Title

Developing the Theory of Dispersion Interactions for Biological Applications

Author

Nguyen, Brian Dinh-Binh

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Developing the Theory of Dispersion Interactions for Biological Applications

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Chemistry with concentration in Chemical and Materials Physics

by

Brian D. Nguyen

Dissertation Committee:
Professor Filipp Furche, Chair
Professor Kieron Burke
Professor Vladimir A. Mandelshtam

2022

DEDICATION

I dedicate this thesis work to my parents, Mac and Mai Anh Nguyen, my sister, Jessica Nguyen, and my brother, Michael Nguyen. Their unconditional love and support have been valuable during the challenges of graduate school and life.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

ACKNOWLEDGMENTS

First, I am grateful for my advisor Dr. Filipp Furche and his unconditional support from day one. I first joined the lab as an undergraduate in 2013 fall quarter. The undergraduate research piqued my interests in quantum mechanics. It inspired my pursuits into computational and theoretical chemistry. As a graduate student, Dr. Furche and his research group accepted my application into the group. Their patience, guidance, and mentorship have been invaluable toward my development as a researcher, mentor, and teacher.

Being in the Furche Research Group, there have been countless memories with postdoctoral and graduate researchers. My first graduate mentor Dr. Brandon Krull had paved the path as a computational chemist when I helped in the development of new luciferin derivatives for bioluminescence. Dr. Asbjörn M. Burow and Matthew M. Agee guided my studies in noncovalent interactions and benchmarking the random phase approximation. In addition, I am grateful for the computational and theoretical discussions on the random phase approximation with Dr. Vamsee Voora, Dr. Guo P. Chen, and Dr. Sree Ganesh Balasubramani. Their insights have been invaluable and inspirational toward my own scientific inquiries. Next, I am thankful for working alongside Dr. Eva von Domaros to understand the electronic structure of MoS_4^{3-}-bridged dilanthanide complexes. Lastly, Dr. Shane Parker, Dr. Mikko Muuronen, Dr. Nambi Mohanam, Jason M. Yu, Samuel Bekoe, Gabriel Phun, Jeffrey Tsai, Ahmad Rajabi, and Sourav Majumdar have all contributed toward developing my scientific curiosities, improving my public speaking skills, and striving for more.

Throughout my graduate career, I have been fortunate enough to mentor high school students and undergraduates. I am thankful to Dr. Furche for providing me the opportunity to lead the Furche High School Outreach program since 2016. The program has resulted in a community of 21 high school students and 12 research mentors. I have personally mentored 4 high school students Matthew P. Tang, Jenny Nguyen, Thanh Huynh, and Natalie Tran. Moreover, I have mentored 4 first generation undergraduates Emily Barragan, Poorvi Rao, Devin J. Hernandez, and Emmanuel V. Flores. Forever I am grateful to work alongside my mentees to expand human understanding of chemical phenomena.

Third, I acknowledge Dr. Thomas Poulos and Dr. Scott A. Hollingsworth from the Poulos Lab. They have been instrumental in growing my interests in protein research. I am grateful for Dr. Hollingsworth who has provided helpful advice and support for my graduate studies.

Finally, I humbly acknowledge the funding agencies and fellowships that supported my research career. The National Science Foundation (CHE-1464828, CHE-1800431, and CHE-2102568) has supported my research projects. The UCI Graduate Chancellor Fellowship in Chemistry provided financial support in my first year of graduate studies. The UCI Dissertation Fellowship in Chemistry gave me the financial independence to focus on writing my dissertation. The UCI School of Physical Sciences Faculty Endowed Fellowship has supported my attendance and oral presentation at the 2022 Spring ACS Conference in San Diego. Lastly, I acknowledge the travel grants in 2019 from the Chemistry Department and the Associated Graduate Students.

The following chapters of the dissertation are reprinted:

Chapter 2 is reprinted with permission from [Hollingsworth, S.A.*; Nguyen, B.D.*; Chreifi, G.; Arce, A.P.; Poulos, T.L. Insights into the Dynamics and Dissociation Mechanism of a Protein Redox Complex Using Molecular Dynamics. *J. Chem. Info. Model.* **2017**, 57(9), 2344–2350]. Copyright 2017 American Chemical Society.

Chapter 3 is reprinted with permission from [Nguyen, B.D.; Chen, G.P.; Agee, M.M.; Burow, A.M.; Tang, M.P.; Furche, F. Divergence of Many-Body Perturbation Theory for Noncovalent Interactions of Large. *J. Chem. Theory Comput.* **2020**, 16(4), 2258–2273]. Copyright 2020 American Chemical Society.

Chapter 4 is reprinted with permission from [Nguyen, B.D.*; Hernandez, D.J.*; Flores, E.; Furche, F. Dispersion Size-Consistency. *Electron. Struct.* **2021**, (4), 040902]. © IOP Publishing. Reproduced with permission. All rights reserved.

* Indicates authors contributed equally.

VITA

Brian D. Nguyen

EDUCATION

Doctor of Philosophy in Chemistry **2022**
University of California, Irvine *Irvine, CA*

Bachelor of Science in Chemistry **2015**
University of California, Irvine *Irvine, CA*

Bachelor of Science in Biology **2015**
University of California, Irvine *Irvine, CA*

RESEARCH EXPERIENCE

Graduate Research **2016–2022**
University of California, Irvine *Irvine, CA*

Undergraduate Research **2013–2016**
University of California, Irvine *Irvine, CA*

TEACHING EXPERIENCE

Honor General Chemistry Teaching Assistant **2022**
University of California, Irvine *Irvine, CA*

Computational Chemistry Lab Graduate Teaching Assistant **2020**
University of California, Irvine *Irvine, CA*

Quantum Chemistry Graduate Teaching Assistant **2020**
University of California, Irvine *Irvine, CA*

Computational Chemistry Graduate Teaching Assistant **2019**
University of California, Irvine *Irvine, CA*

General Chemistry Teaching Assistant **2016–2017**
University of California, Irvine *Irvine, CA*

MENTORING EXPERIENCE

Furche High School Outreach Program **2016–2022**
University of California, Irvine *Irvine, CA*

UCI Undergraduate Research Mentorship **2017–2022**
University of California, Irvine *Irvine, CA*

UCI Competitive Edge Peer Mentor **2017**
University of California, Irvine *Irvine, CA*

Orange County Regional Science Olympiad **2017–2020**
University of California, Irvine *Irvine, CA*

REFEREED JOURNAL PUBLICATIONS

1. **Nguyen, B.D.***; Hernandez, D.J.*; Flores, E.; Furche, F. Dispersion Size-Consistency. *Electron. Struct.* **2021**, (4), 040902.

2. Yu, J.M.*;**Nguyen, B.D.***; Tsai, J.; Furche, F. Selfconsistent Random Phase Approximation Methods. *J. Chem. Phys.* **2021**, *155*(4), 040902.

3. Darago, L.E.; Boshart, M.D.; **Nguyen, B.D.**; Perlt, E.; Ziller, J.W.; Lukens, W.W.; Furche, F.; Evans, W.J.; Long, J.R. Strong Ferromagnetic Exchange Coupling and Single-Molecule Magnetism in MoS_4^{3-}-Bridged Dilanthanide Complexes. *J. Am. Chem. Soc.* **2021**, *143*(22), 8465–8475.

4. Balasubramani, S.G.; [and 36 others, including **Nguyen, B.D.**] TURBOMOLE: Modular Program Suite for Ab Initio Quantum-Chemical and Condensed-Matter Simulations. *J. Chem. Phys.* **2020**, *152*(18), 184017.

5. **Nguyen, B.D.**; Chen, G.P.; Agee, M.M.; Burow, A.M.; Tang, M.P.; Furche, F. Divergence of Many-Body Perturbation Theory for Noncovalent Interactions of Large Molecules. *J. Chem. Theory Comput.* **2020**, *16*(4), 2258–2273.

6. Hollingsworth, S.A.*; **Nguyen, B.D.***; Chreifi, G.; Arce, A.P.; Poulos, T.L. Insights into the Dynamics and Dissociation Mechanism of a Protein Redox Complex Using Molecular Dynamics. *J. Chem. Info. Model.* **2017**, *57*(9), 2344–2350.

7. Steinhardt, R.C.; Rathbun, C.M.; Krull, B.T.; Yu, J.; Yang, Y.; **Nguyen, B.D.**; Kwon, J.; McCutcheon, D.C.; Jones, K.A.; Furche, F.; Prescher, J.A. Brominated Luciferins Are Versatile Bioluminescent Probes. *ChemBioChem* **2016**, *18*(1), 96–100.

8. Hollingsworth, S.A.*; Batabyal, D.*; **Nguyen, B.D.**; Poulos, T.L. Conformational Selectivity in Cytochrome P450 Redox Partner Interactions. *Proc. Natl. Acad. Sci.* **2016**, *113*(31), 8723–8728.

9. Furche, F.; Krull, B.T.; **Nguyen, B.D.**; Kwon, J. Accelerating Molecular Property Calculations with Nonorthonormal Krylov Space Methods. *J. Chem. Phys.* **2016**, *144*(17), 174105.

* Indicates authors contributed equally.

CONFERENCE PRESENTATIONS

Dispersion Size-Consistency **Mar 2022**
2022 Spring ACS National Meeting

Divergence of Many-Body Perturbation Theory **April 2020**
2020 Fall ACS National Meeting - Oral

Size dependence of noncovalent interactions within RPA **May 2019**
2019 Southern California Theoretical Chemistry Symposium - Poster

Accuracy of RPA to Predict Large Weakly Interacting Systems **June 2018**
2018 Conference on Excited State Processes - Poster

INVITED TALK

Is dispersion considered weak? **Apr 2021**
University of Vienna, Austria

AWARDS AND HONORS

UCI Grad Slam Semifinalist **Feb 2022**
University of California, Irvine

UCI School of Physical Sciences Faculty Endowed Fellowship **Jun 2021**
University of California, Irvine

UCI Dissertation Fellowship in Chemistry **Jun 2021**
University of California, Irvine

UCI Graduate Chancellor Fellowship in Chemistry **Sept 2016**
University of California, Irvine

UCI Chancellor's Undergraduate Award of Distinction **Jun 2015**
University of California, Irvine

Phi Beta Kappa **May 2015**
University of California, Irvine

Phi Lambda Upsilon **May 2015**
University of California, Irvine

Hypercube Scholar Award **Jun 2014**
Hypercube Inc.

OC American Chemical Society Undergraduate Award **Apr 2014**
OC American Chemical Society Local Chapter

ABSTRACT OF THE DISSERTATION

Developing the Theory of Dispersion Interactions for Biological Applications

By

Brian D. Nguyen

Doctor of Philosophy in Chemistry with concentration in Chemical and Materials Physics

University of California, Irvine, 2022

Professor Filipp Furche, Chair

Noncovalent interactions (NIs) are present in the properties and functions of all matter, from solid state to soft materials. These interactions can range from a few kcal/mol to several hundreds or even thousands of kcal/mol. Ubiquitous to all molecules is the presence of dispersion interactions, which is a force allowing geckos to stick onto walls. However, dispersion remains challenging to intuitively understand and accurately predict. In this thesis, I investigate the dissociation of *Leishmania major* peroxidase using atomistic molecular dynamics (MD), revisit the current understanding of many-body perturbation theory for NIs, introduce an exact constraint known as the dispersion size-consistency along with its importance for electronic structure methods, and apply that knowledge to understand the novel face-on halogen-π interaction between the lissoclimide family and eukaryotic 80S ribosome.

In chapter 2, I investigate the dissociation of *Leishmania major* peroxidase from MD and contribute to the understanding of the interactions between heme proteins and its electron-transferring redox partner. Peroxidases function to catalyze the reduction of peroxide into water and molecular oxygen. This allows parasites such as *Leishmania major* to evade reactive oxygen species (ROS) from the host's defenses. *Leishmania major* peroxidase (LmP) and cytochrome c (Cytc) form a complex that mediates interprotein electron transfer (ET) and reduces ROS. Such a complex must navigate between fast turnover and tight binding.

A previous Brownian dynamics study showed that LmCytc associates with the LmP by first interacting with helix A of LmP and then moving toward the ET site. Critical to this association is the intermolecular Arg-Asp ion pair at the center of the interface. In anticipation, the dissociation process is the reverse by breaking the Asp-Arg ion pair and follow by the movement towards LmP helix A. To test this, I performed multiple MD simulations along with *in silico* mutation of the LmP Asp211 to Ala211 and observed the dissociation process consistent with the experiment.

In chapters 3 and 4, I revisit the many-body perturbation (MBPT) for NIs due to alarmingly large binding energy errors obtained from MBPT. Currently, NIs have traditionally been understood as "weak" relative to covalent bonds on the order of several magnitudes. Due to this weakness, MBPT has been expected to accurately model NIs. This was observed with small complexes from the S22 testset. I reassess the performance of the second-order Møller-Plesset MBPT (MP2) and compare the results to spin-scaled MP2, dispersion-corrected semilocal density functional approximations (DFAs), and post-Kohn–Sham random phase approximation (RPA). These methods is benchmarked against the S66, L7, and S30L testsets for predicted binding energies. All binding energies are extrapolated to the complete basis set limit, corrected for basis set superposition errors, and compared to the reference results of the domain-based local pair-natural orbital coupled-cluster (DLPNO-CCSD(T)) or better quality. The results reveal that MP2 significantly overestimates the binding energies. In some cases, the MP2 relative errors are over 100%. Spin-scaled MP2, while an improvement, still inherits the limitations of MP2. RPA and dispersion-corrected DFAs have similar performance ranging between 5% to 10% errors. A regression analysis shows that MP2 binding energy errors grew with the system size by at rate of $\sim 0.1\%$ per valence electron, whereas RPA and dispersion-corrected DFA errors remain constant.

To understand the errors, I develop an asymptotic adiabatic connection symmetry-adapted perturbation theory (AC-SAPT). The theory considers a supersystem in terms of the non-

overlapping monomers at full coupling whose ground-state density is constrained to the ground-state of the supersystem. Using the fluctuation-dissipation theorem, a nonperturbative "screened second-order" expression for the dispersion energy in terms of the monomer basis is obtained. AC-SAPT expansion of the interaction energy reveals that the source of binding energy errors come from missing or an incomplete "electrodynamic" screening of the Coulomb interaction due to induced particle-hole pairs between electrons in different monomers. MP2 and higher-ordered perturbation theories lack this property leading to a divergent series within the AC-SAPT framework, whereas RPA converges.

Furthermore, extension of the AC-SAPT framework to the thermodynamic limit establishes the dispersion size-consistency, which states that the total dispersion energy of an N-monomer system is independent of any partitioning into subsystems. MBPT is found to violate this condition and RPA does not. This is due to the additive separability of the dispersion energy results from multiplicative separability of the generalized screening factor defined as the inverse generalized dielectric function. Based on the computational and theoretical results, MBPT may not be qualitatively and quantitatively adequate for prediction of NIs. Nonperturbative methods such as RPA or coupled cluster methods should be used.

In chapter 5, I collaborated with the Vanderwall Lab to understand the novel face-on halogen-π interaction between the lissoclimide family and eukaryotic 80S ribosome. The model is based on the structure-activity relationship (SAR) of the lissoclimide family inhibition of protein synthesis obtained from the X-ray co-crystal structure of the 80S ribosome and chlorolissoclimide. I took advantage of the SAR and instead of modeling the $\sim 400,000$ atoms from the crystal structure, the region where the chlorolissoclimide binds is studied instead. The system includes the lissoclimide family, two guanine nucleobases, and the phospahte backbone. This becomes a manageable 122 atoms system for electronic structure methods. The remaining protein environment is approximated using the implicit conductor-like solvation model. I verify the model by correlating the experimental half-maximum

inhibitor concentration (IC_{50}) and RPA predicted binding energies between the lissoclimide derivatives and the eukaryotic 80S ribosome. The relationship reveals a negative correlation consistent with the anticipation that stronger binding leads to a more potent inhibitor. Based on these results, I proposed additional inhibitors and contributed to an expanded SAR knowledge of the lissoclimide family inhibition.

Chapter 1

Introduction: Noncovalent Interactions

Noncovalent interactions (NIs) play a role determining the chemical and physical properties of materials,[1–3] biological systems,[4,5] and large supramolecular complexes.[6–8] Present in all these systems are dispersion interactions, which have been considered the weakest kind of NIs. Yet, the dispersion interations allow geckos to stick onto ceilings and walk along walls.[9] In a sense, geckos are literally turning on and off dispersion interactions. This brings into question our current understanding of dispersion interactions.

Johannes Diderik van der Waals was the first to observe these weak attractive forces when he was establishing the relationship between the pressure, volume, and temperature of gases and liquids. Shortly after, Fritz London first recognized the importance of quantum mechanics to capture the correct physics and suggested that dispersion interactions arise from electronic fluctuation, or random electronic motions, within the molecules.[10] These random motions can cause "instantneous dipoles" that are attractive in nature and follow the $1/R^6$ power law. The discovery has been widely accepted in academia and taught in chemistry courses.

However, in practice, NIs remain a challenge to intuitively understand and accurately predict. Controlling and manipulating these interactions have been based on an understanding of the electrostatics and dispersion in small molecules. For instance, the current electronic structure method of choice is the many-body perturbation theory (MBPT) due to the relative weakness of NIs compared to covalent bonds. Binding energies of small molecules on the size of tens-of-atoms were accurately obtained from the second-order Møller-Plesset MBPT (MP2).[11–13] Furthermore, numerical results of the high-order MP series were observed to converge for the He, H_2, LiH, H_2O, and HF dimers.[14] Combined with the MP theory being size-extensive and size-consistent,[15] the MBPT appeared to be appropriate for modeling NIs. Size-consistency has been suggested by Pople[16] and Barlett[17] to have a great importance for predicting chemical properties.

This perception has made it safe to assume that the MBPT should work for modeling NIs in more complex systems e.g supramolecular complexes and proteins. Unfortunately, in the early 2000s, MP2 showed inconsistencies and yielded unusually large binding energy errors. One of the first examples is the coronene dimer in which MP2 significantly overbinds by 100% compared to the the quadratic configuration interaction reference data.[18,19]

In the thesis, I aim to investigate the source of the MBPT's inconsistencies for NIs, determine appropriate electronic structure methods for modeling NIs, and apply the knowledge towards biological applications. Prior to investigating MP2 binding energy errors, chapter 2 focuses on the dissociation mechanism of *Leishmania major* peroxidase using the molecular dynamic simulation. The mechanism is facilitated with mainly ionic interactions that corroborates with the association mechanism presented by previous Brownian dynamic simulations.[5,20] While molecular dynamics are appropriate for capturing ionic interactions, dispersion interactions remain challenging for force fields. A quantum mechanic approach should be used. However, not all electronic structure methods are appropriate to model NIs as previously mentioned.

Hence, in chapter 3, I revist the MBPT for NIs and determine the source of these errors by building the adiabatic connection symmetry adapted perturbation theory (AC-SAPT) that establishes the importance of the electrodynamic screening. With the AC-SAPT framework, in chapter 4, I explicitly establish the exact constraint coined as dispersion size-consistency, which has been implied by prior works.[14,21–24] I show that this condition is important for modeling NIs using electronic structure methods. Lastly, in chapter 5, I apply the knowledge toward expanding the structure activity relationship of the lissoclimide family inhibition of protein synthesis and take advantage of the novel face-on halogen-π interaction.

Chapter 2

Insights into the Dynamics and Dissociation Mechanism of a Protein Redox Complex Using Molecular Dynamics

Chapter 2 is reprinted with permission from [Hollingsworth, S.A.*; Nguyen, B.D.*; Chreifi, G.; Arce, A.P.; Poulos, T.L. Insights into the Dynamics and Dissociation Mechanism of a Protein Redox Complex Using Molecular Dynamics. *J. Chem. Info. Model.* **2017**, 57(9), 2344–2350]. (* Indicates authors contributed equally.)

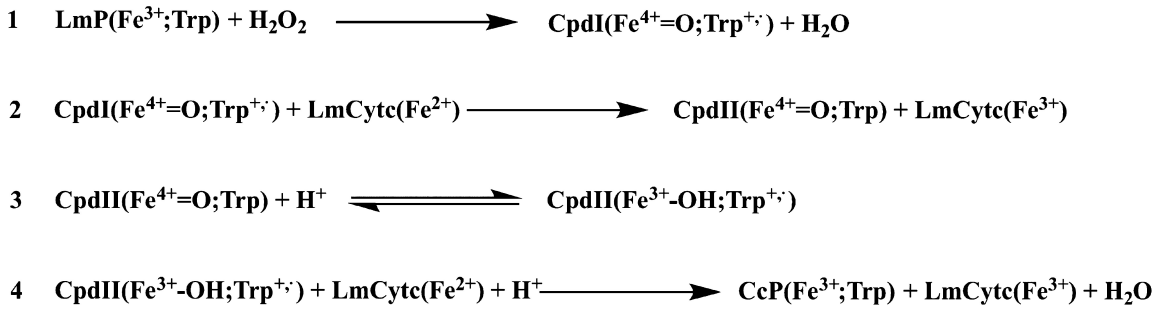
Contribution Statement: The following are my contributions to the project. I modeled and performed molecular dynamics simulations for the *Leishmania major* peroxidase using both CHARMM and AMBER packages. I created plots using Grace that analyzed the simulations by looking at bond distances and root mean square deviations. I contributed to

the writing and prepared the submission into the *J. Chem. Info. Model.*

2.1 Introduction

Interprotein electron transfer requires formation of protein-protein complexes.[25,26] Such complexes often are quite specific, since the electron transfer (ET) rates between redox centers decreases exponentially with distance, making it necessary to minimize the donor–acceptor distance and/or provide an appropriate ET path. Nature, however, must balance the requirement of specific binding with rapid turnover, so protein redox complexes also are designed to rapidly dissociate. Therefore, while the association rate often is quite fast, the dissociation rate also is fast, which often means that neither association or dissociation is rate-limiting. The resulting moderate stability of protein redox complexes is one reason why there are very few such complex crystal structures. One of the exceptions is the well-studied yeast Cytochrome c peroxidase (CCP)–cytochrome c (Cytc) complex that has long served as a paradigm for interprotein ET studies.[27,28]

Recently, a second CCP–Cytc complex has been characterized. The human pathogen *Leishmania major* has a peroxidase (LmP) that is mechanistically and structurally similar to yeast CCP.[29–32] Mechanistically, these systems have been shown to be almost identical (Scheme 2.1).[30–32]



Scheme 2.1: Interprotein electron transfer

In step 1, the peroxidase is first oxidized by hydrogen peroxide to produce the ferryl compound **I** and a Trp radical (Fe^{IV}=O; $Trp^{+\cdot}$).[33,34] Through an electron transfer event with the related Cytochrome c (step 2), the $Trp^{+\cdot}$ radical is reduced to give compound **II** (Fe^{IV}=O; Trp). An intramolecular proton-coupled electron transfer (PCET, step 3) from the Trp to Fe^{IV}=O gives Fe^{III}–OH; $Trp^{+\cdot}$,[35,36] before a second electron transfer event (step 4) with Cytochrome c reduces Trp^{+} to return the peroxidase to its resting state. In addition, the structure of the LmP–LmCytc complex is strikingly similar to the experimentally determined yeast complex.[31]

Despite their mechanistic and structural similarities, however, the LmP–LmCytc redox pair is kinetically unlike the CCP–Cytc system, which suggests that the association and dissociation of these redox pairs may be different. First, while the LmP system obeys simple Michaelis–Menten kinetics,[31,32] CCP does not.[37] Second, the crystal structure of each complex shows that the LmP–LmCytc complex is stabilized by specific intermolecular ion pairs (Figure 2.1),[32] while the CCP/Cytc interface has no intermolecular ion pairs but instead appears to be stabilized by nonpolar interactions at the interface.[27] Further emphasizing the importance of electrostatic interactions in the LmP system, we recently documented a secondary binding site for LmCytc on LmP.[5] This noncatalytic site is composed of four nearly consecutive negatively charged residues on helix A, adjacent to but separate from the active site. A combination of computational and experimental results has shown that helix A influences complex formation and dissociation for the LmP system, but not for CCP. (14) In the Brownian dynamics simulation, LmCytc initially docks to LmP helix A and then migrates toward the ET-active position. For the proper ET complex to form, the critically important $R24_{LmCytc}$–$D211_{LmP}$ ion pair must form at the center of the complex. The importance of this ion pair is underscored by the D211N mutant, which exhibits $\sim$ 8% wild-type activity.[5] This decrease is due to the rate-limiting step switching from the Trp-to-Fe(IV)=O intramolecular ET to the rate of association of the LmP–LmCytc complex itself.[30] For the $R24_{LmCytc}$–$D211_{LmP}$ interaction to form, the intramolecular interaction

between $R24_{LmCytc}$ and $E101_{LmCytc}$ must be broken, thus freeing $R24_{LmCytc}$ to adopt a new rotameric conformation in order to interact with $D211_{LmP}$.[32] These computational results guided the mutagenesis studies, where removing three negative charges on helix A was found to lower k_{cat} by $\sim$ 3-fold and the rate of association of the two proteins by $\sim$ 6-fold.[5]

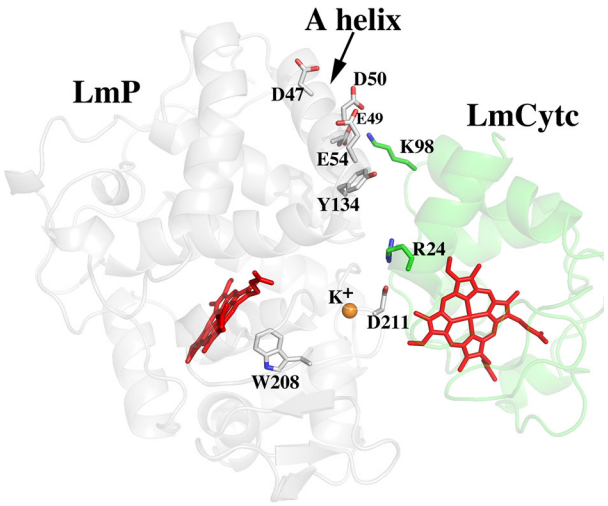


Figure 2.1: Schematic of the LmP–LmCytc complex highlighting key interactions. [Legend: LmP is depicted in faint gray, LmCytc is depicted in faint green, and the respective heme groups are shown as red sticks.] The catalytically important intermolecular ion pair between D211 of LmP and R24 of LmCytc defines the ET-active binding site for LmCytc, while the negatively charged residue of D47, E49, D50, and E54 of LmP helix A constitutes the secondary binding site. LmP Y134 is positioned directly between the ET-active and secondary binding sites.

In order to study the dynamics of the LmP–LmCytc redox pair complex and provide a deeper understanding of a possible dissociation mechanism, we have performed molecular dynamics (MD) simulations of the experimentally determined complex structure. These simulations have revealed a clear visualization of the dynamics of the important intermolecular and intramolecular ionic interactions and a point of comparison to other well-studied heme protein redox partner systems.

2.2 Methods: Molecular Dynamics Simulations

In order to the study the dynamics of the LmP–LmCytc complex, we conducted an initial set of three atomistic molecular dynamics (MD) simulations of the experimental co-crystal structure of LmCytc in complex with LmP (PDB ID: 4GED).[32] The preparation of the MD simulation is similar to our previous study of the LmP–LmCytc system[5] and described briefly here. Hydrogen atoms were added to the crystal structure, using the psfgen plugin of VMD 1.9.1.[38] Patches were employed to connect the ferric high-spin heme with the coordinating His residues in both LmP and LmCytc, while an extra bond parameter was added to describe the Met-heme coordination in LmCytc, as well as the coordination of the ions present in the co-crystal structure. The surrounding orthogonal solvent box was constructed with a 20 Å cushion in all directions around the proteins. The resulting complex system contained $73,815$ atoms.

The initial three MD simulations were performed using NAMD,[39] version 2.10, on the green-planet cluster at UC Irvine and the XSEDE Stampede computing cluster. The CHARMM22[40] force field was employed for the proteins and co-factor. The TIP3P model[41] was used to model the solvent. Each system underwent 1000 steps of conjugate gradient energy minimization at a constant pressure of 1 atm and 300 K, using a Nosé–Hoover–Langevin piston for pressure control and a Langevin dynamics for temperature control, respectively.[42,43] A time step of 1 fs was employed for the first 10 ns of each simulation before being increased to 2 fs for the remainder of the trajectory, while a multiple time step algorithm was employed to integrate the equations for motion, as described previously.[44] The electrostatic interactions were treated using a smooth particle mesh Ewald algorithm[45] and the real space part of the Ewald sum and the Lennard-Jones interactions were switched off between 10Å and 12Å, while the all of the bonds to H atoms were constrained usingthe SHAKE algorithm.[46] Analyses was performed using VMD[38] and PyMOL (http://www.pymol.org), as well as locally developed analysis tools.

Based on the findings from the three NAMD simulations, we performed two additional simulations, using AMBER. The NAMD runs showed that once the $Arg24_{LmCytc}$–$Asp211_{LmP}$ ion pair at the LmP/Cytc interface breaks and Cytc moves toward helix A of LmP. We therefore generated the in silico D211A LmP mutant, thus eliminating this intermolecular ion pair, to see if LmCytc will undergo the same dissociation process as that observed in the NAMD runs. As a control, we also included a simulation with the wild-type complex. The systems were prepared similarly to the NAMD runs, using a solvent box of the same size. Ferric high-spin heme parameters were taken from Collins and Loew,[47] and, for the protein, the AMBER ff99SB force field was used. Hydrogen mass repartitioning (HMR) through parmed was employed to redistribute the mass of the hydrogens, allowing for the use of a 4 fs time step for all AMBER simulations. As part of another study, we compared simulations of Cytochrome P450 with and without HMR and found little difference in the dynamics. All AMBER runs were run on the GPU clusters at the San Diego Supercomputer Center.

2.3 Results and Discussion

2.3.1 LmP–LmCytc Complex Dynamics

In order to study the stability of the LmP–LmCytc complex, we performed three 650 ns CHARMM MD simulations, using the experimentally determined co-crystal structure[32] as a starting point. We also performed an AMBER simulation of the wild-type complex plus a second simulation, where $D211_{LmP}$ has been converted to an Ala. This mutant mimics rupturing of the $R24_{LmCytc}$–$D211_{LmP}$ ion, which is a critical intermolecular ion pair and the center of the complex (see Figure 2.1). Root-mean-square deviation (RMSD) analysis (Figure S2.1 in the Supporting Information) showed that both LmP and LmCytc undergo large backbone deviations, because of variations in surface loops. The deviations are much

less when confined to regular elements of secondary structure. This, together with visual inspection of the trajectories, shows that the individual structures are quite stable. However, visual inspection of the trajectories clearly showed that LmCytc moves away from the ET-active site toward helix A (see Figure 2.2). From the crystal structure, the distance between the center of mass of LmCytc and helix A is $\sim$ 26Å and substantially shortens during the simulations (Table 2.1) as LmCytc migrates toward LmP helix A. In CHARMM replicate 1, LmCytc first moves toward helix A, dissociates, and then reassociates with helix A. The main reason for this larger motion in replicate 1 is that the $R24_{LmCytc}$–$D211_{LmP}$ ion pair dynamically breaks and reforms throughout the simulation but remains broken after $\sim$ 550 ns, which frees LmCytc to move further toward LmP helix A. The models in Figure 2.2 show that, toward the end of the simulations, LmCytc has moved closer to helix A in all five simulations. In the D211A mutant AMBER simulation, LmCytc moves to helix A much more quickly and remains there for the remainder of the simulation. The AMBER simulation of the wild-type complex behaves similar to CHARMM replicate 1. The $R24_{LmCytc}$–$D211_{LmP}$ ion pair breaks at $\sim$ 2.2μs and remains broken for the remainder of the simulation (see Figure 2.3, as well as Figure S2.2 in the Supporting Information). As in CHARMM replicate 1, breaking of the ion pair frees LmCytc to move closer toward helix A. Taken together, these results show that LmCytc favors moving toward helix A but is restrained by the $R24_{LmCytc}$–$D211_{LmP}$ ion pair, which prevents the full motion of LmCytc to transition from the ET-active conformation to the secondary binding site of LmP helix A. Once this ion pair breaks, however, as observed in both CHARMM replicate 1 and the AMBER simulations, LmCytc is free to fully transition to the secondary binding site previously observed in Brownian dynamics simulations (helix A of LmP) and ultimately dissociates.

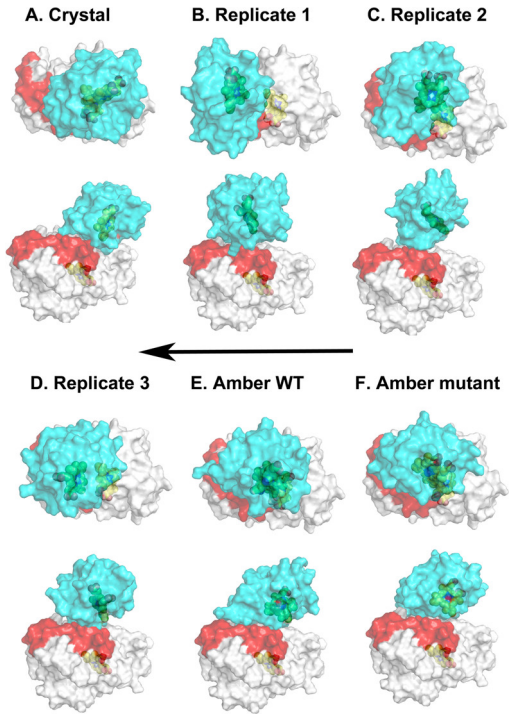


Figure 2.2: Schematic of the LmP–LmCytc complex toward the end of simulation. Snapshots near the end of each respective MD simulation are displayed, where LmCytc is cyan, LmP is white (except helix A, which is shown in red). Two views are shown: one viewed down on the LmCytc docking site and a side view. Images shown represent (A) crystal structure, (B) CHRAMM replicate 1, (C) CHARMM replicate 2, (D) CHARMM replicate 3, (E) AMBER wild type, and (F) AMBER D211A mutant. Relative to the crystal structure, LmCytc moves toward the A helix in all simulations. However, in replicate 1 (panel (B)) and the Amber simulations (panels (E) and (F)), where the Asp–Arg intermolecular ion pair breaks, LmCytc now is free to form closer interactions with helix A. The arrow indicates the direction of motion of LmCytc.

Table 2.1: LmCytc–Helix A Minimum Center of Mass Distance

	Distance (Å)
Crystal	26
CHARMM1	20.2
CHARMM2	22.4
CHARMM3	21.0
AMBER WT	20.8
AMBER Mutant	19.7

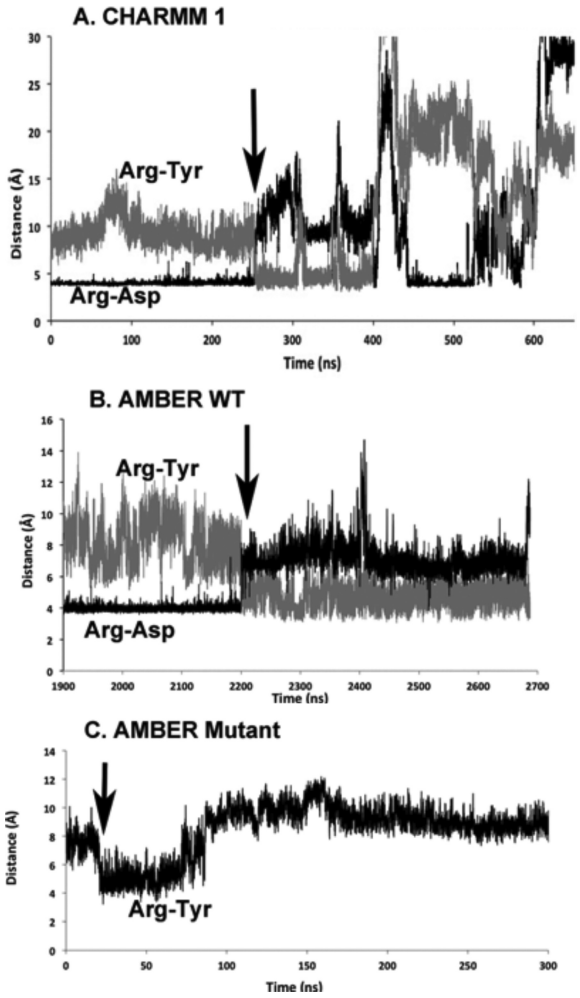


Figure 2.3: Plots showing the $R24_{LmCytc}$–$D211_{LmP}$ and $R24_{LmCytc}$–$Y134_{LmP}$ distances as a function of simulation time: (A) CHARMM replicate 1, (B) AMBER wild type, and (C) AMBER D211A mutant. The arrows indicate where the he $R24_{LmCytc}$–$D211_{LmP}$ breaks and the $R24_{LmCytc}$–$Y134_{LmP}$ interaction forms. In the AMBER mutant (panel (C)), the $R24_{LmCytc}$–$Y134_{LmP}$ interaction forms quickly and remains stable for $\sim$ 50 ns before this interaction is lost as LmCytc moves closer to helix A.

2.3.2 LmP–LmCytc Dissociation

We next focus on CHARMM replicate 1, since the $R24_{LmCytc}$–$D211_{LmP}$ ion pair breaks relatively early and thus provides the most detailed picture of the dissociation process in the wild-type complex. Once the $R24_{LmCytc}$–$D211_{LmP}$ ion breaks and LmCytc begins to slide further toward helix A, $R24_{LmCytc}$ forms a new interaction with $Y134_{LmP}$. As shown in Figure 2.3 for both the CHARMM replicate 1 and AMBER WT simulations, the $R24_{LmCytc}$–$Y134_{LmP}$ pair forms as soon as the $R24_{LmCytc}$–$D211_{LmP}$ ion breaks. In the D211A AMBER simulation, $R24_{LmCytc}$ also forms an interaction with $Y134_{LmP}$. Thus, in all three simulations where the intermolecular ion pair breaks or is not present, because of in silico mutagenesis, LmCytc begins its "crawl" toward helix A by initially interacting with $Y134_{LmP}$. $Y134_{LmP}$ is located between the $D211_{LmP}$ at the ET-active site and helix A (Figure 2.1) and thus provides both a new hydrogen-bonding partner to $R24_{LmCytc}$ as LmCytc moves toward helix A, as well as preventing reformation of the ET-active ion pair with $D211_{LmP}$.

The energetic incentive for LmCytc moving toward helix A during this process are two surface-exposed Lys residues of LmCytc (K16 and K19) that approach E49, D50, and E54 in helix A of LmP and aid to pull LmCytc away from the ET-active site. Interestingly, following breakage of this new transient interaction in CHARMM replicate 1, $R24_{LmCytc}$ again reorients and reforms an intramolecular ion pair with $E101_{LmCytc}$ that is observed in the LmCytc crystal structure in the absence of LmP (see Figure 2.4 and Figure 2.5, as well as Figure S2.3 and Figure S2.4 in the Supporting Information). In the wild-type AMBER simulation, the $R24_{LmCytc}$–$E101_{LmCytc}$ distance also decreases (Figure S2.3) but is not close enough to reform the ion pair. Once this change occurs, LmCytc effectively dissociates and moves away from helix A to a point where the LmCytc center of mass reaches 28 Å from its starting position in the complex crystal structure. Continuing, LmCytc then returns to LmP, where $R24_{LmCytc}$ moves back and forth between $D211_{LmP}$ and $Y134_{LmP}$.

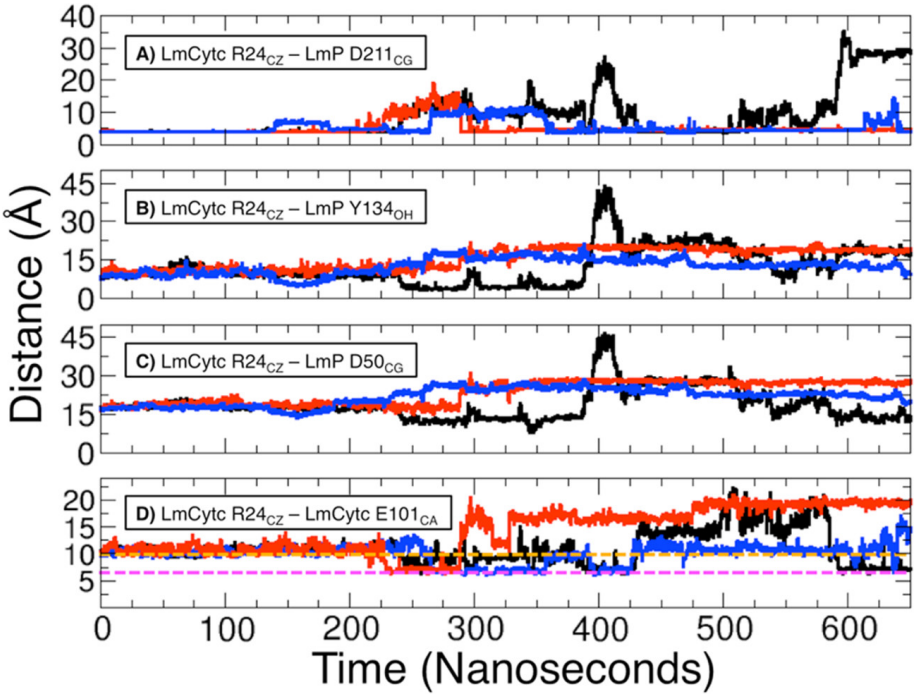


Figure 2.4: Distance evolution of the LmP–LmCytc complex interactions. The distance between interacting residues that were found to be important during the dissociation of the LmP–LmCytc complex over the course of each 650 ns MD simulation CHARMM replicates are highlighted in panels (A)–(D), where replicate 1 is shown in black, replicate 2 is shown in red, and replicate 3 is shown in blue. The highlighted interactions are shown as follows: (A) the electron transfer (ET)-active interprotein ion pair of LmCytc R24(CZ) and LmP D211(CG), (B) the transient interaction between LmCytc R24(NE) and LmP Y134(OH), (C) the distances between LmCytc R24(CZ) and D50 of LmP helix A, and (D) the intramolecular ion pair between LmCytc R24(CZ) and LmCytc E101(CA). In panel (D), the experimentally observed distance for the intramolecular ion pair in the individual structure (PDB ID: 4DY9) is shown in pink and the equivalent distance in the co-crystal structure (PDB ID: 4GED) is shown in orange.

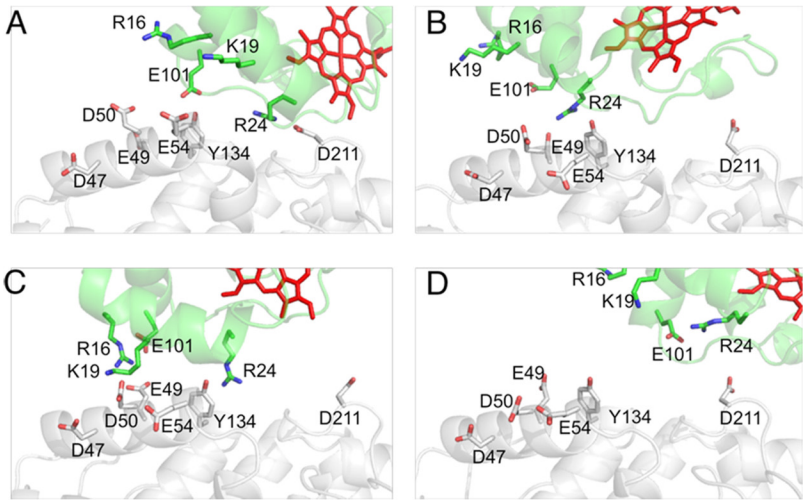


Figure 2.5: Snapshots of the dissociation process of the LmP–LmCytc complex. Molecular snapshots of the dissociation of the LmP–LmCytc complex, as observed through the unbiased MD simulation replicate 1 taken at (A) $\sim$ 0 ns, (B) $\sim$ 290 ns, (C) $\sim$ 340 ns, and (D) $\sim$ 400 ns.

2.3.3 Kinetics of the LmP Y134F Mutant

Since simulations across not only wild-type and mutant complexes but also CHARMM and AMBER force fields predict that $Y134_{LmP}$ plays a role in the association/dissociation reactions, we generated the experimental $Y134F_{LmP}$ mutant and determined its kinetic parameters. Figure 2.6 shows a comparison of our previously determined wild-type kinetics[5] with that of the $Y134F_{LmP}$ mutant. The $Y134F_{LmP}$ mutant exhibits a simple hyperbolic behavior, with a 3.6-fold increase in Km, when compared to the wild type (Figure 2.6), suggesting that the $Y134F_{LmP}$ mutation decreases the affinity of LmCytc for the LmP mutant. Wild-type rates are restored with saturating amounts of LmCytc, as seen by the measured k_{cat} value being almost identical to that of the wild type. This indicates that (i) the rate-limiting step of the reaction at steady state has not changed, but remains the intramolecular proton-coupled ET PCET from the Trp to Fe^{IV}, as previously shown for wild-type LmP,[30] but (ii) the binding of LmCytc to LmP has been weakened. Together, these results suggest that the association rate constant of the LmPY134F–LmCytc complex, although impaired, remains greater than the intramolecular PCET rate constant of $\sim$ 400 s^{-1}. This is consistent with

the prediction of $Y134_{LmP}$ playing an important, but not vital, role in complex turnover.

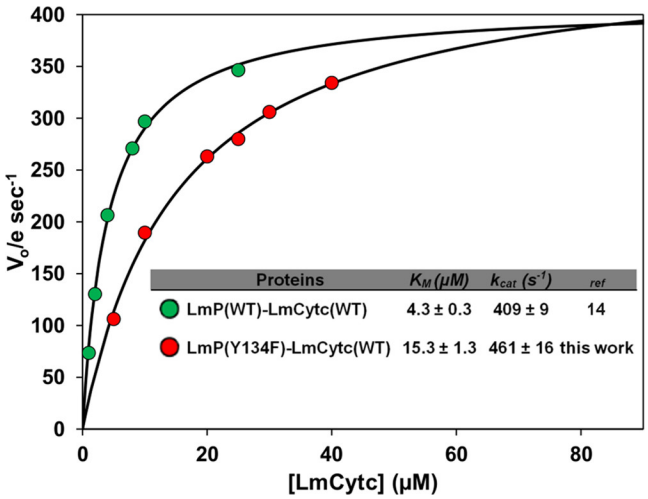


Figure 2.6: Experimental steady-state kinetic analysis of $Y134F_{LmP}$. A plot of V_0/e vs LmCytc concentration for the $Y134F_{LmP}$–LmCytc complex (shown in red) is superimposed onto the data from the wild-type LmP–LmCytc complex (shown in green). (Data taken from ref 5.)

2.3.4 Predicted Electron Transfer Rates in the LmP–LmCytc Complex

These new computational and experimental results, coupled with our previous Brownian dynamics studies,[5] indicate that helix A provides a secondary nonspecific electrostatic surface to which LmCytc can rapidly bind. However, this also raises the question of whether or not LmCytc delivers electrons while hovering near helix A, or if LmCytc must move to the position observed in the crystal structure in order to transfer an electron. Since ET rates are quite sensitive to the distance between donor and acceptor, comparing ET distances over the course of a trajectory can provide insights into which complexes are active and inactive (see Figure 2.7). The various parameters that are required for this application of Marcus theory[48,49] were taken from ref 50: ΔG, which is the difference in redox potential between the Trp208LmP radical and hemeLmCytc (-0.5 eV), and a reorganization energy of 1.0 eV.[50] In the crystal structure, the ET distance is ~ 15Å, which gives a rate of $\sim 2 \times 10^5$ s^{-1}. The

observed rate of electron transfer from Cytc to the Trp radical, using laser flash photolysis in the CCP-Cytc complex, is $\sim 2 \times 10^6$ s^{-1}.[50] Given the close similarity between the CCP-Cytc and LmP–LmCytc complexes, we can expect the LmP–LmCytc complex to exhibit a similar rate. It is important to note that this rate cannot be compared to the steady-state rate of $\sim 460 s^{-1}$, which is the intramolecular ET from the Trp radical to Fe(IV)=O. The Cytc-to-Trp radical ET rate also cannot be compared to second-order rates measured by stopped flow kinetics, since stopped flow mixing experiments measure the rate of association and not intermolecular ET from the LmCytc heme to the LmP Trp radical. Therefore, any computed intermolecular ET rate well below $105 - 106 s^{-1}$ is not compatible with the observed kinetic behavior of LmP and, thus, does not represent an ET active complex. In CHARMM replicates 2 and 3 (red and blue, respectively), the ET rate decreases by 3 orders of magnitude below the experimental values of 2×10^6 s^{-1} as LmCytc moves toward helix A. In replicate 1, after LmCytc moves to helix A of LmP, the closest ET distance is ~ 20Å, which gives a rate of ~ 100 s^{-1}. This is well below both the flash photolysis rate[50] and even k_{cat},[30] which provides additional evidence that the helix-A-bound complex is likely inactive, as predicted by previous BD and experimental results.[5]

2.4 Conclusions

A longstanding problem in understanding biological ET reactions is the requirement for balancing specificity with a high rate of turnover. In order to maintain rapid kinetics, the formation of protein complexes must be relatively weak. On the other hand, rapid intermolecular ET requires bringing the donor and acceptor relatively close and, in some cases, provide the proper intervening medium for rapid ET.[26] Since the interface that must align properly for ET is small, compared to the total surface area available to each protein, the probability of forming the ET-active complex via random intermolecular collisions is small.

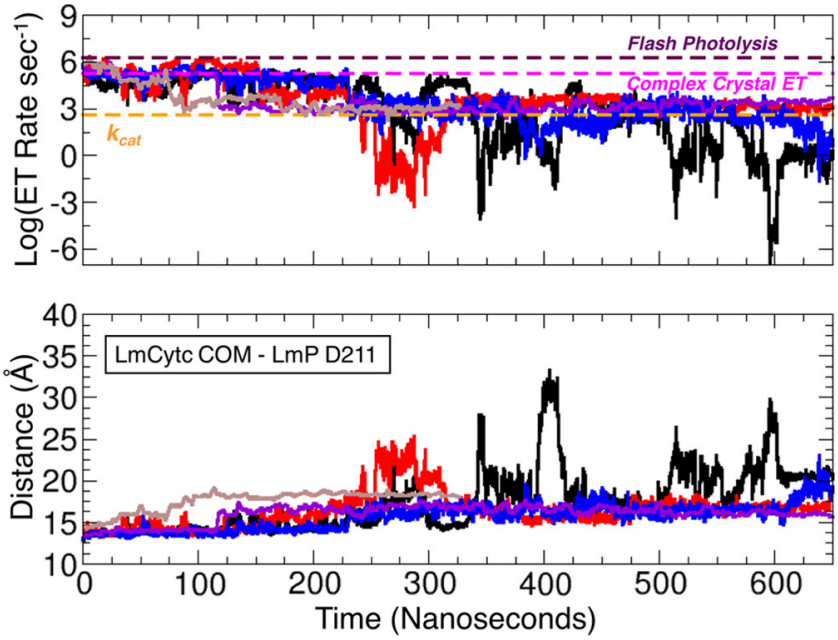


Figure 2.7: Predicted electron transfer distance and rate evolution for the LmP–LmCytc complex. (Top) The logarithm of the calculated ET rate, plotted as a function of time, where CHARMM replicate 1 is black, 2 is red, and 3 is blue, while the Amber WT simulation is purple and the mutant simulation is shown in light brown. The ET rate was calculated using Marcus theory as described in the text and the distance between the closest atom in the rings of $\mathrm{Trp208_{LmP}}$ and the LmCytc heme. The dashed lines indicate the ET rate obtained from laser flash photolysis experiments for the yeast CCP–Cytc system,[50] the computed rate using the Trp208LmP–hemeLmCytc closest distance obtained from the LmP–Cytc structure,[32] and k_{cat} obtained from steady-state kinetics.[30] (Bottom) The distance between the center of mass of LmCytc and the LmP active site residue D211 over the course of each replicate is tracked using the same color scheme as described above.

This problem has given rise to the "bind and crawl" or "velcro" model of ET,[51] which more recently has witnessed experimental support for transient redox complexes.[52,53] Here, the redox partners initially interact via nonspecific complementary electrostatic surfaces. Next, the partners sample each other's surface in a rapid two-dimensional search until the more energetically favorable ET-active complex is obtained. However, to ensure rapid dissociation, the difference in stability between the ET-active and inactive complexes must be small. The LmP–Cytc system has provided the most detailed molecular level picture on this process. Our previous Brownian dynamics work indicates that LmCytc initially forms a nonspecific complex with helix A and remains there until the $R24_{LmCytc}$–$E101_{LmCytc}$ intramolecular ion pair breaks, which enables $R24_{LmCytc}$ to form an intermolecular ion pair with $D211_{LmP}$. In the present work, we find that exactly the reverse happens in the dissociation reaction. The intermolecular $R24_{LmCytc}$–$D211_{LmP}$ ion pair first must break, which then enables LmCytc to slide toward the helix A prior to full dissociation. In the CHARMM simulations, we observe the rupture of the $R24_{LmCytc}$–$D211_{LmP}$ ion pair in only one of the three simulations, but observe significant movement toward helix A in all three replicates. Further supporting these findings, using AMBER force fields we observe the same motions toward helix A, including breakage of the $R24_{LmCytc}$–$D211_{LmP}$ ion pair, followed by the formation of $R24_{LmCytc}$–$Y134_{LmP}$. The fact that these same specific interactions, as well as the broader motions toward helix A, occur across different simulation conditions increases confidence that the dynamic tug of war between the broad electronegative surface of helix A and the ET-active complex controlled by the $R24_{LmCytc}$–$D211_{LmP}$ ion pair is an accurate picture of the binding/dissociation reactions. In summary, this work, coupled with our previous Brownian dynamics study, presents a consistent picture on the dynamics of both the association and dissociation reactions and provides further support for the "bind and crawl/velcro" model of ET protein–protein interactions and more recent advances on transient redox partner complexes.[52,53]

2.5 Supporting Information

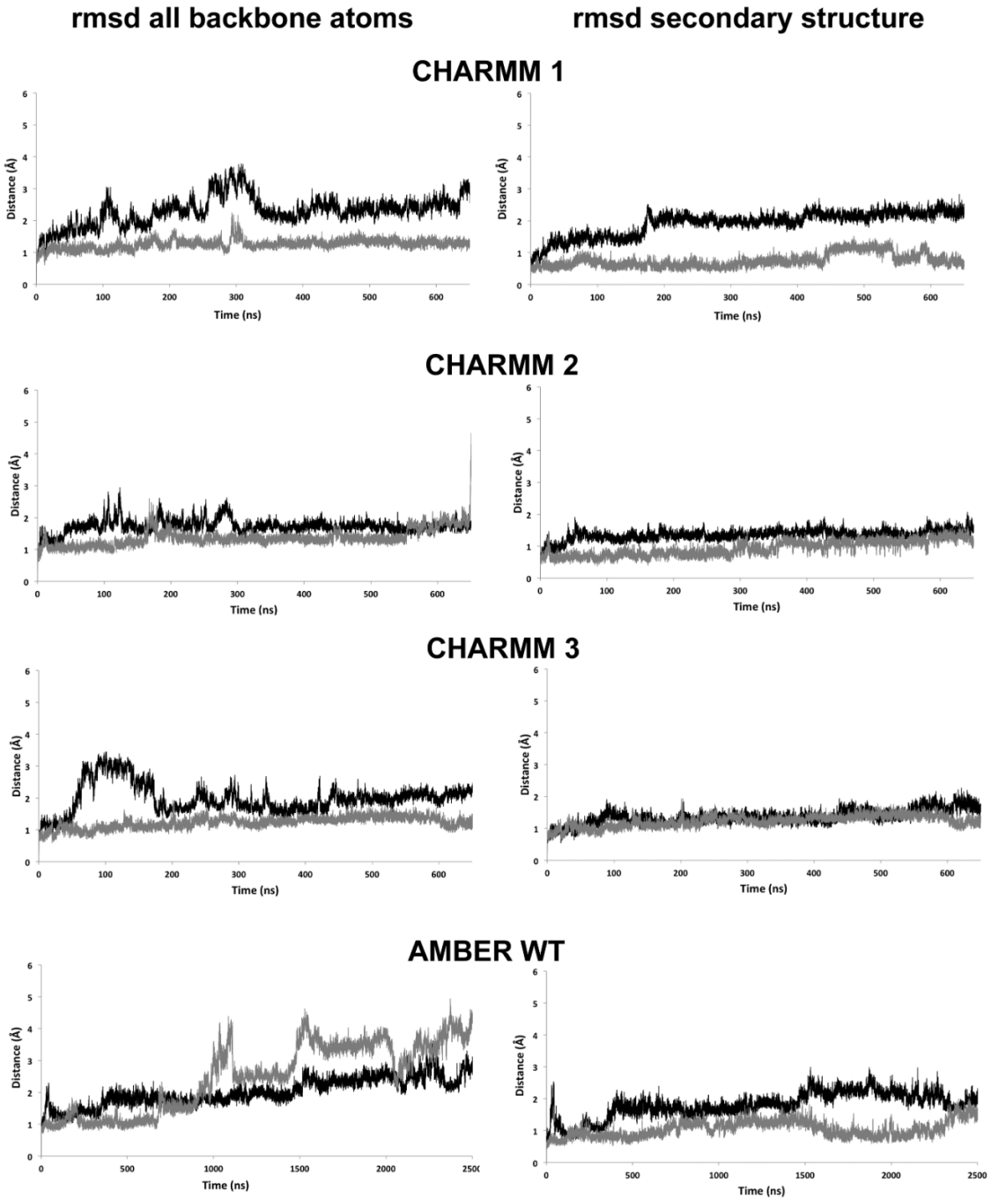


Figure S2.1: Root mean square deviation of all backbone atoms (CA, C, N) relative to the first frame for all 4 wild type MD trajectories. Panels on the left are for the entire protein while on the right only residues in regular secondary structure (helices and sheets) are included. Black lines correspond to LmCytc while gray lines correspond to LmP.

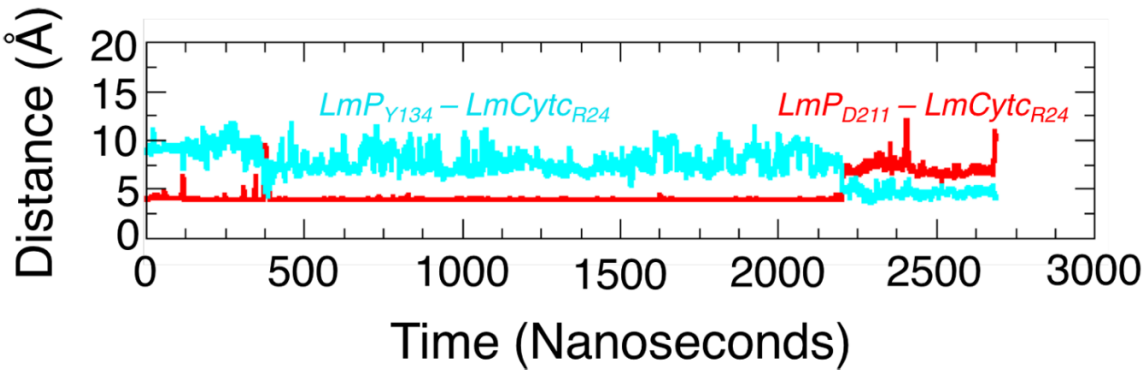


Figure S2.2: The distances between LmP Y134 and LmCytc R24 (cyan) and the electron transfer (ET) active ion-pair of LmP D211 and LmCytc R24 (red) are displayed for the multi-microsecond Amber simulation. Breakage of the ET ion-pair is followed by formation of the intermediate LmP Y134-LmCytc R24 interaction, identical to that observed in CHARMM replicate 1 and Amber mutant simulations.

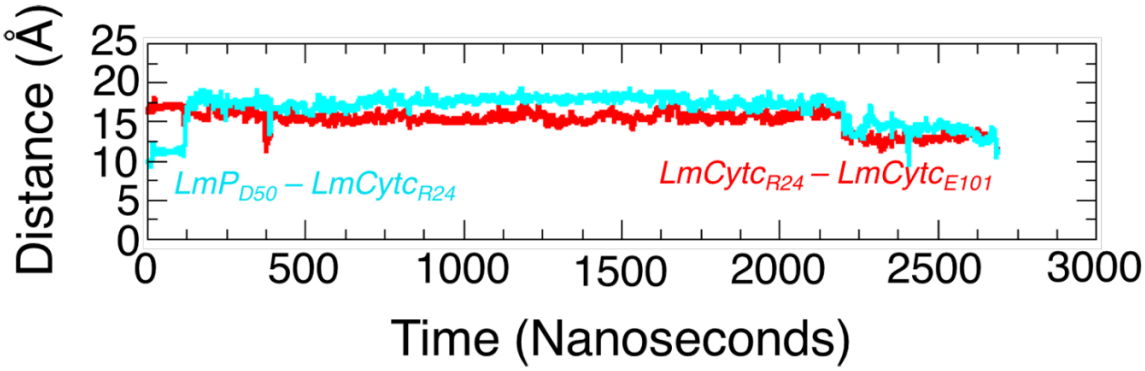


Figure S2.3: The distances between LmP D50 and LmCytc R24 (cyan) and LmCytcR24 and LmCytcE101 (red) for the multi-microsecond Amber simulation replicate.

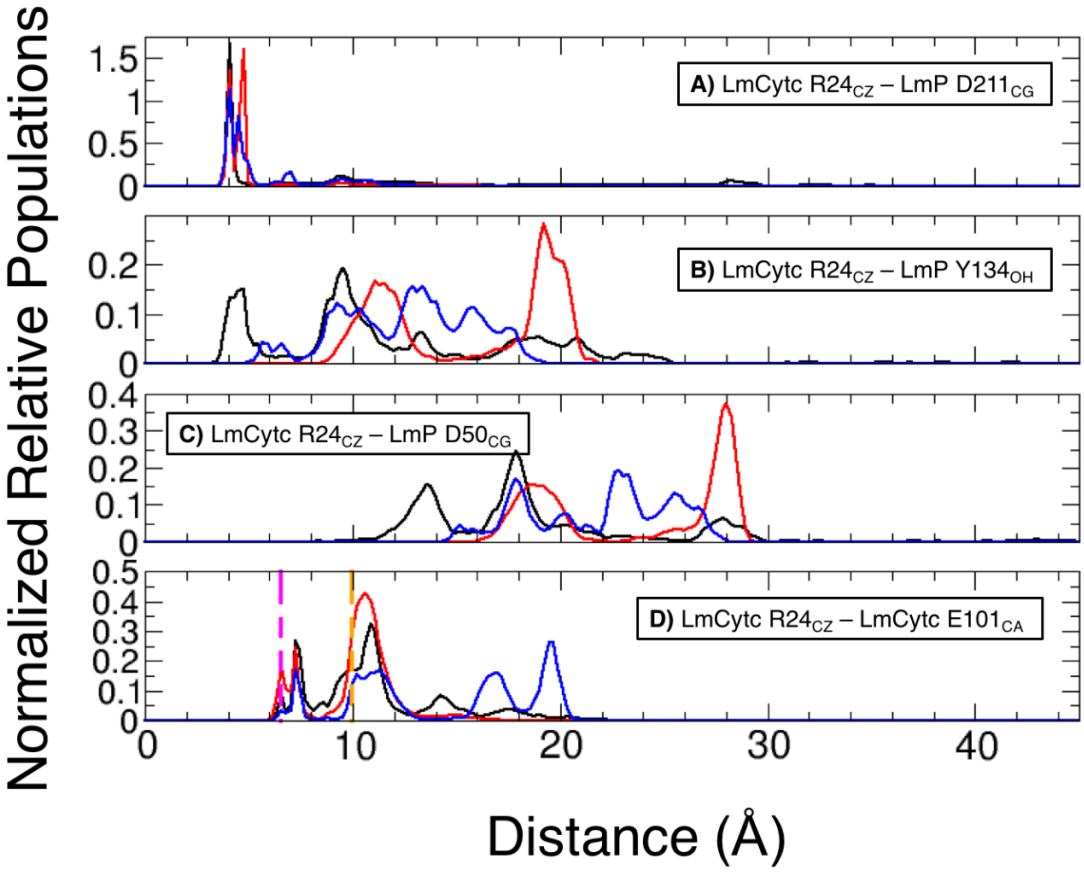


Figure S2.4: Interpolations of the normalized distance distributions between interacting residues are highlighted in A through D where CHARMM replicate 1 is shown in black, CHARMM replicate 2 in red and CHARMM replicate 3 in blue. The highlighted interactions are as follows; A) The electron transfer active interprotein ion-pair of LmCytc R24(CZ) and LmP D211(CG), B) the transient interaction between LmCytc R24(NE) and LmP Y134(OH), C) the distances between LmCytc R24(CZ) and D50 of LmP helix A and, D) the intramolecular ion pair between LmCytc R24(CZ) and LmCytc E101(CA). In D, the experimentally observed distance for the intramolecular ion pair in the individual structure (PDBID 4DY9) is shown in pink and the equivalent distance in the co-crystal structure (PDBID 4GED) in orange.

Chapter 3

Divergence of Many-Body Perturbation Theory for Noncovalent Interactions of Large Molecules

Contribution Statement: The following are my contributions to the project. I revisited and confirmed the large binding energy errors of the second-order Møller-Plesset many-body perturbation theory (MP2) by benchmarking against the S66, S30L, and L7 testsets. I then compared the MP2 results against the random phase approximation (RPA), spin-scaled MP2, and dispersion corrected density functional approximations (DFAs) ensuring that all results were at the complete basis limit. In addition, I performed RPA geometry optimization of the ROT34 testset to assess the performance of RPA for intramolecular interactions. Next, I

analyzed the results by developing the adiabatic connection symmetry adapted perturbation theory (AC-SAPT) at the asymptotic limit. To support the theoretical results, I adapted the TURBOMOLE source code to provide numerical results of the AC-SAPT expansion confirming my observations. Lastly, I contributed to the writing, created all the plots using R, and prepared the submission into the *J. Chem. Theory Comput.*

3.1 Introduction

While covalent bonding is a central paradigm of chemical theory, noncovalent interactions (NIs)[54,55] are often considered secondary due to their "weakness." For small molecules with 10 or less atoms, NIs are at least one order of magnitude smaller than covalent bonds, and their low chemical specificity makes them difficult to detect and control. However, it has long been recognized that NIs are pairwise nonadditive and can grow superlinearly with system size.[56–58] Indeed, NIs are key factors determining conformation, tertiary structure, and other properties of molecular aggregates and complexes,[8,59,60] materials,[2,3] or molecular crystals.[61,62] Recent advances in experimental techniques such as molecular beam spectroscopy[63] have made NIs readily observable in larger molecular systems, and even areas focused on covalent bonding such as synthetic chemistry and catalysis increasingly use NIs to fine-tune reactivity and selectivity.[64,65]

Perhaps with the exception of density functional theory, most electronic structure methods have been developed and tested for small molecules. A central assumption underlying this "bottom-up" approach is that methods performing well for small systems may be scaled up to larger ones without deterioration in accuracy. Size consistency and size extensivity[15,66,67] are often assumed to be sufficient to ensure that the accuracy of an electronic structure method for chemical processes is approximately independent of the system size. Møller–Plesset (MP) many-body perturbation theory (MBPT),[68] which is based on the Fock operator as a zeroth-

order Hamiltonian, has enjoyed much popularity as one of the least expensive yet useful ab-initio correlated electronic methods; size-extensivity of the energy is an often cited advantage of MP theory.[15] Moreover, unlike semilocal density functional approximations (DFAs), whose performance for NIs can be erratic,[69–71] MBPT has widely been considered a qualitatively suitable starting point for modeling NIs, particularly in systems too large to be tractable by more advanced methods.[72–74] This view appears to have emerged from the correct $1/R^6$ behavior of the dispersion energy obtained from MBPT as well as early favorable convergence for small closed-shell systems.[11,14] In a landmark 1993 paper,[14] Moszynski, Jeziorski, and Szalewicz investigated the convergence of the MBPT expansion of the dispersion energy for several small weakly bound complexes and concluded that convergence of the series is "very fast." This may be contrasted with covalent interactions, where MBPT is known to diverge in many systems of chemical interest.[75–77] The assumption that "weak" closed-shell interactions between distant electron pairs are accurately captured by MBPT is implicit in many applications as well as theoretical approaches such as local correlation methods.[78] Against the backdrop of the qualitative inability of semilocal DFAs to capture long-range NIs, this assumption has also motivated the development of efficient computational methods to apply MBPT to systems with 100 and more atoms.[79–83]

However, with the expanding scope of MBPT applications during the past two decades, an increasing number of examples were reported that shows substantial overestimation of NI energies by second-order MP MBPT (MP2).[84] In 2010, Pulay and co-workers pointed out that MP2 overbinds coronene dimer by almost 100% compared to the quadratic configuration interaction reference data.[18,19] Comparing initially to solution phase thermodynamic data[85] and later to coupled cluster singles, doubles, and perturbative triples (CCSD(T)) calculations,[86] Grimme noted that the accuracy of MP2 severely deteriorates for supramolecular systems.[87] Initially, these deviations were viewed as the result of a quantitative rather than qualitative shortcomings of MBPT, which led to the development of empirical correction schemes such as spin-component-scaled MP2 methods[88,89] and MP2.5.[73] For large molecular

complexes, the shortcomings of pairwise additive methods such as MP2 have been ascribed to missing three- and higher-body dispersion interactions,[87] triggering the development of dispersion corrections including MP2C,[23,24,90] MP2D,[91,92] as well as empirical three-,[87,93] and many-body[94] dispersion corrections for DFAs. Meanwhile, Dobson and co-workers showed that the simple addition of pairwise $1/R^6$ interactions yields qualitatively incorrect asymptotic power laws for dispersion interactions between macroscopic solids. For example, the interaction between two large parallel graphene sheets decays as $1/D^3$ with the distance D between the sheets, whereas finite-order MBPT yields a $1/D^4$ behavior.[95,96] These troubling inconsistencies and the sheer magnitude of the errors raise the question whether and to what extent MBPT is fundamentally adequate for NIs of large but finite molecules.

To address this question, we revisit the performance of MBPT with particular emphasis on large molecules with 100 and more atoms, and contrast it with the random phase approximation (RPA) to the ground-state correlation energy in a density functional context.[97–99] Particle–hole RPA may be viewed as a resummation of ring diagrams[100] which correspond to direct Coulomb interactions between particle–hole pairs and constitute the most long-ranged correlation contribution to the interaction energy between closed-shell systems.[55,101] Indeed, the accuracy of RPA for NIs in small molecules,[102,103] rare-gas solids,[104] and layered materials[96] is well documented, but few results for intra- and intermolecular NIs in large systems are available.[105,106] Building on efficient RPA implementations for energy[107] and analytic derivatives,[108] we investigate the size-dependence of MBPT and RPA interaction energies using the S66,[13,109] L7,[110] and S30L[111] benchmarks in Section 3.3. These benchmarks contain systems ranging from 6 to 204 atoms, with binding energies between 1 kcal/mol and 136 kcal/mol. To evaluate whether our observations for energetics also hold for structures, we also compare rotational constants from MP2, RPA, and Grimme's dispersion corrected DFA-D3[87] structure optimizations to gas-phase spectroscopy data using organic molecules with 18–35 atoms from the ROT34 benchmark set.[112,113]

In Section 3.4, these results are analyzed in detail using a symmetry-adapted perturbation theory (SAPT) type[114,115] asymptotic theory of NIs between closed-shell fragments whose ground-state density is constrained to the supersystem ground-state density using a local one-body potential in the spirit of the adiabatic connection (AC)[97,98,116,117] in density functional theory (DFT). This leads to a compact, non-perturbative expression for the dispersion energy, as well as explicit estimates for the convergence radius of the AC-SAPT expansion. We investigate how the estimated convergence radii correlate with errors of MP2 and RPA calculations of NIs and the system size. A numerical model for the divergence of the AC-SAPT expansion for moderately large and polarizable systems is obtained by re-expansion of the RPA correlation energy into powers of the interaction. Conclusions for electronic structure theory and computational practice are presented in Sec. 3.5.

3.2 Methods

3.2.1 Computational Details

MP2 energies were evaluated using on self-consistent Hartree–Fock (HF) orbitals. Variants of MP2 such as spin-component-scaled MP2 (SCS-MP2)[88] and scaled opposite-spin MP2 (SOS-MP2)[89] were also assessed. All MP2 calculations were performed with the RI approximation (RI-MP2) using the `ricc2` module[118] of TURBOMOLE.[119]

RPA energies and analytic derivatives were obtained in a post-Kohn–Sham (KS) fashion, i.e., a KS calculation using a semilocal DFA was first performed to obtain the KS orbitals, and subsequently the exact exchange energy and the RPA correlation energy were evaluated non-self-consistently. The Perdew–Burke–Ernzerhof (PBE)[120] and Tao–Perdew–Staroverov–Scuseria (TPSS)[121] DFAs were used for the KS calculations; the ensuing RPA calculations, dubbed RPA(PBE) and RPA(TPSS), respectively, employed the resolution-of-the-identity

(RI) approximation and the imaginary frequency integration technique as implemented in the `rirpa` module[107] of TURBOMOLE.[119] Perturbative order-by-order analysis of the RPA correlation energies was carried out using a modified version of TURBOMOLE 7.3.

Tight convergence criteria of 10^{-9} Hartrees for the energy and 10^{-7} atomic units (a.u.) for the root mean square change of the one-particle density matrix were used for the KS and HF self-consistent field iterations. DFA quadrature grids of m5 quality[122] were used for the KS reference calculations. Imaginary frequency grids of 100 points were employed for the RPA energy calculations. Interaction energies were computed based on the supermolecular approach and extrapolated to the complete basis set (CBS) limit, as detailed in the next subsection. RPA structural optimizations[108] of molecules in the ROT34 benchmark set were converged to a maximum Cartesian gradient norm $\leq 10^{-4}$ a.u. and 10^{-7} a.u. in the energy change. Fine imaginary frequency grids of 200 points were used for RPA gradient calculations and structure optimizations.

The interaction energies were benchmarked against CCSD(T) values for the S66 benchmark set[13,109] and against the domain-based local pair-natural orbital (DLPNO) based CCSD(T) calculations for the L7 and S30L benchmark sets.[123,124] DLPNO-CCSD(T) results can vary significantly for weakly bound complexes depending on the truncation of the PNO basis and domain size.[125,126] Moreover, even with tight truncation thresholds, the results may differ by up to $\sim$2 kcal/mol based on the choice of basis set and basis-set extrapolation scheme as seen by Refs. 124 and 126. For the present study, the DLPNO-CCSD(T) reference values were taken from Brandenburg *et al*[124] employing the "TightPNO" truncation thresholds[125] and the basis-set extrapolation scheme from Ref. 127 for both the L7 and S30L benchmarks. Throughout the paper, signed errors are defined as differences between calculated and reference values; for example, a positive error in binding energies signifies underbinding.

3.2.2 Basis Set Convergence

The RPA total energy is the sum of the energy expectation value of the KS determinant, i.e., the sum of the zeroth- and first-order energies, and the RPA correlation energy. These two parts of the energy exhibit qualitatively different basis set convergence.[128] The energy expectation value of the KS determinant was evaluated within the RI-JK algorithm with the corresponding optimized auxiliary basis sets.[129] Karlsruhe segmented-contracted polarized quadruple-ζ (def2-QZVP) basis sets[130,131] were chosen for the KS reference calculations because they were found to yield significantly faster convergence of the KS energy expectation value than the corresponding correlation-consistent basis sets using generalized contractions.[132–134]

RPA correlation energies were evaluated using Dunning's correlation-consistent polarized valence basis sets[132,133] in conjunction with the corresponding auxiliary correlation-consistent basis sets optimized for RI-MP2.[134,135] Small core relativistic effective core potentials[136,137] were used for the halogen atoms in the S30L[111] benchmark set. The frozen core approximation was employed for the RPA correlation energy calculations. Basis set superposition error was estimated by 50% counterpoise (CP) correction as recommended by Risthaus and coworkers;[138] the CP correction was only applied to the RPA correlation energy.

The CBS limit of the RPA correlation energy was estimated using the two-point $1/X^3$ extrapolation, where $X = 3$ (triple-ζ), 4 (quadruple-ζ), etc.[128,139] Dunning's correlation-consistent polarized triple- (cc-pVTZ) and quadruple-ζ (cc-pVQZ) valence basis sets[132,133] were employed for the 3-4 extrapolation. The basis set dependence of the S30L interaction energies is displayed in Figure 3.1. To assess the residual basis set error, exploratory calculations were performed using Dunning's cc-pV5Z basis sets[134] for the pincer complex with 2,4,7-trinitro-9-fluorenone as the guest molecule (TNF@tweezer2, compound 5).[111] The 4-5 extrapolation was found to be within 0.20 kcal/mol compared to the 3-4 extrapolation, see Figure S3.1 in

the Supporting Information.

Similarly, correlation energies from MP2 and its variants were obtained using the frozen core approximation, 50% CP correction, and the 3-4 extrapolation using cc-pVTZ and cc-pVQZ basis sets.[132,133] HF energies were computed using def2-QZVP basis set.[130,131]

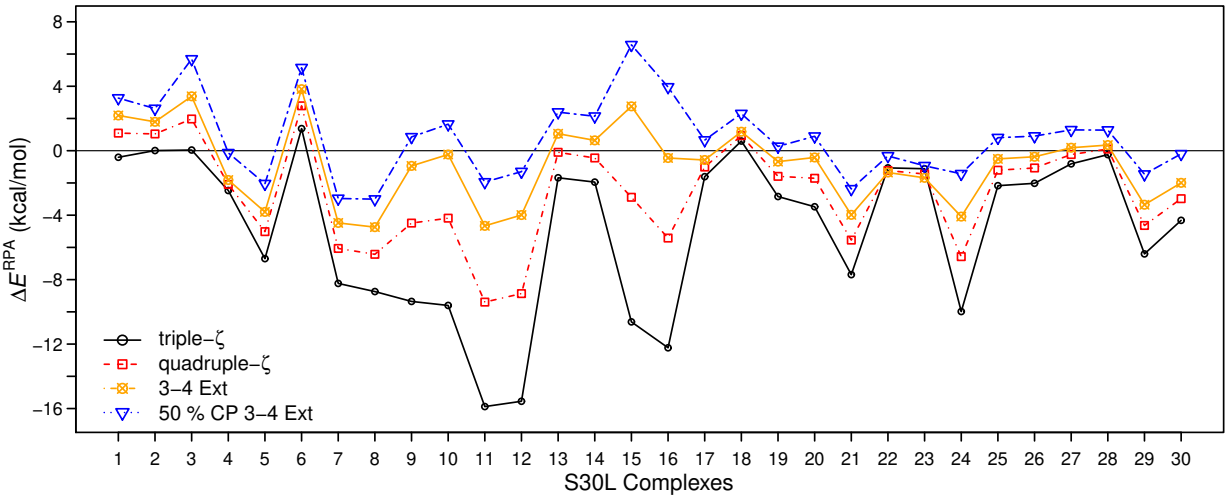


Figure 3.1: S30L interaction energy errors (ΔE^{RPA}) computed using RPA(PBE) with correlation-consistent triple-ζ and quadruple-ζ basis sets, as well as using triple-quadruple (3-4) extrapolation with 50% and without counterpoise (CP) correction. DLPNO-CCSD(T) reference values are from Caldeweyher *et al.*[123]

For the ROT34 benchmark,[112,113] def2-QZVP basis sets were used for both the KS expectation value and the RPA correlation; core electrons were treated explicitly. This approach is expected to yield RPA structures of near basis-set limit quality.[108] Indeed, changing the basis sets from def2-TZVP to def2-QZVP only yields a small but systematic decrease in the error in the RPA rotational constants, see Supporting Information.

Table 3.1: Mean absolute errors (MAE), mean errors (ME), and absolute minimum-maximum error range (MinMax) in kcal/mol of various methods at complete basis set limit for the S66, L7, and S30L test sets. Positive ME corresponds to underbinding.

	S66			L7			S30L		
Methods	MAE	ME	MinMax	MAE	ME	MinMax	MAE	ME	MinMax
RPA(PBE)	0.61	0.61	1.00	1.72	1.47	3.53	2.03	0.82	6.44
RPA(TPSS)	0.63	0.63	1.02	1.92	1.73	3.75	2.05	1.33	8.64
MP2	0.35	−0.54	2.30	8.77	−8.77	17.51	18.74	−18.74	66.07
MP3[a]	0.47	0.47	1.93	6.71	6.26	13.13	−−	−−	−−
SCS-MP2	0.32	0.64	1.92	2.49	−1.41	5.77	7.43	−4.68	33.47
SOS-MP2	0.65	1.23	2.57	2.28	2.27	5.83	6.26	2.35	16.26
PBE-D3[b]	0.34	−0.24	1.93	1.91	0.92	4.39	3.06	1.83	9.80
PBE-D4[b]	0.34	−0.30	1.66	1.71	0.16	2.44	2.94	−0.77	9.30
PW6B95-D3[b]	−−	−−	−−	1.39	1.19	2.20	1.82	0.32	4.70
PW6B95-D4[b]	−−	−−	−−	1.80	1.40	3.30	2.02	0.96	4.10

[a] MP3 values from Refs. 109 and 110.
[b] Dispersion corrected DFA values for L7 and S30L from Ref. 123.

3.3 Results

3.3.1 S66, L7, and S30L Interaction Energy Benchmarks

The S66 benchmark consists of binding energies of 66 complexes ranging from 6 to 34 atoms; the average binding energy is 5.50 kcal/mol.[13,109] This benchmark set is divided into groups featuring hydrogen bonding, π–π stacking, aliphatic–aliphatic interactions, π–aliphatic interactions, and other nonspecific interactions, respectively. For the binding energies of these complexes, MBPT is accurate: MP2 and third-order MP MBPT (MP3) yield mean absolute errors (MAEs) of 0.35 kcal/mol and 0.47 kcal/mol, respectively. The errors of dispersion-corrected DFT and MP2 variants are comparable. The good performance of dispersion corrected DFAs is hardly surprising here since S66 is part of commonly used training sets for parameter estimation.[93]

The L7 benchmark contains binding energies of seven complexes: Octadecane dimer, guanine trimer, circumcoronene–adenine dimer, coronene dimer, guanine–cytosine dimer, circumcoronene–

guanine–cytosine dimer, and an amyloid fragment trimer containing phenylalanine residues; the average binding energy is 16.7 kcal/mol.[110,124] As shown in Table Table 3.1, MP2 performs poorly with an MAE of 8.10 kcal/mol, which is an order of magnitude more larger for S66. Aside from the dependence of errors on system size discussed in Section 3.4.6, MP2 is known to systematically overestimate π–π stacking interactions even in smaller systems.[109] The inclusion of higher orders does not systematically improve the MBPT results. For both, the S66 and L7 benchmark sets, MP2 and MP3 mean errors are on the same orders of magnitude but of opposite sign. Empirically, one observes that odd MBPT orders tend to produce underbinding, whereas even orders produce overbinding.[76,77] The poor performance of MBPT for L7 and especially S30L is in sharp contrast to the one observed for RPA, dispersion corrected PBE-D3, and the recently developed PBE-D4,[93] which all yield MAEs in the range of 2 kcal/mol. The RPA L7 results reported here are $\sim 50\%$ more accurate than the ones previously obtained using def2-TZVP basis sets,[140] underlining the importance of basis set extrapolation for RPA interaction energy benchmarks.[128]

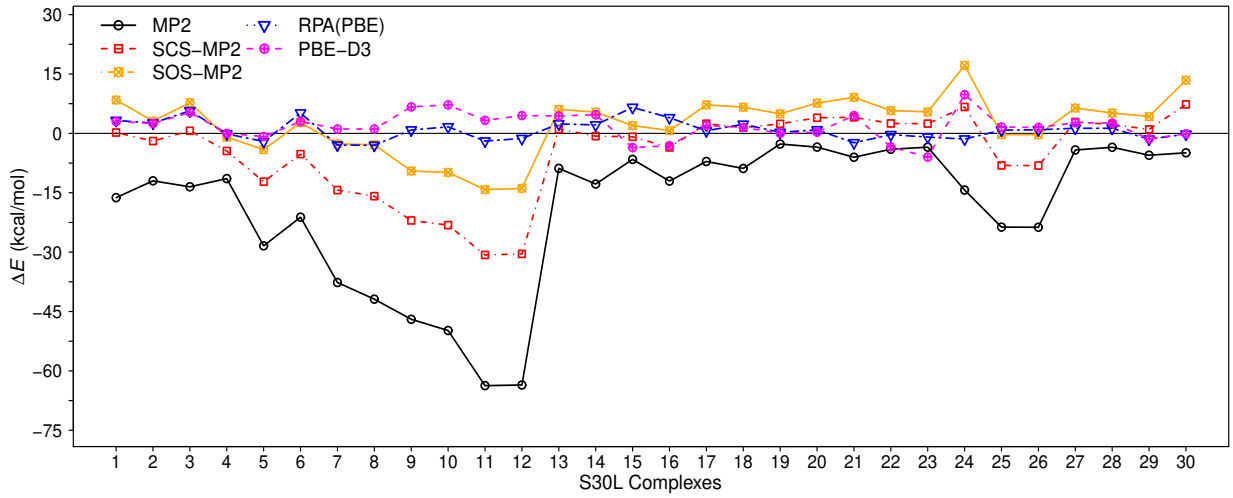


Figure 3.2: S30L interaction energy errors (ΔE) for MP2 variants, RPA(PBE), and dispersion corrected PBE-D3. MP2 and RPA(PBE) results are 3-4 extrapolated, and PBE-D3 results use def2-QZVP basis sets. DLPNO-CCSD(T) reference values are from Caldeweyher *et al.*[123]

The S30L benchmark set contains binding energies of 30 large supramolecular complexes including π stacking and CH–π interactions, hydrogen and halogen bonding, and charged

species; the average binding energy is 37.5 kcal/mol.[111,124] MP2 exhibits severe overbinding for most species, producing spectacular errors > 60 kcal/mol for the π stacked complexes 11 and 12, see Figure 3.2 and Table Table 3.1. With MAEs of 7.43 and 6.26 kcal/mol, SCS-MP2 and the SOS-MP2 inherit the shortcomings of MP2 to a significant degree. RPA yields MAEs close to 2 kcal/mol regardless of the KS reference, consistent with its performance observed for L7. Dispersion corrected DFAs show comparable MAEs, but behave less systematically than RPA, as evidenced by somewhat larger absolute minimum-maximum error ranges. The present S30L results for RPA are more than twice as accurate as the ones reported by Heßelmann for the S12L subset;[105] this is likely a consequence of the improved CCSD(T) reference values[123] used here as well as better controlled basis set errors.

3.3.2 ROT34: Intramolecular Interactions

To investigate whether the strong performance of RPA for intermolecular binding energies translates to other properties such as molecular structures of larger flexible molecules, the equilibrium structures of 34 organic molecules contained in the ROT34 benchmark[112,113] were optimized using RPA with def2-QZVP basis sets; rotational constants were calculated in the rigid rotor approximation. Rotational constants are a sensitive measure of intramolecular mid- and long-range interactions, and accurate experimental values are available from gas-phase rotational spectroscopy.[112,113]

As displayed in Figure 3.3, the PBE DFA produces a MAE of 18.2 MHz, even with D3 dispersion correction; the Minnesota DFA M06L[141] and the strongly constrained and appropriately normed (SCAN)[142] DFA perform significantly better:[143] For SCAN, SCAN-D3, and M06L MAEs of 3.7 MHz, 3.3 MHz, and 4.0 MHz were reported.[143] The MAEs of MP2 and SCS-MP2, on the other hand, are 5.5 MHz and 5.4 MHz,[113] respectively, comparing unfavorably with the RPA(PBE) result of 3.1 MHz; using the TPSS instead of the PBE

DFA to generate the KS reference yields an almost identical MAE of 3.0 MHz. Thus, even for the moderately sized systems contained in ROT34, the sub-par performance of MP2 is notable.

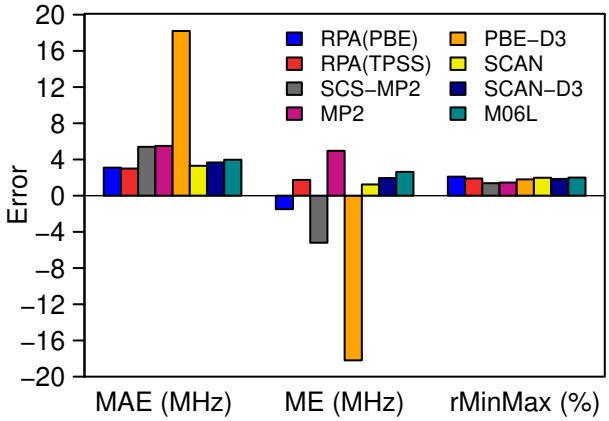


Figure 3.3: Errors in ROT34 computed rotational constants compared to experiment.[113] Mean absolute errors (MAE) and mean errors (ME) are in MHz, and relative minimum-maximum error ranges (rMinMax) are in %. MP2 and SCS-MP2 results are from Ref. 113; and SCAN, SCAN-D3, PBE-D3, and M06L results are from Ref. 143.

3.4 AC-SAPT Analysis

3.4.1 Statement of the Problem

In the following, the results of the previous sections are analyzed in two steps: First, we derive asymptotically exact expressions for the dispersion energy at large separation. We rely on the general SAPT partitioning of the Hamiltonian,[115] but unlike prior DFT-SAPT methods[144,145] or van-der-Waals inclusive frozen density embedding,[146,147] the present AC-SAPT approach uses density functional theory only for the inter-fragment interaction and constrains the density of the monomers, leading to compact closed-form expressions for the interaction energy and a separation of dispersion and induction effects. We derive an upper bound for the convergence radius of the resulting AC-SAPT expansion. Second, we show that

the RPA dispersion energy corresponds to a partial resummation of the AC-SAPT expansion, whereas the MBPT dispersion energy corresponds to a finite-order approximation. When the monomers are treated within RPA, the AC-SAPT converges for nondegenerate monomers, whereas it is found to be susceptible to spurious divergence for MBPT treatments of the monomers.

We consider a molecular supersystem or complex A–B consisting of two non-overlapping subsystems or fragments A and B, i.e., the inter-fragment distance R is assumed to be so large that the overlap between the two ground states is exponentially small. Since there is vanishing charge transfer in this limit, the fragments have integer electron numbers $N_{\mathrm{A}}, N_{\mathrm{B}}$ with ground states resembling the lowest-energy separated fragment (dissociation) limit of A–B. The Born–Oppenheimer Hamiltonian of the supersystem at coupling strength α is

$$\hat{H}^{\alpha} = \hat{H}_{\mathrm{A}} + \hat{H}_{\mathrm{B}} + \alpha \hat{V}_{ee\,\mathrm{int}} + \hat{V}^{\alpha}_{s\,\mathrm{int}}[\rho], \tag{3.1}$$

where $\hat{H}_{\mathrm{A}}$ and $\hat{H}_{\mathrm{B}}$ denote the N_{A}- and N_{B}-electron Hamiltonians of the isolated subsystems, and $\hat{V}_{ee\,\mathrm{int}}$ is the operator of the electron-electron Coulomb interaction *between* A and B. The supersystem eigenstates $|\Psi^{\alpha}_{n}\rangle$, which are constrained to be antisymmetric under any permutations of electrons in the supersystem as in conventional SAPT, and their energies E^{α}_{n} are defined by

$$\hat{H}^{\alpha}\,|\Psi^{\alpha}_{n}\rangle = E^{\alpha}_{n}\,|\Psi^{\alpha}_{n}\rangle\,. \tag{3.2}$$

$\hat{V}^{\alpha}_{s\,\mathrm{int}}[\rho]$ represents a local one-electron potential which constrains the density of the supersystem ground state $|\Psi^{\alpha}_{0}\rangle$ to the physical ground-state density $\rho = \rho^{1} = \rho^{\alpha}|_{\alpha=1}$. Since $\hat{V}^{\alpha}_{s\,\mathrm{int}}[\rho]$ is a unique functional of ρ,[148] so are $|\Psi^{\alpha}_{n}\rangle$ and the corresponding energies E^{α}_{n}. Throughout this paper, the ground state $|\Psi^{\alpha}_{0}\rangle$ is assumed to be nondegenerate for finite R, a mild condition typically satisfied for interacting closed-shell fragments. The use of symmetry adaption

implies that the present approach is valid only at large R as in conventional SAPT,[149] but this is sufficient for asymptotic analysis.

In analogy with the conventional KS potential,

$$\hat{V}^{\alpha}_{s\,\mathrm{int}}[\rho] = \hat{V}_{ne\,\mathrm{int}} + V_{nn\,\mathrm{int}} + \hat{V}^{\mathrm{HXC}}_{\mathrm{int}}[\rho] - \hat{V}^{\alpha\mathrm{HXC}}_{\mathrm{int}}[\rho] \tag{3.3}$$

may be defined as a sum of the "external" one-electron nucleus-electron potential $\hat{V}_{ne\,\mathrm{int}}$ and the constant nucleus-nucleus attraction $V_{nn\,\mathrm{int}}$ between the fragments plus a remainder accounting for Hartree-, exchange-, and correlation (HXC) effects. At full coupling ($\alpha = 1$), the HXC part of $\hat{V}^{\alpha}_{s\,\mathrm{int}}[\rho]$ vanishes, whereas at $\alpha = 0$, $\hat{V}^{\alpha}_{s\,\mathrm{int}}[\rho]$ equals the KS potential arising from the interaction between the fragments. Since the KS potential is spatially local, it additively separates into A and B parts for large R, giving rise to a unique partitioning of the supersystem density into a sum of subsystem densities, i.e., $\rho(x) = \rho_{\mathrm{A}}(x) + \rho_{\mathrm{B}}(x)$. Thus, the present approach is closely related to partition DFT (PDFT).[150,151] However, PDFTs[144,145] and related embedding schemes[146,147] typically start from a KS-DFT calculation of the supersystem, whereas here the fragment ground states corresponding to $\alpha = 0$ include the full intra-fragment electron-electron interaction and therefore are generally not Slater determinants.

3.4.2 Interaction Energy

We define the A–B interaction energy as the difference in the ground state energies at full and zero coupling,

$$E_{\mathrm{int}}[\rho] = E^{1}_{0}[\rho] - E^{0}_{0}[\rho]. \tag{3.4}$$

Using the Hellman–Feynman theorem, the interaction energy may be expressed as a coupling strength average of the potential energy of interaction,[151]

$$E_{\text{int}}[\rho] = \int_0^1 d\alpha\, W^\alpha[\rho], \tag{3.5}$$

where $W^\alpha[\rho] = dE_0^\alpha/d\alpha = \langle\Psi_0^\alpha|\hat{V}_{ee\,\text{int}}|\Psi_0^\alpha\rangle$; one-electron contributions vanish upon coupling strength integration due to the density constraint. The AC-SAPT expansion of the interaction energy is obtained by expansion of the interacting ground state and the corresponding energy into powers of α, analogous to Görling–Levy perturbation theory.[152,153] Equivalently, the coupling strength integrand W^α may be expanded around $\alpha = 0$, yielding the AC-SAPT series expansion of the interaction energy,

$$E_{\text{int}}[\rho] = \int_0^1 d\alpha \sum_{k=0}^{\infty} \alpha^k W^{(k)}[\rho] = \sum_{k=0}^{\infty} \frac{1}{k+1} W^{(k)}[\rho]. \tag{3.6}$$

The first-order interaction energy results from evaluating the integrand at $\alpha = 0$,

$$E_{\text{int}}^{(1)}[\rho] = \langle\Psi_0^0|\hat{V}_{ee\,\text{int}}|\Psi_0^0\rangle = E_{\text{int}}^{\text{HX}}[\rho] = \int dx_1 dx_2 \frac{\rho_A(x_1)\rho_B(x_2) - \gamma_A(x_1,x_2)\gamma_B(x_2,x_1)}{|\mathbf{r}_1 - \mathbf{r}_2|}, \tag{3.7}$$

where $|\Psi_0^0\rangle$ is the antisymmetrized product of the two fragment density-constrained ground-state wavefunctions with one-particle density matrices γ_A, γ_B. As in standard SAPT, the first-order interaction energy is electrostatic and corresponds to the sum of the Hartree and exchange (HX) interactions between the two fragments; unlike in standard SAPT, and as discussed in detail below, the interaction arises from the electrons only. The exchange term is exponentially small for non-degenerate ground states in the large R limit.

The remaining correlation part of the interaction energy,

$$E_{\text{int}}^{\text{C}}[\rho] = \int_0^1 d\alpha \left(\langle\Psi_0^\alpha|\hat{V}_{ee\,\text{int}}|\Psi_0^\alpha\rangle - \langle\Psi_0^0|\hat{V}_{ee\,\text{int}}|\Psi_0^0\rangle\right) \tag{3.8}$$

is due to dispersion and does not contain any terms describing changes in the fragment densities as the interaction is turned on. The purely dispersive character of the interaction becomes apparent from a factorization of $\hat{V}_{ee\,\mathrm{int}}$ in a fashion analogous to Eqs. (6)-(13) of Ref. 154 and neglecting exponentially small exchange terms, yielding

$$E_{\mathrm{int}}^{\mathrm{C}}[\rho] = \int_0^1 d\alpha \int dx_1 dx_2 \frac{\langle \Psi_0^\alpha | \Delta\hat{\rho}_A(x_1) \Delta\hat{\rho}_B(x_2) | \Psi_0^\alpha \rangle - \langle \Psi_0^0 | \Delta\hat{\rho}_A(x_1) \Delta\hat{\rho}_B(x_2) | \Psi_0^0 \rangle}{|\mathbf{r}_1 - \mathbf{r}_2|}, \quad (3.9)$$

where $\Delta\hat{\rho}_A(x) = \hat{\rho}_A(x) - \rho_A(x)$ is the density fluctuation operator associated with fragment A, $\hat{\rho}_A(x)$ is the corresponding density operator, and $\Delta\hat{\rho}_B(x)$ is defined analogously. As opposed to standard SAPT (including DFT-SAPT), which is based on a partitioning of the Hamiltonian that includes induction effects to all orders, the interaction energy in the present approach does not contain any induction terms and is exclusively due to electrostatic (Hartree plus exchange) and dispersion. Factorization of the density operator product in Eq. (3.9) using the completeness of the eigenstates of $\hat{H}^\alpha$ yields a spectral sum over Hartree interactions between the A and B parts of ground-to-n-th excited state transition densities $\rho_{0n}^\alpha(x) = \rho_{n0}^{\alpha *}(x)$,

$$E_{\mathrm{int}}^{\mathrm{C}}[\rho] = \sum_{n \neq 0} \int_0^1 d\alpha \int dx_1 dx_2 \frac{\rho_{0n\,A}^\alpha(x_1) \rho_{n0\,B}^\alpha(x_2) - \rho_{0n\,A}^0(x_1) \rho_{n0\,B}^0(x_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}. \quad (3.10)$$

The $\alpha = 0$ term vanishes for large R since excitations of the non-interacting monomers are localized on either monomer, but is included here to emphasize the analogy to general RPA theory.[155] In this sense, the A–B dispersion energy is given exactly by the Hartree interaction between "electrodynamic" density fluctuations of the monomers. While similar ideas are implicit, e.g., in molecular quantum electrodynamics,[156] the present approach yields a compact, exact expression valid beyond perturbation theory. The induction energy, on the other hand, appears as the difference between the non-interacting ($\alpha = 0$) monomer ground

state energies with and without density constraint,

$$E_{\text{ind}} = E^0[\rho] - E_{\text{A}} - E_{\text{B}}. \tag{3.11}$$

The total dissociation energy of the complex A–B is thus

$$D_{\text{A–B}} = E_{\text{int}}[\rho] + E_{\text{ind}}. \tag{3.12}$$

Induction effects are comparatively small for large R;[157] hence, we will focus on $E_{\text{int}}[\rho]$ in the following.

If the zero-point fluctuation–dissipation theorem (FDT) is invoked[55,97,98,116,117] to factorize the products of fluctuation operators in Eq. (3.9), the dispersion energy may be expressed as

$$E_{\text{int}}^{\text{C}}[\rho] = -\frac{1}{2}\int_0^1 d\alpha \int_{-\infty}^{\infty} \frac{dz}{2\pi i} \left\langle \left(\mathbf{\Pi}^\alpha(z) - \mathbf{\Pi}^0(z)\right) \mathbf{V}_{\text{int}} \right\rangle. \tag{3.13}$$

Here, $\mathbf{\Pi}^\alpha(z)$ is the time-ordered supersystem polarization propagator at coupling strength α and imaginary frequency $z = i\omega \in i\mathbf{R}$ defined as[158]

$$\mathbf{\Pi}^\alpha(z) = -\sum_{n\neq 0} \left\{ \frac{\boldsymbol{\gamma}_{0n}^\alpha \otimes \boldsymbol{\gamma}_{0n}^{\alpha\dagger}}{z - \Omega_n^\alpha + i0^+} - \frac{\boldsymbol{\gamma}_{0n}^{\alpha\dagger} \otimes \boldsymbol{\gamma}_{0n}^\alpha}{z + \Omega_n^\alpha - i0^+} \right\}. \tag{3.14}$$

$\Omega_n^\alpha = E_n^\alpha - E_0^\alpha$ is the energy of an excitation from $|\Psi_0^\alpha\rangle$ to $|\Psi_n^\alpha\rangle$, and $\boldsymbol{\gamma}_{0n}^\alpha$ denotes the corresponding one-particle transition density matrix. Eq. (3.14) shows that, for non-degenerate monomer ground states, $\mathbf{\Pi}^\alpha(z)$ is a self-adjoint and negative semidefinite operator on the tensor-product space of one-particle operators. $\mathbf{V}_{\text{int}}$ represents the bare inter-fragment electron-electron Coulomb interaction or the Hartree kernel on the same space.

Eqs. (3.9)-(3.10) could also be used to define the dispersion energy in conjunction with MP-

style partitioning of the Hamiltonian, using the analogous, albeit approximate, HF-based AC framework.[154] Formally, this corresponds to replacing the local exchange-correlation potential in Eq. (3.3) with the non-local HF exchange potential. The $\alpha = 0$ reference of this approach is equivalent to monomers with full intra-fragment electron-electron interaction whose inter-fragment interaction is treated at the Hartree plus exchange level. In the HF-based AC framework, the interaction energy obtained from coupling strength integration contains additional induction effects resulting from changes in the fragment densities due to inter-fragment correlation which are not captured by the FDT, but the present conclusions for the dispersive part of the interaction energy remain valid.

3.4.3 Dispersion Energy

Expression (3.13) for the dispersion energy can be further re-cast by noting that any contributions from charge transfer excitations between the fragments vanish exponentially due to exponentially vanishing overlap at large R; it therefore suffices to consider $\mathbf{\Pi}^{\alpha}(z)$ on the domain of fragment-centered excitations only. Thus, $\mathbf{\Pi}^{\alpha}(z)$ may be partitioned as

$$\mathbf{\Pi}^{\alpha}(z) = \begin{pmatrix} \mathbf{\Pi}^{\alpha}_{\mathrm{AA}}(z) & \mathbf{\Pi}^{\alpha}_{\mathrm{AB}}(z) \\ \mathbf{\Pi}^{\alpha}_{\mathrm{BA}}(z) & \mathbf{\Pi}^{\alpha}_{\mathrm{BB}}(z) \end{pmatrix}, \tag{3.15}$$

where indices AA refer to the 4-index tensor space spanned by products of transition density matrices centered on fragment A, etc. In the non-interacting ($\alpha = 0$) case, all excitations are either excitations of A or B only, hence $\mathbf{\Pi}^{0}_{\mathrm{AB}}(z) = \mathbf{\Pi}^{0}_{\mathrm{BA}}(z) = \mathbf{0}$, and the diagonal parts reduce to the fragment polarization propagators (at full intra-fragment coupling). The inter-fragment Coulomb interaction may be partitioned as

$$\mathbf{V}_{\mathrm{int}} = \begin{pmatrix} \mathbf{0} & \mathbf{V}_{\mathrm{AB}} \\ \mathbf{V}_{\mathrm{BA}} & \mathbf{0} \end{pmatrix}, \tag{3.16}$$

where the diagonal blocks must vanish to recover the correct monomer limit at large R.

It is instructive to introduce the dielectric operator $\boldsymbol{\epsilon}^{\alpha}(z)$ (also called generalized dielectric function or matrix[159,160]) via

$$\boldsymbol{\Pi}^{\alpha}(z) = \boldsymbol{\epsilon}^{\alpha}(z)^{-1}\boldsymbol{\Pi}^{0}(z). \tag{3.17}$$

In the spirit of the Bethe–Salpeter equation,[158] $\boldsymbol{\epsilon}^{\alpha}(z)$ may be expressed as

$$\boldsymbol{\epsilon}^{\alpha}(z) = \mathbf{1} - \boldsymbol{\Pi}^{0}(z)\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z), \tag{3.18}$$

where $\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z)$ denotes the imaginary-frequency-dependent HXC kernel at coupling strength α for the intersystem interaction. Similar to the Hartree kernel, the elements in the diagonal blocks of $\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z)$ vanish, because $\boldsymbol{\Pi}^{0}(z)$ contains the full intra-fragment interaction.

Since $\mathbf{V}_{\mathrm{int}}$ and $\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}$ have vanishing diagonal blocks, these can be used to further simplify the dispersion energy: Noting that the diagonal blocks of $\boldsymbol{\Pi}^{0}(z)\mathbf{V}_{\mathrm{int}}$ vanish, and using Eq. (3.17), the FDT (3.13) takes the form

$$E_{\mathrm{int}}^{\mathrm{C}}[\rho] = -\frac{1}{2}\int_{0}^{1} d\alpha \int_{-\infty}^{\infty} \frac{dz}{2\pi i} \left\langle \boldsymbol{\epsilon}^{\alpha}(z)^{-1}\boldsymbol{\Pi}^{0}(z)\mathbf{V}_{\mathrm{int}} \right\rangle. \tag{3.19}$$

For similar reasons, all even orders in the geometric series expansion of $\boldsymbol{\epsilon}^{\alpha}(z)^{-1}$ with respect to $\boldsymbol{\Pi}^{0}(z)\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z)$ do not contribute to the dispersion energy. It is hence convenient to define the second-order generalized dielectric function

$$\boldsymbol{\kappa}^{\alpha}(z) = \boldsymbol{\epsilon}^{\alpha}(z)\left(\mathbf{2} - \boldsymbol{\epsilon}^{\alpha}(z)\right) = \mathbf{1} - \left(\boldsymbol{\Pi}^{0}(z)\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z)\right)^{2}, \tag{3.20}$$

which is block diagonal with exponentially vanishing off-diagonal blocks. Substituting Eq. (3.20) into Eq. (3.19) and using the vanishing trace of $\boldsymbol{\kappa}^{\alpha}(z)^{-1}\boldsymbol{\Pi}^{0}(z)\mathbf{V}_{\mathrm{int}}$, we arrive at a central

theoretical result of this paper,

$$E_{\rm int}^{\rm C}[\rho] = -\frac{1}{2}\int_0^1 d\alpha \int_{-\infty}^{\infty} \frac{dz}{2\pi i} \left\langle \boldsymbol{\kappa}^{\alpha}(z)^{-1}\boldsymbol{\Pi}^0(z)\mathbf{K}_{\rm int}^{\alpha{\rm HXC}}(z)\boldsymbol{\Pi}^0(z)\mathbf{V}_{\rm int}\right\rangle . \tag{3.21}$$

Eq. (3.21) "exactifies" the well-known Longuet-Higgins Zaremba-Kohn[161,162] expression for the second-order dispersion energy, denoted LHZK(2) in the following. Eq. (3.21) goes beyond LHZK(2) by including (i) exchange and correlation effects in the AB interaction through $\mathbf{K}_{\rm int}^{\alpha{\rm HXC}}$, and (ii) screening of the bare Coulomb interaction $\mathbf{V}_{\rm int}$ by $\boldsymbol{\kappa}^{\alpha}(i\omega)$. Indeed, the replacements $\mathbf{K}_{\rm int}^{\alpha{\rm HXC}} \to \alpha\mathbf{V}_{\rm int}$ and $\boldsymbol{\kappa}^{\alpha}(z) \to \mathbf{1}$ recover LHZK(2). Figure 3.4 displays a diagrammatic representation of Eq. (3.21).

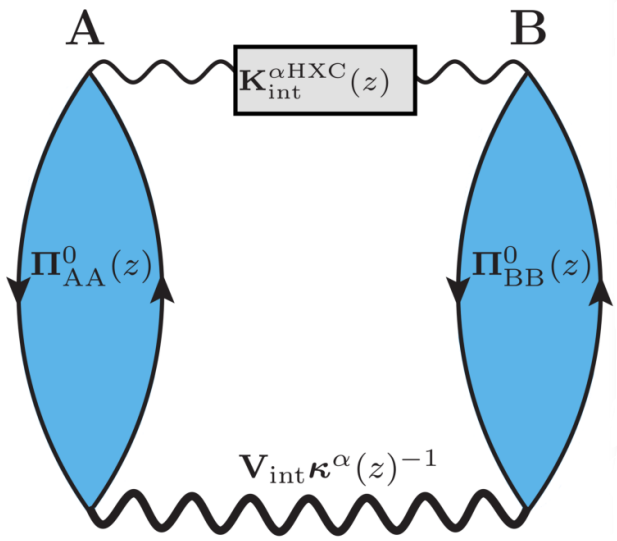


Figure 3.4: Diagrammatic representation of the A–B dispersion energy Eq. (3.21). Blue-shaded rings with upward–downward arrows denote particle–hole propagators of the monomers containing the full intra-monomer interaction, whereas horizontal wavy lines represent interactions between the monomers. Coupling strength and frequency integration are implied.

3.4.4 Convergence Radius of the AC-SAPT Series

The convergence of the AC-SAPT series, Eq. (3.6), is governed by the analyticity of the coupling strength integrand $W^\alpha[\rho]$ in the complex α plane: To guarantee convergence, the coupling strength integrand must be analytic for $|\alpha| \leq 1$. Using Eq. (3.19), this implies that $\boldsymbol{\epsilon}^\alpha(z)^{-1}$ must give rise to an analytic coupling strength integrand for $|\alpha| \leq 1$. A necessary condition for the latter is that $\boldsymbol{\epsilon}^\alpha(z)^{-1}$ is free of poles, i.e., by Eq. (3.17),

$$\left\|\boldsymbol{\Pi}^0(z)\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z)\right\|_2 = \left\|\mathbf{1} - \boldsymbol{\epsilon}^\alpha(z)\right\|_2 < 1. \tag{3.22}$$

Here, the spectral norm $\|\cdot\|_2$ equals the largest singular value of an operator; $\|\cdot\|_2$ is induced by the scalar product on the tensor product space and thus a natural choice. An upper bound for the convergence radius of the AC-SAPT series is thus

$$\alpha_c = \min_\alpha \left\{ |\alpha| : \left\|\boldsymbol{\Pi}^0(z)\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z)\right\|_2 = 1 \right\}. \tag{3.23}$$

α_c is generally an upper bound for the convergence radius, since $\boldsymbol{\epsilon}^\alpha(z)^{-1}$ could exhibit additional, non-algebraic singularities inside the complex α unit circle.

The convergence criterion (3.22) also implies that the geometric series

$$\boldsymbol{\kappa}^\alpha(z)^{-1} = \mathbf{1} + \left(\boldsymbol{\Pi}^0(z)\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z)\right)^2 + \left(\boldsymbol{\Pi}^0(z)\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z)\right)^4 + \ldots \tag{3.24}$$

converges, since, by the definition of the spectral norm and the negative definiteness of $\boldsymbol{\Pi}^0(z)$,

$$\left\|\mathbf{1} - \boldsymbol{\kappa}^\alpha(z)\right\|_2 = \left\|\left(\boldsymbol{\Pi}^0(z)\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z)\right)^2\right\|_2 = \left\|\boldsymbol{\Pi}^0(z)\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z)\right\|_2^2 < 1. \tag{3.25}$$

Necessary conditions equivalent to Eq. (3.22) are thus that $\boldsymbol{\kappa}^\alpha(z)$ be positive definite or

$\boldsymbol{\epsilon}^{\alpha}(z)^{-1}$ have eigenvalues < 2.

3.4.5 Approximations within AC-SAPT

In the following, we determine the asymptotic expressions for the dispersion energy corresponding to supermolecular RPA and MBPT calculations, and analyze the consequences for the AC-SAPT expansion.

Random Phase Approximation

The RPA polarization propagator at zero intermonomer interaction,

$$\boldsymbol{\Pi}^{\mathrm{RPA}}(z) = \left(\mathbf{1} - \boldsymbol{\Pi}_0^0(z)\mathbf{V}_0\right)^{-1} \boldsymbol{\Pi}_0^0(z), \tag{3.26}$$

is defined in terms of the bare KS polarization propagator of the supersystem $\boldsymbol{\Pi}_0^0$, which does not include any electron-electron interactions, and the intramonomer interaction $\mathbf{V}_0$ with the matrix representation

$$\mathbf{V}_0 = \begin{pmatrix} \mathbf{V}_{AA} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{BB} \end{pmatrix} \tag{3.27}$$

in the large R limit. A supermolecular RPA calculation corresponds to the replacements

$$\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z) \to \alpha\mathbf{V}_{\mathrm{int}} \quad \text{and} \quad \boldsymbol{\Pi}^0(z) \to \boldsymbol{\Pi}^{\mathrm{RPA}}(z) \tag{3.28}$$

in Eqs. (3.13)-(3.23). In other words, the intramolecular electron correlation and the screening factor $\boldsymbol{\kappa}^{\alpha}(z)^{-1}$ in Eq. (3.21) are treated non-perturbatively, whereas the intermolecular HXC kernel is replaced by its first-order approximation.

Condition (3.22) implies that, within RPA, the AC-SAPT series will converge if

$$|\alpha| \left\| \mathbf{\Pi}^{\mathrm{RPA}}(z) \mathbf{V}_{\mathrm{int}} \right\|_2 < 1. \tag{3.29}$$

This condition is necessary and sufficient for RPA, because the only singularities of the RPA coupling strength integrand in the complex α plane result from zeros of

$$\boldsymbol{\kappa}^{\alpha\mathrm{RPA}}(z) = \mathbf{1} - \alpha^2 \left(\mathbf{\Pi}^{\mathrm{RPA}}(z) \mathbf{V}_{\mathrm{int}} \right)^2. \tag{3.30}$$

The convergence radius thus has the lower bound

$$\alpha_c^{\mathrm{RPA}} = \left\| \mathbf{\Pi}^{\mathrm{RPA}}(z) \mathbf{V}_{\mathrm{int}} \right\|_2^{-1} \geq \left\| \mathbf{\Pi}^{\mathrm{RPA}}(z) \mathbf{V}_0 \right\|_2^{-1} \left\| \mathbf{V}_0^{-1} \mathbf{V}_{\mathrm{int}} \right\|_2^{-1}, \tag{3.31}$$

where the inequality follows from the submultiplicativity of the spectral norm. Since $\mathbf{\Pi}^0(z)\mathbf{V}_0$ is negative definite,

$$\left\| \mathbf{\Pi}^{\mathrm{RPA}}(z) \mathbf{V}_0 \right\|_2 \leq 1 \tag{3.32}$$

by Eq. (3.26). Moreover, $\left\| \mathbf{V}_0^{-1} \mathbf{V}_{\mathrm{int}} \right\|_2 \leq 1$ follows[163] from the fact that both $\mathbf{V}_0$ and $\mathbf{V}_0 + \mathbf{V}_{\mathrm{int}}$, corresponding to the monomer-only and supersystem Hartree kernels, are positive definite. Consequently,

$$\alpha_c^{\mathrm{RPA}} \geq 1, \tag{3.33}$$

where the inequality holds as long as $\left\| \mathbf{\Pi}_0^0(z) \mathbf{V}_0 \right\|_2 < \infty$. This condition is satisfied for monomers with finite KS gap, where $\mathbf{\Pi}_0^0(z)$ is bounded, but may be violated, e.g., for infinite one-dimensional metals, see below. Hence, the AC-SAPT series always converges within RPA for nondegenerate monomers.

RPA permits analytic integration over coupling strength in Eq. (3.21), yielding a compact expression for the dispersion energy within RPA,

$$E_{\text{int}}^{\text{C RPA}}[\rho] = \frac{1}{4}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\langle \ln \boldsymbol{\kappa}^{\text{RPA}}(z)\rangle. \tag{3.34}$$

Eq. (3.34) illustrates how the analytic structure of $\ln \boldsymbol{\kappa}^{\text{RPA}}(z)$ governs the convergence of the AC-SAPT expansion at full coupling within RPA. For nondegenerate monomers, $\boldsymbol{\kappa}^{\text{RPA}}(z)$ is positive definite with eigenvalues between 0 and 1, and thus the Taylor expansion around $\boldsymbol{\kappa}^{\text{RPA}}(z) = \mathbf{1}$, which generates the AC-SAPT series, converges. However, if $\boldsymbol{\kappa}^{\text{RPA}}(z)$ has zero eigenvalues, the series diverges due to the essential singularity of the natural logarithm at zero.

Dobson and Gould (DG) obtained Eq. (3.34) by a coupling strength integration argument without density constraint, and showed that it reduces to the non-retarded Lifshitz formula for macroscopic slab systems, which is accurate for dispersion interactions between macroscopic objects.[55] DG also identified conditions for which Eq. (3.34) predicts unconventional power laws of dispersion interactions that cannot be obtained from AC-SAPT: Systems must be macroscopic in at least one dimension, allowing for infinite-wavelength density fluctuations, finite in at least one other dimension, and exhibit zero electronic gap. This is precisely when $\boldsymbol{\kappa}^{\text{RPA}}(z)$ can have zero eigenvalues, causing AC-SAPT to diverge. The unconventional power laws observed[95] for these systems cannot be obtained from a Taylor series with respect to α and thus are examples of physical systems exhibiting divergence of AC-SAPT.

Many-Body Perturbation Theory

In the present framework, supersystem MBPT calculations correspond to perturbatively expanding the coupling strength integrand $W^{\alpha}[\rho]$ with respect to both, the inter- and intramonomer interaction, such that the resulting total interaction energy is consistent to a given

finite order. In the following, we consider supersystem MBPT(2) theory, which is by far the most commonly used MBPT approach for NIs in large molecular systems. Supersystem MBPT(2) corresponds to the replacements

$$\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z) \to \alpha\mathbf{V}_{\mathrm{int}} \quad \text{and} \quad \boldsymbol{\Pi}^0(z) \to \boldsymbol{\Pi}_0^0(z) \quad \text{and} \quad \boldsymbol{\kappa}_\alpha(z) \to \mathbf{1} \tag{3.35}$$

in Eqs. (3.13)-(3.23). These replacements are sufficient to make the coupling strength integrand correct to first order, corresponding to MBPT(2) for the (coupling strength integrated) correlation energy. In the present constant-density AC approach, this is equivalent to second-order Görling–Levy perturbation theory treatment of the supersystem; analogous considerations apply to MP2. The replacements (3.35) amount to the LHZK(2) limit. This implies that the geometric series (3.24) is truncated after the first term, which is likely to result in large errors unless the series converges very rapidly.

Within the MBPT(2) approximation to monomer correlation, the convergence radius of the AC-SAPT series is, according to Eq. (3.23),

$$\alpha_c^{\mathrm{PT2}} = \left\| \boldsymbol{\Pi}_0^0(z)\mathbf{V}_{\mathrm{int}} \right\|_2^{-1}. \tag{3.36}$$

However, unlike the RPA propagator, $\boldsymbol{\Pi}_0^0\mathbf{V}_{\mathrm{int}}$ is generally not bounded by 1, which may cause unphysical divergence of the series.

Since exchange effects vanish exponentially for large R, MBPT(2) coincides with the second-order perturbative limit of RPA, enabling us to alternatively consider the behavior of the coupling-strength integrated AC-SAPT expansion within MBPT(2) via Eqs. (3.30) and (3.34). Clearly, this argument can only be used as long as RPA itself is reasonably accurate, a conclusion supported by our results. MBPT(2) corresponds to replacing $\boldsymbol{\kappa}^{\mathrm{RPA}}(z)$, the

generalized second-order RPA dielectric function at full intra-monomer coupling with

$$\boldsymbol{\kappa}^{\mathrm{PT2}}(z) = \mathbf{1} - \left(\boldsymbol{\Pi}_0^0(z)\mathbf{V}\right)^2, \tag{3.37}$$

and truncating the Taylor expansion of $\ln \boldsymbol{\kappa}^{\mathrm{PT2}}(z)$ after the first order. However, since $\boldsymbol{\Pi}_0^0(z)$ is unbounded, $\boldsymbol{\kappa}^{\mathrm{PT2}}(z)$ may exhibit eigenvalues ≤ 0 even for finite, nondegenerate monomers, corresponding to nonanalytic behavior of $\ln \boldsymbol{\kappa}^{\mathrm{PT2}}(z)$. In this scenario, the Taylor expansion of the natural logarithm around $\boldsymbol{\kappa}^{\mathrm{PT2}}(z) = \mathbf{1}$ spuriously diverges, and the first-order approximation may be expected to yield a poor approximation to the RPA dispersion energy.

MP2C: Partial Resummation of MBPT

Heßelmann's MP2C method[23,24,90] replaces the (uncoupled) LHZK(2) part of the MP2 interaction energy with its (coupled) time-dependent DFT counterpart. In the large-R asymptotic limit, time-dependent DFT reduces to RPA, and hence MP2C corresponds to the replacements

$$\mathbf{K}_{\mathrm{int}}^{\alpha\mathrm{HXC}}(z) \to \alpha\mathbf{V}_{\mathrm{int}} \quad \text{and} \quad \boldsymbol{\Pi}^0(z) \to \boldsymbol{\Pi}^{\mathrm{RPA}}(z) \quad \text{and} \quad \boldsymbol{\kappa}_\alpha(z) \to \mathbf{1} \tag{3.38}$$

in Eqs. (3.13)-(3.23). Equivalently, MP2C may be understood as a low-order approximation to the RPA dispersion energy resulting from first-order truncation of the Taylor expansion of $\ln \boldsymbol{\kappa}^{\mathrm{RPA}}(z)$ in Eq. (3.34) around $\boldsymbol{\kappa}^{\mathrm{RPA}}(z) = \mathbf{1}$,

$$E_{\mathrm{int}}^{\mathrm{C\ PT2C}}[\rho] = \frac{1}{4}\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \langle \boldsymbol{\kappa}^{\mathrm{RPA}}(z) - \mathbf{1} \rangle. \tag{3.39}$$

In other words, MP2C includes intramonomer screening effects to infinite order, but the intermonomer interaction is second order only. The convergence of this partially resummed

MBPT series is more benign compared to standard MBPT, because, for non-degenerate monomers, the eigenvalues of $\boldsymbol{\kappa}^{\mathrm{RPA}}(z)$ are between 0 and 1, and hence the Taylor expansion of $\ln \boldsymbol{\kappa}^{\mathrm{RPA}}(z)$ converges. Also, since $\ln x \leq x - 1$ for $0 < x \leq 1$, the MP2C dispersion energy is an upper bound for the RPA dispersion energy. However, with decreasing eigenvalues of $\boldsymbol{\kappa}^{\mathrm{RPA}}(z)$, i.e., for large and polarizable monomers, this Taylor series converges increasingly slowly (and eventually diverges under DG conditions), and hence this bound deteriorates rapidly. This is consistent with the observation that MP2C underestimates binding energies of larger complexes contained in the L7 benchmark.[110] Unlike the previously discussed RPA and MBPT approximations, the MP2C method does not possess a "seamless" supermolecular equivalent, i.e., it requires an SAPT-style partitioning into monomers, because the inter- and intramonomer interactions are treated at different levels. From a computational viewpoint, truncation of the Taylor expansion of $\ln \boldsymbol{\kappa}^{\mathrm{RPA}}(z)$ offers little advantage compared to full RPA using Eq. (3.34).

Fig. Figure 3.5 summarizes the approximations to the AC-SAPT dispersion energy discussed in this section in diagrammatic form.

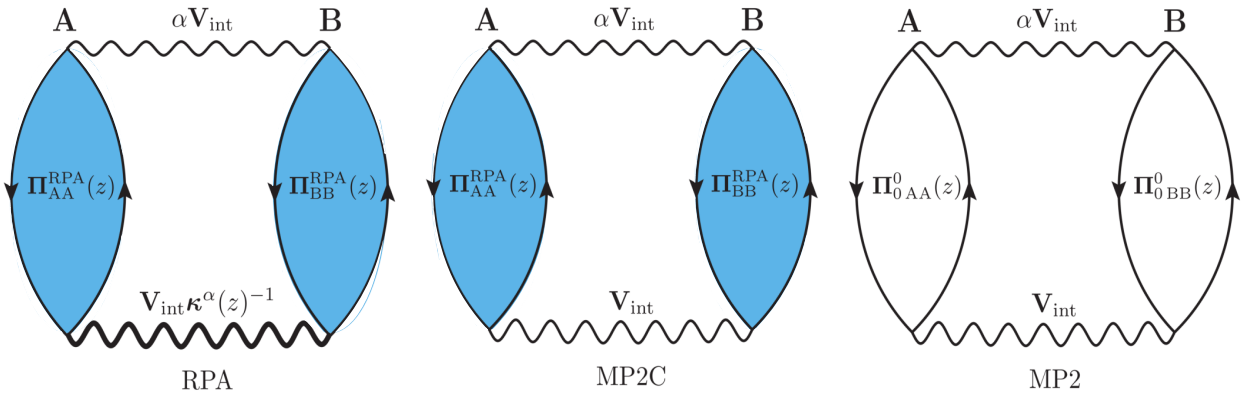


Figure 3.5: Diagrammatic representations of the A–B dispersion energy within RPA, MP2C, and MP2. Blue-shaded and empty rings with upward–downward arrows denote particle–hole propagators of the monomers containing the full and zero intra-monomer interaction, respectively, whereas horizontal wavy lines represent interactions between the monomers. Coupling strength and frequency integration are implied.

3.4.6 Numerical Validation

Convergence Estimates

The previous sections suggest that the AC-SAPT convergence radius α_c depends critically on the level of theory used to describe the monomers (through $\mathbf{\Pi}_0$). Here we numerically evaluate the upper bounds for α_c^{PT2} and investigate whether these asymptotic bounds can serve as meaningful convergence estimates for large but finite R.

Starting from Eq. (3.36) and using the same inequalities as in Sec. 3.4.5, we obtain

$$\alpha_c^{\mathrm{PT2}}(z) \geq \left\| \mathbf{\Pi}_0^0(z) \mathbf{V}_0 \right\|_2^{-1}. \tag{3.40}$$

Since the spectral norm is invariant under similarity transformations,

$$\left\| \mathbf{\Pi}_0^0(z) \mathbf{V}_0 \right\|_2 = \|\mathbf{Q}(z)\|_2, \tag{3.41}$$

where $\mathbf{Q}(z) = -\mathbf{L}^T \mathbf{\Pi}^0(z) \mathbf{L}$ and $\mathbf{L}$ is the Cholesky factor of $\mathbf{V}^0$. $\mathbf{Q}(z)$ is routinely computed in efficient RPA and beyond-RPA implementations.[107,108,164] For purely imaginary z, $\|\mathbf{Q}(z)\|_2$ has a maximum at $z = 0$ and is otherwise monotonous, as may be demonstrated, e.g., using the spectral representation of $\mathbf{\Pi}^0(z)$. Thus,

$$\alpha_c^{\mathrm{PT2}}(z) \geq \underline{\alpha}_c^{\mathrm{PT2}} = \|\mathbf{Q}(0)\|_2^{-1}. \tag{3.42}$$

Figure 3.6 shows the correlation between absolute errors in MP2 interaction energies and the inverse convergence radius $1/\underline{\alpha}_c^{\mathrm{PT2}}$, evaluated using a PBE KS reference. As long as $1/\underline{\alpha}_c^{\mathrm{PT2}}$ is close to 1, the MP2 errors are small, but they increase rapidly once $1/\underline{\alpha}_c^{\mathrm{PT2}}$ increases above ~ 3. The correlation is quite convincing since (i) divergence at α values close to 1 may not

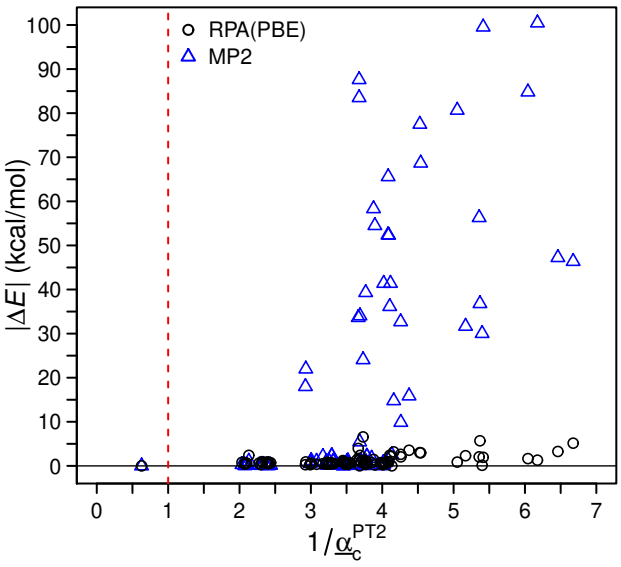


Figure 3.6: Absolute MP2 and RPA interaction energy errors $|\Delta E|$ for the S66,[13,109] L7,[110] and S30L[111] benchmarks as well as helium dimer as functions of the inverse convergence radius $1/\underline{\alpha}_c^{\mathrm{PT2}}$. $\underline{\alpha}_c^{\mathrm{PT2}}$ was evaluated using a PBE KS reference and cc-pVTZ basis sets.

translate into large MP2 errors, (ii) a KS reference yields $1/\underline{\alpha}_c^{\mathrm{PT2}}$ values 2-3 times larger than a HF reference, see Fig Figure 3.6. With decreasing convergence radius, MP2 and the LHZK(2) dispersion energy become increasingly incorrect estimates of the exact dispersion energy (3.8). Figure 3.6 also shows that the RPA interaction energies are uncorrelated with $\underline{\alpha}_c^{\mathrm{PT2}}$, as expected from the estimate (3.33).

The convergence behavior of the AC-SAPT expansion in relation to $\underline{\alpha}_c^{\mathrm{PT2}}$ is further illustrated by re-expansion of the RPA interaction energy in powers of α, see Figure 3.7. Only for helium dimer ($\underline{\alpha}_c^{\mathrm{PT2}} = 1.59$ with a PBE reference), the series converges with respect to the spectral norm. For all other cases, increasingly large oscillations are observed at higher orders. This behavior is characteristic of asymptotic series and has been observed for MBPT ground state energies.[75,165,166] With decreasing $\underline{\alpha}_c^{\mathrm{PT2}}$, the oscillations become more pronounced lower orders, causing significant error even at $n = 2$. While the convergence radii for some of the smaller systems are closer to 1 with a HF reference compared to a PBE reference, their AC-SAPT series eventually diverge as well.

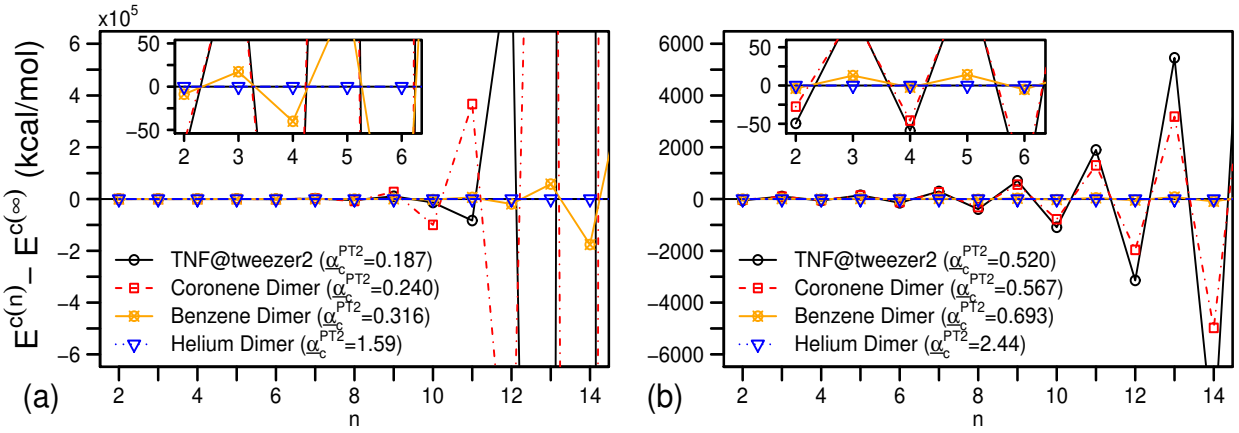


Figure 3.7: Errors (kcal/mol) of MBPT(n) interaction energies within the ring approximation (i.e. RPA) using (a) a KS and (b) a HF reference and cc-pVTZ basis sets. Insets show lower orders on a smaller energy scale.

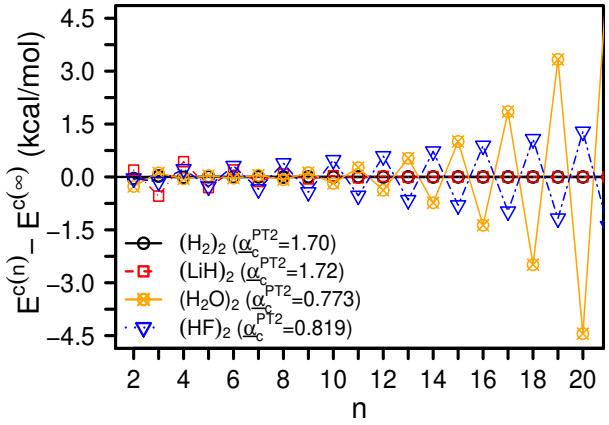


Figure 3.8: Errors (kcal/mol) of MBPT(n) interaction energies within the ring approximation (i.e. RPA) using a HF reference and cc-pVTZ basis sets for the dimers taken from Ref. 14.

In their 1993 work,[14] Moszynski, Jeziorski, and Szalewicz considered the convergence of MBPT supermolecular dispersion energies using a HF reference for He_2, $(H_2)_2$, $(LiH)_2$, $(H_2O)_2$, and $(HF)_2$ within the ring approximation, which is equivalent to RPA. Based on numerical results up to order $n = 10$, they concluded that the convergence of the MBPT expansion for these systems is "very fast." While we confirm this conclusion for He_2, the present results suggest that the MBPT expansion of the dispersion energy indeed diverges for modestly larger systems. For example, our results for $(H_2O)_2$ agree well with those of Moszynski, Jeziorski, and Szalewicz up to $n = 10$, but the $\underline{\alpha}_c^{\mathrm{PT2}}$ value of 0.773 suggests that

the series is divergent. Indeed, oscillations of increasing magnitude are observed when orders up to $n = 20$ are considered, see Figure 3.8.

Size Dependence of Errors

The dependence of the convergence radius estimate $\underline{\alpha}_c^{\mathrm{PT2}}$ on the size of the complex is displayed in Figure 3.9. Only for the helium dimer, $\underline{\alpha}_c^{\mathrm{PT2}} > 1$, whereas $\underline{\alpha}_c^{\mathrm{PT2}}$ is significantly smaller than 1 for all systems in the S66,[13,109] L7,[110] and S30L[111] benchmarks. Clearly, as opposed to $\mathbf{\Pi}_0^{\mathrm{RPA}}(0)$, $\mathbf{\Pi}^0(0)$ is not necessarily bounded. Whether $\|\mathbf{\Pi}^0(0)\mathbf{V}_{\mathrm{int}}\|_2$ saturates or becomes infinite in the thermodynamic limit is system-dependent; nevertheless, the convergence criterion Eq. (3.22) suggests that perturbative calculations of NIs start to diverge already for fairly small system sizes with few tens of atoms and comparatively large HOMO–LUMO gap. Indeed, the HOMO–LUMO gap is a fairly poor estimator of the interaction energy error, see Supporting Information.

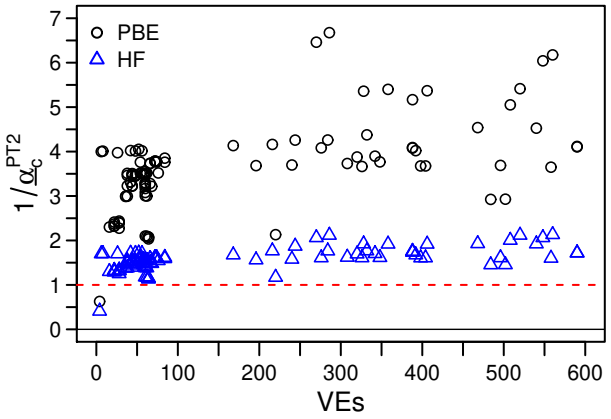


Figure 3.9: Inverse of the convergence radius ($1/\underline{\alpha}_c^{\mathrm{PT2}}$) for MBPT(2) with PBE and HF references vs. number of valence electrons (VEs) for the S66,[13,109] L7,[110] and S30L[111] benchmarks, and helium dimer. $\underline{\alpha}_c^{\mathrm{PT2}}$ values were computed using cc-pVTZ basis sets. A $1/\underline{\alpha}_c^{\mathrm{PT2}}$ value ≥ 1 indicates divergence of the AC-SAPT series.

Relative errors in NIs as a function of system size are shown in Figure 3.10. While the correlation of the errors with the number of valence electrons (VEs) is less strong than the one observed for the convergence estimates, there are clearly discernible trends: Whereas

percentage errors in binding energies are virtually constant within RPA, they increase linearly for MP2, at a rate of approximately 0.1%, per VE on average. For slightly over 700 valence electrons, the MP2 relative error regression fit reaches 100% for the systems tested here. SCS-MP2 has an approximately 5 times lower slope of 0.025%, per VE (but notably higher y-intercept), and PBE-D3 relative errors grow at a rate of slightly less than 0.01%, for the present benchmarks, see Table Table 3.2. The largest MP2 errors occur for systems with strong π–π stacking interactions such as complexes 3 to 12 from the S30L test set.[111]

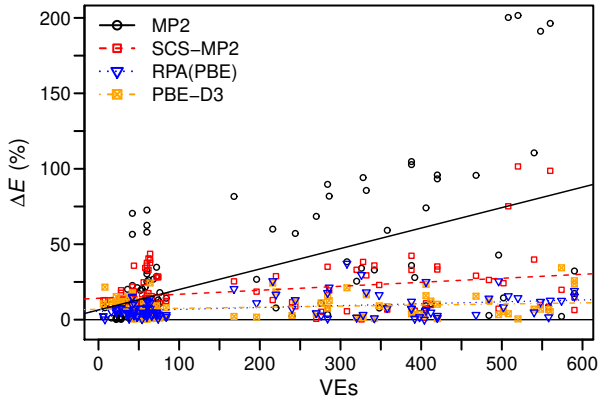


Figure 3.10: Relative errors (ΔE) of MP2, SCS-MP2, RPA(PBE), and PBE-D3 interaction energies in the S66,[13,109] L7,[110] and S30L[111] benchmarks vs. number of valence electrons (VEs).

Table 3.2: Parameters of the linear regression fits displayed in Figure 3.10. The slope corresponds to the average relative interaction energy error (%) per valence electron (VE), and the y-intercept corresponds to the average relative interaction energy error (%) in the limit of zero VEs.

Method	Slope (%/VE)	y-intercept (%)
MP2	0.1219	7.30
SCS-MP2	0.0251	14.10
RPA(PBE)	0.0030	6.24
PBE-D3	0.0083	6.79

3.4.7 Physical Interpretation

The present analytical and numerical results show that the convergence of the AC-SAPT series for NIs is strongly dependent on the level of theory used for computing the monomer polarization propagator $\mathbf{\Pi}^0(z)$. Thus, it is helpful to consider the generalized dielectric function

$$\boldsymbol{\epsilon}_0(z) = \mathbf{1} - \mathbf{\Pi}^0(z)\mathbf{V}_0, \tag{3.43}$$

which characterizes the response of the monomers at frequency z. Within RPA, Eq. (3.32) implies that the eigenvalues of $\boldsymbol{\epsilon}_0$ are bounded from above by 2, reflecting the fact that the RPA response to external fields is reduced ("screened") by the creation of induced electron-pairs. As a result, the effective interaction "seen" by an electron of subsystem A due to electrons in subsystem B decays rapidly within B. This effective interaction is indeed "weak" in the sense that it affords a convergent AC-SAPT expansion for nondegenerate monomers.

For systems satisfying the DG conditions, there is no screening in the monomers in at least one dimension, and the largest eigenvalue of $\boldsymbol{\epsilon}_0(z)$ may equal 2, even within RPA, causing divergence of the AC-SAPT expansion and unconventional power laws of the dispersion interaction. Since the HXC kernel is dominated by the Hartree kernel for large R, this behavior of the RPA is expected to be correct, at least qualitatively.

As opposed to these physical divergences of the AC-SAPT expansion, unphysical divergences may result if the intra-monomer electron interaction is treated perturbatively. In this case, $\boldsymbol{\epsilon}_0$ is not bounded, with largest eigenvalues around 5 (KS reference) or 2 (HF reference) for typical systems studied here, see Figure 3.6 and Figure 3.9. This reflects a much stronger perturbation of the monomers by the intersystem interaction due to incomplete screening resulting primarily from the neglect of higher-order particle–hole ring diagrams. Hence, the MBPT effective interaction is too strong for intermolecular perturbation theory, causing

spurious divergence of the AC-SAPT expansion. This effect is not seen in the smallest and least polarizable monomers such as He atoms, but it becomes noticeable for even moderately large monomers with a few atoms, where the neglect of screening due to multiple induced particle–hole pairs by finite-order MBPT produces significant over- or underestimations of NIs. For the large π systems in the S30L, the lack of screening produces maximum eigenvalues of $\boldsymbol{\epsilon}_0$ around 7 (KS reference) or 3 (HF reference), which suggests rapid divergence of the AC-SAPT expansion, providing a plausible rationale for the spectacular errors in MP2 binding energies observed for these systems.

3.5 Conclusions

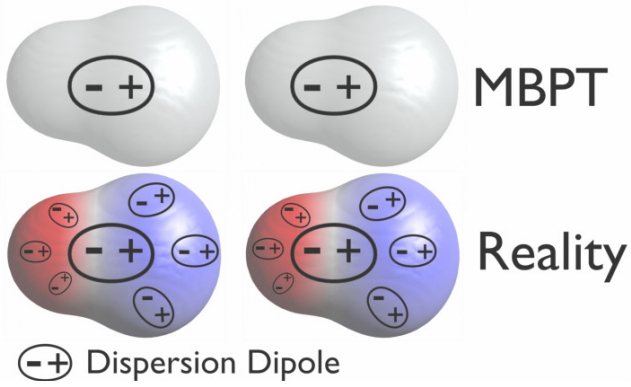


Figure 3.11: An illustration of the electrodynamic screening present between dimers.

A key result of this study is that dispersion interactions cannot be considered "weak" unless intra-monomer screening effects are taken into account at least at the level of RPA. Similar to electrostatic screening, electrodynamic screening due to induced density fluctuations is inadequately captured by finite-order MBPT-type approaches, except for the smallest and least polarizable systems, see Figure 3.11. As a result, unscreened perturbation theories produce exaggerated responses of the monomers to external perturbation and divergent estimates for NIs even in moderately large systems. The conventional wisdom that MBPT is useful for accurate calculations of NIs in even moderately large systems is incorrect: Numer-

ically small interaction energies compared to covalent interactions do not imply "weakness" in the sense of a bounded response or convergent intermolecular perturbation theory. Consequently, finite-order MBPT results tend to worsen systematically with system size, as was demonstrated by the basis-set extrapolated MP2 results for complexes with up to 600 VEs. Given the computational efficiency and popularity of MBPT implementations, this is a sobering result: Size-extensivity and (conventional) size-consistency are insufficient conditions for accurate predictions of NIs. While there may be a place for MBPT calculations of NIs in complexes of small, hard monomers, MBPT estimates of NIs cannot be considered reliable for most systems of chemical interest, much less for nanomaterials, metallic systems, or soft matter. In light of the current results, empirically scaled MP2 methods, and particularly MP2.5, appear as simple regularizations of an asymptotic series; while this strategy is clearly successful for some systems, it does not address the underlying physical problem, limiting predictive power and robustness.

The MP2C method can be viewed as a partial MBPT resummation which includes intra-monomer screening and exhibits better analytical properties than bare MBPT, but truncates intermonomer interactions at second order, and has no obvious supermolecular equivalent. For increasingly large and polarizable monomers, MP2C underestimates the magnitude of dispersion interactions progressively due to missing many-body dispersion.

A qualitatively correct treatment of intra-monomer screening to all orders is possible with RPA at little extra cost compared to MBPT approaches. Indeed, RPA produces constant relative errors in NIs that are virtually independent of the system size and type, consistent with its qualitatively correct treatment of electrodynamic polarization; this may be viewed as a manifestation of "Casimir-Polder size consistency."[21] Similar conclusions may hold for more elaborate non-perturbative coupled cluster methods, which include the ring diagrams corresponding to RPA.[167] In particular, the accuracy of ring-coupled-cluster methods for NIs[168] supports the view that ring diagrams dominate in the long-range limit of NIs. It is

remarkable that ring diagrams also constitute the main part of the correlation energy for the uniform electron gas at high density,[100] a similarity first noted by Dzyaloshinskii, Lifshitz, and Pitaevskii.[169] Both, dispersion interactions in finite systems and electron correlation in the high-density electron gas emerge from collective density fluctuations[170,171] caused by the long-range Coulomb interaction. In this sense, the expectation that RPA energy differences should be universally accurate[172,173] appears to hold as long as long-range correlation dominates.

The present results raise the question whether the perturbative triples correction of CCSD(T) inherits any of the limitations of MBPT, even though it includes some screening at the level of the amplitudes. Whereas CCSD(T) errors for interaction energies of small molecular complexes were determined to be on the order of 1%,[110] and typical deviations between RPA and CCSD(T) binding energies are on the order of 5-10% for the benchmarks studied here, the linear scaling domain-based pair natural orbital CCSD(T) binding energies for water on small graphene flakes are significantly larger than the corresponding RPA and diffusion Monte Carlo ones,[174] and the CCSD(T) perturbative triples correction diverges for the correlation energy of the uniform electron gas.[175]

The AC-SAPT formalism developed here affords separate, non-perturbative definitions of dispersion and induction effects in NIs. While it has mainly been used to rationalize the results of supermolecular calculations here, the analytical expressions obtained from AC-SAPT could be evaluated using monomer calculations given suitable approximations to the intermolecular KS potential $\hat{V}_{s\,\mathrm{int}}[\rho]$. Beyond-RPA perturbation methods known from supermolecular calculations such as second-order screened exchange (SOSEX)[106,176–178] or approximate exchange kernel (AXK)[164,179] may prove useful for this purpose. Our conclusions regarding the divergence of MBPT for NIs rely in part on the assumption that the RPA dispersion energy is qualitatively accurate, which is supported by the close agreement between the RPA and the benchmark results. This agreement is remarkable given that RPA

is parameter-free aside from using a KS reference from a semilocal DFA. Moreover, the estimated AC-SAPT convergence radii are upper bounds only, but they show strong correlation to the NI errors of MBPT methods for the benchmarks studied here.

Our results show that the convergence of the AC-SAPT expansion for dimers consisting of large monomers depends critically on the accuracy of intra-monomer correlation. Specifically, electrodynamic screening should be included in the monomer correlation treatment at least at the time-dependent HF level. Apart from unphysical divergences, AC-SAPT does converge more slowly for large and polarizable monomers, and exhibits physical divergence for systems satisfying the DG conditions such as one-dimensional metals. Under these circumstances, the traditional LHZK(2) picture of dispersion breaks down and must replaced by the RPA one embodied in Eq. (3.34). Importantly, Eq. (3.34) correctly recovers both, the small, hard monomer LHZK(2) limit where MBPT can converge, and the macroscopic Lifshitz limit. Eq. (3.34) is largely scale invariant and therefore a far better starting point for computing and conceptualizing dispersion interactions than LHZK(2).

The breakdown of MBPT for NIs also has important implications for the development of approximations such as van-der-Waals density functionals or force fields. The present results cast further doubt on the validity of empirical $1/R^6$ corrections for very large and polarizable monomers – even though some dispersion-corrected DFAs admittedly perform remarkably well for large systems. An accurate description of NIs for such systems may require methods including Lifshitz-type physics and electrodynamic polarization effects. Quantum Drude models[180] or many-body dispersion methods[94,181] may be considered coarse-grained RPA approaches suitable for this purpose.

Modern RPA implementations are insignificantly more expensive than the most advanced MP2 approaches from a computational viewpoint, and RPA calculations for molecules with hundreds of atoms on workstation clusters are now routine[107,140,182–188] — although the present results also show that accurate RPA binding energies for NIs require triple- to

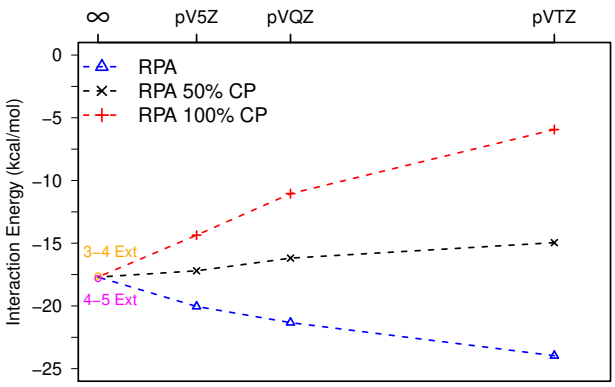
quadruple-ζ basis set extrapolation, making some of the proposed low-scaling methods less effective. Taken together with the superior accuracy of RPA for large and polarizable systems without empirical adjustments, MP2 can be safely and efficiently replaced by RPA for calculations of NIs in most systems of chemical interest. If MP2 results are nevertheless desired, diagnostic $\underline{\alpha}_c^{\mathrm{PT2}}$ values should be used to gauge their reliability. Similarly, the present results support the use of RPA calculations to calibrate dispersion-corrected DFA results.

RPA calculations of NIs benefit from variational optimization of the reference,[189] but the improvement appears to be most pronounced for small systems such as rare gas dimers and diminish with increasing monomer size. With average interaction energy errors consistently in the 5–10% range, RPA is accurate enough for a wide range of applications, irrespective of system size, gap size, or empirical training sets. The accuracy of RPA for NIs may also contribute to its recent successful application to activation energies of sterically crowded transition states.[65,190]

3.6 Supporting Information

3.6.1 RPA Basis Set Convergence Study

To investigate the effect of the basis set incompleteness on RPA interaction energies, a basis set convergence study for the correlation contribution to the interaction energy was carried out for the pincer complex with 2,4,7-trinitro-9-fluorenone as the guest from S30L (compound number 5),[111] see Figure S3.1. The 50% counterpoise corrected 3-4 and 4-5 extrapolated results agreed within 0.20 kcal/mol, validating the use of 3-4 extrapolated RPA correlation energies with 50% counterpoise correction.



.

Figure S3.1: Basis set convergence of the RPA interaction energy using a PBE KS reference for compound number 5 of S30L[111] is reported. The energy expectation value of the KS determinant was computed using def2-QZVP[131] basis sets. X denotes the cardinal number of Dunning's correlation consistent polarized cc-pVXZ (X=T,Q,5) basis sets.[132,133] All RPA correlation energies were computed within the frozen core approximation.

3.6.2 S30L-CI: Charged Species with Counterions

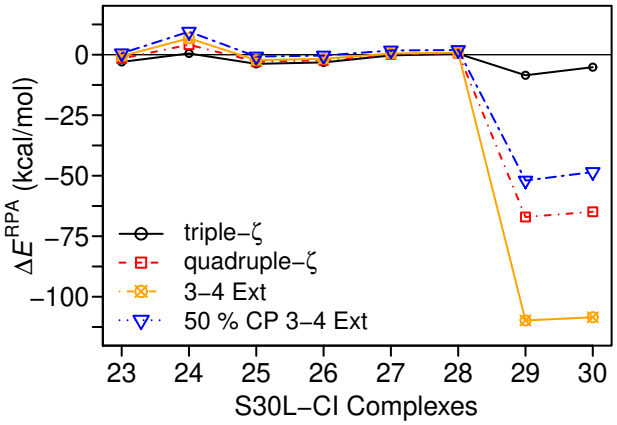


Figure S3.2: RPA(PBE) interaction energy errors (ΔE^{RPA}) for S30L complexes with counterions.[111]

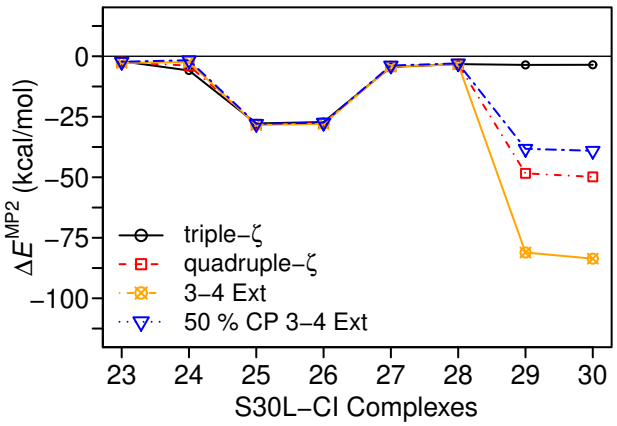


Figure S3.3: MP2 interaction energy errors (ΔE^{MP2}) for S30L complexes with counterions.[111]

Within the S30L testset, counterions were investigated for the MP2 and RPA calculations. These complexes were denoted the S30L-CI, see Figure S3.2 and Figure S3.3. The poor basis set convergence observed for complexes 29 and 30 is due to large BSSE, which is ascribed to the fact that the sodium basis sets are optimized for neutral atoms. Due to these basis set artifacts, counterions were excluded from the benchmark calculations.

3.6.3 HOMO-LUMO Gap Effect on Interaction Energies

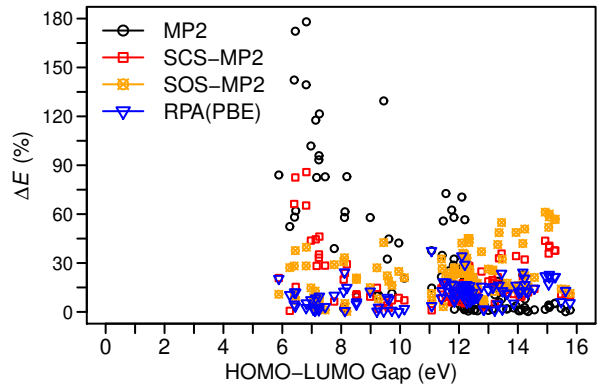


Figure S3.4: Percentage errors of interaction energy (ΔE) for the S66,[13,109] L7,[110] and S30L[111] benchmark vs. the HF HOMO-LUMO gap.

A potential source of binding energy error is the small HOMO-LUMO gap. For the complexes in this study, the HOMO-LUMO gaps were observed to be at least 6 eV and binding errors do not strongly correlate with the HOMO-LUMO gaps, see Figure S3.4.

Chapter 4

Dispersion Size-Consistency

Contribution Statement: The following are my contributions to the project. I contributed toward developing the multivariate adiabatic connection symmetry adapted perturbation theory (MAC-SAPT) and established the exact constraint dispersion size-consistency. Next, I analyzed and determined whether the many-body perturbation theory, random phase approximation, and SAPT satisfy the dispersion size-consistency. Lastly, I contributed to the writing and prepared the submission into the *Electron. Struct.*

4.1 Introduction

Noncovalent interactions (NIs) play an important role determining the physical and chemical properties of supramolecular chemistry,[8,191] molecular crystals,[61,62,192] and materials.[2,3,193]

Perturbation theory has a long history of applications to NIs, and may appear to be a logical choice based on the relative "weakness" of NIs compared to covalent bonding, and successful applications to relatively small molecular systems[78,115] seemed to support this notion. In particular, Møller–Plesset (MP)[68] many-body perturbation theory (MBPT) has been popular to use due to its relative inexpensive cost and size extensivity.[15]

However, during the last 1-2 decades, a number of theoretical[55,194] and computational[18,19,87] results have cast doubt on the suitability of perturbative approaches for NIs. Recently, Nguyen and co-workers[195] showed that, contrary to previous assumptions, supermolecular MBPT diverges for the vast majority of weakly bound dimers. Moreover, relative errors of supermolecular MP2 binding energies were found to increase with system size at a rate of approximately 0.1% per valence electron, leading to relative errors ¿100% for dimers containing a few hundred atoms.

Here, we extend the dimer adiabatic connection (AC) approach developed in reference[195] to supermolecular systems consisting of N nonoverlapping monomers. To this end, we introduce a multivariate adiabatic connection (MAC) which enables "turning on" electronic interactions between individual monomers. As in the dimer case, the MAC provides a complete separation of dispersion from induction and electrostatic interactions. This leads to a concise definition of dispersion size-consistency, a term recently proposed by Ángyán, Dobson, and Gould[196] as an "infinite-order" generalization of Casimir-Polder size-consistency;[197] the idea of dispersion size-consistency has been implicit in earlier work for some time[14,21–24] but does not appear to have been explicitly defined or shown to be an exact constraint.

In view of its fundamental importance for electronic structure theory, it is surprising that several different definitions of size-consistency are being used in the literature, some of them conflicting with each other.[15,66,67,198] In this work, we consider a method to be size-extensive if and only if it yields an energy per (electrostatically neutral) monomer that has a well-defined

thermodynamic limit, i.e., the energy asymptotically behaves as

$$E(N) = e_m N + O(N^{2/3}) \tag{4.1}$$

for large N with finite energy per monomer e_m. On the other hand, a method will be called size-consistent if and only if the error of $E(N)/N$ is approximately independent of N. This relatively strict definition of size-consistency implies size-extensivity, but it also guarantees that relative errors in energies of finite systems be approximately constant, which is critical for applications to molecular systems.

Table 4.1: Key symbols and terms

Symbol	Definition
N	Number of monomers
$I, J, \ldots$	Fragments or monomers
$\boldsymbol{\alpha}$	Intermonomer coupling constant
$\hat{H}^{\alpha}$	Born–Oppenheimer Hamiltonian at coupling strength α
$\hat{T}$	Kinetic energy
ρ	Ground-state density
$\hat{V}_s^{\alpha}[\rho]$	Local one-electron potential
$\hat{\mathbf{V}}^{ee}$	Electron–electron Coulomb interaction
$\hat{V}^{\text{ext}}$	External potential
$\hat{V}^{ne}$	External nucleus–electron potential
V^{nn}	Nucleus–nucleus potential
$\hat{V}^{\boldsymbol{\alpha}\text{HXC}}[\rho]$	Hartree-, exchange-, and correlation potential
$\mathbf{J}$	Matrix of ones
$\lvert\Psi_m^{\boldsymbol{\alpha}}\rangle$	Supersystem eigenstate
$E_m^{\boldsymbol{\alpha}}$	Supersystem energy eigenvalue
$\mathbf{W}^{\boldsymbol{\alpha}}$	Coupling strength integrand
E^{HXC}	Ground-state Hartree-, exchange-, and correlation energy
$C_{\mathbf{0}}^{\mathbf{J}}$	Interaction path
$E_{\text{int}}^{\text{C}}$	Dispersion energy
$\Delta\hat{\rho}$	Density fluctuation operator
$\boldsymbol{\Pi}^{\boldsymbol{\alpha}}(z)$	Interacting polarization propagator
$\gamma(x_1, x_2)$	one-particle density matrix
$\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}(z)$	Generalized dielectric function
$\mathbf{K}^{\alpha\text{HXC}}(z)$	Hartree-, exchange-, and correlation kernel
$\mathbf{V}^{\alpha}$	$\boldsymbol{\alpha}$-scaled bare electron–electron Coulomb interaction

4.2 Multivariate Adiabatic Connection Approach to Dispersion

4.2.1 Statement of the Problem

We consider a molecular supersystem consisting of N non–overlapping, electrostatically neutral subsystems (also called fragments or monomers) at large but finite interfragment separations R_{IJ}, each of which is assumed to be in their non-degenerate ground state at infinite separation. All distances R_{IJ} are assumed to be so large that exchange effects are exponentially small such that the fragments are distinguishable. We will use capital indices $I, J, \ldots$ to label fragments.

Since the non-relativistic many-electron Hamiltonian only contains at most pairwise interactions and there are $N(N-1)/2$ pairs of monomers, it is expedient to introduce real coupling constants $0 \leq \alpha_{IJ} \leq 1$ which scales the strength of the electron-electron interaction between each pair. In the present framework, it is straightforward to include intramonomer interactions as non-zero diagonal elements α_{II}, whereas previous work on dimers used a scalar coupling constant for the intermonomer interaction only.[195] The matrix $\boldsymbol{\alpha}$ is symmetric due to the symmetry of the electron-electron interaction; we will refer to $\boldsymbol{\alpha}$ as "coupling strength" for brevity in the following, even though $\boldsymbol{\alpha}$ contains $N(N+1)/2$ independent elements. The Born–Oppenheimer Hamiltonian of the supersystem at coupling strength $\boldsymbol{\alpha}$ is

$$\hat{H}^{\boldsymbol{\alpha}} = \sum_{I=1}^{N} \left(\hat{T}_I + \hat{V}_{sI}^{\boldsymbol{\alpha}}[\rho] \right) + \frac{1}{2} \sum_{I,J=1}^{N} \hat{\mathbf{V}}_{IJ}^{\boldsymbol{\alpha} ee}, \tag{4.2}$$

where $\hat{T}_I$ denotes the kinetic energy operator for electrons belonging to fragment I, $\hat{V}_{sI}^{\boldsymbol{\alpha}}[\rho]$ is a local coupling-strength dependent potential for electrons belonging to fragment I, and $\hat{V}_{IJ}^{\boldsymbol{\alpha} ee} = \alpha_{IJ} \hat{V}_{IJ}$ is the operator of the $\boldsymbol{\alpha}$-scaled electron–electron Coulomb interaction for

electrons on fragments I and J. In analogy to the scalar version of the AC,[98,116,117] the potentials $\hat{V}_{sI}^{\boldsymbol{\alpha}}[\rho]$ constrain the ground-state density of the supersystem at coupling strength $\boldsymbol{\alpha}$, $\rho^{\boldsymbol{\alpha}}$, to the physical ground-state density at full coupling,

$$\rho(x) = \rho^{\mathbf{J}}(x), \tag{4.3}$$

where $\mathbf{J}$ denotes the $N \times N$ matrix of ones and $x = (\mathbf{r}, \sigma)$ stands for space-spin coordinates. In the following, full coupling ($\boldsymbol{\alpha} = \mathbf{J}$) will be implied for all coupling-strength dependent quantities unless explicitly labeled. $\sum_I \hat{V}_{sI}^{\boldsymbol{\alpha}}$ is a unique functional of ρ by the Hohenberg-Kohn theorem.[148] Because the total density is uniquely decomposable into N fragment densities ρ_I at large separation, the partitioning of the sum into the fragment parts $\hat{V}_{sI}^{\boldsymbol{\alpha}}[\rho]$ is also unique. As in the scalar AC, each fragment potential

$$\hat{V}_{sI}^{\boldsymbol{\alpha}}[\rho] = \hat{V}_I^{\mathrm{ext}} + \hat{V}_I^{\mathrm{HXC}}[\rho] - \hat{V}_I^{\boldsymbol{\alpha}\mathrm{HXC}}[\rho] \tag{4.4}$$

may be decomposed into an "external" part $\hat{V}_I^{\mathrm{ext}}$ and a coupling-strength dependent Hartree-, exchange-, and correlation (HXC) part. In the present context, the external potential contains the Coulomb attraction between the electrons belonging to fragment I and the nuclei at all J, $\hat{V}_{IJ}^{ne}$, as well as the corresponding part of the nucleus-nucleus repulsion energy V^{nn},

$$\hat{V}_I^{\mathrm{ext}} = \sum_{J=1}^{N} \left(\hat{V}_{IJ}^{ne} + V_{IJ}^{nn} \right). \tag{4.5}$$

The HXC part of $\hat{V}_{sI}^{\boldsymbol{\alpha}}[\rho]$ vanishes at full coupling, while $\hat{V}_I^{\boldsymbol{\alpha}\mathrm{HXC}}[\rho]$ vanishes at $\boldsymbol{\alpha} = \mathbf{0}$, recovering the Kohn–Sham (KS) potential[199] for electrons belonging to fragment I. In the following, the functional dependence of the Hamiltonian and all derived quantities on the ground-state density ρ will be implied for notational clarity.

4.2.2 Multivariate Adiabatic Connection

The eigenstates and energy eigenvalues of the Hamiltonian $\hat{H}^{\boldsymbol{\alpha}}$ satisfy the coupling-strength dependent electronic Schrödinger equation for the supersystem,

$$\hat{H}^{\boldsymbol{\alpha}} \left|\Psi_m^{\boldsymbol{\alpha}}\right\rangle = E_m^{\boldsymbol{\alpha}} \left|\Psi_m^{\boldsymbol{\alpha}}\right\rangle, \quad m = 0, 1, \ldots. \tag{4.6}$$

where the supersystem eigenstates $|\Psi_m^{\boldsymbol{\alpha}}\rangle$ is required to be antisymmetric for all electron permutations within each monomer. Alternatively, the $|\Psi_m^{\boldsymbol{\alpha}}\rangle$ could be required to be antisymmetric with respect to any electron permutations in the spirit of symmetry-adapted perturbation theory (SAPT),[200–202] but the resulting exchange effects are exponentially small in the nonoverlapping limit. The physical ground state $|\Psi_0^{\mathbf{J}}\rangle = |\Psi_0\rangle$ includes all monomer interactions, and its energy is $E_0^{\mathbf{J}} = E_0$, whereas the KS ground-state $|\Psi_0^{\mathbf{0}}\rangle$ is a single Slater determinant with energy eigenvalue $E_0^{\mathbf{0}}$. Invoking the Hellman-Feynman theorem, we define the MAC coupling-strength integrand in analogy to the scalar case,

$$\mathbf{W}^{\boldsymbol{\alpha}} = \langle\Psi_0^{\boldsymbol{\alpha}}|\hat{\mathbf{V}}^{\mathrm{ee}}|\Psi_0^{\boldsymbol{\alpha}}\rangle = \frac{d}{d\boldsymbol{\alpha}}\left(E_0^{\boldsymbol{\alpha}} + \int dx\, V^{\boldsymbol{\alpha}\mathrm{HXC}}(x)\rho(x)\right). \tag{4.7}$$

$\mathbf{W}^{\boldsymbol{\alpha}}$ is the gradient of a scalar function of $\boldsymbol{\alpha}$, and $\langle\mathbf{W}^{\boldsymbol{\alpha}} d\boldsymbol{\alpha}\rangle$ is a total differential of the ground-state HXC energy of the supersystem, since

$$\int_{C_{\mathbf{0}}^{\mathbf{J}}} \langle\mathbf{W}^{\boldsymbol{\alpha}} d\boldsymbol{\alpha}\rangle = \left(E_0^{\boldsymbol{\alpha}} + \int dx\, V^{\boldsymbol{\alpha}\mathrm{HXC}}(x)\rho(x)\right)\Big|_{\boldsymbol{\alpha}=\mathbf{0}}^{\boldsymbol{\alpha}=\mathbf{J}} = E_0^{\mathbf{J}} + \int dx\, V^{\mathrm{HXC}}(x)\rho(x) - E_0^{\mathbf{0}} = E^{\mathrm{HXC}}. \tag{4.8}$$

In the last step, it was used that $V^{\mathbf{J}\mathrm{HXC}}(x) = V^{\mathbf{J}\mathrm{HXC}}(x)$, whereas $V^{\mathbf{0}\mathrm{HXC}}(x) = 0$. Single brackets $\langle\cdot\rangle$ are used to denote traces over monomer indices, whereas double brackets $\langle\langle\cdot\rangle\rangle$ denote traces over all electronic degrees of freedom. $C_{\mathbf{0}}^{\mathbf{J}}$ is *any* piece-wise smooth curve on the $N(N+1)$-dimensional domain of $\boldsymbol{\alpha}$ starting at $\mathbf{0}$ (zero coupling) and ending at $\mathbf{J}$

(full coupling). This arbitrariness of $C_{\mathbf{0}}^{\mathbf{J}}$ will allow us to choose MAC integration paths particularly suited for studying molecular interactions and for discussing size-consistency.

4.2.3 MAC Interaction Path

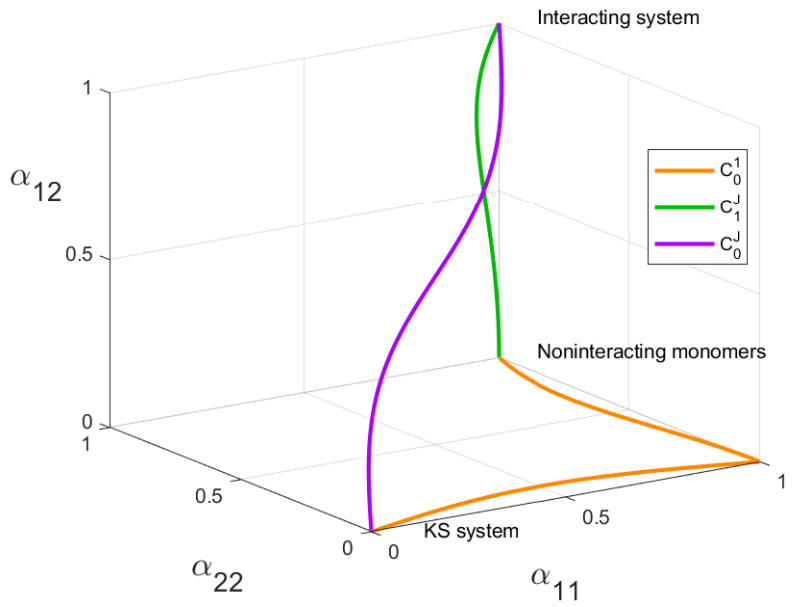


Figure 4.1: The full interaction between two monomers is illustrated by the decomposition of the MAC interaction path into the intermonomer interaction (α_{12}) and the intramonomer interactions (α_{11} and α_{22}).

To define the dispersion energy, we decompose the interaction path as

$$C_{\mathbf{0}}^{\mathbf{J}} = C_{\mathbf{0}}^{\mathbf{1}} \cup C_{\mathbf{1}}^{\mathbf{J}}, \tag{4.9}$$

where $C_{\mathbf{0}}^{\mathbf{1}}$ turns on the intramonomer interaction (corresponding to the diagonal elements of $\boldsymbol{\alpha}$) only, and $C_{\mathbf{1}}^{\mathbf{J}}$ subsequently turns on the intermonomer interaction, see Figure 4.1. Hence,

$$E_{\mathrm{diag}}^{\mathrm{HXC}} = \int_{C_{\mathbf{0}}^{\mathbf{1}}} \langle \mathbf{W}^{\boldsymbol{\alpha}} d\boldsymbol{\alpha} \rangle \tag{4.10}$$

corresponds to the HXC energy associated with the intramonomer electron-electron interaction, whereas

$$E_{\text{int}}^{\text{HXC}} = \int_{C_{\mathbf{1}}^{\mathbf{J}}} \langle \mathbf{W}^{\boldsymbol{\alpha}} d\boldsymbol{\alpha} \rangle \tag{4.11}$$

is the HXC energy associated with the intermonomer interaction, starting from the monomers at full coupling. The total supersystem ground-state energy is thus

$$E_0 = E_0^{\mathbf{0}} - \int dx\, V^{\text{HXC}}(x)\rho(x) + E_{\text{diag}}^{\text{HXC}} + E_{\text{int}}^{\text{HXC}}, \tag{4.12}$$

where $E_{\text{diag}}^{\text{HXC}}$ contains all intramonomer ($I = J$) parts of the total energy due to electron interaction, whereas the HXC interaction energy $E_{\text{int}}^{\text{HXC}}$ accounts for all contributions due to interactions of different monomers ($I \neq J$). In analogy to the dimer case,[195] $E_{\text{int}}^{\text{HXC}}$ contains no induction effects, because the ground-state density is kept fixed at the ground-state density of the supersystem. The induction energy can be obtained by considering the isolated monomers without density constraint. Moreover, the entire one-electron part of the interaction energy is contained in $E_0^{\mathbf{0}}$. The Hartree and exchange part of $E_{\text{int}}^{\text{HXC}}$ is first order in $\boldsymbol{\alpha}$ and amounts to the electrostatic part of the electron-electron interaction,

$$E_{\text{int}}^{\text{HX}} = \int_{C_{\mathbf{1}}^{\mathbf{J}}} \langle \mathbf{W}^{\mathbf{1}} d\boldsymbol{\alpha} \rangle = \langle \mathbf{W}^{\mathbf{1}}(\mathbf{J} - \mathbf{1}) \rangle = \frac{1}{2} \sum_{I \neq J} \langle \Psi_0^{\mathbf{1}} | \hat{V}_{IJ}^{\text{ee}} | \Psi_0^{\mathbf{1}} \rangle = \frac{1}{2} \sum_{I \neq J} \int dx_1 dx_2 \frac{\rho_I(x_1)\rho_J(x_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}; \tag{4.13}$$

the exchange part vanishes exponentially with the intermonomer distance and therefore has been omitted. $|\Psi_0^{\mathbf{1}}\rangle$ is the ground state of the supersystem with full intramonomer interaction, but zero intermonomer electron interaction. The remainder,

$$E_{\text{int}}^{\text{C}} = \int_{C_{\mathbf{1}}^{\mathbf{J}}} \left\langle \left(\mathbf{W}^{\boldsymbol{\alpha}} - \mathbf{W}^{\mathbf{1}} \right) d\boldsymbol{\alpha} \right\rangle, \tag{4.14}$$

is the correlation part of the electron-electron interaction and hence the dispersion energy.

4.2.4 Fluctuation-Dissipation Theorem

Following the argument in the dimer case,[195] the dispersion energy, equation (4.14), may be expressed as the difference between correlated and uncorrelated intermonomer density fluctuations by factorizing the coupling strength integrand,

$$E_{\text{int}}^{\text{C}} = \frac{1}{2}\sum_{IJ=1}^{N}\int_{C_{\mathbf{1}}^{\mathbf{J}}} d\alpha_{IJ}\int \frac{dx_1 dx_2}{|\mathbf{r}_1-\mathbf{r}_2|}(\langle\Psi_0^{\boldsymbol{\alpha}}|\Delta\hat{\rho}_I(x_1)\Delta\hat{\rho}_J(x_2)|\Psi_0^{\boldsymbol{\alpha}}\rangle - \langle\Psi_0^{\mathbf{1}}|\Delta\hat{\rho}_I(x_1)\Delta\hat{\rho}_J(x_2)|\Psi_0^{\mathbf{1}}\rangle). \tag{4.15}$$

(Since $d\alpha_{II} = 0$ along $C_{\mathbf{1}}^{\mathbf{J}}$, the intramonomer correlations are excluded.) Here, $\Delta\hat{\rho}_I = \hat{\rho}_I(x) - \rho_I(x)$ is the density fluctuation operator associated with fragment I, and $\hat{\rho}_I(x)$ is the corresponding density operator. As in the dimer case, using the completeness of the eigenstates $\{|\Psi_m^{\boldsymbol{\alpha}}\rangle\}$ yields the analog of the fluctuation-dissipation theorem for the dispersion energy,

$$E_{\text{int}}^{\text{C}} = -\frac{1}{2}\int_{C_{\mathbf{1}}^{\mathbf{J}}}\int_{-\infty}^{\infty}\frac{dz}{2\pi i}\left\langle\!\!\left\langle \left(\boldsymbol{\Pi}^{\boldsymbol{\alpha}}(z) - \boldsymbol{\Pi}^{\mathbf{1}}(z)\right) d\mathbf{V}^{\boldsymbol{\alpha}} \right\rangle\!\!\right\rangle. \tag{4.16}$$

$\boldsymbol{\Pi}^{\boldsymbol{\alpha}}(z)$ is the time-ordered supersystem polarization propagator at coupling strength $\boldsymbol{\alpha}$ and imaginary frequency $z = i\omega \in i\mathbb{R}$. Using the Lehmann representation, $\boldsymbol{\Pi}^{\boldsymbol{\alpha}}(z)$ may be expressed in terms of the eigenpairs $\{|\Psi_m^{\boldsymbol{\alpha}}\rangle, E_m^{\boldsymbol{\alpha}}\}$,

$$\boldsymbol{\Pi}_{IJ}^{\boldsymbol{\alpha}}(z) = \sum_{m\neq 0}\left\{\frac{\boldsymbol{\gamma}_{0mI}^{\boldsymbol{\alpha}}\otimes\boldsymbol{\gamma}_{0mJ}^{\boldsymbol{\alpha}\dagger}}{z-\Omega_m^{\boldsymbol{\alpha}}+i0^+} - \frac{\boldsymbol{\gamma}_{0mJ}^{\boldsymbol{\alpha}\dagger}\otimes\boldsymbol{\gamma}_{0mI}^{\boldsymbol{\alpha}}}{z+\Omega_m^{\boldsymbol{\alpha}}-i0^+}\right\}. \tag{4.17}$$

$\Omega_m^{\boldsymbol{\alpha}} = E_m^{\boldsymbol{\alpha}} - E_0^{\boldsymbol{\alpha}}$ is the energy of an excitation from $|\Psi_0^{\boldsymbol{\alpha}}\rangle$ to $|\Psi_m^{\boldsymbol{\alpha}}\rangle$, and $\gamma_{0mI}^{\boldsymbol{\alpha}}$ denotes the diagonal block of the corresponding one-particle transition density matrix – off-diagonal blocks corresponding to intermonomer charge-transfer excitations may be neglected, because their

contribution to the energy decays exponentially with the overlap between the monomers. As in the dimer case, $\mathbf{\Pi}^{\boldsymbol{\alpha}}(z)$ is a self-adjoint and negative semidefinite operator on the tensor-product space of one-particle operators, but it now has $N(N+1)/2$ blocks, one for each monomer pair. $\mathbf{V}^{\boldsymbol{\alpha}}$ is the $\boldsymbol{\alpha}$-scaled bare electron–electron Coulomb interaction (or Hartree kernel) on the same space; because the Hartree kernel is spatially local, $\mathbf{V}^{\boldsymbol{\alpha}}$ consists of $N(N+1)/2$ blocks $V_{IJ}^{\boldsymbol{\alpha}} = \alpha_{IJ} V_{IJ}$ corresponding to intra- and intermonomer parts of the interaction. The differential applies to the coupling constant, i.e., $[d\mathbf{V}^{\boldsymbol{\alpha}}]_{IJ} = V_{IJ} d\alpha_{IJ}$.

4.2.5 Generalized Dielectric Function

It is impractical to obtain $\mathbf{\Pi}^{\boldsymbol{\alpha}}$ from equation (4.17) because knowledge of the supersystem eigenpairs $\{|\Psi_m^{\boldsymbol{\alpha}}\rangle, E_m^{\boldsymbol{\alpha}}\}$ at finite $\boldsymbol{\alpha}$ would be required. Instead, it is desirable to connect $\mathbf{\Pi}^{\boldsymbol{\alpha}}$ to its noninteracting KS counterpart, $\mathbf{\Pi}^{\mathbf{0}}$, according to

$$\mathbf{\Pi}^{\boldsymbol{\alpha}}(z) = [\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}(z)]^{-1} \mathbf{\Pi}^{\mathbf{0}}(z). \tag{4.18}$$

$\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}(z)$, the generalized dielectric function at coupling strength $\boldsymbol{\alpha}$, is further related to $\mathbf{\Pi}^{\mathbf{0}}$ according to[185,203]

$$\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}(z) = \mathbf{1} - \mathbf{\Pi}^{\mathbf{0}}(z) \mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}(z), \tag{4.19}$$

where $\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}(z)$ denotes the HXC kernel at coupling strength α and imaginary frequency z. (In the following, the frequency argument will be implied for clarity where possible.) $\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}$ is spatially local and has the same block structure as $\mathbf{V}$. The inverse $[\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}]^{-1}$ may be interpreted as generalized dielectric screening factor and will be just referred to as "screening factor" in the following. Existence of the inverse is guaranteed as long as the ground state $|\Psi_0^{\boldsymbol{\alpha}}\rangle$ is stable and nondegenerate.

$\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}$ may be decomposed into a diagonal (intramonomer) part at full coupling, $\mathbf{K}^{\mathbf{1}\mathrm{HXC}}$, and a remainder $\mathbf{K}_{\mathrm{int}}^{\boldsymbol{\alpha}\mathrm{HXC}}$, which accounts for the intermonomer interaction,

$$\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}} = \mathbf{K}^{\mathbf{1}\mathrm{HXC}} + \mathbf{K}_{\mathrm{int}}^{\boldsymbol{\alpha}\mathrm{HXC}}. \tag{4.20}$$

It follows that

$$\boldsymbol{\epsilon}^{\boldsymbol{\alpha}} = \mathbf{1} - \boldsymbol{\Pi}^{\mathbf{0}}\mathbf{K}^{\mathbf{1}\mathrm{HXC}} - \boldsymbol{\Pi}^{\mathbf{0}}\mathbf{K}_{\mathrm{int}}^{\boldsymbol{\alpha}\mathrm{HXC}} = \boldsymbol{\epsilon}^{\mathbf{1}}\boldsymbol{\epsilon}_{\mathrm{int}}^{\boldsymbol{\alpha}}, \quad \boldsymbol{\alpha} \in C_{\mathbf{1}}^{\mathbf{J}}, \tag{4.21}$$

where $\boldsymbol{\epsilon}^{\mathbf{1}} = \mathbf{1} - \boldsymbol{\Pi}^{\mathbf{0}}\mathbf{K}^{\mathbf{1}\mathrm{HXC}}$ is the generalized dielectric function accounting for intramonomer interaction, and

$$\boldsymbol{\epsilon}_{\mathrm{int}}^{\boldsymbol{\alpha}} = \mathbf{1} - \boldsymbol{\Pi}^{\mathbf{1}}\mathbf{K}_{\mathrm{int}}^{\boldsymbol{\alpha}\mathrm{HXC}}, \tag{4.22}$$

the generalized dielectric function for the intermonomer interaction, contains the *monomer-screened* polarization propagator

$$\boldsymbol{\Pi}^{\mathbf{1}} = [\boldsymbol{\epsilon}^{\mathbf{1}}]^{-1}\boldsymbol{\Pi}^{\mathbf{0}} \tag{4.23}$$

at full intramonomer interaction.

4.2.6 Dispersion Energy

Insertion of equation (4.18) into the fluctuation-dissipation theorem for the dispersion energy (4.16) yields

$$E_{\mathrm{int}}^{\mathrm{C}} = -\frac{1}{2}\int_{C_{\mathbf{1}}^{\mathbf{J}}}\int_{-\infty}^{\infty}\frac{dz}{2\pi i}\left\langle\!\!\left\langle \left([\boldsymbol{\epsilon}_{\mathrm{int}}^{\boldsymbol{\alpha}}(z)]^{-1} - \mathbf{1}\right)\boldsymbol{\Pi}^{\mathbf{1}}(z)d\mathbf{V}^{\boldsymbol{\alpha}} \right\rangle\!\!\right\rangle. \tag{4.24}$$

This equation may be further transformed in two different ways: First, by exploiting the zero trace of $\mathbf{\Pi^1} d\mathbf{V}^{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha} \in C_{\mathbf{1}}^{\mathbf{J}}$, we arrive at a screened-exchange like expression for the dispersion energy,

$$E_{\text{int}}^{\text{C}} = -\frac{1}{2} \int_{C_{\mathbf{1}}^{\mathbf{J}}} \int_{-\infty}^{\infty} \frac{dz}{2\pi i} \left\langle\!\!\left\langle [\boldsymbol{\epsilon}_{\text{int}}^{\boldsymbol{\alpha}}(z)]^{-1} \mathbf{\Pi^1}(z) d\mathbf{V}^{\boldsymbol{\alpha}} \right\rangle\!\!\right\rangle. \tag{4.25}$$

Second, by noting that

$$[\boldsymbol{\epsilon}_{\text{int}}^{\boldsymbol{\alpha}}]^{-1} - \mathbf{1} = [\boldsymbol{\epsilon}_{\text{int}}^{\boldsymbol{\alpha}}]^{-1} (\mathbf{1} - \boldsymbol{\epsilon}_{\text{int}}^{\boldsymbol{\alpha}}) = [\boldsymbol{\epsilon}_{\text{int}}^{\boldsymbol{\alpha}}]^{-1} \mathbf{\Pi^1} \mathbf{K}_{\text{int}}^{\boldsymbol{\alpha}\text{HXC}}, \tag{4.26}$$

we arrive at the screened second-order-like expression

$$E_{\text{int}}^{\text{C}} = -\frac{1}{2} \int_{C_{\mathbf{1}}^{\mathbf{J}}} \int_{-\infty}^{\infty} \frac{dz}{2\pi i} \left\langle\!\!\left\langle [\boldsymbol{\epsilon}_{\text{int}}^{\boldsymbol{\alpha}}(z)]^{-1} \mathbf{\Pi^1}(z) \mathbf{K}_{\text{int}}^{\boldsymbol{\alpha}\text{HXC}}(z) \mathbf{\Pi^1}(z) d\mathbf{V}^{\boldsymbol{\alpha}} \right\rangle\!\!\right\rangle. \tag{4.27}$$

While equation (4.27) might be further simplified in the dimer case by exploiting the fact that only odd orders in the coupling constant have nonzero trace, this simplification is no longer possible in the N-monomer case, where even orders can be nonzero.

4.3 Partitioning of the Dispersion Energy

4.3.1 Subsystem Partitioning of the Interaction Path

To investigate the behavior of the dispersion energy as a function of N, we partition the MAC interaction path $C_{\mathbf{1}}^{\mathbf{J}}$ into $N-1$ segments $C_2, C_3, \ldots, C_N$, such that

$$C_{\mathbf{1}}^{\mathbf{J}} = \bigcup_{K=2}^{N} C_K. \tag{4.28}$$

Each segment C_K turns on the interaction between the K-th monomer and a $K-1$-mer consisting of already interacting monomers. This may be formalized by introducing the $N \times N$ matrices $\mathbf{J}_K, K = 1, \ldots, N$, defined by recursion $\mathbf{J}_1 = \mathbf{1}$,

$$\mathbf{J}_K - \mathbf{J}_{K-1} = \Delta\mathbf{J}_K = \sum_{L=1}^{K-1} \left(\mathbf{e}_L \otimes \mathbf{e}_K + \mathbf{e}_K \otimes \mathbf{e}_L\right), \tag{4.29}$$

where $\{\mathbf{e}_I, I = 1, \ldots, N\}$ are unit vectors. It follows that

$$\mathbf{J}_K = \mathbf{1} + \sum_{K=2}^{K} \Delta\mathbf{J}_K \tag{4.30}$$

and $\mathbf{J}_N = \mathbf{J}$. Along each segment C_K, the coupling constant changes from $\mathbf{J}_{K-1}$ to $\mathbf{J}_K$, i.e.,

$$C_K = C^{\mathbf{J}_K}_{\mathbf{J}_{K-1}}. \tag{4.31}$$

Using the MAC interaction path partitioning (4.28), the dispersion energy, equation (4.27), decomposes into a sum of $N-1$ terms,

$$E^{\mathrm{C}}_{\mathrm{int}} = \sum_{K=2}^{N} \Delta E^{\mathrm{C}}_{\mathrm{int}K}, \tag{4.32}$$

where

$$\Delta E^{\mathrm{C}}_{\mathrm{int}K} = -\frac{1}{2} \int_{C_K} \int_{-\infty}^{\infty} \frac{dz}{2\pi i} \left\langle\!\!\left\langle [\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}}(z)]^{-1} \boldsymbol{\Pi}^{\mathbf{1}}(z) d\mathbf{V}^{\boldsymbol{\alpha}} \right\rangle\!\!\right\rangle. \tag{4.33}$$

Intuitively, each increment $\Delta E^{\mathrm{C}}_{\mathrm{int}K}$ should amount to the dispersion energy gained when the K-th monomer is allowed to interact with an existing $K-1$-mer.

The interaction path (4.28) is by no means unique – any (non-identity) permutation of the N monomers may give rise to a different path, but the final interaction energy is path-independent and therefore invariant under such permutations. The following results are

nevertheless general in that they do not depend on a specific ordering of monomers.

4.3.2 Partitioning of the HXC Kernel

Along the MAC interaction path (4.28), the HXC interaction kernel $\mathbf{K}_{\text{int}}^{\boldsymbol{\alpha}\text{HXC}}$ changes from $\mathbf{0}$ to its value at full coupling. This suggests a partitioning into $N-1$ increments

$$\mathbf{K}_{\text{int}}^{\boldsymbol{\alpha}\text{HXC}} = \sum_{K=2}^{N} \Delta\mathbf{K}_{\text{int}K}^{\boldsymbol{\alpha}\text{HXC}}, \tag{4.34}$$

where each increment accounts for the change of $\mathbf{K}_{\text{int}}^{\boldsymbol{\alpha}\text{HXC}}$ along C_K,

$$\Delta\mathbf{K}_{\text{int}K}^{\boldsymbol{\alpha}\text{HXC}} = \begin{cases} \mathbf{0}; & \boldsymbol{\alpha} \in C_{\mathbf{1}}^{\mathbf{J}_{K-1}} \\ \mathbf{K}_{\text{int}}^{\boldsymbol{\alpha}\text{HXC}} - \mathbf{K}_{\text{int}}^{\mathbf{J}_{K-1}\text{HXC}}; & \boldsymbol{\alpha} \in C_K \\ \mathbf{K}_{\text{int}}^{\mathbf{J}_K\text{HXC}} - \mathbf{K}_{\text{int}}^{\mathbf{J}_{K-1}\text{HXC}}; & \boldsymbol{\alpha} \in C_{\mathbf{J}_K}^{\mathbf{J}} \end{cases} . \tag{4.35}$$

The thus defined $\Delta\mathbf{K}_{\text{int}K}^{\boldsymbol{\alpha}\text{HXC}}$ exhibit $\boldsymbol{\alpha}$ dependence only for $\boldsymbol{\alpha} \in C_K$, and become either zero or assume their value at full coupling elsewhere. Again, this partitioning is not unique, but it produces the correct behavior in the large separation limit, as shown below.

4.3.3 Partitioning of the Generalized Dielectric Function

In Sec. 4.2.5, we observed that additive separability of the HXC kernel leads to multiplicative separability of the generalized dielectric function. Given the additive separation of the HXC interaction kernel from equation (4.34), we separate $\boldsymbol{\epsilon}_{\text{int}}^{\boldsymbol{\alpha}}$ as

$$\boldsymbol{\epsilon}_{\text{int}}^{\boldsymbol{\alpha}} = \boldsymbol{\epsilon}_{\text{int}N}^{\boldsymbol{\alpha}} = \prod_{K=1}^{N} \Delta\boldsymbol{\epsilon}_{\text{int}K}^{\boldsymbol{\alpha}}, \tag{4.36}$$

where $\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}1} = \mathbf{1}$ and

$$\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K} = \mathbf{1} - [\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}N-1}]^{-1}\boldsymbol{\Pi}^{\mathbf{1}}\Delta\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}K}. \tag{4.37}$$

Equation (4.36) implies that the screening factor is multiplicatively separable as well,

$$[\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}}]^{-1} = [\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}N}]^{-1} = \prod_{K=N}^{1}[\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K}]^{-1}. \tag{4.38}$$

Existence of the inverses is again guaranteed as long as the ground state remains stable and nondegenerate along the interaction path. The multiplicative increments $\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K}$ exhibit $\boldsymbol{\alpha}$ dependence only for $\boldsymbol{\alpha} \in C_K$, and either equal the identity (no screening) or assume their value at full coupling elsewhere: Using the definition of $\Delta\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}K}$ (4.35), one finds

$$\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K} = \begin{cases} \mathbf{1}; & \boldsymbol{\alpha} \in C^{\mathbf{J}_{K-1}}_{\mathbf{1}} \\ \mathbf{1} - [\boldsymbol{\epsilon}^{\mathbf{J}}_{\mathrm{int}N-1}]^{-1}\boldsymbol{\Pi}^{\mathbf{1}}\Delta\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}K}; & \boldsymbol{\alpha} \in C_K \\ \mathbf{1} - [\boldsymbol{\epsilon}^{\mathbf{J}}_{\mathrm{int}N-1}]^{-1}\boldsymbol{\Pi}^{\mathbf{1}}\Delta\mathbf{K}^{\mathbf{J}\mathrm{HXC}}_{\mathrm{int}K}; & \boldsymbol{\alpha} \in C^{\mathbf{J}}_{\mathbf{J}_K} \end{cases}. \tag{4.39}$$

In particular, for $\boldsymbol{\alpha} \in C_K$, the generalized dielectric constant takes the intuitive form

$$\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}} = \boldsymbol{\epsilon}^{\mathbf{J}}_{\mathrm{int}K-1}\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K}, \boldsymbol{\alpha} \in C_K. \tag{4.40}$$

Insertion into equation (4.33) yields

$$\Delta E^{\mathrm{C}}_{\mathrm{int}K} = -\frac{1}{2}\int_{C_K}\int_{-\infty}^{\infty}\frac{dz}{2\pi i}\left\langle\!\!\left\langle [\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K}(z)]^{-1}[\boldsymbol{\epsilon}^{\mathbf{J}}_{\mathrm{int}K-1}(z)]^{-1}\boldsymbol{\Pi}^{\mathbf{1}}(z)d\mathbf{V}^{\boldsymbol{\alpha}} \right\rangle\!\!\right\rangle. \tag{4.41}$$

This suggests that $[\boldsymbol{\epsilon}^{\mathbf{J}}_{\mathrm{int}K-1}]^{-1}\boldsymbol{\Pi}^{\mathbf{1}}$ is the equivalent of the $K-1$-mer polarization propagator, whereas $[\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K}]^{-1}$ accounts for screening due to interaction between the K-th monomer and the $K-1$-mer.

4.3.4 Proof that $\Delta E^{\mathrm{C}}_{\mathbf{int}K}$ is a Dimer Dispersion Energy

To prove that our intuitive interpretation of equation (4.41) is correct, we need to show that $\Delta E^{\mathrm{C}}_{\mathrm{int}K}$ indeed equals the dispersion interaction between the (fully interacting) $K-1$-mer and the K-th monomer, and hence $\Delta E^{\mathrm{C}}_{\mathrm{int}K}$ reduces to the expression for the dimer interaction energy obtained by Nguyen and co-workers,[195] with one monomer corresponding to monomer K and the other monomer corresponding to the compound $K-1$-mer.

We start by observing that, at long range, the I,J-block of $\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}}$ depends on α_{IJ} only,

$$K^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}IJ} = K^{\alpha_{IJ}\mathrm{HXC}}_{\mathrm{int}IJ}. \tag{4.42}$$

Although we will try to weaken this assumption later, it is obviously true because $\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}}$ reduces to $\mathbf{V}^{\boldsymbol{\alpha}}_{\mathrm{int}}$ in this limit. For the following, it is convenient to partition the underlying N-dimensional vector space into a direct sum of a $K-1$-dimensional space associated with the $K-1$-mer, a one-dimensional space associated with monomer K, and an $N-K$ dimensional subspace (which may be zero) associated with the remaining monomers. We further adopt a short-hand bracket notation for (sub)vectors and (sub)matrices, e.g., $\boldsymbol{\alpha}_{[K]L}$ stands for a column vector containing the first K elements of the L-th row of $\boldsymbol{\alpha}$, etc. The HXC kernel increment thus takes the simple form

$$\Delta\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}K} = \begin{pmatrix} \mathbf{0} & \mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}[K-1]K} & \mathbf{0} \\ \mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}K[K-1]} & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{4.43}$$

As expected, $\Delta\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}K}$ only changes for $\boldsymbol{\alpha} \in C_K$, which turns on all $\alpha_{KL}, L = 1,\ldots K-1$. As a result, only the first $K\times K$ sub-block of $\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}}$ is nonzero for $\boldsymbol{\alpha} \in C_k$, and hence only the first $K\times K$ sub-block of $\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}}$ is different from $\mathbf{1}$ under these conditions. The (block-diagonal)

monomer polarization propagator may be partitioned as

$$\mathbf{\Pi^1} = \begin{pmatrix} \mathbf{\Pi^1_{[K-1][K-1]}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Pi^{\mathbf{1}}_{KK} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Pi^1_{[N-K][N-K]}} \end{pmatrix}, \tag{4.44}$$

where $\mathbf{\Pi^1_{[K-1][K-1]}}$ is the polarization propagator of the $K-1$-mer including all intramonomer but no intermonomer interaction, and $\Pi^{\mathbf{1}}_{KK}$ is the polarization propagator of the K-th monomer at full coupling. Moreover, for $\boldsymbol{\alpha} \in C_k$,

$$\boldsymbol{\epsilon}^{\mathbf{J}}_{\mathrm{int}K-1} = \begin{pmatrix} \boldsymbol{\epsilon}^{\mathbf{J}}_{\mathrm{int}[K-1][K-1]} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix}, \tag{4.45}$$

and therefore $[\boldsymbol{\epsilon}^{\mathbf{J}}_{\mathrm{int}K-1}]^{-1}\mathbf{\Pi^1}$ contains

$$\mathbf{\Pi^J_{[K-1][K-1]}} = [\boldsymbol{\epsilon}^{\mathbf{J}}_{\mathrm{int}[K-1][K-1]}]^{-1}\mathbf{\Pi^1_{[K-1][K-1]}}, \tag{4.46}$$

the $K-1$-mer polarization propagator at *full* (intra- and intermonomer) interaction in its first $(K-1)\times(K-1)$ block. The incremental dielectric function thus can be written as

$$\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K} = \begin{pmatrix} \mathbf{1} & \mathbf{\Pi^J_{[K-1][K-1]}}\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{[K-1]K} & \mathbf{0} \\ \Pi^{\mathbf{1}}_{KK}\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{K[K-1]} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix}. \tag{4.47}$$

Finally, for $\boldsymbol{\alpha} \in C_k$,

$$d\mathbf{V}^{\boldsymbol{\alpha}} = \begin{pmatrix} \mathbf{0} & d\mathbf{V}^{\boldsymbol{\alpha}}_{[K-1]K} & \mathbf{0} \\ d\mathbf{V}^{\boldsymbol{\alpha}}_{K[K-1]} & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{4.48}$$

Inserting equations (4.44)-(4.48) into the expression for the incremental dispersion energy (4.41), we observe that there is zero contribution to the trace from monomers $K+1,\dots N$, and therefore

$$\begin{aligned}\Delta E^{\mathrm{C}}_{\mathrm{int}K} = &-\frac{1}{2}\int_{C_K}\int_{-\infty}^{\infty}\frac{dz}{2\pi i}\\ &\times\left\langle\!\!\left\langle \begin{pmatrix} \mathbf{1} & \mathbf{\Pi}^{\mathbf{J}}_{[K-1][K-1]}\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{[K-1]K} \\ \Pi^{\mathbf{1}}_{KK}\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{K[K-1]} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{\Pi}^{\mathbf{J}}_{[K-1][K-1]} & \mathbf{0} \\ \mathbf{0} & \Pi^{\mathbf{1}}_{KK} \end{pmatrix}\right.\right.\\ &\left.\left.\times \begin{pmatrix} \mathbf{0} & d\mathbf{V}^{\boldsymbol{\alpha}}_{[K-1]K} \\ d\mathbf{V}^{\boldsymbol{\alpha}}_{K[K-1]} & 0 \end{pmatrix}\right\rangle\!\!\right\rangle.\end{aligned} \tag{4.49}$$

If the integration path C_K is chosen such that $d\boldsymbol{\alpha}_{[K-1]K} = d\alpha$, where α is a scalar integration variable, equation (4.49) recovers the expression by Nguyen and co-workers[195] for the dispersion interaction of a dimer consisting of a compound $K-1$-mer and the K-th monomer, which completes this proof.

It might be possible to further generalize this result by not assuming a pairwise block structure of $\mathbf{K}^{\boldsymbol{\alpha}\mathrm{HXC}}_{\mathrm{int}}$. This would lead to a more general expression for the dimer interaction including changes to the intra-monomer correlation energies due to the interaction. The only hard requirements for proving that $\Delta E^{\mathrm{C}}_{\mathrm{int}K}$ is an interaction energy are (i) additivity of the HXC kernel increments along the MAC interaction path and (ii) invertibility of the resulting generalized dielectric functions.

4.4 Dispersion Size-Consistency

The results in the previous section afford a succinct and general statement of dispersion size-consistency, which apply to any physical N-monomer system at large but finite distance:

The dispersion energy of a K-monomer subsystem equals the dispersion energy

of a composite $K-1$-mer plus the interaction energy of the $K-1$-mer and the K-th monomer.

An equally useful formulation is:

The total dispersion energy of a N-monomer system is independent of any partitioning into subsystems.

The latter statement can be understood as a direct consequence of the independence of the interaction energy on the interaction path. While these statement may appear trivial, they impose stringent conditions violated by many common approximate electronic structure methods, as shown below. Unlike other size-consistency definitions,[67] the above conditions neither apply at infinite separation or in the thermodynamic limit only, but hold for finite systems interacting at finite but large separations.

4.5 Approximate Electronic Structure Methods

4.5.1 Random Phase Approximation

Within the random phase approximation (RPA), $\mathbf{K}^{\alpha\mathrm{HXC}}$ is replaced by its first-order approximation $\mathbf{V}^{\boldsymbol{\alpha}}$, the bare Hartree kernel at coupling strength $\boldsymbol{\alpha}$. Since this approximation preserves all properties we have used to derive dispersion size-consistency, it follows immediately that RPA is dispersion size-consistent. This means that it does not matter whether the dispersion energy of an N-monomer system is obtained from a supermolecular calculation or by calculating interaction energies from its constituent fragments as long as the RPA is used in all steps.

The relatively simple structure of the RPA makes it possible to verify this result explicitly.

By parameterizing the interaction path $C_{\mathbf{1}}^{\mathbf{J}}$ using a scalar coupling constant α, i.e., $d\alpha_{IJ} = d\alpha, I \neq J$, the coupling strength integral within RPA may be performed analytically, yielding

$$E_{\text{int}}^{\text{C RPA}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{dz}{2\pi i}\Big\langle\!\Big\langle \ln[\boldsymbol{\epsilon}_{\text{int RPA}}(z)]^{-1}\Big\rangle\!\Big\rangle, \tag{4.50}$$

in complete analogy to the dimer case.[195] The interaction part of the generalized dielectric function within RPA is, according to equation (4.22),

$$\boldsymbol{\epsilon}_{\text{int RPA}}^{\boldsymbol{\alpha}} = \mathbf{1} - \boldsymbol{\Pi}_{\text{RPA}}^{\mathbf{1}}\mathbf{V}_{\text{int}}^{\boldsymbol{\alpha}}, \tag{4.51}$$

where $\mathbf{V}_{\text{int}}^{\boldsymbol{\alpha}} = \mathbf{V}^{\boldsymbol{\alpha}} - \mathbf{V}^{\mathbf{1}}$, and $\boldsymbol{\Pi}_{\text{RPA}}^{\mathbf{1}}$ is the polarization propagator including the full intramonomer screening within RPA caused by $\mathbf{V}^{\mathbf{1}}$. Next, we observe that the multiplicative separability of the generalized dielectric function according to equation (4.36) carries over to RPA; in particular, at full coupling ($\boldsymbol{\alpha} = \mathbf{J}$),

$$\boldsymbol{\epsilon}_{\text{int RPA}} = \boldsymbol{\epsilon}_{\text{int RPA}N} = \prod_{K=1}^{N}\Delta\boldsymbol{\epsilon}_{\text{int RPA}K}, \tag{4.52}$$

where the increments $\Delta\boldsymbol{\epsilon}_{\text{int RPA}K}$ are defined as before with $\mathbf{K}_{\text{int}}^{\alpha\text{HXC}}$ replaced by $\mathbf{V}_{\text{int}}^{\boldsymbol{\alpha}}$. Inserting the factorization (4.52) in equation (4.50), we obtain an additive decomposition of the RPA dispersion energy,

$$E_{\text{int}}^{\text{C RPA}} = \sum_{K=1}^{N}\Delta E_{\text{int}K}^{\text{C RPA}}, \tag{4.53}$$

where

$$\Delta E_{\text{int}K}^{\text{C RPA}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{dz}{2\pi i}\Big\langle\!\Big\langle \ln[\Delta\boldsymbol{\epsilon}_{\text{int RPA}K}(z)]^{-1}\Big\rangle\!\Big\rangle; \tag{4.54}$$

in the last step, it was used that $\langle\ln(AB)\rangle = \langle\ln A\rangle + \langle\ln B\rangle$ holds even for non-commuting

operators A, B.

Equation (4.53) explicitly verifies the dispersion size-consistency of RPA: The first ($K = 1$) term in the sum is zero, the second ($K = 2$) equals the RPA dispersion energy between monomers 1 and 2, the third ($K = 3$) equals the RPA dispersion energy between the compound 2-mer consisting of 1 and 2 and monomer 3, etc.

4.5.2 Supermolecular MBPT

In supermolecular MBPT, the total energy of the N-monomer system (and any fragments) are evaluated using a power series approximation of the coupling strength integrand $\mathbf{W}^{\boldsymbol{\alpha}}$ around $\boldsymbol{\alpha} = \mathbf{0}$. In the following, we will consider MBPT(2) for illustration, which amounts to a special case of second-order Görling-Levy perturbation theory[152,153] in the present context.

Since the coupling strength dependence enters the coupling strength integrand through the screening factor only, the latter is expanded to first order in $\boldsymbol{\alpha}$ (corresponding to a second-order energy after coupling strength integration). A particularly useful way to perform this expansion is to start from the factorization

$$[\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}]^{-1} = [\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}}]^{-1}[\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{diag}}]^{-1} \tag{4.55}$$

and apply the product rule, which yields

$$[\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}]^{-1(1)} = [\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}}]^{-1(0)}[\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{diag}}]^{-1(1)} + [\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}}]^{-1(1)}[\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{diag}}]^{-1(0)}, \tag{4.56}$$

where orders in $\boldsymbol{\alpha}$ are indicated by superscript $(0), (1), \ldots$. To order zero in $\boldsymbol{\alpha}$ (no interaction), all screening factors are unity. Hence the MBPT(2) screening factor for the supermolecule is a *sum*, rather than a *product*, of the first-order screening factors for the intramolecular and the intermolecular interaction. The first term of the sum in equation

(4.56) gives rise to the intramonomer MBPT(2) energy, whereas the second term leads to the supermolecular MBPT(2) dispersion energy,

$$E_{\text{int}}^{\text{C PT2}} = -\frac{1}{2}\int_{C_1^{\text{J}}}\int_{-\infty}^{\infty}\frac{dz}{2\pi i}\left\langle\!\!\left\langle [\boldsymbol{\epsilon}_{\text{int}}^{\boldsymbol{\alpha}}(z)]^{-1(1)}\boldsymbol{\Pi}^{\mathbf{0}}(z)d\mathbf{V}^{\boldsymbol{\alpha}}\right\rangle\!\!\right\rangle. \tag{4.57}$$

However, because the intra-monomer screening factor enters to order zero only, the non-interacting (bare, uncoupled) KS polarization propagator appears in the supermolecular MBPT(2) expression for the dispersion energy rather than the first-order polarization propagator at full coupling, which would be expected in a dispersion size-consistent method. Whereas the intermonomer interaction is correctly included to first order in equation (4.57), the intra-monomer interaction is not, and therefore supermolecular MBPT(2) is not dispersion size-consistent. Indeed, the screening factor that would give rise to dispersion size-consistency,

$$[\boldsymbol{\epsilon}_{\text{int}}^{\boldsymbol{\alpha}}]^{-1(1)}[\boldsymbol{\epsilon}_{\text{diag}}^{\boldsymbol{\alpha}}]^{-1(1)}, \tag{4.58}$$

is *second* order in $\boldsymbol{\alpha}$ and thus be obtained from supermolecular MBPT(2), which admits only first-order screening factors. The missing intra-monomer screening explains the dramatic overestimation of dispersion interactions observed in supermolecular MBPT(2) calculations in an intuitive way.

4.5.3 SAPT

As opposed to supermolecular MBPT, SAPT uses perturbation theory for the intermolecular part of the interaction only, whereas the — presumably stronger — intramolecular interaction is previously addressed, e.g., using correlated wavefunction methods, RPA, or DFT.[14,115,145,204–206] AC-SAPT corresponds to using a correlated polarization propagator $\boldsymbol{\Pi}^{\mathbf{1}}$ in equation (4.25), followed by Taylor expansion of $\boldsymbol{\alpha}$ around $\mathbf{1}$ (full monomer coupling),

as opposed to a Taylor expansion around the coupling-strength origin in supermolecular MBPT.

In the dimer ($N = 2$) case, AC-SAPT exhibits fairly benign convergence behavior in conjunction with approximations to $\mathbf{\Pi^1}$ that accurately account for intra-monomer correlation,[195] in keeping with the generally high accuracy of dimer interaction energies obtained from SAPT.[207–210] However, this desirable behavior of SAPT does not necessarily carry over to the many-monomer case, because many-monomer SAPT lacks dispersion size-consistency: Writing the generalized dielectric function for the interaction of a composite $K-1$-mer with monomer K as

$$\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K} = \boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K-1}\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K}, \tag{4.59}$$

SAPT(2) requires first-order expansion of the corresponding screening factor around $\boldsymbol{\alpha} = \mathbf{1}$,

$$\begin{aligned}[\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K}]^{-1(1)} &= [\Delta\boldsymbol{\epsilon}^{\mathbf{1}}_{\mathrm{int}K}]^{-1}[\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K-1}]^{-1(1)} + [\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K}]^{-1(1)}[\boldsymbol{\epsilon}^{\mathbf{1}}_{\mathrm{int}K-1}]^{-1} \\ &= [\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K-1}]^{-1(1)} + [\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K}]^{-1(1)},\end{aligned} \tag{4.60}$$

where it was used that $\Delta\boldsymbol{\epsilon}^{\mathbf{1}}_{\mathrm{int}K} = \boldsymbol{\epsilon}^{\mathbf{1}}_{\mathrm{int}K-1} = \mathbf{1}$ (no intermolecular screening). Analogous to the supermolecular MBPT case, the multiplicative separability of the screening factor is lost, but only for the intermolecular part: The first term in the above sum accounts for the intramolecular dispersion interaction of the $K-1$-mer, whereas the second term gives rise to the dispersion interaction between the compound $K-1$-mer and the K-th monomer,

$$E^{\mathrm{C\ SAPT2}}_{\mathrm{int}K} = -\frac{1}{2}\int_{C^{\mathrm{J}}_1}\int_{-\infty}^{\infty}\frac{dz}{2\pi i}\Big\langle\!\Big\langle[\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_{\mathrm{int}K}(z)]^{-1(1)}\mathbf{\Pi^1}(z)d\mathbf{V}^{\boldsymbol{\alpha}}.\Big\rangle\!\Big\rangle. \tag{4.61}$$

Comparing to the exact result, equation (4.41), it is obvious that SAPT(2) misses the screening factor accounting for intra-$K-1$-mer interaction altogether. Therefore, while again the

interaction between the $K-1$-mer and the K-th monomer is treated consistently to first order order, the intra-$K-1$-mer interaction is not, as reflected by the appearance of $\mathbf{\Pi^1}$. While this is perfectly fine in the dimer case, with increasing $K > 2$ SAPT(2) will miss increasingly more intra-$K-1$-mer screening, leading to increasing overestimation of interaction energies. This also applies to SAPT-derived perturbative schemes such as MP2C.[23,24]

The lack of dispersion size-consistency of SAPT becomes obvious when one considers the (more costly) alternative of treating the $K-1$-mer as a compound monomer treated at the same high level as the monomers. In this case the $K-1$-mer polarization propagator would be properly screened, and the resulting interaction energy would indeed be of comparable accuracy as a dimer interaction energy. Thus, the accuracy of the dispersion energy of the K-monomer system depends on how the system is partitioned into subsystems, in violation of dispersion size-consistency.

4.6 Conclusions

The MAC provides a convenient framework to analyze the behavior of intermolecular interactions in many-monomer systems. Dispersion size-consistency is a strong constraint which applies to finite and infinite *interacting* systems of nonoverlapping monomers. A critical requirement for dispersion size-consistency is multiplicative separability of the total screening factor of an N-monomer system into increments $[\Delta\boldsymbol{\epsilon}^{\boldsymbol{\alpha}}_K]^{-1}$ corresponding to additional screening as the K-th monomer is added. With increasing system size, the incremental dispersion energy $E^{\mathrm{C}}_{\mathrm{int}K}$ will converge to the cohesive energy of the bulk many-monomer system, which could be realized in, e.g., a van-der-Waals (nano)crystal or a weakly interacting fluid.

Finite-order many-body perturbation theory is not dispersion size-consistent because it cannot produce multiplicatively separable screening factors. In particular, MBPT(2) grossly

overestimates dispersion energies of large polarizable many-monomer systems, because the incremental dispersion energy $E_{\text{int}K}^{\text{C PT2}}$ is simply the sum of the MBPT(2) interaction energies of the K-th monomer with monomer $K-1,\ldots,1$ *in the absence* of screening by all other monomers. The bulk limit of $E_{\text{int}K}^{\text{C PT2}}$ can be infinite, but even if it is finite and therefore the MBPT(2) correlation energy is size-extensive, $E_{\text{int}K}^{\text{C PT2}}$ may become less and less accurate with increasing K. This illustrates why size-extensivity alone is too weak a criterion to guarantee uniform accuracy of electronic structure methods independent of system size: Size-extensivity merely requires a finite limit of $E_{\text{int}K}^{\text{C PT2}}$ for $K \to \infty$, whereas size-consistency would require that the error of $E_{\text{int}K}^{\text{C PT2}}$ is independent of K. In particular, the good accuracy of low-order MBPT observed in relatively small molecular systems may not carry over to larger systems and the thermodynamic limit. This might explain why MBPT is still relatively popular in applications to small molecules but produces large errors, e.g., in cohesion energies of molecular crystals with moderately polarizable monomers.[211] SAPT methods can be very accurate for dimers, but may eventually suffer from loss of accuracy for large N. The relatively uniform accuracy of RPA dispersion energies independent of system size,[102,195,212,213] on the other hand, may be viewed as a consequence of dispersion size-consistency. Recent algorithmic improvements have made RPA calculations possible at only marginally higher computational cost than low-order MBPT, making RPA preferable for computing dispersion energies in most applications.[102,182,184,186,214–216]

The violation of dispersion size-consistency by MBPT is connected to the physical origin of its divergence for NIs in large molecules: The lack of screening in the MBPT dispersion energy leads to divergence of the geometric series expansion of the generalized screening factor which generates the intermolecular interaction energy expansion.[195] Nevertheless, the concept of dispersion size-consistency developed here does not rely on perturbative arguments (except, perhaps, for the strong assumption of the HXC kernel reducing to a pairwise interaction at large separation) and applies to any electronic structure method.

The limitations of MBPT in the large N limit also have implications for methods derived from it, such as long-range van-der-Waals corrections.[85,217–223] The present results suggest that methods including "dispersion polarization" such as many-body dispersion[94,224] or quantum Drude models[180,225,226] are distinctly preferable to conventional dispersion corrections for applications to large and polarizable systems, including nanocrystals, metallic nanoparticle or nanowire arrays, or large polarizable biomolecules.

Chapter 5

Modeling the Interactions between the Lissoclimide Family and Eukaryotic 80S Ribosome

Contribution Statement: The following are my contributions to the project. I designed and verified the computational model to understand the interactions between the lissoclimide family and eukaryotic 80S ribosome. I discussed with experimentalists from the Vanderwall Lab at the University of California, Irvine, and applied the structure-activity relationship (SAR) of the lissoclimide family inhibtion of protein synthesis to develop the model. Next, I performed geometry optimization of the lissoclimide model, computed binding energies, and performed a correlation analysis between the predicted binding energies and half of the maximum inhibitory concentration (IC_{50}). Based on the SAR and correlation plot, I proposed potential candidates for inhibiting protein synthesis. Furthermore, I structurally analyzed the lissoclimide models such as bond distances, dihedral angles, and root mean square deviation using visual molecular dynamics. Lastly, I contributed to the writing, created all plots with R, and am currently preparing a manuscript for submission.

5.1 Introduction

Natural products have been important sources for developing new potential cancer therapeutics.[227,228] They have also been used to provide insight into the biological functions and mechanisms of proteins.[229–232] In the early 1990s, chlorolissoclimide (CL) and dichlorolissoclimide (DCL) were extracted from an ascidian *Lissoclinum voeltzkowi*.[233] These are powerful cytotoxic bicyclic diterpene alkaloids that inhibit eukaryotic translation by preventing the transfer ribonucleic acid (tRNA) from exiting the ribosome and interfering with the elongation step of protein synthesis.[234] The interference leads to polysomal accumulation and eventual cell death.

Until recently, Könst and coworkers[235] expanded the understanding of the structure-activity relationship (SAR) for the lissoclimide family inhibition of protein synthesis. An X-ray co-crystal structure of CL with the eukaryotic 80S ribosome showed a novel halogen–π interaction with the guanine residues (G2793 and G2794) in a face–on geometry.[235,236]

A simplified computational model of the interactions between the chlorine from CL and guanine nucleotides was set up by removing the phosphate backbone between the nucleobases and replacing the CL with chloromethane.[235] This simplification revealed a dispersion-based stabilization by up to 1.1 kcal/mol based on the gas phase dispersion corrected semilocal Perdew–Burke–Ernzerhof (PBE) functional.[235] However, the face–on halogen–π interaction with a nucleobase is not well understood. This understanding may be valuable for rationale drug design based on structure and function.

In previous studies, halogen substituents have been recognized to influence intermolecular interactions and contribute to the stability between the ligand and target.[235,237–240] These interactions have been difficult to computationally predict using force fields.[241]

With expanded knowledge of the SAR for the lissocimide family, an improved model is pro-

posed to better capture the interactions between the lissoclimide analogues and ribosome. The simplified model by Könst and co-workers[235] was adapted to include the phosphate backbone and the drug along with representing the remaining protein environment by an implicit solvation model. One potential model is the conductor-like screening model (COSMO),[242] which has been a valuable tool for drug discovery.[243–246] Altogether, the improved lissoclimide model has laid the foundation for rationalizing SAR and provided an approach to expand potential cancer drug candidates within the lissoclimide family. Benchmark results of the X40 testset[247] are also presented to confirm electronic structure methods for modeling the halogen–π interaction. These results can provide a better understanding of the unique halogen–π interaction.

5.2 Computational Details

5.2.1 Modeling the Lissoclimide-Ribosome Interaction

The co-crystal structure of the CL and ribosome determined by Könst and coworkers[235] was used to understand the halogen–π interaction. Since the interaction is localized in the E-site of the ribosome, only the CL, the guanine nucleotides (residues G2794 and G2793) of the ribonucleic acid (RNA), and the phosphate backbone were included in the calculation. The surrounding was approximated using the COSMO.[242] Previous studies have estimated the average dielectric constant inside a protein to be about $6 - 7$ and can reach up to about $20 - 30$ at the protein's surface.[248] Based on visual inspection of the model protein site, a dielectric constant of 10 was used. All calculations were performed on version TURBOMOLE 7.5.

The lissoclimide model was geometry optimized using the Tao–Perdew–Staroverov–Scuseria hybrid functional (TPSSh)[121,249] and dispersion corrections with the D3 method, denoted

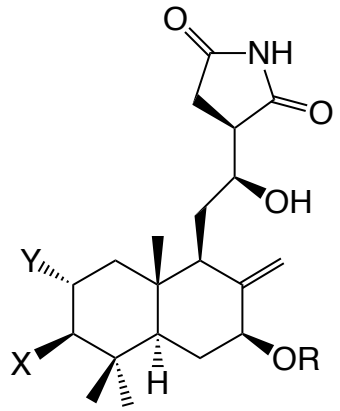
as TPSSh-D3.[86,250] The Karlsruhe segmented-contracted polarized split valence (def2-SVP) basis sets[130] and resolution-of-the-identity (RI) approximation[251] were chosen to explore the interactions between the lissoclimide inhibitors and guanine nucleotides. Various inhibitors were included with a similar structure as the CL, see Figure 5.1. These include the DCL, haterumaimide family, and substitution of the chlorine in CL with the halogen, cyano, trifluoromethyl, nitro, and methyl functional groups.

Tight convergences were applied for the geometry optimization. These include energy convergence criterion of 10^{-7} Hartree, 10^{-7} atomic units (au) for the root mean square change of the one-particle density matrix, the exchange-correlation employing gridsize 4,[122] and weight derivatives. All geometries were verified to be a minima on the potential energy surface by harmonic vibrational analysis.

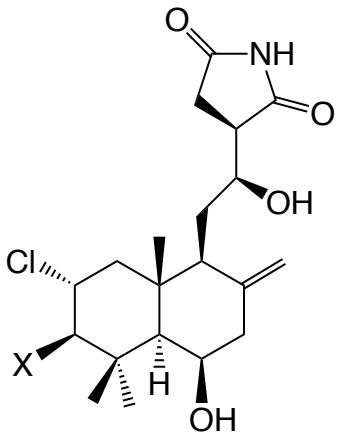
Structural analysis was performed by looking at the bond distances between the halogen functional group and guanine nucleotides as well as comparing the dihedral angles and root mean square deviation (RMSD) between the geometry optimized model and co-crystal structure. Dihedral angles were measured along the N9 and C3′ of G2393 to C5′ and N9 of G2394. These were chosen to capture the relative orientations of the guanine bases, see Figure 5.2. Bond distances were included for the interaction between the halogens and guanine bases. RMSD values were taken between the RNAs of the optimized model and co-crystal structure.

5.2.2 X40 Benchmark

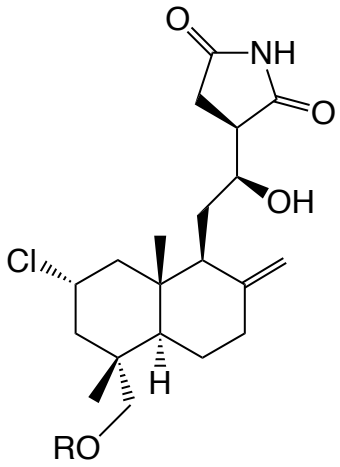
To verify appropriate electronic structure methods for predicting the halogen–π interaction, the random phase approximation (RPA) and the semilocal PBE[120] were benchmarked against the X40 testset.[247] Dispersion corrections to PBE were included known as D3[86,250] and D4.[93,123] This testset includes 40 halogen complexes ranging from $6-24$ atoms and reference values are based on the coupled cluster singles, doubles, and perturbative triples (CCSD(T))



Compound	X	Y	R
Dichlorolissoclimide (DCL, **1**)	Cl	Cl	H
Haterumaimide A (HatA, **2**)	Cl	Cl	Ac
Haterumaimide N (HatN, **3**)	H	Cl	Ac
Haterumaimide Q (HatQ, **4**)	H	H	H
Fluorolissoclimide (FL, **5**)	H	F	H
Chlorolissoclimide (CL, **6**)	H	Cl	H
Bromolissoclimide (BL, **7**)	H	Br	H
Iodolissoclimide (IL, **8**)	H	I	H
Cyanolissoclimide (CNL, **9**)	H	CN	H
Trifluoromethyllissoclimide (TFL, **10**)	H	CF_3	H
Nitrolissoclimide (NOL, **11**)	H	NO_2	H
Methyllissoclimide (CHL, **12**)	H	CH_3	H

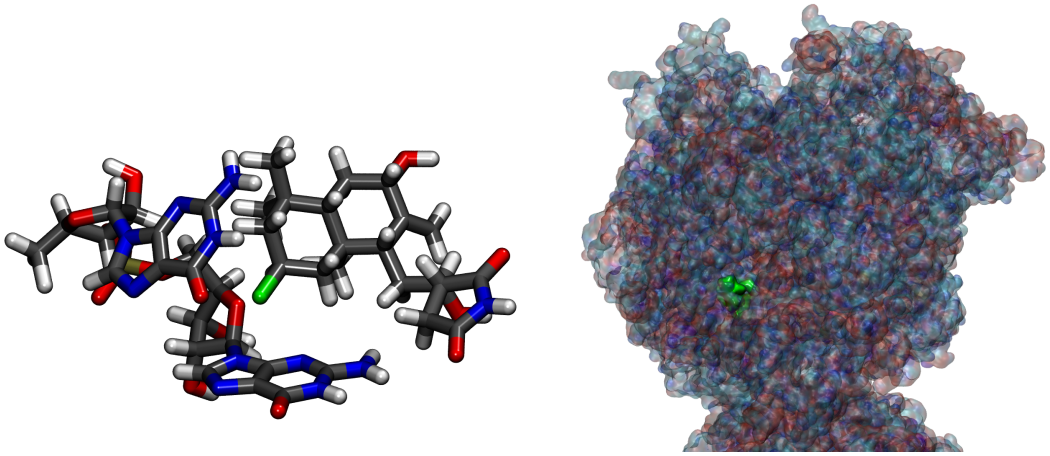


Compound	X
Haterumaimide E (HatE, **13**)	Cl
Haterumaimide F (HatF, **14**)	H



Compound	R
Haterumaimide J (HatJ, **15**)	H
Haterumaimide K (HatK, **16**)	Ac

Figure 5.1: Various natural and proposed inhibitors were computationally modeled in this computational study.



a) CL Model b) Co-crystal Structure

Figure 5.2: a) Proposed lissoclimide model to study the SAR. Color scheme: H - white, Cl - green, C - gray, P - brown, N - blue, and O - red. b) Region of interest is shown in green.

ranging from $0.5 - 15$ kcal/mol.

Interaction energies (ΔE) were computed using the supramolecular approach defined as,

$$\Delta E = E_{\mathrm{A}} + E_{\mathrm{B}} - E_{\mathrm{AB}} \tag{5.1}$$

where E_{AB} is the total energy of the complex while E_{A} and E_{B} are the total energies of the fragments. The choice of basis sets for RPA is based on previous work by Nguyen and coworkers[195] and summarized here. RPA energies are obtained in a post-Kohn–Sham (KS) fashion. The PBE reference orbitals[120] were chosen for RPA calculations, denoted as RPA(PBE). The reference orbitals were computed with quadrature grids of m5 quality[122] and converged at 10^{-7} Hartree and 10^{-7} au for the root mean square change of the one-particle density matrix. For the RPA energy calculations, imaginary frequency grids of 100 points were employed.[107]

The Karlsruhe segmented-contracted polarized quadruple-ζ (def2-QZVP) basis sets[130] were chosen for the expectation value of the KS determinant evaluated within the RI algorithm. The RPA correlation energies were evaluated using the Dunning's correlation-consistent po-

larized valence basis sets[132,133] and frozen core approximation. Basis set superposition error was estimated by 50% counterpoise (CP) correction as recommended by Risthaus and Grimme.[138] The complete basis set (CBS) limit of the RPA correlation energy was estimated using the two-point $1/X^3$ extrapolation, where $X = 3$ (triple-ζ), $X = 4$ (quadruple-ζ), etc.

Based on the convergences, it is recommended to include $(n-1)$d subvalence whenever appropriate. This is consistent with a previous studying showing that the $(n-1)$d subvalence shell can contribute up to 0.9 kcal/mol.[252] As for PBE dispersion corrected methods, the CBS limit is determined using 50% CP correction and def2-QZVP basis sets[130] based on Risthaus and Grimme's recommendation.[138]

5.2.3 Correlation and Potency Analysis of the Lissoclimide Model

The potency of the inhibitors were assessed based on the computed interaction energies between the inhibitor and guanine nucleotides. RPA, dispersion corrected semilocal Tao–Perdew–Staroverov–Scuseria (TPSS) and TPSSh were used to assess the drug's potency. RPA binding energies were computed with TPSS reference orbitals[121] and extrapolated to the CBS limit based on Section 5.2.2. The subvalence $(n-1)$d basis functions to the Dunning's correlation consistent basis sets were included whenever applicable, see Section 5.5.1. Moreover, dispersion corrected TPSS and TPSSh, denoted as TPSS-D3 and TPSSh-D3, binding energies were computed at def2-QZVP basis set[130] with 50% CP corrections.[138] A correlation plot is created between the predicted interaction energies and the reported half-maximum inhibitory concentration (IC_{50}) values.[253–258]

5.3 Results

5.3.1 X40 Interaction Energy Benchmark

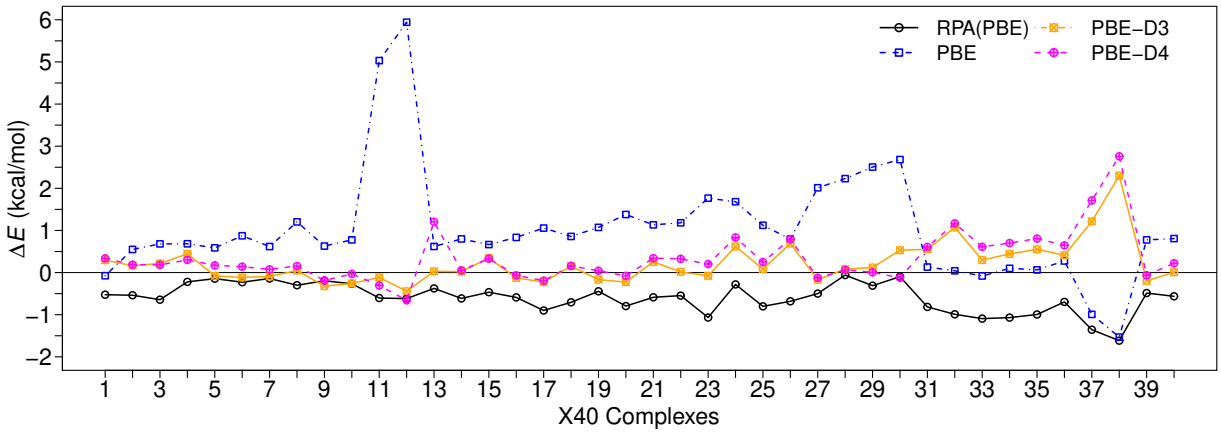


Figure 5.3: Reported binding energy errors (kcal/mol) for RPA(PBE), PBE, PBE-D3, and PBE-D4 against the X40 testset. All calculations are at the CBS limit. Positive values indicate overbinding.

To begin modeling these interactions, the X40 testset was chosen to analyze the performance of RPA(PBE) and dispersion-corrected semilocal PBE functional. The benchmark contains 40 halogenated complexes that include London dispersion (complexes $1-4$), induction (complexes $5-8$), dipole-dipole interaction (complexes $9-10$), $\pi-\pi$ stacking (complexes $11-12$), halogen bonds (complexes $13-26$), halogen-π interactions (complexes $27-30$), and hydrogen bonds (complexes $31-40$).[247]

For the binding energies of these complexes, RPA(PBE) underbinds in all cases, see Figure 5.3. The underbinding behavior has been observed in S66, L7, and S30L benchmarks.[195] The mean error (ME) and mean absolute error (MAE) are both 0.60 kcal/mol, see Table 5.1. Since the halogen-π interaction is of interest, the MAE of RPA(PBE) for complexes $27-30$ is 0.23 kcal/mol. The agreement between RPA(PBE) and the reference value is remarkable that supports the use of RPA(PBE) for computing the binding energies of the lissoclimide model.

Table 5.1: Mean errors (ME), mean absolute errors (MAE), absolute minimum-maximum error range (MinMax) and standard deviation (STDDEV) in kcal/mol of various methods at CBS limit. Positive ME corresponds to overbinding.

	RPA(PBE)	PBE	PBE-D3	PBE-D4
ME	0.60	1.04	0.34	0.21
MAE	0.60	1.17	0.43	0.34
MinMax	1.56	7.48	3.42	2.74
STDDEV	0.35	1.32	0.60	0.50

In comparison, semilocal PBE functional overbinds for nearly all complexes and performs the worst with MAE of 1.17 kcal/mol. Dispersion corrections PBE-D3 and PBE-D4 improve upon PBE with MAEs of 0.43 and 0.34 kcal/mol, respectively. While the improvements are expected with dispersion corrections, the absolute minimum-maximum error ranges (MinMax) for PBE, PBE-D3, and PBE-D4 are significantly larger than RPA(PBE). Furthermore, for halogen-π complexes, the MAEs of PBE-D3, and PBE-D4 are 0.08 kcal/mol and 0.23 kcal/mol, respectively. The results may suggest that the dispersion corrections are not systematic in the treatment of NIs since the X40 testset was not part of the training sets for parameter estimation.[87,93] There is some variability in the accuracy of the dispersion corrected semilocal density functional approximations (DFAs). This observation is consistent with a previous study by Kouzuch and Martin showing that dispersion corrected density functional approximations do not sufficiently predict the geometries and dissociation energies of halogen bonds.[259] In comparison, the smaller MinMax and standard deviation (STDDEV) from RPA(PBE) suggest that RPA(PBE) is robust and can confidently be used to predict reliable binding energies of halogenated complexes.

Moreover, complex 1 stood out for RPA(PBE), PBE, PBE-D3, and PBE-D4, see Figure 5.4. This complex contains methane interacting with fluorine and the CCSD(T) binding energy reference is 0.49 kcal/mol. PBE yielded the most accurate binding energy for this system within 5.5% error. Meanwhile, dispersion corrected PBE-D3 and PBE-D4 have percent errors of 58.9% and 48.7%, respectively. RPA(PBE) predicted that complex 1 does not bind

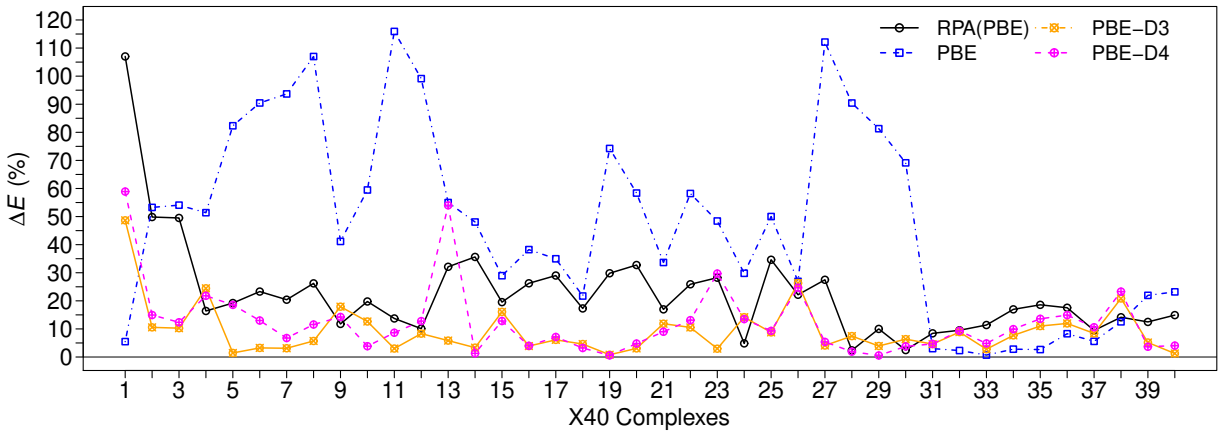


Figure 5.4: Reported binding energy percentage errors (%) for RPA(PBE), PBE, PBE-D3, and PBE-D4 against the X40 testset. All calculations are at the CBS limit.

and has an 107.0% error.

One potential reason for the large errors is that the complexes from the X40 testset were geometry optimized with the CP corrected second-order Møller–Plesset many-body perturbation theory (MP2) and triple-ζ basis sets.[247] Based on this geometry, the CCSD(T) binding energies are interpolated along an intermolecular axis and a new minimum distance was determined. MP2 geometries may be the source of the error since the halogen group is highly polarizable which cannot be accurately modeled with MP2.[195,260] This consideration may need to be taken into account and to date, there are few testsets that involve NIs of halogenated compounds.[111,247,259]

5.3.2 Lissoclimide Analogues

Table 5.2: Parameters of the linear regression lines displayed in Figure 5.5

Method	Slope	y-int	R^2
RPA(TPSS)	−0.796	29.8	0.627
TPSS-D3	−0.683	29.2	0.581
TPSSh-D3	−0.697	30.0	0.577

Evaluation of the lissoclimide model against the experimentally reported IC_{50} of the DCL,

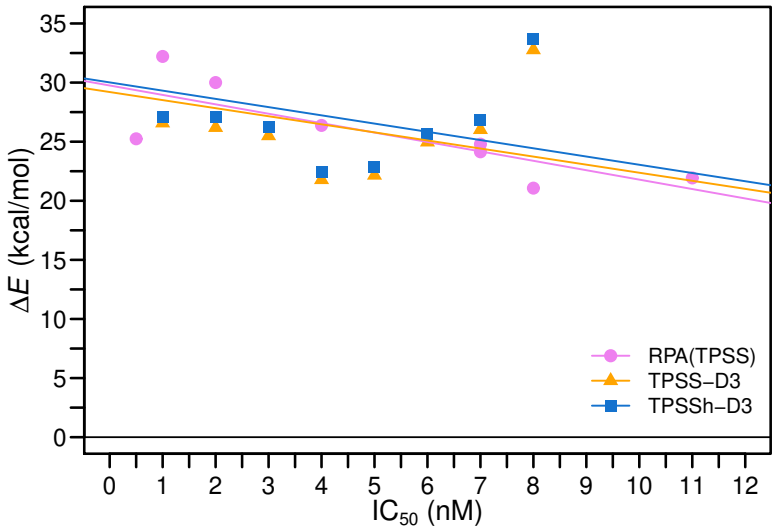
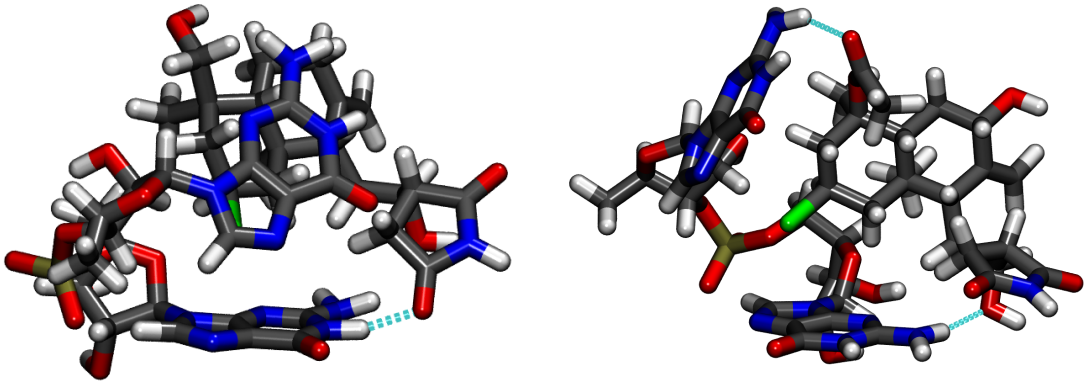


Figure 5.5: Correlation plot between the predicted RPA(TPSS), TPSS-D3, and TPSSh-D3 binding energies at CBS limit (ΔE) in kcal/mol and experimental IC_{50} in nM.[253–258]

CL, and haterumaimide family revealed a consistent agreement with the predicted binding energies. A correlation plot revealed a moderately negative correlation ($R^2 = 0.627$) between the predicted RPA(TPSS) binding energies and the IC_{50} of P388 murine leukaemia cells, see Figure 5.5 and Table 5.2. Furthermore, TPSS-D3 and TPSSh-D3 yielded similar trends where the binding energies are within $\sim$ 4 kcal/mol of the RPA(TPSS) results, see Section 5.5.2. Knowledge of the CL binding site from the X-ray co-crystal structure combined with these predicted binding energies has made it possible to expand the understanding of SAR in the lissoclimide family of translation-inhibition cytotoxins.

To further confirm the model, binding energies of DCL and CL suggest that DCL is the more potent cytotoxin consistent with the IC_{50}, see Table 5.3. While the correlation is promising, there are a few inconsistencies. The first is that haterumaimide J (HatJ) does not have the highest predicted binding energy contradicting the subnanomolar IC_{50}. Haterumaimide K (HatK) has the largest binding energy of 32.2 kcal/mol but marginally higher IC_{50} than HatJ by 0.5 nM. Upon visualization of the optimized HatJ and HatK models, it revealed that HatK forms two hydrogen bonds with G2793 and G2794 as compared to HatJ only forming

a hydrogen bond with G2794. The acetyl group of HatK seems to play a role providing the additional stabilization.



a) HatJ b) HatK

Figure 5.6: Visualization of the optimized HatJ and HatK model showing the presence of hydrogen bonds (cyan). The cutoff for hydrogen bonds is within 4 Å. (Color scheme: H - white, Cl - green, C - gray, P - brown, N - blue, and O - red)

Another notable discrepancy is the haterumaimide Q (HatQ) yielding a similar binding energy as compared to haterumaimide A (HatA) and haterumaimide N (HatN), see Table 5.3. Yet, HatQ has an IC_{50} of 120 nM compared to HatA and HatN with both having IC_{50} of 7 nM. Another indicator for potency is to look at the RMSD value for HatQ. It is observed to have the largest RMSD of 5.08 Å. This coarse correlation between predicted binding energies and IC_{50} showed that structural data may need to be considered.

Excluding the discrepancies, structural analysis for this set remains fairly consistent. For instance, RMSD values of the DCL, CL, and the haterumaimide family range between $\sim 4-5$ Å, see Table 5.3. Dihedral angle was chosen to observe the relative orientations of the guanine nucleotides. This included the N9 and C3′ of G2793 to C5′ and N9 of G2794. Based on the geometry optimized structures, the dihedral angles deviate up to $\sim 30°$ relative to the crystal structure. Given these results, the lissoclimide model provides confidence that this is an appropriate model to explore the face-on geometry of the halogen-π interaction and propose potential candidates for cancer therapeutics.

Table 5.3: Calculated RPA(TPSS) binding energies at CBS limit (ΔE) in kcal/mol and root means square deviation (RMSD) in Å relative to the guanine nucleotides with the phosphate backbone from the co-crystal structure. Cytotoxicity towards P388 murine leukaemia cells (IC_{50}) are compiled from Malochet-Grivois *et al.*, Uddin *et al.*, and Fu et al..[253–258] Dihedral angles (τ) in degrees are measured along N9 and C3′ of G2793 to C5′ and N9 of G2794.

Compound	ΔE (kcal/mol)	RMSD (Å)	τ (°)	P388 IC_{50} (nM)
Crystal	–	–	50.1	–
DCL	30.0	4.67	65.8	2
FL	15.0	4.14	64.2	–
CL	26.4	4.40	70.5	4
BL	24.1	4.33	70.6	–
IL	26.3	4.15	75.9	–
CNL	22.6	3.36	73.4	–
TFL	23.2	4.43	73.3	–
NOL	17.5	4.38	55.9	–
CHL	25.9	4.32	77.5	–
HatA	24.8	4.35	64.4	7
HatN	24.1	3.94	66.2	7
HatQ	20.8	5.08	69.7	120
HatE	21.1	4.48	65.8	8
HatF	21.9	4.53	63.4	11
HatJ	25.2	4.09	69.6	0.5
HatK	32.2	4.24	80.5	1

The face-on halogen-π interaction has been documented for protein-ligand interactions. These included aromatic side chains of Phe, Tyr, Trp, and His.[236] Most recently, Könst and coworkers discovered one of the first face-on interactions between the halogen and nucleotide base.[235] Little is known about this interaction. With the model, the chlorine from the CL was substituted with fluorine, bromine, and iodine to predict potential trends. The fluorolissoclimide (FL) appears to be the least potent out of the four halogens and has a predicted binding energy of 15.0 kcal/mol, see Table 5.3.

A potential reason for the small binding energy from FL may be due to fluorine being the least polarizable out of the halogen group. It has been known in the past that the chemical and electronic environment can make the fluorine polarizable.[261–264] One route is to have additional electron withdrawing groups that could affect the intermolecular contacts involving fluorine. To improve the binding energy, the trifluoromethyllissoclimide (TFL) was proposed and yielded a significantly stronger binding energy of 23.2 kcal/mol, see Table 5.3. The dihedral angle of TFL increased to 73.3° which is within the range of the CL, bromolissoclimide (BL), and iodolissoclimide (IL). In addition, the bond distances between the fluorines and guanine nucleobases are within the range as the FL.

In comparison, the binding energies of CL, BL, and IL were 26.4 kcal/mol, 24.1 kcal/mol, and 26.3 kcal/mol, respectively. Bond distances between the halogens and the guanine residues increased with increasing radii. Comparison to the crystal structure, the CL model bond distances are smaller up to 0.5 Å, see Table 5.4. Here, the chemical environment could affect these structural differences and adjusting the dielectric constant parameters can achieve bond distances qualitatively closer to the crystal structure, see Section 5.5.3. This is a limitation of the model since the dielectric constant of the protein can fluctuate depending on the environment.[248,265,266]

Based on the expanded knowledge of the SAR in the lissoclimide family, cyanolissoclimide (CNL), nitrolissoclimide (NOL) and the methyllissoclimide (CHL) are potential drug candi-

Table 5.4: Bond distances in Å from the halogen (X = F, Cl, Br, and I) to the guanine residues (G2793 and G2794) are reported.

Compound	X–G2793 (Å)	X–G2794 (Å)
Crystal	3.44	3.66
FL	2.91	2.91
CL	3.06	3.20
BL	3.38	3.38
IL	3.53	3.62
TFL	3.09, 2.97	2.86
DCL	3.12, 3.39	3.33, 5.34

dates for translation inhibition. Cyano, nitro, and methyl functional groups have a diverse set of biological functions. These can help to polarize the chemical environment[267] and in other cases, they can be a key component for molecular recognition.[268–271] The motivation to choose the cyano and nitro functional groups is the known $\pi - \pi$ interactions capabilities.[272–274] For methyl, CH-π have been detected in proteins[275] and is dominated by dispersion.[270,276,277] The predicted binding energies of CNL, NOL, and CHL are 22.6 kcal/mol, 17.5 kcal/mol, and 25.9 kcal/mol, respectively. Out of the three, NOL is most likely the least potent. CNL and CHL present themselves as potential drug inhibitors with binding energies that are within the range of other potent lissoclimide analogs.

5.4 Conclusions

With the expanded understanding of SAR in the lissoclimide family of translation-inhibition cytostatic agents, a computational approach is developed that provides an approach and tool for rational drug design. Rather than modeling the complete X-ray co-crystal structure ($\sim 410,000$ atoms), the isolation of the region of interest may be adequate to build on top of the SAR knowledge. Here, an improved model totaling 122 atoms has been introduced that includes the lissoclimide analogs, guanine nucleotides, and the phosphate backbone. Approximating the protein environment with COSMO allow greater understanding of SAR

in the lissoclimide family.

While the correlation between the predicted binding energy and the IC_{50} was moderate, the model did allow greater understanding of the structural effects from these functional groups. For instance, Konst and coworkers anticipated a synthesized compound featuring both DCL and HatJ to have a greater potency but did not.[235] From the model, HatJ and HatK have close IC_{50} values of 0.5 nM and 1 nM, respectively. However, HatK has a significantly stronger binding energy and this is potentially attributed to being able to form two hydrogen bonds with G2793 and G2974. This suggests that the new compound should combine features from DCL and HatK. It may be worth exploring this possiblity. Keep in mind, the model does not include all amino acids surrounding the lissoclimide derivatives which allows greater freedom for molecules to move.

Furthermore, TFL, CHL, and CNL have presented themselves as potential inhibitors. This is based on understanding the dispersion interactions with the conjugated-π of the guanine nucleobases. These interactions cannot easily be modeled with force fields. Electronic structure methods are required in this instance. RPA provides adequate confidence to capture the trends between the IC_{50} and the binding energies. The sensitivity to differentiate the potency of DCL and CL is observed where semilocal TPSS-D3 and TPSSh-D3 are unable to, see Section 5.5.2.

Moreover, RPA(PBE), PBE, PBE-D3, and PBE-D4 were benchmarked against the X40 testset showing that RPA has a smaller MinMax and STDDEV. Whereas, PBE and dispersion corrected PBE do not systematically lead to greater accuracy with larger MinMax and STDDEV. This is remarkable for RPA, which is relatively parameter-free aside from using a KS reference from a semilocal DFA. Most modern RPA implementations are comparable to the second order Møller–Plessett many-body perturbation theory (MP2). In addition, MP2 is not adequate for polarizable molecules. This is indicated by the X40 testset which is geometry optimized with MP2. For instance, RPA, PBE-D3, and PBE-D4 yielded an

unusually large error for compound 1. It may be worth developing additional halogen bond benchmarks since halogens are prominent in drug design.

Altogether, the combination of the lissoclimide model and electronic structure methods have permitted the study of the SAR and expanded the understanding of the structural basis for binding energies in the lissoclimide family. The computational study has shown the possibility of using RPA to model dispersion interactions and provides an approach that focuses on the region of interest without needing to completely model the co-crystal structure of the lissoclimide family and 80S ribosome.

5.5 Supporting Information

5.5.1 Basis Set Convergence of X40 Testset and DCL Model

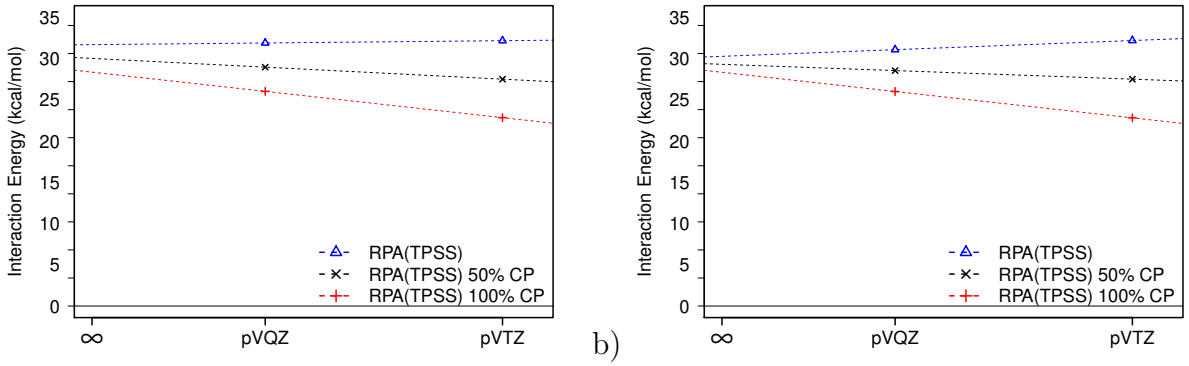


Figure S5.1: Basis set convergence of the RPA interaction energy using a TPSS KS reference for the DCL model. Dunning's correlation consistent basis sets are compared with and without $(n-1)$d orbital contribution in part a) and b), respectively. X denotes the cardinal number of Dunning's correlation consistent polarized cc-pVXZ (X=T,Q) basis sets.[132,133]

The importance of the $(n-1)$d subvalence basis functions is confirmed for studying the halogen-π interactions of moderately sized systems. The binding energies of the DCL model were computed using the TPSS KS reference where the expectation value of the KS deter-

minant is evaluated with the def2-QZVP basis sets.[130] The RPA correlation energies are computed using the frozen core approximation and the Dunning's basis sets with and without the $(n-1)$d subvalence basis functions. In Figure S5.1, the CBS limit shows that the $(n-1)$d orbitals contribute ~ 1 kcal/mol to the binding energy. This has been noted by Kesharwani *et al*[252] for the binding energies of halogenated complexes.

5.5.2 Binding Energies of Lissoclimide Analog from Dispersion Corrected DFAs

Table S5.1: Binding energies of lissoclimide derivatives in kcal/mol for RPA(TPSS), TPSS-D3, and TPSSh-D3.

Drug	RPA(TPSS)	TPSS-D3	TPSSh-D3
DCL	30.0	26.6	27.1
FL	15.0	15.5	16.1
CL ($\epsilon = 7$)	19.0	19.7	20.3
CL ($\epsilon = 10$)	26.4	26.2	27.1
BL	24.1	24.9	25.7
IL	26.3	26.4	27.3
CN	22.6	22.7	23.4
TFL	23.2	21.7	22.5
NL	17.5	15.7	16.3
CHL	25.9	27.1	27.8
HatA	24.8	25.5	26.2
HatE	21.1	21.8	22.5
HatF	21.9	22.2	22.8
HatJ	25.2	26.0	26.8
HatN	24.1	25.0	25.7
HatQ	20.8	22.0	22.6
HatK	32.2	32.8	33.7

Comparison of the correlation between the IC_{50} and binding energies for the dispersion corrected TPSS and TPSSh,[86,250] denoted as TPSS-D3 and TPSSh-D3, suggests that the model is adequate to predict the drug's potency. The TPSS-D3 and TPSSh-D3 binding energies of the lissoclimide analogs were computed with def2-QZVP basis sets,[130] energy convergence criterion of 10^{-7} Hartree, 10^{-7} atomic units (au) for the root mean square

change of the one-particle density matrix, and the exchange-correlation employing gridsize 4.[122] Basis sets superposition error is corrected using 50% CP as recommended by Risthaus and Grimme.[138]

RPA(TPSS), TPSS-D3, and TPSSh-D3 show close agreement up to $\sim$ 4 kcal/mol, see Table S5.1. Furthermore, binding energies of CL model at different dielectric constants are consistent between the methods. Moreover, a closer inspection of binding energies between CL and DCL reveals that RPA(TPSS) is able to differentiate the potency between CL and CL. Meanwhile, TPSS-D3 and TPSSh-D3 showed near identical binding energies between CL and DCL. Taking these results together, the lissoclimide model along with using RPA(TPSS) suggests that this is an appropriate approach and predictor for expanding the lissoclimide family for protein synthesis inhibition.

5.5.3 Effect of Dielectric Constant on the Geometry Optimization of CL

Table S5.2: CL model was optimized at different dielectric constants (ϵ). Predicted RPA(TPSS) binding energies at CBS limit (ΔE) in kcal/mol, RMSD in Å relative to the crystal structure guanine nucleotides and phosphate backbone, and bond distances in Å from the chlorine (Cl) to the guanine residues (G2793 and G2794) are reported. Dihedral angles (τ) in degrees are measured along N9 and C3′ of G2793 to C5′ and N9 of G2794.

ϵ	ΔE (kcal/mol)	RMSD (Å)	Cl–G2793 (Å)	Cl–G2794 (Å)	τ (°)
7	19.0	4.40	3.28	3.28	62.52
10	26.4	4.40	3.06	3.20	70.47

The optimal choice of dielectric constant for proteins has been an outstanding problem.[278] The values can vary depending on the protein environment. Here, the dielectric constants for the geometry optimization of the CL model were explored for the values of 7 and 10. The binding energies differ up to 7.4 kcal/mol and are significantly be influenced by the dielectric constant, see Table S5.2. The closer Cl-guanine bond distances with dielectric

constant of 10 may contribute to the larger binding energy. RMSD relative to the RNA backbone of the crystal yields exactly the same value. An overlay of the optimized CL models at dielectric constants 7 and 10 have visually similar orientation relative to the crystal structure, see Figure S5.2. Noting some of thsee differences may suggest that there is a dependence on polarization effects in the local environment of the halogen.[241,279] Based on these observations, all geometry optimizations were performed with dielectric constant of 10 for consistency.

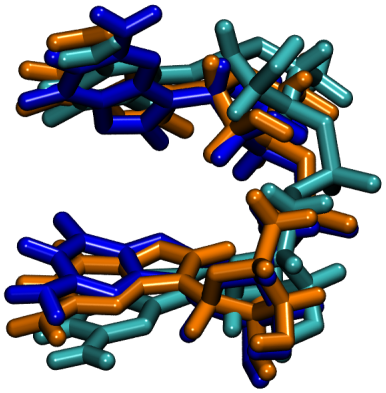


Figure S5.2: The RNAs from the crystal structure (cyan) and optimized CL models at dielectric constants 7 (orange) and 10 (blue) are superimposed. RMSD values of the CL models relative to the crystal structure are both 4.40 Å.

Bibliography

[1] McClements, D. J. Non-covalent interactions between proteins and polysaccharides. *Biotechnol. Adv.* **2006**, *24*, 621–625.

[2] Liu, K.; Kang, Y.; Wang, Z.; Zhang, X. 25th Anniversary Article: Reversible and Adaptive Functional Supramolecular Materials: "Noncovalent Interaction" Matters. *Adv. Mat.* **2013**, *25*, 5530–5548.

[3] Georgakilas, V.; Tiwari, J. N.; Kemp, K. C.; Perman, J. A.; Bourlinos, A. B.; Kim, K. S.; Zboril, R. Noncovalent Functionalization of Graphene and Graphene Oxide for Energy Materials, Biosensing, Catalytic, and Biomedical Applications. *Chem. Rev.* **2016**, *116*, 5464–5519.

[4] Hollingsworth, S. A.; Nguyen, B. D.; Chreifi, G.; Arce, A. P.; Poulos, T. L. Insights into the Dynamics and Dissociation Mechanism of a Protein Redox Complex Using Molecular Dynamics. *J. Chem. Info. Model.* **2017**, *57*, 2344–2350.

[5] Fields, J. B.; Hollingsworth, S. A.; Chreifi, G.; Heyden, M.; Arce, A. P.; Magaña-Garcia, H. I.; Poulos, T. L.; Tobias, D. J. "Bind and Crawl" Association Mechanism of Leishmania major Peroxidase and Cytochrome c Revealed by Brownian and Molecular Dynamics Simulations. *Biochemistry* **2015**, *54*, 7272–7282.

[6] Fa, S.; Sakata, Y.; Akine, S.; Ogoshi, T. Non-Covalent Interactions Enable the Length-Controlled Generation of Discrete Tubes Capable of Guest Exchange. *Angew. Chem.* **2020**, *132*, 9395–9399.

[7] Tkatchenko, A.; Alfè, D.; Kim, K. S. First-Principles Modeling of Non-Covalent Interactions in Supramolecular Systems: The Role of Many-Body Effects. *J. Chem. Theory Comput.* **2012**, *8*, 4317–4322.

[8] Huang, Z.; Qin, K.; Deng, G.; Wu, G.; Bai, Y.; Xu, J.-F.; Wang, Z.; Yu, Z.; Scherman, O. A.; Zhang, X. Supramolecular Chemistry of Cucurbiturils: Tuning Cooperativity with Multiple Noncovalent Interactions from Positive to Negative. *Langmuir* **2016**, *32*, 12352–12360.

[9] Autumn, K.; Sitti, M.; Liang, Y. A.; Peattie, A. M.; Hansen, W. R.; Sponberg, S.; Kenny, T. W.; Fearing, R.; Israelachvili, J. N.; Full, R. J. Evidence for van der Waals adhesion in gecko setae. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12252–12256.

[10] London, F. The general theory of molecular forces. *Trans. Faraday Soc.* **1937**, *33*, 8b–26.

[11] Moszynski, R.; Jeziorski, B.; Ratkiewicz, A.; Rybak, S. Many-body perturbation theory of electrostatic interactions between molecules: Comparison with full configuration interaction for four-electron dimers. *J. Chem. Phys.* **1993**, *99*, 8856–8869.

[12] Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

[13] Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466–3470.

[14] Moszynski, R.; Jeziorski, B.; Szalewicz, K. Møller–Plesset expansion of the dispersion energy in the ring approximation. *Int. J. Quantum Chem.* **1993**, *45*, 409–431.

[15] Pople, J. A.; Binkley, J. S.; Seeger, R. Theoretical models incorporating electron correlation. *Int. J. Quantum Chem.* **1976**, *10*, 1–19.

[16] Pople, J. A.; Head-Gordon, M.; Raghavachari, K. Quadratic configuration interaction. A general technique for determining electron correlation energies. *J. Chem. Phys.* **1987**, *87*, 5968–5975.

[17] Bartlett, R. J.; Shavitt, I. Determination of the size-consistency error in the single and double excitation configuration interaction model. *Int. J. Quantum Chem.* **1977**, *12*, 165–173.

[18] Janowski, T.; Ford, A. R.; Pulay, P. Accurate correlated calculation of the intermolecular potential surface in the coronene dimer. *Mol. Phys.* **2010**, *108*, 249–257.

[19] Janowski, T.; Pulay, P. A benchmark quantum chemical study of the stacking interaction between larger polycondensed aromatic hydrocarbons. *Theo. Chem. Acc.* **2011**, *130*, 419–427.

[20] Gabdoulline, R. R.; Wade, R. C. Protein-protein association: investigation of factors influencing association rates by Brownian dynamics simulations1 1Edited by B. Honig. *J. Mol. Biol.* **2001**, *306*, 1139–1155.

[21] Gould, T.; Toulouse, J.; Ángyán, J. G.; Dobson, J. F. Casimir–Polder Size Consistency: A Constraint Violated by Some Dispersion Theories. *J. Chem. Theory Comput.* **2017**, *13*, 5829–5833.

[22] Dobson, J. F.; Ángyán, J. G.; Gould, T. Does the exchange–correlation kernel fxc have a very long-ranged dependence on the groundstate electron density? *Theor. Chem. Acc.* **2018**, *137*, 167.

[23] Heßelmann, A. Improved supermolecular second order Møller–Plesset intermolecular interaction energies using time-dependent density functional response theory. *J. Chem. Phys.* **2008**, *128*, 144112.

[24] Pitoňák, M.; Heßelmann, A. Accurate Intermolecular Interaction Energies from a Combination of MP2 and TDDFT Response Theory. *J. Chem. Theory Comput.* **2010**, *6*, 168–178.

[25] Poulos, T. L. Heme Enzyme Structure and Function. *Chem. Rev.* **2014**, *114*, 3919–3962.

[26] Winkler, J. R.; Gray, H. B. Electron Flow through Metalloproteins. *Chem. Rev.* **2014**, *114*, 3369–3380.

[27] Pelletier, H.; Kraut, J. Crystal Structure of a Complex Between Electron Transfer Partners, Cytochrome c Peroxidase and Cytochrome c. *Science* **1992**, *258*, 1748–1755.

[28] Volkov, A. N.; Nicholls, P.; Worrall, J. A. The complex of cytochrome c and cytochrome c peroxidase: The end of the road? *Biochim. Biophys. Acta. Bioenerg.* **2011**, *1807*, 1482–1503.

[29] Adak, S.; Datta, A. Leishmania major encodes an unusual peroxidase that is a close homologue of plant ascorbate peroxidase: a novel role of the transmembrane domain. *Biochem. J.* **2005**, *390*, 465–474.

[30] Chreifi, G.; Hollingsworth, S. A.; Li, H.; Tripathi, S.; Arce, A. P.; Magaña-Garcia, H. I.; Poulos, T. L. Enzymatic Mechanism of Leishmania major Peroxidase and the Critical Role of Specific Ionic Interactions. *Biochemistry* **2015**, *54*, 3328–3336.

[31] Jasion, V. S.; Doukov, T.; Pineda, S. H.; Li, H.; Poulos, T. L. Crystal structure of the ¡i¿Leishmania major¡/i¿ peroxidase–cytochrome ¡i¿c¡/i¿ complex. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 18390–18394.

[32] Jasion, V. S.; Poulos, T. L. Leishmania major Peroxidase Is a Cytochrome c Peroxidase. *Biochemistry* **2012**, *51*, 2453–2460.

[33] Coulson, A. F.; Yonetani, T. Oxidation of cytochrome c peroxidase with hydrogen peroxide: Identification of the 'Endogenous donor'. *Biochem. Biophys. Res. Commun.* **1972**, *49*, 391–398.

[34] Sivaraja, M.; Goodin, D. B.; Smith, M.; Hoffman, B. M. Identification by ENDOR of Trp^{191} as the Free-Radical Site in Cytochrome c Peroxidase Compound ES. *Science* **1989**, *245*, 738–740.

[35] Liu, R.-Q.; Miller, M. A.; Han, G. W.; Hahm, S.; Geren, L.; Hibdon, S.; Kraut, J.; Durham, B.; Millett, F. Role of methionine 230 in intramolecular electron transfer between the oxyferryl heme and tryptophan 191 in cytochrome c peroxidase compound II. *Biochemistry* **1994**, *33*, 8678–8685.

[36] Miller, M. A. A Complete Mechanism for Steady-State Oxidation of Yeast Cytochrome c by Yeast Cytochrome c Peroxidase†. *Biochemistry* **1996**, *35*, 15791–15799.

[37] Kanh, C. H.; Brautigan, D. L.; Osheroff, N.; Margoliash, E. Definition of cytochrome c binding domains by chemical modification. reaction of carboxydinitrophenyl and trinitrophenyl cytochromes c with baker's yeast cytochrome c peroxidase. *J. Bio. Chem.* **1978**, *253*, 6502–6510.

[38] VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

[39] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem. 26*, 1781–1802.

[40] MacKerell, A. D. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

[41] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

[42] Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

[43] Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

[44] Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Mol. Simul.* **1991**, *6*, 121–142.

[45] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

[46] Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

[47] Collins, J. R.; Loew, G. H. Comparison of computational models for simulating heme proteins: A study of cytochrome C peroxidase. *Int. J. Quantum Chem. 44*, 87–107.

[48] Marcus, R. A. On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. I. *J. Chem. Phys.* **1956**, *24*, 966–978.

[49] Marcus, R. A.; Sutin, N. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta. Bioenerg.* **1985**, *811*, 265–322.

[50] Wang, K.; Mei, H.; Geren, L.; Miller, M. A.; Saunders, A.; Wang, X.; Waldner, J. L.; Pielak, G. J.; Durham, B.; Millett, F. Design of a Ruthenium-Cytochrome c Derivative To Measure Electron Transfer to the Radical Cation and Oxyferryl Heme in Cytochrome c Peroxidase. *Biochemistry* **1996**, *35*, 15107–15119.

[51] McLendon, G. Control of biological electron transport via molecular recognition and binding: The "velcro" model. Long-Range Electron Transfer in Biology. Berlin, Heidelberg, 1991; pp 159–174.

[52] Schilder, J.; Ubbink, M. Formation of transient protein complexes. *Curr. Opin. Struct. Biol.* **2013**, *23*, 911–918.

[53] Volkov, A. N. Structure and Function of Transient Encounters of Redox Proteins. *Acc. Chem. Res.* **2015**, *48*, 3036–3043.

[54] Hobza, P.; Zahradník, R.; Müller-Dethlefs, K. The World of Non-Covalent Interactions: 2006. *Collect. Czech. Chem. Commun.* **2006**, *71*, 443–531.

[55] Dobson, J. F.; Gould, T. Calculation of dispersion energies. *J. Phys. Condens. Matter* **2012**, *24*, 073201.

[56] Dobson, J. F. Beyond pairwise additivity in London dispersion interactions. *Int. J. Quantum Chem.* **2014**, *114*, 1157–1161.

[57] Axilrod, B. M.; Teller, E. Interaction of the van der Waals Type Between Three Atoms. *J. Chem. Phys.* **1943**, *11*, 299–300.

[58] Muto, Y. Force between nonpolar molecules. *J. Phys. Math. Soc. Jpn* **1943**, *17*, 629–631.

[59] Biedermann, F.; Schneider, H.-J. Experimental Binding Energies in Supramolecular Complexes. *Chem. Rev.* **2016**, *116*, 5216–5300.

[60] Britz, D. A.; Khlobystov, A. N. Noncovalent interactions of molecules with single walled carbon nanotubes. *Chem. Soc. Rev.* **2006**, *35*, 637–659.

[61] Moulton, B.; Zaworotko, M. J. From Molecules to Crystal Engineering: Supramolecular Isomerism and Polymorphism in Network Solids. *Chem. Rev.* **2001**, *101*, 1629–1658.

[62] Beran, G. J. O. Modeling Polymorphic Molecular Crystals with Electronic Structure Theory. *Chem. Rev.* **2016**, *116*, 5567–5613.

[63] Becucci, M.; Melandri, S. High-Resolution Spectroscopic Studies of Complexes Formed by Medium-Size Organic Molecules. *Chem. Rev.* **2016**, *116*, 5014–5037.

[64] Zuend, S.; Jacobsen, E. N. Mechanism of Amido-Thiourea Catalyzed Enantioselective Imine Hydrocyanation: Transition State Stabilization via Multiple Non-Covalent Interactions. *J. Am. Chem. Soc.* **2009**, *131*, 15358–15374.

[65] Tao, D. J.; Muuronen, M.; Slutskyy, Y.; Le, A.; Furche, F.; Overman, L. E. Diastereoselective Coupling of Chiral Acetonide Trisubstituted Radicals with Alkenes. *Chem. Eur. J.* **2016**, *22*, 8786–8790.

[66] Bartlett, R. J. Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359–401.

[67] Hirata, S. Thermodynamic limit and size-consistent design. *Theor. Chem. Acc.* **2011**, *129*, 727–746.

[68] Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

[69] Kristyán, S.; Pulay, P. Can (semi)local density functional theory account for the London dispersion forces? *Chem. Phys. Lett.* **1994**, *229*, 175–180.

[70] Pérez-Jordá, J.; Becke, A. D. A density-functional study of van der Waals forces: rare gas diatomics. *Chem. Phys. Lett.* **1995**, *233*, 134 – 137.

[71] Hobza, P.; Šponer, J.; Reschel, T. Density functional theory and molecular clusters. *J. Comput. Chem.* **1995**, *16*, 1315–1325.

[72] Antony, J.; Grimme, S. Is Spin-Component Scaled Second-Order Møller–Plesset Perturbation Theory an Appropriate Method for the Study of Noncovalent Interactions in Molecules? *J. Phys. Chem. A* **2007**, *111*, 4862–4868.

[73] Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. Scaled MP3 Non-Covalent Interaction Energies Agree Closely with Accurate CCSD(T) Benchmark Data. *ChemPhysChem* **2009**, *10*, 282–289.

[74] Riley, K. E.; Platts, J. A.; Řezáč, J.; Hobza, P.; Hill, J. G. Assessment of the Performance of MP2 and MP2 Variants for the Treatment of Noncovalent Interactions. *J. Phys. Chem. A* **2012**, *116*, 4159–4169.

[75] Olsen, J.; Christiansen, O.; Koch, H.; Jørgensen, P. Surprising cases of divergent behavior in Møller–Plesset perturbation theory. *J. Chem. Phys.* **1996**, *105*, 5082–5090.

[76] Cremer, D.; He, Z. Sixth-Order Møller–Plesset Perturbation Theory–On the Convergence of the MPn Series. *J. Phys. Chem.* **1996**, *100*, 6173–6188.

[77] Leininger, M. L.; Allen, W. D.; Schaefer, H. F.; Sherrill, C. D. Is Mo/ller–Plesset perturbation theory a convergent ab initio method? *J. Chem. Phys.* **2000**, *112*, 9213–9222.

[78] Werner, H.-J. Eliminating the domain error in local explicitly correlated second-order Møller–Plesset perturbation theory. *J. Chem. Phys.* **2008**, *129*, 101103.

[79] Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151 – 154.

[80] Schütz, M.; Hetzer, G.; Werner, H.-J. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *J. Chem. Phys.* **1999**, *111*, 5691–5705.

[81] Manzer, S.; Horn, P. R.; Mardirossian, N.; Head-Gordon, M. Fast, accurate evaluation of exact exchange: The occ-RI-K algorithm. *J. Chem. Phys.* **2015**, *143*, 024113.

[82] Koch, H.; Sánchez de Merás, A.; Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J. Chem. Phys.* **2003**, *118*, 9481–9484.

[83] Witte, J.; Mardirossian, N.; Neaton, J. B.; Head-Gordon, M. Assessing DFT-D3 Damping Functions Across Widely Used Density Functionals: Can We Do Better? *J. Chem. Theory Comput.* **2017**, *13*, 2043–2052.

[84] Hohenstein, E. G.; Sherrill, C. D. Wavefunction methods for noncovalent interactions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 304–326.

[85] Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

[86] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

[87] Grimme, S. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chem. Eur. J.* **2012**, *18*, 9955–9964.

[88] Grimme, S. Improved Second-Order Møller–Plesset Perturbation Theory by Separate Scaling of Parallel- and Antiparallel-Spin Pair Correlation Energies. *J. Chem. Phys.* **2003**, *118*, 9095.

[89] Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. Scaled opposite-spin second order Møller–Plesset correlation energy: An economical electronic structure method. *J. Chem. Phys.* **2004**, *121*, 9793–9802.

[90] Huang, Y.; Shao, Y.; Beran, G. J. O. Accelerating MP2C dispersion corrections for dimers and molecular crystals. *J. Chem. Phys.* **2013**, *138*, 224112.

[91] Řezáč, J.; Greenwell, C.; Beran, G. J. O. Accurate Noncovalent Interactions via Dispersion-Corrected Second-Order Møller-Plesset Perturbation Theory. *J. Chem. Theory Comput.* **2018**, *14*, 4711–4721.

[92] Greenwell, C.; McKinley, J. L.; Zhang, P.; Zeng, Q.; Sun, G.; Li, B.; Wen, S.; Beran, G. J. O. Overcoming the difficulties of predicting conformational polymorph energetics in molecular crystals via correlated wavefunction methods. *Chem. Sci.* **2020**, *11*, 2200–2214.

[93] Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 dispersion coefficient model. *J. Chem. Phys.* **2017**, *147*, 034112.

[94] Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.

[95] Dobson, J. F.; White, A.; Rubio, A. Asymptotics of the Dispersion Interaction: Analytic Benchmarks for van der Waals Energy Functionals. *Phys. Rev. Lett.* **2006**, *96*, 073201.

[96] Lebègue, S.; Harl, J.; Gould, T.; Ángyán, J. G.; Kresse, G.; Dobson, J. F. Cohesive properties and asymptotics of the dispersion interaction in graphite by the random phase approximation. *Phys. Rev. Lett.* **2010**, *105*, 196401.

[97] Langreth, D. C.; Perdew, J. P. The Exchange-Correlation Energy of a Metallic Surface. *Solid State Commun.* **1975**, *17*, 1425–1429.

[98] Langreth, D. C.; Perdew, J. P. Exchange-correlation energy of a metallic surface: Wavevector analysis. *Phys. Rev. B* **1977**, *15*, 2884–2901.

[99] Hult, E.; Rydberg, H.; Lundqvist, B. I.; Langreth, D. C. Unified treatment of asymptotic van der Waals forces. *Phys. Rev. B* **1999**, *59*, 4708–4713.

[100] Gell-Mann, M.; Brueckner, K. A. Correlation Energy of an Electron Gas at High Density. *Phys. Rev.* **1957**, *106*, 364–368.

[101] Dobson, J. F.; Wang, J.; Dinte, B. P.; McLennan, K.; Le, H. M. Soft cohesive forces. *Int. J. Quantum Chem.* **2005**, *101*, 579–598.

[102] Eshuis, H.; Furche, F. A Parameter-Free Density Functional That Works for Noncovalent Interactions. *J. Phys. Chem. Lett.* **2011**, *2*, 983–989.

[103] Lu, D.; Nguyen, H.-V.; Galli, G. Power series expansion of the random phase approximation correlation energy: The role of the third- and higher-order contributions. *J. Chem. Phys.* **2010**, *133*, 154110.

[104] Harl, J.; Kresse, G. Accurate bulk properties from approximate many-body techniques. *Phys. Rev. Lett.* **2009**, *103*, 056401.

[105] Heßelmann, A. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O., DiLabio, G. A., Eds.; Elsevier, 2017; pp 65 – 136.

[106] Beuerle, M.; Graf, D.; Schurkus, H. F.; Ochsenfeld, C. Efficient calculation of beyond RPA correlation energies in the dielectric matrix formalism. *J. Chem. Phys.* **2018**, *148*, 204104.

[107] Eshuis, H.; Yarkony, J.; Furche, F. Fast computation of molecular random phase approximation correlation energies using resolution of the identity and imaginary frequency integration. *J. Chem. Phys.* **2010**, *132*, 234114.

[108] Burow, A. M.; Bates, J. E.; Furche, F.; Eshuis, H. Analytical First-Order Molecular Properties and Forces within the Adiabatic Connection Random Phase Approximation. *J. Chem. Theory Comput.* **2014**, *10*, 180–194.

[109] Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.

[110] Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364–3374.

[111] Sure, R.; Grimme, S. Comprehensive Benchmark of Association (Free) Energies of Realistic Host-Guest Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 3785–3801.

[112] Grimme, S.; Steinmetz, M. Effects of London dispersion correction in density functional theory on the structures of organic molecules in the gas phase. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031–16042.

[113] Risthaus, T.; Steinmetz, M.; Grimme, S. Implementation of nuclear gradients of range-separated hybrid density functionals and benchmarking on rotational constants for organic molecules. *J. Comput. Chem.* **2014**, *35*, 1509–1516.

[114] Szalewicz, K.; Jeziorski, B. Symmetry-adapted double-perturbation analysis of intramolecular correlation effects in weak intermolecular interactions. *Mol. Phys.* **1979**, *38*, 191–208.

[115] Rybak, S.; Jeziorski, B.; Szalewicz, K. Many-body symmetry-adapted perturbation theory of intermolecular interactions. H_2O and HF dimers. *J. Chem. Phys.* **1991**, *95*, 6576–6601.

[116] Gunnarsson, O.; Lundqvist, B. I. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B* **1976**, *13*, 4274–4298.

[117] Gunnarsson, O.; Lundqvist, B. I. Erratum: Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B* **1977**, *15*, 6006–6006.

[118] Hättig, C.; Hellweg, A.; Köhn, A. Distributed memory parallel implementation of energies and gradients for second-order Møller–Plesset perturbation theory with the resolution-of-the-identity approximation. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1159–1169.

[119] Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *WIREs Comput. Mol. Sci.* **2014**, *4*, 91–100.

[120] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[121] Tao, J.; Perdew, J.; Staroverov, V.; Scuseria, G. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

[122] Treutler, O.; Ahlrichs, R. Efficient molecular numerical integration schemes. *J. Chem. Phys.* **1995**, *102*, 346.

[123] Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, *150*, 154122.

[124] Brandenburg, J. G.; Bannwarth, C.; Hansen, A.; Grimme, S. B97–3c: A revised low-cost variant of the B97–D density functional method. *J. Chem. Phys.* **2018**, *148*, 064104.

[125] Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *J. Chem. Theory Comput.* **2015**, *11*, 1525–1539.

[126] Carter-Fenk, K.; Lao, K. U.; Liu, K.-Y.; Herbert, J. M. Accurate and Efficient ab Initio Calculations for Supramolecular Complexes: Symmetry-Adapted Perturbation Theory with Many-Body Dispersion. *J. Phys. Chem. Lett.* **2019**, *10*, 2706–2714.

[127] Kruse, H.; Mladek, A.; Gkionis, K.; Hansen, A.; Grimme, S.; Sponer, J. Quantum Chemical Benchmark Study on 46 RNA Backbone Families Using a Dinucleotide Unit. *J. Chem. Theory Comput.* **2015**, *11*, 4972–4991.

[128] Eshuis, H.; Furche, F. Basis set convergence of molecular correlation energy differences within the random phase approximation. *J. Chem. Phys.* **2012**, *136*, 084105.

[129] Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

[130] Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

[131] Weigend, F.; Furche, F.; Ahlrichs, R. Gaussian basis sets of quadruple zeta valence quality for atoms H-Kr. *J. Chem. Phys.* **2003**, *119*, 12753.

[132] Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007.

[133] Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

[134] Hättig, C. Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core-valence and quintuple-ζ basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.

[135] Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.

[136] Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16–18 elements. *J. Chem. Phys.* **2003**, *119*, 11113–11123.

[137] Peterson, K. A.; Shepler, B. C.; Figgen, D.; Stoll, H. On the Spectroscopic and Thermochemical Properties of ClO, BrO, IO, and Their Anions. *J. Phys. Chem. A* **2006**, *110*, 13877–13883.

[138] Risthaus, T.; Grimme, S. Benchmarking of London Dispersion-Accounting Density Functional Theory Methods on Very Large Molecular Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 1580–1591.

[139] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis–set convergence in correlated calculations on Ne, N_2, and H_2O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

[140] Luenser, A.; Schurkus, H. F.; Ochsenfeld, C. Vanishing-Overhead Linear-Scaling Random Phase Approximation by Cholesky Decomposition and an Attenuated Coulomb-Metric. *J Chem. Theory Comput.* **2017**, *13*, 1647–1655.

[141] Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125*, 194101.

[142] Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

[143] Brandenburg, J. G.; Bates, J. E.; Sun, J.; Perdew, J. P. Benchmark tests of a strongly constrained semilocal functional with a long-range dispersion correction. *Phys. Rev. B* **2016**, *94*, 115144.

[144] Jansen, G. Symmetry-adapted perturbation theory based on density functional theory for noncovalent interactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 127–144.

[145] Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. Dispersion Energy from Density-Functional Theory Description of Monomers. *Phys. Rev. Lett.* **2003**, *91*, 033201.

[146] Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050–8053.

[147] Kevorkyants, R.; Eshuis, H.; Pavanello, M. FDE-vdW: A van der Waals inclusive subsystem density-functional theory. *J. Chem. Phys.* **2014**, *141*, 044127.

[148] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.

[149] Ćwiok, T.; Jeziorski, B.; Kol/os, W.; Moszynski, R.; Szalewicz, K. On the convergence of the symmetrized Rayleigh-Schrödinger perturbation theory for molecular interaction energies. *J. Chem. Phys.* **1992**, *97*, 7555–7559.

[150] Cohen, M. H.; Wasserman, A. On Hardness and Electronegativity Equalization in Chemical Reactivity Theory. *J. Stat. Phys.* **2006**, *125*, 1121–1139.

[151] Elliott, P.; Burke, K.; Cohen, M. H.; Wasserman, A. Partition density-functional theory. *Phys. Rev. A* **2010**, *82*, 024501.

[152] Görling, A.; Levy, M. Correlation-energy functional and its high-density limit obtained from a coupling-constant perturbation expansion. *Phys. Rev. B* **1993**, *47*, 13105–13113.

[153] Görling, A.; Levy, M. Hardness of molecules and the band gap of solids within the Kohn–Sham formalism: A perturbation-scaling approach. *Phys. Rev. A* **1995**, *52*, 4493–4499.

[154] Eshuis, H.; Bates, J. E.; Furche, F. Electron correlation methods based on the random phase approximation. *Theor. Chem. Acc.* **2012**, *131*, 1084.

[155] Furche, F. Developing the random phase approximation into a practical post-Kohn-Sham correlation model. *J. Chem. Phys.* **2008**, *129*, 114105.

[156] Craig, D. P.; Thirunamachandran, T. *Molecular quantum electrodynamics: an introduction to radiation-molecule interactions*; Courier Corporation, 1998.

[157] Szalewicz, K. Symmetry-adapted perturbation theory of intermolecular forces. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 254–272.

[158] Fetter, A. L.; Walecka, J. D. *Quantum Theory of Many-Particle Systems*; McGraw-Hill, 1971.

[159] Cappellini, G.; Del Sole, R.; Reining, L.; Bechstedt, F. Model dielectric function for semiconductors. *Phys. Rev. B* **1993**, *47*, 9892–9895.

[160] Brawand, N. P.; Vörös, M.; Govoni, M.; Galli, G. Generalization of Dielectric-Dependent Hybrid Functionals to Finite Systems. *Phys. Rev. X* **2016**, *6*, 041002.

[161] Longuet-Higgins, H. C. Spiers Memorial Lecture. Intermolecular forces. *Discuss. Faraday Soc.* **1965**, *40*, 7–18.

[162] Zaremba, E.; Kohn, W. Van der Waals interaction between an atom and a solid surface. *Phys. Rev. B* **1976**, *13*, 2270–2285.

[163] Horn, R. A.; Mathias, R. Cauchy-Schwarz inequalities associated with positive semidefinite matrices. *Linear Algebra and its Applications* **1990**, *142*, 63 – 82.

[164] Chen, G. P.; Agee, M. M.; Furche, F. Performance and Scope of Perturbative Corrections to Random-Phase Approximation Energies. *J. Chem. Theory Comput.* **2018**, *14*, 5701–5714.

[165] Kato, T. *Perturbation theory for linear operators*; Springer Science & Business Media, 2013; Vol. 132.

[166] Olsen, J.; Jørgensen, P.; Helgaker, T.; Christiansen, O. Divergence in Møller–Plesset theory: A simple explanation based on a two-state model. *J. Chem. Phys.* **2000**, *112*, 9736–9748.

[167] Scuseria, G. E.; Henderson, T. M.; Sorensen, D. C. The ground state correlation energy of the random phase approximation from a ring coupled cluster doubles approach. *J. Chem. Phys.* **2008**, *129*, 231101.

[168] Toulouse, J.; Zhu, W.; Savin, A.; Jansen, G.; Ángyán, J. G. Closed-shell ring coupled cluster doubles theory with range separation applied on weak intermolecular interactions. *J. Chem. Phys.* **2011**, *135*, 084119.

[169] Dzyaloshinskii, I.; Lifshitz, E.; Pitaevskii, L. The general theory of van der Waals forces. *Adv. Phys.* **1961**, *10*, 165–209, cited By 1257.

[170] Bohm, D.; Pines, D. A Collective Description of Electron Interactions. I. Magnetic Interactions. *Phys. Rev.* **1951**, *82*, 625–634.

[171] Pines, D.; Bohm, D. A Collective Description of Electron Interactions: II. Collective vs Individual Particle Aspects of the Interactions. *Phys. Rev.* **1952**, *85*, 338–353.

[172] Yan, Z.; Perdew, J. P.; Kurth, S. Density functional for short-range correlation: Accuracy of the random-phase approximation for isoelectronic energy changes. *Phys. Rev. B* **2000**, *61*, 16430–16439.

[173] Yan, Z.; Perdew, J. P.; Kurth, S. Erratum: Density functional for short-range correlation: Accuracy of the random-phase approximation for isoelectronic energy changes [Phys. Rev. B 61, 16430 (2000)]. *Phys. Rev. B* **2010**, *81*, 169902.

[174] Ajala, A. O.; Voora, V.; Mardirossian, N.; Furche, F.; Paesani, F. Assessment of Density Functional Theory in Predicting Interaction Energies between Water and Polycyclic Aromatic Hydrocarbons: from Water on Benzene to Water on Graphene. *J. Chem. Theory and Comput.* **2019**, *15*, 2359–2374.

[175] Shepherd, J. J.; Grüneis, A. Many-Body Quantum Chemistry for the Electron Gas: Convergent Perturbative Theories. *Phys. Rev. Lett.* **2013**, *110*, 226401.

[176] Grüneis, A.; Marsman, M.; Harl, J.; Schimka, L.; Kresse, G. Making the random phase approximation to electronic correlation accurate. *J. Chem. Phys.* **2009**, *131*, 154115.

[177] Paier, J.; Janesko, B. G.; Henderson, T. M.; Scuseria, G. E.; Grüneis, A.; Kresse, G. Hybrid functionals including random phase approximation correlation and second-order screened exchange. *J. Chem. Phys.* **2010**, *132*, 094103.

[178] Paier, J.; Janesko, B. G.; Henderson, T. M.; Scuseria, G. E.; Grüneis, A.; Kresse, G. Erratum: "Hybrid functionals including random phase approximation correlation and second-order screened exchange" [J. Chem. Phys. 132, 094103 (2010)]. *J. Chem. Phys.* **2010**, *133*, 179902.

[179] Bates, J. E.; Furche, F. Communication: Random phase approximation renormalized many-body perturbation theory. *J. Chem. Phys.* **2013**, *139*, 171103.

[180] Jones, A. P.; Crain, J.; Sokhan, V. P.; Whitfield, T. W.; Martyna, G. J. Quantum Drude oscillator model of atoms and molecules: Many-body polarization and dispersion interactions for atomistic simulation. *Phys. Rev. B* **2013**, *87*, 144103.

[181] Hermann, J.; DiStasio, R. A.; Tkatchenko, A. First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications. *Chem. Rev.* **2017**, *117*, 4714–4758.

[182] Heßelmann, A. Random-phase-approximation correlation method including exchange interactions. *Phys. Rev. A* **2012**, *85*, 012517.

[183] Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, 034106.

[184] Kállay, M. Linear-scaling implementation of the direct random-phase approximation. *J. Chem. Phys.* **2015**, *142*, 204105.

[185] Chen, G. P.; Voora, V. K.; Agee, M. M.; Balasubramani, S. G.; Furche, F. Random-Phase Approximation Methods. *Annu. Rev. Phys. Chem.* **2017**, *68*, 421–445.

[186] Schurkus, H. F.; Ochsenfeld, C. Communication: An effective linear-scaling atomic-orbital reformulation of the random-phase approximation using a contracted double-Laplace transformation. *J. Chem. Phys.* **2016**, *144*, 031101.

[187] Beuerle, M.; Ochsenfeld, C. Low-scaling analytical gradients for the direct random phase approximation using an atomic orbital formalism. *J. Chem. Phys.* **2018**, *149*, 244111.

[188] Graf, D.; Beuerle, M.; Ochsenfeld, C. Low-Scaling Self-Consistent Minimization of a Density Matrix Based Random Phase Approximation Method in the Atomic Orbital Space. *J. Chem. Theory Comput.* **2019**, *15*, 4468–4477.

[189] Voora, V. K.; Balasubramani, S. G.; Furche, F. Variational generalized Kohn-Sham approach combining the random-phase-approximation and Green's-function methods. *Phys. Rev. A* **2019**, *99*, 012518.

[190] Muuronen, M.; Deglmann, P.; Tomović, Ž. Design Principles for Rational Polyurethane Catalyst Development. *J. Org.* **2019**, *84*, 8202–8209.

[191] Zhang, S.-Y.; Shi, W.; Cheng, P.; Zaworotko, M. J. A Mixed-Crystal Lanthanide Zeolite-like Metal–Organic Framework as a Fluorescent Indicator for Lysophosphatidic Acid, a Cancer Biomarker. *J. Am. Chem. Soc.* **2015**, *137*, 12203–12206.

[192] Price, S. L. Computed Crystal Energy Landscapes for Understanding and Predicting Organic Crystal Structures and Polymorphism. *Acc. Chem. Res.* **2009**, *42*, 117–126.

[193] Choi, E.-Y.; Han, T. H.; Hong, J.; Kim, J. E.; Lee, S. H.; Kim, H. W.; Kim, S. O. Noncovalent functionalization of graphene with end-functional polymers. *J. Mater. Chem.* **2010**, *20*, 1907–1912.

[194] Gould, T.; Gray, E.; Dobson, J. F. van der Waals dispersion power laws for cleavage, exfoliation, and stretching in multiscale, layered systems. *Phys. Rev. B* **2009**, *79*, 113402.

[195] Nguyen, B. D.; Chen, G. P.; Agee, M. M.; Burow, A. M.; Tang, M. P.; Furche, F. Divergence of Many-Body Perturbation Theory for Noncovalent Interactions of Large Molecules. *J. Chem. Theory Comput.* **2020**, *16*, 2258–2273.

[196] Ángyán, J.; Dobson, J.; Jansen, G.; Gould, T. *London Dispersion Forces in Molecules, Solids and Nano-structures*; Theoretical and Computational Chemistry Series; The Royal Society of Chemistry, 2020.

[197] Casimir, H. B. G.; Polder, D. The Influence of Retardation on the London-van der Waals Forces. *Phys. Rev.* **1948**, *73*, 360–372.

[198] Nooijen, M.; Shamasundar, K. R.; Mukherjee, D. Reflections on size-extensivity, size-consistency and generalized extensivity in many-body theory. *Mol. Phys.* **2005**, *103*, 2277–2298.

[199] Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

[200] Jeziorski, B.; Szalewicz, K.; Chałasiński, G. Symmetry forcing and convergence properties of perturbation expansions for molecular interaction energies. *Int. J. Quantum Chem.* **1978**, *14*, 271–287.

[201] Patkowski, K.; Jeziorski, B.; Korona, T.; Szalewicz, K. Symmetry-forcing procedure and convergence behavior of perturbation expansions for molecular interaction energies. *J. Chem. Phys.* **2002**, *117*, 5124–5134.

[202] Szalewicz, K.; Patkowski, K.; Jeziorski, B. In *Intermolecular Forces and Clusters II*; Wales, D. J., Ed.; Springer Berlin Heidelberg, 2005; pp 43–117.

[203] Onida, G.; Reining, L.; Rubio, A. Electronic excitations: density-functional versus many-body Green's-function approaches. *Rev. Mod. Phys.* **2002**, *74*, 601–659.

[204] Hohenstein, E. G.; Sherrill, C. D. Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory. *J. Chem. Phys.* **2010**, *133*, 014101.

[205] Hapka, M.; Przybytek, M.; Pernal, K. Symmetry-Adapted Perturbation Theory Based on Multiconfigurational Wave Function Description of Monomers. *J. Chem. Theory Comput.* **2021**, *17*, 5538–5555.

[206] Williams, H. L.; Chabalowski, C. F. Using Kohn–Sham Orbitals in Symmetry-Adapted Perturbation Theory to Investigate Intermolecular Interactions. *J. Phys. Chem. A* **2001**, *105*, 646–659.

[207] Heßelmann, A.; Jansen, G. The helium dimer potential from a combined density functional theory and symmetry-adapted perturbation theory approach using an exact exchange–correlation potential. *Phys. Chem. Chem. Phys.* **2003**, *5*, 5010–5014.

[208] Heßelmann, A.; Jansen, G. Intermolecular dispersion energies from time-dependent density functional theory. *Chem. Phys. Lett.* **2003**, *367*, 778–784.

[209] Tekin, A.; Jansen, G. How accurate is the density functional theory combined with symmetry-adapted perturbation theory approach for CH-π and π-π interactions? A comparison to supermolecular calculations for the acetylene–benzene dimer. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1680–1687.

[210] Pitoňák, M.; Riley, K. E.; Neogrády, P.; Hobza, P. Highly Accurate CCSD(T) and DFT–SAPT Stabilization Energies of H-Bonded and Stacked Structures of the Uracil Dimer. *ChemPhysChem* **2008**, *9*, 1636–1644.

[211] Zen, A.; Brandenburg, J. G.; Klimeš, J.; Tkatchenko, A.; Alfè, D.; Michaelides, A. Fast and accurate quantum Monte Carlo for molecular crystals. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 1724–1729.

[212] Harl, J.; Kresse, G. Cohesive energy curves for noble gas solids calculated by adiabatic connection fluctuation-dissipation theory. *Phys. Rev. B* **2008**, *77*, 045136.

[213] Li, Y.; Lu, D.; Nguyen, H.-V.; Galli, G. van der Waals Interactions in Molecular Assemblies from First-Principles Calculations. *J. Phys. Chem. A* **2010**, *114*.

[214] Ren, X.; Rinke, P.; Blum, V.; Wieferink, J.; Tkatchenko, A.; Sanfilippo, A.; Reuter, K.; Scheffler, M. Resolution-of-identity approach to Hartree–Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *N. J. Phys.* **2012**, *14*, 053020.

[215] Kaltak, M.; Klimeš, J.; Kresse, G. Low Scaling Algorithms for the Random Phase Approximation: Imaginary Time and Laplace Transformations. *J. Chem. Theory Comput.* **2014**, *10*, 2498–2507.

[216] Del Ben, M.; Schütt, O.; Wentz, T.; Messmer, P.; Hutter, J.; VandeVondele, J. Enabling simulation at the fifth rung of DFT: Large scale RPA calculations with excellent time to solution. *Comput. Phys. Commun.* **2015**, *187*, 120–129.

[217] Johnson, E. R.; Becke, A. D. A post-Hartree–Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123*, 024101.

[218] Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101.

[219] Johnson, E. R.; Becke, A. D. A post-Hartree–Fock model of intermolecular interactions: Inclusion of higher-order corrections. *J. Chem. Phys.* **2006**, *124*, 174104.

[220] Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the dispersion interaction revisited. *J. Chem. Phys.* **2007**, *127*, 154108.

[221] Schwabe, T.; Grimme, S. Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397–3406.

[222] Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.

[223] Bučko, T.; Lebègue, S.; Hafner, J.; Ángyán, J. G. Tkatchenko-Scheffler van der Waals correction method with and without self-consistent screening applied to solids. *Phys. Rev. B* **2013**, *87*, 064110.

[224] Ambrosetti, A.; Reilly, A. M.; DiStasio, R. A.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys.* **2014**, *140*, 18A508.

[225] Odbadrakh, T. T.; Jordan, K. D. Dispersion dipoles for coupled Drude oscillators. *J. Chem. Phys.* **2016**, *144*, 034111.

[226] Sommerfeld, T.; Jordan, K. D. Quantum Drude Oscillator Model for Describing the Interaction of Excess Electrons with Water Clusters: An Application to $(H_2O)_{13}$. *J. Phys. Chem. A* **2005**, *109*, 11531–11538.

[227] Dias, D. A.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, *2*, 303–336.

[228] Butler, M. S. The Role of Natural Product Chemistry in Drug Discoveryi†. *J. Nat. Prod.* **2004**, *67*, 2141–2153.

[229] Vasquéz, D. *Inhibitors of Protein Biosyntehsis*; Springer, 1979.

[230] Belle, A.; Tanay, A.; Bitincka, L.; Shamir, R.; O'Shea, E. K. Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. U.S.A* **2006**, *103*, 13004–13009.

[231] Ingolia, N.; Lareau, L.; Weissman, J. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **2011**, *147*, 789–802.

[232] Stern-Ginossar, N.; Weisburd, B.; Michalski, A.; Le, V. T. K.; Hein, M. Y.; Huang, S.-X.; Ma, M.; Shen, B.; Qian, S.-B.; Hengel, H.; Mann, M.; Ingolia, N. T.; Weissman, J. S. Decoding Human Cytomegalovirus. *Science* **2012**, *338*, 1088–1093.

[233] Malochet-Grivois, C.; Cotelle, P.; Biard, J.; Hénichart, J.; Debitus, C.; Roussakis, C.; Verbist, J. Dichlorolissoclimide, a new cytotoxic labdane derivative from Lissoclinum voeltzkowiMichaelson (Urochordata). *Tetrahedron Lett.* **1991**, *32*, 6701–6702.

[234] Robert, F.; Gao, H. Q.; Donia, M.; Merrick, W. C.; Hamann, M. T.; Pelletier, J. Chlorolissoclimides: New inhibitors of eukaryotic protein synthesis. *RNA* **2006**, *12*, 717–725.

[235] Könst, Z. A.; Szklarski, A. R.; Pellegrino, S.; Michalak, S. E.; Meyer, M.; Zanette, C.; Cencic, R.; Nam, S.; Voora, V. K.; Horne, D. A.; Pelletier, J.; Mobley, D. L.; Yusupova, G.; Yusupov, M.; Vanderwal, C. D. Synthesis facilitates an understanding of the structural basis for translation inhibition by the lissoclimides. *Nat. Chem.* **2017**, *9*, 1140–1149.

[236] Imai, Y. N.; Inoue, Y.; Nakanishi, I.; Kitaura, K. Cl-π Interactions in Protein-Ligand Complexes. *Protein Sci.* **2008**, *17*, 1129–1137.

[237] Riley, K. E.; Hobza, P. Strength and Character of Halogen Bonds in Protein–Ligand Complexes. *Cryst. Growth Des.* **2011**, *11*, 4272–4278.

[238] 3-Iodo-4-phenoxypyridinones (IOPY's), a new family of highly potent non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4309–4312.

[239] Himmel, D. M.; Das, K.; Clark, A. D.; Hughes, S. H.; Benjahad, A.; Oumouch, S.; Guillemont, J.; Coupa, S.; Poncelet, A.; Csoka, I.; Meyer, C.; Andries, K.; Nguyen, C. H.; Grierson, D. S.; Arnold, E. Crystal Structures for HIV-1 Reverse Transcriptase in Complexes with Three Pyridinone Derivatives: A New Class of Non-Nucleoside Inhibitors Effective against a Broad Range of Drug-Resistant Strains. *J. Med. Chem.* **2005**, *48*, 7582–7591.

[240] Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; Plancher, J.-M.; Hartmann, G.; Banner, D. W.; Haap, W.; Diederich, F. Systematic Investigation of Halogen Bonding in Protein–Ligand Interactions. *Angew. Chem. Int. Ed. 50*, 314–318.

[241] Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **2013**, *56*, 1363–1388.

[242] Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

[243] Kem, W. R.; Mahnir, V. M.; Prokai, L.; Papke, R. L.; Cao, X.; LeFrancois, S.; Wildeboer, K.; Prokai-Tatrai, K.; Porter-Papke, J.; Soti, F. Hydroxy Metabolites of the Alzheimer's Drug Candidate 3-[(2,4-Dimethoxy)Benzylidene]-Anabaseine Dihydrochloride (GTS-21): Their Molecular Properties, Interactions with Brain Nicotinic Receptors, and Brain Penetration. *Mol. Pharmacol.* **2004**, *65*, 56–67.

[244] Cavasotto, C. N.; Di Filippo, J. I. In silico Drug Repurposing for COVID-19: Targeting SARS-CoV-2 Proteins through Docking and Consensus Ranking. *Mol. Inform. 40*, 2000115.

[245] Margiotta, N.; Marzano, C.; Gandin, V.; Osella, D.; Ravera, M.; Gabano, E.; Platts, J. A.; Petruzzella, E.; Hoeschele, J. D.; Natile, G. Revisiting [PtCl2(cis-1,4-DACH)]: An Underestimated Antitumor Drug with Potential Application to the Treatment of Oxaliplatin-Refractory Colorectal Cancer. *J. Med. Chem.* **2012**, *55*, 7182–7192.

[246] Barone, G.; Guerra, C. F.; Gambino, N.; Silvestri, A.; Lauria, A.; Almerico, A. M.; Bickelhaupt, F. M. Intercalation of Daunomycin into Stacked DNA Base Pairs. DFT Study of an Anticancer Drug. *J. Biomol. Struct. Dyn.* **2008**, *26*, 115–129.

[247] Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.

[248] Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.

[249] Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119*, 12129–12137.

[250] Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem. 32*, 1456–1465.

[251] Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143 – 152.

[252] Kesharwani, M. K.; Manna, D.; Sylvetsky, N.; Martin, J. M. L. The X40×10 Halogen Bonding Benchmark Revisited: Surprising Importance of (n–1)d Subvalence Correlation. *J. Phys. Chem. A* **2018**, *122*, 2184–2197.

[253] Malochet-Grivois, C.; Roussakis, C.; Robillard, N.; Biard, J. F.; Riou, D.; Débitus, C.; Verbist, J. F. Effects in vitro of two marine substances, chlorolissoclimide and dichlorolissoclimide, on a non-small-cell bronchopulmonary carcinoma line (NSCLC-N6). *Anticancer Drug Des* **1992**, *7*, 493–502.

[254] Uddin, J.; Ueda, K.; Siwu, E. R.; Kita, M.; Uemura, D. Cytotoxic labdane alkaloids from an ascidian Lissoclinum sp.: Isolation, structure elucidation, and structure–activity relationship. *Bioorg. Med. Chem.* **2006**, *14*, 6954–6961.

[255] Fu, X.; Palomar, A. J.; Hong, E. P.; Schmitz, F. J.; Valeriote, F. A. Cytotoxic Lissoclimide-Type Diterpenes from the Molluscs Pleurobranchus albiguttatus and Pleurobranchus forskalii. *J. Nat. Prod.* **2004**, *67*, 1415–1418.

[256] Uddin, M. J.; Kokubo, S.; Ueda, K.; Suenaga, K.; Uemura, D. Haterumaimides J and K, Potent Cytotoxic Diterpene Alkaloids from the Ascidian Lissoclinum Species. *Chem. Lett.* **2002**, *31*, 1028–1029.

[257] Uddin, M. J.; Kokubo, S.; Ueda, K.; Suenaga, K.; Uemura, D. Haterumaimides F-I, Four New Cytotoxic Diterpene Alkaloids from an Ascidian Lissoclinum Species. *J. Nat. Prod.* **2001**, *64*, 1169–1173.

[258] Uddin, M. J.; Kokubo, S.; Suenaga, K.; Ueda, K.; Uemura, D. Haterumaimides A-E, Five New Dichlorolissoclimide-Type Diterpenoids from an Ascidian, *Lissoclinum* Species. *Heterocycles* **2022**, *54*, 1039–1047.

[259] Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem Theory Comput.* **2013**, *9*, 1918–1931.

[260] Nguyen, B. D.; Hernandez, D. J.; Flores, E. V.; Furche, F. Dispersion size-consistency. *Electron. Struct.* **2022**, *4*, 014003.

[261] Hathwar, V. R.; Chopra, D.; Panini, P.; Guru Row, T. N. Revealing the Polarizability of Organic Fluorine in the Trifluoromethyl Group: Implications in Supramolecular Chemistry. *Cryst. Growth Des.* **2014**, *14*, 5366–5369.

[262] Hathwar, V. R.; Guru Row, T. N. Charge Density Analysis of Heterohalogen (Cl $\cdots$ F) and Homohalogen (F $\cdots$ F) Intermolecular Interactions in Molecular Crystals: Importance of the Extent of Polarizability. *Crys. Growth Des.* **2011**, *11*, 1338–1346.

[263] Hathwar, V. R.; Row, T. N. G. Nature of Cl $\cdots$ Cl Intermolecular Interactions via Experimental and Theoretical Charge Density Analysis: Correlation of Polar Flattening Effects with Geometry. *J. Phys. Chem. A* **2010**, *114*, 13434–13441.

[264] Hathwar, V. R.; Thakur, T. S.; Dubey, R.; Pavan, M. S.; Guru Row, T. N.; Desiraju, G. R. Extending the Supramolecular Synthon Based Fragment Approach (SBFA) for Transferability of Multipole Charge Density Parameters to Monofluorobenzoic Acids and their Cocrystals with Isonicotinamide: Importance of C–H $\cdots$ O, C–H $\cdots$ F, and F $\cdots$ F Intermolecular Regions. *J. Phys. Chem. A* **2011**, *115*, 12852–12863.

[265] Kukic, P.; Farrell, D.; McIntosh, L. P.; García-Moreno E., B.; Jensen, K. S.; Toleikis, Z.; Teilum, K.; Nielsen, J. E. Protein Dielectric Constants Determined from NMR Chemical Shift Perturbations. *J. Am. Chem. Soc.* **2013**, *135*, 16968–16976.

[266] Schutz, C. N.; Warshel, A. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins: Struct., Funct., Bioinf. 44*, 400–417.

[267] Xiu, X.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. Nicotine binding to brain receptors requires a strong cation–π interaction. *Nature* **2009**, *458*, 534–537.

[268] Fleming, F. F.; Yao, L.; Ravikumar, P. C.; Funk, L.; Shook, B. C. Nitrile-Containing Pharmaceuticals: Efficacious Roles of the Nitrile Pharmacophore. *J. Med. Chem.* **2010**, *53*, 7902–7917.

[269] Nepali, K.; Lee, H.-Y.; Liou, J.-P. Nitro-Group-Containing Drugs. *J. Med. Chem.* **2019**, *62*, 2851–2893.

[270] Kobayashi, K.; Asakawa, Y.; Kato, Y.; Aoyama, Y. Complexation of hydrophobic sugars and nucleosides in water with tetrasulfonate derivatives of resorcinol cyclic tetramer having a polyhydroxy aromatic cavity: importance of guest-host CH-.pi. interaction. *J. Am. Chem. Soc.* **1992**, *114*, 10307–10313.

[271] Amabilino, D. B.; Ashton, P. R.; Brown, C. L.; Cordova, E.; Godinez, L. A.; Goodnow, T. T.; Kaifer, A. E.; Newton, S. P.; Pietraszkiewicz, M. a. Molecular Meccano. 2. Self-Assembly of [n]Catenanes. *J. Am. Chem. Soc.* **1995**, *117*, 1271–1293.

[272] Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. Intermolecular interactions of nitrobenzene-benzene complex and nitrobenzene dimer: Significant stabilization of slipped-parallel orientation by dispersion interaction. *J. Chem. Phys.* **2006**, *125*, 124304.

[273] Helttunen, K.; Lehtovaara, L.; Häkkinen, H.; Nissinen, M. Crystal Structures and Density Functional Theory Calculations of o- and p-Nitroaniline Derivatives: Combined Effect of Hydrogen Bonding and Aromatic Interactions on Dimerization Energy. *Crys. Growth Des.* **2013**, *13*, 3603–3612.

[274] Croes, K.; Steffens, A.; Marchand, D.; Snyder, L. Relevance of $\pi - \pi$ and dipole-dipole interactions for retention on cyano and phenyl columns in reversed-phase liquid chromatography. *J. Chromatogr. A* **2005**, *1098*, 123–130.

[275] Plevin, M. J.; Bryce, D. L.; Boisbouvier, J. Direct detection of CH/π interactions in proteins. *Nat. Chem.* **2010**, *2*, 466–471.

[276] CHπ interaction in the conformation of organic compounds. A database study. *Tetrahedron* **1999**, *55*, 10047–10056.

[277] Tsuzuki, S.; Honda, K.; Fujii, A.; Uchimaru, T.; Mikami, M. CH/π interactions in methane clusters with polycyclic aromatic hydrocarbons. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2860–2865.

[278] Kato, M.; Pisliakov, A. V.; Warshel, A. The barrier for proton transport in aquaporins as a challenge for electrostatic models: The role of protein relaxation in mutational calculations. *Proteins: Struct., Funct., Bioinf. 64*, 829–844.

[279] Hennemann, M.; Murray, J. S.; Politzer, P.; Riley, K. E.; Clark, T. Polarization-induced σ-holes and hydrogen bonding. *J. Mol. Model.* **2012**, *18*, 2461–2469.