# INVESTIGATION OF NONCOVALENT INTERACTIONS IN COMPLEX SYSTEMS USING EFFECTIVE FRAGMENT POTENTIAL METHOD

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Pradeep Kumar Gurunathan

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2018

Purdue University

West Lafayette, Indiana

# THE PURDUE UNIVERSITY GRADUATE SCHOOL
# STATEMENT OF DISSERTATION APPROVAL

Dr. Lyudmila V. Slipchenko

Department of Chemistry

Dr. Nikolai R. Skrynnikov

Department of Chemistry

Dr. Garth J. Simpson

Department of Chemistry

Dr. Adam Wasserman

Department of Chemistry

**Approved by:**

Dr. Christine A. Hrycyna

Head of the School Graduate Program

# ACKNOWLEDGMENTS

My sincere and heartfelt thanks to my advisor Prof. Lyudmila Slipchenko for her continuous support throughout the entire duration of my graduate career. It has been a true honor to have been mentored by you. You have been a true role-model in every aspect of your life and an unwavering support during my ebbs and flows. You have inculcated a vibrant and positive atmosphere within the research group, which helped me tremendously. I hope I can repay your kindness by becoming the successful scientist you want me to be.

I wish to express my sincere gratitude to my collaborators - Prof. Anna Krylov, Prof. Ksenia Bravaya, and Prof. Debashree Ghosh for their constant support and encouragement. It was a tremendous pleasure collaborating with you and your research groups.

I thank my committee members, Prof. Garth Simpson, Prof. Nikolai Skrynnikov and Prof. Adam Wasserman for their insightful comments and encouragement.

I am deeply indebted to my colleagues (Ben Nebgen, Carlos Borca, Yen Bui, Claudia Viquez-Rojas, Yongbin Kim, Danil Kaliakin, Sungmin Hong and Hanjing Xu) for their great support and invaluable advice. It has been a tremendously enjoyable experience working with you!

I would like to thank my colleague, Ms. Cynthia Everhart for her moral support throughout my graduate career. Your genuine care and support means a lot to me.

Thanks to all my friends who made me feel at home in a new country and a new environment. You have become an integral part of my life, and my time at Purdue has been memorable and joyful because of you!

A special thanks to Neha Agrawal, whose unwavering support throughout my graduate career helped me immensely. This dissertation would not have been possible without your support and encouragement.

Lastly, I would like to thank my parents who have stood by me every step of the way in my career. You have supported me in every possible way you could, and then some.

TABLE OF CONTENTS

Page

# LIST OF TABLES

# LIST OF FIGURES

Figure Page

Figure Page

# ABBREVIATIONS

| | |
|---|---|
| AO | Atomic orbital |
| BioEFP | Biomolecular Effective Fragment Potential method |
| CADD | Computer-aided drug design |
| CCSD | Coupled cluster - single and double excitations |
| CIS(D) | Configuration Interaction - single excitations and perturbative doubles correction |
| CPHF | Coupled Perturbed Hartree-Fock |
| CT | Charge-transfer |
| DFT | Density functional theory |
| EFP | Effective Fragment Potential |
| EM | Energy minimization |
| EOM | Equation of motion |
| F-SAPT | Functional symmetry adapted perturbation theory |
| GFP | Green fluorescent protein |
| HF | Hartree-Fock |
| HOMO | Highest occupied molecular orbital |
| IE | Ionization energy |
| IP | Ionization potential |
| LCAO-MO | Linear combination of atomic orbitals - molecular orbitals |
| LMO | Localized molecular orbital |
| LP | Lone pair |
| LUMO | Lowest unoccupied molecular orbital |
| mEFP | Macromolecular Effective Fragment Potential method |
| MD | Molecular dynamics |

| | |
|---|---|
| MFCC | Molecular fractionation with conjugate caps |
| MM | Molecular mechanics |
| MO | Molecular orbital |
| MTA | Molecular tailoring approach |
| $OH^{\bullet}$ | Hydroxyl radical |
| P-SAPT | Pairwise symmetry adapted perturbation theory |
| QM | Quantum mechanics |
| SAPT | Symmetry adapted perturbation theory |
| SOMO | Singly occupied molecular orbital |
| SOS-CIS(D) | Scaled opposite spin configuration interaction - single excitations and perturbative doubles correction |
| TD-DFT | Time-dependent density functional theory |
| UV | Ultraviolet |
| VDE | Vertical detachment energy |
| VEA | Vertical electron affinity |

# ABSTRACT

Gurunathan, Pradeep Kumar Ph.D., Purdue University, December 2018. Investigation of Noncovalent Interactions in Complex Systems Using Effective Fragment Potential Method. Major Professor: Lyudmila V. Slipchenko.

Computational Chemistry has proven to be an effective means of solving chemical problems. The two main tools of Computational Chemistry - quantum mechanics and molecular mechanics, have provided viable avenues to probe such chemical problems at an electronic or molecular level, with varying levels of accuracy and speed. In this work, attempts have been made to combine the speed of molecular mechanics and the accuracy of quantum mechanics to work across multiples scales of time and length, effectively resulting in simulations of large chemical systems without compromising the accuracy.

The primary tool utilized for methods development and application in this work is the Effective Fragment Potential (EFP) method. The EFP method is a computational technique for studying non-covalent interactions in complex systems. EFP is an accurate *ab initio* force field, with accuracy comparable to many Density Functional Theory (DFT) methods, at significantly lower computational cost. EFP decomposes intermolecular interactions into contributions from four terms: electrostatics, polarization, exchange-repulsion and dispersion.

In the first chapter, the possibility of applying EFP method to study large radical-water clusters is probed. An approximate theoretical model in which the transition dipole moments of excitations are computed using the information from the ground state orbitals is implemented.

A major challenge to broaden the scope of EFP is to overcome its limitation in describing only small and rigid molecules such as water, acetone, etc. In the second

chapter, the extension of EFP method to large covalently bound biomolecules and polymers such as proteins, lipids etc., is described. Using this new method, referred to as BioEFP/mEFP, it is shown that the effect of polarization is non-negligible and must be accounted for when modeling photochemical and electron-transfer processes in photoactive proteins.

Another area of interest is the development of novel drug-target binding models, in which a chemically active part of the ligand is modified via functional group modification, while the rest of the system remains intact. In the third chapter, the development and application of a drug-target binding model is explained.

Lastly, in the fourth and final chapter, we show the derivation for working equations corresponding to the coupling gradient term describing the dispersion interactions between quantum mechanical and effective fragment potential regions.

The primary focus of this work is to explore and expand the boundaries of multiscale QM/MM simulations applied to chemical and biomolecular systems. We believe that the work described here leads to exciting pathways in the future in terms of modeling novel systems and processes such as heterogeneous catalysis, QSAR, crystal structure prediction, etc.

# 1. CHARGE TRANSFER STATES OF HYDROXYL RADICAL IN BULK WATER AND AT THE AIR-WATER INTERFACE

The hydroxyl radical ($OH^{\bullet}$) is one of the most important oxidants in the atmosphere. As $OH^{\bullet}$ is attracted to water-containing aerosols, its chemistry and photochemistry are strongly affected by interactions with water. The absorption spectrum of hydrated $OH^{\bullet}$ exhibits a broad charge transfer peak corresponding to the electron transfer from nearby water molecules to the singly-occupied molecular orbital of the hydroxyl radical, which results in a charge-separated pair of hydroxyl anion ($OH^{-}$) and cationic water ($H_2O^{+}$). In the present work, we investigate the nature of charge transfer states of the hydroxyl radical floated near a water-vacuum interface, using a combination of Koopmans' theorem and hybrid QM/MM simulations based on the polarizable effective fragment potential (EFP) method. Our results indicate that the charge transfer to the hydroxyl radical preferentially happens from the interfacial water molecules. These long-range charge transfer transitions occur within the UV range accessible in the atmosphere and present intriguing implications such as increased acidity of water surface and ionization of atmospheric aerosols.

## 1.1 Introduction

The existence of hydroxyl radical in the atmosphere and its importance for chemical make-up of atmosphere have been recognized for over 40 years [1–5]. However, high reactivity of the hydroxyl radical makes it difficult to monitor the kinetics and thermodynamics of its reaction pathways, and hence a comprehensive understanding of the underlying chemistry has been elusive [4, 6–8]. Even more challenging task is to understand photochemical processes occurring in the atmosphere. For instance,

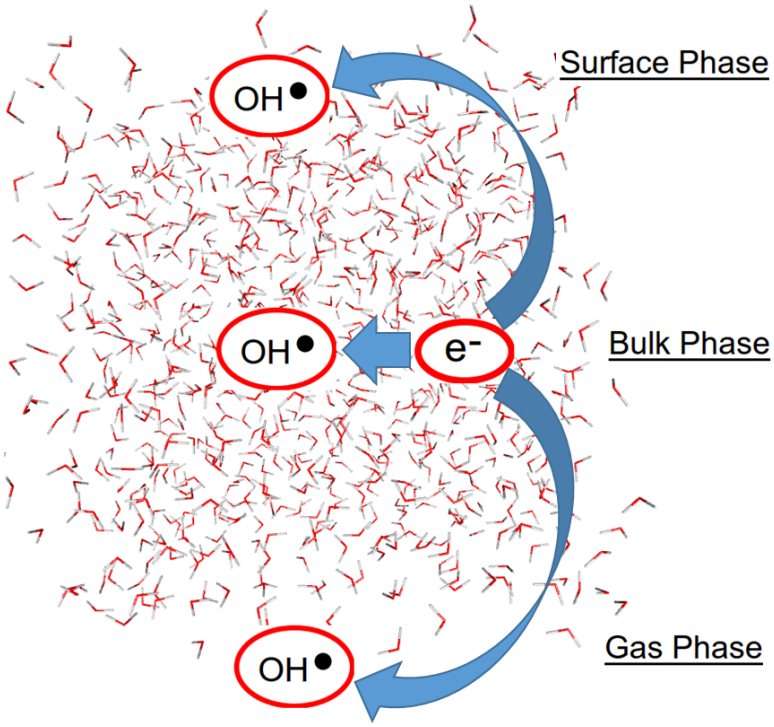



Fig. 1.1. Representative image: Transfer of electron from a water molecule to hydroxyl radicals situated in gas-phase, surface phase and in bulk phase.

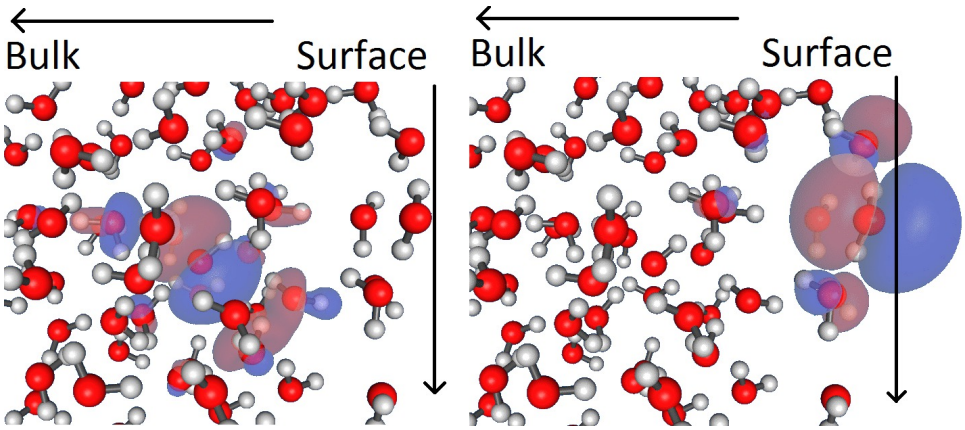



Fig. 1.2. Representative image of HOMO of hydroxyl radical situated in bulk (left) and lone pair orbital of water situated at the surface.

the interaction of the hydroxyl radical with water has been of interest due to its propensity to water-containing aerosols, however, the excited state properties of such complexes have not been well studied [9–11].

Hydroxyl radical is known to interact with one or few water molecules forming a cluster, or it may dissolve in a water droplet preferentially staying near droplet surface [12–14]. While it might be expected that various water environments alter the excited state properties of the hydroxyl radical, detailed understanding of these phenomena is missing.

The absorption spectrum of hydroxyl radical in bulk water has two distinct characteristic features, a relatively less bright, broad shoulder that is centered at around 340 nm, and a narrower, but brighter peak near 230 nm [15, 16]. The broad shoulder at 300-350 nm is thought to be due to the valence $n \rightarrow \pi^*$ transition, which can be observed in the gas phase excitation spectrum as well [16, 17]. The peak at 230 nm is assigned to charge transfer (CT) excitation, caused by the transfer of an electron from surrounding water to the hydroxyl radical. This CT transition results in formation of a positive hole on the surrounding water(s), and a hydroxide anion ($OH^-$) [16].

Earlier, we studied the interaction of hydroxyl radical with water clusters containing one to four water molecules, and their charge-transfer behavior [18]. We sampled sufficient number of charge transfer states in these clusters to understand a transformation of the CT spectrum with increasing number of water molecules. Specifically, we noticed that with the increase in size of water cluster, the number of CT states and their density increase, while their excitation energy decreases. We also observed a number of CT states in which the excitation is delocalized over more than one water molecule and expands beyond the first hydration shell.

In the present work, we set a more challenging goal of modeling the CT spectra of the hydroxyl radical solvated in a water environment mimicking atmospheric aerosols.

Specifically, we address the following questions:

1. Which water molecules contribute the most to the CT band? For example, are the strongly-contributing water molecules located in the first hydration shell of the hydroxyl radical or are they further separated? Are they involved in H-bonding with the hydroxyl radical? How long-range are water-to-$OH^{\bullet}$ CT transitions?

2. What are the spectroscopic signatures unique to the hydroxyl radical situated at the surface of water droplet and in bulk water?

To answer these questions, we performed hybrid quantum mechanics / Effective Fragment Potential (QM/EFP) simulations on snapshots extracted from classical MD trajectories from Ref. [12]. Four hydroxyl radicals are solvated in a periodic water slab, such that solvation in both bulk water and at the vacuum-water interface are explored. In QM/EFP simulations, the QM region includes all water molecules within 6 Å of the hydroxyl radical (i.e., approximately two full hydration shells), such that long-range CT states, if present, are attainable. The EFP method provides polarizable embedding which has been shown to be essential in describing electronic states with charge-transfer character [19, 20].

## 1.2 Methods and Computational details

Molecular dynamics (MD) simulations trajectories were obtained from the previous work by Roeselova et. al [12]. For the sake of completeness, a brief description of these MD simulations is presented here. A rectangular periodic box of 3 x 3 x 10 nm was constructed with 864 water molecules and 4 hydroxyl radicals. Along the x and y axes, the box is completely filled with water molecules, but a layer of vacuum separates water slabs along the z-axis, such that an air-water interface is present along z direction. Water molecules are described using POL3 water model, a polarizable water model developed by Caldwell and Kollman [21]. The parameters corresponding to the hydroxyl radical were developed by Roeselova et. al. [12]. After equilibration, a 2 ns long NVE simulation was performed on the system.

We extracted 94 snapshots from the above trajectory, at 20 ps intervals. In every snapshot, a single hydroxyl radical was randomly chosen for inclusion in the QM region. The remaining three hydroxyl radicals are described as EFP fragments. The hydroxyl radical in the QM part is then centered along the X and Y axes of the box, leaving the z-axis intact. A sphere with radius of 0.6 nm is drawn from the bond midpoint of the hydroxyl radical. All water molecules falling within this range are included in the QM region, while all the other water molecules are described as EFP fragments. Water molecules that have only one or two atoms situated within the QM sphere are completely included in the QM region. Practically, a sphere of 0.6 nm ensures inclusion of about two shells of water molecules around the hydroxyl radical. Even though inclusion of this number of water molecules into the QM region is computationally challenging, this is necessary for exploring long-range nature of the CT states.

The effective fragment potential (EFP) method [22–25] provides a robust description of the surroundings for a QM subsystem in a quantum mechanics / molecular mechanics (QM/MM) type simulations. EFP is a polarizable model potential for describing non-covalent interactions [25]. EFP represents the interaction energy between

species as a sum of electrostatic, polarization, dispersion and exchange-repulsion terms, all of which are derived from first principles. Combined with QM, EFP provides explicit contributions from electrostatic and polarization terms of the environment [23]. Electrostatic term in EFP is represented by multipoles up to octopoles centered at atoms and bond mid-points. Polarization is obtained from interactions of the electric field with anisotropic polarizability tensors centered at localized molecular orbitals of the fragments [26]. Both terms contribute to the QM Hamiltonian through one-electron integrals, corresponding to polarization embedding [27].

The EFP parameters for water and hydroxyl radical were obtained using mixed-basis set simulations [28]: 6-31G(d) for electrostatic multipoles and 6-31G(3df,2p) for other parameters. Equation of motion coupled cluster for ionization potentials (EOM-IP-CCSD) [29–31] calculations for a cluster containing one $OH^{\bullet}$ and four water molecules were performed in 6-31G basis in Q-Chem 4.2 package [32]. Calculations utilizing Koopmans' theorem [33] for all the snapshots were computed at HF/6-31+G(d) level of theory using GAMESS-US package (v.2014R1) [34, 35]. The absorption spectra for liquid phase simulations are obtained as averages of spectra computed for individual snapshots. Gaussians with a full width at half maximum of 3.73 nm were used to broaden the charge transfer excitations to obtain the gaussian broadened spectra.

### 1.3 Results and Discussions

To analyze the propensity of the hydroxyl radical to the vacuum-water interface, the distribution of the number of the water molecules within 6 Å of $OH^{\bullet}$, corresponding to water molecules included in the QM region, is plotted in a histogram as shown in Figure 1.3. In several snapshots, the hydroxyl radical is surrounded by only a few water molecules such that the hydroxyl radical is more exposed to the vacuum than to the bulk. We refer to these snapshots (n<17) to as hydroxyl radicals in 'gas-phase region'. In a few other snapshots, the hydroxyl radicals are buried deep within the

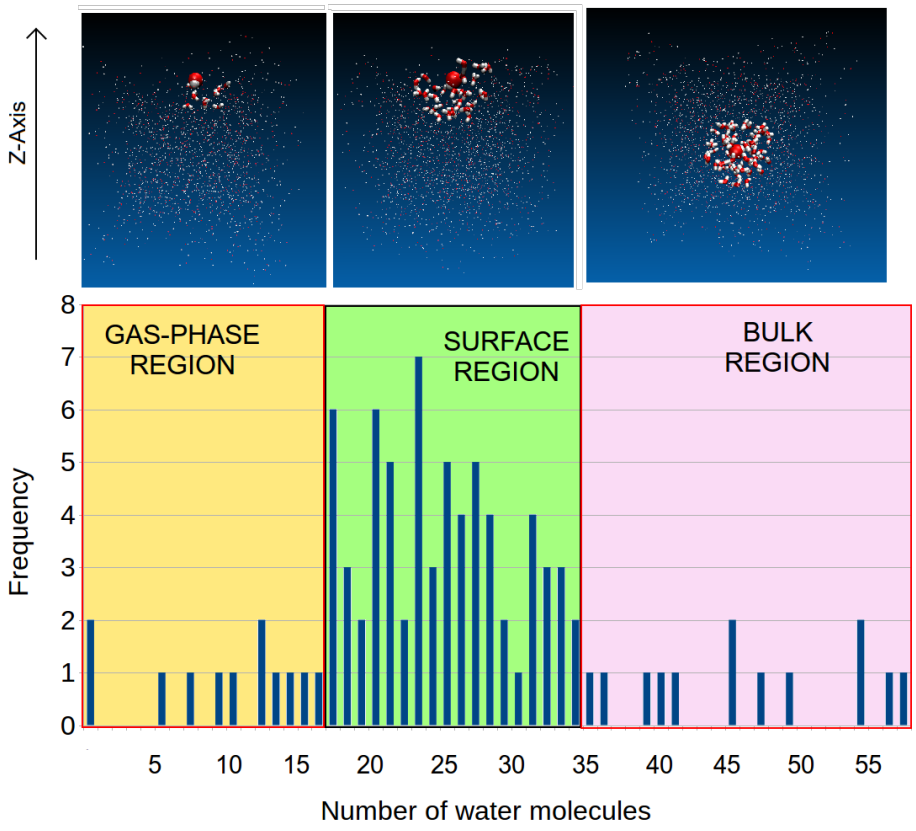



Fig. 1.3. Distribution of snapshots containing $OH^{\bullet}$ in gas-phase, at the vacuum-water interface, and in bulk water. Top: Representative snapshots of hydroxyl radical in gas, surface and bulk phases. van der Waals representation denotes the hydroxyl radical, thick bonds denote QM waters and dots denote the positions of EFP atoms of water.

water layer (n>34) and water molecules within the 6 Å sphere are not exposed to the vacuum layer. Such snapshots are termed as hydroxyl radical in 'bulk region'. It is to be noted that in the bulk region, the hydroxyl radicals are located at least two layers of water molecules underneath the surface. The region that lies in between, referred to as 'surface region', comprises of structures in which the hydroxyl radicals are partially exposed to the surface, or are not well buried inside the bulk water (18<n<34). In these snapshots, there are at the most two hydration shells between the air-water interface and the hydroxyl radical.

As discussed in detail in Ref. [12], and is well reproduced by a selection of snapshots for the present study, the hydroxyl radical is preferentially located in the interfacial region. Indeed, the hydroxyl radical is situated in bulk water in less than 15% of all the snapshots. In Ref. [12], the surface preference of hydroxyl radical is rationalized using potential of mean force calculations. In the present work, we do not reinvestigate the question of surface preference of the hydroxyl radical but rather focus on analysis of the CT transitions specific to the hydroxyl radical in various water environments.

For computing CT states of the hydroxyl radical solvated by small water clusters in Ref. [18], we used equation-of-motion coupled cluster method truncated at double excitations for ionized potentials (EOM-IP-CCSD). In EOM-IP-CCSD, coupled cluster equations are solved for a closed-shell anion, i.e., $OH^-(H_2O)_n$, and ionized states of the anion, i.e., electronic states of the neutral radical, are computed as electron-removing excitations in the following step. As a result, the formalism is free of spin-contamination and provides a balanced description of excited states in radicals. However, it is practically not viable to employ correlated excited state methods such as EOM-CCSD for computing clusters containing more than a dozen of water molecules. To circumvent this issue, in the present work we obtain CT excitations in hydroxyl radical - water droplets using the Koopmans' theorem applied to the anionic system.

Koopmans' theorem [33] states that the ionization potential (IP) from a given molecular orbital is equal to the negative of the orbital energy. Then, the excited state energies in a radical molecule can be estimated as a difference in IPs of the closed-shell anion. Specifically, excitations to a singly-occupied molecular orbital (SOMO) of the radical are related to differences in IP from corresponding orbitals of the anion, i.e., the highest occupied molecular orbital (HOMO) and other occupied anion's orbitals $n$:

$$E^{neutral}_{n\rightarrow SOMO} \approx IP^{anion}_{n} - IP^{anion}_{HOMO} \approx -\epsilon^{anion}_{n} + \epsilon^{anion}_{HOMO} \tag{1.1}$$

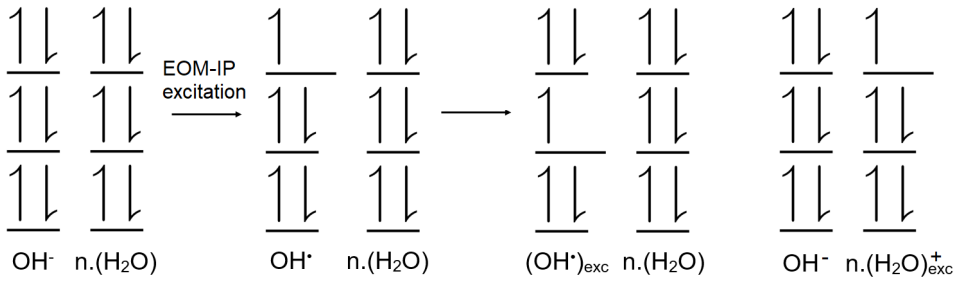



Fig. 1.4. Hartree-Fock and EOM-IP-CCSD reference orbitals in the hydroxyl radical-water system. Note: Orbital energies and ordering are not to scale.

In case of the OH-water system, $n$ corresponds to a lone-pair orbital of one of the surrounding water molecules and HOMO is a lone-pair orbital of $OH^-$. Similarly, in the EOM-IP-CCSD method, the excited state energies are computed as differences in the ionization potentials corresponding to removing an electron from different occupied orbitals. However, in EOM-IP-CCSD, ionized states are correlated by contributions from other electronic transitions. Thus, Koopmans' theorem can be considered as a frozen-orbital approximation to EOM-IP methodology for predicting excited states of radical species.

Graphical representation of electronic states in $OH^\bullet$ - $(H_2O)_n$ cluster is shown in Fig. 1.4. Thus, the IP excitations from OH− - $(H_2O)_n$ cluster produce ground and excited states of hydroxyl radical as well as CT states corresponding to water to $OH^\bullet$ excitations.

Using Koopmans' theorem, oscillator strengths of the excited states in the radical can be obtained from transition dipole moments $\mu$ of involved Hartree-Fock (HF) orbitals of the anion, i.e.:

$$\mu_\alpha^{n\rightarrow HOMO} = \int dr \psi_n^*(r)\ \hat{\mu}_\alpha\ \psi_{HOMO}(r) \tag{1.2}$$

where $\mu_\alpha$ is a projection of the transition dipole moment on a coordinate axis $\alpha = x, y, z$. $\psi_n(r)$ and $\psi_{HOMO}(r)$ are the Hartree-Fock orbitals of the anion. The

oscillator strengths can be obtained from the transition dipole moments using the following relation [36]:

$$f_L = \frac{2}{3}(\mu_\alpha^{n\to HOMO})^2 \Delta E \tag{1.3}$$

where $\Delta E = \epsilon_{HOMO} - \epsilon_n$ is the difference in orbital energies.

Obviously, Koopmans' approximation is only valid for singly-excited states dominated by a single transition, which is the case for CT states in $OH^\bullet$-water systems, as discussed below. The accuracy of CT excited state energies and oscillator strengths using Koopmans' theorem (equations 1.1 - 1.3) and EOM-IP-CCSD is compared for a single snapshot containing 4 water molecules interacting with one hydroxyl radical (see Fig. 1.5).

The excitation energies and oscillator strengths obtained from Koopmans' theorem and EOM-IP-CCSD calculations are plotted in 1.5. The first three states of the system are the electronic states of the hydroxyl radical. They are not shown in Fig. 1.5, as they are not relevant to this particular study. Out of those states, the first two correspond to the near-degenerate components of Π state of the hydroxyl radical and are formally obtained as ionization from one of the two lone pair orbitals on the hydroxyl anion. The lowest of this pair of states is the ground state of the system. The third state, corresponding to $\sigma \to LP$ excitation on the OH radical, is due to ionization from $\sigma$ orbital. The next eight states correspond to the excitation of an electron from occupied LP orbitals of water molecules to the singly-occupied lone-pair orbital of hydroxyl radical, obtained formally as ionization from LP orbitals of waters. These excitations result in the formation of a positively charged hole on water molecules, while the hydroxyl radical becomes negatively charged. As each water molecule possesses two lone-pair orbitals, and four water molecules are included in the calculation, eight charge-transfer states are observed. CT states originating from $\sigma$-orbitals on waters as well as local electronic excitations on waters are higher in energy than a manifold of $H_2O_{LP} \to OH^\bullet_{LP}$ CT transitions. As energetically those

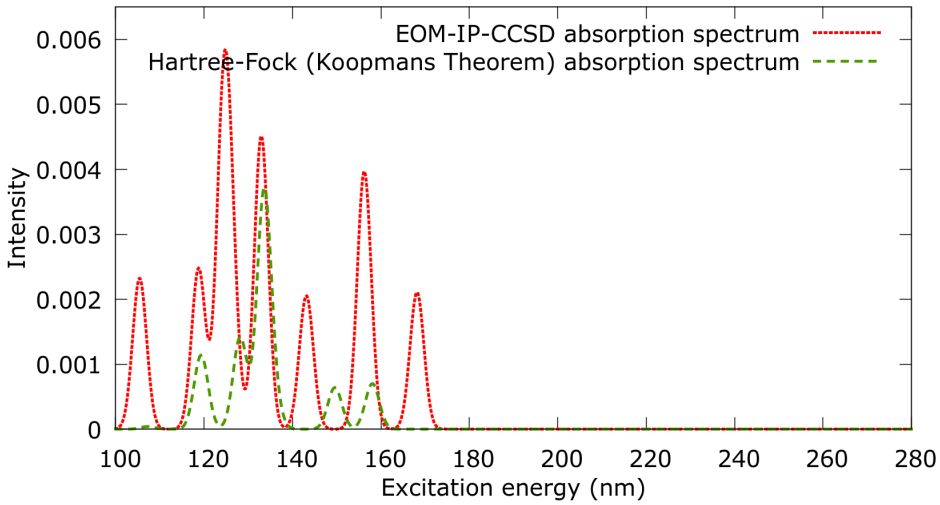



Fig. 1.5. Comparison of excitation energies and oscillator strengths obtained using EOM-IP-CCSD and Koopmans' Theorem (HF).

states fall into UV region dominated by absorption of water, they are less relevant for atmospheric chemistry and are not discussed in this work.

As follows from Figure 1.5, excitation energies obtained from Koopmans' theorem overestimate EOM-IP-CCSD energies in most cases but do not differ by more than 0.5 eV. The ordering of the CT states matches in both methods. A good agreement between the two schemes can be rationalized by the observation that in EOM-IP-CCSD, all CT states, i.e., ionized states from lone pairs on water molecules are dominated by a single electron transition with amplitudes greater than 0.95. Thus, correlation plays a relatively minor role in these CT states and the HF description of the system is reliable.

Figure 1.5 also shows that the oscillator strengths obtained using HF orbitals are underestimated in most cases as compared to the oscillator strengths from the EOM-IP-CCSD method. However, barring a couple of states, oscillator strengths from both methods show very similar trends. Thus, these results suggest that Koopmans' theo-

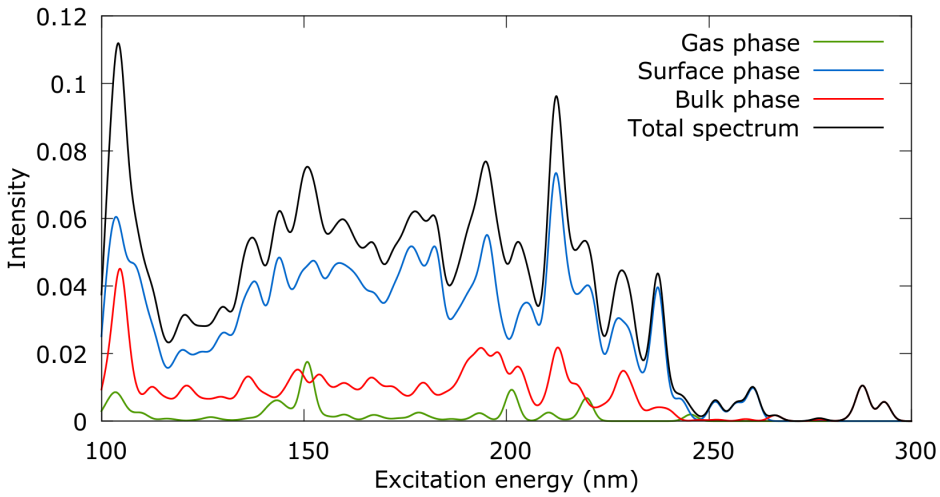



Fig. 1.6. Excitation spectrum of hydroxyl radical in gas phase (green), surface phase (blue), bulk phase (red), and the combined total excitation spectrum of hydroxyl radical taken from all snapshots.

rem can be used as a viable alternative to EOM-IP-CCSD for a qualitative assessment of the nature of the CT states in $OH^{\bullet}$-water complexes.

Now let's turn to the analysis of the full system, in which hydroxyl radical is solvated in a slab of water. Here we use QM/EFP in which the QM subsystem consists of $OH^{-}$-water cluster of 6 Å in radius and is described by the HF/6-31+G(d) method with further application of Koopmans' theorem. Remaining waters (further than 6 Å from the oxygen of the hydroxyl radical) are described as EFP fragments. Figure 1.6 depicts the CT band of the absorption spectrum produced as an average of 94 snapshots selected from the MD trajectory. Additionally, the total spectrum is separated into contributions from snapshots in which the hydroxyl radical is located in the gas phase, at the surface, and in the bulk.

The resulting CT spectrum is a broad band extending to 250 nm, with additional peaks red-shifted to as far as 290 nm. Taking into account that the Koopmans' theorem description of excitation energies is blue-shifted with respect to the EOM-IP-CCSD description, it might be anticipated that, if EOM-IP-CCSD would be

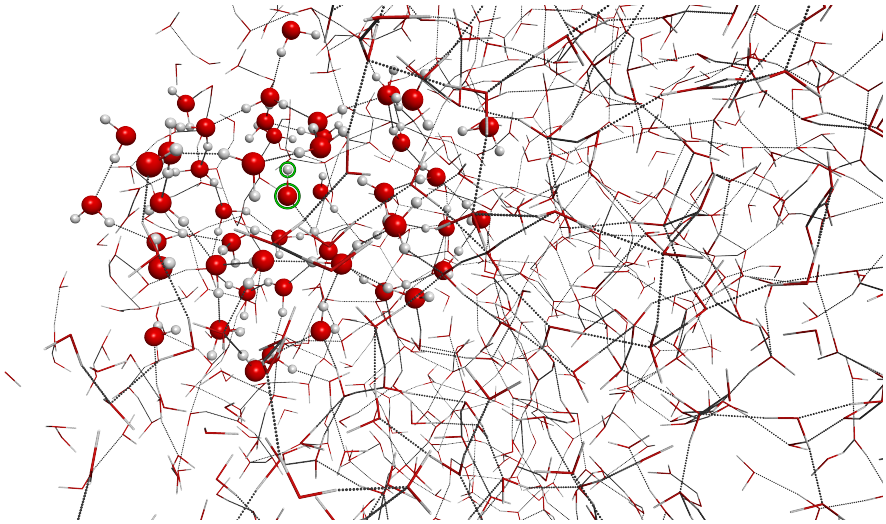


Fig. 1.7. Hydroxyl radical partially immersed in bulk. Hydroxyl radical is highlighted in green, while the water molecules within a 6 Å radius are represented in ball and stick representation. Water molecules beyond the 6 Å radius are represented with thin sticks.

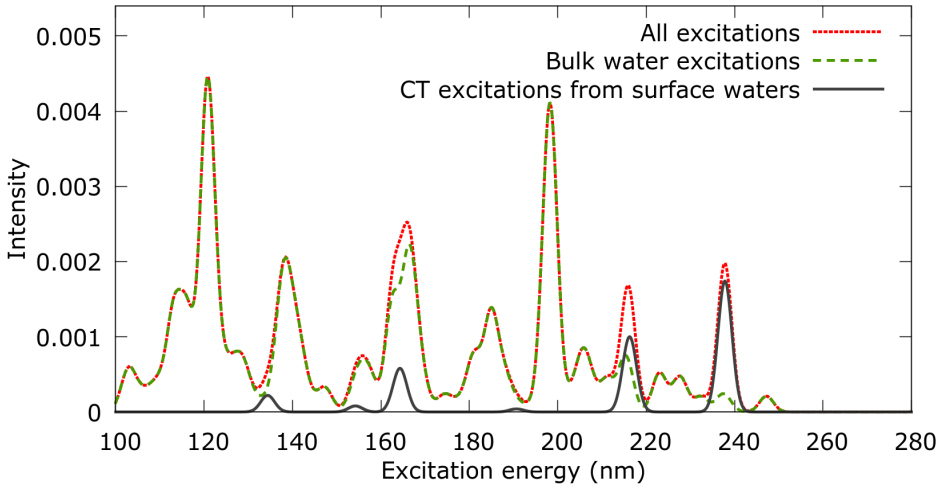



Fig. 1.8. Contribution of surface waters and bulk waters to the CT spectrum of hydroxyl radical partially immersed in bulk.

employed, the overall spectrum could additionally shift to the red by approximately 10-20 nm. As follows from Figure 1.6, the contributions from the gas-phase and bulk phase hydroxyl radical to the total CT spectrum are lower as compared to the contribution from the hydroxyl radical at the surface. However, this effect is primarily due to the higher number of snapshots representing the hydroxyl radical in the surface phase, as against the other two.

The most striking observation in this study is the presence of a few charge-transfer excitations that extend as far as 290 nm, bordering the near-UV region. Interestingly, these red-shifted excitations originate from snapshots in which the hydroxyl radical is buried inside the bulk water.

To understand further the origin of such low-energy transitions, we considered a representative snapshot (Fig. 1.7) in which a hydroxyl radical is located about two water shells beneath the surface. The resulting CT spectrum is plotted in (Fig. 1.8). In this case, we distinguished the contributions of waters in different regions of the simulation cell. As can be seen in the Fig. 1.7, two water molecules at the surface with dangling hydrogens are only partially hydrogen-bonded to other water molecules. We term such water molecules as 'surface waters'. All the other water molecules in the QM region are referred to as 'bulk waters'.

It can be seen that the surface waters contribute to low-energy region of the CT spectrum. This is because the lone-pair orbitals on waters that lack stabilization due to hydrogen bonding are more prone to donate the electron to the hydroxyl radical than the bulk waters that are fully H-bonded to other water molecules. Thus, the interplay of H-bond network, along with the position of the hydroxyl radical in the water slab, seems to play a major role in low-energy CT transitions. Another striking observation is that such low-energy transitions are intense, even though the initial and final orbitals are situated far from each other. Thus, our analysis suggests that distorted H-bonding network at the air-water interface leads to long-range photo-activated charge transfer in near-UV range.

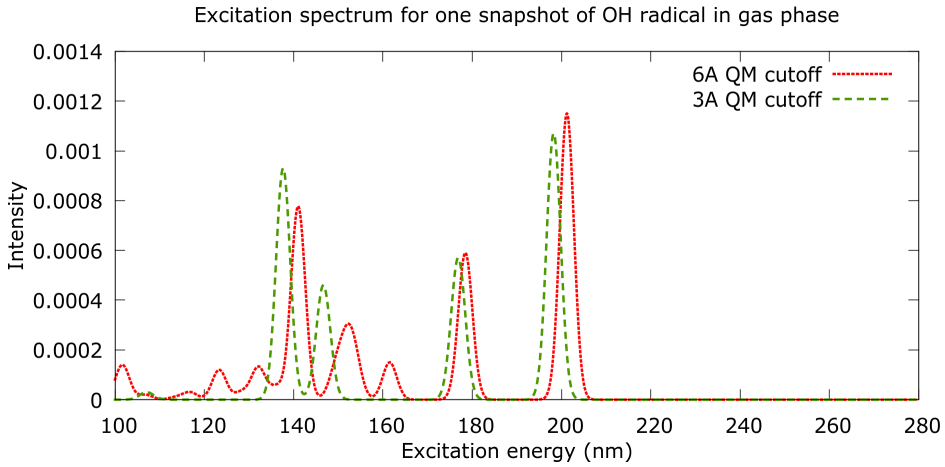



Fig. 1.9. Comparison of small and larger QM regions for hydroxyl radical in gas phase.

### 1.3.1 Effect of cutoffs in modeling the system

This section analyzes validity of a selected model system (explicit QM water droplet of 0.6 nm immersed in a bigger polarizable water droplet described by EFP waters of the size of MD simulation unit cell) for predicting CT absorption spectrum. The first test explores effect of diminishing the size of the QM region from 0.6 nm to 0.3 nm (Figs. 1.9 - 1.11 and Table 1.1). The second set of calculations tests how description of classical waters (polarizable EFP waters versus non-polarizable waters versus no waters) influences the CT spectrum (Fig. 1.12 - 1.14 and Table 1.2). For both tests, we selected one snapshot from each gas phase, surface phase and bulk phase subsets and conducted comparisons of produced CT spectra.

The biggest deviation in spectra is observed for the case of OH radical in bulk water, where the intense peak at around 200 nm is present only in the spectrum with larger QM region. The reason is that this peak is due to a transition from a water molecule that is about 0.5 nm away from the midpoint of the hydroxyl radical. Hence, it is essential to include a large shell of water into QM simulations in order to account for long-range CT excitations.

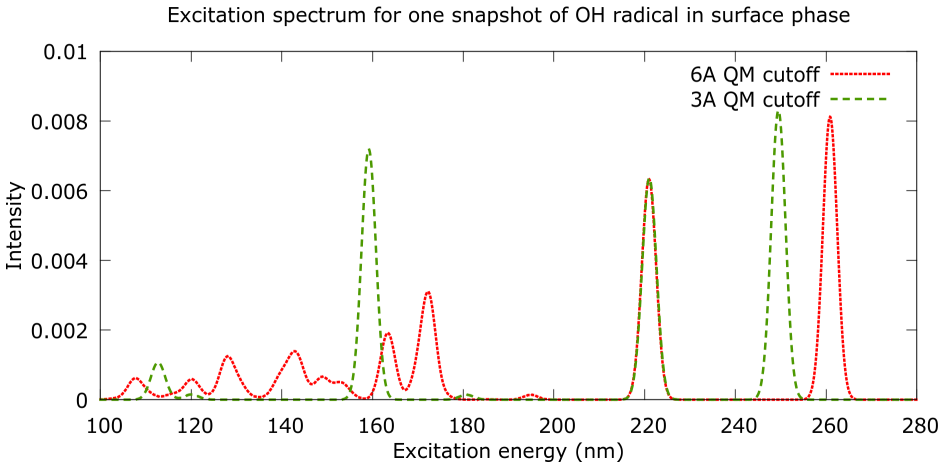



Fig. 1.10. Comparison of small and larger QM regions for hydroxyl radical in surface phase.

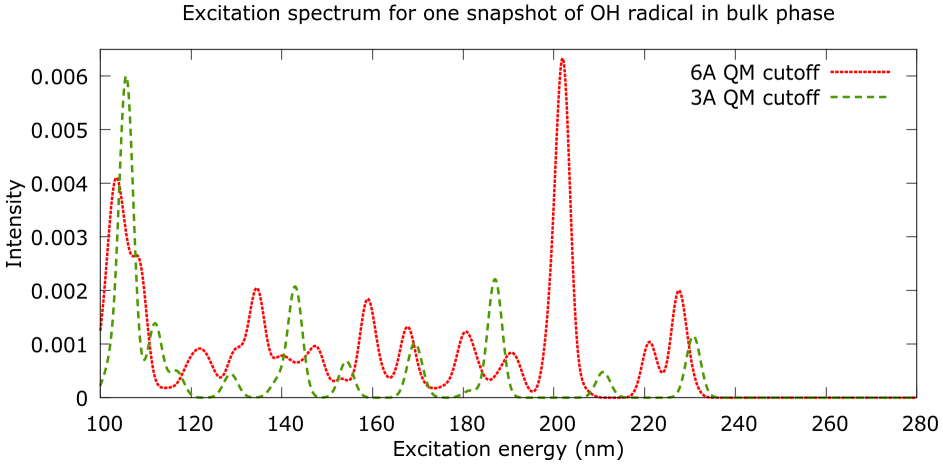



Fig. 1.11. Comparison of small and larger QM regions for hydroxyl radical in bulk phase.

Table 1.1.
Comparison of CT transitions in surface phase hydroxyl radical

| Surface Phase simulations | |
|---|---|
| Peak in 0.3 nm cutoff (nm) | Peak in 0.6 nm cutoff (nm) |
| 248.7 | 260.0 |
| 221.0 | 221.0 |
| 180.1 | 194.7 |
| 158.9 | 162.8 |
| 119.8 | 119.8 |
| 112.7 | 107.4 |

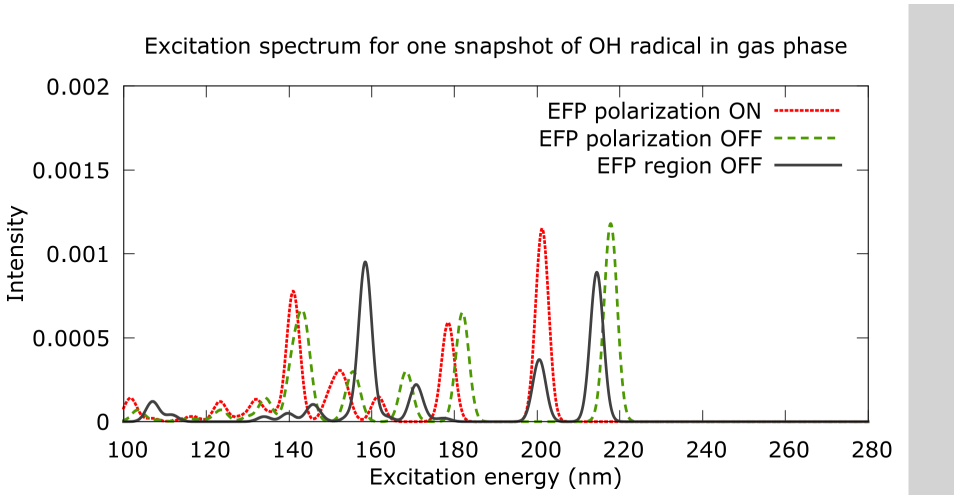



Fig. 1.12. Comparison of different levels of embedding for hydroxyl radical in gas phase.

### 1.3.2 Comparison between polarizable embedding and electrostatic embedding treatments

Figs. 1.12 - 1.14 compare spectra in which water environment is described at different levels of embedding, i.e, polarizable embedding (with QM/EFP model), electrostatic embedding (with QM/EFP electrostatic part only) and no embedding.

The lack of description of the water environment in no-embedding model produces prominent red shifts in the CT excitations. This is particularly noticeable in the case of the gas phase spectrum, where the first CT transition is red-shifted by over 15 nm. It can also be seen that the lack of polarization can also cause a red-shift to the tune of about 15 nm, even though the discrepancies between spectra computed with electrostatic and polarizable embeddings are modest. Overall, hydroxyl radical in bulk phase produces the most complex spectra with varying effects of polarizable embedding on different excitations.

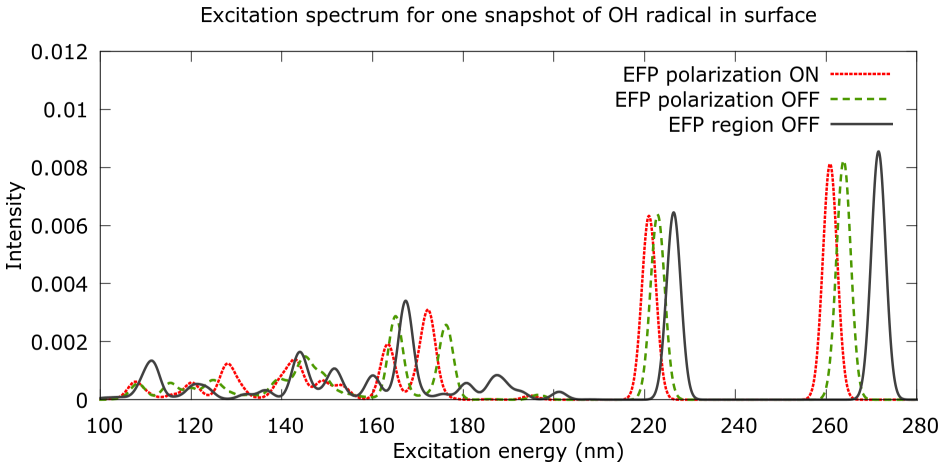



Fig. 1.13. Comparison of different levels of embedding for hydroxyl radical in surface phase.

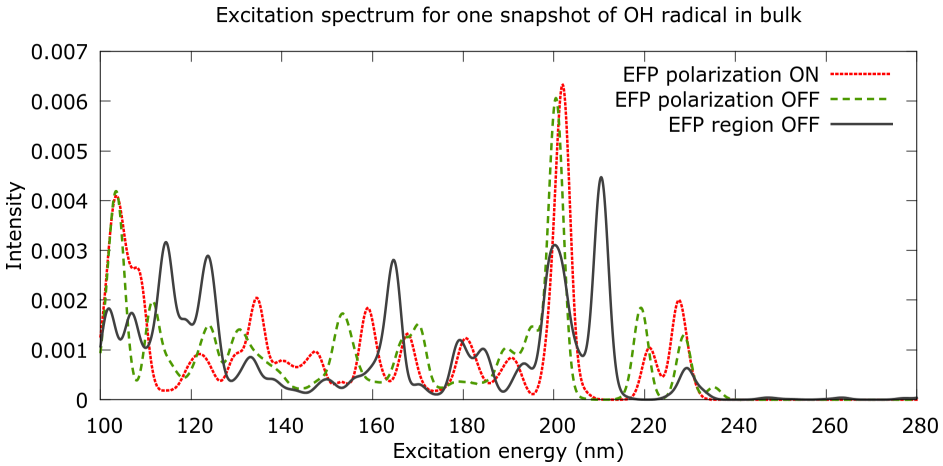



Fig. 1.14. Comparison of different levels of embedding for hydroxyl radical in bulk phase.

Table 1.2.
Comparison of different embedding levels in gas phase hydroxyl radical CT spectrum

| Gas Phase Simulations | | |
|---|---|---|
| QM/EFP (elec+pol) | QM/EFP (elec) | QM (no embedding) |
| 201.6 | 217.9 | 275.6 |
| 178.7 | 182.3 | 213.8 |
| 161.7 | 168.2 | 174.0 |
| 154.1 | 156.2 | 170.7 |
| 152.0 | 155.1 | 163.4 |
| 151.5 | 148.1 | 150.0 |

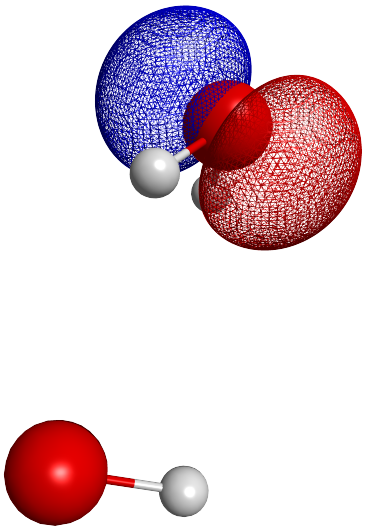


Fig. 1.15. $P_z$ orbital of water hemibonded to hydroxyl radical in bulk

### 1.3.3 Contribution of hemibonded waters

Hemibonding can be qualitatively defined as the interaction due to a 2-center 3-electron partial bond of the singly occupied pi-orbital of the hydroxyl radical and the lone pair orbital of water [37]. Hemibonded complexes of hydroxyl radical and water have been studied extensively [13, 37], and we probe the contribution of such hemibonded waters to the total electronic spectrum obtained via QM/EFP simulations. While hemibonded structures of water-hydroxyl radical clusters do not correspond to global minimum geometries, they produce one of the numerous possible local minima conformations. The occurrence and extent of hemibonding are largely dependent on the applied computational method (correlated wave function methods vs DFT vs classical force fields) [38, 39]. While hemibonded structures occur rarely along MD trajectory, we selected two such snapshots for spectral analysis (see Figs. 1.15 and 1.16.

One interesting observation from the hydroxyl radical in bulk is the following: while the hydroxyl radical is directly hydrogen bonded to two other water molecules, the lowest CT excitation originates from the hemibonded water, although the intensity

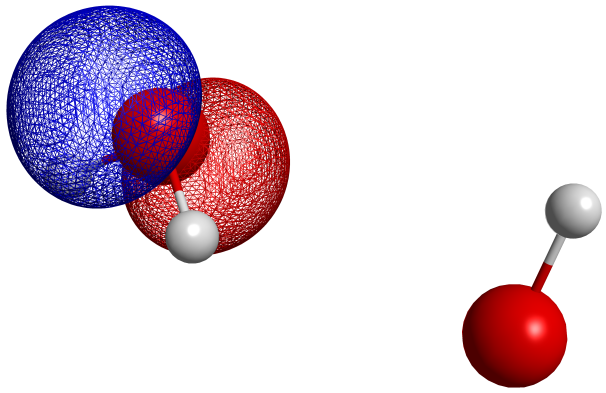


Fig. 1.16. $P_z$ orbital of water hemibonded to hydroxyl radical on surface

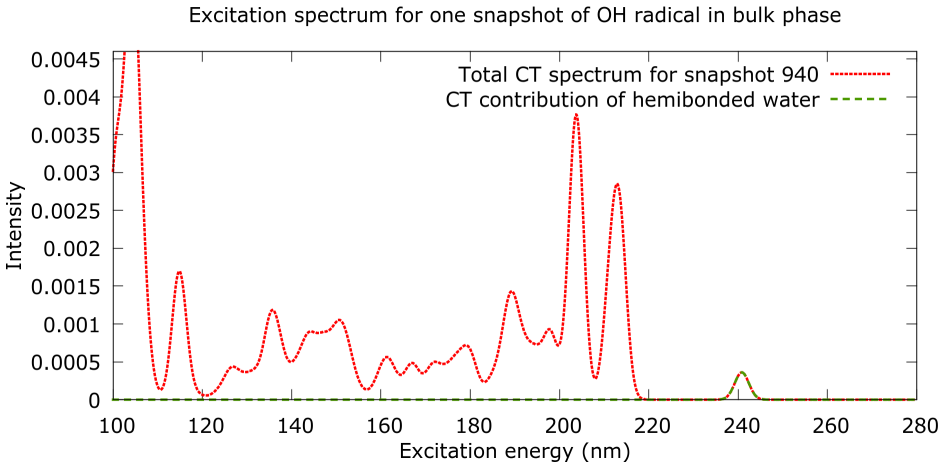



Fig. 1.17. CT spectrum of one snapshot of hydroxyl radical in bulk water (54 water molecules).

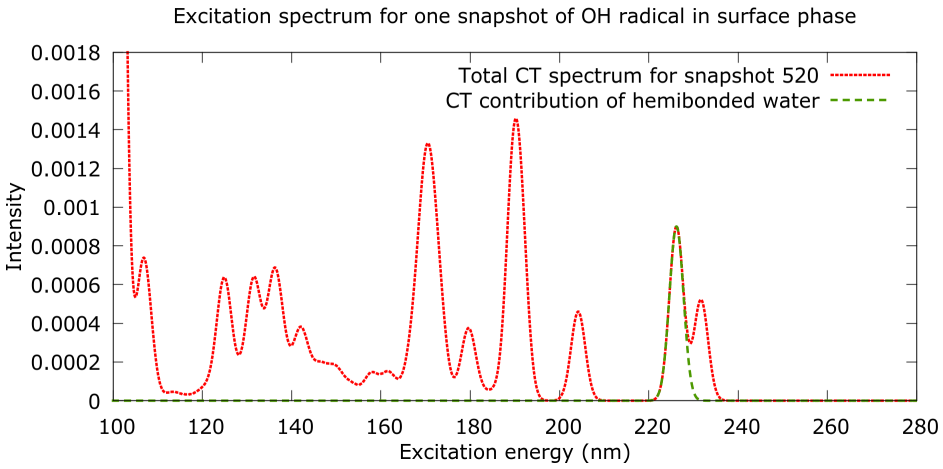



Fig. 1.18. CT spectrum of one snapshot of hydroxyl radical in surface water (29 water molecules).

is fairly weak. On the other hand, with respect to the surface hydroxyl radical(Fig. 1.16), the lowest CT excitation originates from a directly hydrogen bonded water, although the hemibonded water contributes to a very strong CT excitation pretty close to the lowest CT excitation. Hence, with the limited number of snapshots available, we can still deduce the low-lying and bright nature of CT states originating from hemibonded waters. However, based on our analysis, CT excitations due to hemiboned waters are found below 240 nm range and thus do not contribute to the most red-shifted feature of the CT spectrum.

## 1.4 Conclusions

The question of how a single hydroxyl radical interacts with large water molecule clusters is addressed here. Earlier MD simulations have given insights on the preference of hydroxyl radicals in water. We have extended this study further to understand the long-range charge transfer behavior of large clusters. We have made use of Koopmans' theorem to model the CT excitations of water molecules on hydroxyl radicals. The CT contributions to gas-phase, surface and bulk hydroxyl radicals were differentiated and studied in detail. We believe that the hydroxyl radical situated in bulk is involved in a few bright CT excitations that are red shifted by about 30 nm as compared to hydroxyl radicals situated in the surface and in the bulk. This leads to the total CT spectrum of the hydroxyl radicals to be red-shifted very close to the near-UV region of the spectrum. To further understand the role of water molecules, we probed a single snapshot of hydroxyl radical immersed in the bulk. In this case, we were able to distinguish the contributions of surface waters against the bulk waters. The surface waters contributed to a few bright CT transitions that were lower in energy in comparison to the other CT excitations from the bulk water. We believe that the roles of hydroxyl radical dissolved in bulk water and water molecules present at the surface have a very interesting photochemical behavior that could be involved in numerous atmospheric processes.

# 2. DEVELOPMENT OF MACROMOLECULAR FORCE FIELD (MEFP) AND APPLICATION TO SIMULATION OF EXCITED STATES

*The following chapter is published in a peer reviewed journal.*

*DOI: 10.1021/acs.jpcb.6b04166*

The Effective Fragment Potential (EFP) approach, which can be described as a non-empirical polarizable force field, affords an accurate first-principles treatment of non-covalent interactions in extended systems. EFP can also describe the effect of the environment on the electronic properties (e.g., electronic excitation energies, ionization and electron-attachment energies) of the subsystem via the QM/EFP (quantum mechanics/EFP) polarizable embedding scheme. The original formulation of the method assumes that the system can be separated, without breaking covalent bonds, into closed-shell fragments, such as solvent and solute molecules. Here we present an extension of the EFP method to macromolecules (mEFP). Several schemes for breaking a large molecule into small fragments described by EFP are presented and benchmarked. We focus on the electronic properties of molecules embedded into a protein environment and consider ionization, electron-attachment, and excitation energies (single-point calculations only). The model systems include chromophores of green and red fluorescent proteins surrounded by several nearby amino-acid residues and phenolate bound to the T4 lysozyme. All mEFP schemes show robust performance and accurately reproduce the reference full QM calculations. For further applications of mEFP we recommend either the scheme in which the peptide is cut along the $C_\alpha$-C bond giving rise to one fragment per amino-acid, or the scheme with two cuts per amino-acid, along $C_\alpha$-C and $C_\alpha$-N bonds. When using these fragmentation schemes, the errors in solvatochromic shifts in electronic energy differences

(excitation, ionization, electron detachment, or electron attachment) do not exceed 0.1 eV. The largest error of QM/mEFP against QM/EFP (no fragmentation of the EFP part) is 0.06 eV (in most cases, the errors are 0.01-0.02 eV). The errors in the QM/MM calculations with standard point charges can be as large as 0.3 eV.

## 2.1 Introduction

Electronic processes in complex environments are at the heart of numerous phenomena of fundamental and societal importance, such as catalysis, solar energy harvesting, and photovoltaics. Predictive computational modeling is instrumental for advancing our mechanistic understanding of the redox and photo-induced processes in condensed phase; it requires a combination of quantum mechanical methods and an appropriate description of the environment (solvent, protein, molecular solids, etc). The effect of the environment is multifaceted. First, it spatially confines the reacting species: for example, the protein matrix controls the structures of reaction centers in enzymes and restricts the range of motions of chromophores in photoactive proteins. Second, the environment serves as a thermal bath. Third, the environment often strongly perturbs the electronic structure of the system by local and long-range electric fields. Preferential stabilization of some electronic states relative to others leads to solvatochromism, solvent-induced shifts of electronic excitation energies, which can be as large as 1 eV [40, 41]. The effect on ionization/electron attachment energies (quantities determining the redox potentials) is even more pronounced — shifts of several electron-volt in polar solvents and in proteins are rather common [42–44]. Fourth, the environment itself can be perturbed by the solute: the changes in electronic structure of an active center (for example, ionization) induce both structural and electronic response of the solvent. In sum, even when the electronic processes are confined to a well defined domain (such as, for example, solvated or protein-bound chromophores), electronic interactions between the solute and the solvent need to be carefully accounted for.

Of course, full quantum mechanical (QM) treatment of the entire solvent-solute system would correctly describe these effects. However, such brute-force approach is impractical due to a steep computational scaling of electronic-structure methods [45]. For example, robust and reliable equation-of-motion coupled-cluster methods [46, 47] with single and double substitutions (EOM-CCSD) can only be applied for moderate-size systems (20-30 heavy atoms), even when combined with efficient parallelization and other algorithmic enhancements [48, 49]. Low-cost time-dependent density functional (TD-DFT) and scaled-opposite-spin configuration interaction singles with doubles correction, SOS-CIS(D), methods [50–52] can treat considerably larger systems (a hundred of heavy atoms), but not sufficiently large to model, for example, bulk solvation or an entire protein.

To overcome this hurdle, one can employ a more approximate description of the environment while treating the solute quantum mechanically. Several strategies of various degree of sophistication have been developed towards this end. The simplest one is to describe the solvent by a polarizable continuum model. Multiple flavors of these methods exists [53–55]; some have been shown to be capable of capturing solvatochromic effects [56] and yielding reasonably accurate redox potentials [57] in aqueous solutions. However, the description of specific solvent-solute interactions (such as hydrogen bonding) and inhomogeneous environments (such as proteins and interfaces) are beyond the reach of these models because of their implicit solvent nature. Thus, explicit solvent models, which are based on more detailed representation of the solvent, are needed to tackle proteins, interfaces, and complex molecular solids.

A popular QM/MM approach [58–61] allows one to combine an arbitrary complex QM treatment of the solute (reaction center) with an atomistic description of the environment. The most common variants of QM/MM use standard (non-polarizable) force fields in which charges on the MM atoms are fixed. Thus, electronic interactions between the QM and MM parts are described as perturbation of the QM part by the (fixed) electrostatic field of the environment. The electronic response of the environment to the changes in electronic structure of the solute is neglected, which

may introduce undesirable errors in relative state energies. Moreover, the charges used in these force fields are known to be too large (to compensate for the lack of polarization); this can lead, for example, to overestimation of reorganization energies in electron-transfer calculations [62]. Finally, an empirical nature of the MM force-fields limits their predictive power.

Effective Fragment Potential (EFP) approach [63–70] has been developed with an aim to address these limitations of QM/MM. EFP is based on a rigorous representation of different components of inter-molecular interactions (electrostatics, polarization, dispersion, exchange-repulsion and optional charge-transfer) based on perturbative expansion; it can be described as a non-empirical parameter-free polarizable force-field. In QM/EFP, the interactions between the QM and EFP subsystems include both electrostatics and mutual polarization. The extension of EFP to electronically excited and ionized states [42, 71–73] includes polarization response of the EFP environment to changes in the electronic distribution of the solute. The existing implementations allow one to combine EFP with many popular electronic structure methods including CC, EOM-CC, TD-DFT, CIS(D), and SOS-CIS(D).

EFP is similar to polarizable embedding (PE) approach [74–77]. The description of electrostatics and polarization is nearly identical in EFP and PE; however, EFP features more rigorous treatment of dispersion and exchange-repulsion. For modeling intermolecular interactions, similar strategies based on multipolar expansion have been utilized [78–80]. Detailed reviews of various flavors of fragment-based methods can be found in Refs. 81, 82.

The original formulation of the EFP method assumes that the system can be separated, without breaking covalent bonds, into closed-shell rigid fragments, such as solvent and solute molecules. In order to extend EFP to macromolecules, one needs to figure out how to break large molecules into fragments that could then be described by EFP. A similar problem often arises in traditional QM/MM calculations when one needs to separate the QM part from the rest of the system by breaking covalent bonds. These broken bonds need to be saturated (or capped) for subsequent

calculations. Various approaches including hydrogen atom caps and frozen molecular orbitals have been developed [59, 83–86]. In the context of fragmentation-based methods [82], schemes such as link atoms, molecular tailoring approach (MTA) [87] and molecular fractionation with conjugate caps (MFCC) [88–90] have been used. Flexible EFP schemes have also been proposed [91–94].

In this work, we break the protein into amino-acid fragments capped by hydrogen link atoms [83]. We benchmark several schemes of breaking a macromolecule into the effective fragments: breaking (i) along the peptide bond, (ii) along either the $C_\alpha$-C or $C_\alpha$-N bonds, or (iii) along both the $C_\alpha$-C and $C_\alpha$-N bonds.

We focus on the electronic effects (such as excitation, ionization, and electron-attachment energies) of molecules embedded in the protein matrix. Our goal here is to enable single-point calculations of electronic properties and energy differences rather than structure optimization. The test cases include excitation, ionization, and electron-attachment energies of common chromophores in the presence of a protein backbone.

The structure of the paper is as follows. The next section presents a brief overview of the EFP scheme and then introduces different approaches for splitting macromolecules into effective fragments. Computational details are given in Sec. 2.3. Sec. 2.4 presents the results of benchmark calculations and discusses relative merit of different fragmentation schemes. Our concluding remarks are given in Sec. 2.5.

## 2.2 Methodology

### 2.2.1 Effective Fragment Potential scheme: QM-EFP interactions

The EFP method [63–70] describes non-covalent interactions by using perturbation theory starting from the non-interacting (unperturbed) fragments. The total interaction energy ($E_{\mathrm{EFP-EFP}}$) between the effective fragments is decomposed into

four contributions: electrostatic ($E^{\mathrm{elec}}$), polarization ($E^{\mathrm{pol}}$), dispersion ($E^{\mathrm{disp}}$), and exchange-repulsion ($E^{\mathrm{exrep}}$):

$$E_{\mathrm{EFP-EFP}} = E^{\mathrm{elec}} + E^{\mathrm{pol}} + E^{\mathrm{disp}} + E^{\mathrm{exrep}}. \tag{2.1}$$

The interactions between the QM part and the effective fragments are computed by the polarization embedding approach in which the Coulomb and polarization parts of the EFP sub-system contribute to the quantum Hamiltonian ($H$) via one-electron terms:

$$\hat{H} = \hat{H}_0 + \left\langle p \left| \hat{V}^{\mathrm{Coul}} + \hat{V}^{\mathrm{pol}} \right| q \right\rangle p^{\dagger} q, \tag{2.2}$$

where $H_0$ is an unperturbed Hamiltonian of the QM part, $\hat{V}^{\mathrm{Coul}}$ and $\hat{V}^{\mathrm{pol}}$ are electrostatic and polarization perturbations and $|p\rangle$, $|q\rangle$ are the atomic orbitals in the QM part. The electrostatic and polarization terms are the most important ones, as far as the effect of the EFP environment on the electronic properties of the QM solute are concerned. In the discussion below, we focus on these terms; the complete details of our EFP implementation including the treatment of dispersion and exchange-repulsion contributions can be found in Refs. 66, 67, 69, 70.

Both terms represent classic electrostatic interactions; $\hat{V}^{\mathrm{Coul}}$ is a Coulomb potential due to the fragments' nuclear charges and their electron density represented by the multipole expansion, whereas $\hat{V}^{\mathrm{pol}}$ describes electrostatic field due to the induced dipoles (thus, it depends on the polarizabilities of the fragments). The induced dipoles are found by an iterative self-consistent procedure, such that the converged dipoles are fully consistent with each other and with the electronic wave function of the QM part.

Neglecting the dispersion and exchange-repulsion contribution terms, the total ground-state (or, more precisely, reference-state) energy of the QM/EFP system is:

$$E_{\mathrm{ref}}^{\mathrm{QM/EFP}} = \langle \Phi_{\mathrm{ref}} | \hat{H}_0 + \hat{V}^{\mathrm{Coul}} + \hat{V}_{\mathrm{ref}}^{\mathrm{pol}} | \Phi_{\mathrm{ref}} \rangle + E_{\mathrm{Coul}} + E_{\mathrm{pol,ref}}, \tag{2.3}$$

where $\Phi_{\mathrm{ref}}$ is the reference-state wave function, $\hat{V}^{\mathrm{Coul}}$ and $\hat{V}_{\mathrm{ref}}^{\mathrm{pol}}$ are Coulomb and polarization EFP contributions to the Hamiltonian. Subscript *ref* indicates that the

induced dipoles correspond to the electronic density of the reference state, i.e., ground electronic state in excited-state calculations, a closed-shell neutral state in the EOM-IP calculations of radical-cations, etc. $E_{\mathrm{Coul}}$ is the electrostatic fragment-fragment interaction energy and $E_{\mathrm{pol,ref}}$ is the self-consistent reference-state polarization energy of the QM/EFP system; it is computed using converged induced dipoles of the fragments and the fields due to the static fragment multipoles and the nuclei and electrons of the quantum region. Note that the polarization contributions appear both in the quantum Hamiltonian through $\hat{V}^{\mathrm{pol}}$ and in the EFP energy as $E_{\mathrm{pol,ref}}$; this is because self-consistency precludes the separation of the QM-EFP and EFP-EFP polarization contributions [63].

Electrostatics is the leading term in the total interaction energy in hydrogen-bonded and polar systems. $E^{\mathrm{Coul}}$ between the effective fragments consists of charge - charge, charge - dipole, charge - quadrupole, charge - octupole, dipole - dipole, dipole - quadrupole, quadrupole - quadrupole terms. At close separation between the fragments (or between a fragment and the QM region), the charge penetration may become significant; to correct classical multipoles for possible charge-penetration, several types of damping functions can be used [95]. Here we employ an exponential damping in which the charge-charge interaction energy is damped using the following equation [96]:

$$f_{kl}^{ch-ch} = 1 - \frac{b^2}{b^2 - a^2} \exp\left(-ar_{kl}\right) - \frac{a^2}{a^2 - b^2} \exp\left(-br_{kl}\right), \tag{2.4}$$

where $r_{kl}$ is the distance between multipole points $k$ and $l$ and damping parameters $a$ and $b$ are for the multipole points $k$ and $l$.

For the QM-EFP interactions, the electrostatic potential of the molecule (QM part) at point $x$ is expressed by the multipole expansion around $k$ points located at the atomic centers and bond midpoints:

$$V_k^{\mathrm{Coul}}(x) = q_k T(r_{kx}) - \sum_{a}^{x,y,z} \mu_a^k T_a(r_{kx}) + \frac{1}{3} \sum_{a,b}^{x,y,z} \Theta_{a,b}^k T_{a,b}(r_{kx}) - \frac{1}{15} \sum_{a,b,c}^{x,y,z} \Omega_{a,b,c}^k T_{a,b,c}(r_{kx}), \tag{2.5}$$

where $q$, $\mu$, $\Theta$ and $\Omega$ are the net charge, dipole, quadrupole and octupole located at $k$ points, respectively. $T$ are the electrostatic tensors of ranks zero to three. Interaction of the QM electronic density with multipole charges are also augmented by gaussian-type damping function, such that Eq. (2.5) becomes

$$V_k^{\mathrm{Coul}}(x) = (q_k^{nuc} + q_k^{ele}(1 - \exp(-\alpha_k r_{kx}^2))T(r_{kx}) - \sum_a^{x,y,z} \mu_a^k T_a(r_{kx}) + \frac{1}{3}\sum_{a,b}^{x,y,z} \Theta_{a,b}^k T_{a,b}(r_{kx}) - \frac{1}{15}\sum_{a,b,c}^{x,y,z} \Omega_{a,b,c}^k T_{a,b,c}(r_{kx}) \quad (2.6)$$

where $q_k^{nuc}$ is the nuclear charge and $q_k^{ele}$ is the electronic charge on multipole point $k$, respectively. Thus, only the electronic charges are damped (smeared) by gaussians. Damping parameters $a$, $b$, and $\alpha$ in Eqns. (2.4) and (2.6) are determined by minimizing the difference between the electrostatic potentials from the damped multipole expansion and the electronic wave function in the parameter-generating calculation for each fragment.

Polarization is a many-body term, which is computed self-consistently because the induced dipoles of one fragment depend on the static electric field and on the induced dipoles of other fragments. Polarization energy of the QM-EFP system is computed as

$$E_{\mathrm{pol,ref}} = \frac{1}{2}\sum_k (-\mu^k(F^{mult,k} + F^{nuc,k}) + \tilde{\mu}^k F^{ai,k}), \quad (2.7)$$

where $\mu^k$and $\tilde{\mu^k}$ are the induced and conjugated induced dipoles at the distributed point, $k$ (at the centers of the localized molecular orbitals); $F^{mult,k}$ is the external field due to the static multipoles and nuclei of other fragments; $F^{nuc,k}$ and $F^{ai,k}$ are the fields due to the nuclei and electrons of the QM region ($F^{ai,k}$ is computed using the reference-state density). The induced dipole at each polarizability point $k$ is given by

$$\mu^k = \sum_k \alpha^k (F^{mult,k} + F^{ind,k} + F^{nuc,k} + F^{ai,k}), \quad (2.8)$$

where $F^{ind,k}$ is the field due to other induced dipoles. The many-body contribution is solved iteratively, as the induced dipole on a particular fragment depends on the

values of the induced dipoles of all other fragments and the wave function of the QM region. The distributed polarizabilities ($\alpha^k$) are located at the centers of the localized molecular orbitals (LMO).

In order to avoid polarization collapse when effective fragments approach each other (which is always the case in macromolecules, see below), the polarization damping functions are applied to the electric fields produced by multipoles and induced dipoles in Eqns. (2.7) and (2.8): [95]

$$f_{kl}^{pol} = 1 - \exp\left(-\sqrt{c_k d_l} r_{kl}^2\right)\left(1 + \sqrt{c_k d_l} r_{kl}^2\right), \tag{2.9}$$

where $c_k$ and $d_l$ are the damping parameters on polarizability points $k$ and $l$, respectively. The default values of the polarization damping parameters are 0.6 for non-covalent neutral fragments and 0.1 for small anions and cations such as halide and alkali ions. Appropriate values of the polarization damping parameters are explored in Sec. 2.4.

### 2.2.2 Breaking macromolecules into fragments

In its original formulation, the EFP method cannot be applied to macromolecules such as biopolymers, peptides, proteins, lipids, or DNA, because it was designed as a rigid-fragment model for treating interactions between (small) molecules in clusters and liquids. The building block of EFP (the so called "fragment") is a molecule rather than an atom (as typical in classical force-fields). The EFP scheme does not include covalent interactions between the fragments.

In order to enable conformational flexibility of macromolecules while keeping the calculations affordable, the polymer chain should be split into small effective fragments and the parameters for each of these fragments need to be generated. We prepare the parameters for individual fragments using the following procedure. First, we saturate the dangling bonds of each fragment using a capping group, to obtain well-behaved closed-shell structures. In this work, we use hydrogens as the capping groups. Once the parameters for the capped closed-shell molecule are generated, we

remove the capping groups and the associated parameters (multipoles and polarizabilities) from the fragment. This scheme enables calculation of the electrostatic and polarization energies in the macromolecule represented by effective fragments as well as electrostatic polarizable embedding using QM/EFP. The specific details of the application of this scheme to polypeptides (such as positions of the cuts) are discussed below. Schemes for other macromolecules will be developed in future work, using similar strategies.

There are different ways to break macromolecules such as proteins or DNA into fragments, depending on the position of the cut between two covalently bound residues [97]. For polypeptides, we consider the following cutting schemes (shown in Fig. 2.1):

- cutting along the peptide link;
- cutting along the $C_\alpha$-C bonds;
- cutting along the $C_\alpha$-N bonds;
- cutting along both the $C_\alpha$-C and $C_\alpha$-N bonds.

In the first three schemes, each fragment consists of a single amino-acid, whereas the last one yields two fragments per amino-acid, one fragment containing the peptide group and another containing the residue, see Fig. 2.1.

The advantage of fractioning the protein along the peptide bond is that it yields "symmetric" fragments. However, in this scheme highly polarized bonds are broken, which may lead to unphysical multipole and polarizability values. One may expect that the fragments obtained by cutting either the $C_\alpha$-C or $C_\alpha$-N bonds produce a more accurate representation of multipoles and polarizabilities near the cuts. Furthermore, smaller fragments and cutting both the $C_\alpha$-C and $C_\alpha$-N bonds will aid future extension of this scheme to dynamics, as this scheme ensures flexibility along the two most important degrees of freedom in proteins, dihedral $\phi$ and $\psi$ angles defining the conformation.

In all schemes, the capped fragments mimic the protein; however the neighboring fragments have duplicated points (overlapping area) due to the hydrogen caps. The

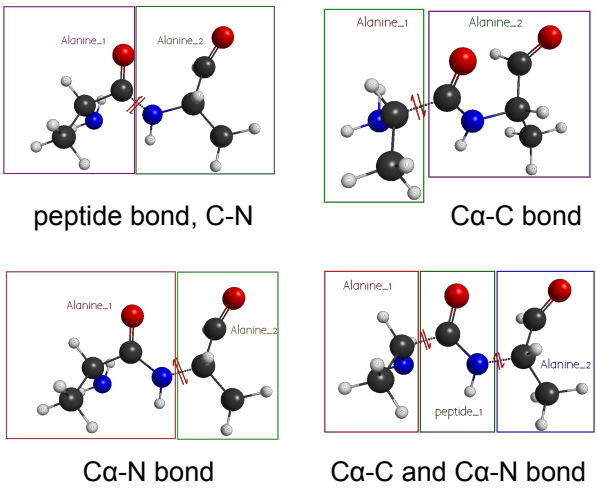



Fig. 2.1. Various cutting schemes for polypeptides.

multipole expansion points and polarizability expansion points are extended on each isolated capped fragment by the standard procedure (see Section 2.3) and the multipoles and polarizabilities at the duplicated expansion points are removed. In order to maintain the net integer charge on each fragment, the monopole expansion of each cap is redistributed to the nearest atom [98]. This method is called Expand-Remove-Redistribute (ERR, see Figure 2.2). Polarizability points located through the 'cut' bonds are removed to avoid over-polarization of the neighboring fragments.

We have implemented the mEFP approach in libefp [69], which is interfaced with the *Q-Chem* package [99, 100].

## 2.3 Computational details

The benchmark set was chosen with the focus on electronic properties. In particular, we consider the effect of the protein scaffold on the electronically excited states of chromophores and on their redox properties. We first quantify the effect of the protein environment by comparing the properties of the bare chromophores (small QM) with the full QM calculations. We then test different fragmentation strate-

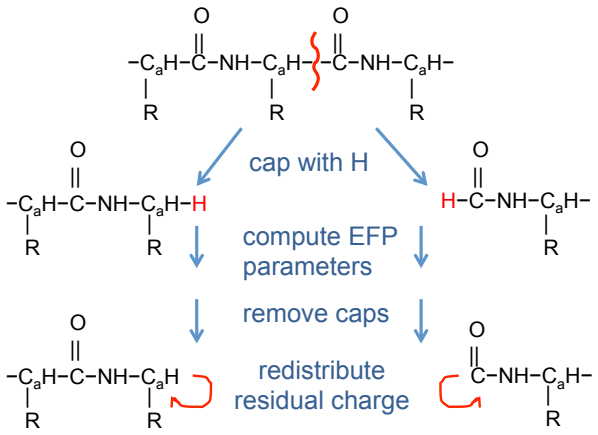



Fig. 2.2. The Expand-Remove-Redistribute procedure (ERR) for generating parameters of the effective fragments in a macromolecule.

gies and compare the results of QM/mEFP against full QM calculations and against QM/EFP (when no breaking along covalent bonds in the EFP part is performend). Finally, we consider QM/MM calculations with fixed point charges using standard force-fields [101, 102].

While setting up such benchmark calculations, one should keep in mind several important points. The QM/MM and QM/mEFP calculations can only be compared with the full QM calculation when the target electronic properties can be described as the electronic properties of the QM region perturbed by the environment. In the calculations of electronically excited states, that means that one should only consider the states in which both the initial and the target molecular orbitals are confined within the QM part. In calculations of ionization or electron attachment energies, the spin density of the target state needs to be localized in the QM region. Within Koopmans theorem, the HOMO (highest occupied molecular orbital) of the initial state is representative of the density of the unpaired electron; however, we found several examples where the shape of the HOMO is quite different from the spin-density of the ionized/electron detached state. Since spin density represents the actual shape of the hole (or the unpaired electron), one needs to always consider spin density,

whereas the shape of the HOMO is only relevant for accessing the validity of Koopmans' description of ionization. In a similar vein, one should consider spin density when analyzing electron-attached states. In addition to the above consideration, one should also assess whether the states in question are bound or unbound with respect to electron ejection. This is particularly relevant for electron-attached states and excited states of anionic systems.

The following model systems were used in the benchmark calculations:

1. GFP chromophore in its anionic (deprotonated) and neutral forms surrounded by four nearby amino-acid residues (Fig. 2.3);

2. The anionic form of the mPlum chromophore with nearby amino-acids and one water molecule (Fig. 2.3);

3. A phenolate molecule with four amino-acids from the T4 lysozyme M102E/L99A mutant (Fig. 2.4).

**GFP model systems.** We constructed the model system as follows. We begin from the X-ray structure of enhanced GFP (pdb id:1EMA). We added hydrogens following the protonation states determined in Ref. 103. We then optimized the structure using CHARMM27 force-field [101]. The parameters of the chromophore are from Ref. 102. From the optimized structure, we extracted the model system comprising the chromophore and four other amino-acid residues, VAL93, GLN94, GLU95, and ARG96. Since these residues constitute a single sequence, we only capped $\alpha$-carbons of VAL93 and ARG96 (these capping hydrogens were not included in the MM point charges in the QM/MM calculations). Also, two capping hydrogens were added to the chromophore.

We constructed two model systems, one with the anionic (deprotonated) chromophore, and another with the neutral chromophore. In the model system with the anionic chromophore, GLU95 is deprotonated and ARG96 is protonated. The total charge of the model system is -1 (negatively charged chromophore and GLU95, and positively charged ARG96). The model structure is shown in Fig. 2.3.

The model system with the neutral GFP chromophore was prepared following the same protocol, except that GLU95 was protonated (this is necessary for finite-cluster calculations to suppress the ionization and electronic excitation from the MO localized on GLU95). The total charge of the the model system with the neutral GFP chromophore is +1.

In small QM, QM/MM, and QM/EFP calculations, only the chromophore constitutes the QM part. Vertical detachment and ionization energies (VDEs and VIEs, respectively) and were calculated with $\omega$B97X-D/aug-cc-pVDZ. We note that using range-separated functional is important for these charged systems. Vertical excitation energies were computed with SOS-CIS(D)/aug-cc-pVDZ. To investigate basis-set effects, we also performed calculations with a smaller basis set, 6-31G(d).

**mPlum model system.** The mPlum model system features a larger anionic chromophore, so one can assess whether the effect of the environment depends on the extent of $\pi$-conjugation in the chromophore. The system was constructed from the PDB ID:2QLG structure of mPlum following the protocol from Ref. 104. The model system was extracted from the QM/MM optimized structure (as described in Ref. 104). The model system comprises the extended conjugated anionic chromophore along with three amino-acids (ARG88-VAL89-MET90) and one water molecule. The total charge of the system is zero. We rearranged the water molecule to facilitate hydrogen bonding between phenolate's oxygen of the chromophore and water (this makes the radical-dianion state bound with respect to electron detachment). ARG89 and MET90 were capped with H-atoms at $C_\alpha$.

In small QM, QM/MM, and QM/EFP calculations, only the chromophore constitutes the QM part. VDE and excitation energies were calculated using the same methods as in the GFP model systems, and the attachment (VEA) energies were computed using $\omega$B97X-D/aug-cc-pVDZ.

**T4 lysozyme model system.** The model system consists of the tyrosine residue in an apolar cavity in the T4 lysozyme [105]. The following residues were retained

from the crystal structure (3GUO): phenolate, ALA99, ILE100, GLU102 (protonated at the side chain) and VAL103. We choose phenolate rather than phenol to ensure the spin density to be localized in the QM region (phenolate). ALA99 and VAL103 were capped with hydrogens at the peptide bond (C-N). Hydrogen positions were optimized by $\omega$B97X-D/6-31+G(d). All residues except for the phenolate are neutral; the total charge of the model system is -1. In small QM, QM/MM, and QM/EFP calculations, only the phenolate constitutes the QM part. We computed VDE using $\omega$B97X-D/aug-cc-pVDZ.

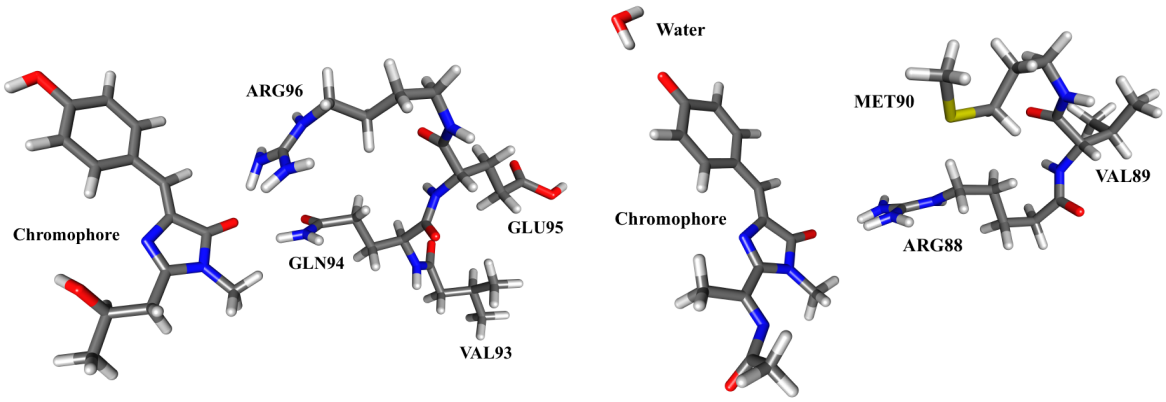



Fig. 2.3. Left: Protonated GFP chromophore and the VAL93, GLN94, GLU95, and ARG96 terapeptide string; GLU95 is protonated. The structure of the model system with the anionic chromophore is the same. Right: Model mPlum system with a deprotonated extended chromophore. The protein is represented by the ARG88, VAL89, and MET90 tripeptide string and one water molecule.

### 2.3.1 QM/MM and QM/mEFP calculations

The EFP parameters for the peptide residues were prepared following the procedure outlined in Sec. 2.2.2. In addition to the four fragmentation schemes described above, we also considered the super-fragment scheme in which all covalently linked peptides constitute a single EFP fragment. That is, in the GFP and mPlum model

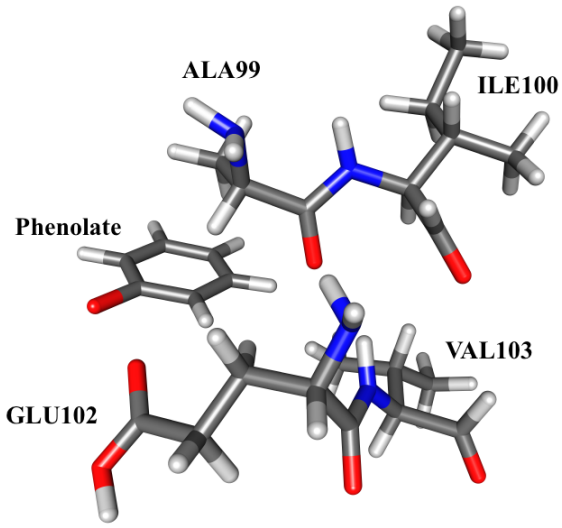
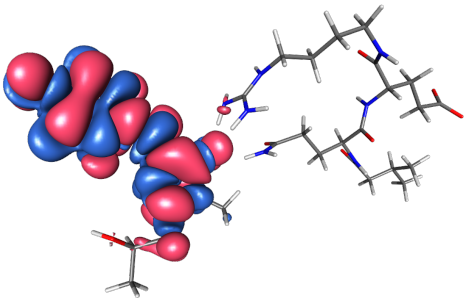



Fig. 2.4. Phenolate embedded into the apolar cavity consisting of four amino-acids in the T4 lysozyme.

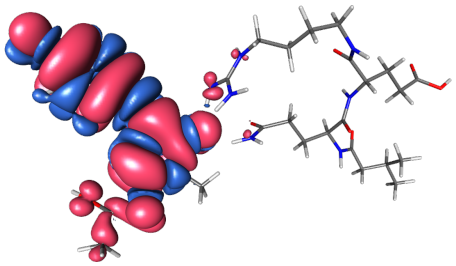
systems, all residues form one super-fragment; in the T4 lysozyme model system, there are two super-fragments.

The EFP parameters for each isolated capped residue and super-fragments have been generated using the `MAKEFP` job (`RUNTYP=MAKEFP`) of the *GAMESS* program [106]. Only the parameters responsible for electrostatic and polarization (multipoles expansion, damping parameters and static polarizabilities) were computed at the Hartree-Fock level of theory and with the 6-31G(d) basis set. The 6-311++G(3df,2p) basis set is usually recommended for computing polarization, dispersion and exchange-repulsion terms.

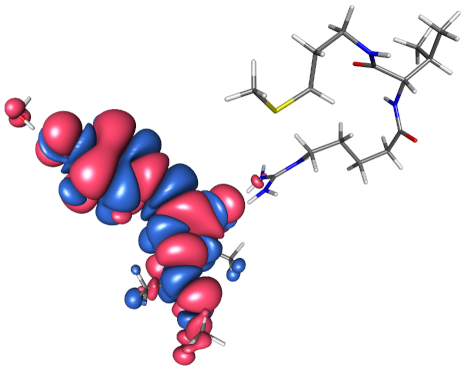
Exponential electrostatic damping of charges (SCREEN2) is employed between the fragments [95]; gaussian-type damping of charges (SCREEN) is used to mitigate the charge-penetration errors in the QM-EFP interactions [64]. Gaussian-type polarization damping controlled by a damping parameter is employed between the fragments, but not between the QM and EFP regions [95]. Effects of electrostatic screening and polarization damping on the QM/EFP energies is discussed below. The



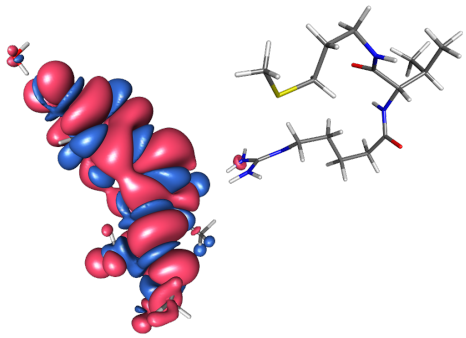
(a) Neutral radical-anion state of deprotonated GFP chromophore



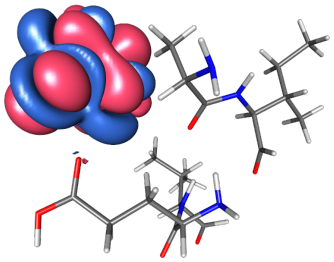
(b) Radical-cation state of neutral GFP chromophore



(c) Neutral radical state of mPlum chromophore



(d) Radical-anion state of mPlum chromophore



(e) Neutral radical state of phenolate in T4 lysozyme

Fig. 2.5. Spin densities of open-shell states for four model systems. In all cases, the spin density is localized on the residue of interest constituting the QM part.

QM region is described at the same level of theory as in the full QM and QM/MM calculations.

**QM/MM calculations.** In the QM/MM calculations, we tested three different schemes. Following the same strategy as in QM/mEFP, we do not include the capping H-atom in the MM part. We use CHARMM's force-field parameters for the MM point charges [101]. This creates an additional charge (e.g., +0.16 in the neutral GFP model system) in the MM part, making the total charge of the MM subsystem non-integer. There are several ways to handle this artifact of QM/MM. For example, we can add an extra -0.16 charge to the next atom ($C_\alpha$ in our case), or we can distribute that extra charge over the same residue. In QM/mEFP, the charge is added to the next carbon (see Fig. 2.1). We compared these two schemes and the QM/MM calculations with non-integer charges for protonated GFP model system:

1. QM/MM; scheme1: charge distributed over all atoms of the entire residue.

2. QM/MM; scheme2: charge added to the next atom.

3. QM/MM; scheme3: no re-distribution of charge; non-integer total MM charge.

Table 2.1 shows that all QM/MM calculations overestimate VDEs relative to full QM. Importantly, scheme1 and scheme2 produce very similar IEs, which indicates that IEs are not very sensitive to the details of charge redistribution, as long as the total charge is conserved. Since QM/mEFP uses the protocol similar to scheme2, in all QM/MM calculations reported below we employed scheme2.

## 2.4 Results

### 2.4.1 Electronic properties of model GFP chromophores

Table 2.2 summarizes the results for the GFP model system with the anionic chromophore. The VDE of the bare chromophore is 2.70 eV. The environment strongly stabilizes the anionic chromophore leading to the significant blue shift (1.65 eV) in

Table 2.1.
Verticle ionization energies (VDEs) of the neutral GFP chromophore surrounded by four nearby amino-acid residues computed with $\omega$B97X-D/aug-cc-pVDZ using three different QM/MM schemes.

| Calculation | VDE, eV | Koopmans IP (DFT), eV |
|---|---|---|
| Full QM | 9.887 | 10.095 |
| scheme1 | 9.953 | 10.095 |
| scheme2 | 9.943 | 10.068 |
| scheme3 | 10.104 | 10.231 |

Table 2.2.
Electronic properties of the anionic GFP chromophore surrounded by four nearby amino-acid residues.

| Calculation | VDE, eV | $E_{ex}$, eV ($f_L$) |
|---|---|---|
| Small QM | 2.70 | 2.76 (0.94) |
| Full QM | 4.35 | 2.56 (1.52) |
| QM/MM | 4.39 | 2.57 (1.45) |
| QM/EFP: superfragment | 4.39 | 2.60 (1.45) |
| QM/mEFP: $C_\alpha$-C | 4.39 | 2.60 (1.45) |
| QM/mEFP: $C_\alpha$-N | 4.37 | 2.60 (1.45) |
| QM/mEFP: C-N | 4.38 | 2.60 (1.45) |
| QM/mEFP: $C_\alpha$-C and $C_\alpha$-N | 4.38 | 2.60 (1.45) |

VDEs computed using $\omega$B97X-D/aug-cc-pVDZ. Excitation energies computed with SOS-CIS(D)/aug-cc-pVDZ; $f_l$ computed at the CIS level. The lowest bright state is shown (see text).

VDE. The VDE value in the full QM calculation is 4.35 eV. As illustrated in Fig. 2.5, the spin density in the full QM calculations is localized on the chromophore, which means that QM/MM and QM/mEFP calculations with the QM part comprising the

Table 2.3.
Electronic properties of the neutral GFP chromophore surrounded by four nearby amino-acid residues.

| Calculation | VIE, eV | $E_{ex}$, eV ($f_L$) |
|---|---|---|
| Small QM | 7.40 | 3.80 (1.05) |
| Full QM | 9.89 | 3.56 (1.09) |
| QM/MM | 9.94 | 3.58 (1.08) |
| QM/EFP: superfragment | 9.90 | 3.55 (1.03) |
| QM/mEFP: $C_\alpha$-C | 9.91 | 3.55 (1.03) |
| QM/mEFP: $C_\alpha$-N | 10.03 | 3.54 (1.04) |
| QM/mEFP: C-N | 9.92 | 3.55 (1.03) |
| QM/mEFP: $C_\alpha$-C and $C_\alpha$-N | 9.89 | 3.55 (1.04) |

VIEs computed using $\omega$B97X-D/aug-cc-pVDZ. Excitation energies computed with SOS-CIS(D)/aug-cc-pVDZ; $f_L$ computed at the CIS level. The lowest bright state is shown (see text).

chromophore can be directly compared to the full QM and small QM calculations. The QM/MM calculations capture the effect of the stabilization of the anionic chromophore reasonably well, yielding VDE of 4.39 eV, which is blue-shifted relative to the full QM value by 0.04 eV. Similarly to QM/MM, the QM/mEFP and QM/EFP VDEs and excitation energies are within 0.02-0.04 eV from the the full QM results. Importantly, neither detachment nor excitation energies are sensitive to the choice of the mEFP fragmentation schemes.

The analysis of the excited-state calculations requires care. There are several types of excited states of chromophores in condensed phase [43,107–109]: (i) local excitations in which both the electron and the hole are confined to the chromophore; (ii) charge-transfer states in which the electron is transferred between the chromophore and another neighboring group [108, 109]; (iii) charge-transfer-to-solvent states in which the electron is transferred from the chromophore to a cavity forming states

resembling solvated electrons [110, 111] (in finite clusters, one often observes surface states in which the target orbital corresponds to an electron on the surface). In addition to these physical states, one should also be aware of pseudo-continuum states corresponding to the electron detached from the system. In CIS calculations, these artificial states will appear above the Koopmans IE/DE when the basis is sufficiently diffuse [112, 113]; they may spoil the description of the bright states that lie above the Koopmans onset.

Obviously, in the QM/MM and QM/EFP calculations with the QM part comprising the chromophore alone, only local excitations corresponding to bound excited states can be described correctly. Thus, it is important to carefully analyze the character of the states in such calculations. Here we focus on the bright $\pi\pi^*$ state of the chromophore, which can be easily identified by its large oscillator strength and by inspecting the target MO.

The Koopmans IE of the isolated anionic GFP chromophore is 2.86 eV (HOMO energy, HF/aug-cc-pVDZ). Thus, all CIS excited states above this energy are embedded in the electron-detachment continuum, which spoils their description [113, 114]. Full QM CIS and SOS-CIS(D) calculations yield two lowest states of different character. One state is relatively dark and has charge-transfer character, whereas the other state is bright and corresponds to the local $\pi\pi^*$ excitation on the chromophore. We carefully checked that nature of the HOMO and the target virtual orbital involved in the bright transition remains the same among different schemes of calculation. Table 2.2 shows the energies of the bright state.

Table 2.3 summarizes the results for the GFP model system with the neutral chromophore. The VDE of the bare chromophore is 7.40 eV. The interaction with the neighboring residues leads to a blue shift of 2.5 eV (small QM versus full QM). Relative to full QM, the QM/MM calculations overestimate the ionization energy by 0.05 eV. QM/mEFP with the double-fragmentation scheme performs well, matching the blue shift in VDE. We also note that QM/mEFP with $C_\alpha$-N fragmentation scheme yields the largest deviation (+0.14 eV) relative to full QM.

In this system, the lowest excited state is the bright $\pi\pi^*$ state; as in the example before, they are of the same character in different calculations). The effect of the environment on the excitation energies is noticeable: the energy of the bright state is red-shifted by 0.24 eV in the full QM calculation relative to the bare chromophore (small QM). The QM/MM calculation overestimates the excitation energy by 0.02 eV. The QM/EFP super-fragment calculation underestimates the excitation energy by 0.01 eV. Different QM/mEFP schemes show similar performance; they underestimate the full QM excitation energy by 0.01-0.02 eV.

Using the anionic and neutral forms of the GFP, we investigate the effects of electrostatic and polarization damping on electronic properties. The results for VDEs are collected in Table 2.4. In Tables 2.2 and 2.3 discussed above, the QM/mEFP values were computed using polarization damping between the EFP fragments with the default parameter value of 0.6 and electrostatic damping between the QM and EFP regions given Eq. (2.6) with parameters precomputed as described in Section 2.3. In Table 2.4, the results with the default polarization and electrostatic damping are compared with the values obtained with (i) reduced polarization damping parameter, which corresponds to a stronger damping of polarization energies between the fragments, (ii) polarization completely turned off, which corresponds to electrostatic embedding, and (iii) electrostatic QM-EFP damping turned off. Although the default value of the polarization damping parameter is 0.6, we find that smaller values of this parameter, i.e., stronger damping, are occasionally necessary for avoiding the polarization catastrophe. In particular, this often happens for strongly-interacting ionic species [115]. Although in the systems that we considered here, we did not encounter difficulties converging polarization energies with the default damping values, we explore the effect of polarization damping on IEs/DEs, in case that the modification of damping parameter is required in more complex systems. Note that polarization damping is applied only to fragment-fragment polarization; the QM→EFP and EFP←QM polarization interactions are not damped.

Comparing IEs computed with polarization turned-off entirely or partially, we observe several interesting effects. The first observation is that polarization of the environment plays a more significant role in the anionic than in the neutral chromophore, as illustrated by the VIE/VDE differences between the default and polarization-off values in the superfragment calculations, which are 0.18 eV and 0.11 eV for the anionic and neutral chromophores, respectively. The second observation is that polarization effect is smaller in fragmented (mEFP) than in the superfragment (EFP) calculations. For example, in the anionic form, the change in VDEs due to polarization decreases from 0.18 eV in QM/EFP to 0.06-0.08 eV in QM/mEFP, and similarly in the neutral form. This is an interesting observation, which suggests that neighboring fragments "depolarize" each other. This effect occurs in QM/mEFP but not in super-fragment calculations. (In superfragment calculations, the QM and EFP subsystems polarize each other, but the EFP super-fragment does not polarize itself.) Finally, a moderate decrease in polarization damping to 0.3 does not affect the results within 0.01 eV. However, decreasing the damping parameter to 0.1 effectively supresses polarization, i.e., these results become identical to those with the polarization turned off.

Electrostatic damping takes care of charge penetration energy; it has been explored in detail for the EFP-EFP interactions [95,96]. The effect of the electrostatic damping on the QM-EFP interactions is analyzed in this work for the first time. For both the anionic and neutral chromophores, the QM-EFP electrostatic damping shifts energy differences by 0.01-0.05 eV, improving the agreement between the QM/EFP results and the reference QM data. Similar shifts are observed for different fragmentation schemes.

To summarize, our analysis suggests that both electrostatic damping and polarization interactions are important for accurate description of VIEs/VDEs. For example, for the C$\alpha$-C scheme in the anionic chromophore, the best (with default electrostatic and polarization damping) QM/mEFP calculations underestimate VDE by 0.09 eV, whereas switching off electrostatic damping and polarization increases the errors further, by 0.04 eV and 0.07 eV, respectively. The C$\alpha$-C scheme in the

neutral chromophore underestimates VDEs by 0.06 eV, and turning off electrostatic damping and polarization introduces additional errors of 0.01 eV each.

Table 2.4.
VDE and VIE (in eV) of the anionic (top) and neutral (bottom) GFP chromophore surrounded by four nearby amino-acid residues as a function of electrostatic and polarization damping in the EFP region. $\omega$B97X-D/6-31G(d) is used in all calculations.

| Calculation | pol damp=0.6 (default) | pol damp=0.3 | pol damp=0.1 | pol off | elec[a] damp off |
|---|---|---|---|---|---|
| Anionic GFP[b] | | | | | |
| QM/EFP: super-fragment | 3.93 | | | 3.75 | 3.88 |
| QM/mEFP: $C_\alpha$-C | 3.94 | 3.94 | 3.87 | 3.87 | 3.90 |
| QM/mEFP: $C_\alpha$-N | 3.92 | 3.92 | 3.86 | 3.86 | 3.87 |
| QM/mEFP: C-N | 3.93 | 3.93 | 3.85 | 3.85 | 3.89 |
| QM/mEFP: $C_\alpha$-C and $C_\alpha$-N | 3.93 | 3.93 | 3.87 | 3.87 | 3.88 |
| Neutral GFP[c] | | | | | |
| QM/EFP: super-fragment | 9.67 | | | 9.56 | 9.66 |
| QM/mEFP: $C_\alpha$-C | 9.68 | 9.68 | 9.67 | 9.67 | 9.67 |
| QM/mEFP: $C_\alpha$-N | 9.81 | 9.81 | 9.79 | 9.79 | 9.80 |
| QM/mEFP: C-N | 9.69 | 9.68 | 9.66 | 9.66 | 9.68 |
| QM/mEFP: $C_\alpha$-C and $C_\alpha$-N | 9.66 | 9.66 | 9.65 | 9.65 | 9.65 |

[a] Default polarization (with damp=0.6) is used.

[b] Full QM:VDE = 4.03 eV, QM/MM:VDE = 3.97 eV

[c] Full QM:VDE = 9.74 eV, QM/MM:VDE = 9.74 eV

Table 2.5.
Electronic properties of the anionic mPlum chromophore surrounded by four nearby amino-acid residues and water.

| Calculation | VDE, eV | $E_{ex}$, eV ($f_L$) | VEA, eV |
|---|---|---|---|
| Small QM | 3.20 | 2.30 (1.51) | 2.03 |
| Full QM | 5.57 | 2.65 (0.91) | -0.21 |
| | | 2.81(0.64) | |
| QM/MM | 5.61 | 2.53 (1.53) | -0.20 |
| QM/EFP: super-fragment | 5.60 | 2.53 (1.54) | -0.22 |
| QM/mEFP: $C_\alpha$-C | 5.63 | 2.52 (1.54) | -0.24 |
| pol off | 5.51 | 2.51 (1.54) | -0.08 |
| QM/mEFP: $C_\alpha$-N | 5.61 | 2.53 (1.54) | -0.22 |
| pol off | 5.52 | 2.51 (1.54) | -0.07 |
| QM/mEFP: C-N | 5.59 | 2.52 (1.54) | -0.21 |
| pol off | 5.52 | 2.51 (1.54) | -0.08 |
| QM/mEFP: $C_\alpha$-C and $C_\alpha$-N | 5.61 | 2.53 (1.54) | -0.22 |
| pol off | 5.51 | 2.51 (1.54) | -0.07 |

VDE and VEA computed with $\omega$B97X-D/aug-cc-pVDZ. Excitation energies computed with SOS-CIS(D)/aug-cc-pVDZ; $f_l$ computed at the CIS level.

### 2.4.2 mPlum model system

Electronic properties for the mPlum model system are collected in Table 2.5. The mPlum chromophore features an extended $\pi$-system in which the conjugation extends beyond the imidazolinone ring and into the side-chain. The extended $\pi$-system is responsible for red-shifted absorption relative to the GFP chromophore [107]. The extended conjugation also leads to higher detachment energy of mPlum, as compared to deprotonated GFP (Table 2.2); the reason for that was described by Ghosh *et*

*al.* [116]. In this system, the chromophore is also anionic and the environment strongly stabilizes its ground state. Thus, we observe similar trends in Tables 2.2 and 2.5.

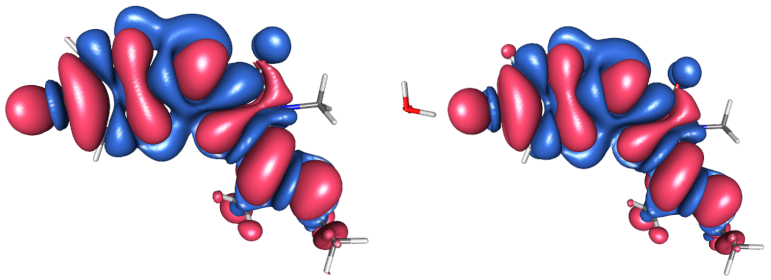
We begin by considering the trends in VDE. From Table 2.5, we once again observe that the protein environment affects the detachment energy significantly, resulting in 2.4 eV blue shift in full QM/MM relative to small QM. QM/MM captures the effect of the environment well; VDE is overestimated by +0.04 eV relative to full QM. QM/mEFP also performs very well. We note that within the same fragmentation scheme, QM/mEFP overestimates VDE when polarization is on (highest deviation +0.06 eV compared to full QM) and underestimates VDE with polarization switched off (the largest deviation from the full QM result is -0.06 eV).

In the bare chromophore (small QM), the lowest excited state is the bright excited state of the $\pi \rightarrow \pi^*$ character and $E_{ex}$=2.30 eV. In the full QM calculations, we observe two blue-shifted excited states of similar character (the oscillator strength of the bright transition is distributed between the two). All QM/MM and QM/mEFP calculations yield only one bright $\pi\pi^*$ state (which is the lowest excited state) carrying large oscillator strength. The excitation energy is rather insensitive to the fragmentation scheme (2.52-2.53 eV).

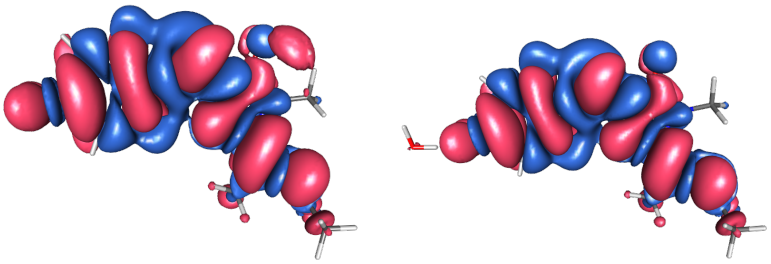
In this model system, we also consider electron-attached states. Although in the bare chromophore, such radical-dianion state is not bound electronically (i.e., it has positive electron affinity), it can be stabilized by interactions with the protein. In a recent study of KillerRed (which has the same chromophore as mPlum) such dianion-radical states were found to be stable [109]. Fig. 2.5(d) and Fig. 2.6 show spin densities for the electron-attached states. As one can see from the data in Table 2.5, this state has similar character across different calculation schemes. In Fig. 2.6, inclusion of water in the QM part of small QM and QM/MM calculations, does not affect the spin density and thereby confirming the effect of water to be purely electrostatic.

In this model system, the protein stabilizes the dianion state leading to 2.24 eV change in VEA between the small QM and full QM calculations. The QM/MM and

QM/mEFP results are within 0.01 eV of the full QM value. The effect of polarization is more pronounced in the case of VEA: the effect of switching off polarization can be as large as 0.14 eV.



(a) small QM (b) (small QM+water)



(c) QM/MM (d) (QM+water)/MM

Fig. 2.6. Spin densities on the electron-attached dianionic mPlum chromophore model system in different calculations (see text and Table 2.5).

### 2.4.3 T4 lysozyme model system

Table 2.6 compares VDEs computed with small QM, QM/MM, full QM, QM/EFP and QM/mEFP with different fragmentation schemes. In this system we observe that QM/MM performs poorly and underestimates VDE by 0.33 eV, as compared to full QM. On the other hand, the QM/mEFP schemes perform really well with the largest error being -0.08 eV relative to full QM. It is also noteworthy that in QM/EFP and

Table 2.6.
VDEs (eV) of the phenolate bound to the T4 lysozyme system, $\omega$B97X-D/aug-cc-pVDZ.

| System | VDE, eV | Koopmans IE (DFT) |
|---|---|---|
| Small QM | 2.12 | 2.14 |
| Full QM | 1.90 | 2.12 |
| QM/MM | 1.57 | 1.58 |
| QM/mEFP: super-fragment | 1.85 | 2.04 |
| pol off | 1.57 | 1.58 |
| QM/mEFP: $C_\alpha$-C | 1.82 | 2.01 |
| pol off | 1.51 | 1.52 |
| QM/mEFP: $C_\alpha$-N | 1.92 | 2.12 |
| pol off | 1.64 | 1.63 |
| QM/mEFP: C-N | 1.84 | 2.01 |
| pol off | 1.61 | 1.61 |
| QM/mEFP: $C_\alpha$-C and $C_\alpha$-N | 1.90 | 2.10 |
| pol off | 1.60 | 1.61 |

QM/mEFP polarization plays a significant role here. Indeed, when polarization is turned off completely, the QM/EFP and QM/mEFP results change by 0.2-0.3 eV and become very similar to the QM/MM data. A significant contribution of polarization can be rationalized by the charge density of the phenolate being largely localized on $O^-$ which leads to strong polarizing interactions with the GLU102 residue. Overall, the scheme with two fragments per amino acid ($C_\alpha$-C and $C_\alpha$-N) shows the best performance for this system: it reproduces the full QM result exactly.

## 2.5 Conclusions

The extension of the EFP method to polypetides, called mEFP, has been developed and validated by computing excitation, ionization, and electron attachment energies of three biologically relevant systems, GFP/mPlum chromophores and phenolate in their natural surroundings. In the mEFP scheme, the polypeptide is split into smaller fragments, and the EFP parameters for each fragment are computed independently of other fragments. Four different fragmentation schemes have been tested; the most consistent results were obtained with the schemes in which the polypeptide is split either along C$\alpha$-C bond or along both C$\alpha$-C and C$\alpha$-N bonds.

In all systems considered here, QM/mEFP accurately reproduces excitation, ionization and electron attachment energies, provided that the electronic process is localized in the QM subsystem. Discrepancies between the full QM calculations and QM/mEFP calculations in all but one case do not exceed 0.1 eV. Polarization interactions have a noticeable effect on electronic properties of biological chromophores. Turning off polarization in QM/mEFP degrades the accuracy and leads to additional errors of up to 0.3 eV. Short-range damping of electrostatic interactions between the QM and EFP subsystems, which corrects classical multipole expansion for charge-penetration energy, also brings the QM/mEFP results into a closer agreement with the reference full QM values. In most cases, the errors of QM/mEFP against QM/EFP (no fragmentation of the EFP part) are 0.01-0.02 eV; the largest error is 0.06 eV.

The QM/mEFP approach provides a rigorous way to incorporate polarization embedding into studies of biological systems. The developed mEFP algorithm cab be generalized to other polymers and flexible molecules, which will be exploited in future studies.

# 3. PREDICTION OF PREFERENTIAL LIGAND-PROTEIN BINDING USING BIOMOLECULAR EFFECTIVE FRAGMENT POTENTIAL METHOD

The primary objective of rational drug design is to estimate the affinity of a molecule or a set of molecules to bind to a target both qualitatively and quantitatively. Molecular mechanics based force fields are considered the methods of choice for computing the drug-target binding by estimating the strength of intermolecular interactions, mainly due to the speeds that can be achieved in screening a large number of compounds as well as the ease in performing dynamics simulations. Such methods utilize a local model in which the interactions that are localized to the drug binding pocket are taken into account, while the interaction of the rest of the protein is usually neglected. Here we propose an Effective Fragment Potential (EFP) method based energy decomposition analysis scheme for application in quantifying the drug-target interactions in proteins. We apply this scheme to understand the substituent effects in a -chloro vs -methyl aryl substitution in factor Xa inhibitor drugs. Our results indicate that the substituents are stabilized/destabilized due to interactions well beyond the binding pocket. While the primary contributor to the stability of the substituent functional group is electrostatic interactions, other terms such as exchange-repulsion play an important role as well. The extended contact model presented here accounts for interactions that are typically neglected in a local model, and we believe that this extended contact model marks an improvement over the latter in terms of accurate prediction of interaction energies.

## 3.1 Introduction

Recent advances in computational chemistry have opened exciting applications in drug discovery and have shortened the timeline for drugs to reach the market [117–122]. Understanding the underlying chemistry of drug-target interactions enables faster development of drugs, and computational chemistry has assisted medicinal chemists for decades in achieving this goal [123]. In the recent years, due to major advances in supramolecular chemistry and crystallography, obtaining detailed information on the 3D structures of large macromolecules have become routine [124]. Computational chemistry has played an important role in determining the viability of ligand molecules to be used as drugs for a given target [125]. While molecular mechanics (MM) methods have been widely employed in computer aided drug design (CADD), due to the advent of high performance computing and faster, efficient codes, quantum mechanical (QM) methods are gaining importance in this field [126–128].

In order to accurately predict the binding energy between two molecules in solvent phase, many factors need to be considered: the interaction energy between the two molecules, desolvation penalties due to the removal of solvent molecules occupied by the ligand, other solvent effects, temperature corrections, etc [129, 130]. A simpler means of modeling such a system involving a ligand-protein complex would be to start from a well-studied ligand-protein system, followed by performing substitution to the functional group(s) present in the ligand to measure the effects of substitution. If the only goal is to estimate the relative binding energies, this approach is simpler and faster.

In simulating the ligand-protein interactions, a common assumption made is to account only for local interactions. This local model takes into account only the interaction of the ligand with transferable contacts such as hydrogen bond interactions, pi-pi interactions, ion-pi interactions, halogen bond interactions etc, and is usually restricted to the drug binding pocket.

Factor Xa is an activated form of thrombokinase, an enzyme that participates in the blood coagulation cascade. Antithrombotic agents corresponding to this enzymes have been developed and studied in detail [131, 132]. The active site of factor Xa consists of four subpockets: S1 - S4. The S1 subpocket plays a major role in selectivity and binding of the factor Xa. Earlier, Sherrill and coworkers [133] probed the effect of functional group modification in the factor Xa ligand using state of the art symmetry adapted perturbation theory (SAPT) methods. The interaction between neutral ligand and the anionic S1 pocket that comprises all the local contacts was computed using Functional SAPT (F-SAPT) [134, 135] and cut-and-cap SAPT (will be referred to as P-SAPT for simplicity) simulations truncated at the zero order (SAPT0) [136,137]. The aim of the work was to understand the preferential binding of Cl- substituted ligands as compared to Me- substituted ligands. To achieve this aim, the interaction energies between all the sidechain residues and peptide backbone fragments and the ligand were computed ($\Delta E_{int}^{Cl}$ and $\Delta E_{int}^{Me}$) in a completely interacting system using F-SAPT and a pairwise interacting system using P-SAPT. Calculating the difference in binding energies($\Delta\Delta G$) using quantum mechanical methods is a computationally tedious process, hence the difference between the interaction energies ($\Delta\Delta E_{int} = \Delta E_{int}^{Cl} - \Delta E_{int}^{Me}$) can be used as an approximate measure of understanding the former quantity. This approximation can be deemed valid because of two reasons: 1. The vdW radii of the two functional groups are similar if not the same, hence the geometric effects due to substitution can be neglected; and 2. The polarities between the two functional groups are not vastly different, hence the desolvation penalties can be neglected as well. For ligand modifications involving substitution of vastly different functional group, this approximation is expected to break down.

The effective fragment potential (EFP) method [22–25] is an ab-initio force field used to describe the non-covalent interactions between fragments in an interacting system of molecules, as well as to model effects of non-covalent interactions on properties of quantum-mechanical region. Recently, we extended this method to model and simulate covalently bonded molecules such as amino acids, lipids etc. [138]. Our

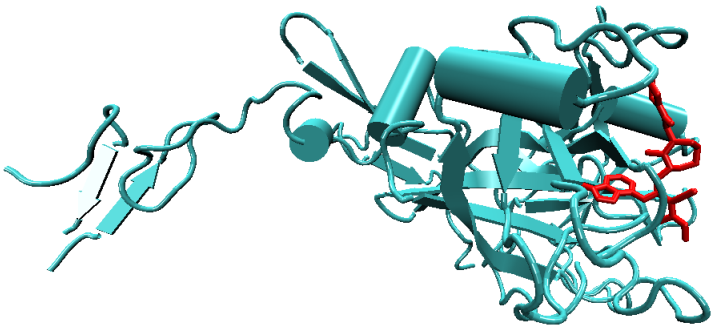


Fig. 3.1. Structure of the 3ENS-Cl molecule (red) bound to factor Xa.

intention here is to compare the performance of the biomolecular effective fragment potential method (BioEFP) with SAPT0 and to assess the validity of the local interaction model in predicting ligand-protein interactions.

## 3.2 Methods

### 3.2.1 Effective Fragment Potential Method

The Effective Fragment Potential method [22–25] describes the interactions between various interacting fragments using perturbation theory. The overall interaction energy of a given system is the sum of electrostatics, polarization, dispersion and exchange-repulsion terms, as given below:

$$E_{\mathrm{EFP-EFP}} = E^{\mathrm{elec}} + E^{\mathrm{pol}} + E^{\mathrm{disp}} + E^{\mathrm{exrep}}. \tag{3.1}$$

The electrostatic interactions are described by the means of a multipole expansion, centered at all the atoms and the midpoints of the bonds connecting the atoms in the fragment [139]. The electrostatic multipole expansion is truncated at the level of octupoles. The electrostatic interaction energy is computed between all the multipole expansion points of one fragment and the multipole expansion points of all other fragments in the system. The electrostatic interaction energy term consists of charge-charge, charge-dipole, charge-quadrupole, charge-octupole, dipole-quadrupole, and quadrupole-quadrupole interaction terms. The parameters corresponding to the multipole expansion are obtained using Distributed Multipole Analysis (DMA) method by Stone [140].

In order to tackle the poor description of short-range electrostatic interactions due to charge penetration, we employ an exponential screening function that damps the charge-charge interactions at close separations [141]. It must be noted that this damping function does not significantly alter the electrostatic interactions at mid- and long- range separations. A detailed description of this damping function is available in ref. [141].

The EFP method computes many-body polarization in a self-consistent, iterative fashion. The static dipole-dipole anisotropic polarizability tensors are computed at the centroids of the non-canonical localized molecular orbitals (LMOs) by solving for the Coupled Hartree-Fock equations (CPHF).

The effective fragment potential method employs a rigorous approach towards computing the dispersion energies. Dynamic polarizability tensors centered at LMO centroids are computed at Time-dependent Hartree-Fock (TDHF) level.

The repulsive interactions that arise due to Pauli's exclusion between two fragments is described using the exchange-repulsion term in the effective fragment potential method. It must be noted that the electostatics, dispersion and exchange-repulsion terms in EFP are computed in a pairwise fashion, i.e., the interaction between two fragments is not altered by the introduction of a third fragment, which is strictly not the case with wave function based methods. Earlier, Slipchenko et. al., benchmarked the performance of EFP methods with respect to accurate quantum mechanical methods and force fields, and found that EFP provides a balanced description of electrostatic- as well as dispersion-dominant interactions [142].

### 3.2.2 Biomolecular Effective Fragment Potential Method (BioEFP)

The biomolecular effective fragment potential method (BioEFP), also known as macromolecular effective fragment potential method (mEFP), is an offshoot of the original implementation of the Effective fragment potential method (EFP) and is designed for modeling large covalent molecules such as proteins, lipids, etc [138]. While the original EFP method was constructed with an intent to describe the effect of solvent environment on the properties of the solute, the BioEFP method extends the idea to covalently-connected molecules by systematically fragmenting the macromolecule into smaller fragments.

A protein is a chain of repeating residues connected by covalent bonds. By making use of the idea of repeating residues, we can fragment the chain at designated

bonds, cap them using capping fragments (hydrogens, in this case) and obtain the parameters corresponding to the capped fragments. We then remove the parameters that correspond to the capped atoms to obtain the parameters corresponding to the uncapped fragments. Once the parameters corresponding to all the uncapped fragments are obtained, we may perform an EFP simulation that corresponds to the entire protein. A detailed summary of this procedure is available in chapter 3, as well as in reference [138].

The electrostatics parameters corresponding to the capped atoms and the bond midpoints connected to the capped atoms are removed. As a result of the removal of these parameters, the total charge of the fragment is now non-zero (or non-integer in the case of charged fragments). The excess charge resulting due to the removal of capped atoms is then added and redistributed to the nearest neighboring atom using the expand-remove-redistribute procedure [138]. Hence, the net charge of individual fragments and the overall system remains the same before and after this procedure.

As described earlier, the polarization and dispersion parameters are centered at the LMO centroids instead of the atoms themselves. Hence, removal of a capped atom results in the removal of the closest LMO, which is usually the LMO that corresponds to the sigma bond between hydrogen and the neighboring atom. This step is highly necessary to avoid polarization collapse, as the sigma bond LMOs corresponding to the capped hydrogens in neighboring fragments are positioned close to each other and could potentially overpolarize each other. This could result in a larger than expected or diverged polarization energy.

The only computed parameters corresponding to the exchange repulsion term are the non-canonical localized molecular orbital coefficients, basis set coefficients and the Fock matrix. The localized molecular orbitals are then constructed as a linear combination of pre-defined atomic orbital basis functions. When fragmenting the system for the purpose of computing parameters, one can remove the localized molecular orbitals corresponding to the fragmented atoms and bonds, or keep them as-is, since exchange-repulsion is computed in a pairwise fashion. While the former

scenario could potentially underestimate the exchange-repulsion energies, the latter could overestimate it. An alternate strategy could be to remove either the orbitals corresponding to $C_\alpha$ carbon or the peptide carbon and maintain this consistency for the entire protein. For the purpose of this study, we have decided not to modify the exchange-repulsion parameters to maintain simplicity.

### 3.2.3 Prediction of Ligand-Target binding

The protein conformation was obtained from Ref. [133]. Briefly, the following procedure describes the modification done by Sherrill et. al. to obtain the protonated ligand-protein structure : The geometry for factor Xa (the protein) in complex with methyl(2Z)-3-[(3-chloro-1H-indol-7-yl)amino]-2-cyano-3-[(3S)-2-oxo-1-(2-oxo-2-pyrrolidin-1-ylethyl)azepan-3-yl]aminoacrylate (the drug containing Cl- functional group) was obtained from RCSB database [143]. The Cl- functional group in the ligand was replaced by a carbon atom to obtain the methylated form of the ligand. Structures for both these analogues were prepared using the Protein Preparation Utility in Maestro (Schrodinger), which provides a rational estimate of optimal torsions, protonation states and the orientation of crystal water molecules. Following this, a constrained optimization was performed, to avoid obvious clashes and to minimize steric hindrances.

Earlier studies on the ligand indicate that following the substitution with the methyl ligand, the electron density shifts from the chlorine group to the farther end of the indole ring (Fig. 3.3). This would mean that the change in the nature of stabilizing/destabilizing interactions are not localized to the functional group alone. Further, it was shown that the single substitution enhanced the in-vitro $IC_{50}$ (the concentration of the inhibitor at which the binding is reduced by half) by roughly 50 times [143]. Major contributions to this enhanced efficacy were attributed to the interactions with Tyr228, Asp189, Gly219 and Cys220 residues [143–145].

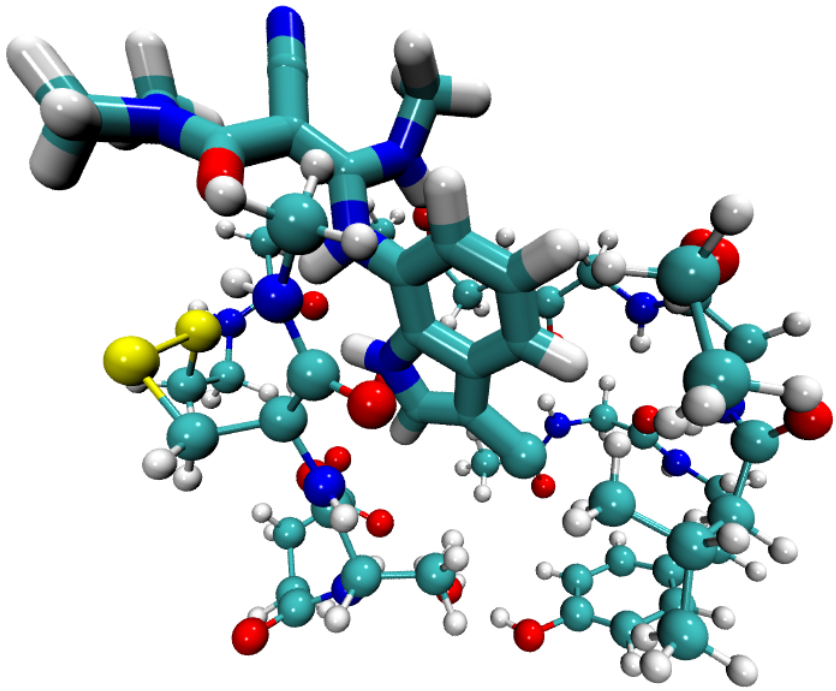


Fig. 3.2. S1 binding pocket of the 3ENS-Cl ligand-protein system. The pocket is shown in a ball and stick representation, while the ligand is shown using thick sticks.

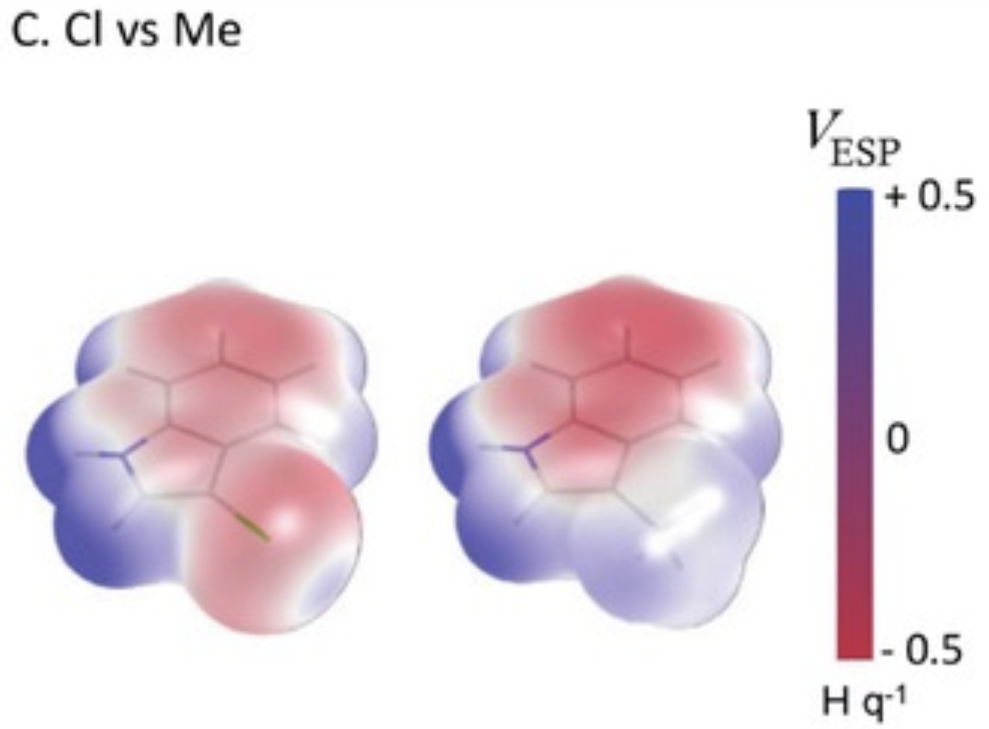



Fig. 3.3. Electron density map of 3ENS ligands with Cl- and Me- substitutions. Adapted from [133].

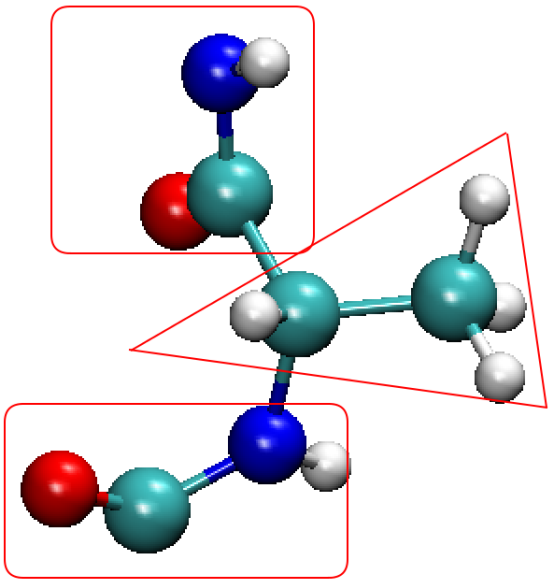


Fig. 3.4. Cut-and-cap strategy used in BioEFP modeling of ligand-S1 pocket. Each amino acid residue is fragmented across two sites, resulting in a sidechain fragment and a peptide fragment.

For the purpose of this study, we considered two model systems: 1. A smaller version of the ligand interacting with the fragments in the S1 pocket (to compare the performance of EFP with SAPT methods); and 2. The unmodified ligand interacting with all the fragments in the protein (to assess the validity of local model and compare it with our extended model). The S1 pocket consists of fragments that directly interact and bind with the ligand.

For the simulation of the S1 pocket with the ligand, only a small portion of the drug molecule that directly interacts with the pocket is included. While it is possible to include the whole drug molecule in the simulation, we decided to fragment the drug molecule and cap it with a hydrogen atom. SAPT0 results available in the literature were performed with this smaller version of the ligand, possibly due to practical restrictions in the simulation time, and hence following a similar strategy would enable a one-to-one comparison of our method with SAPT0. The amino acid residues were fragmented in such a way that the parameters for the sidechain and peptide groups were computed separately. In other words, the residues were fragmented along the C$\alpha$-C bond as well as the C$\alpha$-N bond (Fig. 3.4). This effectively results in two fragments per amino acid residue. The amino acid residues included in the S1 pocket are: ASP189, ALA190, CYS191, GLN192, SER195, VAL213, SER214, TRP215, GLY216, GLU217, GLY218, CYS220, GLY226, ILE227, and TYR228. All the fragmented residues were capped with hydrogen atoms. The disulfide bond between the cysteine residues were fragmented as well. As a result, the contribution of the disulfide bond to the total interaction energy is not explicitly computed using a separate disulfide fragment in BioEFP method. However, it must be noted that SAPT0 simulations were computed in such a way that the cysteine residue interactions and disulfide bond interactions were calculated separately. While it is not possible to directly compare the interaction energy components corresponding to these fragments, the sum of interaction energies due to the pair of cysteine EFP fragments can be compared to the sum of interaction energies due to cysteine and disulfide SAPT fragments.

All the EFP parameters were computed using HF/6-31G(d) basis set except for the exchange repulsion term, which was computed using a larger basis set (HF/6-31+G(d)) for better accuracy. All the EFP-EFP simulations were performed using libefp package [146,147] using the pairwise energy decomposition feature implemented recently. It must be noted that the electrostatics, dispersion and exchange repulsion terms are pairwise additive, and are directly comparable to dimer interaction energies. The polarization term, however, is not pairwise additive, and hence must be computed in a fully interacting system. The energy contributions due to polarization are then obtained from the converged induced dipole moments centered at the localized molecular orbitals of each fragment.

To model the protein as a whole, we employed the following strategy: Both the Cl- and Me- substituted ligands were modeled as is, with no fragmentation or modification. The Cl- functional group was replaced by a Me- group within the binding pocket to obtain the conformation for the Me- substituted ligand-protein system. The surrounding protein chain was fragmented along the $C_\alpha$-C bond alone, resulting in one effective fragment per amino acid residue. Thus, decomposing the energy terms into peptide fragment and sidechain fragment contributions is not possible. However, this fragmentation scheme is in line with the original BioEFP scheme and is expected to produce accurate interfragment interaction energies and energy components.

### 3.3 Results and Discussion

In order to assess the performance of various interaction terms in the BioEFP method, we compare them directly with P-SAPT or F-SAPT data from ref. [133]. Figure 3.5 shows the contribution of electrostatic interactions to the $\Delta(\Delta E)$ term. Positive $\Delta(\Delta E)$ contributions indicate a preference of the methylated ligand over the chlorinated ligand, while negative $\Delta(\Delta E)$ contributions indicate the opposite. BioEFP correctly predicts the stronger preferential interactions due to peptide bond contacts, especially due to residues 190, 215 and 219. Since the electrostatics term

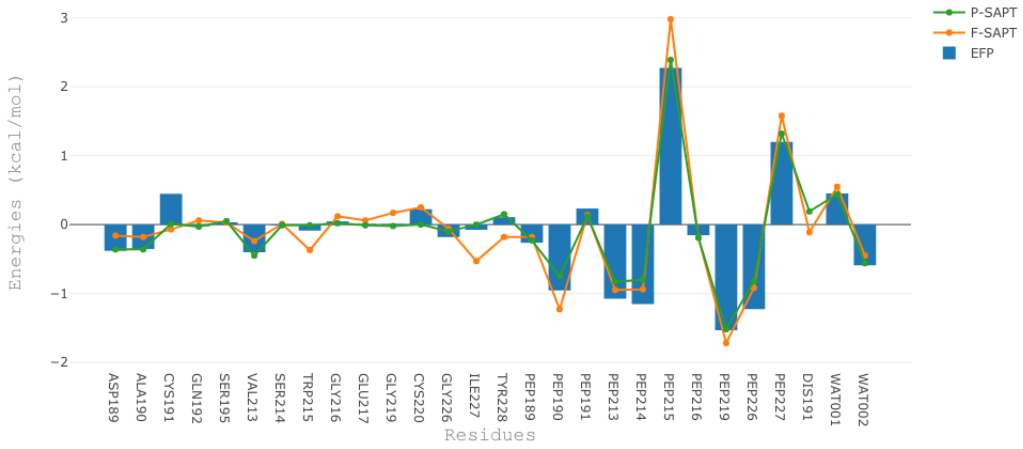



Fig. 3.5. Contribution of electrostatic interactions to $\Delta(\Delta E)$ term in the small ligand - S1 pocket model.

in EFP is computed in a pairwise fashion, it is directly comparable to P-SAPT simulations. It can be seen that EFP predicts the electrostatic interaction energies to within 0.5 kcal/mol in comparison to the P-SAPT simulations. The discrepancies found in $\Delta(\Delta E_{elec})$ in the CYS191, CYS220 and disulfide fragments are due to the different fragmentation schemes employed in EFP and SAPT methods. Electrostatic interactions are the major contribution to the preferential binding energies, in some cases as high as 2 kcal/mol.

Figure 3.6 shows the contribution of polarization interactions to the $\Delta(\Delta E)$ term. For polarization interactions, it is prudent to compare the EFP results with the F-SAPT results, as both the methods obtain interaction energies in a 'fully interacting' system. Again, the qualitative prediction of BioEFP method is reasonable, however, the difference in interaction energies are less than 0.4 kcal/mol in most fragments. In many cases, EFP seems to overestimate the stabilization/destabilization due to binding energy differences, but the errors do not exceed 0.25 kcal/mol. Damping the polarization using screening functions does not affect the results by much, and hence

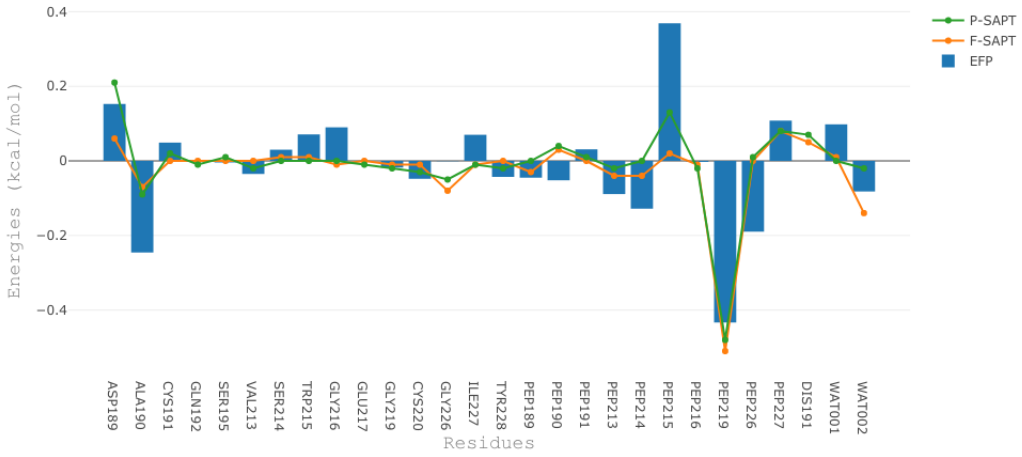



Fig. 3.6. Contribution of polarization interactions to $\Delta(\Delta E)$ term in the small ligand - S1 pocket model.

it can be concluded that the polarization interactions do not contribute much to the difference in binding energies, as predicted by SAPT results as well.

Figure 3.7 shows the contribution of dispersion interactions to the $\Delta(\Delta E)$ term. For dispersion interactions, we revert back to comparing the EFP results with the P-SAPT results, as both the methods obtain interaction energies in a pairwise manner. The contribution of dispersion to the $\Delta(\Delta E)$ term is understandably low (less than 0.2 kcal/mol), as the substitution of Cl- with a Me- group does not significantly affect the $\pi - \pi$ interactions or CH-$\pi$ interactions within the ligand-S1 pocket. This finding is in line with what has been observed in earlier as well [133].

Figure 3.8 shows the contribution of exchange repulsion interactions to the $\Delta(\Delta E)$ term. For the exchange-repulsion term, again we compare the EFP results with the P-SAPT results, as both the methods obtain interaction energies in a pairwise manner. Formally, the exchange-repulsion term decays in an exponential manner, and we can assume that only the closest residues that directly interact with the indole ring would contribute to this term. This is evident in the case of ALA190 and VAL213, which

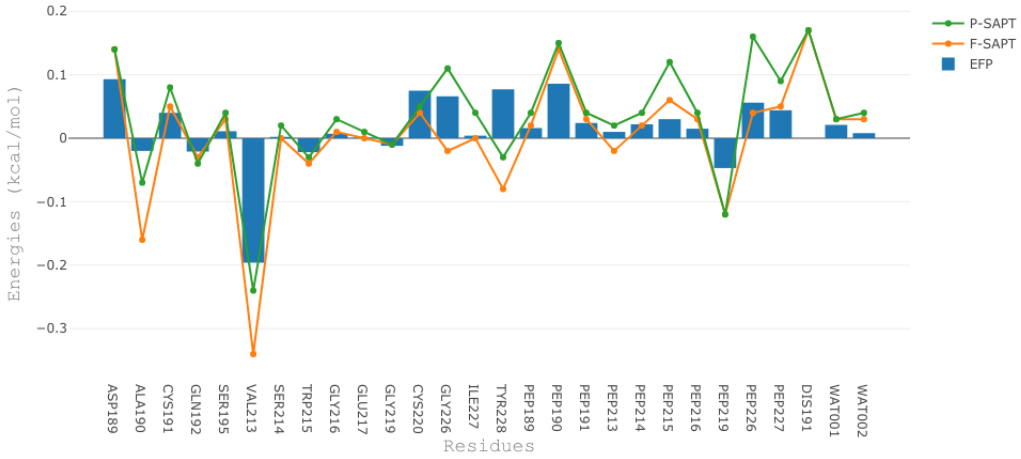



Fig. 3.7. Contribution of dispersion interactions to $\Delta(\Delta E)$ term in the small ligand - S1 pocket model.

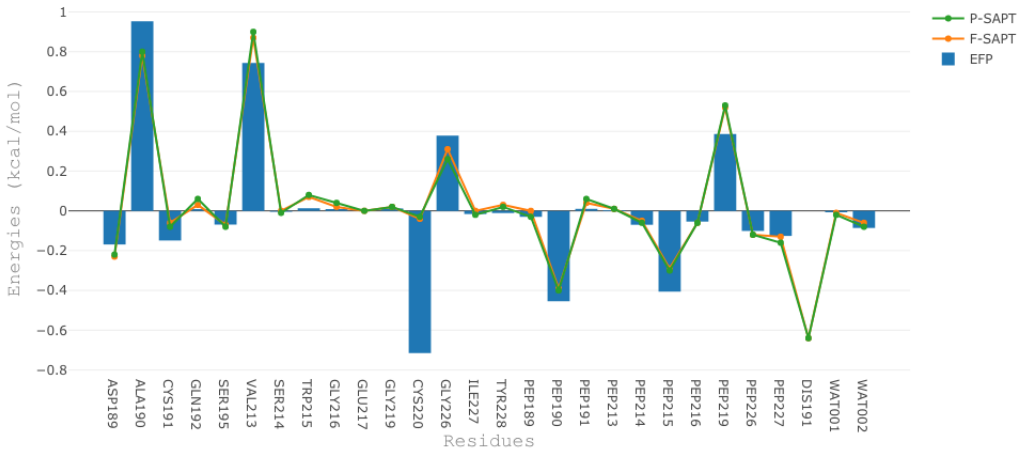



Fig. 3.8. Contribution of exchange-repulsion interactions to $\Delta(\Delta E)$ term in the small ligand - S1 pocket model.

is captured by EFP as well as SAPT methods. As explained earlier, the discrepancy in DIS fragment interactions caused by fragmentation are captured in the two CYS fragments.

Figure 3.9 shows the sum of all the interaction energy components to the $\Delta(\Delta E)$ term. Figure 3.10 individual contributions of all the four terms that contribute to the $\Delta(\Delta E)$. A point to note here is that while the $\Delta(\Delta E_{total})$ for a single fragment could be negligible, but the contributions of individual interaction energy terms could be non-negligible. Case in point: The contribution of ALA190 and VAL213 residues to the $\Delta(\Delta E)$ is less than 0.5 kcal/mol, while the electrostatic contribution to these residues are closer to 1 kcal/mol, which are then countered by other terms.

Another question we are trying to answer here is the validity of using the S1 pocket as a representative model for simulating the ligand-protein interactions. Figure 3.11 shows the convergence of $\Delta(\Delta E)$ components as a function of distance between centroids of individual residues in the S1 binding pocket. As one can expect, electrostatic interaction term converges very slowly, as the formal decay of charge-dominant inter-

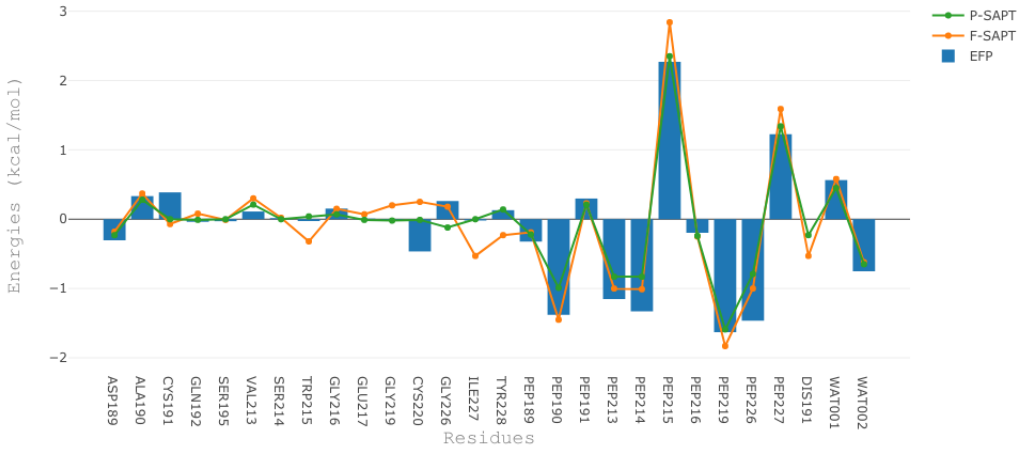



Fig. 3.9. Contribution of total interaction energies to $\Delta(\Delta E)$ term in the small ligand - S1 pocket model.

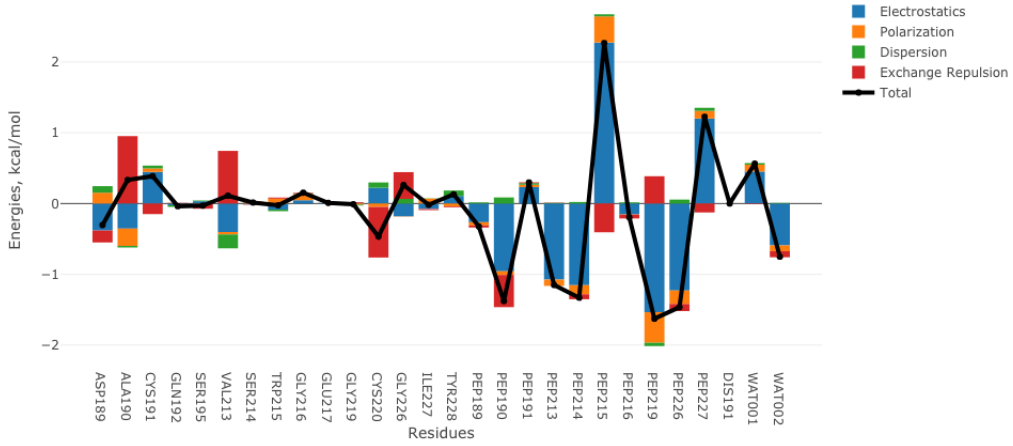



Fig. 3.10. Total interaction energies and energy component contributions of individual fragments to $\Delta(\Delta E)$ term in the small ligand-S1 pocket model.

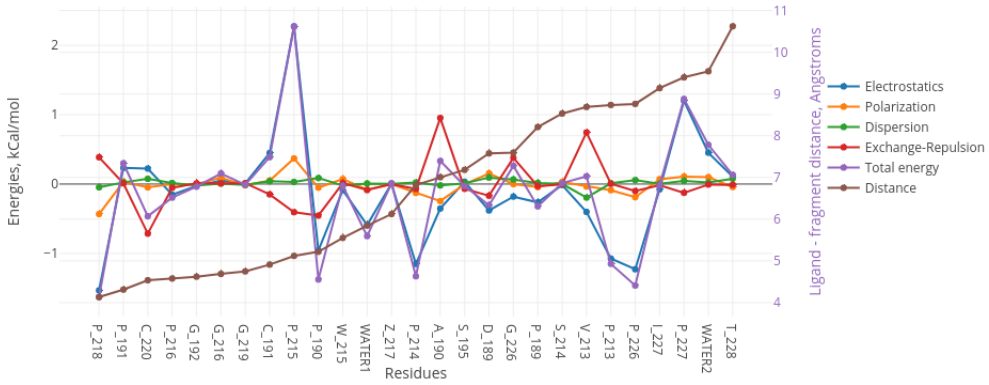



Fig. 3.11. Pairwise contributions to $\Delta(\Delta E)$ term in small ligand-S1 pocket model as a function of distances between ligand and fragments.

actions as a function of distance is $1/r$. Fragments such as Peptide226 and Peptide227 contribute well over 1 kcal/mol to the $\Delta(\Delta E)$ term, even though they are located 8-9 Å away from the ligand. This indicates that the ligand-S1 pocket may not be representative of all the significant interactions in the ligand-protein system.

To test this hypothesis, we simulated the entire protein-ligand system using BioEFP. Figure 3.12 shows individual fragment contributions to the $\Delta(\Delta E)$ term as a function of the separation between these fragments and the ligand. We can observe that the electrostatic contributions converge slowly, presenting a few interactions greater than 0.4 kcal/mol at distances beyond 1.5 nm. This is an indication that the substitution in the ligand is stabilized/destabilized by interactions with residues located well beyond the pocket.

Finally, we test the performance of BioEFP method in accurately predicting the preferential binding energies as a result of substitution. Table 3.3 lists the $\Delta\Delta G$ computed using various methods and systems. While the F-SAPT method predicts the total $\Delta\Delta G$ accurately to within 0.2 kcal/mol, P-SAPT underestimates the preferential Cl-ligand binding by 1.1 kcal/mol due to accumulation of errors. BioEFP

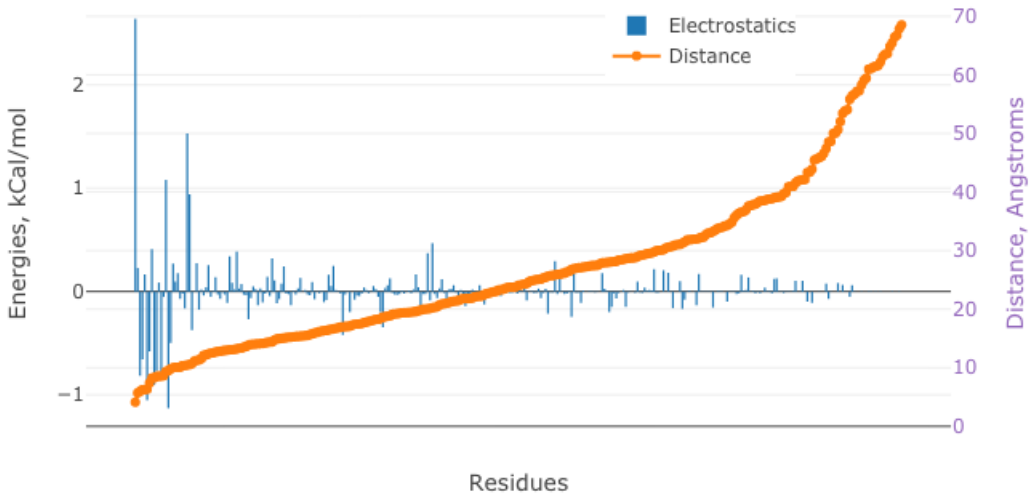



Fig. 3.12. Convergence of $\Delta(\Delta E_{elec})$ term in the ligand - protein complex as a function of distance.

Table 3.1.
Difference in binding energies between Cl- and Me- ligands, computed using different methods.

| **Method** | $\Delta\Delta E$ **(kcal/mol)** |
|---|---|
| **F-SAPT (small)** [133] | -2.464 |
| **P-SAPT (small)** [133] | -1.208 |
| **EFP (small)** | -3.314 |
| **EFP (large)** | -2.436 |
| **Experiment** [133] | -2.3 |

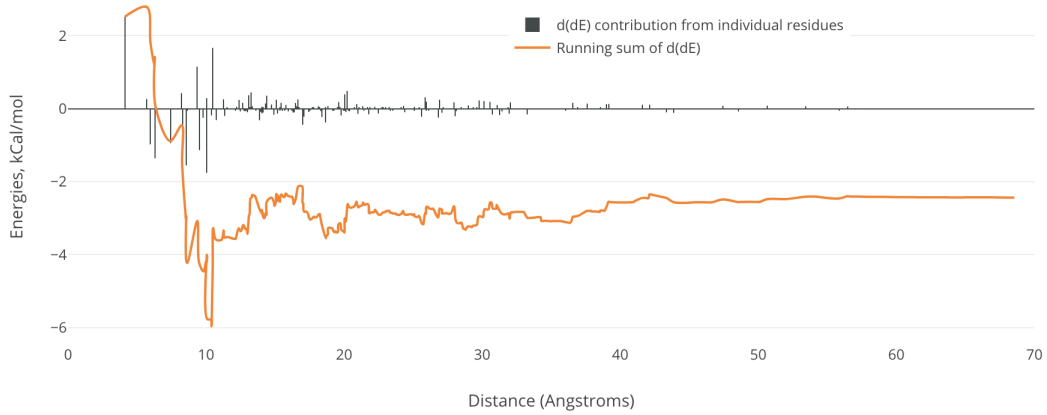



Fig. 3.13. Convergence of $\Delta(\Delta E_{total})$ term in the ligand - protein complex as a function of distance. A running sum of the $\Delta(\Delta E_{total})$ term is plotted (orange line).

overestimates the Cl-ligand binding by 1 kcal/mol for the smaller ligand-S1 pocket model, and this can be attributed to the lack of convergence in $\Delta\Delta G$ components for the smaller system. This is more evident in Fig. 3.13, where the presence of very strong stabilizing interactions can be noticed at around 1 nm from the ligand. These interactions are then countered by several destabilizing interactions beyond the 1 nm sphere, and the $\Delta\Delta G$ converges at around 4 nm from the ligand.

## 3.4 Conclusions

The performance of BioEFP as a tool for computing the intermolecular interaction energies and binding energy differences is probed. The results presented here are in good agreement with accurate first principles methods. Simulations of the ligand in the whole protein indicate that the electrostatic energy differences due to varying substituents do converge at distances far exceeding the size of the pocket previously used for modeling $\Delta(\Delta G)$ in this system. Performing simulation of the ligand with protein ensures that the interactions unaccounted for in the smaller system are now properly accounted for. BioEFP provides a viable option for performing such simulations at a much faster timeframe. Additional effects that need to be considered for obtaining rigorous binding energy differences are solvent effects and configurational sampling, which will be explored in future work.

# 4. METHOD DEVELOPMENT: ANALYTIC GRADIENTS OF THE QM/EFP DISPERSION TERM

Accurate modeling of dispersion interactions is essential for a proper and balanced description of intermolecular interactions. Here we show the working equations for the dispersion gradient term describing the coupling between quantum mechanical (QM) and classical mechanical regions (described using effective fragment potential (EFP) method). A detailed derivation of the equations governing the dispersion gradient is presented here. These equations involve orbitals in the quantum mechanical region and the dynamic polarizability tensors centered at the localized orbitals of the EFP region. Until now, the development and application of molecular dynamics (MD) and energy minimization (EM) routines involving EFP description was hindered due to the lack of mathematical equations governing dispersion and exchange-repulsion interactions. Once the codes corresponding to the equations are implemented, MD and EM simulations at the full-embedding level can be performed with QM/EFP methods.

## 4.1 Introduction

Electrostatic interactions involving charges, dipoles and higher order multipoles account for the majority of interactions between molecular clusters of most types - ions, polar molecules, biomolecular interactions, etc [139, 148–150]. Electrostatic interactions are responsible for the molecules (or molecular clusters in general) to exist in solid, liquid or gaseous phases [139]. However, molecules that do not have dominant electrostatic interactions such as non-polar molecules do exhibit similar behavior [151–153]. This indicates that there are some intermolecular interactions that are not particularly strong as simple electrostatic interactions, but do govern impor-

tant physical and chemical properties in molecules. Van der Waals forces is the sum of attractive and repulsive forces between two different atoms from same or different molecule which are caused due to factors other than covalent, ionic and hydrogen bonds [154, 155]. The van der Waals forces can be thought of as a superposition of dispersion and repulsive interactions in a given system. It is highly vital to include dispersion effects from the surrounding environment for accurate modeling of solute properties in a solvent [155, 156].

A detailed description of the effective fragment potential method [22]- [25] is given in the earlier chapters (chapter 1 and 2). For the purpose of this study, we can start by looking at the interaction terms [157] in the EFP method for a EFP/EFP type simulation:

$$E_{EFP-EFP} = E_{elec} + E_{pol} + E_{disp} + E_{ex-rep} + E_{CT} \tag{4.1}$$

For a hybrid, multiscale QM/EFP simulation, the total Hamiltonian is given by [23]:

$$\widehat{H}_{Total} = \widehat{H}_{QM} + \widehat{H}_{QM/EFP} + E_{EFP} \tag{4.2}$$

where $H_{QM/EFP}$ is the coupling term that augments the original Hamiltonian ($H_{QM}$), resulting in the total Hamiltonian ($H_{Total}$) that now includes the perturbation due to the environment.

In the current implementation of Effective Fragment Potential method, only the electrostatics and polarization terms account for the QM/EFP energy contribution in the ground and excited states. A framework for extending the dispersion [158] and exchange-repulsion [159] interactions to ground state QM/EFP simulations were published recently. Hence, at the polarizable embedding level, following the addition of dispersion correction, the coupling term is now a sum of three contributions:

$$\widehat{H}_{QM/EFP} = \widehat{H}^{elec}_{QM/EFP} + \widehat{H}^{pol}_{QM/EFP} + E^{disp}_{EFP} \tag{4.3}$$

where the diagonal QM/EFP dispersion term is given by:

$$E_{QM-EFP}^{disp,diag} = -\frac{1}{\pi}\sum_{j\in B}^{occ}\sum_{k}^{vir}\sum_{r}^{x,y,z}\sum_{\beta}\langle k|\, T_{\beta}^{j}\,|r\rangle\,\langle r|\, T_{\beta}^{j}\,|k\rangle\int_{0}^{\infty}\frac{\omega_{rk}^{A}}{(\omega_{rk}^{A})^{2}+\omega^{2}}\alpha_{\beta\beta}^{j}(i\omega)d\omega \quad (4.4)$$

In this work, we present the development of the gradient term corresponding to the above equation.

## 4.2 Theory

### 4.2.1 The QM/EFP dispersion term

In this section, the energy expression for QM/EFP dispersion term is explained briefly. For a more detailed derivation and benchmarks, please refer to [158].

The dispersion interactions between two molecules $A$ and $B$ can be described using the second order Rayleigh-Schrodinger Perturbation Theory (RSPT2) as follows [139]:

$$E_{disp} = -\sum_{m\neq 0, n\neq 0}\frac{\langle 00|\,\widehat{H'}\,|mn\rangle\,\langle mn|\,\widehat{H'}\,|00\rangle}{W_{m}^{A}+W_{n}^{B}-W_{0}^{A}-W_{0}^{B}} \quad (4.5)$$

where $m$ and $n$ are the electronic states of molecules $A$ and $B$ respectively; $\widehat{H'}$ is the preturbed Hamiltonian that is expressed as a function of the charge density operators for molecules $A$ and $B$. The subscripts 0 correspond to the electronic ground state of the molecules. $W$ corresponds to the energy of the electronic state of the molecule $A$ or $B$. This expression can be then expanded using the electric field expansion, and after assuming separability of the states on different molecules, we get the following expression:

$$E^{disp} = \sum_{m\neq 0, n\neq 0} T_{\alpha\beta}^{AB}T_{\gamma\delta}^{AB}\frac{\langle 0^{A}|\,\widehat{\mu}_{\alpha}^{A}\,|m\rangle\,\langle m|\,\widehat{\mu}_{\gamma}^{A}\,|0^{A}\rangle\,\langle 0^{B}|\,\widehat{\mu}_{\beta}^{B}\,|n\rangle\,\langle n|\,\widehat{\mu}_{\delta}^{B}\,|0^{B}\rangle}{E_{m0}^{A}+E_{n0}^{B}} \quad (4.6)$$

where $T_{\alpha\beta}$ is the second order electrostatic tensor, $\mu^{A/B}$ are the dipole moment operators corresponding to fragment $A$ or $B$, $E_{m0}^{A}$ is the excitation energy corresponding to a $m \leftarrow 0$ transition of molecule A.

Further, by applying the McLachlan identity [160] and by utilizing a distributed approach in which the dynamic polarizabilities are distributed at different points throughout the molecule, the expression for EFP-EFP dispersion is obtained [161]:

$$E^{disp}_{EFP-EFP} = -\frac{3}{\pi}\sum_{k\in A}\sum_{j\in B}\frac{1}{R^6_{kj}}\int_0^\infty \overline{\alpha}^k(i\omega)\overline{\alpha}^j(i\omega)d\omega \tag{4.7}$$

where $k$ and $j$ are points centered at localized molecular orbitals in EFP fragments $A$ and $B$ respectively. The dynamic polarizability tensor $\alpha$ at a given frequency $\omega$ is given by:

$$\alpha_{\alpha\beta}(\omega) = 2\sum_m \frac{\omega_{m0}\left\langle 0\right|\widehat{\mu}_\alpha\left|m\right\rangle\left\langle m\right|\widehat{\mu}_\beta\left|0\right\rangle}{\hbar(\omega^2_{m0}-\omega^2)} \tag{4.8}$$

It must be noted that the dynamic polarizability tensor in 4.8 is anisotropic, while the polarizability tensors in 4.7 utilizes isotropic approximation, and is hence a scalar quantity.

Similarly, the energy expression for QM/EFP dispersion interactions is given by [158]:

$$E^{disp,anisotropic}_{QM-EFP} = -\frac{1}{\pi}\sum_{j\in B}\sum_k^{occ}\sum_r^{vir}\sum_\beta^{x,y,z}\left\langle k\right|T^j_\alpha\left|r\right\rangle\left\langle r\right|T^j_\beta\left|k\right\rangle\int_0^\infty\frac{\omega^A_{rk}}{(\omega^A_{rk})^2+\omega^2}\alpha^j_{\alpha\beta}(i\omega)d\omega \tag{4.9}$$

where $T_\alpha$ is the electrostatic field tensor of order one, $j$ are the localized molecular orbital (LMO) expansion points in fragment $B$, $k$ and $r$ are occupied and virtual orbitals in the QM region respectively, subscript $\alpha$ and $\beta$ are the cartesian axes corresponding to the electrostatic field tensor, $\alpha$ is the polarizability tensor at point $j \in B$, and $\omega^A_{rk}$ is the orbital energy difference between virtual orbital $r$ and occupied orbital $k$ of the quantum mechanical region, depicted by $A$. It must be noted that 4.9 follows an orbital-based summation, while earlier equations make use of a sum-over-states expansion.

By setting the off-diagonal terms of $\alpha$ to zero we obtain the isotropic expression for the dispersion term, as given in 4.4. Following this, one can apply a spherical

approximation in which the total polarizability computed at a point is approximated as the sum of the diagonal terms ($\overline{\alpha} = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$, we obtain the following expression:

$$E^{disp}_{QM-EFP} = -\frac{1}{\pi}\sum_{j\in B}\sum_{k}^{occ}\sum_{r}^{vir}\sum_{\beta}^{x,y,z}\langle k|\, T^j_\beta\,|r\rangle\,\langle r|\, T^j_\beta\,|k\rangle\int_0^\infty \frac{\omega^A_{rk}}{(\omega^A_{rk})^2+\omega^2}\overline{\alpha}^j(i\omega)d\omega \quad (4.10)$$

### 4.2.2 Definitions

In this section, some basic definitions for orbital derivatives and the resulting expressions are explored briefly. The energy expression for second order perturbation theory (MP2) is given by [162]:

$$E_{MP2} = \frac{1}{4}\sum_{ij}^{occ}\sum_{ab}^{vir}\left((ia||jb)\right)T^{ab}_{ij} \quad (4.11)$$

where the amplitude $T^{ab}_{ij}$ is given by:

$$T^{ab}_{ij} = (ia||jb)/D^{ab}_{aj} \quad (4.12)$$

In the above equation, $(D^{ab}_{ij} = \epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b)$.

Instantly, one can notice the similarities between equations 4.11 and 4.5, where the numerator terms are one/two electron integrals, and the denominator term contain orbital energies.

The derivative of a molecular orbital can be written as [163]:

$$\frac{\partial\,|p\rangle}{\partial x} = |p\rangle^x = \left(\sum_\mu^{AO} C_{\mu p}\,|\mu\rangle\right)^x = \sum_\mu^{AO} C_{\mu p}\,|\mu\rangle^x + \sum_\mu^{AO} C^x_{\mu p}\,|\mu\rangle$$

$$= |p\rangle^{(x)} + \sum_\mu^{AO}\sum_q^{all} U^x_{qp}C_{\mu q}\,|\mu\rangle = |p\rangle^{(x)} + \sum_q^{all} U^x_{qp}\,|q\rangle \quad (4.13)$$

The above equation describes the derivative of a molecular orbital due to perturbation along x axis. The basis set expansion coefficients are depicted by $C$; the

superscript $x$ indicates the derivative with respect to a nuclear derivative along the x-direction, while the superscript in paranthesis $(x)$ describes the derivatives of atomic orbitals only. Separating the molecular orbital using LCAO-MO [164] approximation and splitting the contributions into derivatives of orbital functions $(|p\rangle^{(x)}$ and the derivatives of expansion coefficients (i.e. the orbital response - $U^x_{pq}$ ) enables us to simplify the problem and concentrate on the latter terms.

We can rewrite the two-electron integrals in MO basis into two-electron integrals in AO basis as follows:

$$(pq|rs) = \sum_{\mu\nu\lambda\sigma} C_{\mu p} C_{\nu q} C_{\lambda r} C_{\sigma s} (\mu\nu|\lambda\sigma)$$

The gradient for the above two-electron integral can be written as

$$(pq|rs)^x = (pq|rs)^{(x)} + \sum_t^{all} U^x_{tp}(tq|rs) + \sum_t^{all} U^x_{tq}(pt|rs) + \sum_t^{all} U^x_{tr}(pq|ts) + \sum_t^{all} U^x_{ts}(pq|rt) \tag{4.14}$$

Due to the orthonormality constraint in the overlap between two MOs ($S_{pq} = \delta_{pq}$) the derivative of this constraint can be expressed the following way:

$$(S_{pq})^x = U^x_{pq} + U^x_{qp} + S^{(x)}_{pq} = 0 \tag{4.15}$$

Hence the orbital response involving the same MOs can be expressed as:

$$U^x_{pp} = -\frac{1}{2} S^{(x)}_{pp} \tag{4.16}$$

From a mathematical point of view, the orbital derivatives are easy to derive and are readily available in most quantum chemistry packages. However, the orbital response term needs to be computed using coupled perturbed Hartree-Fock (CPHF) method by solving for the coupled equations described below [165].

$$U^x_{pq} = \frac{1}{(\epsilon_q - \epsilon_p)} Q^x_{pq} \tag{4.17}$$

The quantity $Q^x_{pq}$ is given by:

$$Q^x_{pq} = (B^x_{pq} + \sum_{c}^{vir}\sum_{k}^{occ} U^x_{ck} A_{pqck}) \tag{4.18}$$

$$A_{pqck} = 2(pq|ck) - (pc|qk) - (pk|qc) \tag{4.19}$$

$$B^x_{pq} = F^{(x)}_{pq} - S^{(x)}_{pq}\epsilon_q - \frac{1}{2}\sum_{kl}^{occ} S^{(x)}_{kl} A_{pqlk} \tag{4.20}$$

$$F^{(x)}_{pq} = H^{(x)}_{pq} + \sum_{k}^{occ}\left[(pq|kk)^{(x)} - (pk|qk)^{(x)}\right] \tag{4.21}$$

For $p = q$, equation 4.18 can be simplified as $Q^x_{pp} = \epsilon^x_p$. It can be seen that the orbital response term is now expressed as a function of simpler overlap integral derivatives, one electron Hamiltonian integral derivatives, and orbital energies, all of which can be obtained by solving the coupled Hartree-Fock equations:

$$\sum_{i}^{occ}\sum_{a}^{vall}\left[\delta_{ab}\delta_{ij}(\epsilon_i - \epsilon_a) - A_{aibj}\right] U^x_{ai} = B^x_{bj} \tag{4.22}$$

One can notice that the terms $U$ and $B$ are coupled to each other and need to be solved for self consistently. For a more detailed derivation of the second order perturbation theory gradient expression, please refer to [165].

### 4.2.3 The QM/EFP dispersion gradient term

We rewrite the energy expression for the diagonal QM/EFP dispersion energy term from 4.9:

$$E^{disp}_{QM-EFP} = -\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta} \langle k|\, T^j_\beta \,|r\rangle \langle r|\, T^j_\beta \,|k\rangle \int_0^\infty \frac{\omega^A_{rk}}{(\omega^A_{rk})^2 + \omega^2}\alpha^j_{\beta\beta}(i\omega)d\omega \tag{4.23}$$

The derivative of the above expression can be written as:

$$\frac{d}{dx}E_{disp} = E^x_{disp} = -\frac{1}{\pi}\sum_j\sum_k\sum_r$$

$$\Bigg[\sum_\beta \langle k|\, T^j_\beta\, |r\rangle^x \langle r|\, T^j_\beta\, |k\rangle \int_0^\infty \frac{\omega^A_{rk}}{(\omega^A_{rk})^2+\omega^2}\alpha^j_{\beta\beta}(i\omega)d\omega \tag{4.24a}$$

$$+\sum_\beta \langle k|\, T^j_\beta\, |r\rangle \langle r|\, T^j_\beta\, |k\rangle^x \int_0^\infty \frac{\omega^A_{rk}}{(\omega^A_{rk})^2+\omega^2}\alpha^j_{\beta\beta}(i\omega)d\omega \tag{4.24b}$$

$$+\sum_\beta \langle k|\, T^j_\beta\, |r\rangle \langle r|\, T^j_\beta\, |k\rangle \left(\int_0^\infty \frac{\omega^A_{rk}}{(\omega^A_{rk})^2+\omega^2}\alpha^j_{\beta\beta}(i\omega)d\omega\right)^x\Bigg] \tag{4.24c}$$

We can call 4.24a as Part A, 4.24b as Part B and 4.24c as Part C. One can note that the parts A and B are derivatives on the field integrals, while the part C includes derivative terms from both QM as well as EFP regions. Now, let us first consider part A and expand the terms:

$$\sum_\beta \langle k|\, T_\beta\, |r\rangle^x \langle r|\, T_\beta\, |k\rangle \int_0^\infty \frac{\omega^A_{rk}}{(\omega^A_{rk})^2+\omega^2}\alpha^j_{\alpha\beta}(i\omega)d\omega$$

$$= \Bigg(\langle k|\, T_\beta\, |r\rangle^{(x)} +$$

$$\sum_t U^x_{tk}\langle t|\, T_\beta\, |r\rangle + \sum_t U^x_{tr}\langle k|\, T_\beta\, |t\rangle\Bigg) \tag{4.25a}$$

$$\langle r|\, T_\beta\, |k\rangle \int_0^\infty \frac{\omega^A_{rk}}{(\omega^A_{rk})^2+\omega^2}\alpha^j_{\alpha\beta}(i\omega)d\omega$$

Let us call 4.25a as Part A1. It can be noticed that the rest of the terms in 4.25 are either derivatives of orbital basis functions or unmodified terms from the original dispersion expression. The summation index $t$ for the terms in part A1 includes all molecular orbitals - occupied as well as virtual. We can now separate A1 into two sets of occupied and virtual orbitals as follows:

$$\sum_t U^x_{tk}\langle t|\, T_\beta\, |r\rangle + \sum_t U^x_{tr}\langle k|\, T_\beta\, |t\rangle$$

$$= \sum_{i}^{occ} U_{ik}^{x} \langle i| T_\beta |r\rangle + \tag{4.26a}$$

$$\sum_{c}^{vir} U_{ck}^{x} \langle c| T_\beta |r\rangle +$$

$$\sum_{i}^{occ} U_{ir}^{x} \langle k| T_\beta |i\rangle +$$

$$\sum_{c}^{vir} U_{cr}^{x} \langle k| T_\beta |c\rangle \tag{4.26b}$$

Let us call 4.26a as Part A1a and 4.26b as Part A1b. Now, let us combine part A1a with rest of part A:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\left(\sum_{i}^{occ} U_{ik}^{x} \langle i| T_\beta |r\rangle\right) \langle r| T_\beta |k\rangle \int_0^\infty \frac{\omega_{rk}^{A}}{(\omega_{rk}^{A})^2+\omega^2}\alpha_{\beta\beta}^{j}(i\omega)d\omega$$

$$= -\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\left(\sum_{i}^{occ} U_{ik}^{x} \langle i| T_\beta |r\rangle\right) \langle r| T_\beta |k\rangle I_{rk} \tag{4.27}$$

where the integral is represented in a simpler notation: $I_{rk}$. From 4.15, we know that $(S_{pq})^x = U_{pq}^x + U_{qp}^x + S_{pq}^{(x)} = 0$. This is now applied in the above equation to expand as follows:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\left(\sum_{i}^{occ} U_{ik}^{x} \langle i| T_\beta |r\rangle\right) \langle r| T_\beta |k\rangle I_{rk}$$

$$= -\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\left(\sum_{i}^{occ}\frac{1}{2}\left[U_{ik}^{x} \langle i| T_\beta |r\rangle - U_{ki}^{x} \langle i| T_\beta |r\rangle - S_{ik}^{(x)} \langle i| T_\beta |r\rangle\right]\right) \langle r| T_\beta |k\rangle I_{rk} \tag{4.28}$$

Now, we can isolate the $U_{kk}$ term by separating it into three parts:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\frac{1}{2}\left[\left(\sum_{i>k}^{occ} U_{ik}^{x} \langle i| T_\beta |r\rangle + \sum_{i<k}^{occ} U_{ik}^{x} \langle i| T_\beta |r\rangle + U_{kk}^{x} \langle k| T_\beta |r\rangle\right)\right.$$

$$-\left(\sum_{i>k}^{occ} U^x_{ki} \langle i| T_\beta |r\rangle + \sum_{i<k}^{occ} U^x_{ki} \langle i| T_\beta |r\rangle + U^x_{kk} \langle k| T_\beta |r\rangle \right) - \sum_{i}^{occ} S^{(x)}_{ik} \langle i| T_\beta |r\rangle \Bigg] \Bigg) \langle r| T_\beta |k\rangle I_{rk} \tag{4.29}$$

We can see that the terms highlighted in red cancel each other exactly. Now, interchanging the indices $i$ and $k$ in the terms highlighted in blue results in:

$$-\frac{1}{\pi}\sum_j\sum_k\sum_r\sum_\beta \frac{1}{2}\Bigg[\left(\sum_{i>k}^{occ} U^x_{ik} \langle i| T_\beta |r\rangle \langle r| T_\beta |k\rangle I_{rk} + \sum_{i>k}^{occ} U^x_{ki} \langle k| T_\beta |r\rangle \langle r| T_\beta |i\rangle I_{ri}\right) - \left(\sum_{i>k}^{occ} U^x_{ki} \langle i| T_\beta |r\rangle \langle r| T_\beta |k\rangle I_{rk} + \sum_{i>k}^{occ} U^x_{ik} \langle k| T_\beta |r\rangle \langle r| T_\beta |i\rangle I_{ri}\right) - \sum_{i}^{occ} S^{(x)}_{ik} \langle i| T_\beta |r\rangle \langle r| T_\beta |k\rangle I_{rk}\Bigg] \tag{4.30}$$

Collecting like terms, we get:

$$-\frac{1}{\pi}\sum_j\sum_k\sum_r\sum_\beta \frac{1}{2}\Bigg[\left(\sum_{i>k}^{occ} U^x_{ik} \langle i| T_\beta |r\rangle \langle r| T_\beta |k\rangle I_{rk} - \sum_{i>k}^{occ} U^x_{ik} \langle k| T_\beta |r\rangle \langle r| T_\beta |i\rangle I_{ri}\right) - \left(\sum_{i>k}^{occ} U^x_{ki} \langle i| T_\beta |r\rangle \langle r| T_\beta |k\rangle I_{rk} - \sum_{i>k}^{occ} U^x_{ki} \langle k| T_\beta |r\rangle \langle r| T_\beta |i\rangle I_{ri}\right) - \sum_{i}^{occ} S^{(x)}_{ik} \langle i| T_\beta |r\rangle \langle r| T_\beta |k\rangle I_{rk}\Bigg] \tag{4.31}$$

Combining the terms containing the same field integrals and orbital responses, we get:

$$-\frac{1}{\pi}\sum_j\sum_k\sum_r\sum_\beta \frac{1}{2}\Bigg[\left(\sum_{i>k}^{occ} U^x_{ik} \langle i| T_\beta |r\rangle \langle r| T_\beta |k\rangle (I_{rk} - I_{ri})\right) - \left(\sum_{i>k}^{occ} U^x_{ki} \langle i| T_\beta |r\rangle \langle r| T_\beta |k\rangle (I_{rk} - I_{ri})\right)$$

$$-\sum_{i}^{occ} S_{ik}^{(x)} \langle i| T_\beta |r\rangle \langle r| T_\beta |k\rangle I_{rk}\Bigg] \tag{4.32}$$

Now, let us expand the $I_{rk} - I_{ri}$ term:

$$I_{rk} - I_{ri} = \int_0^\infty \frac{\omega_{rk}^A}{(\omega_{rk}^A)^2 + \omega^2} \alpha_{\beta\beta}^j(i\omega) d\omega - \int_0^\infty \frac{\omega_{ri}^A}{(\omega_{ri}^A)^2 + \omega^2} \alpha_{\beta\beta}^j(i\omega) d\omega \tag{4.33}$$

Dividing and multiplying by common factors, we obtain:

$$= \int_0^\infty \frac{\omega_{rk}^A \Big[(\omega_{ri}^A)^2 + \omega^2\Big]}{\Big[(\omega_{rk}^A)^2 + \omega^2\Big]\Big[(\omega_{ri}^A)^2 + \omega^2\Big]} \alpha_{\beta\beta}^j(i\omega) d\omega - \int_0^\infty \frac{\omega_{ri}^A \Big[(\omega_{rk}^A)^2 + \omega^2\Big]}{\Big[(\omega_{ri}^A)^2 + \omega^2\Big]\Big[(\omega_{rk}^A)^2 + \omega^2\Big]} \alpha_{\beta\beta}^j(i\omega) d\omega \tag{4.34}$$

combining both the terms due to common denominator, we get:

$$= \int_0^\infty \frac{\omega_{rk}^A \Big[(\omega_{ri}^A)^2 + \omega^2\Big] - \omega_{ri}^A \Big[(\omega_{rk}^A)^2 + \omega^2\Big]}{\Big[(\omega_{rk}^A)^2 + \omega^2\Big]\Big[(\omega_{ri}^A)^2 + \omega^2\Big]} \alpha_{\beta\beta}^j(i\omega) d\omega \tag{4.35}$$

$$= \int_0^\infty \frac{\omega_{rk}^A (\omega_{ri}^A)^2 + \omega_{rk}^A \omega^2 - \omega_{ri}^A (\omega_{rk}^A)^2 - \omega_{ri}^A \omega^2}{\Big[(\omega_{rk}^A)^2 + \omega^2\Big]\Big[(\omega_{ri}^A)^2 + \omega^2\Big]} \alpha_{\beta\beta}^j(i\omega) d\omega \tag{4.36}$$

$$= \int_0^\infty \frac{\omega_{rk}^A \omega_{ri}^A (\omega_{ri}^A - \omega_{rk}^A) - \omega^2 (\omega_{ri}^A - \omega_{rk}^A)}{\Big[(\omega_{rk}^A)^2 + \omega^2\Big]\Big[(\omega_{ri}^A)^2 + \omega^2\Big]} \alpha_{\beta\beta}^j(i\omega) d\omega \tag{4.37}$$

$$= \int_0^\infty \frac{(\omega_{ri}^A - \omega_{rk}^A)(\omega_{ri}^A \omega_{rk}^A - \omega^2)}{\Big[(\omega_{rk}^A)^2 + \omega^2\Big]\Big[(\omega_{ri}^A)^2 + \omega^2\Big]} \alpha_{\beta\beta}^j(i\omega) d\omega \tag{4.38}$$

We can now rewrite $\omega_{ri}$ and $\omega_{rk}$ in terms of orbital energies, which can be taken out of the integral.:

$$= (\epsilon_k - \epsilon_i) \int_0^\infty \frac{((\omega_{ri}^A \omega_{rk}^A - \omega^2)}{\Big[(\omega_{rk}^A)^2 + \omega^2\Big]\Big[(\omega_{ri}^A)^2 + \omega^2\Big]} \alpha_{\beta\beta}^j(i\omega) d\omega \tag{4.39}$$

Combining 4.39 with 4.32, we get:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\frac{1}{2}\Bigg[\Bigg(\sum_{i>k}^{occ}U_{ik}^{x}\left\langle i\right|T_{\beta}\left|r\right\rangle\left\langle r\right|T_{\beta}\left|k\right\rangle$$
$$(\epsilon_k-\epsilon_i)\int_0^{\infty}\frac{((\omega_{ri}^{A}\omega_{rk}^{A}-\omega^2)}{\Big[(\omega_{rk}^{A})^2+\omega^2\Big]\Big[(\omega_{ri}^{A})^2+\omega^2\Big]}\alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg)$$
$$-\Bigg(\sum_{i>k}^{occ}U_{ki}^{x}\left\langle i\right|T_{\beta}\left|r\right\rangle\left\langle r\right|T_{\beta}\left|k\right\rangle$$
$$(\epsilon_i-\epsilon_k)\int_0^{\infty}\frac{((\omega_{ri}^{A}\omega_{rk}^{A}-\omega^2)}{\Big[(\omega_{rk}^{A})^2+\omega^2\Big]\Big[(\omega_{ri}^{A})^2+\omega^2\Big]}\alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg)$$
$$-\sum_{i}^{occ}S_{ik}^{(x)}\left\langle i\right|T_{\beta}\left|r\right\rangle\left\langle r\right|T_{\beta}\left|k\right\rangle\int_0^{\infty}\frac{\omega_{rk}^{A}}{(\omega_{rk}^{A})^2+\omega^2}\alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg] \tag{4.40}$$

Now, we expand the orbital responses in terms of known quantities as described in 4.17 - 4.21:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\frac{1}{2}\Bigg[\Bigg(\sum_{i>k}^{occ}\frac{Q_{ik}^{x}}{(\epsilon_i-\epsilon_k)}\left\langle i\right|T_{\beta}\left|r\right\rangle\left\langle r\right|T_{\beta}\left|k\right\rangle$$
$$(\epsilon_k-\epsilon_i)\int_0^{\infty}\frac{((\omega_{ri}^{A}\omega_{rk}^{A}-\omega^2)}{\Big[(\omega_{rk}^{A})^2+\omega^2\Big]\Big[(\omega_{ri}^{A})^2+\omega^2\Big]}\alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg)$$
$$-\Bigg(\sum_{i>k}^{occ}\frac{Q_{ki}^{x}}{(\epsilon_k-\epsilon_i)}\left\langle i\right|T_{\beta}\left|r\right\rangle\left\langle r\right|T_{\beta}\left|k\right\rangle$$
$$(\epsilon_i-\epsilon_k)\int_0^{\infty}\frac{((\omega_{ri}^{A}\omega_{rk}^{A}-\omega^2)}{\Big[(\omega_{rk}^{A})^2+\omega^2\Big]\Big[(\omega_{ri}^{A})^2+\omega^2\Big]}\alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg)$$
$$-\sum_{i}^{occ}S_{ik}^{(x)}\left\langle i\right|T_{\beta}\left|r\right\rangle\left\langle r\right|T_{\beta}\left|k\right\rangle\int_0^{\infty}\frac{\omega_{rk}^{A}}{(\omega_{rk}^{A})^2+\omega^2}\alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg] \tag{4.41}$$

The orbital energy difference term cancels out exactly, as given below:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\frac{1}{2}\Bigg[\Bigg(\sum_{i>k}^{occ}Q_{ik}^{x}\left\langle i\right|T_{\beta}\left|r\right\rangle\left\langle r\right|T_{\beta}\left|k\right\rangle$$
$$\int_0^{\infty}\frac{((\omega_{ri}^{A}\omega_{rk}^{A}-\omega^2)}{\Big[(\omega_{rk}^{A})^2+\omega^2\Big]\Big[(\omega_{ri}^{A})^2+\omega^2\Big]}\alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg)$$

$$+\Bigg(\sum_{i>k}^{occ} Q_{ki}^{x} \langle i|\, T_\beta\, |r\rangle \langle r|\, T_\beta\, |k\rangle \int_0^\infty \frac{((\omega_{ri}^A \omega_{rk}^A - \omega^2)}{\Big[(\omega_{rk}^A)^2+\omega^2\Big]\Big[(\omega_{ri}^A)^2+\omega^2\Big]} \alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg)$$

$$-\sum_{i}^{occ} S_{ik}^{(x)} \langle i|\, T_\beta\, |r\rangle \langle r|\, T_\beta\, |k\rangle \int_0^\infty \frac{\omega_{rk}^A}{(\omega_{rk}^A)^2+\omega^2} \alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg] \tag{4.42}$$

Let us introduce a shorter notation for the terms as below:

$$\langle i|\, T_\beta\, |r\rangle = T_\beta^{ir}$$

$$\int_0^\infty \frac{((\omega_{ri}^A \omega_{rk}^A - \omega^2)}{\Big[(\omega_{rk}^A)^2+\omega^2\Big]\Big[(\omega_{ri}^A)^2+\omega^2\Big]} \alpha_{\beta\beta}^{j}(i\omega)d\omega = J_{rk-ri}$$

This simplifies 4.42 into the following form:

$$-\frac{1}{\pi}\sum_j\sum_k\sum_r\sum_\beta \frac{1}{2}\Bigg[\Bigg(\sum_{i>k}^{occ} Q_{ik}^{x} T_\beta^{ir} T_\beta^{rk} J_{rk-ri}\Bigg)+\Bigg(\sum_{i>k}^{occ} Q_{ki}^{x} T_\beta^{kr} T_\beta^{ri} J_{rk-ri}\Bigg)-\sum_{i}^{occ} S_{ik}^{(x)} T_\beta^{ir} T_\beta^{rk} I_{rk}\Bigg] \tag{4.43}$$

Similarly, after expanding part A1b which contains the summation of the orbital responses involving virtual orbitals, we get the following expression:

$$-\frac{1}{\pi}\sum_j\sum_k\sum_r\sum_\beta \frac{1}{2}\Bigg[\Bigg(\sum_{c>r}^{vir} Q_{cr}^{x} \langle k|\, T_\beta\, |c\rangle \langle r|\, T_\beta\, |k\rangle \int_0^\infty \frac{((\omega_{rk}^A \omega_{ck}^A - \omega^2)}{\Big[(\omega_{rk}^A)^2+\omega^2\Big]\Big[(\omega_{ck}^A)^2+\omega^2\Big]} \alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg)$$

$$+\Bigg(\sum_{c>r}^{vir} Q_{rc}^{x} \langle k|\, T_\beta\, |r\rangle \langle c|\, T_\beta\, |k\rangle \int_0^\infty \frac{((\omega_{ck}^A \omega_{rk}^A - \omega^2)}{\Big[(\omega_{ck}^A)^2+\omega^2\Big]\Big[(\omega_{rk}^A)^2+\omega^2\Big]} \alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg)$$

$$-\sum_{i}^{occ} S_{cr}^{(x)} \langle k|\, T_\beta\, |c\rangle \langle r|\, T_\beta\, |k\rangle \int_0^\infty \frac{\omega_{rk}^A}{(\omega_{rk}^A)^2+\omega^2} \alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg] \tag{4.44}$$

Simplifying the terms leads to the following expression:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\frac{1}{2}\left[\left(\sum_{c>r}^{vir}Q^{x}_{cr}T^{kc}_{\beta}T^{rk}_{\beta}J_{rk-ck}\right)+\left(\sum_{c>r}^{vir}Q^{x}_{rc}T^{kr}_{\beta}T^{ck}_{\beta}J_{rk-ck}\right)-\sum_{c}^{vir}S^{(x)}_{cr}T^{kc}_{\beta}T^{rk}_{\beta}I_{rk}\right] \tag{4.45}$$

Adding the parts A1a, A1b and the remaining terms in A1 results in the following expression:

$$\begin{aligned}-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\frac{1}{2}\Bigg[&\left(\sum_{i>k}^{occ}Q^{x}_{ik}T^{ir}_{\beta}T^{rk}_{\beta}J_{rk-ri}\right)+\left(\sum_{i>k}^{occ}Q^{x}_{ki}T^{kr}_{\beta}T^{ri}_{\beta}J_{rk-ri}-\sum_{i}^{occ}S^{(x)}_{ik}T^{ir}_{\beta}T^{rk}_{\beta}I_{rk}\right)\\&+\left(\sum_{c>r}^{vir}Q^{x}_{cr}T^{kc}_{\beta}T^{rk}_{\beta}J_{rk-ck}\right)+\left(\sum_{c>r}^{vir}Q^{x}_{rc}T^{kr}_{\beta}T^{ck}_{\beta}J_{rk-ck}-\sum_{c}^{vir}S^{(x)}_{cr}T^{kc}_{\beta}T^{rk}_{\beta}I_{rk}\right)\\&+\sum_{c}^{vir}U^{x}_{ck}T^{cr}_{\beta}+\sum_{i}^{occ}U^{x}_{ir}T^{ki}_{\beta}\Bigg]\end{aligned} \tag{4.46}$$

From 4.18, we know that $Q^{x}_{pq}=B^{x}_{pq}+\sum_{c}^{vir}\sum_{k}^{occ}U^{x}_{ck}A_{pqck}$. We can now use this information in 4.46:

$$\begin{aligned}-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\frac{1}{2}\Bigg[&\left(\sum_{i>k}^{occ}\left(B^{x}_{ik}+\sum_{b}^{vir}\sum_{l}^{occ}U^{x}_{bl}A_{ikbl}\right)T^{ir}_{\beta}T^{rk}_{\beta}J_{rk-ri}\right)\\&+\left(\sum_{i>k}^{occ}\left(B^{x}_{ki}+\sum_{b}^{vir}\sum_{l}^{occ}U^{x}_{bl}A_{kibl}\right)T^{kr}_{\beta}T^{ri}_{\beta}J_{rk-ri}-\sum_{i}^{occ}S^{(x)}_{ik}T^{ir}_{\beta}T^{rk}_{\beta}I_{rk}\right)\\&+\left(\sum_{c>r}^{vir}\left(B^{x}_{cr}+\sum_{b}^{vir}\sum_{l}^{occ}U^{x}_{bl}A_{crbl}\right)T^{kc}_{\beta}T^{rk}_{\beta}J_{rk-ck}\right)\\&+\left(\sum_{c>r}^{vir}\left(B^{x}_{rc}+\sum_{b}^{vir}\sum_{l}^{occ}U^{x}_{bl}A_{rcbl}\right)T^{kr}_{\beta}T^{ck}_{\beta}J_{rk-ck}-\sum_{c}^{vir}S^{(x)}_{cr}T^{kc}_{\beta}T^{rk}_{\beta}I_{rk}\right)\\&+\sum_{c}^{vir}U^{x}_{ck}T^{cr}_{\beta}+\sum_{i}^{occ}U^{x}_{ir}T^{ki}_{\beta}\Bigg]\end{aligned} \tag{4.47}$$

By visual inspection, one can see that the solutions to parts A and B are exactly the same, and hence the above equation is multiplied by a factor of 2.

$$parts(A+B) =$$

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\Bigg[\Bigg(\sum_{i>k}^{occ}\Bigg(B_{ik}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{ikbl}\Bigg)T_{\beta}^{ir}T_{\beta}^{rk}J_{rk-ri}\Bigg)$$
$$+\Bigg(\sum_{i>k}^{occ}\Bigg(B_{ki}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{kibl}\Bigg)T_{\beta}^{kr}T_{\beta}^{ri}J_{rk-ri}-\sum_{i}^{occ}S_{ik}^{(x)}T_{\beta}^{ir}T_{\beta}^{rk}I_{rk}\Bigg)$$
$$+\Bigg(\sum_{c>r}^{vir}\Bigg(B_{cr}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{crbl}\Bigg)T_{\beta}^{kc}T_{\beta}^{rk}J_{rk-ck}\Bigg)$$
$$+\Bigg(\sum_{c>r}^{vir}\Bigg(B_{rc}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{rcbl}\Bigg)T_{\beta}^{kr}T_{\beta}^{ck}J_{rk-ck}-\sum_{c}^{vir}S_{cr}^{(x)}T_{\beta}^{kc}T_{\beta}^{rk}I_{rk}\Bigg)$$
$$+\sum_{c}^{vir}U_{ck}^{x}T_{\beta}^{cr}+\sum_{i}^{occ}U_{ir}^{x}T_{\beta}^{ki}\Bigg] \tag{4.48}$$

Part C of the dispersion energy gradient is expanded as follows:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}T_{\beta}^{kr}T_{\beta}^{rk}\frac{d}{dx}\Bigg(\int_{0}^{\infty}\frac{\omega_{rk}^{A}}{(\omega_{rk}^{A})^{2}+\omega^{2}}\alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg) \tag{4.49}$$

Taking the derivative by parts, we obtain:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}T_{\beta}^{kr}T_{\beta}^{rk}\Bigg(\int_{0}^{\infty}\frac{(\omega_{rk}^{A})^{x}}{(\omega_{rk}^{A})^{2}+\omega^{2}}\alpha_{\beta\beta}^{j}(i\omega)d\omega$$
$$-\int_{0}^{\infty}\frac{2(\omega_{rk}^{A})^{2}(\omega_{rk}^{A})^{x}}{((\omega_{rk}^{A})^{2}+\omega^{2})^{2}}\alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg) \tag{4.50}$$

$$=-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\langle k|\,T_{\beta}\,|r\rangle\,\langle r|\,T_{\beta}\,|k\rangle\Bigg(\int_{0}^{\infty}\frac{(\omega_{rk}^{A})^{x}\Big[\omega^{2}-(\omega_{rk}^{A})^{2}\Big]}{((\omega_{rk}^{A})^{2}+\omega^{2})^{2}}\alpha_{\beta\beta}^{j}(i\omega)d\omega\Bigg) \tag{4.51}$$

$(\omega_{rk}^{A})^{x}$ is not dependent on $\omega$ and hence can be separated out of the integral.

$$= -\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\langle k|\,T_\beta\,|r\rangle\,\langle r|\,T_\beta\,|k\rangle\,(\omega_{rk}^{A})^{x}\left(\int_0^\infty \frac{\left[\omega^2-(\omega_{rk}^{A})^2\right]}{((\omega_{rk}^{A})^2+\omega^2)^2}\alpha^{j}_{\beta\beta}(i\omega)d\omega\right) \tag{4.52}$$

From eq. 4.17, we know that $(\omega_{rk}^{A})^{x} = (\epsilon_r - \epsilon_k)^{x} = (\epsilon_r^{x} - \epsilon_k^{x}) = (Q_{rr}^{x} - Q_{kk}^{x})$. Hence, the above equation can be rewritten as:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}T_\beta^{kr}T_\beta^{rk}(Q_{rr}^{x}-Q_{kk}^{x})\left(\int_0^\infty \frac{\left[\omega^2-(\omega_{rk}^{A})^2\right]}{((\omega_{rk}^{A})^2+\omega^2)^2}\alpha^{j}_{\beta\beta}(i\omega)d\omega\right) \tag{4.53}$$

Using eq. 4.17 - 4.20, we can simplify the above expression with known terms:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}T_\beta^{kr}T_\beta^{rk}\left(\left(B_{rr}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{rrbl}\right)-\left(B_{kk}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{kkbl}\right)\right)$$

$$\left(\int_0^\infty \frac{\left[\omega^2-(\omega_{rk}^{A})^2\right]}{((\omega_{rk}^{A})^2+\omega^2)^2}\alpha^{j}_{\beta\beta}(i\omega)d\omega\right) \tag{4.54}$$

Collecting all the three parts - A, B and C, we get:

$$parts(A+B+C) = E_{disp}^{x} =$$

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\Bigg[\left(\sum_{i>k}^{occ}\left(B_{ik}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{ikbl}\right)T_\beta^{ir}T_\beta^{rk}J_{rk-ri}\right)$$

$$+\left(\sum_{i>k}^{occ}\left(B_{ki}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{kibl}\right)T_\beta^{kr}T_\beta^{ri}J_{rk-ri}-\sum_{i}^{occ}S_{ik}^{(x)}T_\beta^{ir}T_\beta^{rk}I_{rk}\right)$$

$$+\left(\sum_{c>r}^{vir}\left(B_{cr}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{crbl}\right)T_\beta^{kc}T_\beta^{rk}J_{rk-ck}\right)$$

$$+\left(\sum_{c>r}^{vir}\left(B_{rc}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{rcbl}\right)T_\beta^{kr}T_\beta^{ck}J_{rk-ck}-\sum_{c}^{vir}S_{cr}^{(x)}T_\beta^{kc}T_\beta^{rk}I_{rk}\right)$$

$$+\sum_{c}^{vir}U_{ck}^{x}T_\beta^{cr}+\sum_{i}^{occ}U_{ir}^{x}T_\beta^{ki}+T_\beta^{kr}T_\beta^{rk}\left(\left(B_{rr}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{rrbl}\right)-\left(B_{kk}^{x}+\sum_{b}^{vir}\sum_{l}^{occ}U_{bl}^{x}A_{kkbl}\right)\right)$$

$$\left(\int_0^\infty \frac{\left[\omega^2 - (\omega_{rk}^A)^2\right]}{((\omega_{rk}^A)^2+\omega^2)^2}\alpha_{\beta\beta}^j(i\omega)d\omega\right)\Bigg] \tag{4.55}$$

The off-diagonal terms highlighted in red above can be converted into a more desirable format by changing the summation indices $c, r \to b$ and $i, k \to l$:

$$\sum_c^{vir} U_{ck}^x T_\beta^{cr} + \sum_i^{occ} U_{ir}^x T_\beta^{ki} = \sum_b^{vir} U_{bl}^x T_\beta^{br} + \sum_l^{occ} U_{lb}^x T_\beta^{kl} \tag{4.56}$$

Further, we can utilize the expression 4.15 once again to convert terms that are obtainable by solving the coupled HF equations:

$$\sum_b^{vir} U_{bl}^x T_\beta^{br} + \sum_l^{occ} U_{lb}^x T_\beta^{kl} = \sum_b^{vir} U_{bl}^x T_\beta^{br} - \sum_l^{occ}\left(U_{bl}^x T_\beta^{kl} + S_{lb}^{(x)} T_\beta^{kl}\right) \tag{4.57}$$

Inserting the newly obtained terms from 4.57 in 4.55, we get:

$$\begin{aligned}
&E_{disp}^x = \\
&-\frac{1}{\pi}\sum_j\sum_k\sum_r\sum_\beta\Bigg[\left(\sum_{i>k}^{occ}\left(B_{ik}^x + \sum_b^{vir}\sum_l^{occ} U_{bl}^x A_{ikbl}\right)T_\beta^{ir}T_\beta^{rk}J_{rk-ri}\right) \\
&+\left(\sum_{i>k}^{occ}\left(B_{ki}^x + \sum_b^{vir}\sum_l^{occ} U_{bl}^x A_{kibl}\right)T_\beta^{kr}T_\beta^{ri}J_{rk-ri} - \sum_i^{occ} S_{ik}^{(x)}T_\beta^{ir}T_\beta^{rk}I_{rk}\right) \\
&+\left(\sum_{c>r}^{vir}\left(B_{cr}^x + \sum_b^{vir}\sum_l^{occ} U_{bl}^x A_{crbl}\right)T_\beta^{kc}T_\beta^{rk}J_{rk-ck}\right) \\
&+\left(\sum_{c>r}^{vir}\left(B_{rc}^x + \sum_b^{vir}\sum_l^{occ} U_{bl}^x A_{rcbl}\right)T_\beta^{kr}T_\beta^{ck}J_{rk-ck} - \sum_c^{vir} S_{cr}^{(x)}T_\beta^{kc}T_\beta^{rk}I_{rk}\right) \\
&+\sum_b^{vir} U_{bl}^x T_\beta^{br} - \sum_l^{occ}\left(U_{bl}^x T_\beta^{kl} + S_{lb}^{(x)}T_\beta^{kl}\right) + T_\beta^{kr}T_\beta^{rk}\Bigg(\left(B_{rr}^x + \sum_b^{vir}\sum_l^{occ} U_{bl}^x A_{rrbl}\right) \\
&-\left(B_{kk}^x + \sum_b^{vir}\sum_l^{occ} U_{bl}^x A_{kkbl}\right)\Bigg)\left(\int_0^\infty \frac{\left[\omega^2 - (\omega_{rk}^A)^2\right]}{((\omega_{rk}^A)^2+\omega^2)^2}\alpha_{\beta\beta}^j(i\omega)d\omega\right)\Bigg]
\end{aligned} \tag{4.58}$$

For simplicity, the integral highlighted in red can be written as $Q_{rk}$. Rewriting the above equation in factors of $U_{bl}$, $B^x$ and $S^{(x)}$ gives us:

$$-\frac{1}{\pi}\sum_{j}\sum_{k}\sum_{r}\sum_{\beta}\Bigg[\Bigg(\sum_{i>k}^{occ}\Bigg((B^x_{ik}+B^x_{ki})T^{ir}_{\beta}T^{rk}_{\beta}J_{rk-ri}+(B^x_{cr}+B^x_{rc})T^{kc}_{\beta}T^{rk}_{\beta}J_{rk-ck}\Bigg)\Bigg)$$

$$+(B^x_{rr}-B^x_{kk})T^{kr}_{\beta}T^{rk}_{\beta}Q_{rk}-T^{rk}_{\beta}I_{rk}\Bigg(\sum_{i}^{occ}S^{(x)}_{ik}T^{ir}_{\beta}+\sum_{c}^{vir}S^{(x)}_{cr}T^{kc}_{\beta}\Bigg)-\sum_{i}^{occ}S^{(x)}_{lb}T^{kl}_{\beta}$$

$$+\sum_{b}^{vir}\sum_{l}^{occ}U^x_{bl}\Bigg[\sum_{i>k}^{occ}T^{rk}_{\beta}T^{ir}_{\beta}J_{rk-ri}(A_{ikbl}+A_{kibl})+\sum_{c>r}^{vir}T^{rk}_{\beta}T^{ck}_{\beta}J_{rk-ck}(A_{crbl}+A_{rcbl}+\sum_{b}^{vir}T^{br}_{\beta}$$

$$-\sum_{b}^{vir}T^{kl}_{\beta}+T^{kr}_{\beta}T^{rk}_{\beta}Q_{rk}(A_{rrbl}+A_{kkbl})\Bigg] \tag{4.59}$$

which is the final expression for the QM/EFP dispersion gradient term. While $B^x$ and $S^{(x)}$ terms are computed on the fly in a gradient simulation, the $U^x_{bl}$ term must be solved for using CPHF equations [166]. The Lagrangian term (highlighted in red above) is solved in the following way:

$$\sum_{ai}U^x_{ai}L_{ai}\equiv \boldsymbol{L^T U}^x \tag{4.60a}$$

$$\boldsymbol{A'U}^x=\boldsymbol{B}^x \tag{4.60b}$$

$$\boldsymbol{U}^x=(\boldsymbol{A'})^{-1}\boldsymbol{B}^x \tag{4.60c}$$

$$\boldsymbol{L^T U}^x=\boldsymbol{L^T}(\boldsymbol{A'})^{-1}\boldsymbol{B}^x=\boldsymbol{Z^T B}^x \tag{4.60d}$$

$$\boldsymbol{Z^T}=\boldsymbol{L^T}(\boldsymbol{A'})^{-1} \tag{4.60e}$$

$$(\boldsymbol{A'})^{\boldsymbol{T}}\boldsymbol{Z}=\boldsymbol{L} \tag{4.60f}$$

$$\sum_{ai}U^x_{ai}L_{ai}=\sum_{ai}B^x_{ai}Z_{ai} \tag{4.60g}$$

$$\sum_{ai}B^x_{ai}Z_{ai}=\sum_{ai}B^x_{ai}P^{(disp)}_{ai} \tag{4.60h}$$

where $P^{(disp)}_{ai}$ is the correction to the density matrix. The final Lagrangian equation is given by:

$$\sum_{b}^{vir}\sum_{j}^{occ}\Big[A_{aibj}+\delta_{ab}\delta_{ij}(\epsilon_b-\epsilon_j)\Big]P_{ai}^{(disp)}=-L_{ai} \tag{4.61}$$

## 4.3 Conclusions

Working equations corresponding to the diagonal QM/EFP dispersion gradient term have been developed. Calculation of the gradient term involves evaluation of field integrals, gradients corresponding to one-electron Hamiltonian terms, overlap matrix and evaluation of Coupled Hartree-Fock equations at every step. Once the codes corresponding to these equations are implemented, dynamics simulations at full embedding level with QM/EFP can be performed.

# REFERENCES

# REFERENCES


[1] T. P. Chamberlain and D. M. Hunten, *Theory of planetary atmospheres: an introduction to their physics and chemistry*. Academic Press, 1990, vol. 36.

[2] B. J. Finlayson-Pitts and J. N. Pitts Jr, *Chemistry of the upper and lower atmosphere: theory, experiments, and applications*. Elsevier, 1999.

[3] E. S. Kwok and R. Atkinson, "Estimation of hydroxyl radical reaction rate constants for gas-phase organic compounds using a structure-reactivity relationship: an update," *Atmospheric Environment*, vol. 29, no. 14, pp. 1685–1696, 1995.

[4] R. Atkinson, "Kinetics and mechanisms of the gas-phase reactions of the hydroxyl radical with organic compounds under atmospheric conditions," *Chemical Reviews*, vol. 86, no. 1, pp. 69–201, 1986.

[5] X. Liu, H. E. Jeffries, and K. G. Sexton, "Hydroxyl radical and ozone initiated photochemical reactions of 1, 3-butadiene," *Atmospheric Environment*, vol. 33, no. 18, pp. 3005–3022, 1999.

[6] L. M. Dorfman and G. E. Adams, "Reactivity of the hydroxyl radical in aqueous solutions." National Standard Reference Data System, Tech. Rep., 1973.

[7] B. Halliwell and J. M. Gutteridge, *Free radicals in biology and medicine*. Oxford University Press, USA, 2015.

[8] R. Xiao, I. Zammit, Z. Wei, W.-P. Hu, M. MacLeod, and R. Spinney, "Kinetics and mechanism of the oxidation of cyclic methylsiloxanes by hydroxyl radical in the gas phase: an experimental and theoretical study," *Environmental science & technology*, vol. 49, no. 22, pp. 13 322–13 330, 2015.

[9] C. T. Rettner, D. J. Auerbach, J. C. Tully, and A. W. Kleyn, "Chemical dynamics at the gassurface interface," *The Journal of Physical Chemistry*, vol. 100, no. 31, pp. 13 021–13 033, 1996. [Online]. Available: https://doi.org/10.1021/jp9536007

[10] R. B. Gerber, "Molecular scattering from surfaces: theoretical methods and results," *Chemical Reviews*, vol. 87, no. 1, pp. 29–79, 1987. [Online]. Available: https://doi.org/10.1021/cr00077a003

[11] J. Barker and D. Auerbach, "Gassurface interactions and dynamics; thermal energy atomic and molecular beam studies," *Surface Science Reports*, vol. 4, no. 1, pp. 1 – 99, 1984. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0167572984900050

[12] M. Roeselov, P. Jungwirth, D. J. Tobias, and R. B. Gerber, "Impact, trapping, and accommodation of hydroxyl radical and ozone at aqueous salt aerosol surfaces. a molecular dynamics study," *The Journal of Physical Chemistry B*, vol. 107, no. 46, pp. 12 690–12 699, 2003. [Online]. Available: https://doi.org/10.1021/jp030592i

[13] P. Cabral do Couto, R. Guedes, B. Costa Cabral, and J. Martinho Simoes, "The hydration of the oh radical: Microsolvation modeling and statistical mechanics simulation," *The Journal of chemical physics*, vol. 119, no. 14, pp. 7344–7355, 2003.

[14] S. Hamad, S. Lago, and J. Mejias, "A computational study of the hydration of the oh radical," *The Journal of Physical Chemistry A*, vol. 106, no. 39, pp. 9104–9113, 2002.

[15] G. Czapski and B. H. Bielski, "Absorption spectra of the. oh and o. -radicals in aqueous solutions," *Radiation Physics and Chemistry*, vol. 41, no. 3, pp. 503–505, 1993.

[16] D. M. Chipman, "Absorption spectrum of oh radical in water," *The Journal of Physical Chemistry A*, vol. 112, no. 51, pp. 13 372–13 381, 2008.

[17] J. K. Thomas, J. Rabani, M. S. Matheson, E. J. Hart, and S. Gordon, "Absorption spectrum of the hydroxyl radical1," *The Journal of Physical Chemistry*, vol. 70, no. 7, pp. 2409–2410, 1966. [Online]. Available: https://doi.org/10.1021/j100879a503

[18] G. J. Hoffman, P. K. Gurunathan, J. S. Francisco, and L. V. Slipchenko, "Excited states of oh-(h2o) n clusters for n= 1–4: An ab initio study," *The Journal of chemical physics*, vol. 141, no. 10, p. 104315, 2014.

[19] D. Kosenkov and L. V. Slipchenko, "Solvent effects on the electronic transitions of p-nitroaniline: A qm/efp study," *The Journal of Physical Chemistry A*, vol. 115, no. 4, pp. 392–401, 2010.

[20] S. Chibani, D. Jacquemin, and A. D. Laurent, "Modelling solvent effects on the absorption and emission spectra of constrained cyanines with both implicit and explicit qm/efp models," *Computational and Theoretical Chemistry*, vol. 1040, pp. 321–327, 2014.

[21] J. W. Caldwell and P. A. Kollman, "Structure and properties of neat liquids using nonadditive molecular dynamics: water, methanol, and n-methylacetamide," *The Journal of Physical Chemistry*, vol. 99, no. 16, pp. 6208–6219, 1995.

[22] P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, and D. Cohen, "An effective fragment method for modeling solvent effects in quantum mechanical calculations," *J. Chem. Phys.*, vol. 105, p. 1968, 1996.

[23] M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, "The effective fragment potential method: a qm-based mm approach to modeling environmental effects in chemistry," *J. Phys. Chem. A*, vol. 105, p. 293, 2001.

[24] D. Ghosh, D. Kosenkov, V. Vanovschi, C. Williams, J. Herbert, M. S. Gordon, M. Schmidt, L. V. Slipchenko, and A. I. Krylov, "Non-covalent interactions in extended systems described by the effective fragment potential method: theory and application to nucleobase oligomers," *J. Phys. Chem. A*, vol. 114, p. 12739, 2010.

[25] M. S. Gordon, Q. A. Smith, P. Xu, and L. V. Slipchenko, "Accurate first principles model potentials for intermolecular interactions," *Annu. Rev. Phys. Chem.*, vol. 64, p. 553, 2013.

[26] M. S. Gordon, L. Slipchenko, H. Li, and J. H. Jensen, "The effective fragment potential: a general method for predicting intermolecular interactions," *Annual reports in computational chemistry*, vol. 3, pp. 177–193, 2007.

[27] H. Lin and D. G. Truhlar, "Qm/mm: what have we learned, where are we, and where do we go from here?" *Theoretical Chemistry Accounts*, vol. 117, no. 2, p. 185, 2007.

[28] N. Dubinets and L. V. Slipchenko, "Effective fragment potential method for h-bonding: How to obtain parameters for nonrigid fragments," *The Journal of Physical Chemistry A*, vol. 121, no. 28, pp. 5301–5312, 2017.

[29] D. Sinha, S. Mukhopadhyay, R. Chaudhuri, and D. Mukherjee, "The eigenvalue-independent partitioning technique in fock space: An alternative route to open-shell coupled-cluster theory for incomplete model spaces," *Chemical physics letters*, vol. 154, no. 6, pp. 544–549, 1989.

[30] J. F. Stanton and J. Gauss, "Analytic energy derivatives for ionized states described by the equation-of-motion coupled cluster method," *The Journal of chemical physics*, vol. 101, no. 10, pp. 8938–8944, 1994.

[31] J. F. Stanton and R. J. Bartlett, "The equation of motion coupled-cluster method. a systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties," *The Journal of chemical physics*, vol. 98, no. 9, pp. 7029–7039, 1993.

[32] Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng *et al.*, "Advances in molecular quantum chemistry contained in the q-chem 4 program package," *Molecular Physics*, vol. 113, no. 2, pp. 184–215, 2015.

[33] T. Koopmans, "Über die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms," *Physica*, vol. 1, no. 1-6, pp. 104–113, 1934.

[34] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su *et al.*, "General atomic and molecular electronic structure system," *Journal of computational chemistry*, vol. 14, no. 11, pp. 1347–1363, 1993.

[35] M. S. Gordon and M. W. Schmidt, "Advances in electronic structure theory: Gamess a decade later," in *Theory and applications of computational chemistry*. Elsevier, 2005, pp. 1167–1189.

[36] H. Okabe *et al.*, *Photochemistry of small molecules*. Wiley New York, 1978, vol. 431.

[37] D. M. Chipman, "Hemibonding between hydroxyl radical and water," *The Journal of Physical Chemistry A*, vol. 115, no. 7, pp. 1161–1171, 2011.

[38] M.-K. Tsai, K. Kowalski, M. Valiev, and M. Dupuis, "Signature oh absorption spectrum from cluster models of solvation: A solvent-to-solute charge transfer state," *The Journal of Physical Chemistry A*, vol. 111, no. 42, pp. 10 478–10 482, 2007.

[39] J. VandeVondele and M. Sprik, "A molecular dynamics study of the hydroxyl radical in solution applying self-interaction-corrected density functional methods," *Physical Chemistry Chemical Physics*, vol. 7, no. 7, pp. 1363–1367, 2005.

[40] C. Reichardt and T. Welton, *Solvents and solvent effects in organic chemistry*. John Wiley & Sons, 2011.

[41] A. DeFusco, N. Minezawa, L. V. Slipchenko, F. Zahariev, and M. S. Gordon, "Modeling solvent effects on electronic excited states," *The Journal of Physical Chemistry Letters*, vol. 2, no. 17, pp. 2184–2192, 2011.

[42] D. Ghosh, O. Isayev, L. V. Slipchenko, and A. I. Krylov, "Effect of solvation on the vertical ionization energy of thymine: from microhydration to bulk," *The Journal of Physical Chemistry A*, vol. 115, no. 23, pp. 6028–6038, 2011.

[43] K. B. Bravaya, M. G. Khrenova, B. L. Grigorenko, A. V. Nemukhin, and A. I. Krylov, "Effect of protein environment on electronically excited and ionized states of the green fluorescent protein chromophore," *The Journal of Physical Chemistry B*, vol. 115, no. 25, pp. 8296–8303, 2011.

[44] D. Ghosh, A. Roy, R. Seidel, B. Winter, S. Bradforth, and A. I. Krylov, "First-principle protocol for calculating ionization energies and redox potentials of solvated molecules and ions: Theory and application to aqueous phenol and phenolate," *The Journal of Physical Chemistry B*, vol. 116, no. 24, pp. 7269–7280, 2012.

[45] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular electronic-structure theory*. John Wiley & Sons, 2014.

[46] R. J. Bartlett, "Coupled-cluster theory and its equation-of-motion extensions," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 1, pp. 126–138, 2012.

[47] A. I. Krylov, "Spin-flip equation-of-motion coupled-cluster electronic structure method for a description of excited states, bond breaking, diradicals, and triradicals," *Accounts of chemical research*, vol. 39, no. 2, pp. 83–91, 2006.

[48] E. Epifanovsky, M. Wormit, T. Kuś, A. Landau, D. Zuev, K. Khistyaev, P. Manohar, I. Kaliman, A. Dreuw, and A. I. Krylov, "New implementation of high-level correlated methods using a general block tensor library for high-performance electronic structure calculations," *Journal of computational chemistry*, vol. 34, no. 26, pp. 2293–2309, 2013.

[49] E. Epifanovsky, D. Zuev, X. Feng, K. Khistyaev, Y. Shao, and A. I. Krylov, "General implementation of the resolution-of-the-identity and cholesky representations of electron repulsion integrals within coupled-cluster and equation-of-motion methods: Theory and benchmarks," *The Journal of chemical physics*, vol. 139, no. 13, p. 134105, 2013.

[50] S. Hirata and M. Head-Gordon, "Time-dependent density functional theory within the tamm–dancoff approximation," *Chemical Physics Letters*, vol. 314, no. 3-4, pp. 291–299, 1999.

[51] Y. M. Rhee and M. Head-Gordon, "Scaled second-order perturbation corrections to configuration interaction singles: efficient and reliable excitation energy methods," *The Journal of Physical Chemistry A*, vol. 111, no. 24, pp. 5314–5326, 2007.

[52] A. Dreuw and M. Head-Gordon, "Single-reference ab initio methods for the calculation of excited states of large molecules," *Chemical reviews*, vol. 105, no. 11, pp. 4009–4037, 2005.

[53] J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," *Chemical reviews*, vol. 105, no. 8, pp. 2999–3094, 2005.

[54] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, "Perspective on foundations of solvation modeling: The electrostatic contribution to the free energy of solvation," *Journal of Chemical Theory and Computation*, vol. 4, no. 6, pp. 877–887, 2008.

[55] C. J. Cramer and D. G. Truhlar, "Reply to comment on a universal approach to solvation modeling," *Accounts of Chemical Research*, vol. 42, no. 4, pp. 493–497, 2009.

[56] J.-M. Mewes, Z.-Q. You, M. Wormit, T. Kriesche, J. M. Herbert, and A. Dreuw, "Experimental benchmark data and systematic evaluation of two a posteriori, polarizable-continuum corrections for vertical excitation energies in solution," *The Journal of Physical Chemistry A*, vol. 119, no. 21, pp. 5446–5464, 2015.

[57] A. V. Marenich, J. Ho, M. L. Coote, C. J. Cramer, and D. G. Truhlar, "Computational electrochemistry: prediction of liquid-phase reduction potentials," *Physical Chemistry Chemical Physics*, vol. 16, no. 29, pp. 15 068–15 106, 2014.

[58] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," *Journal of molecular biology*, vol. 103, no. 2, pp. 227–249, 1976.

[59] U. C. Singh and P. A. Kollman, "A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the ch3cl+ cl- exchange reaction and gas phase protonation of polyethers," *Journal of Computational Chemistry*, vol. 7, no. 6, pp. 718–730, 1986.

[60] G. Monard and K. M. Merz, "Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems," *Accounts of chemical research*, vol. 32, no. 10, pp. 904–911, 1999.

[61] H. Lin and D. G. Truhlar, "Qm/mm: what have we learned, where are we, and where do we go from here?" *Theoretical Chemistry Accounts*, vol. 117, no. 2, p. 185, 2007.

[62] M. Sulpizi, S. Raugei, J. VandeVondele, P. Carloni, and M. Sprik, "Calculation of redox properties: Understanding short-and long-range effects in rubredoxin," *The Journal of Physical Chemistry B*, vol. 111, no. 15, pp. 3969–3976, 2007.

[63] P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, and D. Cohen, "An effective fragment method for modeling solvent effects in quantum mechanical calculations," *The Journal of chemical physics*, vol. 105, no. 5, pp. 1968–1986, 1996.

[64] M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, "The effective fragment potential method: A qm-based mm approach to modeling environmental effects in chemistry," *The Journal of Physical Chemistry A*, vol. 105, no. 2, pp. 293–307, 2001.

[65] M. Gordon, L. Slipchenko, H. Li, J. Jensen, D. Spellmeyer, and R. Wheeler, "Annual reports in computational chemistry," *Annual Reports in Computational Chemistry*, vol. 3, p. 177, 2007.

[66] D. Ghosh, D. Kosenkov, V. Vanovschi, C. F. Williams, J. M. Herbert, M. S. Gordon, M. W. Schmidt, L. V. Slipchenko, and A. I. Krylov, "Noncovalent interactions in extended systems described by the effective fragment potential method: Theory and application to nucleobase oligomers," *The Journal of Physical Chemistry A*, vol. 114, no. 48, pp. 12 739–12 754, 2010.

[67] D. Ghosh, D. Kosenkov, V. Vanovschi, J. Flick, I. Kaliman, Y. Shao, A. T. Gilbert, A. I. Krylov, and L. V. Slipchenko, "Effective fragment potential method in q-chem: A guide for users and developers," *Journal of computational chemistry*, vol. 34, no. 12, pp. 1060–1070, 2013.

[68] M. S. Gordon, Q. A. Smith, P. Xu, and L. V. Slipchenko, "Accurate first principles model potentials for intermolecular interactions," *Annual review of physical chemistry*, vol. 64, pp. 553–578, 2013.

[69] I. A. Kaliman and L. V. Slipchenko, "Libefp: A new parallel implementation of the effective fragment potential method as a portable software library," *Journal of computational chemistry*, vol. 34, no. 26, pp. 2284–2292, 2013.

[70] ——, "Hybrid mpi/openmp parallelization of the effective fragment potential method in the libefp software library," *Journal of computational chemistry*, vol. 36, no. 2, pp. 129–135, 2015.

[71] P. Arora, L. V. Slipchenko, S. P. Webb, A. DeFusco, and M. S. Gordon, "Solvent-induced frequency shifts: configuration interaction singles combined with the effective fragment potential method," *The Journal of Physical Chemistry A*, vol. 114, no. 25, pp. 6742–6750, 2010.

[72] L. V. Slipchenko, "Solvation of the excited states of chromophores in polarizable environment: orbital relaxation versus polarization," *The Journal of Physical Chemistry A*, vol. 114, no. 33, pp. 8824–8830, 2010.

[73] D. Kosenkov and L. V. Slipchenko, "Solvent effects on the electronic transitions of p-nitroaniline: A qm/efp study," *The Journal of Physical Chemistry A*, vol. 115, no. 4, pp. 392–401, 2010.

[74] J. M. Olsen, K. Aidas, and J. Kongsted, "Excited states in solution through polarizable embedding," *Journal of Chemical Theory and Computation*, vol. 6, no. 12, pp. 3721–3734, 2010.

[75] K. Sneskov, T. Schwabe, O. Christiansen, and J. Kongsted, "Scrutinizing the effects of polarization in qm/mm excited state calculations," *Physical Chemistry Chemical Physics*, vol. 13, no. 41, pp. 18 551–18 560, 2011.

[76] K. Sneskov, T. Schwabe, J. Kongsted, and O. Christiansen, "The polarizable embedding coupled cluster method," *The Journal of Chemical Physics*, vol. 134, no. 10, p. 03B608, 2011.

[77] J. M. H. Olsen, C. Steinmann, K. Ruud, and J. Kongsted, "Polarizable density embedding: A new qm/qm/mm-based computational strategy," *The Journal of Physical Chemistry A*, vol. 119, no. 21, pp. 5344–5355, 2015.

[78] R. M. Richard and J. M. Herbert, "A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory," *The Journal of Chemical Physics*, vol. 137, no. 6, p. 064113, 2012.

[79] R. M. Richard, K. U. Lao, and J. M. Herbert, "Aiming for benchmark accuracy with the many-body expansion," *Accounts of chemical research*, vol. 47, no. 9, pp. 2828–2836, 2014.

[80] K. U. Lao and J. M. Herbert, "Accurate and efficient quantum chemistry calculations for noncovalent interactions in many-body systems: The xsapt family of methods," *The Journal of Physical Chemistry A*, vol. 119, no. 2, pp. 235–252, 2014.

[81] S. Wen, K. Nanda, Y. Huang, and G. J. Beran, "Practical quantum mechanics-based fragment methods for predicting molecular crystal properties," *Physical Chemistry Chemical Physics*, vol. 14, no. 21, pp. 7578–7590, 2012.

[82] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, "Fragmentation methods: a route to accurate calculations on large systems," *Chemical reviews*, vol. 112, no. 1, pp. 632–672, 2011.

[83] I. Antes and W. Thiel, "Adjusted connection atoms for combined quantum mechanical and molecular mechanical methods," *The Journal of Physical Chemistry A*, vol. 103, no. 46, pp. 9290–9295, 1999.

[84] Y. Shao and J. Kong, "Yinyang atom: a simple combined ab initio quantum mechanical molecular mechanical model," *The Journal of Physical Chemistry A*, vol. 111, no. 18, pp. 3661–3671, 2007.

[85] X. Assfeld and J.-L. Rivail, "Quantum chemical computations on parts of large molecules: the ab initio local self consistent field method," *Chemical physics letters*, vol. 263, no. 1-2, pp. 100–106, 1996.

[86] R. B. Murphy, D. M. Philipp, and R. A. Friesner, "A mixed quantum mechanics/molecular mechanics (qm/mm) method for large-scale modeling of chemistry in protein environments," *Journal of Computational Chemistry*, vol. 21, no. 16, pp. 1442–1457, 2000.

[87] N. Sahu and S. R. Gadre, "Molecular tailoring approach: a route for ab initio treatment of large clusters," *Accounts of chemical research*, vol. 47, no. 9, pp. 2739–2747, 2014.

[88] D. W. Zhang and J. Zhang, "Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy," *The Journal of chemical physics*, vol. 119, no. 7, pp. 3599–3605, 2003.

[89] D. W. Zhang, X. Chen, and J. Z. Zhang, "Molecular caps for full quantum mechanical computation of peptide–water interaction energy," *Journal of computational chemistry*, vol. 24, no. 15, pp. 1846–1852, 2003.

[90] P. Soderhjelm and U. Ryde, "How accurate can a force field become? a polarizable multipole model combined with fragment-wise quantum-mechanical calculations," *The Journal of Physical Chemistry A*, vol. 113, no. 3, pp. 617–627, 2008.

[91] B. L. Grigorenko, A. V. Nemukhin, I. A. Topol, and S. K. Burt, "Modeling of biomolecular systems with the quantum mechanical and molecular mechanical method based on the effective fragment potential technique: Proposal of flexible fragments," *The Journal of Physical Chemistry A*, vol. 106, no. 44, pp. 10 663–10 672, 2002.

[92] R. M. Minikis, V. Kairys, and J. H. Jensen, "Accurate intraprotein electrostatics derived from first principles: an effective fragment potential method study of the proton affinities of lysine 55 and tyrosine 20 in turkey ovomucoid third domain," *The Journal of Physical Chemistry A*, vol. 105, no. 15, pp. 3829–3837, 2001.

[93] P. A. Molina, H. Li, and J. H. Jensen, "Intraprotein electrostatics derived from first principles: Divide-and-conquer approaches for qm/mm calculations," *Journal of computational chemistry*, vol. 24, no. 16, pp. 1971–1979, 2003.

[94] C. Steinmann, D. G. Fedorov, and J. H. Jensen, "The effective fragment molecular orbital method for fragments connected by covalent bonds," *PLoS One*, vol. 7, no. 7, p. e41117, 2012.

[95] L. V. Slipchenko and M. S. Gordon, "Damping functions in the effective fragment potential method," *Molecular Physics*, vol. 107, no. 8-12, pp. 999–1016, 2009.

[96] ——, "Electrostatic energy in the effective fragment potential method: Theory and application to benzene dimer," *Journal of computational chemistry*, vol. 28, no. 1, pp. 276–291, 2007.

[97] Other schemes exists such as the IUPAC definition with a the $C_\alpha$-C(=O) bond cut or a definition splitting each amino-acid into the backbone and the side chain moiety. This fragmentation leads to asymmetrical fragments, which are not convenient for computing DMA distribution because they result in the delocalized multipoles. However, the cut is between atoms of same chemical nature. Natural fragmentation solves the problem of asymmetry, but an extremely polar bond is cut.

[98] Other methods exist in the literature to guarantee a net integer charge, the Expand-Remove and Scale (ERS) is similar to our method up to the removal step. The monopoles located on all caps are then scaled to reflect the net integer charge of the entire protein. Two other methods developed by the same group are the Remove and Expand Method (RE) and the Expand-Collect-and-Correct (ECC) [93]. Comparing these four methods will be done in future work.

[99] Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng *et al.*, "Advances in molecular quantum chemistry contained in the q-chem 4 program package," *Molecular Physics*, vol. 113, no. 2, pp. 184–215, 2015.

[100] A. I. Krylov and P. M. Gill, "Q-chem: an engine for innovation," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 3, no. 3, pp. 317–326, 2013.

[101] N. Foloppe and A. D. MacKerell, Jr, "All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data," *Journal of computational chemistry*, vol. 21, no. 2, pp. 86–104, 2000.

[102] N. Reuter, H. Lin, and W. Thiel, "Green fluorescent proteins: Empirical force field for the neutral and deprotonated forms of the chromophore. molecular dynamics simulations of the wild type and s65t mutant," *The Journal of Physical Chemistry B*, vol. 106, no. 24, pp. 6310–6321, 2002.

[103] A. M. Bogdanov, A. Acharya, A. V. Titelmayer, A. V. Mamontova, K. B. Bravaya, A. B. Kolomeisky, K. A. Lukyanov, and A. I. Krylov, "Turning on and off photoinduced electron transfer in fluorescent proteins by $\pi$-stacking, halide binding, and tyr145 mutations," *Journal of the American Chemical Society*, vol. 138, no. 14, pp. 4807–4817, 2016.

[104] S. Faraji and A. I. Krylov, "On the nature of an extended stokes shift in the mplum fluorescent protein," *The Journal of Physical Chemistry B*, vol. 119, no. 41, pp. 13 052–13 062, 2015.

[105] L. Liu, W. A. Baase, M. M. Michael, and B. W. Matthews, "Use of stabilizing mutations to engineer a charged group within a ligand-binding hydrophobic cavity in t4 lysozyme," *Biochemistry*, vol. 48, no. 37, pp. 8842–8851, 2009.

[106] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su *et al.*, "General atomic and molecular electronic structure system," *Journal of computational chemistry*, vol. 14, no. 11, pp. 1347–1363, 1993.

[107] K. B. Bravaya, B. L. Grigorenko, A. V. Nemukhin, and A. I. Krylov, "Quantum chemistry behind bioimaging: insights from ab initio studies of fluorescent proteins and their chromophores," *Accounts of chemical research*, vol. 45, no. 2, pp. 265–275, 2011.

[108] B. L. Grigorenko, A. V. Nemukhin, D. I. Morozov, I. V. Polyakov, K. B. Bravaya, and A. I. Krylov, "Toward molecular-level characterization of photoinduced decarboxylation of the green fluorescent protein: Accessibility of the charge-transfer states," *Journal of chemical theory and computation*, vol. 8, no. 6, pp. 1912–1920, 2012.

[109] R. B. Vegh, K. B. Bravaya, D. A. Bloch, A. S. Bommarius, L. M. Tolbert, M. Verkhovsky, A. I. Krylov, and K. M. Solntsev, "Chromophore photoreduction in red fluorescent proteins is responsible for bleaching and phototoxicity," *The Journal of Physical Chemistry B*, vol. 118, no. 17, pp. 4527–4534, 2014.

[110] R. Ballard, J. Jones, E. Sutherland, and C. B. Lee, "Photoelectron spectra of anions in solutions and the ionisation energy of the solvated electron," *Chemical Physics Letters*, vol. 97, pp. 419–421, 1983.

[111] S. Gorelsky, V. Y. Kotov, and A. Lever, "Vertical ionization energies and electron affinities of ions in solution from outer-sphere charge transfer transition energies," *Inorganic chemistry*, vol. 37, no. 18, pp. 4584–4588, 1998.

[112] E. Epifanovsky, I. Polyakov, B. Grigorenko, A. Nemukhin, and A. I. Krylov, "Quantum chemical benchmark studies of the electronic properties of the green fluorescent protein chromophore. 1. electronically excited and ionized states of the anionic chromophore in the gas phase," *Journal of chemical theory and computation*, vol. 5, no. 7, pp. 1895–1906, 2009.

[113] D. Zuev, K. B. Bravaya, T. D. Crawford, R. Lindh, and A. I. Krylov, "Electronic structure of the two isomers of the anionic form of p-coumaric acid chromophore," *The Journal of Chemical Physics*, vol. 134, no. 3, p. 034310, 2011.

[114] K. B. Bravaya, D. Zuev, E. Epifanovsky, and A. I. Krylov, "Complex-scaled equation-of-motion coupled-cluster method with single and double substitutions for autoionizing excited states: Theory, implementation, and examples," *The Journal of chemical physics*, vol. 138, no. 12, p. 124106, 2013.

[115] B. M. Rankin, M. D. Hands, D. S. Wilcox, K. R. Fega, L. V. Slipchenko, and D. Ben-Amotz, "Interactions between halide anions and a molecular hydrophobic interface," *Faraday discussions*, vol. 160, pp. 255–270, 2013.

[116] D. Ghosh, A. Acharya, S. C. Tiwari, and A. I. Krylov, "Toward understanding the redox properties of model chromophores from the green fluorescent protein family: an interplay between conjugation, resonance stabilization, and solvent effects," *The Journal of Physical Chemistry B*, vol. 116, no. 41, pp. 12 398–12 405, 2012.

[117] N. Homeyer and H. Gohlke, "Free energy calculations by the molecular mechanics poisson- boltzmann surface area method," *Molecular Informatics*, vol. 31, no. 2, pp. 114–122, 2012.

[118] N. Huang, C. Kalyanaraman, K. Bernacki, and M. P. Jacobson, "Molecular mechanics methods for predicting protein–ligand binding," *Physical Chemistry Chemical Physics*, vol. 8, no. 44, pp. 5166–5177, 2006.

[119] C. McInnes, "Virtual screening strategies in drug discovery," *Current opinion in chemical biology*, vol. 11, no. 5, pp. 494–502, 2007.

[120] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: methods and applications," *Nature reviews Drug discovery*, vol. 3, no. 11, p. 935, 2004.

[121] Y. Chen and D. Zhi, "Ligand–protein inverse docking and its potential use in the computer search of protein targets of a small molecule," *Proteins: Structure, Function, and Bioinformatics*, vol. 43, no. 2, pp. 217–226, 2001.

[122] W. L. Jorgensen, "The many roles of computation in drug discovery," *Science*, vol. 303, no. 5665, pp. 1813–1818, 2004.

[123] G. R. Marshall, "Computer-aided drug design," *Annual review of pharmacology and toxicology*, vol. 27, no. 1, pp. 193–213, 1987.

[124] A. Wlodawer, W. Minor, Z. Dauter, and M. Jaskolski, "Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures," *The FEBS journal*, vol. 275, no. 1, pp. 1–21, 2008.

[125] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa, "Prediction of drug–target interaction networks from the integration of chemical and genomic spaces," *Bioinformatics*, vol. 24, no. 13, pp. i232–i240, 2008.

[126] K. Raha, M. B. Peters, B. Wang, N. Yu, A. M. Wollacott, L. M. Westerhoff, and K. M. Merz Jr, "The role of quantum mechanics in structure-based drug design," *Drug discovery today*, vol. 12, no. 17-18, pp. 725–731, 2007.

[127] M. De Vivo, "Bridging quantum mechanics and structure-based drug design," *optimization*, vol. 7, p. 8, 2011.

[128] M. W. van der Kamp and A. J. Mulholland, "Combined quantum mechanics/molecular mechanics (qm/mm) methods in computational enzymology," *Biochemistry*, vol. 52, no. 16, pp. 2708–2728, 2013.

[129] M. K. Gilson and H.-X. Zhou, "Calculation of protein-ligand binding affinities," *Annual review of biophysics and biomolecular structure*, vol. 36, 2007.

[130] K. M. Merz Jr, "Limits of free energy computation for protein- ligand interactions," *Journal of chemical theory and computation*, vol. 6, no. 5, pp. 1769–1776, 2010.

[131] N. R. Patel, D. V. Patel, P. R. Murumkar, and M. R. Yadav, "Contemporary developments in the discovery of selective factor xa inhibitors: a review," *European journal of medicinal chemistry*, vol. 121, pp. 671–698, 2016.

[132] D. J. Pinto, J. M. Smallheer, D. L. Cheney, R. M. Knabb, and R. R. Wexler, "Factor xa inhibitors: next-generation antithrombotic agents," *Journal of medicinal chemistry*, vol. 53, no. 17, pp. 6243–6274, 2010.

[133] R. M. Parrish, D. F. Sitkoff, D. L. Cheney, and C. D. Sherrill, "The surprising importance of peptide bond contacts in drug–protein interactions," *Chemistry–A European Journal*, vol. 23, no. 33, pp. 7887–7890, 2017.

[134] R. M. Parrish and C. D. Sherrill, "Spatial assignment of symmetry adapted perturbation theory interaction energy components: The atomic sapt partition," *The Journal of chemical physics*, vol. 141, no. 4, p. 044115, 2014.

[135] R. M. Parrish, T. M. Parker, and C. D. Sherrill, "Chemical assignment of symmetry-adapted perturbation theory interaction energy components: the functional-group sapt partition," *Journal of chemical theory and computation*, vol. 10, no. 10, pp. 4417–4431, 2014.

[136] B. Jeziorski, R. Moszynski, and K. Szalewicz, "Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes," *Chemical Reviews*, vol. 94, no. 7, pp. 1887–1930, 1994.

[137] K. Szalewicz, "Symmetry-adapted perturbation theory of intermolecular forces," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 2, pp. 254–272, 2012.

[138] P. K. Gurunathan, A. Acharya, D. Ghosh, D. Kosenkov, I. Kaliman, Y. Shao, A. I. Krylov, and L. V. Slipchenko, "Extension of the effective fragment potential method to macromolecules," *The Journal of Physical Chemistry B*, vol. 120, no. 27, pp. 6562–6574, 2016.

[139] A. Stone, *The theory of intermolecular forces*. OUP Oxford, 2013.

[140] ——, "Distributed multipole analysis, or how to describe a molecular charge distribution," *Chemical Physics Letters*, vol. 83, no. 2, pp. 233–239, 1981.

[141] L. V. Slipchenko and M. S. Gordon, "Damping functions in the effective fragment potential method," *Molecular Physics*, vol. 107, no. 8-12, pp. 999–1016, 2009.

[142] J. C. Flick, D. Kosenkov, E. G. Hohenstein, C. D. Sherrill, and L. V. Slipchenko, "Accurate prediction of noncovalent interaction energies with the effective fragment potential method: Comparison of energy components to symmetry-adapted perturbation theory for the s22 test set," *Journal of Chemical Theory and Computation*, vol. 8, no. 8, pp. 2835–2843, 2012. [Online]. Available: http://pubs.acs.org/doi/abs/10.1021/ct200673a

[143] Y. Shi, D. Sitkoff, J. Zhang, H. E. Klei, K. Kish, E. C.-K. Liu, K. S. Hartl, S. M. Seiler, M. Chang, C. Huang *et al.*, "Design, structure- activity relationships, x-ray crystal structure, and energetic contributions of a critical p1 pharmacophore: 3-chloroindole-7-yl-based factor xa inhibitors," *Journal of medicinal chemistry*, vol. 51, no. 23, pp. 7541–7551, 2008.

[144] H. Matter, M. Nazaré, S. Güssregen, D. W. Will, H. Schreuder, A. Bauer, M. Urmann, K. Ritter, M. Wagner, and V. Wehner, "Evidence for c cl/c br $\pi$ interactions as an important contribution to protein–ligand binding affinity," *Angewandte Chemie International Edition*, vol. 48, no. 16, pp. 2911–2916, 2009.

[145] H. G. Wallnoefer, T. Fox, K. R. Liedl, and C. S. Tautermann, "Dispersion dominated halogen–$\pi$ interactions: energies and locations of minima," *Physical Chemistry Chemical Physics*, vol. 12, no. 45, pp. 14 941–14 949, 2010.

[146] I. A. Kaliman and L. V. Slipchenko, "Libefp: A new parallel implementation of the effective fragment potential method as a portable software library," *Journal of computational chemistry*, vol. 34, no. 26, pp. 2284–2292, 2013.

[147] ——, "Hybrid mpi/openmp parallelization of the effective fragment potential method in the libefp software library," *Journal of computational chemistry*, vol. 36, no. 2, pp. 129–135, 2015.

[148] S. Cabani, P. Gianni, V. Mollica, and L. Lepori, "Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution," *Journal of Solution Chemistry*, vol. 10, no. 8, pp. 563–595, 1981.

[149] K. A. Sharp and B. Honig, "Electrostatic interactions in macromolecules: theory and applications," *Annual review of biophysics and biophysical chemistry*, vol. 19, no. 1, pp. 301–332, 1990.

[150] V. Mohan, M. Davis, J. McCammon, and B. M. Pettitt, "Continuum model calculations of solvation free energies: accurate evaluation of electrostatic contributions," *The Journal of Physical Chemistry*, vol. 96, no. 15, pp. 6428–6431, 1992.

[151] J. N. Israelachvili, *Intermolecular and surface forces*. Academic press, 2011.

[152] A. Nicholls, K. A. Sharp, and B. Honig, "Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons," *Proteins: Structure, Function, and Bioinformatics*, vol. 11, no. 4, pp. 281–296, 1991.

[153] A. Buckingham, "Molecular quadrupole moments," *Quarterly Reviews, Chemical Society*, vol. 13, no. 3, pp. 183–214, 1959.

[154] A. D. McNaught and A. D. McNaught, *Compendium of chemical terminology*. Blackwell Science Oxford, 1997, vol. 1669.

[155] O. Anatole von Lilienfeld and A. Tkatchenko, "Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids," *The Journal of Chemical Physics*, vol. 132, no. 23, pp. –, 2010. [Online]. Available: http://scitation.aip.org/content/aip/journal/jcp/132/23/10.1063/1.3432765

[156] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, "Fragmentation methods: A route to accurate calculations on large systems," *Chemical Reviews*, vol. 112, no. 1, pp. 632–672, 2012. [Online]. Available: http://pubs.acs.org/doi/abs/10.1021/cr200093j

[157] M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, "The effective fragment potential method: A qm-based mm approach to modeling environmental effects in chemistry," *The Journal of Physical Chemistry A*, vol. 105, no. 2, pp. 293–307, 2001.

[158] L. V. Slipchenko, M. S. Gordon, and K. Ruedenberg, "Dispersion interactions in qm/efp," *The Journal of Physical Chemistry A*, vol. 121, no. 49, pp. 9495–9507, 2017.

[159] C. I. Viquez Rojas, J. Fine, and L. V. Slipchenko, "Exchange-repulsion energy in qm/efp," *The Journal of chemical physics*, vol. 149, no. 9, p. 094103, 2018.

[160] A. McLachlan, "Retarded dispersion forces between molecules," *Proc. R. Soc. Lond. A*, vol. 271, no. 1346, pp. 387–401, 1963.

[161] I. Adamovic and M. S. Gordon*, "Dynamic polarizability, dispersion coefficient c6 and dispersion energy in the effective fragment potential method," *Molecular Physics*, vol. 103, no. 2-3, pp. 379–387, 2005.

[162] J. Pople, R. Krishnan, H. Schlegel, and J. S. Binkley, "Derivative studies in hartree-fock and møller-plesset theories," *International Journal of Quantum Chemistry*, vol. 16, no. S13, pp. 225–241, 1979.

[163] Y. Yamaguchi, *A new dimension to quantum chemistry: analytic derivative methods in ab initio molecular electronic structure theory.* Oxford University Press, USA, 1994.

[164] L. Pauling, "The application of the quantum mechanics to the structure of the hydrogen molecule and hydrogen molecule-ion and to related problems." *Chemical Reviews*, vol. 5, no. 2, pp. 173–213, 1928.

[165] C. M. Aikens, S. P. Webb, R. L. Bell, G. D. Fletcher, M. W. Schmidt, and M. S. Gordon, "A derivation of the frozen-orbital unrestricted open-shell and restricted closed-shell second-order perturbation theory analytic gradient expressions," *Theoretical Chemistry Accounts*, vol. 110, no. 4, pp. 233–253, 2003.

[166] R. McWeeny, *Methods of molecular quantum mechanics.* Academic press, 1992.

VITA

# VITA

I, Pradeep Kumar Gurunathan, was born in Madurantakam, a small rural town in the state of Tamil Nadu in India. I went for elementary and secondary school education at Buvana Krishnan Matriculation School in Kelambakkam, Tamil Nadu. Following this, I went tp pursue high school studies in St. Patrick's Anglo Indian Higher Secondary School, Adyar, Tamil Nadu. I chose to go to Pondicherry University for a combined undergraduate-graduate program: five years Integrated Masters program in Chemistry. I did my graduate research there under the guidance of Prof. Binoy Krishna Saha on experimental and computational prediction of crystal co-former stability in dimer complexes. This experience gave me a platform to continue my further graduate studies in Purdue University.

At Purdue, I joined the research group of Prof. Lyudmila Slipchenko in Fall 2012. I started working on electronic structure investigation of atmospherically relevant complexes in surfaces and in bulk. Slowly, I started diverting my interests to biologically relevant applications. I tested and developed a variant of the Effective Fragment Potential method applicable to biomolecules, named BioEFP. I have also worked on machine-learning based parameter prediction models and a drug-target binding model using BioEFP. Apart from my research, I have also worked and managed at a student-run consulting club (PSC) at Purdue, helping a few local startups with their initial development phases. As a teaching assistant, I have taught General Chemistry (CHM115), Physical Chemistry lecture (CHM370). Physical Chemistry laboratory (CHM376) and Computational Chemistry (CHM673) courses.

In the future, I can be reached at geepradeep@gmail.com.