developed simplified time dependent density functional theory (sTD-DFT) approach.[413] Single-point calculations with the global hybrid BHLYP[97] and the range-separated hybrid functional CAM-B3LYP[107] together with the def2-TZVP basis set
citedef2-tzvp were performed on TPSS[93]-D3(BJ)[36,37]/def2-TZVP optimized structures. Both functionals reproduce the main features of the experimental spectra, but CAM-B3LYP shows the overall better agreement. The range-separation technique which alleviates so called self-interaction errors in the density functional improves the quality of the calculated spectra even if CAM-B3LP is not asymptotically correct as it contains 65% exact exchange in the long-range limit. In case of (–)-(S_P)-**4**, (+)-(S_P)-**6** and (+)-(S_P)-**13** a small shift of the computed vertical excitation energies by -0.2 to -0.3 eV is observed, which is typical for this functional together with the sTD-DFT approach (spectra for all four compounds including the rotatory strengths for CAM-B3LYP are provided in the Supporting Information). For a more detailed discussion of the CD spectra of [2.2]paracyclophanes, see ref[505].

Given the fact that all (–)-enantiomers described above were found to be (R_P)-configurated we were initially surprised to find that that comparison of the experimental and

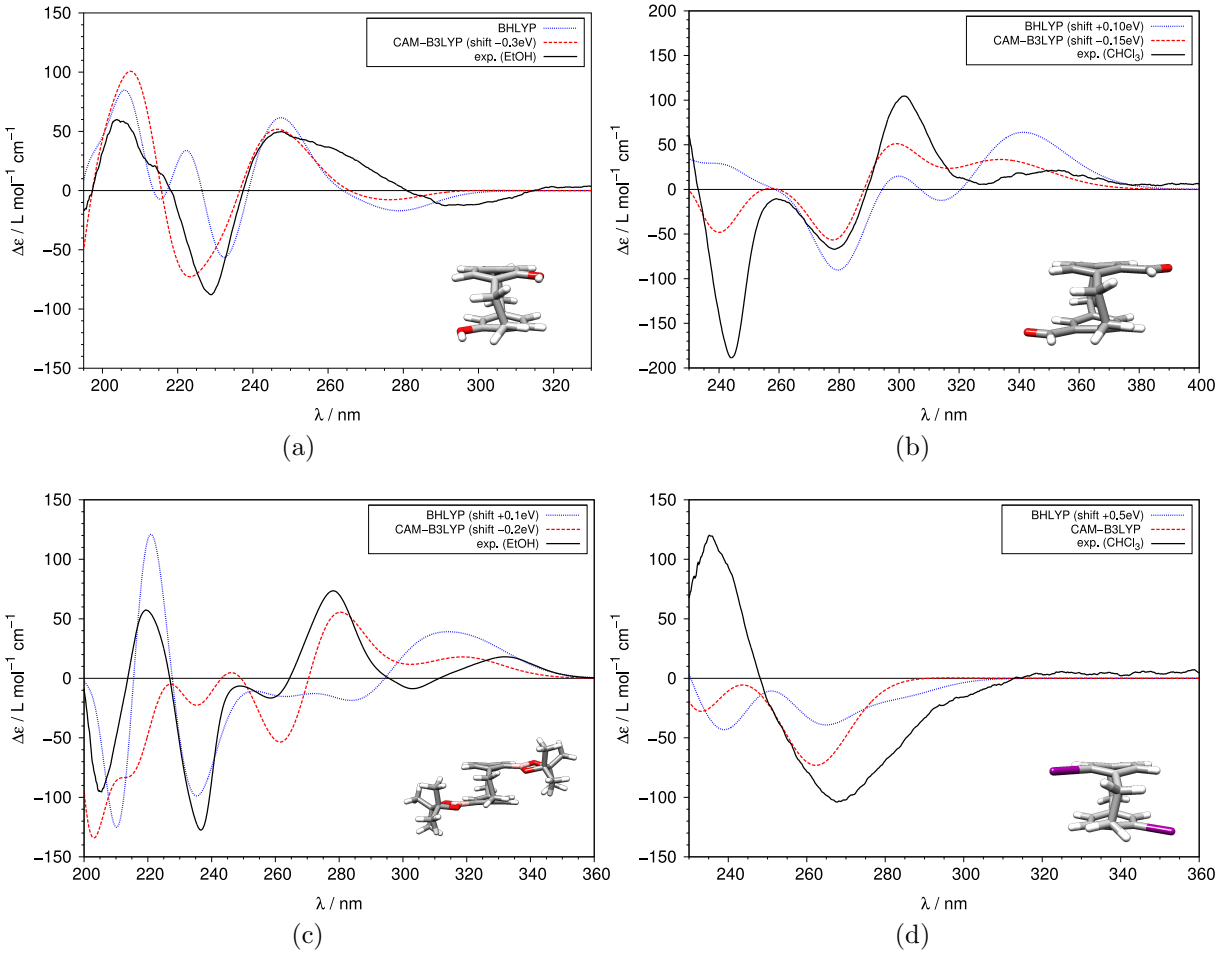


Figure 9.8.: Experimental and simulated CD spectra of (a) (–)-(S_P)-**4**, (b) (+)-(S_P)-**6** (c) (–)-(R_P)-**10**, and (d) (+)-(S_P)-**13**.

simulated spectra clearly proved that the (–)-enantiomer of compound **4** is indeed (S_P)-configurated. Hence, we decided to validate our theoretical approach by also applying it to the simulation of the spectra of **6** and **10** whose configuration we could already assign via single crystal X-ray diffraction. These calculations nicely agree with the experimental spectra and corroborate the assignment made according to the X-ray diffraction analysis. Having proved that the applied quantum chemical method is indeed reliable, we were also able to determine the absolute configuration of boronic ester **13** showing that the (+)-enantiomer is again (S_P)-configurated.

9.3. Conclusion

In summary we have synthesized 12 planar chiral 4,15-difunctionalized [2.2]paracyclophanes. Some of them (**2**, **4**–**6**, **9**, and **12**) have been prepared in racemic form before, but some of them (**8**, **10**, **11**, **13**, and **14**) have not been reported yet. By improving

the efficiency of the bromine-lithium exchange we have been able to dramatically increase the yields of **4**–**6** and **12**. Furthermore, we have been able to resolve five of these compounds (**4**, **6**, **7**, **12**, and **13**) by HPLC techniques on an analytical, semipreparative and preparative scale by using CHIRALPAK IA and CHIRALPAK IB stationary phases. The absolute configuration could be assigned by X-ray crystal structure analysis of **7**, **10**, and **14**, which also implies the absolute configuration of **5**, **6**, and **12**. The absolute configuration of **4** and **13** was determined by comparison of experimental CD-spectra with quantum chemically calculated ones. This method was validated by comparing experimental and simulated CD spectra of **6** and **10** with known configuration. Most of the derivatives carry versatile functional groups that offer the possibility to integrate the 4,15-difunctionalized [2.2]paracyclophane skeleton into more sophisticated (supra-)molecular architectures with well-defined stereochemical properties.

9.4. Computational details

The geometries of the investigated molecules were optimized on the DFT level using the TURBOMOLE 6.4 program package[261] and employing the TPSS functional[93] together with the D3(BJ)[36,37] dispersion correction and the def2-TZVP[237] basis set. The resolution-of-identity (RI) approximation for the Coulomb integrals[74] was applied using matching default auxiliary basis sets[238] and for the integration of the exchange-correlation contribution the numerical quadrature grid *m4* was employed.[239] Single-point calculations on the optimized geometries were performed with the global hybrid functional BHLYP[97] and the range-separated hybrid functional CAM-B3LYP[107] together with the def2-TZVP basis set utilizing the development version of ORCA 3.0 (precursor of 3.0.1 release).[233,234] The RI approximation for the Coulomb integrals was used in combination with the chain-of-spheres (COSX) approximation.[538,539] Rotatory strengths values for the electronic transitions from ground to singly excited states were obtained at the sTD-DFT level,[413,414] and all excitations up to a threshold of 10 eV were included. The molecular circular dichroism ($\Delta\epsilon$) values were calculated by convoluting Gaussian functions with a width of $\sigma = 0.4$ eV which are centered at the wavelength of the electronic transitions and multiplied by the corresponding rotatory strength (vertical transitions). For all spectra the origin independent velocity rotatory strength was used.

Acknowledgement

Financial support for this work from the Deutsche Forschungsgemeinschaft (SFB 624) is gratefully acknowledged. We would like to thank Prof. A. C. Filippou for providing the infrastructure for the the XRD experiments.

Part V.

Final summary and conclusions

In this thesis, the application and development of computationally cost-efficient DFT and HF based methods for noncovalent interactions in large systems and the evaluation of multilevel quantum mechanical methodologies for the accurate description of the thermochemistry of supramolecular host-guest systems have been presented.

It has been shown that semi-local density functional approximations and HF can yield good results for interaction energies and geometries of noncovalently bound systems in a small atomic-orbital basis set if the two major error sources are eliminated. These are the missing London dispersion and the BSSE. Several modern correction schemes have been tested and they all outperform the plain semi-local DFAs or plain HF. Dispersion correcting effective potentials, i.e. B3LYP-DCP performs excellently for water, small noncovalent dimers, and ice polymorphs, but cannot accurately treat larger organic systems. The combination of the D3 dispersion scheme and the geometrical counterpoise correction, e.g. B3LYP-D3-gCP provides accurate results for small noncovalent dimers, large supramolecular systems and organic molecular crystals, but fails for water and ice. The recently developed PBEh-3c functional is always one of the two best performing methods, and yields reasonable results for water and ice and accurate energies and geometries for small dimers, supramolecular complexes, and molecular crystals.

Further, a minimal basis set HF based method (HF-3c) with three atom-pairwise corrections for dispersion, BSSE, and BSIE, which has been developed and tested within this thesis. HF-3c is slower than conventional semiempirical methods, but generally more accurate, robust, numerically stable, and less empirical. Compared to DFT, HF is inherently free of the self-interaction error and provides numerically noiseless analytical frequencies. HF-3c yields excellent interaction energies for small noncovalently bound dimers and good association energies for the S12L set of supramolecular host-guest complexes. Further, it provides accurate geometries of small organic molecules, noncovalent dimers, and supramolecular systems. HF-3c optimized geometries of small proteins with up to 550 atoms yield good results in standard protein structure health checks and have a reasonable agreement with experimental structures. HF-3c also provides good thermostatistical corrections from energy to free energies and with its good cost-accuracy ratio it is ideally suited for sampling free energies of different binding conformations for large systems, e.g. supramolecular host-guest complexes.

Within this thesis, HF-3c has been employed in a nondynamic single-structure multilevel approach to compute association free energies ΔG_a of host-guest complexes in solution. The association energy ΔE in the gas phase has been computed with dispersion-corrected DFT including three-body dispersion terms, i.e. on the DFT(hybrid)-D3$^{\mathrm{ATM}}$/QZ level. HF-3c and other semiempirical methods have been employed to calculate the thermostatistical corrections from energy to free energy ΔG_{RRHO} and a continuum solvation model has been used to include solvation effects ($\Delta\delta G_{solv}$). The general procedure

has been illustrated for a case study of eight typical complexes: two tweezer complexes, two pseudorotaxanes with either one or two crownethers as wheels, two cucurbit[7]uril complexes, and two complexes of the fullerene C_{70} in cycloparaphenylenes. For these eight systems the PW6B95-D3/QZ energies with HF-3c thermostatistical corrections and COSMO-RS(12) solvation contributions yield an MD of -2.1 and an MAD of 4.2 kcal mol^{-1} compared to the experimental values. Errors larger than 6 kcal mol^{-1} have been observed for the fullerene-cycloparaphenylene complexes. Excluding them from the statics results in an MD and MAD of 0.4 and 1.8 kcal mol^{-1}, respectively.

Although the methodology did not always yield highly accurate results, its predictive power has been be demonstrated. With the participation in the SAMPL4 blind test challenge, it has been put to a realistic evaluation. Relative association free energies have been predicted for a cucurbit[7]uril host and 14 guest molecules that contain either one or two ammonia groups. HF-3c has been applied to sample possible binding conformations and the final ΔG_a have been calculated on the PW6B95-D3$^{\mathrm{ATM}}$/QZ level with HF-3c thermal corrections and COSMO-RS(12) solvation contributions with and without the inclusion of chloride counterions. The results with counterions have been submitted to the organizing committee and the MAD and RMSD compared to the subsequently published experimental values are only 2.0 and 2.6 kcal mol^{-1}, respectively. In comparison with computational results obtained with other methods, theses predictions rank in the top three of all statistical accuracy measurements. For almost all complexes the inclusion of counterions improve the calculated ΔG_a.

Errors in ΔG_a appeared during the case study, as well as in the SAMPL4 participation, especially for the second host which was a flexible, octa-acid basket-shaped molecule with an unclear charge state. Thus, further testing of the multilevel approach seemed mandatory. The S12L set of supramolecular host-guest complex has been extended to 30 realistic host-guest systems. The new S30L set contains larger systems with up to 200 atoms, a broader variety of interaction motifs, slightly more flexible compounds, and higher charges (anions and cations). The experimental reference values are in the range of -0.7 to -24.7 kcal mol^{-1}. The influence of counterions for the charged systems has been investigated further by employing chloride for cations and sodium ions for anions (S30L-CI). For the two sets, S30L and S30L-CI the theoretical best estimate for ΔG_a has been obtained. Different dispersion-corrected density functionals (PBE-D3, TPSS-D3, B3LYP-D3, PW6B95-D3 and ωB97X-D3) together with a quadruple-zeta basis set have been tested for ΔE. They have been combined with the semiempirical methods HF-3c, PM6-D3 and DFTB3-D3 to calculate ΔG_{RRHO}, and COSMO-RS with different parametrizations as well as SMD to compute $\Delta\delta G_{solv}$. For the S30L the best association free energies are obtained with PW6B95-D3$^{\mathrm{ATM}}$/QZ energies, HF-3c thermostatistical corrections, and COSMO-RS(12/13-fine) (for non-polar solvents/water) solvation free en-

ergies. The resulting MD and MAD are -0.1 and 2.4 kcal mol^{-1}, respectively. For the systems including counterions ωB97X-D3$^{\mathrm{ATM}}$/QZ outperforms PW6B95-D3$^{\mathrm{ATM}}$/QZ for the energies and the COSMO-RS(13-fine) yields the best $\Delta\delta G_{solv}$, giving an MAD of 2.1 kcal mol^{-1}.

In collaboration with the group of Prof. Gansäuer, the presented approach to calculate ΔG_a in solution has been used to investigate adducts of titanocene(III) catalysts and hydrochloride additives (Py $\cdot$ HCl, Coll $\cdot$ HCl and $Et_3N \cdot$ HCl). It has been shown that the stabilities of these adducts are determined by the extent of hydrogen bonding between the catalyst and the ammonium cation and can be tuned by the steric bulk and acidity of the additive cation.

In the collaboration with Prof. Lützen and his group, quantum chemical calculations of electronic chircular dichroism spectra have helped to identify the absolute configuration of several 4,15-difunctionalized [2.2]paracyclophanes. These difunctionalized paracyclophanes are precursors for ligands that form double- or triple-stranded helicates with transition metal ions via self-assembly processes, whose investigation is ongoing. As the theoretical spectra for two of the three possible diastereomers are often similar, the computation of relative association free energies is used as a second indicator to identify the experimentally obtained helicates.

With the current approach to compute ΔG_a the average error for association energies of host-guest systems is about 2 kcal mol^{-1}, which has been found during the SAMPL4 blind test participation as well as during the S30L studies. The non-dynamic methodology makes the drastic assumption that only one conformer with one specific binding motif is important and all other structures are neglected. For testing purposes, HF-3c has been employed to sample about a hundred different conformers for two different cucurbit[7]uril complexes. Weighting all their ΔG_a values according to the Boltzmann statistics lowers the association fee energy obtained from only one structure by about 1 kcal mol^{-1}. Cucurbiturils are rigid molecules and the error due to neglecting other conformers could be larger for more flexible systems. A quantum mechanically derived force field (QMDFF[540]) was developed recently and coupled to a simulated annealing approach. This could simplify the sampling process and could be used to identify the structures with the lowest ΔG_a for further treatment with quantum chemical methods.

Already for the single-structure ΔG_a value several error sources are present. The DFT-D3$^{\mathrm{ATM}}$/QZ association energies are inherently off by about 5 %. Depending on the system, ΔE lies in the range of -20 to -80 kcal mol^{-1} which results in an error of about 1 to 4 kcal mol^{-1}. For routine applications, dispersion corrected DFT is the only option and thus, this error has to be tolerated. The thermostatistical corrections are based on a lot of thermodynamic approximations, and their error is estimated to be about 5 %, i.e.

about 1 kcal mol^{-1}. Anharmonic effects might be important and should be investigated. However, computing the third and forth derivatives of the energy with respect to the nuclear coordinates are too expensive with standard quantum chemical methods. The QMDFF chould be used to compute anharmonic corrections to the ΔG_{RRHO} values in order to investigate their significance. The largest error source is likely the treatment of solvation effects. Continuum solvation models like COSMO-RS neglect all explicit solvation effects. Further, the structural change and the change in the molecular vibrations when going from gas phase to solution are disregarded. Sometimes, COSMO-RS yields $\Delta\delta G_{solv}$ values that differ by up to 5 kcal mol^{-1} depending on the parametrization. The estimated error in the COSMO-RS $\Delta\delta G_{solv}$ is 10 % which amounts to 1–3 kcal mol^{-1} depending on the system. A generalized Born solvation model is currently developed in the Grimme group and can hopefully be used to obtain better molecular structures in solution as well as more accurate $\Delta\delta G_{solv}$.

Overall, the sum of all estimated errors is much larger than the average error in ΔG_a of about 2 kcal mol^{-1} which indicates fortunate error compensation due to the different signs of the contributions. This thesis has clearly demonstrated that the nondynamic single-structure multilevel approach can nevertheless be used routinely to accurately compute and predict association free energies in solution. It can be used to design new supramolecular complexes and help to guide experimentalists in finding suitable guest molecules for a specific host.

Bibliography

[1] J.-M. Lehn, *Angew. Chem. Int. Ed.* **1988**, *27*, 89–112.

[2] J.-M. Lehn, *Science* **1993**, *260*, 1762–1763.

[3] A. J. Stone, *The Theory of Intermolecular Forces*, Oxford University Press: Oxford, U.K., 2nd ed., **2013**.

[4] I. G. Kaplan, *Intermolecular Interactions*, John Wiley & Sons: Chichester, U.K., **2006**.

[5] C. R. H. A. Villiers, *Seances Acad. Sci.* **1891**, *112*, 536–538.

[6] A. Werner, *Z. Anorg. Chem.* **1893**, *3*, 267–330.

[7] E. Fischer, *Ber. Deut. Chem. Ges.* **1894**, *27*, 2985–2993.

[8] W. M. Latimer, W. H. Rodebush, *J. Am. Chem. Soc.* **1920**, *42*, 1419–1433.

[9] D. E. Koshland Jr., *Proc. Nat. Acad. Sci.* **1958**, *44*, 92–97.

[10] W. J. D., F. H. C. Crick, *Nature* **1953**, *171*, 737–738.

[11] http://www.nobelprize.org/nobel_prizes/medicine/laureates/1962/.

[12] C. J. Pedersen, *J. Am. Chem. Soc.* **1967**, *89*, 7017–7036.

[13] C. J. Pedersen, *Angew. Chem. Int. Ed.* **1988**, *27*, 1021–1027.

[14] http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1987/.

[15] T. Park, S. C. Zimmerman, S. Nakashima, *J. Am. Chem. Soc.* **2005**, *127*, 6520–1.

[16] J.-M. Lehn, *Supramolecular chemistry: Concepts and perspectives*, VCH: Weinheim, Germany, **1995**.

[17] J. L. Atwood, J. Steed, *Supramolecular Chemistry*, John Wiley & Sons: Chichester, U.K., 2nd ed., **2009**.

[18] I. on Molecular Recognition, *Chem. Rev.* **1997**, *97*, 1231–1734.

[19] M. Hardouin-Lerouge, P. Hudhomme, M. Sallé, *Chem. Soc. Rev.* **2011**, *40*, 30–43.

[20] M. M. Conn, J. Rebek, *Chem. Rev.* **1997**, *97*, 1647–1668.

[21] F. M. Raymo, J. F. Stoddart, *Chem. Rev.* **1999**, *99*, 1643–1664.

[22] D. J. Cram, *Angew. Chem. Int. Ed.* **1988**, *27*, 1009–1020.

[23] R. Eisenschitz, F. London, *Z. Phys.* **1930**, *60*, 491–527.

[24] F. London, *Z. Phys.* **1930**, *63*, 245–279.

[25] K. S. Kim, P. Tarakeshwar, J. Y. Lee, *Chem. Rev.* **2000**, *100*, 4145–4185.

[26] K. E. Riley, P. Hobza, *WIREs Comput. Mol. Sci.* **2011**, *1*, 3–17.

[27] E. G. Hohenstein, C. D. Sherrill, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 304–326.

[28] E. A. Meyer, R. K. Castellano, F. Diederich, *Angew. Chem. Int. Ed.* **2003**, *42*, 1210–1250.

[29] N. Kannan, S. Vishveshwara, *Protein Eng.* **2000**, *13*, 753–761.

[30] B. Kirchner, M. Reiher in *Analytical Methods in Supramolecular Chemisty* (Ed.: C. Schalley), Wiley-VCH: Weinheim, Germany, 2nd ed., **2012**, pp. 743–794.

[31] F. London, *Trans. Faraday Soc.* **1937**, *33*, 8–26.

[32] S. Grimme, *WIREs Comput. Mol. Sci.* **2011**, *1*, 211–228.

[33] J. Klimeš, A. Michaelides, *J. Chem. Phys.* **2012**, *137*, 120901.

[34] L. A. Burns, A. Vázquez-Mayagoitia, B. G. Sumpter, C. D. Sherrill, *J. Chem. Phys.* **2011**, *134*, 084107.

[35] L. Goerigk, *J. Chem. Theory Comput.* **2014**, *10*, 968–980.

[36] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.

[37] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[38] S. Miertuš, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, *55*, 117–129.

[39] S. Miertuš, J. Tomasi, *Chem. Phys.* **1982**, *65*, 239–245.

[40] A. Klamt, G. Schüürmann, *J. Chem. Soc. Perkin Trans. 2* **1993**, 799–805.

[41] A. Klamt, *J. Phys. Chem.* **1995**, *99*, 2224–2235.

[42] F. Eckert, A. Klamt, *AIChE J.* **2002**, *48*, 369–385.

[43] S. Ehrlich, J. Moellmann, S. Grimme, *Acc. Chem. Res.* **2013**, *46*, 916–926.

[44] S. Grimme, *ChemPhysChem* **2012**, *13*, 1407–1409.

[45] G. Ohlendorf, C. W. Mahler, S.-S. Jester, G. Schnakenburg, S. Grimme, S. Höger, *Angew. Chem. Int. Ed.* **2013**, *52*, 12086–12090.

[46] S. Grimme, *Chem. Eur. J.* **2012**, *18*, 9955–9964.

[47] H. Kruse, S. Grimme, *J. Chem. Phys.* **2012**, *136*, 154101.

[48] See http://sampl.eyesopen.com/.

[49] H. S. Muddana, A. T. Fenley, D. L. Mobley, M. K. Gilson, *J. Comput.-Aided Mol. Des.* **2014**, *28*, 305–317.

[50] S. Liu, C. Ruspic, P. Mukhopadhyay, S. Chakrabarti, P. Y. Zavalij, L. Isaacs, *J.*

Am. Chem. Soc. **2005**, *127*, 15959–15967.

[51] J. W. Lee, S. Samal, N. Selvapalam, H.-J. Kim, K. Kim, *Acc. Chem. Res.* **2003**, *36*, 621–630.

[52] J. Lagona, P. Mukhopadhyay, S. Chakrabarti, L. Isaacs, *Angew. Chem. Int. Ed.* **2005**, *44*, 4844–4870.

[53] M. V. Rekharsky, T. Mori, C. Yang, Y. H. Ko, N. Selvapalam, H. Kim, D. Sobransingh, A. E. Kaifer, S. Liu, L. Isaacs, W. Chen, S. Moghaddam, M. K. Gilson, K. Kim, Y. Inoue, *Proc. Nat. Acad. Sci.* **2007**, *104*, 20737–20742.

[54] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

[55] A. Gansäuer, B. Rinker, M. Pierobon, S. Grimme, M. Gerenkamp, C. Mück-Lichtenfeld, *Angew. Chem. Int. Ed.* **2003**, *42*, 3687–3690.

[56] B. M. Trost, H. C. Shen, J.-P. Surivet, *J. Am. Chem. Soc.* **2004**, *126*, 12565–12579.

[57] A. Gansäuer, B. Rinker, N. Ndene-Schiffer, M. Pierobon, S. Grimme, M. Gerenkamp, C. Mück-Lichtenfeld, *Eur. J. Org. Chem.* **2004**, *2004*, 2337–2351.

[58] D. Leca, L. Fensterbank, E. Lacôte, M. Malacria, *Angew. Chem.Int. Ed.* **2004**, *43*, 4220–4222.

[59] A. Gansäuer, M. Behlendorf, D. von Laufenberg, A. Fleckhaus, C. Kube, D. V. Sadasivam, R. A. Flowers, *Angew. Chem. Int. Ed.* **2012**, *51*, 4739–4742.

[60] A. Lützen, M. Hapke, J. Griep-Raming, D. Haase, W. Saak, *Angew. Chem. Int. Ed.* **2002**, *41*, 2086.

[61] J. Bunzen, R. Hovorka, A. Lützen, *J. Org. Chem.* **2009**, *74*, 5228–5236.

[62] J. Bunzen, T. Bruhn, G. Bringmann, A. Lützen, *J. Am. Chem. Soc.* **2009**, *131*, 3621–30.

[63] C. Gütz, R. Hovorka, N. Struch, J. Bunzen, G. Meyer-Eppler, Z.-W. Qu, S. Grimme, F. Topić, K. Rissanen, M. Cetina, M. Engeser, A. Lützen, *J. Am. Chem. Soc.* **2014**, *136*, 11830–8.

[64] C. Gütz, R. Hovorka, C. Klein, Q.-Q. Jiang, C. Bannwarth, M. Engeser, C. Schmuck, W. Assenmacher, W. Mader, F. Topić, K. Rissanen, S. Grimme, A. Lützen, *Angew. Chem. Int. Ed.* **2014**, *53*, 1693–1698.

[65] G. Meyer-Eppler, PhD thesis, Universiät Bonn, **2014**.

[66] E. Schrödinger, *Ann. Phys.* **1926**, *384*, 361–376.

[67] M. Born, R. Oppenheimer, *Ann. Phys.* **1927**, *389*, 457–484.

[68] J. Slater, *Phys. Rev.* **1929**, *34*, 1293–1322.

[69] D. R. Hartree, *Math. Proc. Cam. Phil. Soc.* **2008**, *24*, 89–110.

[70] D. R. Hartree, *Math. Proc. Cam. Phil. Soc.* **1928**, *24*, 111–132.

[71] V. Fock, *Z. Phys.* **1930**, *61*, 126–148.

[72] C. C. J. Roothaan, *Rev. Mod. Phys.* **1951**, *23*, 69–89.

[73] G. G. Hall, *Proc. Roy. Soc. A* **1951**, *205*, 541–552.

[74] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *242*, 652–660.

[75] M. Schütz, H.-J. Werner, *J. Chem. Phys.* **2001**, *114*, 661.

[76] M. Schütz, H.-J. Werner, *Chem. Phys. Lett.* **2000**, *318*, 370–378.

[77] C. Riplinger, F. Neese, *J. Chem. Phys.* **2013**, *138*, 034106.

[78] C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese, *J. Chem. Phys.* **2013**, *139*, 134101.

[79] L. H. Thomas, *Proc. Cambridge Phil. Soc.* **1927**, *23*, 542–548.

[80] E. Fermi, *Z.Physik* **1928**, *48*, 73–79.

[81] P. Hohenberg, W. Kohn, *Phys. Rev. B* **1964**, *136*, 864–871.

[82] R. G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press: Oxford, U.K., Oxford, **1989**.

[83] W. Koch, M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley-VCH: New York, U.S., **2001**.

[84] W. Kohn, L. J. Sham, *Phys. Rev. A* **1965**, *140*, 1133–1138.

[85] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, G. I. Csonka, *J. Chem. Phys.* **2005**, *123*, 062201.

[86] L. Goerigk, S. Grimme, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.

[87] J. Slater, *Phys. Rev.* **1951**, *81*, 385–390.

[88] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200–1211.

[89] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[90] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

[91] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100.

[92] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.

[93] J. Tao, J. Perdew, V. Staroverov, G. Scuseria, *Phys. Rev. Lett.* **2003**, *91*, 146401.

[94] A. M. Teale, S. Coriani, T. Helgaker, *J. Chem. Phys.* **2010**, *132*, 164115.

[95] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[96] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623–11627.

[97] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 1372.

[98] C. Adamo, V. Barone, *J. Chem. Phys* **1999**, *110*, 6158.

[99] Y. Zhao, N. E. Schultz, D. G. Truhlar, *J. Chem. Theory Comput.* **2006**, *2*, 364–382.

[100] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215–241.

[101] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 5656–5667.

[102] T. Leininger, H. Stoll, H.-J. Werner, A. Savin, *Chem. Phys. Lett.* **1997**, *275*, 151–160.

[103] H. Iikura, T. Tsuneda, T. Yanai, K. Hirao, *J. Chem. Phys-* **2001**, *115*, 3540.

[104] J.-D. Chai, M. Head-Gordon, *J. Chem. Phys.* **2008**, *128*, 084106.

[105] J.-D. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–20.

[106] Y.-S. Lin, G.-D. Li, S.-P. Mao, J.-D. Chai, *J. Chem. Theory Comput.* **2013**, *9*, 263–272.

[107] T. Yanai, D. P. Tew, N. C. Handy, *Chem. Phys. Lett* **2004**, *393*, 51–57.

[108] A. Görling, M. Levy, *Phys. Rev. B* **1993**, *47*, 13105–13113.

[109] A. Görling, M. Levy, *Phys. Rev. A* **1994**, *50*, 196–204.

[110] H. Eshuis, J. E. Bates, F. Furche, *Theor. Chem. Acc.* **2012**, *131*,.

[111] S. Grimme, *J. Chem. Phys.* **2006**, *124*, 034108.

[112] G. Seifert, E. H., W. Bieger, *Z. Phys. Chem.* **1986**, *267*, 529–539.

[113] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, *Phys. Rev. B* **1998**, *58*, 7260–7268.

[114] M. Elstner, *Theor. Chem. Acc.* **2005**, *116*, 316–325.

[115] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, John Wiley & Sons Ltd: West Sussex, U.K., 2nd ed., **2004**.

[116] F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons Ltd: West Sussex, U.K., 2nd ed., **2007**.

[117] M. J. S. Dewar, W. Thiel, *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.

[118] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

[119] J. J. P. Stewart, *J. Comput. Chem.* **1989**, *10*, 209–220.

[120] D. N. Nanda, K. Jug, *Theor. Chim. Acta* **1980**, *57*, 95–106.

[121] B. Ahlswede, K. Jug, *J. Comput. Chem.* **1999**, *20*, 563–571.

[122] K. Jug, G. Geudtner, T. Homann, *J. Comput. Chem.* **2000**, *21*, 974–987.

[123] T. Bredow, G. Geudtner, K. Jug, *J. Comput. Chem.* **2001**, *22*, 861–887.

[124] J. J. P. Stewart, *J. Mol. Mod.* **2007**, *13*, 1173–213.

[125] J. J. P. Stewart, *J. Mol. Mod.* **2013**, *19*, 1–32.

[126] W. Weber, W. Thiel, *Theor. Chem. Acc.* **2000**, *103*, 495–506.

[127] Y. Yang, H. Yu, D. York, Q. Cui, M. Elstner, *J. Phys. Chem. A* **2007**, *111*, 10861–73.

[128] M. Gaus, A. Goez, M. Elstner, *J. Chem. Theory Comput.* **2013**, *9*, 338–354.

[129] M. Kubillus, T. Kubař, M. Gaus, J. Řezáč, M. Elstner, *J. Chem. Theory Comput.* **2015**, *11*, 332–342.

[130] S. Grimme, *J. Comput. Chem.* **2004**, *25*, 1463–1473.

[131] S. Grimme, *J. Comp. Chem.* **2006**, *27*, 1787–1799.

[132] H. B. G. Casimir, D. Polder, *Phys. Rev.* **1948**, *73*, 360–372.

[133] A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **2005**, *123*, 154101.

[134] E. R. Johnson, A. D. Becke, *J. Chem. Phys.* **2005**, *123*, 24101.

[135] R. A. DiStasio Jr, V. V. Gobre, A. Tkatchenko, *J. Phys.: Condens. Matter* **2014**, *26*, 213202.

[136] J. F. Dobson, *Int. J. Quant. Chem.* **2014**, *114*, 1157–1161.

[137] M. R. Kennedy, A. R. McDonald, A. E. DePrince, M. S. Marshall, R. Podeszwa, C. D. Sherrill, *J. Chem. Phys.* **2014**, *140*, 121104.

[138] B. M. Axilrod, E. Teller, *J. Chem. Phys.* **1943**, *11*, 299–300.

[139] Y. Muto, *Proc. Phys. Math. Soc. Jpn.* **1944**, *17*, 629.

[140] R. F. Ribeiro, A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2011**, *115*, 14556–14562.

[141] Y. Zhao, D. G. Truhlar, *Phys. Chem. Chem. Phys.* **2008**, *10*, 2813–2818.

[142] C. J. Cramer, D. G. Truhlar, *Chem. Rev.* **1999**, *99*, 2161–2200.

[143] A. Klamt, *WIREs Comput. Mol. Sci.* **2011**, *1*, 699–709.

[144] R. Sedlak, T. Janowski, M. Pitoňák, J. Řezáč, P. Pulay, P. Hobza, *J. Chem. Theory Comput.* **2013**, *9*, 3364–3374.

[145] H. Kruse, L. Goerigk, S. Grimme, *J. Org. Chem.* **2012**, *77*, 10824–10834.

[146] L. Goerigk, C. A. Collyer, J. R. Reimers, *J. Phys. Chem. B* **2014**, *118*, 14612–26.

[147] P. Hohenberg, W. Kohn, *Physical Review* **1964**, *136*, B864–B871.

[148] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133–A1138.

[149] Y. Zhao, D. G. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 157–167.

[150] R. Peverati, D. G. Truhlar, *Phil. Trans. R. Soc. A* **2014**, *372*, 20120476.

[151] M. Bühl, C. Reimann, D. A. Pantazis, T. Bredow, F. Neese, *J. Chem. Theory Comput.* **2008**, *4*, 1449–1459.

[152] S. Grimme, M. Steinmetz, *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031–16042.

[153] T. Risthaus, M. Steinmetz, S. Grimme, *J. Comp. Chem.* **2014**, *35*, 1509–1516.

[154] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.

[155] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.

[156] Scifinder 2015, Chemical Abstracts Service: Columbus, OH, https://scifinder.cas.org (accessed June 2, 2015).

[157] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *The Journal of Chemical Physics* **1980**, *72*, 650.

[158] S. L. Price, *Chem. Soc. Rev.* **2014**, *43*, 2098–2111.

[159] M. A. Neumann, F. J. J. Leusen, J. Kendrick, *Angew. Chem. Int. Ed.* **2008**, *47*, 2427–2430.

[160] C. C. Pantelides, C. S. Adjiman, A. V. Kazantsev, *Top Curr Chem* **2014**, *345*, 25–58.

[161] S. Grimme, J. Antony, T. Schwabe, C. Mück-Lichtenfeld, *Org. Biomol. Chem.* **2007**, *5*, 741–758.

[162] E. R. Johnson, I. D. Mackie, G. A. DiLabio, *J. Phys. Org. Chem.* **2009**, *22*, 1127–1135.

[163] K. Müller-Dethlefs, P. Hobza, *Chem. Rev.* **2000**, *100*, 143–167.

[164] Y. Ikabata, H. Nakai, *Int. J. Quant. Chem.* **2015**, *115*, 309–324.

[165] J. P. Wagner, P. R. Schreiner, *Angew. Chem. Int. Ed.* **2015**, *54*, 12274–12296.

[166] S. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553–566.

[167] L. M. Mentel, E. J. Baerends, *J. Chem. Theory Comput.* **2014**, *10*, 252–267.

[168] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, J. H. van Lenthe, *Chem. Rev.* **1994**, *94*, 1873–1885.

[169] M. Gutowski, G. Chałasiński, *J. Chem. Phys.* **1993**, *98*, 5540.

[170] J. Katriel, E. R. Davidson, *Proc. Natl. Acad. Sci.* **1980**, *77*, 4403–4406.

[171] C.-O. Almbladh, U. von Barth, *Phys. Rev. B* **1985**, *31*, 3231–3244.

[172] J. Řezáč, K. E. Riley, P. Hobza, *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.

[173] L. F. Holroyd, T. van Mourik, *Chem. Phys. Lett.* **2007**, *442*, 42–46.

[174] H. Valdés, V. Klusák, M. Pitonák, O. Exner, I. Starý, P. Hobza, L. Rulísek, *J. Comput. Chem.* **2008**, *29*, 861–70.

[175] T. van Mourik, P. G. Karamertzanis, S. L. Price, *J. Phys. Chem. A* **2006**, *110*, 8–12.

[176] T. Helgaker, J. Gauss, P. Jorgensen, J. Olsen, *J. Chem. Phys.* **1997**, *106*, 6430.

[177] S. Grimme, J. G. Brandenburg, C. Bannwarth, A. Hansen, *J. Chem. Phys.* **2015**, *143*, 054107.

[178] Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2008**, *4*, 1849–1868.

[179] O. A. Vydrov, T. Van Voorhis, *J. Chem. Phys.* **2010**, *133*, 244103.

[180] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist, *Phys. Rev. Lett.* **2004**, *92*, 246401.

[181] J. Klimeš, D. R. Bowler, A. Michaelides, *J. Phys.: Condens. Matter* **2010**, *22*, 022201.

[182] A. Heßelmann, G. Jansen, *Chem. Phys. Lett.* **2002**, *357*, 464–470.

[183] A. Heßelmann, G. Jansen, *Chem. Phys. Lett.* **2002**, *362*, 319–325.

[184] I. D. Mackie, G. A. Dilabio, *Phys. Chem. Chem. Phys.* **2010**, *12*, 6092–8.

[185] E. Torres, G. A. DiLabio, *J. Phys. Chem. Lett.* **2012**, *3*, 1738–1744.

[186] G. A. DiLabio, M. Koleini, E. Torres, *Theor. Chem. Acc.* **2013**, *132*, 1389.

[187] O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, D. Sebastiani, *Phys. Rev. Lett.* **2004**, *93*, 153004.

[188] O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, D. Sebastiani, *Phys. Rev. B* **2005**, *71*, 195119.

[189] L. Goerigk, H. Kruse, S. Grimme, *ChemPhysChem.* **2011**, *12*, 3421–33.

[190] J. C. Faver, M. L. Benson, X. He, B. P. Roberts, B. Wang, M. S. Marshall, M. R. Kennedy, C. D. Sherrill, K. M. Merz, *J. Chem. Theory Comput.* **2011**, *7*, 790–797.

[191] M. S. Marshall, L. A. Burns, C. D. Sherrill, *J. Chem. Phys.* **2011**, *135*, 194102.

[192] T. Risthaus, S. Grimme, *J. Chem. Theory Comp.* **2013**, *9*, 1580–1591.

[193] J. A. van Santen, G. A. DiLabio, *J. Phys. Chem. A* **2015**, *119*, 6703–13.

[194] J. P. Perdew, A. Zunger, *Phys. Rev. B* **1981**, *23*, 5048–5079.

[195] B. Baumeier, P. Krüger, J. Pollmann, *Phys. Rev. B* **2006**, *73*, 195205.

[196] F. Jensen, *J. Chem. Theory Comput.* **2010**, *6*, 100–106.

[197] A. Galano, J. R. Alvarez-Idaboy, *J. Comput. Chem.* **2006**, *27*, 1203–1210.

[198] T. Yoshida, T. Hayashi, A. Mashima, H. Chuman, *Bioorg. Med. Chem. Lett.* **2015**,

25, 4179–4184.

[199] P. Jurecka, J. Šponer, J. Černy, P. Hobza, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–193.

[200] L. Goerigk, J. R. Reimers, *J. Chem. Theory Comput.* **2013**, *9*, 3240–3251.

[201] H. Valdes, K. Pluháckova, M. Pitonák, J. Rezác, P. Hobza, *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747–57.

[202] H. J. Kulik, N. Luehr, I. S. Ufimtsev, T. J. Martinez, *J. Phys. Chem. B* **2012**, *116*, 12501–12509.

[203] J. G. Brandenburg, M. Alessio, B. Civalleri, M. F. Peintinger, T. Bredow, S. Grimme, *J. Phys. Chem. A* **2013**, *117*, 9282–92.

[204] A. M. Reilly, A. Tkatchenko, *J. Chem. Phys.* **2013**, *139*, 024705.

[205] R. Sure, S. Grimme, *J. Comput. Chem.* **2013**, *34*, 1672–1685.

[206] R. Sure, S. Grimme, *J. Chem. Theory Comput.* **2015**, *11*, 3785–3801.

[207] J. G. Brandenburg, T. Maas, S. Grimme, *J. Chem. Phys.* **2015**, *142*, 124104.

[208] V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin, W. A. Goddard, *J. Chem. Theory Comput.* **2009**, *5*, 1016–1026.

[209] T. Anacker, J. Friedrich, *J. Comput. Chem.* **2014**, *35*, 634–43.

[210] A. Otero-de-la Roza, E. R. Johnson, *J. Chem. Phys.* **2012**, *137*, 054103.

[211] A. Hansen, F. Neese, S. Grimme.

[212] U. R. Fogueri, S. Kozuch, A. Karton, J. M. L. Martin, *J. Phys. Chem. A* **2013**, *117*, 2269–2277.

[213] J. Witte, M. Goldey, J. B. Neaton, M. Head-Gordon, *J. Chem. Theory Comput.* **2015**, *11*, 1481–1492.

[214] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, C. D. Sherrill, *J. Chem. Phys.* **2010**, *132*, 144104.

[215] J. G. Brandenburg, S. Grimme, *Top. Curr. Chem.* **2014**, *345*, 1–23.

[216] S. Block, C. E. Weir, G. J. Piermarini, *Science* **1970**, *169*, 586–587.

[217] B. P. van Eijck, A. L. Spek, W. T. M. Mooij, J. Kroon, *Acta Crystallogr. B* **1998**, *54*, 291–299.

[218] W. B. Schweizer, J. D. Dunitz, *J. Chem. Theory Comput.* **2006**, *2*, 288–291.

[219] A. Ringer, C. Sherrill, *Chem. Eur. J.* **2008**, 2542–2547.

[220] O. Bludský, M. Rubeš, P. Soldán, *Phys. Rev. B* **2008**, *77*, 092103.

[221] S. Wen, G. J. O. Beran, *J. Chem. Theory Comput.* **2011**, *7*, 3733–3742.

[222] R. Podeszwa, B. M. Rice, K. Szalewicz, *Phys. Rev. Lett.* **2008**, *101*, 115503.

[223] D. J. Carter, A. L. Rohl, *J. Chem. Theory Comput.* **2014**, *10*, 3423–3437.

[224] G. J. O. Beran, K. Nanda, *J. Phys. Chem. Lett.* **2010**, *1*, 3480–3487.

[225] J. Moellmann, S. Grimme, *J. Phys. Chem. C* **2014**, *118*, 7615–7621.

[226] G. Raabe, *Z. Naturforsch. A* **2004**, *59a*, 609.

[227] J. G. Brandenburg, S. Grimme, *J. Phys. Chem. Lett.* **2014**, *5*, 1785–1789.

[228] J. G. Brandenburg, M. Hochheim, T. Bredow, S. Grimme, *J. Phys. Chem. Lett.* **2014**, *5*, 4275–4284.

[229] P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953.

[230] G. Kresse, J. Joubert, *Phys. Rev. B* **1999**, *59*, 1758.

[231] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, G. K.-L. Chan, *Science* **2014**, *345*, 640–643.

[232] A. Klamt, *WIREs Comput. Mol. Sci.* **2011**, *1*, 699–709.

[233] F. Neese, "ORCA: An ab initio, density functional and semiempirical program package", Version 3.0 (Current Development Version), Max Planck Institute for Chemical Energy Conversion, Germany, **2014**.

[234] F. Neese, ORCA - an ab initio, density functional and semiempirical program package, Ver. 2.9 (Rev 0), Max Planck Institute for Chemical Energy Conversion, Germany, 2011.

[235] R. Ahlrichs, M. K. Armbruster, M. Bär, H.-P. Baron, R. Bauernschmitt, N. Crawford, P. Deglmann, M. Ehrig, K. Eichkorn, S. Elliott, F. Furche, F. Haase, M. Häser, C. Hättig, A. Hellweg, H. Horn, C. Huber, U. Huniar, M. Kattannek, C. Kölmel, M. Kollwitz, K. May, P. Nava, C. Ochsenfeld, H. Öhm, H. Patzelt, D. Rappoport, O. Rubner, A. Schäfer, U. Schneider, M. Sierka, O. Treutler, B. Unterreiner, M. von Arnim, F. Weigend, P. Weis, H. Weiss, TURBOMOLE 7.0, Universität Karlsruhe **2015**. See also: http://www.turbomole.com.

[236] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, F. Weigend, *WIREs Comput. Mol. Sci.* **2014**, *4*, 91–100.

[237] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–305.

[238] F. Weigend, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–65.

[239] O. Treutler, R. Ahlrichs, *J. Chem. Phys.* **1995**, *102*, 346–354.

[240] R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. De La Pierre, P. D'Arco, Y. Noël, M. Causà, M. Rérat, B. Kirtman, *Int. J. Quantum Chem.* **2014**, *114*, 1287–1317.

[241] See http://www.thch.uni-bonn.de/tc/?section=downloads&lang=english.

[242] J. Antony, S. Grimme, *J. Comput. Chem.* **2012**, *33*, 1730–1739.

[243] Y. Zhang, W. Yang, *J. Chem. Phys.* **1998**, *109*, 2604–2608.

[244] O. Gritsenko, B. Ensing, P. R. T. Schipper, E. J. Baerends, *J. Phys. Chem. A* **2000**, *104*, 8558–8565.

[245] S. Grimme, W. Hujo, B. Kirchner, *Phys. Chem. Chem. Phys.* **2012**, *14*, 4875–4883.

[246] E. Rudberg, *J. Phys.: Condens. Matter* **2012**, *24*, 072202.

[247] M. Gaus, A. Goez, M. Elstner, *J. Chem. Theory Comput.* **2013**, *9*, 338–354.

[248] *Computational Methods for Large Systems*, (Ed.: J. R. Reimers), Wiley: New Jersy, U.S., **2011**.

[249] M. Korth, *Chem. Phys. Chem* **2011**, *12*, 3131–3142.

[250] L. Goerigk, H. Kruse, S. Grimme, *Chem. Phys. Chem* **2011**, *12*, 3421–3433.

[251] E. D. Murray, K. Lee, D. C. Langreth, *J. Chem. Theory Comput.* **2009**, *5*, 2754–2762.

[252] J. A. Pople, *Modern Theoretical Chemistry*, Plenum: New York, U.S., 4th ed., **1976**.

[253] E. R. Davidson, D. Feller, *Chem. Rev.* **1986**, *86*, 681–696.

[254] W. Kołos, *Theor. Chim. Acta* **1979**, *51*, 219–240.

[255] A. Bauza, D. Quinonero, P. M. Deya, A. Frontera, *Phys. Chem. Chem. Phys.* **2012**, *14*, 14061–14066.

[256] A. Antony, C. Hakanoglu, A. Asthagiri, J. F. Weaver, *J. Chem. Phys.* **2012**, *136*, 054702.

[257] J. Granatier, M. Pitoň]ák, P. Hobza, *J. Chem. Theory Comp.* **2012**, *8*, 2282–2292.

[258] H. Tatewaki, S. Huzinaga, *J. Comput. Chem.* **1980**, *1*, 205–228.

[259] A. Schäfer, H. Horn, R. Ahlrichs, *J. Chem. Phys.* **1992**, *97*, 2571–2577.

[260] K. A. Peterson, D. Figgen, E. Goll, H. Stoll, M. Dolg, *J. Chem. Phys.* **2003**, *119*, 11113–11123.

[261] TURBOMOLE 6.4: R. Ahlrichs, M. K. Armbruster, M. Bär, H.–P. Baron, R. Bauernschmitt, N. Crawford, P. Deglmann, M. Ehrig, K. Eichkorn, S. Elliott, F. Furche, F. Haase, M. Häser, C. Hättig, A. Hellweg, H. Horn, C. Huber, U. Huniar, M. Kattannek, C. Kölmel, M. Kollwitz, K. May, P. Nava, C. Ochsenfeld, H. Öhm, H. Patzelt, D. Rappoport, O. Rubner, A. Schäfer, U. Schneider, M. Sierka, O. Treutler, B. Unterreiner, M. von Arnim, F. Weigend, P. Weis and H. Weiss. Universität Karlsruhe 2012. See also: http://www.turbomole.com.

[262] M. Korth, M. Pitoňák, J. Rezáč, P. Hobza, *J. Chem. Theory Comput.* **2010**, *6*,

344–352.

[263] J. J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, 2012. See http://OpenMOPAC.net.

[264] F. Eckert and A. Klamt, COSMOtherm, Version C2.1, Release 01.11; COSMOlogic GmbH & Co. KG, Leverkusen, Germany, 2010.

[265] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822–8824.

[266] A. Schäfer, C. Huber, R. Ahlrichs, *J. Chem. Phys.* **1994**, *100*, 5829–5835.

[267] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[268] E. A. Coutsias, C. Seok, K. A. Dill, *J. Comput. Chem.* **2004**, *25*, 1849–1857.

[269] D. Reha, H. Valdés, J. Vondrásek, P. Hobza, A. Abu-Riziq, B. Crews, M. S. de Vries, *Chem. Eur. J.* **2005**, *11*, 6803–6817.

[270] D. Gruzman, A. Karton, J. M. L. Martin, *J. Phys. Chem. A* **2009**, *113*, 11974–11983.

[271] G. I. Csonka, A. D. French, G. P. Johnson, C. A. Stortz, *J. Chem. Theory Comput.* **2009**, *5*, 679–692.

[272] J. J. Wilke, M. C. Lind, H. F. Schaefer, A. G. Császár, W. D. Allen, *J. Chem. Theory Comput.* **2009**, *5*, 1511–1523.

[273] J. Řezáč, K. E. Riley, P. Hobza, *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.

[274] J. Řezáč, J. Fanfrlik, D. Salahub, P. Hobza, *J. Chem. Theory Comp.* **2009**, *5*, 1749–1760.

[275] M. Kamieth, U. Burkert, P. S. Corbin, S. J. Dell, S. C. Zimmerman, F.-G. Klärner, *Eur. J. Org. Chem.* **1999**, 2741–2749.

[276] J. Graton, J.-Y. Le Questel, B. Legouin, P. Uriac, P. van de Weghe, D. Jacquemin, *Chem. Phys. Lett.* **2012**, *522*, 11–16.

[277] C. Mück-Lichtenfeld, S. Grimme, L. Kobryn, A. Sygula, *Phys. Chem. Chem. Phys.* **2010**, *12*, 7091–7097.

[278] C. Allott, H. Adams, C. A. Hunter, J. A. Thomas, P. L. Bernad Jr., C. Rotger, *Chem. Commun.* **1998**, 2449–2450.

[279] W. L. Mock, N. Y. Shih, *J. Am. Chem. Soc.* **1989**, *111*, 2697–2699.

[280] S. Moghaddam, C. Yang, M. Rekharsky, Y. H. Ko, K. Kim, Y. Inoue, M. K. Gilson, *J. Am. Chem. Soc.* **2011**, *133*, 3570–3581.

[281] C. Mück-Lichtenfeld, S. Grimme, L. Kobryn, A. Sygula, *Phys. Chem. Chem. Phys.* **2010**, *12*, 7091–7097.

[282] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, *J. Mol. Biol.* **1977**, *112*, 535–542.

[283] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[284] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Cryst. D* **2010**, *66*, 12–21.

[285] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, D. C. Richardson, *Nucleic Acids Res.* **2007**, *35*, W375–W383.

[286] I. W. Davis, L. W. Murray, J. S. Richardson, D. C. Richardson, *Nucleic Acids Res.* **2004**, *32*, W615–W619.

[287] S. Kristyán, P. Pulay, *Chem. Phys. Lett.* **1994**, *229*, 175–180.

[288] P. Hobza, J. Šponer, T. Reschel, *J. Comput. Chem.* **1995**, *16*, 1315–1325.

[289] M. J. Allen, D. J. Tozer, *J. Chem. Phys.* **2002**, *117*, 11113–11120.

[290] Y. Cho, W. J. Cho, I. S. Youn, G. Lee, N. J. Singh, K. S. Kim, *Acc. Chem. Res.* **2014**, *47*, DOI: 10.1021/ar400326q.

[291] C. Corminboeuf, *Acc. Chem. Res.* **2014**, *47*, DOI: 10.1021/ar400303a.

[292] L. Kronik, A. Tkatchenko, *Acc. Chem. Res.* **2014**, *47*, DOI: 10.1021/ar500144s.

[293] A. Heßelmann, *J. Chem. Phys.* **2012**, *136*, 014104.

[294] O. A. Vydrov, T. Van Voorhis, *J. Chem. Theory Comput.* **2012**, *8*, 1929–1934.

[295] A. T. Fenley, N. M. Henriksen, H. S. Muddana, M. K. Gilson, *J. Chem. Theory Comput.* **2014**, *10*, 4069–4078.

[296] G. Ohlendorf, C. W. Mahler, S.-S. Jester, G. Schnakenburg, S. Grimme, S. Höger, *Angew. Chem. Int. Ed.* **2013**, 12086–12090.

[297] A. Tkatchenko, R. A. DiStasio, R. Car, M. Scheffler, *Phys. Rev. Lett.* **2012**, *108*, 236402.

[298] A. Otero-de-la Roza, E. R. Johnson, *J. Chem. Phys.* **2013**, *138*, 054103.

[299] A. Ambrosetti, D. Alfè, R. A. DiStasio, A. Tkatchenko, *J. Phys. Chem. Lett.* **2014**, *5*, 849–855.

[300] M. Elstner, *J. Phys. Chem. A* **2007**, *111*, 5614–5621.

[301] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim,

S. Suhai, G. Seifert, *Phys. Rev. B* **1998**, *58*, 7260–7268.

[302] J. Z. Vilseck, J. Tirado-Rives, W. L. Jorgensen, *J. Chem. Theory Comput.* **2014**, *10*, 2802–2812.

[303] T. Kloss, J. Heil, S. M. Kast, *J. Phys. Chem. B* **2008**, *112*, 4337–4343.

[304] F. Hoffgaard, J. Heil, S. M. Kast, *J. Chem. Theory Comput.* **2013**, *9*, 4718–4726.

[305] W. Iali, P. Petrovic, M. Pfeffer, S. Grimme, J.-P. Djukic, *Dalton Trans.* **2012**, *41*, 12233–12243.

[306] R. Sure, J. Antony, S. Grimme, *J. Phys. Chem. B* **2014**, *118*, 3431–3440.

[307] J. Graton, J.-Y. L. Questel, B. Legouin, P. Uriac, P. van de Weghe, D. Jacquemin, *Chem. Phys. Lett.* **2012**, *522*, 11–16.

[308] J. Graton, B. Legouin, F. Besseau, P. Uriac, J.-Y. Le Questel, P. van de Weghe, D. Jacquemin, *J. Phys. Chem. C* **2012**, *116*, 23067–23074.

[309] Y. Zhao, D. G. Truhlar, *Phys. Chem. Chem. Phys.* **2008**, *10*, 2813–2818.

[310] M. Yanney, A. Sygula, *Tetrahedron Lett.* **2013**, *54*, 2604–2607.

[311] T. Iwamoto, Y. Watanabe, T. Sadahiro, T. Haino, S. Yamago, *Angew. Chem. Int. Ed.* **2011**, *50*, 8342–8344.

[312] T. Iwamoto, Y. Watanabe, H. Takaya, T. Haino, N. Yasuda, S. Yamago, *Chem. Eur. J.* **2013**, *19*, 14061–14068.

[313] N. Mardirossian, M. Head-Gordon, *J. Chem. Theory Comput.* **2013**, *9*, 4453–4461.

[314] J. Kessler, M. Jakubek, B. Dolenský, P. Bouř, *J. Comput. Chem.* **2012**, *33*, 2310–2317.

[315] P. Mikulskis, D. Cioloboc, M. Andrejić, S. Khare, J. Brorsson, S. Genheden, R. Mata, P. Söderhjelm, U. Ryde, *J. Comput.-Aided Mol. Des.* **2014**, *28*, 375–400.

[316] G. Casella, A. Bagno, G. Saielli, *Phys. Chem. Chem. Phys.* **2013**, *15*, 18030–18038.

[317] R. V. Pinjari, S. P. Gejji, *J. Phys. Chem. A* **2008**, *112*, 12679–12686.

[318] R. V. Pinjari, S. P. Gejji, *J. Phys. Chem. A* **2010**, *114*, 2338–2343.

[319] M. Sundararajan, V. Sinha, T. Bandyopadhyay, S. K. Ghosh, *J. Phys. Chem. A* **2012**, *116*, 4388–4395.

[320] M. Sundararajan, *J. Phys. Chem. B* **2013**, *117*, 13409–13417.

[321] M. Shaikh, J. Mohanty, M. Sundararajan, A. C. Bhasikuttan, H. Pal, *J. Phys. Chem. B* **2012**, *116*, 12450–12459.

[322] M. L. McKee, *J. Phys. Chem. A* **2013**, *117*, 2365–2372.

[323] S. M. Bachrach, *J. Phys. Chem. A* **2013**, *117*, 8484–8491.

[324] D. Benitez, E. Tkatchouk, I. Yoon, J. F. Stoddart, W. A. Goddard, *J. Am. Chem. Soc.* **2008**, *130*, 14928–14929.

[325] M. Juriíĉek, J. C. Barnes, E. J. Dale, W.-G. Liu, N. L. Strutt, C. J. Bruns, N. A. Vermeulen, K. C. Ghooray, A. A. Sarjeant, C. L. Stern, Y. Y. Botros, W. A. Goddard, J. F. Stoddart, *J. Am. Chem. Soc.* **2013**, *135*, 12736–12746.

[326] Q. Zhang, T. Pankewitz, S. Liu, W. Klopper, L. Gan, *Ang. Chem. Int. Ed.* **2010**, *49*, 9935–9938.

[327] O. Hampe, T. Karpuschkin, M. Vonderach, P. Weis, Y. Yu, L. Gan, W. Klopper, M. M. Kappes, *Phys. Chem. Chem. Phys.* **2011**, *13*, 9818–9823.

[328] A. Dieckmann, K. N. Houk, *J. Chem. Theory Comput.* **2012**, *8*, 5064–5071.

[329] A. Dieckmann, K. N. Houk, *Chem. Sci.* **2013**, *4*, 3591–3600.

[330] G. Haberhauer, A. Pintér, S. Woitschetzki, *Nat. Commun.* **2013**, *4*, 2945.

[331] G. Haberhauer, S. Woitschetzki, H. Bandmann, *Nat. Commun.* **2014**, *5*, 3542.

[332] A. Tkatchenko, D. Alfè, K. S. Kim, *J. Chem. Theory Comput.* **2012**, *8*, 4317–4322.

[333] A. Ambrosetti, D. Alfè, R. A. DiStasio, A. Tkatchenko, *J. Phys. Chem. Lett.* **2014**, *5*, 849–855.

[334] A. Ambrosetti, A. M. Reilly, R. A. DiStasio, A. Tkatchenko, *J. Chem. Phys.* **2014**, *140*, 18A508.

[335] A. Heßelmann, T. Korona, *J. Chem. Phys.* **2014**, *141*, 094107.

[336] P. A. Denis, *Chem. Phys. Lett.* **2011**, *516*, 82–87.

[337] P. A. Denis, *Chem. Phys. Lett.* **2014**, *591*, 323–327.

[338] P. A. Denis, *RSC Adv.* **2013**, *3*, 25296–25305.

[339] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999–3094.

[340] B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *106*, 5151–5158.

[341] B. Mennucci, E. Cancès, J. Tomasi, *J. Phys. Chem. B* **1997**, *101*, 10506–10517.

[342] R. Cammi, B. Mennucci, J. Tomasi, *J. Phys. Chem. A* **1999**, *103*, 9100–9108.

[343] R. Cammi, B. Mennucci, J. Tomasi, *J. Phys. Chem. A* **2000**, *104*, 5631–5637.

[344] M. Cossi, G. Scalmani, N. Rega, V. Barone, *J. Chem. Phys.* **2002**, *117*, 43–54.

[345] R. Joseph, E. Masson, *Eur. J. Org. Chem.* **2014**, *2014*, 105–110.

[346] F.-G. Klärner, B. Kahlert, *Acc. Chem. Res.* **2003**, *36*, 919–932.

[347] T. Chang, A. M. Heiss, S. J. Cantrill, M. C. T. Fyfe, A. R. Pease, S. J. Rowan, J. F. Stoddart, D. J. Williams, *Org. Lett.* **2000**, *2*, 2943–2946.

[348] S. Moghaddam, C. Yang, M. Rekharsky, Y. H. Ko, K. Kim, Y. Inoue, M. K. Gilson,

J. Am. Chem. Soc. **2011**, *133*, 3570–3581.

[349] M. K. Gilson, K. K. Irikura, *J. Phys. Chem. B* **2010**, *114*, 16304–16317.

[350] J. Liu, C. P. Kelly, A. C. Goren, A. V. Marenich, C. J. Cramer, D. G. Truhlar, C.-G. Zhan, *J. Chem. Theory Comput.* **2010**, *6*, 1109–1117.

[351] D. Feller, *J. Chem. Phys.* **1992**, *96*, 6104–6114.

[352] D. Feller, *J. Chem. Phys.* **1993**, *98*, 7059–7071.

[353] S. Zhong, E. C. Barnes, G. A. Petersson, *J. Chem. Phys.* **2008**, *129*, 184116.

[354] F. Neese, E. F. Valeev, *J. Chem. Theory Comput.* **2011**, *7*, 33–43.

[355] Private communication with M. Elstner, 2014.

[356] F. Eckert, A. Klamt, COSMOtherm, Version C3.0, Release 14.01; COSMOlogic GmbH & Co. KG, Leverkusen, Germany, 2014.

[357] T. Williams, C. Kelley et. al, Gnuplot 4.4: an interactive plotting program, 2011. See also: http://www.gnuplot.info/.

[358] L. Cao, M. Sekutor, P. Y. Zavalij, K. Mlinarić-Majerski, R. Glaser, L. Isaacs, *Angew. Chem. Int. Ed.* **2014**, *53*, 988–93.

[359] L. Cao, L. Isaacs, *Supramol. Chem.* **2014**, *26*, 251–258.

[360] C. L. D. Gibb, B. C. Gibb, *J. Am. Chem. Soc.* **2004**, *126*, 110408–11409.

[361] S. Liu, B. C. Gibb, *Chem. Commun.* **2008**, 3709–3716.

[362] C. L. D. Gibb, B. C. Gibb, *J. Comput.-Aided Mol. Design* **2014**, *28*, 319–325.

[363] J. Ewell, B. C. Gibb, S. W. Rick, *J. Phys. Chem. B* **2008**, *112*, 10272–10279.

[364] J. Clayden, N. Greeves, S. Warren, P. Wothers, *Organic Chemistry*, Oxford University Press: Oxford, U.K., **2001**.

[365] P. Mikulskis, D. Cioloboc, M. Adrejić, S. Khare, J. Brorsson, S. Genheden, R. Mata, P. Söderhjelm, U. Ryde, *J. Comput.-Aided Mol. Design* **2014**, *28*, 375–400.

[366] J. Pérez-Jordá, A. Becke, *Chem. Phys. Lett.* **1995**, *233*, 134–137.

[367] J. Antony, R. Sure, S. Grimme, *Chem. Comm.* **2015**, *51*, 1764–1774.

[368] A. Heß elmann, T. Korona, *J. Chem. Phys.* **2014**, *141*, 094107.

[369] A. Tkatchenko, *Adv. Funct. Mater.* **2014**, *25*, 1054–2061.

[370] J. Moellmann, S. Grimme, *J. Phys. Chem. C* **2014**, *118*, 7615–7621.

[371] Z.-W. Qu, A. Hansen, S. Grimme, *J. Chem. Theory Comput.* **2015**, *11*, 1037–1045.

[372] C. Bannwarth, A. Hansen, S. Grimme, *Isr. J. Chem.* **2015**, *55*, 235–242.

[373] A. Hansen, C. Bannwarth, S. Grimme, P. Petrović, C. Werlé, J.-P. Djukic, *ChemistryOpen* **2014**, *3*, 177–189.

[374] C. N. Nguyen, A. Cruz, M. K. Gilson, T. Kurtzman, *J. Chem. Theory Comput.* **2014**, *10*, 2769–2780.

[375] H. S. Muddana, A. T. Fenley, D. L. Mobley, M. K. Gilson, *J. Comput.-Aided Mol. Des.* **2014**, *28*, 305–17.

[376] A. Petitjean, R. G. Khoury, N. Kyritsakas, J.-M. Lehn, *J. Am. Chem. Soc.* **2004**, *126*, 6637–6647.

[377] T. Kawase, Y. Nishiyama, T. Nakamura, T. Ebi, K. Matsumoto, H. Kurata, M. Oda, *Angew. Chem. Int. Ed.* **2007**, *46*, 1086–8.

[378] P. E. Georghiou, A. H. Tran, S. Mizyed, M. Bancu, L. T. Scott, *J. Org. Chem.* **2005**, *70*, 6158–63.

[379] J. Hornung, D. Fankhauser, L. D. Shirtcliff, A. Praetorius, W. B. Schweizer, F. Diederich, *Chem. Eur. J.* **2011**, *17*, 12362–71.

[380] S. H. Jungbauer, D. Bulfield, F. Kniep, C. W. Lehmann, E. Herdtweck, S. M. Huber, *J. Am. Chem. Soc.* **2014**, *136*, 16740–3.

[381] Y. Liu, E.-C. Yang, Y.-W. Yang, H.-Y. Zhang, Z. Fan, F. Ding, R. Cao, *J. Org. Chem.* **2004**, *69*, 173–80.

[382] B. A. Blight, C. A. Hunter, D. A. Leigh, H. McNab, P. I. T. Thomson, *Nat. Chem.* **2011**, *3*, 244–48.

[383] M. Juríček, J. C. Barnes, E. J. Dale, W.-G. Liu, N. L. Strutt, C. J. Bruns, N. A. Vermeulen, K. C. Ghooray, A. A. Sarjeant, C. L. Stern, Y. Y. Botros, W. A. Goddard, J. F. Stoddart, *J. Am. Chem. Soc.* **2013**, *135*, 12736–46.

[384] K.-C. Chang, T. Minami, P. Koutnik, P. Y. Savechenkov, Y. Liu, P. Anzenbacher, *J. Am. Chem. Soc.* **2014**, *136*, 1520–5.

[385] B. L. Feringa, *Molecular Switches*, Wiley-VCH: Weinheim, Germany, 2nd ed., **2011**.

[386] D. Ugarte, *Nature* **1992**, *359*, 707–9.

[387] B. W. Smith, M. Monthioux, D. E. Luzzi, *Nature* **1998**, *396*, 323–324.

[388] H. W. Kroto, K. McKay, *Nature* **1988**, *331*, 328–331.

[389] T. Kawase, K. Tanaka, N. Fujiwara, H. R. Darabi, M. Oda, *Angew.Chem. Int. Ed.* **2003**, *42*, 1624–8.

[390] L. P. Wolters, P. Schyman, M. J. Pavan, W. L. Jorgensen, F. M. Bickelhaupt, , S. Kozuch, *WIREs Comput. Mol. Sci.* **2014**, *4*, 523–540.

[391] K. E. Riley, P. Hobza, *Phys. Chem. Chem. Phys.* **2013**, *15*, 17742–51.

[392] S. Kozuch, J. M. L. Martin, *J. Chem. Theory Comput.* **2013**, *9*, 1918–1931.

[393] M. V. Rekharsky, Y. Inoue, *Chem. Rev.* **1998**, *98*, 1875–1918.

[394] D. Benitez, E. Tkatchouk, I. Yoon, J. F. Stoddart, W. A. Goddard, *J. Am. Chem. Soc.* **2008**, *130*, 14928–9.

[395] W. Iali, P. Petrović, M. Pfeffer, S. Grimme, J.-P. Djukic, *Dalton trans.* **2012**, *41*, 12233–122343.

[396] C. Kind, M. Reiher, J. Neugebauer, B. A. Hess, SNF Version 2.2.1, Universität Erlangen, **2002**.

[397] B. Aradi, B. Hourahine, T. Frauenheim, *J. Phys. Chem. A* **2007**, *111*, 5678–84.

[398] M. Korth, *J. Chem. Theory Comp.* **2010**, *6*, 3808–3816.

[399] J. Řezáč, P. Hobza, *J. Chem. Theory Comput.* **2012**, *8*, 141–151.

[400] *MNDO2005, Version 7.0, W. Thiel, MPI für Kohlenforschung, Mülheim, Germany*.

[401] F. Eckert, A. Klamt, COSMOtherm, Version C3.0, Release 14.01; COSMOlogic GmbH & Co. KG, Leverkusen, Germany, **2014**.

[402] D. Rappoport, F. Furche, *J. Chem. Phys.* **2010**, *133*, 134105.

[403] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. a. De Jong, *Comput. Phys. Commun.* **2010**, *181*, 1477–1489.

[404] J. Liu, C. P. Kelly, A. C. Goren, A. V. Marenich, C. J. Cramer, D. G. Truhlar, C.-G. Zhan, *J. Chem. Theory Comput.* **2010**, *6*, 1109–1117.

[405] A. Ruzsinszky, J. P. Perdew, J. Tao, G. I. Csonka, J. M. Pitarke, *Phys. Rev. Lett.* **2012**, *109*, 233203.

[406] J. P. Perdew, A. Ruzsinszky, J. Sun, S. Glindmeyer, G. I. Csonka, *Phys. Rev. A* **2012**, *86*, 062714.

[407] J. Antony, B. Brüske, S. Grimme, *Phys. Chem. Chem. Phys.* **2009**, *11*, 8440–8447.

[408] A. Li, H. S. Muddana, M. K. Gilson, *J. Chem. Theory Comput.* **2014**, *10*, 1563–1575.

[409] J. Hostaš, J. Řezáč, P. Hobza, *Chem. Phys. Lett.* **2013**, *568-569*, 161–166.

[410] A. Li, H. S. Muddana, M. K. Gilson, *J. Chem. Theory Comput.* **2014**, *10*, 1563–1575.

[411] J. Witte, M. Goldey, J. B. Neaton, M. Head-Gordon, *J. Chem. Theory Comput.* **2015**, *11*, 150317130807007.

[412] A. Gansäuer, D. von Laufenberg, C. Kube, T. Dahmen, A. Michelmann, M. Behlendorf, R. Sure, M. Seddiqzai, S. Grimme, D. V. Sadasivam, G. D. Fianu, R. A. Flowers Ii, *Chem. Eur. J.* **2015**, *21*, 280–289.

[413] S. Grimme, *J. Chem. Phys.* **2013**, *138*, 244104.

[414] C. Bannwarth, S. Grimme, *Comp. Theor. Chem.* **2014**, *1040-1041*, 45–53.

[415] B. M. Trost, *Science* **1991**, *254*, 1471–1477.

[416] B. M. Trost, *Angew. Chem. Int. Ed.* **1995**, *34*, 259–281.

[417] B. M. Trost, *Acc. Chem. Res.* **1996**, *29*, 355–364.

[418] S. Z. Zard, *Radical Reactions in Organic Synthesis*, Oxford University Press: Oxford, U.K., **2003**.

[419] P. Renaud, M. P. Sibi, *Radicals in Organic Synthesis*, Wiley-VCH: Weinheim, Germany, **2001**.

[420] D. P. Curran, N. A. Procter, B. Giese, *Stereochemistry of Radical Reactions*, Wiley-VCH: Weinheim, Germany, **1996**.

[421] U. Jahn in *Radicals in Synthesis III*, *Vol. 320* (Eds.: M. Heinrich, A. Gansäuer), **2012**, pp. 121–189.

[422] U. Jahn in *Radicals in Synthesis III*, *Vol. 320* (Eds.: M. Heinrich, A. Gansäuer), **2012**, pp. 191–322.

[423] U. Jahn in *Radicals in Synthesis III*, *Vol. 320* (Eds.: M. Heinrich, A. Gansäuer), **2012**, pp. 323–451.

[424] H. Matsumoto, T. Nakano, Y. Nagai, *Tetrahedron Lett.* **1973**, *14*, 5147–5150.

[425] L. Quebatte, K. Thommes, K. Severin, *J. Am. Chem. Soc* **2006**, *128*, 7440–7441.

[426] K. Severin, *Chimia* **2012**, *66*, 386–388.

[427] J.-S. Wang, K. Matyjaszewski, *J. Am. Chem. Soc.* **1995**, *117*, 5614–5615.

[428] T. E. Patten, J. Xia, T. Abernathy, K. Matyjaszewski, *Science* **1996**, *272*, 866–868.

[429] K. Matyjaszewski, , J. Xia, *Chem. Rev.* **2001**, *101*, 2921–2990.

[430] D. M. Smith, M. E. Pulling, , J. R. Norton, *J. Am. Chem. Soc.* **2007**, *129*, 770–771.

[431] J. Hartung, M. E. Pulling, D. M. Smith, m. D. X. Yang, J. R. Nortonm, *Tetrahedron* **2008**, *64*, 11822–11830.

[432] G. Li, A. Han, M. E. Pulling, D. P. Estes, J. R. Norton, *J. Am. Chem. Soc.* **2012**, *134*, 14662–14665.

[433] A. Gansäuer, T. Lauterbach, S. Narayan, *Angew. Chem. Int. Ed.* **2003**, *42*, 5556–5573.

[434] J. M. Cuerva, J. Justicia, J. L. Oller-López, J. Oltra in *Radicals in Synthesis II*, *Vol. 264*, **2006**, pp. 63–91.

[435] A. Gansäuer, J. Justicia, C.-A. Fan, D. Worgull, F. Piestert in *Metal Catalyzed Reductive C–C Bond Formation*, *Vol. 279* (Ed.: M. J. Krische), **2007**, pp. 25–52.

[436] A. Gansäuer, A. Fleckhaus, M. A. Lafont, A. Okkel, K. Kotsis, A. Anoop, F. Neese, *J. Am. Chem. Soc.* **2009**, *131*, 16989–16999.

[437] A. Suzuki, *Angew. Chem. Int. Ed.* **2011**, *50*, 6722–6737.

[438] E.-i. Negishi, *Angew. Chem. Int. Ed.* **2011**, *50*, 6738–6764.

[439] A. J. Bard, L. R. Faulkner, *Electrochemical Methods: Fundamentals and Applications*, Wiley: New York, U.S., 2nd ed., **2001**.

[440] A. Jutand, *Chem. Rev.* **2008**, *108*, 2300–2347.

[441] W. E. Geiger, *Coord. Chem. Rev.* **2013**, *257*, 1459–1471.

[442] R. J. Enemærke, G. H. Hjøllund, K. Daasbjerg, T. Skrydstrup, *C. R. Acad. Sci.* **2001**, *4*, 435–438.

[443] R. J. Enemærke, J. Larsen, T. Skrydstrup, K. Daasbjerg, *Organometallics* **2004**, *23*, 1866–1874.

[444] R. J. Enemærke, J. Larsen, T. Skrydstrup, K. Daasbjerg, *J. Am. Chem. Soc* **2004**, *126*, 7853–7864.

[445] R. J. Enemærke, J. Larsen, G. H. Hjøllund, T. Skrydstrup, K. Daasbjerg, *Organometallics* **2005**, *24*, 1252–1262.

[446] J. Larsen, R. J. Enemærke, T. Skrydstrup, K. Daasbjerg, *Organometallics* **2006**, *25*, 2031–2036.

[447] Y. Mugnier, C. Moise, L. E., *J. Organomet. Chem.* **1981**, *204*, 61–66.

[448] S. Edmond, V. Jacques, *Organometallics* **1989**, *8*, 237–241.

[449] A. Gansäuer, A. Barchuk, F. Keller, M. Schmitt, S. Grimme, M. Gerenkamp, C. Mück-Lichtenfeld, K. Daasbjerg, H. Svith, *J. Am. Chem. Soc.* **2007**, *129*, 1359–1371.

[450] R. F. Johnston, R. E. Borjas, J. L. Furilla, *Electrochim. Acta* **1995**, *40*, 473–477.

[451] J. Langmaier, Z. Samec, V. Varga, M. Horáček, K. Mach, *Journal of Organometallic Chemistry* **1999**, *579*, 348–355.

[452] M. Rudolph, S. W. Feldberg, DigiSim, version 3.03, Bioanalytical Systems, Inc.: West Lafayette, IN.

[453] M. Rudolph, *J. Electroanal. Chem.* **1991**, *314*, 13–22.

[454] M. Rudolph, *J. Electroanal. Chem.* **1992**, *338*, 85–98.

[455] E. Cesarotti, H. B. Kagan, R. Goddard, C. Krüger, *J. Organomet. Chem.* **1978**, *162*, 297–309.

[456] A. Gansäuer, H. Bluhm, M. Pierobon, M. Keller, *Organometallics* **2001**, *20*, 914–919.

[457] M. Klahn, P. Arndt, A. Spannenberg, A. Gansäuer, U. Rosenthal, *Organometallics* **2008**, *27*, 5846–5851.

[458] S. L. Hart, D. J. Duncalf, J. J. Hastings, A. McCamley, P. C. Taylor, *J. Chem. Soc. Dalton Trans.* **1996**, 2843–2849.

[459] R. A. Howie, G. P. McQuillan, D. Thompson, *J. Organomet. Chem.* **1984**, *268*, 149–154.

[460] B. G. Conway, M. D. Rausch, *Organometallics* **1985**, *4*, 688–693.

[461] M. D. Rausch, J. F. Lewison, W. P. Hart, *J. Organomet. Chem.* **1988**, *358*, 161–168.

[462] W. C. Finch, E. V. Anslyn, R. H. Grubbs, *J. Am. chem. Soc.* **1988**, *110*, 2406–2413.

[463] M. R. M. Bruce, A. Sclafani, D. R. Tyler, *Inorg. Chem.* **1986**, *25*, 2546–2549.

[464] N. G. Connelly, W. E. Geiger, *Chem. Rev.* **1996**, *96*, 877–910.

[465] C. Hansch, A. Leo, R. W. Taft, *Chem. Rev.* **1991**, *91*, 165–195.

[466] A. Gansäuer, T. Lauterbach, H. Bluhm, M. Noltemeyer, *Angew. Chem. Int. Ed.* **1999**, *38*, 2909–2910.

[467] A. Gansäuer, H. Bluhm, T. Lauterbach, *Adv. Synth. Catal.* **2001**, *343*, 785–787.

[468] A. Gansäuer, H. Bluhm, B. Rinker, S. Narayan, M. Schick, T. Lauterbach, M. Pierobon, *Chem. Eur. J.* **2003**, *9*, 531–542.

[469] A. Gansäuer, C.-A. Fan, F. Keller, J. Keil, *J. Am. Chem. Soc.* **2007**, *129*, 3484–3485.

[470] A. Gansäuer, C.-A. Fan, F. Keller, P. Karbaum, *Chem. Eur. J.* **2007**, *13*, 8084–8090.

[471] A. Gansäuer, L. Shi, M. Otte, *J. Am. Chem. Soc.* **2010**, *132*, 11858–11859.

[472] J. Heinze, *Angew. Chem. Int. Ed.* **1984**, *23*, 831–847.

[473] O. W. Webster, *J. Am. Chem. Soc.* **1966**, *88*, 3046–3050.

[474] A. Gansäuer, C.-A. Fan, F. Piestert, *J. Am. Chem. Soc.* **2008**, *130*, 6916–6917.

[475] J. Streuff, M. Feurer, P. Bichovski, G. Frey, U. Gellrich, *Angew.Chem. Int. Ed.* **2012**, *51*, 8661–8664.

[476] W. A. Nugent, T. V. RajanBabu, *J. Am. Chem. Soc.* **1988**, *110*, 8561–8562.

[477] T. V. RajanBabu, W. A. Nugent, *J. Am. Chem. Soc.* **1989**, *111*, 4525–4527.

[478] T. V. RajanBabu, W. A. Nugent, M. S. Beattie, *J. Am. Chem. Soc.* **1990**, *112*, 6408–6409.

[479] T. V. RajanBabu, W. A. Nugent, *J. Am. Chem. Soc.* **1994**, *116*, 986–997.

[480] A. Gansäuer, M. Pierobon, H. Bluhm, *Angew.Chem. Int. Ed.* **1998**, *37*, 101–103.

[481] A. Gansäuer, H. Bluhm, M. Pierobon, *J. Am. Chem. Soc.* **1998**, *120*, 12849–12859.

[482] A. F. Barrero, A. Rosales, J. M. Cuerva, J. E. Oltra, *Org. Lett.* **2003**, *5*, 1935–1938.

[483] J. Justicia, A. Rosales, E. Buñuel, J. L. Oller-López, M. Valdivia, A. Haïdour, J. E. Oltra, A. F. Barrero, D. J. Cárdenas, J. M. Cuerva, *Chem. Eur. J.* **2004**, *10*, 1778–1788.

[484] J. Justicia, J. E. Oltra, J. M. Cuerva, *J. Org. Chem.* **2004**, *69*, 5803–5806.

[485] J. Justicia, J. L. Oller-López, A. G. Campañna, J. E. Oltra, J. M. Cuerva, E. Buñuel, D. J. Cárdenas, *J. Am. Chem. Soc.* **2005**, *127*, 14911–14921.

[486] A. Gansäuer, J. Justicia, A. Rosales, D. Worgull, B. Rinker, J. M. Cuerva, J. E. Oltra, *Eur. J. Org. Chem.* **2006**, *2006*, 4115–4127.

[487] A. Gansäuer, A. Rosales, J. Justicia, *Synlett* **2006**, 927–929.

[488] J. Justicia, A. Campaña, B. Bazdi, R. Robles, J. Cuerva, J. E. Oltra, *Adv. Synth. Catal.* **2008**, *350*, 571–576.

[489] T. Jiménez, S. P. Morcillo, A. Martín-Lasanta, D. Collado-Sanz, D. J. Cárdenas, A. Gansäuer, J. Justicia, J. M. Cuerva, *Chem. Eur. J.* **2012**, *18*, 12825–12833.

[490] A. Gansäuer, T. Lauterbach, D. Geich-Gimbel, *Chem. Eur. J.* **2004**, *10*, 4983–4990.

[491] J. Friedrich, M. Dolg, A. Gansäuer, D. Geich-Gimbel, T. Lauterbach, *J. Am. Chem. Soc.* **2005**, *127*, 7071–7077.

[492] J. Friedrich, K. Walczak, M. Dolg, F. Piestert, T. Lauterbach, D. Worgull, A. Gansäuer, *J. Am. Chem. Soc.* **2008**, *130*, 1788–1796.

[493] A. Gansäuer, D. Worgull, K. Knebel, I. Huth, G. Schnakenburg, *Angew. Chem. Int. Ed.* **2009**, *48*, 8882–8885.

[494] A. Gansäuer, K. Knebel, C. Kube, M. van Gastel, A. Cangönül, K. Daasbjerg, T. Hangele, M. Hülsen, M. Dolg, J. Friedrich, *Chem. Eur. J.* **2012**, *18*, 2591–2599.

[495] A. Gansäuer, M. Pierobon, *Synlett* **2009**, 1357–1359.

[496] A. Gansäuer, M. Pierobon, H. Bluhm, *Angew. Chem. Int. Ed.* **2002**, *41*, 3206–3208.

[497] J. Y. Cha, J. T. S. Yeoman, S. E. Reisman, *J. Am. Chem. Soc.* **2011**, *133*, 14964–14967.

[498] J. T. S. JYeoman, V. W. Mak, S. E. Reisman, *J. Am. Chem. Soc.* **2013**, *135*, 11764–11767.

[499] F. Vögtle, *Cyclophan-Chemie*, Teubner: Stuttgart, Germany, **1990**.

[500] H. Hopf, R. Gleiter, *Modern Cyclophane Chemistry*, Wiley-VCH: Weinheim, Germany, **2004**.

[501] A. A. Aly, A. B. Brown, *Tetrahedron* **2009**, *65*, 8055–8089.

[502] S. I. on Cyclophanes, *Isr. J. Chem.* **2012**, *52*, 1–192.

[503] H. Wolf, D. Leusser, Mads R V Jørgensen, R. Herbst-Irmer, Y.-S. Chen, E.-W. Scheidt, W. Scherer, B. B. Iversen, D. Stalke, *Chem. Eur. J.* **2014**, *20*, 7048–7053.

[504] S. Grimme, *Chem. Eur. J.* **2004**, *10*, 3423–3429.

[505] S. Grimme, A. Bahlmann in *Modern Cyclophane Chemistry* (Eds.: R. Gleiter, H. Hopf), Wiley-VCH: Weinheim, Germany, **2004**, p. 311.

[506] G. P. Bartholomew, G. C. Bazan, *Acc. Chem. Res.* **2001**, *34*, 30–39.

[507] G. C. Bazan, *J. Org. Chem.* **2007**, *72*, 8615–835.

[508] H. Hopf, *Angew. Chem. Int. Ed.* **2008**, *47*, 9808–9812.

[509] Y. Morisaki, Y. Chujo, *Bull. Chem. Soc. Jap.* **2009**, *82*, 1070–1082.

[510] E. Elacqua, L. R. MacGillivray, *Chem. Eur. J.* **2010**, *2010*, 6883–6894.

[511] Y. Morisaki, Y. Chujo, *Polymer Chemistry* **2011**, *2*, 1249–1257.

[512] A. Marrocchi, I. Tomasi, L. Vaccaro, *Isr. J. Chem.* **2012**, *52*, 41–52.

[513] E. Elacqua, T. Friščić, L. R. MacGillivray, *Isr. J. Chem.* **2012**, *52*, 53–59.

[514] S. E. Gibson, J. D. Knight, *Org. Biomol. Chem.* **2003**, *1*, 1256–1269.

[515] V. Rozenberg, E. Sergeeva, H. Hopf in *Modern Cyclophane Chemistry* (Eds.: R. Gleiter, H. Hopf), Wiley-VCH: Weinheim, Germany, **2004**, p. 435.

[516] S. Bräse in *Asymmetric Synthesis with Chemical and Biological Methods* (Eds.: D. Enders, K.-E. Jaeger, G. Helmchen), Wiley-VCH, Weinheim, Germany, **2007**, p. 196.

[517] S. Bräse in *Asymmetric Synthesis – The Essentials* (Eds.: M. Christenmann, S. Bräse), Wiley-VCH: Weinheim, Germany, 2nd ed., **2008**, p. 67.

[518] G. J. Rowlands, *Isr. J. Chem.* **2012**, *52*, 60–75.

[519] J. Paradies, *Synthesis* **2011**, *2011*, 3749–3766.

[520] J. F. Schneider, R. Fröhlich, J. Paradies, *Isr. J. Chem.* **2012**, *52*, 76–91.

[521] H. J. Reich, D. J. Cram, *J. Am. Chem. Soc.* **1969**, *91*, 3527–3533.

[522] N. V. Vorontsova, V. I. Rozenberg, E. V. Sergeeva, E. V. Vorontsov, Z. A. Starikova, K. A. Lyssenko, H. Hopf, *Chem. Eur. J.* **2008**, *14*, 4600–4617.

[523] H. Allgeier, M. G. Siegel, R. C. Helgeson, E. Schmidt, D. J. Cram, *J. Am. Chem. Soc.* **1975**, *97*, 3782–3789.

[524] L. Bondarenko, I. Dix, H. Hinrichs, H. Hopf, *Synthesis* **2004**, *2004*, 2751–2759.

[525] F. Thiemann, T. Piehler, D. Haase, W. Saak, A. Lützen, *Eur. J. Org. Chem.* **2005**, *2005*, 1991–2001.

[526] U. Kiehne, T. Bruhn, G. Schnakenburg, R. Fröhlich, G. Bringmann, A. Lützen,

Chem. Eur. J. **2008**, *14*, 4246–4255.

[527] C. Benkhäuser-Schunk, B. Wezisla, K. Urbahn, U. Kiehne, J. Daniels, G. Schnakenburg, F. Neese, A. Lützen, *ChemPlusChem* **2012**, *77*, 396–403.

[528] G. Meyer-Eppler, E. Vogelsang, C. Benkhäuser, A. Schneider, G. Schnakenburg, A. Lützen, *Eur. J. Org. Chem.* **2013**, *2013*, 4523–4532.

[529] N. Dalla Favera, U. Kiehne, J. Bunzen, S. Hytteballe, A. Lützen, C. Piguet, *Angew. Chem. Int. Ed.* **2010**, *49*, 125–128.

[530] C. Gütz, R. Hovorka, G. Schnakenburg, A. Lützen, *Chem. Eur. J.* **2013**, *19*, 10890–10894.

[531] C. Klein, C. Gütz, M. Bogner, F. Topić, K. Rissanen, A. Lützen, *Angew. Chem. Int. Ed.* **2014**, *53*, 3739–3742.

[532] G. Meyer-Eppler, F. Topić, G. Schnakenburg, K. Rissanen, A. Lützen, *Eur. J. Inorg. Chem.* **2014**, *2014*, 2495–2501.

[533] S. Müller, B. Liepold, G. J. Roth, H. J. Bestmann, *Synlett* **1996**, *1996*, 521–522.

[534] S. Ohira, *Synth.Commun.* **2006**, *19*, 561–564.

[535] T. Ishiyama, M. Murata, N. Miyaura, *J. Org. Chem.* **1995**, *60*, 7508–7510.

[536] M. W. Andersen, B. Hildebrandt, G. Köster, R. W. Hoffmann, *Chem. Ber.* **1989**, *122*, 1777–1782.

[537] R. H. Wallace, K. Zong, *Tetrahedron Lett.* **1992**, *33*, 6941–6944.

[538] F. Neese, F. Wennmohs, A. Hansen, U. Becker, *Chem. Phys.* **2009**, *356*, 98–109.

[539] T. Petrenko, S. Kossmann, F. Neese, *J. Chem. Phys.* **2011**, *134*, 054116.

[540] S. Grimme, *J. Chem. Theory Comput.* **2014**, *10*, 140825004103004.

[541] R. H. Blessing, *Acta Crystallogr.* **1995**, *51*, 33–38.

[542] G. M. Sheldrick, *Acta Crystallogr.* **1990**, *46*, 467–473.

[543] G. M. Sheldrick, *SHELXL-97, Program for Crystal Structure Analysis*, Universität Göttingen, Germany, **1997**.

Part VI.

Appendix

A. Supporting Information to Chapter 3

Appendix A contains:

- List of the organic molecules in the fitting set
- Total energies for the S22 test set
- Thermal corrections to Gibbs free energies for small molecules and non-covalent complexes
- Free enthalpies of association of the S12L test set
- Health criteria for all proteins

List of the organic molecules in the fitting set

Table A.1.: List of molecules in the fitting set including the RMSD relative to the B3LYP-D3(BJ)/def2-TZVPP reference geometries for HF-3c and PM6 optimized structures.

No.	name	sum formula	RMSD(HF-3c)	RMSD(PM6)
molecules containing C H				
1	dihydrogen	H2	0.018	0.010
2	methane	CH4	0.004	0.004
3	ethyne	C2H2	0.009	0.018
4	ethene	C2H4	0.005	0.028
5	ethane	C2H6	0.006	0.004
6	cyclopropane	C3H6	0.012	0.034
7	cyclopropene	C3H4	0.011	0.034
8	allene	C3H4	0.005	0.030
9	cyclobutane	C4H8	0.030	0.005
10	butadiene	C4H6	0.174	0.106
11	c-butadiene	C4H4	0.009	0.068
12	neo-pentane	C5H12	0.008	0.013
13	cyclohexane	C6H12	0.018	0.016
14	benzene	C6H6	0.003	0.011
15	norbornadiene	C7H8	0.008	0.034
16	cyclooctatetraene	C8H8	0.101	0.156
17	naphtalene	C10H8	0.008	0.017
18	biphenyl	C16H10	0.096	0.241
molecules containing C H F				
19	difluorine	F2	0.070	0.015
20	hydrogen fluoride	HF	0.011	0.022
21		CF2	0.010	0.003
22	tetraflouroethylene	C2F4	0.017	0.054

23	hexafluoroethane	C2F6	0.067	0.076
24	1,1,1-trifluoroethane	CH3CF3	0.037	0.056
25	hexaflourbenezene	C6F6	0.020	0.013
molecules containing C H O (F)				
26	dioxygen	O2	0.005	0.048
27	carbon monooxide	CO	0.005	0.007
28	carbon dioxide	CO2	0.009	0.009
29	water	H2O	0.012	0.012
30	hydrogen peroxide	H2O2	0.069	0.274
31	oxygen difluoride	F2O	0.059	0.045
32	formyl flouride	HCOF	0.020	0.041
33	formaldeyde	CH2O	0.009	0.007
34	fomic acid	HCOOH	0.027	0.041
35	acetone	CO(CH3)2	0.023	0.019
36	acetyl (radical)	CO(CH3)*	0.031	0.075
37	glyoxal	C2H2O2	0.021	1.376
38	EtOH	C2H5OH	0.017	0.036
39	dimethylether	CH3OCH3	0.021	0.045
40	oxirane	C2H4O	0.022	0.034
41	furane	C4H40	0.013	0.031
42	3,3,4,4-tetramethyl-1,2-dioxetane	C6H12O2	0.047	0.070
43	methyl acetate	CH3COOCH3	0.027	0.494
44	a-D-Glocupyranose	C6H12O6	0.051	0.051
molecules containing C H N (O F)				
45	dinitrogen	N2	0.004	0.013
46	hydrogen cyanide	HCN	0.005	0.006
47	flouramine	NH2F	0.045	0.018
48		NH2OF	0.111	0.041
49	ammonia	NH3	0.029	0.006
50	hydrazine	N2H4	0.042	0.543
51	trimethylamine	N(CH3)3	0.047	0.038
52	aziridine	C2H5N	0.023	0.053
53	pyridine	C5H5N	0.006	0.024
54	tetrazine	C2H2N4	0.074	0.067
55	methaneimine	CH2NH	0.017	0.063
56	succinimide	C4H5NO2	0.021	0.025
57	urea	CO(NH2)2	0.029	1.315
58	alanin	CH3CH(NH2)COOH	0.144	0.069
59	nitrobenzene	C6H5NO2	0.011	0.019
60	phenylamine	C6H5NH2	0.078	0.025
61	2-(2-aminophenoxy)phenol	C12H11O2N	0.095	0.092
62	adenine	C5H5N5	0.391	0.396
63	cytosine	C4H5N3O	0.042	0.051
molecules containing C H Cl (N O F)				
64	dichlorine	Cl2	0.011	0.017
65	hydrogen chloride	Hcl	0.022	0.009
66	chlorine monofluoride	Fcl	0.030	0.001
67	dichlorine monoxide	Cl2O	0.022	0.024
68	carbon tetrachloride	Ccl4	0.015	0.030
69	tetrachloroethylene	C2Cl4	0.016	0.023
70	hexachlorobenzene	C6Cl6	0.012	0.012
71	phosgene	COCl2	0.019	0.027
72	chloramine	NH2Cl	0.042	0.020
molecules containing C H S (O F)				
73	disulfur	S2	0.004	0.014

74	hydrogen sulfide	H2S	0.025	0.024
75	hydrogen disulfide	H2S2	0.028	0.040
76	sulfur dioxide	SO2	0.028	0.040
77	sulfur difluoride	SF2	0.022	0.146
78	sulfoxide	CH2S	0.034	0.031
79	dimethylsulfide	S(CH3)2	0.010	0.018
80	ethanthiol	C2H5SH	0.022	0.049
81	dithietane	C2S2H4	0.018	0.036
82	thiophene	C4H4S	0.015	0.025
83	thiirane	C2H4S	0.022	0.026
molecules containing C H P				
84	phosphorous	P4	0.005	0.032
85	phosphane	PH3	0.022	0.094
86	diphosphane	P2H4	0.033	0.855
87	methylidynephosphine	HCP	0.013	0.022
88	methylenephosphine	CH2PH	0.011	0.072
89	methylphosphine	PH2CH3	0.017	0.070
90	phosphinous amide	PH2NH2	0.113	0.656
91	phenylphosphine	PhPH2	0.017	0.053
molecules containing C H B (F)				
92	borane	BH3	0.001	0.008
93	diborane	B2H6	0.051	0.036
94	trifluoroborane	BF3	0.006	0.005
95		BH2CH3	0.034	0.032
96	ammonia borane	BH3NH3	0.022	0.036
97		BH3NCH	0.052	0.031
98	borazine	B3H6N3	0.098	0.078
99	boronic acid	B(OH)3	0.058	0.048
100	trimethylborane	B(CH3)3	0.027	0.022
101	tris(pentafluorophenyl)borane	B(C6F5)3	0.032	0.198
molecules containing C H Si				
102	silane	SiH4	0.003	0.007
103	disilane	Si2H6	0.010	0.052
104		CH2SiH2	0.011	0.028
105	phenylsilane	PhSiH3	0.011	0.028
106	silabenzene	C5SiH6	0.008	0.027
107	TMS	Si(CH3)4	0.014	0.014

Total energies for the S22 test set

Table A.2.: Contributions $E_{disp}^{D3(BJ)}$, E_{BSSE}^{gCP} and E_{SRB} to the energy as well as final total energies E_{tot}^{HF-3c} for HF-3c single-point calculation on the reference geometries of the S22 complexes. All values are given in a.u..

no.	complex	$E_{disp}^{D3(BJ)}$	E_{BSSE}^{gCP}	E_{SRB}	E_{tot}^{HF-3c}
1	Ammonia dimer	-0.0135677	0.0593118	-0.0545009	-0.0087568
2	Water dimer	-0.0074538	0.0606613	-0.0700433	-0.0168358
3	Formic acid dimer	-0.0299646	0.0857724	-0.2152092	-0.1594014
4	Formamide dimer	-0.0376455	0.0878221	-0.1860949	-0.1359184
5	Uracil dimer h-bonded	-0.1212656	0.1636129	-0.4324284	-0.3900810
6	2-pyridoxine 2-aminopyridine complex	-0.1252456	0.1483840	-0.2767470	-0.2536085
7	Adenine thymine Watson-Crick complex	-0.1567821	0.1956715	-0.4590304	-0.4201410
8	Methane dimer	-0.0179855	0.0500821	-0.0378408	-0.0057442
9	Ethene dimer	-0.0327958	0.0588616	-0.0619226	-0.0358568
10	Benzene - Methane complex	-0.0671056	0.0847888	-0.1010838	-0.0834007
11	Benzene dimer parallel displaced	-0.1203611	0.1189375	-0.1638166	-0.1652402
12	Pyrazine dimer	-0.1039268	0.1080350	-0.2346852	-0.2305771
13	Uracil dimer stack	-0.1291107	0.1609975	-0.4343544	-0.4024677
14	Indole benzene complex stack	-0.1522298	0.1454883	-0.2238504	-0.2305920
15	Adenine thymine complex stack	-0.1698002	0.1950683	-0.4593097	-0.4340416
16	Ethene ethyne complex	-0.0243677	0.0460663	-0.0623129	-0.0406143
17	Benzene water complex	-0.0619781	0.0890309	-0.1164851	-0.0894322
18	Benzene ammonia complex	-0.0646446	0.0889949	-0.1091647	-0.0848144
19	Benzene HCN complex	-0.0662910	0.0723309	-0.1296924	-0.1236525
20	Benzene dimer T-shaped	-0.1165023	0.1190570	-0.1636935	-0.1611388
21	Indole benzene T-shape complex	-0.1465955	0.1457880	-0.2236032	-0.2244107
22	Phenol dimer	-0.1279341	0.1530626	-0.2435771	-0.2184486

Thermal corrections to Gibbs free energies for small molecules and non-covalent complexes

Table A.3.: Thermal corrections for 10 molecules from the fitting set, 4 complexes from the S22 and 6 complexes from the S66 test set. All values are given in kcal/mol.

molecule/complex	**B3LYP-D3**	**HF-3c**	**PM6**
nitrobenzene	43.39	42.62	41.35
tiophene	23.87	24.58	22.28
norbornadiene	60.88	63.59	58.53
naphtalene	70.86	71.97	68.05
glucose	97.29	97.05	84.03
trifluoroethane	14.84	14.55	13.02
methylacetate	35.98	36.43	31.47
succinimide	23.58	23.34	21.75
ammoniaborane	28.34	29.53	28.64
tetramethylsilan	70.51	74.79	58.45
adenenine-thymine	111.79	111.27	101.87
indole-benzene	115.48	117.26	109.69
phenol dimer	103.32	104.23	91.84
formic acid dimer	23.86	22.58	16.54
ethene-pentane	105.37	109.99	97.21

benzene dimer (T-shaped)	98.49	91.44	86.80
acetic acid dimer	54.11	54.26	45.58
peptide-MeOH	35.64	36.29	26.04
benzene-MeOH	70.77	71.31	61.73
cyclopentane-neopentane	157.84	165.26	140.50

Free enthalpies of association of the S12L test set

Table A.4.: Gas phase interaction energies ΔE, thermal corrections ΔG_{RRHO}, free enthalpies of solvation $\Delta\delta G_{solv}$ and final enthalpies of association ΔG_a for the S12L test set at the HF-3c level. All values are given in kcal/mol.

no.	complex	ΔE	ΔG_{RRHO}	$\Delta\delta G_{solv}$	ΔG_a
1a	TCNA@Tweezer	-31.19	15.80	9.02	-6.37
1b	DCB@Tweezer	-21.67	15.02	4.33	-2.32
2a	π-Syst1@Pincer	-26.50	18.28	4.44	-3.78
2b	π-Syst2@Pincer	-19.68	17.03	2.81	+0.15
3a	C_{60}@Catcher	-37.19	14.99	5.79	-16.41
3b	C_{70}@Catcher	-39.24	15.14	6.79	-17.31
4a	GLH@Mcycle	-31.38	17.22	9.02	-5.14
4b	BQ@Mcycle	-22.02	17.00	4.58	-0.44
5a	$BuNH_3$@CB6	-91.73	15.29	57.07	-19.38
5b	$PrNH_3$@CB6	-88.93	14.81	56.36	-17.76
6a	FECP@CB7	-142.23	20.77	89.56	-31.91
6b	ADOH@CB7	-27.66	16.27	-8.97	-20.36

Health criteria for all proteins

Standard health checks include: (1) clashcores or steric overlaps greater than 0.4 Å per 1000 atoms, (2) percentage of bad side-chain dihedrals or rotamers, (3) number of β-carbon deviations greater than 0.25 Å from the expected position based on the backbone coordinates, (4) percentage of backbone dihedrals that fall into a favoured region on a Ramachandran plot and (5) percentage of those, which are Ramachandran outliers, (6) percentage of bad bonds and (7) percentage of bad angles. Also the MolProbity Score and the backbone RMSD relative to the experimental starting structure.

Table A.5.: Health criteria, MolProbity score and backbone RMSD for the HFx optimized protein structures.

No.	**PDB ID**	**clash cores**	**poor rotam.**	**Rama favoured**	**Rama outlier**	**C^β dev**	**bad bond**	**bad angle**	**MolProbity score**	**backbone RMSD**
1	1AQG	38	30	0	78	0	0	0	3.95	0.773
2	1EMZ	10	8	0	100	0	10	5	2.18	1.438
3	1LB0	35	17	0	82	0	8	0	3.67	2.133
4	1LB7	24	23	0	100	0	6	6	2.9	1.761
5	1LBJ	30	10	5	85	0	5	5	3.37	1.092
6	1LCX	31	17	0	82	0	8	0	3.61	2.383
7	1LVQ	19	25	0	86	0	0	22	3.48	1.454
8	1LVR	44	38	0	86	0	0	0	3.97	2.343
9	1LVZ	38	10	0	100	0	0	0	2.82	0.725
10	1MZI	31	17	0	82	0	8	0	3.61	2.107

11	1O53	68	14	0	69	0	13	0	4.03	2.749
12	1ODP	30	13	6	89	0	5	5	3.38	2.656
13	1RIJ	18	11	5	95	0	4	4	2.87	1.123
14	1T2Y	55	0	9	78	0	4	0	2.97	1.893
15	1UAO	36	0	0	100	0	10	0	2.04	0.597
16	1V46	49	0	14	86	0	22	0	2.81	1.046
17	1Y03	12	0	3	91	0	6	3	2.12	4.198
18	1Y49	41	29	14	58	0	0	0	4.13	1.174
19	1YJP	37	17	0	80	0	29	14	3.72	0.877
20	1YT6	90	13	0	75	1	10	10	4.05	1.077
21	2AP7	13	7	6	78	1	10	5	3.04	0.627
22	2CEH	28	19	18	47	0	0	0	3.88	4.149
23	2CSA	38	19	18	53	0	5	5	3.98	1.795
24	2E4E	54	0	13	88	0	30	20	2.82	1.156
25	2EVQ	30	0	0	90	0	8	0	2.52	1.231
26	2FBU	46	18	10	70	0	8	0	3.94	2.111
27	2FXY	28	14	6	82	0	6	0	3.53	2.177
28	2FXZ	29	10	9	91	0	0	0	3.23	2.011
29	2I9M	24	11	0	100	0	11	6	2.67	1.746
30	2JOF	53	7	0	94	0	15	5	3.2	1.294
31	2JTA	45	22	0	63	0	30	20	4.05	1.216
32	2JXF	22	15	4	93	0	3	0	3.19	5.163
33	2K59	15	7	8	92	0	4	4	2.82	2.181
34	2KNP	14	16	0	94	0	3	3	2.99	0.998
35	2KUX	10	8	0	90	0	0	0	2.76	0.565
36	2KVX	24	20	0	88	0	0	0	3.45	0.654
37	2NX6	12	4	0	76	7	4	0	2.81	2.055
38	2NX7	14	0	4	92	0	4	0	2.12	1.285
39	2OL9	11	17	0	100	0	17	17	2.48	0.46
40	2ONW	61	0	25	50	0	0	0	3.23	4.034
41	2OQ9	22	13	0	82	0	4	0	3.39	4.963
42	2PJV	13	0	5	86	0	4	4	2.27	5.211
43	2PV6	34	24	0	90	0	5	0	3.61	2.909
44	2RLJ	18	0	7	86	0	6	6	2.4	2.574
45	2RMW	41	21	25	46	1	8	8	4.08	1.858
46	3FTK	70	17	20	40	0	14	14	4.26	3.291
47	3FTR	66	0	0	75	0	0	0	3.08	3.548
48	3FVA	53	50	25	50	0	17	0	4.45	2.308
49	3NJW	22	7	6	94	0	5	0	2.86	0.451
50	3NVG	51	20	50	50	0	33	17	4.15	3.117
	average	34	13	6	81	0.2	8	4	3.26	2.015

Table A.6.: Health criteria, MolProbity score and backbone RMSD for the protein structures optimized with HFx including COSMO.

No.	PDB ID	clash cores	poor rotam.	Rama favoured	Rama outlier	C^{β} dev	bad bond	bad angle	MolProbity score	backbone RMSD
14	1ODP	6	13	6	94	9	5	0	2.54	0.956
29	2EVQ	20	0	0	100	0	8	0	1.8	0.759
30	2FBU	23	0	0	80	0	0	0	2.59	0.881
35	2JTA	38	11	0	88	0	10	0	3.47	0.789
50	2RLJ	0	0	0	93	0	0	0	0.95	1.411

Table A.7.: Health criteria, MolProbity score and backbone RMSD for the PM6 optimized protein structures.

No.	PDB ID	clash cores	poor rotam.	Rama favoured	Rama outlier	C^{β} dev	bad bond	bad angle	MolProbity score	backbone RMSD
1	1AQG	33	30	0	67	0	0	0	3.99	0.577
2	1EMZ	58	8	5	84	0	5	0	3.58	4.847
3	1LB0	44	17	0	82	0	0	0	3.76	2.392
4	1LB7	55	15	0	79	0	0	0	3.87	2.606
5	1LBJ	55	40	0	85	0	0	0	4.09	2.923
6	1LCX	49	17	0	72	0	0	0	3.91	3.313
7	1LVQ	63	38	0	71	0	0	0	4.3	2.900
8	1LVR	44	75	14	57	0	0	0	4.48	2.060
9	1LVZ	-	-	-	-	-	-	-	-	-
10	1MZI	27	25	9	72	0	8	0	3.79	4.143
11	1O53	52	36	15	52	0	7	0	4.28	1.606
12	1ODP	49	19	11	67	0	0	0	4	3.930
13	1RIJ	60	17	0	86	0	0	0	3.82	1.339
14	1T2Y	89	17	4	57	0	0	0	4.28	2.126
15	1UAO	51	14	0	88	0	10	0	3.67	1.717
16	1V46	66	14	29	57	0	0	0	4.1	2.138
17	1Y03	-	-	-	-	-	-	-	-	-
18	1Y49	49	14	14	71	0	0	0	3.87	2.065
19	1YJP	37	17	0	60	0	0	0	3.9	2.171
20	1YT6	67	13	13	75	0	10	0	3.92	1.878
21	2AP7	50	14	0	83	0	0	0	3.74	1.716
22	2CEH	63	31	29	35	0	0	0	4.44	4.791
23	2CSA	95	25	12	47	0	0	0	4.49	4.486
24	2E4E	47	17	0	75	0	10	10	3.87	2.033
25	2EVQ	46	20	10	80	0	0	0	3.86	2.974
26	2FBU	23	27	0	50	0	25	8	3.92	3.611
27	2FXY	57	43	13	88	0	0	0	4.08	3.169
28	2FXZ	62	10	18	73	0	0	0	3.84	4.022
29	2I9M	45	33	0	73	0	0	0	4.1	4.474
30	2JOF	42	20	6	89	0	0	0	3.67	1.588
31	2JTA	-	-	-	-	-	-	-	-	-
32	2JXF	-	-	-	-	-	-	-	-	-
33	2K59	78	22	8	77	0	0	0	4.16	2.683
34	2KNP	-	-	-	-	-	-	-	-	-
35	2KUX	64	12	0	89	0	0	0	3.65	0.692
36	2KVX	54	24	0	88	0	4	0	3.84	0.973
37	2NX6	66	16	20	72	0	4	0	4.03	2.190
38	2NX7	-	-	-	-	-	-	-	-	-
39	2OL9	32	33	0	100	0	0	0	3.15	1.628

40	2ONW	121	0	25	75	0	17	0	3.34	4.115
41	2OQ9	83	19	5	72	0	0	0	4.17	8.808
42	2PJV	54	9	0	77	0	4	0	3.7	5.752
43	2PV6	-	-	-	-	-	-	-	-	-
44	2RLJ	-	-	-	-	-	-	-	-	-
45	2RMW	-	-	-	-	-	-	-	-	-
46	3FTK	30	17	0	60	0	0	0	3.81	4.101
47	3FTR	53	0	25	75	0	0	0	2.99	4.022
48	3FVA	42	0	50	50	0	0	0	3.07	4.394
49	3NJW	56	20	6	82	0	0	5	3.91	1.210
50	3NVG	21	20	0	25	0	17	0	3.87	3.047
	average	54	21	8	71	0	3	0.6	3.89	2.956

Table A.8.: Health criteria and MolProbity score for the experimentally obtained protein structures.

No.	**PDB ID**	**clash cores**	**poor rotam.**	**Rama favoured**	**Rama outlier**	**C^{β} dev**	**bad bond**	**bad angle**	**MolProbity score**
1	1AQG	22	30	11	67	1	0	9	3.82
2	1EMZ	6	8	0	100	0	0	0	2.03
3	1LB0	9	33	0	100	0	0	0	2.63
4	1LB7	0	8	0	100	0	0	0	1.17
5	1LBJ	0	15	5	85	0	0	0	1.74
6	1LCX	0	33	0	100	0	0	0	1.66
7	1LVQ	25	50	0	86	4	22	44	3.83
8	1LVR	0	38	0	57	2	0	0	2.63
9	1LVZ	11	0	0	100	0	0	0	1.55
10	1MZI	40	8	9	91	0	0	0	3.3
11	1O53	36	14	8	61	0	0	0	3.82
12	1ODP	6	19	6	78	0	0	0	3.06
13	1RIJ	21	17	0	100	0	0	0	2.74
14	1T2Y	162	33	26	43	0	0	0	4.83
15	1UAO	51	29	0	100	0	0	0	3.29
16	1V46	156	50	29	29	0	0	0	5.01
17	1Y03	5	54	0	94	0	0	0	2.95
18	1Y49	180	14	29	29	0	0	0	4.65
19	1YJP	0	0	0	100	0	0	0	0.5
20	1YT6	7	38	13	50	0	0	0	3.58
21	2AP7	50	29	6	94	0	0	0	3.66
22	2CEH	3	25	24	35	1	0	0	3.24
23	2CSA	23	13	18	29	0	0	0	3.74
24	2E4E	16	17	0	75	0	0	0	3.42
25	2EVQ	0	0	0	90	0	0	0	1.05
26	2FBU	23	0	20	60	0	0	0	3.5
27	2FXY	10	36	6	82	0	0	0	3.44
28	2FXZ	24	20	0	91	0	0	0	3.38
29	2I9M	20	33	0	73	0	0	0	3.77
30	2JOF	0	13	0	100	0	0	0	1.35
31	2JTA	90	56	0	50	0	0	0	4.71
32	2JXF	2	27	4	93	0	0	0	2.51
33	2K59	70	4	8	92	0	0	0	3.22
34	2KNP	40	24	3	90	0	0	0	3.66
35	2KUX	35	8	0	100	0	0	0	2.69
36	2KVX	35	8	4	92	0	0	0	3.19
37	2NX6	43	20	4	92	0	0	0	3.59
38	2NX7	33	14	0	92	0	0	0	3.34

39	2OL9	0	0	0	100	0	0	0	0.5
40	2ONW	0	0	0	100	0	0	0	0.5
41	2OQ9	26	31	0	86	0	0	0	3.67
42	2PJV	3	0	0	77	0	0	0	1.9
43	2PV6	84	38	0	85	0	0	0	4.25
44	2RLJ	22	0	0	64	0	0	0	2.73
45	2RMW	48	21	8	42	0	0	0	4.17
46	3FTK	0	0	0	100	0	0	0	0.5
47	3FTR	0	0	0	100	0	0	0	0.5
48	3FVA	0	0	0	100	0	0	0	0.5
49	3NJW	0	0	0	88	0	0	0	1.09
50	3NVG	0	0	0	100	0	0	0	0.5
	average	29	19	5	81	0.2	0.5	1	2.74

B. Supporting Information to Chapter 4

Appendix B contains:

- Contributions to ΔG_a and final ΔG_a values for the eight complexes

Contributions to ΔG_a and final ΔG_a values

Table B.1.: Contributions to ΔG_a and final ΔG_a values on the PW6B95-D3/def2-QZVP(-g,-f)/COSMO-RS//TPSS-D3/def2-TZVP/HF-3c(freq.) for all eight complexes including Cl^- counterions for **3**, **4**, **5** and **6**. All values are given kcal mol^{-1}.

	ΔE	$\Delta E^{(2)}_{disp}$	$\Delta E^{(3)}_{disp}$	ΔG^T_{RRHO}	$\Delta\delta G^T_{solv}$	ΔG_a
1	-32.58	-20.51	1.73	15.42	8.45	-6.98
2	-26.24	-18.01	1.77	15.94	6.56	-1.96
3	-27.75	-17.40	1.64	20.33	-3.20	-8.99
4	-38.99	-22.94	2.35	21.64	12.15	-2.86
5	-38.37	-23.34	3.75	17.58	1.48	-15.57
6	-46.30	-23.90	3.86	18.52	5.52	-18.39
7	-52.94	-43.95	4.50	18.26	11.8	-18.44
8	-45.80	-31.97	3.87	16.02	11.1	-14.80

C. Supporting Information to Chapter 5

Appendix C contains:

- Contributions to ΔG_a values for the **CB7** complexes
- Submitted final relative ΔG_a values for the C**B7** complexes
- Contributions to ΔG_a and final ΔG_a values for the **OA** complexes
- Comparison of the computed with the experimental binding free energies for the OA hsot with one Na^+ counterion for the guest and four or six Na^+ counterions for the host

Contributions to ΔG_a and final ΔG_a values for the CB7 Complexes

Table C.1.: Contributions to ΔG_a and final ΔG_a values on the PW6B95-D3/def2-QZVP(-g,-f)/COSMO-RS//TPSS-D3-cosmo/def2-TZVP/HF-3c(freq.) for all 14 complexes including Cl^- counter ions (kcal mol^{-1}).

	ΔE	$\Delta E^{(2)}_{disp}$	$\Delta E^{(3)}_{disp}$	ΔG^T_{RRHO}	$\Delta\delta G^T_{solv}$	ΔG_a
1@CB7	-55.1	-18.1	2.9	19.1	22.2	-10.9
2@CB7	-30.2	-14.1	2.3	15.6	0.4	-11.8
3@CB7	-36.4	-14.0	2.4	17.6	8.0	-8.3
4@CB7	-52.8	-15.5	2.6	18.0	23.4	-8.8
5@CB7	-35.7	-12.7	1.8	18.8	10.2	-4.9
6@CB7	-26.3	-18.8	3.0	17.1	2.8	-3.4
7@CB7	-28.6	-15.4	2.5	16.2	-1.1	-11.1
8@CB7	-31.2	-18.0	2.9	16.7	-2.1	-13.7
9@CB7	-35.6	-19.6	3.2	16.9	-0.2	-15.7
10@CB7	-40.5	-18.7	3.1	17.8	14.1	-5.5
11a@CB7	-28.4	-17.1	2.8	16.0	0.6	-9.0
11b@CB7	-32.7	-16.5	2.7	15.7	0.3	-14.0
12@CB7	-36.8	-24.0	3.8	15.7	-0.6	-17.8
13@CB7	-33.0	-20.1	3.3	16.9	-4.5	-17.3
14@CB7	-39.8	-23.4	3.8	19.2	4.9	-11.9

Table C.2.: Submitted relative ΔG_a ($\Delta\Delta G_{a,rel}$) values (reference compound is 1@CB7) in kcal mol^{-1}.

	$\Delta\Delta G_{a,rel}$		$\Delta\Delta G_{a,rel}$
2@CB7	-1	**10@CB7**	5.4
3@CB7	2.6	**11@CB7**	-3.1
4@CB7	2.1	**11a@CB7**	1.9
5@CB7	6	**11b@CB7**	-3.1
6@CB7	7.5	**12@CB7**	-6.9
7@CB7	-0.2	**13@CB7**	-6.4
8@CB7	-2.8	**14@CB7**	-1.1
9@CB7	-3.9		

Contributions to ΔG_a and final ΔG_a values for the OA Complexes

Table C.3.: Contributions to ΔG_a and final ΔG_a values on the PW6B95-D3/def2-QZVP(-g,-f)/COSMO-RS//TPSS-D3-cosmo/def2-TZVP/HF-3c(freq.) for all nine complexes including nine Na^+ counter ions and 26 water molecules (kcal mol^{-1}). The TPSS-D3(BJ)/def2-TZVP to PW6B95-D3(BJ)/def2-QZVP' correction $\Delta\Delta E_{tq}$ is taken from the fully TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) optimized structure of the anionic guest with the neutral host. Results obtained by calculating single-point energies after re-substituting eight Na^+ counter ions by H and removing the 24 coordinating water molecules are given in parentheses. The hydrogen atoms are placed in a relative geometry derived from a fully HF-3c optimized structure of the neutral host: d(H-O)= 1.0 Å, α(H-O-C)= 110 degrees, θ(H-O-C-O)= 0 degrees.

	ΔE	$\Delta E^{(2)}_{disp}$	$\Delta E^{(3)}_{disp}$	$\Delta\Delta E_{tq}$	ΔG^T_{RRHO}	$\Delta\delta G^T_{solv}$	ΔG_a
15@OA	-34.0 (-32.4)	-33.5 (-33.0)	2.9 (2.9)	1.1	13.9	8.7 (7.1)	-7.4 (-7.5)
16@OA	-38.5 (-37.8)	-36.5 (-36.0)	3.0 (2.9)	1.1	17.2	10.2 (9.2)	-7.0 (-7.4)
17@OA	-35.1 (-34.7)	-38.2 (-37.7)	3.3 (3.2)	-0.5	17.8	4.8 (3.7)	-9.7 (-10.5)
18@OA	-39.4 (-37.7)	-34.9 (-34.5)	2.8 (2.7)	1.3	16.4	10.0 (8.5)	-8.9 (-8.7)
19@OA	-32.4 (-30.5)	-33.5 (-33.0)	2.9 (2.8)	0.8	15.0	6.3 (4.5)	-7.4 (-7.4)
20@OA	-29.6 (-28.9)	-31.4 (-31.0)	3.1 (3.0)	2.4	16.0	5.9 (4.6)	-2.2 (-2.8)
21@OA	-34.0 (-33.4)	-35.2 (-34.8)	3.5 (3.4)	-1.2	16.6	5.2 (4.3)	-10.0 (-10.3)
22@OA	-34.6 (-32.0)	-36.0 (-35.5)	3.3 (3.2)	-0.3	15.8	10.7 (8.4)	-5.0 (-4.7)
23@OA	-31.5 (-30.8)	-36.5 (-36.0)	3.5 (3.4)	1.6	16.5	4.4 (3.2)	-5.5 (-6.0)

Table C.4.: Contributions to ΔG_a and final ΔG_a values on the PW6B95-D3/def2-QZVP(-g,-f)/COSMO-RS//TPSS-D3-cosmo/def2-TZVP/HF-3c(freq.) for all nine complexes including seven Na^+ counter ions and 20 water molecules (kcal mol^{-1}). The TPSS-D3(BJ)/def2-TZVP to PW6B95-D3(BJ)/def2-QZVP' correction $\Delta\Delta E_{tq}$ is taken from the fully TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) optimized structure of the anionic guest with the neutral host.

	ΔE	$\Delta E^{(2)}_{disp}$	$\Delta E^{(3)}_{disp}$	$\Delta\Delta E_{tq}$	ΔG^{T}_{RRHO}	$\Delta\delta G^{T}_{solv}$	ΔG_a
15@OA	-36.5	-33.5	2.7	1.1	13.9	10.6	-8.2
16@OA	-32.5	-34.2	2.9	1.1	17.2	4.9	-6.5
17@OA	-33.1	-38.6	3.2	-0.5	17.8	3.0	-9.6
18@OA	-33.3	-33.8	2.6	1.3	16.4	4.7	-8.2
19@OA	-30.9	-33.1	2.7	0.8	15.0	4.3	-8.1
20@OA	-27.3	-30.3	2.9	2.4	16.0	3.0	-3.0
21@OA	-28.1	-33.4	3.3	-1.2	16.6	0.8	-8.6
22@OA	-27.8	-34.5	3.1	-0.3	15.8	5.8	-3.4
23@OA	-38.0	-35.3	3.1	1.6	16.5	9.2	-7.6

Table C.5.: Contributions to ΔG_a and final ΔG_a values on the PW6B95-D3/def2-QZVP(-g,-f)/COSMO-RS//TPSS-D3-cosmo/def2-TZVP/HF-3c(freq.) for all nine complexes including five Na^+ counter ions and 14 water molecules (kcal mol^{-1}). The TPSS-D3(BJ)/def2-TZVP to PW6B95-D3(BJ)/def2-QZVP' correction $\Delta\Delta E_{tq}$ is taken from the fully TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) optimized structure of the anionic guest with the neutral host.

	ΔE	$\Delta E^{(2)}_{disp}$	$\Delta E^{(3)}_{disp}$	$\Delta\Delta E_{tq}$	ΔG^{T}_{RRHO}	$\Delta\delta G^{T}_{solv}$	ΔG_a
15@OA	-31.6	-32.4	2.7	1.1	13.9	6.8	-7.2
16@OA	-32.9	-34.9	2.9	1.1	17.2	5.3	-6.3
17@OA	-31.9	-38.2	3.2	-0.5	17.8	2.3	-9.1
18@OA	-31.8	-35.2	2.7	1.3	16.4	4.6	-6.8
19@OA	-30.3	-33.1	2.7	0.8	15.0	3.8	-8.1
20@OA	-28.2	-30.4	2.9	2.4	16.0	3.9	-3.1
21@OA	-29.0	-33.4	3.3	-1.2	16.6	1.1	-9.2
22@OA	-30.6	-35.0	3.1	-0.3	15.8	8.3	-3.6
23@OA	-30.3	-35.3	3.2	1.6	16.5	2.9	-6.1

Table C.6.: Contributions to ΔG_a and final ΔG_a values on the PW6B95-D3/def2-QZVP(-g,-f)/COSMO-RS//TPSS-D3-cosmo/def2-TZVP/HF-3c(freq.) for all nine complexes including one Na^+ counter ion and two water molecules (kcal mol^{-1}). The position of the Na^+ counterion and two explicit water molecules relative to the guest is obtained by TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) geometry optimization of guest **15** + Na^+ + $2H_2O$: d(Na-O)= 2.4 Å, α(Na-O-C)= 90 degrees, θ(Na-O-C-O)= 0 degrees, d(O-Na)= 2.4 Å, α(O-Na-O)= 162 degrees, θ(O-Na-O-C)= 4 degrees, d(H-O)=1.0 Å, α(H-O-Na)= 120 degrees, θ(H-O-Na-O)= -129 degrees, d(H-O)= 1.0 Å, α(H-O-Na)= 120 degrees, θ(H-O-Na-O)= 2 degrees, d(O-Na)= 2.4 Å, α(O-Na-O)= 105 degrees, θ(O-Na-O-C)= -179 degrees, d(H-O)= 1.0 Å, α(H-O-Na)= 120 degrees, θ(H-O-Na-O)= -130 degrees, d(H-O)= 1.0 Å, α(H-O-Na)= 120 degrees, θ(H-O-Na-O)= 1 degree. The TPSS-D3(BJ)/def2-TZVP to PW6B95-D3(BJ)/def2-QZVP' correction $\Delta\Delta E_{tq}$ is taken from the fully TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) optimized structure of the anionic guest with the neutral host. The neutral to ionic correction $\Delta\Delta E_{na}$ is calculated on the TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) level of theory using a structure of the host where the protons of the four benzoic acid carboxylate groups are replaced by Na^+ ions in a relative geometry derived from a fully HF-3c optimized anionic structure of the host with twelve explicit water molecules: d(Na-O)= 2.2 Å are the distances to the coordinating oxygen atoms, α(Na-O-C)= 109/90/96/93 degrees are the bond angles of the 1st/2nd/3rd/4th Na^+, and θ(Na-O-C-O)= -1/3/1/6 degrees are the corresponding dihedrals.

	ΔE	$\Delta E^{(2)}_{disp}$	$\Delta E^{(3)}_{disp}$	$\Delta\Delta E_{tq}$	$\Delta\Delta E_{na}$	ΔG^{T}_{RRHO}	$\Delta\delta G^{T}_{solv}$	ΔG_a
15@OA	-34.7	-34.6	2.7	1.1	0.3	13.9	9.0	-7.6
16@OA	-32.6	-34.3	2.8	1.1	0.6	17.2	6.0	-4.9
17@OA	-35.3	-39.1	3.1	-0.5	0.6	17.8	5.7	-8.6
18@OA	-32.3	-33.2	2.5	1.3	0.5	16.4	4.6	-7.0
19@OA	-36.5	-35.2	2.6	0.8	0.1	15.0	9.9	-8.1
20@OA	-33.3	-31.2	2.9	2.4	0.7	16.0	8.3	-3.0
21@OA	-31.1	-32.7	3.2	-1.2	0.3	16.6	2.8	-9.4
22@OA	-19.6	-37.5	3.2	-0.3	0.5	15.8	15.1	14.9
23@OA	-35.6	-35.4	3.2	1.6	0.7	16.5	7.7	-5.9

Table C.7.: Contributions to ΔG_a and final ΔG_a values on the PW6B95-D3/def2-QZVP(-g,-f)/COSMO-RS//TPSS-D3-cosmo/def2-TZVP/HF-3c(freq.) for all nine complexes of neutral host and anionic guest (kcal mol^{-1}). The neutral to ionic correction $\Delta\Delta E_{na}$ is calculated on the TPSS-D3(BJ)/def2-TZVP + COSMO (ϵ=78) level of theory using a structure of the host where the protons of the four benzoic acid carboxylate groups are replaced by Na^+ ions in a relative geometry derived from a fully HF-3c optimized anionic structure of the host with twelve explicit water molecules: d(Na-O)= 2.2 Å are the distances to the coordinating oxygen atoms, α(Na-O-C)= 109/90/96/93 degrees are the bond angles of the 1st/2nd/3rd/4th Na^+, and θ(Na-O-C-O)= -1/3/1/6 degrees are the corresponding dihedrals. ΔG_a values submitted to SAMPLE4 are given in parentheses. The main difference to the ΔG_a values presented in here is the missing TPSS-D3(BJ)/def2-TZVP to PW6B95-D3(BJ)/def2-QZVP' correction $\Delta\Delta E_{tq}$ given in C.3, C.4, C.5, and C.6. Furthermore the $\Delta\Delta E_{na}$ term in C.6 and this Table is replaced by 1.36 kcal mol^{-1} which was the average value of the neutral to ionic corrections for the first seven complexes (15@OA to 21@OA) upon submission. Finally, for complexes 21@OA and 23@OA a structure with more negative ΔG_a found after submission is considered here, and for complex 22@OA instead of the structure with lowest ΔG_a a less favourable structure is taken in which the carboxylate group of cyclopentanecarboxylic acid points out of instead of into the binding pocket.

	ΔE	$\Delta E^{(2)}_{disp}$	$\Delta E^{(3)}_{disp}$	$\Delta\Delta E_{na}$	ΔG^{T}_{RRHO}	$\Delta\delta G^{T}_{solv}$	ΔG_a
15@OA	-45.8	-30.0	2.5	1.8	13.9	27.6	0.1 (-1.5)
16@OA	-45.0	-32.9	2.7	0.9	17.2	22.5	-1.7 (-2.4)
17@OA	-48.6	-37.8	3.1	1.1	17.8	21.6	-5.0 (-4.2)
18@OA	-44.0	-31.9	2.5	1.1	16.4	20.2	-3.9 (-4.8)
19@OA	-44.9	-30.3	2.5	1.7	15.0	24.4	-1.3 (-2.5)
20@OA	-45.5	-28.4	2.7	1.9	16.0	27.5	2.7 (-0.3)
21@OA	-43.4	-31.4	3.1	0.8	16.6	16.4	-6.5 (-1.5)
22@OA	-47.8	-32.9	2.8	2.1	15.8	40.2	13.1 (11.6)
23@OA	-47.9	-33.1	3.0	1.8	16.5	25.6	-0.9 (-0.9)

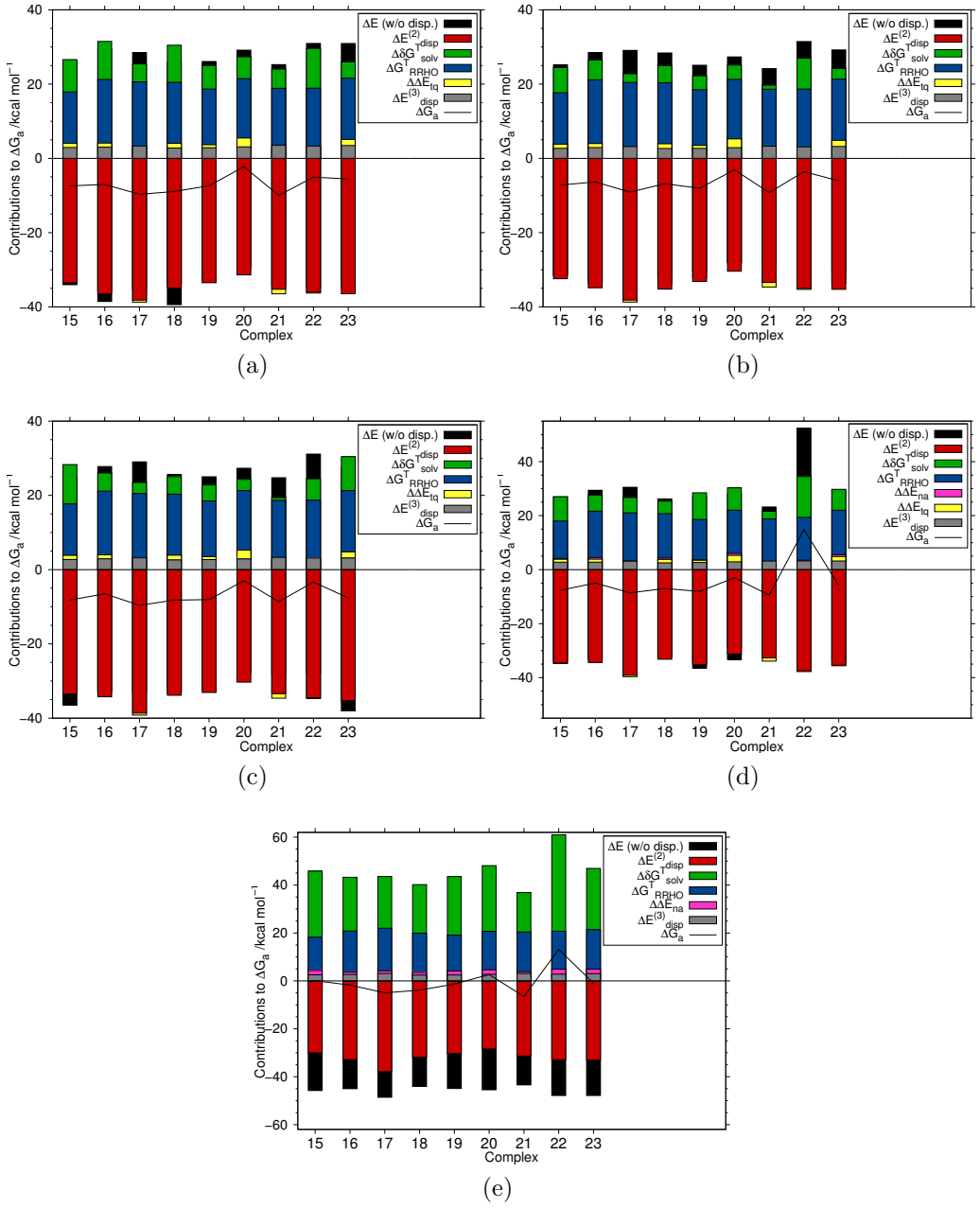


Figure C.1.: Contributions to ΔG_a and final ΔG_a values on the PW6B95-D3/def2-QZVP(-g,-f)/COSMO-RS//TPSS-D3-cosmo/def2-TZVP/HF-3c(freq.) for all nine **OA** complexes in kcal mol^{-1} of fully deprotonated anionic host and anionic guest including nine Na^+ counter ions and 26 water molecules (a), partially deprotonated anionic host and anionic guest including five Na^+ counter ions and 14 water molecules (b, artially deprotonated anionic host and anionic guest including seven Na^+ counter ions and 20 water molecules (c, neutral host and anionic guest including one Na^+ counter ion and two water molecules (d), and neutral host and anionic guest (e).

Comparison of the computed with the experimental binding free energies

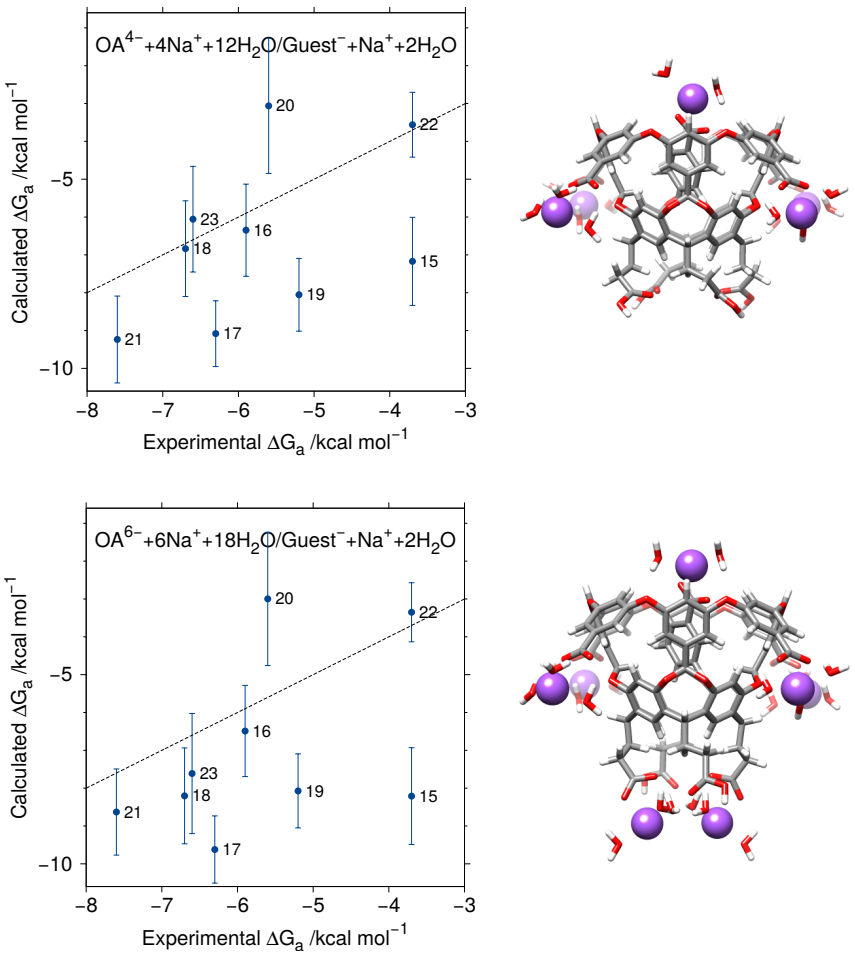


Figure C.2.: Computed vs. experimental binding free energy (kcal mol^{-1}) for the anionic guests with one Na^+ counterion in the anionic octa-acid host with four (top) and six (bottom) Na^+ counterions. Geometries of the **15@OA** complex are shown on the right.

D. Supporting Information to Chapter 6

Appendix C contains:

- Electronic energies of all S30L complexes for all used density functionals and semiempirical methods
- Thermal corrections of all S30L complexes for all used methods
- Solvation contributions of all S30L complexes for all used methods
- Validation of thermal corrections obtained with semiempirical methods
- Comparison of the reference binding energies of 11 complexes with other methods
- Statistics (MAD, MD) without charged systems

Electronic energies of all S30L complexes for all used density functionals and semiempirical methods

Table D.1.: ΔE obtained with all tested functionals for the S30L set. The two body dispersion contribution $\Delta E^{(2)}_{disp}$ is included in ΔE, the three body dispersion $\Delta E^{(3)}_{disp}$ is given independently. In case of S30L-CI counter ions were included for the complexes indicated with a "CI".

	PW6B95-D3	TPSS-D3	B3LYP-D3	PBE-D3	ωB97xD3	$\Delta E^{(3)}_{disp}$
1	-32.08	-33.011	-35.09	-29.77	-33.18	1.83
2	-21.72	-21.798	-23.01	-19.43	-21.78	1.27
3	-26.17	-22.047	-27.92	-19.97	-21.47	1.95
4	-21.07	-19.900	-21.78	-19.28	-20.35	0.74
5	-35.19	-34.818	-38.05	-30.85	-34.66	2.28
6	-31.38	-33.556	-35.85	-24.15	-24.86	1.88
7	-34.93	-35.952	-39.58	-31.63	-35.39	1.99
8	-39.73	-41.283	-45.35	-36.46	-41.00	2.22
9	-34.98	-37.468	-37.46	-30.58	-32.14	3.30
10	-36.05	-38.339	-38.77	-31.33	-33.67	3.61
11	-43.42	-46.703	-46.94	-37.95	-39.25	5.12
12	-42.88	-46.069	-46.48	-37.16	-39.19	5.07
13	-28.82	-26.917	-30.18	-25.77	-31.55	2.93
14	-31.32	-29.294	-33.21	-26.90	-33.00	3.04
15	-18.17	-22.362	-21.21	-21.41	-15.91	0.33
16	-24.51	-27.923	-26.92	-25.20	-20.09	0.49
17	-32.92	-32.803	-36.87	-33.29	-37.22	1.05
18	-21.41	-22.039	-25.73	-22.25	-24.58	1.03
19	-16.64	-16.468	-17.26	-16.91	-19.11	1.63
20	-20.18	-19.875	-21.45	-20.50	-23.75	2.33
21	-28.98	-27.347	-30.95	-26.81	-32.54	3.42

22	-33.89	-37.786	-38.52	-38.80	-37.39	0.11
23	-58.67	-57.312	-50.45	-67.86	-64.47	-0.26
24	-139.23	-133.727	-143.77	-132.30	-146.92	5.88
25	-33.17	-35.672	-37.78	-29.71	-34.39	2.70
26	-33.14	-35.709	-37.80	-29.80	-34.50	2.74
27	-84.41	-83.021	-87.67	-82.67	-89.02	2.33
28	-80.68	-79.641	-83.44	-79.02	-84.80	1.92
29	-54.79	-53.830	-56.00	-54.77	-56.77	0.52
30	-50.59	-50.187	-52.96	-50.40	-52.60	0.73
23-CI	-63.11	-71.391	-69.46	-72.44	-65.68	-0.05
24-CI	-74.16	-68.629	-77.60	-55.23	-81.44	5.77
25-CI	-35.82	-35.672	-37.78	-31.20	-35.68	2.30
26-CI	-35.62	-35.709	-37.80	-34.31	-35.50	1.90
27-CI	-34.63	-34.458	-37.83	-34.31	-39.34	2.49
28-CI	-29.41	-29.701	-32.40	-29.44	-33.74	2.53
29-CI	-47.18	-44.473	-49.07	-44.18	-46.72	0.49
30-CI	-47.70	-45.667	-51.08	-44.71	-46.46	1.01

Table D.2.: $\Delta E_{el}^{semiemp}$ obtained with all tested semiempirical methods for the S30L set. The two body dispersion contribution $\Delta E_{disp}^{(2)}$ is included. In case of S30L-CI counter ions were included for the complexes indicated with a "CI".

	HF-3c	DFTB-D3	PM6-D3	PM6-D3H+	PM6-D3H2	PM6-D3H4	PM7	OM2-D3
1	-30.89	-28.62	-29.42	-29.42	-29.42	-31.53	-38.49	-32.28
2	-21.26	-19.88	-19.64	-19.64	-19.64	-21.75	-26.70	-20.47
3	-20.78	-23.94	-25.74	-24.23	-24.94	-27.85	-32.04	-24.94
4	-19.22	-18.79	-20.17	-18.93	-19.97	-22.28	-24.23	-
5	-34.38	-34.07	-34.92	-34.74	-34.74	-37.03	-45.56	-34.64
6	-25.71	-24.61	-26.04	-26.01	-26.01	-28.14	-35.51	-23.59
7	-39.88	-38.78	-31.20	-31.20	-31.20	-33.31	-49.45	-30.04
8	-45.32	-44.03	-35.17	-35.17	-35.17	-37.28	-57.31	-33.89
9	-37.37	-36.37	-30.39	-30.39	-30.39	-32.50	-57.19	-32.71
10	-39.26	-38.10	-31.82	-31.82	-31.82	-33.92	-60.79	-33.68
11	-39.28	-44.49	-41.49	-41.48	-41.48	-43.59	-75.20	-
12	-39.69	-44.33	-41.69	-41.69	-41.69	-43.80	-75.97	-
13	-28.06	-26.75	-28.17	-27.77	-27.63	-30.27	-35.81	-25.45
14	-29.68	-30.64	-29.90	-29.68	-29.68	-32.01	-36.04	-
15	-38.97	-36.63	-31.82	-31.82	-35.38	-35.56	-4.03	-
16	-52.12	-42.40	-41.36	-41.35	-48.25	-45.79	-8.91	-
17	-28.56	-31.15	-43.63	-42.53	-45.00	-45.74	-45.09	-36.47
18	-20.47	-23.08	-33.03	-32.42	-32.32	-35.14	-34.48	-26.67
19	-16.41	-17.15	-19.66	-21.89	-22.03	-21.77	-19.26	-15.43
20	-19.79	-22.62	-25.62	-27.38	-27.52	-27.72	-25.23	-20.75
21	-27.62	-28.22	-28.12	-28.33	-27.22	-30.23	-33.16	-23.82
22	-39.52	-33.76	-42.37	-42.22	-44.57	-44.47	-49.26	-34.99
23	-66.17	-41.69	-57.24	-58.79	-57.22	-59.35	-72.11	-50.40
24	-144.10	-162.05	-166.42	-164.68	-164.68	-168.53	-183.77	-162.21
25	-34.08	-29.38	-24.33	-24.33	-24.33	-26.44	-50.48	-27.11
26	-34.28	-29.49	-24.11	-24.11	-24.11	-26.22	-50.85	-26.97
27	-92.41	-95.17	-95.32	-94.92	-92.76	-97.43	-112.65	-94.98
28	-88.76	-89.88	-92.37	-91.95	-90.13	-94.48	-110.49	-91.20
29	-68.72	-58.39	-41.06	-49.80	-59.37	-77.21	-53.64	-89.93
30	-65.11	-56.79	-73.22	-46.69	-59.77	-76.41	-53.72	-147.72
23-CI	-72.46	-47.47	-110.74	-55.88	-112.28	-124.59	-68.76	-

24-CI	-64.82	-87.86	-82.35	-81.14	-81.14	-84.60	-83.11	-
25-CI	-37.76	-34.71	-27.86	-27.86	-27.86	-30.10	-50.81	-
26-CI	-37.80	-34.77	-27.57	-27.58	-27.58	-29.82	-51.37	-
27-CI	-38.65	-42.97	-42.85	-42.20	-39.77	-45.09	-58.42	-
28-CI	-33.89	-36.83	-39.59	-38.28	-37.44	-41.83	-56.81	-
29-CI	-71.57	-61.74	-51.78	-47.08	-52.94	-60.20	-41.36	-
30-CI	-81.65	-61.16	-58.02	-55.89	-57.27	-63.58	-42.98	-

Thermal corrections of all S30L complexes for all used methods

Table D.3.: ΔG_{RRHO} obtained with all tested semiempirical methods for the S30L set. In case of S30L-CI counter ions were included for the complexes indicated with a "CI".

	HF-3c	DFTB-D3	PM6-D3
1	15.80	14.54	13.35
2	15.01	14.63	14.79
3	16.65	23.12	16.01
4	15.48	15.04	15.22
5	17.46	22.52	15.28
6	15.71	15.42	14.18
7	21.54	22.53	20.91
8	23.87	24.45	21.58
9	16.83	13.59	16.53
10	17.45	13.77	16.56
11	16.59	14.33	18.55
12	17.42	14.36	18.99
13	14.94	14.52	15.82
14	15.52	15.65	15.43
15	17.84	16.41	19.96
16	19.31	16.75	19.61
17	17.18	16.87	16.29
18	16.15	16.13	15.62
19	15.54	14.97	16.63
20	16.50	16.20	16.11
21	16.26	17.46	18.74
22	19.60	19.31	23.78
23	18.27	16.07	17.90
24	21.39	20.77	26.71
25	17.60	18.17	17.95
26	17.84	17.51	16.59
27	15.29	18.30	13.03
28	14.82	17.05	12.16
29	19.54	15.73	16.10
30	20.13	17.18	22.17
23-CI	18.27	16.07	17.90
24-CI	21.39	20.77	26.71
25-CI	17.60	18.17	17.95
26-CI	17.84	17.51	16.59
27-CI	15.29	18.30	13.03
28-CI	14.82	17.05	12.16
29-CI	19.54	15.73	16.10
30-CI	20.13	17.18	22.17

Solvation contributions of all S30L complexes for all used methods

Table D.4.: ΔG_{solv} obtained with all tested continuum solvation models for the S30L set. In case of S30L-CI counter ions were included for the complexes indicated with a "CI".

	COSMO-RS	12	13	13 fine	14	14 fine	SMD
1		9.04	9.08	9.28	9.31	9.8	12.97
2		4.37	4.34	4.35	4.48	5.18	8.71
3		5.39	5.62	5.51	5.77	6.5	6.62
4		2.99	3.44	2.94	3.35	3.1	5.57
5		6.33	6.47	6.48	6.70	8.91	11.04
6		5.19	5.39	5.50	5.50	7.67	10.39
7		8.02	8.23	8.78	8.33	10.69	10.93
8		10.72	10.88	11.57	11.05	13.84	12.09
9		6.25	5.52	5.75	7.06	8.35	14.68
10		7.23	6.33	6.58	8.05	9.75	16.25
11		11.96	7.80	8.07	11.13	13.08	18.21
12		12.30	7.99	8.21	11.41	14.25	17.71
13		5.89	5.36	5.38	6.08	6.37	6.58
14		6.81	6.16	5.60	6.97	7.28	7.42
15		-1.16	-1.27	-1.18	-0.95	-1.71	-0.31
16		0.71	0.39	0.47	0.89	-0.09	1.48
17		7.90	7.74	7.47	8.09	8.03	12.82
18		3.86	4.08	4.20	4.09	5.33	11.54
19		-1.38	-2.51	-1.08	-1.43	-1.62	0.54
20		-2.56	-3.59	-2.15	-2.55	-2.9	-0.70
21		-9.11	-7.95	-6.15	-8.21	-9.24	3.43
22		11.3	10.78	8.80	11.47	8.78	13.18
23		25.75	25.69	23.96	26.25	23.65	34.76
24		83.85	86.27	89.42	85.54	81.27	88.17
25		3.96	3.17	2.22	3.70	4.86	3.27
26		3.73	3.00	1.85	3.60	4.49	4.20
27		54.33	57.14	59.99	56.09	52.55	48.75
28		54.05	56.79	59.59	55.77	53.26	50.43
29		25.80	27.40	28.60	27.04	26.901	29.74
30		21.46	23.22	24.61	22.97	23.278	23.35
23-CI		27.428	33.92	31.71	34.69	30.94	20.37
24-CI		27.21	28.00	29.32	27.27	27.4	26.43
25-CI		8.10	7.42	7.07	7.96	7.76	7.43
26-CI		7.89	7.25	7.34	7.78	7.51	8.23
27-CI		10.59	10.94	14.12	10.44	10	7.05
28-CI		8.45	8.67	11.44	8.20	8.77	8.19
29-CI		22.08	21.03	19.77	21.63	20.12	16.17
30-CI		26.74	25.74	24.26	26.70	25.08	16.71

Validation of thermal corrections obtained with semiempirical methods

Table D.5.: Comparison of ΔG_{RRHO} obtained with semi-empirical methods with those obtained from TPSS-D3/def2-TZVP calculations for eight complexes. The MAD and MD is given w.r.t. TPSS-D3.

	TPSS-D3	HF-3c	DFTB-D3	PM6-D3
1	14.4	15.8	16.8	13.3
3	16.8[a]	18.3	23.6	16.0
7	21.4	21.5	21.8	19.1
9	16.8[a]	17.0	18.7	16.8
15	18.3	17.8	16.4	20.0
17	17.4[a]	17.2	17.3	16.3
22	19.2	18.7	16.1	17.9
25	18.1	18.1	18.2	17.9
MAD		0.4	1.6	1.1
MD		0.1	0.3	0.4

[a] Taken from: S. Grimme *Chem. Eur. J.*, **2012**, 18, 9955-9964.

Comparison of the reference binding energies of 11 complexes with other methods

Table D.6.: Comparison of the reference binding energies ΔE^{emp} for the complexes that were already contained in the S12L set, with the old empirical reference values and binding energies obtained from DFT-SAPT and DQMC calculations.For convenience we also provide PBE-MBD* and PBE-D3+ATM values.

	ΔE^{emp}	ΔE^{emp}_{old} [a]	ΔE (DFT-SAPT)[b]	ΔE (DQMC)[c]	ΔE (PBE-MBD*)[c]	ΔE (PBE-D3+ATM)
1	-29.0	-30.0	-30.00	-27.2	-29.0	-27.9
2	-20.8	-20.4	-19.3	-17.2	-18.8	-18.2
3	-23.5	-24.8	-27.0	-	-	-18.0
4	-20.3	-20.7	-22.0	-	-	-18.5
9	-28.4	-27.9	-33.9	-25.8	-28.3	-27.3
10	-29.8	-29.3	-35.2	-	-	-27.7
17	-33.4	-34.8	-33.0	-33.4	-33.8	-32.2
18	-23.3	-23.2	-21.5	-	-	-21.2
21	-24.2	-22.6	-27.7	-24.1	-27.4	-23.4
27	-82.2	-77.4	-82.4	-81.0	-82.1	-80.3
28	-80.1	-77.1	-78.7	-	-	-77.1

[a] Taken from: T. Risthaus, S. Grimme *J. Chem. Theory Comp.*, **2013**, *9*, 1580–1591.
[b] Taken from: A. Heßelmann, T. Korona *J. Chem. Phys.*, **2014**, *141*, 094107.
[c] Taken from: A. Ambrosetti, D. Alfè, R. A. DiStasio, A. Tkatchenko *J. Phys. Chem. Lett.*, **2014**, *5*, 849–855.

Statistics (MAD, MD) without charged systems

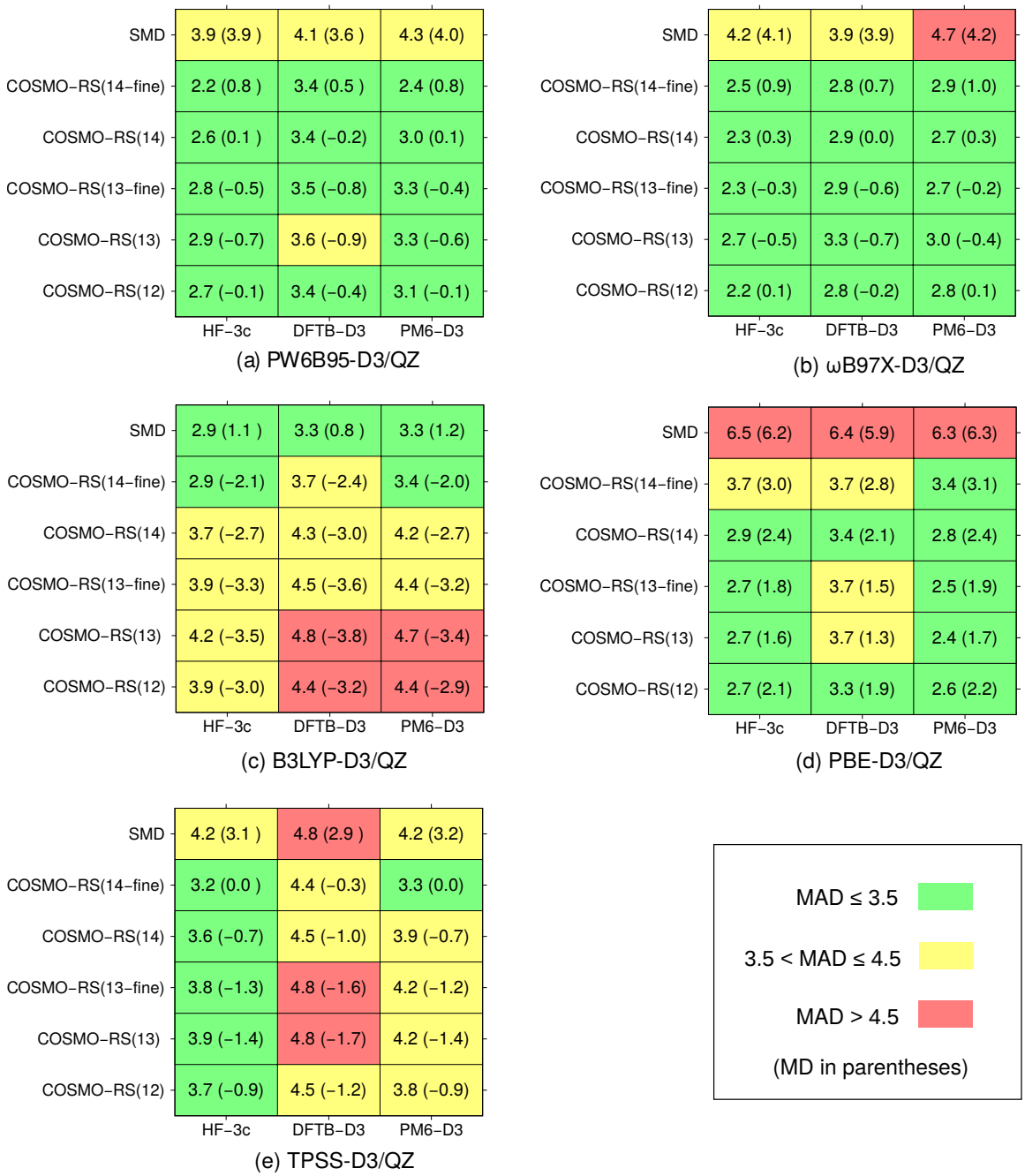


Figure D.1.: MADs (and MDs) for several combination of functionals (PW6B95-D3, B3LYP-D3, TPSS-D3 and PBE-D3 for ΔE), semi-empirical methods for frequencies (HF-3c, DFTB-D3 and PBE-D3 for ΔG^{T}_{RRHO}), and continuum solvation models (SMD and COSMO-RS with several parametrizations for $\Delta\delta G_{solv}$) w.r.t. to experimental ΔG_a for the S30L test set without the charged system **23** to **30**. The three body dispersion contribution $\Delta E^{(3)}_{disp}$ is included in ΔE.

E. Supporting Information to Chapter 7

Appendix E contains:

- Experimental details for the kinetic studies
- Experimental details for the cyclic voltammetry
- Details on the DigiSim Simulations
- Contributions to the reaction free energies of the hydrochloride adduct formation

Experimental details for the kinetic studies

General procedure for RPKA runs:
All kinetic runs were carried out using a Mettler-Toledos ReactIR 4000 fitted with SiComp probe and running iCIR software 4.2.26. To a two neck rbf manganese (9.5 mg, 0.173 mmol) and $Hex_3 \cdot HCl$ (60.5 mg, 0.197 mmol) were added and attached with a reflux condenser inside glove box. This was taken out and fixed to reactIR probe and flushed with argon. An air background (64 scans) was obtained and 11.5 ml of THF was added to the rbf through septa and started heating with an oil bath previously heated to 60.0 °C for 15 min and the iCIR software was initiated to obtain data. IR spectra were collected every minute till the completion of the reaction. The Cp_2TiCl_2 (21.4 mg, 0.0859 mmol) in 2.0 ml of THF was added after 4.0–5.0 minutes of starting the reaction and washed with 0.25 ml of THF. The reaction mixture was refluxed until the appearance of green color indicative of formation of Ti(III) and epoxide (**10**) (100.1 mg, 0.418 mmol) in 1 ml of THF was added and washed with 0.2 ml THF. The contents were continued refluxing till the end of reaction as monitored by reactIR.

Table E.1.: Conditions for Reaction with $Hex_3N \cdot HCl$ salt.

10 / M	$Hex_3N \cdot HCl$ / M	Mn / M	Cp2TiCl2 / M	Temperature / °C
0.0275	0.0132	0.0115	0.0057	60.0 ± 5
0.0275	0.0264	0.0115	0.0057	60.0 ± 5
0.0275	0.0528	0.0115	0.0057	60.0 ± 5

Experimental details for the cyclic voltammetry

Materials: THF was distilled over potassium under an atmosphere of argon. Tetrabutylammoniumhexafluorophosphate, Bu_4NPF_6, and Tetrabutylammoniumiodide, Bu_4NI, were commercially available in electrochemical grade from Aldrich and were stored in a glovebox under an atmosphere of argon. The additives LiCl, $nBu_4N^+Cl^-$, $nHex_3N{\cdot}HCl$, 2,4,6-collidinehydrochloride

and pyridinehydrochloride were dried under vacuum and stored in a glovebox under an atmosphere of argon.

Apparatus: All handling of chemicals and the cyclic voltammetric experiments were performed in a glovebox under an atmosphere of argon. The cyclic voltammograms were recorded by a 600D or a 1140D Electrochemical Analyzer/Workstation (CH-Instruments). The working electrode was a glassy carbon disk of diameter 1 mm. The electrode surface was polished using 0.25 μm diamond paste (Struers A/S), followed by cleaning in an ethanol bath. The counter electrode consisted of a platinum coil melted into glass, while a Ag/AgI electrode (silver wire immersed in a Pyrex tube containing 0.2 M Bu_4NPF_6 + 0.02 M Bu_4NI in THF) separated from the main solution by a ceramic frit served as the reference electrode. All potentials were reported against the Fc^+/Fc redox couple, the potential of which is equal to 0.52 V vs. SCE in 0.2 M Bu_4NPF_6/THF. All cyclic voltammograms were recorded using the IR compensation mode of the CH-Instrument Electrochemical Analyzer/Workstation.

Procedure: In the cyclic voltammetric experiments 0.77 g of Bu_4NPF_6 (2.0 mmol) and a small magnetic bar were added to the electrochemical cell. 9 mL of freshly distilled THF, 1 mL of the appropriate standard solution containing the complex or the zinc-reduced complex and the appropriate amount of additive were added to the cell and the solution was stirred. At the end of each series of experiments a small amount of ferrocene was added and the potential of the Fc^+/Fc redox couple was measured.

Details on the DigiSim-Simulations

$(C_5H_4CN)CpTiCl_2$: Simulation parameters and input values (using DigiSim syntax) for the simulation of the voltammograms of $(C_5H_4CN)CpTiCl_2$.

Heterogeneous reactions (electrode reactions)	$E^{[a]}$	α	$k_s^{[b]}$
$(C_5H_4CN)CpTiCl_2$ + e = $(C_5H_4CN)CpTiCl_2^-$	-1.00±0.03	0.5	6×10^{-3}
$[CpTiCl_2]_2^+$ + e = $[CpTiCl_2]_2$	-0.87±0.03	0.5	1×10^{-3}
$CpTiCl_2^+$ + e = $CpTiCl_2$	-0.85±0.03	0.5	6×10^{-3}
Homogeneous reactions	$K^{[c]}$	$k_f^{[d]}$	$k_b^{[d]}$
2 $CpTiCl_2$ = $[CpTiCl_2]_2$	1	10	10
$(C_5H_4CN)CpTiCl_2^-$ = $CpTiCl_2$ + $(C_5H_4CN)^-$	1×10^4	10	1×10^{-3}

[a] In V vs. Fc^+/Fc. [b] In cm s^{-1} [c] In Mx according to the reaction in question. [d] In M-x s^{-1} according to the reaction in question. [e] Equilibrium constants are determined automatically from square scheme thermodynamics in DigiSim.

Other input parameters (in DigiSim syntax):

E_{start} (V): 0
E_{switch} (V): -1.80
E_{end} (V): 0
Temperature (K): 298.2
Ru (Ohms): 0
Cdl (F): 0

cycles: 1
electrode geometry: planar
area (cm^2): 0.00785
diffusion: semi-infinite
pre-equilibrium: disabled

Diffusion coefficients:

$(C_5H_4COOMe)CpTiCl_2$, $(C_5H_4COOMe)CpTiCl_2^-$, $CpTiCl_2^+$: 6×10^{-6} $cm^s s^{-1}$
$(C_5H_4COOMe)^-$, $CpTiCl_2$, $[CpTiCl_2]_2$, $[CpTiCl_2]_2^+$: 1×10^{-5} $cm^s s^{-1}$

$(C_5H_4Cl)CpTiCl_2$: Simulation parameters and input values (using DigiSim syntax) for the simulation of the voltammograms of $(C_5H_4Cl)CpTiCl_2$.

Heterogeneous reactions (electrode reactions)	$E^{[a]}$	α	$k_s^{[b]}$
Heterogeneous reactions (electrode reactions)	E^a	α	k_s^b
$(C_5H_4Cl)CpTiCl_2$ + e = $(C_5H_4Cl)CpTiCl_2^-$	–1.20±0.03	0.5	6×10^{-3}
$(C_5H_4Cl)CpTiCl$ + + e = $(C_5H_4Cl)CpTiCl$	–0.73±0.03	0.5	1×10^{-3}

Homogeneous reactions	K^c	k_f^d	k_b^d
$(C_5H_4Cl)CpTiCl_2^-$ = $(C_5H_4Cl)CpTiCl$ + Cl^-	6×10^{-4}	2×10^4	3.3×10^7
$(C_5H_4Cl)CpTiCl_2$ = $(C_5H_4Cl)CpTiCl^+$ + Cl^-	4.5×10^{-11}	1	2.2×10^{10}

[a] In V vs. Fc^+/Fc. [b] In cm s^{-1} [c] In Mx according to the reaction in question. [d] In M-x s^{-1} according to the reaction in question. [e] Equilibrium constants are determined automatically from square scheme thermodynamics in DigiSim.

Other input parameters (in DigiSim syntax):

E_{start} (V): 0
E_{switch} (V): -2
E_{end} (V): 0
Temperature (K): 298.2
Ru (Ohms): 0
Cdl (F): 0

cycles: 1
electrode geometry: planar
area (cm^2): 0.00785
diffusion: semi-infinite
pre-equilibrium: disabled

Diffusion coefficients: 1.3×10^{-5} $cm^s s^{-1}$

$(C_5H_4Cl)_2TiCl_2$: Simulation parameters and input values (using DigiSim syntax) for the simulation of the voltammograms of $(C_5H_4Cl)_2TiCl_2$.

Heterogeneous reactions (electrode reactions)	E^a	α	k_s^b
$(C_5H_4Cl)_2TiCl_2$ + e = $(C_5H_4Cl)_2TiCl_2^-$	-1.12±0.03	0.5	9×10^{-3}
$(C_5H_4Cl)_2TiCl^+$ + e = $(C_5H_4Cl)_2TiCl$	-0.57±0.03	0.5	1×10^{-3}

Homogeneous reactions	K^c	k_f^d	k_b^d
$(C_5H_4Cl)_2TiCl_2$ = $(C_5H_4Cl)_2TiCl^+$ + Cl^-	4.2×10^{-12}	1	3.5×10^{13}
$(C_5H_4Cl)_2TiCl_2^-$ = $(C_5H_4Cl)_2TiCl$ + Cl^-	7×10^{-4}	2×10^4	2.9×10^{-6}

[a] In V vs. Fc^+/Fc. [b] In cm s^{-1} [c] In Mx according to the reaction in question. [d] In M-x s^{-1} according to the reaction in question. [e] Equilibrium constants are determined automatically from square scheme thermodynamics in DigiSim.

Other input parameters (in DigiSim syntax):

E_{start} (V): 0
E_{switch} (V): -2
E_{end} (V): 0
Temperature (K): 298.2
Ru (Ohms): 0
Cdl (F): 0

cycles: 1
electrode geometry: planar
area (cm^2): 0.00785
diffusion: semi-infinite
pre-equilibrium: disabled

Diffusion coefficients: 1×10^{-5} $cm^s s^{-1}$

$(C_5H_4tBu)CpTiCl_2$ Simulation parameters and input values (using DigiSim syntax) for the simulation of the voltammograms of $(C_5H_4tBu)CpTiCl_2$.

Heterogeneous reactions (electrode reactions)	E^a	α	k_s^b
$(C_5H_4tBu)CpTiCl_2$ + e = $(C_5H_4tBu)CpTiCl_2^-$	-1.36±0.03	0.5	0.005
$(C_5H_4tBu)CpTiCl^+$ + e = $(C_5H_4tBu)CpTiCl$	-0.97±0.03	0.5	6×10^{-5}
Homogeneous reactions	K^c	k_f^d	k_b^d
$(C_5H_4tBu)CpTiCl_2^-$ = $(C_5H_4tBu)CpTiCl$ + Cl^-	8×10^{-5}	2×10^4	2.5×10^8
$(C_5H_4tBu)CpTiCl_2$ = $(C_5H4\,tBu)CpTiCl^+$ + Cl^-	4.5×10^{-11}	1	2.2×10^{10}

[a] In V vs. Fc^+/Fc. [b] In cm s^{-1} [c] In Mx according to the reaction in question. [d] In M-x s^{-1} according to the reaction in question. [e] Equilibrium constants are determined automatically from square scheme thermodynamics in DigiSim.

Other input parameters (in DigiSim syntax):

E_{start} (V): 0
E_{switch} (V): -1.60
E_{end} (V): 0
Temperature (K): 298.2
Ru (Ohms): 0
Cdl (F): 0

cycles: 1
electrode geometry: planar
area (cm^2): 0.00785
diffusion: semi-infinite
pre-equilibrium: disabled

Diffusion coefficients:6×10^{-6} $cm^s s^{-1}$

$(C_5H_4COOMe)_2CpTiCl_2$ Simulation parameters and input values (using DigiSim syntax) for the simulation of the voltammograms of $(C_5H_4COOMe)_2TiCl_2$.

Heterogeneous reactions (electrode reactions)	E^a	α	k_s^b
$(C_5H_4COOMe)_2TiCl_2$ + e = $(C_5H_4COOMe)_2TiCl_2^-$	-1.01±0.03	0.5	6×10^{-2}
$(C_5H_4COOMe)TiCl_2^+$ + e = $(C_5H_4COOMe)TiCl_2$	-0.79±0.03	0.5	1×10^{-2}
$TiCl_{22}^+$ + e = $TiCl_2$	-0.58±0.03	0.5	1×10^{-2}
Homogeneous reactions	K^c	k_f^d	k_b^d
$(C_5H_4COOMe)TiCl_2$ = $TiCl_2^+$ + (C_5H_4COOMe)	1×10^4	0.5	5×10^{-5}
$(C_5H_4COOMe)_2TiCl_2^-$ = $(C_5H_4COOMe)TiCl_2$ + $(C5\,H_4COOMe^-$	1×10^4	0.12	1.2×10^{-5}

[a] In V vs. Fc^+/Fc. [b] In cm s^{-1} [c] In Mx according to the reaction in question. [d] In M-x s^{-1} according to the reaction in question. [e] Equilibrium constants are determined automatically from square scheme thermodynamics in DigiSim.

Other input parameters (in DigiSim syntax):

E_{start} (V): -020
E_{switch} (V): -1.60
E_{end} (V): -0.20
Temperature (K): 298.2
Ru (Ohms): 0
Cdl (F): 0

cycles: 1
electrode geometry: planar
area (cm^2): 0.00785
diffusion: semi-infinite
pre-equilibrium: disabled

Diffusion coefficients:1×10^{-5} $cm^s s^{-1}$

Contributions to the reaction free energies of the hydrochloride adduct formation

Table E.12.: Contributions to the reaction free energies and final ΔG values of the hydrochloride adduct for- mation on the PW6B95-D3/def2-QZVP-COSMO-RS//TPSS-D3-cosmo/def2-TZVP level of theory.

	ΔE	ΔG_{RRHO}	ΔH_{RRHO}	$T\Delta S_{RRHO}$	ΔG_{solv}	ΔH_{solv}	$T\Delta S_{solv}$	ΔG	ΔH	$T\Delta S$
$Cp_2TiCl_2 \cdot PyH$	-26.03	14.5	1.7	-12.79	10.55	19.39	8.83	-0.98	-4.94	-3.96
$Cp_2TiCl_2 \cdot CollH$	-26.64	16.22	1.76	-14.46	7.79	15.34	7.55	-2.63	-9.55	-6.92
$Cp_2TiCl_2 \cdot Et_3NH$	-24.33	14.88	1.51	-13.36	0.91	7.67	6.76	-8.55	-15.15	-6.6
$(C_5H_4Cl)_2TiCl_2 \cdot PyH$	-29.1	15.04	1.82	-13.22	9.84	19.26	9.42	-4.22	-8.02	-3.8
$(C_5H_4Cl)_2TiCl_2 \cdot CollH$	-26.58	16.25	1.62	-14.63	6.61	14.6	7.99	-3.73	-10.37	-6.64
$(C_5H_4Cl)_2TiCl_2 \cdot Et_3NH$	-29.81	16.43	1.71	-14.73	3.1	10.92	7.82	-10.28	-17.19	-6.91
$(C_5H_4tBu)_2TiCl_2 \cdot PyH$	-22.65	15.22	1.62	-13.6	11.93	20.98	9.06	4.5	-0.05	-4.55
$(C_5H_4tBu)_2TiCl_2 \cdot CollH$	-23.95	15.98	1.51	-14.47	10.7	17.05	6.35	2.73	-5.39	-8.12
$(C_5H_4tBu)_2TiCl_2 \cdot Et_3NH$	-22.37	16.35	1.65	-14.7	5.62	11.57	5.95	-0.41	-9.16	-8.75

Table E.13.: Contributions to the reaction free energies and final ΔG values of the hydrochloride adduct for- mation on the B3LYP-D3/def2-QZVP-COSMO-RS//TPSS-D3-cosmo/def2-TZVP level of theory.

	ΔE	ΔG_{RRHO}	ΔH_{RRHO}	$T\Delta S_{RRHO}$	ΔG_{solv}	ΔH_{solv}	$T\Delta S_{solv}$	ΔG	ΔH	$T\Delta S$
$Cp_2TiCl_2 \cdot PyH$	-23.44	14.5	1.7	-12.79	10.55	19.39	8.83	1.61	-2.35	-3.96
$Cp_2TiCl_2 \cdot CollH$	-24.48	16.22	1.76	-14.46	7.79	15.34	7.55	-0.47	-7.39	-6.92
$Cp_2TiCl_2 \cdot Et_3NH$	-21.82	14.87	1.51	-13.36	0.91	7.67	6.76	-6.03	-12.63	-6.6
$(C_5H_4Cl)_2TiCl_2 \cdot PyH$	-26.37	15.04	1.82	-13.22	9.84	19.26	9.42	-1.49	-5.29	-3.8
$(C_5H_4Cl)_2TiCl_2 \cdot CollH$	-25.49	16.25	1.62	-14.63	6.61	14.6	7.99	-2.64	-9.28	-6.64
$(C_5H_4Cl)_2TiCl_2 \cdot Et_3NH$	-25.53	16.43	1.71	-14.73	3.1	10.92	7.82	-6	-12.91	-6.91
$(C_5H_4tBu)_2TiCl_2 \cdot PyH$	-19.97	15.22	1.62	-13.6	11.93	20.98	9.06	7.18	2.63	-4.55
$(C_5H_4tBu)_2TiCl_2 \cdot CollH$	-21.56	15.98	1.51	-14.47	10.7	17.05	6.35	5.12	-3	-8.12
$(C_5H_4tBu)_2TiCl_2 \cdot Et_3NH$	-18.95	16.35	1.65	-14.7	5.62	11.57	5.95	3.02	-5.73	-8.75

F. Supporting Information to Chapter 8

Appendix F contains:

- Synthetic details
- Details on the crystal structure determinations
- Calculated CD spectra including the rotatory strength for CAM-B3LYP

Synthetic details

General Information. All reactions with moisture or air sensitive substances were performed under argon according by using Schlenk techniques with oven-dried glass ware. Thin-layer chromatography was performed with aluminum TCL plates (silica gel 60F254). Detection was carried out under UV light with 254 and 366 nm. Products were purified via column chromatography by using silica gel 60 (70–230 mesh). ^{1}H NMR chemical shifts are reported on the δ scale (ppm) relative to residual nondeuterated solvent as the internal standard. The ^{13}C $\{^{1}$H$\}$ NMR chemical shifts are reported on the δ scale (ppm) relative to deuterated solvent as the internal standard. Signals were assigned on the basis of ^{1}H, ^{13}C, H,H–COSY, HMQC, and HMBC NMR experiments. Mass spectra were recorded as EI or as QToF-ESI spectra. Chiral analytical and semipreparative stationary phases CHIRALPAK IA (column size 0.46 cm × 25 cm respectively 1 cm × 25 cm, equipped with precolumns of the same diameter and 2 cm length) and chiral analytical and preparative stationary phases CHIRALPAK IB (column size 0.46 cm × 25 cm respectively 2.5 cm × 20 cm, equipped with precolumns of the same diameter and 2 cm length) were applied and solvent mixtures of *n*-heptane (HPLC quality) and chloroform (p.a. stabilized with ethanol) and *n*-hexane (HPLC qualitiy) and ethanol (p.a.) and 2-propanol (HPLC quality) were used. Circular dichroism spectroscopy was performed using ethanol or chloroform (p.a.) as solvents. Most solvents were dried, distilled and stored under argon according to standard procedures. All chemicals were used as received from commercial sources. (*rac*)-4,15-Dibromo[2.2]paracyclophane[522] was prepared according to a literature protocol.

(*rac*)-4,15-Dihydroxy[2.2]paracyclophane {(*rac*)-4}. 7.20 mL of *t*BuLi (1.9 M in pentane, 13.70 mmol) were added to 40 mL of dry THF at –78 °C and stirred for 5 min. To the flashy yellow solution (*rac*)-**2** (1.00 g, 2.74 mmol) dissolved in THF (20 mL) was added via a syringe. The mixture is stirred for 1 h at –78 °C turning from flashy yellow to pale yellow. Then B(O*i*Pr)$_3$ (2.06 g, 2.53 mL, 10.96 mmol) was added, and the solution was allowed to slowly warm to room temperature turning from yellow to colorless, and precipitate was formed. Subsequently aqueous KOH (0.5 M, 2.74 mL, 1.38 mmol) and H_2O_2 (35%, 2.00 mL, 21.92 mmol) were added,

and the solution was stirred for 1 h. The reaction mixture was poured into water and was extracted with Et_2O (3 × 60 mL). The combined organic layers were washed with brine and dried over $MgSO_4$. The solvents were evaporated, and crude **4** was purified by column chromatography on silica gel (cyclohexane/ethyl acetate 2:1 (v/v); $R_f = 0.6$), yield 0.53 g (2.22 mmol, 81%). The product is a white solid: mp 227 °C (decomposing); ^{1}H NMR (500.1 MHz, acetone-d_6, 298 K) δ = 2.67–2.76 (m, 2 H, H-1, H-2), 2.79–2.82 (m, 4 H, H-9, H-10), 3.15–3.22 (m, 2 H, H-1, H-2), 5.67 (d, $^4J_{5,7} = {}^4J_{16,12}$= 1.7 Hz, 2 H, H-5, H-16), 6.08 (dd, $^3J_{7,8} = {}^3J_{12,13}$ = 7.6 Hz, $^4J_{7,5} = {}^4J_{12,16}$ = 1.7 Hz, 2 H, H-7, H12), 6.83 (d, $^3J_{8,7} = {}^3J_{13,12}$ = 7.6 Hz, 2 H, H-8, H-13) 7.64 (s, 2 H,O–H) ppm; ^{1}3C {^{1}H} NMR (125.8 MHz, acetone-d_6, 298 K) δ = 30.5 (C-1, C-2), 35.0 (C-9, C-10), 121.71 (C-5, C-16), 124.1 (C-7, C-12), 126.7 (C-3, C-14), 130.8 (C-8, C-13), 141.4 (C-6, C-11), 156.6 (C-4, C-15) ppm; MS (EI) m/z (%) = 240.1 (70) $[C_{16}H_{16}O_2]^{+\cdot}$, 120.1 (100) $[C_7H_8O]^+$, 91 (20) $[C_7H_7]^+$; EI-HRMS m/z calcd. for $[C_{16}H_{16}O_2]^{+\cdot}$ 240.1150, found 240.1157.

Separation of Enantiomers. HPLC [chiral phase (semipreparative): CHIRALPAK IA; *n*-hexane/EtOH (90:10); f = 5.0 mL/min; loading 20 mg of racemic material per run] t_R = 12.80 [(+)-(R_P)-**4** $[\alpha]_D^{20}$ = +37.0 (c = 4.4645 g/mL, THF), >99.9% ee], 15.23 [(–)-(S_P)-**4** $[\alpha]_D^{20}$ = –35.8 (c = 4.4605 g/mL, THF), 99.8% ee] min.

(R_P)- and (S_P)-[2.2]Paracyclophane-4,15-dicarboxylic acid (R_P)- and (S_P)-5. KO*t*Bu (0.337 g, 3.00 mmol) was dissolved in water (0.54 mL, 3.00 mmol), and THF (40 mL) and enantiomerically pure (R_P)- or (S_P)-**8** (0.150 g, 0.25 mmol) was added. The resulting mixture was stirred overnight. The THF was evaporated and water was added. The mixture was acidified with aq. HCl (2 M), and the white precipitated was filtered off and washed with water to give the enantiomerically pure target compound. Yield: 0.057 g (0.19 mmol, 77%). The product is a white solid: mp > 250 °C; ^{1}H NMR (400.1 MHz, DMSO d_6 , 293 K) δ = 2.83–2.94 (m, 2 H, H-1, H-2), 2.94–3.05 (m, 2 H, H-9, H-10), 3.09–3.21 (m, 2H, H-9, H-10), 3.89–4.02 (m, 2 H, H-1, H-2), 6.47 (d, $^3J_{8,7} = {}^3J_{13,12}$ = 7.8 Hz, 2 H, H-8, H-13), 6.62 (dd, $^3J_{7,8} = {}^3J_{12,13}$ = 7.8 Hz, $^4J_{7,5} = {}^4J_{12,16}$ = 2.0 Hz, 2 H, H-7, H12) 7.13 (d, $^4J_{5,7} = {}^4J_{16,12}$ = 2.0 Hz, 2 H, H-5, H-16), 12.59 (s, 2 H, CO2–H) ppm; ^{13}C ^{1}H NMR (100.6 MHz, DMSO d_6, 293 K) δ = 34.1 (C-1, C-2), 35.1 (C-9, C-10), 131.5 (C-4, C-15), 133.8 (C-5, C-16), 135.1 (C-8, C-13), 135.7 (C-7, C-12), 139.8 (C-6, C-11), 142.2 (C-3, C-14), 176.9 (CO_2H) ppm; MS (ESI negative mode) m/z (%) = 295.1 (100) $[C_{18}H_{15}O_4]^-$; ESI-HRMS m/z calcd. for $[C_{18}H_{15}O_4]^-$ 295.0976, found 295.0978.
Compound (+)-(S_P)-**5**: $[\alpha]_D^{20}$ = +65.6 (c = 4.1 mg/mL, EtOH). Compound (–)-(R_P)-**5**: $[\alpha]_D^{20}$ = –68.2 (c = 4.40 mg/mL, EtOH).

(*rac*)-4,15-Diformyl[2.2]paracyclophane (*rac*)-6. 7.20 mL of *t*BuLi (1.9 M in pentane, 13.70 mmol) were added to 40 mL of dry THF at –78 °C and stirred for 5 min. To the flashy yellow solution (*rac*)-**2** (1.00 g, 2.74 mmol) dissolved in THF (20 mL) was added via a syringe. The mixture was stirred for 1 h at –78 °C turning from flashy yellow to pale yellow. Then DMF (0.72 g, 0.76 mL, 10.00 mmol) was added, and the solution was allowed to slowly warm to room temperature turning from yellow to colorless. Subsequent aqueous HCl (4 M, 7.5 mL, 30 mmol) was added, and the mixture was stirred for further 30 min. Water and Et_2O were added, and the phases were separated. The aqueous layer was extracted with Et_2O (3 × 50 mL), and the

combined organic phases were washed with 0.5 M HCl, saturated aqueous $NaHCO_3$ solution, and brine and dried over $MgSO_4$. The solvent was removed under reduced pressure and crude **6** was purified via column chromatography on silica gel (cyclohexane/ ethyl acetate, 5:1, v/v) $R_f = 0.6$), yield 0.645 mg (2.44 mmol, 89%). The product is a white solid. The analytical data were in accordance with the literature data.[524]

Separation of Enantiomers. HPLC [chiral phase (semiprepar-ative): CHIRALPAK IA; *n*-hexane/EtOH (90:10); $f = 5.0$ mL/min; loading 30 mg of racemic material per run] t_R = 14.53 [(+)-(S_P)-**6** $[\alpha]_D^{20} = +231.2$ ($c = 4.645$ g/mL, THF), > 99.9% ee], 18.76 [(–)-(R_P)-**6** $[\alpha]_D^{20}$ = –226.3 ($c = 4.715$ g/mL, THF), > 99.9% ee] min.

(*rac*)-Di(4-bromophenyl)[2.2]paracyclophane-4,15-dixarboxylate (*rac*)-7. (*rac*)-**6** (0.200 g, 0.64 mmol) was dissolved in dry diethyl ether (40 mL). Oxalyl chloride (0.12 mL, 1.48 mmol) and one drop of DMF were added, and the resulting mixture was stirred for 2 h at room temperature. The solvent was evaporated, and the white residue was dissolved in dry dichloromethane (10 mL). Subsequently dry triethylamine (10 mL) was added, and the solution turned red. After that 4-bromophenol (0.463 g, 2.68 mmol) was added, and the solution turned yellow. The solution was stirred at room temperature overnight and then poured into ice water. The mixture was acidified, and the layers were separated. The aqueous layer was extracted with dichloromethane (3 × 10 mL). The combined organic layers were washed with saturated aq. $NaHCO_3$ and brine and dried over Mg_2SO_4. The product was obtained as a white powder. If needed it can be further purified by column chromatography on silica gel (eluent: cyclohexane/ethyl acetate, 2:1 (v/v), $R_f = 0.8$). The product was obtained as a white powder. Yield: 0.356 g (0.59 mmol, 92%); mp 94 °C; ^{1}H NMR (400.1 MHz, $CDCl_3$, 293 K) δ = 3.08–3.17 (m, 4 H, H-1, H-2, H-9, H-10), 3.22–3.33 (m, 2 H, H-9, H-10), 4.08–4.15 (m, 2 H, H-1, H-2), 6.71 (d, 2 H, H-8, H-13, $^3J_{8,7} = {}^3J_{13,12} = 7.9$ Hz), 6.75 (dd, 2 H, H-7, H-12 $^3J_{7,8} = {}^3J_{12,13} = 7.9$ Hz, $^4J_{7,5} = {}^4J_{12,16} = 1.9$ Hz), 7.15 (d, 4 H, H-phenyl $^3J = 8.9$ Hz), 7.39 (d, 2 H, H-5, H-16 $^4J_{5,7} = {}^4J_{16,12} = 1.9$ Hz) 7.59 (d, 4 H, H-phenyl $^3J = 8.9$ Hz) ppm; ^{13}C ^{1}H NMR (100.6 MHz, $CDCl_3$, 293 K) δ = 34.9 (C-9, C-10), 35.8 (C-1, C-2), 119.2 (C-21), 123.7 (C-19), 130.1 (C-4, C-15), 132.8 (C-20), 134.6 (C-8, C-13), 135.7 (C-5, C-12), 137.0 (C-7, C-12), 140.8 (C-6, C-11), 144.3 (C-3, C-14), 150.0 (C-18), 165.0 (C-17) ppm; MS (EI) m/z (%) 606.0 (5) $[C_{30}H_{22}Br_2O_4]^{+\cdot}$, 433.0 (100) $[C_{24}H_{18}BrO_3]^+$, 131.0(70) $[C_9H_7O]^+$; ESI-HRMS m/z calcd. for $[C_{30}H_{22}O_4Br_2]^+$ 603.9885, found 603.9880. Elemental analysis calcd. (%) for $C_{30}H_{22}O_4Br_2$ (606.30) C 59.43, H 3.66. Found: C 59.21, H 3.91.

Separation of Enantiomers. HPLC [chiral phase (semiprepar): CHIRALPAK IA; *n*-hexane/EtOH (80:20); $f = 5.0$ mL/min; loading 20 mg of racemic material per run] t_R = 10.63 [(+)-(S_P)-**7** $[\alpha]_D^{20} = +164.8$ ($c = 3.405$ g/mL, THF), 99.9% ee], 14.27 [(–)-(R_P)-**7** $[\alpha]_D^{20}$ = –161.3 ($c = 4.080$ g/mL, THF), 99.7% ee] min.

(R_P)- and (S_P)-4,15-Di(trifluoromethanesulfonate)[2.2]-paracyclophane (R_P)- and (S_P)-8. Enantiomerically pure (S_P)- or (R_P)-**4** (0.200 g, 0.83 mmol) was dissolved in dry triethylamine (1.15 mL, 8.30 mmol) and dry CH_2Cl_2 (20 mL). The solution was cooled to –78 °C, and triflic anhydride (0.35 mL, 2.08 mmol) was added slowly via syringe. The reaction

mixture was allowed to warm to room temperature. After that the solution was acidified with aq. HCl (2 M), and the layers were separated. The aqueous layer was extracted with CH_2Cl_2 (2 × 20 mL), and the combined organic layers were washed with saturated $NaHCO_3$ and brine and dried with $MgSO_4$. The solvents were evaporated, and the crude product was purified by column chromatography on silica gel (eluent: cyclohexane/ethyl acetate, 2:1 (v/v); R_f = 0.7). The product was obtained a colorless highly viscous oil, whereas (*rac*)-**8** was obtained as a pale yellow solid (mp 77 °C). Yield: 0.352 g (0.70 mmol, 84%); ^{1}H NMR (400.1 MHz, $CDCl_3$, 293 K) δ = 2.92–3.01 (m, 2 H, H-1, H-2), 3.04–3.13 (m, 4 H, H-9, H-10), 3.32–3.41 (m, 2 H, H-1, H-2), 6.23 (d, $^4J_{5,7}$ = $^4J_{16,12}$ = 1.7 Hz, 2 H, H-5, H-16), 6.61 (dd, $^3J_{7,8}$ = $^3J_{12,13}$ = 8.0 Hz, $^4J_{7,5}$ = $^4J_{12,16}$ = 1.7 Hz, 2 H, H-7, H12), 7.02 (d, $^3J_{8,7}$ = $^3J_{13,12}$ = 8.0 Hz, 2 H, H-8, H-13) ppm; ^{13}C ^{1}H NMR (100.6 MHz, $CDCl_3$, 293 K) δ = 30.4 (C-1, C-2), 34.4 (C-9, C-10), 120.4 [CF3, 1JC,F = 318 Hz], 127.5 (C-5, C-16), 132.0 (C-3, C-14), 132.1 (C-7, C-12), 132.7 (C-8, C-13), 142.8 (C-6, C-11), 148.7 (C-4, C-15) ppm; MS (EI) m/z (%) = 504.0 (30) $[C_{18}H_{14}F_6O_6S_2]^{+\cdot}$, 371.0 (40) $[C_{17}H_{14}O_4F_3S]^+$, 252.0 (100) $[C_9H_7O_3F_3S]^+$, 91 (25) $[C_7H_7]^+$; EI-HRMS m/z calcd. for $[C_{18}H_{14}F_6O_6S_2]^{+\cdot}$ 504.0136, found 504.0136.
Compound (+)-(S_P)-**8**: $[\alpha]_D^{20}$ = +16.7 (*c*= 8.050 mg/mL, $CHCl_3$). Compound (–)-(R_P)-**8**: $[\alpha]_D^{20}$ = +17.3 (*c* = 4.180 mg/mL, $CHCl_3$).

(R_P)- and (S_P)-4,15-Diethynyl[2.2]paracyclophane (R_P)- and (S_P)-(9). Enantiomerically pure (R_P)- or (S_P)-**6** (0.500 g, 1.89 mmol) and Cs_2CO_3 (2.407 g, 7.56 mmol) were suspended in anhydrous MeOH (40 mL), and the Bestmann–Ohira reagent (1.390 g, 7.56 mmol) was added. The resulting mixture was stirred for 24 h at room temperature. Subsequently another portion of Cs_2CO_3 (0.722 g, 2.52 mmol) and the Bestmann–Ohira reagent (0.463 g, 2.52 mmol) was added and stirred for further 12 h. After that CH_2Cl_2 and water were added, and the layers were separated. The aqueous layer was extracted CH_2Cl_2 (2 × 20 mL), and the combined organic layers were washed with brine and dried over $MgSO_4$. The solvent was evaporated, and the crude product was purified, if necessary, by column chromatography on silica gel (eluent: 5% of ethyl acetate in cyclohexane, R_f = 0.8). The product was obtained as a pale yellow powder. Yield: 0.475 g (1.85 mmol, 98%). The analytical data were in accordance with the literature data.[524]
Compound (+)-(S_P)-**9**: $[\alpha]_D^{20}$ = +339 (*c*= 2.760 mg/mL, $CHCl_3$). Compound (–)-(R_P)-**9**: $[\alpha]_D^{20}$ = –342 (*c*= 2.355 mg/mL, $CHCl_3$).

(*rac*)-4,15-Diiodo[2.2]paracyclophane (*rac*)-10. 7.20 mL of *t*BuLi (1.9 M in pentane, 13.70 mmol) were added to 40 mL of dry THF at –78 °C and stirred for 5 min. To the flashy yellow solution (*rac*)-**2** (1.000 g, 2.74 mmol) dissolved in THF (20 mL) was added via a syringe. The mixture is stirred for 1 h at –78 °C turning from flashy yellow to pale yellow. Then iodine (1.905 g, 7.50 mmol) was added. The solution was allowed to slowly warm to room temperature. The reaction mixture was diluted with CH_2Cl_2 and water, and the layers separated. The organic layer was washed with saturated aqueous Na_2SO_3, water, and brine and dried over $MgSO_4$. The solvent was evaporated, and the crude product was purified by column chromatography on silica gel (eluent: cyclohexane, R_f = 0.7). The product is a white powder. Yield: 2.060 g (4.34 mmol,

80%); mp 161 °C; ^{1}H NMR (400.1 MHz, $CDCl_3$, 293 K) δ = 2.84–2.92 (m, 2 H, H-1, H-2), 3.03–3.20 (m, 2 H, H-1, H-2), 3.21–3.22 (m, 4 H, H-9,H-10), 6.50 (dd, 2 H, H-7, H-12, $^3J_{7,8}$ = $^3J_{12,13}$ = 7.8 Hz, $^4J_{7,5}$ = $^4J_{12,16}$ = 1.8 Hz), 6.91 (d, 2 H, H-5, H-16, $^4J_{5,7}$ = $^4J_{16,12}$ = 1.8 Hz) 7.20 (d, 2 H, H-8, H-13, $^3J_{8,7}$ = $^3J_{13,12}$ = 7.8 Hz) ppm; ^{13}C ^{1}H NMR (100.6 MHz, $CDCl_3$, 293 K δ = 34.8 (C-1, C-2), 36.9 (C-9, C-10), 104.9 (C-4, C-15), 129.5 (C-8, C-13), 131.9 (C-7, C-12), 141.0 (C-6, C-11), 142.7 (C-3, C-14), 142.8 (C-5, C-16); MS (EI) m/z (%) 459.9 (100) $[C_{16}H_{14}I_2]^+$, 229.9 (60) $[C_8H_7I]^+$; ESI-HRMS m/z calcd. for $[C_{16}H_{14}I_2]^+$ 459.9185, found 459.9192.

(R_P)-4,15-Diiodo[2.2]paracyclophane (R_P)-(10). Enantiomerically pure (R_P)-**12** (0.100 g, 0.43 mmol) was dissolved in conc. HCl (1.5 mL) and diluted with water (5 mL). The stirred solution was cooled to 0 °C and $NaNO_2$ (0.070 g, 1.03 mmol) dissolved in water (2.5 mL) was slowly added. After 30 min KI (0.374 g, 2.25 mmol) dissolved in water (2 mL) was added at 0 °C. The solution was stirred at 0 °C for 30 min and was then heated to 80 °C for 2 h. After cooling to room temperature the solution was extracted with dichloromethane (3 × 20 mL). The combined organic phases were washed with saturated aqueous $NaHSO_3$ solution and brine and dried over $MgSO_4$. The solvent was evaporated under reduced pressure and crude **10** was purified by column chromatography on silica gel (eluent: cyclohexane, R_f = 0.7). The product is a white powder. Yield: 0.152 g (0.32 mmol, 74%). Suitable crystals for X-ray diffraction analysis were grown from a mixture of cyclohexane and ethyl acetate.
Compound (–)-(R_P)-**10**: $[\alpha]_D^{20}$ = –245 (c = 3.20 mg/mL, $CHCl_3$).

***(rac)*-[2.2]Paracyclophane-4,15-diazide [*rac*)-11.** 7.20 mL of tBuLi (1.9 M in pentane, 13.70 mmol) were added to 40 mL of dry THF at –78 °C and stirred for 5 min. To the flashy yellow solution (*rac*)-**3** (1.000 g, 2.74 mmol) dissolved in THF (20 mL) was added via a syringe. The mixture was stirred for 1 h at –78 °C turning from flashy yellow to pale yellow. Then p-toluenesulfonyl azide (1.608 g, 8.16 mmol) in 10 mL of dry THF was added slowly to the stirred solution. The solution was allowed to slowly warm to room temperature, turning from pale yellow to red and then to black. The reaction mixture is poured into saturated ammonium chloride solution, and the aqueous layer was extracted with dichloromethane (3 × 60 mL). The combined organic layers were washed with brine and dried over $MgSO_4$. The solvent was removed under reduced pressure and crude **11** was purified via column chromatography on silica gel (cyclohexane, R_f = 0.4) to remove remaining p-toluenesulfonate. The mixture still contains monoazide byproduct. The product mixture is a pale yellow solid: ^{1}H NMR (400.1 MHz, $CDCl_3$, 293 K) δ = 2.75–2.84 (m, 2 H, H-1, H-2), 2.99–3.12 (m, 4 H, H-1, H-9), 3.14–3.23 (m, 2 H, H-1, H-2), 6.01 (d, 2 H, H-5, H-16 $^4J_{5,7}$ = $^4J_{16,12}$ = 1.7 Hz), 6.38 (dd, 2 H, H-7, H-12, $^3J_{7,8}$ = $^3J_{12,13}$ = 7.9 Hz, $^4J_{7,5}$ = $^4J_{12,16}$ = 1.7 Hz), 6.84 (d, 2 H, H-8, H-13, $^3_{J8,7}$ = $^3J_{13,12}$ = 7.9 Hz) ppm; ^{13}C ^{1}H NMR (100.6 MHz, $CDCl_3$, 293 K) δ = 30.8 (C-1, C-2), 34.7 (C-9, C-10), 123.8 (C-5, C-16), 128.0 (C-7, C-12), 131.1 (C-8, C-13), 133.2 (C-3, C-14), 140.8 (C-6, C-11), 146.2 (C-4, C-15) ppm; MS (EI) m/z (%) = 290.1 (25) $[C_{16}H_{14}N_6]^{+\cdot}$; EI-HRMS m/z calcd. for $[C_{16}H_{14}N_6]^{+\cdot}$ 290.1280, found 290.1282.

***(rac)*-4,15-Diamino[2.2]paracyclophane (*rac*)-12.** A roundbottom flask was charged with tetrabutylammonium iodide (1.272 g, 3.44 mmol) and $NaBH_4$ (2.612 g, 68.80 mmol) under

an argon atmosphere. Subsequently, (*rac*)-**11** (1.000 g, 3.44 mmol), dissolved in 26.5 mL of dry THF, and 21.8 mL of water were added, and the solution was stirred for 48 h at room temperature. Afterwards additional $NaBH_4$ (1.306 g, 34.40 mmol) was added, and the mixture was stirred for further 24 h. The reaction mixture was then poured into water and was extracted with Et_2O (4 × 50 mL). The combined organic layers were washed with brine and dried over $MgSO_4$. The solvent was evaporated under reduced pressure, and crude **12** was purified by column chromatography on silica gel (cyclohexane/ethyl acetate 2:1, v/v + 5% triethylamine, R_f = 0.5), yield 0.729 g (3.06 mmol, 73%). The product is a brownish solid: mp 232 °C; ^{1}H NMR (400.1 MHz, CD_2Cl_2, 293 K) δ = 2.76–2.80 (m, 4 H, H-1*, H-2*), 2.83–2.97 (m, 4 H, H-9*, H-10*), 3.35 (bs, 4H, N–H), 5.45 (d, 2 H, H-5, H-16, $^4J_{5,7} = {}^4J_{16,12}$ = 1.8 Hz), 5.97 (dd, 2 H, H-7, H-12, $^3J_{7,8} = {}^3J_{12,13}$ = 7.7 Hz, $^4J_{7,5} = {}^4J_{12,16}$ = 1.8 Hz), 6.92 (d, 2 H, H-8, H-13, $^3J_{8,7} = {}^3J_{13,12}$ = 7.7 Hz) ppm; ^{13}C ^{1}H NMR (100.6 MHz, CD_2Cl_2, 293 K) δ = 29.7 (C-1*, C-2*), 35.0 (C-9*, C-10*), 120.7 (C-5, C-16), 122.3 (C-7, C-12), 124.0 (C-3, C-14), 128.7 (C-8, C-13), 140.8 (C-6, C-11), 146.2 (C-4, C-15) ppm (* assignment might be interchanged); MS (EI) m/z (%) = 238.1 (50) $[C_{16}H_{18}N_2]^{+\cdot}$, 119.0 (100) $[C_8H_9N]^+$, 91 (10) $[C_7H_7]^+$; EI-HRMS m/z calcd. for $[C_{16}H_{18}N_2]^{+\cdot}$ 238.1470, found 238.1472.

Separation of Enantiomers. HPLC [chiral phase (semipreparative): CHIRALPAK IB; *n*-hexane/EtOH (70:30); f = 20.0 mL/min; loading 40 mg of racemic material per run] t_R = 11.86 [(+)-(S_P)-**12** $[\alpha]_D^{20}$ = +95 (c = 2.55 g/mL, THF), 99.9% ee], 15.55 [(–)-(R_P)-**12** $[\alpha]_D^{20}$ = –97 (c = 3.15 g/mL, THF), 99.9% ee] min.

(*rac*)-4,15-Di-(4,4,5,5-tetramethyl-1,3,2-dioxborolan)-[2.2]- paracylcophane (*rac*)-13. 7.20 mL of *t*BuLi (1.9 M in pentane, 13.70 mmol) were added to 40 mL of dry THF at –78 °C and stirred for 5 min. To the flashy yellow solution (*rac*)-**3** (1.000 g, 2.74 mmol) dissolved in THF (20 mL) was added via a syringe. The mixture was stirred for 1 h at –78 °C turning from flashy yellow to pale yellow. Then 2-isopropoxy-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (1.530 g, 1.66 mL, 8.22 mmol) in 10 mL of dry THF was added, and the solution was allowed to slowly warm to room temperature, thereby first turning to pale green and then colorless. The reaction was quenched by the addition of water and ethyl acetate. The phases were separated, and the aqueous phase was extracted with ethyl acetate (2 × 40 mL). The combined organic phases were washed with water and brine and dried over $MgSO_4$. The crude product was purified by column chromatography on silica gel (cyclohexane/ethyl acetate, 10:1 v/v, R_f = 0.5), yield 1.196 g (2.60 mmol, 95%). The product is a white powder: mp 190 °C; ^{1}H NMR (400.1 MHz, acetone-d_6, 293 K) δ = 1.41 (s, 12 H, CH_3), 2.86–2.94 (m, 2H, H-1, H-2), 2.95–3.05 (m, 2H, H-9, H-10), 3.10–3.20 (m, 2H, H-9, H-10), 3.85–3.95 (m, 2H, H-1, H-2), 6.37 (d, 2H, H-8, H-13, $^3J_{8,7} = {}^3J_{13,12}$ = 7.7 Hz), 6.52 (dd, 2H, H-7, H12, $^3J_{7,8} = {}^3J_{12,13}$ = 7.7 Hz, $^4J_{7,5} = {}^4J_{12,16}$ = 2.0 Hz), 7.00 (d, $^4J_{5,7} = {}^4J_{16,12}$ = 2.0 Hz, 2H, H-5, H-16), ppm; ^{13}C ^{1}H NMR (100.4 MHz, acetone-d_6, 293 K) δ = 24.8 (CH_3), 25.1 (CH_3), 35.39 (C-1, C-2), 36.42 (C-1, C-2), 83.6 ($C_{quaternary}$), 134.3 (C-8, C-13), 135.3 (C-7, C-12), 138.8 (C-6, C-11), 140.9 (C-5, C-16), 148.0 (C-3, C-14) ppm. The carbon connected to the boron cannot be seen in the NMR spectrum because of its low intensity due to the coupling to the boron. ^{11}B ^{1}H NMR (128.4 MHz, acetone-d_6, 293 K) δ = 31.23

(bs) ppm; MS (ESI) m/z (%) = 461.3 (50) $[C_{28}H_{38}B_2O_4{}^+H]^{+\cdot}$, 483.3 (100) $[C_{28}H_{38}B_2O_4{}^+Na]^+$; ESI-HRMS m/z calcd. for $[C_{28}H_{38}B_2O_4{}^+Na]^+$ 483.2858, found 483.2863. Elemental analysis calcd. (%) for $C_{28}H_{38}B_2O_4$ (460.22): C 73.07, H 8.32. Found: C 72.96, H 8.18.

Separation of Enantiomers. HPLC [chiral phase (preparative): CHIRALPAK IB; *n*-hexane/$CHCl_3$ (98:2); f = 9.0 mL/min; loading 10 mg of racemic material per run] t_R = 15.03 [(+)-(S_P)-**13** $[\alpha]_D^{20}$ = +158 (c = 2.010 g/mL, EtOH), > 99.9% ee], 17.48 [(–)-(R_P)-**13** $[\alpha]_D^{20}$ = –158 (c = 2.29 g/mL, EtOH), 98.7% ee] min.

(S_P)-4,15-Di(4-bromophenyl)hydrazone[2.2]paracyclophane (S_P)-14. 4-Bromohydrazine (0.45 g) was dissolved in conc. H_2SO_4 (2 mL) and water (3 mL). EtOH (10 mL) was added to this solution and precipitate was filtered off. (S_P)-**5** (0.100 g, 0.273 mmol) was dissolved in CH_2Cl_2 (2 mL) and added to the 4-bromohydrazine solution. After keeping at room temperature overnight greenish crystals were formed, which were filtered off and carefully washed with water. These crystals were suitable for X-ray diffraction analysis: mp 218 °C; ^{1}H NMR (400.1 MHz, $CDCl_3$, 293 K) δ = 2.87–2.97 (m, 2 H, H-1, H-2), 2.98–3.09 (m, 2 H, H-9, H-10), 3.10–3.21 (m, 2 H, H-9, H-10), 3.69–3.80 (m, 2 H, H-1, H-2) 6.45 (d, 2 H, H-7, H-12, $^3J_{7,8}$ = $^3J_{12,13}$ = 7.7 Hz), 6.59 (d, 2 H, H-8, H-13, $^3J_{8,7}$ = $^3J_{13,12}$ = 7.7 Hz), 6.86 (s, 2 H, H-5, H-16), 7.03 (d, 4 H, H-phenyl, ^3J = 8.8 Hz), 7.41 (d, 4 H, H-phenyl, ^3J = 8.8 Hz), 7.68 (s, 2 H, CHN) ppm; MS (ESI) m/z (%) 601.06 (100) $[C_{30}H_{26}Br_2N_4{}^+H]^+$, 623.0 (35) $[C_{30}H_{26}Br_2N_4{}^+Na]^+$; ESI-HRMS m/z calcd. for $[C_{30}H_{26}Br_2N_4{}^+H]^+$ 601.0597, found 601.0574. Elemental analysis calcd. (%) for $C_{30}H_{26}Br_2N_4$ (602.36) · 2 H_2O: C 56.44, H 4.74. Found: C 56.89, H 4.75.
Compound (+)-(S_P)-**14**: $[\alpha]_D^{20}$ = +1056 (c= 4.42 mg/mL, THF).

Crystal structure determinations

Data were collected on a Nonius KappaCCD diffractometer equipped with a low temperature device (Cryostream, Oxford Cryosystems, 600er series) using graphite monochromated Mo Kα radiation (λ = 0.71073 Å). Intensities were measured by fine-slicing ω- and π-scans and corrected for background, polarization and Lorentzian effects. A semiempirical absorption correction from equivalent reflections was applied for all data sets according to Blessing's method.[541] The structures were solved by direct methods (SHELXL-97) and refined by full-matrix least-squares on F2 (SHELXL-97).[542,543] All non-hydrogen atoms were refined anisotropically. Hydrogen atoms at carbon were placed in calculated positions and refined isotropically using a riding model. For selected details of the crystallographic data see Table F.1. CCDC-1003202 [(–)-(R_P)-**7**], CCDC-1003203 [(–)-(R_P)-**10**], and CCDC-1003204 [(+)-(S_P)-**14**] contain the supplementary data for these structures. These data can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif, or by emailing data_request@ccdc.cam.ac.uk, or by contacting The Cambridge Crystallographic Data Centre, 12, Union Road, Cambridge CB2 1EZ, U.K.; fax: +44 1223 336033.

Table F.1.: Crystallographic Data for (–)-(R_P)-**7**, (–)-(R_P)-**10**, and (+)-(S_P)-**14**

parameters	(–)-(R_P)-**7**	(–)-(R_P)-**10**	(+)-(S_P)-**14**
formula	$C_{30}H_{22}Br_2O_4$	$C_{16}H_{14}I_2$	$C_{30}H_{26}Br_2N4$
M_r	606.3	460.07	602.37
T /K	123(2)	123(2)	123(2)
crystal system	orthorhombic	trigonal	monoclinic
space group	$P22_12_1$	$P3_2$	$C2$
crystal dimensions /mm	0.60 × 0.12 × 0.02	0.36 × 0.24 × 0.18	0.24 × 0.08 × 0.04
a /Å	6.8010(2)	11.6253(2)	33.742(2)
b /Å	16.1651(5)	11.6253(2)	7.9547(3)
c /Å	22.3501(8)	9.0313(2)	9.8768(7)
α /Å	90	90	90
β /Å	90	90	97.566(2)
γ /Å	90	120	90
V /Å^3	2457.14(14)	1057.04(5)	2627.9(3)
Z	4	3	4
ρ /mg m^3	1.639	2.168	1.523
μ /mm^{-1}	3.335	4.442	3.111
θ range /°	2.68–28.00	3.03–28.00	2.44–27.86
completeness /%	98.6	99.9	98
reflections measured	17603	19608	8898
unique reflections (R_{int})	5763 (0.0856)	3396 (0.0404)	5504 (0.0455)
data/restrains/parameters	5763/0/325	11266/19/649	5504/85/325
GoF on F^2	0.996	1.088	0.937
final R indices [$I > 2\sigma(I)$]	$R1 = 0.0389$	$R1 = 0.0238$	$R1 = 0.0382$
	$\omega R2 = 0.0826$	$\omega R2 = 0.0588$	$\omega R2 = 0.0789$
R indices all data	$R1 = 0.0515$	$R1 = 0.0244$	$R1 = 0.0572$
	$\omega R2 = 0.0867$	$\omega R2 = 0.0244$	$\omega R2 = 0.0842$
absolute structure parameter X	–0.013(8)	–0.04(3)	0.000(9)

Calculated CD spectra including the rotatory strengths

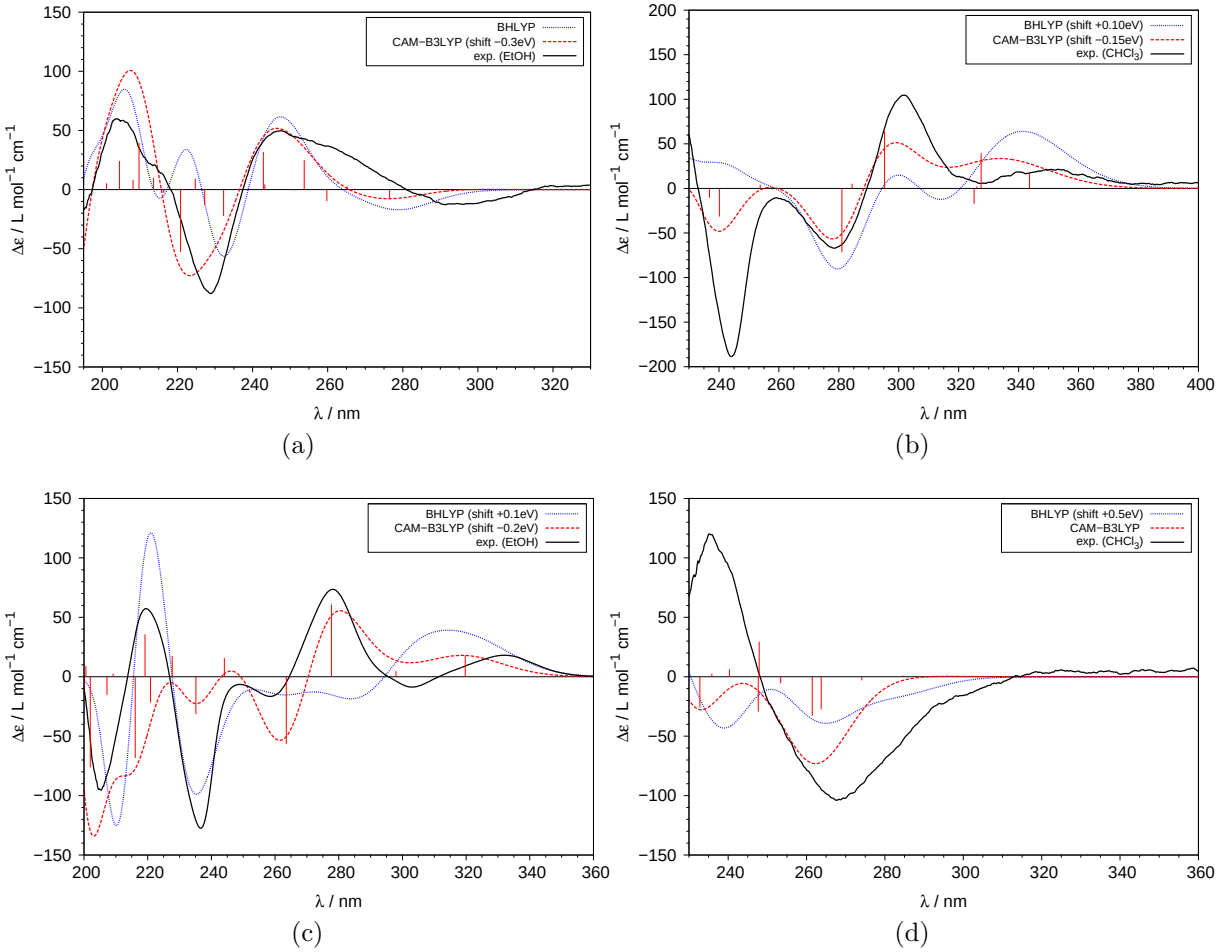


Figure F.1.: Experimental and simulated CD spectra of (a) (–)-(S_P)-**4**, (b) (+)-(S_P)-**6** (c) (–)-(R_P)-**10**, and (d) (+)-(S_P)-**13** including the rotatory strength for CAM-B3LYP.

Acknowledgments

First, I would like to thank my supervisor Prof. Dr. Stefan Grimme for the chance to work on highly interesting research topics during my thesis and for the freedom to chose several of them myself. His constant support, encouragement, and interest provided an excellent working environment.

I also thank Prof. Dr. Thomas Bredow for being my second referee.

I thank Prof. Dr. Arne Lützen and Prof. Dr. Diana Imhof for their agreement to be part of my examination committee.

I like to thank all my collaborators:
Prof. Dr. Gansäuer and his coworkers Dr. Daniel von Laufenberg, Dr. Christian Kube, Tobias Dahmen and Katharina Zimmer (Bonn); Prof. Dr. Arne Lützen and his coworkers Dr. Georg Meyer-Eppler, Lucia Vollbach and Tina Tenten (Bonn); Prof. Dr. Diana Imhof and her coworkers Henning Brewitz and Amelie Wißbrock (Bonn); Prof. Dr. Iris Oppel and her coworker Christian Göb (Aachen); and Dr. Jens Antony, Dr. Gerit Brandenburg and Meriam Seddiqzai from the Grimme group.

I thank the whole Grimme group, including *alumni*, for the excellent working atmosphere with many fruitful scientific discussions and the great non-working atmosphere with lots of bad jokes and useless conversations.
Especially, I am grateful to Dr. Gerit Brandenburg, Christoph Bannwarth, Christoph Bauer, and Dr. Jens Antony for proof-reading parts of this thesis.
I thank Dr. Marc Steinmetz, Meriam Seddiqzai, and Mirko Bauer for the great time we had while sharing the office and Dr. Andreas Hansen for his advice whenever I needed it.

My dearest thanks go to Adrian Gabriel, for his ongoing support, encouragement and patience.

Finally, I would like to thank my parents. Without their constant support, my studies would not have been possible.