

2023

# Theoretical methods for electron-mediated processes

---

<https://hdl.handle.net/2144/47992>

*Boston University*

BOSTON UNIVERSITY  
GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

**THEORETICAL METHODS FOR  
ELECTRON-MEDIATED PROCESSES**

by

**JAMES R. GAYVERT**

B.S., Le Moyne College, 2017

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

2023

© 2023 by  
JAMES R. GAYVERT  
All rights reserved

Approved by

First Reader

---

Ksenia B. Bravaya  
Professor of Chemistry

Second Reader

---

David F. Coker  
Professor of Chemistry

*If I have seen further, it is by standing on the shoulders of Giants.*

## Acknowledgments

First, I would like to express my sincere gratitude to Prof. Ksenia Bravaya for her mentorship and guidance. Her seemingly endless supply of ideas and insights provided me with opportunities to work on so many different projects and learn so much more than I ever could have imagined. I am eternally grateful for the freedom she provided me to explore my own ideas and passions, and grow as both a scientist and a software developer. Lastly, the importance of her leadership during these last few years cannot be overstated, as she was able to keep both my research and the group as a whole on track during some truly unprecedented times.

I would like to extend my heartfelt thanks to all of the members of the Bravaya group over the last five years. It has been a joy to be your classmate, colleague, collaborator, and friend, both in person and over Zoom. The Bravaya group hikes will forever have a place among my most cherished memories of my time in Boston.

This work would not have been possible without the support of my family and friends. I would like to thank my dog, Ivy, for her companionship and words of motivation, my sister, Katie, for her guidance during the trials and tribulations of a PhD, my brothers, Tom and Chris, for their friendship and support, my Mom, for her love and guidance, and my Dad for essential scientific and technical discussions.

And finally, I would like to thank my wife, Dr. Megan Audette. We started a journey together five years ago to become doctors, and I am finally ready to uphold my end of the bargain. She bore the brunt of the 5+ hour drives on I-90, and supported me through all the ups and downs. She makes my life better every day, and I am excited to build our life together.

# **THEORETICAL METHODS FOR ELECTRON-MEDIATED PROCESSES**

**JAMES R. GAYVERT**

Boston University, Graduate School of Arts and Sciences, 2023

Major Professor: Ksenia B. Bravaya  
Professor of Chemistry

## ABSTRACT

Electron-driven processes lie at the core of a large variety of physical, biological, and chemical phenomena. Despite their crucial roles in science and technology, detailed description of these processes remains a significant challenge, and there is a need for the development of accurate and efficient computational tools that enable predictive simulation. This work is focused on the development of novel software tools and methodologies aimed at two classes of electron-mediated processes: (i) electron-molecule scattering, and (ii) charge transfer in proteins.

The first major focus of this thesis is the electronic structure of autoionizing electronic resonances. The theoretical description of these metastable states is intractable by means of conventional quantum chemistry techniques, and specialized techniques are required in order to accurately describe their energies and lifetimes. In this work, we have utilized the complex absorbing potential (CAP) method, and describe three developments which have advanced the applicability, efficiency, and accessibility of the CAP methodology for molecular resonances: (1) implementation and investigation of the smooth Voronoi potential (2) implementation of CAP in the projected scheme, and (3) development of the OpenCAP package, which extends the CAP methodology

to popular electronic structure packages.

The second major focus is the identification of electron and hole transfer (ET) pathways in biomolecules. Both experimental and theoretical inquiries into electron/hole transfer processes in biomolecules generally require targeted approaches, which are complicated by the existence of numerous potential pathways. To this end, we have developed an open-source web platform, eMap, which exploits a coarse-grained model of the protein crystal structure to (1) enable pre-screening of potentially efficient ET pathways, and (2) identify shared pathways/motifs in families of proteins.

Following introductory chapters on motivation and theoretical background, we devote a chapter to each new methodology mentioned above. The open-source software tools discussed herein are under active development, and have been utilized in published work by several unaffiliated experimental and theoretical groups across the world. We conclude the dissertation with a summary and discussion of the outlook and future directions of the OpenCAP and eMap software packages.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Resonances . . . . .	1
1.1.1	Bound states and stationary states . . . . .	1
1.1.2	Quasi-bound states . . . . .	3
1.1.3	Electronic resonances . . . . .	4
1.2	Electron/hole transfer in proteins . . . . .	7
<b>2</b>	<b>Theoretical Background</b>	<b>11</b>
2.1	Electronic structure methods for excited states . . . . .	11
2.1.1	Single reference wave function methods . . . . .	17
2.1.2	Multi-reference wave function methods . . . . .	24
2.1.3	Density Functional Theory . . . . .	31
2.2	Non-Hermitian description of resonances . . . . .	36
2.2.1	Complex Absorbing Potential . . . . .	38
2.2.2	Forms of CAP . . . . .	42
2.2.3	Projected CAP scheme . . . . .	44
2.3	Graph Theory . . . . .	46
2.4	Identifying ET pathways in proteins . . . . .	50
2.4.1	Pathways Model . . . . .	50
2.4.2	Frequent Subgraph Mining . . . . .	52
<b>3</b>	<b>Application of Box and Voronoi CAPs to Molecular Clusters</b>	<b>55</b>
3.1	Overview . . . . .	55

3.2	Introduction . . . . .	56
3.3	Methods . . . . .	60
3.4	Results and Discussion . . . . .	63
3.5	Conclusions . . . . .	76
<b>4</b>	<b>Projected CAP-EOM-CCSD method for electronic resonances</b>	<b>78</b>
4.1	Overview . . . . .	78
4.2	Introduction . . . . .	79
4.3	Theory . . . . .	81
4.4	Computational Details . . . . .	86
4.5	Results and Discussion . . . . .	88
4.6	Summary and Conclusions . . . . .	105
<b>5</b>	<b>OpenCAP: An Open-Source Framework for Computing Resonance Parameters in Molecules</b>	<b>106</b>
5.1	Overview . . . . .	106
5.2	Introduction . . . . .	107
5.3	Theory . . . . .	109
5.3.1	Complex Absorbing Potential . . . . .	110
5.3.2	CAP Subspace Projection Scheme . . . . .	111
5.4	Computational Details . . . . .	113
5.5	Results and Discussion . . . . .	114
5.5.1	Using OpenCAP . . . . .	114
5.5.2	Benchmarks . . . . .	117
5.5.3	Conclusions . . . . .	132
<b>6</b>	<b>eMAP: an online platform for identifying and visualizing electron and hole transfer pathways in proteins</b>	<b>133</b>
6.1	Overview . . . . .	133

6.2	Introduction . . . . .	134
6.3	Model and Implementation . . . . .	136
6.4	Applications . . . . .	143
6.5	Conclusions . . . . .	146
<b>7</b>	<b>eMap 2.0: A web-based platform for identifying electron transfer pathways in proteins and protein families</b>	<b>147</b>
7.1	Introduction . . . . .	148
7.2	Model . . . . .	150
7.2.1	Preliminaries . . . . .	150
7.2.2	Single protein analysis . . . . .	151
7.2.3	Searching for common electron/hole pathways . . . . .	154
7.3	Application examples . . . . .	160
7.3.1	Efficient electron transfer pathway in a single protein . . . . .	160
7.3.2	Shared electron transfer pathways and structural motives . . . . .	164
7.3.3	Guided search . . . . .	165
7.4	Conclusions . . . . .	168
<b>8</b>	<b>Summary and Outlook</b>	<b>170</b>
<b>A</b>	<b>Supporting Information for Chapter 5</b>	<b>173</b>
A.1	Geometries . . . . .	174
A.1.1	$N_2$ . . . . .	174
A.1.2	$C_2H_4$ . . . . .	174
A.1.3	$CO_2$ . . . . .	174
A.1.4	$CH_2O$ . . . . .	174
A.2	CAP Parameters . . . . .	174
A.3	Alternative MRPT schemes . . . . .	174
A.4	Smaller MRCI active spaces . . . . .	176

A.5	Representative $\eta$ -trajectories . . . . .	177
<b>B</b>	<b>Algorithms for Chapter 6</b>	<b>186</b>
B.0.1	Percent-based edge pruning algorithm . . . . .	186
B.0.2	Degree-based edge pruning algorithm . . . . .	186
B.0.3	Identification of non-protein ET active moieties . . . . .	187
<b>C</b>	<b>Supporting Information for Chapter 7</b>	<b>189</b>
C.1	Default Mining Parameters . . . . .	189
C.2	Protein Subgraph Clustering . . . . .	189
C.2.1	Structural Similarity . . . . .	189
C.2.2	Sequence Similarity . . . . .	190
	<b>References</b>	<b>191</b>
	<b>Curriculum Vitae</b>	<b>219</b>

# List of Tables

2.1	DFS codes corresponding to the traversals in Fig. 2-8. . . . .	53
3.1	Resonance positions (widths) of the four states of the $(CO)_2^-$ dimer at the equilibrium geometry, and at $R = 20 \text{ \AA}$ , computed using box and Voronoi CAPs. . . . .	76
4.1	Computed resonance energy (width), $E_R(\Gamma)$ , for $N_2^-$ and $CO^-$ $\pi^*$ shape resonances. All quantities are given in eV. . . . .	91
4.2	Basis set dependence of the computed zero-order resonance parameters for the $N_2^-$ $\pi^*$ shape resonance. All energies are given in eV. . . . .	94
4.3	Basis set dependence of the computed first-order resonance parameters for the $N_2^-$ $\pi^*$ shape resonance. All energies are given in eV. . . . .	95
4.4	Zero-order resonance parameters for the lowest electron-attached $\pi^*$ shape resonances in polyatomic molecules. All energies are given in eV. . . . .	99
4.5	First-order resonance parameters for the lowest electron-attached $\pi^*$ shape resonances in polyatomic molecules. All energies are given in eV. . . . .	99
4.6	Resonance positions and widths of three lowest $\pi^*$ resonances in uracil. All energies are given in eV. . . . .	101
4.7	First-order positions and widths of three lowest $\pi^*$ states of naphthalene and 1- and 2-cyanonaphthalene. All energies are given in eV. . . . .	102
4.8	Comparison of pCAP and fCAP for $\pi^*$ and $\sigma^*$ states of the chloroethene anion. All energies were obtained from zero-order trajectories, and are given in eV. . . . .	104

5.1	Summary of CAP subspace projection methods used in this work. . .	114
5.2	Resonance parameters for the ${}^2B_1$ resonance in $CH_2O^-$ computed using the EOM-EA-CCSD and ADC family of methods. . . . .	118
5.3	Computed CAP/MRPT first-order resonance parameters, along with selected theoretical and experimental results for the $\pi^*$ shape resonances studied in this work. $E_R$ and $\Gamma$ are given in eV. . . . .	124
5.4	Computed CAP/MRPT first-order resonance parameters for the ${}^2B_{2g}$ resonance state of $C_2H_4^-$ using selected active spaces and state averaging schemes. $E_R(\Gamma)$ are given in eV. . . . .	126
5.5	Computed CAP/MR-CI first-order resonance parameters. $E_R(\Gamma)$ are given in eV. . . . .	127
5.6	Computed CAP/MR-CI first-order $N_2^-$ ${}^2\Pi_g$ resonance parameters for selected active spaces. $E_R(\Gamma)$ are given in eV. . . . .	127
5.7	Computed first order resonance parameters for the ${}^2\Pi_g$ resonance in $N_2^-$ computed using various electronic structure methods. $E_R(\Gamma)$ are given in eV. . . . .	131
7.1	Four identified ET pathways from Y122 ( $\beta$ subunit) to C439 ( $\alpha$ subunit) in class Ia RNR from <i>E.coli</i> . . . . .	162
A.1	Box CAP Parameters (in bohr) for systems studied in this work. . . .	175
A.2	Computed CAP/MRPT first-order resonance parameters using IPEA shift of 0.25 a.u. and real shift of 0.2 a.u. $E_R(\Gamma)$ are given in eV. . . .	175

A.3	Computed CAP/MRPT first-order resonance parameters using alternative scheme without diffuse orbital. The orbitals are optimized for the anionic states only, and the neutral ground state energy is obtained using CASCI/CASPT2 with the optimized orbitals. An IPEA shift of 0.25 a.u. and an imaginary shift of 0.2 a.u. are used for the perturbation theory step for both calculations. $E_R(\Gamma)$ are given in eV.	175
A.4	Computed CAP/MRCI first-order resonance parameters using various small active spaces. No stationary point could be identified on either trajectory for MRCISD/MRCISD(Q) for $CO_2$ . The trajectories are provided in Figs. A.9-A.12.	176
C.1	Relevant default parameters for frequent subgraph mining in eMap version 2.0.2.	189

# List of Figures

1.1	Discrete and continuous part of the eigenvalue spectrum of system supporting a resonance. In the case of an electronic shape resonance, A can refer to a closed shell molecule, and B an excess electron which is temporarily localized on the molecule. In the Hermitian formalism, a resonance is manifested by an increased density of states in the continuum. . . . .	5
1.2	Schematic of a shape resonance (left) and a Feshbach resonance (right). The shape resonance is temporarily trapped by the target and decays by a tunneling mechanism. The Feshbach resonance is a bound state of the closed channel, but can decay due to coupling with the open channel. In molecular systems, the electronic resonances can be approximately classified to Feshbach or shape based on the mechanism of their decay (bottom panel). Shape resonances decay in a one-electron manner, whereas the decay of Feshbach resonances is intrinsically a two-electron processes. . . . .	7
1.3	Marcus free energy curves in the diabatic (left) and adiabatic (right) pictures. Eq. 1.13 holds in the non-adiabatic ( $H_{DA} \ll \Delta G^\circ, \lambda$ ) regime, and in an ET event, the system jumps from the donor curve to the acceptor curve. . . . .	9
2.1	Selected variants of EOM-CC which generate different types of target states. . . . .	19



2·2	Schematic illustrating the orbital partitioning in the CASSCF method. The active space is CAS(2,4), which indicates 2 active electrons and 4 active orbitals. . . . .	26
2·3	$\eta$ -trajectories for the ${}^2\Pi_g$ resonance in $N_2^-$ computed using projected CAP-EOM-EA-CCSD[114]. The left panel shows the eigenvalue spectrum for four states, and the right panel shows the uncorrected(red) and corrected (blue) trajectories for the resonance state. . . . .	40
2·4	Comparison of different box and Voronoi CAPs for a model $N_2^-(H_2O)_2$ cluster. Plotted is the isoline $W=1$ hartree for box CAPs of 4 bohr (brown) and 3 bohr (magenta), and smooth Voronoi CAPs of 4 bohr (navy) and 3 bohr (cyan) in the plane of the molecular system. This image is reproduced with permission from Ref. [114]. Copyright 2022 American Chemical Society. . . . .	43
2·5	A graph with node labels $\{X, Y, Z\}$ and edge labels $\{a, b, c\}$ . . . . .	47
2·6	A tree illustrating the order nodes are visited in a depth-first search traversal of a graph. Back edges (edges between nodes which have already been visited) are omitted. . . . .	49
2·7	A graph representing the connectivity network between atoms in a protein. $P'_{ij}$ are penalty functions associated with electron/hole hopping between the sites. This image is reproduced with permission from Ref. [151]. Copyright 2019 American Chemical Society. . . . .	51
2·8	Examples of DFS traversals for a graph with node labels $\{X, Y, Z\}$ , and edge labels $\{a, b, c\}$ . Back edges (edges between nodes which have already been visited) are indicated as dashed lines. . . . .	53

2.9	Schematic illustrating the gSpan algorithm. The DFS codes are sorted in lexicographic ordering, and a DFS traversal of the search space is performed to identify frequent subgraphs. Subtrees of infrequent or non-minimal DFS codes are pruned. . . . .	54
3.1	Geometries of the $N_2(H_2O)_2$ model cluster (A), and the $(CO)_2$ dimer (B). Resonance states are investigated as functions of the intermolecular distance, R. . . . .	63
3.2	Uncorrected (left) and corrected (right) eigenvalue trajectories obtained using various CAPs for the ${}^2\Pi_g$ resonance of $N_2^-$ . V1/B1, V3/B3, V5/B5 correspond to Voronoi/box CAPs with onsets of 1, 3, and 5 bohr, respectively. . . . .	64
3.3	Uncorrected (left) and corrected (right) eigenvalue trajectories obtained using various CAPs for the ${}^2\Pi$ resonance of $CO^-$ . V1/B1, V3/B3, V5/B5 correspond to Voronoi/box CAPs with onsets of 1, 3, and 5 bohr, respectively. . . . .	65
3.4	Dependence of computed resonance parameters for ${}^2\Pi_g$ resonance of $N_2^-$ on the CAP onset: $r_{cut}$ and $r_{box}$ for Voronoi and box CAPs, respectively. Corrected results for box and Voronoi CAPs are shown in blue and red, respectively. . . . .	66
3.5	Dependence of computed resonance parameters for ${}^2\Pi$ resonance of $CO^-$ on the CAP onset: $r_{cut}$ and $r_{box}$ for Voronoi and box CAPs, respectively. Corrected results for box and Voronoi CAPs are shown in blue and red, respectively. . . . .	66

3·6	Comparison of different box and Voronoi CAPs for the $N_2^-(H_2O)_2$ at $R = 5\text{\AA}$ . Plotted is the isoline $W=1$ hartree for box CAPs of 4 bohr (brown) and 3 bohr (magenta), and smooth Voronoi CAPs of 4 bohr (navy) and 3 bohr (cyan) in the plane of the molecular system. . . . .	67
3·7	Resonance positions (top) and widths(bottom) for the in-plane ${}^2B_{1g}$ (indicated as $\bullet$ ) and out-of-plane ${}^2B_{3g}$ (indicated as $\nabla$ ) resonance states of the model $N_2^-(H_2O)_2$ cluster, computed using box CAP (blue) and Voronoi CAP (red). The widths of the isolated $N_2^-$ computed using the corresponding CAP are indicated as dashed lines. The energy of the neutral N-electron ground state is shown as a solid black line. . . . .	69
3·8	Corrected trajectories for the in-plane ${}^2B_{1g}$ state of the model $N_2^-(H_2O)_2$ cluster, computed using a box CAP (blue) and a Voronoi CAP (red). . . . .	70
3·9	Schematics and the real part of complex Dyson orbitals[181]for each of the four resonance states of the $(CO)_2^-$ dimer. The Dyson orbitals were plotted at the equilibrium geometry with an isosurface value of 0.02. . . . .	71
3·10	Corrected resonance positions and widths for the out-of-plane-in-phase ${}^2A_u$ (indicated as $\bullet$ ) and out-of-plane-out-of-phase ${}^2B_g$ (indicated as $*$ ) resonance states of the $(CO)_2^-$ dimer, computed using box CAP (blue) and Voronoi CAP (red). The values for isolated $CO^-$ computed using a Voronoi CAP ( $r_{cut}=3.0$ a.u.) are indicated as dashed lines. . . . .	72
3·11	Corrected resonance positions and widths for the in-plane-in-phase ${}^2A_g$ (indicated as $\bullet$ ) and in-plane-out-of-phase ${}^2B_u$ (indicated as $*$ ) resonance states of the $(CO)_2^-$ dimer, computed using box CAP (blue) and Voronoi CAP (red). The values for isolated $CO^-$ computed using analogous CAPs are indicated as dashed lines. . . . .	74

3·12	CAP-augmented Hamiltonian eigenvalue spectra for Voronoi CAP (left) and box CAP (right) for the in-plane, in-phase ${}^2A_g$ state of of the $(CO)_2^-$ dimer at $R = 8 \text{ \AA}$ . . . . .	74
3·13	Corrected trajectories for the in-plane, in-phase ${}^2A_g$ state of the $(CO)_2^-$ dimer, computed using a box CAP (blue) and a Voronoi CAP (red). . . . .	75
4·1	Comparison of the zero-order (left) and the first-order (right) trajectories for the ${}^2\Pi_g$ shape resonance in $N_2^-$ , computed with pCAP with 3 states (red), 4 states (blue), 30 states (magenta), and fCAP (green). Points associated with $\eta_{opt}$ are indicated by a $\bullet$ . $\eta$ was varied from 0.0 to 0.05. . . . .	88
4·2	Dependence of the zero-order (red) and first-order (blue) resonance parameters on the size of the correlated basis for the ${}^2\Pi_g$ shape resonance in $N_2^-$ . The result obtained from fCAP is indicated by a dashed line. . . . .	89
4·3	Comparison of zero-order (left) and first-order (right) trajectories for the ${}^2\Pi$ shape resonance in $CO^-$ , computed with pCAP with 5 states (red), 7 states (orange), 10 states (blue), 30 states (magenta), and fCAP(green). Points associated with $\eta_{opt}$ are indicated by a $\bullet$ . $\eta$ was varied from 0.0 to 0.02. . . . .	90
4·4	Dependence of the zero-order (red) and first-order (blue) resonance parameters on the size of the correlated basis set for the ${}^2\Pi$ resonance in $CO^-$ . The results obtained from fCAP are indicated by a dashed line. . . . .	91
4·5	Dependence of the zero-order (red) and first-order (blue) resonance parameters on the box size for the ${}^2\Pi$ resonance in $CO^-$ . Results obtained from pCAP are plotted with solid lines, while fCAP results are plotted with dashed lines. . . . .	92

4.6	Comparison of zero-order (left) and first-order (right) trajectories for $N_2^-$ with the aug-cc-pvtz + 3s3p[C] basis set, computed with pCAP (magenta) and fCAP (green). Points associated with $\eta_{opt}$ are indicated by a $\bullet$ . $\eta$ was varied from 0.0 to 0.15. 30 states were included in the subspace for the pCAP trajectories. . . . .	95
4.7	First-order resonance position (left) and width (right) along complex PEC for the $^2\Pi_g$ shape resonance in $N_2^-$ computed by pCAP (magenta) and fCAP (green). For pCAP, 10 states were included in the correlated basis at all points. . . . .	96
4.8	Comparison of zero-order (left) and first-order (right) trajectories for the $^2\Pi_g$ shape resonance in $N_2^-$ at $R(N-N) = 0.80 \text{ \AA}$ . Results computed with pCAP (10 states) are plotted in magenta, and results from fCAP are plotted in green. Points associated with $\eta_{opt}$ determined by Eq. 4.13 are indicated by a $\bullet$ , and points associated with $\eta_{opt}$ using the criteria from Ref. [118] are indicated by a $\star$ . $\eta$ was varied from 0.0 to 0.05. . . . .	98
5.1	CAP-ADC(3) eigenvalue spectrum (left) and $\eta$ -trajectory (right) for the $^2B_1$ resonance in $CH_2O^-$ . The zero- and first-order trajectories are shown in red and blue, respectively. $\eta$ is varied from 0.0 to 0.02 in steps of 0.00005, and the 6th root of the CAP augmented Hamiltonian is tracked to generate the eigenvalue trajectories. $\eta_{opt}$ on the first-order trajectory is indicated by a $\star$ . . . . .	116
5.2	Comparison of the zero-order (left) and the first-order (right) trajectories for the $^2B_1$ resonance in $CH_2O^-$ , computed using CAP/ADC(3) with 9 states (green), 13 states (red), 20 states (blue), and 24 states (magenta). The point associated with $\eta_{opt}$ is indicated by a $\star$ . $\eta$ was varied from 0.0 to 0.05. . . . .	119

5.3	Dependence of resonance half-widths ( $\Gamma/2$ ) and CAP norms ( $ W $ ) on the numerical integration parameters for box CAP for the ${}^2B_1$ resonance in $CH_2O^-$ . The label $N$ for radial precision refers to a precision of $10^{-N}$ . . . . .	120
5.4	Dependence of resonance half-widths ( $\Gamma/2$ ) and CAP norms ( $ W $ ) on the numerical integration parameters for Voronoi CAP ( $r_{cut}=3.0$ bohr) for the ${}^2B_1$ resonance in $CH_2O^-$ . The label $N$ for radial precision refers to a precision of $10^{-N}$ . . . . .	121
5.5	CAS(3e,5o) active space for $C_2H_4^-$ . The fake IP orbital is not shown, and the orbitals are visualized with the isosurface value of 0.005. Each of the four model states in the SA(4)-CASSCF(3e,5o) calculation is dominated by a configuration corresponding to electron attachment to one of the $b_{2g}$ orbitals, $\Phi_0 - \Phi_3$ . The state dominated by $\Phi_0$ represents the neutral. . . . .	123
5.6	Comparison of the zero-order (left) and the first-order (right) trajectories for the ${}^2B_1$ resonance in $CH_2O^-$ , computed using CAP/XMS-CASPT2 with the following active spaces: SA(5)-CAS(3e,6o)(green), SA(6)-CAS(3e,7o)(red), and SA(7)-CAS(3e,8o)(blue). The point associated with $\eta_{opt}$ is indicated by a $\star$ . $\eta$ was varied from 0.0 to 0.05. . .	125
5.7	First-order resonance positions (left) and widths (right) along the complex PEC for the ${}^2\Pi_g$ shape resonance in $N_2^-$ , computed by CAP-EOM-CCSD (magenta) and CAP-TDDFT(HFELYP) (green). . . . .	131

6.1	A graph representing the connectivity network between the aromatic sites in a protein. $P'_{ij}$ is a penalty function associated with electron/hole hopping between the sites (see text for more details). Surface-exposed residues are indicated as squares, while buried residues are indicated as circles. . . . .	136
6.2	Schematic representing the dependencies between different modules in eMap. . . . .	139
6.3	Options specifying pairwise distance map construction: specification of the chains included in the analysis, algorithm used to identify surface-exposed residues, intersite distance evaluation scheme, and standard ETA sites (a); selection of the non-protein ETA sites included in the analysis (b); tuning thresholds, cutoffs, and penalty function parameters for drawing graph edges (c). . . . .	141
6.4	Results of the eMap analysis for Cry1 protein: shortest path connecting flavin of FAD (FAD510-2) and the terminal Trp from the Trp triad (W324) visualized in 2D (a) and 3D (b). . . . .	144
6.5	The results of eMap analysis for 6MJS structure: (a) the shortest path connecting Re (CUST-2) and Cu (CUST-1) atoms identified using the default parameters for the graph edges generation; (b) the shortest path connecting Re (CUST-2) and Cu (CUST-1) identified using tighter Distance Cutoff (13 Å); (c) 3D image of the pathway identified with 13 Å cutoff. . . . .	145
7.1	The workflow of the eMap single protein analysis (left panel) and common subgraph mining for a series of proteins (right panel). . . . .	153

- 7.2 Flowchart illustrating the three major steps of frequent subgraph mining with eMap. The first step (a) generates a graph data set which consists of protein graphs created by the single protein version of eMap. In the second step (b), the graph data set is mined for frequently occurring subgraph patterns. For each identified pattern, all instances of said pattern within the data set are found using graph matching, which yields a set of *protein subgraphs*. In the final step (c), the protein subgraphs matching a selected pattern are clustered into groups based on similarity. Shown in the bottom of panel C is the web interface which enables users to browse through clusters of similar protein subgraphs, visualized in 2D and in the 3D protein structure. . . . . 155
- 7.3 Example of DFS codes for a graph with node labels  $\{X, Y, Z\}$ , and edge labels  $\{a, b, c\}$ . Back edges (edges between nodes which have already been visited) are indicated as dashed lines. Many different DFS codes can be generated for a given graph by choosing different starting nodes and making different choices at each step of the traversal. By the gSpan lexicographic ordering scheme, the ordering of these three codes is  $3 < 2 < 1$ , i.e. 3 (the most right column) is the minimal DFS code among these three[143].  $X, Y, Z$ , etc. correspond to the chemical nature of the redox-active site (e.g. aromatic amino acids) and  $a, b, c$  label edges classified based on their length. . . . . 157



7.4	Illustration of depth-first traversal of the search space by the gSpan algorithm for an example graph data set. Starting from 1-edge graphs, the search space is expanded by adding additional edges, those which are infrequent are pruned (indicated by a dashed circle). The DFS code tree is constructed using the canonical lexicographic ordering, so when a non-minimum DFS code is encountered, the entire branch below can be pruned[143]. . . . .	158
7.5	Illustration of algorithm used to cluster a set of protein subgraphs $\{V\}$ which match a particular subgraph pattern. The largest cluster in the super graph $G(V,E)$ contains 16 protein subgraphs which are deemed structurally or sequentially similar. . . . .	160
7.6	Structure of the class Ia RNR from <i>E.coli</i> (a) and inter-unit radical transfer pathways involved in the catalysis, $\beta Y122-\beta W48-\beta Y356-\alpha Y731-\alpha Y730-\alpha C439$ (b). . . . .	161
7.7	The results of the eMap analysis for the 6W4X structure. Only B and C chains were included in the analysis, all T,W, and C residues were considered as ETA sites. All aromatic moieties of co-factors and non-standard aminoacids were considered as ETA sites. The default parameters used for the simulation are provided in Table C1 in Appendix C. The graph structures of the class Ia RNR from <i>E.coli</i> and inter-unit radical transfer pathways involved in the catalysis, $\beta Y122-\beta W48-\beta Y356-\alpha Y731-\alpha Y730-\alpha C439$ , are shown in (a) and (b), and the top ranked pathway is visualized in the crystal structure in (c). .	163

7.8	Relevant shared graph patterns and subgraphs for a test set of 27 flavo- proteins, containing cryptochromes, photolyases, and non-cryptochromes and non-photolyases proteins. Only the shared patterns with clusters with more than 15 proteins based on structural similarity criteria are shown. Groups 8/9 and 21/22 only differ by presence (8 and 21) or absence (9 and 22) of adenine of FAD co-factor, which is represented by $\square^*$ in the figure. The residue IDs correspond to <i>Drosophila</i> Cryp- tochrome, PDB ID 4GU5. . . . .	166
7.9	Residues involved in photo-induced electron transfer (canonical Trp triad, Trp420, Trp397, Trp342) as well as other Trp residues from shared subgraphs identified by eMap. The numbering and structure correspond to PDB ID 4GU5. . . . .	167
7.10	Active and inactive forms of chain A in the di-haem Cytochrome C Peroxidase in <i>Marinobacter nauticus</i> (PDB structures:1RZ5, 1RZ6). Shown are overlays of the active and inactive system, where Trp94 is shifted out of the central location between hemes in the inactive form. This inactive structure was not identified as a HEC-Trp-HEC pathway by eMap. . . . .	168
A.1	CAP trajectories for $CH_2O^-$ computed using EOM-EA-CCSD and the ADC(N) family of methods. . . . .	177
A.2	CAP-XMS-CASPT2 trajectories for the $\pi^*$ resonances for the active spaces listed in Table 4 of the main text. . . . .	178
A.3	CAP-MRCIS trajectories for the $\pi^*$ resonances for the active spaces listed in Table 6 of the main text. . . . .	178
A.4	CAP-MRCISD trajectories for the $\pi^*$ resonances for the active spaces listed in Table 6 of the main text. . . . .	179

A·5	CAP-MRCISD(Q) trajectories for the $\pi^*$ resonances for the active spaces listed in Table 6 of the main text. . . . .	179
A·6	CAP-TDDFT trajectories for $N_2^-$ computed using various functionals. . . . .	180
A·7	CAP trajectories for $N_2^-$ computed using the CIS(D) family of methods. . . . .	181
A·8	Trajectories for the $N_2^-$ potential energy curve computed using CAP-TDDFT (HFE_BLYP). . . . .	182
A·9	CAP-MRCIS trajectories for the $\pi^*$ resonances for the active spaces listed in Table A.4. . . . .	183
A·10	CAP-MRCISD trajectories for the $\pi^*$ resonances for the active spaces listed in Table A.4. . . . .	184
A·11	CAP-MRCISD(Q) trajectories for the $\pi^*$ resonances for the active spaces listed in Table A.4. . . . .	185
A·12	CAP-XMS-CASPT2 trajectories for the $\pi^*$ resonances for the active spaces listed in Table A.4. . . . .	185

# List of Abbreviations

ADC	.....	Algebraic Diagrammatic Construction Theory
AO	.....	Atomic Orbital
BO	.....	Born Oppenheimer Approximation
CAP	.....	Complex Absorbing Potential
CASPT2	.....	Complete Active Space Perturbation Theory (2nd Order)
CASSCF	.....	Complete Active Space Self Consistent Field
CC	.....	Coupled Cluster
CI	.....	Configuration Interaction
DFS	.....	Depth First Search
DFT	.....	Density Functional Theory
EOM-CCSD	.....	Equations of Motion Coupled Cluster with Singles and Doubles Substitutions
ET	.....	Electron Transfer
FCI	.....	Full Configuration Interaction
FSM	.....	Frequent Subgraph Mining
gSpan	.....	Graph-based Substructure Pattern Mining
HF	.....	Hartree Fock
MO	.....	Molecular Orbital
MRCI	.....	Multi-Reference Configuration Interaction
MRPT	.....	Multi-Reference Perturbation Theory
MS-CASPT2	.....	Multi-State Complete Active Space Perturbation Theory (2nd Order)
TDDFT	.....	Time Dependent Density Functional Theory
TDSE	.....	Time Dependent Schrödinger Equation
TISE	.....	Time Independent Schrödinger Equation

# Chapter 1

## Introduction

In this chapter, we give the theoretical foundations for the electron-mediated processes studied in this thesis. First we discuss resonances, a class of metastable quantum states that often serve as key intermediates in chemical conversions. Second we describe electron and hole transfer reactions, which are essential components of a wide variety of biological processes.

### 1.1 Resonances

Below, we outline the quantum theory of resonances, focusing on their key properties that distinguish them from bound states. We also provide the foundation for the non-Hermitian description of resonances, which is described in Chapter 2.

#### 1.1.1 Bound states and stationary states

In quantum mechanics, the time evolution of a system is given by the time-dependent Schrödinger equation (TDSE)[1]:

$$\hat{H}|\Psi_n(r, t)\rangle = i\hbar\frac{\partial}{\partial t}|\Psi_n(r, t)\rangle \quad (1.1)$$

where  $\hat{H}$  is the Hamiltonian operator, and  $\Psi_n(r, t)$  is the wave function which depends on position  $\mathbf{r}$  and time  $t$ . Assuming the separation of variables  $\Psi_n(r, t) = \Psi_n(r)f_n(t)$ ,

for a time-independent Hamiltonian, one arrives to the time-independent Schrödinger equation:

$$\hat{H} |\Psi_n(r)\rangle = E_n |\Psi_n(r)\rangle. \quad (1.2)$$

Substituting  $\Psi_n(r, t) = \Psi_n(r)f_n(t)$  into Eq. 1.1 and taking into account Eq. 1.2, the time-dependent wave-function becomes:

$$f_n(t) = \exp(-iE_n t/\hbar) \quad (1.3)$$

Thus, the resulting solution to the time-dependent Schrödinger equation has the form:

$$|\Psi_n(r, t)\rangle = |\Psi_n(r)\rangle \exp(-iE_n t/\hbar) \quad (1.4)$$

While the wave-function itself depends on time, it represents a *stationary state*, i.e. the associated probability density,  $P(r, t) = |\Psi_n(r, t)|^2$ , and the observables are independent of time:

$$\begin{aligned} P(r, t) &= |\Psi_n(r, t)|^2 \\ &= |e^{-iE_n t/\hbar} \Psi_n(r, 0)|^2 \\ &= |e^{-iE_n t/\hbar}|^2 |\Psi_n(r, 0)|^2 \\ &= |\Psi_n(r, 0)|^2 \end{aligned} \quad (1.5)$$

There are two types of solutions to the TISE: bound states, which are associated with discrete eigenvalues of the Hamiltonian, and scattering states, which belong to the continuous part of the spectrum. Electronic structure theory is primarily concerned with bound states, which tend to remain localized in one or more regions of space. Such states are elements of the  $L^2$  Hilbert space of square integrable functions,

and thus are normalizable under the standard Hermitian inner product:

$$\langle \phi_i | \phi_i \rangle = \int \phi_i^*(r) \phi_i(r) dr = 1, \quad (1.6)$$

and satisfy the following boundary conditions:

$$\Psi_n(x \rightarrow \pm\infty) = 0. \quad (1.7)$$

In contrast to bound states, scattering states do not remain localized, and thus are not normalizable, and cannot be represented using the square-integrable basis functions which are commonly used in electronic structure calculations for bound states.

### 1.1.2 Quasi-bound states

A quantum system does not always exist in a stationary state. Any superposition of stationary states is a valid solution to the TDSE, which can yield quantum states whose observables are dependent on time. These *non-stationary* states do not have a well defined energy, and obey the time-energy version of the Heisenberg uncertainty principle:

$$\Delta E \Delta t \geq \frac{\hbar}{2} \quad (1.8)$$

Long lived non-stationary states that are primarily localized in a region of space are often referred to as “quasi-bound” states[2]. While they still do not have a well-defined energy, the uncertainty in the energy is relatively small, and so quasi-bound states can be considered to possess a discretized energy similar to a bound state, but also possess a *width*,  $\Gamma = 2\Delta E$ . It follows that a quasi-bound state becomes a bound state when its width becomes 0.

A *resonance* is a quasi-bound state that has sufficient energy to break up into

subsystems, and exists long enough to be observed experimentally. Formally, resonances can be defined as poles of the scattering matrix[3]. They are ubiquitous in chemistry and physics, often existing as intermediate compound states[4]. The most familiar example can be found in nuclear physics; the well-known  $\alpha$ -decay process involves a metastable nuclid decaying by emission of an  $\alpha$ -particle[5]. The ubiquity and similarity of resonances to bound states has led to numerous theoretical and experimental efforts to characterize them. One of the most convenient theoretical frameworks is derived from the Siegert formalism, which associates a resonance with a discrete solution of the TISE that satisfies purely outgoing boundary conditions, and has a complex Siegert energy[6]:

$$E = E_R - i\Gamma/2 \tag{1.9}$$

where  $\Gamma = \frac{\hbar}{\tau}$  and  $\tau$  corresponds to the lifetime. This notion of a complex energy has also been derived using the Feshbach projection operator formalism, which describes a resonance as a bound state coupled to the continuum[7], and scattering theory, which associates resonances with complex poles of the S-matrix[3]. The utility of a complex energy can be easily understood by plugging it into the stationary state condition (Eq. 1.4):

$$\Psi_R(r, t) = e^{-\Gamma t/2\hbar} e^{-iE_R t/\hbar} \Psi_R(r, 0), \tag{1.10}$$

which shows that a complex energy with a negative imaginary part yields a stationary state that decays exponentially in time, and whose lifetime depends on the width  $\Gamma$ .

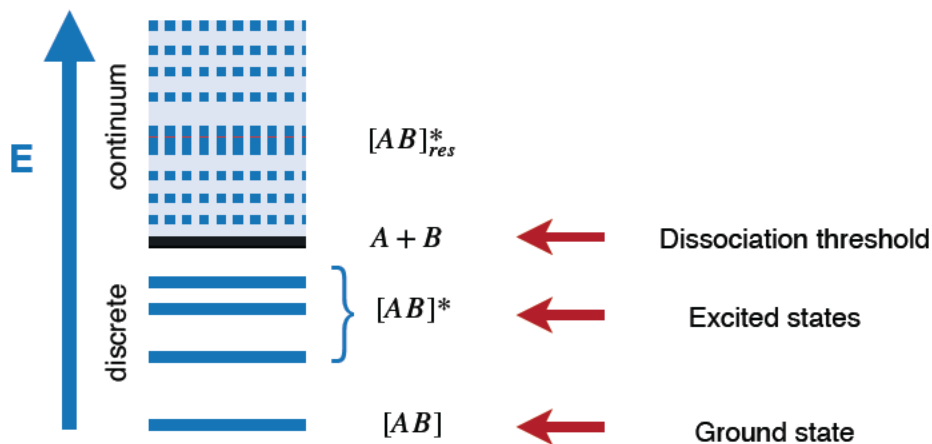
### 1.1.3 Electronic resonances

This thesis is concerned with a particular class of resonances known *electronic resonances*. Electronic resonances are metastable scattering states embedded in the ion-



ization/detachment continuum of the electronic Hamiltonian (Fig. 1.1) that decay by spontaneous electron emission[8].

These molecular electronic states that lie above the ionization or detachment thresholds can be accessed in various ways, including electron impact, photo-excitation, and through core-ionization or core-excitation. They can be probed experimentally using approaches such as electron-molecule scattering experiments (e.g. electron impact and electron transmission spectroscopies[9–13]) and photodetachment spectroscopy[14]. The width of a resonance can sometimes be determined from the width of the resonance peak in experimental cross sections, though doing so in a way which does not depend on nuclear degrees of freedom is generally not possible.



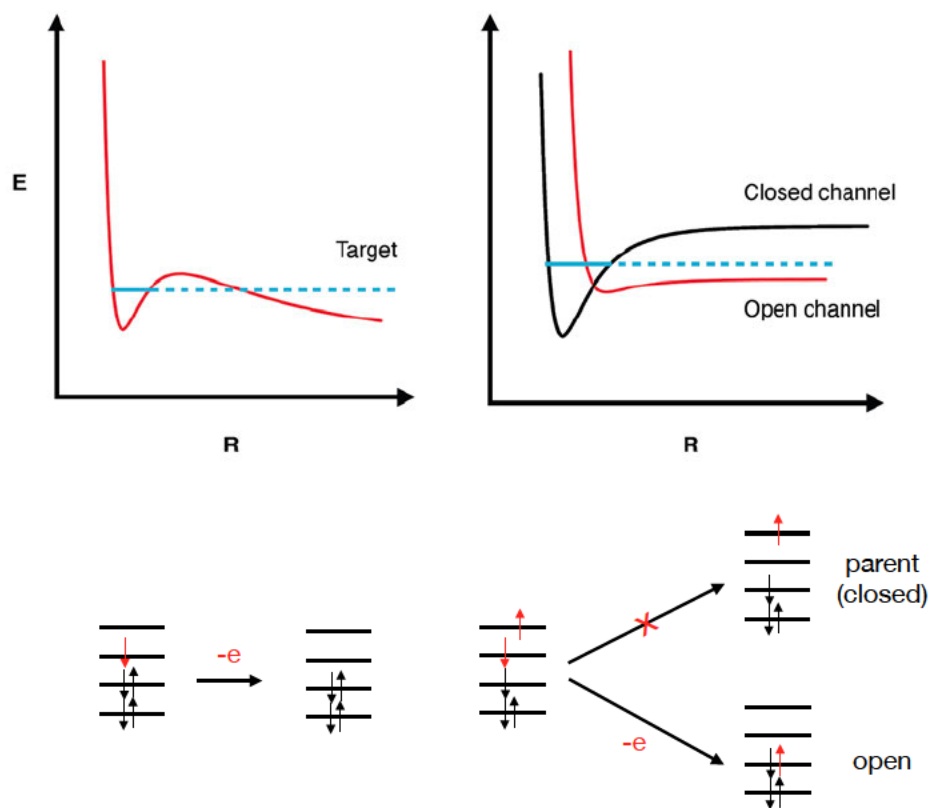
**Figure 1.1:** Discrete and continuous part of the eigenvalue spectrum of system supporting a resonance. In the case of an electronic shape resonance, A can refer to a closed shell molecule, and B an excess electron which is temporarily localized on the molecule. In the Hermitian formalism, a resonance is manifested by an increased density of states in the continuum.

Electronic resonances are usually classified as either shape resonances or Feshbach resonances based on their decay mechanism[9, 15](Fig. 1.2). Shape resonances are formed when an electron becomes temporarily trapped due to the shape of the

potential, and they decay by one-electron tunneling mechanisms. (N+1)-electron shape resonances lie above the parent N-electron state, which can be the ground electronic state or an excited electronic state (the latter are known as “core-excited” shape resonances), and the parent state remains unchanged after the excess electron tunnels out[9, 15, 16]. Feshbach resonances decay by two-electron processes, e.g. electron emission and de-excitation. Feshbach resonances generally lie below the parent electronic state, and are metastable due to the presence of one or many open decay channels[15]. Shape resonances tend to exhibit lifetimes in the femtosecond-picosecond range, while Feshbach resonances are generally longer lived[8]. While the primary focus of this work has been the study of shape resonances, the methodologies developed in Chapter 3 and Chapter 5 can be used for the study of Feshbach resonances.

The relatively long lifetimes of molecular electronic resonances implies that they can play important roles in chemical reactions. Electronic resonances are often implicated as intermediates in processes such as radiolysis and damage to DNA by secondary electrons[17, 18], plasmonic catalysis[19], and interstellar chemistry[20, 21]. They are also important in X-ray and attosecond spectroscopies[22–24], for example, the Auger effect is a resonant process involving a core-ionized Feshbach resonance[25].

Despite their importance, reliable calculation of energies and lifetimes of electronic resonances in molecules remains a significant challenge. The specialized methods commonly used for computing resonance parameters are discussed in Sec. 2.2. The work carried out in this thesis, described in Chapters 3-5, has been aimed at enhancing the applicability, efficiency, and accessibility of modern electronic structure methodologies for the description of electronic resonances in order to facilitate further development of this exciting and under-explored field.

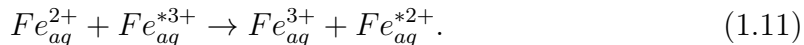


**Figure 1.2:** Schematic of a shape resonance (left) and a Feshbach resonance (right). The shape resonance is temporarily trapped by the target and decays by a tunneling mechanism. The Feshbach resonance is a bound state of the closed channel, but can decay due to coupling with the open channel. In molecular systems, the electronic resonances can be approximately classified to Feshbach or shape based on the mechanism of their decay (bottom panel). Shape resonances decay in a one-electron manner, whereas the decay of Feshbach resonances is intrinsically a two-electron processes.

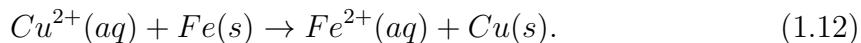
## 1.2 Electron/hole transfer in proteins

Electron transfer (ET) reactions are a fundamental class of naturally occurring chemical reactions that play central roles in various physical processes. The simplest ET reactions are isotopic self-exchange reactions[26, 27], where the reactants and prod-

ucts are the same except for an isotopic label, e.g.



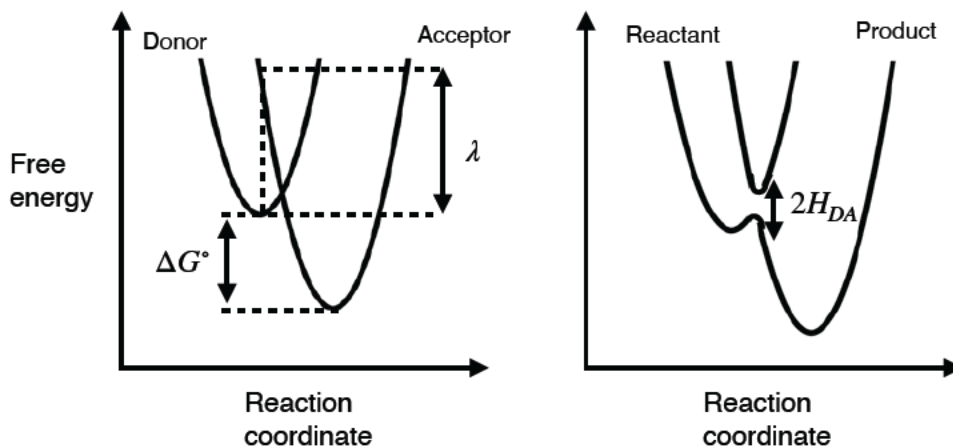
Far more common are cross-reactions[26, 27] occurring between different chemical species. Generations of general chemistry students are familiar with redox cross-reactions in aqueous solution, such as the reduction of  $Cu^{+2}$  ion by iron:



The theoretical foundation of ET theory comes from the work of Rudolph Marcus, for which he was awarded the Nobel Prize in 1992. Marcus Theory was originally formulated in the 1950s for ‘outer sphere’ electron transfer[28–31], where two chemical species change their oxidation state due to the transfer of an electron but do not undergo large structural change, and was later refined to include other types of transformations[32]. In the Marcus picture (Fig. 1.3), the free energy surfaces of the reactant and product state are represented as parabolas that evolve along a shared “reaction coordinate” that contains the nuclear coordinates of the entire system, including the solvent. In the semi-classical formulation[33], the ET rate depends on three reaction parameters: (i) the standard reaction free-energy change  $\Delta G^\circ$ , (2) the structural and solvent reorganization energy associated with the electron transfer  $\lambda$ , and (3) the strength of the electronic coupling between the donor and acceptor molecules  $H_{DA}$ . The Marcus rate constant is obtained as[33]:

$$k_{ET} = \sqrt{\frac{4\pi^3}{h^2\lambda k_B T}} H_{DA}^2 \exp\left[-\frac{(\Delta G^\circ + \lambda)^2}{4\lambda k_B T}\right] \quad (1.13)$$

where  $k_B$  is the Boltzmann constant, and  $h$  is the Planck constant.



**Figure 1.3:** Marcus free energy curves in the diabatic (left) and adiabatic (right) pictures. Eq. 1.13 holds in the non-adiabatic ( $H_{DA} \ll \Delta G^\circ, \lambda$ ) regime, and in an ET event, the system jumps from the donor curve to the acceptor curve.

The advent of Marcus Theory, along with advancements in structural biology[34], sparked significant interest in the study of ET in biological systems[27, 35, 36], and today, the fundamental importance of ET in life-sustaining processes such as cellular respiration and photosynthesis is well understood. ET in the photosynthetic and respiratory machinery often proceeds through long range electron tunneling between metal centers or other redox co-factors separated by 10-20 Å[27, 36, 37]. Over the years, long-range electron/hole hopping has been implicated in a variety of other biological processes, including DNA repair, magnetoreception in migratory birds, and protection against oxidative stress[38–43]. For the latter, Gray and Winkler have strongly emphasized the role of tryptophan (Trp) and tyrosine (Tyr) residues[35, 42, 43]. Their 2015 study of the protein structural database showed that approximately one third of all characterized proteins at the time contained Tyr/Trp chains composed of three or more residues[42]. Extended chains of aromatic residues have been shown to participate in electron/hole shuttling in various families of proteins including ribonucleotide reductases, photolyases and cryptochromes, and cytochrome c peroxidases[35, 44].

Experimental evidence on ET in biological systems comes from a variety of sources, including protein electrochemistry and crystallography experiments. A particularly useful technique is site-directed mutagenesis, which can be used to elucidate the importance of individual residues for ET pathways and mechanisms[36, 45]. The theoretical description of ET in biological systems is a challenging but feasible task with modern computational chemistry methods, provided that the key players are already known from experimental data or simpler theoretical models. For complex biological systems, the quantum mechanics/molecular mechanics (QM/MM) method with polarizable embedding has been suggested to be “gold standard” for the accurate calculation of free energies[46], which are needed to compute observables such as Marcus rate constants. QM/MM methods require partitioning the system into a QM part (typically, a chromophore and its local environment) that is treated quantum mechanically, and a MM part that is treated classically[47], and thus require some prior knowledge of the relevant residues and co-factors.

Targeted experimental and theoretical methodologies are greatly aided by simple theoretical models that can predict likely pathways which warrant further investigation. The development of eMap, a web application aimed at pre-screening probable ET pathways in proteins, is described in Chapter 6. The extension of eMap to the identification of shared pathways/motifs in families of proteins is discussed in Chapter 7.

## Chapter 2

# Theoretical Background

### 2.1 Electronic structure methods for excited states

Electronic structure theory methods are aimed at describing the quantum states of many-electron systems in the electrostatic field of the nuclei in atoms and molecules. Like any quantum system, electronic wave functions are governed by the TDSE, and the stationary solutions can be found by solving the TISE. Electronic structure calculations are most often performed within the Born-Oppenheimer (BO) approximation[48], where it is assumed that the electronic problem can be solved for a fixed nuclear configuration due to the fact that electrons are much lighter and move much faster than the nuclei. Accordingly, one solves the electronic TISE with the nuclei fixed at positions  $\mathbf{R}$ :

$$\hat{H}^e |\Psi_k^e(\mathbf{r}; \mathbf{R})\rangle = E_k^e(\mathbf{R}) |\Psi_k^e(\mathbf{r}; \mathbf{R})\rangle, \quad (2.1)$$

which is often referred to as the clamped-nuclei approximation. Here,  $\hat{H}^e$  is the electronic Hamiltonian,  $E_k^e$  are energies of the electronic states, and  $|\Psi_k^e(\mathbf{r}; \mathbf{R})\rangle$  are electronic wave-functions that depend on the nuclear coordinates  $\mathbf{R}$  only parametrically.

The electronic Hamiltonian includes the electron kinetic energy ( $\hat{T}^e$ ), electron-nuclear attraction ( $\hat{V}^{\text{en}}$ ), electron-electron repulsion ( $\hat{V}^{\text{ee}}$ ) operators and a nuclear repulsion term ( $V^{\text{nn}}$ ). The total *non-relativistic* molecular Hamiltonian is given below

in atomic units[49]:

$$\begin{aligned}
 \hat{H} &= \hat{T}^n + \hat{H}^e \\
 \hat{H}^e &= - \sum_i \frac{1}{2} \nabla_i^2 - \sum_{i,A} \frac{Z_A}{r_{iA}} + \sum_{i>j} \frac{1}{r_{ij}} + \sum_{B>A} \frac{Z_A Z_B}{R_{AB}} \\
 \hat{T}^n &= - \sum_A \frac{1}{2M_A} \nabla_A^2
 \end{aligned} \tag{2.2}$$

where  $\hat{T}^n$  is the nuclear kinetic energy operator. The indices  $i, j$  refer to electrons, the indices  $A, B$  refer to nuclei, and  $M, Z$  denote nuclear masses and charges, respectively. The  $\hat{T}^e$  and  $\hat{V}^{en}$  terms are often written as a sum of one-electron operators  $\hat{h}_i$ :

$$\hat{T}^e + \hat{V}^{en} = \sum_i^{N_e} \hat{h}_i. \tag{2.3}$$

The solution of the non-relativistic, electronic TISE gives the electronic wave functions  $|\Psi_k^e(\mathbf{r}; \mathbf{R})\rangle$ , and the electronic energies  $E_k^e(\mathbf{R})$ . Solving Eq. 2.1 for a range of nuclear configurations ( $\mathbf{R}$ ) gives the potential energy surface, whose topography provides insight into molecular geometries and chemical reaction dynamics. Below, we describe the approximate electronic structure methods used in this thesis to solve Eq. 2.1, whose solutions of are ultimately used in combination with the CAP method to represent resonance wave functions (see Sec. 2.2.1).

## The Hartree-Fock Method

The starting point for most of the many-body electronic structure methods used in this work is the Hartree-Fock (HF) method[49]. The electronic wave function is



approximated by a single Slater determinant:

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \cdots & \psi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & \cdots & \psi_N(\mathbf{x}_N) \end{vmatrix} \quad (2.4)$$

where  $\psi_i$  are spin-orbitals (called molecular orbitals, MOs, in the case of molecules),  $N$  is the number of electrons, and  $x_1, x_2, \dots, x_N$  denote spatial and spin coordinates of electrons 1, 2,  $\dots$ ,  $N$ . The orbitals are optimized according to the variational principle, which states that the expectation value of  $\hat{H}$  calculated with an approximate wave function is greater than or equal to the exact ground state energy:

$$E_0^{exact} \leq \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (2.5)$$

MOs  $|p\rangle$ , the single electron functions defined above, are expanded in terms of atomic orbital (AO) basis functions  $|\mu\rangle$ , which in conventional electronic structure calculations, are typically atom-centered Gaussians:

$$|p\rangle = \sum_{\mu} C_{\mu p} |\mu\rangle \quad (2.6)$$

where  $C_{\mu p}$  are elements of the MO-coefficient matrix. In the AO basis, the HF energy is written as:

$$E_{\text{HF}} = \frac{1}{2} \sum_{\mu\nu} D_{\mu\nu}^{\text{HF}} (h_{\mu\nu} + F_{\mu\nu}) + V_{\text{nn}}. \quad (2.7)$$

where the density matrix  $D^{HF}$  is derived from the MO coefficients:

$$D_{\mu\nu}^{\text{HF}} = \sum_i C_{\mu i}^* C_{\nu i}, \quad (2.8)$$

$h_{\mu\nu}$  are the one-electron integrals  $h_{\mu\nu} = \langle \mu | \hat{h}_1 | \nu \rangle$ , with  $\hat{h}_1$  defined in Eq. 2.3, and  $\mathbf{F}$  is the Fock matrix:

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\sigma\rho} D_{\sigma\rho}^{\text{HF}} \langle \mu\sigma || \nu\rho \rangle \quad (2.9)$$

which contains the antisymmetrized two-electron integrals  $\langle \mu\sigma || \nu\rho \rangle = \langle \mu\sigma | \frac{1}{|\mathbf{r}_{12}|} | \nu\rho \rangle - \langle \mu\sigma | \frac{1}{|\mathbf{r}_{12}|} | \rho\nu \rangle$ . In Eq. 2.9, we have used the so-called physicists notation:  $\langle \mu\sigma || \nu\rho \rangle$ ; in subsequent sections of this thesis, we will primarily use chemists notation:  $(\mu\nu || \sigma\rho)$ .

The two are linked by the following relation:

$$\langle \mu\sigma || \nu\rho \rangle = (\mu\nu || \sigma\rho). \quad (2.10)$$

The optimal MO coefficients that minimize the HF energy, under the constraint that the resulting MOs form an orthonormal set, are determined by iterative solution of the Roothan-Hall equations:

$$\mathbf{FC} = \mathbf{SC}\epsilon \quad (2.11)$$

where  $\mathbf{S}$  is the overlap matrix  $S_{\mu\nu} = \langle \mu | \nu \rangle$  and  $\epsilon$  is a diagonal matrix that contains the orbital energies, and  $\mathbf{C}$  is the MO-coefficient matrix whose elements are defined in Eq. 2.6.

## Electronic correlation

Since the Hartree-Fock energy approximates the total wave function as a single Slater determinant, according to the variational principle it only provides an upper bound to the total energy of a system. For a given one-electron basis set, the exact energy can be obtained from full configuration interaction (FCI), where the electronic wave

function is represented as a linear combination of all possible Slater determinants[49]:

$$\Psi_{FCI} = C_0\Phi_0 + \sum_{ia} C_i^a\Phi_i^a + \sum_{\substack{i<j \\ a<b}} C_{ij}^{ab}\Phi_{ij}^{ab} + \dots \quad (2.12)$$

where  $\Phi_0$  is the reference HF determinant. The MO indices  $i, j, \dots$  are used for occupied orbitals,  $a, b, \dots$  for virtual orbitals. Later in this section, the indices  $p, q, \dots$  are used to denote general orbitals, which can be either occupied or virtual. This notation will be used throughout the remainder of this section. The CI coefficients  $C_i^a, C_{ij}^{ab}$  etc. are obtained by diagonalization of the molecular Hamiltonian in the basis of all possible Slater determinants. For a system of  $N$  electrons and  $K$  spin orbitals, the number of Slater determinants can be computed according to the binomial equation:

$$\binom{2K}{N} = \frac{2K!}{N!(2K-N)!} \quad (2.13)$$

which quickly becomes intractable for anything but small molecules. As such, approximations to the FCI expansion are essential.

The *correlation energy* is defined as the difference between the exact FCI energy and the HF energy (calculated with a complete basis)[49]:

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}} \quad (2.14)$$

It is generally broken up into two components: **dynamic** correlation and **static** correlation[50], though we emphasize that the distinction is not well-defined, and that both are recovered by the inclusion of additional Slater determinants. Dynamic correlation refers to the failure of the HF model to account for the instantaneous interactions of electrons with each other, as the model treats every electron as inter-

acting with the average field created by all other electrons. Methods which recover dynamic correlation will be well suited for situations where the wave-function is dominated by a single Slater determinant. Static correlation refers to the situation where the electronic wave function is poorly represented by a single Slater determinant, and can only be accurately described as a linear combination of multiple (often nearly degenerate) Slater determinants. Methods which recover static correlation are essential in situations such as di-radicals and bond-breaking. In many cases, both types of correlation are needed to obtain qualitative and quantitative results.

### Reduced one-electron density matrices (1RDM)

In the many-body theories described below, the state and transition properties are often calculated as traces of the respective integrals with reduced density matrices[51, 52]. Here, we are focused on the one-particle density matrices:

$$\gamma^{IF}(x_1, x'_1) = N \int \dots \int \Psi_I^*(x_1, x_2, \dots, x_n) \Psi_F(x'_1, x_2, \dots, x_n) dx_2 \dots dx_n \quad (2.15)$$

where  $N$  is the number of electrons, and the labels  $I, F$  indicate the initial and final state. When  $I = F$ ,  $\gamma$  is a one-particle *state* density matrix, otherwise it is a one-particle *transition* density matrix between states  $I$  and  $F$ . In second quantization, this operator is written as:

$$\gamma^{IF}(x_1, x'_1) = \langle \Psi_I | \sum_{pq} a_p^\dagger a_q | \Psi_F \rangle \quad (2.16)$$

where the sum  $pq$  runs over all molecule orbitals, and  $a_p^\dagger, a_q$  represent creation and annihilation operators. With these quantities, the expectation value or transition

matrix element of any one-particle operator  $\hat{O}$  can be computed as follows:

$$\langle \Psi_I | \hat{O} | \Psi_F \rangle = \text{Tr} [\gamma^{IF} O] \quad (2.17)$$

### 2.1.1 Single reference wave function methods

Single reference wave function methods improve upon the HF solution by including a large number of *excited determinants*, which are generated by replacing MOs that are occupied in the reference determinant by MOs that are unoccupied. As such, they mostly recover dynamic correlation, and can yield very accurate energies when the target state is well described by a single reference determinant. Below, we review each of the single reference wave function methods utilized in this work.

#### EOM-CC

Coupled cluster (CC) is one of the most accurate electronic structure methods, and possesses desirable properties such as size-extensivity. The central principle of the CC method is to improve upon the HF solution by including all excitations of a given type (singles, doubles, etc.) to infinite order[50]. This is accomplished through the use of an exponential ansatz[53]

$$\Psi_{CC} = e^{\hat{T}} |\Phi_0\rangle \quad (2.18)$$

$$e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2} \hat{T}^2 + \dots = \sum_{k=0}^{\infty} \frac{1}{k!} \hat{T}^k$$

where  $\hat{T}$  is the cluster operator, and  $|\Phi_0\rangle$  is the reference HF determinant.  $\hat{T}$  is a sum of excitation operators,  $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots + \hat{T}_N$ , where

$$\begin{aligned}\hat{T}_1 |\Phi_0\rangle &= \sum_i^{occ} \sum_a^{vir} t_i^a |\Phi_i^a\rangle \\ \hat{T}_2 |\Phi_0\rangle &= \sum_{i<j}^{occ} \sum_{a<b}^{vir} t_{ij}^{ab} |\Phi_{ij}^{ab}\rangle\end{aligned}\tag{2.19}$$

Typically, the CC energy is not optimized variationally, and instead the reference and excited determinants are used to determine the amplitudes  $t$  of the cluster operator. The equations are commonly written using the *similarity transformed Hamiltonian*  $\bar{H}$ :

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}}\tag{2.20}$$

which preserves the spectrum of the original Hamiltonian, but is no longer Hermitian. The CC TISE is written as:

$$\hat{H} |\Psi_{CC}\rangle = E_{CC} |\Psi_{CC}\rangle\tag{2.21}$$

where  $E_{CC}$  is the coupled-cluster energy. Using  $\bar{H}$ , Eq. 2.21 can be written as:

$$\begin{aligned}e^{-\hat{T}} \hat{H} |\Psi_{CC}\rangle &= e^{-\hat{T}} E_{CC} |\Psi_{CC}\rangle \\ e^{-\hat{T}} \hat{H} e^{\hat{T}} |\Phi_0\rangle &= E_{CC} e^{-\hat{T}} e^{\hat{T}} |\Phi_0\rangle \\ \bar{H} |\Phi_0\rangle &= E_{CC} |\Phi_0\rangle\end{aligned}\tag{2.22}$$

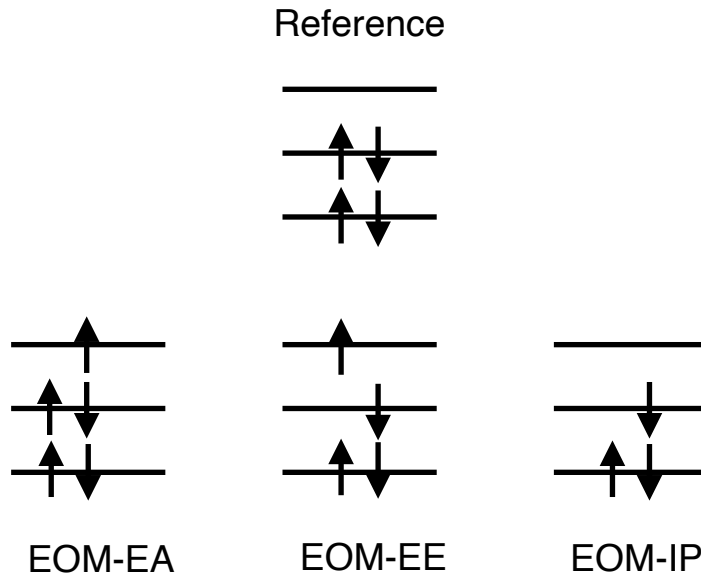
The expression for the ground state CC energy is obtained by projecting from the left onto the reference determinant.

$$E_{CC} = \langle \Phi_0 | e^{-\hat{T}} H e^{\hat{T}} | \Phi_0 \rangle = \langle \Phi_0 | \bar{H} | \Phi_0 \rangle \quad (2.23)$$

To determine the CC energy, one must solve for the expansion coefficients in Eq. 2.19, which are known as the  $t$ -amplitudes. This is done by multiplying Eq. 2.22 from the left by excited determinants:

$$\begin{aligned} \langle \Phi_i^a | \bar{H} | \Phi_0 \rangle &= 0 \\ \langle \Phi_{ij}^{ab} | \bar{H} | \Phi_0 \rangle &= 0 \end{aligned} \quad (2.24)$$

which yields a set of non-linear equations for the  $t$ -amplitudes that are solved iteratively. The coupled cluster singles and doubles method (CCSD) truncates the excitation operator at double excitations (i.e.  $\hat{T} = \hat{T}_1 + \hat{T}_2$ ) [54].



**Figure 2-1:** Selected variants of EOM-CC which generate different types of target states.

The equations of motion method (EOM-CC) extends the coupled cluster formalism to the calculation of excited or open-shell states (ionized, electron-attached, spin-flipped)[55, 56]. After solving for the  $t$ -amplitudes of the ground state, the target state wave function is parameterized as:

$$|\Psi_{EOM}\rangle = \hat{R}|\Psi_{CC}\rangle \quad (2.25)$$

where  $\hat{R}$  is a generalized excitation operator  $\hat{R} = \hat{R}_1 + \hat{R}_2 + \dots$ . Inserting  $\Psi_{EOM}$  into the TISE and using the fact that  $\hat{R}$  and  $\hat{T}$  commute leads to the following eigenvalue problem

$$\bar{H}\hat{R}|\Phi_0\rangle = E_{exc}\hat{R}|\Phi_0\rangle. \quad (2.26)$$

Target state energies and eigenvectors are found by diagonalizing the similarity transformed Hamiltonian  $\bar{H}$  in the basis of determinants generated by the  $\hat{R}$  operator acting on the reference determinant. The  $\hat{R}$  operator expansion is typically truncated at the same level as the coupled cluster  $\hat{T}$  operator, e.g. EOM-CCSD implies  $\hat{R} = \hat{R}_1 + \hat{R}_2$ .

Importantly, one can generate different types of target states depending on the choice of  $\hat{R}$ , including those with different number of electrons than the reference determinant (Fig. 2.1). Of particular importance for electronic resonances is the variant for electron affinities (EOM-EA), which can be used to calculate (N+1)-electron states, allowing for a correlated description of anionic excited states.

### **Algebraic Diagrammatic Construction**

The Algebraic Diagrammatic Construction(ADC) method for electronically excited states originates from the polarization propagator of many-body Green's function theory, which describes the time evolution of the polarization of a many-electron system[57]. The spectral representation of the polarization propagator expressed as



a matrix function is given as:

$$\begin{aligned} \Pi_{pq,rs}(\omega) = & \sum_{n \neq 0} \frac{\langle \psi_0 | c_q^\dagger c_p | \psi_n \rangle \langle \psi_n | c_r^\dagger c_s | \psi_0 \rangle}{\omega + E_0^N - E_n^N} \\ & + \sum_{n \neq 0} \frac{\langle \psi_0 | c_r^\dagger c_s | \psi_n \rangle \langle \psi_n | c_q^\dagger c_p | \psi_0 \rangle}{-\omega + E_0^N - E_n^N} \end{aligned} \quad (2.27)$$

where  $\psi_0$  is the ground state wave function of the molecule with energy  $E_0^N$ , and  $c_q^\dagger$  and  $c_p$  represent creation and annihilation operators typically associated with canonical HF orbitals. The summation is carried out over all electronically excited states with total energy  $E_n^N$ . The poles  $\omega_n = E_n^N - E_0^N$  correspond to vertical excitation energies, and the residues are transition probabilities of the corresponding excitation. In matrix notation, Eq. 2.27 can be written as:

$$\Pi(\omega) = \mathbf{x}^\dagger (\omega - \Omega)^{-1} \mathbf{x} \quad (2.28)$$

where  $\Omega$  is the diagonal matrix of vertical excitation energies  $\omega_n$  and  $\mathbf{x}$  is the matrix of transition amplitudes. A more general representation which is used for the derivation of approximate ADC schemes is written as:

$$\Pi(\omega) = \mathbf{f}^\dagger (\omega - \mathbf{M})^{-1} \mathbf{f} \quad (2.29)$$

where  $\mathbf{M}$  is a nondiagonal matrix of ‘effective’ interaction and  $\mathbf{f}$  is the matrix of effective transition moments. The  $\mathbf{M}$  and  $\mathbf{f}$  matrices are expanded according to the typical Møller-Plesset perturbation expansion:

$$\begin{aligned} \mathbf{M} &= \mathbf{M}^{(0)} + \mathbf{M}^{(1)} + \mathbf{M}^{(2)} + \dots \\ \mathbf{f} &= \mathbf{f}^{(0)} + \mathbf{f}^{(1)} + \mathbf{f}^{(2)} + \dots \end{aligned} \quad (2.30)$$

The expressions for the matrix elements of  $\mathbf{M}$  and  $\mathbf{f}$  are most conveniently derived using the so-called intermediate state representation[57, 58]. Starting from a correlated ground-state wave function  $\psi_0$ , a correlated excited-state basis  $\{\psi_J^0\}$  can be generated by the action of excitation operators  $\{\hat{C}_J\} \equiv \{\hat{c}_a^\dagger \hat{c}_k, \hat{c}_a^\dagger \hat{c}_b^\dagger c_k \hat{c}_l, \dots\}$  on a reference state  $\psi_0$ :

$$\psi_J^0 = \hat{C}_J \psi_0. \quad (2.31)$$

The correlated excited states  $\{\psi_J^0\}$  are not orthogonal in general, and are typically orthogonalized via Gram-Schmidt orthogonalization, which yields the orthogonal intermediate state basis  $\{\tilde{\psi}_J^0\}$ . The matrix elements of  $\mathbf{M}$  and  $\mathbf{f}$  can be expressed in the intermediate-state basis as follows:

$$(\mathbf{M})_{IJ} = \langle \tilde{\psi}_I | \hat{H} - E_0^N | \tilde{\psi}_J \rangle, \quad (2.32)$$

$$(\mathbf{f})_{J,pq} = \langle \tilde{\psi}_J | \hat{c}_p^\dagger \hat{c}_q | \psi_0 \rangle. \quad (2.33)$$

Now that explicit expressions for  $\mathbf{M}$  are available, the poles of the polarization propagator (and therefore, the excitation energies) can be obtained by diagonalization of  $\mathbf{M}$  at a desired order of perturbation theory. One solves the Hermitian eigenvalue problem:

$$\mathbf{M}\mathbf{Y} = \mathbf{Y}\Omega; \mathbf{Y}^\dagger \mathbf{Y} = 1 \quad (2.34)$$

which yields vertical excitation energies  $\omega_n$  and eigenvectors  $\mathbf{y}$ , which can be used to compute transition amplitudes  $x$ :

$$\mathbf{x} = \mathbf{y}^\dagger \mathbf{f}. \quad (2.35)$$

It is typical in ADC calculations to choose the  $n$ th order Møller-Plesset ground state as the starting point for generation of the intermediate-state basis, and this choice

defines the order of the ADC(n) approximation scheme. For example, the ADC(2) method starts from an MP2 ground state[57].

### CIS/CIS(D) methods

The configuration interaction singles (CIS) is one of the simplest excited state methods[59, 60]. Starting from a single reference HF determinant, the wave function is expanded as a linear combination of all possible singly excited determinants:

$$\Psi_{\text{CIS}} = \sum_{ia} c_i^a \Psi_i^a \quad (2.36)$$

The coefficients  $c_i^a$  are obtained by diagonalizing the Hamiltonian in the space of all single substitutions:

$$\mathbf{A}\mathbf{X} = \omega\mathbf{X} \quad (2.37)$$

where the matrix elements of  $\mathbf{A}$  are expressed as[60]:

$$A_{ia,jb} = (E_{HF} + \varepsilon_a - \varepsilon_i) \delta_{ij} \delta_{ab} + (ia||jb) \quad (2.38)$$

where  $E_{HF}$  is the ground state Hartree-Fock energy,  $\varepsilon$  are orbital energies, and  $(ia||jb)$  are two-electron integrals. The excitation energies are simply the difference between  $E_{HF}$  and the CIS excited state energies  $\omega$ .

CIS excitation energies often show a notable deviation from experiment. A common correction to CIS is the CIS(D) method, which approximately introduces effects of double excitations using a perturbative approach[61]. The CIS(D) energy correction for a given CIS excitation energy  $\omega$  is given by[61]:

$$\omega^{\text{CIS(D)}} = -\frac{1}{4} \sum_{ijab} \frac{(u_{ij}^{ab})^2}{\Delta_{ij}^{ab} - \omega} + \sum_{ia} b_i^a v_i^a \quad (2.39)$$

where

$$\begin{aligned}
\Delta_{ij}^{ab} &= \varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j \\
u_{ij}^{ab} &= \sum_c [(ab||cj)c_i^c - (ab||ci)c_j^c] + \sum_k [(ka||ij)c_k^b - (kb||ij)c_k^a], \\
v_i^a &= \frac{1}{2} \sum_{jkb} (jk||bc) (c_i^a a_{jk}^{ca} + c_j^a a_{ik}^{cb} + 2c_j^b a_{ik}^{ac})
\end{aligned} \tag{2.40}$$

and  $a_{ij}^{ab}$  are the MP2 amplitudes:

$$a_{ij}^{ab} = -\frac{(ij||ab)}{\Delta_{ij}^{ab}}. \tag{2.41}$$

Further improvements to the CIS(D) model include the spin-opposite scaled SOS-CIS(D) method[62], and the quasi-degenerate CIS( $D_0$ ) method[63] the latter which improves the description when multiple excited states are close in energy.

### 2.1.2 Multi-reference wave function methods

Multi-reference wave function methods generally involve two steps: an initial step which recovers static correlation, and a second step which recovers dynamic correlation. The former is necessary to obtain a qualitatively correct description of the wave function, the latter is necessary to obtain accurate quantities such as transition energies for excited states.

#### CASSCF

Similar to how the HF determinant is the starting point for single reference methods, the starting point for the multi-reference methods discussed in this thesis is the Complete Active Space Self Consistent Field (CASSCF) method[50, 64]. In CASSCF, a subset of the electrons and orbitals are chosen to define an ‘active space’, and the FCI

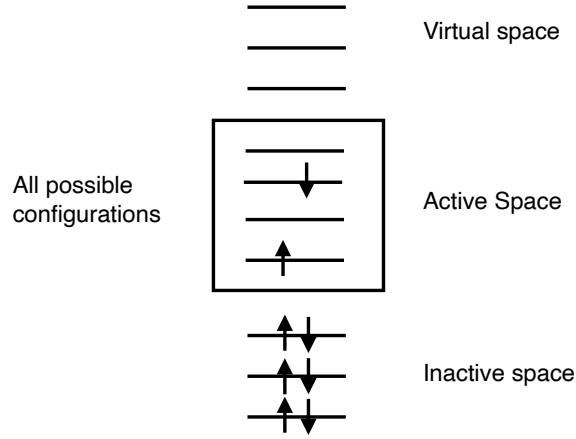
problem is solved exactly in the space defined by all possible configurations within the active space. The remaining occupied virtuals are either frozen or deemed *inactive*, and the remaining unoccupied orbitals are denoted as virtual (Fig. 2.2). The CASSCF wave function is written as:

$$|\Psi_{CASSCF}\rangle = \sum_I C_I |I\rangle \quad (2.42)$$

where  $\{|I\rangle\}$  are the complete set of Slater determinants defined by the choice of the active space, corresponding to all possible distributions of active electrons over active orbitals. The orbitals and CI coefficients  $C_I$  are variationally optimized to minimize the CASSCF energy. Simultaneous description of multiple states (e.g. ground and excited) can be incorporated by using state-averaged CASSCF (SA-CASSCF), which is aimed at achieving a balanced description of multiple states. In this case, optimization is performed for the state-averaged energy:

$$E = \sum_P w_P E_P \quad (2.43)$$

where  $w_P$  is the weighted coefficient for an individual state  $P$ . While the optimized orbitals are not necessarily optimal for each state individually, the SA-CASSCF approach has several advantages, including simplicity, a balanced description between ground and excited states, and avoiding convergence issues caused by root flipping (i.e. changes in the energetic ordering of states during optimization).



**Figure 2-2:** Schematic illustrating the orbital partitioning in the CASSCF method. The active space is CAS(2,4), which indicates 2 active electrons and 4 active orbitals.

### MS-CASPT2 methods

CASSCF methods mostly capture static correlation, which is insufficient to obtain accurate excitation energies. Here, we discuss a family of methods for recovering dynamic correlation known as multi-state complete active space second-order perturbation theory (MS-CASPT2)[65], which belongs to a broader class of methods known as multi-reference perturbation theory (MRPT)[66].

As usual in perturbation theory, the full Hamiltonian is partitioned into a zeroth-order part  $\hat{H}_0$  with known eigenfunctions, and a perturbation operator  $\hat{V} = \hat{H} - \hat{H}_0$ . The zeroth-order Hamiltonian in the MS-CASPT2 method is defined as[65]:

$$\hat{H}^{(0)} = \hat{P}\hat{f}\hat{P} + \hat{Q}\hat{f}\hat{Q} \quad (2.44)$$

where  $\hat{P} = \sum_N |N\rangle\langle N|$  is the projector onto the ‘reference space’  $\mathcal{P}$ , and  $\hat{Q} = 1 - \hat{P}$  is the projector onto the complementary space. The reference space is spanned by the set of CASSCF states  $\{|N\rangle\}$  which are included in the MS-CASPT2 calculation

and are linear combinations of the reference configurations included in the active space (Eq. 2.42).  $\hat{f}$  is the Fock operator (Eq. 2.50), which can be either state-specific or state-averaged, and this choice leads to different variants of MS-CASPT2, which will be discussed more in detail later on.

In internally contracted CASPT2, the first-order wave function for the reference state  $|N\rangle$  is parameterized as [67, 68]:

$$|\tilde{\Psi}_N\rangle = |N\rangle + |\Psi_N^{(1)}\rangle \quad (2.45)$$

$$\begin{aligned} |\Psi_N^{(1)}\rangle &= \sum_I t_N^I |I\rangle + \sum_{S,a} t_{a,N}^S |S^a\rangle \\ &+ \frac{1}{2} \sum_{ij,ab,M} t_{ab,NM}^{ij} |\Phi_{ij,M}^{ab}\rangle \end{aligned} \quad (2.46)$$

where  $|S^a\rangle, |\Phi_{ij,M}^{ab}\rangle$ , are internally contracted singly and doubly excited external configuration state functions (i.e. those involving occupation of orbitals outside of the active space) generated from the reference function  $|N\rangle$ . The amplitudes  $t$  are determined by the stationary condition of the Hylleraas functional:

$$\langle \Omega | \hat{H}^{(0)} - E_N^{(0)} + E_{\text{shift}} | \Psi_N^{(1)} \rangle + \langle \Omega | \hat{H} | \tilde{N} \rangle = 0 \quad (2.47)$$

where  $\Omega$  spans the manifold of  $|I\rangle, |S^a\rangle, |\Phi_{ij,M}^{ab}\rangle$ , and  $E_{\text{shift}}$  is a real or imaginary level shift, which is used to avoid intruder states [69, 70]. After solving for the amplitudes, the second order effective Hamiltonian is constructed as follows:

$$(H_{\text{eff}})_{MN} = \frac{1}{2} \left( \langle \tilde{\Psi}_M | \hat{H} | N \rangle + \langle M | \hat{H} | \tilde{\Psi}_N \rangle \right). \quad (2.48)$$

Diagonalization of  $H_{\text{eff}}$  gives the MS-CASPT2 energies, and the final wave functions:

$$|\Psi_N\rangle = \sum_M |\tilde{\Psi}_M\rangle T_{MN} \quad (2.49)$$

where  $T$  are the normalized eigenvectors of  $H_{\text{eff}}$ .

Returning to the zeroth order Hamiltonian, the Fock operator  $\hat{f}$  has the form:

$$\begin{aligned} \hat{f} &= \sum_{rs} f_{rs} \hat{E}_{rs} \\ f_{rs} &= h_{rs} + \sum_{ij} D_{ij} \left( J_{rs}^{ij} - \frac{1}{2} K_{rs}^{ij} \right) \end{aligned} \quad (2.50)$$

where  $h_{rs}$  are the one-electron integrals,  $J_{rs}^{ij}, K_{rs}^{ij}$  are two-electron integrals, and  $D_{ij}$  are elements of the one-particle density matrix. The one-particle density matrix  $\mathbf{D}$  can be state specific, or it can be a weighted average of the density matrices of the reference states, depending on the implementation[71, 72]:

$$\bar{\mathbf{D}}^\alpha = \sum_{\beta \in \mathcal{P}} \omega_\alpha^\beta \mathbf{D}^\beta. \quad (2.51)$$

In general, the Fock operator is not diagonal in the space of reference wavefunctions, but in the original formulation of MS-CASPT2, the elements  $\langle M | \hat{f} | N \rangle$  in Eq. 2.44 were arbitrarily set to zero. This *diagonal approximation* led to an issue known as non-invariance, which manifests as unphysical bumps on potential energy surfaces in the vicinity of conical intersections[73]. The solution to this problem is to use a set of rotated reference states  $\{|\tilde{N}\rangle\}$  that diagonalize the state-averaged Fock operator[71–74]:

$$\mathbf{U}^\dagger \mathbf{f}^{\text{SA}} \mathbf{U} = \tilde{\mathbf{f}} \quad (2.52)$$



$$|\tilde{N}\rangle = \sum_M |M\rangle U_{MN} \quad (2.53)$$

Methods that use rotated reference states are often referred to as ‘extended methods’, and Eqs. 2.44-2.48 proceed in exactly the same way, substituting  $\{|\tilde{N}\rangle\}$  for  $\{|N\rangle\}$ .

Finally, as mentioned previously, the zeroth order Hamiltonian (Eq. 2.44) can use either state-specific or state-averaged Fock operators, which defines different flavors of extended MS-CASPT2. The original extended XMS-CASPT2[71] variant uses the state averaged Fock operator (equal weights for each state in Eq. 2.51) for all states, which results in a single  $\hat{H}^{(0)}$ :

$$\hat{H}^{(0)} = \hat{P}\hat{f}^{SA}\hat{P} + \hat{Q}\hat{f}^{SA}\hat{Q}. \quad (2.54)$$

The use of state-specific Fock operators results in a separate partitioning for each rotated reference state:

$$\hat{H}_\alpha^{(0)} = \hat{P}\hat{f}^\alpha\hat{P} + \hat{Q}\hat{f}^\alpha\hat{Q} \quad (2.55)$$

The extended dynamically weighted (XDW)[72] and rotated (RMS)[74] variants differ in their construction of  $\hat{f}^\alpha$ . XDW-CASPT2 dynamically adjusts the weights of an averaged density matrix (Eq. 2.51) for each state in such a way that the Fock operator  $\hat{f}^\alpha$  for each state has the following property:

$$\hat{f}^\alpha \approx \begin{cases} \hat{f}^\alpha & \text{if } |\tilde{\alpha}\rangle \text{ weakly interacts with other model states} \\ \hat{f}^{sa} & \text{if } |\tilde{\alpha}\rangle \text{ strongly interacts with other model states} \end{cases}. \quad (2.56)$$

In this way, XDW-CASPT2 is meant to interpolate between the XMS-CASPT2 and the original MS-CASPT2, behaving more like the former near conical intersections, and more like the latter otherwise[72]. RMS-CASPT2 simply uses the state specific

Fock matrix  $\hat{f}^\alpha$  for each rotated reference state, and thus behaves like the original MS-CASPT2 but for rotated reference states[74].

### Multi-reference configuration interaction

The multi-reference configuration interaction (MRCI)[66, 75, 76] method is another approach which recovers dynamic correlation following a SA-CASSCF calculation. The MRCI wave-function is expanded in the basis of all configurations generated by considering excitations out of the reference CASSCF space. These excitations are typically truncated to singles or doubles, which results in the MR-CIS and MR-CISD methods. The MR-CISD wave function, for example, is written as[66, 76]:

$$\begin{aligned}
 |\Psi^{\text{MRCISD}}\rangle &= \sum_I c_I |I\rangle + \sum_I \sum_i \sum_a c_{ai}^I \hat{C}_{ai} |I\rangle \\
 &+ \sum_I \sum_{ij} \sum_{ab} c_{ai,bj}^I \hat{C}_{ai,bj} |I\rangle
 \end{aligned}
 \tag{2.57}$$

where  $\hat{C}$  are the usual excitation operators and  $|I\rangle$  are Slater determinants, typically those which make up the reference space.

Truncated MRCI methods are not size-consistent ( $E_{AB}(r = \infty) = E_A + E_B$ ) and are not size-extensive (correct scaling with the number of electrons). These errors need to be corrected in order to obtain a high level of accuracy. There are several correction schemes, the most common being the *a posteriori* Davidson correction[77]:

$$\Delta E_Q = (1 - c_0^2) (E_{\text{CISD}} - E_{\text{HF}})
 \tag{2.58}$$

where  $c_0$  is the coefficient of the HF determinant in the CISD expansion. The Davidson correction estimates the energy up to quadruple excitations (CISDTQ) from the energy of MR-CISD, and as such the results are frequently referred to as

MR-CISD(Q). While the Davidson correction improves both the size consistency and size extensivity of MRCI methods, the MR-CISD(Q) energies are not formally size-consistent or size-extensive.

### 2.1.3 Density Functional Theory

Density functional theory (DFT) is an alternative framework to the wave function based methods described previously, which uses the electron density as the basic quantity to characterize the electronic structure of a system[50, 78]. Due to its superior computational scaling, DFT has become the method of choice for much of the quantum chemistry community, as it enables practical calculations on large systems.

The electron density  $\rho(r)$  is a physical observable that describes the distribution of electrons in a given system:

$$\rho(\mathbf{r}) = N \int d^3\mathbf{r}_2 \cdots \int d^3\mathbf{r}_N \Psi^*(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \Psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (2.59)$$

where  $N$  is the total number of electrons. The theoretical foundation for DFT was provided by Hohenberg-Kohn theorems[79], which demonstrated the relationship between electron density and the ground-state properties of a system. The first theorem shows that the electron density  $\rho(\mathbf{r})$  of a system of interacting particles uniquely determines the external potential  $v_{ext}(r)$  (i.e. the  $\hat{V}^{en}$  term in Eq. 2.2), and therefore, the total energy. The second provides a variational principle, and shows that the exact ground state density minimizes the total energy.

While these theorems do not provide a way of determining the ground state density, the Kohn-Sham (KS) formalism offers a practical foundation upon which modern DFT implementations are built[80]. KS-DFT introduces a fictitious system of non-interacting electrons that generate the same density as the real system. The wave

function is a Slater determinant of Kohn-Sham orbitals  $\varphi_i$ , and the total energy is expressed as a functional of the density:

$$E[\rho] = T_s[\rho] + \int d\mathbf{r} v_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) + E_C[\rho] + E_{xc}[\rho] \quad (2.60)$$

Here,  $T_s[\rho]$  is the Kohn–Sham kinetic energy

$$T_s[\rho] = \sum_{i=1}^N \int d\mathbf{r} \varphi_i^*(\mathbf{r}) \left( -\frac{\hbar^2}{2m} \nabla^2 \right) \varphi_i(\mathbf{r}), \quad (2.61)$$

$v_{\text{ext}}$  is the external potential,  $E_C[\rho]$  is the Coulomb self-interaction of the electron density

$$E_C[\rho] = \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}', \quad (2.62)$$

and  $E_{xc}[\rho]$  is the *exchange correlation* functional, which is unknown.  $E_{xc}[\rho]$  is meant to recover the remaining electronic energy not accounted for by the non-interacting kinetic energy and Coulomb terms. Common approximations to  $E_{xc}[\rho]$  are briefly described in Sec. 2.1.3. Finally, the density that minimizes the total energy is expressed in terms of Kohn-Sham orbitals  $\varphi_i$ :

$$\rho(\mathbf{r}) = \sum_i^N |\varphi_i(\mathbf{r})|^2 \quad (2.63)$$

which are obtained through iterative solution of the Kohn-Sham equations:

$$\left( -\frac{\hbar^2}{2m} \nabla^2 + v_{\text{eff}}(\mathbf{r}) \right) \varphi_i(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}) \quad (2.64)$$

where  $v_{\text{eff}}$  is the Kohn-Sham potential in which the fictitious non-interacting particles

move:

$$v_{\text{eff}}(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})}. \quad (2.65)$$

## Time-dependent density functional theory

The time-dependent extension of DFT, TDDFT, offers a means of calculating the excited states of a system[81–83]. It is based on the response of the ground state density to a time-varying applied electric field, where the poles of the linear response function correspond to excitation energies. The theoretical foundation for TDDFT is provided by the Runge-Gross Theorem[84], which shows that there exists a unique mapping between the time-dependent external potential  $v_{\text{ext}}(r, t)$  of a system and its time-dependent density  $\rho(r, t)$ .

In the linear response formulation, the excitation energies  $\omega$  are found by solving the following generalized eigenvalue equation:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \omega \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \quad (2.66)$$

The matrices A and B are defined as:

$$\begin{aligned} A_{ia,jb} &= \delta_{ij} \delta_{ab} (\epsilon_a - \epsilon_i) + (ia||jb) + (ia|f_{\text{xc}}|jb) \\ B_{ia,jb} &= (ia||bj) + (ia|f_{\text{xc}}|bj) \end{aligned} \quad (2.67)$$

where  $i, j, a, b$  are Kohn-Sham orbital indices,  $\epsilon$  are Kohn-Sham orbital energies,  $(ia||bj)$  are two-electron integrals, and  $f_{\text{xc}}$  is the exchange correlation potential. The

integral  $(ia | f_{xc} | bj)$  is given by[83]:

$$(ia | f_{xc} | bj) = \int d^3r d^3r' \varphi_i^*(r) \varphi_a(r) \frac{\delta^2 E_{xc}}{\delta \rho(r) \delta \rho(r')} \varphi_b^*(r') \varphi_j(r') \quad (2.68)$$

where  $\varphi$  are Kohn-Sham orbitals.

$X$  and  $Y$  above refer to the excitation and de-excitation transition vectors respectively. In the Tamm-Dancoff approximation (TDA)[85–87], the de-excitation amplitudes  $Y$  are neglected, and the  $B$  matrix not needed, which reduces the problem to a Hermitian eigenvalue problem:

$$\mathbf{A}\mathbf{X} = \omega\mathbf{X}. \quad (2.69)$$

Compared to other excited state methods, TDDFT has a computational cost roughly similar to that of CIS, but can offer (depending on the functional) an improved description compared to CIS due to electronic correlation implicitly included in the exchange correlation functional.

### Exchange correlation

As mentioned previously, the true form of the ‘exchange correlation’ functional is unknown, and must be approximated. Hundreds of functionals have been developed, and they are loosely arranged into one of 5 rungs on ‘Jacob’s ladder’ of functionals, where each higher rung of the ladder represents additional complexity and (hopefully but not necessarily) a step closer towards chemical accuracy[88]. The simplest approximation, and lowest rung on Jacob’s ladder, is the local density approximation (LDA), which typically applies the exchange and correlation energy expressions derived for a uniform electron gas. Other approximations include the generalized

gradient (GGA) and random phase (RPA) approximations.

Exchange correlation functionals can generally be decomposed into exchange and correlation terms:

$$E_{xc} = E_x + E_c \quad (2.70)$$

which has led to extensive mixing and matching of *ab initio* and empirically derived approaches. An important class of functionals which are used in this thesis are hybrid functionals, which incorporate some Hartree-Fock or ‘exact’ exchange

$$E_x^{\text{HF}} = -\frac{1}{2} \sum_{i,j} \iint \psi_i^*(\mathbf{r}_1) \psi_j^*(\mathbf{r}_2) \frac{1}{r_{12}} \psi_j(\mathbf{r}_1) \psi_i(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2. \quad (2.71)$$

For example, the popular PBE0 functional[89], is defined as follows:

$$E_{xc}^{\text{PBE0}} = \frac{1}{4} E_x^{\text{HF}} + \frac{3}{4} E_x^{\text{PBE}} + E_c^{\text{PBE}} \quad (2.72)$$

where  $E_x^{\text{HF}}$  is the exact exchange functional,  $E_x^{\text{PBE}}$  is the Perdew–Burke–Ernzerhof (PBE) exchange functional, and  $E_c^{\text{PBE}}$  is the PBE correlation functional[90]. An important category of hybrid functionals are those which include long-range corrections to account for the non-Coulomb part of the exchange dying off too quickly. Examples include the CAM-B3LYP functional[91], and the long-range correction of Hirao and coworkers[92] which can be combined with many pure correlation functionals.

## 2.2 Non-Hermitian description of resonances

In section 1.1.2, we introduced the Siegert formulation of a resonance as a discrete state with a complex energy[6]:

$$E_{res} = E_R - i\Gamma/2.$$

In this framework, a resonance appears as a single solution of the TISE, and the TDSE does not need to be solved. While the usefulness of this description can easily be seen from Eq. 1.10, how one calculates such a state is much less obvious, as the immediate question arises of how the Hamiltonian, a Hermitian operator, could admit a complex eigenvalue. For practitioners of electronic structure theory, the description of a resonance as a single square-integrable state with a complex energy is particularly promising, as one can attempt to extend the methods for bound states to the treatment of resonances.

The theoretical approaches for resonances can be classified into two major groups: (i) those that account for the scattering part of the resonance wave function explicitly, and (ii) those that operate within the  $L^2$  representation, and thus account for the continuum implicitly. Examples of the former include R-matrix theory[93] and the Schwinger multichannel method[94], and examples of the latter include the stabilization method[95–97], Stieltjes imaging technique[98, 99], and Non-Hermitian quantum mechanics (NHQM) formalisms which we describe more in detail below.

NHQM methods involve transformation of the resonance wave function into a square-integrable form[15]. One of the most rigorous approaches is the complex coordinate or direct complex scaling (CS) method, which analytically continues the Hamiltonian into the complex plane by scaling the coordinates by a complex number



$r \rightarrow re^{i\theta}$ [4, 100]. The complex scaled Hamiltonian retains the same bound spectrum as the original Hamiltonian, and continuum states are rotated into the lower complex plane by an angle of  $2\theta$ . As  $\theta$  is increased from 0 to  $\pi/2$ , new discrete eigenvalues may appear at the complex Siegert energies of resonance states, and correspond to square-integrable eigenfunctions[4].

The application of the complex coordinate method to molecular systems is not straightforward due to complications which arise from the BO approximation[15]. This has led to the development of alternative approaches that can be applied to molecules, such as the complex basis functions method[101, 102], and exterior complex scaling[103]. The complex absorbing potential (CAP) method[104], described in Sec. 2.2.1, has been shown to be related to the latter[105, 106], and shares some important properties outlined below.

Due to the non-Hermiticity of the CS Hamiltonian, the usual Hermitian inner product is replaced by the c-product[15, 107]

$$(\Phi_i | \Phi_j) = \int \Phi_i(\tau)\Phi_j(\tau)d\tau \quad (2.73)$$

which is typically denoted by the use of round brackets, which indicates the lack of complex conjugation on the bra vector. We caution the reader not to confuse the c-product with the chemists notation for two-electron integrals (Eq. 2.10). When there are no degeneracies, normalization under the c-product generates a complete set of orthogonal eigenfunctions for the CS/CAP Hamiltonian. When degeneracies are present, the Hamiltonian becomes defective due to the self-orthogonality phenomenon[15]. Such non-Hermitian degeneracies are known as exceptional points[108], and are briefly touched upon in Sec. 4. Lastly, for a c-normalizable trial wave func-

tion, there is an analogous complex variational principle[107]:

$$\bar{E} = \frac{(\Phi | \hat{H}_{cx} | \Phi)}{(\Phi | \Phi)} \quad (2.74)$$

where  $H_{cx}$  indicates a complex scaled or CAP Hamiltonian, which justifies the use of standard electronic structure methods for resonances. Unlike the standard variational principle (Eq. 2.5),  $\bar{E}$  is not an upper bound, and is instead a stationary approximation to the true Siegert eigenvalue  $E_{res}$ .

### 2.2.1 Complex Absorbing Potential

In the CAP method[104], the physical Hamiltonian is augmented with a negative, complex potential which is intended to absorb the outgoing tail of resonance wave functions. The non-Hermitian, complex-symmetric CAP-augmented Hamiltonian is written as:

$$\hat{H}(\eta) = \hat{H} - i\eta W \quad (2.75)$$

where  $\eta$  is the CAP strength parameter, and  $W$  is typically a real and positive potential which vanishes near the system (e.g. an isolated molecule) and grows with distance. The presence of the absorbing potential changes the boundary conditions, and transforms a resonance into a square integrable eigenstate of the CAP-Hamiltonian, associated with the proper complex Siegert eigenvalue.

The ancestry of CAPs can be traced to optical potentials[109], but the first appearance of CAPs as we know them occurred in time-dependent calculations, where they were introduced by Kosloff and Kosloff[110] as tools to reduce reflections from the edge of the grid in wave packet propagation simulations. CAPs were first used for computing resonance states by Jolicard and Austin[111], who also introduced key

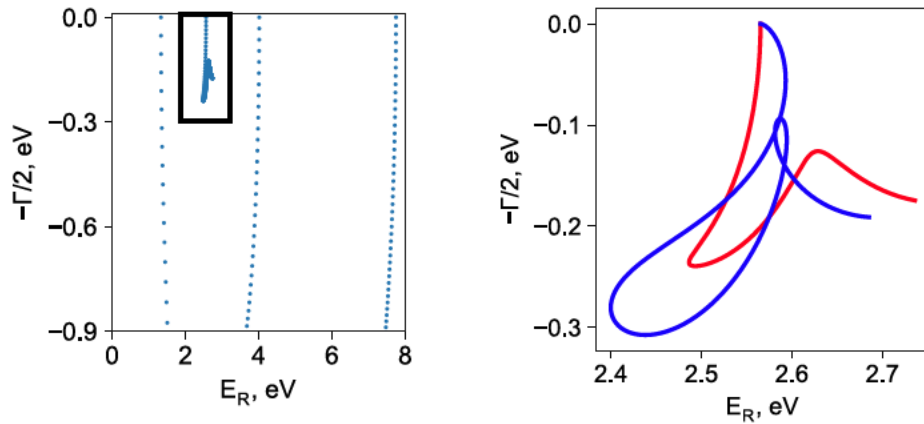
ideas such as varying the strength or location of the potential in order to minimize the perturbations caused by the presence of the artificial potential. These concepts were refined in the seminal work by Riss and Meyer[104], which provided a mathematical foundation for the CAP method, described the requirements for the form of the CAP that produces discrete resonance states, and established many of the practical aspects of performing CAP calculations that are still in use today.

As shown by Riss and Meyer[104], with a complete basis set, the exact resonance position and width are obtained in the limit of an infinitesimal CAP strength  $\eta \rightarrow 0^+$ , since there is no perturbation to the system from the CAP. In finite basis set calculations,  $\eta = 0$  simply yields a real eigenvalue of the original physical Hamiltonian, and offers little useful information about resonance states. Thus, in practical calculations with finite basis sets,  $E_{res}$  must either be obtained from a finite CAP strength, or by performing analytic continuation to the  $\eta \rightarrow 0^+$  limit, which can be done, for example, using Padé approximants[112]. A formula for identifying a finite optimal value of  $\eta$  was proposed by Riss and Meyer[104]:

$$\eta_{opt} = \min \left| \frac{\partial E}{\partial \ln(\eta)} \right| = \min \left| \eta \frac{\partial E}{\partial \eta} \right|. \quad (2.76)$$

This lowest “speed” in the logarithmic velocity minimizes the first-order term in a Taylor series expansion of the energy  $E(\eta)$ , and strikes a balance between minimizing the perturbation due to the CAP (which increases with the CAP strength) and minimizing the basis set error (which increases when the eigenfunction  $\psi(\eta)$  is more delocalized, i.e. smaller values of  $\eta$ ). Unfortunately,  $\eta_{opt}$  varies with the system, basis set, and form of CAP, and cannot be determined *a priori*. The typical strategy is to perform a series of at least 50-100 calculations to generate the so-called  $\eta$ -trajectories(Fig. 2·3). Trajectories which correspond to resonances can be dis-

tinguished from discretized continuum states by their slow variation with the CAP strength and pronounced stabilization behavior near  $\eta_{opt}$ . The calculated trajectories can vary quite widely in their structure, and interpretation can become challenging when there are multiple points which appear to satisfy Eq. 2.76, or none at all. These issues become magnified when accurate resonance energies and widths are not reported in the literature[113].



**Figure 2.3:**  $\eta$ -trajectories for the  ${}^2\Pi_g$  resonance in  $N_2^-$  computed using projected CAP-EOM-EA-CCSD[114]. The left panel shows the eigenvalue spectrum for four states, and the right panel shows the uncorrected(red) and corrected (blue) trajectories for the resonance state.

The majority of CAP practitioners (especially in more recent studies) apply correction schemes in order to diminish the artifacts introduced by the CAP, typically up to the first order. The first-order corrected energy proposed by Riss and Meyer has the form[104]:

$$U(\eta) = E(\eta) - \eta \frac{dE}{d\eta} \quad (2.77)$$

The derivative  $\frac{dE}{d\eta}$  can be recast as the expectation value of the CAP by employing the generalized Hellman-Feynman theorem[104, 115]:

$$\frac{dE}{d\eta} = -i(\psi(\eta)|W|\psi(\eta)) \quad (2.78)$$

which has the simple physical interpretation of  $U(\eta)$  removing the contribution of the CAP to the energy to first order in perturbation theory. Eq. 2.77 can equivalently be expressed in terms of the reduced one-particle density matrix  $\gamma$ [116]:

$$U(\eta) = E(\eta) + i\eta \text{Tr}[\gamma W] \quad (2.79)$$

which can eliminate the need for numerical differentiation. Multiple schemes have been proposed for the analysis of corrected trajectories. A common strategy is to identify the minimum of the logarithmic velocity on the corrected trajectory[117]:

$$\min \left| \eta^2 \frac{d^2 E}{d\eta^2} \right| = \min \left| \eta \frac{dU}{d\eta} \right| \quad (2.80)$$

Alternatively, the real and imaginary parts of the corrected energies can be analyzed independently, resulting in two separate values of  $\eta_{opt}$ [116]. The use of correction schemes is not only physically justified, but also has significant practical benefits. First order results have been shown to be less dependent on the choice of CAP and yield more reliable complex potential energy surfaces[114, 118–120]. Additionally, in many (but not all) cases, the corrected trajectories are easier to interpret than the raw  $E(\eta)$  trajectories because they do not suffer as much from spurious or poorly stabilized stationary points.

The relative simplicity of the CAP method has made it a popular choice for computing resonance parameters, and has resulted in its combination with numerous electronic structure methods, including configuration interaction[121–124], symmetry-adapted cluster-configuration interaction[125], Fock-space coupled cluster[126], density functional theory [127], algebraic diagrammatic construction[128, 129], and multi-reference perturbation theory[130–132]. Unfortunately, very few of these studies have

resulted in publicly available software implementations, a problem which we have addressed in Chapter 5 with the development of the OpenCAP package.

### 2.2.2 Forms of CAP

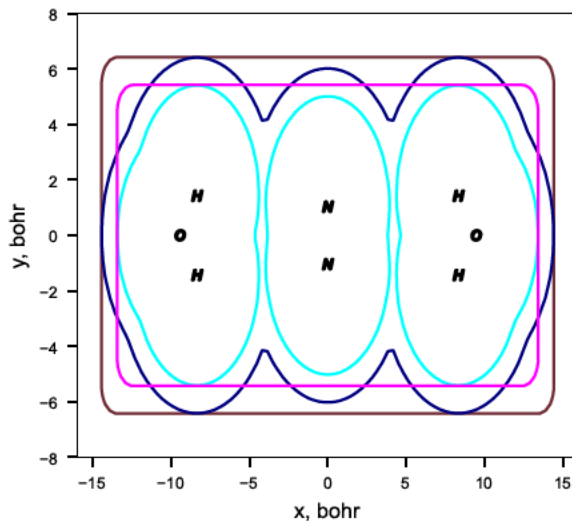
In their seminal work on the CAP method[104], Riss and Meyer laid out requirements for  $W(r)$  which ensure that the eigenvalues of  $H(\eta)$  converge towards the true Siegert energy in the limit  $\eta \rightarrow 0^+$  with a complete basis set. These requirements allow for a fairly broad family of CAPs, and early studies employed simple functional forms such as quadratic and cubic potentials centered about the center of mass of the molecular system. In more recent years, the community has largely settled on two major types of the CAP, the box and smooth Voronoi CAP, one or both of which are used in nearly every publication on molecular resonances today (see Fig. 2.4). Below, we provide details of their functional form. The applicability of performance of these two CAPs for different types of molecular shape resonances is investigated in Chapter 3.

The quadratic “box CAP”, is defined by three cutoff parameters ( $R_x^0, R_y^0, R_z^0$ ) that specify the onset of the box in each Cartesian coordinate [8, 133].

$$W = W_x + W_y + W_z$$

$$W_\alpha = \begin{cases} 0, & |r_\alpha| < R_\alpha^0 \\ (r_\alpha - R_\alpha^0)^2, & |r_\alpha| > R_\alpha^0 \end{cases} \quad (2.81)$$

where  $\alpha = x, y,$  or  $z$ . This convenient functional form allows for the one-electron CAP integrals [133] to be computed analytically, and analytical gradients [134] for CAP-EOM-CCSD calculations which use a box CAP have also been reported. However, its rigid structure makes it less suitable for molecular systems which do not efficiently fill the space.



**Figure 2.4:** Comparison of different box and Voronoi CAPs for a model  $N_2^-(H_2O)_2$  cluster. Plotted is the isoline  $W=1$  hartree for box CAPs of 4 bohr (brown) and 3 bohr (magenta), and smooth Voronoi CAPs of 4 bohr (navy) and 3 bohr (cyan) in the plane of the molecular system. This image is reproduced with permission from Ref. [114]. Copyright 2022 American Chemical Society.

The smooth Voronoi CAP was proposed by Sommerfeld and Ehara [135, 136]. It wraps uniformly around any molecule, and is specified by a single cutoff radius  $r_{cut}$ . The edges between Voronoi cells are smoothed out in order to make the CAP more amenable to numerical integration. Its functional form is given as:

$$W(\vec{r}) = \begin{cases} 0, & r_{WA} \leq r_{cut} \\ (r_{WA}(\vec{r}) - r_{cut})^2, & r_{WA} > r_{cut} \end{cases} \quad (2.82)$$

$r_{WA}(\vec{r})$  in Eq. 2.82 is the effective distance to the molecular system evaluated as a weighted average of the distances to all nuclei:

$$r_{WA}(\vec{r}) = \sqrt{\frac{\sum_i w_i |\vec{r} - \vec{R}_i|^2}{\sum_i w_i}} \quad (2.83)$$

$R_i$  is the distance to the  $i^{\text{th}}$  nucleus and  $w_i$  is the weight defined as follows:

$$w_i = \frac{1}{(|\vec{r} - \vec{R}_i|^2 - r_{min}^2 + 1 \text{ a.u.})^2} \quad (2.84)$$

where  $r_{min}$  is the distance to the closest atom:  $r_{min} = \min_i |\vec{r} - \vec{R}_i|$ . The smooth Voronoi CAP, by definition, shares exactly the symmetry of the molecular system, and is flexible to changing nuclear configurations, for example in dynamical simulation. The immediate price one pays for this flexibility is a less convenient functional form; the one-electron CAP integrals must be evaluated numerically, and there are currently no reported implementations of analytical gradients for metastable states which use a Voronoi CAP.

### 2.2.3 Projected CAP scheme

Since the CAP is simply a one-electron operator, it is conceptually straightforward to calculate and incorporate into electronic structure calculations. Depending on the implementation, the most challenging task is often the adaptation of standard machineries such as eigensolvers (e.g., Davidson) and optimization procedures (e.g., DIIS) to handle the complex algebra with the c-product. Further, the absorbing potential is generally intended to act on the excited states of a system and, in principle, should not be affecting the description of the ground state. These considerations have resulted in multiple strategies employed by different authors for including CAPs in their calculations. For example, for the EOM-CC method, there are reported implementations of CAP where it is applied at the HF level [119], the CC level [137, 138], the EOM level [138, 139], and *a posteriori*, where the CAP is applied to a small number of real EOM-CC eigenstates of the physical Hamiltonian [120]. The latter is by far the most efficient, since it requires only a single EOM-CC calculation to generate



$\eta$ -trajectories, and is known as the “projected scheme”.

The projected scheme was introduced by Sommerfeld and Santra [122] for CAP/CI calculations, and has been adapted for various electronic structure methods, including MRPT,[130, 132], ADC [129], MRCI, and symmetry adapted cluster-configuration interaction[125]. The projected scheme can alternately be viewed as a “subspace scheme”, where the CAP is represented on a subspace spanned by a small set of eigenvectors, or as a quasi-degenerate perturbative scheme, where those eigenvectors are allowed to mix under the influence of the absorbing potential. In any case, the first step of a projected CAP calculation is to generate a set of  $M$  target states which comprise the “correlated basis” or “subspace”. While the exact number of states needed is generally not known ahead of time, practical experience has shown that relatively few (typically less than 10-15) are needed to converge low lying resonances. The next step is to compute the CAP matrix in orbital basis (which can be done independently of the electronic structure calculation), and then project it onto the space spanned by the correlated basis (CB):

$$W_{ij}^{CB} = (i|W|j) \quad (2.85)$$

where  $|i\rangle, |j\rangle$  are eigenstates of the physical Hamiltonian. These expressions can easily be evaluated using the one-electron reduced density matrices ( $\rho$ ) for each state, and the set of one-electron transition density matrices ( $\gamma$ ) between each pair of states obtained from the bound state calculation:

$$W_{ij}^{CB} = \begin{cases} Tr [W\gamma^{ij}], & i \neq j \\ Tr [W\rho^i], & i = j \end{cases} \quad (2.86)$$

Finally, the last step is to diagonalize the CAP-augmented Hamiltonian in the space

spanned by the correlated basis:

$$H^{CAP} = H_0^{CB} - i\eta W^{CB} \quad (2.87)$$

over a range of values of  $\eta$  in order to generate the  $\eta$ -trajectory and identify the  $\eta_{opt}$ .  $H_0^{CB}$  in Eq. 2.87 is an appropriate zero-order Hamiltonian for the chosen correlated basis, e.g. the diagonalized similarity transformed Hamiltonian from CAP-EOM-CCSD or the second order effective Hamiltonian in CAP-XMS-CASPT2[130].

To summarize, the basic ingredients of any projected CAP calculation are:

- Choice of  $M$ -dimensional correlated basis set and zero-order Hamiltonian
- Complete set of reduced one-particle density matrices and transition density matrices between each pair of basis states
- CAP matrix projected onto correlated basis set
- Diagonalization of Eq. 2.87 over a range of  $\eta$ -values to identify  $\eta_{opt}$

In Chapter 4, the implementation of the projected CAP-EOM-CCSD method is described, and the results of benchmark calculations for various shape resonances is presented. In Chapter 5, we describe the implementation of the OpenCAP package, which utilizes the projected CAP scheme to extend the functionality of popular electronic structure packages to the description of molecular resonances, and present the results of benchmark calculations using a variety of electronic structure methods.

## 2.3 Graph Theory

Below, we switch gears to discuss important concepts pertaining to the models used to identify and predict ET pathways in proteins.

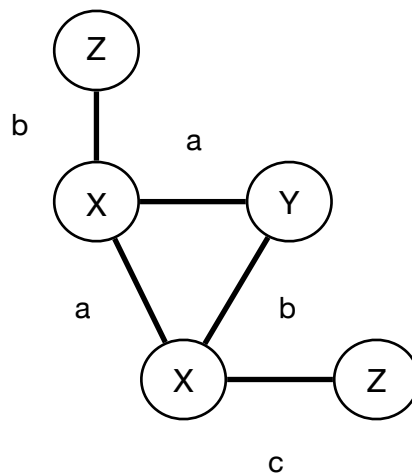
In mathematics and computer science, graphs are mathematical structures used to represent complex, structured data. Below, we establish some key definitions which are commonly used in graph theory[140].

## Graph

A graph  $G(N,E)$  is an ordered pair consisting of two sets:

- $N$  is a set of *vertices* or *nodes*
- $E$  is a set of *edges*, which connect pairs of nodes

In *labeled* graphs, the nodes and/or edges are assigned distinguishable labels. If the edges represent ordered pairs of nodes, the graph is *directed*, otherwise it is an *undirected* graph. One can also assign numerical values to each edge in the graph, in which case the graph is considered to be *weighted*. Otherwise, it is considered an *unweighted* graph. A *tree* is a special type of undirected graph in which any two vertices are connected by exactly one path, i.e. the graph contains no cycles.



**Figure 2·5:** A graph with node labels  $\{X, Y, Z\}$  and edge labels  $\{a, b, c\}$ .

## Subgraph

If  $G'=(N',E')$  is a subgraph of  $G(N,E)$

- $N'$  is a subset of  $N$
- $E'$  is a subset of  $E$

## Isomorphism

Two graphs  $G(N, E)$  and  $G' = (N', E')$  are isomorphic if and only if there exists a bijection  $\varphi$  (one-to-one correspondence) from  $N$  to  $N'$  such that:

$$\{u, v\} \in E \Leftrightarrow \{\varphi(u), \varphi(v)\} \in E'$$

where  $u, v \in N$ . In other words, two graphs are isomorphic if all nodes and edges are preserved, and there is a one-to-one correspondence between the respective node/edge labels.

## Subgraph isomorphism

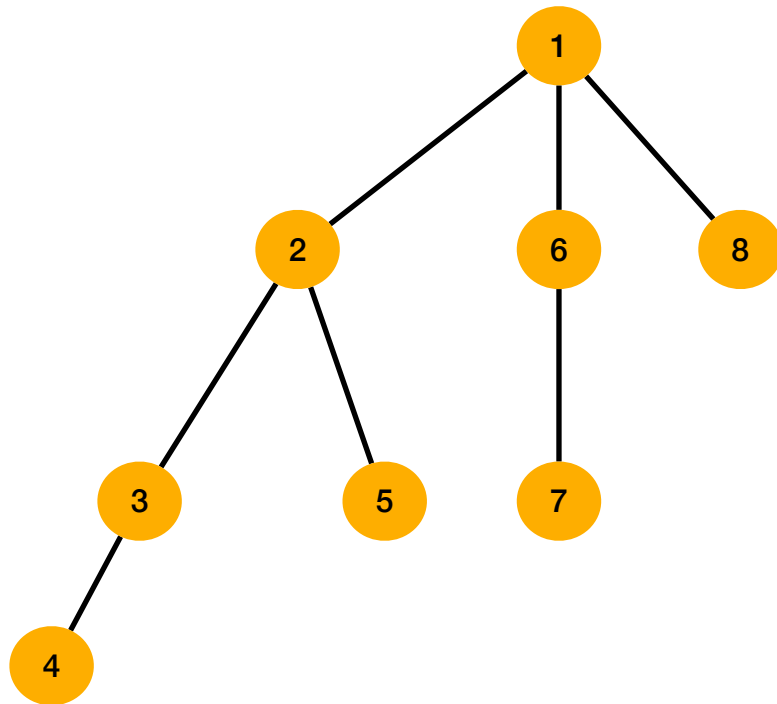
For two graphs  $G(V, E)$  and  $H(V', E')$ ,  $G$  and  $H$  are subgraph isomorphic if there exists a  $G'(V_0, E_0)$  such that[\[141, 142\]](#):

- $G'$  is a subgraph of  $G$
- $G'$  is isomorphic to  $H$

The task of identifying all such  $G'$  is known as the *subgraph matching* problem, and we refer to individual  $G'$  as a subgraph isomorphisms.

## Depth First Search

Depth-first search (DFS)[140] is an algorithm for traversing the nodes and edges of graphs. The algorithm starts at a (typically arbitrarily chosen) root node, and explores as far as possible along a branch (see Fig. 2·6) before backtracking to nodes which have already been visited.



**Figure 2·6:** A tree illustrating the order nodes are visited in a depth-first search traversal of a graph. Back edges (edges between nodes which have already been visited) are omitted.

## DFS Code

Given a DFS tree  $T$  for a graph  $G$ , a *DFS code*  $(G, T)$  is an edge sequence that describes the order in which the edges of a graph were traversed in a depth-first search[143]. One can represent each edge as a 5-tuple,  $(i, j, l_i, l_{ij}, l_j)$ , where  $i$  and  $j$  are indices which indicate the order in which nodes  $v_i$  and  $v_j$  were first visited,  $l_{i/j}$  is

the label of node  $v_{i/j}$ , and  $l_{ij}$  is the label of the edge connecting nodes  $v_i$  and  $v_j$ .

## Support

For a given set of graphs  $D = \{G_1, G_2, \dots, G_n\}$ , the  $support(P)$  is the number of graphs  $G \in D$  for which  $P$  is a subgraph[143].

## 2.4 Identifying ET pathways in proteins

### 2.4.1 Pathways Model

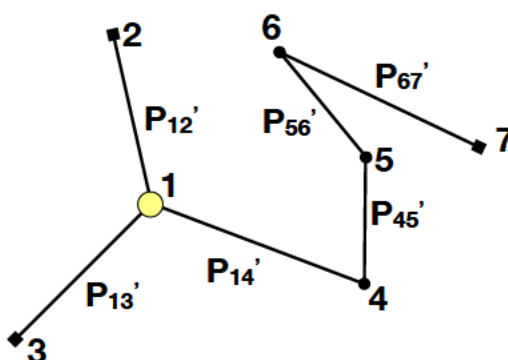
The *Pathways* model is a theoretical framework for identifying efficient ET pathways based on the tertiary protein structure[144–146]. It is focused on estimating the donor-acceptor coupling  $H_{DA}$  from Marcus Theory (Eq. 1.13), and assumes that ET occurs as a sequence of tunneling events. Each step can be mediated by a covalent bond, hydrogen bond, or vacuum, and is characterized by a *penalty function*. The total coupling  $H_{AB}$  is calculated as a product of penalties for each step [145, 146]:

$$H_{DA} = const \times \prod_{i \in space} \epsilon_i \times \prod_{j \in bond} \epsilon_j \times \prod_{k \in H-bond} \epsilon_k \quad (2.88)$$

In most parameterizations, the penalty for covalent bond mediated tunneling is  $\epsilon_j = 0.6$ [147], the penalty for through space tunneling is  $\epsilon_i = 0.6 \times \exp(\beta^S(R_i - 1.4))$ , and the penalty for hydrogen bond mediated tunneling is  $\epsilon_k = 0.6 \times \exp(\beta^S(R_k - 2.8))$ , where  $R_{i/k}$  is the interatomic distance, and the decay factor  $\beta^S$  is  $1.7 \text{ \AA}^{-1}$ [145, 148, 149]. Later studies have also used decay factors of  $1.1 \text{ \AA}^{-1}$  for  $\beta$ -sheets and  $1.4 \text{ \AA}^{-1}$  for  $\alpha$ -helical structures[145, 150].

With this parameterization of  $H_{DA}$ , the task of identifying the most efficient pathways is equivalent to maximizing the product in Eq. 2.88. To solve this problem,

the Pathways model employs a graph theory based strategy (Fig 2.7). The protein is represented as an undirected graph, where each atom corresponds to a node, and the edges are weighted using modified  $P' = -\log(\epsilon)$  penalty functions [144–146, 148]. This ensures non-negative edge weights, and transforms the problem from maximizing the product of  $\epsilon$  to minimizing the sum of  $P'$ . In other words, the shortest pathways (which can be routinely calculated using standard graph theory algorithms) in the graph correspond to the largest values of  $H_{DA}$ , and therefore, the most efficient ET pathways.



**Figure 2.7:** A graph representing the connectivity network between atoms in a protein.  $P'_{ij}$  are penalty functions associated with electron/hole hopping between the sites. This image is reproduced with permission from Ref. [151]. Copyright 2019 American Chemical Society.

The eMap model, described in Chapter 6, can be considered a coarse grained adaptation of the Pathways model which only considers through-space tunneling events. The graph of the protein structure in the eMap model consists of nodes corresponding to aromatic residues and other user selected redox sites and co-factors, and the edge weights correspond to purely distance dependent penalty functions.

### 2.4.2 Frequent Subgraph Mining

In the eMap software, the task of identifying shared pathways among a family of proteins is formulated as searching for frequently occurring subgraphs among a family of protein graphs. In graph theory, this problem is known as *frequent subgraph mining* (FSM) [152, 153]. FSM techniques are commonly used in areas which require identification of patterns in structured data, such as analysis of XML documents and web usage. FSM has also been applied to various problems in the bio- and cheminformatics communities, including drug discovery [154], identification of protein-ligand interfaces [155, 156] and the study of evolutionary relationships through mining phylogenetic trees [157–159].

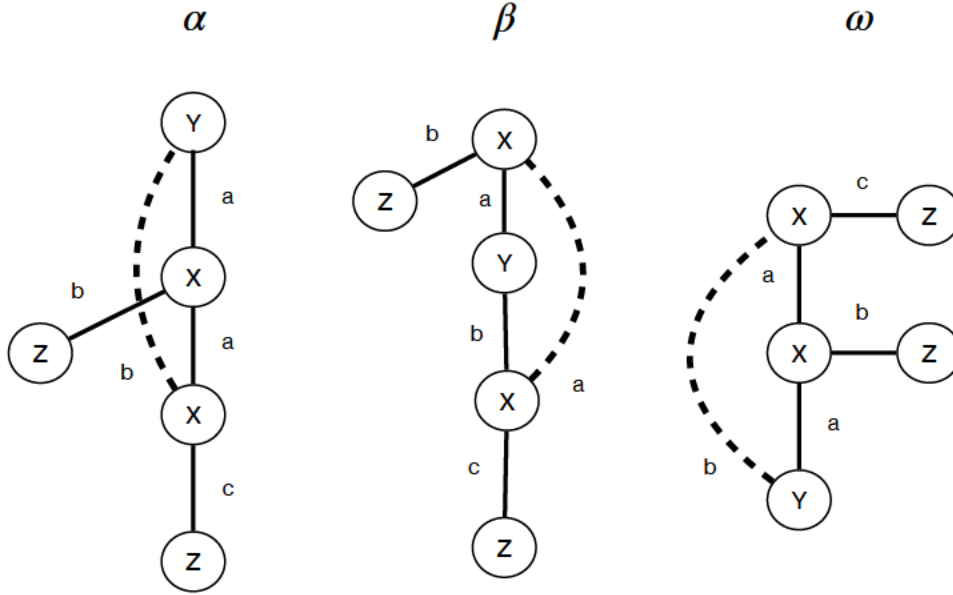
FSM in the eMap software relies on the well-known gSpan algorithm [143]. gSpan is a complete algorithm, which means it identifies all possible subgraphs whose support is above a given threshold. The algorithm relies on a lexicographic ordering scheme to determine *minimum DFS codes*. Consider the following example. Let  $\alpha = \text{code}(G_\alpha, T_\alpha) = (a_0, a_1, \dots, a_m)$  and  $\beta = \text{code}(G_\beta, T_\beta) = (b_0, b_1, \dots, b_n)$ . In the DFS lexicographic ordering,  $\alpha \geq \beta$  if and only if either of the following is true:

1.  $\exists t, 0 \leq t \leq \min(m, n), a_k = b_k$  for  $k < t, a_t < b_t$
2.  $a_k = b_k$  for  $0 \leq k \leq m$ , and  $n \geq m$

Fig. 2.8 and Table 2.1 show an example of 3 possible DFS codes for the same graph. According to the DFS lexicographic ordering,  $\omega < \beta < \alpha$ , i.e.  $\omega$  is the minimal DFS code among these three [143].

The mining algorithm is illustrated schematically in Fig 2.9, and is described more in detail in Ref. [143]. Briefly, the algorithm utilizes a depth first search traversal of the search space, only exploring subgraphs which are frequent (i.e. support greater





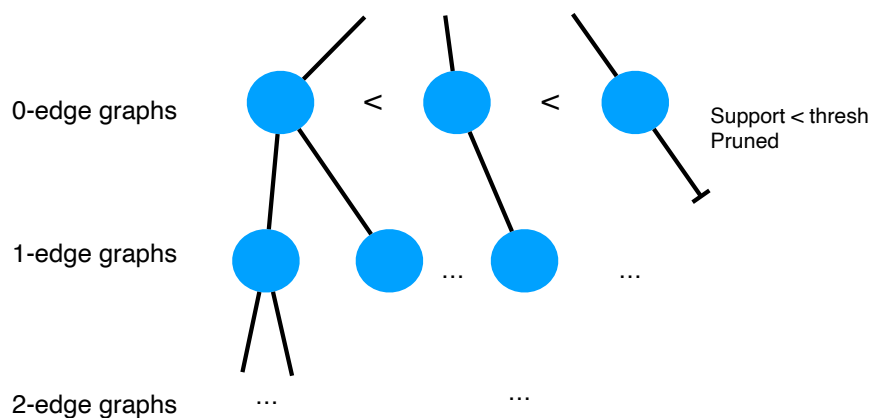
**Figure 2-8:** Examples of DFS traversals for a graph with node labels  $\{X, Y, Z\}$ , and edge labels  $\{a, b, c\}$ . Back edges (edges between nodes which have already been visited) are indicated as dashed lines.

**Table 2.1:** DFS codes corresponding to the traversals in Fig. 2-8.

Edge	$\alpha$	$\beta$	$\omega$
1	(1,2,Y,a,X)	(1,2,X,a,Y)	(1,2,X,a,X)
2	(2,3,X,a,X)	(2,3,Y,b,X)	(2,3,X,a,Y)
3	(3,4,X,c,Z)	(3,4,X,c,Z)	(3,1,Y,b,X)
4	(3,1,X,b,Y)	(3,1,X,a,X)	(2,4,X,b,Z)
5	(2,5,X,b,Z)	(1,5,X,b,Z)	(1,5,X,c,Z)

than the chosen threshold) and associated with minimum DFS codes. Importantly, gSpan does not rely on explicit generation of candidate subgraphs, and instead grows frequent  $(k+1)$ -edge DFS codes from frequent  $k$ -edge DFS codes. When the algorithm encounters a DFS code which is infrequent or not minimal, the entire branch below it is pruned, which substantially reduces the size of the search space while maintaining the completeness of algorithm up to the chosen support threshold[143].

The final step of any FSM application is to match the mined patterns to subgraph isomorphisms, and, optionally, group them based on some metric of importance. In



**Figure 2-9:** Schematic illustrating the gSpan algorithm. The DFS codes are sorted in lexicographic ordering, and a DFS traversal of the search space is performed to identify frequent subgraphs. Subtrees of infrequent or non-minimal DFS codes are pruned.

this work, we use the VF2 matching algorithm; the details of the algorithm will not be described here and can be found elsewhere[141, 142, 160]. Since there can be hundreds of matches for a given pattern, a novel clustering algorithm (see Chapter 7 and Appendix C) based on sequence or structural similarity is used to group similar pathways.

The implementation and application of FSM techniques for identifying shared ET pathways in the eMap web application is presented in Chapter 7.

## Chapter 3

# Application of Box and Voronoi CAPs to Molecular Clusters

### 3.1 Overview

The complex absorbing potential (CAP) approach offers a practical tool for characterization of energies and lifetimes of metastable electronic states, such as temporary anions and core ionized states. Here, we present an implementation of the smooth Voronoi CAP combined with equation-of-motion coupled cluster with single and double substitutions method for metastable states. The performance of the smooth Voronoi CAP and box CAP is compared for different classes of systems: resonances in isolated molecules, and localized and delocalized resonances in molecular clusters. The benchmark calculations show that the Voronoi CAP is generally more robust when applied to molecular clusters, but box CAPs are equally reliable for localized resonances or in the cases when the resonance does not exhibit significant electron density delocalization into the intramolecular region. As such, the choice of the CAP shape and onset should be guided by the character of the metastable states.

---

The work reported in this Chapter is reproduced with permission from Ref. [114]. Copyright 2022 American Chemical Society. The Supporting Information for this Chapter is available free of charge at <https://pubs.acs.org/doi/full/10.1021/acs.jpca.2c04892>

## 3.2 Introduction

Electronic resonances are metastable electronic states with finite lifetimes lying in the ionization/detachment continuum. Common examples include transient anions formed by electron attachment, and core-excited and core-ionized states that can undergo Auger decay or related relaxation pathways. These states are key players in a wide variety of processes ranging from those that occur in high energy environments (plasmonic photocatalysis, attosecond and X-ray spectroscopies) to low energy electron-molecule scattering (DNA damage from secondary electrons, interstellar chemistry) [17, 21–24, 161]. Resonances belong to the continuous spectrum of the electronic Hamiltonian, so they are not part of the usual Hilbert space of square integrable functions [4]. Theoretical description of these states is generally not possible by means of conventional quantum chemistry methods developed for bound states, and one has to use special techniques to obtain accurate resonance energies and lifetimes.

Non-Hermitian quantum mechanics (NHQM) techniques provide an appealing approach that allows one to leverage existing quantum chemistry methodology to treat metastable electronic states [15, 162, 163]. In NHQM formalisms, a resonance appears as a single square-integrable eigenstate of a non-Hermitian Hamiltonian. The resonance parameters can be extracted from the corresponding complex eigenvalue:

$$E = E_R - \frac{i\Gamma}{2} \quad (3.1)$$

where the real part of the energy ( $E_R$ ) is the *resonance position*, and the imaginary part ( $-\Gamma/2$ ) is the negative of the *half-width*, which is inversely proportional to the lifetime of the metastable state ( $\tau = \frac{\hbar}{\Gamma}$ ) [4, 104]. One of the NHQM formalisms that has been successfully used to study resonances in molecular systems is the *complex absorbing potential* (CAP) method. CAPs are imaginary potentials added to the

Hamiltonian, originally devised as a numerical technique to absorb outgoing wave packets near the boundaries of finite grids [110]. CAPs have also been applied in the time-independent framework to study various problems, including the Stark effect, reactive scattering, and quantum transport in molecular devices [164–166], and are now routinely used for evaluation of metastable states parameters [8, 125, 127–130, 132, 135, 167].

In this work, we explored the effects of the CAP shape on computed resonance parameters in molecular systems. In the context of electronic structure of metastable electronic states, CAPs are used to transform the resonance into a single square integrable state, and to render the state accessible by means of standard bound-state techniques [104]. To this end, the electronic Hamiltonian is augmented with an imaginary potential ( $-i\eta W$ )

$$H_{CAP}(\eta) = H - i\eta W \quad (3.2)$$

where  $\eta$  is the CAP strength parameter, and  $W$  is a real potential which vanishes in the vicinity of the molecular system and grows with distance [104]. An ideal CAP must satisfy the following two criteria. First, it should not perturb the system in the inner, molecular region, i.e. it should not introduce reflections. Secondly, the CAP should absorb the outgoing tail of the resonance wave function [104, 135]. Several approaches for minimizing these residual reflections have been proposed. Examples include “reflection-free” and “transformative” CAPs [168–170], and eliminating the artificial effect of CAP using the Padé approximant [112]. In most practical applications, simple quadratic CAPs such as those discussed here are used, as they can easily be combined with existing electronic structure theory codes.

Since the CAP-augmented Hamiltonian depends on the strength of the CAP (Eq. 3.2), an optimal value of  $\eta$  which provides the best estimate of the resonance po-

sition and width, must be identified. With a complete one-electron basis set, the exact resonance position and width are obtained in the limit of an infinitesimally weak CAP ( $\eta \rightarrow 0^+$ ) [104]. In practice, when finite bases are used, the optimal CAP strength  $\eta_{opt}$  is found by locating a stationary point on the eigenvalue trajectory  $E(\eta)$ , for example using the minimum of the logarithmic velocity criterion ( $|\eta \frac{dE}{d\eta}| \rightarrow min$ ) [104]. The structure of this trajectory, and the best estimate of resonance position and width for a given basis set/method, depends on the choice of the CAP.

The most commonly used form of the CAP for electronic structure calculations is the quadratic ‘‘box CAP’’, which is defined by three cutoff parameters ( $R_x^0, R_y^0, R_z^0$ ) that specify the onset of the box in each Cartesian coordinate [8, 133].

$$W = W_x + W_y + W_z$$

$$W_\alpha = \begin{cases} 0, & |r_\alpha| < R_\alpha^0 \\ (r_\alpha - R_\alpha^0)^2, & |r_\alpha| > R_\alpha^0 \end{cases} \quad (3.3)$$

where  $\alpha = x, y$ , or  $z$ .

Practical recipes for choosing the box size have been suggested, and correction schemes have been shown to reduce the dependence of the results on the box size [116, 119]. However, the rigid box-like shape is not always reflective of the detailed geometry of a molecular system, and it can be difficult to apply this type of CAP to systems which do not efficiently fill the space.

The idea of defining a CAP using each atom’s Voronoi cell was proposed by Sommerfeld and Ehara [135, 136]. The Voronoi absorbing potential wraps uniformly around any molecule, and is specified by a single cutoff radius  $r_{cut}$ . Following the work by Sommerfeld and Ehara [135, 136] we use the ‘‘smooth Voronoi’’ potential, which smooths out the edges between Voronoi cells, making the resulting CAP

more amenable to numerical integration [135]. In this case, the CAP is defined as follows [135]:

$$W(\vec{r}) = \begin{cases} 0, & r_{WA} \leq r_{cut} \\ (r_{WA}(\vec{r}) - r_{cut})^2, & r_{WA} > r_{cut} \end{cases} \quad (3.4)$$

$r_{WA}(\vec{r})$  in Eq. 3.4 is the effective distance to the molecular system evaluated as a weighted average of the distances to all nuclei:

$$r_{WA}(\vec{r}) = \sqrt{\frac{\sum_i w_i |\vec{r} - \vec{R}_i|^2}{\sum_i w_i}} \quad (3.5)$$

$R_i$  is the distance to the  $i^{th}$  nucleus and  $w_i$  is the weight defined as follows:

$$w_i = \frac{1}{(|\vec{r} - \vec{R}_i|^2 - r_{min}^2 + 1 \text{ a.u.})^2} \quad (3.6)$$

where  $r_{min}$  is the distance to the closest atom:  $r_{min} = \min_i |\vec{r} - \vec{R}_i|$ .

In contrast to the box CAP, the Voronoi CAP shares exactly the symmetry of the molecular system, and is flexible, i.e. is easily adjustable to changing nuclear configurations, for example in dynamical simulation. The immediate price one pays for this flexibility is a less convenient functional form; CAP integrals [133] and gradients [134] for metastable states can be obtained analytically for box CAP, but currently must be evaluated numerically using Voronoi CAPs. Another consequence of this flexible functional form is that unlike the box CAP, the Voronoi CAP will leak into the empty spaces in molecular clusters as it wraps around each molecule. As such, one can expect differences in behavior of these two types of CAP for different types of systems.

To shed light on these differences, we have implemented the smooth Voronoi CAP in the Q-Chem program package [171] and compared its performance to that of the

standard box CAP for shape resonances in several representative systems. To explore the effects of CAP type on resonances in compact molecules we considered  $\pi^*$  resonances in  $N_2^-$  and  $CO^-$ . Following the work by Ehara and Sommerfeld [135], we have explored the effects of CAP shape on a localized resonance in an  $N_2^-(H_2O)_2$  cluster. In addition, we have considered an example of delocalized electron-attached resonances in an anion of the experimentally observed [172–174] carbon monoxide dimer  $((CO)_2)$ . We employed the CAP equation-of-motion coupled cluster method with single and double substitutions for electron attachment (CAP-EOM-EA-CCSD) [8, 116, 119] to compute the resonance energies and widths of all model systems. Specifically, we exploit CAP-EOM-EA-CCSD in the projected scheme, where the CAP-augmented Hamiltonian is represented on a reduced subspace spanned by a small number of EOM-EA-CCSD eigenstates [120, 122, 125, 129, 130, 132].

The structure of this Chapter is as follows. We outline the main features of the CAP method in Sec. 3.3 and discuss the technical details in Sec. 3.3. The performance of box and smooth Voronoi CAPs for description of resonance parameters in diatomics, localized and delocalized resonances in clusters is discussed in Secs. 3.4, 3.4, and 3.4, respectively.

### 3.3 Methods

#### Complex Absorbing Potentials

In the CAP method, resonance parameters are obtained as complex eigenvalues (Eq. 3.1) of the CAP-augmented electronic Hamiltonian (Eq. 3.2). The CAP-augmented Hamiltonian is complex symmetric ( $H(\eta)^\dagger = H(\eta)^*$ ), and so the usual Hermitian inner product is replaced with the *c-product* [104, 175]

$$(\phi_i(\vec{r})|\phi_j(\vec{r})) = \int d\vec{r}\phi_i(\vec{r})\phi_j(\vec{r})$$



In a finite basis representation, an optimal  $\eta_{opt}$  which provides the best estimate of resonance position and width must be identified. The  $\eta_{opt}$  is system- and state-dependent, and varies with the electronic structure method, basis set, and the form of the CAP used. To search for  $\eta_{opt}$ , the calculation is repeated over a range of  $\eta$  values, generating complex eigenvalue trajectories which depend on  $\eta$ . When using the logarithmic velocity criterion [104],  $\eta_{opt}$  corresponds to the CAP strength on the eigenvalue trajectory where  $|\eta \frac{dE}{d\eta}|$  has its minimum.

Results obtained in this fashion are sensitive to the CAP onset, due to unphysical perturbations of the resonance wave function induced by the CAP [116]. Several correction schemes have been proposed to improve the accuracy and stability of CAP augmented calculations [116, 117]. Here we employ the scheme based on first order “de-perturbation” of the complex energy [104]. The real,  $U^R$ , and imaginary,  $U^I$ , parts of the resulting corrected complex energy can be expressed as follows [116]:

$$\begin{aligned} U^R &= E^R - \eta Tr [W \gamma^I] \\ U^I &= E^I + \eta Tr [W \gamma^R] \end{aligned} \quad (3.7)$$

where  $W$  is the CAP matrix (Eq. 3.2), and  $\gamma_R/\gamma_I$  are the real and imaginary parts of the state reduced one particle density matrix. The  $\eta_{opt}$  for corrected trajectories can be obtained by minimizing the logarithmic velocity of  $U(\eta)$ :

$$\left| \eta \frac{dU}{d\eta} \right| \rightarrow min \quad (3.8)$$

Alternatively, the real and imaginary parts of the corrected energy can be analyzed independently, resulting in two values of  $\eta_{opt}$ [116].

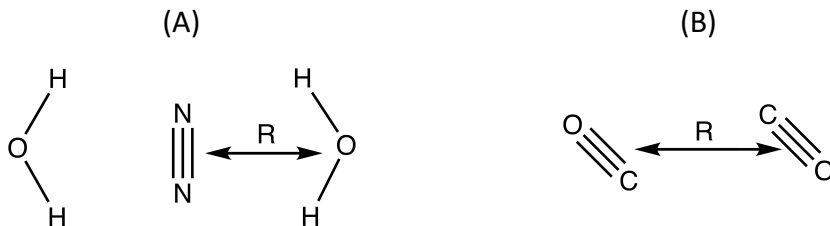
Regardless of which scheme is used, analysis of  $\eta$ -trajectories can pose a significant challenge when there are multiple apparent stationary points. Naively following Eq. 3.8, for example, can result in sudden jumps from switching between station-

ary points, which is undesirable for the purpose of obtaining smooth potential energy curves, and for the purpose of comparing results from different CAPs on equal footing. Strategies such as continuum remover CAP [113] aim to identify the “true” resonance stabilization point, but ultimately still rely on graphical searches, which can be inconclusive. In the reported numerical results, the value of  $\eta_{opt}$  for each calculation was selected based on Eq. 3.8, following the same stationary point (when possible) for different CAPs at each geometry.

### Computational Details

The smooth Voronoi CAP was implemented in the Q-Chem program package, and all calculations were performed using Q-Chem version 5.4.1. In this work, we employ the CAP-EOM-EA-CCSD method in the projected scheme[120, 122]. 30 EOM-EA-CCSD states in the appropriate irreducible representation were included in the correlated basis for the projected CAP calculations.

The basis sets used in all calculations consisted of Dunning’s correlation consistent cc-pVTZ [176] basis augmented by additional even-tempered basis functions. For the  $N_2$  monomer and the  $N_2(H_2O)_2$  cluster, the basis set was chosen to match that used by Ehara and Sommerfeld in Ref. [135]. Starting from the cc-pVTZ basis set, the basis set for nitrogen was augmented with a (2s5p2d) set of even-tempered diffuse functions (with scaling factors of 1/2 for s-type and d-type functions, and 2/3 for p-type basis functions), and the basis set for the oxygen atoms on the water monomers was augmented with a set of (1s1p1d) diffuse functions (with a scaling factor of 1/3 for all angular momenta). For the  $CO$  monomer and the dimer, the aug-cc-pVTZ basis set was augmented by an additional (3s3p3d) set of diffuse even-tempered basis functions (with a scaling factor of 1/2) located at the center of mass of each monomer[119]. The grid used for computing the CAP integrals was a Becke-type grid of 500 radial points and 3470 angular Lebedev points [177, 178].



**Figure 3.1:** Geometries of the  $N_2(H_2O)_2$  model cluster (A), and the  $(CO)_2$  dimer (B). Resonance states are investigated as functions of the intermolecular distance,  $R$ .

Bond lengths of 2.0740 and 2.1316 bohr were used for  $N_2$  and  $CO$ , respectively. For the model  $N_2(H_2O)_2$  cluster, experimental equilibrium geometries were used for the monomers  $N_2$  and  $H_2O$ , while the  $N_2$ -oxygen distance was varied, as indicated in Fig 3.1A. For the  $CO$  dimer, the computed equilibrium geometry of the lower energy “a-state” isomer from Ref. [179] was chosen as the starting point, and the distance between centers of mass of the two monomers was varied (Fig. 3.1B). Representative geometries of the clusters are given in the SI (Sec. S1).

Throughout the discussion, the box CAP size is specified by a single onset parameter  $r_{box}$  in such a way that the onset in each dimension ( $R_\alpha^0$  in Eq. 3.3,  $\alpha = x, y, z$ ) is obtained by adding  $r_{box}$  to the maximum value of the nuclear coordinate in that dimension ( $x, y, \text{ or } z$ ), centered at the molecular center of mass. Voronoi CAPs are specified by the  $r_{cut}$  cutoff radius (see Eq. 3.4). For all CAP calculations on the model  $N_2^-(H_2O)_2$  and the  $(CO)_2^-$  dimer, the CAP onset was chosen as 3.0 bohr for box and Voronoi CAPs.

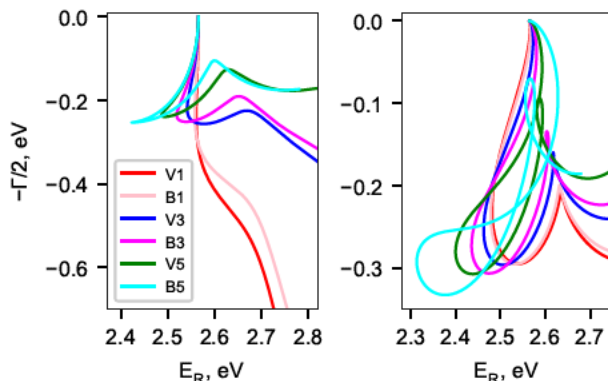
### 3.4 Results and Discussion

In this section, we compare computed eigenvalue trajectories for box and Voronoi CAPs for three types of systems: resonances in diatomic molecules (3.4), localized

resonances in a model cluster (3.4), and delocalized resonance states in a dimer (3.4).

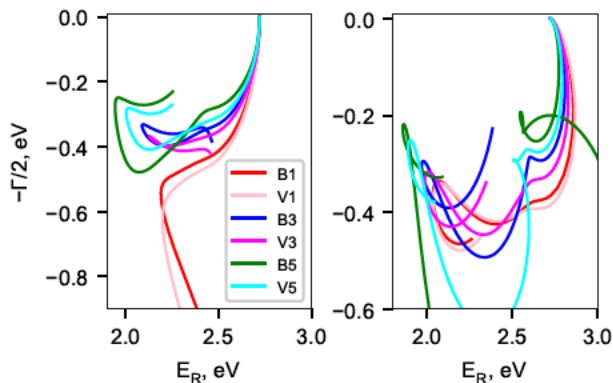
### Diatomics: $N_2^-$ , $CO^-$

The  $\pi^*$  resonance states of  $CO^-$  and  $N_2^-$  arise from electron attachment to the lowest unoccupied valence molecular orbital of the neutral diatomic molecule. As shown in Figs. 3-2 and 3-3, similarly defined box and Voronoi CAPs (i.e. with the same onset) yield nearly identical trajectories, but the structure of the zero-order trajectories can vary substantially with the onset. CAPs placed further away from the system show a tendency to exhibit additional spurious stationary regions at smaller values of  $\eta$ , while CAPs placed too close to the system can result in trajectories which lack a clear stabilization behavior. These differences are mostly eliminated when switching to corrected trajectories, and each CAP shows a single prominent stationary region.



**Figure 3-2:** Uncorrected (left) and corrected (right) eigenvalue trajectories obtained using various CAPs for the  ${}^2\Pi_g$  resonance of  $N_2^-$ . V1/B1, V3/B3, V5/B5 correspond to Voronoi/box CAPs with onsets of 1, 3, and 5 bohr, respectively.

Corrected resonance positions and widths computed using both types of CAP are shown in Fig. 3-4 and 3-5. The effects of the CAP shape on the computed resonance parameters are rather minor for these two states, and box and Voronoi CAPs show a similar monotonic dependence of the computed resonance parameters on the CAP onset, as both parameters smoothly decrease as the CAP is placed further away from

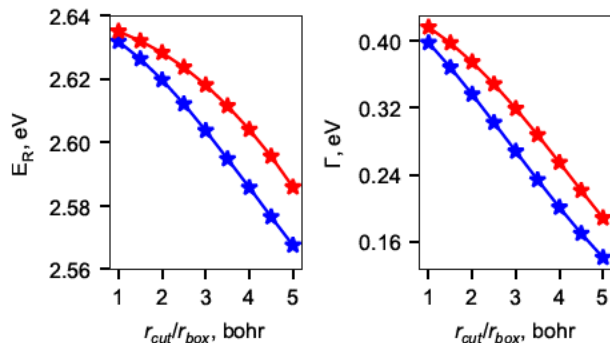


**Figure 3-3:** Uncorrected (left) and corrected (right) eigenvalue trajectories obtained using various CAPs for the  ${}^2\Pi$  resonance of  $CO^-$ . V1/B1, V3/B3, V5/B5 correspond to Voronoi/box CAPs with onsets of 1, 3, and 5 bohr, respectively.

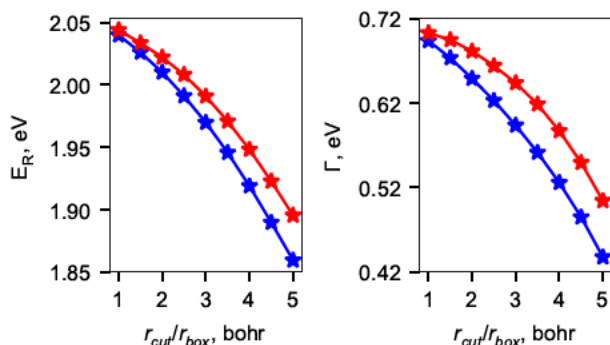
the molecular system (i.e. increasing the onset). The slightly greater resonance energy and width computed with Voronoi CAPs in comparison to those evaluated using analogous box CAPs can be attributed to the Voronoi CAPs having an effectively smaller volume of the region where the CAP is zero, i.e. an effectively smaller CAP onset. The reported trends for both states are consistent with previous CAP studies for both types of CAP [116, 119, 135]. Therefore, as expected, Voronoi and box CAPs perform similarly for small molecules, as the choice of CAP shape presents no unique challenges for these types of systems.

### Localized resonance in a cluster: $N_2^-(H_2O)_2$

We turn our attention now to molecular clusters. A model  $N_2^-(H_2O)_2$  cluster was used as a pilot application of the Voronoi CAP by Sommerfeld and Ehara [135]. In this cluster, the two water molecules provide an environment which splits the  ${}^2\Pi_g$  resonance of  $N_2^-$  into in-plane ( ${}^2B_{1g}$ ) and out-of-plane ( ${}^2B_{3g}$ ) components [135]. In both states, the resonance state is primarily localized on the  $N_2$  monomer. The interaction with the water molecules stabilizes both states, which is reflected in lower resonance energies and widths relative to the isolated  $N_2^-$ . The environment has a



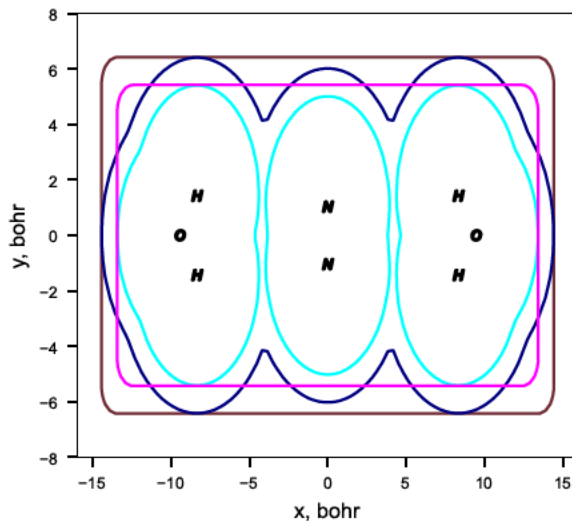
**Figure 3-4:** Dependence of computed resonance parameters for  ${}^2\Pi_g$  resonance of  $N_2^-$  on the CAP onset:  $r_{cut}$  and  $r_{box}$  for Voronoi and box CAPs, respectively. Corrected results for box and Voronoi CAPs are shown in blue and red, respectively.



**Figure 3-5:** Dependence of computed resonance parameters for  ${}^2\Pi$  resonance of  $CO^-$  on the CAP onset:  $r_{cut}$  and  $r_{box}$  for Voronoi and box CAPs, respectively. Corrected results for box and Voronoi CAPs are shown in blue and red, respectively.

larger effect on the in-plane component, lowering its energy and width relative to the out-of-plane component[135]. This model system serves as a test case to illustrate a key difference in the behavior of box and Voronoi CAPs. As the water molecules are moved further from  $N_2$ , the Voronoi CAP will leak into the cavity between molecules (Fig. 3-6), while the box CAP will contain regions of “dead space” [135] which are poorly represented by atom-centered one-electron basis functions, and can potentially introduce artifacts into CAP calculations.

Real and imaginary potential energy curves computed using the two CAPs are



**Figure 3-6:** Comparison of different box and Voronoi CAPs for the  $N_2^-(H_2O)_2$  at  $R = 5\text{\AA}$ . Plotted is the isoline  $W=1$  hartree for box CAPs of 4 bohr (brown) and 3 bohr (magenta), and smooth Voronoi CAPs of 4 bohr (navy) and 3 bohr (cyan) in the plane of the molecular system.

shown in Fig. 3-7. The CAP shape has little effect on the resonance position for all considered values of  $R$ , and the energy gap between the two states narrows as the system is stretched. We focus our attention on the widths, which show clear differences in behavior for the two states. For the out-of-plane  ${}^2B_{3g}$  state, both CAPs reproduce the width of the isolated  $N_2^-$  monomer (computed using an analogously defined CAP) in the asymptotic region, however, only the Voronoi CAP produces degeneracy between the two states in the imaginary part of the energy (Table S2). While the difference in the widths between the two states computed by box CAP is minor ( $\sim 0.05$  eV at  $R = 12\text{\AA}$ ), it clearly illustrates an important consideration for choosing a CAP: the onset parameters in each dimension ( $r_{x,y,z}$  in Eq. 3.3) are not equally important for a given state [119], which can introduce an imbalance in treatment of different resonances. In this case, as  $R$  is increased, the box CAP onsets in the dimensions perpendicular to the molecular axis remain unchanged, but the volume of the space where box CAP is zero is increased by introducing empty

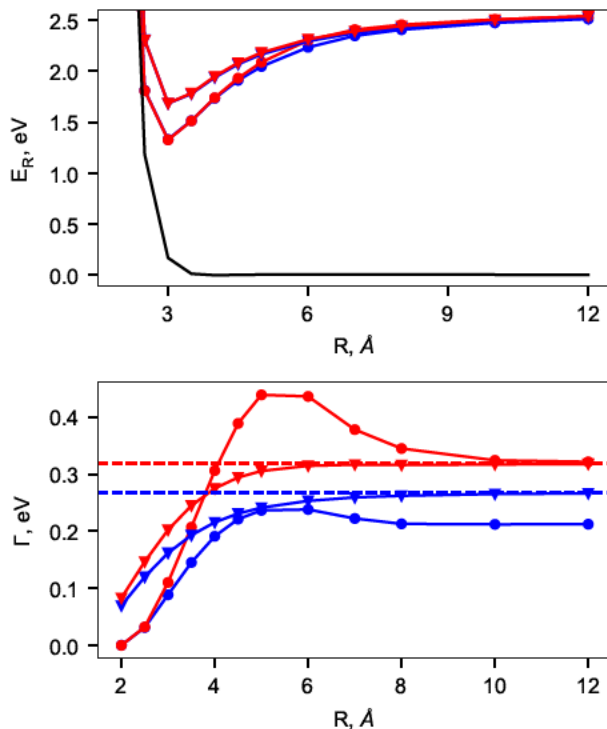
space in the intermolecular region. As a result, the in-plane state ( ${}^2B_{1g}$ ) experiences a weaker CAP as  $R$  is increased, while the out-of-plane state ( ${}^2B_{3g}$ ) is largely unaffected by this empty space. Based on the trends observed for the  $N_2^-$  monomer, one could expect a lower resonance energy and width for the in-plane ( ${}^2B_{1g}$ ) state relative to the out-of-plane ( ${}^2B_{3g}$ ), and indeed, the asymptotic value of the resonance widths for the  ${}^2B_{1g}$  state is  $\sim 0.05$  eV below that for the out-of-plane  ${}^2B_{3g}$  state and the isolated  $N_2^-$  monomer. This also explains why the  ${}^2B_{1g}$  state is more sensitive to the CAP shape from 4 Å to 8 Å; the in-plane state feels the presence of Voronoi CAP in the cavity more strongly, while the box CAP becomes progressively weaker, resulting in a larger difference in the computed widths between the two CAPs for this state.

In the initial benchmarks for Voronoi CAP, Ehara and Sommerfeld [135] reported that a box CAP resulted in larger values of  $\Gamma$  than a Voronoi CAP for the in-plane  ${}^2B_{1g}$  state as the system was stretched from  $R \sim 4 - 5$  Å, which at first appears to contradict our results. The authors of Ref. [135] attributed this behavior to artifacts stemming from the “dead space”. This disagreement likely stems from analysis of the trajectories. The trajectories admit multiple stationary points, and depending on the choice of the stationary point, the width for the box CAP in this region can be either smaller or greater than that for the Voronoi CAP (see Figs 3-8 and S14). However, at the dissociation limit, there is only one stationary point (where box CAP exhibits a smaller width), which is the one we chose to follow throughout the potential energy curve. To support our conclusions, we repeated the calculations with the projected CAP/EA-ADC(2) method [129], and were able to reproduce the same key trends (see Figs. S13, S14, and Table S3).

### Delocalized resonance: $(CO)_2^-$

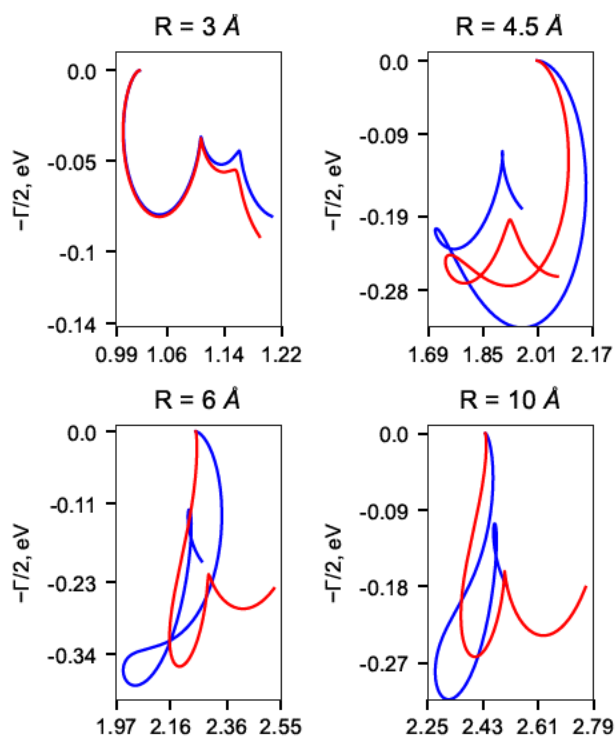
Lastly, we examine the effect of CAP shape on delocalized resonance states. As a model system we use an anion of the carbon monoxide dimer. The neutral dimer has





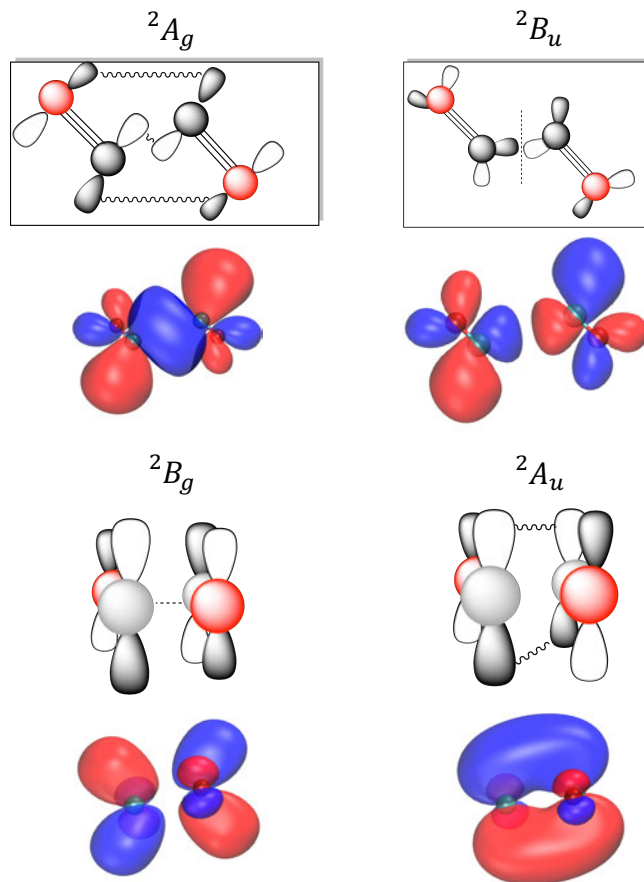
**Figure 3-7:** Resonance positions (top) and widths (bottom) for the in-plane  ${}^2B_{1g}$  (indicated as  $\bullet$ ) and out-of-plane  ${}^2B_{3g}$  (indicated as  $\nabla$ ) resonance states of the model  $N_2^-(H_2O)_2$  cluster, computed using box CAP (blue) and Voronoi CAP (red). The widths of the isolated  $N_2^-$  computed using the corresponding CAP are indicated as dashed lines. The energy of the neutral N-electron ground state is shown as a solid black line.

two stable isomers separated by  $0.88\text{ cm}^{-1}$  that have been characterized both computationally and experimentally [172–174, 179]. We explored the lower energy “a-state” isomer, for which the carbon atoms are arranged closer together in a slipped parallel geometry [172, 179]. The  ${}^2\Pi$  resonance in  $CO^-$  splits into four states in the dimer, which arise from in-phase/out-phase and in-plane/out-of-plane combinations of CO  $\pi^*$  orbitals (Fig. 3-9 [180]). At the equilibrium geometry (Table 3.1), the resonance positions for the out-of-phase states ( ${}^2B_g$  and  ${}^2B_u$ ) are located higher in energy than the isolated  $CO^-$ , while the in-phase states ( ${}^2A_u$  and  ${}^2A_g$ ) are stabilized by the  $\pi^*$  orbital overlap. In addition to the empty space in the intermolecular region affecting



**Figure 3-8:** Corrected trajectories for the in-plane  ${}^2B_{1g}$  state of the model  $N_2^-(H_2O)_2$  cluster, computed using a box CAP (blue) and a Voronoi CAP (red).

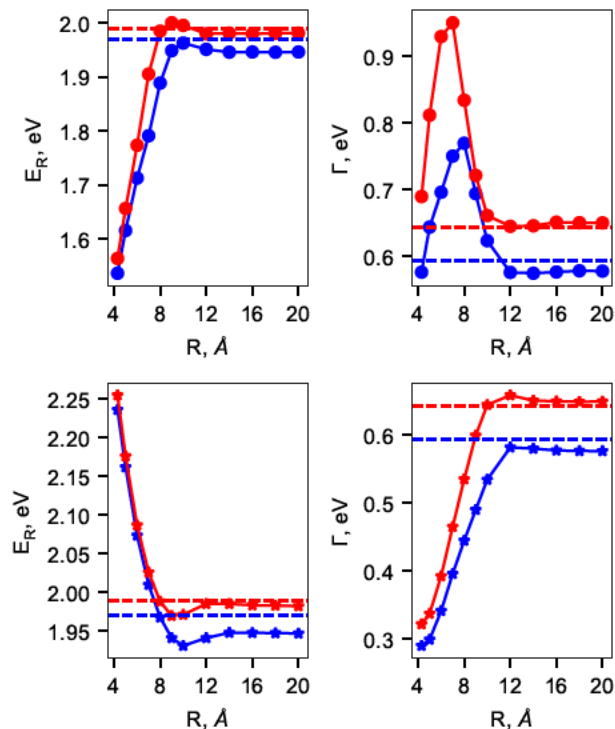
the performance of the box CAP (weaker CAP along one of the axes), there is also the possibility that the Voronoi CAP leaking into the cavity will overlap with the valence part of these delocalized resonances and perturb the resonance wave function. To investigate these competing issues, we computed eigenvalue trajectories for various distances  $R$  (Fig. 3-1) between the centers of mass of the two monomers.



**Figure 3.9:** Schematics and the real part of complex Dyson orbitals<sup>[181]</sup> for each of the four resonance states of the  $(CO)_2^-$  dimer. The Dyson orbitals were plotted at the equilibrium geometry with an isosurface value of 0.02.

Potential energy curves for the out-of-plane states for the  ${}^2B_g$  (out-of-phase) and  ${}^2A_u$  (in-phase) states of the  $(CO)_2^-$  dimer are shown in Fig. 3.10. Both CAPs yield similar trajectories at each geometry (see Figs. S4 and S5), and we were able to obtain smooth real and imaginary potential energy curves which become degenerate at the dissociation limit, and approximately reproduce the resonance parameters of the isolated  $CO^-$  monomer (see Table 3.1). The in-phase  ${}^2A_u$  state is more sensitive to CAP shape for both the real and imaginary parts of the energy, with resonance positions from the two CAPs varying by as much as 0.1 eV (e.g. at  $R$  to 7 and 8

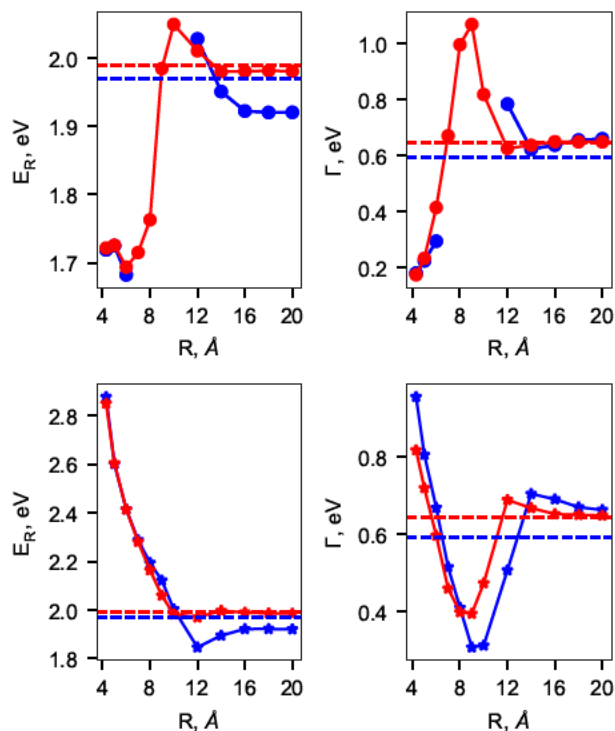
Å). Since the state is still stabilized relative to the isolated monomer in this region, this could be the result of overlap between Voronoi CAP and the valence part of the delocalized resonance, but the effect remains rather moderate, and the eigenvalue trajectories do not show any major artifacts otherwise.



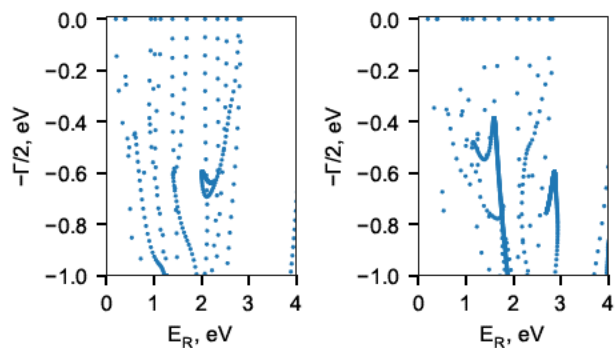
**Figure 3-10:** Corrected resonance positions and widths for the out-of-plane-in-phase  ${}^2A_u$  (indicated as  $\bullet$ ) and out-of-plane-out-of-phase  ${}^2B_g$  (indicated as  $*$ ) resonance states of the  $(CO)_2^-$  dimer, computed using box CAP (blue) and Voronoi CAP (red). The values for isolated  $CO^-$  computed using a Voronoi CAP ( $r_{cut}=3.0$  a.u.) are indicated as dashed lines.

Finally, we consider the in-plane states, which are shown in Fig. 3-11. As expected, these states are more affected by the shape of CAP, which leads to notable differences as the system is stretched. For the in-phase  ${}^2A_g$  state, the stationary points on the trajectories become increasingly less sharp as the system is stretched from the equilibrium geometry, and near  $8 \text{ \AA}$  the box CAP starts to fail entirely. Examination of the eigenvalue spectra, shown in Fig. 3-12) provides insight into the problem. While

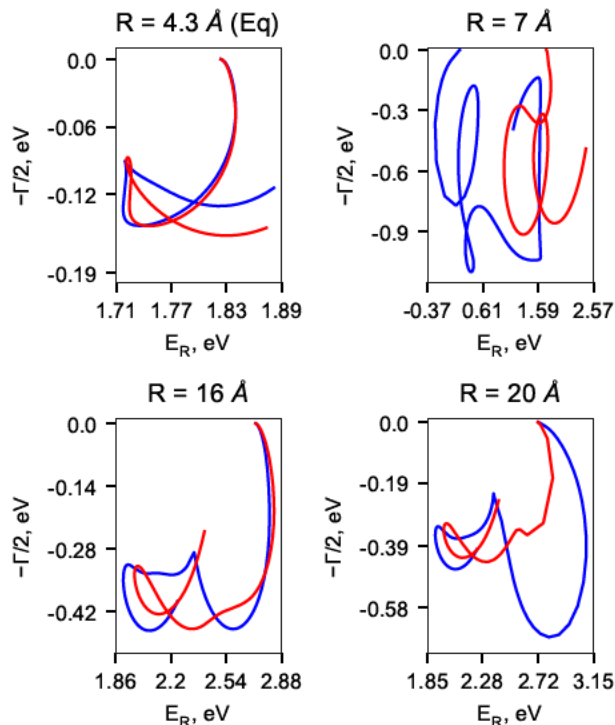
the Voronoi CAP clearly stabilizes a single resonance trajectory, the box CAP splits into two stabilized regions near 1.5 eV and 2.9 eV, neither of which yields reliable stationary points on the corrected trajectory. This second region eventually comes down in energy and becomes stabilized on the corrected trajectory starting near 14 Å, and as shown in Fig. 3-13, trajectories from the two CAPs start to resemble each other again as the dissociation limit is approached. This problematic behavior in the intermediate region can likely be attributed to the poor basis set description in the cavity. This issue is expected to be especially problematic for this state in particular, which arises from an in-plane in-phase combination of  $\pi^*$  orbitals. In principle, a more flexible basis set could resolve the issue, but our attempts to do so by including additional diffuse functions in the cavity region were unsuccessful, which underlines the challenge of describing such a state with primarily atom centered basis functions. Notably, both Voronoi and box CAP exhibit a large 0.2 eV jump in resonance position from  $R = 8 \text{ \AA}$  to  $R = 9 \text{ \AA}$ . The out-of-phase  ${}^2B_u$  state is not as problematic, and clear stabilization points for both CAPs are present on the corrected trajectories at each geometry.



**Figure 3-11:** Corrected resonance positions and widths for the in-plane-in-phase  ${}^2A_g$  (indicated as  $\bullet$ ) and in-plane-out-of-phase  ${}^2B_u$  (indicated as  $*$ ) resonance states of the  $(CO)_2^-$  dimer, computed using box CAP (blue) and Voronoi CAP (red). The values for isolated  $CO^-$  computed using analogous CAPs are indicated as dashed lines.



**Figure 3-12:** CAP-augmented Hamiltonian eigenvalue spectra for Voronoi CAP (left) and box CAP (right) for the in-plane, in-phase  ${}^2A_g$  state of the  $(CO)_2^-$  dimer at  $R = 8 \text{ \AA}$ .



**Figure 3-13:** Corrected trajectories for the in-plane, in-phase  $^2A_g$  state of the  $(CO)_2^-$  dimer, computed using a box CAP (blue) and a Voronoi CAP (red).

Returning to the comparison between the two types of CAP, at the dissociation limit the Voronoi CAP produces degeneracy in the real and imaginary parts of the energy between all four states (see Table 3.1), but for box CAP, the widths for the in-plane states lie  $\sim 0.08$  eV above those obtained for the out-of-plane states. The difference in treatment between the two types of states by box CAP is consistent with what we observed for the  $N_2^-(H_2O)_2$  cluster. Moreover, in this case, the definition of the box CAP in the asymptotic region is different from the monomer not only due to the "dead space" in the intermolecular region, but also due to the tilted orientation of the dimer. These differences in the definition of CAP likely explain the slight deviations from degeneracy of all four states for box CAP in the asymptotic region. Nonetheless, our results clearly show that the Voronoi CAP is able to provide a

balanced description of all four states, while the box CAP struggles to do so due to different states being sensitive to different components of the CAP onset. Lastly, to support our conclusions, we repeated the calculations using the projected CAP/EA-ADC(2) method[129], which showed similar trends for each of the states (see Figs S6-S9 and Table S1).

**Table 3.1:** Resonance positions (widths) of the four states of the  $(CO)_2^-$  dimer at the equilibrium geometry, and at  $R = 20 \text{ \AA}$ , computed using box and Voronoi CAPs.

	Eq.	$R = 20 \text{ \AA}$
	Box	
$^2A_g$	1.719(0.181)	1.921(0.660)
$^2B_u$	2.877(0.957)	1.921(0.663)
$^2A_u$	1.536(0.576)	1.946(0.577)
$^2B_g$	2.236(0.289)	1.947(0.576)
Monomer	-	1.970(0.593)
	Voronoi	
$^2A_g$	1.721(0.174)	1.981(0.648)
$^2B_u$	2.850(0.819)	1.985(0.647)
$^2A_u$	1.563(0.689)	1.981(0.650)
$^2B_g$	2.255(0.321)	1.982(0.650)
Monomer	-	1.991(0.644)
Experiment <sup>1</sup>	-	1.50(0.8) <sup>182</sup>

<sup>1</sup> - estimate extracted as a half-width of the peaks in elastic ( $\nu'=0$ ) and inelastic ( $\nu=1$ ) cross-sections[182].

### 3.5 Conclusions

In summary, we have analyzed the performance of box and Voronoi CAPs for three different types of systems. For diatomics, box and Voronoi CAPs were shown to generate eigenvalue trajectories which share the same key features, and display a similar dependence on CAP onset. For localized resonances in the  $N_2^-(H_2O)_2$  cluster,



Voronoi CAPs yield systematically larger widths in the intermediate region when the state is sensitive to the presence of CAP in the cavity, but ultimately produce the proper behavior at the dissociation limit. The box CAP performed well for both localized resonance states, but did not fully reproduce their degeneracy at the dissociation limit, due to the in-plane state experiencing a weaker CAP. For delocalized resonances in the  $(CO)_2^-$  dimer, the two types of CAP perform equally well for the out-of-plane states, but the flexibility of the Voronoi CAP was crucial for robust description of the in-plane in-phase state, where the box CAP struggled to stabilize the resonance as the system was stretched due to reflections from the edge of the basis set. Thus, the choice of CAP for molecular clusters should be guided by the character of the state of interest. The convenience of box CAPs can be exploited when the state is less sensitive to the onset in the dimension where box CAP is absent (e.g. the clustering axis), but a Voronoi CAP should be used when a more flexible CAP is needed.

## Chapter 4

# Projected CAP-EOM-CCSD method for electronic resonances

### 4.1 Overview

The complex absorbing potential equation-of-motion coupled-cluster (CAP-EOM-CC) method is routinely used to investigate metastable electronic states in small molecules. However, the requirement of evaluating eigenvalue trajectories presents a barrier to larger simulations, as each point corresponding to a different value of the CAP strength parameter requires a unique eigenvalue calculation. Here, we present a new implementation of CAP-EOM-CCSD which uses a subspace projection scheme to evaluate resonance positions and widths at the overall cost of a single electronic structure calculation. We analyze the performance of the projected CAP-EOM-CC scheme against the conventional scheme, where the CAP is incorporated starting from the Hartree-Fock level, for various small and medium sized molecules, and investigate its sensitivity to various parameters. Finally, we report resonance parameters for a set of molecules commonly used for benchmarking CAP-based methods, and we report estimates of resonance energies and widths for 1- and 2-cyanonaphthalene, molecules which were recently detected in the interstellar medium.

---

The work below is reproduced from Ref. [120], with the permission of AIP Publishing. The Supplementary Material for this Chapter is available free of charge at <https://aip.scitation.org/doi/10.1063/5.0082739>

## 4.2 Introduction

Electronic resonances are (N+1)-electron metastable electronic states which lie energetically above the associated N-electron system and decay through auto-ionization, and thus have intrinsically finite lifetimes. Common examples include transient anions formed by electron attachment to closed-shell species[9, 183], and core ionized or core excited states which decay through Auger decay or other related processes[25]. These states play important roles in a wide variety of processes, ranging from low energy electron-molecule scattering (e.g. DNA damage by secondary electrons, interstellar chemistry) [17, 21] to electron-molecule interactions occurring in highly energetic environments (e.g. plasmonic photocatalysis, attosecond and X-ray spectroscopies)[19, 22–24]. Resonances belong to the continuous spectrum of the Hamiltonian, and therefore, the associated eigenstates are not square-integrable[4]. Accurate calculation of resonance energies and lifetimes is a challenging task due to the non-stationary nature of metastable electronic states, which makes them formally not tractable with the conventional quantum chemistry techniques developed for bound states. Among the theoretical approaches developed for resonances, two groups of methods can be distinguished: those based on scattering theory which account for the continuum explicitly [7, 93, 184], and  $L^2$  methods which implicitly treat the scattering nature of the decaying state[4, 8, 104, 125]. In this work, the focus is on the complex absorbing potential (CAP) method, an  $L^2$  method which can be readily applied to the study of resonances in molecular systems[104, 111]. The CAP method is a non-Hermitian quantum mechanics  $L^2$ -formalism, in which the physical Hamiltonian is augmented with a complex-valued, artificial absorbing potential. Resonances appear as square-integrable eigenstates of the resulting Hamiltonian, and are associated with complex Siegert-Gamow eigenvalues:  $E = E_R - i\Gamma/2$ , where  $E_R$  and  $\Gamma$  correspond to resonance energy and width, respectively[6, 15]. Other examples of  $L^2$  methods include

complex scaling[175], complex basis functions[101], stabilization[95–97], and the analytical continuation of the coupling constant (ACCC) method[185]. The performance of the CAP, stabilization, and analytic continuation methods for a model potential was recently studied by Davis and Sommerfeld[162].

CAPs were originally introduced as tools to avoid reflections from the edge of the grid in wave packet propagation simulations[110], but were later adapted for computing resonance states by Jolicard and Austin[111], and placed on a firm theoretical footing for this purpose by Riss and Meyer[104]. They have since been combined with several many-body electronic structure methods, including configuration interaction (CI)[121–123], symmetry-adapted cluster-configuration interaction (SAC-CI)[125], Fock-space coupled cluster[126], density functional theory (DFT) [127], algebraic diagrammatic construction (ADC)[128, 129], and multi-reference perturbation theory[130–132].

In this work, we present an implementation of the CAP equation-of-motion coupled-cluster (CAP-EOM-CC) method which uses the CAP subspace projection scheme [122, 125, 129, 130, 132] to investigate shape resonances in molecular systems. Our main focus is on the variant of EOM-CCSD for electron attachment energies (EOM-EA-CCSD), however, our implementation is completely general, and can be applied to other types of target states (e.g. EE/SF/IP etc.) and EOM-CC models which incorporate higher levels of excitation, provided that one-particle density and transition density matrices are available. Here, we show that projected CAP-EOM-CCSD provides a computationally efficient alternative to the “full” CAP-EOM-CCSD when the CAP is introduced at the Hartree-Fock stage[116, 119], with practically no loss in accuracy for zero-order results, and relatively minor differences between the two methods for first-order results.

The structure of this Chapter is as follows: In Sec. 4.3 we briefly outline the CAP

formalism (Sec. 4.3), and then discuss CAP-EOM-CCSD and its implementation in the projected scheme (Sec. 4.3). In Sec. 4.4 we describe the relevant computational details. In Sec. 4.5 we present the comparison of the projected and full CAP-EOM-CCSD methods for describing resonance positions and widths of well-known  $\pi^*$  shape resonances of  $N_2^-$  and  $CO^-$ . Sec. 4.5 reports the resonance parameters obtained for polyatomic molecules, including 1- and 2-cyanonaphthalene, which were recently identified in the interstellar medium [186]. Finally, the results are summarized in Sec. 4.6.

### 4.3 Theory

#### Complex Absorbing Potential

The CAP method is aimed at computing the complex Siegert-Gamow energy[6]

$$E = E_R - i\Gamma/2 \tag{4.1}$$

of a resonance, where  $E_R$  and  $\Gamma$  are the resonance position and width, respectively. To this end, the physical Hamiltonian  $H$  is augmented with a complex potential ( $-i\eta W$ ) which absorbs the outgoing tail of the resonance wave function, transforming it into a square-integrable eigenstate of the non-Hermitian Hamiltonian [104]. The resulting CAP-augmented Hamiltonian is written as follows:

$$H_{CAP} = H - i\eta W, \tag{4.2}$$

where  $\eta$  is the CAP strength parameter, and  $W$  is a real potential which vanishes in the vicinity of the molecular system and grows with distance. In this work,  $W$  is chosen as either a quadratic box-potential [133] or smooth Voronoi potential [135]. The box potential has the following form:

$$W = W_x + W_y + W_z \tag{4.3}$$

$$W_\alpha = \begin{cases} 0 & |r_\alpha| \leq R_\alpha^0 \\ (r_\alpha - R_\alpha^0)^2 & |r_\alpha| > R_\alpha^0 \end{cases}$$

where  $\alpha = x, y$ , or  $z$ , and the parameters  $R_\alpha^0$  define the onsets of the CAP along the Cartesian axes [8, 133]. The smooth Voronoi potential wraps around the molecular system at a specified cutoff radius  $r_{cut}$ ; details of its functional form can be found in Ref. [135].

In a complete one-electron basis set, the eigenvalue associated with a resonance converges to the exact Siegert-Gamow energy in the limit of an infinitesimal CAP strength [104]. With a finite basis set, the limit  $\eta \rightarrow 0^+$  yields only the real energy of a discretized continuum state. As such, in practical calculations employing finite basis sets, the approximate Siegert-Gamow energy is obtained by analyzing the parametric dependence of the eigenvalues of the CAP-augmented Hamiltonian on the CAP strength parameter  $\eta$ . The resonance position and width are typically extracted by analyzing the so-called  $\eta$ -trajectory: the series of eigenvalues associated with the resonance computed for a range of  $\eta$  values. The estimate of the resonance energy and width is obtained from an optimal value of the CAP strength parameter,  $\eta_{opt}$ , for which the trajectory exhibits a pronounced stabilization [104]. The  $\eta_{opt}$  for a given resonance cannot be determined *a priori*. It is state-specific, and depends on both the one-particle basis set and the CAP onset. Thus, evaluation of a single resonance position and width requires multiple eigenvalue calculations (typically at least 50-100). Specific schemes that can be used to identify  $\eta_{opt}$  and, therefore, the best estimate of the resonance position and width, are outlined in Sec. 4.3.

## CAP-EOM-CC

To accurately compute energies and lifetimes of electronic resonances, the CAP method must be combined with a reliable electronic structure method. Here, we use the EOM-CC method with single and double substitutions (EOM-CCSD) [55, 56]. In the EOM-CC framework, the target states are parameterized as:

$$|\Psi_I\rangle = \hat{R}_I e^{\hat{T}} |\Phi_0\rangle \quad (4.4)$$

where  $|\Phi_0\rangle$  is the Hartree-Fock reference,  $\hat{T}$  is the cluster operator from the coupled cluster ansatz, and  $\hat{R}_I$  is a generalized excitation operator [56].

The CAP can be introduced at different stages of the EOM-CC theory, which has led to multiple implementations. The implementation of CAP-EOM-CCSD described by Zuev et al.[119] introduces the CAP starting from the Hartree-Fock (HF) level (herein referred to as “fCAP” or “full” CAP). Implementations which introduce CAP at the CC and EOM levels, which have the advantage of not perturbing the SCF ground state, have been described by Ghosh et al.[138, 139]. In the present work, we utilize the projected CAP scheme (herein referred to as “pCAP” or projected CAP), in which the CAP-augmented Hamiltonian is represented on a subspace spanned by a small number of real eigenstates of the physical Hamiltonian [122]. This approach was first used in the context of CAP augmented configuration interaction (CAP-CI) by Sommerfeld and Santra [122], and later adapted for CAP-based multi-reference perturbation theory (MRPT)[130, 132], algebraic diagrammatic construction [129], and symmetry adapted cluster-configuration interaction (SAC-CI)[125], the latter of which is closely related to EOM-CC. While the majority of implementations of CAPs for resonances have either included it at the post-HF stage or have utilized the projected CAP scheme, we have chosen to use the fCAP implementation of Zuev et al.[119], which has been implemented in the Q-Chem program package[171], as our

reference for comparison.

In the projected CAP-EOM-CC method, one first obtains a set of  $M$  real left and right eigenstates of the similarity transformed Hamiltonian (often less than 20-30 states) which comprise the subspace or “correlated basis set”. In addition, the reduced one-particle density and transition density matrices between each pair of states must be evaluated. The CAP matrix in atomic orbital basis is calculated separately and then transformed into the “correlated basis set” as follows:

$$W_{ij}^{CB} = \begin{cases} Tr [W^{AO}\gamma^{ij}], & i \neq j \\ Tr [W^{AO}\gamma^i], & i = j \end{cases} \quad (4.5)$$

where  $\gamma^i$  is the reduced one-particle density matrix for state  $i$ , and  $\gamma^{ij}$  is the one-particle transition density matrix between states  $i$  and  $j$ . The similarity transformed Hamiltonian  $\bar{\mathbf{H}}$  is diagonal in the  $M \times M$  block of the basis states, which we refer to as  $H_0^{CB}$ . The complex, non-Hermitian projected CAP Hamiltonian is then defined as:

$$H^{CAP} = H_0^{CB} - i\eta W^{CB} \quad (4.6)$$

The search for  $\eta_{opt}$  therefore requires a single real-valued EOM-CC calculation, followed by the relatively trivial task of repeated diagonalization of the  $M \times M$   $H^{CAP}$  defined by Eq. 4.6 for different values of  $\eta$ . Thus, the entire  $\eta$ -trajectory, and therefore the best estimate of resonance position and width, is obtained at the cost of a single electronic structure calculation.

Introducing a CAP to the physical Hamiltonian (Eq. 4.2) makes the results sensitive to the CAP parameters, as the presence of the artificial potential introduces perturbations (often referred to as reflections)[104, 170]. To diminish the artifacts introduced by the CAP, one can employ  $\eta$ -trajectories corrected for the presence of CAP, typically up to the first order. The zero-order trajectory ( $E(\eta)$ ) corresponds to the bare eigenvalues of the CAP-augmented Hamiltonian. In the first order, the



complex energy has the form [104]:

$$U(\eta) = E(\eta) - \eta \frac{dE}{d\eta}, \quad (4.7)$$

The derivative  $\frac{dE}{d\eta}$  can be evaluated using finite differences, or it can be recast as the expectation value of the CAP via the generalized Hellman-Feynman theorem[104, 115]:

$$\frac{dE}{d\eta} = -i (\psi(\eta)|W|\psi(\eta)) \quad (4.8)$$

which leads to the convenient expression:

$$U(\eta) = E(\eta) + i\eta \text{Tr}[\gamma W], \quad (4.9)$$

where  $\gamma$  is the reduced one-particle density matrix[116]. Note that the parentheses in Eq. 4.8 indicate the use of the complex symmetric scalar product (c-product) instead of the usual Hermitian inner-product[104].

The  $\eta_{opt}$  for zero-order trajectories can be found using the logarithmic velocity criterion:

$$\left| \eta \frac{dE}{d\eta} \right| \rightarrow \min \quad (4.10)$$

The same approach can be used for finding the  $\eta_{opt}$  for the first-order corrected trajectories by minimizing the logarithmic velocity of  $U(\eta)$ . A different scheme has been proposed by Jagau et al.[116], where separate values of  $\eta_{opt}$  for real and imaginary parts are obtained by locating stationary points for the real and imaginary parts of  $U(\eta)$  independently:

$$U^R(\eta) = E^R(\eta) - \eta \text{Tr} [\gamma^I W] \quad (4.11)$$

$$U^I(\eta) = E^I(\eta) + \eta \text{Tr} [\gamma^R W] \quad (4.12)$$

In this work, we obtain a single  $\eta_{opt}$  for first-order trajectories from the minimum of

the logarithmic velocity of the first-order trajectory.

$$\left| \eta \frac{dU}{d\eta} \right| \rightarrow \min \quad (4.13)$$

The real and imaginary parts of  $U(\eta_{opt})$  determine the resonance energy and width, respectively.

#### 4.4 Computational Details

We have implemented the projected CAP-EOM(EA/EE/SF/IP) family of methods in the Q-Chem program package, and all calculations were performed using version 5.4 [171]. In this work, we have focused on the EOM-EA-CCSD variant for electron affinities, which is well-suited for describing (N+1)-electron target states. We have used both the “full” CAP-EOM-EA-CCSD method of Zuev et al.[119][119] and the new projected CAP-EOM-EA-CCSD. For the rest of the paper, these methods will be referred to as “fCAP” and “pCAP”, respectively.

Experimental ground state geometries obtained from the NIST Computational Chemistry Comparison and Benchmark Database were used for  $N_2$ ,  $CO$ ,  $CO_2$ ,  $HCN$ , formaldehyde ( $CH_2O$ ), and ethylene( $C_2H_4$ ) [187]. For *para*-benzoquinone (pBQ), the optimized geometry of the ground state radical anion from Ref. [188] was used. For chloroethene, the geometries from Ref. [189] were used. For all remaining systems, the optimized geometry of the neutral ground state at the B3LYP/cc-pVTZ level of theory was used [190, 191]. The frozen-core approximation was utilized for calculations on uracil, pBQ, and the naphthalene derivatives, while all orbitals were correlated in the remaining calculations.

The basis sets used in CAP-EOM-CCSD calculations consist of Dunning’s correlation consistent (aug)-cc-pVXZ [176] basis sets augmented by additional even-tempered diffuse basis functions. Following Ref. [119], we explored two different

types of augmentation. The first, denoted as [A], augments the basis set for all heavy atoms, and the second, denoted as [C], places a set of diffuse functions with exponents averaged over those for different chemical elements located at the center of mass of the molecule[119]. In most cases, the exponents for the additional diffuse functions were obtained as one half of the exponent of the preceding function of the same angular momentum, starting from the most diffuse function in the parent basis set[119]. The exception was the 2s5p2d[A] basis set[125], for which we generated the exponents for the additional basis functions in a similar way, but used a scaling factor of 1.5 for p-type functions, and 2.0 for s and d type functions.

Unless otherwise specified, we used the aug-cc-pVTZ + 3s3p3d[C] basis set for calculations on  $N_2$ ,  $CO$ ,  $CO_2$ ,  $HCN$ ,  $CH_2O$ , and  $C_2H_4$ . For pBQ, we used the cc-pVTZ + 8gh [3s] basis set from Ref. [188]. For chloroethene, we used the aug-cc-pVDZ + 3p basis from Ref [189]. For uracil, naphthalene, and 1- and 2-cyanonaphthalene, we used the cc-pVDZ + 2s5p2d[A] basis set[104, 191]. All CAP integrals were evaluated through numerical quadrature using a Becke-type grid of 99 radial points and 500 angular Lebedev points. [177, 192].

Multiple forms of CAP were used in this work, and the CAP parameters for each system are summarized in Table S1 in the supplementary material. Unless otherwise specified, a quadratic box-potential[133] was used, and the onset in each Cartesian coordinate was chosen as the square root of the expectation value  $\langle \alpha^2 \rangle$  ( $\alpha = x, y, z$ ) for the ground state calculated at the CCSD level of theory[119]. For calculations on uracil, naphthalene, and 1- and 2-cyanonaphthalene, a smooth Voronoi potential[135] with a cutoff radius of 4.0 bohr was used.

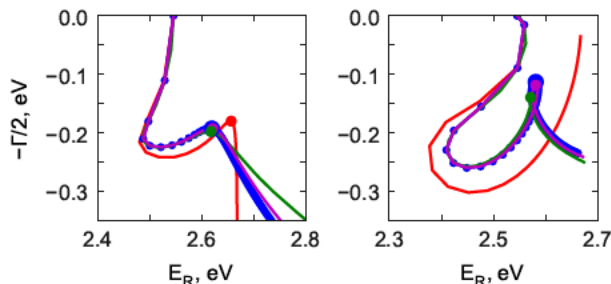
For zero-order trajectories,  $\eta_{opt}$  was determined following Eq. 4.10. For first-order results, we computed the first-order correction using Eq. 4.9, and a single  $\eta_{opt}$  for the first-order trajectory was obtained using Eq. 4.13. For pCAP, the density matrix

$\gamma$  in Eq. 4.9 was computed as a linear combination of the densities and transition densities for the zero-order states, which can be evaluated using only the  $\eta$ -dependent eigenvectors and the matrix elements of the CAP in the correlated basis (see Eq. S1 in the supplementary material).

## 4.5 Results and Discussion

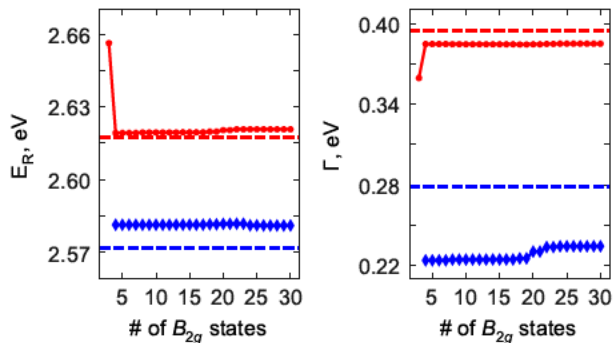
### $N_2^-$ and $CO^-$ $\pi^*$ shape resonances.

The well known  $\pi^*$  shape resonances of  $N_2^-$  and  $CO^-$  were used to thoroughly test the performance of pCAP and to analyze the dependence on computational parameters, such as CAP onset and choice of one-electron basis set.



**Figure 4.1:** Comparison of the zero-order (left) and the first-order (right) trajectories for the  $^2\Pi_g$  shape resonance in  $N_2^-$ , computed with pCAP with 3 states (red), 4 states (blue), 30 states (magenta), and fCAP (green). Points associated with  $\eta_{opt}$  are indicated by a  $\bullet$ .  $\eta$  was varied from 0.0 to 0.05.

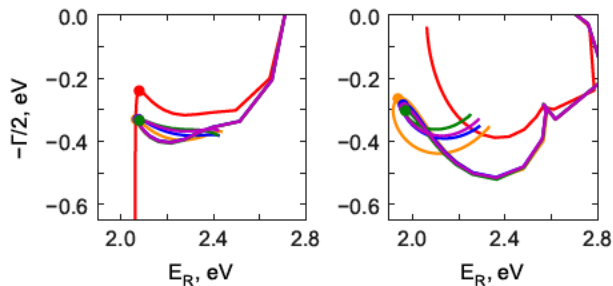
Zero-order trajectories obtained from pCAP and fCAP for the  $^2\Pi_g$  shape resonance in  $N_2^-$  are shown in Fig. 4.1 (left panel). We first compare the overall shape of the trajectories obtained with the two methods, and then proceed to discussing the effects of the size of the correlated basis. Using the logarithmic velocity criterion (Eq. 4.10), the zero-order resonance parameters are  $E_R = 2.619$  and  $\Gamma = 0.385$  eV at  $\eta_{opt} = 0.007$  for pCAP, and  $E_R = 2.617$  and  $\Gamma = 0.395$  eV at  $\eta_{opt} = 0.007$  for fCAP. The zero-order trajectories are nearly identical at smaller CAP strengths, and only begin to differ



**Figure 4-2:** Dependence of the zero-order (red) and first-order (blue) resonance parameters on the size of the correlated basis for the  ${}^2\Pi_g$  shape resonance in  $N_2^-$ . The result obtained from fCAP is indicated by a dashed line.

noticeably beyond the stationary point, after which both trajectories move linearly into the complex plane. In the right panel of Fig. 4-1, we plot first-order trajectories. For fCAP, the first-order resonance parameters are  $E_R = 2.572$  and  $\Gamma = 0.279$  at  $\eta_{opt} = 0.0175$ . For pCAP, the first-order resonance parameters are  $E_R = 2.581$  and  $\Gamma = 0.226$  at  $\eta_{opt} = 0.0172$ . The slight differences between these two results can be explained by the differences in the shape of the zero-order trajectories beyond the stabilization point, as the steeper slope of the zero-order trajectory obtained from pCAP leads to a slightly larger correction to the energy. Lastly, we note that computing the derivative in Eq. 4.7 by finite differences or by using Eq. 4.9 yields nearly identical first-order trajectories (see Fig. S1 in the supplementary material).

Next, we explored the effect of the the size of the correlated basis on the computed parameters of the  $N_2^-$  and  $CO^- \pi^*$  resonances. An ideal subspace is one for which the total overlap between the true resonance wave function and the set of zero-order states is close to 1.[122] It is generally not possible to know *a priori* which or how many zero-order states are needed, and, therefore, a common procedure is to include all states up to some chosen energy threshold in the correlated basis set[125, 130]. Here, we scanned over a range of 2 to 30 eigenstates of the appropriate target irreducible



**Figure 4.3:** Comparison of zero-order (left) and first-order (right) trajectories for the  ${}^2\Pi$  shape resonance in  $CO^-$ , computed with pCAP with 5 states (red), 7 states (orange), 10 states (blue), 30 states (magenta), and fCAP (green). Points associated with  $\eta_{opt}$  are indicated by a  $\bullet$ .  $\eta$  was varied from 0.0 to 0.02.

representation (e.g.  $B_{2g}$  for  $N_2^-$  assuming  $D_{2h}$  point group).

Starting with  $N_2^-$  (Fig. 4.2), a resonance trajectory with a clear stationary point first appears upon inclusion of the lowest 3 left and right eigenstates of  $\bar{H}$ , which span an energy range of  $\sim 1.3$  eV to  $\sim 4.1$  eV for  $\eta = 0$ . While the zero-order trajectory does show stationary behavior with 3 states in the correlated basis, the first-order trajectory does not, and instead accelerates towards the real axis. Inclusion of the next lowest eigenstate ( $\sim 8.1$  eV) resolves this issue. Beyond this point, the change in the eigenvalue trajectories upon inclusion of an additional state is nearly imperceptible, and as a result, the zero-order and first-order parameters essentially converge after inclusion of the fourth state. Since this is a shape resonance, one would expect the resonance wave function to have predominantly one-particle (1p) character with respect to the closed shell reference, and unsurprisingly, the four lowest states which prove crucial to the description of the resonance wave function are indeed all dominated by 1p terms. The remaining 26 states have predominantly two-particle-one-hole (2p1h) character, so their inclusion does little to improve the correlated basis set. This is similar to the trends reported in Ref. [129] for projected CAP-EA-ADC.

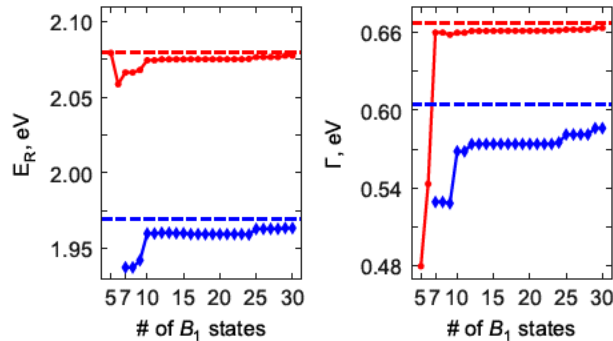
For  $CO^-$  (Figs. 4.3 and 4.4), a clear stationary point on the zero-order trajectory

**Table 4.1:** Computed resonance energy (width),  $E_R$  ( $\Gamma$ ), for  $N_2^-$  and  $CO^-$   $\pi^*$  shape resonances. All quantities are given in eV.

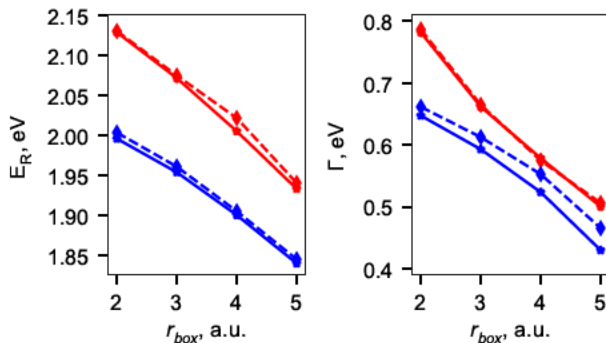
	$N_2^-$	$CO^-$
pCAP		
# states	4	7
Zero-order	2.619(0.385)	2.066(0.660)
First-order	2.581(0.226)	1.937(0.530)
fCAP		
Zero-order	2.617(0.395)	2.080(0.668)
First-order	2.572(0.279)	1.969(0.605)
Experiment	2.32(0.41)*[193]	1.50(0.8)[182]

\* Estimated using theoretical model fitted to experimental data

first appears upon inclusion of 5 states, and first-order trajectories begin displaying stationary behavior once 7 states are included in the correlated basis set. Similar to  $N_2^-$ , these 7 lowest lying states have predominantly 1p character, and represent the minimal subspace required to sufficiently represent the resonance wave function. However, unlike  $N_2^-$ , convergence with respect to the number of states is not as immediate. A noticeable jump in both position and width (which is more pronounced



**Figure 4-4:** Dependence of the zero-order (red) and first-order (blue) resonance parameters on the size of the correlated basis set for the  $^2\Pi$  resonance in  $CO^-$ . The results obtained from fCAP are indicated by a dashed line.



**Figure 4-5:** Dependence of the zero-order (red) and first-order (blue) resonance parameters on the box size for the  ${}^2\Pi$  resonance in  $CO^-$ . Results obtained from pCAP are plotted with solid lines, while fCAP results are plotted with dashed lines.

for the first-order trajectory) appears upon inclusion of the 10th state, which has mixed 1p/2p1h character. This state non-trivially overlaps with the initial set of 1p states, and provides a tangible improvement to the description of the resonance wave function, which manifests in a more accurate shape of the zero-order trajectory, and therefore, a more accurate first-order width. All remaining states do not contribute substantially, and the resonance parameters change very little with adding more states beyond the 10th to the correlated basis. The  $\eta_{opt}$  for zero-order and first-order trajectories does not change as the dimension of the correlated basis set is increased from 7 to 30 states, despite the jump in energy upon inclusion of the 10th state.

The results obtained from the smallest correlated basis set required to achieve stabilized first-order trajectories are summarized in Table 4.1, and the difference from the largest 30 state subspace is less than 0.1 eV for both zero-order and first-order resonance parameters.

Considering that the convergence of the resonance parameters is achieved at a small size of the correlated basis, the pCAP approach leads to dramatic decrease in the computational time. For example, a real EOM-EA-CCSD/aug-cc-pVTZ+3s3p3d[C] calculation for  $N_2^-$  done on a single 2.90 GHz Intel<sup>®</sup> Xeon<sup>®</sup> E5-E5-2690 processor



takes 54 and 524 seconds for 4 states and 30 states, respectively. For pCAP, since the EOM-EA-CCSD eigenvectors are only computed once, these estimates essentially represent the entire cost of generating an eigenvalue trajectory for any range of  $\eta$  values for the corresponding subspace (4 or 30 eigenvectors). For fCAP, each value of  $\eta$  requires an independent calculation. For this example, with 200  $\eta$ -points and approximately 300 seconds per  $\eta$  value (e.g. the timings for  $\eta = 0, 0.015$  and  $0.05$  were 212 s, 357 s, and 388 s, respectively, when requesting 4 states), generating the fCAP eigenvalue trajectory required a total of  $\sim 17$  CPU hours.

Next, we investigated the effect of the CAP onset on computed resonance parameters. Zero-order and first-order results for  $\text{CO}^-$  with different box sizes are shown in Fig. 4.5. The box size in Fig. 4.5 is specified by a single onset parameter,  $r_{box}$ , such that the onset in each dimension,  $R_\alpha^0$  in Eq. 4.3, is obtained by adding  $r_{box}$  to the maximum value of the nuclear coordinate in that dimension, with the origin placed at the center of mass. The two methods show a similar monotonic dependence on the CAP onset for both real and imaginary parts of the energy, and the resonance parameters computed from pCAP and fCAP lie within 0.01 eV and 0.05 eV of each other for the position and width respectively. Importantly, these differences are less than the shifts of the results due to the variation in the CAP onset, which changes the resonance position by as much as 0.19 eV, and the width by as much as 0.28 eV as the box is moved from 2.0 bohr to 5.0 bohr. Similar trends were observed for  $\text{N}_2^-$  (see Fig. S2 in the supplementary material), indicating that the relative performance of the two methods is weakly dependent on the CAP onset.

Next, we investigated the sensitivity of the results for  $\text{N}_2^-$  to the valence and diffuse parts of the one-electron basis set. For each basis set, we calculated up to 30 states (the corresponding trajectories and resonance parameters dependence on the size of the correlated basis set are given in the supplementary material, Sec. SV). The

**Table 4.2:** Basis set dependence of the computed zero-order resonance parameters for the  $N_2^-$   $\pi^*$  shape resonance. All energies are given in eV.

Basis set	# States	pCAP			fCAP		
		$E_R$	$\Gamma$	$\eta_{opt}$	$E_R$	$\Gamma$	$\eta_{opt}$
aug-cc-pVDZ+3s3p3d[C]	3	2.775	0.373	0.0055	2.778	0.381	0.0056
aug-cc-pVTZ+3s3p3d[C]	4	2.619	0.385	0.0070	2.617	0.395	0.0070
aug-cc-pVQZ+3s3p3d[C]	4	2.507	0.360	0.0039	2.508	0.369	0.0038
aug-cc-pV5Z+3s3p3d[C]	5	2.454	0.377	0.0026	2.456	0.384	0.0024
aug-cc-pVTZ+3s3p3d[A]	7	2.526	0.500	0.0015	2.524	0.494	0.0015
aug-cc-pVTZ+6s6p6d[C]	7	2.536	0.423	0.0023	2.536	0.425	0.0023
cc-pVTZ+2s5p2d[A]	4	2.518	0.499	0.0016	2.518	0.498	0.0016
aug-cc-pVTZ+3s3p[A]	5	2.601	0.360	0.0059	2.604	0.370	0.0060
aug-cc-pVTZ+3s3p[C]	2	2.901	0.758	0.0400	2.922	0.707	0.0410

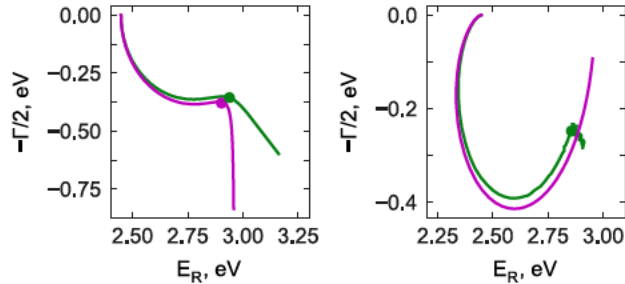
number of states in Tables 4.2 and 4.3 refers to the initial set of eigenstates with 1p character which appear before the first 2p1h state, and the reported data corresponds to the correlated basis set defined by these states.

In all cases except the aug-cc-pVTZ + 3s3p[C] basis set, the eigenvalue trajectories obtained from pCAP and fCAP have nearly identical structures, and yield zero-order resonance parameters and first-order resonance positions within 0.01 eV. First-order resonance widths obtained from pCAP are systematically lower than those from fCAP, though there is no clear trend between the size of the basis set and the agreement between the two methods. Increasing the size of the correlated basis set (and thereby including additional states of mixed 1p/2p1h character) does typically lead to better agreement with the fCAP result, though even with the minimal subspace, the first-order widths obtained from the two methods differ by less than 0.1 eV. For this system, increasing the quality of the valence basis set does lead to better agreement of the first-order results (though not the zero-order ones) with the values from Ref. [193], which were obtained using a parameterized model fitted to the experimental

**Table 4 3:** Basis set dependence of the computed first-order resonance parameters for the  $N_2^-$   $\pi^*$  shape resonance. All energies are given in eV.

Basis set	# States	pCAP			fCAP		
		$E_R$	$\Gamma$	$\eta_{opt}$	$E_R$	$\Gamma$	$\eta_{opt}$
aug-cc-pVDZ+3s3p3d[C]	3	2.727	0.200	0.0135	2.718	0.228	0.0124
aug-cc-pVTZ+3s3p3d[C]	4	2.581	0.226	0.0172	2.572	0.279	0.0175
aug-cc-pVQZ+3s3p3d[C]	4	2.455	0.256	0.0079	2.458	0.298	0.0072
aug-cc-pV5Z+3s3p3d[C]	5	2.394	0.322	0.0048	2.399	0.348	0.0044
aug-cc-pVTZ+3s3p3d[A]	7	2.519	0.310	0.0085	2.508	0.348	0.0075
aug-cc-pVTZ+6s6p6d[C]	7	2.601	0.228	0.0210	2.585	0.290	0.0180
cc-pVTZ+2s5p2d[A]	4	2.594	0.274	0.0132	2.589	0.342	0.0108
aug-cc-pVTZ+3s3p[A]	5	2.560	0.191	0.0141	2.556	0.235	0.0120
aug-cc-pVTZ+3s3p[C] <sup>a</sup>	2	–	–	–	2.860	0.499	0.0750
Berman et al <sup>193</sup>	2.32(0 41)						

<sup>a</sup> no stationary point could be located for first-order pCAP trajectory.



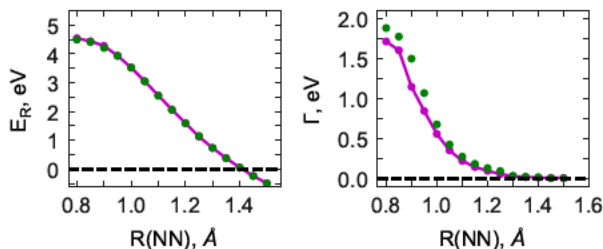
**Figure 4-6:** Comparison of zero-order (left) and first-order (right) trajectories for  $N_2^-$  with the aug-cc-pvtz + 3s3p[C] basis set, computed with pCAP (magenta) and fCAP (green). Points associated with  $\eta_{opt}$  are indicated by a  $\bullet$ .  $\eta$  was varied from 0.0 to 0.15. 30 states were included in the subspace for the pCAP trajectories.

$N_2-e^-$  cross section, and are often referred to as “experimental”. However, we did not see improved agreement with the estimates extracted from the experimental cross section[194] for  $CO^-$  when we increased the quality of the valence basis set (see Table S3 in the supplementary material).[182]

The projection scheme breaks down in the case of the aug-cc-pVTZ + 3s3p[C]

basis set, which was the smallest diffuse basis set we explored. While reasonable agreement (though poorer than the other basis sets) was found for zero-order results, no stationary point could be identified for the first-order trajectory for pCAP. As shown in Fig. 4-6, pCAP qualitatively fails to reproduce the  $\eta$ -dependence of fCAP beyond the stationary point of the zero-order trajectory, resulting in a first-order trajectory which accelerates towards the real axis rather than displaying any stationary behavior. Increasing the size of the correlated basis has little effect; among the lowest 30 eigenstates (spanning an energy range up to 13.8 eV), only the first two are of primarily 1p character, resulting in a poor description of the resonance wave function.

Resonance positions and widths obtained from CAP-EOM-CCSD can be used to obtain complex potential energy surfaces. In Fig. 4-7, we compare the complex potential energy curves for  $N_2^-$  computed by first-order pCAP and fCAP. Following Ref. [118], we used a fixed CAP onset (see Table S1 in the supplementary material), and focus on first-order results, which are necessary to yield smooth curves in regions where the resonance is narrow (e.g. near the crossing point with the neutral ground state)[118]. From 1.0-1.4 Å (where the resonance becomes bound), pCAP first-order results and trajectories show excellent agreement with fCAP, but at smaller bond distances, the trajectories become more difficult to interpret, and the strategy used

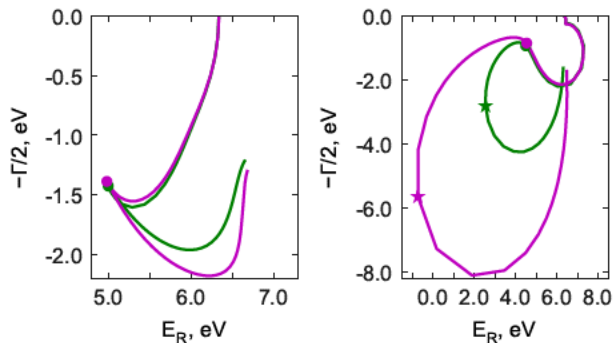


**Figure 4-7:** First-order resonance position (left) and width (right) along complex PEC for the  ${}^2\Pi_g$  shape resonance in  $N_2^-$  computed by pCAP (magenta) and fCAP (green). For pCAP, 10 states were included in the correlated basis at all points.

for determining  $\eta_{opt}$  becomes crucial. Jagau et al. [118] obtained smooth complex potential energy curves for this system using fCAP and the same basis set by evaluating  $\eta_{opt}$  for the real and imaginary parts of the energy separately, with the width taken as a maximum of  $U^I(\eta)$ , and the resonance position a minimum of  $U^R(\eta)$  (see Eqs. 4.11 and 4.12). While this strategy works well for pCAP from 1.0-1.4 Å, at smaller bond distances, the structures of the first-order pCAP trajectories begin to noticeably differ from fCAP (see Fig. S12 in the supplementary material), leading to drastically different results between pCAP and fCAP for the resonance position if one uses this criteria. For example, as shown in Fig. 4.8, at a bond distance of 0.80 Å, the pCAP resonance position obtained by locating a minimum of  $U^R(\eta)$  would be -0.80 eV, a value which is obviously not physically meaningful, and the fCAP resonance position would be evaluated as  $\sim 2.5$  eV, a value which is  $\sim 2$  eV less than the zero-order resonance position. When the first-order energy for both pCAP and fCAP is evaluated using Eq. 4.13, there is no such discrepancy, and the two methods yield similar first-order results even at bond distances less than 1.0 Å. Lastly, we note that zero-order results and trajectories for pCAP and fCAP are very similar for all points we explored, but resonance parameters cannot be obtained for either method beyond 1.3 Å, as the resonance becomes too narrow to determine an optimal CAP strength using Eq. 4.10 (see Fig. S11 in the supplementary material).

### Shape resonances in polyatomic molecules

Below, we discuss the performance of the projected CAP-EOM-EA-CCSD scheme for evaluation of the resonance position and width for low-lying shape-resonances in several representative polyatomic systems. Computed zero-order and first-order resonance positions and widths are listed in Tables 4.4 and 4.5. In all cases, eigenvalue trajectories obtained from the two methods have the same general structure once the initial set of 1p-dominated states (the second column in Table 4.4) are included in



**Figure 4.8:** Comparison of zero-order (left) and first-order (right) trajectories for the  ${}^2\Pi_g$  shape resonance in  $N_2^-$  at  $R(\text{N-N}) = 0.80 \text{ \AA}$ . Results computed with pCAP (10 states) are plotted in magenta, and results from fCAP are plotted in green. Points associated with  $\eta_{opt}$  determined by Eq. 4.13 are indicated by a  $\bullet$ , and points associated with  $\eta_{opt}$  using the criteria from Ref. [118] are indicated by a  $\star$ .  $\eta$  was varied from 0.0 to 0.05.

the correlated basis. The agreement between pCAP and fCAP for zero-order and first-order resonance positions is excellent; the two methods yield values that differ by no more than 0.02 eV. Expanding the subspace only slightly improves the first-order widths, which tend to be underestimated by pCAP relative to fCAP. Increasing the size of correlated basis led to greater shifts in computed resonance parameters for polar molecules ( $CO$ ,  $HCN$ ,  $CH_2O$ ) than for non-polar ones ( $N_2$ ,  $CO_2$ ,  $C_2H_4$ ,  $pBQ$ ) due to the presence of higher lying states with 1p character for the former. Note, however, that our results obtained for  $N_2^-$  and  $CO^-$  with different one-electron basis sets clearly show that the dependence on the size of correlated basis is itself dependent on the choice one electron basis set. Nonetheless, it is clear that the strategy of including the initial set of 1p states works quite well for these isolated shape resonances in small to medium size molecules.

In addition to isolated resonances, the subspace projection scheme is able to handle the case of multiple resonances belonging to the same symmetry. In Table 4.6, we summarize pCAP results for the  $\pi^*$  resonances of uracil, which all belong to  $A''$

**Table 4.4:** Zero-order resonance parameters for the lowest electron-attached  $\pi^*$  shape resonances in polyatomic molecules. All energies are given in eV.

System	# states	Projected		Full	
		$E_R$	$\Gamma$	$E_R$	$\Gamma$
<i>HCN</i>	8	2.145	1.197	2.163	1.239
<i>CO<sub>2</sub></i>	6	4.017	0.178	4.006	0.171
<i>CH<sub>2</sub>O</i>	7	1.202	0.322	1.205	0.328
<i>C<sub>2</sub>H<sub>4</sub></i>	3	2.012	0.432	2.012	0.439
<i>pBQ<sup>a</sup></i>	4	2.871	0.034	2.89 <sup>188</sup>	0.027

<sup>a</sup>  $E_R$  is reported relative to CCSD energy of ground state anion.

**Table 4.5:** First-order resonance parameters for the lowest electron-attached  $\pi^*$  shape resonances in polyatomic molecules. All energies are given in eV.

System	Projected		Full		Experiment	
	$E_R$	$\Gamma$	$E_R$	$\Gamma$	$E_R$	$\Gamma$
<i>HCN</i>	1.969	0.820	1.949	0.957	2.260 <sup>195</sup>	-
<i>CO<sub>2</sub></i>	4.000	0.111	3.989	0.116	3.6 <sup>196</sup>	-
<i>CH<sub>2</sub>O</i>	1.154	0.233	1.155	0.268	0.86 <sup>11, 197</sup>	-
<i>C<sub>2</sub>H<sub>4</sub></i>	1.932	0.306	1.932	0.347	1.8 <sup>198, 199</sup>	0.7
pBQ	2.867	0.012	2.87 <sup>188</sup>	0.013	2.5 <sup>200</sup>	0.025

symmetry in the  $C_s$  point group. The results we report in Table 4.6 were obtained using a smooth Voronoi<sup>[135]</sup> CAP with a cutoff radius of 4 bohr, and we included up to 55 states in our calculations.

Starting with the  $\pi_1^*$  state, our reported pCAP resonance position lies 0.15 eV above the projected CAP/SAC-CI resonance position reported by Sommerfeld et al. <sup>[136]</sup>. However, our obtained pCAP resonance parameters agree well with the fCAP results reported by Thodika et al. <sup>[191]</sup>. The minor differences between the fCAP and pCAP results in Table 4.6 are simply due to the different choice of CAP; pCAP nearly perfectly reproduces fCAP resonance parameters when the same CAP

is used (see Table S11 in the supplementary material). For the  $\pi_2^*$  state, EOM-CC and SAC-CI show much better agreement, but the resonance position still lies 0.6 eV above the experimental ETS result and the results obtained from electron scattering calculations [201–203]. The present deviations of 0.5–0.6 eV for the  $\pi_1^*$  and  $\pi_2^*$  states are consistent with previous CAP/SAC-CI studies on double-bonded and heteroaromatic compounds, and such deviations have been attributed to geometry relaxation effects and the incomplete one-electron basis set [136, 190]. With the chosen one-electron basis set, the first 22 eigenstates have primarily 1p character, and this subspace is more than sufficient to describe these first two states (see Sec. SX in the supplementary material). A satisfactory first-order trajectory appears with 7 states for the  $\pi_1^*$  state, and 19 states for the  $\pi_2^*$ , and the resonance parameters change very little for both states when the size of the subspace increases any further.

The behavior of the  $\pi_3^*$  state is more complicated, and exhibits a strong dependence on the subspace size. One can expect that the projected scheme applied to anions can converge slower for resonances lying above or near the first excited state of the neutral molecule owing to interaction with another continuum associated with the excited state decay channel in addition to the ground state one. When 20 states are included, a stabilized first-order trajectory originating ( $\eta = 0$ ) at  $\sim 5.0$  eV can possibly be attributed to the  $\pi_3^*$  resonance (see Fig. S22 in the supplementary material), and the zero-order and first-order trajectories obtained from this subspace strongly resemble the projected CAP/SAC-CI trajectories reported in Ref. [136]. However, inclusion of the next two states, which both have 1p-character, leads to pronounced changes in the trajectories: the zero-order trajectory, which previously was not stabilized, shows stationary behavior near 5.1 eV ( $\Gamma \approx 1.2$  eV), and the stationary point on the first-order trajectory vanishes. Inclusion of the 23rd state, which has 2p1h-character, shifts the zero-order resonance position even further to  $\sim 5.3$  eV, and brings down the



**Table 4.6:** Resonance positions and widths of three lowest  $\pi^*$  resonances in uracil. All energies are given in eV.

Method	$\pi_1^*$		$\pi_2^*$		$\pi_3^*$	
	$E_R$	$\Gamma$	$E_R$	$\Gamma$	$E_R$	$\Gamma$
Zero-order	0.731	0.05	2.284	0.232	5.31	0.818
First-order	0.726	0.042	2.258	0.17	-	-
fCAP (First-order) <sup>191</sup>	0.718	0.033	-	-	-	-
CAP/SAC-CI (First-order) <sup>136</sup>	0.57	0.05	2.21	0.10	4.82	0.58
SMCPP <sup>202</sup>	0.14	0.005	1.76	0.15	4.83	0.78
R-Matrix <sup>201</sup>	0.13	0.003	1.94	0.17	4.95	0.38
Experiment <sup>203</sup>	0.22	-	1.58	-	3.83	-

resonance width to 0.82 eV. Including additional states beyond this point has little effect, and we were unable to obtain a stabilized first-order trajectory when including up to 55 states. This 23rd state is actually a decay channel which corresponds to the first  $A'$  excited state of the neutral molecule, which at the EOM-EE-CCSD level of theory has an excitation energy of 5.56 eV, but this is shifted to 7.36 eV in EOM-EA-CCSD, likely due to higher levels of excitation being required for the EOM-EA-CC ansatz to accurately represent this state. While further studies are necessary to determine the nature of interaction between  $\pi_3^*$  resonance and the excited decay channel, in the present work, we consider the zero-order result in Table 4.6 to be our best estimate of the resonance parameters of the  $\pi_3^*$  state using this method and one-electron basis set.

As a pilot application of the pCAP method, we explore low-lying  $\pi^*$  resonances in polycyclic aromatic hydrocarbons (PAHs). In astrochemical environments, PAHs are believed to be widely abundant, and are thought to carry up to 25 percent of all carbon in the universe [21, 204]. The formation of PAH anions in space has been a frequent topic of study, due to their perceived abundance and their vacant  $\pi^*$  orbitals, which could give them the ability to soak up free electrons. Low-lying shape

**Table 4.7:** First-order positions and widths of three lowest  $\pi^*$  states of naphthalene and 1- and 2-cyanonaphthalene. All energies are given in eV.

Molecule	$\pi_1^*(^2B_{2g})$	$\pi_2^*(^2B_{1g})$	$\pi_3^*(^2B_{3u})$
1-cyanonaphthalene	-0.226	0.759(0.019)	1.423(0.119)
2-cyanonaphthalene	-0.146	0.500(0.016)	1.551(0.241)
naphthalene	0.674(0.015)	1.335(0.042)	2.150(0.384)
Experiment <sup>13</sup>	0.19	0.90	1.67

resonances can potentially serve as gateway states to stable anions, particularly in species which lack a sufficient dipole moment to support dipole-bound states, which are widely believed to be involved in such processes [21]. Recently, the first two PAHs, 1- and 2-cyanonaphthalene, were conclusively identified in the TMC-1 molecular cloud [186]. The dipole-bound states of these two molecules were recently investigated by Ref. [205], but to our knowledge, no experimental or theoretical data exists for their resonances. In Table 4.7 we report first-order positions and widths for three  $\pi^*$  states in 1- and 2-cyanonaphthalene and in the parent naphthalene, along with the experimental electron transmission spectroscopy (ETS) data for naphthalene by Burrow et al. [13]. The symmetry labels in Table 4.7 correspond to the assignments in Ref. [13] of the 1p states of naphthalene. The states of 1- and 2-cyanonaphthalene all belong to  $A''$  symmetry in the  $C_s$  point group, and we group them with the analogous states of the parent naphthalene based on their character. Starting with naphthalene, the resonance energies obtained by Burrow et al. [13] were originally presumed to be associated with vertical transitions, but later studies have ascribed those values to adiabatic transitions, thus making geometry relaxation effects and zero point vibrational energy (ZPVE) crucial [206–208]. Our computed resonance energy for the  $^2B_{2g}$  state of naphthalene agrees well with the EOM-EA-CCSD/aug-cc-pVDZ value of 0.63 eV by Shelton et al. [207], and when the "pseudo adiabatic" corrections

for naphthalene from Ref. [208] are applied, we obtain a resonance position of 0.33 eV. Indeed, our computed resonance positions for all three resonances of naphthalene are  $\sim 0.5$  eV above the ETS measurements, and so we attribute these deviations to the aforementioned geometry relaxation effects and ZPVE. The metastable  ${}^2B_{2g}$  state of the naphthalene anion becomes bound in the cyano-derivatives due to the electron withdrawing CN groups. Experimental measurements on these species have yielded adiabatic electron affinities of 0.68 eV and 0.65 eV for 1- and 2-cyanonaphthalene respectively[206], which is fairly consistent with our vertical estimates of 0.146 eV and 0.226 eV. The lifetimes of the  $\pi_2^*$  and  $\pi_3^*$  states of both derivatives agree well with the lifetimes of the corresponding states in naphthalene, and in both cases the presence of the cyano group brings down the vertical energies by more than 0.5 eV. While the roles of these states in astrochemistry and other physical processes has yet to be investigated, here we emphasize that our new methodology is capable of computing resonances in these large molecules in a computationally feasible manner. With the chosen one-electron basis set, and exploiting Cholesky decomposition, computing the lowest 40  $A''$  states of 1- and 2-cyanonaphthalene each took approximately 38 hours in wall time, and computing 10 states in each of the three irreducible representations for the parent naphthalene took approximately 2.5 hours in wall time using 16 threads on an 8-core 2.90 GHz Intel<sup>®</sup> Xeon<sup>®</sup> E5-E5-2690 processor.

As a final test, we consider the performance of pCAP near exceptional points, the analogues of conical intersections between metastable electronic states on complex potential energy surfaces. Our test case is the chloroethene anion, which belongs to a class of widely studied unsaturated halogenated compounds that are believed to undergo dissociative electron attachment (DEA) processes[209–212]. A proposed mechanism involves initial formation of a long lived  $\pi^*$  resonance, which subsequently couples to a  $\sigma^*$  resonance, leading to C-Cl bond dissociation[213]. This system was

**Table 4.8:** Comparison of pCAP and fCAP for  $\pi^*$  and  $\sigma^*$  states of the chloroethene anion. All energies were obtained from zero-order trajectories, and are given in eV.

Equilibrium geometry				
	Projected		Full <sup>189</sup>	
State	$E_R$	$\Gamma$	$E_R$	$\Gamma$
$\pi^*$	1.800	0.257	1.801	0.260
$\sigma^*$	2.273	1.042	2.268	1.115
MEEP				
	Projected		Full <sup>189</sup>	
State	$E_R$	$\Gamma$	$E_R$	$\Gamma$
$\pi^*$	1.396	0.240	1.381	0.267
$\sigma^*$	1.378	0.249	1.388	0.279

recently studied by Phung et al. using projected CAP/SAC-CI and projected CAP-XMS-CASPT2[130], and the minimum energy exceptional point (MEEP) between the  $\pi^*$  and  $\sigma^*$  resonance states was identified by Benda et al.[189] using analytical gradient techniques they have developed for fCAP[134]. In Table 4.8 we compare zero-order fCAP and pCAP results at the equilibrium geometry of neutral chloroethene, and the MEEP geometry identified in Ref. [189]. At the equilibrium geometry, which belongs to the  $C_s$  point group, we computed 60  $A'$  states for the  $\sigma^*$  state and 43  $A''$  states for the  $\pi^*$  state. At the MEEP geometry, where the Cl atom is bent out of the plane of the molecule (thereby allowing the two states to interact), we computed 65  $A$  states. Data from the minimal 1p subspace is provided in Table S15 in the supplementary material. We have also chosen to optimize  $\eta$  separately for both states at both geometries, while Ref. [189] used the same value of  $\eta$  for both states at the MEEP because they were unable to identify a stationary point for the  $\sigma^*$  state. Although pCAP does not quite reproduce the near exact degeneracy of fCAP at the MEEP, we are clearly able to distinguish two eigenvalue trajectories

with nearly degenerate stationary points, and agreement with fCAP at the MEEP and the equilibrium geometries suggest that the performance of the projected scheme does not significantly deteriorate near exceptional points. These results, as well as the results of Ref [130], show that projected CAP schemes are robust enough to generate complex potential energy surfaces involving multiple interacting resonance states, provided that the underlying electronic structure method is capable of accurately describing the target states.

## 4.6 Summary and Conclusions

We have presented the projected CAP-EOM-CC-CCSD method for electronic resonances, and benchmarked its performance relative to full CAP-EOM-CCSD. We have shown that a small number of EOM-EA-CCSD states (typically the initial set of 1p-dominated states) is sufficient to produce zero-order projected CAP-EOM-EA-CCSD trajectories which closely approximate those obtained from full CAP-EOM-EA-CCSD. There are more noticeable differences between the first-order trajectories obtained from projected and full CAP-EOM-EA-CCSD, but in most cases, the first-order resonance parameters differ by less than 0.1 eV. We have also demonstrated that the projected and full CAP-EOM-CC are similarly dependent on CAP onset and choice of one-electron basis set, and that the differences in results obtained from the two methods are small relative to the sensitivity of the results to these parameters. Lastly, our results show that the projected CAP-EOM-CCSD scheme can be readily applied to larger molecules and potential energy surfaces, two situations in which the drastic improvement in computational efficiency can be particularly advantageous.

## Chapter 5

# OpenCAP: An Open-Source Framework for Computing Resonance Parameters in Molecules

### 5.1 Overview

The complex absorbing potential (CAP) is a flexible tool for computing resonance parameters of metastable electronic states. In this work, we describe OpenCAP, an open-source package for conducting and analyzing CAP-augmented electronic structure calculations. The OpenCAP software exploits the projected CAP scheme, which allows for interfaces with a wide variety of electronic structure methods. The approach allows one to calculate resonance parameters by a straightforward post-processing of the results obtained from standard bound-state calculations. To demonstrate the capabilities of the software, we present the results of benchmark calculations for molecular shape resonances computed using various electronic structure methods with commonly used quantum chemistry packages. We discuss the technical details of setting up, running, and analyzing the results of CAP-based calculations with the OpenCAP software, as well as advantages and limitations of specific electronic structure methods for describing resonance parameters.

---

This manuscript has been submitted to the Journal of Physical Chemistry A. Copyright American Chemical Society. The OpenCAP software, described below, is publicly available on [GitHub](#), and pre-compiled wheels are distributed on [PyPI](#).

## 5.2 Introduction

Electronic resonances are (N+1)-electron metastable states which lie energetically above an N-electron decay channel, and consequently decay through electron emission. These transient states live long enough to be characterized experimentally, and play key roles in a wide variety of processes, including low energy electron-molecule scattering (e.g. DNA damage by secondary electrons, interstellar chemistry) [17, 21], and electron-molecule interactions occurring in highly energetic environments (e.g. plasmonic photocatalysis, attosecond and X-ray spectroscopies)[19, 22–24]. Resonances can be accessed when a molecule is excited above its ionization threshold, and are commonly formed by electron attachment to closed-shell species[9, 183], or through core-ionization or core-excitation[25].

Because resonances are embedded in the ionization/detachment continuum, the associated eigenstates are not square-integrable[4]. Their non-stationary nature implies that conventional quantum chemistry techniques which have been developed for bound states cannot be applied in a straightforward way[8]. Two major groups of methods to characterize electronic resonances have emerged: scattering theory approaches which account for the continuum explicitly [7, 93, 94], and  $L^2$  methods which treat the scattering part of the metastable state implicitly[4, 8, 104, 125]. Examples of the first class include R-matrix theory [93] and Feshbach projection operator approaches [7, 214], and examples of the latter include stabilization[95–97] and complex scaling[175]. Due to their highly specialized nature, such calculations are far from routine, and there is a very limited number of electronic structure methods that can be used.

The complex absorbing potential (CAP)[104, 111] method is one of the most commonly used non-Hermitian approaches, and belongs to the second group of  $L^2$  methods. In the CAP method, the physical Hamiltonian is augmented with a complex-

valued, artificial potential which absorbs the scattering tail of the resonance wave function. Resonances appear as square-integrable eigenstates of the CAP-augmented Hamiltonian associated with complex Siegert-Gamow eigenvalues:  $E = E_R - i\Gamma/2$ , where  $E_R$  and  $\Gamma$  correspond to resonance energy and width, respectively[6, 15].

Several properties of the CAP approach make it particularly attractive for describing electronic resonances in molecular systems. The first is its ease of implementation. The CAP itself is simply a one-electron operator, and the corresponding one-electron integrals can be computed analytically[133] or using standard numerical integration techniques, depending on the choice of CAP. Second, there is flexibility in how one chooses to incorporate the CAP into the calculation[119, 122, 126, 138, 139]. And finally, CAP, in principle, is compatible with any many-body electronic structure method as long as the method can accurately describe the target states of interest. This has led to numerous implementations with various electronic structure methods, including configuration interaction (CI)[121–124], symmetry-adapted cluster-configuration interaction (SAC-CI)[125], Fock-space coupled cluster[126], density functional theory (DFT) [127], algebraic diagrammatic construction (ADC)[128, 129], and multireference perturbation theory (MRPT) [130–132].

Despite a renewed interest in CAPs in the last decade, very few electronic structure packages possess the capability to perform CAP-augmented electronic structure calculations out of the box. To our knowledge, Q-Chem[171], which supports CAP for the equation-of-motion coupled-cluster with single and double substitutions (EOM-CCSSD) [119, 120] and for ADC[129] calculations, is the only quantum chemistry package which includes documented CAP functionality as part of its distribution. To enhance the accessibility of CAP methodologies, we have developed OpenCAP; a free open-source package which is aimed at extending the CAP methodology to a variety of electronic structure methods and software packages. OpenCAP is a hybrid



C++/Python package built using pybind11[215], and is primarily distributed through its Python interface: PyOpenCAP. OpenCAP leverages the growing trends in quantum chemistry of open-source software and enhanced post-processing capabilities, and utilizes the CAP subspace projection scheme[122] to perform CAP augmented multi-state electronic structure calculations through post-processing of data generated by conventional quantum chemistry calculations. The data can be specified through standard post-processing formats (Molden[216], checkpoint, HDF5 etc.) or in RAM within the Python environment, and then Python-based analysis tools are used to extract resonance position and width. OpenCAP officially supports interfaces with the OpenMolcas[217], PySCF[218], PSI4[219], Q-Chem[171], and Columbus[76] packages, and in principle results from any multi-state electronic calculation can be supported by supplying properly formatted one-particle density matrices.

Below we discuss the main theoretical concepts pertaining to the OpenCAP approach, its implementation, and its performance. First, we outline important aspects of the CAP theory, and the CAP subspace projection formalism which provides the foundation for the OpenCAP approach (Sec. 5.3). Next, we present the relevant details of the computational set-up (Sec. 5.4). The structure and functionality of the OpenCAP package are described in Sec. 5.5.1. Finally, we present the results of benchmark calculations and discuss important technical aspects of performing projected CAP calculations with several electronic structure methods. Specifically, we examine single reference wave function based methods (Sec 5.5.2), multi-reference methods (Sec. 5.5.2), and a pilot application of projected CAP combined with time-dependent density functional theory (TDDFT) (Sec. 5.5.2).

### 5.3 Theory

### 5.3.1 Complex Absorbing Potential

In the CAP method, the physical Hamiltonian  $H$  is augmented with a complex potential ( $-i\eta W$ ) which absorbs the outgoing tail of a resonance wave function and transforms it into a square-integrable form[104]

$$H_{CAP} = H - i\eta W, \quad (5.1)$$

Here,  $\eta$  is the CAP strength parameter, and  $W$  is a real potential which vanishes in the vicinity of the molecular system and grows with distance. A resonance appears as a discrete eigenstate of the non-Hermitian CAP-augmented Hamiltonian, and is associated with a complex Siegert-Gamow energy of a resonance[6]

$$E = E_R - i\Gamma/2 \quad (5.2)$$

where  $E_R$  and  $\Gamma$  are the resonance position and width, respectively. In a complete one-electron basis, the exact resonance position and width is formally obtained in the limit of an infinitesimal CAP strength ( $\eta \rightarrow 0^+$ ). In finite basis set calculations, the limit  $\eta \rightarrow 0^+$  yields a real eigenvalue corresponding to the original Hermitian Hamiltonian, and thus yields no useful information about the resonance state. In practical calculations, the approximate Siegert-Gamow energy is instead found by locating an optimal value of the CAP strength parameter,  $\eta_{opt}$ , where the so called  $\eta$ -trajectory exhibits a pronounced stabilization. The  $\eta_{opt}$  can be found using the logarithmic velocity criterion[104]:

$$\left| \eta \frac{dE}{d\eta} \right| \rightarrow \min \quad (5.3)$$

Further, it is typical in CAP calculations to apply a first-order correction to the raw CAP energies which diminishes the artifacts due the CAP[104, 117]

$$U(\eta) = E(\eta) - \eta \frac{dE}{d\eta}, \quad (5.4)$$

The corrected  $\eta$ -trajectories have been shown to reduce the dependence on the CAP onset[116, 119], and exhibit more stable behavior for applications such as potential energy curves[118, 120]. In this work, the  $\eta_{opt}$  for the first-order corrected trajectories can be found by minimizing the logarithmic velocity,

$$\left| \eta \frac{dU}{d\eta} \right| \rightarrow \min \quad (5.5)$$

of  $U(\eta)$ [117].

### 5.3.2 CAP Subspace Projection Scheme

OpenCAP exploits the CAP subspace projection scheme, in which the CAP-augmented Hamiltonian is represented on a subspace spanned by a small number of real eigenstates of the physical Hamiltonian [122]. This approach was first used in the context of CAP augmented configuration interaction (CAP-CI) by Sommerfeld and Santra [122], and has been adapted for various electronic structure methods, including MRPT [130, 132], ADC [129], SAC-CI [125], and EOM-CCSD [120]. Previous studies[120, 122], have shown that the projected CAP scheme serves as a very accurate approximation to approaches which include the CAP in the full many-body Hamiltonian, and is robust enough to yield smooth potential energy surfaces and handle the case of multiple resonances belonging to the same symmetry.

The first step of any projected CAP calculation is to generate a set of  $M$  target states which comprise the “correlated basis” or “subspace” by performing a conventional (in the absence of CAP) multistate electronic structure calculation. While it

is generally not possible to know *a priori* which or how many zero-order states are needed, practical experience has shown that relatively few (typically less than 10-15) states are needed to converge low lying resonance states. The next step is to compute the CAP matrix in orbital basis, and then project it onto the space spanned by the correlated basis:

$$W_{ij}^{CB} = (i|W|j) \quad (5.6)$$

where  $i$  and  $j$  label the eigenstates of the physical Hamiltonian ( $i, j = 1, \dots, N$ ) and the parentheses emphasize the use of  $c$ -product. Since the CAP is a one-particle operator, these expressions can easily be evaluated using the one-electron reduced density matrices ( $\rho$ ) for each state, and the set of transition density matrices ( $\gamma$ ) between each pair of states that are obtained from the bound-state calculation:

$$W_{ij}^{CB} = \begin{cases} Tr [W\gamma^{ij}], & i \neq j \\ Tr [W\rho^i], & i = j \end{cases} \quad (5.7)$$

Finally, the last step is to diagonalize the CAP-augmented Hamiltonian in the space spanned by the correlated basis:

$$H^{CAP} = H_0^{CB} - i\eta W^{CB} \quad (5.8)$$

over a range of values of  $\eta$  in order to generate the  $\eta$ -trajectory and identify the  $\eta_{opt}$ .  $H_0$  in Eq. 5.8 is an appropriate zero-order Hamiltonian for the chosen correlated basis, e.g. the diagonalized similarity transformed Hamiltonian from CAP-EOM-CCSD or the second order effective Hamiltonian  $H_{eff}^{(2)}$  from CAP-XMS-CASPT2.

To summarize, the basic ingredients of any projected CAP calculation are:

- Choice of  $M$ -dimensional correlated basis set and zero-order Hamiltonian
- Complete set of reduced 1-particle density matrices and transition density matrices between each pair of basis states

- CAP matrix projected onto correlated basis set
- Diagonalization of Eq. 5.8 over range of  $\eta$ -values to find  $\eta_{opt}$

Since the CAP matrix can be computed independently, the projected CAP approach can be accomplished entirely through post-processing of data generated by a single electronic structure calculation.

## 5.4 Computational Details

Experimental ground state geometries of the neutral obtained from the NIST Computational Chemistry Comparison and Benchmark Database were used for  $N_2$ ,  $CO_2$ , formaldehyde ( $CH_2O$ ), and ethylene ( $C_2H_4$ ), the geometries are given in Sec. A1 in Appendix A. Unless otherwise specified, we used the quadratic box-potential, and computed the integrals analytically[133]. The Box CAP parameters for each molecule are provided in Sec. A2 in Appendix A.

In all calculations, we used Dunning’s correlation consistent aug-cc-pVTZ [176] basis set augmented by an additional set of even-tempered 3s3p3d diffuse basis functions placed on a ghost atom in the center of the molecule. The exponents for the additional diffuse functions were obtained as one half of the exponent of the preceding function of the same angular momentum, starting from the most diffuse function in the parent basis set[119]. For the MRPT, ADC, and TDDFT calculations, an additional diffuse function with an exponent of  $1 \times 10^{-8}$  was placed on the ghost atom to mimic ionization[132]. The frozen core approximation was utilized for multi-reference CI (MR-CI) calculations. To avoid intruder states IPEA shift[220] of 0.25 a.u. and an imaginary shift[69] of 0.2 a.u. were used in MRPT simulations. For XDW-CASPT2[72], the squared energy difference was used to compute the weights, and the parameter  $\zeta$  was chosen as 50 for all calculations.

**Table 5.1:** Summary of CAP subspace projection methods used in this work.

Method	$H_0^{CB}$	1RDMs	Package
EOM-CCSD	Diagonal	Unrelaxed EOM-CCSD	PySCF
ADC(n)	Diagonal	Unrelaxed ADC(n)	ADCC
MRPT	$H_{eff}^{(2)}$	Unrelaxed CASSCF	OpenMolcas
MR-CI	Diagonal	Unrelaxed MR- CI	COLUMBUS
TDDFT	Diagonal	TDA amplitudes	PySCF

Multiple quantum chemistry packages have been used for the various electronic methods employed in this work (summarized in Table 5.1). For the EOM-EA-CCSD calculations, the PySCF package [218, 221, 222] version 2.01 was used. For the ADC(N) calculations, we used the PySCF interface to the ADC-connect package [223, 224], version 0.15.14. For the MRPT calculations, we used the OpenMolcas [217] package revision v21.06. For the MR-CI calculations, we used the COLUMBUS package [76], version 7.0.2. For the TDDFT calculations, we used the Libxc [225] library as implemented in PySCF version 2.01. For the CIS and CIS(D) methods, we used a locally modified version of the Q-Chem [171] package, version 5.4. All projected CAP calculations were carried out with OpenCAP version 1.2.6, and all results reported herein, unless otherwise specified, refer to first-order or ‘corrected’ results computed using Eqs. 5.4 and 5.5.

## 5.5 Results and Discussion

### 5.5.1 Using OpenCAP

The OpenCAP package is designed to post-process data from commonly used electronic structure packages in order to conduct projected CAP calculations. There is built-in support for post-processing files in common formats such as Molden, format-

ted checkpoint, and HDF5, from popular packages such as Q-Chem and OpenMolcas. However, the interface is flexible enough to support any package provided that the user can prepare the necessary data in a supported manner. Below, we briefly outline the three major components of the OpenCAP package: the *System*, the *CAP*, and the *CAPHamiltonian*.

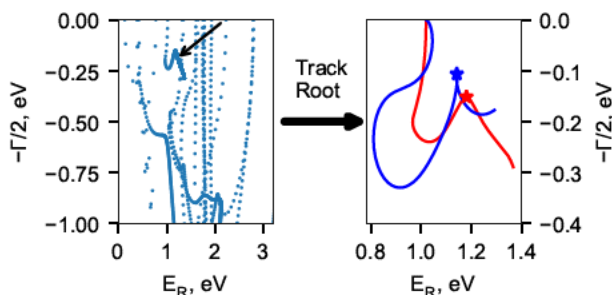
**System:** The *System* contains the molecular geometry and one-electron basis set. It is most conveniently created by specifying a Molden or other post-processing file, but the geometry and basis set can also be specified using an inline format. The overlap matrix computed by the *System* can be used to verify the ordering of the atomic orbital basis functions as is required for the next section.

**CAP:** The *CAP* component computes the CAP integrals, first in the atomic orbital basis, and then projected into the CAP subspace using the one-particle density matrices. There is built in support for the commonly used box[133] and smooth Voronoi CAPs[135], and support for custom CAP functions defined by the user in the PyOpenCAP interface. CAP integrals in the AO basis are computed analytically by default for Box CAPs, and numerically for all other CAPs. For numerical integration, we exploit standard DFT quadrature, computed using the NumGrid package[226]. The radial grid is computed using the scheme from Ref. 227, which is basis set dependent and tuned using a single radial precision parameter (set to  $1 \times 10^{-16}$  by default). The angular grid is generated using the SPHERE.LEBEDEV.RULE library[228], which is done according the scheme from Lebedev and Laikov [192] and requires users to specify an allowed number of grid points (590 by default). We briefly explore the dependence of computed resonance positions and widths on these numerical integration parameters in Sec. 5.5.2. Alternatively, users can specify a numerical integration grid of their own choice, which is done atom by atom.

The next step is to process the one-particle densities in order to perform the CAP

subspace projection. This can be done by specifying a supported post-processing file, and in some cases the  $H_0$  can be parsed as well by providing a supported output file. The most flexible option is to use the PyOpenCAP interface and pass the densities directly as formatted arrays in atomic orbital basis. In this case, the ordering of Cartesian/solid harmonic Gaussian type basis functions must be specified using the name of a supported electronic structure package (e.g. PySCF, PSI4) or standard format (e.g. Molden).

**CAPHamiltonian:** Finally, the *CAPHamiltonian* component contains Python-based tools for generating and analyzing eigenvalue trajectories (Fig. 5-1). For appropriate  $H_0^{CB}$  and  $W^{CB}$  matrices (Eq. 5.8), users specify a range of  $\eta$  values to generate the  $\eta$ -dependent eigenvalue spectrum. Trajectories corresponding to resonances will vary slowly with the CAP strength, while unphysical discretized continuum states not corresponding to resonances will accelerate quickly into the complex plane. Once the resonance eigenvector is identified, the  $\eta$ -trajectory is tracked using energetic or eigenvector overlap criterion, and the first-order trajectory and  $\eta_{opt}$  are computed.



**Figure 5-1:** CAP-ADC(3) eigenvalue spectrum (left) and  $\eta$ -trajectory (right) for the  ${}^2B_1$  resonance in  $CH_2O^-$ . The zero- and first-order trajectories are shown in red and blue, respectively.  $\eta$  is varied from 0.0 to 0.02 in steps of 0.00005, and the 6th root of the CAP augmented Hamiltonian is tracked to generate the eigenvalue trajectories.  $\eta_{opt}$  on the first-order trajectory is indicated by a  $\star$ .



### 5.5.2 Benchmarks

In this section, we present the results of benchmark calculations for various projected CAP methods obtained using OpenCAP and popular electronic structure packages, and provide practical insights for computing shape resonances using different electronic structure methods.

#### Single reference wave-function methods

As an initial proof of concept, we consider the  ${}^2B_1$  resonance in formaldehyde using two electronic structure methods which have frequently been combined with CAP: EOM-CCSD[55, 56] and ADC[57]. Specifically, for EOM-CCSD, we utilize the EOM-CCSD for electron affinities (EOM-EA-CCSD) as implemented in PySCF, and for ADC we utilize the standard variant for excitation energies as implemented in the freely available ADC-connect package [223, 224]. In the latter case, we “trick” the method into computing electron affinities by including a diffuse continuum-like orbital with an exponent of  $1.0 \times 10^{-8}$  (“fake-ip” orbital) in the basis set so the lowest energy state corresponds to the electron occupying this continuum orbital and representing the neutral. The results for CAP/EOM-EA-CCSD and different variants of CAP/ADC are presented in Table 5.2. Since the description of this shape resonance relies primarily upon accurate description of one-particle states, the second-order ADC(2) scheme is able to provide a similar level of accuracy to ADC(3) and EOM-CCSD[129]. The *ad hoc* ADC(2)-X scheme provides an unbalanced description of the excitation spectrum, which leads to an underestimated resonance position[57].

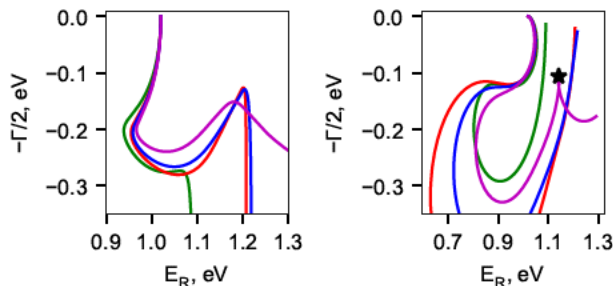
For all methods, 30 states were computed, and the resonance appears as the 6<sup>th</sup> eigenvector of the CAP augmented Hamiltonian. The resonance eigenvector is strongly coupled to five other low-lying states through the CAP matrix. These six states form the minimal correlated basis set required to represent the state, and

**Table 5.2:** Resonance parameters for the  ${}^2B_1$  resonance in  $CH_2O^-$  computed using the EOM-EA-CCSD and ADC family of methods.

Method	$E_R$	$\Gamma$	$\eta_{opt}$
EOM-EA-CCSD	1.159	0.225	0.005
ADC(2)	0.996	0.176	0.005
ADC(2)-X	0.757	0.123	0.0052
ADC(3)	1.142	0.213	0.0051
Electron transmission spectroscopy [11]	0.87		

correspond to the lowest six one-particle ( $1p$ )-states of the proper symmetry. There is no straightforward way to determine the size of this minimal subspace ahead of time, and it varies with the system and the choice of one-electron basis set [120]. The structures of the zero-order and first-order trajectories can serve as a useful diagnostic tool for this purpose. As shown in Fig. 5.2, the zero-order trajectory first becomes stabilized when 13 states (which among other states contain three states that are strongly coupled to the resonance eigenvector) are included; however, the curvature beyond the stationary point rapidly accelerates into the complex plane, and as a result, no clear stationary point appears on the first-order trajectory. This is an indication that more states are needed, and it’s not until 24 states (which corresponds to adding two additional strongly coupled states) where a stationary point finally appears on the first-order trajectory. Previous studies have shown that increasing the size of the subspace beyond this point to include higher lying states which couple to the resonance can slightly improve the accuracy of first-order parameters [120, 124, 129], however, the differences are minor once stabilization on the first-order trajectory is achieved, and so we will not discuss this point any further.

Lastly, we explore the convergence of numerical integration of the CAP integrals with respect to the size of the radial and angular components of the grid. In Figs. 5.3 and 5.4, we consider the results for EOM-EA-CCSD, focusing on the CAP norm ( $|W|$ ) and zero-order half-width ( $\Gamma/2$ ). Both types of CAP show a similar convergence behavior with respect to the numerical integration parameters. For a given radial

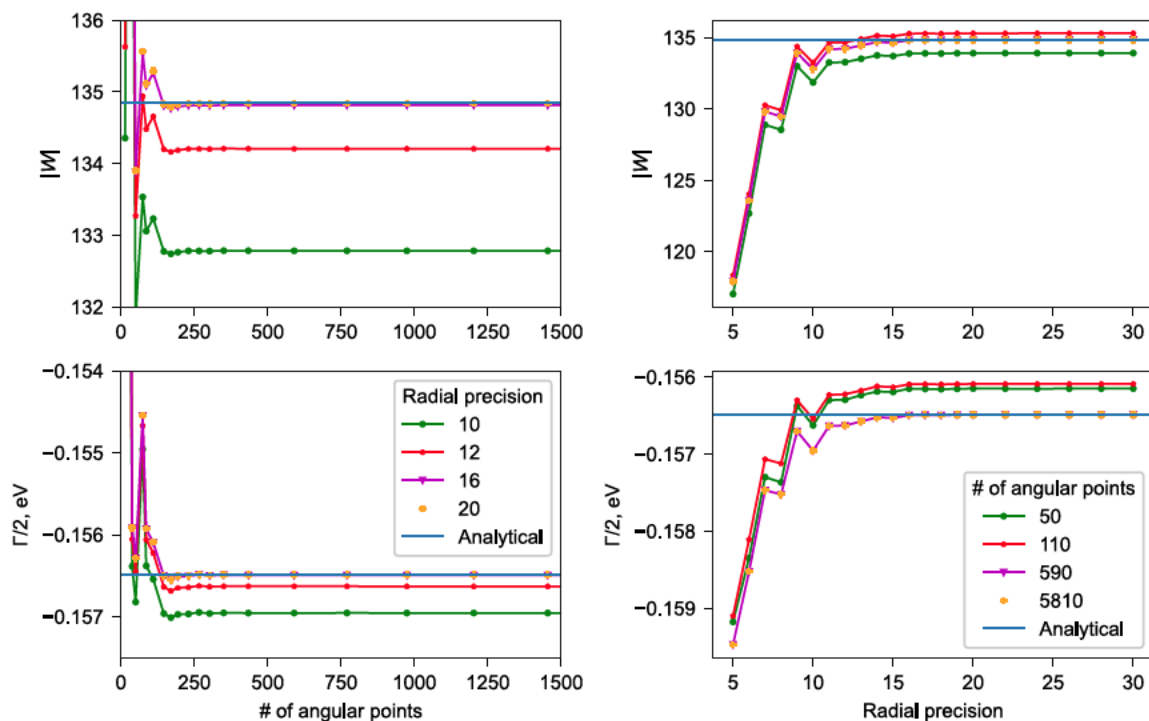


**Figure 5.2:** Comparison of the zero-order (left) and the first-order (right) trajectories for the  ${}^2B_1$  resonance in  $CH_2O^-$ , computed using CAP/ADC(3) with 9 states (green), 13 states (red), 20 states (blue), and 24 states (magenta). The point associated with  $\eta_{opt}$  is indicated by a  $\star$ .  $\eta$  was varied from 0.0 to 0.05.

precision, the norm is converged within  $1 \times 10^{-4}$  a.u. at the default 590 angular points for both types of CAP, and the half-width within  $1 \times 10^{-7}$  eV and  $1 \times 10^{-6}$  eV for box and Voronoi CAPs respectively. For a given number of angular points, using the the default radial precision of  $10^{-16}$  yields a CAP norm which is converged within  $1 \times 10^{-2}$ , and a half-width within  $1 \times 10^{-6}$  eV. Such variations are minuscule compared to those associated with changing the CAP onset, basis set, or electronic structure method, which indicates that there is little benefit to employing anything larger than the default grid allocated by OpenCAP.

### Multi-reference methods

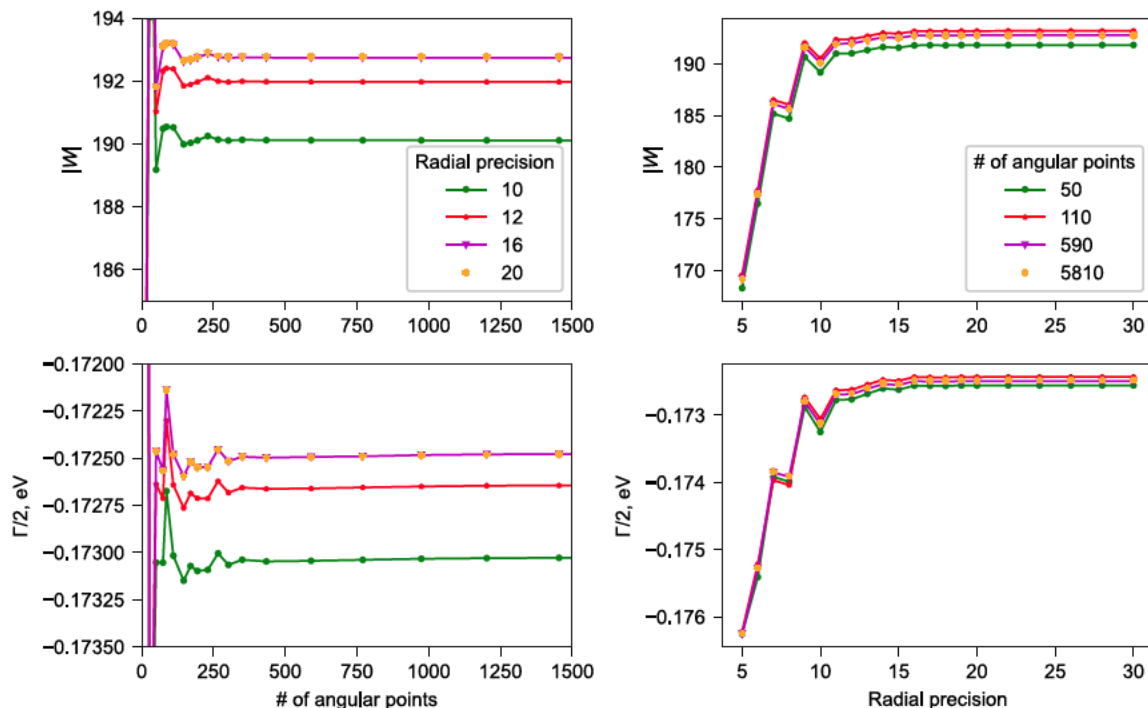
Next, we consider multi-reference electronic structure methods, which are crucial tools in quantum chemistry which provide a balanced description of states dominated by varying levels of excitations, and provide appropriate description when the states of interest have intrinsically multireference character. In the context of resonances, such situations include Feshbach resonances, which are often dominated by multiply excited electronic configurations, and potential energy surfaces near the bond dissociation limit. In this work, we focus on two methods which are implemented in free electronic structure packages: multi-state multi-reference perturbation theory



**Figure 5.3:** Dependence of resonance half-widths ( $\Gamma/2$ ) and CAP norms ( $|W|$ ) on the numerical integration parameters for box CAP for the  ${}^2B_1$  resonance in  $CH_2O^-$ . The label  $N$  for radial precision refers to a precision of  $10^{-N}$ .

(MRPT)[65, 66, 229], and multi-reference configuration interaction (MR-CI)[75, 76]. For both methods, static correlation is recovered with an initial state-averaged complete active space self-consistent field (SA-CASSCF[64]) step, which results in a set of reference states, including the ground state and multiple low-lying excited states, as well as the optimized orbitals. Dynamic correlation is then recovered in a second step, either through a perturbative approach starting from CASSCF zero-order states (MRPT)[65, 229], or variationally using CASSCF optimized orbitals and considering excitations from the reference space, which are typically truncated at single (MR-CI) or single and double (MR-CISD) level[75].

Starting with MRPT, we explore three types of multi-state multi-reference com-



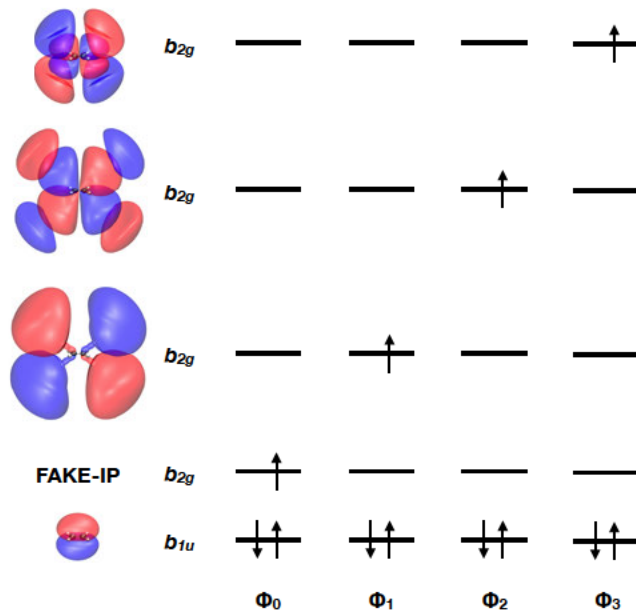
**Figure 5.4:** Dependence of resonance half-widths ( $\Gamma/2$ ) and CAP norms ( $|W|$ ) on the numerical integration parameters for Voronoi CAP ( $r_{cut}=3.0$  bohr) for the  ${}^2B_1$  resonance in  $CH_2O^-$ . The label  $N$  for radial precision refers to a precision of  $10^{-N}$ .

plete active space second order perturbation theory (MS-CASPT2) which are suitable for quasi-degenerate calculations[73]: extended MS-CASPT2 (XMS-CASPT2[71]), extended dynamically weighted MS-CASPT2 (XDW-CASPT2[72]) and rotated MS-CASPT2 (RMS-CASPT2[74]), using the OpenMolcas package [217]. For the CAP calculations, we employ the strategy used in Refs. 130 and 132, defining  $H_0$  in Eq. 5.8 as the second-order effective Hamiltonian matrix ( $H_{eff}^{(2)}$ ) computed using (XMS/RMS/XDW)-CASPT2, which is symmetric and non-diagonal, and computing the matrix elements  $W_{ij}^{CB}$  using the (XMS/RMS/XDW)-rotated CASSCF wave functions which diagonalize the state-averaged Fock matrix[71, 72].

Multi-reference methods require the user to choose an active space, a subset of the

total electrons and orbitals in the system, and a state-averaging scheme. These choices are commonly guided by chemical intuition, although there are several automated approaches that have been proposed recently[230, 231]. Choosing the active space and state-averaging scheme is especially challenging in the case of electronic resonances, where special care should be taken to account for relevant diffuse orbitals in order to generate a model space which can represent the discretized continuum. For the MRPT calculations discussed below, the active spaces for each calculation include an occupied  $\pi$  orbital and a virtual  $\pi^*$  orbital, augmented by additional virtual orbitals belonging to the target symmetry. Additionally, a diffuse orbital with an exponent of  $1.0 \times 10^{-8}$  is added to the active space in order to include the neutral ground state in the state-averaging and subsequent MS-CASPT2 treatment, thereby achieving a balance in the energies of the neutral and anionic states[132]. For example, a CAS(3,5) active space for  $C_2H_4^-$ , consists of one  $\pi$   $b_{1u}$  orbital, one very diffuse  $b_{2g}$  orbital to capture the ground state, and three additional  $b_{2g}$  orbitals (Fig. 5.5). Unless otherwise specified, the model space for the MRPT step consisted of all reference CASSCF states generated by the preceding SA-CASSCF calculation. Note that in the case of the linear molecules  $CO_2^-$  and  $N_2^-$ , we included both sets of degenerate  $\pi$  and  $\pi^*$  orbitals in the active space, but only averaged over states in one irreducible representation.

CAP/MRPT results can be strongly dependent on the choice of active and model spaces[130, 132]. In our calculations, the model spaces which gave the best results consisted of the ground state (closed shell + electron occupying fake IP orbital) and a one-electron state for each virtual orbital included in the active space. While it is not always straightforward to generate such a model space without imposing further restrictions (i.e. generalized active space approaches[232]), a strategy which seemed to work well was to average over the same number of states as the number of virtual



**Figure 5-5:** CAS(3e,5o) active space for  $C_2H_4^-$ . The fake IP orbital is not shown, and the orbitals are visualized with the isosurface value of 0.005. Each of the four model states in the SA(4)-CASSCF(3e,5o) calculation is dominated by a configuration corresponding to electron attachment to one of the  $b_{2g}$  orbitals,  $\Phi_0 - \Phi_3$ . The state dominated by  $\Phi_0$  represents the neutral.

orbitals, with equal weights assigned to each state in the averaging procedure.

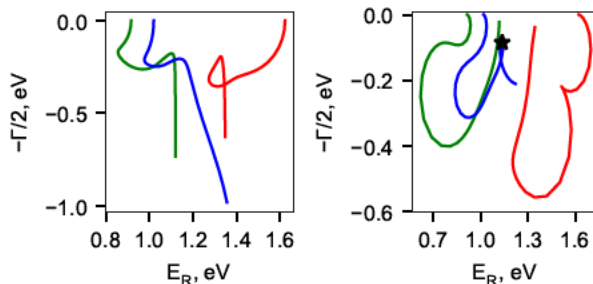
In Table 5.3, we present CAP/MRPT results for  $\pi^*$  shape resonances in  $CO_2^-$ ,  $N_2^-$ ,  $CH_2O^-$ , and  $C_2H_4^-$ . For all four systems, our CAP/MRPT strategy shows good agreement with results from CAP/EOM-EA-CCSD, which were obtained with the same basis sets, CAPs, and the same number of states in the correlated basis. Of the three types of extended MS-CASPT2, the RMS-CASPT2 method tends to give slightly higher resonance energies and widths than XMS/XDW-CASPT2. Similar results for all three methods are obtained with a real shift[70] for MRPT (Table S2), and with an alternative strategy which does not include the diffuse orbital (Table S3). Thus, slightly higher resonance energies and widths for considered resonances are intrinsic for RMS-CASPT2 rather than originating from the specifics of the computa-

**Table 5.3:** Computed CAP/MRPT first-order resonance parameters, along with selected theoretical and experimental results for the  $\pi^*$  shape resonances studied in this work.  $E_R$  and  $\Gamma$  are given in eV.

Method		$E_R$	$\Gamma$
	$CO_2^-$ SA(7)-CAS(5e,15o)		
XMS		4.000	0.125
RMS		4.065	0.138
XDW		3.992	0.129
CAP/EOM-EA-CCSD[120]		4.017	0.178
Electron transmission spectroscopy [233]		3.58	-
	$N_2^-$ SA(5)-CAS(5e,11o)		
XMS		2.394	0.195
RMS		2.474	0.224
XDW		2.409	0.206
CAP-EOM-EA-CCSD[120]		2.581	0.226
Feshbach projection formalism based on experimental data [193]		2.32	0.41
	$C_2H_4^-$ SA(4)-CAS(3e,5o)		
XMS		1.733	0.233
RMS		1.790	0.269
XDW		1.730	0.241
CAP-EOM-EA-CCSD[120]		2.012	0.432
Electron scattering [198]		1.8	0.7
	$CH_2O^-$ SA(7)-CAS(3e,8o)		
XMS		1.137	0.170
RMS		1.252	0.216
XDW		1.147	0.177
CAP-EOM-EA-CCSD[120]		1.202	0.322
Electron transmission spectroscopy [11]		0.87	-

tional setup. Importantly, the model spaces used in Table 5.3 are the minimal model spaces required to represent each respective resonance. As illustrated for  $CH_2O^-$  in Fig. 5.6, including fewer model states results in no clear stationary point appearing on the first-order trajectory, similar to what we showed in Fig. 5.2 for EOM-CCSD. Further, the minimum number of model states, excluding the ground state, is equal to the minimum number of EOM-EA-CCSD states required for the same basis set (see Ref. 120), which points to the minimal CAP subspace dimension likely being a property of the one-electron basis set rather than the electronic structure method, assuming a similar character of the low-lying  $1p$ -states in both methods.





**Figure 5-6:** Comparison of the zero-order (left) and the first-order (right) trajectories for the  ${}^2B_1$  resonance in  $CH_2O^-$ , computed using CAP/XMS-CASPT2 with the following active spaces: SA(5)-CAS(3e,6o)(green), SA(6)-CAS(3e,7o)(red), and SA(7)-CAS(3e,8o)(blue). The point associated with  $\eta_{opt}$  is indicated by a  $\star$ .  $\eta$  was varied from 0.0 to 0.05.

In Table 5.4, we explore different state averaging schemes for the  ${}^2B_{2g}$  resonance in  $C_2H_4^-$ . Averaging over more states than virtual orbitals generates additional reference states that do not have one-electron attached character, and their inclusion in the model space led to notable shifts in the resonance parameters for XMS-CASPT2 and XDW-CASPT2. Of the three MRPT schemes, RMS-CASPT2 was by far the most resistant to this problem, which is consistent with previous studies that have suggested RMS-CASPT2 to be the best choice when averaging over a large number of states is unavoidable[74, 234]. Another approach in cases of unbalanced active space/model space calculations would be to manually select the states of the  $1p$ -character and keep only those in the model space. For a SA(10)-CAS(3e,5o) calculation this choice corresponds to four states in the model space (the last row in Table 5.4), and it leads to improved results for XMS-CASPT2 and XDW-CASPT2 in comparison to those obtained with all 10 states in the model space (first row in Table 5.4). Lastly, averaging over fewer states than the number of virtual orbitals leads to a notable increase in the resonance parameters. A similar observation for these ‘unbalanced active spaces’ was noted in Ref. 132 for CAP/XMCQDPT2 calculations on  $N_2^-$ .

Next, we consider the CAP/MR-CI method [122, 124] implemented through in-

**Table 5.4:** Computed CAP/MRPT first-order resonance parameters for the  ${}^2B_{2g}$  resonance state of  $C_2H_4^-$  using selected active spaces and state averaging schemes.  $E_R(\Gamma)$  are given in eV.

Nr. of states	Nr. of $b_{2g}$ orbs.	XMS	RMS	XDW
10	4	1.415(0.171)	1.771(0.252)	1.584(0.216)
10	5	1.528(0.208)	1.843(0.315)	1.670(0.277)
4	5	1.963(0.316)	1.995(0.344)	1.958(0.321)
4	6	1.960(0.315)	1.992(0.342)	1.955(0.320)
4	4	1.733(0.233)	1.790(0.269)	1.730(0.241)
5	5	1.741(0.227)	1.809(0.266)	1.744(0.236)
6	6	1.727(0.221)	1.807(0.266)	1.746(0.237)
10(4)*	4	1.664(0.168)	1.768(0.210)	1.702(0.187)

\* SA-CASSCF was performed with 10 states, but only the 4 one-electron attached states were included in the model space at the MRPT step.

terface with the COLUMBUS package[76]. There are a few key differences in the CAP/MR-CI scheme from what we used previously for MRPT. First, the matrix elements  $W_{ij}^{CB}$  and  $H_0^{CB}$  are expressed in the basis of MR-CI states (rather than CASSCF), and accordingly,  $H_0^{CB}$  is diagonal, and  $W_{ij}^{CB}$  is computed using MR-CI densities. Second, the COLUMBUS package has the capability of including states of different spin, symmetry, and with different number of electrons in the same SA-CASSCF calculation, so the state-averaging between the neutral and anionic states can be done explicitly without using an extra diffuse orbital. We use the notation SA(1+N) to indicate that the state averaging was performed with the neutral ground state in the totally symmetric irreducible representation, and  $N$  anionic states in the target irreducible representation. Lastly, at the MR-CI step, we compute 5-10 anionic states (enough for a stationary point to appear on the first-order trajectory) and the neutral ground state in separate calculations using the same set of optimized orbitals.

In Table 5.5, we present CAP/MR-CI results for active spaces equivalent to those discussed in Table 5.3 for MRPT. In general, both CAP/MR-CIS and CAP/MR-CISD overestimate the resonance parameters compared to MRPT and EOM-CCSD,

**Table 5.5:** Computed CAP/MR-CI first-order resonance parameters.  $E_R(\Gamma)$  are given in eV.

Molecule	Method	MR-CIS	MR-CISD	MR-CISD(Q)
$C_2H_4^-$	SA(1+3)-CAS(3e,4o)	2.375(0.575)	2.177(0.455)	2.034(0.373)
$N_2^-$	SA(1+4)-CAS(5e,10o)	3.100(0.345)	2.941(0.329)	2.721(0.320)
$CH_2O^-$	SA(1+6)-CAS(3e,7o)	1.877(0.547)	1.715(0.469)	1.532(0.294)
$CO_2^-$	SA(1+6)-CAS(5e,14o)	4.578(0.163)	4.388(0.153)	4.212(0.168)

**Table 5.6:** Computed CAP/MR-CI first-order  $N_2^- \ ^2\Pi_g$  resonance parameters for selected active spaces.  $E_R(\Gamma)$  are given in eV.

Method	MR-CIS	MR-CISD	MR-CISD(Q)
$C_2H_4^-$			
SA(1+9)-CAS(3e,5o)	2.231(0.466)	2.248(0.547)	2.072(0.424)
SA(1+3)-CAS(3e,5o)	2.241(0.469)	2.202(0.512)	2.062(0.413)
SA(1+4)-CAS(3e,5o)	2.323(0.543)	2.219(0.520)	2.071(0.409)
$N_2^-$			
SA(1+3)-CAS(5e,8o)	3.262(0.377)	3.227(0.188)	3.009(0.425)*
SA(1+9)-CAS(5e,10o)	2.977(0.325)	3.017(0.357)	2.759(0.341)
SA(1+4)-CAS(5e,12o)	2.903(0.269)	2.8873(0.304)	2.686(0.300)
SA(1+5)-CAS(5e,12o)	3.034(0.317)	2.934(0.345)	2.719(0.339)

\* Uncorrected result; no stationary point could be identified on first-order trajectory.

which can be due to multiple factors, including an imbalance in correlation between the neutral anionic states and the well known size extensivity problem of truncated CI methods[76, 124]. For MR-CISD, the latter can be approximately accounted for using *a posteriori* corrections to the energy, and here we employ the Davidson correction[77], which is commonly referred to as MR-CISD(Q). The MR-CISD(Q) resonance positions are generally shifted lower (between 0.1 and 0.2 eV), and thus agree better with the MRPT and EOM-CCSD results summarized in Table 5.3.

The lack of an explicit dependence on a model space for MR-CI leads to some important differences from what we previously discussed for MRPT, which are demonstrated using results from selected active spaces and state-averaging schemes in Table 5.6. While there is some dependence on the state averaging scheme, which also affects the optimized orbitals, CAP/MR-CI does not suffer from the marked shifts which arise

from ‘unbalanced’ active spaces in CAP/MRPT. Another important difference is that one can use smaller active spaces with MR-CI, for example a CAS(5e,8o) active space for  $N_2^-$ . It has been shown that a minimum of four one-electron-attached discretized continuum states (each corresponding to occupation of a different virtual orbital in the target irreducible representation) must be included in the correlated basis set to achieve stabilization on the first order CAP-EOM-EA-CCSD trajectory[120] for the basis set used. The (5e,8o) active space includes three pairs of degenerate virtual orbitals of symmetries corresponding to the two degenerate resonances), thus it is not possible to generate reference states for four discretized continuum states with four different virtual orbitals being singly-occupied with this active space. This leads to the failure of MRPT to yield stabilized trajectories in this case (see Fig. A11 in Appendix A). With MR-CI methods, the question instead becomes whether the discretized continuum states corresponding to occupation of orbitals outside of the active space are described in a balanced enough way. At the MR-CIS level, we are able to achieve a stabilized first-order trajectory with a CAS(5e,8o) active space which resembles those from larger active spaces (Fig. A9 in Appendix A), but at the MR-CISD level, the stationary point on the first-order resonance trajectory is far less pronounced (Fig. A10). However, using such small active spaces with fewer virtual orbitals than the CAP subspace dimension is not advisable, as it leads to even further overestimation of the resonance parameters (Table A4 in Appendix A).

Lastly, we note that a recent CAP/MR-CI study[124] reported MR-CISD overestimating the resonance position by 0.4 eV compared to MR-CIS for  $N_2^-$ , which we did not observe in our calculations. There are two likely causes of this discrepancy. The first is that the authors of Ref. 124 did not include the ground and anionic states in the same SA-CASSCF calculation, and instead used orbitals optimized for the anionic states only. The second is that they used a CAS(5e,8o) active space, which as

noted previously is problematic at the MR-CISD level. Overall, we recommend that users exercise caution when using CAP/MR-CI methods due to their tendency to overestimate resonance parameters, and suggest to use active spaces that include at least as many virtual orbitals as the minimum number of states required to represent the resonance.

### **Time dependent density functional theory**

Finally, as a pilot application of OpenCAP, we explore CAP combined with time dependent density functional theory (TDDFT)[82] for the first time. DFT is one of the most commonly used methods in quantum chemistry, but its application to anions has historically been controversial[235–239]. For bound anions, concerns were raised early on that DFT calculations often yielded positive HOMO energies for the anion (thus describing an unbound electron[235–237]), a discrepancy that has been attributed to strong self-interaction errors[236, 238, 240]. Despite these formal difficulties, numerous studies have demonstrated that few problems arise in practice with reasonable finite basis sets[239, 241], and thus it has been argued that conventional DFT can give reliable and useful results for electron affinities when calculated self-consistently as a difference between the total energies of the anionic and neutral species [236, 237].

For temporary anions, a CAP complex DFT (CAP-CODFT) implementation was described by Zhou and Ernzerhof in Ref. 127, but to our knowledge, there has been little development since then in combining CAP with density functional theory for this purpose. DFT has seen much wider use in conjunction with stabilization methods [95, 97, 242–249], including studies on resonances in large systems with biological relevance such as quinones[250] and polyaromatic hydrocarbons[251].

In the present CAP-TDDFT calculations, we have used the PySCF package [218, 221, 222, 225], employing the diffuse orbital trick to compute the (N+1)-electron states, supplying the Kohn-Sham orbitals of the closed-shell neutral molecule as an

initial guess for the reference DFT calculation to aid convergence. The Tamm-Dancoff approximation (TDA) is employed [85–87], and the density matrices are computed by taking the TDA amplitudes as CIS coefficients. We consider several functionals: long-range corrected functionals LC functionals which employ exact exchange at long electron-electron distances without re-parametrization of the parent functional (LC\_BLYP[92, 252, 253] and LC\_ $\omega$ PBE[254]), commonly-used range-separated functionals ( $\omega$ b97XD[255] and CAM-B3LYP[91]), and functionals which use exact exchange at all electron-electron distances (HFE\_LYP[253] and HFE\_PBE[90]). The use of exact exchange functionals follows the work of Falcetta et al.[249], who found that such functionals provide good results for unbound anions in the context of stabilization calculations. Importantly, all range-separated functionals used in this work, except CAM-B3LYP, employ 100% asymptotic HF exchange for the long-range.

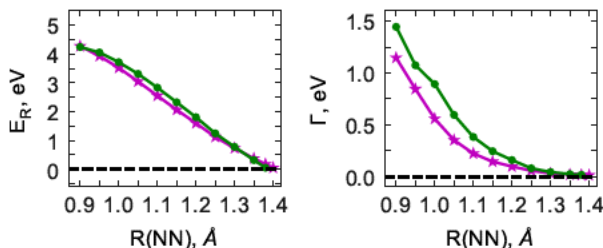
In Table 5.7, we present CAP/TDDFT results for the  ${}^2\Pi_g$  resonance in  $N_2^-$ , along with the stabilization results reported in Ref. 249, the CAP/CO-DFT results from Ref.127, and results from the CAP/CIS(D)[60–63] family of methods, which involve perturbative corrections to the CIS energies to recover additional correlation. Our findings for CAP-TDDFT are very similar to the data obtained with the stabilization technique reported by Falcetta et al. [249]. The range separated hybrids severely underestimate both position and width, the long range corrected functionals fare slightly better, and the exact exchange functionals provide by far the best performance. With the lone exception of the LC\_BLYP functional, there is good agreement for the resonance positions between CAP-TDDFT and the stabilization calculations, while CAP-TDDFT generally predicts lower widths. There is also good agreement between exact exchange CAP-TDDFT and the CAP-CIS(D) family of methods, which all slightly overestimate the resonance position and width compared to more accurate electronic structure methods such as CAP/EOM-CCSD.

**Table 5.7:** Computed first order resonance parameters for the  ${}^2\Pi_g$  resonance in  $N_2^-$  computed using various electronic structure methods.  $E_R$  ( $\Gamma$ ) are given in eV.

Method	CAP	Stabilization[249]
LC_BLYP	1.464(0.099)	1.920(0.358)
LC_WPBE	1.933(0.121)	1.960(0.178)
$\omega$ b97XD	1.321(0.056)	1.285(0.094)*
CAM-B3LYP	0.529(0.032)	0.84(0.118)
HFE_LYP	2.857(0.389)	2.86(0.76)
HFE_PBE	2.957(0.343)	3.078(0.54)
CIS	3.58(0.582)	3.77(1.14)
CIS(D)	3.01(0.375) <sup>†</sup>	2.760(0.04)
SOS-CIS( $D_0$ )	2.80(0.280)	-
SOS-CIS(D)	3.13(0.408)	-
CO-DFT[127]	3.39(0.506)	-

\* Computed using Koopman’s theorem (KT) instead of TDDFT. See Ref. 249.

<sup>†</sup> Computed using the resolution of identity (RI) approximation.



**Figure 5.7:** First-order resonance positions (left) and widths (right) along the complex PEC for the  ${}^2\Pi_g$  shape resonance in  $N_2^-$ , computed by CAP-EOM-CCSD (magenta) and CAP-TDDFT(HFE\_LYP) (green).

Lastly, we consider the application of CAP-TDDFT with the HFE\_LYP functional to the potential energy curve of  $N_2^-$ . In Fig. 5.7, we compare first-order results from CAP/EOM-CCSD[120] and CAP-TDDFT over a range of bond lengths starting from 0.9 Å up to 1.4 Å, using a fixed CAP onset. From 0.9 Å to 1.3 Å, the two methods predict similar curves, with CAP-TDDFT slightly overestimating the resonance positions and widths. At 1.39 Å, we were not able to achieve a reliable description of the resonance parameters owing to the poor convergence of the reference DFT cal-

ulation, which can possibly originate from the near-degeneracy between the neutral and the anionic states treated within a single calculation using 'fake-ip' approach. The resonance parameters at 1.38 Å are  $E_R = 0.05$  eV and  $\Gamma = 0.01$ eV, this can reasonably be interpreted as the crossing point with the neutral ground state. This result agrees very well with CAP/EOM-CCSD, which predicts the resonance to become bound at 1.41 Å[118, 120]. The difficulty in representing the bound anionic state with CAP/TDDFT beyond the crossing region is a drawback of the current approach, but the results are internally consistent (i.e. the resonance width approaches zero as the energy of the electron-attached state approaches the energy of the neutral), and show overall good agreement with the results from CAP/EOM-CCSD.

### 5.5.3 Conclusions

In summary, we have reported the development and application of the open-source OpenCAP package, which extends the capability of electronic structure methods to electronic resonances using the projected CAP methodology. We have described the basic structure and usage of the code, and how it is integrated with several popular electronic structure packages and methods implemented within. Additionally, we have benchmarked various CAP methodologies, including single reference electronic structure methods, multi-reference methods, and a pilot application of CAP-TDDFT, providing practical insight on how each methodology can be applied to shape resonances in molecular systems. We hope that the free of charge OpenCAP package can provide a platform for further explorations of CAP-based methodologies, and make performing such calculations accessible to anyone who wishes to do so with the method of their choosing.



## Chapter 6

# eMAP: an online platform for identifying and visualizing electron and hole transfer pathways in proteins

### 6.1 Overview

The work reported in this Chapter is reproduced with permission from Ref. [151]. Copyright 2019 American Chemical Society. The Supporting Information for this Chapter is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpccb.9b04816>.

Since this publication, the back end to the eMap web application has been released as an open-source Python package known as PyeMap, distributed on PyPI. Additionally, several features not mentioned below have been added, include automatic detection of iron-sulfur clusters and metal ions, and inclusion of other standard amino acid residues. The novel algorithms which were developed to generate the protein graphs and identify aromatic moieties of cofactors are described in Appendix B.

eMap is a web-based platform for identifying and visualizing electron or hole transfer pathways in proteins based on their crystal structures. The underlying model can be viewed as a coarse-grained version of the *Pathways* model, where each tunneling step between hopping sites represented by electron transfer active (ETA) moieties is described with one effective decay parameter that describes protein-mediated tunneling. ETA moieties include aromatic amino acid residue side chains and aromatic fragments of cofactors that are automatically detected, and, in addition, electron/hole residing sites that can be specified by the users. The software searches for the shortest paths connecting the user-specified electron/hole source to either all surface-exposed ETA residues or to the user-specified target. The identified pathways are ranked based on their effective length. The pathways are visualized in 2D as a graph, in which each node represents an ETA site, and in 3D using available protein visualization tools. Here, we present the capability and user interface of eMap 1.0, which is available at <https://emap.bu.edu>.

## 6.2 Introduction

Electron transfer (ET) in proteins lies at the core of many biological processes including respiration, photosynthesis, DNA photodamage repair, and possibly magnetosensing by migratory birds. [38–41] Many proteins, including photosystem II, ribonucleotide reductase, galactose reductase, cytochrome *c* oxidase, photolyase, and cryptochrome, utilize redox-active aromatic Tyrosine (Tyr or Y) and Tryptophan (Trp or W) amino acid residues to shuttle electrons or holes. [44] The recent analysis by Gray and Winkler has revealed that one third of structurally characterized proteins exhibit extended chains of Tyr and Trp residues, which they hypothesize to serve as universal oxidative stress protection mechanism in proteins. [42, 43] Understanding the mechanistic features of proteins and their stability, therefore, often

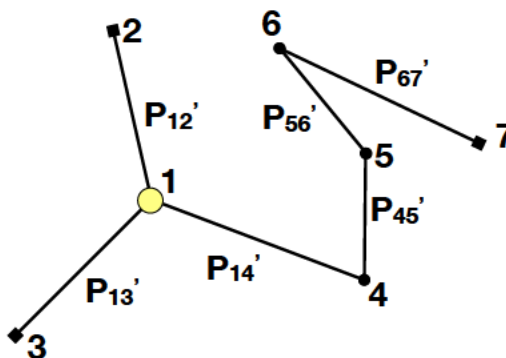
depends on availability of detailed information on the electron/hole transfer pathways through aromatic sites. Here, we present eMap, a robust computational tool aimed at qualitative mapping of electron hopping pathways in proteins based on their tertiary structures. The primary purpose of this software is to efficiently identify possible electron hopping channels to be used in further quantitative studies. The performance of the model is illustrated on two representative proteins.

Existing software for predicting paths of electron transfer in proteins ranges from empirical models (e.g. Pathways plugin [149] for VMD [256]) to more sophisticated electronic structure based schemes (e.g. Electron Tunneling in Proteins Program or ETP [257]). The *Pathways* model [146, 258] by Beratan and Onuchic describes electron transfer as a collection of pathways, each of which is defined as a sequence of through-space, through-covalent bond, and through-hydrogen bond hopping events with multiplicative penalty functions. The ETP software developed by Stuchebrukhov and co-workers [257] evaluates and visualizes electron tunneling current based on multi-level electronic structure calculations including semi-empirical simulation on an entire protein. However, this sophisticated analysis is computationally demanding. eMap provides an inexpensive chemistry-inspired alternative, and can be viewed as a coarse-grained version of the *Pathways* model, limiting the analysis to protein-mediated tunneling between hopping sites represented by automatically identified Electron Transfer Active (ETA) moieties.

The structure of this Chapter is as follows. The empirical electron hopping model used in eMap is described in Sec. 6.3. The structure of the software is outlined in Sec. 6.3. The functionality and user interfaces are described in Sec. 6.3. Finally, the results of the eMap analysis for two representative proteins are discussed in Sec. 6.4.

## 6.3 Model and Implementation

### Pathways model



**Figure 6-1:** A graph representing the connectivity network between the aromatic sites in a protein.  $P'_{ij}$  is a penalty function associated with electron/hole hopping between the sites (see text for more details). Surface-exposed residues are indicated as squares, while buried residues are indicated as circles.

The eMap analysis is based on a coarse-grained version of the *Pathways* model [144–146, 258] with only through-space tunneling between aromatic and user-specified sites being considered. In the *Pathways* model, an electron/hole transfer pathway between the specified donor and acceptor is described as a series of through-space, through-covalent bond, and through-hydrogen bond tunneling events. Considering multiplicative penalty functions ( $\epsilon_{i/j/k}$ ) for each tunneling event, the resulting tunneling matrix element ( $T_{DA}$ ), an effective donor-acceptor coupling, has the following form: [146]

$$T_{DA} = const \times \prod_{i \in space} \epsilon_i \times \prod_{j \in bond} \epsilon_j \times \prod_{k \in H-bond} \epsilon_k \quad (6.1)$$

The problem of finding the most efficient pathway, is, therefore, reduced to maximizing  $T_{DA}$  matrix elements. [144–146, 258] The penalty functions for through-covalent bond tunneling ( $\epsilon_j = 0.6$ ) are constant regardless of the nature of the atoms

and bond length. The penalty functions for through-space ( $\epsilon_i = 0.6 \times \exp(-1.7(R_i - 1.4))$ ) and through-hydrogen bond ( $\epsilon_k = 0.6 \times \exp(-1.7(R_k - 2.8))$ ) tunneling are empirical and distance-dependent, where  $R_{i/k}$  is the interatomic distance in bohr. Smaller values of the prefactor in the exponent ( $1.1 \text{ \AA}^{-1}$  and  $1.4 \text{ \AA}^{-1}$  for  $\beta$ -sheet and for  $\alpha$ -helical structures, respectively) were suggested in later works by Beratan and coworkers[145, 259].

The protein then can be viewed as an undirected graph, with each atom representing a node and the edges connecting the nodes being associated with the penalty functions. To make the problem tractable with graph theory methods, it is convenient to operate with non-negative edge lengths and additive parameters rather than multiplicative penalty functions. This can be achieved using modified penalty functions,  $P' = -\log \epsilon$ . Maximizing the product of  $\epsilon$  penalty functions, which determines  $T_{DA}$  coefficient (Eq. 6.1) is equivalent to minimizing the sum of the corresponding modified  $P'$  functions. Therefore, finding the path with the largest  $T_{DA}$  value is equivalent to finding a shortest path on a graph with edge lengths defined by modified penalty functions,  $P'$ . [144–146, 258] This can be efficiently done using graph theory algorithms, for example, Dijkstra’s. The analysis relies on the available tertiary structure and a specified electron donor and acceptor.

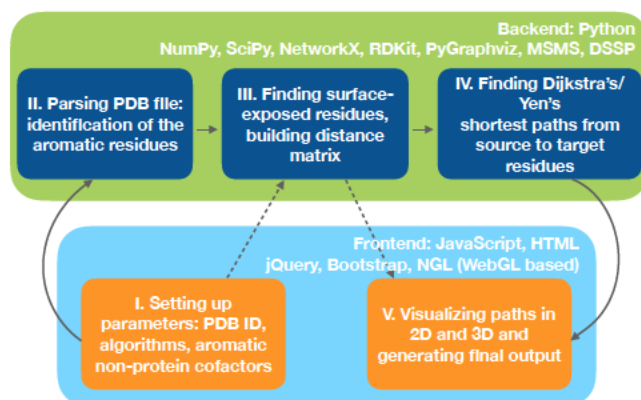
The model employed by eMap can be viewed as a coarse-grained extension of the *Pathways* model where only protein-mediated through-space hopping events between ETA sites are considered. The unique features that distinguish eMap from other available software are discussed below.

### **eMap: coarse-grained hopping model**

eMap relies on a chemistry-inspired tunneling model extending *Pathways* approach to electron hopping through aromatic and/or user-specified ETA sites. While having an atomistic description of electron flow in proteins is often desirable, the key chemically

relevant information, e.g. whether a specific mutation is going to enhance or inhibit electron transfer, can be obtained without an atomic-resolution picture. eMap is aimed at identifying aromatic residues playing key roles in hopping events, such as serving as intermediate electron or hole traps, or serving as terminal acceptors of a hole or electron at the protein surface. The coarse-grained description of long-range electron or hole hopping distinguishes eMap from the *Pathways* model implementation, and allows the user to obtain an intuitive picture for long-range hopping channels.

eMap translates a problem of finding efficient electron/hole hopping pathways into finding the shortest paths between the electron transfer active (ETA) moieties based solely on the geometry information obtained from a PDB/CIF file. The nodes on the eMap graph represent ETA sites, which are side chains of aromatic residues and/or user-defined redox-active centers. The edge lengths are defined as modified penalty functions,  $P' = -\log \epsilon$ , with  $\epsilon = \alpha \exp[-\beta(R - R_{\text{offset}})]$ , where  $\alpha$ ,  $\beta$ , and  $R_{\text{offset}}$  are tunneling parameters, similar to the through-space tunneling penalty function in the *Pathways* model. The default values of  $\alpha$ ,  $\beta$ , and  $R_{\text{offset}}$  are 1.0,  $2.3 \text{ \AA}^{-1}$ , and  $0.0 \text{ \AA}$ , respectively.  $R$  corresponds either to the distance between centers of mass of the two ETA sites or to the distance between two closest atoms belonging to the two ETA sites (see Sec. 6.3 for more details). For the default parameters, the edge lengths are equal to the distances between the hopping sites (multiplied by a prefactor of  $2.3 \times \log(e) \approx 1$ ). The software then finds the shortest paths connecting a given electron/hole source (node 1 in Fig. 6.1) to either all surface-exposed residues of the protein (nodes 2, 3, and 7 in Fig. 6.1) or to a user-specified target residue, which can be buried or surface-exposed. More details on different modes of using eMap are given in Sec. 6.3.



**Figure 6.2:** Schematic representing the dependencies between different modules in eMap.

## Implementation details

eMap is implemented as a web application and takes information on the protein structure (PDB) and user-defined algorithmic parameters as the input. The eMap architecture is schematically shown in Fig. 6.2. The front-end part is responsible for direct communication with the user. The initial step of the analysis is to set up the input for eMap, which includes specifying the protein structure using PDB ID or PDB/CIF files and defining the algorithmic parameters (see Sec. 6.3). This step is fully performed on the client side (Module I in Fig. 6.2). Once all input parameters are specified, the back-end part carries out the preliminary analysis based on the specified input.

The first step on the back end is to parse the protein structure file, which mainly relies on the open-source Biopython package. [260, 261] eMap locates all ETA sites, including aromatic amino acid residues, aromatic moieties of cofactors, and user-specified ETA sites (Module II in Fig. 6.2). The next step is to identify the surface-exposed ETA sites (Module III in Fig. 6.2). ETA residues are classified as buried or surface-exposed using residue depth [262–264] or relative solvent accessibility [265–

[267] criteria (see Sec. 6.3). Biopython, [260, 261] together with the MSMS [262] or the DSSP [265, 268] software, is utilized to perform this classification. A pairwise distance matrix is then constructed for the ETA sites, and an image of a graph with each node representing an ETA site and the edge lengths being defined by the distances between sites is returned to the front end (Module III in Fig. 6.2). The final step on the back end is to search for the shortest paths connecting a specified electron/hole source to each surface-exposed residue or to a single user-specified target (Module IV in Fig. 6.2). The analysis is done using the NetworkX python package, [160] and the graphs are visualized using PyGraphviz. [269] The graph images, along with the results of the analysis (the identified paths ranked by their effective lengths) are then passed to the client side, where the pathways are also visualized in 3D using the NGL Viewer, [270, 271](Module V in Fig. 6.2). Some of the relevant algorithmic details and features are discussed below.

The full list of external packages, references, and licenses is given in the [Supporting Information](#).



## Features and the user interface

**(a)** Parameters

General Additional Residues Advanced

**Protein Chains**

A  
 Choose All

**Distance Options**  
 Center of Mass  
 Closest Atom

**Surface Definition**  
 Residue Depth  
 Solvent Accessibility

**Aromatic Amino Acids**  
 Trp (W)  Phe (F)  
 Tyr (Y)  His (H)

Process File ⓘ

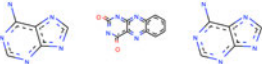
---

**(b)** Parameters

General Additional Residues Advanced

Non-Protein ET Active Moieties

FAD510-1  FAD510-2  ANP511



Choose All

**Custom Atom Range:**  
 See Help page for syntax  
 ex. 160-164

Process File ⓘ

---

**(c)** Parameters

General Additional Residues Advanced

Graph Parameters

**Edges per Vertex, %**

**Distance Cutoff, Å**

**SD Cutoff**

Penalty Function Parameters

Process File ⓘ

**Figure 6-3:** Options specifying pairwise distance map construction: specification of the chains included in the analysis, algorithm used to identify surface-exposed residues, intersite distance evaluation scheme, and standard ETA sites (a); selection of the non-protein ETA sites included in the analysis (b); tuning thresholds, cutoffs, and penalty function parameters for drawing graph edges (c).

Below we discuss the key input parameters and algorithmic details used to construct the graph and predict the most efficient electron transfer channels.

***Specifying ETA sites.*** Once the protein structure uploaded by the user (using either PDB ID or PDB/CIF file) has been parsed, the user can specify the sites to be considered as ETAs, and therefore, to be included in the analysis. This is done using options illustrated in Fig. 6-3a and 6-3b. By default, all of the Tyr and Trp residues from every chain of the protein and all of the automatically identified aromatic moieties of cofactors will be included into the analysis. The “Additional Residues” tab (Fig. 6-3b) shows identified non-amino acid aromatic sites. The user can also manually specify ETA sites atom-by-atom (from PDB atom serial number) using the Custom Atom Range option (Fig. 6-3b).

***Identifying surface-exposed residues.*** The user can choose one of two algorithms to classify residues as surface-exposed or buried (Fig. 6-3a). The default option is the residue depth criterion. Residue depth is defined as the average distance of the residue’s atoms to the computed solvent-excluded surface of the protein. [262–264] Alternatively, the user can choose relative solvent accessibility, [265–268] which is defined as the ratio of the calculated solvent accessible surface area to the tabulated maximum solvent accessible surface area (MaxASA) for this residue type. [267] Relative solvent accessibility cannot be evaluated for non-protein ETA sites due to the lack of pre-computed MaxASA values for non-protein residues.

***Graph and shortest path search parameters.*** The general parameters panel sets the distance measure used in the penalty function: either the distance between the centers of mass of the two ETA sites or the shortest distance between two atoms of the ETA sites (Fig. 6-3a). In addition, the “Advanced” tab enables tuning threshold parameters (Fig. 6-3c) for how the graph is constructed. Edges with distances greater than Distance Cutoff (Fig. 6-3c) are immediately discarded. The density of

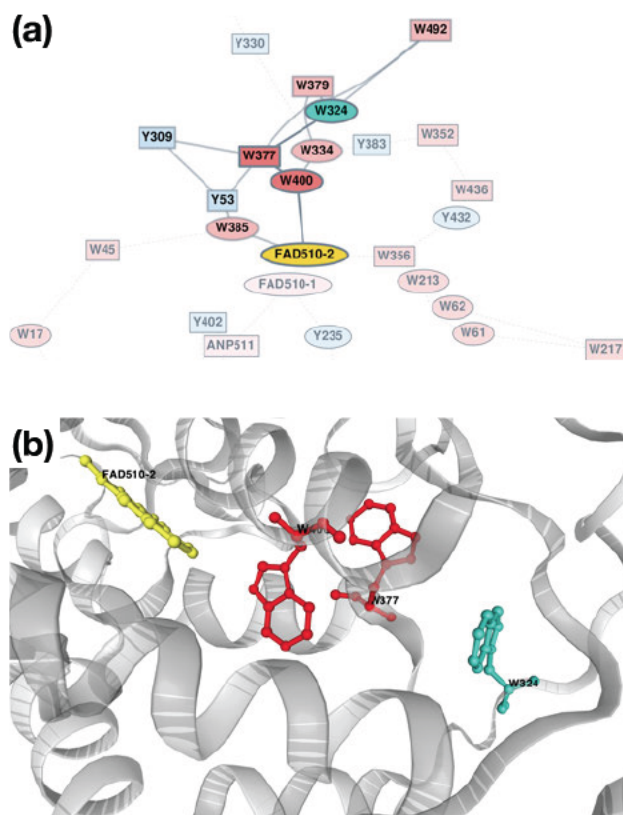
the graph can be tuned using the Edges per Vertex and Standard Deviation (SD) sliders (Fig. 6.3c). The former specifies the percentage of the shortest edges that are kept per node, with a minimum of 2 edges being preserved as long as they satisfy the distance cutoff. Among the remaining edges, only those with length  $l \leq \bar{l}_{node} + \sigma_{node}$  are kept, where  $\sigma_{node}$  is determined by the SD parameter, and  $\bar{l}_{node}$  is the average length of the edges for a given node.

***Specifying electron/hole donor and acceptor.*** Once all of the input parameters are specified and the structure has been processed, the user then specifies the electron or hole source. The target(s) can be selected to be a single site, or the collection of all surface-exposed residues. The shortest paths are evaluated using the NetworkX package. [160] For a single target, the five shortest paths connecting the source and target are identified on the basis of Yen’s algorithm. [272] The shortest paths connecting the source to each surface-exposed residue are identified using Dijkstra’s algorithm (single path per single target residue). The paths are then grouped based on the first surface-exposed residue reached during the path, and ranked according to their effective length. The pathways are further visualized in 3D using the NGL viewer. [270, 271]

## 6.4 Applications

Below we illustrate the capabilities of eMap in predicting electron hopping pathways in proteins using *Arabidopsis Thaliana* Cryptochrome 1 (Cry1) and *Pseudomonas aeruginosa* azurin as examples.

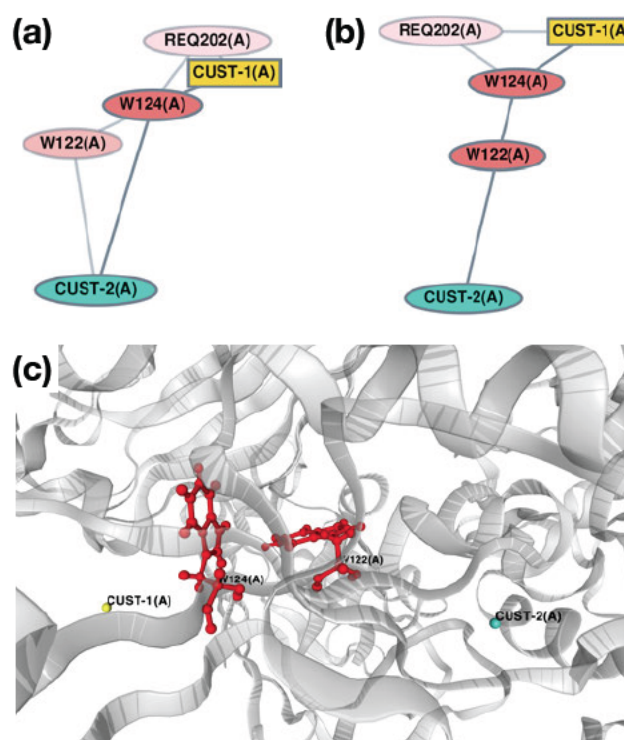
***Mapping electron transfer in Cryptochrome.*** Cryptochromes are photoactive flavoproteins with diverse biological functions, including being an integral part of a circadian clock machinery and likely being involved in magnetoreception by avian birds. [273] Photoactivation of the plant Cry1 proceeds via photoinduced electron



**Figure 6-4:** Results of the eMap analysis for Cry1 protein: shortest path connecting flavin of FAD (FAD510-2) and the terminal Trp from the Trp triad (W324) visualized in 2D (a) and 3D (b).

transfer to flavin of FAD cofactor from a neighboring Trp residue. [273–275] The resulting hole then propagates to the surface of the protein via so-called Trp triad, three conserved Trp residues (W400 – W377 – W324 in Cry1). [273, 274, 276, 277] Here, we analyze Cry1 structure (PDB ID 1U3D [278]) with eMap. Once the structure is parsed by eMap, three non-protein aromatic sites are identified: FAD510-1, FAD510-2, and ANP511 in addition to standard side chains of the aromatic residues. The two former sites are adenine and flavin of FAD cofactor, whereas ANP511 is adenine of AMP-PNP bound to the protein. All three non-amino acid aromatic sites as well as all Trp and Tyr residues have been included into the analysis. Default values for all thresholds have been used in this simulation. After the structure has

been processed by eMap, the source and target sites have been specified. The source of the electron/hole transfer is chosen to be flavin of FAD (FAD510-2) and W324 is selected as the hole target. With this selection of the source and target, eMap successfully identifies the electron transfer channel active upon photoactivation of Cry1: FAD510-2 – W400 – W377 – W324. The identified path is shown in bold in the 2D graph (Fig. 6.4a) and in 3D using the NGL viewer [270, 271] (Fig. 6.4b).



**Figure 6.5:** The results of eMap analysis for 6MJS structure: (a) the shortest path connecting Re (CUST-2) and Cu (CUST-1) atoms identified using the default parameters for the graph edges generation; (b) the shortest path connecting Re (CUST-2) and Cu (CUST-1) identified using tighter Distance Cutoff (13 Å); (c) 3D image of the pathway identified with 13 Å cutoff.

*Electron transfer in azurin mutant.* As shown above eMap can efficiently predict electron transfer pathways in proteins. Yet, simple distance-dependent penalty functions also impose some limitations. In particular, the intermediate nodes might be

missing in the predicted shortest path. To illustrate this, an example of *Pseudomonas aeruginosa* azurin mutant [279] is considered. The mutant was originally designed to demonstrate the increase in the long-range electron transfer rate by introducing two intermediate Trp residues between hole donor (Re) and hole acceptor (Cu). [279] The crystallographic structure with PDB ID 6MJS [279] was used for the analysis. The Custom Atom Range feature was used to specify a hole donor (*Re*-atom) and acceptor (*Cu*-atom). The results of the eMap analysis are shown in Fig. 6.5. CUST-2 and CUST-1 ETA sites represent Re and Cu atoms, respectively. One can see that if the default parameters for edge generation are used (Fig. 6.5a), the edge between W124 and CUST-1 is present on the graph, and, therefore, the shortest path found by eMap corresponds to CUST-2–W124–CUST-1 hopping pathway, rather than to a pathway involving two Trp residues. Thus, if a direct hopping is allowed in the model (i.e. the corresponding edge is present in the graph), the resulting one-step hopping will be always preferred by eMap over a multi-step hopping, which may or may not be the case in the actual protein system. If the Distance Cutoff criterion is tightened (13 Å) the edge between W124 and CUST-1 site is discarded, and now the most efficient pathway is CUST-2–W124–W122–CUST-1 (Fig. 6.5b, c), the pathway verified and characterized experimentally for this protein.

## 6.5 Conclusions

The eMap web application is a user-friendly tool for predicting and visualizing electron hopping pathways in proteins based on their crystal structure. On the basis of an empirical tunneling model, the software predicts shortest tunneling pathways connecting the specified electron or hole donor with either user-specified acceptor or all surface-exposed residues. The software is publicly available at <https://emap.bu.edu> free of charge.

## Chapter 7

# **eMap 2.0: A web-based platform for identifying electron transfer pathways in proteins and protein families**

The manuscript below has been prepared for submission to the journal Wiley Interdisciplinary Reviews: Computational Molecular Science. The developments described therein are implemented in the [eMap](#) software, and the open source PyeMap backend is publicly available at [GitHub](#) and is distributed on [PyPI](#). The Supporting Information for this article is provided in Appendix C.

## 7.1 Introduction

Biological processes, such as respiration [280, 281], photosynthesis [282, 283], and DNA photo-damage repair [284, 285] utilize electron or hole transfer. Structural analysis of the protein database has shown that approximately one third of the structurally characterized proteins exhibit extensive chains of Trp and Tyr residues. This observation led to the suggestion that these chains serve as protective channels against oxidative stress by guiding potentially chemically reactive holes away from the active site towards the surface of the protein [42]. Indeed, Trp and Tyr have been shown to participate in electron/hole shuttling in various protein families [286] including but not limited to: cryptochromes [287, 288], photolyases [285, 289], cytochrome c peroxidase [290], blue light receptors [291], Photosystem II [292], oxalate decarboxylases [293], and ribonucleotide reductases [294].

Theoretical description of electron transfer in proteins that yields quantitative observables such as electron transfer rates is a challenging but often a feasible task provided that the key players are known either from experimental data (e.g. mutagenesis studies), or from preliminary simulations [294–297]. Detailed characterization of electron transfer requires accounting for the dynamic behavior of a protein and careful description of nuclear and electronic degrees of freedom, especially for proton-coupled electron transfer events [298]. While this can be achieved for a single protein system when the participating residues and co-factors are known, rigorous quantitative characterization of electron transfer process without any *a priori* knowledge is seldom feasible computationally.

Several theoretical models have been proposed to enable blind screening for ET pathways in proteins using the protein structure as an input. The Pathways model [146] proposed by Beratan and Onuchik describes ET as a sequence of through-bond, through-space and through-H-bond atom-to-atom tunneling events. Each electron



hop is parameterized by an empirical penalty function exponentially decaying with the distance. The model relies on a graph theory algorithms to find the most efficient electron transfer pathway as a shortest path on a connected graph. Each protein atom represents a node in the graph, and the edge lengths are specified by the penalty function [146]. The outcome of the simulation is a series of pathways ranked based on the predicted tunneling matrix element values. A more rigorous approach based on semi-empirical quantum-chemical calculations has been proposed by Stuchebrukhov and co-workers [257]. The model and the corresponding software, Electron Tunneling in Proteins program (ETP), [257] quantifies and visualizes electron tunneling current through a protein. The recently proposed EHPATH model [299] by Beratan and co-workers targets electron/hole hopping through redox-active sites in the protein. EHPATH identifies and ranks electron transfer pathways in the protein based on pathway's mean residence time. EHPATH relies on a kinetic model parametrized using Marcus rate constants for the electron hopping events, and, thus inevitably requires information on the Gibbs free energies, reorganization energy, and electronic couplings [299]. The eMap model [151] proposed earlier by us can be viewed as an intermediate between Pathways and EHPATH approaches. Similar to EHPATH the model targets electron/hole hopping processes through redox-active sites, avoiding atomistic electron flow description, however it otherwise follows the Pathways approach, keeping only through-space tunneling penalty functions. The approach was used successfully in predicting existing electron transfer pathways in cryptochromes, and more examples are given below. The simplicity of the model allows for a quick screening for efficient electron transfer pathways based solely on the protein crystal structure. The model has been implemented as a web application with an intuitive interface that is available at <https://emap.bu.edu>. Here we describe the new features implemented in the current release of eMap 2.0, in particular the ability to screen

series of proteins for shared electron transfer pathways. We discuss multiple modes of how one can use eMap 2.0 to analyze a single protein or protein families, and present test cases which show the potential of the software to predict individual and shared ET pathways.

## 7.2 Model

Following the Pathways model [146], eMap relies on graph theory to identify candidates for efficient electron transfer pathways in a protein. In eMap 2.0, we apply graph mining approaches to search for shared electron transfer pathways (or structural motifs) among a group of protein structures. Before describing the eMap model, we first establish some definitions which will be used throughout the paper.

### 7.2.1 Preliminaries

Below we introduce relevant graph theory definitions used in the single protein and multiple protein analysis models.

**Graph.** A graph  $G(N,E)$  is an ordered pair consisting of two sets:

- $N$  is a set of *vertices* or *nodes*
- $E$  is a set of *edges*, which connect pairs of nodes

In a *labeled* graph, the nodes and edges are assigned distinguishable labels. If the edges represent ordered pairs of nodes, the graph is *directed*, otherwise it is an *undirected* graph. One can also assign numerical values to each edge in the graph, in which case the graph is considered to be a *weighted* graph. Otherwise, it is considered an *unweighted* graph. The eMap model employs weighted, undirected graphs.

**Subgraph.** If  $G'=(N',E')$  is a subgraph of  $G(N,E)$

- $N'$  is a subset of  $N$
- $E'$  is a subset of  $E$

**Isomorphism.** If  $G'=(N',E')$  is isomorphic to  $G(N,E)$

- there exists a one-to-one and onto mapping between  $N$  and  $N'$
- there exists a one-to-one and onto mapping between  $E$  and  $E'$

**Subgraph isomorphism.** For two graphs  $G(V,E)$  and  $H(V',E')$ ,  $G$  and  $H$  are subgraph isomorphic if there exists a  $G'(V_0,E_0)$  such that:

- $G'$  is a subgraph of  $G$
- $G'$  is isomorphic to  $H$

The task of identifying all such  $G'$  is known as the *subgraph matching* problem, and we refer to individual  $G'$  as subgraph isomorphisms[300].

### 7.2.2 Single protein analysis

The eMap model of a single protein can be viewed as a coarse-grained adaptation of the *Pathways* model [144–146, 258], where only through-space tunneling between aromatic residues and other user-specified sites is considered. In the original *Pathways* model, an electron/hole transfer pathway between a specified donor and acceptor is described as a series of through-space, through-covalent bond, and through-hydrogen bond tunneling events. When multiplicative penalty functions ( $\epsilon_{i/j/k}$ ) for each tunneling event are employed, the tunneling matrix element ( $T_{DA}$ ), an effective donor-acceptor coupling, has the following form: [146]

$$T_{DA} = const \times \prod_{i \in space} \epsilon_i \times \prod_{j \in bond} \epsilon_j \times \prod_{k \in H-bond} \epsilon_k \quad (7.1)$$

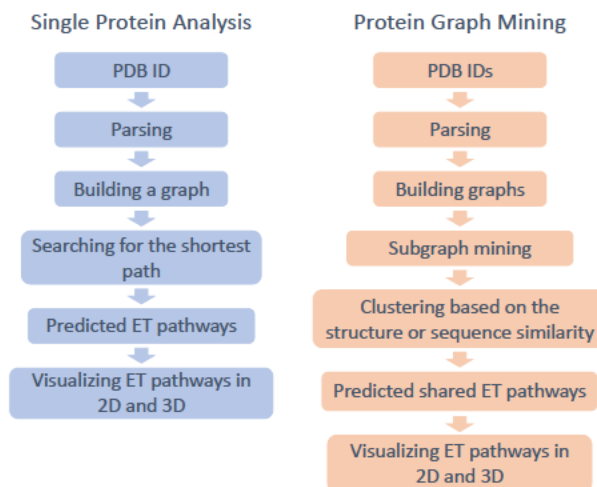
Thus the task of finding the most efficient pathways can be formulated as identifying those which maximize the matrix element  $T_{DA}$ .

In the eMap model, the protein crystal structure is used to construct an undirected weighted *protein graph*, where each aromatic residue or other user-specified site represents a node, and each edge has a weight associated with a through-space tunneling penalty function. To make use of standard graph theory methods, it is convenient to operate with non-negative edge weights and additive parameters rather than multiplicative penalty functions. This is done using modified penalty functions,

$$P' = -\log \epsilon, \quad (7.2)$$

with  $\epsilon = \alpha \exp[-\beta(R - R_{\text{offset}})]$ , where  $\alpha$ ,  $\beta$ , and  $R_{\text{offset}}$  are empirical parameters, and  $R$  is the distance between the sites defined either as the distance between the centers of mass of the sites or as the distance between the closest atoms. Since minimizing the sum of the modified functions  $P'$  is equivalent to maximizing the product of the penalty functions  $\epsilon$ , finding the path with the largest  $T_{DA}$  value is equivalent to finding the shortest path in a graph with edge lengths defined by the modified penalty functions,  $P'$ . [144–146, 258] For graphs with non-negative weights, this problem is easily solved using Dijkstra’s algorithm[301] to find the optimal path, or Yen’s algorithm[272] to find deviations from the optimal path, for a given source and target.

Currently, the default values for  $\alpha$  (=1.0),  $\beta$  (=2.3 Å<sup>-1</sup>), and  $R_{\text{offset}}$  (=0.0 Å) give edge weights which are equal to the scaled distances between hopping sites, resulting in a purely distance dependent model. To reduce the complexity of the protein graphs, and therefore provide a more meaningful network structure, the edges of the graph are pruned, favoring those with the smallest weights[302]. Lastly, nodes in the graph are classified as buried or surface-exposed using either the residue depth [262–264] or



**Figure 7-1:** The workflow of the eMap single protein analysis (left panel) and common subgraph mining for a series of proteins (right panel).

relative solvent accessibility [265–267] criteria. Details of the implementation can be found in Refs. 151 and 302.

eMap 2.0 introduces only minor changes in the single-protein analysis routine, such as an expanded set of automatically identified redox-active sites. The workflow of the eMap single protein analysis is illustrated in Fig. 7-1. The user can use PDB ID or structure files (PDB or CIF) as an input for the software. The structure is then parsed, and redox-active sites are identified. By default aromatic side-chains of Tyr and Trp are considered active along with aromatic fragments of co-factors. In addition to aromatic fragments, redox-active metals and Fe-S clusters are automatically identified and can be included in the analysis as redox-active sites. In the next step, a graph representing the protein structure is generated. The user then specifies an electron/hole donor and optionally acceptor, and the software searches for the shortest paths connecting the donor and acceptor or donor and all residues on the protein surface. The shortest pathways are visualized on a graph (in 2D) and in 3D using NGL viewer [270, 271].

### 7.2.3 Searching for common electron/hole pathways

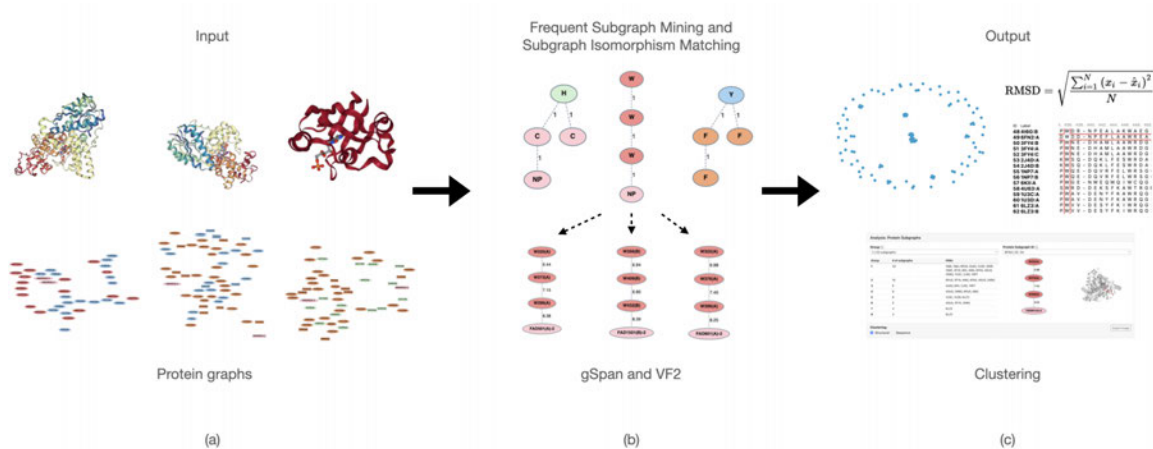
Single protein analysis results in a protein graph for the protein of interest, and each pathway identified by eMap can be viewed as a *subgraph* of the protein graph. Therefore, using the eMap model, the task of identifying pathways (and other structural motifs) which are shared among a group of proteins can be formulated as identifying frequently occurring subgraphs in a data set of protein graphs. This is a well-studied problem in graph theory known as *frequent subgraph mining* (FSM)[152, 153].

FSM techniques are aimed at the discovery of patterns in graphs that exhibit a particular structure which is deemed interesting or relevant[303]. What makes a pattern “interesting” or not depends on the application. FSM techniques are useful where there are complex relationships between data entities, and are commonly used in areas such as analysis of XML documents and web usage[304]. In the scientific community, FSM has seen numerous applications in bio- and cheminformatics contexts, including drug discovery[154], identification of protein-ligand interfaces[155, 156] and the study of evolutionary relationships through mining phylogenetic trees[157–159]. While others have used FSM techniques to identify conserved three-dimensional structures or motifs in proteins[155, 305–307], to our knowledge, eMap is the first to do so with an emphasis on electron transfer.

The FSM problem has a large degree of computational complexity. The most basic task, determining whether a graph  $G$  contains a pattern  $P$ , is an instance of the subgraph isomorphism problem, which is NP-complete[143]. Further, searching over all possible patterns  $P$  becomes rapidly intractable for large and dense graph datasets. Several approaches have been developed to efficiently tackle this problem, where the goal is typically to identify patterns which occur more often than a given threshold[303]. Here, we use the gSpan algorithm, implemented in the gspan-mining Python package[143, 308, 309]. gSpan is a popular and complete algorithm which

identifies all subgraphs  $P$  that satisfy a minimum **support** threshold. For a graph dataset,  $D = \{G_1, G_2, \dots, G_n\}$ , the  $support(P)$  is the number of graphs  $G \in D$  for which  $P$  is a subgraph. The details of the algorithm can be found in Ref. 143; here we focus on the aspects which relate to its usage in the eMap software.

Frequent subgraph mining in eMap can be divided into 3 major steps: (1) graph generation (2) mining and matching (3) clustering of identified subgraphs (Figure 7·2)[303]. Below, we outline how each of these steps is implemented in the eMap-gSpan model for identifying shared pathways and motifs in protein graphs.



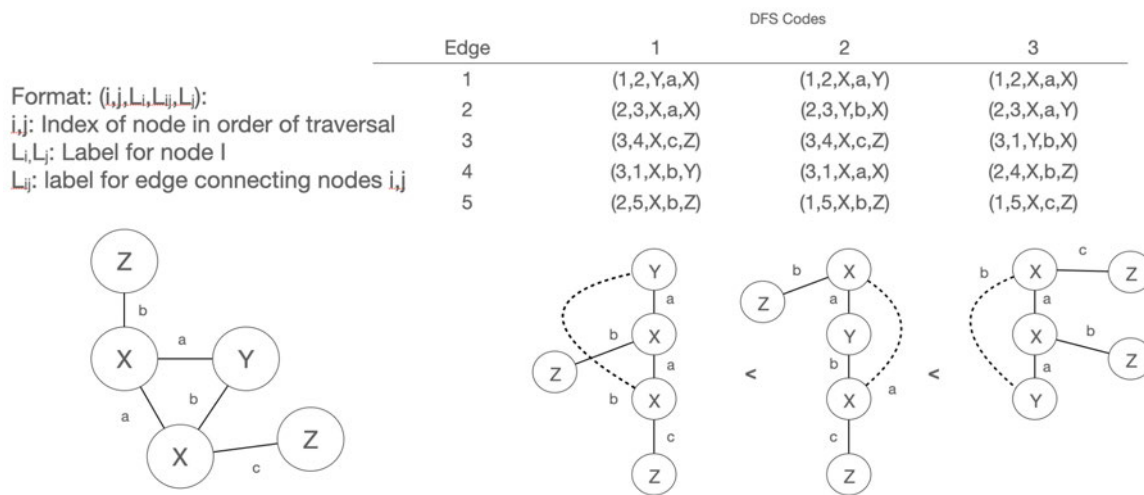
**Figure 7·2:** Flowchart illustrating the three major steps of frequent subgraph mining with eMap. The first step (a) generates a graph data set which consists of protein graphs created by the single protein version of eMap. In the second step (b), the graph data set is mined for frequently occurring subgraph patterns. For each identified pattern, all instances of said pattern within the data set are found using graph matching, which yields a set of *protein subgraphs*. In the final step (c), the protein subgraphs matching a selected pattern are clustered into groups based on similarity. Shown in the bottom of panel C is the web interface which enables users to browse through clusters of similar protein subgraphs, visualized in 2D and in the 3D protein structure.

## Graph Generation

Processing a PDB structure by eMap yields a graph with labeled nodes, where each node is an electron-transfer-active (ETA) site, and edges weighted based on the distances between sites. When the analysis is performed on  $n$  proteins the result will be a set of  $n$  protein graphs  $D = \{G_1, G_2, \dots, G_n\}$ . gSpan exploits depth-first-search (DFS) together with a uniquely defined canonical labeling system to efficiently discover common subgraphs in the dataset,  $D$ . It relies on DFS codes, which are sequences of edges visited during a depth-first search traversal. Since there are many possible DFS codes for a graph, gSpan uses a lexicographic ordering scheme to generate *minimum DFS codes* (Fig. 7.3). In the simplest implementation, all nodes and edges are of the same type, and thus the algorithm will identify frequent patterns of connectivity in the graph database. Both the efficiency and descriptive power of gSpan are enhanced when the DFS code distinguishes between different types of nodes and edges.

The protein graphs generated by eMap naturally lend themselves to classifications of nodes and edges. By default, each standard amino acid type included in the graph is assigned a label, and all non-standard residues included in the analysis are labeled as ‘NP’ for non-protein. To add additional flexibility, users can specify a group of standard amino-acid residue types to be given the label ‘X’ (the usual notation for unknown residue type), which enables these residues to be deemed equivalent in isomorphism testing. The edges can also be labeled based on their weights by specifying cutoff thresholds to define different edge types. For example, setting the thresholds 8 Å and 12 Å defines three categories of edges: 0-8 Å, 8-12 Å, and  $\geq 12$  Å.

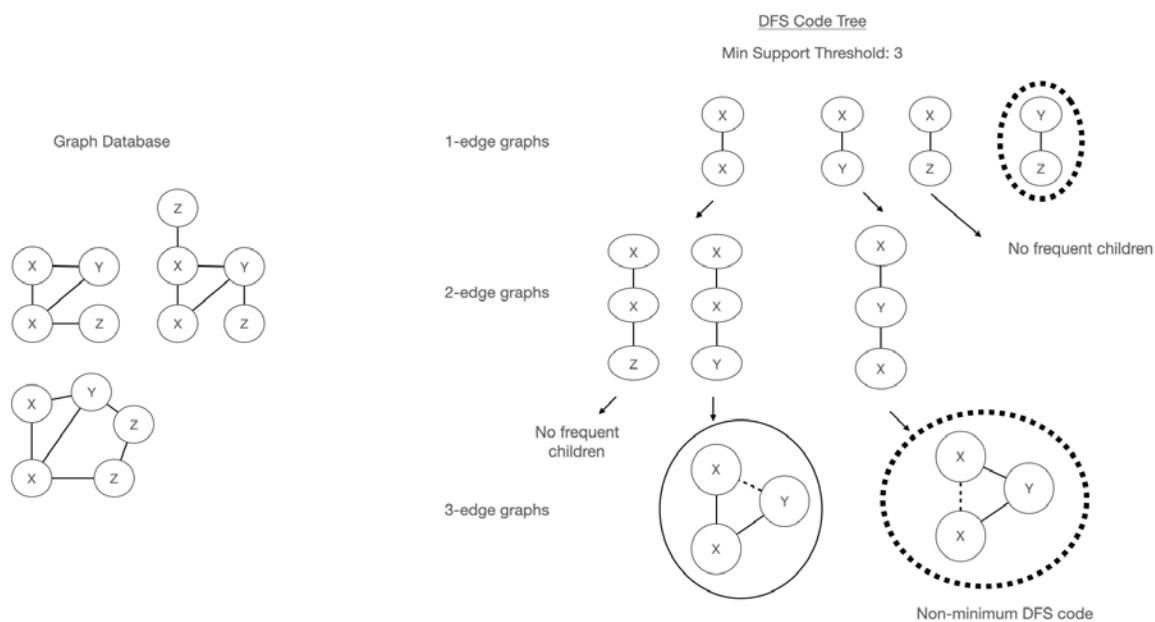




**Figure 7.3:** Example of DFS codes for a graph with node labels  $\{X, Y, Z\}$ , and edge labels  $\{a, b, c\}$ . Back edges (edges between nodes which have already been visited) are indicated as dashed lines. Many different DFS codes can be generated for a given graph by choosing different starting nodes and making different choices at each step of the traversal. By the gSpan lexicographic ordering scheme, the ordering of these three codes is  $3 < 2 < 1$ , i.e. 3 (the most right column) is the minimal DFS code among these three[143].  $X, Y, Z$ , etc. correspond to the chemical nature of the redox-active site (e.g. aromatic amino acids) and  $a, b, c$  label edges classified based on their length.

## Mining and Matching

Unlike earlier algorithms, gSpan does not rely on explicit generation of candidate subgraphs, and instead grows frequent  $(k+1)$ -edge DFS codes from frequent  $k$ -edge DFS codes, which essentially combines the creation of candidate subgraphs with testing for isomorphism. This allows for an aggressive reduction of the search space through pruning of infrequent and non-minimum DFS codes while maintaining the completeness of the algorithm[143] (see Fig. 7.4) up to a specified support threshold. The end result of gSpan is a set of minimum DFS codes  $\{P\}$  corresponding to the subgraph patterns which meet the minimum support threshold, and the support of each subgraph pattern within the data set.



**Figure 7-4:** Illustration of depth-first traversal of the search space by the gSpan algorithm for an example graph data set. Starting from 1-edge graphs, the search space is expanded by adding additional edges, those which are infrequent are pruned (indicated by a dashed circle). The DFS code tree is constructed using the canonical lexicographic ordering, so when a non-minimum DFS code is encountered, the entire branch below can be pruned[143].

In some contexts, simply identifying the patterns and frequencies is sufficient, but in the case of eMap, it is crucial to identify the actual pathways (i.e. the specific residues involved) in each PDB. For this purpose, the VF2 graph matching algorithm, as implemented in NetworkX, is employed[141, 142, 160, 310]. For each protein graph  $G$  which supports a given subgraph  $P$ , all subgraph isomorphisms  $G'$  are identified. When done for the entire graph dataset, this gives a set of *protein subgraphs* that all share the same connectivity and node/edge labels.

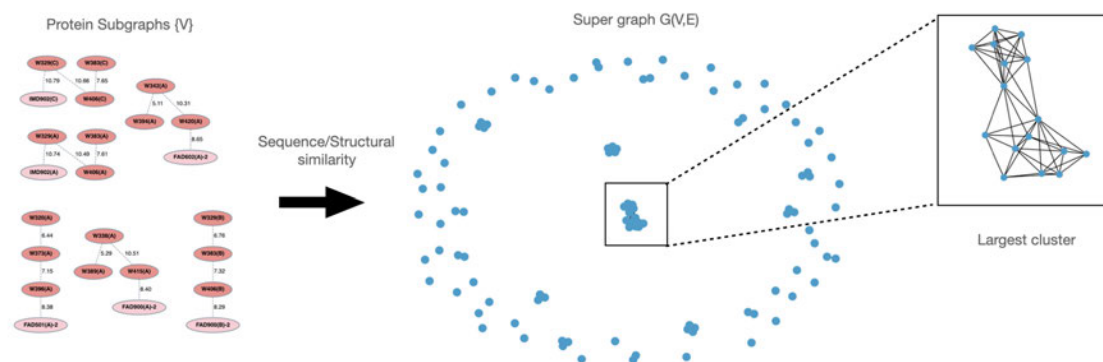
In addition to blind searches performed with the gSpan algorithm, one can also perform guided searches to find protein subgraphs which match a previously known pattern  $P$ . In this case, the problem is reduced to one of subgraph matching, and we

use the VF2 algorithm[141, 142, 160] to search each protein graph  $G$  for subgraphs which are isomorphic to  $P$ .

## Clustering

For a given frequent subgraph pattern  $P$ , there can be hundreds of matches, and thus a meaningful analysis requires grouping them based on some metric of relevance. Since the objective of eMap is to identify shared pathways, we perform a clustering analysis to group protein subgraphs by their similarity. eMap employs two metrics of similarity: structural similarity, which is based on a modified root mean squared distance (RMSD) between residues, and sequence similarity, which is based on a multiple sequence alignment performed using the MUSCLE package[311]. Details of the similarity metrics are described in Sec C2 of Appendix C.

The clustering algorithm is as follows (see Fig. 7-5). For a given subgraph pattern  $P$ , we have a set of protein subgraphs  $\{V\}$  which correspond to specific pathways in protein structures that match the pattern  $P$ . We construct a supergraph  $G(V,E)$ , where each node corresponds to a protein subgraph, and two nodes share an edge if and only if they are the corresponding protein subgraphs are deemed sufficiently similar by the chosen similarity metric. The supergraph  $G$  is composed of one or multiple connected components, and each connected component corresponds to a cluster of similar protein subgraphs. As such, structurally or sequentially ‘conserved’ pathways will belong to the same cluster of protein subgraphs. The end result of the analysis is that users can browse through different subgraph patterns, and then for each subgraph pattern, examine the different groups of structurally or sequentially similar protein subgraphs.



**Figure 7-5:** Illustration of algorithm used to cluster a set of protein subgraphs  $\{V\}$  which match a particular subgraph pattern. The largest cluster in the super graph  $G(V,E)$  contains 16 protein subgraphs which are deemed structurally or sequentially similar.

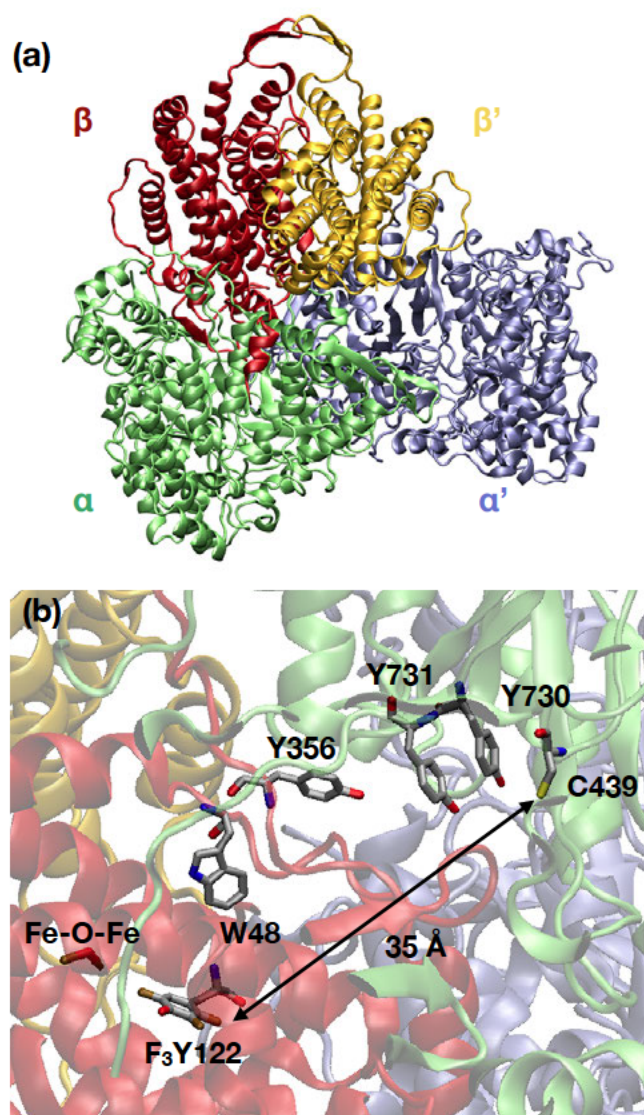
### 7.3 Application examples

Below we discuss several different modes of how the eMap 2.0 software can be used to explore electron transfer pathways in a single protein and protein families. For multiple protein analysis, we present two test cases. The first corresponds to a blind search, where no information on the shared pathway/structural feature is given as an input, and the second is an illustration of a search for a shared electron transfer pathway between a user specified donor and acceptor.

#### 7.3.1 Efficient electron transfer pathway in a single protein

eMap has been used before to successfully identify an existing ET transfer channel in a single cryptochrome protein [151]. Here, we present a case study of radical transfer pathway in class Ia ribonucleotide reductase (RNR) from *Escherichia coli*. RNRs catalytically reduce ribonucleotides into deoxyribonucleotides and, thus, play a key role in DNA biosynthesis. Ribonucleotides reduction by class Ia RNRs proceeds via formation of a thiyl radical generated by a long-range radical transfer from diferric-tyrosyl radical co-factor in a neighboring unit (Fig. 7-6). The ET pathways involves

six residues from the neighboring  $\alpha$  and  $\beta$  subunits:  $\beta$ Y122– $\beta$ W48– $\beta$ Y356– $\alpha$ Y731– $\alpha$ Y730– $\alpha$ C439 [312].



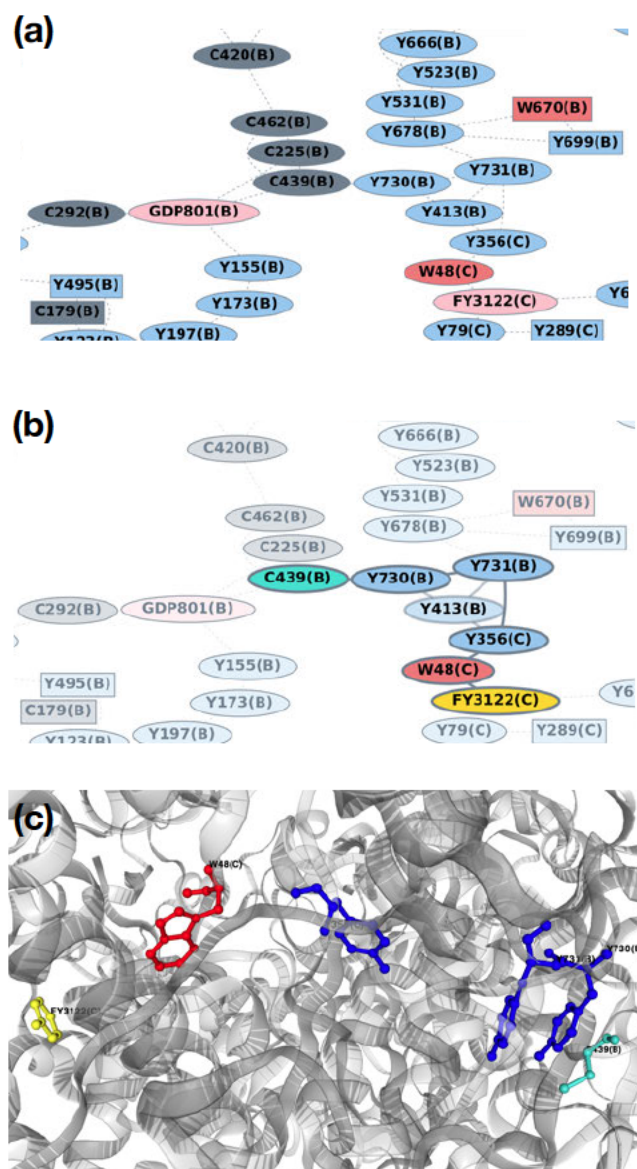
**Figure 7.6:** Structure of the class Ia RNR from E.coli (a) and inter-unit radical transfer pathways involved in the catalysis,  $\beta$ Y122– $\beta$ W48– $\beta$ Y356– $\alpha$ Y731– $\alpha$ Y730– $\alpha$ C439 (b).

Below we discuss the results of the eMap analysis for the protein using the recently resolved electron cryo-microscopy (cryo-EM) structure (PDB ID: 6W4X) [313] as the input. As shown in Fig. 7.7a, the software detects all aromatic moieties of co-factors

**Table 7.1:** Four identified ET pathways from Y122 ( $\beta$  subunit) to C439 ( $\alpha$  subunit) in class Ia RNR from *E.coli*.

Source	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Score
FY3122(C)	W48(C)	Y356(C)	Y731(B)	Y730(B)	C439(B)		39.26
FY3122(C)	W48(C)	Y356(C)	Y413(B)	Y730(B)	C439(B)		41.23
FY3122(C)	W48(C)	Y356(C)	Y731(B)	Y413(B)	Y730(B)	C439(B)	45.83
FY3122(C)	W48(C)	Y356(C)	Y413(B)	Y731(B)	Y730(B)	C439(B)	46.61

(thymidine triphosphate, TTP, and guanosine diphosphate, GDP) and non-standard aminoacids (3-fluoro tyrosine,  $F_3Y$ ). Including Y, W, C, and all non-aminoacid aromatic sites into the analysis and considering only chains B ( $\alpha$ ) and C( $\beta$ ) yields the graph shown in Fig. 7.7b. Specifying FY3122(C) as a source and C439(B) as a target of charge transfer, and using default simulation parameters, the experimentally observed pathway is the top ranked pathway (Table 7.1), i.e. is predicted to be the most efficient ET pathway channel.



**Figure 7.7:** The results of the eMap analysis for the 6W4X structure. Only B and C chains were included in the analysis, all T,W, and C residues were considered as ETA sites. All aromatic moieties of co-factors and non-standard aminoacids were considered as ETA sites. The default parameters used for the simulation are provided in Table C1 in Appendix C. The graph structures of the class Ia RNR from *E.coli* and inter-unit radical transfer pathways involved in the catalysis,  $\beta$ Y122– $\beta$ W48– $\beta$ Y356– $\alpha$ Y731– $\alpha$ Y730– $\alpha$ C439, are shown in (a) and (b), and the top ranked pathway is visualized in the crystal structure in (c).

### 7.3.2 Shared electron transfer pathways and structural motives

The main new feature of the eMap 2.0 software is screening a group of PDB files for the existence of common electron transfer pathways or structural motifs. There are two primary ways of using eMap to explore electron transfer in protein families. The first is a blind search of a shared pathways in a series of proteins. The other is a guided search when some mechanistic information on the participating residues is available. Below we present case studies illustrating these two applications.

#### Blind search

**Cryptochromes and photolyases.** To illustrate the graph mining by eMap we have first considered a series of 27 flavoprotein crystal structures (1U3D, 1U3C, 6LZ3, 2J4D, 6FN2, 1NP7, 3ZXS, 4GU5, 4I6G, 6PU0, 1IQR, 4U63, 6KII, 3FY4, 1DNP, 1QNF, 1IQU, 2WB2, 6RKF, 1O96, 1EFP, 1O97, 1X0P, 2Z6C, 1G28, 4EER, 2IYG). The set includes 7 plant cryptochromes, 3 animal cryptochromes, 8 photolyases, and 9 non-cryptochrome and non-photolyases proteins. Light absorption by cryptochromes and photolyases initiates photo-induced electron transfer involving FAD cofactor and conserved Trp residues [273, 314–316]. Plant cryptochromes ET pathways involve three conserved Trp residues (one replaced by Tyr in some cases), the so-called Trp-triad [273, 314]. The ET pathway in animal cryptochromes is extended by forth aromatic residues (Tyr or Trp) [317]. To test the capability of eMap to identify the shared electron transfer pathway in cryptochromes and photolyases, we performed a blind search for all frequent subgraphs. The search was performed using the default parameters and the minimum support of 18. The mining yielded 25 shared subgraph patterns with the support above 18. However, only 5 of those contained shared subgraph groups with significant structural similarity (Fig. 7·8). The established electron transfer pathway corresponds to the subgraph patterns 21 and 22 in Fig.



7.8 and is found by eMap in 16 proteins, those that are either cryptochromes or photolyases. The only two cryptochrome structures which did not contain the Trp triad are 6LZ3 and 3ZXS. 6LZ3 is a cryo-EM structure of a tetramer of CRY1c from *Zea mays* with the second Trp the Trp-triad replaced by Ala (W368A mutation following the residue indexing in 6LZ3) [318]. 3ZXS is an X-ray structure of CRYb from *Rhodobacter Sphaeroides* that contains Tyr (Y387) in place of the first Trp in the triad [319]. Thus, the eMap software correctly identified all proteins (cryptochromes and photolyases) that exhibit the conserved Trp triad.

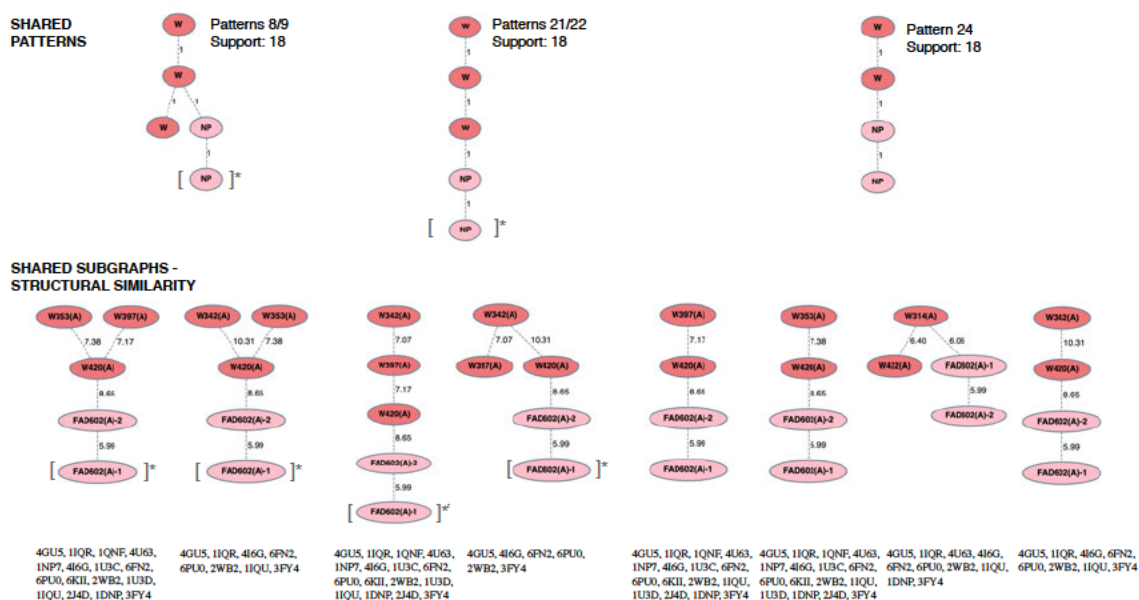
Interestingly, some of the other shared subgraphs identified by eMap could also have mechanistic importance. Subgraph patterns 8 and 9 include adenine and flavin moieties of the FAD cofactor as well three Trp residues. Two of which belong to the conserved Trp triad (W420 and W397 in PDB ID 4GU5, see Fig. 7.9), while the third Trp (W353) is also highly conserved among 18 considered photolyases and cryptochromes, and appears in 16 of them. Importantly, the remaining subgraph pattern only contain clusters with 8 or less proteins, which indicates that they are unlikely correspond to mechanistic pathways common for cryptochrome or photolyases families of proteins.

### 7.3.3 Guided search

#### **Heme-to-heme electron transfer in bacterial cytochrome c peroxidases**

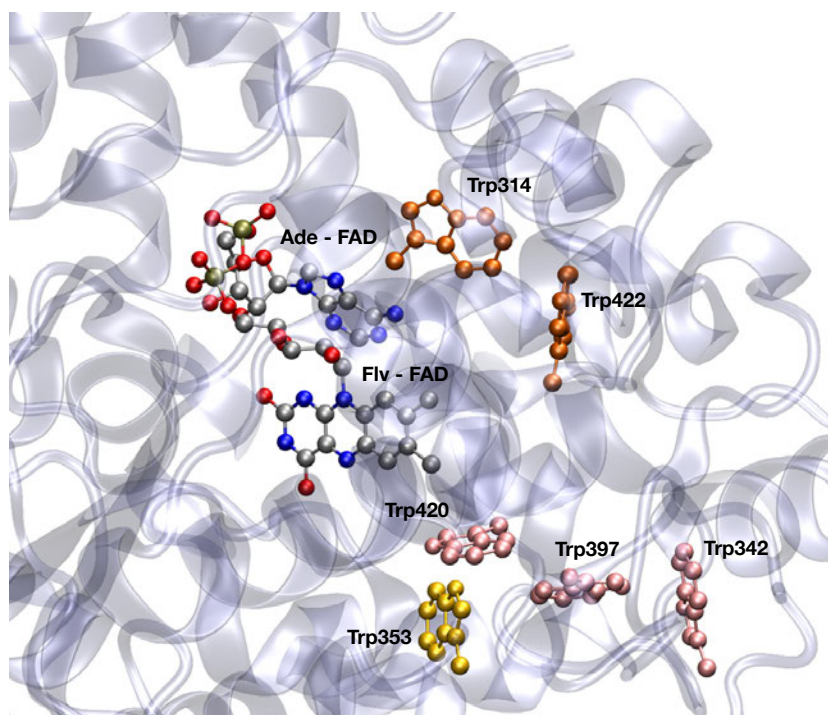
To illustrate eMap’s ability to make evolutionary suggestions given limited mechanistic information, we look towards the bacterial cytochrome c peroxidase (bCcPs) family of proteins, which catalyze the reduction of hydrogen peroxide, and are part of the oxidative stress protection machinery of many Gram-negative prokaryotes [320].

The catalytic cycles of some of the reductively activated bCcPs rely on electron transfer between the *heme* sites. One of the proposed pathways for electron shuttling between the two heme center involves a Trp94 residue [320]. To test whether eMap



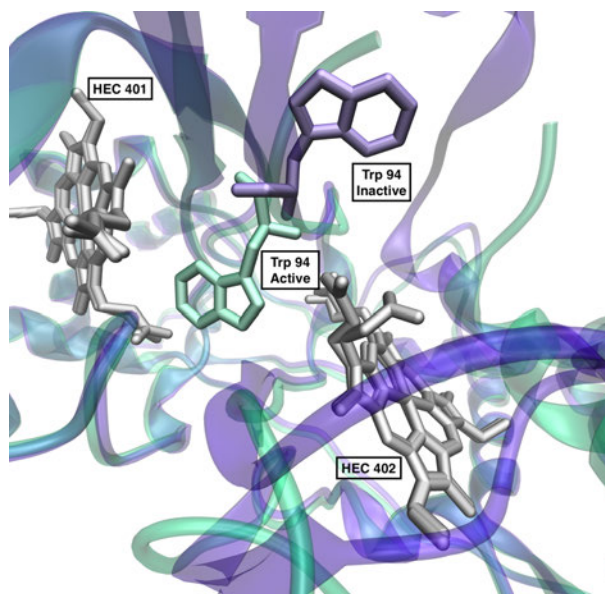
**Figure 7-8:** Relevant shared graph patterns and subgraphs for a test set of 27 flavoproteins, containing cryptochromes, photolyases, and non-cryptochromes and non-photolyases proteins. Only the shared patterns with clusters with more than 15 proteins based on structural similarity criteria are shown. Groups 8/9 and 21/22 only differ by presence (8 and 21) or absence (9 and 22) of adenine of FAD co-factor, which is represented by  $[]^*$  in the figure. The residue IDs correspond to *Drosophila* Cryptochrome, PDB ID 4GU5.

analysis supports this channel, we performed a search of Heme\*–Heme pathways, where '\*' indicates a wild card character, i.e. any residue. We considered a subset of 11 structures belonging to the bCcP superfamily that contain two hemes: 6V59, 6NX0, 1EB7, 2VHD, 1RZ5, 2C1U, 2C1V, 1IQC, 1RZ6, 1NML, and 1ZZH. The search was performed using the closest-atom distance criteria between residues, and the default settings were used for all other parameters (Table S1). The Heme-Trp-Heme path was found in 7 structures: 1EB7, 2VHD, 1RZ5, 2C1U, 2C1V, 1IQC, and 1ZZH. This pathway was correctly not identified in the other four structures screened. In the cases of 6V59 and 6NX0, both are native to *Burkholderia thailandensis*, and are similar to MauG proteins in that they contain a Trp to Ser mutation[321] that eliminates this pathway. 1RZ6, 1NML, and 1RZ5 are especially interesting cases because they are all



**Figure 7.9:** Residues involved in photo-induced electron transfer (canonical Trp triad, Trp420, Trp397, Trp342) as well as other Trp residues from shared subgraphs identified by eMap. The numbering and structure correspond to PDB ID 4GU5.

*Marinobacter* structures, but the pathway was only identified in the latter by eMap. The absence of the Heme-Trp-Heme pathway in the 1RZ6 and 1NML structures is explained by the fact that neither of these structures were crystallized in the active form, and the protein loop containing the Trp residue is shifted out of alignment to a position where it is no longer directly between the donor and acceptor hemes[322] (Fig. 7-10). The ability of eMap to make this distinction between inactive and active redox pathways highlights the predictive power of the relatively simple underlying model of electron transfer. Finally, all seven identified pathways were deemed both structurally and sequentially similar, which indicates an evolutionary relationship that is revealed by eMap analysis.



**Figure 7.10:** Active and inactive forms of chain A in the dihaem Cytochrome C Peroxidase in *Marinobacter nauticus* (PDB structures:1RZ5, 1RZ6). Shown are overlays of the active and inactive system, where Trp94 is shifted out of the central location between hemes in the inactive form. This inactive structure was not identified as a HEC-Trp-HEC pathway by eMap.

## 7.4 Conclusions

The eMap 2.0 web application is a user-friendly tool which enables prediction and visualization of electron and hole hopping pathways in single proteins, and identification of shared pathways in families of proteins. For single proteins, eMap has been used to quickly screen for efficient pathways based on an empirical tunneling model, connecting a user-specified electron or hole donor with either a user-specified acceptor or all surface-exposed residues. For families of proteins, we have demonstrated the capability of the software to identify shared pathways in cryptochromes and photolyases based solely on their crystal structures, and to provide evolutionary insights into bcCPs based on limited mechanistic information. The software is publicly available at <https://emap.bu.edu> free of charge, and the open-source backend PyeMap

is available at <https://github.com/gayverjr/pyemap/>.

## Chapter 8

# Summary and Outlook

The new methods and software discussed in this dissertation have advanced the computational infrastructure available to researchers in the field of electron-molecule interactions, and have provided researchers studying charge transfer in biological systems with an interactive platform for identifying relevant pathways.

In Chapter 3, we examined the performances of the box and smooth Voronoi CAPs for shape resonances in different types of systems[114]. We found that the Voronoi CAP provides a more flexible and robust treatment of states with different symmetries, but the advantages of box CAP can be exploited when the system is less sensitive to the CAP onset in the dimension where box CAP is absent, e.g. the clustering axis.

In Chapter 4, we described the projected CAP-EOM-CCSD approach for molecular resonances, and analyzed its performance for shape resonances in small to medium sized molecules. Additionally, we investigated the sensitivity of the new method to parameters such as the basis set and CAP onset. Our results show that a small number of EOM-EA-CCSD states is sufficient to produce projected CAP-EOM-CCSD results which closely approximate those obtained from conventional CAP-EOM-CCSD. The method can be readily applied to larger molecules and potential energy surfaces, two situations where the drastic improvement in computational efficiency is particularly advantageous.

In Chapter 5, we described the OpenCAP package, and explored the advantages

and limitations of several electronic structure methods for calculating resonance parameters. The software is publicly available on [GitHub](#), and pre-compiled wheels are distributed on [PyPI](#). OpenCAP is under active development, and there are multiple directions we intend to explore in future work. First, we plan on developing interfaces with more electronic structure packages, particularly focusing on those which are capable of performing calculations which go beyond double excitations. One possibility is the CFOUR package of Stanton et al[323], which is capable of performing advanced calculations such as EOM-CCSDT. These methods are crucial for the description of strongly correlated systems such as dipole-supported resonances[324], which cannot be accurately described with methods which truncate at double excitations, and are very challenging to characterize using multi-reference methods. Another direction is dynamical simulation of resonance states using CAP-based approaches. We are currently developing an interface with the Surface Hopping including Arbitrary Couplings (SHARC) molecular dynamics program[325–327], with the goal of performing surface leaking fewest switches surface hopping (SL-FSSH)[328, 329], which is able to incorporate resonance lifetimes into non-adiabatic molecular dynamics simulations. Finally, we intend to explore other forms of CAP and alternate numerical integration strategies for computing CAP integrals, two areas which have seldom been explored in the literature, but can now be easily investigated using the PyOpenCAP interface.

In Chapters 6 and 7, we described the development of the eMap software, and presented example applications for identifying pathways in single proteins, and for identifying shared pathways in families of proteins. The [eMap](#) web application is available free of charge, and the open-source PyeMap backend is hosted on [GitHub](#) and [PyPI](#). The most immediate goal for the future of the software is to improve the underlying model of electron transfer. Current efforts are focused on computing Marcus-like rate constants to describe hopping between aromatic sites. In order to

use shortest path algorithms, the edge weights are defined as the mean residence time associated with hopping from a donor residue to an acceptor residue[299]:

$$\tau_{\text{approx}} \cong \frac{1}{k_{D \rightarrow A}} \quad (8.1)$$

If we neglect all backward ET rates, the average time spent on the hopping chain (i.e. the efficiency of the pathway) is approximately the sum of the mean residence times for each step of the pathway:

$$\tau_{\text{approx}} \cong \sum_{n=0}^N \frac{1}{k_{n \rightarrow n+1}} \quad (8.2)$$

Several developments are underway to approximate the necessary Marcus parameters from the crystal structure. For the electronic coupling  $H_{DA}$ , we are developing machine learning models which use a structural feature set[330–332], and are trained on couplings computed at the DFT level using the frontier molecular orbital (FMO) approach[333]. We are also exploring approaches for estimating the free energy which include the effects of the solvation[334, 335] and the local electrostatic environment.



## Appendix A

# Supporting Information for Chapter 5

## A.1 Geometries

All coordinates are given in Å.

### A.1.1 $N_2$

N	0.000000	0.000000	0.548757
N	0.000000	0.000000	-0.548757
Gh	0.000000	0.000000	0.000000

### A.1.2 $C_2H_4$

C	0.669462	0.000000	0.000000
C	-0.669462	0.000000	0.000000
H	-1.232083	0.928897	0.000000
H	1.232083	0.928897	0.000000
H	-1.232083	-0.928897	0.000000
H	1.232083	-0.928897	0.000000
Gh	0.000000	0.000000	0.000000

### A.1.3 $CO_2$

C	0.000000	0.000000	0.000000
O	0.000000	0.000000	-1.162100
O	0.000000	0.000000	1.162100

### A.1.4 $CH_2O$

C	0.000000	-0.000000	0.533358
H	0.942900	0.000000	1.120983
H	-0.942900	0.000000	1.120983
O	0.000000	0.000000	-0.671617
Gh	0.000000	0.000000	-0.069317

## A.2 CAP Parameters

## A.3 Alternative MRPT schemes

**Table A.1:** Box CAP Parameters (in bohr) for systems studied in this work.

System	$R_X^0$	$R_Y^0$	$R_Z^0$
$N_2$	2.76	2.76	4.88
$CO_2$	3.331	3.331	9.539
$CH_2O$	3.872	2.958	6.086
$C_2H_4$	7.077	4.639	3.422

**Table A.2:** Computed CAP/MRPT first-order resonance parameters using IPEA shift of 0.25 a.u. and real shift of 0.2 a.u.  $E_R(\Gamma)$  are given in eV.

Molecule	Method	XMS	RMS	XDW
$CO_2$	SA(7)-CAS(5e,15o)	4.051(0.127)	4.105(0.139)	4.040(0.131)
$N_2$	SA(5)-CAS(5e,11o)	2.494(0.211)	2.560(0.240)	2.503(0.222)
$C_2H_4$	SA(4)-CAS(3e,5o)	1.787(0.253)	1.835(0.286)	1.780(0.259)
$CH_2O$	SA(7)-CAS(3e,8o)	1.172(0.162)	1.320(0.226)	1.189(0.172)

**Table A.3:** Computed CAP/MRPT first-order resonance parameters using alternative scheme without diffuse orbital. The orbitals are optimized for the anionic states only, and the neutral ground state energy is obtained using CASCI/CASPT2 with the optimized orbitals. An IPEA shift of 0.25 a.u. and an imaginary shift of 0.2 a.u. are used for the perturbation theory step for both calculations.  $E_R(\Gamma)$  are given in eV.

Molecule	Method	XMS	RMS	XDW
$CO_2$	SA(6)-CAS(5e,14o)	3.969(0.123)	4.056(0.134)	3.984(0.126)
$N_2$	SA(4)-CAS(5e,10o)	2.378(0.206)	2.483(0.224)	2.410(0.209)
$C_2H_4$	SA(3)-CAS(3e,4o)	1.717(0.253)	1.775(0.263)	1.725(0.254)
$CH_2O$	SA(6)-CAS(3e,7o)	1.070(0.170)	1.196(0.210)	1.087(0.174)

## A.4 Smaller MRCI active spaces

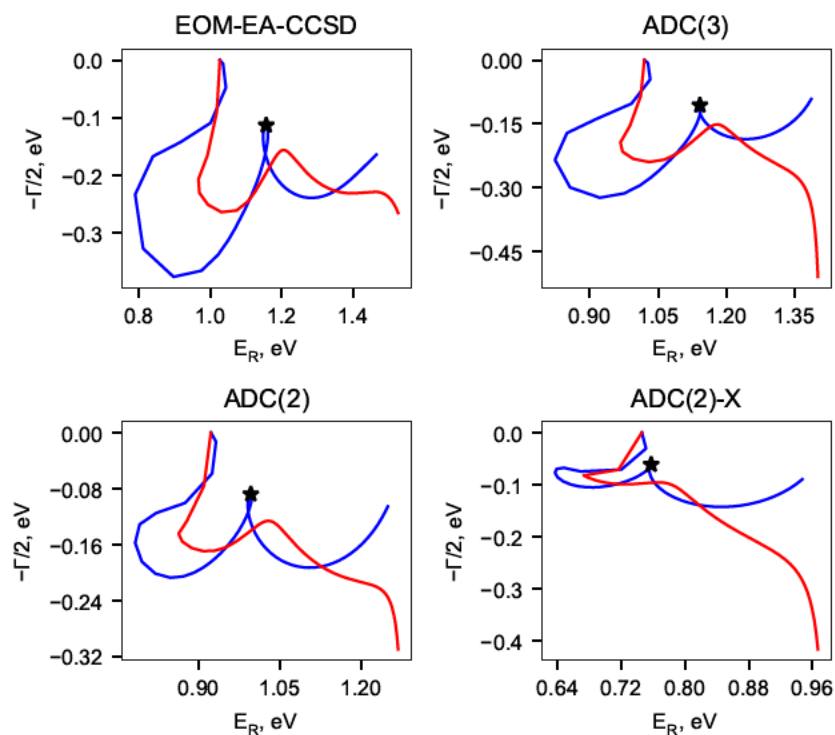
**Table A.4:** Computed CAP/MRCI first-order resonance parameters using various small active spaces. No stationary point could be identified on either trajectory for MRCISD/MRCISD(Q) for  $CO_2$ . The trajectories are provided in Figs. [A.9-A.12](#).

Molecule	Method	MR-CIS	MR-CISD	MR-CISD(Q)
$C_2H_4$	SA(1+2)-CAS(3e,3o)	2.702(0.619)	2.448(0.143)	2.277(0.127)
$N_2$	SA(1+3)-CAS(5e,8o)	3.262(0.377)	3.227(0.188)	3.009(0.425)*
$CH_2O$	SA(1+5)-CAS(3e,6o)	2.333(0.532)	2.224(0.155)	1.831(0.124)
$CO_2$	SA(1+5)-CAS(5e,12o)	4.733(0.092)	4.500(0.165)*	4.306(0.186)*

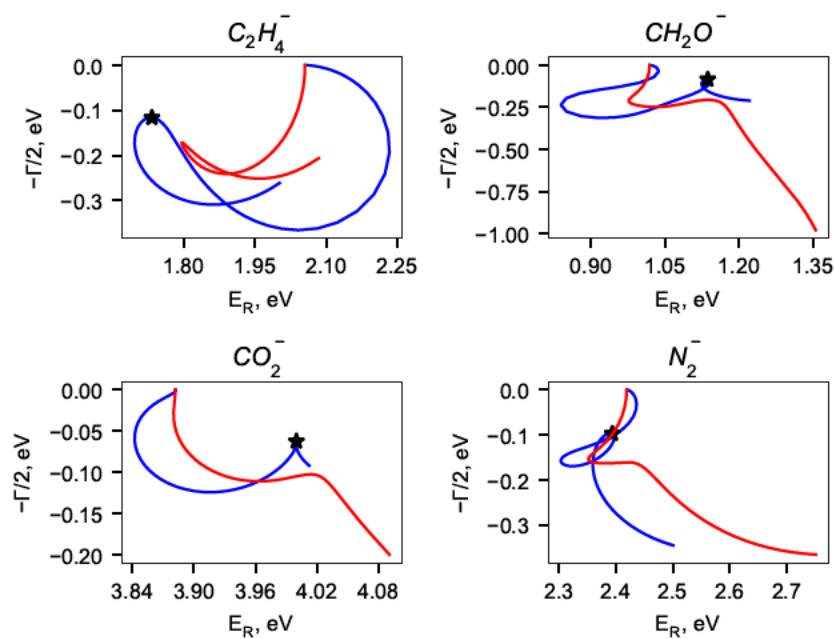
\* Uncorrected result; no stationary point could be identified on first-order trajectory.

## A.5 Representative $\eta$ -trajectories

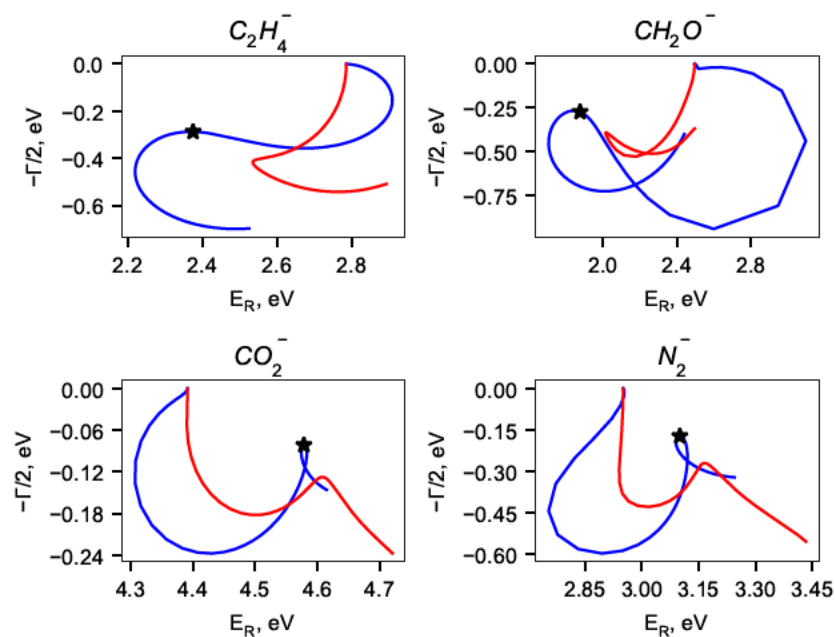
$\eta$ -trajectories for the various CAP calculations are presented below. Uncorrected trajectories are shown in red, corrected trajectories in blue. The point associated with  $\eta_{opt}$  on the corrected trajectory is indicated by a  $\star$ .



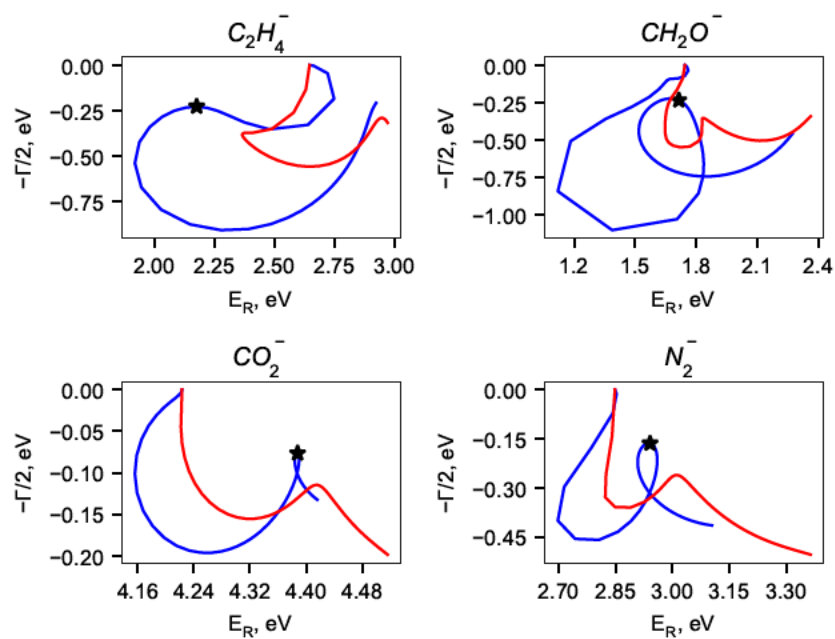
**Figure A-1:** CAP trajectories for  $CH_2O^-$  computed using EOM-EA-CCSD and the ADC(N) family of methods.



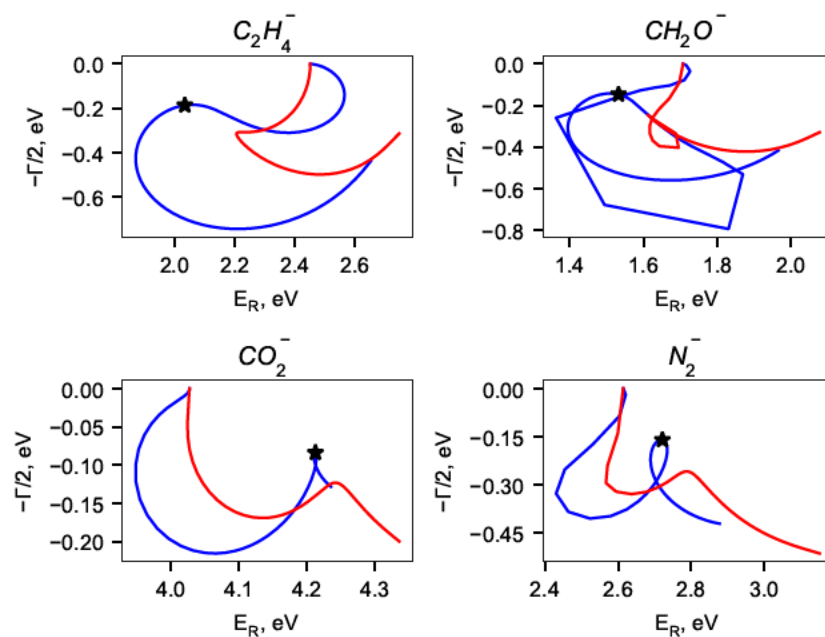
**Figure A-2:** CAP-XMS-CASPT2 trajectories for the  $\pi^*$  resonances for the active spaces listed in Table 4 of the main text.



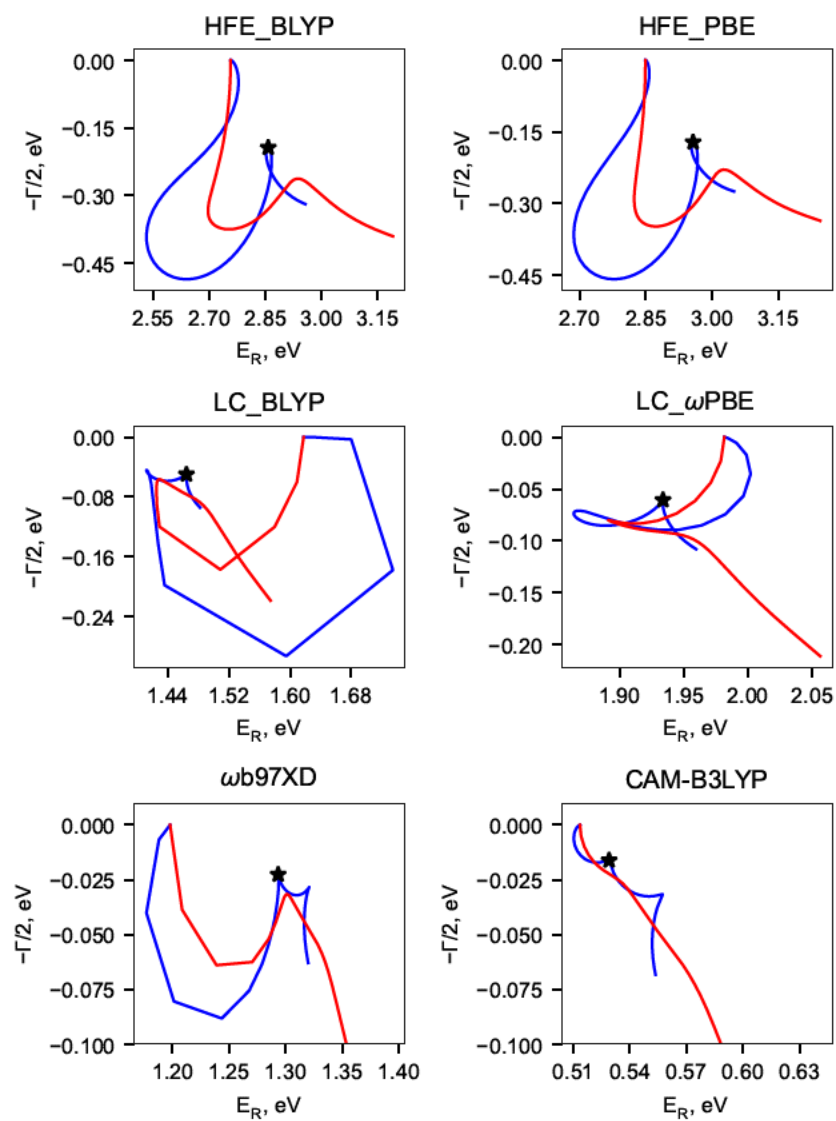
**Figure A-3:** CAP-MRCIS trajectories for the  $\pi^*$  resonances for the active spaces listed in Table 6 of the main text.



**Figure A-4:** CAP-MRCISD trajectories for the  $\pi^*$  resonances for the active spaces listed in Table 6 of the main text.

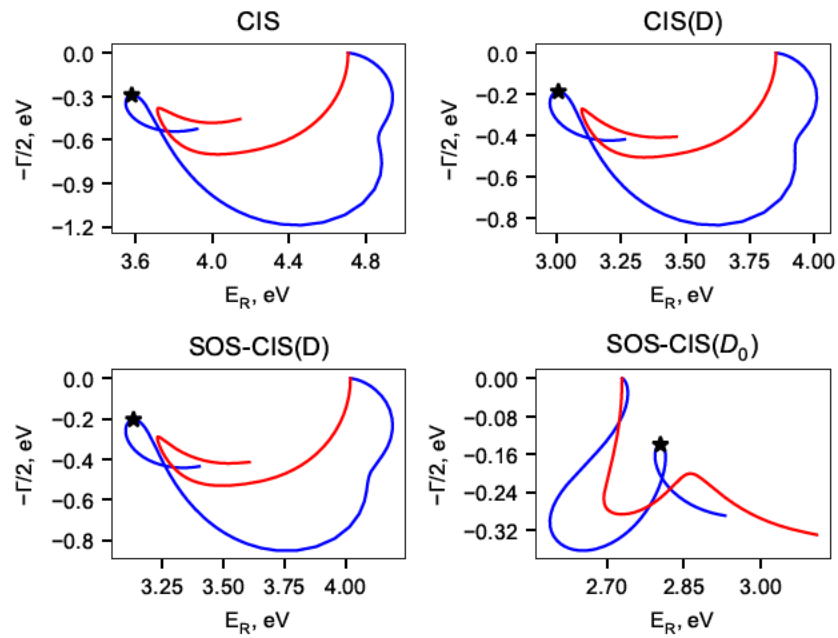


**Figure A-5:** CAP-MRCISD(Q) trajectories for the  $\pi^*$  resonances for the active spaces listed in Table 6 of the main text.

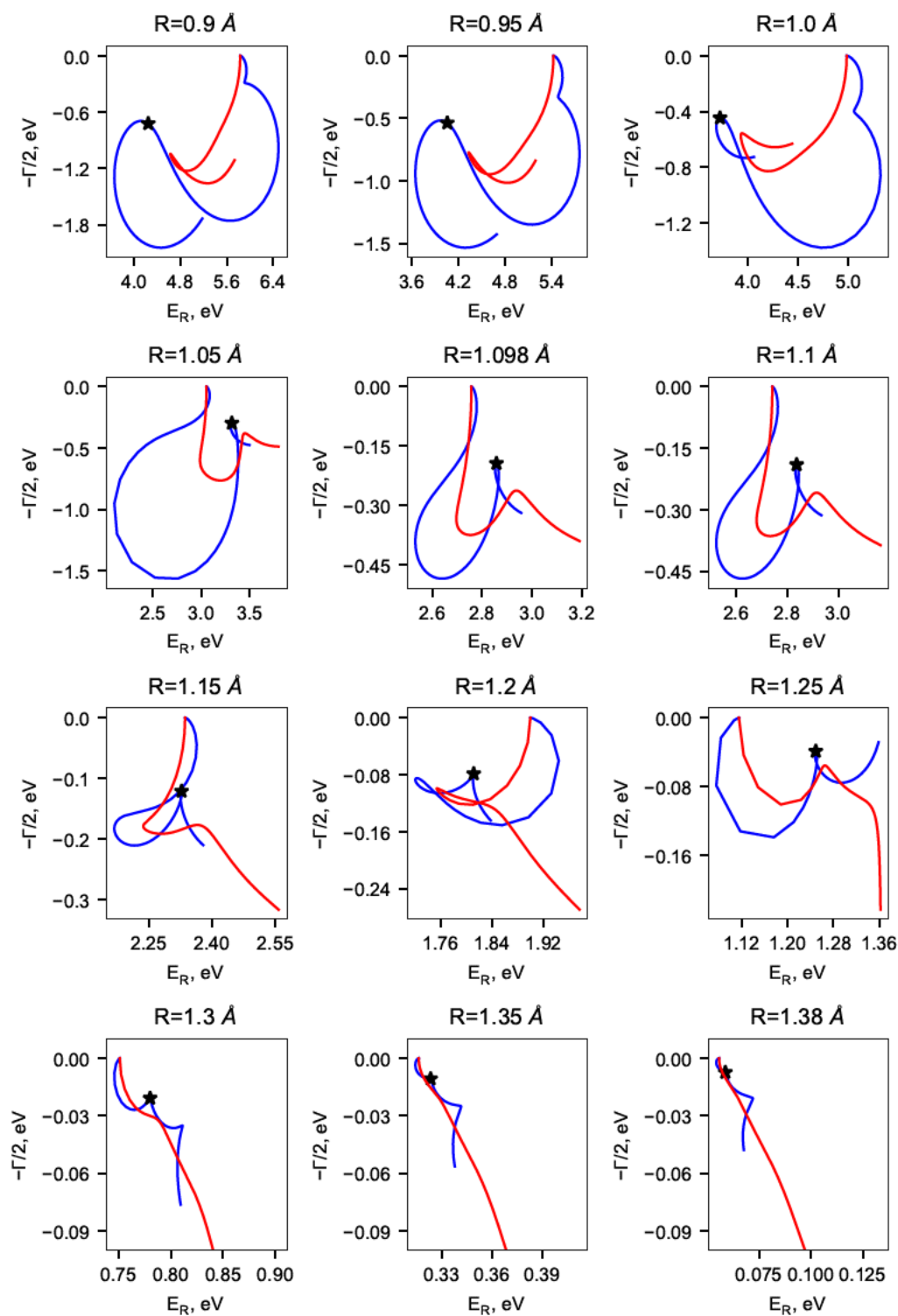


**Figure A-6:** CAP-TDDFT trajectories for  $N_2^-$  computed using various functionals.

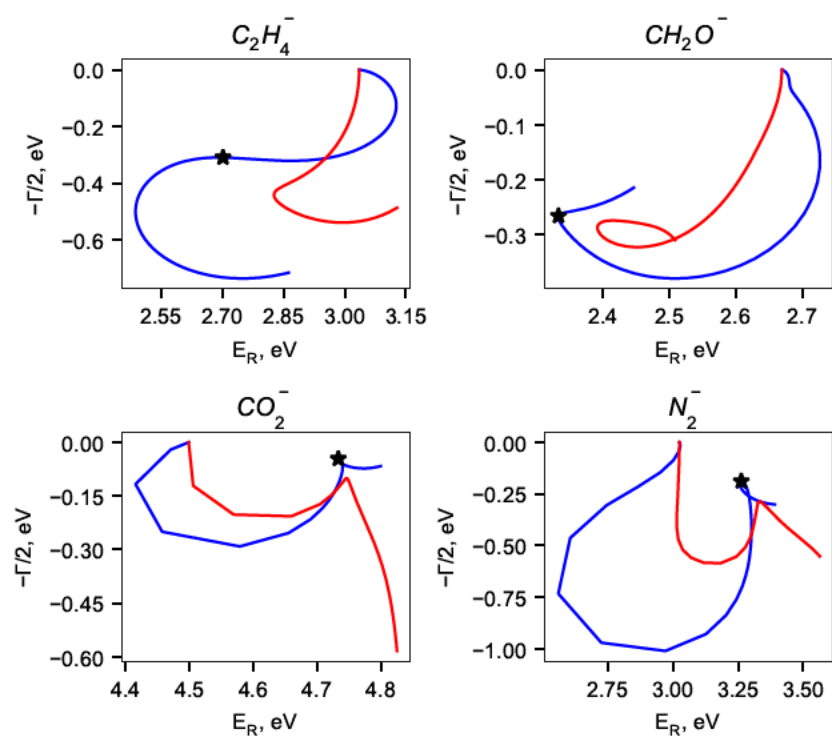




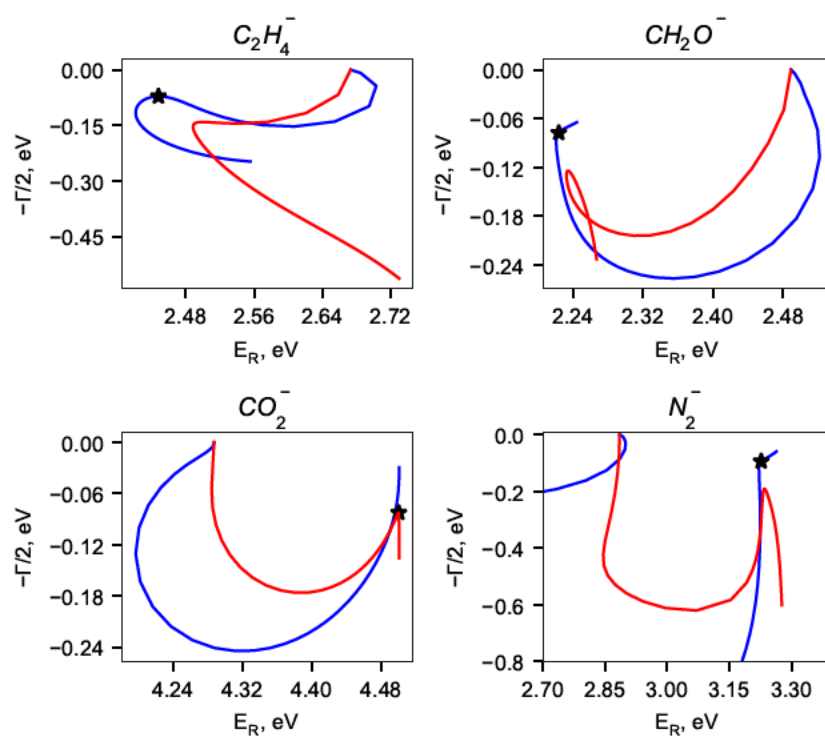
**Figure A.7:** CAP trajectories for  $N_2^-$  computed using the CIS(D) family of methods.



**Figure A-8:** Trajectories for the  $N_2^-$  potential energy curve computed using CAP-TDDFT (HFE\_BLYP).



**Figure A.9:** CAP-MRCIS trajectories for the  $\pi^*$  resonances for the active spaces listed in Table A.4.



**Figure A.10:** CAP-MRCISD trajectories for the  $\pi^*$  resonances for the active spaces listed in Table A.4.

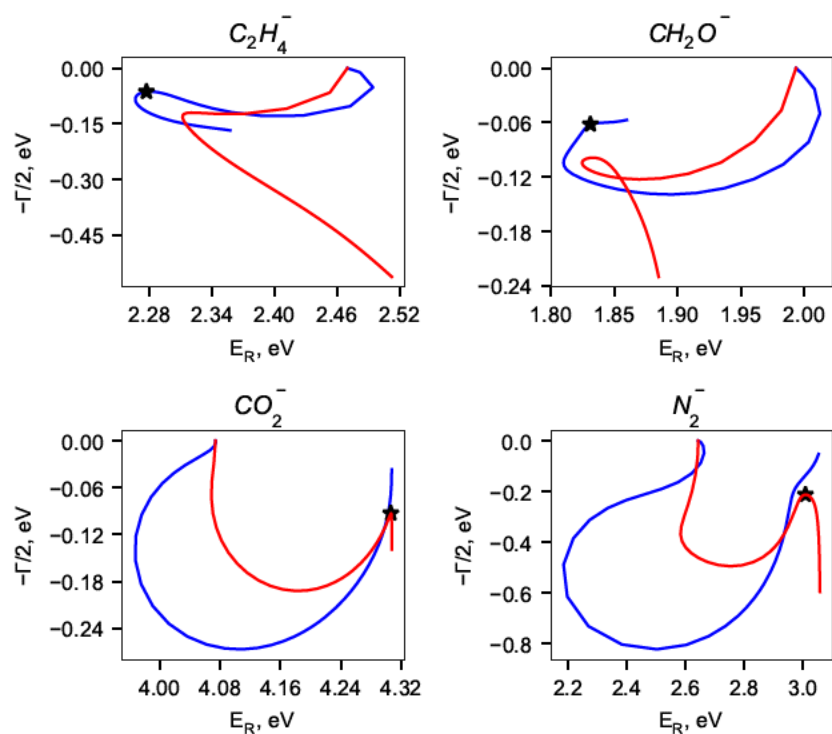


Figure A.11: CAP-MRCISD(Q) trajectories for the  $\pi^*$  resonances for the active spaces listed in Table A.4.

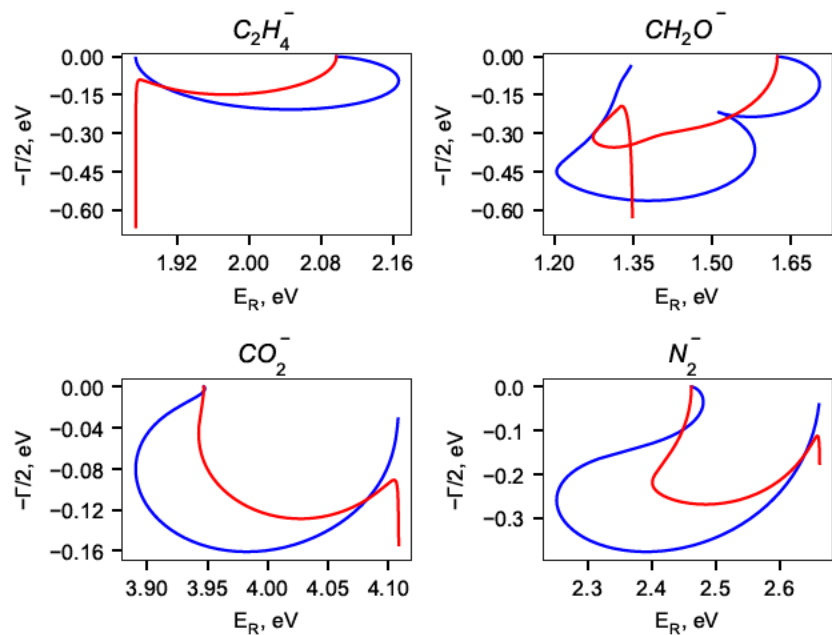


Figure A.12: CAP-XMS-CASPT2 trajectories for the  $\pi^*$  resonances for the active spaces listed in Table A.4.

## Appendix B

### Algorithms for Chapter 6

#### B.0.1 Percent-based edge pruning algorithm

This algorithm considers only the smallest *Percent Edges* % of edges by weight per node, and then prunes based on the mean and standard deviation of the weights of the remaining edges.

---

#### Algorithm 1 Prune by Percent

---

```

procedure PRUNE( $G(V,E)$ , percent_edges, num_st_dev_edges, distance_cutoff)
  for  $v$  in  $V$  do
    for  $e$  in  $v$ .edges do
      if  $e$ .distance  $\geq$  distance_cutoff or  $e$ .weight  $\geq$  per-
centileweight(percent_edges) then
         $G$ .remove( $e$ )
      end if
    end for
     $\bar{l}$  = mean_weight( $v$ .edges)
     $\sigma$  = st_dev_weight( $v$ .edges)
    for  $e$  in  $v$ .edges do
      if  $e$ .weight  $\geq$   $\bar{l}$  + num_st_dev_edges  $\cdot$   $\sigma$  then
         $G$ .remove( $e$ )
      end if
    end for
  end for
end procedure

```

---

#### B.0.2 Degree-based edge pruning algorithm

This algorithm greedily prunes the largest edges by weight of the graph until each node has at most *Max Degree* neighbors.

---

**Algorithm 2** Prune by Degree

---

```

procedure PRUNE(G(V,E), max_degree, distance_cutoff)
  removal_candidates = []
  for e in E do
    if e['distance'] > distance_cutoff then
      G.remove(e)
    end if
  end for
  for v in V do
    if degree(v) > D then
      removal_candidates.append(v.edges)
    end if
  end for
  sort_by_weight_descending(removal_candidates)
  for e(u,v) in removal_candidates do
    if degree(u) > max_degree or degree(v) > max_degree then
      G.remove(e)
    end if
  end for
end procedure

```

---

**B.0.3 Identification of non-protein ET active moieties**

After initial parsing, non-protein residues are analyzed for detection of ET active moieties. For each non-standard residue, a chemical graph is constructed using the NetworkX library, consisting of the O, C, N, P and S atoms in the residue [160]. To isolate the conjugated systems, an edge is only drawn between two atoms  $j$  and  $k$  if:

$$r_{jk} \leq \bar{x} - 2\sigma_{\bar{x}}$$

where  $\bar{x}$  is the mean single-bond distance between those two elements, and  $\sigma_{\bar{x}}$  is the standard deviation. The data was obtained from the online CRC Handbook of Chemistry and Physics [336].

If there are any conjugated systems, the resulting chemical graph will contain one or more connected components. Each connected component that contains a cycle, or

consists of 10 or more atoms will be considered a non-protein ET active moiety, and can be selected for the analysis.



## Appendix C

# Supporting Information for Chapter 7

### C.1 Default Mining Parameters

**Table C.1:** Relevant default parameters for frequent subgraph mining in eMap version 2.0.2.

Parameter	Value
Standard AA	Tyr, Trp
Edge pruning	Max Degree <a href="#">302</a> = 4
Edge thresholds	12 Å
Distance criteria	Center of mass

### C.2 Protein Subgraph Clustering

#### C.2.1 Structural Similarity

The structural similarity between two protein subgraphs in eMap is computed by superimposing the two sets of atoms and computing the root mean squared distance (RMSD). We make the following approximations to the true RMSD:

- Only the alpha carbon (CA) is considered for standard amino acid residues. If one or both are missing the alpha carbon, the residues will not be connected in the supergraph.
- For non-standard amino acids, we use the first atom type both residues have in common. If no shared atom type is found, the residues will not be connected in the supergraph.

The threshold used for determining whether two subgraphs are connected in the supergraph is 0.5 Å.

### C.2.2 Sequence Similarity

Sequence similarity in eMap relies on a multiple sequence alignment, which is performed by the [MUSCLE](#) package[311]. Starting from a one-to-one mapping between the residues, the sequence similarity between two protein subgraphs is defined as the sum of the differences in the residue numbers with respect to the multiple sequence alignment. Non-protein ET active moieties are not considered in sequence similarity, only standard amino acids. The threshold used for determining whether two subgraphs are connected in the supergraph is  $N$ , where  $N$  is the number of nodes, which allows for slight misalignments.

## References

- [1] Shankar, R. *Principles of Quantum Mechanics*; Springer, 1994.
- [2] Sakurai, J. *Modern Quantum Mechanics*; Addison-Wesley Publishing Company, 1995.
- [3] Taylor, J. *Scattering Theory: The Quantum Theory of Nonrelativistic Collisions*; Dover Books on Engineering; Dover Publications, 2012.
- [4] Reinhardt, W. P. Complex coordinates in the theory of atomic and molecular structure and dynamics. *Annual Review of Physical Chemistry* **1982**, *33*, 223–255.
- [5] Gamow, G. The quantum theory of the atom nucleus. *Zeitschrift für Physik* **1928**, *51*, 204.
- [6] Siegert, A. J. F. On the derivation of dispersion formula for nuclear reactions. *Physical Review* **1939**, *56*, 750.
- [7] Feshbach, H. A UNIFIED THEORY OF NUCLEAR REACTIONS. II. *Annals of Physics* **1962**, *19*, 287–313.
- [8] Jagau, T.; Bravaya, K.; Krylov, A. Extending Quantum Chemistry of Bound States to Electronic Resonances. *Annual Review of Physical Chemistry* **2017**, *68*, 525–553.
- [9] Schulz, G. J. Resonances in electron impact on diatomic molecules. *Reviews of Modern Physics* **1973**, *45*, 423.
- [10] Jordan, K. D.; Burrow, P. D. Studies of the temporary anion states of unsaturated hydrocarbons by electron transmission spectroscopy. *Accounts of Chemical Research* **1978**, *11*, 341–348.
- [11] Burrow, P. D.; Michejda, J. A. Electron transmission study of the formaldehyde electron affinity. *Chemical Physics Letters* **1976**, *42*, 223–226.
- [12] Sanche, L.; Schulz, G. J. Electron transmission spectroscopy: Resonances in triatomic molecules and hydrocarbons. *The Journal of Chemical Physics* **1973**, *58*, 479.

- [13] Burrow, P. D.; Michejda, J. A.; Jordan, K. D. Electron transmission study of the temporary negative ion states of selected benzenoid and conjugated aromatic hydrocarbons. *The Journal of Chemical Physics* **1987**, *86*, 9–24.
- [14] Sanov, A.; Mabbs, R. Photoelectron imaging of negative ions. *International Reviews in Physical Chemistry* **2008**, *27*, 53–85.
- [15] Moiseyev, N. *Non-Hermitian quantum mechanics*; Cambridge University Press, 2011.
- [16] Simons, J. Molecular anions. *The Journal of Physical Chemistry A* **2008**, *112*, 6401–6511.
- [17] Boudaïffa, B.; Cloutier, P.; Hunting, D.; Huels, M. A.; Sanche, L. Resonant Formation of DNA Strand Breaks by Low-Energy (3 to 20 eV) Electrons. *Science* **2000**, *287*, 1658–1660.
- [18] Simons, J. How do low-energy (0.1-2 eV) electrons cause DNA-strand breaks? *Accounts of Chemical Research* **2006**, *39*, 772–779.
- [19] Mukherjee, S.; Libisch, F.; Large, N.; Neumann, O.; Brown, L.; Lassiter, J. C. J.; Carter, E.; Nordlander, P.; Halas, N. Hot Electrons Do the Impossible: Plasmon-Induced Dissociation of H<sub>2</sub> on Au. *Nano Letters* **2013**, *13*, 240–247.
- [20] Thomas, R. When electrons meet molecular ions and what happens next: dissociative recombination from interstellar molecular clouds to internal combustion engines. *Mass Spectrometry Reviews* **2008**, *27*, 485–530.
- [21] Millar, T. J.; Walsh, C.; Field, T. A. Negative Ions in Space. *Chemical Reviews* **2017**, *117*, 1765–1795.
- [22] Strelkov, V. Role of autoionizing state in resonant high-order harmonic generation and attosecond pulse production. *Physical Review Letters* **2010**, *104*, 123901.
- [23] Corkum, P.; Krausz, F. Attosecond science. *Nature Physics* **2007**, *3*, 381–387.
- [24] Bressler, C.; Chergui, M. Ultrafast X-ray absorption spectroscopy. *Chemical Reviews* **2004**, *104*, 1781–1812.
- [25] Åberg, T. Theory of the Radiative Auger Effect. *Physical Review A* **1971**, *4*, 1735–1740.
- [26] *Introduction to Marcus Theory of Electron Transfer Reactions*; WORLD SCIENTIFIC, 2019; pp 1–39.
- [27] Gray, H. B.; Winkler, J. R. ELECTRON TRANSFER IN PROTEINS. *Annual Review of Biochemistry* **1996**, *65*, 537–561.

- [28] Marcus, R. On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. I. *The Journal of Chemical Physics* **1956**, *24*, 966.
- [29] Marcus, R. Electrostatic Free Energy and Other Properties of States Having Nonequilibrium Polarization. I. *The Journal of Chemical Physics* **1956**, *24*, 979.
- [30] Marcus, R. Theory of Oxidation-Reduction Reactions Involving Electron Transfer. 4. A Statistical-Mechanical Basis for Treating Contributions from Solvent, Ligands, and Inert Salt. *Discussions of the Faraday Society* **1960**, *29*, 21.
- [31] Marcus, R. On the Theory of Electron-Transfer Reactions. VI. Unified Treatment of Homogeneous and Electrode Reactions. *The Journal of Chemical Physics* **1965**, *43*, 679.
- [32] Hush, N. S. Adiabatic theory of outer sphere electron-transfer reactions in solution. *Transactions of the Faraday Society* **1961**, *57*, 557–580.
- [33] Marcus, R.; Sutin, N. Electron Transfers in Chemistry and Biology. *Biochimica et Biophysica Acta* **1985**, *811*, 265–322.
- [34] Berman, H. M. The Protein Data Bank: a historical perspective. *Acta Crystallographica Section A* **2008**, *64*, 88–95.
- [35] Gray, H. B.; Winkler, J. R. Functional and protective hole hopping in metalloenzymes. *Chemical Science* **2021**, *12*, 13988–14003.
- [36] Winkler, J. R.; Gray, H. B. Long-Range Electron Tunneling. *Journal of the American Chemical Society* **2014**, *136*, 2930–2939.
- [37] Winkler, J. R.; Gray, H. B. Electron Flow through Metalloproteins. *Chemical Reviews* **2014**, *114*, 3369–3380.
- [38] Lukacs, A.; Eker, A.; Byrdin, M.; Brettel, K.; Vos, M. H. Electron Hopping through the 15 Å Triple Tryptophan Molecular Wire in DNA Photolyase Occurs within 30 ps. *Journal of the American Chemical Society* **2008**, *130*, 14394–14395.
- [39] Kao, Y.-T.; Saxena, C.; Wang, L.; Sancar, A.; Zhong, D. Direct observation of thymine dimer repair in DNA by photolyase. *Proceedings of the National Academy of Sciences* **2005**, *102*, 16128–16132.
- [40] Gegear, R.; Casselman, A.; Waddel, S.; Reppert, S. Cryptochrome mediates light-dependent magnetosensitivity in *Drosophila*. *Nature* **2008**, *454*, 1014–1018.

- [41] Maeda, K.; Henbest, K.; Cintolesi, F.; Kuprov, I.; Rodgers, C.; Liddell, P.; Gust, D.; Timmel, C.; Hore, P. Chemical compass model of avian magnetoreception. *Nature* **2008**, *453*, 387–390.
- [42] Gray, H.; Winkler, J. Hole hopping through tyrosine/tryptophan chains protects proteins from oxidative damage. *Proceedings of the National Academy of Sciences* **2015**, *112*, 10920–10925.
- [43] Winkler, J. R.; Gray, H. B. Electron flow through biological molecules: Does hole hopping protect proteins from oxidative damage? *Quarterly Reviews of Biophysics* **2015**, *48*, 411–420.
- [44] Migliore, A.; Polizzi, N.; Therien, M.; Beratan, D. Biochemistry and Theory of Proton-Coupled Electron Transfer. *Chemical Reviews* **2014**, *114*, 3381–3365.
- [45] Gray, H.; Winkler, J. Electron tunneling through proteins. *Quarterly Reviews of Biophysics* **2003**, *36*, 341.
- [46] Bondanza, M.; Nottoli, M.; Cupellini, L.; Lipparini, F.; Mennucci, B. Polarizable embedding QM/MM: the future gold standard for complex (bio)systems? *Physical Chemistry Chemical Physics* **2020**, *22*, 14433–14448.
- [47] Senn, H.; Thiel, W. QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition* **2009**, *48*, 1198–1229.
- [48] Born, M.; Oppenheimer, R. On the quantum theory of molecules. *Annals of Physics* **1927**, *84*, 0457–0484.
- [49] Szabo, A.; Ostlund, N. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; McGraw-Hill: New York, 1989.
- [50] Jensen, F. *Introduction to Computational Chemistry* John Wiley & Sons, 2017.
- [51] Helgaker, T.; Coriani, S.; Jørgensen, P.; Kristensen, K.; Olsen, J.; Ruud, K. Recent Advances in Wave Function-Based Methods of Molecular-Property Calculations. *Chemical Reviews* **2012**, *112*, 543–631.
- [52] Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular electronic structure theory*; John Wiley & Sons, 2000.
- [53] Cizek, J. On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods. *The Journal of Chemical Physics* **1966**, *45*, 4256–4266.
- [54] Purvis, G.; Bartlett, R. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of Chemical Physics* **1982**, *76*, 1910–1918.

- [55] Stanton, J.; Bartlett, R. 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* **1993**, *98*, 7029–7039.
- [56] Krylov, A. Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: The hitchhiker’s guide to Fock space. *Annual Review of Physical Chemistry* **2008**, *59*, 433–462.
- [57] Dreuw, A.; Wormit, M. The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2015**, *5*, 82–95.
- [58] Schirmer, J.; Trofimov, A. B. Intermediate state representation approach to physical properties of electronically excited molecules. *The Journal of Chemical Physics* **2004**, *120*, 11449–11464.
- [59] Bene, J. E. D.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. X. Molecular Orbital Studies of Excited States with Minimal and Extended Basis Sets. *The Journal of Chemical Physics* **1971**, *55*, 2236–2241.
- [60] Foresman, J.; Head-Gordon, M.; Pople, J.; Frisch, M. Toward a systematic molecular orbital theory for excited states. *The Journal of Physical Chemistry* **1992**, *96*, 135–149.
- [61] Head-Gordon, M.; Rico, R.; Oumi, M.; Lee, T. A doubles correction to electronic excited states from configuration interaction in the space of single substitutions. *Chemical Physics Letters* **1994**, *219*, 21–29.
- [62] Rhee, Y. M.; Head-Gordon, M. Scaled Second-Order Perturbation Corrections to Configuration Interaction Singles: Efficient and Reliable Excitation Energy Methods. *The Journal of Physical Chemistry A* **2007**, *111*, 5314–5326.
- [63] Casanova, D.; Rhee, Y. M.; Head-Gordon, M. Quasidegenerate scaled opposite spin second order perturbation corrections to single excitation configuration interaction. *The Journal of Chemical Physics* **2008**, *128*, 164106.
- [64] Roos, B. *Ab initio methods in quantum chemistry, II*; John Wiley & Sons, 1987; pp 399–446.
- [65] Finley, J.; Malmqvist, P.-A.; B.O.Roos,; Serrano-Andrés, L. The multi-state CASPT2 method. *Chemical Physics Letters* **1998**, *288*, 299–306.
- [66] Park, J. W.; Al-Saadon, R.; MacLeod, M. K.; Shiozaki, T.; Vlaisavljevich, B. Multireference Electron Correlation Methods: Journeys along Potential Energy Surfaces. *Chemical Reviews* **2020**, *120*, 5878–5909.

- [67] Andersson, K.; Malmqvist, P.-Å.; Roos, B.; Sadlej, A.; Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *The Journal of Physical Chemistry* **1990**, *94*, 5483–5488.
- [68] Andersson, K.; Malmqvist, P.-Å.; Roos, B. Second-order perturbation theory with a complete active space self-consistent field reference function. *The Journal of Chemical Physics* **1992**, *96*, 1218.
- [69] Forsberg, N.; Malmqvist, P.-Å. Multiconfiguration perturbation theory with imaginary level shift. *Chemical Physics Letters* **1997**, *274*, 196–204.
- [70] Multiconfigurational perturbation theory with level shift —the Cr2 potential revisited. *Chemical Physics Letters* **1995**, *245*, 215–223.
- [71] Shiozaki, T.; Győrffy, W.; Celani, P.; Werner, H.-J. Communication: Extended multi-state complete active space second-order perturbation theory: Energy and nuclear gradients. *The Journal of Chemical Physics* **2011**, *135*, 081106.
- [72] Battaglia, S.; Lindh, R. Extended Dynamically Weighted CASPT2: The Best of Two Worlds. *Journal of Chemical Theory and Computation* **2020**, *16*, 1555–1567.
- [73] Granovsky, A. Extended multi-configuration quasi-degenerate perturbation theory: The new approach to multi-state multi-reference perturbation theory. *The Journal of Chemical Physics* **2011**, *134*, 214113.
- [74] Battaglia, S.; Lindh, R. On the role of symmetry in XDW-CASPT2. *The Journal of Chemical Physics* **2021**, *154*, 034102.
- [75] Werner, H.; Knowles, P. An efficient internally contracted multiconfiguration-reference configuration interaction method. *The Journal of Chemical Physics* **1988**, *89*, 5803–5814.
- [76] Lischka, H. et al. The generality of the GUGA MRCI approach in COLUMBUS for treating complex quantum chemistry. *The Journal of Chemical Physics* **2020**, *152*, 134110.
- [77] Langhoff, S.; Davidson, E. Configuration interaction calculations on the nitrogen molecule. *International Journal of Quantum Chemistry* **1974**, *8*, 61–72.
- [78] Ziegler, T. Approximate density functional theory as a practical tool in molecular energetics and dynamics. *Chemical Reviews* **1991**, *91*, 651.
- [79] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **1964**, *136*, B864–B871.



- [80] Kohn, W.; Sham, L. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140*, A1133–A1138.
- [81] Petersilka, M.; Grossman, U.; Gross, E. Excitation Energies from Time-Dependent Density-Functional Theory. *Physical Review Letters* **1996**, *76*, 1212–1215.
- [82] Bauernschmitt, R.; Ahlrichs, R. Treatment of electronic excitations within the adiabatic approximation of time dependent functional theory. *Chemical Physics Letters* **1996**, *256*, 454–464.
- [83] Dreuw, A.; Head-Gordon, M. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chemical Reviews* **2005**, *105*, 4009 – 4037.
- [84] Runge, E.; Gross, E. Density-Functional Theory for Time-Dependent Systems. *Physical Review Letters* **1984**, *52*, 997–1000.
- [85] Mandelstam, L.; Tamm, I.G. The Uncertainty Relation Between Energy and Time in Non-relativistic Quantum Mechanics. *USSR Journal of Physics* **1945**, *9*, 449.
- [86] Dancoff, S. Non-adiabatic meson theory of nuclear forces. *Physical Review* **1950**, *78*, 382–385.
- [87] Hirata, S.; Head-Gordon, M. Time-dependent density functional theory within the Tamm-Dancoff approximation. *Chem. Phys. Lett.* **1999**, *314*, 291–299.
- [88] Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. *The Journal of Chemical Physics* **2005**, *123*, 062201.
- [89] Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics* **1999**, *110*, 6158–6170.
- [90] Perdew, J.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865–3868.
- [91] Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical Physics Letters* **2004**, *393*, 51–57.
- [92] Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *The Journal of Chemical Physics* **2001**, *115*, 3540.
- [93] Tennyson, J. Electron-molecule collision calculations using the R-matrix method. *Physics Reports* **2010**, *491*, 29–76.

- [94] da Costa, R. F.; Varella, M. T. d. N.; Bettega, M. H. F.; Lima, M. A. P. Recent advances in the application of the Schwinger multichannel method with pseudopotentials to electron-molecule collisions. *The European Physical Journal D* **2015**, *69*, 159.
- [95] Hazi, A.; Taylor, H. Stabilization Method of Calculating Resonance Energies: Model Problem. *Physical Review A* **1970**, *1*, 1109–1120.
- [96] Bhattacharya, D.; Ben-Asher, A.; Haritan, I.; Pawlak, M.; Landau, A.; Moiseyev, N. Polyatomic ab Initio Complex Potential Energy Surfaces: Illustration of Ultracold Collisions. *The Journal of Chemical Theory and Computation* **2017**, *13*, 1682–1690.
- [97] Carlson, B. J.; Falcetta, M. F.; Slimak, S. R.; Jordan, K. D. A Fresh Look at the Role of the Coupling of a Discrete State with a Pseudocontinuum State in the Stabilization Method for Characterizing Metastable States. *The Journal of Physical Chemistry Letters* **2021**, *12*, 1202–1206.
- [98] Hazi, A. U. A purely  $L^2$  method for calculating resonance widths. *Journal of Physics B* **1978**, *11*, L259–L264.
- [99] Kopelke, S.; Gokhberg, K.; Cederbaum, L. S.; Tarantelli, F.; Averbukh, V. Autoionization widths by Stieltjes imaging applied to Lanczos pseudospectra. *The Journal of Chemical Physics* **2011**, *134*, 024106.
- [100] Aguilar, J.; Combes, J. M. A class of analytic perturbations for one-body Schrödinger Hamiltonians. *Communications in Mathematical Physics* **1971**, *22*, 269–279.
- [101] McCurdy, C.; Rescigno, T. Extension of the Method of Complex Basis Functions to Molecular Resonances. *Physical Review Letters* **1978**, *41*, 1364–1368.
- [102] White, A. F.; Epifanovsky, E.; McCurdy, C. W.; Head-Gordon, M. Second order Möller-Plesset and coupled cluster singles and doubles methods with complex basis functions for resonances in electron-molecule scattering. *The Journal of Chemical Physics* **2017**, *146*, 234107.
- [103] Simon, B. The definition of molecular resonance curves by the method of exterior complex scaling. *Physics Letters* **1979**, *71A*.
- [104] Riss, U. V.; Meyer, H.-D. Calculation of resonance energies and widths using the complex absorbing potential method. *Journal of Physics B* **1993**, *26*, 4503–4536.
- [105] Santra, R. Why complex absorbing potentials work: A discrete-variable-representation perspective. *Physical Review A* **2006**, *74*, 034701–.

- [106] Moiseyev, N. Derivations of universal exact complex absorption potentials by the generalized complex coordinate method. *Journal of Physics B* **1998**, *31*, 1431.
- [107] Moiseyev, N.; Certain, P.; Weinhold, F. Resonance properties of complex-rotated hamiltonians. *Molecular Physics* **1978**, *36*, 1613–1630.
- [108] Heiss, W. D. The physics of exceptional points. *Journal of Physics A* **2012**, *45*, 444016.
- [109] Berman, M.; Walter, O.; Cederbaum, L. S. Electron-Molecule Scattering in the Optical-Potential Approach: Surpassing Second Order. *Physical Review Letters* **1983**, *50*, 1979–1982.
- [110] Kosloff, R.; Kosloff, D. Absorbing boundaries for wave propagation problems. *Journal of Computational Physics* **1986**, *63*, 363–376.
- [111] Jolicard, G.; Austin, E. J. Optical potential stabilisation method for predicting resonance levels. *Chemical Physics Letters* **1985**, *121*, 106–110.
- [112] Landau, A.; Moiseyev, N. Molecular resonances by removing complex absorbing potentials via Padé; Application to CO- and N<sub>2</sub>-. *The Journal of Chemical Physics* **2016**, *145*, 164111.
- [113] Sajeev, Y.; Vysotskiy, V.; Cederbaum, L.; Moiseyev, N. Continuum remover-complex absorbing potential: Efficient removal of the nonphysical stabilization points. *The Journal of Chemical Physics* **2009**, *131*, 211102.
- [114] Gayvert, J. R.; Bravaya, K. B. Application of Box and Voronoi CAPs for Metastable Electronic States in Molecular Clusters. *The Journal of Physical Chemistry A* **2022**, *126*, 5070–5078.
- [115] Jolicard, G.; Leforestier, C.; Austin, E. J. Resonance states using the optical potential model. Study of Feshbach resonances and broad shape resonances. *The Journal of Chemical Physics* **1988**, *88*, 1026–1031.
- [116] Jagau, T.-C.; Zuev, D.; Bravaya, K.; Epifanovsky, E.; Krylov, A. A fresh look at resonances and complex absorbing potentials: Density matrix based approach. *The Journal of Physical Chemistry Letters* **2014**, *5*, 310–315.
- [117] Santra, R.; Cederbaum, L. Non-Hermitian electronic theory and applications to clusters. *Physics Reports* **2002**, *368*, 1–117.
- [118] Jagau, T.-C.; Krylov, A. I. Complex Absorbing Potential Equation-of-Motion Coupled-Cluster Method Yields Smooth and Internally Consistent Potential Energy Surfaces and Lifetimes for Molecular Resonances. *The Journal of Physical Chemistry Letters* **2014**, *5*, 3078–3085.

- [119] Zuev, D.; Jagau, T.-C.; Bravaya, K. B.; Epifanovsky, E.; Shao, Y.; Sundstrom, E.; Head-Gordon, M.; Krylov, A. Complex absorbing potentials within EOM-CC family of methods: Theory, Implementation, and Benchmarks. *The Journal of Chemical Physics* **2014**, *141*, 024102.
- [120] Gayvert, J. R.; Bravaya, K. B. Projected CAP-EOM-CCSD method for electronic resonances. *The Journal of Chemical Physics* **2022**, *156*, 094108.
- [121] Sommerfeld, T.; Riss, U.V.; Meyer, H.-D.; Cederbaum, L.S.; Engels, B.; Suter, H.U. Temporary anions - calculation of energy and lifetime by absorbing potentials: the  $N_2^-$   $^2\Pi_g$  resonance. *Journal of Physics B* **1998**, *31*, 4107–4122.
- [122] Sommerfeld, T.; Santra, R. Efficient method to perform CAP/CI calculations for temporary anions. *International Journal of Quantum Chemistry* **2001**, *82*, 218–226.
- [123] Santra, R.; Cederbaum, L. An efficient combination of computational techniques for investigating electronic resonance states in molecules. *The Journal of Chemical Physics* **2001**, *115*, 6853.
- [124] Thodika, M.; Matsika, S. Projected Complex Absorbing Potential Multireference Configuration Interaction Approach for Shape and Feshbach Resonances. *Journal of Chemical Theory and Computation* **2022**, *18*, 3377–3390.
- [125] Ehara, M.; Sommerfeld, T. CAP/SAC-CI method for calculating resonance states of metastable anions. *Chemical Physics Letters* **2012**, *537*, 107–112.
- [126] Basumallick, S.; Sajeev, Y.; Pal, S.; Vaval, N. Negative Ion Resonance States: The Fock-Space Coupled-Cluster Way. *The Journal of Physical Chemistry A* **2020**, *124*, 10407–10421.
- [127] Zhou, Y.; Ernzerhof, M. Calculating the Lifetimes of Metastable States with Complex Density Functional Theory. *The Journal of Physical Chemistry Letters* **2012**, *3*, 1916–1920.
- [128] Santra, R.; Cederbaum, L. Complex absorbing potentials in the framework of electron propagator theory. I. General formalism. *The Journal of Chemical Physics* **2002**, *117*, 5511–5521.
- [129] Dempwolff, A. L.; Belogolova, A. M.; Sommerfeld, T.; Trofimov, A. B.; Dreuw, A. CAP/EA-ADC method for metastable anions: Computational aspects and application to  $\pi^*$  resonances of norbornadiene and 1, 4-cyclohexadiene. *The Journal of Chemical Physics* **2021**, *155*, 054103.

- [130] Phung, Q. M.; Komori, Y.; Yanai, T.; Sommerfeld, T.; Ehara, M. Combination of a Voronoi-Type Complex Absorbing Potential with the XMS-CASPT2 Method and Pilot Applications. *Journal of Chemical Theory and Computation* **2020**, *16*, 2606–2616.
- [131] Al-Saadon, R.; Shiozaki, T.; Knizia, G. Visualizing Complex-Valued Molecular Orbitals. *The Journal of Physical Chemistry A* **2019**, *123*, 3223–3228.
- [132] Kunitsa, A.; Granovsky, A.; Bravaya, K. CAP-XMCQDPT2 method for metastable electronic states. *The Journal of Chemical Physics* **2017**, *146*, 184107.
- [133] Santra, R.; Cederbaum, L.; Meyer, H.-D. Electronic decay of molecular clusters: non-stationary states computed by standard quantum chemistry methods. *Chemical Physics Letters* **1999**, *303*, 413–419.
- [134] Benda, Z.; Jagau, T.-C. Communication: Analytic gradients for the complex absorbing potential equation-of-motion coupled-cluster method. *The Journal of Chemical Physics* **2017**, *146*, 031101.
- [135] Sommerfeld, T.; Ehara, M. Complex Absorbing Potentials with Voronoi Isosurfaces Wrapping Perfectly around Molecules. *Journal of Chemical Theory and Computation* **2015**, *11*, 4627–4633.
- [136] Kanazawa, Y.; Ehara, M.; Sommerfeld, T. Low-Lying  $\pi^*$  Resonances of Standard and Rare DNA and RNA Bases Studied by the Projected CAP/SAC-CI Method. *The Journal of Physical Chemistry A* **2016**, *120*, 1545–1553.
- [137] Ghosh, A.; Vaval, N.; Pal, S. Equation-of-motion coupled-cluster method for the study of shape resonance. *The Journal of Chemical Physics* **2012**, *136*, 234110.
- [138] Ghosh, A.; Karne, A.; Pal, S.; Vaval, N. CAP/EOM-CCSD method for the study of potential curves of resonant states. *Physical Chemistry Chemical Physics* **2013**, *15*, 17915–17921.
- [139] Ghosh, A.; Vaval, N.; Pal, S.; Bartlett, R. J. Complex absorbing potential based equation-of-motion coupled cluster method for the potential energy curve of CO<sub>2</sub>- anion. *The Journal of Chemical Physics* **2014**, *141*, 164113.
- [140] Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. *Introduction to algorithms*; MIT press, 2022.
- [141] Cordella, L.; Foggia, P.; Sansone, C.; Vento, M. A (sub)graph isomorphism algorithm for matching large graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2004**, *26*, 1367–1372.

- [142] Cordella, L. P.; Foggia, P.; Sansone, C.; Vento, M. An improved algorithm for matching large graphs. In: 3rd IAPR-TC15 Workshop on Graph-based Representations in Pattern Recognition, Cuen. 2001; pp 149–159.
- [143] Xifeng Yan.; Jiawei Han, gSpan: graph-based substructure pattern mining. 2002 IEEE International Conference on Data Mining, 2002. Proceedings. 2002; pp 721–724.
- [144] Beratan, D. N.; Onuchic, J. N.; Hopfield, J. J. Electron tunneling through covalent and noncovalent pathways in proteins. *The Journal of Chemical Physics* **1987**, *86*, 4488–4498.
- [145] Beratan, D.; Betts, J.; Onuchic, J. Protein electron transfer rates set by the bridging secondary and tertiary structure. *Science* **1991**, *252*, 1285–1288.
- [146] Onuchic, J. N.; Beratan, D. N.; Winkler, J. R.; Gray, H. B. Pathway analysis of protein electron-transfer reactions. *Annual Review of Biophysics and Biomolecular Structure* **1992**, *21*, 349–377.
- [147] Closs, G. L.; Miller, J. R. Intramolecular Long-Distance Electron Transfer in Organic Molecules. *Science* **1988**, *240*, 440–447.
- [148] Betts, J. N.; Beratan, D. N.; Onuchic, J. N. Mapping electron tunneling pathways: an algorithm that finds the "minimum length"/maximum coupling pathway between electron donors and acceptors in proteins. *Journal of the American Chemical Society* **1992**, *114*, 4043–4046.
- [149] Balabin, I. A.; Hu, X.; Beratan, D. N. Exploring biological electron transfer pathway dynamics with the *Pathways* Plugin for VMD. *Journal of Computational Chemistry* **2012**, *33*, 906–910.
- [150] Jones, M. L.; Kurnikov, I. V.; Beratan, D. N. The Nature of Tunneling Pathway and Average Packing Density Models for Protein-Mediated Electron Transfer. *The Journal of Physical Chemistry A* **2002**, *106*, 2002–2006.
- [151] Tazhigulov, R. N.; Gayvert, J. R.; Wei, M.; Bravaya, K. B. eMap: A Web Application for Identifying and Visualizing Electron or Hole Hopping Pathways in Proteins. *The Journal of Physical Chemistry B* **2019**, *123*, 6946–6951.
- [152] Han, J.; Cheng, H.; Xin, D.; Yan, X. Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery* **2007**, *15*, 55–86.
- [153] Jiang, C.; Coenen, F.; Zito, M. A survey of frequent subgraph mining algorithms. *The Knowledge Engineering Review* **2013**, *28*, 75–105.
- [154] Takigawa, I.; Mamitsuka, H. Graph mining: procedure, application to drug discovery and recent advances. *Drug Discovery Today* **2013**, *18*, 50–57.

- [155] Queiroz, F. C.; Vargas, A. M. P.; Oliveira, M. G. A.; Comarela, G. V.; Silveira, S. A. ppiGReMLIN: a graph mining based detection of conserved structural arrangements in protein-protein interfaces. *BMC Bioinformatics* **2020**, *21*, 143.
- [156] Ribeiro, V. S.; Santana, C. A.; Fassio, A. V.; Cerqueira, F. R.; da Silveira, C. H.; Romanelli, J. P. R.; Patarroyo-Vargas, A.; Oliveira, M. G. A.; Gonçalves-Almeida, V.; Izidoro, S. C.; de Melo-Minardi, R. C.; Silveira, S. d. A. vis-GReMLIN: graph mining-based detection and visualization of conserved motifs at 3D protein-ligand interface at the atomic level. *BMC Bioinformatics* **2020**, *21*, 80.
- [157] Ramu, A.; Kahveci, T.; Burleigh, J. G. A scalable method for identifying frequent subtrees in sets of large phylogenetic trees. *BMC Bioinformatics* **2012**, *13*, 256.
- [158] Zhang, S.; Wang, J. T. Discovering Frequent Agreement Subtrees from Phylogenetic Data. *IEEE Transactions on Knowledge and Data Engineering* **2008**, *20*, 68–82.
- [159] Deepak, A.; Fernández-Baca, D.; Tirthapura, S.; Sanderson, M. J.; McMahon, M. M. EvoMiner: frequent subtree mining in phylogenetic databases. *Knowledge and Information Systems* **2014**, *41*, 559–590.
- [160] Hagberg, A.; Swart, P.; S Chult, D. *Exploring Network Structure, Dynamics, and Function Using NetworkX*; 2008.
- [161] Mukherjee, S.; Libisch, F.; Large, N.; Neumann, O.; Brown, L.; Cheng, J.; Lassiter, J.; Carter, E.; Norlander, P.; Halas, N. Hot electrons do impossible: plasmon-induced dissociation of H<sub>2</sub> on Au. *Nano Lett.* **2013**, *13*, 240.
- [162] Davis, J. U.; Sommerfeld, T. Computing resonance energies directly: method comparison for a model potential. *European Physical Journal D* **2021**, *75*, 316.
- [163] Jagau, T.-C. Theory of electronic resonances: fundamental aspects and recent advances. *Chemical Communications* **2022**, *58*, 5205–5224.
- [164] Sahoo, S.; Ho, Y. K. The complex absorbing potential method (CAP) to study the Stark effect in hydrogen and lithium. *Journal of Physics B* **2000**, *33*, 2195–2206.
- [165] Muga, J.; Palao, J.; Navarro, B.; Egusquiza, I. Complex absorbing potentials. *Physics Reports* **2004**, *395*, 357 – 426.
- [166] Zhang, L.; Chen, J.; Wang, J. First-principles investigation of transient current in molecular devices by using complex absorbing potentials. *Physical Review B* **2013**, *87*, 205401.

- [167] Ghosh, A.; Vaval, N.; Pal, S. Equation-of-motion coupled-cluster method for the study of shape resonances. *The Journal of Chemical Physics* **2012**, *136*, 234110.
- [168] Sajeev, Y.; Sindelka, M.; Moiseyev, N. Reflection-free complex absorbing potential for electronic structure calculations: Feshbach type autoionization of Helium. *Chemical Physics* **2006**, *329*, 307–312.
- [169] Sajeev, Y.; Moiseyev, N. Reflection-free complex absorbing potential for electronic structure calculations: Feshbach-type autoionization resonances of molecules. *The Journal of Chemical Physics* **2007**, *127*, 034105.
- [170] Riss, U.; Meyer, H.-D. The transformative complex absorbing potential method: a bridge between complex absorbing potentials and smooth exterior scaling. *Journal of Physics B* **1998**, *31*, 2279–2304.
- [171] Epifanovsky, E. et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *The Journal of Chemical Physics* **2021**, *155*, 84801.
- [172] Rezaei, M.; Sheybani-Deloui, S.; Moazzen-Ahmadi, N.; Michaelian, K. H.; McKellar, A. R. W. CO Dimer: The Infrared Spectrum Revisited. *The Journal of Physical Chemistry A* **2013**, *117*, 9612–9620.
- [173] Brookes, M. D.; McKellar, A. The mystery of the CO dimer: assignments from variable-temperature jet-cooled infrared spectra. *Chemical Physics Letters* **1998**, *287*, 365 – 370.
- [174] Brookes, M. D.; McKellar, A. R. W. Infrared spectrum and energy levels of the CO dimer: Evidence for two almost isoenergetic isomers. *The Journal of Chemical Physics* **1999**, *111*, 7321–7328.
- [175] Moiseyev, N. Quantum theory of resonances: calculating energies, widths and cross-sections by complex scaling. *Physics Reports* **1998**, *302*, 212–293.
- [176] Dunning, T. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.
- [177] Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *The Journal of Chemical Physics* **1988**, *88*, 2547–2553.
- [178] Lebedev, V. Values of the nodes and weights of ninth to seventeenth order Gauss-Markov quadrature formulae invariant under the octahedron group with inversion. *USSR Computational Mathematics and Mathematical Physics* **1975**, *15*, 48–54.



- [179] Vissers, G. W. M.; Heßelmann, A.; Jansen, G.; Wormer, P. E. S.; van der Avoird, A. New CO-CO interaction potential tested by rovibrational calculations. *The Journal of Chemical Physics* **2005**, *122*, 54306.
- [180] The Dyson orbitals in Fig. 3·9 were computed at the equilibrium geometry using CAP-EOM-EA-CCSD, with the CAP incorporated starting from Hartree-Fock [119]. A Voronoi CAP of 3.0 bohr was used, and the CAP strength was chosen as 0.005. Thus, the orbitals are not necessarily the best representation of the stabilized resonance, but do provide qualitative insight into the character of the states.
- [181] Jagau, T.-C.; Krylov, A. I. Characterizing metastable states beyond energies and lifetimes: Dyson orbitals and transition dipole moments. *The Journal of Chemical Physics* **2016**, *144*, 054113.
- [182] Ehrhardt, H.; Langhans, L.; Linder, F.; Taylor, H. S. Resonance scattering of slow electrons from H<sub>2</sub> and CO angular distributions. *Physical Review* **1968**, *173*, 222–230.
- [183] Simons, J. Theoretical Studies of Negative Molecular Ions. *Annual Review of Physical Chemistry* **1977**, *28*, 15–45.
- [184] Watson, D. K. In *Schwinger Variational Methods*; Bates, D., Bederson, B., Eds.; Advances in Atomic and Molecular Physics; Academic Press, 1989; Vol. 25; pp 221–250.
- [185] Kukulín, V. I.; Krasnopol'sky, V. M. Description of few-body systems via analytical continuation in coupling constant. *Journal of Physics A* **1977**, *10*, L33–L37.
- [186] McGuire, B. A.; Loomis, R. A.; Burkhardt, A. M.; Lee, K. L. K.; Shingledecker, C. N.; Charnley, S. B.; Cooke, I. R.; Cordiner, M. A.; Herbst, E.; Kalenskii, S.; Siebert, M. A.; Willis, E. R.; Xue, C.; Remijan, A. J.; McCarthy, M. C. Detection of two interstellar polycyclic aromatic hydrocarbons via spectral matched filtering. *Science* **2021**, *371*, 1265–1269.
- [187] NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101. <http://cccbdb.nist.gov/>, released 08/21/2020, 2020.
- [188] Kunitsa, A. A.; Bravaya, K. B. First-Principles Calculations of the Energy and Width of the <sup>2</sup>A<sub>u</sub> Shape Resonance in p-Benzoquinone: A Gateway State for Electron Transfer. *The Journal of Physical Chemistry Letters* **2015**, *6*, 1053–1058.

- [189] Benda, Z.; Jagau, T.-C. Locating Exceptional Points on Multidimensional Complex-Valued Potential Energy Surfaces. *The Journal of Physical Chemistry Letters* **2018**, *9*, 6978–6984.
- [190] Ehara, M.; Fukuda, R.; Sommerfeld, T. Projected CAP/SAC-CI method with smooth Voronoi potential for calculating resonance states. *Journal of Computational Chemistry* **2015**, *37*, 242–249.
- [191] Thodika, M.; Fennimore, M.; Karsili, T. N. V.; Matsika, S. Comparative study of methodologies for calculating metastable states of small to medium-sized molecules. *The Journal of Chemical Physics* **2019**, *151*, 244104.
- [192] Lebedev, V. I.; Laikov, D. A quadrature formula for the sphere of the 131st algebraic order of accuracy. *Doklady Mathematics*. 1999; pp 477–481.
- [193] Berman, M.; Estrada, H.; Cederbaum, L. S.; Domcke, W. Nuclear dynamics in resonant electron-molecule scattering beyond the local approximation: The 2.3-eV shape resonance in N<sub>2</sub>. *Physical Review A* **1983**, *28*, 1363–1381.
- [194] The half-width of 0.4 eV in Ref. [182] was estimated from the half-width of the  $\nu'=0$  resonance in the elastic scattering cross-section, and from the half-width of the inelastic scattering cross-section for the  $\nu=1$  vibrational state. We emphasize that these quantities are dependent on the nuclear coordinates, and that the values  $E_R(R_{eq})$  or  $\Gamma(R_{eq})$  can not be directly extracted from the experiment.
- [195] Burrow, P. D.; Howard, A. E.; Johnston, A. R.; Jordan, K. D. Temporary anion states of hydrogen cyanide, methyl cyanide, and methylene dicyanide, selected cyanoethylenes, benzonitrile, and tetracyanoquinodimethane. *The Journal of Physical Chemistry* **1992**, *96*, 7570–7578.
- [196] Allan, M. Selectivity in the Excitation of Fermi-Coupled Vibrations in CO<sub>2</sub> by Impact of Slow Electrons. *Physical Review Letters* **2001**, *87*, 033201.
- [197] Veen, E. H. V.; Dijk, W. L. V.; Brongersma, H. H. Low-energy electron-impact excitation spectra of formaldehyde, acetaldehyde and acetone. *Chemical Physics* **1976**, *16*, 337–345.
- [198] Walker, I. C.; Stamatovic, A.; Wong, S. F. Vibrational excitation of ethylene by electron impact: 1-11 eV. *The Journal of Chemical Physics* **1978**, *69*, 5532.
- [199] Panajotovic, R.; Kitajima, M.; Tanaka, H.; Jelisavcic, M.; Lower, J.; Campbell, L.; Brunger, M. J.; Buckman, S. J. Electron collisions with ethylene. *Journal of Physics B* **2003**, *36*, 1615.

- [200] Schiedt, J.; Weinkauff, R. Resonant photodetachment via shape and Feshbach resonances: p-benzoquinone anions as a model system. *The Journal of Chemical Physics* **1999**, *110*, 304.
- [201] Dora, A.; Tennyson, J.; Bryjko, L.; van Mourik, T. R-matrix calculation of low-energy electron collisions with uracil. *The Journal of Chemical Physics* **2009**, *130*, 164307.
- [202] Kossoski, F.; Bettega, M. H. F.; Varella, M. T. d. N. Shape resonance spectra of uracil, 5-fluorouracil, and 5-chlorouracil. *The Journal of Chemical Physics* **2014**, *140*, 24317.
- [203] Aflatooni, K.; Gallup, G. A.; Burrow, P. D. Electron Attachment Energies of the DNA Bases. *The Journal of Physical Chemistry A* **1998**, *102*, 6205–6207.
- [204] Tielens, A. Interstellar Polycyclic Aromatic Hydrocarbon Molecules. *Annual Review of Astronomy and Astrophysics* **2008**, *46*, 289–337.
- [205] Santaloci, T. J.; Fortenberry, R. C. Electronically Excited States of Closed-Shell, Cyano-Functionalized Polycyclic Aromatic Hydrocarbon Anions. *Chemistry* **2021**, *3*, 296–313.
- [206] Heinis, T.; Chowdhury, S.; Kebarle, P. Electron affinities of naphthalene, anthracene and substituted naphthalenes and anthracenes. *Organic Mass Spectrometry* **1993**, *28*, 358–365.
- [207] Bhaskaran-Nair, K.; Kowalski, K.; Jarrell, M.; Moreno, J.; Shelton, W. A. Equation of motion coupled cluster methods for electron attachment and ionization potential in polyacenes. *Chemical Physics Letters* **2015**, *641*, 146–152.
- [208] Hajgató, B.; Deleuze, M. S.; Tozer, D. J.; De Proft, F. A benchmark theoretical study of the electron affinities of benzene and linear acenes. *The Journal of Chemical Physics* **2008**, *129*, 84308.
- [209] Modelli, A.; Venuti, M. Temporary  $\pi^*$  and  $\sigma^*$  Anions and Dissociative Electron Attachment in Chlorobenzene and Related Molecules. *The Journal of Physical Chemistry A* **2001**, *105*, 5836–5841.
- [210] Stricklett, K.; Chu, S.; Burrow, P. Dissociative attachment in vinyl and allyl chloride, chlorobenzene and benzyl chloride. *Chemical Physics Letters* **1986**, *131*, 279–284.
- [211] Kaufel, R.; Illenberger, E.; Baumgärtel, H. Formation and dissociation of the chloroethylene anions. *Chemical Physics Letters* **1984**, *106*, 342–346.
- [212] Olthoff, J. K.; Tossell, J. A.; Moore, J. H. Electron attachment by haloalkenes and halobenzenes. *The Journal of Chemical Physics* **1985**, *83*, 5627–5634.

- [213] Burrow, P.; Modelli, A.; Chiu, N.; Jordan, K. Temporary  $\Sigma$  and  $\Pi$  anions of the chloroethylenes and chlorofluoroethylenes. *Chemical Physics Letters* **1981**, *82*, 270–276.
- [214] Skomorowski, W.; Krylov, A. I. Feshbach–Fano approach for calculation of Auger decay rates using equation-of-motion coupled-cluster wave functions. I. Theory and implementation. *The Journal of Chemical Physics* **2021**, *154*, 084124.
- [215] Jakob, W.; Rhineland, J.; Moldovan, D. pybind11 – Seamless operability between C++11 and Python. 2017; <https://github.com/pybind/pybind11>.
- [216] Schaftenaar, G.; Noordik, J. H. Molden: a pre- and post-processing program for molecular and electronic structures\*. *Journal of Computer-Aided Molecular Design* **2000**, *14*, 123–134.
- [217] Fdez. Galván, I. et al. OpenMolcas: From Source Code to Insight. *Journal of Chemical Theory and Computation* **2019**, *15*, 5925–5964.
- [218] Sun, Q. et al. Recent developments in the PySCF program package. *The Journal of Chemical Physics* **2020**, *153*, 024109.
- [219] Turney, J. M. et al. Psi4: An Open-Source *Ab Initio* Electronic Structure Program. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 556–565.
- [220] Ghigo, G.; Roos, B.; Malmqvist, P.-A. *Chemical Physics Letters* **2004**, *396*, 142.
- [221] Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1340.
- [222] Sun, Q. Libcint: An efficient general integral library for Gaussian basis functions. *Journal of Computational Chemistry* **2015**, *36*, 1664–1671.
- [223] Herbst, M. F.; Scheurer, M.; Fransson, T.; Rehn, D. R.; Dreuw, A. adcc: A versatile toolkit for rapid development of algebraic-diagrammatic construction methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2020**, *10*, e1462.
- [224] DOI:10.5281/zenodo.5979220.
- [225] Lehtola, S.; Steigemann, C.; Oliveira, M. J.; Marques, M. A. Recent developments in LIBXC - A comprehensive library of functionals for density functional theory. *SoftwareX* **2018**, *7*, 1–5.

- [226] Bast, R. Numgrid: Numerical integration grid for molecules. 2020; <https://doi.org/10.5281/zenodo.1470276>.
- [227] Lindh, R.; Malmqvist, P.-Å.; Gagliardi, L. Molecular integrals by numerical quadrature. I. Radial integration. *Theoretical Chemistry Accounts* **2001**, *106*, 178–187.
- [228] Burkardt, J. SPHERE\_LEBEDEV\_RULE: Quadrature Rules for the Unit Sphere. 2010; [https://people.sc.fsu.edu/~jburkardt/c\\_src/sphere\\_lebedev\\_rule/sphere\\_lebedev\\_rule.html](https://people.sc.fsu.edu/~jburkardt/c_src/sphere_lebedev_rule/sphere_lebedev_rule.html).
- [229] Nakano, H.; Hirao, K.; Gordon, M. Analytic energy gradients for multiconfigurational self-consistent field second-order quasidegenerate perturbation theory (MC-QDPT). *The Journal of Chemical Physics* **1998**, *108*, 5660–5669.
- [230] Sayfutyarova, E. R.; Hammes-Schiffer, S. Constructing Molecular  $\pi$ -Orbital Active Spaces for Multireference Calculations of Conjugated Systems. *Journal of Chemical Theory and Computation* **2019**, *15*, 1679–1689.
- [231] Stein, C. J.; Reiher, M. autoCAS: A Program for Fully Automated Multiconfigurational Calculations. *Journal of Computational Chemistry* **2019**, *40*, 2216–2226.
- [232] Ma, D.; Li Manni, G.; Gagliardi, L. The generalized active space concept in multiconfigurational self-consistent field methods. *The Journal of Chemical Physics* **2011**, *135*, 044128.
- [233] Aflatooni, K.; Hitt, B.; Gallup, G. A.; Burrow, P. D. Temporary anion states of selected amino acids. *The Journal of Chemical Physics* **2001**, *115*, 6489.
- [234] Nishimoto, Y.; Battaglia, S.; Lindh, R. Analytic First-Order Derivatives of (X)MS, XDW, and RMS Variants of the CASPT2 and RASPT2 Methods. *Journal of Chemical Theory and Computation* **2022**, *18*, 4269–4281.
- [235] Jensen, F. Describing Anions by Density Functional Theory: Fractional Electron Affinity. *Journal of Chemical Theory and Computation* **2010**, *6*, 2726–2735.
- [236] Lee, D.; Furche, F.; Burke, K. Accuracy of Electron Affinities of Atoms in Approximate Density Functional Theory. *The Journal of Physical Chemistry Letters* **2010**, *1*, 2124–2129.
- [237] Anderson, L. N.; Oviedo, M. B.; Wong, B. M. Accurate Electron Affinities and Orbital Energies of Anions from a Nonempirically Tuned Range-Separated Density Functional Theory Approach. *Journal of Chemical Theory and Computation* **2017**, *13*, 1656–1666.
- [238] Rösch, N.; Trickey, S. B. Comment on “Concerning the applicability of density functional methods to atomic and molecular negative ions” [J. Chem. Phys. *105*, 862 (1996)]. *The Journal of Chemical Physics* **1997**, *106*, 8940–8941.

- [239] Galbraith, J. M.; Schaefer, H. F. Concerning the applicability of density functional methods to atomic and molecular negative ions. *The Journal of Chemical Physics* **1996**, *105*, 862–864.
- [240] Shore, H. B.; Rose, J. H.; Zaremba, E. Failure of the local exchange approximation in the evaluation of the  $H^-$  ground state. *Physical Review B* **1977**, *15*, 2858–2861.
- [241] Rienstra-Kiracofe, J. C.; Tschumper, G. S.; Schaefer, H. F.; Nandi, S.; Ellison, G. B. Atomic and Molecular Electron Affinities: Photoelectron Experiments and Theoretical Computations. *Chemical Reviews* **2002**, *102*, 231–282.
- [242] Taylor, H. S.; Hazi, A. U. Comment on the stabilization method: variational calculation of the resonance width. *Physical Review A* **1976**, *14*, 2071–2074.
- [243] Thiam, G.; Rabilloud, F. Multi-Basis-Set (TD-)DFT Methods for Predicting Electron Attachment Energies. *The Journal of Physical Chemistry Letters* **2021**, *12*, 9995–10001.
- [244] Cheng, H.-Y.; Chen, C.-W. Energy and Lifetime of Temporary Anion States of Uracil by Stabilization Method. *The Journal of Physical Chemistry A* **2011**, *115*, 10113–10121.
- [245] Cheng, H.-Y.; Chen, C.-W.; Chang, J.-T.; Shih, C.-C. Application of the Stabilization Method to Temporary Anion States of  $CH_3CN$ ,  $CH_3NC$ ,  $CH_3SCN$ , and  $CH_3NCS$  in Density Functional Theory with Asymptotically Corrected Potentials. *The Journal of Physical Chemistry A* **2011**, *115*, 84–93.
- [246] Cheng, H.-Y.; Chen, C.-W.; Huang, C.-H. Characterization of the Temporary Anion States on Perfluoroalkanes via Stabilized Koopmans' Theorem in Long-Range Corrected Density Functional Theory. *The Journal of Physical Chemistry A* **2012**, *116*, 3224–3236.
- [247] Cheng, H.-Y.; Chen, C.-W. Energies and Lifetimes of Temporary Anion States of Chloromethanes by Stabilized Koopmans' Theorem in Long-Range Corrected Density Functional Theory. *The Journal of Physical Chemistry A* **2012**, *116*, 12364–12372.
- [248] Cheng, H.-Y.; Chen, Y.-C.; Lin, C.-J.; Liu, W.-C.; Hsieh, S.-H. Temporary anion states of radiosensitive halopyrimidines: Shape and core-excited resonances. *Computational and Theoretical Chemistry* **2016**, *1075*, 18–29.
- [249] Falcetta, M. F.; Di Falco, L. A.; Ackerman, D. S.; Barlow, J. C.; Jordan, K. D. Assessment of Various Electronic Structure Methods for Characterizing Temporary Anion States: Application to the Ground State Anions of  $N_2$ ,  $C_2H_2$ ,  $C_2H_4$ , and  $C_6H_6$ . *The Journal of Physical Chemistry A* **2014**, *118*, 7489–7497.

- [250] Cheng, H.-Y.; Huang, Y.-S. Temporary anion states of p-benzoquinone: shape and core-excited resonances. *Physical Chemistry Chemical Physics* **2014**, *16*, 26306–26313.
- [251] Mensa-Bonsu, G.; Lietard, A.; Tozer, D. J.; Verlet, J. R. R. Low energy electron impact resonances of anthracene probed by 2D photoelectron imaging of its radical anion. *The Journal of Chemical Physics* **2020**, *152*, 174303.
- [252] Becke, A. DENSITY-FUNCTIONAL EXCHANGE-ENERGY APPROXIMATION WITH CORRECT ASYMPTOTIC BEHAVIOR. *Physical Review A* **1988**, *38*, 3098–3100.
- [253] Lee, C.; Yang, W.; Parr, R. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Physical Review B* **1988**, *37*, 785–789.
- [254] Vydrov, O. A.; Scuseria, G. E. Assessment of a long-range corrected hybrid functional. *The Journal of Chemical Physics* **2006**, *125*, 234109.
- [255] Chai, J.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion interactions. *Physical Chemistry Chemical Physics* **2008**, *10*, 6615–6620.
- [256] Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.
- [257] Hagräs, M. A.; Stuchebrukhov, A. A. Electron tunneling in proteins program. *Journal of Computational Chemistry* **2016**, *37*, 1388–1395.
- [258] Beratan, D.; Onuchic, J.; Winkler, J.; Gray, H. Electron-tunneling pathways in proteins. *Science* **1992**, *258*, 1740–1741.
- [259] Jones, M. L.; Kurnikov, I. V.; Beratan, D. N. The Nature of Tunneling Pathway and Average Packing Density Models for Protein-Mediated Electron Transfer. *The Journal of Physical Chemistry A* **2002**, *106*, 2002–2006.
- [260] Hamelryck, T.; Manderick, B. PDB file parser and structure class implemented in Python. *Bioinformatics* **2003**, *19*, 2308–2310.
- [261] Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B., et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423.
- [262] Sanner, M. F.; Olson, A. J.; Spehner, J. C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305–320.

- [263] Chakravarty, S.; Varadarajan, R. Residue depth: A novel parameter for the analysis of protein structure and stability. *Structure* **1999**, *7*, 723–732.
- [264] Song, J.; Tan, H.; Mahmood, K.; Law, R. H. P.; Buckle, A. M.; Webb, G. I.; Akutsu, T.; Whisstock, J. C. Prodepth: Predict Residue Depth by Support Vector Regression Approach from Protein Sequences Only. *PLoS ONE* **2009**, *4*, e7072.
- [265] Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.
- [266] Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A.; Krieger, E.; Joosten, R. P.; Vriend, G. A series of PDB-related databanks for everyday needs. *Nucleic Acids Research* **2015**, *43*, D364–D368.
- [267] Tien, M. Z.; Meyer, A. G.; Sydykova, D. K.; Spielman, S. J.; Wilke, C. O. Maximum allowed solvent accessibilities of residues in proteins. *PLoS ONE* **2013**, *8*, e80635.
- [268] Joosten, R. P.; te Beek, T. A. H.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W. W.; Schneider, R.; Sander, C.; Vriend, G. A series of PDB related databases for everyday needs. *Nucleic Acids Research* **2011**, *39*, D411–D419.
- [269] *PyGraphviz License* **2013**,
- [270] Rose, A.; Hildebrand, P. NGL Viewer: A web application for molecular visualization. *Nucleic Acids Research* **2015**, *43*, W576–W579.
- [271] Prlić, A.; Bradley, A. R.; Duarte, J. M.; Rose, P. W.; Rose, A. S.; Valasatava, Y. NGL viewer: Web-based molecular graphics for large complexes. *Bioinformatics* **2018**, *34*, 3755–3758.
- [272] Yen, J. Y. Finding the  $K$  Shortest Loopless Paths in a Network. *Management Science* **1971**, *17*, 712–716.
- [273] Chaves, I.; Pokorny, R.; Byrdin, M.; Hoang, N.; Ritz, T.; Brettel, K.; Essen, L.-O.; van der Horst, G. T. J.; Batschauer, A.; Ahmad, M. Cryptochromes: Blue light photoreceptors in plants and animals. *Annual Review of Plant Biology* **2011**, *62*, 335–364.
- [274] Kao, Y.-T.; Tan, C.; Song, S.-H.; Öztürk, N.; Li, J.; Wang, L.; Sancar, A.; Zhong, D. Ultrafast Dynamics and Anionic Active States of the Flavin Cofactor in Cryptochrome and Photolyase. *Journal of the American Chemical Society* **2008**, *130*, 7695–7701.



- [275] Brazard, J.; Usman, A.; Lacomat, F.; Ley, C.; Martin, M. M.; Plaza, P.; Mony, L.; Heijde, M.; Zabulon, G.; Bowler, C. Spectro-Temporal Characterization of the Photoactivation Mechanism of Two New Oxidized Cryptochrome/Photolyase Photoreceptors. *Journal of the American Chemical Society* **2010**, *132*, 4935–4945.
- [276] Burney, S.; Hoang, N.; Caruso, M.; Dudkin, E. A.; Ahmad, M.; Bouly, J.-P. Conformational change induced by ATP binding correlates with enhanced biological function of *Arabidopsis* cryptochrome. *FEBS Letters* **2009**, *583*, 1427–1433.
- [277] Solov'yov, I. A.; Domratcheva, T.; Moughal Shahi, A. R.; Schulten, K. Decrypting Cryptochrome: Revealing the Molecular Identity of the Photoactivation Reaction. *Journal of the American Chemical Society* **2012**, *134*, 18046–18052.
- [278] Brautigam, C. A.; Smith, B. S.; Ma, Z.; Palnitkar, M.; Tomchick, D. R.; Machius, M.; Deisenhofer, J. Structure of the photolyase-like domain of cryptochrome 1 from *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **2004**, *101*, 12142–12147.
- [279] Takematsu, K.; Williamson, H. R.; Nikolovski, P.; Kaiser, J. T.; Sheng, Y.; Pospíšil, P.; Towrie, M.; Heyda, J.; Hollas, D.; Zálíš, S., et al. Two Tryptophans Are Better Than One in Accelerating Electron Flow through a Protein. *ACS Central Science* **2019**, *5*, 192–200.
- [280] Ramirez, B.; Malmström, B.; Winkler, J.; Gray, H. The currents of life: the terminal electron-transfer complex of respiration. *Proceedings of the National Academy of Sciences* **1995**, *92*, 11949–11951.
- [281] Anraku, Y. Bacterial electron transport chains. *Annual Review of Biochemistry* **1988**, *57*, 101–132.
- [282] Fleming, G.; Martin, J.; Breton, J. Rates of primary electron transfer in photosynthetic reaction centres and their mechanistic implications. *Nature* **1988**, *333*, 190–192.
- [283] Jortner, J. Dynamics of electron transfer in bacterial photosynthesis. *Biochimica et Biophysica Acta* **1980**, *594*, 193–230.
- [284] Weber, S. Light-driven enzymatic catalysis of DNA repair: a review of recent biophysical studies on photolyase. *Biochimica et Biophysica Acta* **2005**, *1707*, 1–22.
- [285] Zhong, D. Electron Transfer Mechanisms of DNA Repair by Photolyase. *Annual Review of Physical Chemistry* **2015**, *66*, 691–715.

- [286] Gray, H.; Winkler, J. Functional and protective hole hopping in metalloenzymes. *Chemical Science* **2021**, *12*, 13988–14003.
- [287] Giovani, B.; Byrdin, M.; Ahmad, M.; Brettel, K. Light-induced electron transfer in a cryptochrome blue-light photoreceptor. *Nature Structural & Molecular Biology* **2003**, *10*, 489–490.
- [288] Karki, N.; Vergish, S.; Zoltowski, B. Cryptochromes: Photochemical and structural insight into magnetoreception. *Protein Science* **2021**, *30*, 1521–1534.
- [289] Kim, S.-T.; Sancar, A. Photochemistry, photophysics, and mechanism of pyrimidine dimer repair by DNA photolyase. *Journal of Photochemistry and Photobiology* *57*, 895–904.
- [290] Yee, E.; Dzikovski, B.; Crane, B. Tuning radical relay residues by proton management rescues protein electron hopping. *Journal of the American Chemical Society* **2019**, *141*, 17571–17587.
- [291] Kennis, J. T. M.; Mathes, T. Molecular eyes: proteins that transform light into biological information. *Interface Focus* **2013**, *3*, 20130005.
- [292] Barry, B. Proton coupled electron transfer and redox active tyrosines in Photosystem II. *The Journal of Physical Chemistry Letters* **2011**, *104*, 60–71.
- [293] Pastore, A.; Teo, R.; Montoya, A.; Burg, M.; Twahir, U.; Bruner, S.; Beratan, D. N.; Angerhofer, A. Oxalate decarboxylase uses electron hole hopping for catalysis. *Journal of Biological Chemistry* **2021**, *297*, 100857.
- [294] Reinhardt, C.; Konstantinovsky, D.; Soudackov, A.; Hammes-Schiffer, S. Kinetic model for reversible radical transfer in ribonucleotide reductase. *Proceedings of the National Academy of Sciences* **2021**, *119*, e2202022119.
- [295] Sjulstok, E.; Magnus, J.; Olsen, H.; Solov'yov, I. Quantifying electron transfer reactions in biological systems: what interactions play the major role? *Scientific Reports* **2016**, *5*, 18446.
- [296] de la Lande, A.; Gillet, N.; Chen, S.; Salahub, D. R. Progress and challenges in simulating and understanding electron transfer in proteins. *Archives of Biochemistry and Biophysics* **2015**, *582*, 28–41.
- [297] Blumberger, J. Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chemical Reviews* **2015**, *115*, 11191–11238.
- [298] Hammes-Schiffer, S. Proton-Coupled Electron Transfer: Moving Together and Charging Forward. *Journal of the American Chemical Society* **2015**, *137*, 8860–8871.

- [299] Teo, R. D.; Wang, R.; Smithwick, E. R.; Migliore, A.; Therien, M. J.; Beratan, D. N. Mapping hole hopping escape routes in proteins. *Proceedings of the National Academy of Sciences* **2019**, *116*, 15811–15816.
- [300] In some sources, the term *subgraph isomorphism* is reserved for when  $G'$  is a node- or edge-induced subgraph, and the term *monomorphism* is preferred for subgraphs which are not induced. Here, we use the term subgraph isomorphism in all cases.
- [301] Dijkstra, E. W. A note on two problems in connexion with graphs. *Numerische mathematik* **1959**, *1*, 269–271.
- [302] Gayvert, J.; Bravaya, K.; Tazhigulov, R. Overview of pyemap. <https://pyemap.readthedocs.io/en/latest/index.html>.
- [303] Mrzic, A.; Meysman, P.; Bittremieux, W.; Moris, P.; Cule, B.; Goethals, B.; Laukens, K. Grasping frequent subgraph mining for bioinformatics applications. *BioData Mining* **2018**, *11*, 20.
- [304] Chi, Y.; Muntz, R. R.; Nijssen, S.; Kok, J. N. Frequent Subtree Mining –An Overview. *Fundamenta Informaticae* **2005**, *66*, 161–198.
- [305] Meysman, P.; Zhou, C.; Cule, B.; Goethals, B.; Laukens, K. Mining the entire Protein DataBank for frequent spatially cohesive amino acid patterns. *BioData Mining* **2015**, *8*, 4.
- [306] Bandyopadhyay, D.; Huan, J.; Liu, J.; Prins, J.; Snoeyink, J.; Wang, W.; Tropsha, A. Structure-based function inference using protein family-specific fingerprints. *Protein Science* **2006**, *15*, 1537–1543.
- [307] Dhifli, W.; Saidi, R.; Nguifo, E. M. Smoothing 3D Protein Structure Motifs Through Graph Mining and Amino Acid Similarities. *Journal of Computational Biology* **2013**, *21*, 162–172.
- [308] Chen, Q. gspan-mining: Implementation of frequent subgraph mining algorithm gSpan. <https://pypi.org/project/gspan-mining/>.
- [309] gSpan-mining License. <https://github.com/betterenvi/gSpan/blob/master/LICENSE>, 2018; MIT.
- [310] NetworkX License. <https://github.com/networkx/networkx/blob/main/LICENSE.txt>, 2004–2022; BSD-3.
- [311] Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **2004**, *32*, 1792–1797.

- [312] Minnihan, E. C.; Nocera, D. G.; Stubbe, J. Reversible, Long-Range Radical Transfer in *E. coli* Class Ia Ribonucleotide Reductase. *Accounts of Chemical Research* **2013**, *46*, 2524–2535.
- [313] Kang, G.; Taguchi, A. T.; Stubbe, J.; Drennan, C. L. Structure of a trapped radical transfer pathway within a ribonucleotide reductase holocomplex. *Science* **2020**, *368*, 424–427.
- [314] Wang, J.; Du, X.; Pan, W.; Wu, W. Photoactivation of the cryptochrome/photolyase superfamily. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* **2015**, *22*, 84–102.
- [315] Liu, Z.; Tan, C.; Guo, X.; Li, J.; Wang, L.; Sancar, A.; Zhong, D. Determining complete electron flow in the cofactor photoreduction of oxidized photolyase. *Proceedings of the National Academy of Sciences* **2013**, *110*, 12966–12971.
- [316] Aubert, C.; Vos, M. H.; Mathis, P.; Eker, A. P.; Brettel, K. Intraprotein radical transfer during photoactivation of DNA photolyase. *Nature* **2000**, *405*, 586–590.
- [317] Nohr, D.; Franz, S.; Rodriguez, R.; Paulus, B.; Essen, L.-O.; Weber, S.; Schleicher, E. Extended Electron-Transfer in Animal Cryptochromes Mediated by a Tetrad of Aromatic Amino Acids. *Biophysical Journal* **2016**, *111*, 301–311.
- [318] Shao, K.; Zhang, X.; Li, X.; Hao, Y.; Huang, X.; Ma, M.; Zhang, M.; Yu, F.; Liu, H.; Zhang, P. The oligomeric structures of plant cryptochromes. *Nature Structural & Molecular Biology* **2020**, *27*, 480–488.
- [319] Geisselbrecht, Y.; Fr uhwirth, S.; Schroeder, C.; Pierik, A.; Klug, G.; Essen, L.-O. CryB from *Rhodobacter sphaeroides*: a unique class of cryptochromes with new cofactors. *EMBO Reports* **2012**, *13*, 223–229.
- [320] Attack, J.; Kelly, D. Structure, Mechanism and Physiological Roles of Bacterial Cytochrome *c* Peroxidases. *Advances in Microbial Physiology* **2007**, *52*, 73–106.
- [321] Rizzolo, K.; Cohen, S. E.; Weitz, A. C.; Muñoz, M. M. L.; Hendrich, M. P.; Drennan, C. L.; Elliott, S. J. A widely distributed diheme enzyme from *Burkholderia* that displays an atypically stable *bis*-Fe (IV) state. *Nature Communications* **2019**, *10*, 1–10.
- [322] Dias, J. M.; Alves, T.; Bonifácio, C.; Pereira, A. S.; Trincão, J.; Bourgeois, D.; Moura, I.; Romão, M. J. Structural Basis for the Mechanism of Ca<sup>2+</sup> Activation of the Di-Heme Cytochrome *c* Peroxidase from *Pseudomonas nautica* 617. *Structure* **2004**, *12*, 961–973.

- [323] Stanton, J. et al. in CFOUR, Coupled Cluster techniques for Computational Chemistry, a quantum-chemical program package ([www.cfour.de](http://www.cfour.de)).
- [324] Li, Z.; Ryszka, M.; Dawley, M. M.; Carmichael, I.; Bravaya, K. B.; Ptasíńska, S. Dipole-Supported Electronic Resonances Mediate Electron-Induced Amide Bond Cleavage. *Physical Review Letters* **2019**, *122*, 073002.
- [325] Mai, S.; Richter, M.; Heindl, M.; Menger, M. F. S. J.; Atkins, A.; Ruckebauer, M.; Plasser, F.; Ibele, L. M.; Kropf, S.; Oppel, M.; Marquetand, P.; González, L. SHARC2.1: Surface Hopping Including Arbitrary Couplings - Program Package for Non-Adiabatic Dynamics. [sharc-md.org](http://sharc-md.org), 2019.
- [326] Richter, M.; Marquetand, P.; González-Vázquez, J.; Sola, I.; González, L. SHARC: ab initio Molecular Dynamics with Surface Hopping in the Adiabatic Representation Including Arbitrary Couplings. *Journal of Chemical Theory and Computation* **2011**, *7*, 1253–1258.
- [327] Mai, S.; Marquetand, P.; González, L. Nonadiabatic Dynamics: The SHARC Approach. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1370.
- [328] Dou, W.; Nitzan, A.; Subotnik, J. E. Surface hopping with a manifold of electronic states. II. Application to the many-body Anderson-Holstein model. *The Journal of Chemical Physics* **2015**, *142*, 084110.
- [329] Ouyang, W.; Dou, W.; Subotnik, J. E. Surface hopping with a manifold of electronic states. I. Incorporating surface-leaking to capture lifetimes. *The Journal of Chemical Physics* **2015**, *142*, 084109.
- [330] Çaylak, O.; Yaman, A.; Baumeier, B. Evolutionary Approach to Constructing a Deep Feedforward Neural Network for Prediction of Electronic Coupling Elements in Molecular Materials. *Journal of Chemical Theory and Computation* **2019**, *15*, 1777–1784.
- [331] Bai, X.; Guo, X.; Wang, L. Machine Learning Approach to Calculate Electronic Couplings between Quasi-diabatic Molecular Orbitals: The Case of DNA. *The Journal of Physical Chemistry Letters* **2021**, *12*, 10457–10464.
- [332] Wang, C.-I.; Braza, M. K. E.; Claudio, G. C.; Nellas, R. B.; Hsu, C.-P. Machine Learning for Predicting Electron Transfer Coupling. *The Journal of Physical Chemistry A* **2019**, *123*, 7792–7802.
- [333] Valeev, E.; Coropceanu, V.; da Silva Filho, D.; Salman, S.; Brédas, J.-L. Effect of electronic polarization on charge-transport parameters in molecular organic semiconductors. *Journal of the American Chemical Society* **2006**, *128*, 9882–9886.

- [334] Harris, R. C.; Pettitt, B. M. Effects of geometry and chemistry on hydrophobic solvation. *Proceedings of the National Academy of Sciences* **2014**, *111*, 14681–14686.
- [335] Paluch, A. S.; Shah, J. K.; Maginn, E. J. Efficient Solvation Free Energy Calculations of Amino Acid Analogs by Expanded Ensemble Molecular Simulation. *Journal of Chemical Theory and Computation* **2011**, *7*, 1394–1403.
- [336] Haynes, W. M. *CRC Handbook of Chemistry and Physics, 96th Edition (Internet Version)*; CRC Press/Taylor and Francis: Boca Raton, FL, United States, 2016.

**CURRICULUM VITAE**

