

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Decompose Everything! Exploring the Fundamental Components of Many-Body Chemical Systems

Permalink

<https://escholarship.org/uc/item/7st8500s>

Author

Egan, Colin Kerry

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Decompose Everything! Exploring the Fundamental Components of Many-Body Chemical
Systems

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Chemistry

by

Colin Kerry Egan

Committee in charge:

Professor Francesco Paesani, Chair

Professor Michael Holst

Professor John Weare

Professor Haim Weizman

Professor Joel Yuen Zhou

2023

Copyright

Colin Kerry Egan, 2023

All rights reserved.

The Dissertation of Colin Kerry Egan is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To my parents, Ana and Brian, for their love and support. I love you both.

EPIGRAPH

*Gar Manches rechnet Erwin schon
Mit seiner Wellenfunktion.
Nur wissen möcht man gerne wohl
Was man sich dabei vorstell'n soll.*
:
*Erwin with his psi can do
Calculations quite a few.
But one thing has not been seen:
Just what does psi really mean?*

—Erich Hückel, 1926 (translated by Felix Bloch)[1]

TABLE OF CONTENTS

Dissertation Approval Page	iii
Dedication	iv
Epigraph	v
Table of Contents	vi
List of Figures	ix
List of Tables	xvii
Acknowledgements	xix
Vita	xx
Abstract of the Dissertation	xxi
Chapter 1 Introduction	1
Chapter 2 Electronic Structure Theory Background	2
2.1 The Born-Oppenheimer Approximation	5
2.2 Wave Function Methods	6
2.2.1 One-Electron Orbitals	6
2.2.2 Basis Set Approximations	7
2.2.3 Inner Products and Bra-Ket Notation	9
2.2.4 Many-Electron Wave Functions	11
2.2.5 Slater-Condon Rules	17
2.2.6 Hartree-Fock Theory	25
2.2.7 The Correlation Energy and The Configuration Interaction Method	30
2.2.8 Size Consistency and Coupled Cluster Methods	32
2.3 Density Functional Theory	38
2.3.1 The Kohn-Sham Equations	40
2.3.2 Density Functional Approximations	42
Chapter 3 Many-Body Interactions	46
3.1 Intermolecular Interactions and The Many-Body Expansion	47
3.1.1 Monomer Distortion Energies and The Binding Energy	47
3.1.2 The Interaction Energy	48
3.1.3 The Many-Body Expansion	51
3.1.4 The SAMBA Method and The Basis Set Superposition Error	55
3.2 Many-Body Interactions in Deprotonated Water Clusters	57
3.2.1 The Minimum Energy $\text{OH}^-(\text{H}_2\text{O})$ and $\text{OH}^-(\text{H}_2\text{O})_2$ Clusters	58
3.2.2 Fragment Groups and Fragment Notation	62

3.2.3	Fragment Analysis and Many-Body Effects in Deprotonated Water Clusters	64
3.2.4	Assessment of DFT in Reproducing Many-Body Interactions in Deprotonated Water Clusters	69
3.3	Many-Body Interactions in Protonated Water Clusters	76
3.3.1	Fragment Analysis With and Without Context	77
3.3.2	The Minimum Energy $\text{H}_3\text{O}^+(\text{H}_2\text{O})$ and $\text{H}_3\text{O}^+(\text{H}_2\text{O})_2$ Clusters	82
3.3.3	Fragment Analysis and Many-Body Effects in Protonated Water Clusters	85
3.3.4	Assessment of DFT in Reproducing Many-Body Interactions in Protonated Water Clusters	92
3.4	Many-Body Interactions in Alkali Ion–Water Clusters	97
3.4.1	2B Interaction Energies	97
3.4.2	3B Interaction Energies	100
3.5	The Scaling of Many-Body Energies in a Lattice	102
3.5.1	Lattices in 1 Spatial Dimension: Return of the Fragment Group	103
3.5.2	Lattices in 2 Spatial Dimensions: Deriving Exact Scaling Relations for Counting Lattice Fragment Groups	105
3.6	Acknowledgements	111
Chapter 4	Energy Decomposition Analysis	112
4.1	Background	112
4.1.1	ALMO-EDA	112
4.2	Energy Decompositions of Halide–Water Clusters	114
4.3	Energy Decompositions of Alkali Ion–Water Clusters And Comparisons With Halide–Water Clusters	119
4.4	Many-Body Energy Decompositions of Protonated and Deprotonated Water Clusters	126
4.5	Acknowledgements	136
Chapter 5	Model Potential Energy Functions	137
5.1	Development of Many-Body PEFs: The MB-nrg Methodology	138
5.1.1	TTM-nrg PEFs: Obtaining the Long-Range Potential	139
5.1.2	MB-nrg PEFs	141
5.1.3	Brief Motivation for Permutationally Invariant Polynomials	143
5.1.4	Permutationally Invariant Polynomials For N Identical Particles	144
5.1.5	“MB-Fitting” Permutationally Invariant Polynomials	146
5.1.6	Fitting Permutationally Invariant Polynomial Gradients	150
5.1.7	Considerations Regarding Switching Functions	152
5.1.8	Developing MB-nrg PEFs	153
5.1.9	Training Set Preparation and Refinement	155
5.1.10	Model Preparation and Refinement: Monomial Connectivity Graphs ...	160
5.2	Evaluating Alkali Ion–Water PEFs	164
5.2.1	2B Energies of Alkali Ion–Water Dimers	165
5.2.2	3B Energies of Alkali Ion–Water Trimers	166

5.2.3	2B Rigid Scans	167
5.3	Acknowledgements	172
Chapter 6	Molecular Dynamics: Simulations of The Zundel Ion in Acetonitrile	173
6.1	The Failure of TTM-nrg For The Flexible Zundel Ion, and The Importance of Evaluating PEFs with Test Sets Taken From Simulations	175
6.2	Analysis of The TTM-nrg Simulations of The Zundel Ion Dissolved in Acetonitrile	180
6.2.1	Simulation Details	180
6.2.2	Calculating Vibrational Spectra and Power Spectrum Decompositions ..	181
6.2.3	The Structure of the Zundel Ion in Acetonitrile in TTM-nrg Simulations ..	184
6.2.4	The Dynamics of the Zundel Ion in Acetonitrile in TTM-nrg Simulations ..	194
Chapter 7	Chemical Reactions: First Steps Toward A Many-Body Valence Bond Model For Aqueous Proton Transfer	204
7.1	Background	204
7.2	MB-VB	206
7.2.1	Model DS PEFs Based on ALMO-EDA Charge Transfer Energies	211
7.2.2	Preliminary Tests of MB-VB on Protonated Water Clusters	214
Chapter 8	Conclusions	221
Appendix A	Electronic Structure Calculation Details	222
A.1	Protonated and Deprotonated Water Clusters	222
A.2	Halide–Water Clusters	224
A.3	Alkali Ion–Water Clusters	224
Appendix B	Energy Decomposition Analysis Comparisons	226
B.1	Halide–Water Clusters	226
B.2	Alkali Ion–Water Clusters	228
Appendix C	Potential Energy Functions	230
C.1	Alkali Ion–Water PEF Descriptions	230
C.1.1	The AMOEBA force field	230
C.1.2	The TTM-nrg PEF	231
C.1.3	The MB-nrg PEF	232
Appendix D	MB-VB	234
D.1	Application of MB-VB Models With Various Switching Parameters on Protonated Water Clusters	234
Appendix E	Acknowledgements	240
Bibliography		241

LIST OF FIGURES

Figure 3.1.	Cartoon depicting the interpretation of the interaction and binding energies of a cluster.	50
Figure 3.2.	Low-lying isomers of the $\text{OH}^-(\text{H}_2\text{O})_3$ cluster shown in order of increasing interaction energy from isomer a to isomer f	58
Figure 3.3.	Low-lying isomers of the $\text{OH}^-(\text{H}_2\text{O})_4$ cluster shown in order of increasing interaction energy from isomer a to isomer d	58
Figure 3.4.	Low-lying isomers of the $\text{OH}^-(\text{H}_2\text{O})_5$ clusters shown in order of increasing interaction energy from isomer a to isomer d	58
Figure 3.5.	Convergence of the many-body expansion of the low-lying isomers of $\text{OH}^-(\text{H}_2\text{O})_3$ shown in Fig. 3.2.	59
Figure 3.6.	Convergence of the many-body expansion of the low-lying isomers of $\text{OH}^-(\text{H}_2\text{O})_4$ shown in Fig. 3.3.	60
Figure 3.7.	Convergence of the many-body expansion of the low-lying isomers of $\text{OH}^-(\text{H}_2\text{O})_5$ shown in Fig. 3.4.	61
Figure 3.8.	$\text{OH}^-(\text{H}_2\text{O})_n$ building blocks. a) $\text{OH}^-(\text{H}_2\text{O})$ global minimum-energy structure, b) $\text{OH}^-(\text{H}_2\text{O})_2$ global minimum-energy structure along with its constituent 2-mer fragments (c-e).	61
Figure 3.9.	The $\{\text{OH}^-, {}^1\text{W}_{\leftarrow C}^A, {}^1\text{W}_{\leftarrow C, AB\leftarrow}^B, {}^2\text{W}_{FG}^C\}$ fragments within the 4-, 5-, and 6-mer GME structures.	68
Figure 3.10.	DFT errors for the hydrogen-bonded 2-mer fragment groups: (a) $\{\text{OH}^-, {}^1\text{W}_{FG}\}$; (b) $\{\text{W}_{\leftarrow B}^A, A\leftarrow \text{W}_{FG}^B\}$	70
Figure 3.11.	DFT errors for the non-hydrogen-bonded 2-mer fragment groups $\{\text{OH}^-, {}^2\text{W}_{FG}\}$ (panel a) and $\{\text{W}_{FG}^A, \text{W}_{FG}^B\}$ (panel b).	72
Figure 3.12.	Relative 2B errors, $\frac{E_{\text{DFT}} - E_{\text{ref}}}{E_{\text{ref}}}$, as a function of the water–water O–O distance for the $\{\text{W}_{FG}^A, \text{W}_{FG}^B\}$ fragments.	73
Figure 3.13.	DFT errors for the anticooperative 3-mer fragment group $\{\text{OH}^-, {}^1\text{W}_{FG}^A, {}^1\text{W}_{FG}^B\}$	74
Figure 3.14.	DFT errors for the cooperative 3-mer fragment groups $\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A\leftarrow {}^2\text{W}_{FG}^B\}$ (panel a) and $\{\text{W}_{\leftarrow B}^A, A\leftarrow \text{W}_{\leftarrow C, B\leftarrow}^B, {}^2\text{W}_{FG}^C\}$ (panel b).	75

Figure 3.15.	Low-lying isomers of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3$, $\text{H}_3\text{O}^+(\text{H}_2\text{O})_4$, and $\text{H}_3\text{O}^+(\text{H}_2\text{O})_5$ clusters, ordered in terms of their interaction energies.	77
Figure 3.16.	Convergence of the many-body expansion of the low-lying isomers of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3$ shown in Fig. 3.15a.	78
Figure 3.17.	Convergence of the many-body expansion of the low-lying isomers of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_4$ shown in Fig. 3.15b.	79
Figure 3.18.	Convergence of the many-body expansion of the interaction energy for the low-lying isomers of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_5$ shown in Fig. 3.15c.	80
Figure 3.19.	Examples of fragment group $\rightarrow$ contextual fragment group notation	80
Figure 3.20.	$\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ building blocks. a) $\text{H}_3\text{O}^+(\text{H}_2\text{O})$ global minimum energy structure. b) $\text{H}_3\text{O}^+(\text{H}_2\text{O})_2$ global minimum-energy structure along with its constituent 2-mer fragments (c–e).	83
Figure 3.21.	a) The Zundel ion shown with the C_2 axis passing through the shared proton (in gray), and one $\text{H}-\text{O}-\text{O}-\text{H}$ dihedral angle between equivalent flanking hydrogens (black arrows). b) The Eigen ion	83
Figure 3.22.	DFT errors for the H-bonded 2-mer fragment groups: a) $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{FG}$ and b) $\{B \leftarrow \text{W}^A, \text{W}_{\leftarrow A}^B\}_{FG}$	93
Figure 3.23.	DFT errors for the three contexts of the H-bonded $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{FG}$ fragment group: a) $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{CFG}$; b) $\{(\text{H}_3\text{O}^+, {}^1\text{W}^A)_{Z^+}\}_{CFG}$; c) $\{(\text{H}_3\text{O}^+)_{CZ^+}, {}^1\text{W}^A\}_{CFG}$	94
Figure 3.24.	DFT errors for the non-H-bonded 2-mer fragment groups: (a) $\{\text{H}_3\text{O}^+, \text{W}^A\}_{FG}$ and (b) $\{\text{W}^A, \text{W}^B\}_{FG}$	96
Figure 3.25.	DFT errors for the two most significant 3-mer fragment groups: a) the anticooperative $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ and b) the cooperative $\{\text{H}_3\text{O}^+, B \leftarrow {}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B\}_{FG}$	97
Figure 3.26.	Deviations from the CCSD(T) 2B interaction energies of the $\text{M}^+(\text{H}_2\text{O})$ dimers with $\text{M} = \text{Li}$ (a), Na (b), K (c), Rb (d), Cs (e), in the corresponding minimum energy configurations calculated with various XC-functionals as well as the AMOEBA, TTM-nrg, and MB-nrg PEFs.	98
Figure 3.27.	Deviations from the CCSD(T) 3B interaction energies of the $\text{M}^+(\text{H}_2\text{O})_2$ trimers with $\text{M} = \text{Li}$ (a), Na (b), K (c), Rb (d), Cs (e), in the corresponding minimum energy configurations calculated with various XC-functionals as well as the AMOEBA, TTM-nrg, and MB-nrg PEFs.	101

Figure 3.28.	Examples of 2-mer fragments in a 12-chain in 1 spatial dimension.	104
Figure 3.29.	All 2-mer, 3-mer, and 4-mer 1D LFGs with $s = 6$	105
Figure 3.30.	Example of the lattice convex hull (the red outline) for a 2D LFG.	107
Figure 4.1.	Deviations from the CCSD(T)-F12b interaction energies of the $X^-(H_2O)$ dimers, with $X = F$ (a), Cl (b), Br (c), and I (d), in the corresponding minimum-energy configurations calculated with various XC-functionals as well as AMOEBA, TTM-nrg, and MB-nrg PEFs.	115
Figure 4.2.	Total frozen energy from ALMO-EDA plotted against interaction energy deviations (relative to CCSD(T)-F12b) for $X^-(H_2O)$ dimers with $X = a)$ F , $b)$ Cl , $c)$ Br , $d)$ I , for each XC-functional, demonstrating the correlation between the two quantities.	118
Figure 4.3.	Correlation between the dipole magnitude of the water molecule from the minimum energy $Li^+(H_2O)$ dimer with the $Li^+(H_2O)$ ALMO-EDA ELEC energy (in kcal/mol).	122
Figure 4.4.	Total frozen energy from ALMO-EDA plotted against interaction energy deviations (relative to CCSD(T)) for $M^+(H_2O)$ dimers with $M = Li$ (a), Na (b), K (c), Rb (d), Cs (e), for each XC-functional, demonstrating the correlation between the two quantities.	127
Figure 4.5.	Analysis of SAMBA 3B interaction energy contributions and ALMO energy decompositions as a function of the average OH^- –water O–O distance in $\{OH^-, {}^1W^A, {}^1W^B\}_{FG}$ fragments.	128
Figure 4.6.	Correlation plot between the sum of 3B $OH^- \rightarrow {}^1W$ CT “electron reorganization” (in millielectrons, $m\bar{e}$) and average OH^- –water O–O distance in $\{OH^-, {}^1W^A, {}^1W^B\}_{FG}$ fragments.	129
Figure 4.7.	Analysis of 3B SAMBA interaction energy contributions and ALMO energy decompositions as a function of $OH^- - {}^1W - {}^2W$ O–O–O angle in $\{OH^-, {}^1W^A_{\leftarrow B}, A \leftarrow {}^2W^B\}_{FG}$ fragments.	130
Figure 4.8.	3B ALMO monomer $\rightarrow$ monomer CT reorganization vs $OH^- - {}^1W - {}^2W$ O–O–O angle for $\{OH^-, {}^1W^A_{\leftarrow B}, A \leftarrow {}^2W^B\}_{FG}$ fragments.	131
Figure 4.9.	Analysis of SAMBA 3B energy contributions and ALMO energy decompositions as a function of the average H_3O^+ –water O–O distance in $\{H_3O^+, {}^1W^A, {}^1W^B\}_{FG}$ fragments.	132

Figure 4.10.	Correlation plot between the sum of pairwise 3B ${}^1\text{W} \rightarrow \text{H}_3\text{O}^+$ CT “electron reorganization,” $\Delta Q_{\text{CT}}^{3\text{B}}$ (in millielectrons, $m\bar{e}$) and average $\text{H}_3\text{O}^+ - {}^1\text{W}$ O–O distance in $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments.	133
Figure 4.11.	Analysis of 3B SAMBA interaction energy contributions and ALMO energy decompositions as a function of the ${}^1\text{W} - {}^2\text{W}$ (H-bonded) donor-H and acceptor-O distance described in the main text in $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B\}_{FG}$ fragments.	134
Figure 4.12.	Pairwise 3B ALMO monomer $\rightarrow$ monomer CT reorganization vs. the ${}^1\text{W} - {}^2\text{W}$ (H-bonded) donor-H and acceptor-O distance described in the main text for $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B\}_{FG}$ fragments.	135
Figure 4.13.	Correlation between the mean 3B energy error and the mean ALMO EDA 3B CT energy of the 3-mer fragment groups: a) $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ and b) $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B\}_{FG}$, calculated with each XC-functional. ...	136
Figure 5.1.	An ethanol monomer with hydrogen (atomic sites) labeled based on their chemical identities. $\text{H}^{(1)}$, $\text{H}^{(2)}$, and $\text{H}^{(3)}$ are all identical, $\text{H}'^{(1)}$ and $\text{H}'^{(2)}$ are identical with each other.	143
Figure 5.2.	The boundary of the outer-cutoff of a switching function dependent on the distance from a single atom. This leads to a PEF that is not permutationally invariant in monomers such as ethane and the Zundel ion.	152
Figure 5.3.	The boundary of the outer-cutoff of a switching function dependent on the average distance from two atoms.	153
Figure 5.4.	The boundary of the outer-cutoff of a switching function dependent on the distance from the geometric center of two atoms.	154
Figure 5.5.	PEF development flow chart.	154
Figure 5.6.	Repulsive (panels a, b) and attractive (panels c, d) acetonitrile-acetonitrile scans used in TTM-nrg training sets.	158
Figure 5.7.	Comparison of NPT densities over time and test set correlation plots between three model PEFs for acetonitrile-acetonitrile interactions. MB-nrg (panels a, b), TTM-nrg fitted to attractive scans (panels c, d), and TTM-nrg fitted to attractive and repulsive scans (panels e, f).	159
Figure 5.8.	Examples of acetonitrile-acetonitrile 2B PIP monomials represented by their monomial connectivity graphs.	161
Figure 5.9.	Acetonitrile-acetonitrile 2B PIP monomial connectivity graphs from Fig. 5.8 with all non-relevant atomic sites grayed out.	162

Figure 5.10.	Examples of two acetonitrile-acetonitrile 2B PIP monomials, represented by their monomial connectivity graphs, that cannot be distinguished without the use of monomial connectivity graphs due to the permutational invariance of methyl hydrogens on the same fragment.	164
Figure 5.11.	Acetonitrile-acetonitrile 2B PIP monomial connectivity graphs from Fig. 5.10 with all non-relevant atomic sites grayed out.	164
Figure 5.12.	Schematic representation of the geometry used in the scans along the R_{M^+O} distance carried out for each of the four $M^+(H_2O)$ dimers, with $M = Li, Na, K, Rb, Cs$. The ion remains in the plane of the water molecule.	165
Figure 5.13.	Comparisons between CCSD(T), ω B97M-V, AMOEBA, TTM-nrg, and MB-nrg interaction energies calculated for radial scans of $M^+(H_2O)$ dimers, with $M = Li$ (a), Na (b), K (c), Rb (d), Cs (e).	169
Figure 5.14.	ELEC and POL energies calculated with each of the three sets of PEFs analyzed in this study, and the ω B97M-V functional (from ALMO-EDA) along the scans described above and depicted in Fig. 5.12.....	171
Figure 6.1.	Training set correlation plot for the Zundel-acetonitrile TTM-nrg PEF. Reference 2B energies were calculated with ω B97M-V/aug-cc-pVTZ. ...	176
Figure 6.2.	Test set correlation plot for the Zundel-acetonitrile TTM-nrg PEF. Reference 2B energies were calculated with ω B97M-V/aug-cc-pVTZ.	176
Figure 6.3.	Test set (the same as in Fig. 6.2) correlation plot for the Zundel-acetonitrile TTM-nrg PEF for configurations with reference energies below 15 kcal/mol. Reference 2B energies were calculated with ω B97M-V/aug-cc-pVTZ. ...	177
Figure 6.4.	Test set correlation plot for the Zundel-acetonitrile TTM-nrg PEF with configurations taken from the TTM-nrg MD simulations. Reference 2B energies were calculated with ω B97M-V/aug-cc-pVTZ.	178
Figure 6.5.	An example Zundel-acetonitrile dimer configuration from the TTM-nrg training set where both the Zundel ion and the acetonitrile molecule are in their (isolated monomer) global minimum energy conformations.	178
Figure 6.6.	A typical Zundel-acetonitrile dimer configuration in our TTM-nrg simulations of the Zundel ion dissolved in liquid acetonitrile.	179
Figure 6.7.	1B Zundel test set correlation plot for configurations taken from our TTM-nrg MD simulations.	179

Figure 6.8.	Internal coordinate basis vectors for Zundel ion H_F velocities and the approximate C_2 axis shown on a typical Zundel ion configuration from our simulations. The basis vectors are separated into: a) radial, b) latitudinal and c) longitudinal components.	183
Figure 6.9.	Free energy with respect to O–O distance and PTC for a) Z_{acn}^+ , b) Z_{gas}^+ (both symmetrized around PTC).	185
Figure 6.10.	Z^+ O–O distance free energy in acetonitrile vs in gas phase.	185
Figure 6.11.	Z_{acn}^+ and Z_{gas}^+ FE with respect to the PTC.	186
Figure 6.12.	Free energy along the Z^+ H_F -O–O angle for Z_{acn}^+ and Z_{gas}^+ , and free energy along the dihedral angle of the Z^+ flanking-H-O-H bisectors for Z_{acn}^+ and Z_{gas}^+	187
Figure 6.13.	N–O and N–H* RDFs.	189
Figure 6.14.	Example RADF schematic.	189
Figure 6.15.	Logarithmic scale radial-angular distribution functions: a) N–O–O, b) C_m –O–O, c) N–O– H_F , d) C_m –O– H_F RADFs.	190
Figure 6.16.	Example coordination of acetonitrile molecules to the Zundel ion from our simulations. Highlighted molecules shown in stick and ball representation, surrounding molecules shown as lines. Black dotted lines representing coordination shown for clarity.	191
Figure 6.17.	C_m –O and C_m –H* RDFs.	192
Figure 6.18.	Logarithmic scale radial-angular distribution functions: a) O–N– C_m RADF; b) O– C_m –N RADF.	194
Figure 6.19.	Logarithmic scale radial-angular distribution functions: a) N–H*– C_2 -axis; b) C_m –H*– C_2 -axis RADFs.	195
Figure 6.20.	Z^+ (in acetonitrile) IR spectra: comparing dipole moment models.	196
Figure 6.21.	Comparing Z^+ (in acetonitrile) and acetonitrile IR spectra to identify peak from induced part of Z^+ spectrum.	197
Figure 6.22.	Comparison of the DMS + induction IR spectra of Z_{acn}^+ with DMS spectrum of Z_{gas}^+	198

Figure 6.23.	Comparison of the DMS + induction IR spectra of Z_{acn}^+ with experimental IR spectrum. The experimental IR spectrum was taken from Fig. 1c in Ref. 2, reproduced with the WebPlotDigitizer.[3]	199
Figure 6.24.	Comparison of the total velocity power spectra of Z_{acn}^+ with Z_{gas}^+	200
Figure 6.25.	Decomposition of the Z_{acn}^+ and Z_{gas}^+ vibrational power spectra into a) H_F radial, b) H_F latitudinal, c) H_F longitudinal, and d) H^* modes. Each H_F mode type is further decomposed, assuming $\mathcal{C}_2$ point group symmetry of the Zundel ion, into A and B irreducible representations.	201
Figure 7.1.	Cartoon depicting the bonding connectivity of the two lowest energy diabatic states and the adiabatic ground state, as well as cartoon plots of the diabatic state energies and the adiabatic ground state energy.	208
Figure 7.2.	Example of the DS PEF extrapolation problem for the protonated water dimer on a rigid scan along the PTC.	210
Figure 7.3.	Logarithmic scale histogram of the ALMO-EDA CT energy with respect to the reference (adiabatic ground state) binding energy and the PTC for all configurations in the Zundel 1B training set	212
Figure 7.4.	Switching function plots (Eq. 7.2) for different values of a , b , and c parameters, and logarithmic scale histograms of the coupling potential, v_{12} with respect to the PTC for each nuclear configuration in the Zundel 1B training set.	215
Figure 7.5.	Example VB Hamiltonian for $H_3O^+(H_2O)_4$ cluster 5β (Fig. 3.15) in kcal/mol using Eq. 7.1 with switch parameters $a = 0.25$, $b = 0.0$, $c = 1.0$. Covalent bond connectivities for each DS are shown on the left, and dimers used in each coupling matrix element calculation are shown on the right.	217
Figure 7.6.	Heat map showing relative accuracy of MB-VB models compared to $\omega B97M-V/ aug-cc-pVTZ$ reference binding energies with various sets of switching function parameters evaluated on the protonated water clusters listed in Table 7.1.	219
Figure B.1.	Energy decompositions for each minimum energy $X^-(H_2O)$ dimer, with $X = F$ (a), Cl (b), Br (c), and I (d), with numerical energy values (in kcal/mol) shown on the right-hand side of each bar.	227
Figure B.2.	Energy decompositions for each minimum energy $M^+(H_2O)$, with $M = Li$ (a), Na (b), K (c), Rb (d), Cs (e), with numerical energy values (in kcal/mol) shown on the right-hand side of each bar.	229

Figure D.1.	Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_4$ cluster, 4β	234
Figure D.2.	Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_4$ cluster, 4δ	235
Figure D.3.	Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_4$ cluster, 4γ	235
Figure D.4.	Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_5$ cluster, 5β	236
Figure D.5.	Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_5$ cluster, 5δ	236
Figure D.6.	Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_5$ cluster, 5γ	237
Figure D.7.	Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_5$ cluster, 5ϵ	237
Figure D.8.	Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_6$ cluster, 6α	238
Figure D.9.	Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_6$ cluster, 6β	238
Figure D.10.	Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_6$ cluster, 6γ	239

LIST OF TABLES

Table 3.1.	Reference interaction energies (in kcal/mol) for the low-lying isomers of $\text{OH}^-(\text{H}_2\text{O})_n$ clusters shown in Figs. 3.8–3.4.	59
Table 3.2.	Statistics for the 2B energies (in kcal/mol) of 2-mer fragments. See main text for details.	65
Table 3.3.	Statistics for the 3B energies (in kcal/mol) of 3-mer fragments. See main text for details.	66
Table 3.4.	Statistics for ‘significant’ 4B energies (in kcal/mol) of 4-mer fragments. See main text for details. See Supporting Information of Ref. 4 for a complete analysis of all 4B energies.	69
Table 3.5.	Statistics for ‘significant’ 5B energies (in kcal/mol) of 5-mer fragments. See main text for details. See Supporting Information of Ref. 4 for a complete analysis of all 5B energies.	70
Table 3.6.	Reference interaction energies (in kcal/mol) for low-lying isomers of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters shown in Figs. 3.20 and 3.15.	76
Table 3.7.	Statistics for 2B energies (in kcal/mol) of 2-mer fragment groups from all $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ ($n \leq 5$) clusters (top) along with examples of the corresponding molecular configurations (bottom).	86
Table 3.8.	Statistics for 2B energies (in kcal/mol) of notable 2-mer fragment groups in context (top) along with examples of the corresponding molecular configurations (bottom).	87
Table 3.9.	Statistics for 3B energies (in kcal/mol) of significant 3-mer fragment groups from all $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ ($n \leq 5$) clusters (top) along with examples of the corresponding molecular configurations (bottom).	89
Table 3.10.	Statistics for 3B energies (in kcal/mol) of notable 3-mer fragment groups in context (top) along with examples of the corresponding molecular configurations (bottom).	90
Table 3.11.	Statistics for ‘Significant’ 4-Body Energies of 4-mer Fragments.	92
Table 4.1.	Halide–water 2B ALMO-EDA statistics of $\text{X}^-(\text{H}_2\text{O})$ dimers ($\text{X} = \text{F}, \text{Cl}$). Signed mean energy values are reported with standard deviations in parentheses (in kcal/mol).	116

Table 4.2.	Halide–water 2B ALMO-EDA statistics of $X^-(H_2O)$ dimers ($X = Br, I$). Signed mean energy values are reported with standard deviations in parentheses (in kcal/mol).	116
Table 4.3.	Alkali ion–water 2B ALMO-EDA statistics of $M^+(H_2O)$ dimers ($M = Li, Na$). Signed mean energy values are reported with standard deviations in parentheses (in kcal/mol).	120
Table 4.4.	Alkali ion–water 2B ALMO-EDA statistics of $M^+(H_2O)$ dimers ($M = K, Rb, Cs$). Signed mean energy values are reported with standard deviations in parentheses (in kcal/mol).	121
Table 5.1.	Example comparison between PIPs defined by different filters.	163
Table 7.1.	MB-VB binding energies (in kcal/mol) for some of the low-lying isomers of $H_3O^+(H_2O)_n$ clusters, and their deviations relative to $\omega B97M-V/aug-cc-pVTZ$ binding energies.	220

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Francesco Paesani for his support as the chair of my committee, and as my advisor and mentor.

I would also like to acknowledge Marc Riera for answering all of the questions, and helping me with a multitude of software challenges.

I would also like to acknowledge the entire Paesani lab for their help, support, and friendship. Thank you all!

Sections 3.2, 3.3, and 4.4 are almost entirely reprints of the material as they appear in “Assessing Many-Body Effects of Water Self-Ions. I: $\text{OH}^-(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2018**, 14 (4), 1982–1997, and “Assessing Many-Body Effects of Water Self-Ions. II: $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2019**, 15 (9), 4816–4833.

Sections 3.4, 4.2, 4.3, and 5.2 are almost entirely reprints of the material as they appear in “Nature of Halide–Water Interactions: Insights from Many-Body Representations and Density Functional Theory” *J. Chem. Theory Comput.*, **2019**, 15 (5), 2983–2995 and “Nature of Alkali–Water Interactions: Insights from Many-Body Representations and Density Functional Theory II” *J. Chem. Theory Comput.*, **2020**, 16 (5), 3055–3072.

Appendices A, B, and C are almost entirely reprints of the material as they appear in “Assessing Many-Body Effects of Water Self-Ions. I: $\text{OH}^-(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2018**, 14 (4), 1982–1997, “Assessing Many-Body Effects of Water Self-Ions. II: $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2019**, 15 (9), 4816–4833, “Nature of Halide–Water Interactions: Insights from Many-Body Representations and Density Functional Theory” *J. Chem. Theory Comput.*, **2019**, 15 (5), 2983–2995 and “Nature of Alkali–Water Interactions: Insights from Many-Body Representations and Density Functional Theory II” *J. Chem. Theory Comput.*, **2020**, 16 (5), 3055–3072.

VITA

- 2015 Bachelor of Arts, Northwestern University
- 2015–2016 Teaching Assistant, Department of Chemistry
University of California San Diego
- 2018 Teaching Assistant, Department of Chemistry
University of California San Diego
- 2023 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

- “Assessing Many-Body Effects of Water Self-Ions. I: $\text{OH}^-(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2018**, 14 (4), 1982–1997
- “Nature of Halide–Water Interactions: Insights from Many-Body Representations and Density Functional Theory” *J. Chem. Theory Comput.*, **2019**, 15 (5), 2983–2995
- “Assessing Many-Body Effects of Water Self-Ions. II: $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2019**, 15 (9), 4816–4833
- “Nature of Alkali–Water Interactions: Insights from Many-Body Representations and Density Functional Theory II” *J. Chem. Theory Comput.*, **2020**, 16 (5), 3055–3072

ABSTRACT OF THE DISSERTATION

Decompose Everything! Exploring the Fundamental Components of Many-Body Chemical Systems

by

Colin Kerry Egan

Doctor of Philosophy in Chemistry

University of California San Diego, 2023

Professor Francesco Paesani, Chair

Chemical systems are inherently difficult to model due to their many-body nature. Difficulties associated with the large number of degrees of freedom are further complicated by the need for a quantum mechanical treatment of the electrons (and the nuclei in many circumstances). Here we employ various types of decompositions in order to dissect the interactions and dynamics in chemical systems, thus facilitating the development of models that can be used in molecular dynamics simulations, as well as facilitating the analysis of the resulting simulations. In particular, the many-body expansion is used to study the intermolecular interactions in protonated and deprotonated water clusters, as well as alkali ion-water clusters. The many-body energies of

protonated and deprotonated water clusters are decomposed into contributions due to so-called fragment groups which are defined by their hydrogen bonding connectivity. Fragment group energies are further decomposed using energy decomposition analysis to investigate the physical interactions that lead to cooperative and anticooperative hydrogen bonding. Additionally, the concept of a fragment group is extended in order to derive scaling relations for many-body interactions in lattices. These investigations of many-body interactions are used in the development and evaluation of model potential energy functions describing intermolecular interactions. Molecular dynamics simulations of the Zundel ion dissolved in liquid acetonitrile, using these many-body potential energy functions, are analyzed with the help of decompositions of the vibrational power spectra. First steps in the development of many-body valence bond models describing aqueous proton transfer are also discussed.

Chapter 1

Introduction

Modeling the behavior of chemical systems is an incredibly difficult task, in general. Beyond the very real complications associated with the need for a quantum mechanical description of subatomic particles, we must grapple with many-body systems of mutually interacting particles (the “many bodies” in this statement being the electrons and nuclei composing a given chemical system). Isaac Newton famously explained in a correspondence with astronomer Edmund Halley that the dynamics of the Earth, Moon, and Sun (a mere three-body system) made his head ache like nothing else, and that the problem kept him awake so often that he stopped thinking about it.[5] It was later proven that the general N-body problem has no closed-form analytical solution, which essentially means that approximate numerical methods must be employed to have any chance of approaching such systems theoretically. The many-body nature of chemical systems is the elephant in every room of chemistry, so it is best to proceed with cautious bravery.

Original contributions include Sections 3.2, 3.3, 3.4, 3.5, 4.2, 4.3, 4.4, 5.1.6, 5.1.7, 5.1.9, 5.1.10, 5.2, Chapter 6, and Chapter 7.

Chapter 2

Electronic Structure Theory Background

Chemical systems are composed of negatively charged electrons and positively charged nuclei, which collectively interact via Coulomb’s law (in natural units):

$$V_{\text{electrostatic}} = \frac{q_1 q_2}{r} \quad (2.1)$$

where q_1 and q_2 are the charges of the two interacting particles (Coulomb’s law is a pairwise interaction), and r is the distance between the two particles. Given the small masses of the relevant particles (in particular the electrons), chemical system are inherently quantum mechanical in nature. We will later see that models based on classical mechanics can offer accurate representations of certain classes of chemical systems under particular circumstances, but ultimately, these models must implicitly account for quantum mechanical effects through fits to quantum chemical calculations, or fits to experimental data.

As in any quantum system, the total energy of a chemical system is described by a *Hamiltonian* operator, usually called the “molecular Hamiltonian”, $\hat{H}_{\text{mol}}$, defined by:[6]

$$\begin{aligned} \hat{H}_{\text{mol}} &= \hat{T}_{\text{nuc}} + \hat{T}_{\text{el}} + \hat{V}_{\text{nuc-el}} + \hat{V}_{\text{el-el}} + \hat{V}_{\text{nuc-nuc}} \\ &= - \sum_{A=1}^{N_{\text{nuc}}} \frac{\hbar^2}{2M_A} \nabla_A^2 - \sum_{a=1}^{N_{\text{el}}} \frac{\hbar^2}{2} \nabla_a^2 - \sum_{A=1}^{N_{\text{nuc}}} \sum_{a=1}^{N_{\text{el}}} \frac{Z_A}{\|\mathbf{R}_A - \mathbf{r}_a\|_2} \\ &\quad + \sum_{A=1}^{N_{\text{nuc}}-1} \sum_{B>A}^{N_{\text{nuc}}} \frac{Z_A Z_B}{\|\mathbf{R}_A - \mathbf{R}_B\|_2} + \sum_{a=1}^{N_{\text{el}}-1} \sum_{b>a}^{N_{\text{el}}} \frac{1}{\|\mathbf{r}_a - \mathbf{r}_b\|_2} \end{aligned} \quad (2.2)$$

where $\hat{T}_{\text{nuc}}$ and $\hat{T}_{\text{el}}$ are the nuclear and electronic kinetic energy operators, $\hat{V}_{\text{nuc-el}}$, $\hat{V}_{\text{el-el}}$, and $\hat{V}_{\text{nuc-nuc}}$ are the potential energy operators describing the electron-nucleus, electron-electron, and nucleus-nucleus potential energies, respectively, N_{nuc} and N_{el} are the number of nuclei and the number of electrons. Capital letters (including indices, A and B) denote nuclear properties (in natural units), M_A , Z_A , $\mathbf{R}_A$, and ∇_A , which represent the mass, charge, position vector, and the gradient with respect to the position of the nucleus indexed by A . Lower case letters, a and b , index the electrons, and $\mathbf{r}_a$ and ∇_a denote the position vector and gradient with respect to the position of electron, a . $\|\mathbf{v}\|_2 = (\mathbf{v} \cdot \mathbf{v})^{\frac{1}{2}}$ is the Euclidian norm of vector, $\mathbf{v}$. The three potential energy operators are due to Coulomb’s law (Eq. 2.1) with the charge of electron in natural units being -1 , and the charge of nucleus, A in natural units being equal to its atomic number, Z_A . Magnetic interactions are present, but are usually small.[7] As a side note, relativistic effects are important in many chemical systems, but are not discussed here.¹

The stationary states of the system—each described by a molecular *wave function*, $\Psi(\mathbf{r}, \mathbf{R})$, which depends on both the electronic positions, $\mathbf{r}$, and the nuclear positions (in the position basis)—can be obtained by solving the corresponding *Time-Independent Schrödinger Equation* (TISE):

$$\hat{H}_{\text{mol}}\Psi(\mathbf{r}, \mathbf{R}) = \mathcal{E}\Psi(\mathbf{r}, \mathbf{R}) \quad (2.3)$$

where $\mathcal{E}$ is the total energy of stationary state, $\Psi(\mathbf{r}, \mathbf{R})$. This equation is extremely difficult to solve in general, having $3(N_{\text{nuc}} + N_{\text{el}})$ coupled degrees of freedom. We will need to be clever to have any chance of solving this equation.

Observing that nuclei are much more massive than electrons, we can imagine that the electrons respond instantaneously to any change in the nuclear configuration. From this perspective, we can “clamp” the nuclei at any given configuration, $\mathbf{R}$, and solve for the stationary states of the electrons, $\Phi_n(\mathbf{r}; \mathbf{R})$ (indexed by n), for that particular nuclear configuration with the

¹However the effective core potentials used in electronic structure calculations involving Br^- , I^- , K^+ , Rb^+ , and Cs^+ discussed later do account for relativistic effects (Secs. A.2 and A.3).

electronic TISE:

$$\hat{H}_{\text{el}}(\mathbf{R})\Phi_n(\mathbf{r};\mathbf{R}) = E_n(\mathbf{R})\Phi_n(\mathbf{r};\mathbf{R}) \quad (2.4)$$

where $E_n(\mathbf{R})$ is the total electronic energy of the n^{th} electronic stationary state, $\Phi_n(\mathbf{r};\mathbf{R})$, at the fixed nuclear configuration, $\mathbf{R}$, and $\hat{H}_{\text{el}}(\mathbf{R})$ is the electronic Hamiltonian for $\mathbf{R}$, defined by:

$$\begin{aligned} \hat{H}_{\text{el}} &= \hat{T}_{\text{el}} + \hat{V}_{\text{nuc-el}} + \hat{V}_{\text{el-el}} \\ &= -\sum_{a=1}^{N_{\text{el}}} \frac{\hbar^2}{2} \nabla_a^2 - \sum_{A=1}^{N_{\text{nuc}}} \sum_{a=1}^{N_{\text{el}}} \frac{Z_A}{\|\mathbf{R}_A - \mathbf{r}_a\|_2} + \sum_{a=1}^{N_{\text{el}}-1} \sum_{b>a}^{N_{\text{el}}} \frac{1}{\|\mathbf{r}_a - \mathbf{r}_b\|_2} \end{aligned} \quad (2.5)$$

Since $\hat{H}_{\text{el}}$ is Hermitian, its eigenvectors, the electronic wave functions, $\Phi_n(\mathbf{r};\mathbf{R})$, (in principle) form a complete (orthogonal) basis for the electronic Hilbert space, $\mathcal{H}_{\text{el}}$, for every nuclear configuration. In practice we must settle for a finite *basis set* approximation, given that the dimension of the Hilbert space is infinite (but separable). Therefore, we can expand the full molecular wave function, $\Psi(\mathbf{r},\mathbf{R})$, as a linear combination of electronic stationary states:

$$\Psi(\mathbf{r},\mathbf{R}) = \sum_n \chi_n(\mathbf{R})\Phi_n(\mathbf{r};\mathbf{R}) \quad (2.6)$$

where the expansion coefficients, $\chi_n(\mathbf{R})$, only depend on the nuclear positions. It turns out that $\chi_n(\mathbf{R})$ can be thought of as the nuclear wave functions.[8]

Plugging the expansion of $\Psi(\mathbf{r},\mathbf{R})$ from Eq. 2.6 into the molecular TISE (Eqs. 2.2 and 2.3), and using the fact that the electronic Hamiltonian, H_{el} doesn't act on $\chi_n(\mathbf{R})$, and the fact that $\hat{V}_{\text{nuc-nuc}}$ is a constant, we obtain:

$$\begin{aligned} \sum_n \hat{T}_{\text{nuc}} [\chi_n(\mathbf{R})\Phi_n(\mathbf{r};\mathbf{R})] + \chi_n(\mathbf{R})\hat{H}_{\text{el}}\Phi_n(\mathbf{r};\mathbf{R}) \\ + \chi_n(\mathbf{R})\Phi_n(\mathbf{r};\mathbf{R})\hat{V}_{\text{nuc-nuc}} = \mathcal{E}\chi_n(\mathbf{R})\Phi_n(\mathbf{r};\mathbf{R}) \end{aligned} \quad (2.7)$$

It is important to remember that while the electronic Hamiltonian does not act on the nuclear wave functions, the nuclear kinetic energy, $\hat{T}_{\text{nuc}}$, does act on the electronic wave functions

because they are (parametrically) dependent on the nuclear positions).

In order to obtain a given nuclear wave functions, indexed by m , we multiply Eq. 2.7 by the complex conjugate of the corresponding electronic wave function, $\Phi_m(\mathbf{r}; \mathbf{R})$, and integrate over the electronic positions, $\mathbf{r}$. After using Eq. 2.4, and defining the *nonadiabaticity operator*, Θ_{mn} , as:[8]

$$\Theta_{mn} = \int \left[\Phi_m^*(\mathbf{r}; \mathbf{R}) \hat{T}_{\text{nuc}} \Phi_n(\mathbf{r}; \mathbf{R}) + \Phi_m^*(\mathbf{r}; \mathbf{R}) \sum_A^{N_{\text{nuc}}} \frac{-\hbar^2}{M_A} \nabla_A \Phi_n(\mathbf{r}; \mathbf{R}) \nabla_A \right] d\mathbf{r} \quad (2.8)$$

and after skipping some steps, we finally obtain the following coupled set of equations for the nuclear wave functions:[8]

$$[\hat{T}_{\text{nuc}} + E_n(\mathbf{R}) + V_{\text{nuc-nuc}}(\mathbf{R}) - \mathcal{E}] \chi_n(\mathbf{R}) = - \sum_{m,n} \Theta_{mn} \chi_n(\mathbf{R}) \quad (2.9)$$

Everything we have done so far has been exact, so the system of equations in Eq. 2.9 are just as hard to solve as Eq. 2.3. The problematic part in Eq. 2.9 is the nonadiabaticity operator, which basically contains all of the terms that we would rather not deal with, dynamically² coupling the different electronic states, $\Phi_n(\mathbf{r}; \mathbf{R})$. In the next section, we (responsibly) throw away the nonadiabaticity operator.

2.1 The Born-Oppenheimer Approximation

The Born-Oppenheimer approximation (BOA) is:

$$\Theta_{ab} \xrightarrow{BOA} 0 \quad (2.10)$$

This simplifies matters considerably. Within the BOA, we can rewrite Eq. 2.9 as:

$$[\hat{T}_{\text{nuc}} + E_n(\mathbf{R}) + V_{\text{nuc-nuc}}(\mathbf{R})] \chi_n(\mathbf{R}) = \mathcal{E} \chi_n(\mathbf{R}) \quad (2.11)$$

²the coupling is dynamic because of the gradient operators.[8]

We can interpret Eq. 2.11 as being the nuclear TISE, identifying the last two terms on the left-hand side as the potential energy operator for the nuclei:

$$E_{\text{PES}}(\mathbf{R}) = E_n(\mathbf{R}) + V_{\text{nuc-nuc}}(\mathbf{R}) \quad (2.12)$$

$E_{\text{PES}}(\mathbf{R})$ is a scalar-valued function of the nuclear positions, and is referred to as the *Born-Oppenheimer potential energy surface* (PES). This is precisely the “electronic energy” calculated in typical electronic structure calculations. The BOA is also called “the adiabatic approximation,”[8] and the resulting electronic states, $\Phi_n(\mathbf{r}; \mathbf{R})$ are referred to as the adiabatic electronic wave functions.

The electronic states are no longer coupled to each other after making the BOA, so the stationary states are simply the product states, $\chi_n(\mathbf{R})\Phi_n(\mathbf{r}; \mathbf{R})$, indexed by n . In other words, the BOA is essentially the approximation that the molecular TISE (Eq. 2.3) is separable with respect to the electronic vs nuclear coordinates.

It’s awfully convenient to be able to just throw away terms that are causing us problems, but there is always a price to pay for such a maneuver, namely that the approximation fails in a wide array of circumstances. In particular, the BOA fails when the electronic energies of two or more adiabatic states get too close. These adiabatic states are no longer good approximate wave functions because we have thrown away the coupling between degenerate, or nearly degenerate reference (basis) states.

2.2 Wave Function Methods

2.2.1 One-Electron Orbitals

For $N = N_{\text{el}}$ electrons, the many-electron wave function, $\Phi(\mathbf{r}; \mathbf{R})$, depends on the positions of each electron, so the corresponding many-electron Hilbert space, $\mathcal{H}_{\text{el}}^{(N)}$, can be thought of as the tensor product of N 1-electron Hilbert spaces, $\mathcal{H}_{\text{el}}^{(N)} = \mathcal{H}_{\text{el}}^{(1)} \otimes \mathcal{H}_{\text{el}}^{(1)} \otimes \cdots \otimes \mathcal{H}_{\text{el}}^{(1)}$. From this perspective, it makes sense to write the many-electron wave function in terms of

1-electron wave functions (elements of $\mathcal{H}_{\text{el}}^{(1)}$) called *orbitals*. The orbital for the a^{th} electron, $\psi(\mathbf{r}_a, \omega_a)$, depends on the position in 3 spatial dimensions, $\mathbf{r}_a$, as well as the spin coordinate, ω_a of the a^{th} electron. In the absence of external fields, ω_a is simply another degree of freedom for electronic wave functions. We define $\psi(\mathbf{r}_a, \omega_a)$ to be the product of a spatial orbital and a spin function:

$$\begin{aligned}\psi(\mathbf{r}_a, \omega_a) &= \psi(\mathbf{x}_a) = \phi(\mathbf{r}_a)\Omega(\omega_a) \\ \Omega(\omega_a) &= \begin{cases} \alpha(\omega_a) & \text{spin up} \\ \beta(\omega_a) & \text{spin down} \end{cases}\end{aligned}\tag{2.13}$$

where $\phi(\mathbf{r}_a)$ is the spatial orbital describing the position of the a^{th} electron (for example, the spherical harmonics describing the positions of the hydrogen atom electron). $\Omega(\omega_a)$ is a placeholder for the spin function, which is either $\alpha(\omega_a)$ or $\beta(\omega_a)$ depending on whether the a^{th} electron is spin up ($\uparrow$) or spin down ($\downarrow$), respectively. Additionally, we define $\mathbf{x}_a = (\mathbf{r}_a, \omega_a)$. $\psi(\mathbf{r}_a, \omega_a) = \psi(\mathbf{x}_a)$ is often called a “spin orbital” to distinguish it from the corresponding spatial orbital.[6]

2.2.2 Basis Set Approximations

The 1-electron Hilbert space, $\mathcal{H}_{\text{el}}^{(1)}$, is infinite dimensional, but separable, so any arbitrary orbital can be expressed as a linear combination of some countable basis for $\mathcal{H}_{\text{el}}^{(1)}$. In practice, we use a finite basis set approximation, which, for practical reasons, usually takes the form of atom-centered gaussians³ (multiplied by spherical harmonics which determine the angular momentum of the orbital), or plane waves (typically used for periodic systems), or some combination of the two. For some basis set approximation (for the 1-electron Hilbert space) composed of K functions, $\{\varphi_1, \varphi_2, \dots, \varphi_K\}$, any spatial orbital, ϕ_n , lying within the span of the basis can be

³Gaussian functions are a convenient choice because the product of two gaussians is another gaussian function, and the Fourier transform of a gaussian is another gaussian.[6]

written as:

$$\phi_n = \sum_{m=1}^K c_{mn} \varphi_m \quad (2.14)$$

where c_{mn} are the expansion coefficients for the linear combination, which are complex numbers in general, but can be taken to be real in most calculations. Note that the matrix representation of the 1-electron Hamiltonian in this basis (approximation) will have a spectrum of K eigenvalues, so in addition to the single occupied orbital (the eigenvector with the smallest eigenvalue), there will be $K - 1$ unoccupied (excited state) orbitals. The accuracy of any particular electronic structure method is limited by the accuracy of this basis set approximation. For example, the Full Configuration Interaction (CI) method, discussed in Sec. 2.2.7, yields the exact, nonrelativistic electronic energies (for both the ground and excited states) for any many-electron system, within the basis set approximation used. If one were to use a small basis set for this kind of calculation, the resulting energies would likely be inaccurate, despite the exactness of the full CI method, due to a poor approximation of the true (countably infinite) basis of the N -electron Hilbert space, $\mathcal{H}_{\text{el}}^{(N)}$. If one were to use a sufficiently large basis set approximation, then the resulting energies would be close to the exact values. If one were to use an even bigger basis set approximation, the resulting energies may improve slightly (if the previous basis set were large enough, this improvement might fall within the floating point precision of the calculation), but at a far greater cost.⁴ The ground electronic state energy calculated with full CI is an upper-bound to the exact ground state energy, so improvements to the basis set approximation will allow the electrons to relax to lower energy, reducing the upper-bound, approaching the exact value.

Thinking about electronic Hilbert spaces can be a little abstract, so a more terrestrial analogy can be helpful. A given basis set approximation is analogous to a house for the electrons, with the size, K , of the basis set roughly corresponding to the size of the house. A small basis set is like a studio apartment that is barely large enough to hold all of the electrons, and not nearly large enough for them to unpack their belongings and relax for the night, while an excessively

⁴It is also worth noting that excessively large atom-centered basis sets may lead to numerical instabilities in the calculations due to near linear dependence between basis functions.

large basis set is analogous to a 20 bedroom mansion that is big enough for the electrons to relax, with 15 extra bedrooms that will likely remain empty and unused (even though they paid for all 20), and an appropriately sized basis set is something in between. In practice, the choice of a basis set approximation can be quite important, but for the remainder of the chapter, we assume the use of an appropriately sized basis set. Details about the basis set approximations used for the calculations presented in later chapters are given in Appendix A.

2.2.3 Inner Products and Bra-Ket Notation

By definition, each Hilbert space has some associated inner product, which is analogous to the dot product in Cartesian vector spaces. Using bra-ket notation, an inner product between two elements of some Hilbert space, $w, v \in \mathcal{H}$, is written as $\langle w|v\rangle$, with $\langle w|$ being the bra, and $|v\rangle$ being the ket. In general, an inner product is a sesquilinear map onto the scalar field, $\mathbb{F}$ underlying the Hilbert space (Hilbert spaces over the complex numbers, $\mathbb{F} = \mathbb{C}$, are used in quantum mechanics) i.e. $\langle \cdot | \cdot \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{F}$. In bra-ket notation, the sesquilinearity of the inner product corresponds to conjugate-linearity in the bra, and linearity in the ket.⁵ Explicitly, for $v_1, v_2, v_3, v_4 \in \mathcal{H}$ and $a, b, c, d \in \mathbb{C}$, bilinearity leads to:

$$\langle av_1 + bv_2 | cv_3 + dv_4 \rangle = a^*c \langle v_1 | v_3 \rangle + a^*d \langle v_1 | v_4 \rangle + b^*c \langle v_2 | v_3 \rangle + b^*d \langle v_2 | v_4 \rangle \quad (2.15)$$

where a^* and b^* are the complex conjugates of a and b , respectively.

To be more precise, in bra-ket notation, the elements of the Hilbert space are denoted by kets, $|v\rangle \in \mathcal{H}$, and elements of the corresponding dual space, $\mathcal{H}^*$, often called “covectors,” are denoted by bras, $\langle v| \in \mathcal{H}^*$. Covectors in the dual space of a given Hilbert space are linear *functionals*, l , that map elements of the Hilbert space to the field, i.e. $l(\cdot) : \mathcal{H} \rightarrow \mathbb{F}$. The Riesz representation theorem states that given some Hilbert space, $\mathcal{H}$, for every linear functional in the corresponding dual space, $l \in \mathcal{H}^*$, there exists a unique element in the Hilbert

⁵Confusingly, this is reversed from the standard mathematical notation for inner products, with $\langle x, y \rangle = \langle y | x \rangle$. In mathematical notation, the inner product is linear in the left-argument, and conjugate-linear in the right-argument.

space, $v_l = |v_l\rangle \in \mathcal{H}$, such that $l(w) = \langle v_l | w \rangle \in \mathbb{F}$ for all $w = |w\rangle \in \mathcal{H}$, where $\langle v_l | w \rangle$ is the inner product between $v_l, w \in \mathcal{H}$. [9] This creates a “direct connection” between unique pairs of bras and kets (elements of $\mathcal{H}^*$ and $\mathcal{H}$) with $\langle v | \leftrightarrow |v\rangle$. Additionally, the Riesz representation theorem guarantees that for each linear operator, $\hat{A}$ (with $\hat{A} : \mathcal{H} \rightarrow \mathcal{H}$), there exists a unique operator $\hat{A}^\dagger$ called the *adjoint* of $\hat{A}$ (with⁶ $\hat{A}^\dagger : \mathcal{H} \rightarrow \mathcal{H}$) such that $\langle v | \hat{A} v \rangle = \langle v | \hat{A} | v \rangle = \langle \hat{A}^\dagger v | v \rangle$.

In the case of $\mathcal{H}_{\text{el}}^{(1)}$, the relevant inner product is integration over the position of the electron (over all of $\mathbb{R}^3$), as well as integration over the spin coordinate of the electron, ω . For $\psi_i(\mathbf{x}) = |\psi_i\rangle, \psi_j(\mathbf{x}) = |\psi_j\rangle \in \mathcal{H}_{\text{el}}^{(1)}$ (and $\langle \psi_i | \in \mathcal{H}_{\text{el}}^{(1)*}$) the corresponding inner product is:

$$\langle \psi_i | \psi_j \rangle = \int \psi_i^*(\mathbf{r}, \omega) \psi_j(\mathbf{r}, \omega) d\mathbf{r} d\omega = \int \psi_i^*(\mathbf{x}) \psi_j(\mathbf{x}) d\mathbf{x} \quad (2.16)$$

Additionally, for some N -electron Hilbert space, $\mathcal{H}_{\text{el}}^{(N)}$, the inner product corresponds to an integral over the positions and spin coordinates of all N electrons. It is important to keep in mind that bra covectors, $\langle \Phi_i | \in \mathcal{H}_{\text{el}}^*$, represent linear functionals corresponding to multiplication by the complex conjugate of $|\Phi_i\rangle \in \mathcal{H}$ and integration over all position and spin coordinates, so one might write $\langle \Phi_i | = \int \Phi_i^*(\mathbf{x}) \cdot d\mathbf{x}$, where the dot, $\cdot$, indicates the input for the functional.

The 1-electron spatial orbitals and spin functions have their own separate inner products as well, corresponding to integration over positions, and integration over spin coordinates, respectively. For now, we can assume that the spatial orbitals form an orthonormal set, such that the inner product between distinct orbitals yields 0 (orthogonality), and the inner product between identical orbitals yields 1 (normalization):

$$\langle \phi_i | \phi_j \rangle = \int \phi_i^*(\mathbf{r}_a) \phi_j(\mathbf{r}_a) d\mathbf{r}_a = \delta_{ij} \quad (2.17)$$

where δ_{ij} is the Kronecker delta which equals 1 when $i = j$, and equals 0 when $i \neq j$. Note that Eq. 2.17 is a 3-dimensional integral over $\mathbb{R}^3$. The spatial functions are also taken to be

⁶For a general linear map from vector space V to vector space W , $\hat{A} : V \rightarrow W$ (not necessarily an operator), the adjoint, $\hat{A}^\dagger : W \rightarrow V$, maps the codomain of $\hat{A}$ to the domain of $\hat{A}$.

orthonormal:

$$\begin{aligned}
\langle \alpha | \alpha \rangle &= \int \alpha^*(\omega_a) \alpha(\omega_a) d\omega_a = 1 \\
\langle \beta | \beta \rangle &= \int \beta^*(\omega_a) \beta(\omega_a) d\omega_a = 1 \\
\langle \alpha | \beta \rangle &= \int \alpha^*(\omega_a) \beta(\omega_a) d\omega_a = 0
\end{aligned} \tag{2.18}$$

It follows from Eqs. 2.17 and 2.18 that the spin orbitals, $\psi(\mathbf{r}_a, \omega_a)$, also form an orthonormal set.

In order to solve the TISE numerically, we will need to represent the electronic Hamiltonian in some basis set approximation. For a given operator, $\hat{O}$, matrix elements, O_{ij} , for the matrix representation of the operator in the basis, $\{ |v_1\rangle, |v_2\rangle, \dots \}$, are calculated by sandwiching the operator between each bra and ket, such that $O_{ij} = \langle v_i | \hat{O} | v_j \rangle$. In the case of a 1-electron Hamiltonian, $\hat{H}$, in the basis of spin orbitals, $\{ |\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_K\rangle \}$, we need to calculate the following integrals:

$$H_{ij} = \langle v_i | \hat{H} | v_j \rangle = \int \psi_i^*(\mathbf{x}) \hat{H} \psi_j(\mathbf{x}) d\mathbf{x} \tag{2.19}$$

2.2.4 Many-Electron Wave Functions

Motivated by our use of 1-electron (spin) orbitals, we can write the electronic Hamiltonian (Eq. 2.5) as a sum of a 1-electron operator, $\hat{h}$, and a 2-electron operator, $\hat{v}$, such that $H_{\text{el}} = \hat{h} + \hat{v}$. The 1-electron operator includes the kinetic energy of each electron, and the electron-nucleus potential energy for each electron:

$$\hat{h} = \sum_{a=1}^N \hat{h}_a = \sum_{a=1}^N \left(-\frac{\hbar^2}{2} \nabla_a^2 - \sum_{A=1}^{N_{\text{nuc}}} \frac{Z_A}{\|\mathbf{R}_A - \mathbf{r}_a\|_2} \right) \tag{2.20}$$

The 2-electron operator represents the electron-electron potential energy between each pair of electrons:

$$\begin{aligned}\hat{v} &= \sum_{a=1}^{N-1} \sum_{b>a}^N \hat{v}_{ab} = \sum_{a=1}^{N-1} \sum_{b>a}^N \frac{1}{\|\mathbf{r}_a - \mathbf{r}_b\|_2} \\ &= \sum_{a=1}^{N-1} \sum_{b>a}^N r_{ab}^{-1}\end{aligned}\tag{2.21}$$

defining r_{ab} to be the distance between electrons a and b .

Discarding the 2-electron operator in Eq. 2.21 from the electronic Hamiltonian, we are left with the 1-electron Hamiltonian (Eq. 2.20), which describes the energy of N noninteracting electrons, since $\hat{v}$ contains all of the coupling interactions between the electrons. Within this noninteracting approximation, the electronic TISE become separable, and can be solved with a many-electron wave function constructed from a product of 1-electron spin orbitals:

$$\Phi(\mathbf{r}; \mathbf{R}) = \Phi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \mathbf{R}) = \psi_i(\mathbf{x}_1; \mathbf{R}) \psi_j(\mathbf{x}_2; \mathbf{R}) \cdots \psi_z(\mathbf{x}_N; \mathbf{R})\tag{2.22}$$

Eq. 2.22 is called the ‘‘Hartree product,’’ and corresponds to an uncorrelated many-electron wave function. This can be seen from the probabilistic interpretation of the wave function (i.e. the Born rule), where the probability of (simultaneously) finding the first electron within ‘‘volume’’ element, $d\mathbf{x}_1 = d\mathbf{r}_1 d\omega_1$, centered at $\mathbf{x}_1$, the second electron within volume element, $d\mathbf{x}_2$, centered at $\mathbf{x}_2$, etc.. is computed with the square-modulus of the position-basis wave function multiplied by the corresponding volume element:

$$\begin{aligned}p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \mathbf{R}) d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \\ &= \Phi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \mathbf{R}) \Phi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \mathbf{R}) d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \\ &= |\Phi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \mathbf{R})|^2 d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N\end{aligned}\tag{2.23}$$

where $p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \mathbf{R})$ is the probability density function for the positions of the N electrons.

When $\Phi(\mathbf{r}; \mathbf{R})$ takes the form of a Hartree product, we have:

$$\begin{aligned}
& p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \mathbf{R}) d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \\
&= |\psi_i(\mathbf{x}_1; \mathbf{R}) \psi_j(\mathbf{x}_2; \mathbf{R}) \cdots \psi_z(\mathbf{x}_N; \mathbf{R})|^2 d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \\
&= |\psi_a(\mathbf{x}_1; \mathbf{R})|^2 d\mathbf{x}_1 |\psi_b(\mathbf{x}_2; \mathbf{R})|^2 d\mathbf{x}_2 \cdots |\psi_z(\mathbf{x}_N; \mathbf{R})|^2 d\mathbf{x}_N
\end{aligned} \tag{2.24}$$

which is equivalent to the product of finding each electron within its own volume element independently of the other electrons, and therefore corresponds to an uncorrelated wave function.

The Hartree product lacks a number of important features of a true many-electron wave function, the most basic of which is the indistinguishability of the electrons. This indistinguishability means, for example, that it is equally likely to find the first electron in the ψ_i orbital as it is to find the second electron in ψ_i , etc.. Therefore, we make use of the permutation operator, $\mathcal{P}_{\mathbf{p}}$, which acts on a single Hartree product, and permutes the electronic coordinates between the orbitals according to $\mathbf{p} \in S_N$, which is an element of the permutation group⁷ of degree N , S_N . For example, if $\mathbf{p} = [3, 1, 2]$, then $\mathcal{P}_{\mathbf{p}}\{\psi_i(\mathbf{r}_1; \mathbf{R}) \psi_j(\mathbf{r}_2; \mathbf{R}) \psi_k(\mathbf{r}_3; \mathbf{R})\} = \psi_i(\mathbf{r}_3; \mathbf{R}) \psi_j(\mathbf{r}_1; \mathbf{R}) \psi_k(\mathbf{r}_2; \mathbf{R})$

With $\mathcal{P}_{\mathbf{p}}$, we can enforce indistinguishability by writing the many-electron wave function as a linear combination of every possible permutation of electrons for a given Hartree product, such that the sum of the square modulus of the expansion coefficients equals 1 (for normalization).

$$\begin{aligned}
\Phi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \mathbf{R}) &= \sum_{\mathbf{p} \in S_N} c_{\mathbf{p}} \mathcal{P}_{\mathbf{p}}\{\psi_i(\mathbf{x}_1; \mathbf{R}) \psi_j(\mathbf{x}_2; \mathbf{R}) \cdots \psi_z(\mathbf{x}_N; \mathbf{R})\} \\
\sum_{\mathbf{p} \in S_N} |c_{\mathbf{p}}|^2 &= 1
\end{aligned} \tag{2.25}$$

For N electrons, there are $N!$ possible permutations, so $|c_{\mathbf{p}}|^2 = (N!)^{-1}$ for all $\mathbf{p}$.

In addition to the indistinguishability of electrons, our many-electron wave function must be antisymmetric with respect to pairwise exchange of electronic (in general: fermionic)

⁷Usually called the “symmetric group.”

coordinates:

$$\Phi(\mathbf{x}_1, \dots, \mathbf{x}_a, \dots, \mathbf{x}_b, \dots, \mathbf{x}_N; \mathbf{R}) = -\Phi(\mathbf{x}_1, \dots, \mathbf{x}_b, \dots, \mathbf{x}_a, \dots, \mathbf{x}_N; \mathbf{R}) \quad (2.26)$$

Notice that the *Pauli exclusion principle*, i.e. the statement that two identical fermions cannot occupy the same quantum state simultaneously, is a consequence of the antisymmetry of fermionic wave functions since setting $\mathbf{x}_a = \mathbf{x}_b$ results in the same function on both sides of Eq. 2.26, which cannot be satisfied by any nonzero wave function. Antisymmetry can be enforced through the expansion coefficients in Eq. 2.25 by setting $c_{\mathbf{p}'} = -c_{\mathbf{p}''}$ for all pairs of permutations, $\mathbf{p}'$ and $\mathbf{p}''$, which differ from each other by an odd permutation (i.e. an odd number of pairwise exchanges of electronic coordinates).

Conveniently, both the indistinguishability and antisymmetry are satisfied by a *Slater determinant*, which is the determinant of a matrix whose rows each correspond to the coordinates of a single electron, and whose columns each correspond to a single spin orbital:

$$\begin{aligned} |\Phi\rangle &= |\psi_1 \psi_2 \cdots \psi_N\rangle \\ &= \Phi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \mathbf{R}) \\ &= \frac{1}{(N!)^{\frac{1}{2}}} \det \begin{vmatrix} \psi_i(\mathbf{x}_1; \mathbf{R}) & \psi_i(\mathbf{x}_2; \mathbf{R}) & \cdots & \psi_i(\mathbf{x}_N; \mathbf{R}) \\ \psi_j(\mathbf{x}_1; \mathbf{R}) & \psi_j(\mathbf{x}_2; \mathbf{R}) & \cdots & \psi_j(\mathbf{x}_N; \mathbf{R}) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_z(\mathbf{x}_1; \mathbf{R}) & \psi_z(\mathbf{x}_2; \mathbf{R}) & \cdots & \psi_z(\mathbf{x}_N; \mathbf{R}) \end{vmatrix} \\ &= \sum_{\mathbf{p} \in S_N} \frac{\sigma(\mathbf{p})}{(N!)^{\frac{1}{2}}} \mathcal{P}_{\mathbf{p}} \{ \psi_i(\mathbf{x}_1; \mathbf{R}) \psi_j(\mathbf{x}_2; \mathbf{R}) \cdots \psi_z(\mathbf{x}_N; \mathbf{R}) \} \end{aligned} \quad (2.27)$$

where ket vectors (and bra vectors), $|\Phi\rangle = |\psi_1 \psi_2 \cdots \psi_N\rangle$, denote a given Slater determinant, and $\sigma(\mathbf{p})$ is the parity of permutation, $\mathbf{p}$, and equals -1 for an odd permutation, and 1 for an even permutation. Note that exchanging any two rows or columns of a determinant flips the sign of the determinant, and setting any two rows or columns equal to each other results in a determinant

of 0, thus satisfying the antisymmetry property.⁸ It is important to note that the antisymmetry is due to exchange of both the spatial and spin coordinates, $\mathbf{x}_a$, meaning that two electrons can be described by identical spatial orbitals, provided that they have different spin functions.

The Slater determinant is a substantial upgrade from the Hartree product, satisfying both the indistinguishability and antisymmetry properties. Furthermore, a many-electron wave function that is antisymmetric with respect to pairwise exchange of electronic coordinates includes a type of electronic correlation known as *exchange* correlation. In a system described by a Hartree product, the behavior of any given electron is completely independent of the behaviors of other electrons (Eq. 2.24), while in a system described by a Slater determinant, all electrons having parallel spins (either all spin up, or all spin down) are correlated such that the probability of finding them at the same position is exactly 0. However, this type of correlation is fairly limited, and only exists between electrons with parallel spins, while electrons with antiparallel spins (i.e. both spin up and spin down) remain entirely uncorrelated, so a Slater determinant is often called an uncorrelated wave function.

Single Slater determinants lie at the center of *Hartree-Fock* (HF) theory, and is often used as a starting point for more accurate calculations, such as full CI. Arbitrary correlated many-electron wave functions (including the exact ground state wave function) can be constructed from a sufficiently large linear combination of Slater determinants, so many correlated electronic structure methods use the ground state HF determinant, $|(\Phi_{\text{HF}})_0\rangle$, as a “reference” state, and attempt to approximate the exact ground state wave function as a linear combination of the HF determinant, and its excited determinants.[6] An M -tuply excited determinant ($1 \leq M \leq N$) relative to $|(\Phi_{\text{HF}})_0\rangle = |\psi_1\psi_2\cdots\psi_i\cdots\psi_j\cdots\psi_N\rangle$ corresponds to removing M electrons from occupied spin orbitals, and adding M electrons to distinct unoccupied (excited) spin orbitals. We refer to these excited determinants by adding subscripts and superscripts to the reference ket vector, which label the occupied and excited orbitals (relative to the reference), respectively.

⁸Identical bosons are symmetric with respect to exchange, and therefore require a permanent, rather than a determinant.

For example, removing electrons from occupied orbitals, ψ_i and ψ_j , and adding electrons to excited orbitals, ψ_p and ψ_q , relative to $|(\Phi_{\text{HF}})_0\rangle$, results in the 2-tuply excited determinant denoted by $|(\Phi_{\text{HF}})_{ij}^{pq}\rangle$.⁹ Using this notation, the ground state full CI N -electron wave function, $|(\Phi_{\text{FCI}})_0\rangle$, which describes the exact nonrelativistic wave function (within the span of the basis set approximation used), can be written as:[6]

$$\begin{aligned} |(\Phi_{\text{FCI}})_0\rangle = & c_0 |(\Phi_{\text{HF}})_0\rangle + \sum_{i,p} c_i^p |(\Phi_{\text{HF}})_{ij}^{pq}\rangle + \sum_{\substack{i<j \\ p<q}} c_{ij}^{pq} |(\Phi_{\text{HF}})_{ij}^{pq}\rangle \\ & + \sum_{\substack{i<j<k \\ p<q<r}} c_{ijk}^{pqr} |(\Phi_{\text{HF}})_{ijk}^{pqr}\rangle + \cdots + \sum_{e_1 < e_2 < \cdots < e_N} c_{o_1 o_2 \cdots o_N}^{e_1 e_2 \cdots e_N} |(\Phi_{\text{HF}})_{o_1 o_2 \cdots o_N}^{e_1 e_2 \cdots e_N}\rangle \end{aligned} \quad (2.28)$$

where c denotes expansion coefficients, the first sum ($\sum_{i,p}$) denotes a double sum over occupied and excited orbitals indexed by i and p , respectively, the second sum denotes a quadruple sum over pairs of occupied orbitals ($i < j$) and pairs of excited orbitals ($p < q$), etc., and the last sum adds all N -tuply excited determinants, with $o_1, o_2, \dots, o_N$ indexing the occupied orbitals (relative to the reference) and $e_1, e_2, \dots, e_N$ indexing the excited orbitals (there are only N occupied orbitals, so there is only one choice of $o_1, o_2, \dots, o_N$ due to the indistinguishability of electrons).

It is convenient to introduce the creation ($\hat{a}^\dagger$) and annihilation ($\hat{a}$) operators, which are defined such that, for a given set of K spin orbitals, $\{\psi_1, \dots, \psi_i, \dots, \psi_p, \dots, \psi_K\}$, the creation operator, $\hat{a}_p^\dagger$, creates an electron in spin orbital p , and the annihilation operator, $\hat{a}_i$, annihilates an electron in spin orbital, i :[6]

$$\hat{a}_p^\dagger |\psi_1 \cdots \psi_N\rangle = |\psi_1 \cdots \psi_N \psi_p\rangle \quad (\text{orbital } p \text{ previously unoccupied}) \quad (2.29a)$$

$$\hat{a}_p^\dagger |\psi_1 \cdots \psi_N \psi_p\rangle = 0 \quad (\text{Pauli exclusion}) \quad (2.29b)$$

$$\hat{a}_i |\psi_1 \cdots \psi_i \cdots \psi_N\rangle = |\psi_1 \cdots \psi_N\rangle \quad (\text{annihilate electron from orbital } i) \quad (2.29c)$$

$$\hat{a}_i |\psi_1 \cdots \psi_N\rangle = 0 \quad (\text{orbital } i \text{ previously unoccupied}) \quad (2.29d)$$

⁹In terms of creation ($\hat{a}^\dagger$) and annihilation ($\hat{a}$) operators, we can write this particular excited determinant as $|(\Phi_{\text{HF}})_{ij}^{pq}\rangle = \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_i \hat{a}_j |(\Phi_{\text{HF}})_0\rangle$.

Note that the creation and annihilation operators are adjoints, with the Riesz representation theorem giving the following direct connection relationship between bra and ket: $\langle \psi_1 \cdots \psi_N | \hat{a}_i^\dagger \leftrightarrow \hat{a}_i | \psi_1 \cdots \psi_N \rangle$.

The creation and annihilation operators are discussed in detail in the context of electronic structure theory in Ref. [6]. Here, we only need to understand how they act on a given determinant, and the following anticommutation relations:

$$\hat{a}_i \hat{a}_j + \hat{a}_j \hat{a}_i = 0 \quad (2.30a)$$

$$\hat{a}_i^\dagger \hat{a}_j^\dagger + \hat{a}_j^\dagger \hat{a}_i^\dagger = 0 \quad (2.30b)$$

$$\hat{a}_i^\dagger \hat{a}_j + \hat{a}_j \hat{a}_i^\dagger = \delta_{ij} \mathbb{1} \quad (2.30c)$$

where all three equations encode the antisymmetry of the electronic wave function with respect to pairwise exchange (e.g. $\hat{a}_i^\dagger \hat{a}_j^\dagger = -\hat{a}_j^\dagger \hat{a}_i^\dagger$), and $\delta_{ij} \mathbb{1} = 1$ (when $i = j$) and $\delta_{ij} \mathbb{1} = 0$ (when $i \neq j$) ensures that bra and ket remain normalized after creating or annihilating an electron.

The full CI wave function (Eq. 2.28) can be written as:

$$\begin{aligned} |(\Phi_{\text{FCI}})_0\rangle &= c_0 |(\Phi_{\text{HF}})_0\rangle + \sum_{i,p} c_i^p \hat{a}_p^\dagger \hat{a}_i |(\Phi_{\text{HF}})_{ij}^{pq}\rangle + \sum_{\substack{i < j \\ p < q}} c_{ij}^{pq} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_j \hat{a}_i |(\Phi_{\text{HF}})_0\rangle \\ &+ \sum_{\substack{i < j < k \\ p < q < r}} c_{ijk}^{pqr} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r^\dagger \hat{a}_k \hat{a}_j \hat{a}_i |(\Phi_{\text{HF}})_0\rangle + \cdots \\ &+ \sum_{e_1 < e_2 < \cdots < e_N} c_{o_1 o_2 \cdots o_N}^{e_1 e_2 \cdots e_N} \hat{a}_{e_1}^\dagger \hat{a}_{e_2}^\dagger \cdots \hat{a}_{e_N}^\dagger \hat{a}_{o_N} \cdots \hat{a}_{o_2} \hat{a}_{o_1} |(\Phi_{\text{HF}})_0\rangle \end{aligned} \quad (2.31)$$

2.2.5 Slater-Condon Rules

Before proceeding, to HF theory and correlated theories, it helps to write down matrix elements of the electronic Hamiltonian in the basis of the HF determinant and its excited determinants (the Slater-Condon rules). In general, for some operator $\hat{O}$, the matrix element corresponding to bra and ket N -electron Slater determinants (Eq. 2.27), $\langle \Phi' | = \Phi'^*(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$

and $|\Phi''\rangle = \Phi''(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$, respectively, is calculated by the integral:

$$\begin{aligned} \langle \Phi' | \hat{O} | \Phi'' \rangle &= \int \Phi'^*(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \hat{O} \Phi''(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N \\ &= \int \sum_{\mathbf{p}' \in S_N} \sum_{\mathbf{p}'' \in S_N} \frac{\sigma(\mathbf{p}') \sigma(\mathbf{p}'')}{N!} \mathcal{P}_{\mathbf{p}'} \{ \psi_i'^*(\mathbf{x}_1) \psi_j'^*(\mathbf{x}_2) \cdots \psi_z'^*(\mathbf{x}_N) \} \\ &\quad \hat{O} \mathcal{P}_{\mathbf{p}''} \{ \psi_i''(\mathbf{x}_1) \psi_j''(\mathbf{x}_2) \cdots \psi_z''(\mathbf{x}_N) \} d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N \end{aligned} \quad (2.32)$$

Borrowing the notation from Ref. 6, we need to consider three types of matrix elements:

1. Bra and ket have identical spin orbitals, $\langle \cdots ij \cdots |$ and $|\cdots ij \cdots \rangle$
2. Bra and ket differ by 1 spin orbital, $\langle \cdots rj \cdots |$ and $|\cdots pj \cdots \rangle$
3. Bra and ket differ by 2 spin orbitals, $\langle \cdots rs \cdots |$ and $|\cdots pq \cdots \rangle$

Additionally, it is helpful to consider the matrix elements of the 1-electron operators (Eq. 2.20) and those of the 2-electron operators (Eq. 2.21) separately. The derivations of the Slater-Condon rules are described in detail in Ref. 6, but the general structure of the results follows from the orthogonality of the spin orbitals (Eqs. 2.17 and 2.18), and from the fact that Hermitian operators (including all terms in the Hamiltonian) in general do not preserve orthogonality. To illustrate this point, we first consider the inner products (often called “overlap integrals”) for the three cases enumerated above. The inner product between determinants $\langle \Phi' |$ and $|\Phi''\rangle$ is obtained by substituting the identity operator, $\mathbb{1}$, for $\hat{O}$ in Eq. 2.32.

For case 1, in which the bra and ket correspond to an identical set of spin orbitals, we have $\langle \cdots ij \cdots | \cdots ij \cdots \rangle = 1$. Notice that each determinant in Eq. 2.32 consists of $N!$ Hartree products differing by permutations of the N electrons within the N occupied spin orbitals, so their product (the integrand in Eq. 2.32) corresponds to a sum of $(N!)^2$ terms in total. Using the orthogonality between spin orbitals (Eqs. 2.17 and 2.18), one can see that the only nonzero terms in the integrand correspond to identical permutations, i.e. $\mathbf{p}' = \mathbf{p}''$, which leaves a single sum over permutations (and a total of $N!$ terms, rather than $(N!)^2$ terms). Additionally, since $\mathbf{p}' = \mathbf{p}''$, the product of the parities of the permutations, $\sigma(\mathbf{p}')\sigma(\mathbf{p}'')$, equals 1 for each permutation (since

odd permutations leads to $(-1)(-1) = 1$, and even permutations leads to $(1)(1) = 1$. This leaves us with a sum of $N!$ terms in the integrand, each integrating to 1 (the spin orbitals are normalized), which gives us $N!/N! = 1$ as a final result.

For cases 2 and 3, we find that the inner products, $\langle \cdots rj \cdots | \cdots pj \cdots \rangle$ and $\langle \cdots rs \cdots | \cdots pq \cdots \rangle$, both equal 0. Following the same reasoning as in the case 1 inner product, there are no permutations in which all bra spin orbitals are matched with the analogous ket spin orbitals, so integration over at least one set of electronic coordinates will yield 0 for each term in the sum. In other words, each term in the sum will have at least one electron occupying a pair of bra and ket spin orbitals that are orthogonal to each other. When computing matrix elements of Hermitian operators, we encounter similar situations. However, the fact that Hermitian operators do not preserve orthogonality will sometimes result in nonzero integrals between orthogonal bras and kets.

Before calculating matrix elements, it is helpful to refactor our expression for the Slater determinant (Eq. 2.27). Using the Laplace (cofactor) expansion of the determinant, we obtain the following identity:

$$\begin{aligned}
|\psi_1 \psi_2 \cdots \psi_N\rangle &= \sum_{\mathbf{p} \in S_N} \frac{\sigma(\mathbf{p})}{(N!)^{\frac{1}{2}}} \mathcal{P}_{\mathbf{p}} \{ \psi_1(\mathbf{x}_1) \psi_2(\mathbf{x}_2) \cdots \psi_N(\mathbf{x}_N) \} \\
&= \sum_{i=1}^N \frac{(-1)^{i+1}}{N^{\frac{1}{2}}} \psi_i(\mathbf{x}_1) \left(\sum_{\mathbf{p}' \in S_{N-1}} \frac{\sigma(\mathbf{p}')}{((N-1)!)^{\frac{1}{2}}} \mathcal{P}_{\mathbf{p}'} \{ \cdots \psi_{n \neq i}(\mathbf{x}_{\mathbf{m} \neq \mathbf{1}}) \cdots \} \right) \\
&= \sum_{i=1}^N \frac{(-1)^{i+1}}{N^{\frac{1}{2}}} |\psi_i\rangle \otimes (\hat{a}_i |\psi_1 \psi_2 \cdots \psi_N\rangle)
\end{aligned} \tag{2.33}$$

where $\{ \cdots \psi_{n \neq i}(\mathbf{x}_{\mathbf{m} \neq \mathbf{1}}) \cdots \}$ denotes the Hartree product of all occupied spin orbitals except for ψ_i , containing all electrons except for electron 1, and $\hat{a}_i |\psi_1 \psi_2 \cdots \psi_N\rangle$ denotes the corresponding Slater determinant with an electron annihilated from (previously) occupied orbital, ψ_i . Note that we need to take the tensor product between the determinants to obtain an element of the N -electron Hilbert space, $\mathcal{H}_{\text{el}}^{(N)} = \mathcal{H}_{\text{el}}^{(1)} \otimes \mathcal{H}_{\text{el}}^{(N-1)}$.

Again, using the Laplace expansion of the determinant, we obtain:

$$\begin{aligned}
|\psi_1 \psi_2 \cdots \psi_N\rangle &= \sum_{\mathbf{p} \in S_N} \frac{\sigma(\mathbf{p})}{(N!)^{\frac{1}{2}}} \mathcal{P}_{\mathbf{p}}\{\psi_1(\mathbf{x}_1) \psi_2(\mathbf{x}_2) \cdots \psi_N(\mathbf{x}_N)\} \\
&= \sum_{i=1}^{N-1} \sum_{j>i}^N \frac{2^{\frac{1}{2}} (-1)^{i+j+1}}{(N(N-1))^{\frac{1}{2}}} \left(\sum_{\mathbf{p}' \in S_2} \frac{\sigma(\mathbf{p}')}{(2!)^{\frac{1}{2}}} \mathcal{P}_{\mathbf{p}'}\{\psi_i(\mathbf{x}_1) \psi_j(\mathbf{x}_2)\} \right) \\
&\quad \left(\sum_{\mathbf{p}'' \in S_{N-2}} \frac{\sigma(\mathbf{p}'')}{((N-2)!)^{\frac{1}{2}}} \mathcal{P}_{\mathbf{p}''}\{\cdots \psi_{n \neq i}(\mathbf{x}_{\mathbf{m} \neq 1}) \psi_{n' \neq j}(\mathbf{x}_{\mathbf{m}' \neq 2}) \cdots\} \right) \\
&= \sum_{i=1}^{N-1} \sum_{j>i}^N \frac{2^{\frac{1}{2}} (-1)^{i+j+1}}{(N(N-1))^{\frac{1}{2}}} |\psi_i \psi_j\rangle \otimes (\hat{a}_j \hat{a}_i |\psi_1 \psi_2 \cdots \psi_N\rangle)
\end{aligned} \tag{2.34}$$

where i and j (in the double sum) index all pairs of distinct orbitals, $\mathbf{p}' \in S_2$ is an element of the permutation group of degree 2 ($S_2 = \{[1, 2], [2, 1]\}$), $\{\cdots \psi_{n \neq i}(\mathbf{x}_{\mathbf{m} \neq 1}) \psi_{n' \neq j}(\mathbf{x}_{\mathbf{m}' \neq 2}) \cdots\}$ denotes the Hartree product of all occupied spin orbitals except for ψ_i and ψ_j , containing all electrons except for electrons 1 and 2, and $\hat{a}_j \hat{a}_i |\psi_1 \psi_2 \cdots \psi_N\rangle$ denotes the corresponding Slater determinant with electrons annihilated from (previously) occupied orbitals, ψ_i and ψ_j . The extra factor of $2^{\frac{1}{2}}$ ensures that $|\psi_i \psi_j\rangle$ is normalized.

Due to the indistinguishability of electrons described by a Slater determinant, we can simplify the 1-electron operator, $\hat{h} = \sum_{a=1}^N \hat{h}_a$, by only considering a single electron, $\sum_{a=1}^N \hat{h}_a \rightarrow N \hat{h}_1$, arbitrarily choosing electron 1. Additionally, indistinguishability allows us to simplify the 2-electron operator, $\hat{v} = \sum_{a=1}^{N-1} \sum_{b>a}^N r_{ab}^{-1}$, by only considering a single pair of electrons, $\hat{v} \rightarrow \frac{1}{2} N(N-1) r_{12}^{-1}$, arbitrarily choosing electrons 1 and 2. With these simplified operators, the use of the expressions for $|\psi_1 \psi_2 \cdots \psi_N\rangle$ in the last lines of Eqs. 2.33 and 2.34 clarifies the computation of the 1- and 2-electron operator matrix elements, respectively, because they allow us to focus on 1- and 2-electron integrals (involving the kets on the left of the tensor products), while the integrals over the coordinates of the remaining electrons (involving the kets on the right of the tensor products) are easily evaluated with the inner product rules.

Evaluating the 1-electron operator ($\hat{h}$) matrix element for case 1 gives:[6]

$$\begin{aligned}
\langle \dots ij \dots | \hat{h} | \dots ij \dots \rangle &= \left(\sum_{i=1}^N \frac{(-1)^{i+1}}{N^{\frac{1}{2}}} \langle \psi_i | \otimes \langle \psi_1 \psi_2 \dots \psi_N | \hat{a}_i^\dagger \right) \\
&\quad (N \hat{h}_1) \left(\sum_{i'=1}^N \frac{(-1)^{i'+1}}{N^{\frac{1}{2}}} | \psi_{i'} \rangle \otimes \hat{a}_{i'} | \psi_1 \psi_2 \dots \psi_N \rangle \right) \\
&= \sum_{i=1}^N \frac{N(-1)^{2i+2}}{N} \langle \psi_i | \hat{h}_1 | \psi_i \rangle \langle \psi_1 \psi_2 \dots \psi_N | \hat{a}_i^\dagger \hat{a}_i | \psi_1 \psi_2 \dots \psi_N \rangle \\
&= \sum_{i=1}^N \int \psi_i^*(\mathbf{x}_1) \hat{h}_1 \psi_i(\mathbf{x}_1) d\mathbf{x}_1 = \sum_{i=1}^N \langle i | \hat{h} | i \rangle
\end{aligned} \tag{2.35}$$

where the double sum collapses into a single sum because the inner product, $\langle \psi_1 \psi_2 \dots \psi_N | \hat{a}_i^\dagger \hat{a}_{i'} | \psi_1 \psi_2 \dots \psi_N \rangle = \delta_{ii'}$, despite the fact that $\langle \psi_i | \hat{h}_1 | \psi_{i'} \rangle$ in general is nonzero, even when $i \neq i'$.

For case 2, we can immediately collapse the double sum into a single sum for the same reason as in case 1. In particular, we index $\langle \dots rj \dots |$ and $| \dots pj \dots \rangle$ such that orbitals ψ_r^* in the bra and ψ_p in the ket are referenced by the same index, i , in the sum. With this in mind, the matrix element is evaluated as:[6]

$$\begin{aligned}
\langle \dots rj \dots | \hat{h} | \dots pj \dots \rangle &= \sum_{i=1}^N \langle \psi_i | \hat{h}_1 | \psi_i \rangle \langle \psi_1 \psi_2 \dots \psi_N | \hat{a}_i^\dagger \hat{a}_i | \psi_1 \psi_2 \dots \psi_N \rangle \\
&= \int \psi_r^*(\mathbf{x}_1) \hat{h}_1 \psi_p(\mathbf{x}_1) d\mathbf{x}_1 \\
&= \langle r | \hat{h} | p \rangle
\end{aligned} \tag{2.36}$$

where all terms in the sum in which $i \neq r$ in the bra and $i \neq p$ in the ket evaluate to 0 because $\langle \psi_1 \psi_2 \dots \psi_N | \hat{a}_{i \neq r}^\dagger \hat{a}_{i \neq p} | \psi_1 \psi_2 \dots \psi_N \rangle = 0$.

In case 3, for each i in the sum, $\langle \psi_1 \psi_2 \dots \psi_N | \hat{a}_i^\dagger$ contains either ψ_r^* or ψ_s^* , and $\hat{a}_i | \psi_1 \psi_2 \dots \psi_N \rangle$ contains either ψ_p or ψ_q , so $\langle \psi_1 \psi_2 \dots \psi_N | \hat{a}_i^\dagger \hat{a}_i | \psi_1 \psi_2 \dots \psi_N \rangle$ always evaluates to 0. We therefore obtain $\langle \dots rs \dots | \hat{h} | \dots pq \dots \rangle = 0$. [6] Note that matrix elements between determinants differing by more than two spin orbitals also evaluate to 0 by the same reasoning.

We can apply a similar approach when calculating the 2-electron operator matrix elements.

We start by evaluating the following 2-electron integral:

$$\begin{aligned}
\langle \psi_r \psi_s | \frac{1}{r_{12}} | \psi_p \psi_q \rangle &= \frac{1}{2} \int (\psi_r^*(\mathbf{x}_1) \psi_s^*(\mathbf{x}_2) - \psi_r^*(\mathbf{x}_2) \psi_s^*(\mathbf{x}_1)) \\
&\quad \frac{1}{r_{12}} (\psi_p(\mathbf{x}_1) \psi_q(\mathbf{x}_2) - \psi_p(\mathbf{x}_2) \psi_q(\mathbf{x}_1)) d\mathbf{x}_1 d\mathbf{x}_2 \\
&= \int \psi_r^*(\mathbf{x}_1) \psi_p(\mathbf{x}_1) \frac{1}{r_{12}} \psi_s^*(\mathbf{x}_2) \psi_q(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\
&\quad - \int \psi_r^*(\mathbf{x}_1) \psi_q(\mathbf{x}_1) \frac{1}{r_{12}} \psi_s^*(\mathbf{x}_2) \psi_p(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\
&= \langle rs | |pq \rangle
\end{aligned} \tag{2.37}$$

where we swapped the integration variables $\mathbf{x}_1 \leftrightarrow \mathbf{x}_2$ in two of the four integrals, resulting in the final two integrals (with factors of $2/2 = 1$ for each), which we abbreviate as $\langle rs | |pq \rangle$. [6] With this result, we can evaluate case 1:

$$\begin{aligned}
\langle \dots ij \dots | \hat{v} | \dots ij \dots \rangle &= \left(\sum_{i=1}^{N-1} \sum_{j>i}^N \frac{2^{\frac{1}{2}} (-1)^{i+j+1}}{(N(N-1))^{\frac{1}{2}}} \langle \psi_i \psi_j | \otimes \langle \psi_1 \psi_2 \dots \psi_N | \hat{a}_i^\dagger \hat{a}_j^\dagger \right) \\
&\quad \left(\frac{N(N-1)}{2} \frac{1}{r_{12}} \left(\sum_{i'=1}^{N-1} \sum_{j'>i'}^N \frac{2^{\frac{1}{2}} (-1)^{i'+j'+1}}{(N(N-1))^{\frac{1}{2}}} | \psi_{i'} \psi_{j'} \rangle \otimes \hat{a}_{j'} \hat{a}_{i'} | \psi_1 \psi_2 \dots \psi_N \rangle \right) \right) \\
&= \sum_{i=1}^{N-1} \sum_{j>i}^N \langle \psi_i \psi_j | \frac{1}{r_{12}} | \psi_i \psi_j \rangle \langle \psi_1 \psi_2 \dots \psi_N | \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_{j'} \hat{a}_{i'} | \psi_1 \psi_2 \dots \psi_N \rangle \\
&= \sum_{i=1}^{N-1} \sum_{j>i}^N \langle ij | |ij \rangle
\end{aligned} \tag{2.38}$$

Taking a closer look at one of the terms in the sum, we have:

$$\begin{aligned}
\langle ij | ij \rangle &= \langle \psi_i \psi_j | \frac{1}{r_{12}} | \psi_i \psi_j \rangle \\
&= \int \psi_i^*(\mathbf{x}_1) \psi_i(\mathbf{x}_1) \frac{1}{r_{12}} \psi_j^*(\mathbf{x}_2) \psi_j(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\
&\quad - \int \psi_i^*(\mathbf{x}_1) \psi_j(\mathbf{x}_1) \frac{1}{r_{12}} \psi_i^*(\mathbf{x}_2) \psi_j(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\
&= \int |\psi_i(\mathbf{x}_1)|^2 \frac{1}{r_{12}} |\psi_j(\mathbf{x}_2)|^2 d\mathbf{x}_1 d\mathbf{x}_2 \\
&\quad - \int \psi_i^*(\mathbf{x}_1) \psi_j(\mathbf{x}_1) \frac{1}{r_{12}} \psi_i^*(\mathbf{x}_2) \psi_j(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\
&= J_{ij} - K_{ij}
\end{aligned} \tag{2.39}$$

J_{ij} is called the *Coulomb integral* due to the fact that it resembles the Coulombic interaction between electron distributions, $|\psi_i(\mathbf{x}_1)|^2$ and $|\psi_j(\mathbf{x}_2)|^2$. This is a sound interpretation of the J_{ij} integrals given that the Born rule (Eq. 2.23) states that $|\psi_i(\mathbf{x}_1)|^2$ is the spatial probability density function for an electron occupying orbital, ψ_i , and the fact that the electron is a charged particle. Unfortunately, this is not an adequate description of electron-electron interactions on its own, since $|\psi_i(\mathbf{x}_1)|^2$ corresponds to a charge-probability density, not a charge density. In other words, while $|\psi_i(\mathbf{x}_1)|^2$ is delocalized over space, a single electron can only exist at a single position at any given time (within the limits of the uncertainty principle). The Coulomb integral treats $|\psi_i(\mathbf{x}_1)|^2$ as a charge distribution where some fraction of the charge exists in some region of space, and another fraction at some other region, etc., while in reality, the charge-probability distribution tells us that there is some probability of finding exactly -1 charge (in natural units) within any given region in space.

The fact that J_{ij} , on its own, is not appropriate for describing electron-electron interactions should not be surprising since $\langle \psi_i \psi_j | r_{12}^{-1} | \psi_i \psi_j \rangle = \langle \psi_i \psi_j | \psi_i \psi_j \rangle = J_{ij} - K_{ij}$, so interpreting J_{ij} , itself, as the Coulomb interaction between the two electrons described by the Slater determinant, $|\psi_i \psi_j\rangle$, neglects the contribution due to the mysterious K_{ij} integral. K_{ij} is called the *exchange integral*, and is a direct consequence of the exchange correlation captured by the use of a Slater

determinant.¹⁰ Part of the ‘mystery’ surrounding the exchange integral comes from the fact that it has no classical analog. While $\psi_i^*(\mathbf{x})\psi_i(\mathbf{x}) = |\psi_i(\mathbf{x})|^2$, in the Coulomb integral, is appropriately interpreted as a probability density function, $\psi_i^*(\mathbf{x})\psi_j(\mathbf{x})$ is rather cryptic. When we re-imagine the Coulomb and exchange interactions in terms of 1-electron operators when discussing HF theory in the next section, we will see that the exchange interaction is inherently nonlocal.

Ultimately, one must be satisfied with the interpretation that K_{ij} quantifies the effect that exchange correlation on the electrostatic interactions between electrons with parallel spin (evaluation of the exchange integral in Eq. 2.39 yields $K_{ij} = 0$ when ψ_i and ψ_j are associated with opposing spin functions). The most apparent feature of exchange correlation is the Pauli exclusion principle, which is demonstrated by the computation of the 2-electron integral (Eq. 2.37 or Eq. 2.39) between determinants composed of identical spin orbitals, i.e. $\langle \psi_i \psi_i | r_{12}^{-1} | \psi_i \psi_i \rangle = \langle ii | ii \rangle = J_{ii} - K_{ii} = 0$. The fact that $J_{ii} = K_{ii}$ allows us to remove the restrictions on the summation index in Eq. 2.38, going from a sum over unique pairs of spin orbitals, to a simple double sum, double counting each pair and including the ‘diagonal terms’:[6]

$$\begin{aligned}
 \langle \cdots ij \cdots | \hat{v} | \cdots ij \cdots \rangle &= \sum_{i=1}^{N-1} \sum_{j>i}^N \langle ij | ij \rangle \\
 &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \langle ij | ij \rangle \\
 &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N J_{ij} - K_{ij}
 \end{aligned} \tag{2.40}$$

where the extra factor of $\frac{1}{2}$ corrects for the double counting in the unrestricted sum, and the diagonal terms (in which $i = j$) can be included because $\langle ii | ii \rangle = 0$. The use of an unrestricted

¹⁰There is an understandable misattribution of the exchange interaction to the antisymmetry of electronic wave functions in the text book by Szabo and Ostlund (although they also correctly ascribe it to exchange correlation, as we do here).[6] This is clearly incorrect given the fact that bosons display exchange interactions as well, differing only by the sign ($\langle \psi_i \psi_j |_{\text{boson}} r_{12}^{-1} | \psi_i \psi_j \rangle_{\text{boson}} = J_{ij} + K_{ij}$) due to their symmetry with respect to pairwise exchange, captured by the use of a permanent (rather than a determinant). By this reasoning, one might say that the exchange interaction is due to the indistinguishability of elementary particles, which is implicit in our attribution to the exchange correlation. It should be noted that this error is of little consequence, and should not deter the reader from the otherwise excellent text.

sum will prove to be convenient when formulating HF theory in terms of an effective 1-electron TISE.

However, while Eq. 2.40 is convenient in the 1-electron picture used in HF, it is important to be careful with our interpretations of the Coulombic part of the expression, given the discussion above about treating the electron distribution as a charge distribution. In particular, after discarding the K_{ij} terms in Eq. 2.40, and interpreting $\rho(\mathbf{x}) = \sum_{i=1}^N |\psi_i(\mathbf{x})|^2$ as a charge density (in natural units), $\mathcal{J}[\rho] = \frac{1}{2} \sum_{i=1}^N J_{ii}$ describes the classical self-repulsion of the charge density with itself.[10] Further elaboration on the consequences of such a transgression can be found in the discussion on the self-interaction error and the delocalization error in density functional theory (DFT) in the literature.[11, 12, 13, 14]

The results from evaluating 2-electron operator matrix elements for cases 2 and 3 are given by:[6]

$$\langle \cdots r j \cdots | \hat{v} | \cdots p j \cdots \rangle = \sum_{j=1}^N \langle r j | | p j \rangle \quad (2.41a)$$

$$\langle \cdots r s \cdots | \hat{v} | \cdots p q \cdots \rangle = \langle r s | | p q \rangle \quad (2.41b)$$

The derivations are left as exercises for the reader.

2.2.6 Hartree-Fock Theory

Despite being incapable of describing a fully-correlated many-electron wave function, a single Slater determinant is a useful starting point when calculating the ground state wave function (corresponding to the lowest energy state of the electrons for some fixed set of nuclear coordinates, $\mathbf{R}$). In particular, the Hartree-Fock approximation leverages the variational principle, minimizing the (total electronic energy) expectation value of the electronic Hamiltonian, $E_{\text{HF},0} = \min_{\Phi} \langle \Phi | \hat{H}_{\text{el}} | \Phi \rangle$, with respect to many-electron wave functions approximated as single Slater determinants, constraining the determinant to be normalized, $\langle \Phi | \Phi \rangle = 1$. In this context, the variational principle states that minimizing E_{HF} with respect to single determinants yields an

upper-bound to the exact electronic energy, and is a consequence of the fact that the exact ground state many-electron wave function minimizes the total electronic energy for a given electronic Hamiltonian (i.e. it corresponds to the smallest eigenvalue of $\hat{H}_{\text{el}}$). Here, we skip most of the details of the HF method (which are described in detail in Ref. 6), and only focus on the major ideas.

Using the results from Eqs. 2.35 and 2.40, we can write the energy expectation value for an N -electron single determinant as:

$$\begin{aligned}
\langle \Phi | \hat{H}_{\text{el}} | \Phi \rangle &= \sum_{i=1}^N \langle i | \hat{h} | i \rangle + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \langle ij | ij \rangle \\
&= \sum_{i=1}^N \int \psi_i^*(\mathbf{x}) \left(-\frac{\hbar^2}{2} \nabla_a^2 - \sum_{A=1}^{N_{\text{nuc}}} \frac{Z_A}{\|\mathbf{R}_A - \mathbf{r}\|_2} \right) \psi_i(\mathbf{x}) d\mathbf{x} \\
&\quad + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left(\int \psi_i^*(\mathbf{x}_1) \psi_i(\mathbf{x}_1) \frac{1}{r_{12}} \psi_j^*(\mathbf{x}_2) \psi_j(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \right. \\
&\quad \left. - \int \psi_i^*(\mathbf{x}_1) \psi_j(\mathbf{x}_1) \frac{1}{r_{12}} \psi_j^*(\mathbf{x}_2) \psi_i(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \right)
\end{aligned} \tag{2.42}$$

Using the method of Lagrange multipliers, the HF approximation can be shown to be equivalent to solving an effective 1-electron TISE where the true electron-electron coupling (Eq. 2.21) is replaced by the average potential, $\hat{v}^{\text{HF}}(\mathbf{x})$, due to the other electrons:[6]

$$\hat{f}(\mathbf{x}_1) \psi_i(\mathbf{x}_1) = (\hat{h}(\mathbf{x}_1) + \hat{v}^{\text{HF}}(\mathbf{x}_1)) \psi_i(\mathbf{x}_1) = \varepsilon_i \psi_i(\mathbf{x}_1) \tag{2.43}$$

where $\hat{f}(\mathbf{x}_1)$ is the effective 1-electron Hamiltonian called the *Fock operator*, $\hat{h}(\mathbf{x}_1)$ is the 1-electron Hamiltonian (Eq. 2.20), $\psi_i(\mathbf{x}_1)$ is the i^{th} eigenvector of $\hat{f}(\mathbf{x}_1)$ corresponding to the i^{th} spin orbital ($i \leq N$ corresponding to occupied orbitals, and $i > N$ corresponding to excited orbitals), and ε_i is the orbital energy of $\psi_i(\mathbf{x}_1)$, arising from the diagonalization of the matrix of Lagrange multipliers.[6]

As expected, the difficult part is obtaining the effective potential, $\hat{v}^{\text{HF}}(\mathbf{x}_1)$, which must

account for the electron-electron interactions in a 1-electron operator. To do this, we define the Coulomb and exchange operators:[6]

$$\hat{v}^{\text{HF}}(\mathbf{x}_1) = \hat{j}(\mathbf{x}_1) - \hat{k}(\mathbf{x}_1) \quad (2.44)$$

with $\hat{j}(\mathbf{x}_1)$ corresponding to the Coulomb potential due to the total charge distribution of the N -electron HF ground state Slater determinant, $\sum_{i=1}^N |\psi_i(\mathbf{x}_2)|^2$:

$$\hat{j}(\mathbf{x}_1) = \sum_{i=1}^N \int |\psi_i(\mathbf{x}_2)|^2 \frac{1}{r_{12}} d\mathbf{x}_2 \quad (2.45)$$

Notice that $\hat{j}(\mathbf{x}_1)$ depends on the electron position, $\mathbf{x}_1$, since r_{12}^{-1} in the integrand depends on $\mathbf{x}_1$, and we only integrate over $\mathbf{x}_2$. $\hat{k}(\mathbf{x}_1)$ describes the exchange interaction:

$$\hat{k}(\mathbf{x}_1) = \sum_{i=1}^N \int \psi_i^*(\mathbf{x}_2) \frac{1}{r_{12}} \mathcal{P}_{[2,1]} \{ \psi_i(\mathbf{x}_2) \cdot \} d\mathbf{x}_2 \quad (2.46)$$

where $\mathcal{P}_{[2,1]}$ permutes the coordinates, $\mathbf{x}_2$ and $\mathbf{x}_1$, and the dot, $\cdot$, indicates that $\mathcal{P}_{[2,1]}$ acts on the operand of $\hat{k}(\mathbf{x}_1)$:

$$\begin{aligned} \hat{k}(\mathbf{x}_1) \psi_j(\mathbf{x}_1) &= \sum_{i=1}^N \int \psi_i^*(\mathbf{x}_2) \frac{1}{r_{12}} \mathcal{P}_{[2,1]} \{ \psi_i(\mathbf{x}_2) \psi_j(\mathbf{x}_1) \} d\mathbf{x}_2 \\ &= \sum_{i=1}^N \int \psi_i^*(\mathbf{x}_2) \frac{1}{r_{12}} \psi_i(\mathbf{x}_1) \psi_j(\mathbf{x}_2) d\mathbf{x}_2 \end{aligned} \quad (2.47)$$

The inclusion of the $\mathcal{P}_{[2,1]}$ makes $\hat{k}(\mathbf{x}_1)$ a *nonlocal operator*, due to the fact that it is not diagonalizable in the position basis,¹¹ i.e. $\langle \mathbf{r} | \hat{k} | \mathbf{r}' \rangle \neq \langle \mathbf{r} | \hat{k} | \mathbf{r} \rangle \delta(\mathbf{r} - \mathbf{r}')$. [10] In other words, the action of $\hat{k}(\mathbf{x}_1)$ on $\psi_j(\mathbf{x}_1)$ depends on ψ_j over all space, not just its value at $\mathbf{x}_1$, due to the fact that we are integrating over $\mathbf{x}_2$ in Eq. 2.47. [6] Additionally, we reiterate that the fact that $J_{ii} = K_{ii}$, which allows us to define the operators, $\hat{j}$ and $\hat{k}$, in a 1-electron picture that is valid for all orbitals

¹¹The kinetic energy operator is nonlocal as well due to the gradient operators.

(we don't need restricted sums over unique pairs of orbitals, i.e. Eq. 2.38 $\rightarrow$ Eq. 2.40).

The HF equations (Eq. 2.43) determine the spin orbitals that compose the optimal Slater determinant that minimizes the total electronic energy (Eq. 2.42). However, the definition of the effective potential (Eqs. 2.44, 2.45, and 2.46) appears to require the knowledge of the solutions up front (since the Coulomb and exchange operators both involve sums over the N occupied spin orbitals, indexed by i). This means that the HF equations need to be solved self-consistently by an iterative method, starting with some initial guess for the ground state determinant, constructing the corresponding Fock operator, solving for the eigenvectors (spin orbitals) of the initial-guess Fock operator, constructing a new Fock operator with the N eigenvectors with the smallest eigenvalues, ε_i , and then solving for the eigenvectors of the new Fock operator, etc., until the spin orbitals converge to some iterative fixed point (i.e. the eigenvectors converge to the orbitals used in the construction of the Fock operator). Once the solutions of the HF equations converge to the optimal spin orbitals, we construct the corresponding Slater determinant (Eq. 2.27), and evaluate the energy expectation value with Eq. 2.42. For this reason, one often calls the HF equations the HF Self Consistent Field (SCF) equations.

In the context of the HF equations (Eq. 2.43), we can express the SCF iterative procedure as follows:

$$\begin{aligned}
& |\psi_1^{(n-1)} \dots \psi_N^{(n-1)}\rangle_{n-1} \rightarrow \hat{f}^{(n)}(\mathbf{x}_1) \\
& \hat{f}^{(n)}(\mathbf{x}_1) \psi_i^{(n)}(\mathbf{x}_1) = \varepsilon_i^{(n)} \psi_i^{(n)}(\mathbf{x}_1) \rightarrow \{\psi_1^{(n)}, \dots, \psi_K^{(n)}\} \\
& \arg \min_{\mathcal{C}_N\{1, \dots, K\} \rightarrow \{\psi_1^{(n)}, \dots, \psi_N^{(n)}\}} \sum_{\psi_i^{(n)} \in \{\psi_1^{(n)}, \dots, \psi_N^{(n)}\}} \langle \psi_i^{(n)} | \hat{f}^{(n)}(\mathbf{x}_1) | \psi_i^{(n)} \rangle \rightarrow |\psi_1^{(n)} \dots \psi_N^{(n)}\rangle_n
\end{aligned} \tag{2.48}$$

where the superscript “ (n) ” denotes the iteration index, a complete set of orthogonal spin orbitals, $\{\psi_1^{(n)}, \dots, \psi_K^{(n)}\}$, are obtained from solving the 1-electron TISE in the n^{th} iteration (line 2 of Eq. 2.48), $\mathcal{C}_N\{1, 2, \dots, K\}$ denotes the set of all N -element combinations of integers, $\{1, 2, \dots, K\}$, and the argmin function picks out the N spin orbitals, $\{\psi_1^{(n)}, \dots, \psi_N^{(n)}\}$, with the lowest (orbital energy) eigenvalues, $\varepsilon_i^{(n)} = \langle \psi_i^{(n)} | \hat{f}^{(n)}(\mathbf{x}_1) | \psi_i^{(n)} \rangle$, within the span of a basis set

approximation composed of K functions (leading to a total of K eigenfunctions and eigenvalues). From Eq. 2.42, the total electronic energy of the n^{th} Slater determinant, $E_0^{(n)}$, can be written as:

$$E_0^{(n)} = \langle \psi_1^{(n)} \psi_2^{(n)} \cdots \psi_N^{(n)} | {}_n \hat{H}_{\text{el}} | \psi_1^{(n)} \psi_2^{(n)} \cdots \psi_N^{(n)} \rangle_n \quad (2.49)$$

and SCF convergence is reached when $|E_0^{(n+1)} - E_0^{(n)}| < \Delta_{\text{conv}}$, where Δ_{conv} is some user-defined convergence threshold.¹²

Before proceeding to the correlated methods, we make a note of an important result from HF theory called *Brillouin's theorem*. Recall that in the correlated methods, we usually use the HF determinant, $|(\Phi_{\text{HF}})_0\rangle$, as a reference state, and construct a correlated wave function as a linear combination of the reference and its excited determinants. Brillouin's theorem states that matrix elements of the electronic Hamiltonian between the HF determinant and its singly-excited determinants evaluate to 0:[6]

$$\langle (\Phi_{\text{HF}})_0 | \hat{H}_{\text{el}} | (\Phi_{\text{HF}})_i^p \rangle = 0 \quad (\text{for all } i, p) \quad (2.50)$$

This is due to the fact that singly-excited determinants relative to the HF ground state differ by a single spin orbital (i.e. $\psi_i \rightarrow \psi_p$ in Eq. 2.50). Using the case 2 matrix elements discussed in Sec. 2.2.5, we find that $\langle (\Phi_{\text{HF}})_0 | \hat{H}_{\text{el}} | (\Phi_{\text{HF}})_i^p \rangle = \langle \psi_i | \hat{f} | \psi_p \rangle$. Since the relevant spin orbitals, ψ_i and ψ_j , are eigenvectors of the Fock operator, $\hat{f}$, $\langle \psi_i | \hat{f} | \psi_p \rangle = 0$ because $\hat{f}$ is Hermitian (its eigenvectors are orthogonal). Note that while the single-excited determinants aren't directly coupled to the (HF) reference state, they are indirectly coupled through coupling to the doubly- and triply-excited determinants.

¹²Other properties like the total density can be used to determine convergence as well.

2.2.7 The Correlation Energy and The Configuration Interaction Method

As discussed above, the exact (nonrelativistic) many-electron wave function within the BOA, and within the basis set approximation used, can be obtained from the Full CI method, resulting in the full CI wave function, $|(\Phi_{\text{FCI}})_0\rangle$, which includes exact electron-electron correlations. We define the (ground state) *correlation energy*, $E_{\text{corr},0}$, to be the difference between the HF ground state energy and the exact energy (from Full CI):

$$E_{\text{corr},0} = E_{\text{FCI},0} - E_{\text{HF},0} = \langle (\Phi_{\text{FCI}})_0 | \hat{H}_{\text{el}} | (\Phi_{\text{FCI}})_0 \rangle - \langle (\Phi_{\text{HF}})_0 | \hat{H}_{\text{el}} | (\Phi_{\text{HF}})_0 \rangle \quad (2.51)$$

For convenience, we write the full CI wave function in “intermediate normalized” form:[6]

$$\begin{aligned} |(\Phi_{\text{FCI}})_0\rangle = & |(\Phi_{\text{HF}})_0\rangle + \sum_{i,p} c_i^p |(\Phi_{\text{HF}})_{ij}^{pq}\rangle + \sum_{\substack{i<j \\ p<q}} c_{ij}^{pq} |(\Phi_{\text{HF}})_{ij}^{pq}\rangle \\ & + \sum_{\substack{i<j<k \\ p<q<r}} c_{ijk}^{pqr} |(\Phi_{\text{HF}})_{ijk}^{pqr}\rangle + \cdots + \sum_{e_1 < e_2 < \cdots < e_N} c_{o_1 o_2 \cdots o_N}^{e_1 e_2 \cdots e_N} |(\Phi_{\text{HF}})_{o_1 o_2 \cdots o_N}^{e_1 e_2 \cdots e_N}\rangle \end{aligned} \quad (2.52)$$

which, itself, is not normalized due to a coefficient of 1 for the HF determinant, $|(\Phi_{\text{HF}})_0\rangle$, leading to $\langle (\Phi_{\text{HF}})_0 | (\Phi_{\text{FCI}})_0 \rangle = 1$. Note that the coefficients, c , differ from those in Eq. 2.52 due to the different normalization. It is important to keep in mind that we are talking about excited determinants relative to the HF ground state, which, by the variational principle, is itself some kind of excited (non-stationary) state relative to the exact (full CI) ground state wave function. Given that the single Slater determinant defining the HF ground state wave function does not correlate the electrons (beyond exchange correlation), the “excitations” of the HF ground state act as correlation corrections to the uncorrelated reference, with double excitations simultaneously correlating pairs of electrons, triple excitations simultaneously correlating 3 electrons, etc.. Reality is slightly more complicated than m -tuply excited determinants strictly

correlating m electrons since determinants corresponding to different orders of excitations are coupled with each other, but it is a decent intuitive picture to keep in mind.

We can act the operator, $\hat{H}_{\text{el}} - E_{\text{HF},0}\mathbb{1}$ on the intermediately normalized wave function to obtain:

$$\begin{aligned} (\hat{H}_{\text{el}} - E_{\text{HF},0}\mathbb{1})|(\Phi_{\text{FCI}})_0\rangle &= (E_{\text{FCI},0} - E_{\text{HF},0})|(\Phi_{\text{FCI}})_0\rangle \\ &= E_{\text{corr}}|(\Phi_{\text{FCI}})_0\rangle \end{aligned} \quad (2.53)$$

If we left-multiply by $\langle(\Phi_{\text{HF}})_0|$, using $\langle(\Phi_{\text{HF}})_0|(\Phi_{\text{FCI}})_0\rangle = 1$, we obtain:

$$\begin{aligned} \langle(\Phi_{\text{HF}})_0|(\hat{H}_{\text{el}} - E_{\text{HF},0}\mathbb{1})|(\Phi_{\text{FCI}})_0\rangle &= E_{\text{corr}}\langle(\Phi_{\text{HF}})_0|(\Phi_{\text{FCI}})_0\rangle = E_{\text{corr}} \\ &= \langle(\Phi_{\text{HF}})_0|(\hat{H}_{\text{el}} - E_{\text{HF},0}\mathbb{1})\left(|(\Phi_{\text{HF}})_0\rangle + \sum_{i,p} c_i^p |(\Phi_{\text{HF}})_i^p\rangle + \sum_{\substack{i<j \\ p<q}} c_{ij}^{pq} |(\Phi_{\text{HF}})_{ij}^{pq}\rangle + \dots\right) \end{aligned} \quad (2.54)$$

Now using the fact that $(\hat{H}_{\text{el}} - E_{\text{HF},0}\mathbb{1})|(\Phi_{\text{HF}})_0\rangle = (E_{\text{HF},0} - E_{\text{HF},0})|(\Phi_{\text{HF}})_0\rangle = 0$, and $\langle(\Phi_{\text{HF}})_0|\hat{H}_{\text{el}}|(\Phi_{\text{HF}})_i^p\rangle = 0$ due to Brillouin's theorem, as well as the fact that matrix elements of $\hat{H}_{\text{el}}$ between $|(\Phi_{\text{HF}})_0\rangle$ and triply- and higher-excited determinants evaluate to 0, we obtain the following result:[6]

$$E_{\text{corr}} = \sum_{\substack{i<j \\ p<q}} c_{ij}^{pq} \langle(\Phi_{\text{HF}})_0|\hat{H}_{\text{el}}|(\Phi_{\text{HF}})_{ij}^{pq}\rangle \quad (2.55)$$

So close, yet so far. This disarmingly simple expression for the correlation energy does not convey the fact that the exact doubly-excited coefficients in the full CI (ground state) wave function depend on the coefficients for the singly- triply- and quadruply-excited determinants, which themselves depend on the 5-tuply, 6-tuply, etc. coefficients. In order to obtain the exact full CI wave function, one must construct the electronic Hamiltonian matrix in the basis of the full set of determinants (or at least some representation of it¹³), and solve for its eigenvectors

¹³The full basis of determinants for even modestly sized systems can be enormous, so storing the full Hamiltonian matrix in memory is not practical. Another approach involves the use of an iterative method that only stores information about how to act the Hamiltonian on a given wave function without needing to store the actual matrix.

and eigenvalues, with the exact ground state corresponding to the eigenvector with the smallest eigenvalue.¹⁴

2.2.8 Size Consistency and Coupled Cluster Methods

The full CI method is intractable for all but the smallest systems (and basis set approximations), so we must use less expensive methods in order to obtain correlation energies for practical applications. The most obvious simplification to full CI is to truncate the CI expansion of the wave function at some order. For example, we might throw away all terms in the expansion except for the (reference) ground state and the singly- and doubly-excited determinants relative to the reference. This kind of truncation leads to CI singles, doubles (CISD), corresponding to the wave function expansion (see Eq. 2.31):

$$\begin{aligned}
 |(\Phi_{\text{CISD}})_0\rangle &= c_0 |(\Phi_{\text{HF}})_0\rangle + \sum_{i,p} c_i^p |(\Phi_{\text{HF}})_i^p\rangle + \sum_{\substack{i<j \\ p<q}} c_{ij}^{pq} |(\Phi_{\text{HF}})_{ij}^{pq}\rangle \\
 &= c_0 |(\Phi_{\text{HF}})_0\rangle + \sum_{i,p} c_i^p \hat{a}_p^\dagger \hat{a}_i |(\Phi_{\text{HF}})_0\rangle + \sum_{\substack{i<j \\ p<q}} c_{ij}^{pq} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_j \hat{a}_i |(\Phi_{\text{HF}})_0\rangle
 \end{aligned} \tag{2.56}$$

This has the appearance of a reasonable approximation, especially given the fact that the exact correlation energy can be written in terms of the Hamiltonian matrix elements between the reference ground state and the doubly-excited determinants (Eq. 2.55). However, serious deficiencies of truncated CI methods are revealed through the consideration of the following example: Suppose a particular system can be partitioned into M noninteracting subsystems (the example given in Ref. [6] is M noninteracting H_2 molecules). Since the subsystems do not interact with each other, the total energy of the full system would be equal to the sum of the subsystem energies, and the wave function of the full system would be the antisymmetrized (direct) product of the subsystem wave functions. If we obtain each subsystem wave function from CISD, each subsystem wave function would take the form of Eq. 2.56. However, the full

¹⁴On the bright side, this procedure can be used to obtain the entire spectrum of excited electronic state energies in addition to the ground state energy.

system wave function obtained through CISD would not be the same as the antisymmetrized product of subsystem CISD wave functions, and as a result, the full system energy from CISD would not be the sum of the subsystem CISD energies (CISD, along with other truncated CI methods, are not *size consistent* for this reason[6, 15, 16]). For example, taking a direct product of two subsystem CISD wave functions results in triply- and quadruply-excited determinants, due to products of singly- with doubly-excited subsystem determinants, and products of doubly- with doubly-excited subsystem determinants, respectively. These triply- and quadruply-excited determinants are due to uncorrelated couplings within separate, noninteracting subsystems, but are nonetheless truncated out of the full system CISD wave function. In the general M subsystem case, the antisymmetrized product of subsystem CISD wave functions will include $2M$ -tuply excited determinants, so the resulting error scales with the system size.[6, 15, 16] Note that the example of noninteracting subsystems is used for illustrative purposes, and these failures of the CISD method would exist between interacting subsystems, and within the subsystems as well.

Each m -tuply excited determinant describes some instantaneous m -electron interaction, capturing effects due to the correlations between the m electrons. However, it turns out that in insulating molecular systems (in the ground electronic state) whose electrons remain relatively localized, these instantaneous m -electron interactions are mostly due to uncorrelated (but simultaneous) 2- and 3-electron interactions,[17, 16] such as those arising from antisymmetrized products of subsystem CISD wave functions. This apparent limitation on the complexity of the many-electron correlations in molecular systems can be explained by considering the short range of the *fluctuation potential*:[17]

$$\mathcal{V} = \sum_{a < b} \frac{1}{||\mathbf{r}_a - \mathbf{r}_b||^2} - (\hat{j}_a(\mathbf{r}_b) - \hat{k}_a(\mathbf{r}_b)) - (\hat{j}_b(\mathbf{r}_a) - \hat{k}_b(\mathbf{r}_a)) \quad (2.57)$$

with $\hat{j}_a(\mathbf{r}_b)$ and $\hat{k}_a(\mathbf{r}_b)$ being the HF Coulomb and exchange potentials on electron b , due to electron a , and $\hat{j}_b(\mathbf{r}_a)$ and $\hat{k}_b(\mathbf{r}_a)$ being the HF Coulomb and exchange potentials on electron a , due to electron b (we need both pairs because of the restricted sum (with summation indices

$a < b$). The fluctuation potential is the correction to the HF Hamiltonian needed to obtain the exact electronic Hamiltonian. Refs. 17 and 18 shows that in the simple case of the Be atom, the fluctuation potential decays to zero within the length scales of the orbital shell structure. In other words, the Pauli exclusion principle separates non-paired pairs of electrons beyond the reach of the fluctuation potential. However, for more complicated molecular systems (than the Be atom example), 3-electron (and sometimes 4-electron) interactions must be considered to achieve sufficient accuracy.[19]

The problem of size consistency is addressed in the *coupled cluster* methods. Defining the excitation operator, $\hat{T}$, as:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_N \quad (2.58)$$

with $\hat{T}_n$ denoting the n^{th} order excitation operator, such that:¹⁵

$$\hat{T}_1 = \sum_{i,p} t_i^p \hat{a}_p^\dagger \hat{a}_i \quad (2.59a)$$

$$\hat{T}_2 = \sum_{\substack{i < j \\ p < q}} t_{ij}^{pq} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_j \hat{a}_i \quad (2.59b)$$

$$\hat{T}_3 = \sum_{\substack{i < j < k \\ p < q < r}} t_{ijk}^{pqr} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r^\dagger \hat{a}_k \hat{a}_j \hat{a}_i \quad (2.59c)$$

$$\hat{T}_N = \sum_{e_1 < e_2 < \cdots < e_N} t_{o_1 o_2 \cdots o_N}^{e_1 e_2 \cdots e_N} \hat{a}_{e_1}^\dagger \hat{a}_{e_2}^\dagger \cdots \hat{a}_{e_N}^\dagger \hat{a}_{o_N} \cdots \hat{a}_{o_2} \hat{a}_{o_1} \quad (2.59d)$$

where the creation and annihilation operators together act as excitation operators, for example with $\hat{a}_p^\dagger \hat{a}_i$ exciting an electron from occupied orbital i to unoccupied (excited) orbital p (see Eq. 2.31), and t_i^p is an expansion coefficient (often called an “amplitude”) for the excitation due to the corresponding $i \rightarrow p$ excitation. Keep in mind that the m^{th} order “excitations” are

¹⁵Here, we don’t need to be careful to sum over occupied and unoccupied orbitals separately (the sum over i vs the sum over p in the definition of $\hat{T}_1$, for example) because the creation and annihilation operators ensure that we cannot create electrons that already exist, and annihilate electrons that do not yet exist (Eqs. 2.29b and 2.29d).

excitations relative to the uncorrelated HF reference state, and therefore each describe some kind of m -electron correlation.

Making use of the definition of the exponential map of an operator, $\hat{A}$:

$$\exp[\hat{A}] = \mathbb{1} + \hat{A} + \frac{1}{2!}\hat{A}^2 + \frac{1}{3!}\hat{A}^3 + \dots \quad (2.60)$$

and with an appropriate choice in t coefficients, we can write the exact full CI wave function as $|(\Phi_{\text{FCI}})_0\rangle = \exp[\hat{T}] |(\Phi_{\text{HF}})_0\rangle$. Approximating the full excitation operator as a sum of the single- and double-excitation operators, $\hat{T}_{\text{SD}} = \hat{T}_1 + \hat{T}_2$, and applying its exponential to the HF reference state, we obtain the coupled cluster singles doubles (CCSD) wave function, $|(\Phi_{\text{CCSD}})_0\rangle$:

$$\begin{aligned} |(\Phi_{\text{CCSD}})_0\rangle &= \exp[\hat{T}_1 + \hat{T}_2] |(\Phi_{\text{HF}})_0\rangle \\ &= [\mathbb{1} + (\hat{T}_1 + \hat{T}_2) + \frac{1}{2!}(\hat{T}_1^2 + 2\hat{T}_1\hat{T}_2 + \hat{T}_2^2) \\ &\quad + \frac{1}{3!}(\hat{T}_1^3 + 3\hat{T}_1^2\hat{T}_2 + 3\hat{T}_1\hat{T}_2^2 + \hat{T}_2^3) \\ &\quad + \frac{1}{4!}(\hat{T}_1^4 + 4\hat{T}_1^3\hat{T}_2 + 6\hat{T}_1^2\hat{T}_2^2 + 4\hat{T}_1\hat{T}_2^3 + \hat{T}_2^4) + \dots] |(\Phi_{\text{HF}})_0\rangle \end{aligned} \quad (2.61)$$

Recognizing that $\hat{T}_n^k$ corresponds to an $(nk)^{\text{th}}$ order excitation, we can rearrange the terms in Eq. 2.61, grouping them by excitation order:

$$\begin{aligned} |(\Phi_{\text{CCSD}})_0\rangle &= \exp[\hat{T}_1 + \hat{T}_2] |(\Phi_{\text{HF}})_0\rangle \\ &= [\mathbb{1} + \hat{T}_1 + (\hat{T}_2 + \frac{1}{2}\hat{T}_1^2) + (\hat{T}_1\hat{T}_2 + \frac{1}{6}\hat{T}_1^3) \\ &\quad + (\frac{1}{2}\hat{T}_2^2 + \frac{1}{2}\hat{T}_1^2\hat{T}_2 + \frac{1}{6}\hat{T}_1^4) + \dots] |(\Phi_{\text{HF}})_0\rangle \end{aligned} \quad (2.62)$$

Taking a closer look at $\frac{1}{2}\hat{T}_1^2$ (for example), and rewriting it in terms of creation and

annihilation operators, we obtain:

$$\begin{aligned}
\frac{1}{2}\hat{T}_1^2 &= \frac{1}{2} \sum_{i,p} \sum_{j,q} t_i^p t_j^q \hat{a}_p^\dagger \hat{a}_i^\dagger \hat{a}_q \hat{a}_j \\
&= \sum_{\substack{i < j \\ p < q}} (t_i^p t_j^q - t_i^q t_j^p) \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_j \hat{a}_i
\end{aligned} \tag{2.63}$$

where we combined the two double sums into a restricted quadruple sum (going from the first line to the second line), leading to an extra factor of 2, and a coefficient (amplitude) of the form, $t_i^p t_j^q - t_i^q t_j^p$, which accounts for the antisymmetry with respect to pairwise exchange of electrons. More explicitly, expanding out the two double sums (in the top line of Eq. 2.63) and focusing on terms involving a particular choice in occupied orbitals labeled by i and j , and unoccupied orbitals labeled by p and q , we have:

$$\begin{aligned}
&t_i^p t_j^q \hat{a}_p^\dagger \hat{a}_i^\dagger \hat{a}_q \hat{a}_j + t_i^q t_j^p \hat{a}_q^\dagger \hat{a}_i^\dagger \hat{a}_p \hat{a}_j + t_j^p t_i^q \hat{a}_p^\dagger \hat{a}_j \hat{a}_q^\dagger \hat{a}_i + t_j^q t_i^p \hat{a}_q^\dagger \hat{a}_j \hat{a}_p^\dagger \hat{a}_i \\
&= t_i^p t_j^q \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_j \hat{a}_i - t_i^q t_j^p \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_j \hat{a}_i - t_j^p t_i^q \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_j \hat{a}_i + t_j^q t_i^p \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_j \hat{a}_i \\
&= (t_i^p t_j^q - t_i^q t_j^p) \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_j \hat{a}_i
\end{aligned} \tag{2.64}$$

where the sign of each term must account for the exchange antisymmetry when rearranging creation and annihilation operators going from the first line to the second line (see Eqs. 2.29a–2.29d, and for more detail, see Ref. [6]), and we arbitrarily absorb a factor of -1 into the coefficients going from the second to third line.

Using Eq. 2.63, and the definition of $\hat{T}_2$ from Eq. 2.59b, we find that the total doubly-excited term in the CCSD operator, $\hat{T}_2 + \frac{1}{2}\hat{T}_1^2$ (Eq. 2.62), can be written as:

$$\hat{T}_2 + \frac{1}{2}\hat{T}_1^2 = \sum_{\substack{i < j \\ p < q}} (t_{ij}^{pq} + t_i^p t_j^q - t_i^q t_j^p) \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_j \hat{a}_i \tag{2.65}$$

The terms due to $\hat{T}_1^2$ are called disconnected double excitations, with amplitudes corresponding to the product of the corresponding single excitation amplitudes, in contrast to the terms due

to $\hat{T}_2$, which are called connected double excitations, and have their own dedicated amplitudes. Singly-excited determinants relative to the HF reference state aren't readily interpreted in terms of electron correlations, because it doesn't make sense to think about single-electron correlations (remember that the singly-excited determinants are important to the full CI ground state due to their coupling to the doubly- and triply-excited determinants). However, using the analogy of literal excitations, the disconnected double excitations can be thought of as arising from independent single excitations occurring simultaneously, whereas the connected double excitations correspond to a true 2-electron interaction.

Returning to the example of noninteracting subsystems which demonstrated the deficiencies of the (truncated) CISD method, we can think about quadruple excitations due to the disconnected $\hat{T}_2^2$ operator in comparison to the connected $\hat{T}_4$ operator. The amplitudes for the disconnected $\hat{T}_2^2$ excitations (for example those in the CCSD wave function) are antisymmetrized products of pairs of (connected) doubly-excited amplitudes, $t_{ij}^{pq} * t_{kl}^{rs} = t_{ij}^{pq} t_{kl}^{rs} - t_{ik}^{pq} t_{jl}^{rs} + t_{il}^{pq} t_{jk}^{rs} - t_{ij}^{pr} t_{kl}^{qs} + \dots + t_{il}^{qr} t_{jk}^{ps}$ (with 18 total terms in the antisymmetrized product[6]). Taking the quadruply-excited CI coefficients, c_{ijkl}^{pqrs} , to be the (antisymmetrized) product of doubly-excited amplitudes (i.e. the approximation that $c_{ijkl}^{pqrs} = t_{ij}^{pq} * t_{kl}^{rs}$) can be interpreted as the approximation that simultaneous 4-electron correlations are (mostly) due to independent, but simultaneous, pairs of 2-electron correlations.

Coupled cluster is an example of a decomposition scheme where resummation of the canonical correlated wave function series (i.e. the CI series) in terms of inherent n -particle correlations (i.e. connected excitations) vastly improves the series convergence (recall that CI isn't even size consistent). Coupled cluster is accurate for the systems that we are interested in here (e.g. aqueous systems in ambient conditions), so we will often use it to calculate reference electronic energies in the following chapters. In particular, we will make use of coupled cluster with single, double, and perturbative triple excitations, i.e., CCSD(T),[287, 288] as well as the explicitly correlated CCSD(T)-f12b method.[290, 291]

2.3 Density Functional Theory

While full CI, coupled cluster, and other correlated wave function methods provide the means for computing electronic energies with high accuracy, they can be prohibitively expensive for even modestly sized systems. *Density functional theory* (DFT) takes a different approach, bypassing the electronic wave function, and instead using the electron density to represent the electronic state of a given system. For some normalized N -electron wave function, $\Phi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$, the corresponding electron density, $\rho(\mathbf{r})$ is obtained by integrating its modulus squared over the position and spin coordinates of electrons 2 through N , as well as the spin coordinate, ω_1 , of electron 1:

$$\rho(\mathbf{r}) = N \int |\Phi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)|^2 d\omega_1 d\mathbf{x}_2 d\mathbf{x}_3 \cdots d\mathbf{x}_N \quad (2.66)$$

where we take the position coordinates of electron 1 to be $\mathbf{r} = \mathbf{r}_1$, which becomes the sole argument for the density, $\rho(\mathbf{r})$, after integrating out all other relevant coordinates. Note that due to the indistinguishability of the electrons, we can arbitrarily choose the position of electron 1 (or the position of any other single electron) such that $\rho(\mathbf{r})$ describes the density of the entire system. Additionally, notice that the factor of N is needed to ensure that integrating $\rho(\mathbf{r})$ over all space evaluates to the total number of electrons, N :

$$\int_{\mathbb{R}^3} \rho(\mathbf{r}) d\mathbf{r} = N \quad (2.67)$$

so that integrating over any arbitrary region of space evaluates to the probability of finding one electron within that region. The electron density depends on three spatial coordinates, and is therefore easier to work with than the electronic wave function, which depends on $3N$ spatial coordinates (and N spin coordinates¹⁶).

Hohenberg and Kohn[20, 10] proved the so-called *Hohenberg-Kohn theorems* (HK) that

¹⁶Many DFT methods handle the electron density of spin up and spin down electrons separately.

there exists a unique functional of the electronic density, $E_{\text{vext}}^{(\text{exact})}[\rho]$, that maps the N -electron density to the corresponding electronic energy, given an “external potential,” $v_{\text{ext}}(\mathbf{r})$ (due to the electron-nucleus attraction energy, as well as any external electric or magnetic fields, etc.). We refer to this functional as the *exact (density) functional*. Furthermore, they proved that minimizing the exact functional with respect to the electron density (constrained to integrate to N , as in Eq. 2.67) yields the exact ground electronic state energy. More specifically, the exact functional can be written as:[20]

$$E_{\text{vext}}^{(\text{exact})}[\rho] = \int v_{\text{ext}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + G[\rho] \quad (2.68)$$

where the first term on the right hand side describes the interaction of the electron density with the external potential (including the nuclei), the second term represents the classical Coulomb interaction between the electron density with itself, $\mathcal{J}[\rho]$ (i.e. the classical self-repulsion energy, as was described in Sec. 2.2.5), and $G[\rho]$ is a universal functional (i.e. it does not need to be tailored to specific systems or applications) which accounts for the electron kinetic energy, and corrects the classical Coulomb energy, recovering the exact exchange and correlation energies for the electron-electron interactions:[20, 10]

$$\begin{aligned} G[\rho] &= T[\rho] + E_{xc}^{(HK)}[\rho] \\ &= T[\rho] + V_{\text{el-el}}^{\text{exact}} - \mathcal{J}[\rho] \end{aligned} \quad (2.69)$$

where $T[\rho]$ describes the total kinetic energy of the N -electron system described by density, ρ , and $E_{xc}^{(HK)}[\rho]$, called the *exchange-correlation (XC) functional*, which describes the total electronic correlation energy including both exchange effects, and the correlation energy captured by the full CI method. It is important to emphasize that in our discussion on wave function methods in the previous sections, we used the term “exchange correlation” to describe electron correlations due to purely exchange effects, while in DFT, we use the term “exchange-correlation” to refer to (the total) exchange + correlation energies. On the second line of Eq. 2.69, we rewrite

$E_{xc}^{(HK)}[\rho]$ explicitly as the difference between exact electron-electron energy functional, $V_{\text{el-el}}^{\text{exact}}$, and the classical self-repulsion energy functional, $\mathcal{J}[\rho]$.

The difficulty with making use of the HK theorems, of course, is in defining $G[\rho]$, which in theory describes all of the effects captured by the full CI method. The proof that such a functional exists is certainly reassuring, however the HK theorems make no statements regarding how one might obtain the exact functional, nor about the difficulty of applying it to real calculations. Given that the electronic TISE is a fundamentally difficult problem (due to its many-body nature), it is hard to believe that solving the exact density functional equations would be significantly easier than approaching the problem with wave function methods. However, approximations to the exact functional have proven to be extremely useful in quantum chemical calculations, being able to capture electron correlation effects at a computational cost that is comparable to HF (which doesn't capture any correlation effects, apart from the exchange energy), with many approximate functionals being even cheaper than HF.

2.3.1 The Kohn-Sham Equations

A major difficulty with obtaining the exact density functional is the presence of nonlocal operators in the Hamiltonian. In particular, the kinetic energy operator, $\hat{T} = (-\hbar^2/2m)\nabla^2$, is nonlocal, $\langle \mathbf{r} | \hat{T} | \mathbf{r}' \rangle \neq \langle \mathbf{r} | \hat{T} | \mathbf{r} \rangle \delta(\mathbf{r} - \mathbf{r}')$, due to the presence of gradient operators, which cannot be diagonalized in the position basis. This poses serious complications in the search for a kinetic energy density functional.[10]

Kohn and Sham proposed a solution to this problem in Ref. 21, in which the electronic kinetic energy of an N -electron density is approximated to be the exact kinetic energy, $T_s[\rho]$, of a fictitious noninteracting N -electron system that has the same density as the real system under question (and feels the same external potential, v_{ext}). The noninteracting system is taken to be described by a single N -electron Slater determinant of 1-electron spin orbitals within the span of some basis set approximation. The evaluation of $T_s[\rho]$ is identical to the evaluation of the kinetic energy in HF theory, once we have deter-

mined the determinant that reproduces the density, $\rho(\mathbf{r})$. Note that for a single determinant, $|\psi_1\psi_2\cdots\psi_N\rangle = |\phi_1(\mathbf{r}_1)\Omega_1(\omega_1), \phi_2(\mathbf{r}_2)\Omega_2(\omega_2), \dots, \phi_N(\mathbf{r}_N)\Omega_N(\omega_N)\rangle$, with each spin orbital, $\psi_i(\mathbf{x}_i)$, equal to the product of a spatial function, $\phi_i(\mathbf{r}_i)$, and a spin function, $\Omega_i(\omega_i)$, the corresponding electron density is equal to (Eq. 2.66):

$$|\phi_1(\mathbf{r}_1)\Omega_1(\omega_1), \phi_2(\mathbf{r}_2)\Omega_2(\omega_2) \cdots \phi_N(\mathbf{r}_N)\Omega_N(\omega_N)\rangle \rightarrow \rho(\mathbf{r}) = \sum_{i=1}^N |\phi_i(\mathbf{r})|^2 \quad (2.70)$$

Based on Eq. 2.70, it is clear that the density resulting from any arbitrary correlated wave function (e.g. from full CI), can be reproduced by a single determinant within the span of the same basis set approximation, given that the density of a single determinant is simply the superposition of the orbital densities. In fact, there are an infinite number of single determinants that can reproduce the density of an arbitrary N -electron wave function, as is discussed in Ref. [10].

Using the construction of Kohn and Sham, we can write down the following density functional (see Eq. 2.68):

$$E_{v_{\text{ext}}}[\rho] = \sum_{i=1}^N \langle \phi_i | \frac{-\hbar^2}{2m} \nabla^2 | \phi_i \rangle + \int v_{\text{ext}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{\text{xc}}[\rho] \quad (2.71)$$

where the first term on the right hand side corresponds to the kinetic energy of the fictitious noninteracting N -electron system described by a single Slater determinant composed of 1-electron spatial orbitals $\phi_1, \phi_2, \dots, \phi_N$, the second and third terms correspond to the energies due to the external potential, and due to the classical (Coulombic) self-repulsion energy of the density, and $E_{\text{xc}}[\rho]$ is the Kohn-Sham XC-functional, which includes the correction to the approximate noninteracting kinetic energy functional, in addition to the XC correction to the electron-electron interaction functional:[10]

$$E_{\text{xc}}[\rho] = T[\rho] - T_s[\rho] + V_{\text{el-el}}^{\text{exact}} - \mathcal{J}[\rho] \quad (2.72)$$

Although the KS approach (usually called KS DFT) sacrifices the simplicity of using

the electron density as the sole representation of the electronic states, the reformulation of the N -electron problem as a 1-electron problem represents a very reasonable compromise, especially given the fact that the correlation energy can be approximately recovered without requiring significant computational cost. In fact, KS DFT would be an exact method if the exact form of $E_{xc}[\rho]$ is used. The objective in KS DFT parallels that of HF theory, namely that we minimize the energy functional, $E_{v_{ext}}[\rho]$, relative to the electron density, $\rho(\mathbf{r})$, subject to the constraint that the integral of the density over all space evaluates to the number of electrons, N (Eq. 2.67). Like in HF theory, the minimization is carried out with the method of Lagrange multipliers.[10]

Skipping the details, we define the effective potential, $v_{eff}(\mathbf{r})$, to be:[10]

$$\begin{aligned} v_{eff}(\mathbf{r}) &= v_{ext}(\mathbf{r}) + \frac{\delta}{\delta\rho(\mathbf{r})} \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \frac{\delta E_{xc}[\rho]}{\delta\rho(\mathbf{r})} \\ &= v_{ext}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + v_{xc} \end{aligned} \quad (2.73)$$

where the second and third terms on the right-hand side of the first line are the functional derivatives (with respect to the electronic density) of the classical Coulomb energy and (KS) exchange-correlation energies, respectively. With this definition of the KS effective potential, we can define the following effective 1-electron TISE:[10]

$$\left(-\frac{\hbar^2}{2m} \nabla^2 + v_{eff}(\mathbf{r}) \right) \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \quad (2.74)$$

Eqs. 2.70, 2.73, and 2.74 are called the KS equations, and must be solved self-consistently by an iterative procedure that parallels that used in HF theory (see Eq. 2.48).[10]

2.3.2 Density Functional Approximations

KS DFT allows for an exact treatment of the electronic TISE, provided the exact KS XC-functional (Eq. 2.72) is used. However, the exact XC-functional is unknown, so we must rely on approximations.

Density functional approximations (DFAs), E_{xc}^{DFA} , to the exact XC-functional can be classified by the ingredients used in their functional forms, forming a hierarchy of *classes*, with different levels of complexity, dubbed “Jacob’s Ladder of Density Functional Approximations.”[22] The dependence on these various ingredients allows for a parameterization of a DFA which satisfies known properties of the exact XC-functional.[22]

At the lowest level of the hierarchy, XC-functionals only depend on the local electron density, $n(\mathbf{r})$. Functionals belonging to this class are called Local Density Approximations (LDA), or in general, Local Spin Density Approximations (LSDA) which depend on the electron density of spin-up and spin-down electrons, $n(\mathbf{r}) = n_{\uparrow}(\mathbf{r}) + n_{\downarrow}(\mathbf{r})$. Correspondingly, their functional form, $E_{xc}^{LDA}[n_{\uparrow}(\mathbf{r}), n_{\downarrow}(\mathbf{r})]$, describes the XC energy of a uniform electron gas at each point in space, since the local density, $n(\mathbf{r}')$, at position $\mathbf{r}'$ contains no information about the neighborhood around $\mathbf{r}'$. Despite their simplicity, LDA functionals perform well on solids and solid surfaces, in part due to some error cancellation between the exchange and correlation energies,[10] but also due to the fact that they satisfy a number of exact properties, such as the Lieb-Oxford lower bound, the exact exchange spin-scaling relation and uniform-density-scaling relation, the exchange-correlation hole sum rule, the exact exchange hole negativity constraint and cusp condition, and has the correct correlation energy scaling in the low-density limit.[22] While LDA can be useful for solids, the use of the uniform electron gas exchange-correlation energy is less applicable for many finite molecular systems, and will not be discussed further.

The second level of the hierarchy corresponds to the generalized gradient approximation (GGA) class of functionals, whose XC-functionals depend on both the local electron density, $n_{\uparrow}(\mathbf{r})$ and $n_{\downarrow}(\mathbf{r})$, as well as the gradient of the electron density, $\nabla n_{\uparrow}(\mathbf{r})$ and $\nabla n_{\downarrow}(\mathbf{r})$. GGA functionals such as BLYP,[23, 24] PBE,[25] and revPBE[26] are widely used in quantum chemical calculations today. The use of the electron density gradient allows GGA functionals to satisfy more exact properties of the true XC-functional than can be satisfied with the LDA, for example properties of the exact correlation energy due to nonuniform scaling of the electron density, and the behavior of the exact correlation energy in the high electron density limit for

uniform scaling.[22, 27] Additionally, the electron density gradient, $\nabla n(\mathbf{r}')$, at position $\mathbf{r}'$ carries information about the electron density in the neighborhood around $\mathbf{r}'$, giving GGA functionals more nonlocal character. The nonlocality is limited to the infinitesimal neighborhood around $\mathbf{r}'$, so GGA functionals are usually classified as *semilocal* functionals.

The third level of the hierarchy corresponds to the meta generalized gradient approximation (mGGA) class of functionals. The mGGA XC-functionals depend on the local electron density, the density gradient, and either the Laplacian (second derivative) of the density or the kinetic energy density, $\tau(\mathbf{r}) = (1/2) \sum_i |\nabla \phi_i|^2$. The use of the kinetic energy density allows for the development of self-correlation-free DFAs.[22] Additionally, the use of the kinetic energy density allows mGGA DFAs to reproduce intermediate-range dispersion interactions.[28] mGGA functionals studied here include SCAN,[29] TPSS,[30] and B97M-rV.[31, 32]

A helpful analogy for understanding differences in the limitations in accuracy inherent to these first three DFA classes is the comparison of different truncations of the Taylor expansion of some function. An analytic function can be expressed exactly, in terms of its derivatives at some expansion point, in the form of a polynomial (the Taylor series). Truncating the Taylor series of a function at a given term (i.e. at a given derivative order) results in an approximation of the function in the neighborhood around the expansion point, with the approximation improving with the inclusion of more terms in the series. In the context of this analogy, LDA, GGA, and mGGA functionals represent increasingly accurate truncations of a Taylor series. This analogy is in fact substantive, given that the inclusion of terms due to the density gradient and Laplacian (or kinetic energy density) allows the functional to reproduce properties from the exact gradient expansions of the exchange and correlation energies.[22]

The fourth class of DFAs are called (global) *hybrid* functionals.[22] Hybrid functionals are called ‘hybrids’ because their functional forms combine that of a pure density functional, which only depends on the electron density and its derivatives (such as DFAs in the first three classes), with the exchange energy from HF (the K_{ij} integral in Eq. 2.39), which, in the context of DFT, is referred to as “exact exchange.” The inclusion of exact exchange makes hybrid

DFA nonlocal functionals (see Sec. 2.2.5), requiring a double integral to evaluate the exact exchange energy. Therefore, hybrid functionals require a greater computational cost than semilocal functionals, but are generally no more expensive than HF. One important feature of exact exchange is that it corrects for the (classical) Coulomb interaction of the 1-electron orbitals with themselves, due to the fact that $K_{ii} = J_{ii}$ in Eq. 2.39, mitigating self-interaction and delocalization errors.[11, 12, 13, 14] Exact exchange can be included in a DFA whose semilocal terms depend on the density, as well as the density gradient, and/or the kinetic energy density, leading to hybrid-GGA and hybrid-mGGA XC-functionals. Hybrid-GGA functionals studied here include B3LYP,[33] PBE0,[34] and revPBE0, and hybrid-mGGA functionals studied here include M06-2X.[35]

The final class considered here is the range-separated hybrid (sometimes called “Coulomb-attenuated” hybrids). Range-separated hybrids use semilocal exchange at close-range, and exact exchange at long-range, smoothly transitioning between the two regimes with the error function and the complementary error function.[36, 37] This is in contrast with the global hybrids discussed in the previous paragraph, which include a fixed fraction of exact exchange at all distances. This feature allows range-separated hybrids to reproduce the correct long-range behavior of the exchange potential,[37] in particular correcting for long-range self interaction errors.[38] As a result, range-separated hybrid DFAs have smaller delocalization errors than semilocal functionals and global hybrids,[14] and therefore yield more accurate charge transfer and dissociation energies.[37, 14] The range-separated hybrid functionals studied here include ω B97X,[39] ω B97X-D,[40] and ω B97M-V.[31, 41]

Chapter 3

Many-Body Interactions

We have seen that in situations where the Born-Oppenheimer approximation is valid, the total electronic energy of a system (for a particular nuclear configuration) can be seen as a potential energy surface which determines the dynamics of the nuclei in the system (similar to how the potential energy of a spring determines the dynamics of a mass attached to the spring). While the total electronic energy is very useful for this reason, it can be a bit cumbersome in very large systems. Therefore, it is desirable to decompose the total electronic energy into smaller, more manageable components, which naturally leads us to the concept of many-body interaction energies.

In order to proceed, we must first determine how to partition our system into “monomer” units (i.e. the ‘bodies’ that collectively constitute the ‘many bodies’ of the full system). There are many ways that we can define the monomers in a system, for example we can define each “atom” to be its own monomer,[42, 43] or alternatively, we can define each molecule to be a monomer. In correlated electronic structure theories, the electrons are considered to be the monomers in a many-electron wave function.¹ Although there is some freedom in the choice of monomer definitions, we must be careful to pick definitions that are best suited for the particular applications that we are interested in. For example, if we wish to develop models to describe the interactions of small molecules with each other (for example, the interactions between two water

¹Electrons have no internal structure, so the interactions are strictly pairwise. Therefore, in the case of many-electron systems (considering electrons to be “monomers”), we are not interested in many-body interactions, only the many-body correlations.

molecules), then it might be sufficient to define each molecule to be a single monomer. However, if we wish to develop models to study biochemical systems, which could involve thousands or millions of atoms, then we need to be more careful in determining which questions that we want to answer. If we are interested in simulating the binding of some drug molecule to a protein, then we might want a very high-resolution model in which we define each amino acid in the protein to be a single monomer. If instead we are interested in simulating larger scale dynamics in a cell involving many proteins and other biomolecules, then the system may be intractable unless we define entire proteins, or even clusters of proteins, to be the monomer units.

Here, we only discuss systems of small molecules, so we will define monomer units to be single molecules (or monoatomic ion monomer units when discussing halide–water and alkali ion–water interactions).

3.1 Intermolecular Interactions and The Many-Body Expansion

3.1.1 Monomer Distortion Energies and The Binding Energy

Once we have defined the monomer units in our system, we can immediately simplify matters by considering a quantity called the *binding energy* rather than the total electronic energy. For simplicity, let's imagine a single monomer in a particular fixed configuration, $\mathbf{R}$. We can perform a geometry optimization of the monomer to find its global minimum energy configuration, $\mathbf{R}^{(0)}$ (i.e. the absolute lowest energy structure). Subtracting the total electronic energy of the optimized monomer, $E_{\text{tot}}(\mathbf{R}^{(0)})$, from the total electronic energy of the monomer in configuration $\mathbf{R}$, $E_{\text{tot}}(\mathbf{R})$, we obtain the distortion energy of the monomer, often referred to as the one-body distortion energy (or 1B energy), $E_{1\text{B}}$:

$$E_{1\text{B}} = E_{\text{tot}}(\mathbf{R}) - E_{\text{tot}}(\mathbf{R}^{(0)}) \quad (3.1)$$

The 1B energy of a monomer in a particular configuration is the energy that it takes to distort the monomer from its lowest-energy structure into the particular configuration under question. Since the energy of the optimized monomer is the smallest possible energy for that monomer, E_{1B} is exactly 0 for the optimized configuration, and strictly positive for any other configuration (assuming there is only one global minimum configuration).

Now we consider an arbitrary system containing N monomers. If we take the total electronic energy of the entire system, $E_{\text{tot}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N)$, where $\mathbf{R}_m$ is the nuclear configuration of monomer m , and subtract the sum of the energies of each optimized monomer, $E_{\text{tot}}(\mathbf{R}_1^{(0)})$, $E_{\text{tot}}(\mathbf{R}_2^{(0)})$, etc., then we obtain the total binding energy of the system, E_{bind} :

$$E_{\text{bind}} = E_{\text{tot}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) - \sum_{m=1}^N E_{\text{tot}}(\mathbf{R}_m^{(0)}) \quad (3.2)$$

Notice that all of the optimized monomer energies are constants, which means that the binding energy differs from the total energy of the full system by a fixed additive constant, which means that the binding energy carries the exact same information as the total electronic energy (assuming the Born-Oppenheimer approximation is valid), so we may choose to use it as our PES for convenience. It is important to note, however, that the binding energy depends on the definition of the monomer units, so these definitions must be consistent for the binding energy to be meaningful.

3.1.2 The Interaction Energy

The next quantity that we need to consider is the *interaction energy*. As the name suggests, the interaction energy, E_{int} , contains information about the interactions between the monomers in a given system. We can obtain E_{int} for a system of monomers in some fixed configuration by subtracting the sum of the 1B monomer distortion energies from E_{bind} :

$$E_{\text{int}} = E_{\text{bind}} - \sum_{m=1}^N E_{1B}(\mathbf{R}_m) \quad (3.3)$$

By inserting Eq. 3.2 into Eq. 3.3, we can see that one may also obtain E_{int} by subtracting the sum of the total energies of the monomers in isolation, $E_{\text{tot}}(R_1)$, $E_{\text{tot}}(R_2)$, etc. from the total energy of the full system:

$$E_{\text{int}} = E_{\text{tot}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) - \sum_{m=1}^N E_{\text{tot}}(\mathbf{R}_m) \quad (3.4)$$

This is analogous to the definition of E_{bind} , except we are subtracting the energies of the monomers in their distorted configurations, rather than subtracting their minimum energies (and as was the case for the binding energy, the interaction energy depends on the definition of the monomer units). Therefore E_{bind} contains contributions from the monomer distortion energies, while E_{int} does not. The rationale for this definition is that the interactions between the monomers are independent of the distortions of the monomers. E_{int} will be equal to E_{bind} when all monomers in the system are in their optimized configurations, or when all monomers are defined to be structureless particles that can not be distorted.

One interpretation of the interaction energy is that it is the energy required to take the system in some configuration, $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N$, and separate all monomers infinitely far from each other (infinitely distant monomers do not interact with each other), or equivalently to take infinitely distant, noninteracting (distorted) monomers and move them into the configuration of interest. Analogously, the binding energy can be thought of in the same way, except we include the energy due to relaxing the monomers into their minimum energy configurations. Based on these interpretations, it is clear that a configuration with a negative binding energy will be in a bound state because that configuration is lower in energy than the dissociated, noninteracting monomers in their minimum energy configurations.

The difference in interaction energies between two configurations of the same system does not quantify the relative stabilities between the two configurations, as is the case for the difference in binding energies. Recall that E_{bind} is effectively equivalent to the total energy since it differs only by an additive constant, containing contributions from all inter-monomer interactions as well as contributions from monomer distortions, while E_{int} contains no contributions from

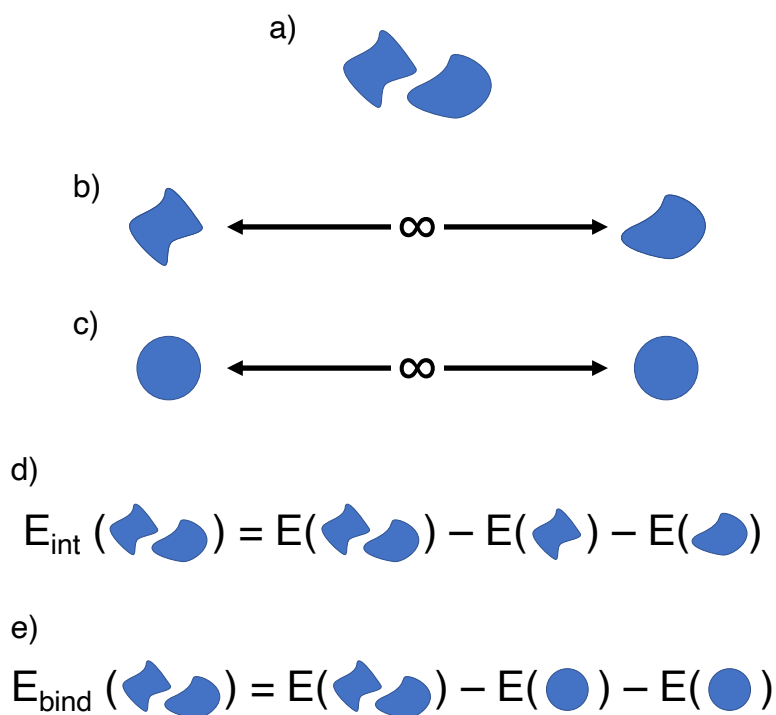


Figure 3.1. Cartoon depicting the interpretation of the interaction energy of a cluster (a dimer in the example) as the energy required to separate the distorted monomer units from the cluster geometry infinitely far apart from each other in the noninteracting limit (panel a $\rightarrow$ b), and the interpretation of the binding energy as the energy required to separate the monomers and relax them into their optimized geometries (panel a $\rightarrow$ c). Circles correspond to optimized monomers, and the distorted shapes correspond to distorted monomers. The equations for the interaction and binding energies are shown pictorially in panels d and e, respectively.

monomer distortions. One will frequently encounter configurations with very attractive (i.e. very negative) interaction energies, but relatively high binding energies. This is often due to the fact that highly-distorted monomers may have unusually large charge separation (for example from stretched covalent bonds), which results in large dipoles that may have strong electrostatic interactions with the surrounding monomers, but the enhanced interaction energies will often be overpowered by large monomer distortion energies.

3.1.3 The Many-Body Expansion

Now that we have defined the interaction energy, E_{int} , we can begin to dissect it into its many-body components. From Eqs. 3.3 and 3.4, we can see that E_{int} quantifies the total interactions of the entire system. However, the entire system might be rather large (containing many monomers), so it is often helpful to first think about interactions between all pairs of monomers. If our system contains N monomers in some fixed configuration, $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N$, then we can imagine removing all monomers except for monomers 1 and 2, which remain in configurations, $\mathbf{R}_1$ and $\mathbf{R}_2$ (i.e. we are considering monomers 1 and 2 in isolation from any other monomers). In this context, we refer to the total system as the *parent* cluster (or parent system), and we refer to its subsystems as *fragments*. More specifically, a subsystem with two monomers is called a dimer (or 2-mer). The interaction energy of this dimer fragment can now be thought of as the 2-body (2B) energy of the dimer:

$$E_{\text{int}}(\mathbf{R}_1, \mathbf{R}_2) = E_{2\text{B}}(\mathbf{R}_1, \mathbf{R}_2) = E_{\text{tot}}(\mathbf{R}_1, \mathbf{R}_2) - E_{\text{tot}}(\mathbf{R}_1) - E_{\text{tot}}(\mathbf{R}_2) \quad (3.5)$$

It is important to keep in mind that the total energy of the dimer is equal to the sum of the monomer energies and the 2B energy:

$$E_{\text{tot}}(\mathbf{R}_1, \mathbf{R}_2) = E_{2\text{B}}(\mathbf{R}_1, \mathbf{R}_2) + E_{\text{tot}}(\mathbf{R}_1) + E_{\text{tot}}(\mathbf{R}_2) \quad (3.6)$$

Therefore, we can interpret the 2B energy as being the correction to the approximation that the monomers in a dimer do not interact with each other, since the energy of a noninteracting dimer would simply be the sum of the monomer energies (calculated with each monomer in isolation). We can imagine dimer fragments composed of each pair of monomers from the original system, each with its own dimer 2B energy. The sum of the 2B energies of each dimer fragment in the

system is defined to be the total 2B energy of the entire system:

$$E_{2B}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = \sum_{m=1}^{N-1} \sum_{n>m}^N E_{\text{tot}}(\mathbf{R}_m, \mathbf{R}_n) - E_{\text{tot}}(\mathbf{R}_m) - E_{\text{tot}}(\mathbf{R}_n) \quad (3.7)$$

In the case of a dimer system, the 2B energy is equal to the interaction energy, but this is not generally true for larger systems. If we imagine a trimer (3-mer) system, the difference in energy between the total interaction energy and the sum of the 2B energies (i.e. the error due to a pairwise approximation of the interaction energy) is referred to as the 3B energy, E_{3B} :

$$E_{3B}(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3) = E_{\text{int}}(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3) - E_{2B}(\mathbf{R}_1, \mathbf{R}_2) - E_{2B}(\mathbf{R}_1, \mathbf{R}_3) - E_{2B}(\mathbf{R}_2, \mathbf{R}_3) \quad (3.8)$$

By definition, the total interaction energy of a 3-mer is equal to the sum of the 2B energies plus the 3B energy. Inserting the definition of the interaction energy and the definition of the 2B energy into Eq. 3.8 the 3B energy of a 3-mer can be written as:

$$E_{3B}(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3) = E_{\text{tot}}(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3) - E_{\text{tot}}(\mathbf{R}_1) - E_{\text{tot}}(\mathbf{R}_2) - E_{\text{tot}}(\mathbf{R}_3) - \left(\sum_{m=1}^2 \sum_{n>m}^3 E_{\text{tot}}(\mathbf{R}_m, \mathbf{R}_n) - E_{\text{tot}}(\mathbf{R}_m) - E_{\text{tot}}(\mathbf{R}_n) \right) \quad (3.9)$$

After expanding the sums and rearranging terms, we obtain:

$$E_{3B}(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3) = E_{\text{tot}}(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3) - E_{\text{tot}}(\mathbf{R}_1, \mathbf{R}_2) - E_{\text{tot}}(\mathbf{R}_1, \mathbf{R}_3) - E_{\text{tot}}(\mathbf{R}_2, \mathbf{R}_3) + E_{\text{tot}}(\mathbf{R}_1) + E_{\text{tot}}(\mathbf{R}_2) + E_{\text{tot}}(\mathbf{R}_3) \quad (3.10)$$

Why did we end up with an expression that adds the sum of the (isolated) monomer energies in the last line of Eq. 3.10? Notice that each 2-mer fragment total energy in the second line of Eq. 3.10 contains contributions from the monomer energy of each monomer in that 2-mer

fragment (Eq. 3.6). Furthermore, each monomer appears in two different total 2-mer energies in the second line of Eq. 3.10. Therefore, simply subtracting each of the total 2-mer energies from the total energy of the 3-mer ends up over-correcting (i.e. under-counting) the contributions due to the monomer energies, so the last line of Eq. 3.10 effectively corrects for this under-counting.

When defining expressions for general K -body energies, this idea of summing corrections for over-counting and under-counting fragment energies becomes very useful. Defining the sum of the total energies of all w -mer fragments of a parent N -mer cluster to be:[44, 45, 46]

$$S_{\text{tot}}(w; N) = \sum_{n_1 < n_2 < \dots < n_w}^N E_{\text{tot}}(\mathbf{R}_{n_1}, \mathbf{R}_{n_2}, \dots, \mathbf{R}_{n_w}) \quad (3.11)$$

The total K -body energy of the N -mer can be written as:[44, 45, 46]

$$E_{\text{KB}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = \sum_{w=1}^K (-1)^{K-w} \binom{N-w}{K-w} S_{\text{tot}}(w; N) \quad (3.12)$$

As stated above, the 3B energy can be seen as a correction to the approximation that monomers in a 3-mer interact in a pairwise (monomer-monomer) way, but what could cause such an approximation to fail? In general terms, 3B interactions arise from the ways in which interactions with a third monomer augment the interactions between other pairs of monomers. This is predominantly a result of polarization interactions, where, for example, the electric field due to the third monomer distorts the electron densities of the other two monomers, which changes the ways that the two monomers interact with each other, relative to their interactions in the absence of the electric field due to the third monomer. The Pauli exclusion principle (due to the antisymmetry of the electronic wave function with respect to exchange) can also contribute to the 3B energy, but we can safely disregard such effects for now.

K -body (KB) energies of an N -mer system can be defined recursively in the same way that we defined 2B and 3B energies (with the recursive relations collapsing into Eq. 3.12), such that the total interaction energy of a N -mer system is the sum of all of the corresponding 2B, 3B,

..., NB energies:

$$E_{\text{int}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = E_{2B} + E_{3B} + \dots + E_{(N-1)B} + E_{NB} \quad (3.13)$$

E_{2B} , E_{3B} , etc. represent the total 2B, 3B, etc. energies of the full N -mer, respectively. Eq. 3.13 is the definition of the *many-body expansion* (MBE) of the interaction energy. One can obtain the MBE of the binding energy by adding the sum of the 1B energies to Eq. 3.13:

$$E_{\text{bind}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = E_{1B} + E_{2B} + E_{3B} + \dots + E_{(N-1)B} + E_{NB} \quad (3.14)$$

We have not made any approximations, so Eq. 3.13 is an exact expression for the interaction energy of the N -mer system, provided that we include all of the K -body energies up to and including the N -body energy, E_{NB} . Since this expression is exact, the MBE on its own doesn't simplify the computation of the interaction energy of a system at all. In fact, computing all terms in the MBE for a given N -mer is actually more costly than calculating the interaction energy directly from Eq. 3.4, since Eq. 3.4 only requires the total energy of the N -mer, and the energies of each isolated monomer, while computing all terms in the MBE requires that we know the energies of every possible K -mer fragment of the system ($1 \leq K \leq N$).

So why do we care about the MBE? It turns out that there are many systems in which the MBE converges rapidly, meaning that the magnitudes of the K -body energies decrease with increasing K , so that the terms in the MBE with large many-body orders are much less important than the lower order terms. This is often the case in systems in which the monomers are insulators (for example neutral water systems).[47, 48, 49, 45, 46, 50] When the monomer units are insulators, their electron densities remain relatively localized, which limits the strength of the polarization effects between large numbers of monomers. On the other hand, when the system is composed of conducting monomer units, the electrons are shared between many monomers, so the polarization effects between large numbers of monomers are important,[44, 45] so for

example, the 6B energy might be on the same order as the 3B energy.

When the MBE converges rapidly in a given system, there are a number of ways in which we may choose to simplify the computation of the total system interaction energy. The simplest approach would be to truncate the MBE at a particular K -body term, approximating that all higher order terms are 0. Truncating the MBE can work well in some circumstances (depending on the system, and on which questions one wishes to answer), but can lead to serious deficiencies if the specific truncation scheme is not chosen carefully, especially given the combinatorics of many-body energies discussed in Sec. 3.5. A safer approach to exploiting the convergence of the MBE in insulating systems is to calculate the important low order terms with an accurate (but expensive) method, while calculating the higher order terms with less costly methods as is discussed in Sec. 3.1.4. Another approach one may take (possibly in conjunction with the previous approach) is to approximate that the higher order terms in the MBE are recovered by classical inductive electrostatics (i.e. classical polarization).[51, 52] This third option (when appropriate for the particular application) can be thought of as combining the best parts of the first two approaches, given that it is relatively cheap (calculating classical polarization energies is more expensive than a truncation of the MBE, but is much cheaper than running electronic structure calculations on large K -body fragments), while still including contributions from all terms in the MBE. We discuss the use of classical polarization for representing high order many-body energies in Chapter 5 when discussing many-body potential energy functions.

3.1.4 The SAMBA Method and The Basis Set Superposition Error

One nice feature of the MBE is that many-body energies for different K -body orders can be calculated independently of each other. This is the main idea in the “stratified approximation” many-body approach (SAMBA) introduced in Ref. 46. In insulating systems for which the MBE converges rapidly, such that low-order terms (2B, 3B, etc.) are much more important than high-order terms (6B, 7B, etc), one may use the most accurate methods available to calculate the low-order terms, and use lower levels of theory for the high-order terms. This is important

because the high-order terms involving a large number of monomer units (and correspondingly, a large number of electrons and basis functions) require more computational resources to compute for a given level of theory. In our studies on protonated and deprotonated water clusters, we make use of SAMBA to scale down the cost of the 3B reference energy calculations relative to the 2B calculations by using a smaller basis set, and to scale down the 4B, 5B and 6B reference energy calculations by using the explicitly correlated CCSD(T)-F12b method rather than the more expensive CCSD(T) calculations used in the 2B and 3B calculations, as is discussed in Sec. A.1.

Additionally, SAMBA can be applied when calculating counterpoise (CP) corrections for the basis set superposition error (BSSE) at different many-body orders. When computing the 2B energy in a dimer, the BSSE results from a mismatch between the basis set used in the dimer calculation relative to the basis sets used in the monomer calculations, since, in the dimer calculation, electrons ‘belonging’ to each monomer can relax into the basis functions centered on the other monomers, while these additional basis functions are not available in the isolated monomer calculations.[53] For 2B energies of dimers, the Boys-Bernardi CP correction scheme involves calculating each of the isolated monomer energies in the basis of the dimer.[53] However, when calculating interaction energies in larger clusters, there are several ways that one may extend the original Boys-Bernardi CP correction scheme, and within the SAMBA framework, multiple BSSE correction schemes can be used for different terms in the MBE for the same system.

The two BSSE correction schemes used in our studies on protonated and deprotonated water clusters are the “site-site function counterpoise” method (SSCP) proposed,[54] and the “K-centered CP” method originally proposed in Ref. 55 and generalized in Ref. 56. The SSCP correction requires the calculation of each (sub-cluster) fragment energy in the basis of the entire cluster.[54] In the K-centered CP correction, each K-mer fragment’s contribution to the total K-body term in the MBE is calculated in the basis of the K-mer fragment.[56] The SSCP-corrected calculations become prohibitively expensive for large clusters and large basis sets, while the

number of K-centered CP corrected calculations grows rapidly with K.

3.2 Many-Body Interactions in Deprotonated Water Clusters

The hydroxide ion, OH^- , plays an important role in aqueous acid/base chemistry, being the fundamental basic unit in aqueous solutions. Given its importance in aqueous systems, it is desirable to develop models describing its interactions with water for use in molecular dynamics simulations. To this end, we investigated the many-body interactions present in low-energy $\text{OH}^-(\text{H}_2\text{O})_n$ ($n \leq 5$) clusters. The clusters investigated in this study were selected to represent various coordination numbers and hydrogen bonding connectivities. Most $\text{OH}^-(\text{H}_2\text{O})_n$ clusters were reoptimized (see Sec. A.1) from coordinates reported in the literature,[57, 58] or adapted from previous studies.[58]

The $\text{OH}^-(\text{H}_2\text{O})_n$ clusters with $n = 3, 4, 5$ are shown in Figs. 3.2–3.4, and the clusters with $n = 1, 2$ are shown in Fig. 3.8a and Fig. 3.8b. The total interaction energies of all clusters are reported in Table 3.1. A discussion of the structures and relative energies of the $\text{OH}^-(\text{H}_2\text{O})_n$ clusters studied here can be found in Ref. 4, and xyz coordinates for all clusters can be found in the corresponding Supporting Information.

Additionally, the convergence of the MBE for the $\text{OH}^-(\text{H}_2\text{O})_n$ clusters with $n = 3, 4, 5$ are shown in Figs. 3.5–3.7. This analysis indicates that 2B and 3B contributions dominate the total interaction energies, which suggests that 2- and 3-mer clusters (and fragments) can effectively be considered as the primary ‘building blocks’ of deprotonated water systems, and will be our primary focus. In Sec. 3.2.2, we introduce the concept of a “fragment group,” and establish notation that will facilitate the discussion of many-body effects in the various fragment groups in Sec. 3.2.3. In Sec. 3.2.1, we discuss the global minimum energy $\text{OH}^-(\text{H}_2\text{O})$ and $\text{OH}^-(\text{H}_2\text{O})_2$ clusters. In Sec. 3.2.4, we discuss the accuracy of various XC-functionals (Sec. A.1) in describing the many-body energies of the $\text{OH}^-(\text{H}_2\text{O})_n$ fragment groups.

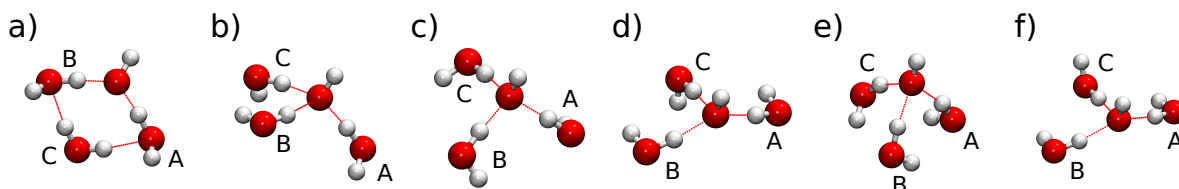


Figure 3.2. Low-lying isomers of the $\text{OH}^-(\text{H}_2\text{O})_3$ cluster shown in order of increasing interaction energy from isomer *a* to isomer *f*.

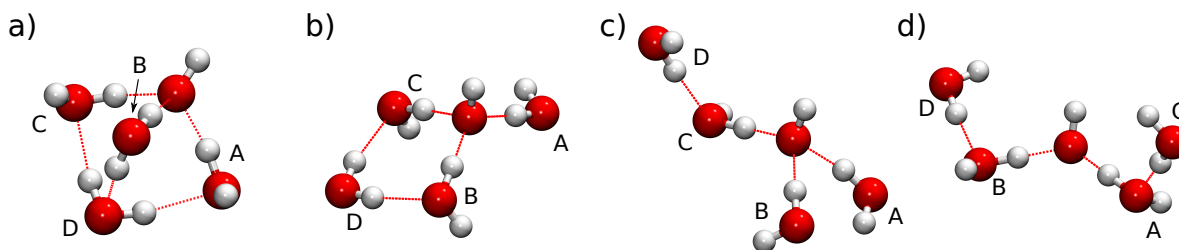


Figure 3.3. Low-lying isomers of the $\text{OH}^-(\text{H}_2\text{O})_4$ cluster shown in order of increasing interaction energy from isomer *a* to isomer *d*.

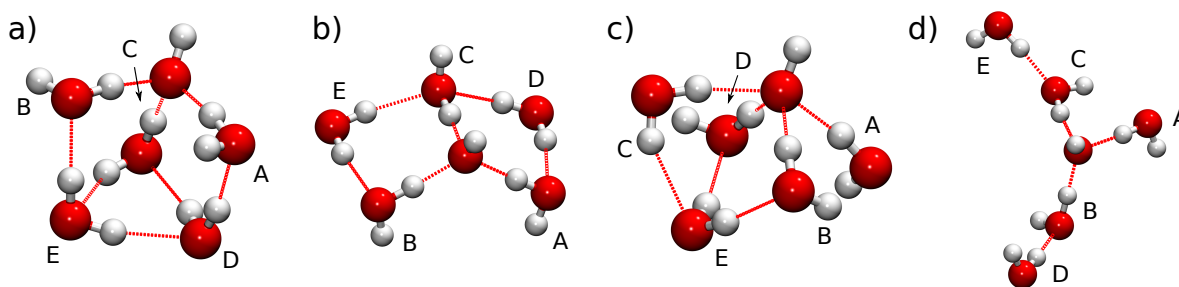


Figure 3.4. Low-lying isomers of the $\text{OH}^-(\text{H}_2\text{O})_5$ clusters shown in order of increasing interaction energy from isomer *a* to isomer *d*.

3.2.1 The Minimum Energy $\text{OH}^-(\text{H}_2\text{O})$ and $\text{OH}^-(\text{H}_2\text{O})_2$ Clusters

The most prevalent and, arguably, most important building block that constitutes all stable $\text{OH}^-(\text{H}_2\text{O})_n$ clusters is the $\text{OH}^-(\text{H}_2\text{O})$ 2-mer in which the water molecule donates a hydrogen bond to OH^- , as is the case in the $\text{OH}^-(\text{H}_2\text{O})$ global minimum-energy (GME) structure (Fig. 3.8). In general, hydrogen bonds, especially when involving charged species, are highly electrostatic in nature and can effectively be seen as “incipient chemical bonds.”[59, 7] This is very much the case in $\text{OH}^-/\text{H}_2\text{O}$ systems, where efficient proton transfer (i.e. incipient chemical bond $\rightarrow$

Table 3.1. Reference interaction energies (in kcal/mol) for the low-lying isomers of $\text{OH}^-(\text{H}_2\text{O})_n$ clusters shown in Figs. 3.8–3.4.

Cluster size	(a)	(b)	(c)	(d)	(e)	(f)
2-mer	−37.09					
3-mer	−54.93					
4-mer	−74.54	−72.83	−72.72	−72.43	−72.08	−71.76
5-mer	−90.72	−89.55	−87.66	−86.90		
6-mer	−106.15	−104.73	−104.45	−99.52		

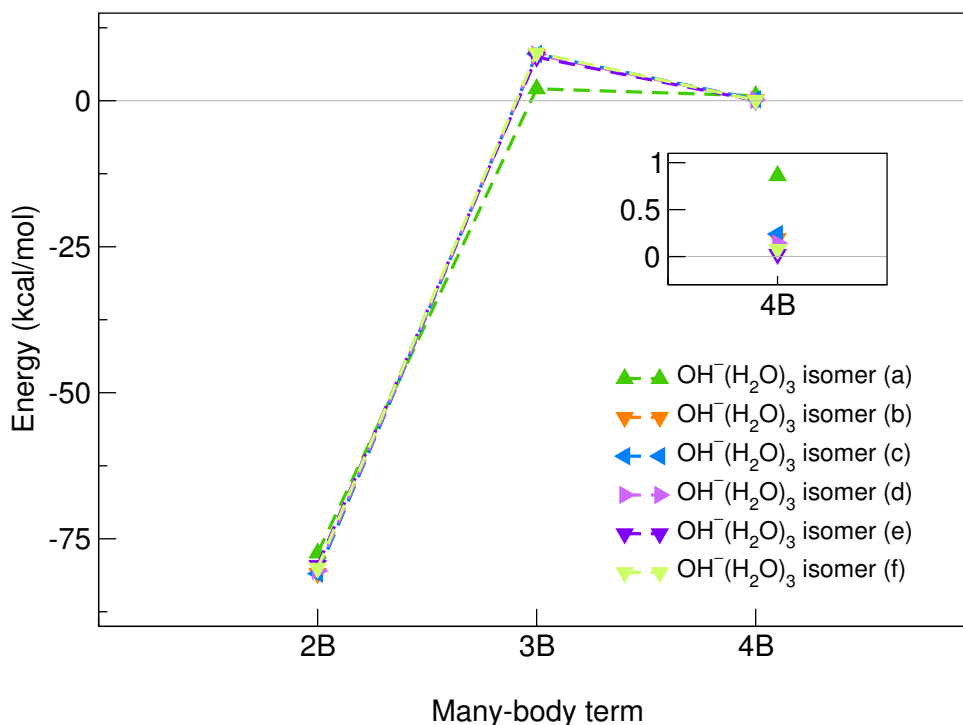


Figure 3.5. Convergence of the many-body expansion of the low-lying isomers of $\text{OH}^-(\text{H}_2\text{O})_3$ shown in Fig. 3.2.

chemical bond) leads to seemingly “anomalously” high ion mobility.[60, 61] In addition to a strongly attractive interaction, a lengthening of the donor-hydrogen (covalent) bond length is characteristic of the hydrogen bond. This structural effect is a result of the electrostatic attraction between the donor-hydrogen and the acceptor atom, as well as charge transfer.[7] These effects

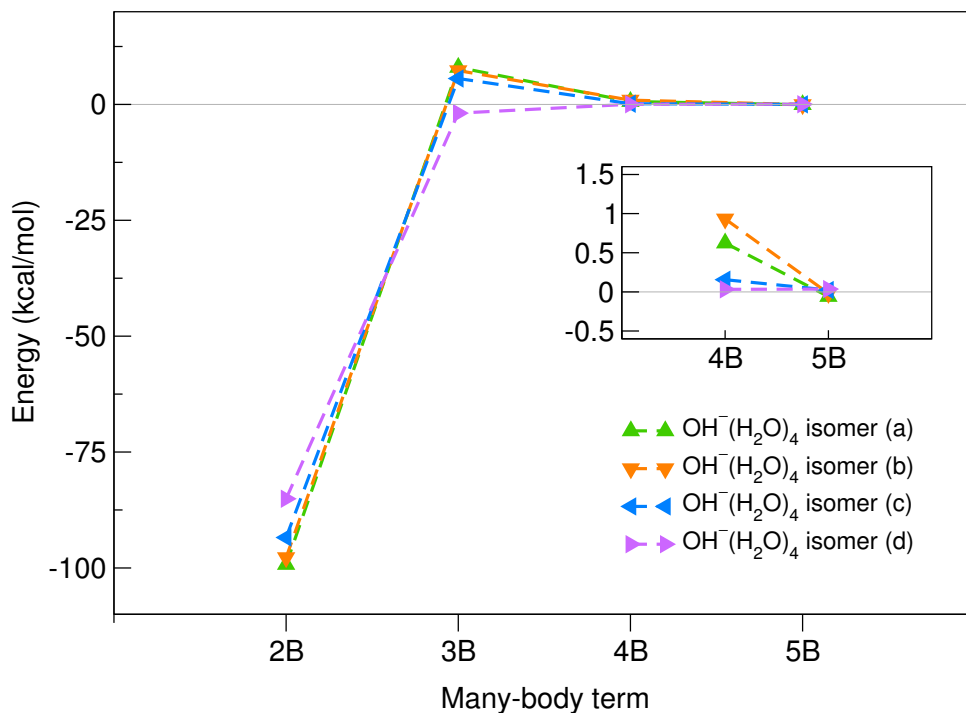


Figure 3.6. Convergence of the many-body expansion of the low-lying isomers of $\text{OH}^-(\text{H}_2\text{O})_4$ shown in Fig. 3.3.

are fully realized in the $\text{OH}^-(\text{H}_2\text{O})$ GME structure for which the interaction energy is -37.092 kcal/mol and the length of the water donor-OH bond is $1.11 \text{ \AA}$. For reference, the equilibrium OH bond length in an isolated water is $0.96 \text{ \AA}$, optimized at the same level of theory (RI-MP2/aug-cc-pVTZ). The relatively large strength of the interaction is emphasized when compared to that of the $\text{F}^-(\text{H}_2\text{O})$ dimer GME structure, which has an interaction energy of -32.704 kcal/mol and a water donor-OH bond length of $1.06 \text{ \AA}$, also obtained at the RI-MP2/aug-cc-pVTZ level of theory.

It is worth noting that the GME structure of $\text{OH}^-(\text{H}_2\text{O})$ reported here, wherein the donor-proton is localized on one oxygen—allowing one to unequivocally identify individual OH^- and H_2O molecular identities within the 2-mer—corresponds to the Born-Oppenheimer equilibrium geometry. However, previous experimental and theoretical studies[62, 63, 64] have shown that the actual quantum-mechanical ground state of $\text{OH}^-(\text{H}_2\text{O})$ is delocalized along the proton transfer coordinate, lying above the associated potential energy barrier. This implies that,

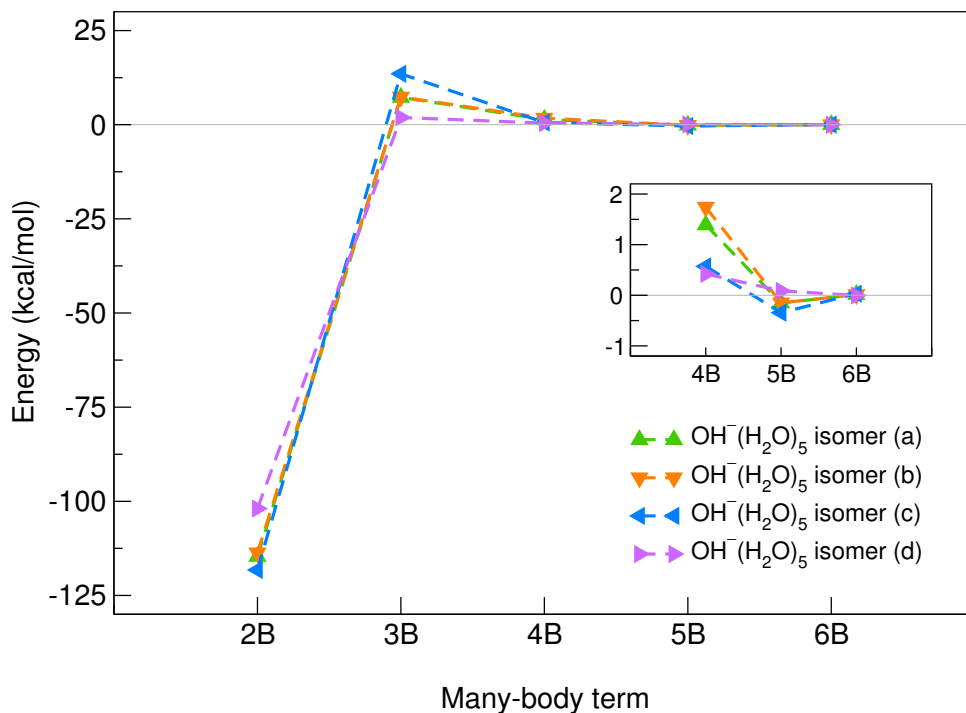


Figure 3.7. Convergence of the many-body expansion of the low-lying isomers of $\text{OH}^-(\text{H}_2\text{O})_5$ shown in Fig. 3.4.

in a fully quantum-mechanical picture of the GME structure, the donor-proton is actually shared equally between the two oxygen atoms.

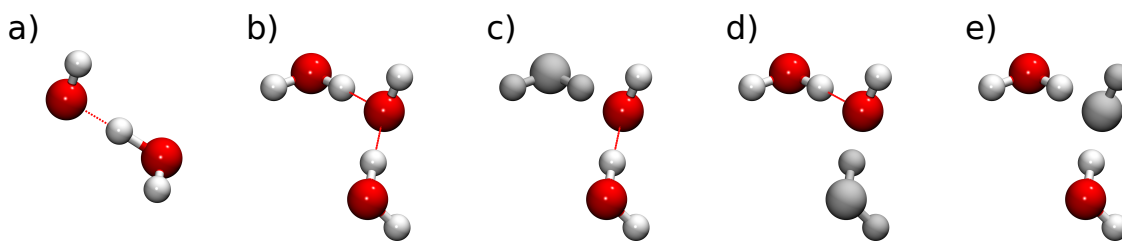


Figure 3.8. $\text{OH}^-(\text{H}_2\text{O})_n$ building blocks. a) $\text{OH}^-(\text{H}_2\text{O})$ global minimum-energy structure, b) $\text{OH}^-(\text{H}_2\text{O})_2$ global minimum-energy structure along with its constituent 2-mer fragments (c-e).

The presence of a second water molecule in $\text{OH}^-(\text{H}_2\text{O})_2$ introduces additional 2B contributions from a second OH^- –water interaction and a single water–water interaction, as well as a 3B contribution associated with the interaction of OH^- with the two water molecules. The total 2B interaction energy for the $\text{OH}^-(\text{H}_2\text{O})_2$ GME structure is -58.333 kcal/mol, with each OH^- –water interaction contributing, on average, -29.576 kcal/mol, which is ~ 7 kcal/mol

higher than in the $\text{OH}^-(\text{H}_2\text{O})$ GME structure. The weakening of the average OH^- –water interaction compared the 2-mer is reflected in a lower degree of water distortions, with both donor-OH bond lengths being equal to ~ 1.03 Å, as well as in the lengthening of the OH^- –water O–O distance from 2.48 Å in the 2-mer to 2.57 Å in the 3-mer. The water–water 2B contribution is relatively small and repulsive, with a magnitude of 0.820 kcal/mol and a water–water O–O distance of 4.59 Å. 3B effects contribute 3.402 kcal/mol to the total interaction energy of $\text{OH}^-(\text{H}_2\text{O})_2$ in the GME structure. An early study based on symmetry-adapted perturbation theory (SAPT) determined that the dominant component of the 3B term is the induction energy (including charge transfer effects) arising from the deformation of the isolated-monomer electron orbitals.[65]

3.2.2 Fragment Groups and Fragment Notation

Important information about the molecular interactions within a given (deprotonated) water cluster is encoded in the hydrogen bonding (H-bonding) connectivity of that cluster.[4, 66, 67, 68, 47, 49, 69, 70, 71, 72, 73, 74, 75] Since H-bonds are directional, their connectivity must be described by directed-graphs (or “digraphs”).[72, 73, 74, 76]

In order to describe various H-bonding connectivities in a compact form, we denote a particular water molecule as $_{A_1A_2\leftarrow}^S\text{W}_{\leftarrow D_1D_2}^X$, where left-superscript, S , a positive integer, denotes the solvation shell in which the water molecule resides (i.e., the smallest number of H-bonds separating it from OH^-), and right-superscript, X , is a capital Roman letter (A, B, C , etc.) which identifies the particular water molecule, for example those labeled in Figs. 3.2–3.4. The left- and right-subscripts, A_1, A_2 , and D_1, D_2 , enumerate the identities (the capital Roman letters) of all water molecules accepting and donating H-bonds, respectively, from/to water molecule X . Left-pointing arrows in the subscripts are included to remind the reader that left-subscripted water molecules accept H-bonds from X , and that right-subscripted water molecules donate H-bonds to X .

For example, if 1st solvation shell water molecules A and B , $^1\text{W}^A$ and $^1\text{W}^B$, each donate

an H-bond to 2nd solvation shell water molecule C , ${}^2W^C$, then the corresponding complete notation would be ${}_{C\leftarrow}{}^1W^A$, ${}_{C\leftarrow}{}^1W^B$, and ${}^2W_{\leftarrow AB}^C$ for the three molecules, respectively. If, in the same cluster, another 1st solvation shell water molecule D , ${}^1W^D$, is not H-bonded to any other water molecules, then it would simply be denoted by ${}^1W^D$, and the notation for the other three molecules would be unchanged. Additionally, OH^- will denote the hydroxide ion.

All water–water and OH^- –water H-bonds are classified according to the following (quadratic) inequality,[77]

$$R_{O_A-O_D} \leq -0.00044 (\theta_{O_A-O_D-H_D})^2 + 3.3 \quad (3.15)$$

where O_A and O_D are the acceptor and donor oxygen atoms, respectively, H_D is the hydrogen atom H-bonded to O_A , $R_{O_A-O_D}$ is the distance (in Å) between O_A and O_D , and $\theta_{O_A-O_D-H_D}$ is the angle (in degrees) between O_A , O_D and H_D . In Eq. 3.15, a pair of molecules are classified as H-bonded if their O_A – O_D distance, $R_{O_A-O_D}$, falls within the inequality.

Since there are over 400 fragments within the 16 $\text{OH}^-(\text{H}_2\text{O})_n$ clusters discussed here, which precludes the possibility of analyzing each fragment separately, all fragments are partitioned into a relatively small number of “fragment groups.” Three properties are used to classify the different fragment groups: 1) presence of OH^- in the fragment, 2) number of molecules in the fragment, and 3) (directional) H-bonding connectivity, not distinguishing between permutations of water molecules. Since the fragment groups are defined only by their H-bonding connectivity (independent of the details about the parent clusters), we only need to specify the solvation shell and connectivity when relevant, and only assign right-superscript letters to distinguish water molecules, rather than to identify particular molecules in a cluster. A subscript “FG” is appended outside the curly braces to specify that the notation refers to a group of fragments, rather than a particular fragment. For example, keeping in mind that OH^- does not donate a hydrogen bond, there are four possible 2-mer fragment groups within $\text{OH}^-(\text{H}_2\text{O})_n$ clusters:

1. water donating a hydrogen bond to OH^- : $\{\text{OH}^-, {}^1W\}_{FG}$

2. water donating a hydrogen bond to water: $\{W_{\leftarrow B}^A, A \leftarrow W^B\}_{FG}$
3. $\text{OH}^-(\text{H}_2\text{O})$ with no hydrogen bond: $\{\text{OH}^-, {}^2W\}_{FG}$
4. $(\text{H}_2\text{O})_2$ with no hydrogen bond: $\{W^A, W^B\}_{FG}$

It is worth noting that the name “fragment group” was chosen independently of any consideration about connections to group theory, and so any connection is coincidental. However, in Sec. 3.5 the concept of a fragment group is extended in a way that hints at some kind of underlying group structure, although we do not explore this idea in detail.

3.2.3 Fragment Analysis and Many-Body Effects in Deprotonated Water Clusters

Table 3.2 reports the statistics for the four 2-mer fragment groups enumerated above. The “freq.” column reports the number of fragments within each group from all clusters. The “max.,” “min.,” and “mean” columns report the signed-maximum, signed-minimum (i.e., the most repulsive and attractive fragment energy contributions, respectively), and signed-mean of the 2B interaction energy contributions within each fragment group. The “RSD” (relative standard deviation) column reports the standard deviation of the 2B interaction energy contributions divided by the absolute value of their signed-mean. Values in the “sig.” column quantify the ‘significance’ of each fragment group, and are calculated by multiplying the unsigned-mean energy (not shown) by the occurrence frequency of that fragment group, and dividing the resulting product by the sum of all frequencies of the fragment groups with the same fragment size. The fragment groups are listed in descending significance order.

Tables 3.2, 3.3, 3.4, and 3.5 show that most fragment groups contain exclusively attractive or repulsive fragments, indicating that the classification adopted in this analysis reveals information about specific many-body interactions within $\text{OH}^-(\text{H}_2\text{O})_n$ clusters. The groups that contain both attractive and repulsive fragments turn out to be structurally ambiguous, since the non-hydrogen bonded region of a given configurational space is much larger than the complementary hydrogen bonded region. The most prevalent example of these ‘structurally ambiguous’

fragment groups (of all sizes) is the non-hydrogen bonded $(\text{H}_2\text{O})_2$ group, $\{B\leftarrow W^A, W_{\leftarrow A}^B\}_{FG}$, whose most attractive and most repulsive fragments contribute -0.525 and 1.720 kcal/mol, respectively, with a RSD of 1.119. However, the majority of the $\{B\leftarrow W^A, W_{\leftarrow A}^B\}_{FG}$ fragments (50 out of 59 fragments) tend to be repulsive, which results in a mean contribution of 0.462 kcal/mol. Although the $\{\text{OH}^-, {}^2\text{W}\}_{FG}$ fragment group also appears to be structurally ambiguous, due to the lack of direct hydrogen bonding, each fragment is associated with a negative (attractive) 2B energy contribution, despite a RSD of 0.200, and a difference of almost 7 kcal/mol between minimum and maximum values. This results from the fact that each fragment considered in this analysis was generated as a ‘byproduct’ of a geometry optimization performed on one of the low-lying $\text{OH}^-(\text{H}_2\text{O})_n$ clusters. This implies that 2B interaction energies associated with particular fragment groups can have a greater influence than others on determining the minimum-energy structures of the parent clusters on the Born-Oppenheimer potential energy surface, thus biasing the type of fragments that can be generated. The remaining two 2-mer fragment groups are hydrogen-bonded $\text{OH}^-(\text{H}_2\text{O})$ and $(\text{H}_2\text{O})_2$ which contain exclusively attractive fragments since hydrogen bonds are attractive, by definition.[59] These two fragment groups have been extensively investigated in previous studies and we refer the reader to the original references for specific details.[78, 62, 63, 79, 80, 81, 82, 83, 84, 85, 64, 85, 86, 87, 88, 68, 89, 90, 77, 7, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100] It should only be noted that the “min.” value for $\{\text{OH}^-, {}^1\text{W}\}_{FG}$ in Table 3.2, corresponding to the GME $\text{OH}^-(\text{H}_2\text{O})$ structure, is significantly lower than

Table 3.2. Statistics for the 2B energies (in kcal/mol) of 2-mer fragments. See main text for details.

Frag. Group	freq.	max.	min.	mean	RSD	sig.
$\{\text{OH}^-, {}^1\text{W}\}_{FG}$	44	-23.9	-37.1	-27.8	0.1	8.7
$\{\text{OH}^-, {}^2\text{W}\}_{FG}$	13	-5.2	-12.0	-9.6	0.2	0.9
$\{B\leftarrow W^A, W_{\leftarrow A}^B\}_{FG}$	24	-2.1	-4.2	-3.3	0.1	0.6
$\{W^A, W^B\}_{FG}$	59	+1.7	-0.5	+0.5	1.1	0.2

Table 3.3. Statistics for the 3B energies (in kcal/mol) of 3-mer fragments. See main text for details.

Frag. Group	freq.	max.	min.	mean	RSD	sig.
$\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$	42	+4.3	+1.9	+2.9	0.2	0.8
$\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A\leftarrow {}^2\text{W}^B\}_{FG}$	20	-0.8	-5.3	-2.2	0.6	0.3
$\{\text{OH}^-, {}^1\text{W}, {}^2\text{W}\}_{FG}$	14	+1.2	+0.3	+0.8	0.4	0.1
$\{\text{W}_{\leftarrow C}^A, \text{W}_{\leftarrow C}^B, AB\leftarrow \text{W}^C\}_{FG}$	8	+1.0	+0.7	+0.9	0.1	0.04
$\{\text{OH}^-, B\leftarrow {}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B\}_{FG}$	3	+2.2	+2.0	+2.1	0.04	0.04
$\{\text{W}_{\leftarrow B}^A, A\leftarrow \text{W}_{\leftarrow C}^B, B\leftarrow \text{W}^C\}_{FG}$	6	-0.3	-0.9	-0.6	0.3	0.02
$\{\text{W}^A, \text{W}^B, \text{W}^C\}_{FG}$	19	-0.005	-0.5	-0.2	0.6	0.02
$\{\text{W}_{\leftarrow B}^A, A\leftarrow \text{W}^B, \text{W}^C\}_{FG}$	27	+0.3	-0.3	+0.1	2.1	0.02
$\{C\leftarrow \text{W}_{\leftarrow B}^A, A\leftarrow \text{W}_{\leftarrow C}^B, B\leftarrow \text{W}_{\leftarrow A}^C\}_{FG}$	1	-1.5	-1.5	-1.5	N/A	0.01
$\{\text{OH}^-, {}^2\text{W}^A, {}^2\text{W}^B\}_{FG}$	3	+0.3	+0.1	+0.3	0.4	0.005
$\{\text{W}_{BC}^A, A\leftarrow \text{W}^B, A\leftarrow \text{W}^C\}_{FG}$	1	+0.6	+0.6	+0.6	N/A	0.004
$\{\text{OH}^-, {}^2\text{W}_{\leftarrow B}^A, A\leftarrow {}^2\text{W}^B\}_{FG}$	1	+0.2	+0.2	+0.2	N/A	0.001

the 2B energy (-31.421 kcal/mol) of the next most attractive fragment. This difference is due to repulsive 3B interactions present in clusters with two or more ${}^1\text{W}$ molecules, which limit how close the ${}^1\text{W}$ molecules can get to OH^- in a stable geometry.

Among the 14 possible 3-mer fragment groups (not counting bifurcated hydrogen bonds, or pairs of water molecules donating hydrogen bonds to each other) that can be generated from the low-lying $\text{OH}^-(\text{H}_2\text{O})_{n=3-5}$ isomers, 12 are represented in the fragment set that is analyzed in Table 3.3. The key concept that shapes the discussion of the 3-mer fragments is that of cooperative vs. anticooperative hydrogen bonding, an idea that has received substantial attention in (neutral) water systems.[67, 68, 47, 49, 69, 70, 71] In general terms, cooperative effects occur when the formation of one hydrogen bond between two molecules promotes further hydrogen bonding with the surrounding molecules due to the redistribution of the electron density, which results in negative 3B (and higher-body) contributions to the total interaction energy. On

the contrary, anticooperativity occurs when the formation of one hydrogen bond between two molecules weakens the capability of these two molecules to engage in further hydrogen bonding, manifesting in positive 3B (and higher-body) energies.[67, 68, 69, 71] In the case of neutral water trimers, cooperativity occurs when at least one water molecule acts as a hydrogen bond donor and acceptor simultaneously, while anticooperativity occurs when a water molecule acts as a double hydrogen bond donor, or as a double acceptor.[67, 68, 47, 49, 69, 70, 71]

The most ‘significant’ 3-mer fragment group among the low-lying isomers of $\text{OH}^-(\text{H}_2\text{O})_{n=3-5}$ is $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$. Fragments in this group have OH^- accepting two hydrogen bonds, which is a common motif in large $\text{OH}^-(\text{H}_2\text{O})_n$ clusters where OH^- tend to be at least 3-coordinated. All fragments in this group are associated with positive 3B contributions to the interaction energy, with a mean value of 2.862 kcal/mol. Furthermore, Fig. 4.5 shows that a nearly linear relationship exists between the 3B energies and the corresponding average OH^- –water O–O distances. Given the inherently repulsive interactions and specific hydrogen bond connectivity, we conclude that OH^- –water hydrogen bonds in this fragment group follow the same anticooperativity trends as water–water hydrogen bonds in the corresponding neutral water analog, $\{\text{W}_{BC}^A, A\leftarrow \text{W}^B, A\leftarrow \text{W}^C\}_{FG}$. [67, 47, 49, 69, 70]

As shown in Table 3.3, the second-most ‘significant’ 3-mer fragment group within $\text{OH}^-(\text{H}_2\text{O})_{n=3-5}$ clusters is $\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A\leftarrow {}^2\text{W}^B\}_{FG}$. 3B energy contributions for these fragments are found to correlate with the $\text{OH}^- - {}^1\text{W} - {}^2\text{W}$ O–O–O angle, with the 3B contributions becoming more attractive as the angle becomes more linear (Fig. 4.7). Furthermore, the $\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A\leftarrow {}^2\text{W}^B\}_{FG}$ fragments analyzed here have exclusively attractive 3B energy contributions and, therefore, are characterized by cooperative hydrogen bonds. This parallels the connectivity trends of cooperative hydrogen bonding in neutral water clusters, since the analogous $\{\text{W}_{\leftarrow B}^A, A\leftarrow \text{W}_{\leftarrow C}^B, B\leftarrow \text{W}^C\}_{FG}$ 3-mer fragments exhibit similar cooperativity.[67, 47, 49, 69, 70] As expected, the interactions in the $\{\text{W}_{\leftarrow B}^A, A\leftarrow \text{W}_{\leftarrow C}^B, B\leftarrow \text{W}^C\}_{FG}$ fragments within our data set are characterized by cooperative hydrogen bonding, with all 3B interaction energies in this set being attractive, with a mean contribution of -0.589 kcal/mol (Table 3.3).

Energy decomposition analysis provides further insight into the nature of these cooperative and anticooperative 3B interactions in $\text{OH}^-(\text{H}_2\text{O})_n$ clusters, as well as an explanation of the correlations of the 3B energies with O–O distances in anticooperative 3-mers, and with the O–O–O angles in cooperative 3-mers in terms of 3B polarization and charge transfer interactions. These energy decompositions are discussed in Sec. 4.4.

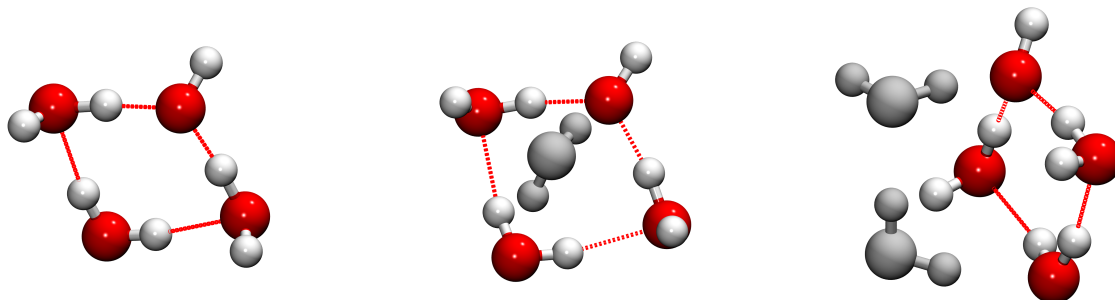


Figure 3.9. The $\{\text{OH}^-, {}^1\text{W}_{\leftarrow C}^A, {}^1\text{W}_{\leftarrow C}^B, AB\leftarrow {}^2\text{W}^C\}_{FG}$ fragments within the 4-, 5-, and 6-mer GME structures.

As shown in Table 3.4, all 4B contributions to the total interaction energies of 4-mer fragments extracted from the low-lying $\text{OH}^-(\text{H}_2\text{O})_{3-5}$ clusters tend to be small (statistics for all 4-mer fragment groups can be found in the Supporting Information of Ref. 4). Nevertheless, it is still important to comment on the most ‘significant’ 4-mer fragment group, $\{\text{OH}^-, {}^1\text{W}_{\leftarrow C}^A, {}^1\text{W}_{\leftarrow C}^B, AB\leftarrow {}^2\text{W}^C\}_{FG}$. Fragments within this group are cyclic with the single ${}^2\text{W}$ molecule donating a hydrogen bond to each ${}^1\text{W}$ molecule, and contains the two most ‘significant’ types of 3-mer fragments, $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ and $\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A\leftarrow {}^2\text{W}^B\}_{FG}$, as sub-fragments. The most notable feature of this group is its presence in each of the GME 4-mer, 5-mer, and 6-mer clusters (Fig. 3.9), despite contributing, on average, a relatively high 4B energy (0.692 kcal/mol).

Finally, it is not surprising that all 5-mer fragment groups are essentially not significant (Table 3.5) given the convergence rate of the MBE for $\text{OH}^-(\text{H}_2\text{O})_n$ clusters shown in Figs. 3.5–3.7. However, it is important to point out that the single instance of the $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B, {}^1\text{W}^C, {}^1\text{W}^D\}_{FG}$ 5-mer fragment group, corresponding to a 4-coordinated OH^- , stands out,

being associated with a relatively large 5B contribution (-0.175 kcal/mol).

3.2.4 Assessment of DFT in Reproducing Many-Body Interactions in Deprotonated Water Clusters

The fragment analysis introduced in section 3.2.3 is used to evaluate the accuracy of various DFT models (see Section A.1) in reproducing the interaction energies of $\text{OH}^-(\text{H}_2\text{O})_{n=1-5}$ clusters and constituting fragments. As discussed above, 2B and 3B terms dominate the MBE, which suggests that the errors for the 2-mer and 3-mer fragments are important indicators of the overall accuracy of different XC-functionals in correctly reproducing the energetics of deprotonated water systems.

Fig. 3.10 shows the errors of the tested exchange-correlation functionals in reproducing the 2B SAMBA energy contributions for the hydrogen bonded 2-mers $\{\text{OH}^-, {}^1\text{W}\}_{FG}$ (panel a) and $\{B\leftarrow \text{W}^A, \text{W}_{\leftarrow A}^B\}_{FG}$ (panel b). These comparisons show that the net negative charge on $\{\text{OH}^-, {}^1\text{W}\}_{FG}$ fragments tends to further emphasize the errors present in the neutral water fragments. This appears to be the case for all functionals, except for B3LYP(-D3), which is slightly more accurate for $\{\text{OH}^-, {}^1\text{W}\}_{FG}$ fragments, as well as for the range-separated hybrids ($\omega\text{B97X-D}$ and $\omega\text{B97M-V}$) which essentially reproduce the reference energies for both fragment groups. These errors may be attributed to well-known deficiencies of several functionals in describing electron density tails (i.e., the valence electrons) of anionic species.[101] This problem is thought to be

Table 3.4. Statistics for ‘significant’ 4B energies (in kcal/mol) of 4-mer fragments. See main text for details. See Supporting Information of Ref. 4 for a complete analysis of all 4B energies.

Frag. Group	freq.	max.	min.	mean	RSD	sig. factor
$\{\text{OH}^-, {}^1\text{W}_{\leftarrow C}^A, {}^1\text{W}_{\leftarrow C}^B, AB\leftarrow {}^2\text{W}^C\}_{FG}$	7	+0.9	+0.5	+0.7	0.1	0.06
$\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B, {}^1\text{W}^C\}_{FG}$	15	+0.3	+0.01	+0.1	0.6	0.03
$\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}_{\leftarrow C}^B, B\leftarrow {}^2\text{W}^C\}_{FG}$	19	+0.2	-0.1	+0.02	3.2	0.01
$\{\text{OH}^-, {}^1\text{W}_{\leftarrow C}^A, C\leftarrow {}^1\text{W}^B, A\leftarrow {}^2\text{W}_{\leftarrow B}^C\}_{FG}$	5	+0.2	-0.1	-0.01	17.9	0.01

Table 3.5. Statistics for ‘significant’ 5B energies (in kcal/mol) of 5-mer fragments. See main text for details. See Supporting Information of Ref. 4 for a complete analysis of all 5B energies.

Frag. Group	freq.	max.	min.	mean	RSD	sig.
$\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B, {}^1\text{W}^C, {}^1\text{W}^D\}_{FG}$	1	-0.2	-0.2	-0.2	N/A	0.01
$\{\text{OH}^-, {}^1\text{W}_{\leftarrow D}^A, {}^1\text{W}_{\leftarrow D}^B, {}^1\text{W}_{AB\leftarrow}^C, {}^2\text{W}_{\leftarrow D}^D\}_{FG}$	5	-0.01	-0.03	-0.03	0.3	0.01
$\{\text{OH}^-, {}^1\text{W}_{D\leftarrow}^A, {}^1\text{W}_{\leftarrow D, C\leftarrow}^B, {}^1\text{W}_{\leftarrow D, C\leftarrow}^C, {}^2\text{W}_{\leftarrow B}^D\}_{FG}$	3	-0.03	-0.05	-0.04	0.2	0.01

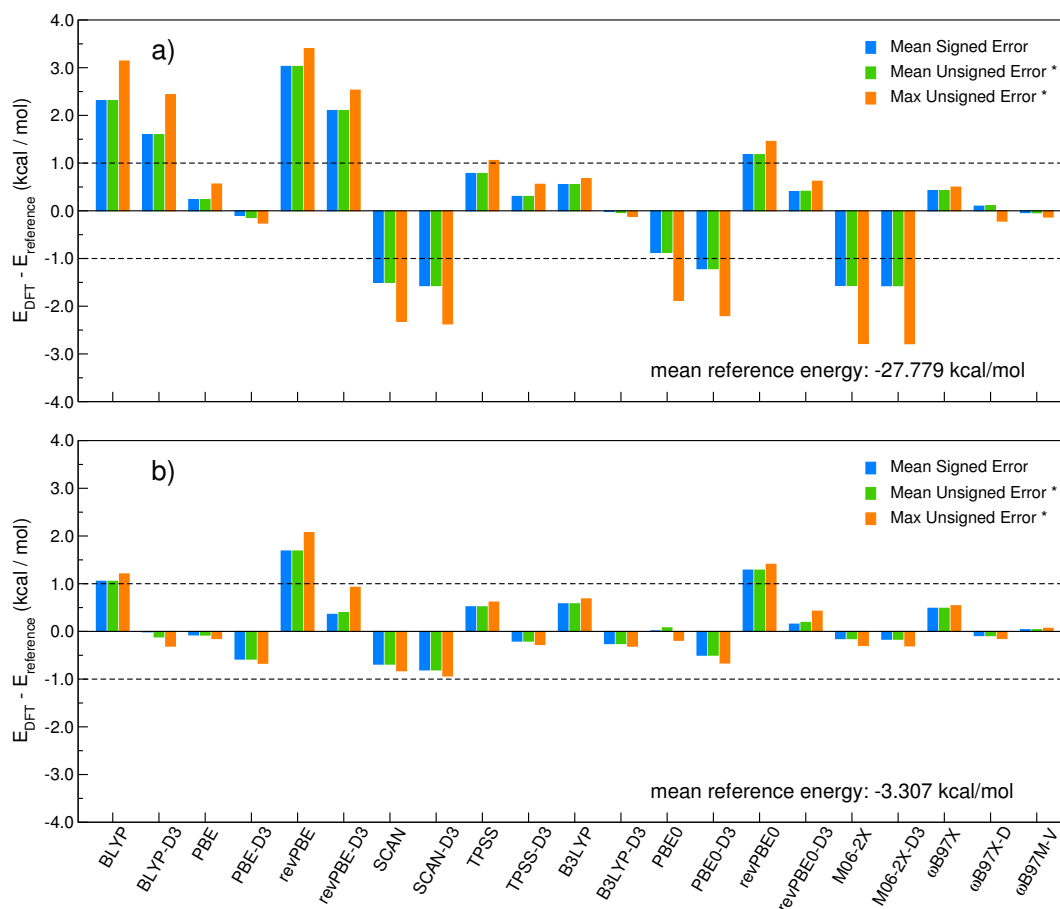


Figure 3.10. DFT errors for the hydrogen-bonded 2-mer fragment groups: (a) $\{\text{OH}^-, {}^1\text{W}\}_{FG}$; (b) $\{\text{W}_{\leftarrow B}^A, \text{A}\leftarrow \text{W}_B^B\}_{FG}$. *Mean Unsigned Errors are arbitrarily given the same sign as the Mean Signed Errors to ease comparison between the two, and the Maximum Unsigned Error is arbitrarily given the sign of the corresponding error value.

due to the overestimation of the HOMO energy of molecular species, resulting from incorrect features of the exchange-correlation functional, particularly in the asymptotic region.[102]

While the error increase is fairly drastic for some functionals, like M06-2X, and quite small for other functionals, such as PBE, the average errors of nearly all functionals for the deprotonated fragments show qualitative correlation with those for the $\{B \leftarrow W^A, W_{\leftarrow A}^B\}_{FG}$ fragments, illustrating the similarity of the interactions present in the two fragment groups. The observed correlation suggests that the magnitude of the error increase may reflect precisely the deficiencies of a given functional in the asymptotic region mentioned above. For example, the tendency of the BLYP and revPBE GGA functionals to underbind the neutral water dimer (Fig. 3.10b) is a result of the strongly-repulsive behavior of their exchange-enhancement factors in the high-density-gradient limit,[103, 104, 105] particularly in regions of large inter-monomer orbital overlap. Since the electron density of OH^- is far more diffuse than that of water, this deficiency is thus emphasized in the net-negatively charged fragments (Fig. 3.10a). A comparison with the corresponding data for H-bonded protonated water dimers in Fig. 3.22 show that the errors of GGA functionals are in fact larger, on average, for the negatively charged deprotonated water dimers than for the protonated water dimers. It should be noted, however, that GGA functionals are not more accurate in describing protonated water dimer 2B energies than neutral water dimer 2B energies (Fig. 3.22), indicating that deficiencies of the models are not strictly due to the diffusivity of the electron density.

Several functionals tend to have smaller average errors for the non-hydrogen bonded 2-mer fragments, $\{\text{OH}^-, {}^2\text{W}\}_{FG}$, and $\{W^A, W^B\}_{FG}$, than for the hydrogen bonded fragments as seen by comparing Fig. 3.11 with Fig. 3.10. In this context, it is important to note the differences between the mean signed errors (MSEs) and mean unsigned errors (MUEs) for the non-hydrogen bonded deprotonated fragments (Fig. 3.11a). These differences quantify the amount of error cancellation present in the 2B estimates, averaged over all clusters (the MUEs, which are technically positive by definition, are arbitrarily given the same sign as the MSEs to ease comparisons between the two sets of results). Error cancellation for the $\{\text{OH}^-, {}^2\text{W}\}_{FG}$ fragments is likely related to the ‘structural ambiguity’ associated with the non-hydrogen bonded fragment groups mentioned in Section 3.2.3 and is significant for BLYP, revPBE, and TPSS,

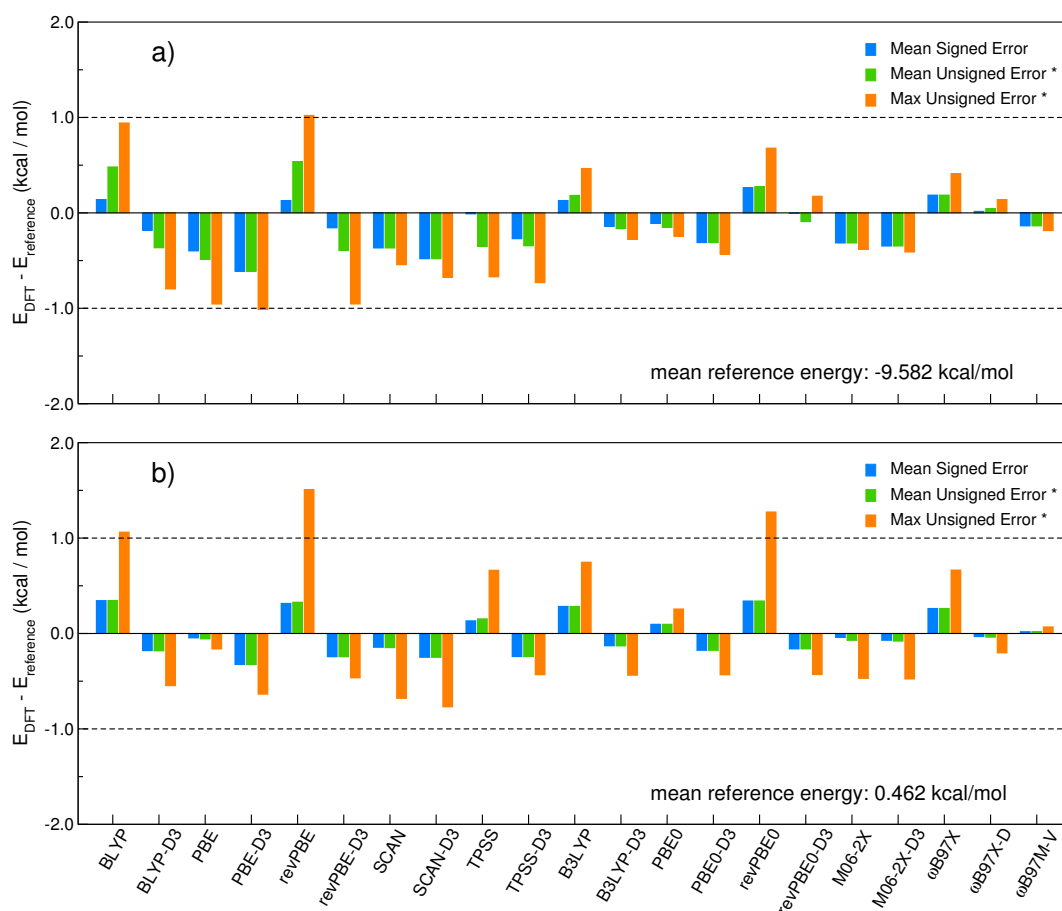


Figure 3.11. DFT errors for the non-hydrogen-bonded 2-mer fragment groups $\{\text{OH}^-, {}^2\text{W}\}_{FG}$ (panel *a*) and $\{\text{W}^A, \text{W}^B\}_{FG}$ (panel *b*).

whose MSEs are deceptively small.

Interestingly, there does not appear to be significant error cancellation for the $\{\text{W}^A, \text{W}^B\}_{FG}$ fragments (Fig. 3.11b). However, several functionals are associated with maximum unsigned errors that are substantially larger in magnitude than the corresponding MUEs (note that the signs of the maximum errors shown in the figure reflect the signs of the particular errors with the largest magnitude for each functional). The cause of these large maximum errors is related to the ‘structural ambiguity’ inherent to this fragment group. Specifically, several functionals are unable to reproduce 2B energy contributions of $\{\text{W}^A, \text{W}^B\}_{FG}$ fragments at small O–O separations. This can be seen in Fig. 3.12, which shows relative errors associated with

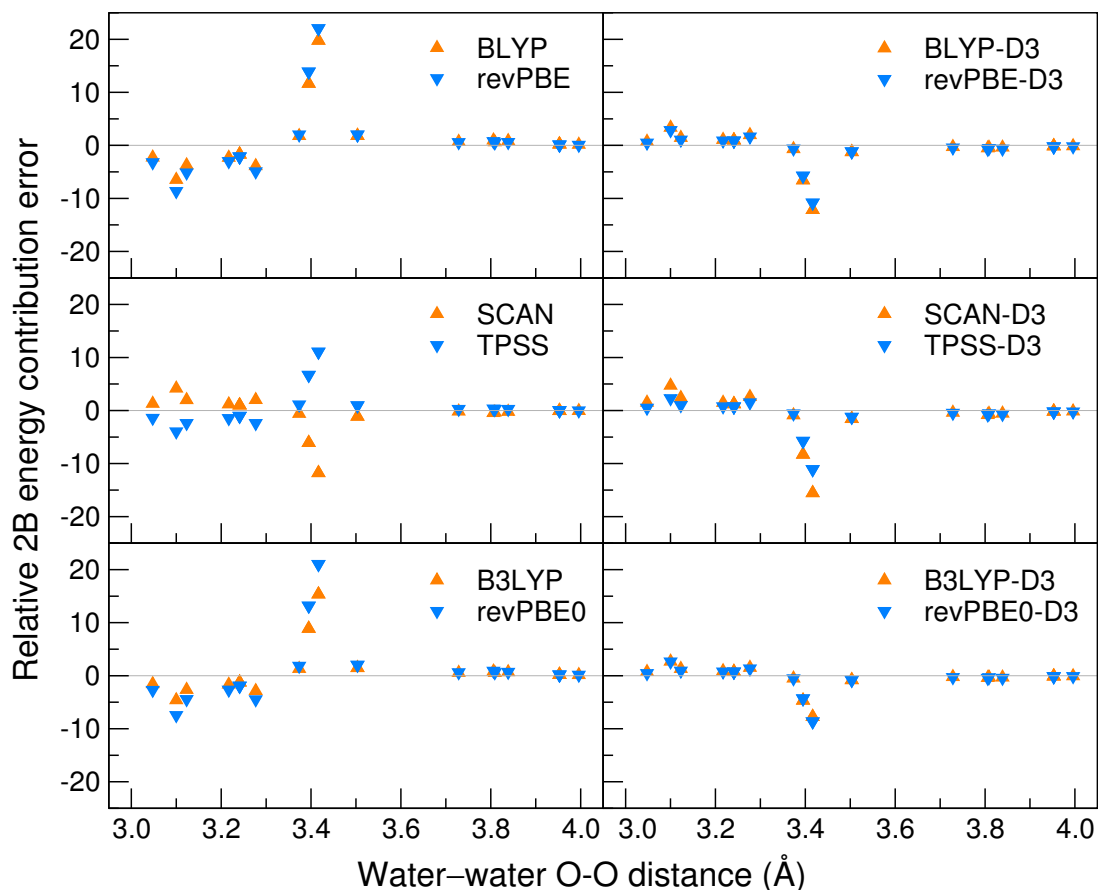


Figure 3.12. Relative 2B errors, $\frac{E_{DFT} - E_{ref}}{E_{ref}}$, as a function of the water–water O–O distance for the $\{W^A, W^B\}_{FG}$ fragments.

selected functionals for $\{W^A, W^B\}_{FG}$ fragments plotted as a function of the $W^A - W^B$ O–O distance from 3 to 4 Å. It should be noted that the reference CCSD(T)/CBS energies and associated errors remain close to zero beyond 4 Å.

The non-uniformity of the relative errors shown in Fig. 3.12 indicates deficiencies in the overall shape of the potential energy surfaces predicted by these functionals, which may thus result in an incorrect description of the local hydration structure of OH^- in aqueous solutions. For some functionals (e.g., BLYP, revPBE, and revPBE0) the largest maximum errors predicted for these fragments are reduced upon inclusion of the GD3 dispersion correction. It should be noted that relatively large maximum errors are also predicted by several functionals for the $\{OH^-, {}^2W\}_{FG}$ fragments, although they tend to be closer to the corresponding MUEs (Fig. 3.11a), thus

indicating, overall, a smaller degree of error cancellation.

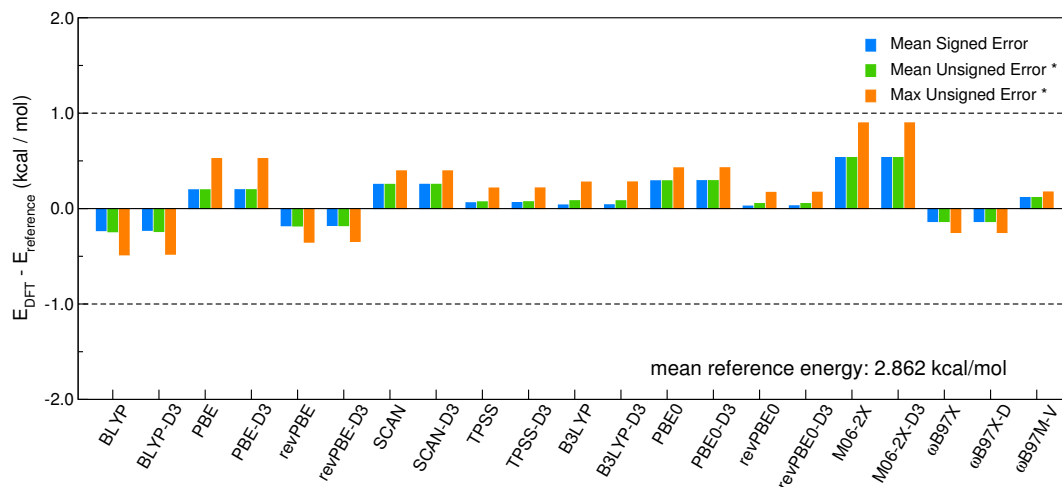


Figure 3.13. DFT errors for the anticooperative 3-mer fragment group $\{\text{OH}^-, {}^1\text{W}^{\text{A}}, {}^1\text{W}^{\text{B}}\}_{\text{FG}}$.

Overall, $\omega\text{B97M-V}$ and $\omega\text{B97X-D}$ exhibit the highest accuracy for all 2-mer fragments considered in this study, followed by the dispersion-corrected, hybrid functionals, B3LYP-D3 and revPBE0-D3. Finally, PBE is the best performer among the GGA functionals. It should be noted, however, that the dispersion-corrected PBE-D3 tends to perform worse than the original PBE, suggesting that the apparent success of the latter is largely incidental. This is also manifested by the rather poor accuracy of PBE in reproducing the 2B energies of the $\{\text{OH}^-, {}^2\text{W}\}_{\text{FG}}$ fragments as well as other higher-body terms (see Supporting Information of Ref. 4).

DFT errors for the $\{\text{OH}^-, {}^1\text{W}^{\text{A}}, {}^1\text{W}^{\text{B}}\}_{\text{FG}}$ 3-mer fragments are shown in Fig. 3.13. As discussed above, OH^- acts as a double-hydrogen bond acceptor in these fragments, leading to anticooperativity which results in exclusively repulsive 3B energies. Among all functionals considered in this study, BLYP and revPBE underestimate, while SCAN, PBE0, and M06-2X overestimate these 3B contributions. This can be related to the tendency of BLYP and revPBE to underbind $\{\text{OH}^-, {}^1\text{W}^{\text{A}}\}_{\text{FG}}$ 2-mer fragments, and the tendency of SCAN, PBE0, and M06-2X to overbind the OH^- –water hydrogen bond (Fig. 3.10).

Errors for the $\{\text{OH}^-, {}^1\text{W}_{\leftarrow\text{B}}^{\text{A}}, \text{A} \leftarrow {}^2\text{W}_{\text{FG}}^{\text{B}}\}$, and $\{\text{W}_{\leftarrow\text{B}}^{\text{A}}, \text{A} \leftarrow \text{W}_{\leftarrow\text{C}}^{\text{B}}, \text{B} \leftarrow \text{W}_{\text{FG}}^{\text{C}}\}$ fragment groups are shown in Fig. 3.14. Both groups are characterized by cooperative hydrogen bonds,

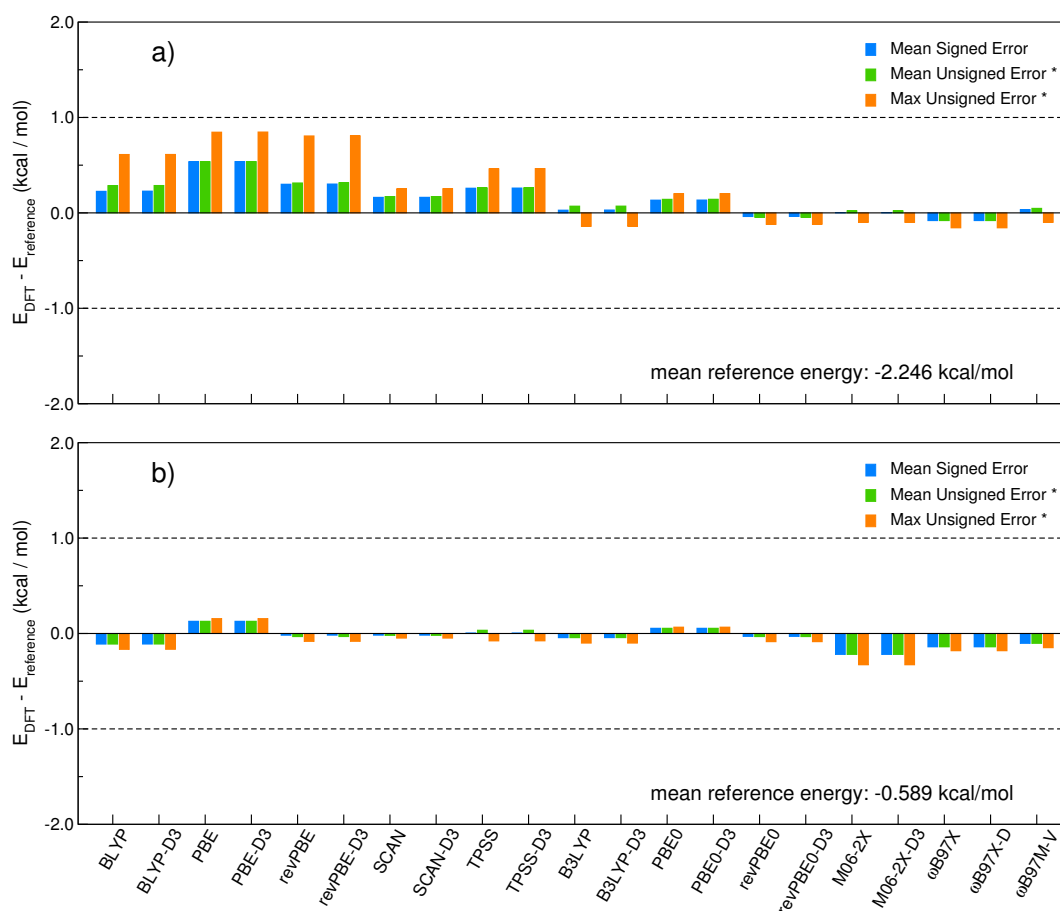


Figure 3.14. DFT errors for the cooperative 3-mer fragment groups $\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A\leftarrow {}^2\text{W}^B\}_{FG}$ (panel a) and $\{\text{W}_{\leftarrow B}^A, A\leftarrow \text{W}_{\leftarrow C}^B, B\leftarrow \text{W}^C\}_{FG}$ (panel b).

resulting in exclusively attractive 3B energies. On average, the errors for the deprotonated water fragments (Fig. 3.14a) tend to be larger than those for the neutral water fragments, particularly for the GGA and meta-GGA functionals which overestimate the 3B contributions of $\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A\leftarrow {}^2\text{W}^B\}_{FG}$ fragments, artificially counteracting cooperative effects that characterize the reference data (Fig. 3.14b). The remaining (hybrid) functionals show a high degree of accuracy for the $\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A\leftarrow {}^2\text{W}^B\}_{FG}$ group, suggesting that the inclusion of exact exchange is needed to correctly reproduce 3B interactions in these fragments. On the other hand, 3B energy contributions in the $\{\text{W}_{\leftarrow B}^A, A\leftarrow \text{W}_{\leftarrow C}^B, B\leftarrow \text{W}^C\}_{FG}$ fragments are reproduced by all functionals considered in this study. This result agrees with previous studies demonstrating that many

functionals are able to semi-quantitatively reproduce cooperative effects in water clusters.[105]

3.3 Many-Body Interactions in Protonated Water Clusters

The hydronium ion, H_3O^+ , like the hydroxide ion, plays an important role in aqueous acid/base chemistry, being the fundamental acidic unit in aqueous solutions. Given its importance in aqueous systems, it is desirable to develop models describing its interactions with water for use in molecular dynamics simulations. To this end, we investigated the many-body interactions present in low-energy $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ ($n \leq 5$) clusters. The $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters were reoptimized (see Sec. A.1) from structures reported in the literature.[73]

The $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters with $n = 3, 4, 5$ are shown in Fig. 3.15, and the clusters with $n = 1, 2$ are shown in Fig. 3.20a and Fig. 3.20b. The total interaction energies of all clusters are reported in Table 3.6. A discussion of the structures and relative energies of the $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters studied here can be found in Ref. 66, and xyz coordinates for all clusters can be found in the corresponding Supporting Information.

Additionally, the convergence of the MBE for the $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters with $n = 3, 4, 5$ are shown in Figs. 3.16–3.18. This analysis indicates that 2B and 3B contributions dominate the total interaction energies, which suggests that 2- and 3-mer clusters (and fragments) can effectively be considered as the primary ‘building blocks’ of protonated water systems, and will

Table 3.6. Reference interaction energies (in kcal/mol) for low-lying isomers of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters shown in Figs. 3.20 and 3.15.

Cluster index:	α	β	γ	δ	ϵ
2-mer	−52.241				
3-mer	−61.312				
4-mer	−92.561	−92.500	−79.028	−78.333	
5-mer	−106.889	−94.828	−94.279	−94.221	−93.952
6-mer	−120.768	−108.322	−107.959	−106.684	

be our primary focus. In Sec. 3.3.1, we extend the concept of a fragment group (introduced in Sec. 3.2.2) to include the context from the parent clusters. In Sec. 3.3.2, we discuss the global minimum energy $\text{H}_3\text{O}^+(\text{H}_2\text{O})$ and $\text{H}_3\text{O}^+(\text{H}_2\text{O})_2$ clusters. In Sec. 3.3.3, we discuss the many-body interactions present in the various $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ fragment groups and contextual fragment groups. In Sec. 3.3.4, we discuss the accuracy of various XC-functionals (Sec. A.1) in describing the many-body energies of the $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ fragment groups.

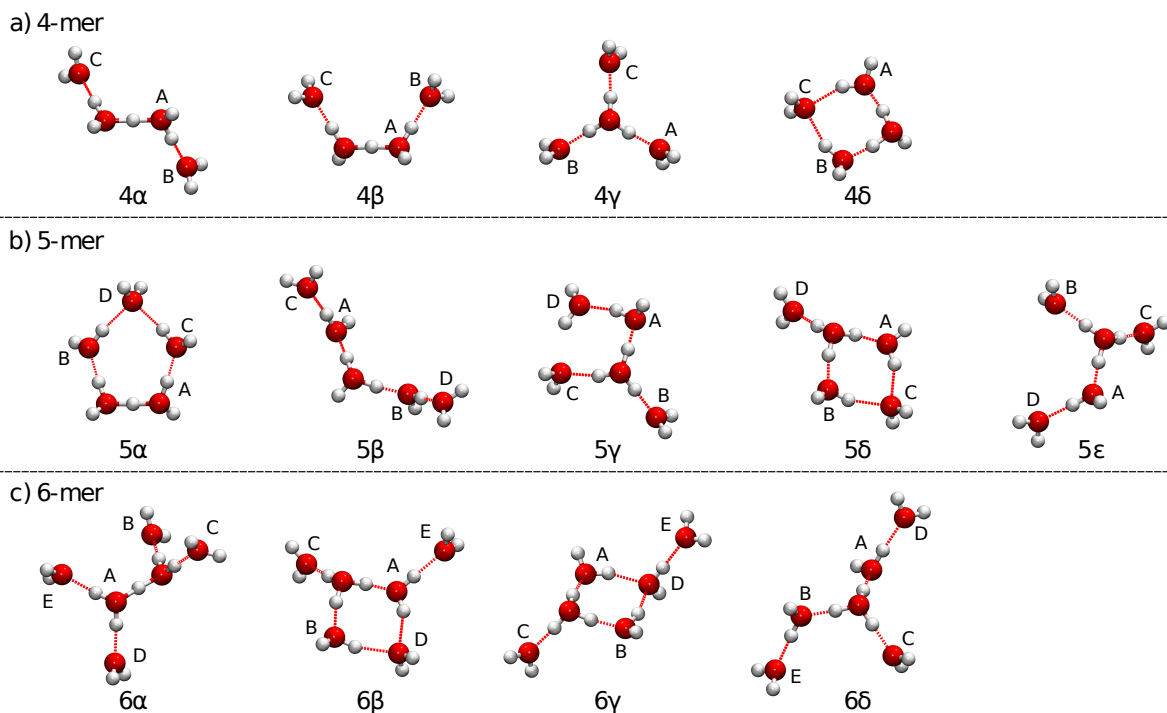


Figure 3.15. Low-lying isomers of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3$, $\text{H}_3\text{O}^+(\text{H}_2\text{O})_4$, and $\text{H}_3\text{O}^+(\text{H}_2\text{O})_5$ clusters, ordered in terms of their interaction energies.

3.3.1 Fragment Analysis With and Without Context

Important information about the molecular interactions within a given (protonated) water cluster is encoded in the H-bonding connectivity of that cluster.[4, 67, 68, 47, 49, 69, 70, 71, 72, 73, 74, 75] Since H-bonds are directional, their connectivity must be described by directed-graphs,[72, 73, 74, 76] which can be represented by the “fragment notation” introduced in the analysis of deprotonated water clusters.[4]

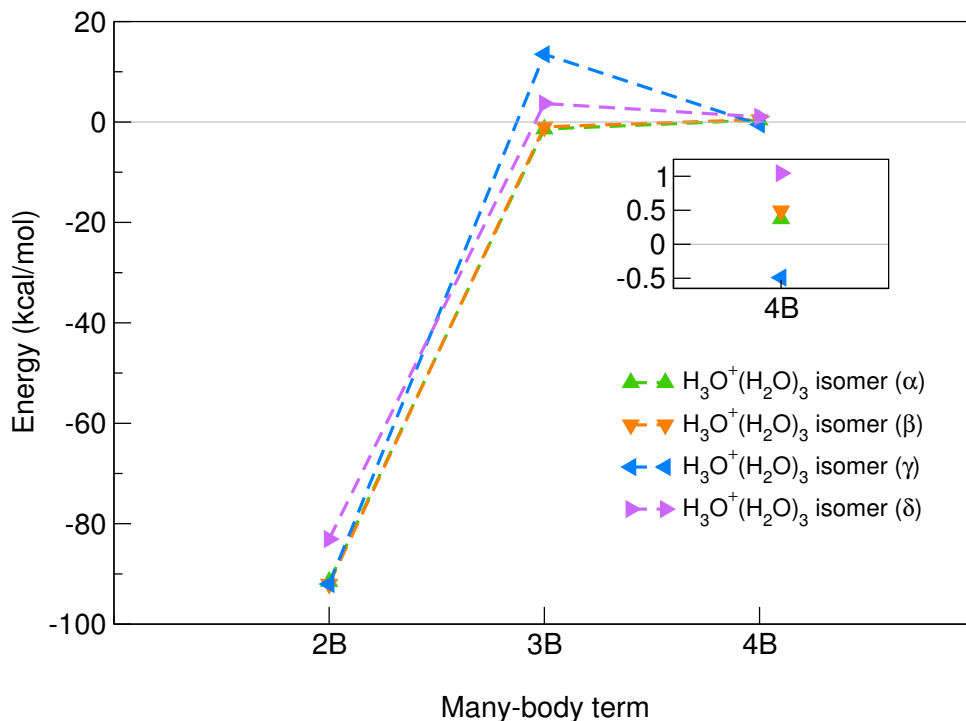


Figure 3.16. Convergence of the many-body expansion of the low-lying isomers of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3$ shown in Fig. 3.15a.

In order to describe various H-bonding connectivities in a compact form, we denote a particular water molecule as ${}_{A_1A_2\leftarrow}{}^S\text{W}_{\leftarrow D_1D_2}^X$, where left-superscript, S , a positive integer, denotes the solvation shell in which the water molecule resides (i.e., the smallest number of H-bonds separating it from H_3O^+), and right-superscript, X , is a capital Roman letter (A, B, C , etc.) which identifies the particular water molecule in Fig. 3.15. The left- and right-subscripts, A_1, A_2 , and D_1, D_2 , enumerate the identities (the capital Roman letters) of all water molecules accepting and donating H-bonds, respectively, from/to water molecule X . Left-pointing arrows in the subscripts are included to remind the reader that left-subscripted water molecules accept H-bonds from X , and that right-subscripted water molecules donate H-bonds to X .

For example, if 1st solvation shell water molecules A and B , ${}^1\text{W}^A$ and ${}^1\text{W}^B$, each donate an H-bond to 2nd solvation shell water molecule C , ${}^2\text{W}^C$, then the corresponding complete notation would be ${}_{C\leftarrow}{}^1\text{W}^A$, ${}_{C\leftarrow}{}^1\text{W}^B$, and ${}^2\text{W}_{\leftarrow AB}^C$ for the three molecules, respectively. If, in the same cluster, another 1st solvation shell water molecule D , ${}^1\text{W}^D$, is not H-bonded to any other

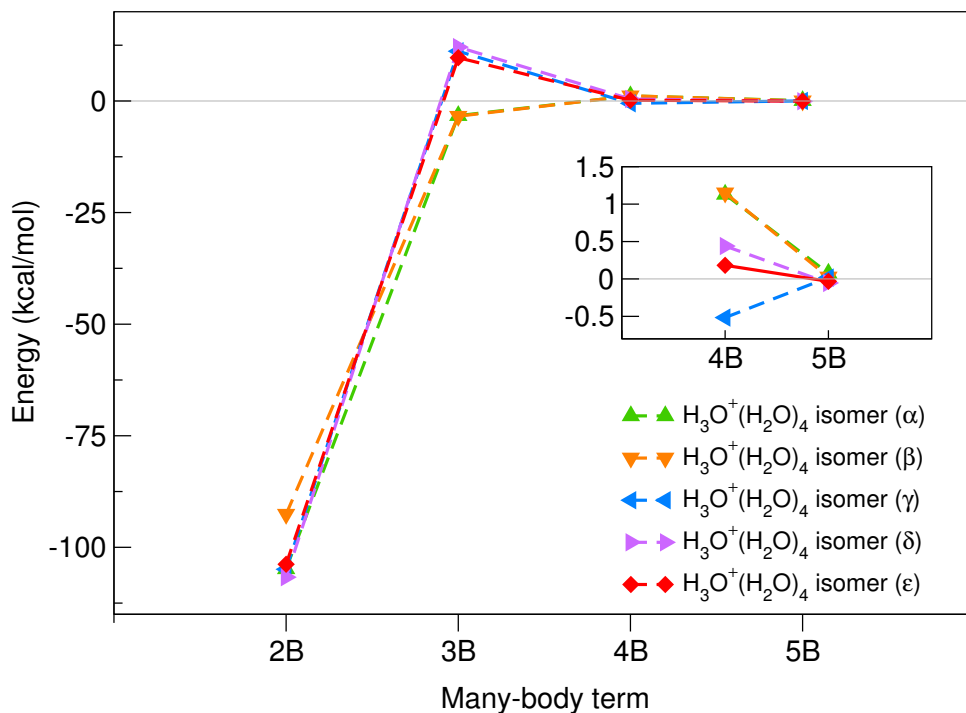


Figure 3.17. Convergence of the many-body expansion of the low-lying isomers of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_4$ shown in Fig.3.15b.

water molecules, then it would simply be denoted by ${}^1\text{W}^D$. Additionally, H_3O^+ will denote the hydronium ion, and Z^+ will denote the Zundel ion (O_2H_5^+ ; see Section 3.3.2 for more detail), when appropriate.

In Sections 3.3.3 and 3.3.4, all $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters and fragments are grouped by their H-bonding connectivity to facilitate the analysis. Each “fragment group” will be denoted by curly-braces with a subscripted “ FG ”, $\{\dots\}_{FG}$, enclosing the fragment notation names for all monomers belonging to that fragment group. For example, the fragment group for the cluster described above would be $\{\text{H}_3\text{O}^+, {}_{C\leftarrow}{}^1\text{W}^A, {}_{C\leftarrow}{}^1\text{W}^B, {}^1\text{W}^D, {}^2\text{W}_{\leftarrow AB}^C\}_{FG}$.

All water–water and H_3O^+ –water H-bonds are classified according to the inequality in Eq. 3.15.[77]

In the interest of analyzing fragment energetics with maximal specificity, we extend our fragment notation to include “context” from the parent clusters by enclosing one or more monomers inside parentheses ornamented with additional scripts. In particular, we use left-

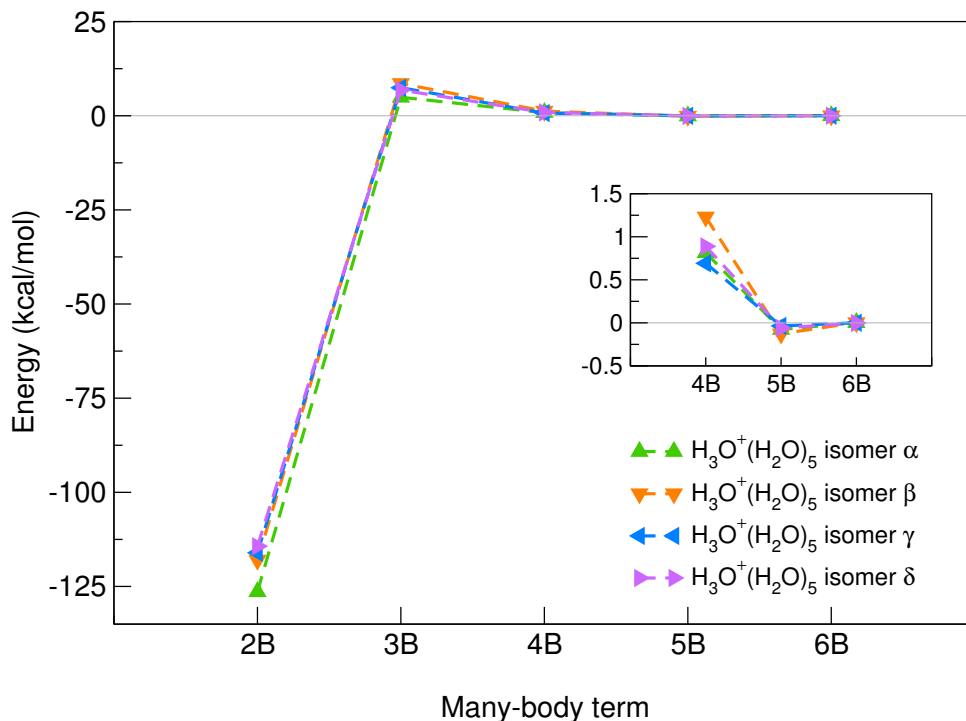


Figure 3.18. Convergence of the many-body expansion of the interaction energy for the low-lying isomers of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_5$ shown in Fig. 3.15c.

superscripts outside the parentheses will indicate solvation shells of water molecules in a cluster-fragment with respect to the parent cluster, when H_3O^+ is not present in that fragment. Additionally, right-subscripts will be used to indicate the presence of the Zundel ion.

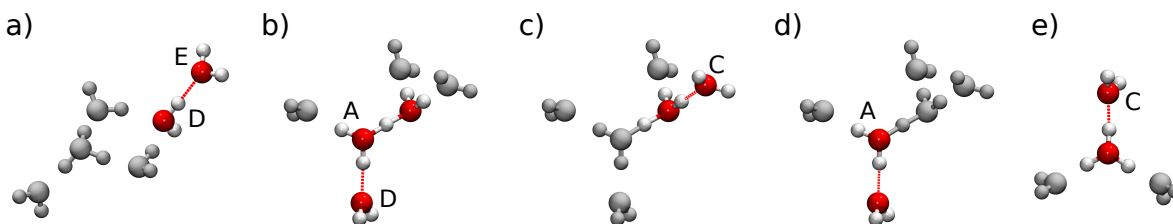


Figure 3.19. Examples of fragment group $\rightarrow$ contextual fragment group notation: a) $\{E \leftarrow W^D, W_{\leftarrow D}^E\}_{FG} \rightarrow \{^2(E \leftarrow W^D), ^3(W_{\leftarrow D}^E)\}_{CFG}$ b) $\{\text{H}_3\text{O}^+, ^1_{D \leftarrow} W^A, ^2W_{\leftarrow A}^D\}_{FG} \rightarrow \{(\text{H}_3\text{O}^+, D \leftarrow ^1W^A)_{Z+}, ^2W_{\leftarrow A}^D\}_{CFG}$ c) $\{\text{H}_3\text{O}^+, ^1W^C\}_{FG} \rightarrow \{(\text{H}_3\text{O}^+)_{CZ+}, ^1W^C\}_{CFG}$ d) $\{D \leftarrow W^A, W_{\leftarrow A}^D\}_{FG} \rightarrow \{(D \leftarrow W^A)_{CZ+}, ^2(W_{\leftarrow A}^D)\}_{CFG}$ e) $\{\text{H}_3\text{O}^+, ^1W^C\}_{FG} \rightarrow \{\text{H}_3\text{O}^+, ^1W^C\}_{CFG}$

Regarding the first usage of context-parentheses: the fragment group notation introduced above will have no indication of the solvation structure of a neutral-water fragment with respect

to H_3O^+ in the parent cluster. For example, the fragment in Fig. 3.19a (not including the grayed-out molecules) belongs to the $\{_{E\leftarrow} \text{W}^D, \text{W}_{\leftarrow D}^E\}_{FG}$ fragment group (note that since the Roman letters are arbitrary, one might call this group $\{_{B\leftarrow} \text{W}^A, \text{W}_{\leftarrow A}^B\}_{FG}$). In “contextual fragment group” notation, each of these water molecules gains parentheses with superscripts indicating their solvation shells within the parent cluster (including the grayed-out molecules). Therefore, the fragment in Fig. 3.19a belongs to the $\{^2_{(E\leftarrow} \text{W}^D), ^3(\text{W}_{\leftarrow D}^E)\}_{CFG}$ contextual fragment group. Note that the braces for a contextual fragment group contain a subscript “CFG” to avoid confusion. This use of context-parentheses may facilitate identifying general trends in the convergence of many-body effects within different solvation shells.

The second usage of context-parentheses allows for comparisons between fragments involving the Zundel ion, with those that do not. One such usage is in referring to the full Zundel ion, O_2H_5^+ , as $(\text{H}_3\text{O}^+, {}^1\text{W}^A)_{Z^+}$. The significance of this is due to the fact that comparisons of many-body energies are simplified when (molecular) monomer units are defined in a consistent way. Therefore, in the following analysis, all many-body energies are computed with respect to H_2O and H_3O^+ monomer units, such that monomers are defined to minimize the total 1B energy. For the clusters presented here, it suffices to define that hydrogens belong to the closest oxygen in the absence of the Zundel ion. However, since the Zundel ion contains a symmetrically-shared proton, each Zundel ion is arbitrarily divided into one H_3O^+ monomer, and one H_2O , i.e. without double-counting. In other words, each Zundel-containing cluster is defined here to contain exactly one $(\text{H}_3\text{O}^+, {}^1\text{W}^A)_{Z^+}$ fragment, and any many-body energies including it will consider H_3O^+ and ${}^1\text{W}^A$ to be distinct monomer units. For example, the fragment in Fig. 3.19b belongs to the $\{(\text{H}_3\text{O}^+, {}_{D\leftarrow} {}^1\text{W}^A)_{Z^+}, {}^2\text{W}_{\leftarrow A}^D\}_{CFG}$ 3-mer contextual fragment group, and the $\{\text{H}_3\text{O}^+, {}_{D\leftarrow} {}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^D\}_{FG}$ 3-mer fragment group.

Additionally, when a fragment contains part of the Zundel ion (either H_3O^+ , or ${}^1\text{W}^A$, exclusively), the monomer that belongs to the Zundel ion in the parent cluster will be distinguished with a right-subscript, “ $\subset Z^+$,” outside of context parentheses. For example, the fragment in Fig. 3.19c belongs to the $\{(\text{H}_3\text{O}^+)_{\subset Z^+}, {}^1\text{W}^C\}_{CFG}$ contextual fragment group, and the fragment

in Fig. 3.19d belongs to the $\{(D \leftarrow W^A)_{CZ^+}, {}^2(W_{\leftarrow A}^D)\}_{CFG}$ contextual fragment group.

Note that when neither usage of context-parentheses applies to a given fragment, the contextual fragment group notation coincides with the fragment group notation, as is the case for the fragment in Fig. 3.19e, which belongs to the $\{H_3O^+, {}^1W^C\}_{CFG}$ contextual fragment group.

3.3.2 The Minimum Energy $H_3O^+(H_2O)$ and $H_3O^+(H_2O)_2$ Clusters

The convergence of the MBE for the hydration of H_3O^+ is shown in Figs. 3.16–3.18 for the low-lying isomers of the $H_3O^+(H_2O)_n$ clusters with $n = 3, 4, 5$. The MBE results indicate that 2B and 3B contributions dominate the total interaction energies of these protonated water clusters. We therefore begin the analysis with the global minimum energy 2- and 3-mer clusters, and their substituent fragments, which serve as the primary building blocks of all $H_3O^+(H_2O)_n$ clusters. Following the expression in Eq. 3.15, the H-bond acceptor and donor-oxygen atoms are labeled as A and D, respectively.

The $H_3O^+(H_2O)$ cluster shown in Fig. 3.20a corresponds to the Zundel ion,[106, 107, 108] which is characterized by a central proton shared equally between two water molecules. As a result, the Zundel ion is C_2 symmetric, having two pairs of equivalent free (flanking) hydrogens (see Fig. 3.21a). The C_2 axis lies perpendicular to the O–O axis, so that the component of the total dipole moment along the O–O axis is null, stabilizing the shared proton to be equidistant from each oxygen. This feature is common to all Zundel-containing clusters in our data set. Within any Zundel-type structure, the equally-shared proton prevents the identification of a unique H_3O^+ molecule. Furthermore, the Eigen ion is characterized by a (3-coordinated) H_3O^+ molecule, which donates an H-bond to three water molecules[109, 110, 111] (see Fig. 3.21b, or cluster 4 γ in Fig. 3.15). In contrast with Zundel-type structures, any Eigen-type structure has a distinct H_3O^+ molecule.

Within the minimum energy $H_3O^+(H_2O)$ cluster, each pair of equivalent flanking hydrogens form H–O–O–H dihedral angles of 100.47° (see Fig. 3.21a). Additionally, the O–O distance is 2.39 Å, and the shared proton is 1.20 Å away from each oxygen, which is nearly colinear with

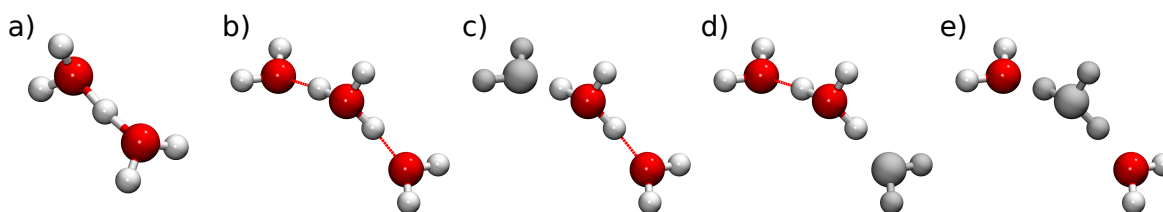


Figure 3.20. $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ building blocks. a) $\text{H}_3\text{O}^+(\text{H}_2\text{O})$ global minimum energy structure. b) $\text{H}_3\text{O}^+(\text{H}_2\text{O})_2$ global minimum-energy structure along with its constituent 2-mer fragments (c–e).

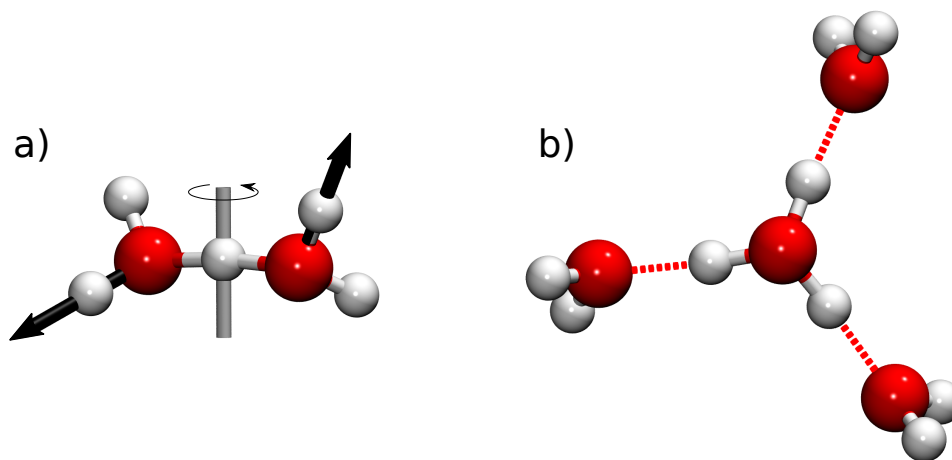


Figure 3.21. a) The Zundel ion shown with the C_2 axis passing through the shared proton (in gray), and one H–O–O–H dihedral angle between equivalent flanking hydrogens (black arrows). b) The Eigen ion

the O–O axis, having an O–H–O angle of 173.57° (corresponding to an acceptor-donor-donor O–O–H angle of 3.22°).

The 2B energy of the $\text{H}_3\text{O}^+(\text{H}_2\text{O})$ cluster is -52.241 kcal/mol, while the corresponding value for the global minimum energy $\text{OH}^-(\text{H}_2\text{O})$ cluster is -37.092 kcal/mol.[4]

A useful metric of the strength of a H-bond ($\text{D}-\text{H}\cdots\text{A}$) is the lengthening of the donor-hydrogen D–H covalent bond resulting from a combination of the electrostatic attraction between the donor-hydrogen and the acceptor-oxygen, as well as charge transfer.[7, 59] In the case of the $\text{OH}^-(\text{H}_2\text{O})$ 2-mer, for example, the charge transfer involved in the H-bond can be thought of a donation of electron density from the OH^- lone pair into the (bound) water H–O antibonding orbital, constituting what may be called an “incipient chemical bond.”[7] However, this picture

clearly differs in the Zundel ion, given that there is no distinct donor, nor acceptor, beckoning the question of whether the corresponding interaction can even be classified as a H-bond. The Quantum Theory of Atoms in Molecules[42, 43] provides a route with which to characterize covalent and noncovalent interactions in terms of the topological properties of the electron density,[112] and has been used to argue that the shared proton in the Zundel ion is bound via a 3-center 4-electron bond with the two oxygens,[112, 113, 114] suggesting that the Zundel ion should be thought of as a single molecule.

In bulk solution at room temperature, the distinction between Zundel-type structures and Eigen-type structures becomes blurred due to molecular motion and vibration.[115, 116, 117] However, such structural motifs are easy to define in optimized clusters, in which the Zundel ion shared proton is equidistant from the two Zundel ion oxygen nuclei. While the primary goal of the present cluster analysis is to extend findings to bulk systems, there is value in considering these two limiting structures. Therefore, we pay particular attention to differences in energetics between Zundel-type and Eigen-type structures, but the energies reported here only consider H_3O^+ to be the protonated monomer unit so that all clusters and cluster-fragments can be compared.

The $\text{H}_3\text{O}^+(\text{H}_2\text{O})_2$ cluster and its substituent fragments are shown in Fig 3.20b–e. The 3-mer lacks a C_2 axis perpendicular to a protonated O–O axis (the cluster has C_s symmetry), so it is not surprising that the Zundel ion is not present. The H_3O^+ bound-O–H bonds are 1.04 Å, and both H_3O^+ –water O–O distances are 2.49 Å. The O–O–O angle is 112.69° , and both O–O–H angles are 3.26° . Both H_3O^+ –water fragments contribute 2B energies of -34.130 kcal/mol, which are more attractive than the average $\{\text{H}_3\text{O}^+, {}^1\text{W}\}_{\text{CFG}}$ 2B energy of -31.930 kcal/mol (see Table 3.8). The water–water 2B energy contribution is $+1.081$ kcal/mol, and the parent 3-mer has a total 2B energy of -67.178 kcal/mol. The 3B energy of this cluster is $+5.866$, which agrees with expectations since H_3O^+ acts as a double-donor (it belongs to $\{\text{H}_3\text{O}^+, {}^1\text{W}^{\text{A}}, {}^1\text{W}^{\text{B}}\}_{\text{FG}}$), which is known to lead to anticooperative 3B energies.[67, 68, 47, 49, 69, 70, 71]

3.3.3 Fragment Analysis and Many-Body Effects in Protonated Water Clusters

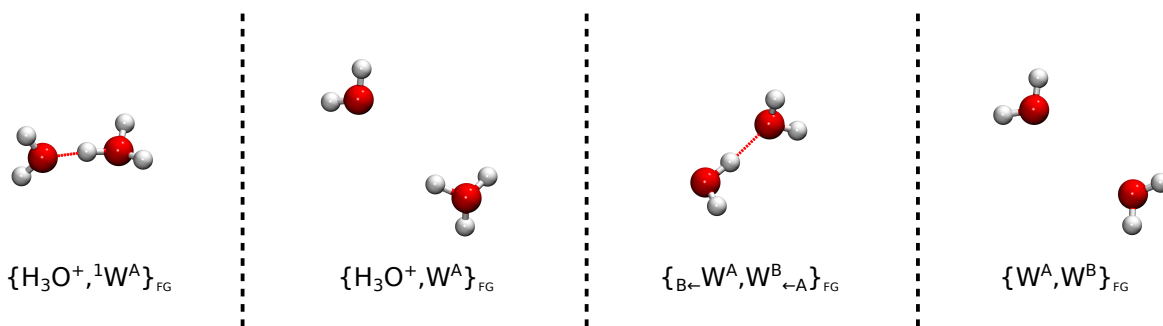
Table 3.7 lists the 2B energy statistics for all 2-mer fragment groups from all clusters analyzed in the previous section. The “freq.” column reports the number of fragments (substituent to all clusters) belonging to the corresponding fragment group. The “max.”, “min.”, and “mean” columns show the signed-maximum, signed-minimum (i.e., the most repulsive and attractive fragment energy contributions, respectively), and signed-mean of the 2B energy contributions among all fragments in each group. The “RSD” (relative standard deviation) column shows the standard deviation of the 2B energy contributions divided by the absolute value of their signed-mean. Values in the “sig.” column quantify the “significance” of each fragment group, and are calculated by multiplying the unsigned-mean energy (not shown) by the occurrence frequency of fragments in that group, divided by the total number of fragments with the same fragment size (from all groups). Fragment groups are listed in descending significance order. In all reported statistics, a single choice of H_3O^+ is considered within each cluster containing the Zundel ion, so that, for example, isomer 4α only contains one $\{(\text{H}_3\text{O}^+)_{\text{CZ}^+}, {}^1\text{W}^{\text{A}}\}_{\text{CFG}}$ substituent fragment.

Tables 3.7 shows that most fragment groups contain exclusively attractive or exclusively repulsive 2B energies, indicating that the classification adopted in this analysis reveals information about specific many-body interactions within $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters. The groups that contain fragments with both attractive and repulsive 2B energies turn out to be structurally ambiguous since the non-H-bonded region of a given configurational space is much larger than the complementary H-bonded region. In particular, the non-H-bonded $\{\text{W}^{\text{A}}, \text{W}^{\text{B}}\}_{\text{FG}}$ group contains fragments with positive and negative 2B energies, and has an RSD of 0.925 kcal/mol. This group happens to have a positive signed mean 2B energy (0.504 kcal/mol), indicating that the $\text{H}_3\text{O}^+(\text{H}_2\text{O})_{n \leq 5}$ clusters are stabilized by the more strongly-interacting H-bonded fragments.

The 2-mer fragments with the most significant 2B energies belong to the $\{\text{H}_3\text{O}^+, {}^1\text{W}^{\text{A}}\}_{\text{FG}}$

Table 3.7. Statistics for 2B energies (in kcal/mol) of 2-mer fragment groups from all $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ ($n \leq 5$) clusters (top) along with examples of the corresponding molecular configurations (bottom).

Frag. Group	freq.	max.	min.	mean	RSD	sig.
$\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{FG}$	37	-27.6	-52.2	-34.2	0.2	9.2
$\{\text{H}_3\text{O}^+, \text{W}^A\}_{FG}$	18	-3.9	-11.0	-8.1	0.2	1.1
$\{B \leftarrow \text{W}^A, \text{W}_{\leftarrow A}^B\}_{FG}$	23	-2.6	-4.3	-3.8	0.1	0.6
$\{\text{W}^A, \text{W}^B\}_{FG}$	60	+1.6	-2.7	+0.5	0.9	0.3

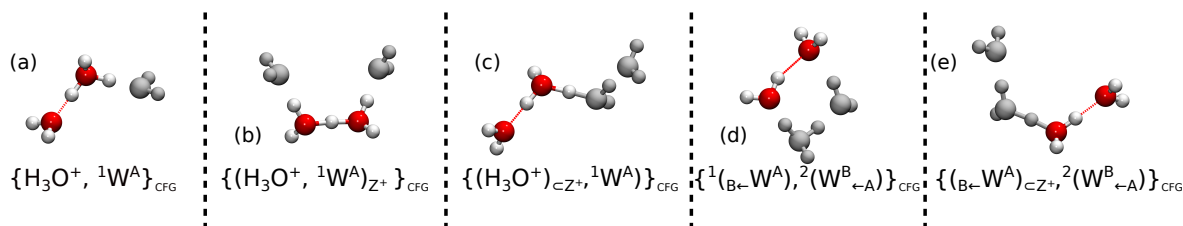


group. The minimum interaction energy fragment in this group corresponds to the minimum-energy structure of the $\text{H}_3\text{O}^+(\text{H}_2\text{O})$ 2-mer, which, as discussed in Section 3.3.2, is the Zundel ion. Since the interactions within the Zundel ion are qualitatively different from those in a typical H_3O^+ –water complex,[112, 113, 114] we further partition the 2-mer fragment set into contextual fragment groups, which distinguish these two contexts of $\text{H}_3\text{O}^+(\text{H}_2\text{O})$ fragments. 2B energy statistics for notable 2-mer contextual fragment groups are shown in Table 3.8 (the full tables of all statistics are reported in the Supporting Information of Ref. 66).

The first three rows of Table 3.8 correspond to different contexts of the $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{FG}$ group. Comparing the statistics of the $\{(\text{H}_3\text{O}^+, {}^1\text{W}^A)_{Z^+}\}_{CFG}$, $\{(\text{H}_3\text{O}^+)_{CZ^+}, {}^1\text{W}^A\}_{CFG}$ and $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{CFG}$ context fragment groups, it is clear that 2B energies on the order of -50 kcal/mol exclusively occur within a Zundel-type structure, corresponding to the covalent bond between the shared proton and one oxygen atom (see Section 3.3.2), while typical H_3O^+ –water H-bonds tend to have 2B energies on the order of -30 kcal/mol. Comparing the range of energies

Table 3.8. Statistics for 2B energies (in kcal/mol) of notable 2-mer fragment groups in context (top) along with examples of the corresponding molecular configurations (bottom).

Frag. Group	freq.	max.	min.	mean	RSD	sig.
$\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{CFG}$	27	-27.6	-36.5	-31.9	0.07	6.2
$\{(\text{H}_3\text{O}^+, {}^1\text{W}^A)_{Z^+}\}_{CFG}$	5	-50.2	-52.2	-51.4	0.01	1.9
$\{(\text{H}_3\text{O}^+)_{CZ^+}, {}^1\text{W}^A\}_{CFG}$	5	-27.7	-29.8	-28.9	0.03	1.0
$\{{}^1(\text{B} \leftarrow \text{W}^A), {}^2(\text{W}^B \leftarrow \text{A})\}_{CFG}$	15	-3.5	-4.1	-3.9	0.04	0.4
$\{(\text{B} \leftarrow \text{W}^A)_{CZ^+}, {}^2(\text{W}^B \leftarrow \text{A})\}_{CFG}$	5	-2.6	-3.8	-3.1	0.2	0.1
$\{{}^2(\text{B} \leftarrow \text{W}^A), {}^3(\text{W}^B \leftarrow \text{A})\}_{CFG}$	1	-4.3	-4.3	-4.3	N/A	0.03



for the latter two CFGs, it appears that $\{(\text{H}_3\text{O}^+)_{CZ^+}, {}^1\text{W}^A\}_{CFG}$ fragments tend to be much less diverse than $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{CFG}$ fragments, with $\{(\text{H}_3\text{O}^+)_{CZ^+}, {}^1\text{W}^A\}_{CFG}$ fragments being among the least attractive fragments within the entire $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{FG}$ group. However, this may be an artifact of the limited size of clusters considered here, in that lower-energy $\{(\text{H}_3\text{O}^+)_{CZ^+}, {}^1\text{W}^A\}_{CFG}$ fragments may be stabilized within larger clusters.

$\{\text{H}_3\text{O}^+, \text{W}^A\}_{FG}$ is the second most significant 2-mer group. Despite the lack of any H-bonding in this group, all $\{\text{H}_3\text{O}^+, \text{W}^A\}_{FG}$ fragments have negative 2B energies, with a mean of -8.103 kcal/mol, which is substantially more attractive than a typical water–water H-bond. This is likely due to the fact that none of the $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters considered here have ${}^1\text{W}$ molecules that accept H-bonds from ${}^2\text{W}$ molecules, resulting exclusively in $\{\text{H}_3\text{O}^+, {}^2(\text{W}^A)\}_{CFG}$ fragments whose W^A oxygens are oriented toward H_3O^+ . This also happens to be the case for the single $\{\text{H}_3\text{O}^+, {}^3(\text{W}^A)\}_{CFG}$ fragment in our data set (the $\{\text{H}_3\text{O}^+, \text{W}^E\}$ fragment within isomer 6γ of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_5$). This structural uniformity of the $\{\text{H}_3\text{O}^+, \text{W}^A\}_{FG}$ group is also reflected in

its RSD of 0.227 kcal/mol.

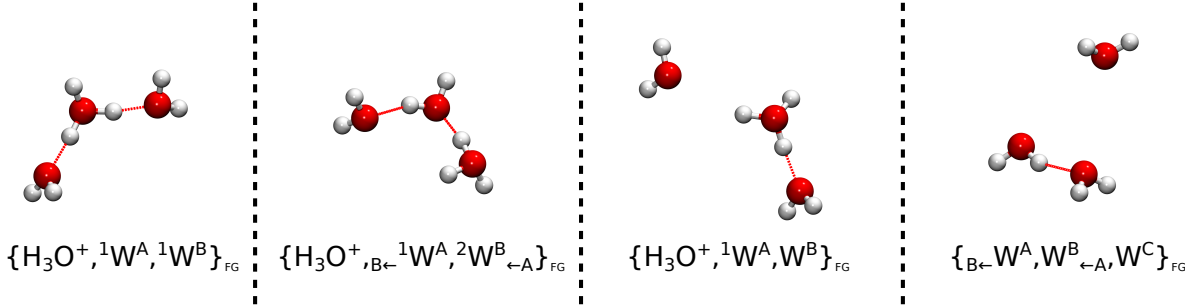
The third most significant 2-mer fragments belong to the H-bonded $\{B \leftarrow W^A, W_{\leftarrow A}^B\}_{FG}$ group. The corresponding contextual fragment group statistics in Table 3.8 show that $\{(B \leftarrow W^A)_{CZ^+}, {}^2(W_{\leftarrow A}^B)\}_{CFG}$ 2B energies tend to be less attractive than the corresponding values of the other contexts of the $\{B \leftarrow W^A, W_{\leftarrow A}^B\}_{FG}$ group, the signed mean 2B energy for which is greater than the signed maximum 2B energy of $\{{}^1(B \leftarrow W^A), {}^2(W_{\leftarrow A}^B)\}_{CFG}$ and $\{{}^2(B \leftarrow W^A), {}^3(W_{\leftarrow A}^B)\}_{CFG}$ fragments. Furthermore, the highest energy $\{B \leftarrow W^A, W_{\leftarrow A}^B\}_{FG}$ fragment belongs to the $\{(B \leftarrow W^A)_{CZ^+}, {}^2(W_{\leftarrow A}^B)\}_{CFG}$ subgroup, having a 2B energy of -2.577 kcal/mol. However, the H-bond definition used here (Eq. 3.15) permits $\{B \leftarrow W^A, W_{\leftarrow A}^B\}_{FG}$ 2B energies that are greater than -2.6 kcal/mol (see Table 3.2), suggesting that the presence of an Eigen-type H_3O^+ within stable $H_3O^+(H_2O)_n$ (with $n \leq 5$) clusters promotes the formation of strong water–water H-bonds within the first two or three hydration shells, in agreement with experiment,[116] whereas H-bonds between the Zundel ion and 1W tend to be weaker. This is consistent with the analysis of the $\{(H_3O^+)_{CZ^+}, {}^1W^A\}_{CFG}$ subgroup.

Statistics for 3B energies of significant 3-mer fragment groups are shown in Table 3.9. Important information regarding the strength and character of 3B interactions in aqueous systems is revealed when considering cooperative and anticooperative H-bonding.[67, 68, 69, 71, 4] In general terms, 3B cooperative effects result when a H-bond between two molecules strengthens a H-bond involving a third molecule, yielding a negative 3B energy between the three molecules. Conversely, 3B anticooperativity results when two H-bonds between three molecules simultaneously weaken each other, yielding a positive 3B energy. Both cooperativity and anticooperativity are consequences of the way in which monomer electron densities are distorted (i.e., polarization and charge transfer) due to the particular H-bonding interactions involved in a given 3-mer fragment, such that fragment configurations with a molecule simultaneously accepting and donating a H-bond result in cooperative effects, while configurations with a double-donor or a double-acceptor molecule result in anticooperative effects.[67, 68, 47, 49, 69, 70, 71, 4]

The most “significant” 3-mer fragment group listed in Table 3.9 is $\{H_3O^+, {}^1W^A, {}^1W^B\}_{FG}$,

Table 3.9. Statistics for 3B energies (in kcal/mol) of significant 3-mer fragment groups from all $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ ($n \leq 5$) clusters (top) along with examples of the corresponding molecular configurations (bottom).

Frag. Group	freq.	max.	min.	mean	RSD	sig.
$\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$	30	+8.3	+3.0	+5.4	0.2	1.1
$\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B\}_{FG}$	21	−1.8	−10.0	−5.1	0.5	0.7
$\{\text{H}_3\text{O}^+, {}^1\text{W}^A, \text{W}^B\}_{FG}$	26	+2.1	−0.3	+0.6	0.7	0.1
$\{{}_{B\leftarrow}\text{W}^A, \text{W}_{\leftarrow A}^B, \text{W}^C\}_{FG}$	32	+0.4	−0.5	+0.1	0.9	0.03
$\{{}_{C\leftarrow}\text{W}^A, {}_{C\leftarrow}\text{W}^B, \text{W}_{\leftarrow AB}^C\}_{FG}$	5	+1.1	+0.8	+0.8	0.1	0.03
$\{{}_{C\leftarrow}\text{W}_{\leftarrow B}^A, {}_{A\leftarrow}\text{W}^B, \text{W}_{\leftarrow A}^C\}_{FG}$	3	−1.0	−1.2	−1.1	0.1	0.02
$\{{}_{BC\leftarrow}\text{W}^A, \text{W}_{\leftarrow A}^B, \text{W}_{\leftarrow A}^C\}_{FG}$	2	+1.6	+1.1	+1.4	N/A	0.02
$\{\text{H}_3\text{O}^+, {}_{B\leftarrow}\text{W}^A, \text{W}_{\leftarrow A}^B\}_{FG}$	2	−0.242	−2.144	−1.193	N/A	0.016



which contains fragments with H_3O^+ acting as a double H-bond donor to two water molecules. The H-bonds in these fragments can be classified as anticooperative since 3B energies for this fragment group are exclusively positive, with a mean value of 5.356 kcal/mol, suggesting that anticooperativity in protonated water follows similar trends as anticooperativity in neutral water since the analogous $\{{}_{BC\leftarrow}\text{W}^A, \text{W}_{\leftarrow A}^B, \text{W}_{\leftarrow A}^C\}_{FG}$ fragments have anticooperative H-bonding.[67, 47, 49, 69, 70, 4] As was the case for the anticooperative $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragment group (Sec. 3.2.3), the 3B energy of $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments correlate with the average $\text{H}_3\text{O}^+ - \text{water O}-\text{O}$ distance. The physical effects which cause this correlation are explained in terms of 3B energy decomposition analysis in Sec. 4.4.

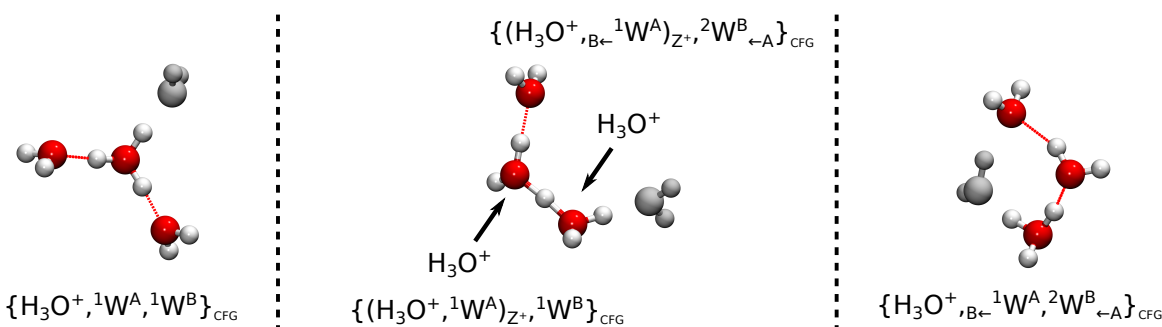
Analysis of the contextual fragment groups associated with $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$

offers another perspective on the nature of anticooperativity in these fragments. Table 3.10 reports the results for three prototypical anticooperative motifs of the 3-mer fragment group. The most significant of these is the $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{CFG}$ subgroup which only includes fragments containing Eigen-type H_3O^+ moieties. The second most significant subgroup is $\{(\text{H}_3\text{O}^+, {}^1\text{W}^B)_{Z^+}, {}^1\text{W}^A\}_{CFG}$ whose fragments contain the Zundel ion. Relating the strong covalent character of the interaction between each oxygen atom and the shared proton within the Zundel ion (see Section 3.3.2) to the particularly repulsive 3B energies of $\{(\text{H}_3\text{O}^+, {}^1\text{W}^B)_{Z^+}, {}^1\text{W}^A\}_{CFG}$ fragments supports the notion of anticooperativity described above.

Table 3.9 shows that $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B\}_{FG}$ is the second-most significant 3-mer fragment group within $\text{H}_3\text{O}^+(\text{H}_2\text{O})_{n=3-5}$ clusters. The 3B energies of these fragments are exclusively negative, with a mean value of -5.126 kcal/mol, which implies that they are characterized by cooperative interactions. This parallels the cooperativity trends in neutral water since the

Table 3.10. Statistics for 3B energies (in kcal/mol) of notable 3-mer fragment groups in context (top) along with examples of the corresponding molecular configurations (bottom).

Frag. Group	freq.	max.	min.	mean	RSD	sig.
$\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{CFG}$	24	+7.5	+3.6	+5.0	0.2	0.8
$\{(\text{H}_3\text{O}^+, {}^1\text{W}^A)_{Z^+}, {}^1\text{W}^B\}_{CFG}$	5	+8.3	+6.6	+7.4	0.09	0.3
$\{(\text{H}_3\text{O}^+)_{CZ^+}, {}^1\text{W}^A, {}^1\text{W}^B\}_{CFG}$	1	+3.0	+3.0	+3.0	N/A	0.02
$\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B\}_{CFG}$	15	-1.8	-6.7	-3.9	0.5	0.4
$\{(\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^A)_{Z^+}, {}^2\text{W}_{\leftarrow A}^B\}_{CFG}$	5	-7.8	-10.0	-9.1	0.09	0.3



analogous $\{C \leftarrow W_{\leftarrow B}^A, A \leftarrow W^B, W_{\leftarrow A}^C\}_{FG}$ fragments also contain cooperative H-bonding. The total 3B energy of $\{H_3O^+, B \leftarrow {}^1W^A, {}^2W_{\leftarrow A}^B\}_{FG}$ fragments is found to correlate with the distance between the 1W – 2W (H-bonded) donor-H and acceptor-O, such that the 3B energy becomes more negative as the H–O distance decreases (Fig. 4.11). This effect is explained in terms of 3B energy decomposition analysis in Sec. 4.4.

Furthermore, Table 3.10 shows that 3B energies of fragments in the $\{(H_3O^+, B \leftarrow {}^1W^A)_{Z^+}, {}^2W_{\leftarrow A}^B\}_{CFG}$ subgroup are far more attractive than then 3B energies of $\{H_3O^+, B \leftarrow {}^1W^A, {}^2W_{\leftarrow A}^B\}_{CFG}$ fragments such that the minimum energy $\{H_3O^+, B \leftarrow {}^1W^A, {}^2W_{\leftarrow A}^B\}_{CFG}$ fragment has a more repulsive 3B energy than the maximum energy $\{(H_3O^+, B \leftarrow {}^1W^A)_{Z^+}, {}^2W_{\leftarrow A}^B\}_{CFG}$ fragment.

Table 3.11 shows the energy statistics for the most “significant” 4-mer fragment groups. As was indicated in Figs. 3.16–3.18, 4B energies tend to be much smaller than 2B and 3B energies in $H_3O^+(H_2O)_n$ clusters, although ref 75 showed that 4B energies are important for recovering relative binding energies of protonated water clusters. The most common 4-mer fragment group, $\{H_3O^+, C \leftarrow {}^1W^A, {}^1W^B, {}^2W_{\leftarrow A}^C\}_{FG}$, is similar to the the cooperative 3-mer group, $\{H_3O^+, B \leftarrow {}^1W^A, {}^2W_{\leftarrow A}^B\}_{FG}$ (discussed in detail above), but includes an additional 1W which is not H-bonded to any other water molecules. Notice that the addition of this 1W results in anticooperative H-bonding, such that fragments in this group have a mean 4B energy of 0.293 kcal/mol. The 4-mer fragment group with the largest 4B energies is $\{H_3O^+, C \leftarrow {}^1W^A, C \leftarrow {}^1W^B, {}^2W_{\leftarrow AB}^C\}_{FG}$, in which two 1W donate H-bonds to a single 2W , forming a cyclic structure (e.g. 4-mer cluster 4δ). All fragments in this group have positive 4B energies, with a mean value of 0.996 kcal/mol. The corresponding cyclic $OH^-(H_2O)_3$ fragment group, $\{OH^-, {}^1W_{\leftarrow C}^A, {}^1W_{\leftarrow C}^B, AB \leftarrow {}^2W^C\}_{FG}$, also has strictly-positive 4B energies, with a mean value of 0.692 kcal/mol (Table 3.4). Note that the H-bonds in the deprotonated fragments point toward OH^- , away from the 2W , while the H-bonds in the protonated fragments point away from H_3O^+ , toward the 2W . Finally, notice that all fragments in the $\{H_3O^+, {}^1W^A, {}^1W^B, {}^1W^C\}_{FG}$ fragment group have attractive 4B energies, in contrast with the corresponding deprotonated fragment group, $\{OH^-, {}^1W^A, {}^1W^B, {}^1W^C\}_{FG}$, which has strictly-positive 4B energies (Table 3.4).

Table 3.11. Statistics for ‘Significant’ 4-Body Energies of 4-mer Fragments.

Frag. Group	freq.	max.	min.	mean	RSD	sig.
$\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B, {}^2\text{W}_{\leftarrow A}^C\}_{FG}$	26	+0.7	+0.01	+0.3	0.6	0.1
$\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B, {}^1\text{W}^C\}_{FG}$	8	-0.5	-0.5	-0.5	0.03	0.1
$\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B, {}^2\text{W}_{\leftarrow AB}^C\}_{FG}$	4	+1.1	+0.9	+1.0	0.1	0.1
$\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B, {}^3\text{W}_{\leftarrow B}^C\}_{FG}$	3	-0.1	-0.8	-0.3	0.8	0.01
$\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B, {}^2\text{W}_{\leftarrow A}^C\}_{FG}$	2	+0.5	+0.3	+0.4	N/A	0.01
$\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B, {}^1\text{W}^C\}_{FG}$	10	+0.06	-0.2	-0.05	1.0	0.01

3.3.4 Assessment of DFT in Reproducing Many-Body Interactions in Protonated Water Clusters

In this Section we evaluate the accuracy of various XC-functionals in reproducing K -body interaction energies of K -mer fragments of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters relative to SAMBA reference data. The error statistics are organized in terms of the fragment groups introduced in Section 3.3.3 so that conclusions may be drawn regarding accuracy with respect to the types of interactions inherent to a given fragment group. As discussed above, 2B and 3B terms dominate the MBE of protonated water clusters, so we limit the discussion on DFT accuracy to significant 2-mer and 3-mer fragments, but we report the complete set of error statistics in the Supporting Information of Ref. 66.

Fig. 3.22 shows the errors of the tested XC-functionals in reproducing the 2B energies of the H-bonded $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{FG}$ and $\{{}^1\text{W}^A, {}^1\text{W}_{\leftarrow A}^B\}_{FG}$ fragments. $\omega\text{B97M-V}$ and B3LYP perform best for $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{FG}$, with both functionals predicting maximum errors of less than 0.5 kcal/mol in magnitude. Since B3LYP predicts 2B energies that are more attractive than CCSD(T) , the D3 correction has a deleterious effect, with B3LYP-D3 having a mean error of -1.1 kcal/mol for this fragment group. Furthermore, most XC-functionals underestimate 2B energies of $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{FG}$, with BLYP , revPBE(-D3) , and revPBE0 being the only functionals to be associated with positive mean errors.

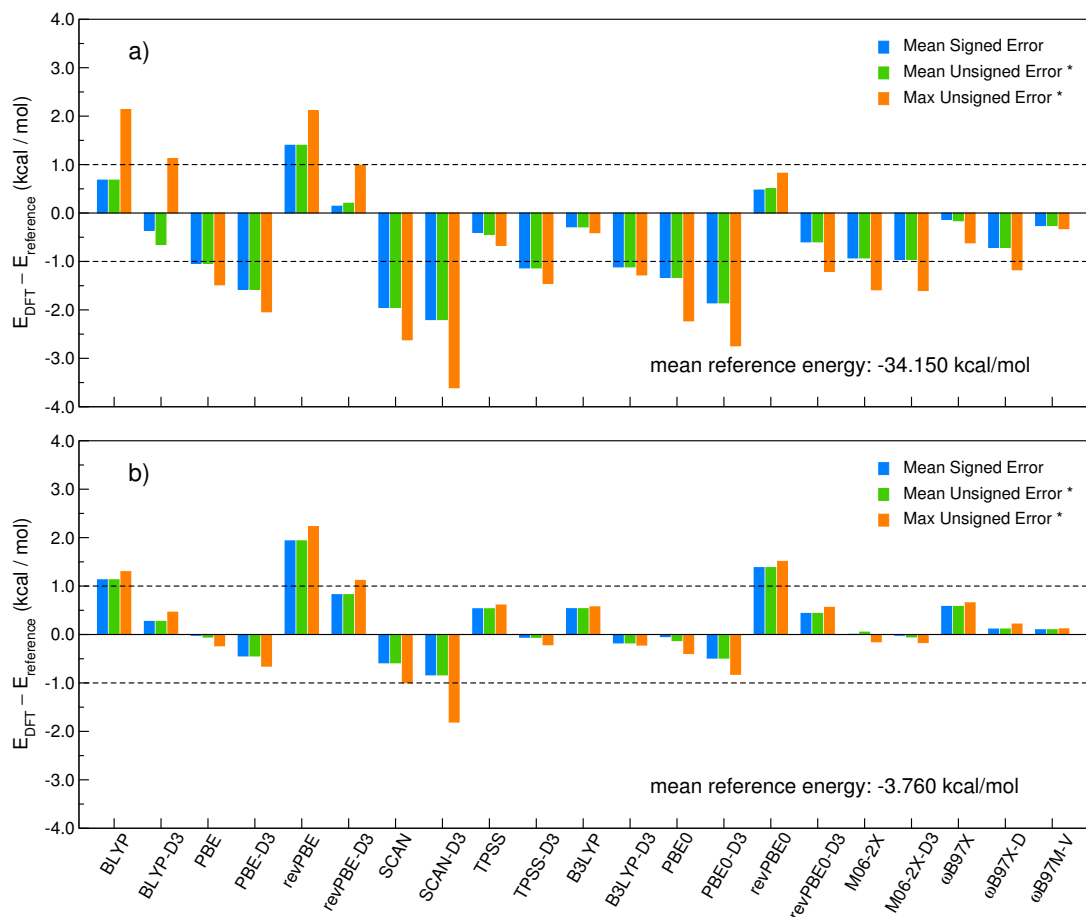


Figure 3.22. DFT errors for the H-bonded 2-mer fragment groups: a) $\{\text{H}_3\text{O}^+, {}^1\text{W}^{\text{A}}\}_{\text{FG}}$ and b) $\{\text{B} \leftarrow \text{W}^{\text{A}}, \text{W}^{\text{B}}_{\leftarrow \text{A}}\}_{\text{FG}}$. *Mean unsigned errors are arbitrarily given the same sign as the mean signed errors to ease comparison between the two, and the maximum unsigned error is arbitrarily given the sign of the corresponding error value.

A notable feature of the data shown in Fig. 3.22a is the discrepancy between maximum unsigned errors and mean errors for BLYP, revPBE, SCAN, PBE0, and M06-2X. Unsurprisingly, this reflects imbalances in the potential energy surfaces calculated by each of these XC-functionals across configuration space. In particular, Fig. 3.23 shows that these XC-functionals display different degrees of accuracy for the the different contextual fragment groups comprising the $\{\text{H}_3\text{O}^+, {}^1\text{W}^{\text{A}}\}_{\text{FG}}$ group. For example, while BLYP 2B energies remain well within chemical accuracy for both $\{\text{H}_3\text{O}^+, {}^1\text{W}^{\text{A}}\}_{\text{CFG}}$ and $\{(\text{H}_3\text{O}^+)_{\text{CZ}^+}, {}^1\text{W}^{\text{A}}\}_{\text{CFG}}$ fragments, a mean error of +2.1 kcal/mol is associated with the corresponding 2B energies calculated for Zundel ion-containing

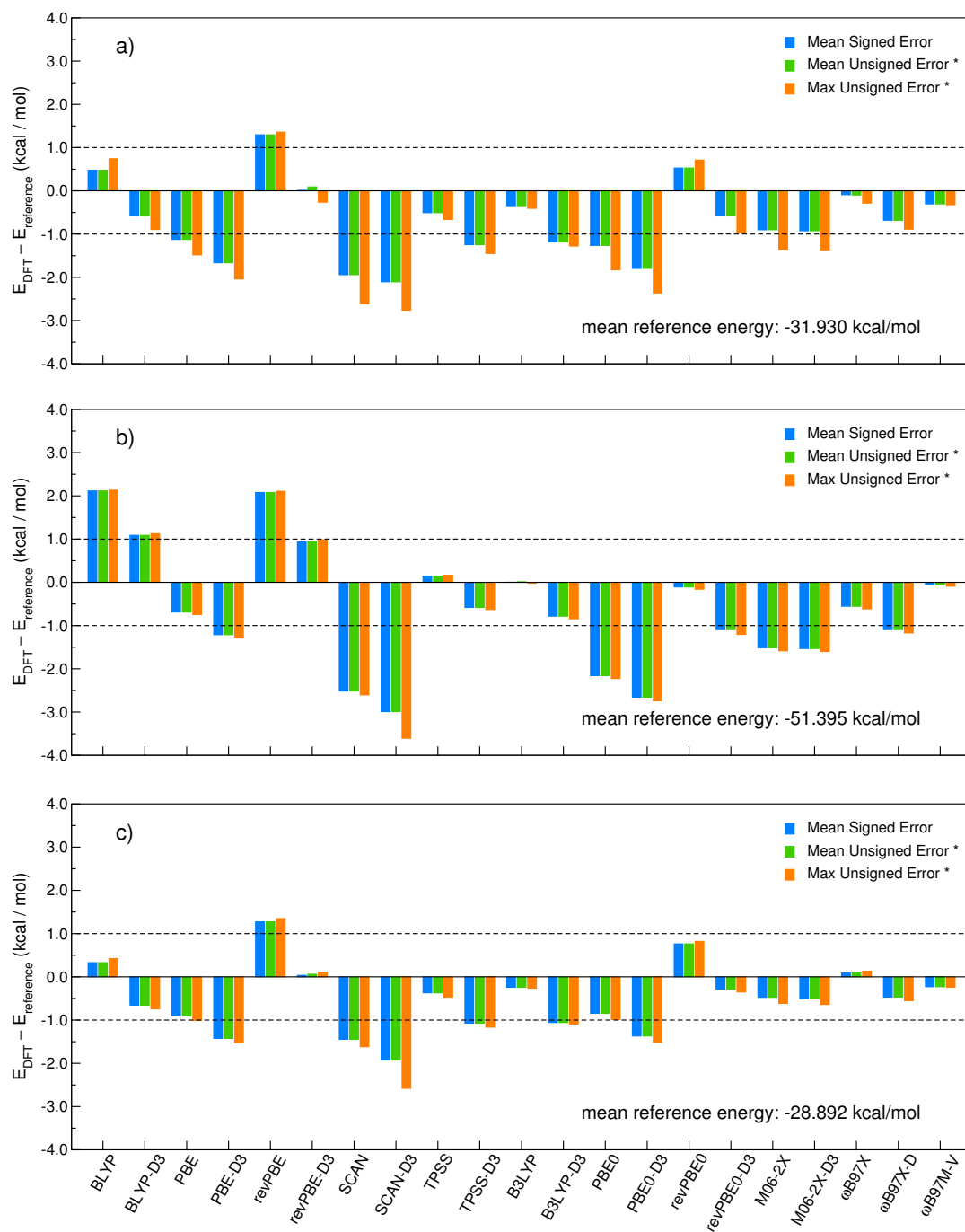


Figure 3.23. DFT errors for the three contexts of the H-bonded $\{H_3O^+, {}^1W^A\}_{FG}$ fragment group: a) $\{H_3O^+, {}^1W^A\}_{CFG}$; b) $\{(H_3O^+, {}^1W^A)_{Z^+}\}_{CFG}$; c) $\{(H_3O^+)_{CZ^+}, {}^1W^A\}_{CFG}$.

fragments, $\{(\text{H}_3\text{O}^+, {}^1\text{W}^A)_{\text{Z}^+}\}_{\text{CFG}}$. PBE0, like BLYP, has difficulties in reproducing the 2B energy of $\{(\text{H}_3\text{O}^+, {}^1\text{W}^A)_{\text{Z}^+}\}_{\text{CFG}}$ fragments, being associated with a mean error of -2.2 kcal/mol, despite displaying mean errors of approximately -1.0 kcal/mol for the other two contextual fragment groups. These discrepancies are particularly troublesome given that proton transfer events are thought to involve interconversion between Eigen-type and Zundel-type structures,[61] suggesting that the potential energy surface imbalances of BLYP and revPBE may artificially hinder proton transfer, while imbalances in the PBE0, SCAN and M06-2X potentials may artificially drive proton transfer.

The error trends of the H-bonded water-dimer group ($\{_{B\leftarrow} \text{W}^A, \text{W}_{\leftarrow A}^B\}_{\text{FG}}$) are markedly different than those of its protonated counterpart. $\omega\text{B97M-V}$ is very accurate for these fragments, as it is for fragments in the $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{\text{FG}}$ group, but high accuracy is also demonstrated by M06-2X(-D3), TPSS-D3, PBE, $\omega\text{B97X-D}$, and PBE0. Interestingly, BLYP, revPBE, and revPBE0 perform worse on average for $\{_{B\leftarrow} \text{W}^A, \text{W}_{\leftarrow A}^B\}_{\text{FG}}$ fragments than for $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{\text{FG}}$ fragments. TPSS and B3LYP have similar accuracies for both H-bonded fragment groups, but have positive mean errors for the neutral $\{_{B\leftarrow} \text{W}^A, \text{W}_{\leftarrow A}^B\}_{\text{FG}}$ fragments, and negative mean errors for the protonated $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{\text{FG}}$ fragments.

Fig. 3.24 shows 2B errors of the non-H-bonded 2-mer fragments. All functionals perform very well for $\{\text{H}_3\text{O}^+, \text{W}^A\}_{\text{FG}}$ fragments, each having mean and maximum errors within 0.5 kcal/mol in magnitude. All functionals therefore show high relative accuracy for this fragment group, which has a mean 2B energy of -8.103 kcal/mol. It should be noted that, due to an unbalanced treatment of the $\text{H}_3\text{O}^+(\text{H}_2\text{O})$ interactions, varying levels of accuracy for $\{\text{H}_3\text{O}^+, \text{W}^A\}_{\text{FG}}$ and $\{\text{H}_3\text{O}^+, {}^1\text{W}^A\}_{\text{FG}}$ fragments would result in a poor representation of the associated molecular dynamics in computer simulations. All of the tested XC-functionals also show small mean errors for 2B energies of the neutral $\{\text{W}^A, \text{W}^B\}_{\text{FG}}$ fragments. However, large maximum errors produced by some functionals indicate poor representations of the water dimer PES, which was shown to depend strongly on the water–water O–O distance.[4]

Fig. 3.25 shows the errors of the tested XC-functionals in reproducing SAMBA 3B

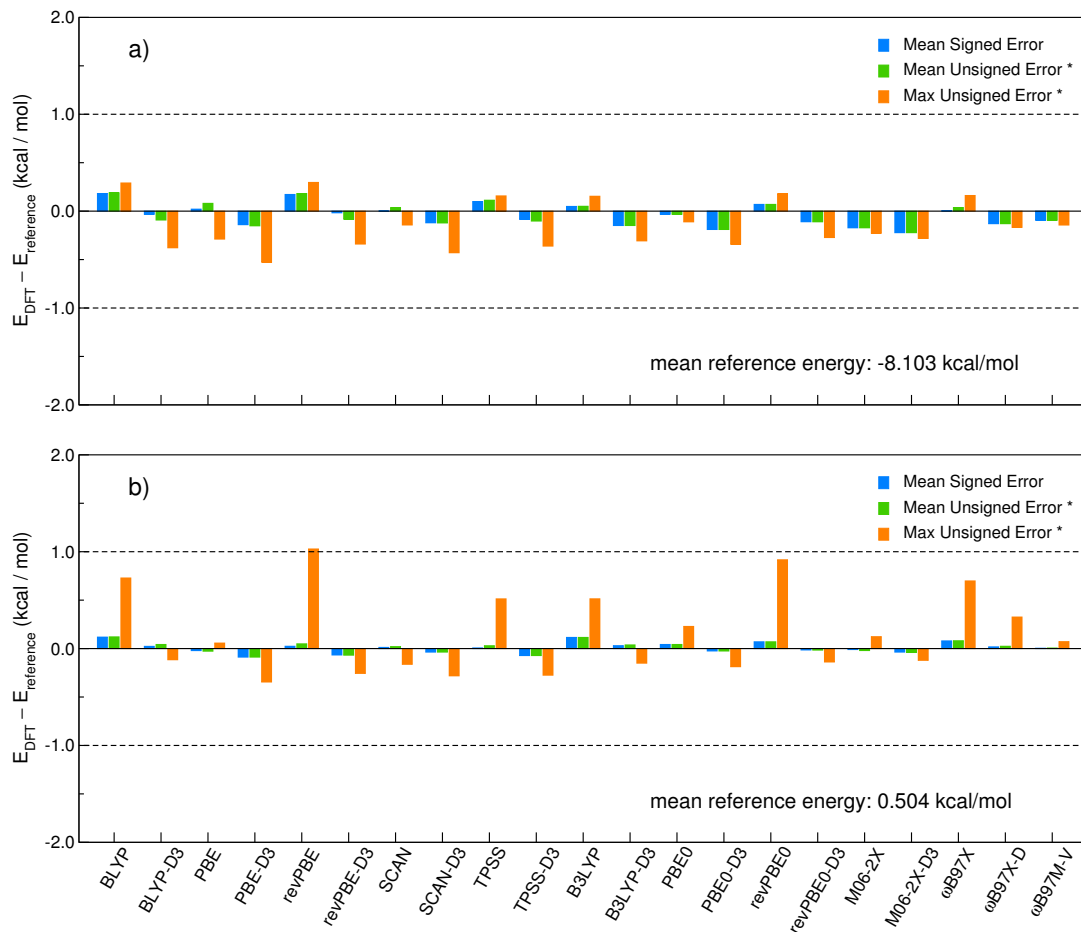


Figure 3.24. DFT errors for the non-H-bonded 2-mer fragment groups: (a) $\{\text{H}_3\text{O}^+, \text{W}^{\text{A}}\}_{\text{FG}}$ and (b) $\{\text{W}^{\text{A}}, \text{W}^{\text{B}}\}_{\text{FG}}$.

energies for fragments from the two most significant $\text{H}_3\text{O}^+(\text{H}_2\text{O})_2$ groups: the anticooperative $\{\text{H}_3\text{O}^+, {}^1\text{W}^{\text{A}}, {}^1\text{W}^{\text{B}}\}_{\text{FG}}$, and the cooperative $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^{\text{A}}, {}^2\text{W}^{\text{B}}_{\leftarrow\text{A}}\}_{\text{FG}}$. Interestingly, all functionals, on average, exaggerate the anticooperativity of the $\{\text{H}_3\text{O}^+, {}^1\text{W}^{\text{A}}, {}^1\text{W}^{\text{B}}\}_{\text{FG}}$ fragments, while also exaggerating the cooperativity of the $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^{\text{A}}, {}^2\text{W}^{\text{B}}_{\leftarrow\text{A}}\}_{\text{FG}}$ fragments. The hybrid functionals tend to have smaller errors than the pure functionals, suggesting that nonlocal effects introduced by the inclusion of Hartree-Fock exchange are needed to accurately reproduce the relevant 3B interactions. A potential explanation for this result based on the delocalization error[12, 13, 14] is discussed in Sec. 4.4.

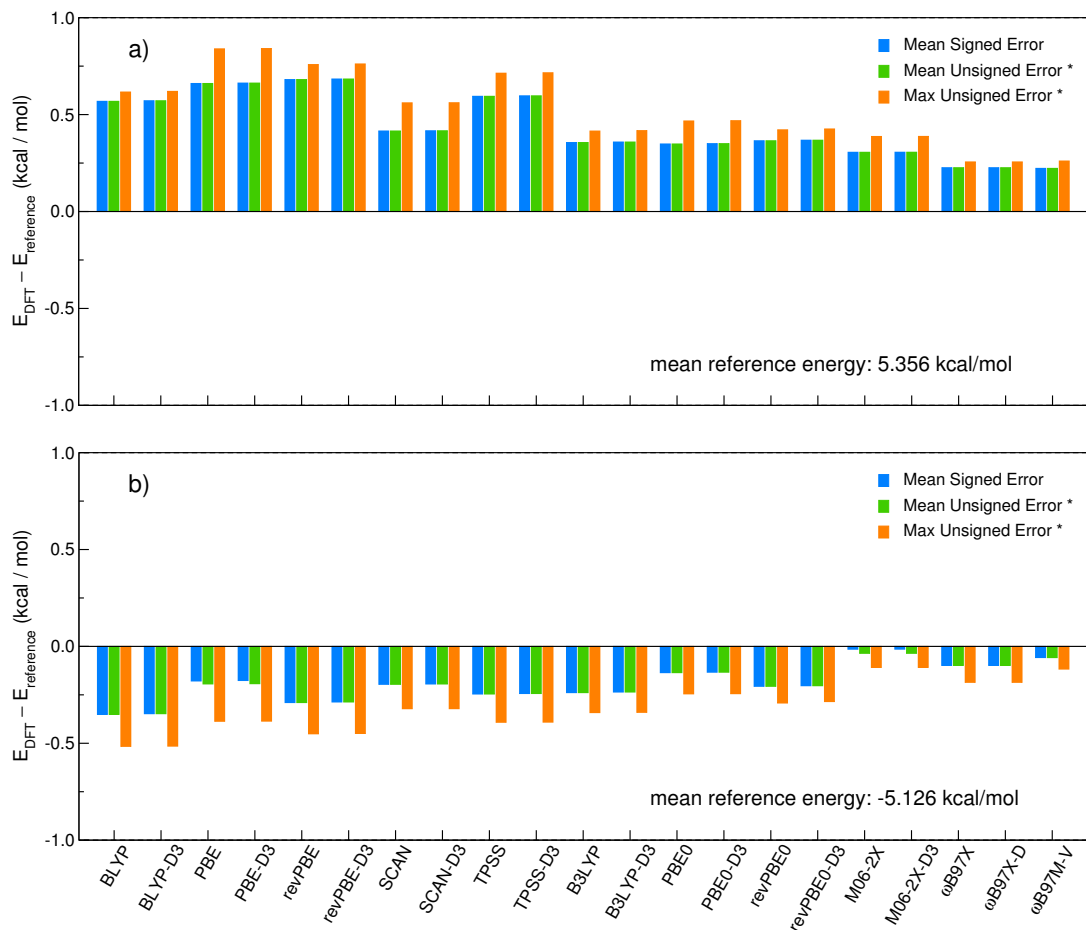


Figure 3.25. DFT errors for the two most significant 3-mer fragment groups: a) the anticooperative $\{\text{H}_3\text{O}^+, {}^1\text{W}^{\text{A}}, {}^1\text{W}^{\text{B}}\}_{\text{FG}}$ and b) the cooperative $\{\text{H}_3\text{O}^+, {}^1\text{W}^{\text{A}}, {}^2\text{W}^{\text{B}}\}_{\text{FG}}$.

3.4 Many-Body Interactions in Alkali Ion–Water Clusters

3.4.1 2B Interaction Energies

Fig. 3.26 shows the deviations of the total 2B interaction energies calculated with each tested model relative to CCSD(T)/CBS for each of the five $\text{M}^+(\text{H}_2\text{O})$ dimers in their corresponding minimum-energy geometries obtained from MP2 optimizations (Section A.3). For reference, a dashed line, indicating the ± 1 kcal/mol threshold that is generally used to define “chemical accuracy”, is also shown in each panel of Fig. 3.26.

Some general trends are apparent when comparing the performance of a given method across different $\text{M}^+(\text{H}_2\text{O})$ dimers. Notably, all non-dispersion-corrected XC-functionals except

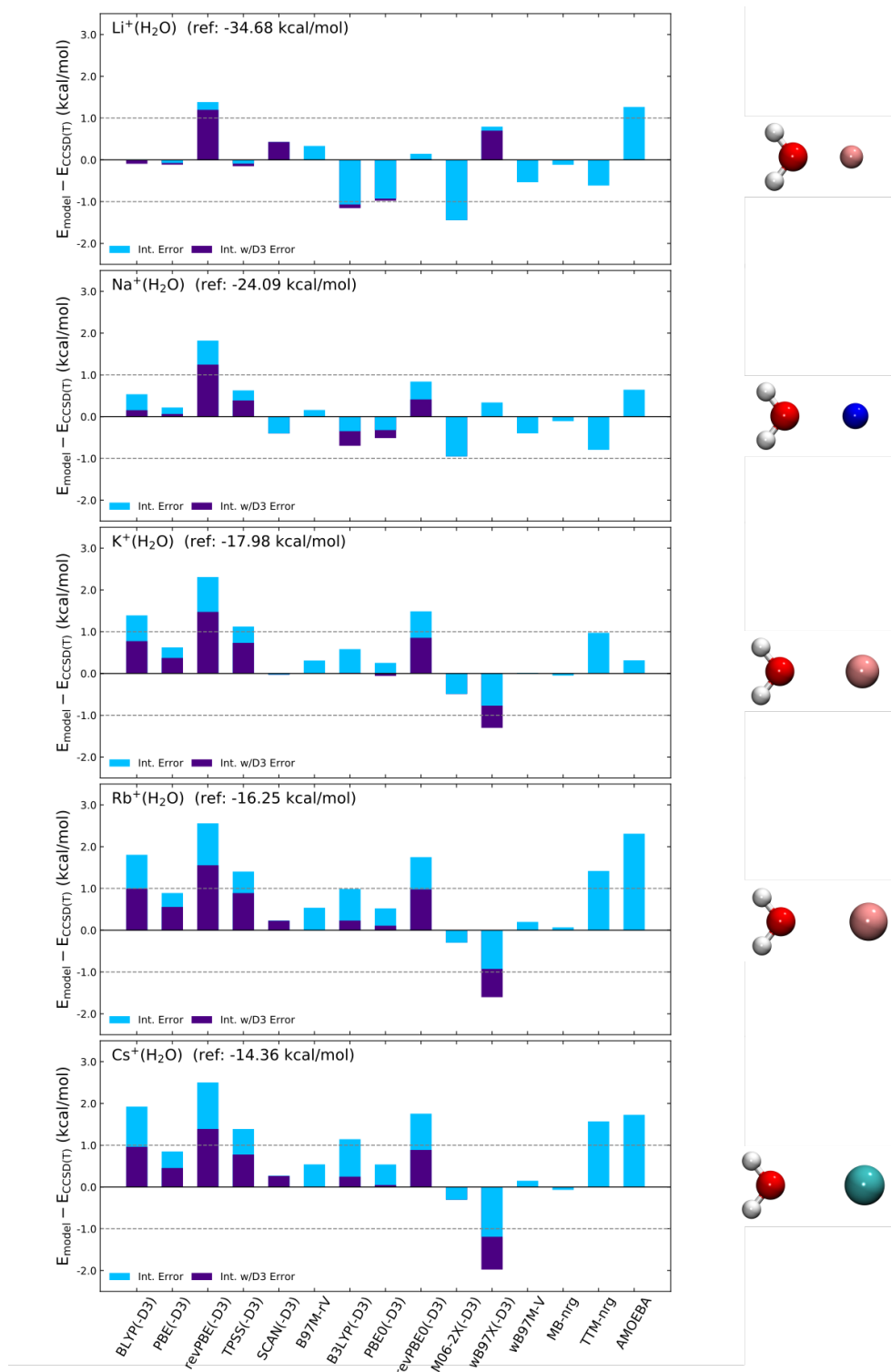


Figure 3.26. Deviations from the CCSD(T) 2B interaction energies of the $M^+(H_2O)$ dimers with $M = Li$ (a), Na (b), K (c), Rb (d), Cs (e), in the corresponding minimum energy configurations (from MP2 as described in Section A.3) calculated with various XC-functionals as well as the AMOEBA, TTM-nrg, and MB-nrg PEFs.

for ω B97X predict 2B energies that tend to become less attractive and/or more repulsive than CCSD(T) with increasing ion size from $\text{Li}^+(\text{H}_2\text{O})$ to $\text{Rb}^+(\text{H}_2\text{O})$, while ω B97X 2B errors become less positive and more negative with increasing ion size for all dimers. Following this trend, XC-functionals such as B3LYP and PBE0 have negative errors for $\text{Li}^+(\text{H}_2\text{O})$ but positive errors for $\text{Cs}^+(\text{H}_2\text{O})$. Since many of the XC-functionals have positive errors for the dimers of the larger ions, the inclusion of D3 corrections appears to substantially improve the 2B energies (notably in B3LYP-D3), although XC-functionals such as BLYP-D3, revPBE-D3, TPSS-D3, and revPBE0-D3 still show errors of ~ 1 kcal/mol. ω B97X has large negative errors for these dimers, so the D3 corrections significantly worsen the 2B energies. ω B97X-D (which, for the reasons mentioned in Section A.3, is not shown in Fig. 3.26) predicts more accurate 2B energies in $\text{K}^+(\text{H}_2\text{O})$ (-17.4 kcal/mol), and in $\text{Rb}^+(\text{H}_2\text{O})$ (-15.7 kcal/mol) than ω B97X-D3, although these energies differ from the CCSD(T) values by more than 0.5 kcal/mol each. The effect of D3(BJ) dispersion corrections[118] on 2B energies of XC-functionals is reported in the Supporting Information of Ref. 119. Since the D3(BJ) corrections are larger in magnitude than D3(0) corrections in all dimer calculations except for ω B97X-D3 for $\text{K}^+(\text{H}_2\text{O})$, $\text{Rb}^+(\text{H}_2\text{O})$, and $\text{Cs}^+(\text{H}_2\text{O})$, D3(BJ) corrections are beneficial when a functional predicts 2B energies that are too repulsive, but are detrimental when a functional predicts 2B energies that are too attractive.

Among the GGA functionals considered in this study, revPBE without D3 corrections significantly underestimates the interaction strength in all dimers, deviating from the CCSD(T) 2B energies by more than 1 kcal/mol. BLYP also underestimates the interaction strength of all dimers other than $\text{Li}^+(\text{H}_2\text{O})$, for which it deviates from CCSD(T) by 0.01 kcal/mol. BLYP has errors greater than 1 kcal/mol for $\text{K}^+(\text{H}_2\text{O})$, and errors just under 2 kcal/mol for $\text{Rb}^+(\text{H}_2\text{O})$, and $\text{Cs}^+(\text{H}_2\text{O})$, but is still more accurate than revPBE for all alkali ion–water dimers. Furthermore, revPBE appears to have, on average, the largest errors among all functionals, deviating from CCSD(T) by over 1 kcal/mol for $\text{Li}^+(\text{H}_2\text{O})$ (for which M06-2X has similar accuracy), and deviating by around 2.5 kcal/mol for $\text{Rb}^+(\text{H}_2\text{O})$ and $\text{Cs}^+(\text{H}_2\text{O})$. PBE performs the best out of the GGA functionals, having errors just below 1 kcal/mol for $\text{Rb}^+(\text{H}_2\text{O})$ and $\text{Cs}^+(\text{H}_2\text{O})$, and

errors below 0.25 kcal/mol for $\text{Li}^+(\text{H}_2\text{O})$ and $\text{Na}^+(\text{H}_2\text{O})$.

Among the meta-GGA functionals considered here, SCAN and B97M-rV perform the best, both deviating from CCSD(T) by at most 0.5 kcal/mol for all dimers. TPSS yields 2B energies that are too repulsive for all dimers other than $\text{Li}^+(\text{H}_2\text{O})$, for which it has a small negative error, leading to significant improvements upon adding the D3 corrections. Interestingly, the accuracy of TPSS-D3 is comparable to that of BLYP-D3 for all minimum energy dimers.

The hybrid functionals have varied performance, with $\omega\text{B97M-V}$ being, on average, the most reliable hybrid functional for all dimers, revPBE0 and $\omega\text{B97X-D3}$ being the least accurate hybrid functionals for dimers of the larger ions, and with M06-2X being the least accurate hybrid functional for $\text{Li}^+(\text{H}_2\text{O})$ and $\text{Na}^+(\text{H}_2\text{O})$. PBE0 tends to benefit from D3 corrections, such that PBE0-D3 deviates from CCSD(T) by around 0.1 kcal/mol for all dimers except for $\text{Li}^+(\text{H}_2\text{O})$ and $\text{Na}^+(\text{H}_2\text{O})$, for which it has errors of approximately -1 and -0.5 kcal/mol, respectively. B3LYP benefits from D3 corrections as well, having errors within 0.25 kcal/mol for all dimers except for $\text{Li}^+(\text{H}_2\text{O})$ and $\text{Na}^+(\text{H}_2\text{O})$, for which its errors are comparable to those for PBE0.

3.4.2 3B Interaction Energies

Fig. 3.27 shows the deviations of the 3B energy calculated with each model relative to CCSD(T)/CBS for each of the five minimum-energy alkali ion–water $\text{M}^+(\text{H}_2\text{O})_2$ trimers. The deviations of each model relative to CCSD(T)/CBS for all stable alkali ion–water $\text{M}^+(\text{H}_2\text{O})_2$ trimers are reported in the Supporting Information of Ref. 119, along with similar comparisons for the other alkali ion–water trimer geometries.

All XC-functionals have deviations within 1 kcal/mol for each trimer. $\text{Li}^+(\text{H}_2\text{O})_2$ has the largest mean absolute 3B errors among the XC-functionals (0.29 kcal/mol), while $\text{Rb}^+(\text{H}_2\text{O})_2$ has the lowest mean absolute error (0.09 kcal/mol). It should be noted that $\text{Li}^+(\text{H}_2\text{O})_2$ also has the largest reference 3B energy of 3.37 kcal/mol, which is a factor of two larger than that of $\text{Na}^+(\text{H}_2\text{O})_2$, the second largest 3B energy. In the analysis of halide–water trimers,[120] it was found that the 3B errors were very uniform between the different systems, such that the

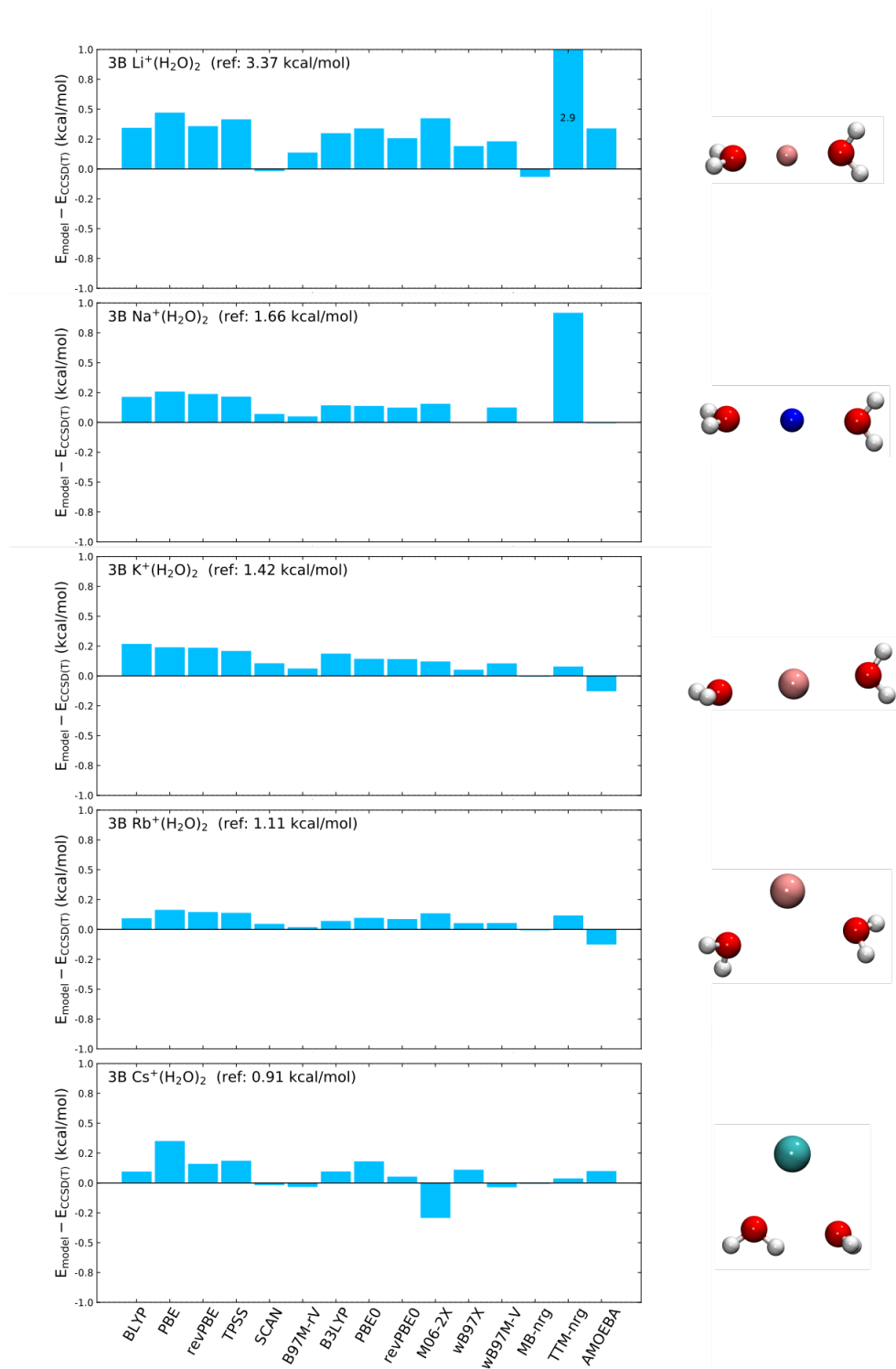


Figure 3.27. Deviations from the CCSD(T) 3B interaction energies of the $M^+(H_2O)_2$ trimers with $M = Li$ (a), Na (b), K (c), Rb (d), Cs (e), in the corresponding minimum energy configurations (from MP2 as described in Section A.3) calculated with various XC-functionals as well as the AMOEBA, TTM-nrg, and MB-nrg PEFs.

$\text{F}^-(\text{H}_2\text{O})_2$ and $\text{Cl}^-(\text{H}_2\text{O})_2$ minimum energy isomers have the largest and smallest mean absolute errors of 0.27 and 0.23 kcal/mol, respectively.

Nearly all XC-functionals predict 3B energies that are slightly more repulsive than the CCSD(T) reference values for all trimers, with the exception being M06-2X for $\text{Cs}^+(\text{H}_2\text{O})_2$. All other negative errors have magnitudes less than 0.05 kcal/mol. It was also found that most of the tested XC-functionals predict halide–water 3B energies that are more repulsive than the CCSD(T)-F12b reference values for the minimum energy halide–water $\text{X}^-(\text{H}_2\text{O})_2$ trimers, such that all negative errors have magnitudes less than 0.05 kcal/mol.[120] Among the XC-functionals, SCAN, B97M-rV, and ωB97X yield the best 3B energies, averaged over the minimum-energy geometries, each deviating from the CCSD(T) reference 3B values, on average, by no more than 0.1 kcal/mol. PBE has the largest mean absolute error among the functionals (0.30 kcal/mol). For the minimum energy halide–water trimers, revPBE0, $\omega\text{B97X-D}$, and $\omega\text{B97M-V}$ predict the most accurate 3B energies, while PBE predicts the least accurate 3B energies.[120]

3.5 The Scaling of Many-Body Energies in a Lattice

The following self-contained topic was motivated by an interesting observation made in Ref. 121 regarding the many-body energies in the $(\text{H}_2\text{O})_N$ clusters. The authors find that the (total) 4B energy represents around 2 percent of the total binding energies in all of the clusters that they studied, regardless of the cluster size (their data set included clusters with $N = 7, 10, 13, 16$ water molecules), leading them to conclude that the total 4B energy is a constant proportion of the total binding energy in all pure water systems. This result makes intuitive sense given that water is an insulator, so high-order terms in the MBE (such as the 4B energy) shouldn't scale with system size in such a way that they overpower the lower-order terms in increasingly larger systems. In other words, the scaling of the total 4B energy with system size should be bounded relative to the scaling of all other many-body terms. Assuming these results do in fact generalize to larger systems, this intuition is apparently correct. And yet, linear scaling

of the total 4B energy (or any many-body term for that matter) is a bit of a puzzling result due to the combinatorics of many-body systems.

The number of K -mer fragments of a system of N monomers can be computed from the corresponding binomial coefficient, $\binom{N}{K} = \frac{N!}{K!(N-K)!}$. From this relation, one can see that the number of K -mer fragments scales with system size, N , like $\mathcal{O}(N^K)$ as $(N \rightarrow \infty)$. Therefore, the ratios of the number of 4-mer fragments with the numbers of 2-mer and 3-mer fragments scale, respectively, like $\mathcal{O}(N^2)$ and $\mathcal{O}(N)$ as $(N \rightarrow \infty)$, so a simple count of the number of K -mer fragments in a given system is clearly insufficient to rationalize this finding.

Here we introduce a simple combinatorial approach to understanding the scaling of many-body energies with system size. In order to make our calculations (and our arguments) exact, we focus on simple model systems, namely: lattices in 1, 2, and 3 spatial dimensions. We only discuss rectangular lattices of point particles, but the arguments can be generalized.

3.5.1 Lattices in 1 Spatial Dimension: Return of the Fragment Group

We start by discussing lattices in 1 spatial dimension (1D). Although a lattice is infinite by definition, in order to investigate the scaling, we will need to consider finite systems of N monomers so that we can obtain the limiting behavior as the system size approaches infinity. We will refer to these finite lattice approximations as *chains*, and more specifically, a 1D chain of $n = N$ monomers as an n -chain, a 2D chain of m rows and n columns of monomers (with $N = mn$) as an (m, n) -chain, etc.

Fig. 3.28 shows 12 examples of 2-mer fragments in a 12-chain, with the squares representing monomers in the chain (which we will call lattice sites), and shaded squares representing the sites composing a particular 2-mer fragment of the chain. The panel on the left shows four examples of 2-mer fragments in which both monomers are directly adjacent. Assuming that all monomers are point particles, each monomer in the chain (lattice) are equivalent, so each of these 2-mer fragments are identical, and therefore all have exactly the same 2B energy. On a finite n -chain, there will be $n - 1$ 2-mer fragments that have adjacent monomers. The middle

panel shows another set of identical 2-mer fragments, with the two monomers being separated by a single lattice site, and the panel on the right shows a third set of identical 2-mer fragments, each with the two monomers being separated by two lattice sites. Again, due to the symmetry of the chain, the 2B energy of a given fragment is uniquely determined by the number of lattice sites separating the monomers. If we define s to be the number of lattice sites between (and including) the monomers (so the three panels in Fig. 3.28 have $s = 2$, $s = 3$, and $s = 4$), then there will be $n - s + 1$ of these fragments on an n -chain.

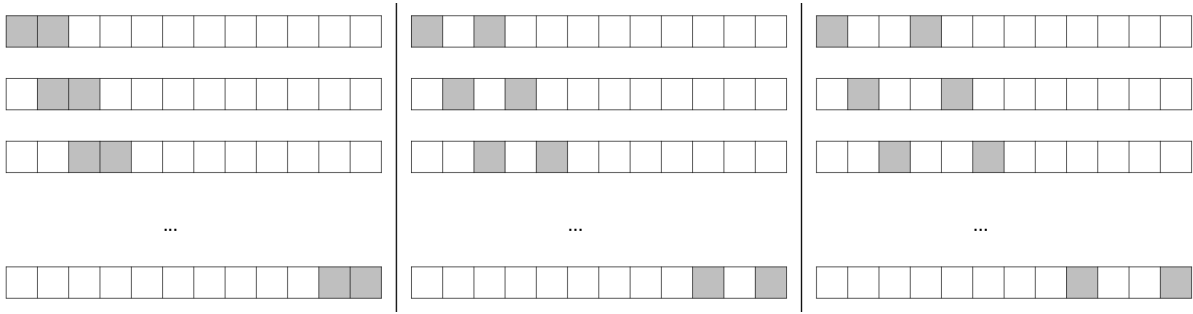


Figure 3.28. Examples of 2-mer fragments in a 12-chain in 1 spatial dimension.

We can find groups of identical K -body fragments in an n -chain for any value of K . These groups will be important, so we shall refer to them as *lattice fragment groups* (LFG). Noticing that LFGs are defined by their connectivity and spacing, there is apparently a connection to the fragment groups used in the analysis of many-body energies in protonated and deprotonated water clusters (Secs. 3.2 and 3.3).²

It turns out that the number of fragments in an LFG for a given n -chain, $g^{(n)}$, (i.e. the order of the group) only depends on the spacing, s , between the left-most and right-most fragments, independently of the many-body order, K . Therefore, if $g_s^{(n)}$ is the order of an LFG with spacing, s , we have:

$$g_s^{(n)} = n - s + 1 \quad (3.16)$$

Notice that the order scales as $\mathcal{O}(n) = \mathcal{O}(N)$ as $(N \rightarrow \infty)$. Therefore, given that the number of

²In the cases of (infinite) lattices or cyclic finite chains, one can easily define a group structure on LFGs, serendipitously giving some justification for our naming.

K -body fragments in an n -chain is $\binom{N}{K}$, the number of possible K -body LFGs on an n -chain scales like $\mathcal{O}(N^{K-1})$ as $(N \rightarrow \infty)$. All possible 2-mer, 3-mer, and 4-mer LFGs with $s = 6$ on an n -chain are shown in Fig. 3.29.

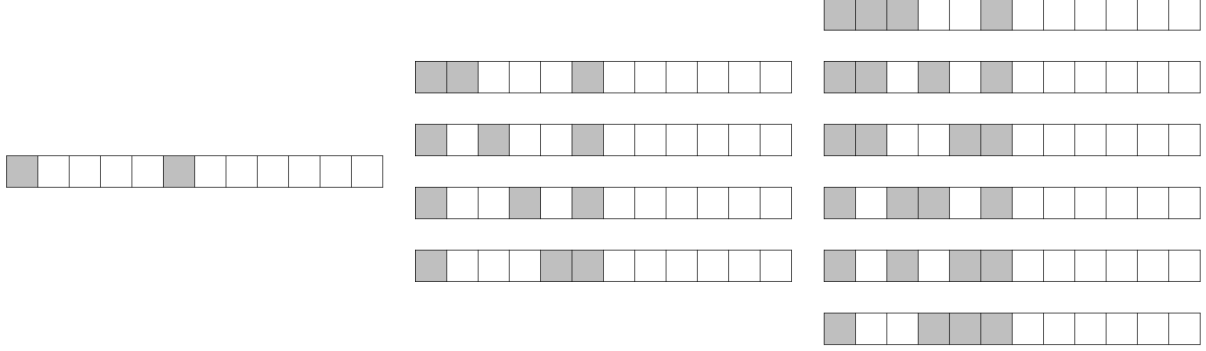


Figure 3.29. All 2-mer, 3-mer, and 4-mer 1D LFGs with $s = 6$.

Why should we care about LFGs? The spacing, s , for LFGs in 1D is meaningful beyond the combinatorics. In the case of 2-mer fragments, s is precisely the distance between the monomers. Additionally, for general K -mer fragments, s bounds the distance between the monomers. Therefore, s is meaningful in terms of the energies of the fragments (most directly in the case of 2-mers). If we define $\epsilon_s^{(K)}$ to be the average K -body energy of all LFGs with a given s in an n -chain, and the corresponding number of distinct K -body LFGs as $\mathcal{N}_s^{(K)}$, then the total K -body energy of the chain is:

$$\begin{aligned}
 E_{\text{KB}} &= \sum_{s=K}^N \mathcal{N}_s^{(K)} g_s^{(n)} \epsilon_s^{(K)} \\
 &= \sum_{s=K}^N \mathcal{N}_s^{(K)} (N - s + 1) \epsilon_s^{(K)}
 \end{aligned} \tag{3.17}$$

3.5.2 Lattices in 2 Spatial Dimensions: Deriving Exact Scaling Relations for Counting Lattice Fragment Groups

The order of 2D LFGs on (m, n) -chains is determined by the dimensions of the smallest rectangular, contiguous subchain that encloses a fragment in the LFG, which we call the *lattice*

convex hull (LCH) of the LFG. Therefore in 2D, we need two parameters, s_x and s_y , to define the spacing in an LFG, while we only need one parameter in 1D. An example of fragments in a (6,6)-chain belonging to an LFG defined by a LCH with $s_x = 4$ and $s_y = 5$ is shown in Fig. 3.30. The order of an LFG on an (m,n) -chain, $g_{s_x,s_y}^{(m,n)}$, is equal to:

$$g_{s_x,s_y}^{(m,n)} = (n - s_x + 1)(m - s_y + 1) \quad (3.18)$$

As was the case with 1D LFGs, the order of 2D LFGs on (m,n) -chains scales linearly with system size, $\mathcal{O}(mn) = \mathcal{O}(N)$ as $(N \rightarrow \infty)$. Furthermore, this finding extends to lattices in any number of spatial dimensions. Given that the order of a given K -body LFG scales linearly with system size, regardless of K , the number of 2D LFGs on an (m,n) -chain must scale like $\mathcal{O}(N^{K-1})$ as $(N \rightarrow \infty)$, as was the case in 1D (and is the case in higher dimensions). This clarifies how it is possible for the total 4B energy in water clusters to represent a constant fraction of the total binding energy, independently of the system size. We can once again define the average energy of all 2D LFGs with LCHs having dimensions s_x and s_y to be $\epsilon_{s_x,s_y}^{(K)}$, and the number of distinct K -body LFGs to be $\mathcal{N}_{s_x,s_y}^{(K)}$, then the total K -body energy, E_{KB} can be calculated in a similar way as in Eq. 3.17. For $E_{KB}/E_{\text{tot}} = c$, constant, we must have $\epsilon_{s_x,s_y}^{(K)}/\epsilon_{s_x,s_y}^{(K')} \rightarrow (s_x s_y)^{K'-K}$ as $s_x s_y \rightarrow \infty$ for all K' , with the special case of $\epsilon_{s_x,s_y}^{(K)} = 0$ when $s_x s_y \geq (s')^2$ for all K , where s' quantifies the finite range of interactions within some system. In this special case defined by finite-range interactions, all many-body terms, E_{KB} , scale linearly with system size ($\mathcal{O}(N)$), because both $\mathcal{N}_{s_x,s_y}^{(K)}$ and $\epsilon_{s_x,s_y}^{(K)}$ are independent of N .

In the interest of completeness, we derive the exact expression for the number of 2D K -body LFGs with s_x and s_y , $\mathcal{N}_{s_x,s_y}^{(K)}$. Derivations of the corresponding expressions in any number of spatial dimensions follow the same reasoning, so we simply present the final results for the 3D case after the derivation in 2D.

The LCH for a given LFG is the smallest contiguous, rectangular subchain that encloses any fragment belonging to the LFG. For the LCH to be large enough to enclose a given fragment,

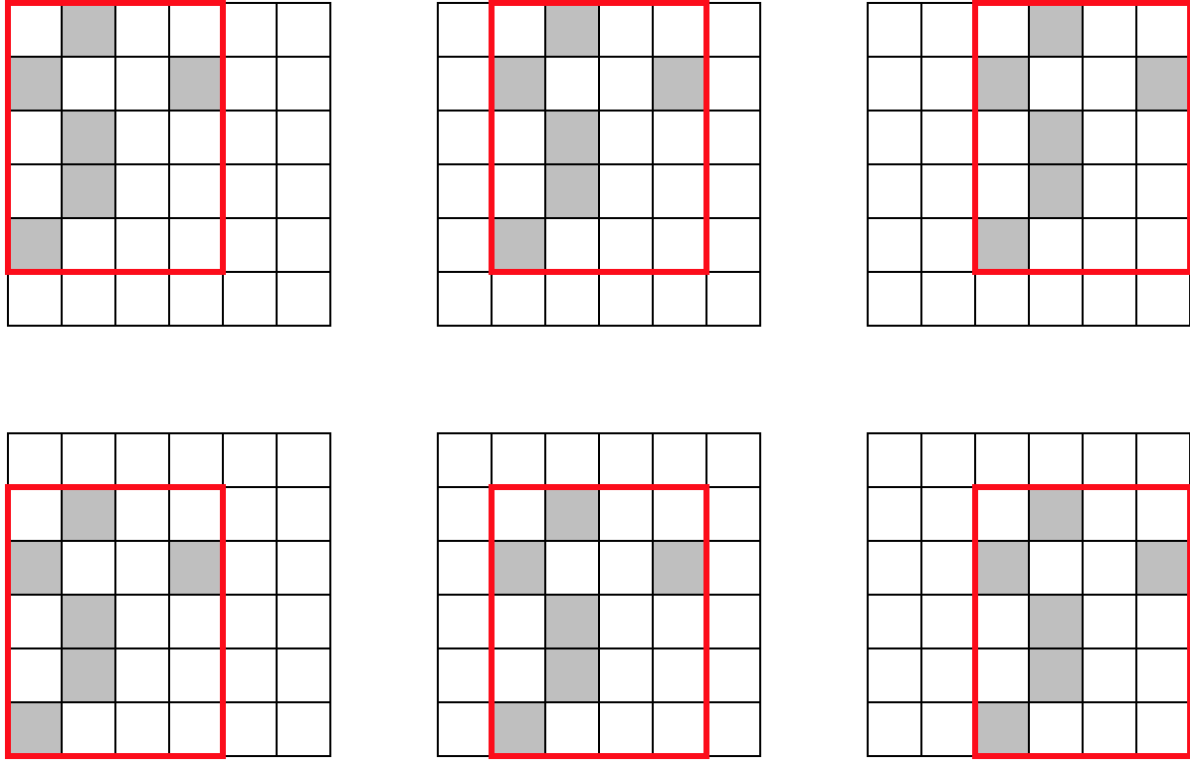


Figure 3.30. Example of the lattice convex hull (the red outline) for a 2D LFG.

s_x and s_y must be equal to the largest distance in the x and y directions, respectively, between two monomers in the fragment. The LCH is a rectangular subchain, so it must enclose $s_x s_y$ lattice sites, so there are $\binom{s_x s_y}{K}$ K -mer fragments that will fit inside the LCH. However, the LCH is the smallest subchain that can enclose a given fragment, so we are over-counting all of the smaller LFGs (fragments) that can be enclosed in smaller LCHs. In fact, if we let s'_x and s'_y be the dimensions of the actual LCH for one of these smaller LFGs, then we would be over-counting by exactly $(s_x - s'_x + 1)(s_y - s'_y + 1)$ fragments for that single LFG (because of Eq. 3.18), so we actually need to be a little careful here.

Let's refer to the LCH defined by s_x and s_y as H (for "hull"). We can begin to correct for the over-counting by considering the number of fragments that will fit within H if we remove its right-most column, yielding a new LCH which we will call R (for "right"). Any LFG that can be enclosed by R (which is defined by dimensions $s_x - 1$ and s_y) was necessarily over-counted

because H would not have been the smallest possible rectangular subchain that could enclose it, so we subtract the number of fragments that fit inside R , $\binom{(s_x-1)s_y}{K}$, from our original count. Additionally, we need to consider the subchain, L , resulting from removing the left-most column from H , because fragments containing monomers on the right-most column of H would not have been included in our count for R . L is defined by dimensions $s_x - 1$ and s_y , as was the case for R , so subtracting the number of K -body fragments enclosed in L from the current count gives us $\binom{s_x s_y}{K} - 2\binom{(s_x-1)s_y}{K}$. In 2D, we need to correct for over-counting in the y -direction as well, so we can define LCHs, U and D , defined by removing the upper- and lower-rows from H , respectively, each of which is defined by dimensions s_x and $s_y - 1$. Applying the same over-counting correction due to U and D that we applied for R and L , we obtain $\binom{s_x s_y}{K} - 2\binom{(s_x-1)s_y}{K} - 2\binom{s_x(s_y-1)}{K}$.

So far so good, but after applying these over-counting corrections, we are now under-counting! There are some number of fragments that are enclosed by two or more of the subchains, R , L , U , and D , so we need to consider the intersections between these four subchains (the sets of lattice sites that are shared by two or more of the subchains) in order to correct our previous correction. In fact, we end up needing to consider all possible intersections between these four subchains to obtain the exact count of the LFGs defined by the LCH, H .

For a given (contiguous, rectangular) subchain, A , defined by dimensions, s_x and s_y , we define the number of enclosed K -body fragments to be $|A|_{KB} = \binom{s_x s_y}{K}$, and for subchains A, B, C , etc., we define the intersection, $A \cap B \cap C \dots$ to be the subchain which contains all lattice sites shared by all of the subchains, A, B, C , etc.. Ultimately, one will obtain the following expression

for the number of K -body LFGs defined by H (with dimensions, s_x and s_y), $\mathcal{N}_{s_x, s_y}^{(K)}$:

$$\begin{aligned}
\mathcal{N}_{s_x, s_y}^{(K)} = & |H|_{KB} - |R|_{KB} - |L|_{KB} - |U|_{KB} - |D|_{KB} \\
& + |R \cap L|_{KB} + |R \cap U|_{KB} + |R \cap D|_{KB} \\
& + |L \cap U|_{KB} + |L \cap D|_{KB} + |U \cap D|_{KB} \\
& - |R \cap L \cap U|_{KB} - |R \cap L \cap D|_{KB} \\
& - |R \cap U \cap D|_{KB} - |L \cap U \cap D|_{KB} \\
& + |R \cap L \cap U \cap D|_{KB}
\end{aligned} \tag{3.19}$$

An even number of intersections corresponds to an over-counting correction (and is negative), and an odd number of intersections corresponds to an under-counting correction. In 1D, we only need to consider R and L subchains, so the analogous expression for $\mathcal{N}_s^{(K)}$ in terms of the intersections of H , R , and L is:

$$\mathcal{N}_s^{(K)} = |H|_{KB} - |R|_{KB} - |L|_{KB} + |R \cap L|_{KB} \tag{3.20}$$

For a LCH in 3D defined by dimensions s_x , s_y , and s_z , we also need to consider the pair of subchains, F and B (for “front” and “back”), which are both defined by dimensions s_x , s_y , and $s_z - 1$, and the corresponding to the expression for $\mathcal{N}_{s_x, s_y, s_z}^{(K)}$ in 3D follows the same pattern as in the 1D and 2D cases.

Rewriting these equations in terms of binomial coefficients involving s_x , s_y , and s_z , and recognizing that the pairs of subchains for each direction (i.e. R and L , U and D , F and B) enclose the same number of fragments, we have (with $s_x = s$ in 1D):

$$\mathcal{N}_{s_x}^{(K)} = \sum_{a=0}^2 (-1)^a \binom{2}{a} \binom{s_x - a}{K} \tag{3.21}$$

$$\mathcal{N}_{s_x, s_y}^{(K)} = \sum_{a=0}^2 \sum_{b=0}^2 (-1)^{a+b} \binom{2}{a} \binom{2}{b} \binom{(s_x-a)(s_y-b)}{K} \quad (3.22)$$

$$\mathcal{N}_{s_x, s_y, s_z}^{(K)} = \sum_{a=0}^2 \sum_{b=0}^2 \sum_{c=0}^2 (-1)^{a+b+c} \binom{2}{a} \binom{2}{b} \binom{2}{c} \binom{(s_x-a)(s_y-b)(s_z-c)}{K} \quad (3.23)$$

It is important to note that these equations can not be used when $s_x = 1$, or $s_y = 1$, or $s_z = 1$, which would involve taking factorials of negative numbers (in the right-most binomial coefficient), so the corresponding lower-dimensional equations must be used instead (e.g. when $s_x = 1$, $s_y \neq 1$, and $s_z \neq 1$ in 3D, one must use the corresponding 2D equation, $\mathcal{N}_{s_y, s_z}^{(K)}$).

Making the substitutions $s_x \rightarrow x$, $s_y \rightarrow y$, and $s_z \rightarrow z$, and expanding the binomial coefficients and simplifying, we obtain many-body scaling polynomials. The first few scaling polynomials are:³

$$\mathcal{N}_x^{(2)} = 1 \quad (3.24a)$$

$$\mathcal{N}_x^{(3)} = x - 2 \quad (3.24b)$$

$$\mathcal{N}_x^{(4)} = \frac{1}{2}x^2 - \frac{5}{2}x + 3 \quad (3.24c)$$

$$\mathcal{N}_x^{(5)} = \frac{1}{6}x^3 - \frac{3}{2}x^2 + \frac{13}{3}x - 4 \quad (3.24d)$$

$$\mathcal{N}_{x,y}^{(2)} = 2 \quad (3.25a)$$

$$\mathcal{N}_{x,y}^{(3)} = 6xy - 6(x+y) + 4 \quad (3.25b)$$

$$\mathcal{N}_{x,y}^{(4)} = 6x^2y^2 - 12(x^2y + xy^2) + 7(x^2 + y^2) + 15xy - 5(x+y) + 1 \quad (3.25c)$$

$$\begin{aligned} \mathcal{N}_{x,y}^{(5)} = & \frac{10}{3}x^3y^3 - 10(x^3y^2 + x^2y^3) + \frac{35}{3}(x^3y + xy^3) \\ & + 18x^2y^2 - 5(x^3 + y^3) - 11(x^2y + xy^2) + x^2 + y^2 + \frac{10}{3}xy \end{aligned} \quad (3.25d)$$

³The expressions in Eqs. 3.21–3.23 were evaluated and simplified with the sympy library.[122]

$$\mathcal{N}_{x,y,z}^{(2)} = 4 \quad (3.26a)$$

$$\mathcal{N}_{x,y,z}^{(3)} = 36xyz - 36(xy + xz + yz) + 36(x + y + z) - 40 \quad (3.26b)$$

$$\begin{aligned} \mathcal{N}_{x,y,z}^{(4)} = & 72x^2y^2z^2 - 144(x^2y^2z + xy^2z^2 + x^2yz^2) \\ & + 84(x^2y^2 + x^2z^2 + y^2z^2) + 288(x^2yz + xy^2z + xyz^2) \\ & - 168(x^2y + x^2z + xy^2 + xz^2 + y^2z + yz^2) - 630xyz \\ & + 98(x^2 + y^2 + z^2) + 390(xy + xz + yz) - 250(x + y + z) + 172 \end{aligned} \quad (3.26c)$$

An example application for these scaling relations is in estimating the total error of a model for many-body interactions in conducting systems in which high-order terms in the MBE are important. After obtaining the average K -body energy as a function of the LCH dimensions, $\mathcal{E}_{s_x, s_y, s_z}^{(K)}$ (in 3D), for a given system (from stochastic sampling, for example), the scaling polynomials allow one to translate the error per fragment (for some model parameterization) into the total error using the analog to Eq. 3.17, which can be used in the optimization cost function when fitting the model.

3.6 Acknowledgements

Sections 3.2, and 3.3, are almost entirely reprints of the material as they appear in “Assessing Many-Body Effects of Water Self-Ions. I: $\text{OH}^-(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2018**, 14 (4), 1982–1997, and “Assessing Many-Body Effects of Water Self-Ions. II: $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2019**, 15 (9), 4816–4833.

Section 3.4 is almost entirely a reprint of the material as they appear in “Nature of Alkali–Water Interactions: Insights from Many-Body Representations and Density Functional Theory II” *J. Chem. Theory Comput.*, **2020**, 16 (5), 3055–3072.

Chapter 4

Energy Decomposition Analysis

4.1 Background

4.1.1 ALMO-EDA

Given a base density functional, $\mathcal{E}^{base}[\cdot]$, ALMO-EDA decomposes the interaction energy calculated with that functional, V_{INT} , into a sum of frozen, POL, and CT energies, V_{FRZ} , V_{POL} , and V_{CT} , respectively:[123]

$$V_{INT} = V_{FRZ} + V_{POL} + V_{CT} \quad (4.1)$$

The V_{FRZ} term can be further decomposed into a sum of ELEC, PAULI, and DISP energies, V_{ELEC} , V_{DISP} , and V_{PAULI} , respectively:[123, 124]

$$V_{FRZ} = V_{ELEC} + V_{DISP} + V_{PAULI} \quad (4.2)$$

Since the ALMO-EDA method is discussed in detail in the original references,[125, 123, 124, 126] we only summarize the main aspects of the method here.

The three terms on the right-hand side of Eq. 4.1 are each interaction energy components for the total system, with each term corresponding to its own total-system electron density, differing from each other in their respective degrees of relaxation (minimization of the total electronic energy). The least-relaxed (highest energy) electronic density, the frozen density, $\rho_{FRZ}(\mathbf{r})$, results from the antisymmetric product of the isolated monomer occupied orbitals

(fully-relaxed in isolation, calculated with the base functional), superimposed into the complex geometry. $\rho_{FRZ}(\mathbf{r})$ is “frozen” in the sense that monomer densities are unchanged relative to their optimized, isolated densities (apart from the antisymmetrization). The frozen energy, V_{FRZ} is taken to be the difference from subtracting the sum of the isolated monomer total energies, E_{1B}^{tot} , from the total energy of $\rho_{FRZ}(\mathbf{r})$ calculated with the base functional.

$$V_{FRZ} = \mathcal{E}^{base}[\rho_{FRZ}(\mathbf{r})] - E_{1B}^{tot} = E_{FRZ}^{tot} - E_{1B}^{tot} \quad (4.3)$$

The decomposition of the frozen energy (Eq. 4.2) is described in detail in refs 123 and 124, but we note here that the calculation of the dispersion energy, V_{DISP} , requires the use of a “dispersion-free” XC-functional, $\mathcal{E}_{DF}^{base}[\cdot]$ specific to the base functional. Ideally, $\mathcal{E}_{DF}^{base}[\cdot]$ would lack any dispersion interaction, but would otherwise be identical to the base functional. Therefore, in the interest of fair comparisons between the functionals, we only consider the sum, $V_{Pauli+Disp} = V_{PAULI} + V_{DISP}$, in the analyses presented in Secs. 4.2 and 4.3. We also only consider the ALMO-EDA 2 “classical” ELEC energy[123, 124] in order to make meaningful comparisons between the decompositions of XC-functionals with those obtained with the PEFs. We also make use of the kinetic energy pressure (KEP) in Sec. 4.3.

The POL energy, V_{POL} , is calculated from the total-system electron density obtained via relaxation with respect to the base functional, such that each monomer polarizes each other monomer, but no electron density is transferred between the monomers. This is accomplished by only allowing monomer electron densities to relax within their own “polarization subspaces. In ALMO-EDA 2, these subspaces are determined with the use of fragment electric field response functions.[126, 123, 127] V_{POL} is then computed by subtracting E_{1B}^{tot} and V_{FRZ} from the total energy of the polarized density, $\rho_{POL}(\mathbf{r})$ calculated with the base functional:

$$V_{POL} = \mathcal{E}^{base}[\rho_{POL}(\mathbf{r})] - E_{1B}^{tot} - V_{FRZ} \quad (4.4)$$

The CT energy, V_{CT} , is calculated from the fully-relaxed total-system electron density, $\rho_{SCF}(\mathbf{r})$, (i.e. what one would obtain from an unconstrained DFT calculation with the base functional). This electron density can be thought of as resulting from mixing each monomer's (polarized) occupied orbitals from $\rho_{POL}(\mathbf{r})$ with the occupied and virtual orbitals of the surrounding monomers (as well as its own virtual orbitals).[126, 123, 127] The total interaction energy, V_{INT} , is calculated by subtracting E_{1B}^{tot} from the total energy of $\rho_{SCF}(\mathbf{r})$ with respect to the base functional. The CT energy, V_{CT} , is finally calculated by subtracting V_{FRZ} and V_{POL} from V_{INT} ,

$$V_{INT} = \min_{\rho_{SCF}(\mathbf{r})} \mathcal{E}^{base}[\rho_{SCF}(\mathbf{r})] - E_{1B} = E_{SCF}^{tot} - E_{1B}^{tot} \quad (4.5)$$

$$V_{CT} = V_{INT} - V_{FRZ} - V_{POL} \quad (4.6)$$

4.2 Energy Decompositions of Halide–Water Clusters

To provide insights into the ability of DFT models to describe halide–water interactions, the ALMO-EDA method is used to decompose the interaction energies of halide–water $X^-(H_2O)$ dimers ($X = F, Cl, Br, I$), calculated by various XC-functionals, into individual contributions defined in Section 4.1.1. For reference, the structures of the $X^-(H_2O)$ dimers as well as the corresponding deviations of 2B energies calculated with each XC-functional relative to reference 2B energies calculated with CCSD(T)-F12b for each dimer are shown in Fig. 4.1. A more detailed analysis of these deviations can be found in Ref. [120].

Fig. B.1 shows the full ALMO-EDA decompositions for each functional for each dimer (as well as decompositions of some model PEFs, as will be discussed in the next chapter). Additionally, 2B ALMO-EDA statistics for each class of XC-functional are shown in Tables 4.1 and 4.2.

Additionally, ALMO-EDA analysis within families of XC-functionals offers some perspective regarding the construction of a DFT model. For example, PBE and revPBE differ from

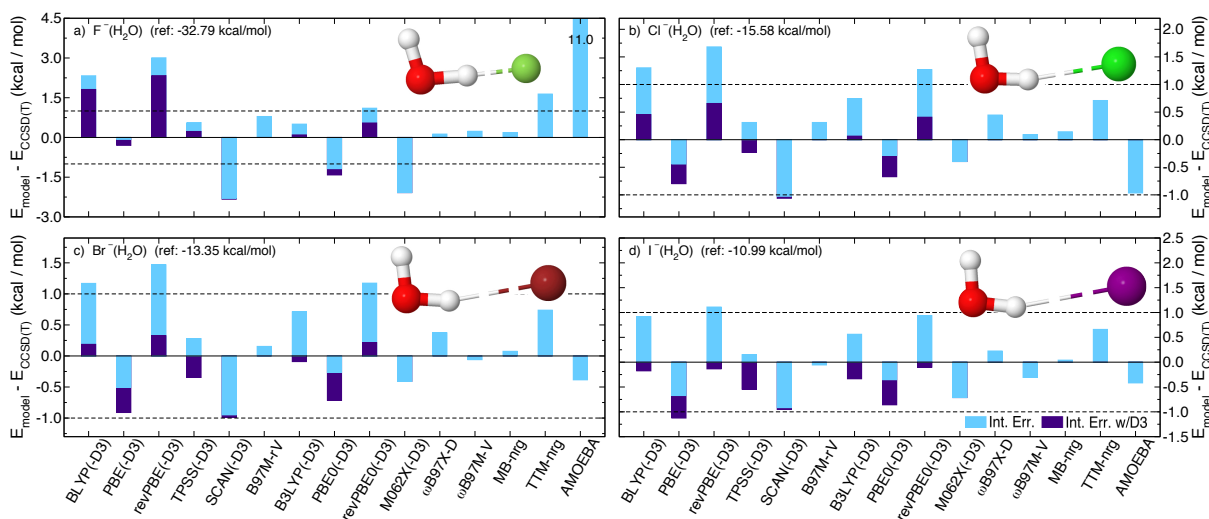


Figure 4.1. Deviations from the CCSD(T)-F12b interaction energies of the $X^-(H_2O)$ dimers, with $X = F$ (a), Cl (b), Br (c), and I (d), in the corresponding RI-MP2/aug-cc-pVTZ(-PP) minimum-energy configurations calculated with various XC-functionals as well as AMOEBa, TTM-nrg, and MB-nrg models.

each other by the value of a single parameter in their exchange functionals (specifically in their exchange enhancement factors),[26] with the revPBE exchange energy being more sensitive to the electron density gradient, and having a lower bound in the large-gradient limit, effectively making revPBE more non-local.[128] Consequently, differences in energy decompositions between PBE and revPBE directly reflect the nature of this single parameter. Moreover, the hybrid functionals, PBE0 and revPBE0 differ from their parent functionals in that their respective GGA exchange term is scaled by 0.75, and Hartree-Fock exchange (scaled by 0.25) is added. Therefore, comparing decompositions of PBE with those of PBE0, and revPBE with revPBE0 will reveal the nature of taking GGA to “GGA0.”

Fig. B.1, shows that all DFT models tend to provide similar values for the ELEC and POL terms for each halide–water dimer, with the largest standard deviation (among all XC-functionals) being 0.3 kcal/mol for both contributions to the interaction energy of the $F^-(H_2O)$ dimer. Consensus among DFT models belonging to all classes suggests that ELEC and POL energies of halide–water interactions are weakly-dependent on non-local effects associated with

Table 4.1. Halide–water 2B ALMO-EDA statistics of $X^-(H_2O)$ dimers ($X = F, Cl$). Signed mean energy values are reported with standard deviations in parentheses (in kcal/mol).

Class	Elec.	Pauli + Disp.	Pol.	CT	KEP	Total 2B
$F^-(H_2O)$						
All	−46.5 (0.3)	48.7 (2.2)	−19.5 (0.3)	−15.3 (1.1)	100.3 (3.0)	−32.6 (1.5)
GGA	−46.6 (0.3)	51.8 (1.4)	−19.4 (0.1)	−16.9 (0.2)	104.7 (1.7)	−31.0 (1.3)
mGGA	−46.2 (0.1)	48.4 (1.3)	−19.5 (0.2)	−15.7 (0.7)	100.0 (1.1)	−33.1 (1.4)
Hybrid	−46.6 (0.3)	47.5 (1.4)	−19.4 (0.3)	−14.5 (0.6)	98.6 (2.0)	−33.1 (1.0)
$Cl^-(H_2O)$						
All	−19.8 (0.2)	14.2 (1.2)	−4.4 (0.1)	−5.4 (0.6)	33.1 (1.0)	−15.3 (0.8)
GGA	−19.9 (0.1)	15.7 (0.9)	−4.4 (0.0)	−6.1 (0.0)	34.5 (0.7)	−14.7 (0.9)
mGGA	−19.6 (0.1)	13.9 (0.7)	−4.4 (0.0)	−5.6 (0.3)	32.8 (0.3)	−15.7 (0.6)
Hybrid	−19.9 (0.1)	13.8 (0.9)	−4.3 (0.1)	−5.0 (0.4)	32.7 (0.8)	−15.4 (0.6)

Table 4.2. Halide–water 2B ALMO-EDA statistics of $X^-(H_2O)$ dimers ($X = Br, I$). Signed mean energy values are reported with standard deviations in parentheses (in kcal/mol).

Class	Elec.	Pauli + Disp.	Pol.	CT	KEP	Total 2B
$Br^-(H_2O)$						
All	−16.8 (0.1)	11.3 (1.1)	−3.4 (0.0)	−4.3 (0.5)	N/A	−13.2 (0.7)
GGA	−16.9 (0.1)	12.5 (0.9)	−3.4 (0.0)	−4.9 (0.0)	N/A	−12.6 (0.9)
mGGA	−16.6 (0.0)	10.9 (0.6)	−3.4 (0.0)	−4.5 (0.2)	N/A	−13.5 (0.6)
Hybrid	−16.8 (0.1)	10.9 (0.8)	−3.3 (0.1)	−3.9 (0.3)	N/A	−13.2 (0.6)
$I^-(H_2O)$						
All	−13.7 (0.1)	8.5 (0.9)	−2.3 (0.0)	−3.5 (0.4)	N/A	−11.0 (0.7)
GGA	−13.8 (0.1)	9.6 (0.8)	−2.3 (0.0)	−4.0 (0.1)	N/A	−10.5 (0.8)
mGGA	−13.6 (0.0)	8.3 (0.5)	−2.3 (0.0)	−3.7 (0.2)	N/A	−11.3 (0.5)
Hybrid	−13.8 (0.1)	8.2 (0.8)	−2.3 (0.0)	−3.1 (0.3)	N/A	−11.0 (0.6)

including the kinetic energy density and/or Hartree-Fock exchange in the XC-functional.

The average values of the charge transfer energy predicted by each class of XC-functionals tend to become less negative from the GGA to the hybrid functionals. This progression can be explained by considering the delocalization error,[14] which is mostly due to the self-interaction

error, which is particularly significant for the GGA functionals. This finding is consistent with earlier work studying intermolecular energy decompositions.[129, 130, 131] All three GGA functionals predict similar values for the CT term, with a standard deviation of at most 0.2 kcal/mol in the case of the $F^-(H_2O)$ dimer. Similarly, the three hybrid-GGA functionals also predict similar CT energies between themselves. It should be noted that Ref. 14 found that meta-GGA functionals have similar delocalization errors as GGAs, so the trends in the CT energies found here may result from a combination of effects.

The largest disagreement among all functionals considered in this study is found for the P+D term, and as with all other contributions to the total interaction energy, the values for this term vary the most for the $F^-(H_2O)$ dimer. Notably, differences between PBE and revPBE almost entirely manifest in their P+D energies, with the revPBE P+D term being at least 1 kcal/mol more repulsive than that of PBE for each minimum energy dimer. BLYP and revPBE differ in their P+D terms by at most 0.4 kcal/mol for all dimers. On average, P+D energies from hybrid functionals are less-repulsive than those from the GGAs, but B3LYP and revPBE0 P+D terms are consistently the most repulsive among the hybrids (note that B3LYP uses BLYP exchange scaled by a factor of 0.72[33]). It is interesting to note that P+D energies calculated with PBE0 and revPBE0 are closer in magnitude than those calculated with PBE and revPBE by a factor of ~ 0.75 , reflecting the scaling of the PBE and revPBE exchange functionals in PBE0 and revPBE0, respectively.

Since, as mentioned above, the disagreement in the CT terms predicted by different classes of XC-functionals is likely due to the delocalization error, it seems reasonable to assume that, among all classes, the hybrid functionals provide the most accurate value for this contribution. Therefore, since both ELEC and POL contributions are of similar magnitude among all XC-functionals, it appears that, within the ALMO-EDA scheme, the P+D term is the most uncertain contribution to the interaction energies of the halide–water dimers. Given the similarity between Eqs. 4.1 and 4.2, it seems likely that the frozen energy (P+D + ELEC) should be correlated with the total interaction energy. To test this hypothesis, Fig. 4.2 shows correlation

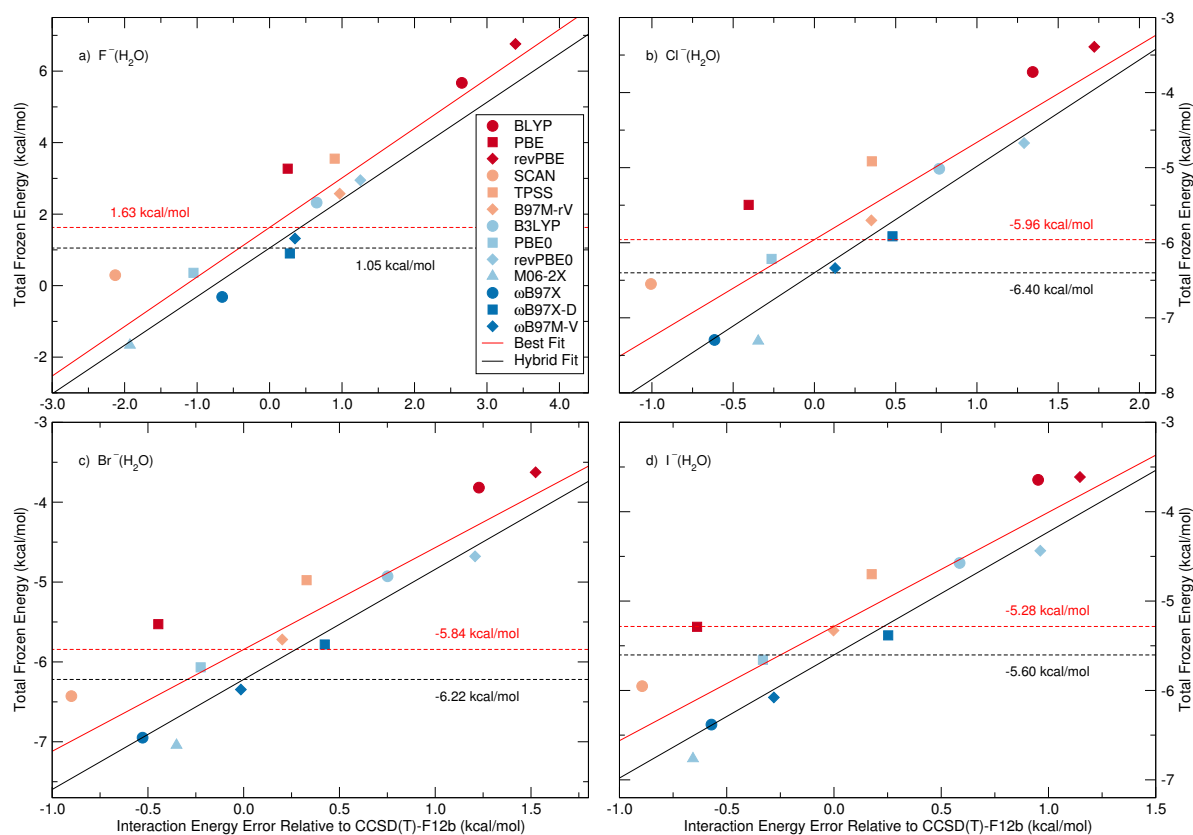


Figure 4.2. Total frozen energy from ALMO-EDA plotted against interaction energy deviations (relative to CCSD(T)-F12b) for $X^-(\text{H}_2\text{O})$ dimers with $X =$ a) F, b) Cl, c) Br, d) I, for each XC-functional, demonstrating the correlation between the two quantities. The best fit lines with respect to all functionals are shown as red solid lines, and the best fit lines with respect to all hybrid functionals are shown as black solid lines. The y-intercepts for each best fit line are also shown, giving an estimate of the “most accurate” total frozen energies for each dimer.

plots between the total frozen energies predicted by the different functionals for each $X^-(\text{H}_2\text{O})$ dimer and the corresponding deviations from the CCSD(T)-F12b total interaction energies. A noticeable correlation indeed exists between these two quantities among all functionals, which becomes even more apparent if only the hybrid functionals are considered. From linear fits to the data shown in Fig. 4.2, it is thus possible to derive an estimate for the “most accurate” total frozen energies associated with each halide–water dimer.

4.3 Energy Decompositions of Alkali Ion–Water Clusters And Comparisons With Halide–Water Clusters

Decompositions of the 2B energies of the $M^+(H_2O)$ dimers are reported in Fig. B.2. Statistics for 2B ALMO-EDA calculations of alkali ion–water $M^+(H_2O)$ dimers are reported Tables 4.3 and 4.4 to aid in the analysis. Interestingly, some of the trends found in the analysis of energy decompositions of 2B energies of halide–water dimers[120] are not reproduced in the analogous decompositions of the 2B energies of the alkali ion–water dimers. Differences in trends between the halide–water and alkali ion–water decompositions could be due to differences in the descriptions of the ion, in particular because anionic species are more difficult to describe with semi-local functionals than neutral or cationic species.[132] Differences between halide–water and alkali ion–water decompositions could also be due to differences in the variations among the functionals in their descriptions of the water hydrogens and the water oxygen, since halide ions accept hydrogen bonds from the water molecule while alkali ions are bound to the water oxygen.

One notable difference between halide–water and alkali ion–water energy decompositions is that the $Li^+(H_2O)$ ELEC energy has a larger standard deviation (0.7 kcal/mol) than any other interaction term among XC-functionals, while the ELEC energy is very consistent among all XC-functionals in all halide–water dimers, with $F^-(H_2O)$ having the largest standard deviation of 0.3 kcal/mol (relative to all functionals from all classes of XC-functionals).[120] The mean ELEC energies (and their standard deviations) in $Li^+(H_2O)$ for GGA, meta-GGA, and hybrid functionals are found to be -31.4 (0.2), -32.0 (0.4), and -32.8 (0.3) kcal/mol, respectively. Relative uniformity within XC-functional classes, but large deviations between the classes, suggests that the use of the kinetic energy density in meta-GGAs and Hartree-Fock exchange in hybrid functionals might be necessary for describing the ELEC energy in $Li^+(H_2O)$. It should be noted, however, that the GGA functionals BLYP and PBE yield more accurate total 2B energies for the $Li^+(H_2O)$ dimer, relative to the CCSD(T) reference, than a number of hybrid

functionals, such as M06-2X, B3LYP, PBE0, and ω B97X, which indicates that the accurate reproduction of the $\text{Li}^+(\text{H}_2\text{O})$ ELEC term is more complicated than simply adding Hartree-Fock exchange.

Table 4.3. Alkali ion–water 2B ALMO-EDA statistics of $\text{M}^+(\text{H}_2\text{O})$ dimers ($\text{M} = \text{Li}, \text{Na}$). Signed mean energy values are reported with standard deviations in parentheses (in kcal/mol).

Class	Elec.	Pauli + Disp.	Pol.	CT	KEP	Total 2B
$\text{Li}^+(\text{H}_2\text{O})$						
All	−32.3 (0.7)	12.2 (0.5)	−10.8 (0.3)	−3.8 (0.2)	17.0 (0.6)	−34.8 (0.8)
GGA	−31.4 (0.2)	12.4 (0.4)	−11.1 (0.1)	−4.1 (0.1)	17.8 (0.4)	−34.2 (0.7)
mGGA	−32.0 (0.4)	12.0 (0.3)	−10.8 (0.2)	−3.8 (0.1)	16.3 (0.4)	−34.5 (0.2)
Hybrid	−32.8 (0.3)	12.1 (0.5)	−10.7 (0.2)	−3.8 (0.1)	17.0 (0.4)	−35.2 (0.8)
$\text{Na}^+(\text{H}_2\text{O})$						
All	−24.9 (0.4)	8.0 (0.5)	−6.4 (0.2)	−0.6 (0.0)	14.5 (0.6)	−23.9 (0.7)
GGA	−24.5 (0.2)	8.6 (0.5)	−6.7 (0.1)	−0.7 (0.0)	15.3 (0.3)	−23.2 (0.7)
mGGA	−24.5 (0.2)	7.7 (0.4)	−6.4 (0.1)	−0.6 (0.0)	13.9 (0.4)	−24.0 (0.4)
Hybrid	−25.2 (0.2)	8.0 (0.4)	−6.3 (0.1)	−0.6 (0.0)	14.4 (0.4)	−24.2 (0.6)

It is likely that the large deviation of the $\text{Li}^+(\text{H}_2\text{O})$ ELEC energy trend from the trends seen in halide–water dimers is at least in part due to the fact that Li^+ has no isoelectronic counterpart in the halide ions. However, we point out that Figs. B.2 and B.1 and Tables 4.3, 4.4, 4.1, and 4.2 show that ELEC energies of the other alkali ion–water dimers have, on average, more variance than the halide–water ELEC energies.[120] Fig. B.2 shows that ELEC energies calculated with GGA and meta-GGA functionals in the other alkali ion–water dimers tend to be very similar, with the two classes having mean values within 0.1 kcal/mol of each other. However, the ELEC energies of all alkali ion–water dimers calculated with the hybrid functionals are, on average, more attractive than those calculated with the semi-local functionals, although differences between the values calculated with hybrid functionals and semi-local functionals become smaller with increasing ion size. Since the ELEC energy is calculated within the ALMO-EDA frozen energy term, differences between functionals reflect differences in their descriptions of the

Table 4.4. Alkali ion–water 2B ALMO-EDA statistics of $M^+(H_2O)$ dimers ($M = K, Rb, Cs$). Signed mean energy values are reported with standard deviations in parentheses (in kcal/mol).

Class	Elec.	Pauli + Disp.	Pol.	CT	KEP	Total 2B
$K^+(H_2O)$						
All	−19.3 (0.3)	7.6 (0.8)	−4.0 (0.1)	−1.7 (0.2)	N/A	−17.4 (0.8)
GGA	−19.1 (0.1)	8.5 (0.5)	−4.0 (0.1)	−1.9 (0.0)	N/A	−16.5 (0.7)
mGGA	−19.0 (0.1)	7.2 (0.6)	−4.0 (0.0)	−1.7 (0.1)	N/A	−17.5 (0.5)
Hybrid	−19.6 (0.1)	7.3 (0.6)	−4.0 (0.1)	−1.5 (0.1)	N/A	−17.8 (0.7)
$Rb^+(H_2O)$						
All	−17.6 (0.3)	7.2 (0.9)	−3.5 (0.1)	−1.6 (0.2)	N/A	−15.4 (0.9)
GGA	−17.4 (0.1)	8.2 (0.6)	−3.5 (0.1)	−1.8 (0.0)	N/A	−14.5 (0.7)
mGGA	−17.3 (0.1)	6.8 (0.6)	−3.5 (0.0)	−1.6 (0.1)	N/A	−15.5 (0.5)
Hybrid	−17.8 (0.1)	6.9 (0.8)	−3.6 (0.1)	−1.4 (0.1)	N/A	−15.9 (0.9)
$Cs^+(H_2O)$						
All	−15.7 (0.2)	6.9 (1.0)	−3.1 (0.1)	−1.6 (0.2)	N/A	−13.6 (1.0)
GGA	−15.6 (0.1)	7.9 (0.6)	−3.1 (0.0)	−1.8 (0.0)	N/A	−12.6 (0.7)
mGGA	−15.5 (0.1)	6.5 (0.6)	−3.1 (0.0)	−1.6 (0.1)	N/A	−13.6 (0.5)
Hybrid	−15.9 (0.1)	6.5 (0.9)	−3.1 (0.1)	−1.4 (0.1)	N/A	−14.0 (1.0)

electronic structures of the isolated monomers. Fig. 4.3 shows that there is a correlation between the magnitude of the dipole moment of the water molecule in $Li^+(H_2O)$ with the $Li^+(H_2O)$ ALMO-EDA ELEC energy, indicating that the large variance in the ELEC energies in $Li^+(H_2O)$ are mostly due to differences in descriptions of the water monomer between the XC-functionals (rather than differences in descriptions of Li^+). The $Li^+(H_2O)$ ELEC energies are more varied than those of the other dimers due to the small distance between Li^+ and the water oxygen (~ 1.8 Å) compared to the other dimers.

In our analysis of the alkali–water dimers, it was found that ALMO-EDA CT energies calculated with GGA functionals tend to be more attractive than those calculated with meta-GGA functionals, which tend to be more attractive than those calculated with hybrid functionals.[120] The same trends seen in halide–water dimer CT energies can also be seen in the alkali ion–water

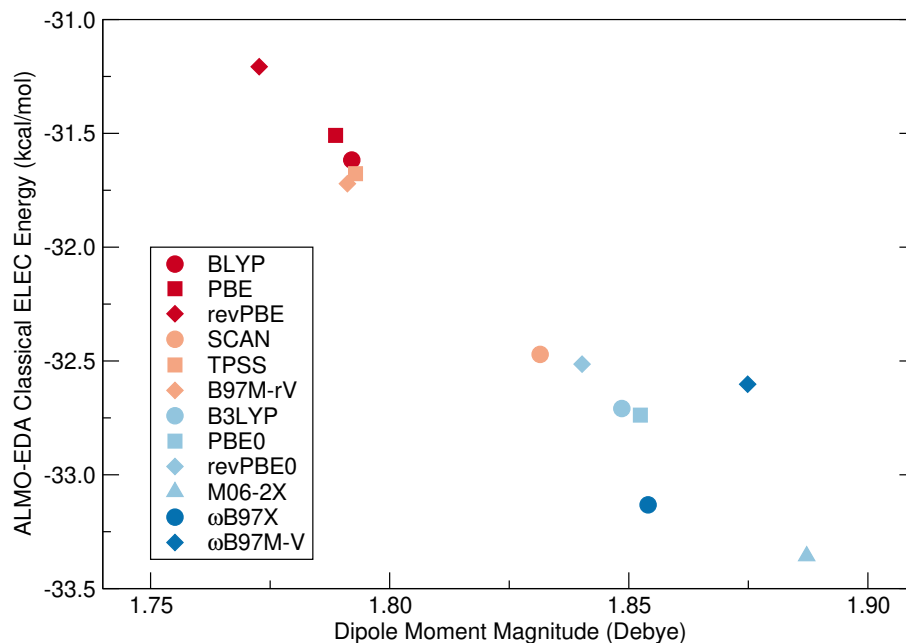


Figure 4.3. Correlation between the dipole magnitude of the water molecule from the minimum energy $\text{Li}^+(\text{H}_2\text{O})$ dimer with the $\text{Li}^+(\text{H}_2\text{O})$ ALMO-EDA ELEC energy (in kcal/mol).

dimers, but to a smaller extent. For example, while the difference in mean CT energies calculated with GGA and hybrid functionals was found to be 2.4 kcal/mol in $\text{F}^-(\text{H}_2\text{O})$, the largest difference in mean CT energies between the GGAs and the hybrid functionals in the alkali ion–water dimers is 0.4 kcal/mol in $\text{Cs}^+(\text{H}_2\text{O})$. Notably, the orderings of the halide–water dimer CT energies produced by the particular XC-functionals tested roughly correlate with the orderings reported in ref 14 regarding the curvature of the electronic energies of the same functionals with respect to fractional electron numbers, a quantity that directly relates to the delocalization error.[14] For example, SCAN yields less attractive ALMO-EDA CT energies than TPSS for all halide–water[120] and alkali ion–water dimers, while it was found that the SCAN electronic energy has less curvature than that of TPSS with respect to fractional electron numbers.[14] Another notable example is M06-2X, which yields less attractive CT energies than the other global hybrid XC-functionals tested for all halide–water[120] and alkali ion–water dimers, except for $\text{Li}^+(\text{H}_2\text{O})$ and $\text{Na}^+(\text{H}_2\text{O})$, in correspondence with its significantly smaller curvature.[14] However, M06-2X also tends to yield less attractive CT energies than the range-

separated hybrid functionals, in contradiction with the trends in curvature,[14] indicating that the delocalization error, is likely not the only contributing factor to errors in ALMO-EDA CT energies. Furthermore, ref 14 indicates that meta-GGA functionals tend to have only marginally smaller curvatures than GGA functionals, suggesting that both classes suffer from comparable delocalization errors, which also seems to contradict the hypothesis that errors in ALMO-EDA CT energies are due to the delocalization error, given that meta-GGA CT energies tend to be less attractive than GGA CT energies. However, the correspondence between the orderings of both quantities (curvatures and ALMO-EDA CT energies) suggest that the two are related, and may both be consequences of the same underlying deficiencies of a given XC-functional. We also point out that monomer properties, alone, are undoubtedly insufficient to describe the trends in CT energies, and that factors such as the difference in the ionization potential of the electron donor and the electron affinity of the electron acceptor are likely significant as well, as is discussed in ref 14.

The Pauli+Disp terms are the most varied among all functionals for all alkali ion–water dimers except for $\text{Li}^+(\text{H}_2\text{O})$. The variance in Pauli+Disp energies increases with increasing ion size, with the standard deviations (among all XC-functionals) for $\text{Li}^+(\text{H}_2\text{O})$ and $\text{Cs}^+(\text{H}_2\text{O})$ being 0.5 and 1.0 kcal/mol, respectively. Pauli+Disp is also the most varied interaction term for all halide–water dimers, although it was found that the standard deviation decreases with increasing ion size, with the standard deviations for $\text{F}^-(\text{H}_2\text{O})$ and $\text{I}^-(\text{H}_2\text{O})$ being 2.2 and 1.0 kcal/mol, respectively.[120] The difference in this variance trend likely reflects the fact that Pauli+Disp energies represent increasing (negative) fractions of the total 2B interaction energy of alkali ion–water dimers with increasing ion size, while representing decreasing (negative) fractions of the total 2B energy of halide–water dimers with increasing ion size.[120]

One notable similarity between energy decompositions of the alkali ion–water dimers and the halide–water dimers is that Pauli+Disp energies predicted by hybrid functionals are, on average, less repulsive than those predicted by GGA functionals.[120] One possible explanation of this trend stems from the inability of GGA functionals to reproduce intermediate range

(attractive) dispersion interactions, which requires a nonlocal treatment of electron correlation (“nonlocal” in the sense of ref 10 as discussed in Sec. 2.3).[133, 28, 134, 135] The fact that the meta-GGA functionals tested here predict Pauli+Disp energies that are comparable to (and sometimes even less repulsive than) those predicted by hybrid functionals supports this hypothesis, since the use of the kinetic energy density adds more nonlocal character to the correlation energy of an XC-functional. In particular, the meta-GGA SCAN functional has been shown to capture intermediate-range dispersion,[29, 28] and the meta-GGA B97M-rV includes the explicitly nonlocal rVV10 correlation functional[134] (note that TPSS yields Pauli+Disp energies that are more repulsive than those from SCAN and B97M-rV for all alkali ion–water dimers except for $\text{Li}^+(\text{H}_2\text{O})$). We also point out that hybrid GGA XC-functionals have been shown to lack dispersion interactions,[28, 135] which appears to be consistent with the fact that B3LYP and revPBE0 tend to yield Pauli+Disp energies that are more repulsive than those from hybrid meta-GGAs (M06-2X, ω B97X, and ω B97M-V), SCAN, and B97M-rV, especially in the dimers of the larger ions. The hybrid GGA PBE0 appears to be an exception to this finding, likely benefiting from error cancellation, resulting from the fact that PBE yields less repulsive Pauli+Disp energies than BLYP and revPBE due to the behavior of its exchange energy with respect to the electronic density gradient.[120] Note that these trends in the Pauli+Disp energies for GGAs, meta-GGAs, hybrid GGAs, and hybrid meta-GGAs can also be seen in the halide–water dimer EDA in ref 120.

A second possible explanation for the trend in the differences between GGA and hybrid Pauli+Disp energies is based on the tendency of semi-local functionals to artificially delocalize electron density while Hartree-Fock exchange localizes the electron density.[13, 14]. Excessive delocalization of the isolated monomer electron densities may result in spurious Pauli repulsion in the frozen-electronic state (Section 4.1.1) of dimers at small intermonomer separations, which would lead to the observed trends. Furthermore, the fact that differences between GGA and hybrid Pauli+Disp energies are more pronounced in the halide–water dimers[120] strengthens this argument since anionic species tend to suffer from artificial delocalization more than neutral

or cationic species.[132] The dominant repulsive interaction contributing to the ALMO-EDA Pauli repulsion energy is the kinetic energy pressure (KEP),[123] which can be thought of as the (primarily kinetic) energetic penalty paid to orthogonalize isolated monomer electron wavefunctions when brought into the complex geometry in the ALMO-EDA frozen state, and therefore quantifies the importance of the second explanation. Although the ALMO-EDA KEP is inaccessible in calculations with ECPs, Tables 4.3, 4.4, 4.1, and 4.2 provide evidence that artificial delocalization contributes to the observed differences between Pauli+Disp energies calculated with GGAs compared to hybrid functionals. It should be noted, however, that the analysis presented in ref 14 indicates that non-hybrid meta-GGA functionals suffer from delocalization errors comparable to those seen in GGA functionals, as was discussed above. This suggests that the trends in Pauli+Disp energies produced by the semilocal meta-GGA functionals (in both alkali ion–water and halide–water dimers) are likely not solely due to the delocalization error, and that intermediate-range dispersion interactions also play a role in the observed trends. Finally, we point out that these trends exist in all dimers, despite the accuracy of the GGAs degrading with increasing alkali ion size (Section 3.4.1), indicating that the inclusion of nonlocal terms is necessary for an XC-functional to be able to correctly reproduce Pauli+Disp energies in alkali ion–water dimers.

Polarization appears to be the contribution that behaves most consistently between the halide–water dimers and alkali ion–water dimers, having the smallest deviation among XC-functionals for all ions. The variance in the POL energy among XC-functionals decreases with increasing alkali ion size, likely reflecting the trend in variance of the ELEC energy. The variance in the POL energies was also found to decrease with increasing halide ion size in the halide–water dimers.[120] In particular, it was found in ref 120 that there is a correlation between the total 2B interaction energies and total 2B ALMO-EDA frozen energies of halide–water dimers predicted by various XC-functionals. It was suggested that this correlation might be due to the similarity in the definitions of these two energies (see eqs 4.5 and 4.3).[120] As shown in Fig. 4.4, we find a similar correlation in alkali ion–water dimers as well. Two best-fit lines are

shown for each dimer, one is fitted to all XC-functionals from all classes, and the other is fitted to the hybrid XC-functionals. Based on each of these best-fit lines, estimates for the total frozen energies (the y-intercepts of each line, where the interaction energy error is 0.0) are also shown in horizontal dotted lines. As was the case for halide–water dimers, the correlation apparently improves with increasing ion size. Additionally, the two sets of best-fit lines tend to converge for alkali ion–water and halide–water dimers with increasing ion size.

4.4 Many-Body Energy Decompositions of Protonated and Deprotonated Water Clusters

As discussed in Sec. 3.2.3, the hydrogen bonding connectivity in a given system can result in cooperative many-body interactions in which one or more hydrogen bonds enhance the hydrogen bonding interactions between the surrounding molecules (resulting in negative 3B energies), or anticooperative interactions in which the surrounding hydrogen bonding is weakened (resulting in positive 3B energies). Here we use 3B energy decomposition analysis to illuminate the physical effects that result in cooperativity and anticooperativity. 1st generation ALMO EDA calculations[136, 137] were performed on the anticooperative $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$, and cooperative $\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, {}^2\text{W}_{A\leftarrow}^B\}_{FG}$ fragments (Sec. 3.2.3) using $\omega\text{B97M-V/def2-TZVPPD}$ (refer to Section A.1 for specific details).

The dependence of the total 3B energy as well as the 3B CT and POL decomposition terms on the average $\text{OH}^- - {}^1\text{W O}-\text{O}$ distance is shown in Fig. 4.5. This shows that $E_{POL}^{ALMO,3B}$ and $E_{CT}^{ALMO,3B}$ calculated for $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments tend to be roughly equal and together dominate the 3B interaction energy, effectively showing the same dependence on the $\text{OH}^- - {}^1\text{W O}-\text{O}$ distances.

The ALMO EDA method was devised in such a way that pairwise, directional (monomer $\rightarrow$ monomer) E_{CT}^{ALMO} (ΔE_{CT}^{ALMO}), and CT “electron reorganization” (ΔQ_{CT}), are well-defined, even in clusters with more than two monomers.[136, 137] Fig. 4.6 shows the correlation between

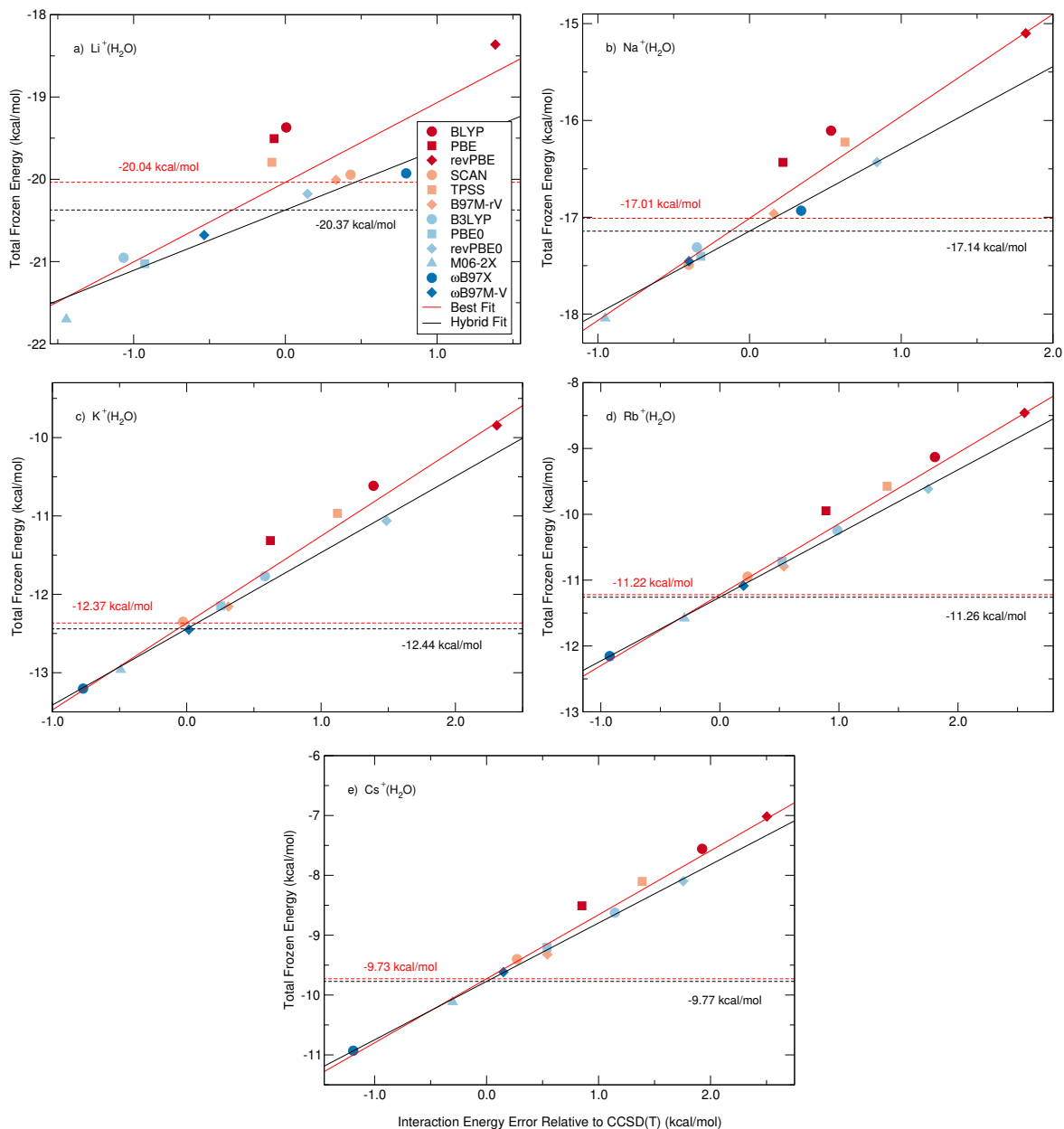


Figure 4.4. Total frozen energy from ALMO-EDA plotted against interaction energy deviations (relative to CCSD(T)) for $M^+(H_2O)$ dimers with $M = Li$ (a), Na (b), K (c), Rb (d), Cs (e), for each XC-functional, demonstrating the correlation between the two quantities. The best fit lines with respect to all functionals are shown as red solid lines, and the best fit lines with respect to all hybrid functionals are shown as black solid lines. The y-intercepts for each best fit line are also shown, giving an estimate of the “most accurate” total frozen energies for each dimer.

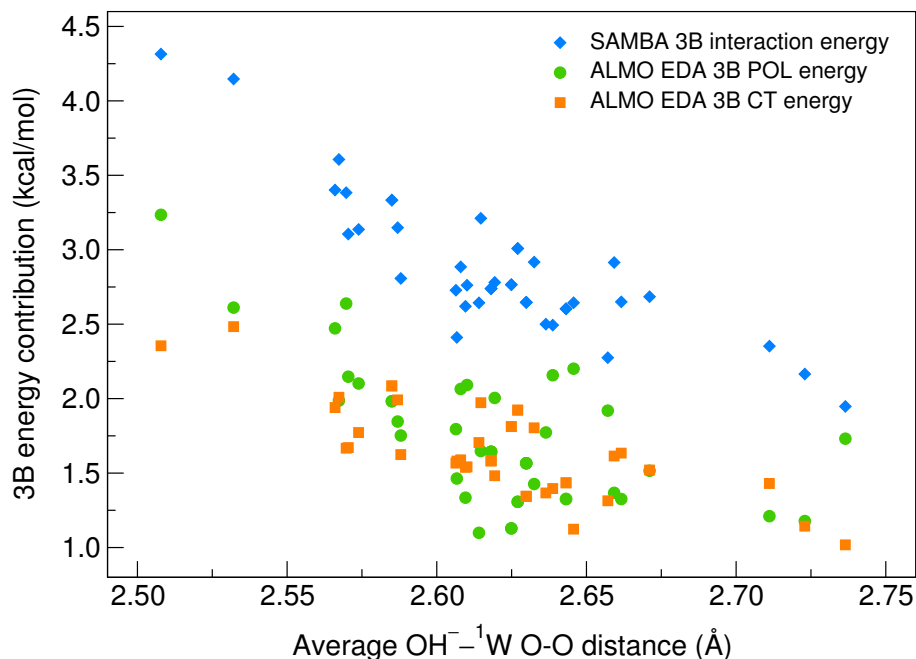


Figure 4.5. Analysis of SAMBA 3B interaction energy contributions and ALMO energy decompositions as a function of the average OH^- –water O–O distance in $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments.

the sum of the 3B CT “electron reorganization,” ΔQ_{CT} (in units of millielectrons, $m\bar{e}$), from OH^- to ${}^1\text{W}$ ($\text{OH}^- \rightarrow {}^1\text{W}$ CT) and the average OH^- – ${}^1\text{W}$ O–O distance calculated for the $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments. Each monomer $\rightarrow$ monomer ΔQ_{CT}^{3B} term was calculated by subtracting the value calculated for the relevant 2-mer from the value of the corresponding term calculated for the full 3-mer fragment, which leads to a negative value when less charge is transferred from the CT donor-monomer to the acceptor-monomer in the presence of the third monomer than when the same 2-mer is considered in isolation.[136, 4, 66] Therefore, a negative ΔQ_{CT}^{3B} implies a positive $E_{CT}^{ALMO,3B}$ contribution since any CT necessarily results in a lower total electronic energy within the ALMO EDA framework, wherein CT interactions correspond to the relaxation of the monomer-specific absolutely-localized molecular orbitals into cluster-wide delocalized molecular orbitals.[136] Summed and averaged quantities are used in this analysis because of the equivalence of the ${}^1\text{W}$ molecules in the definition of the fragment group.

Fig. 4.6 shows that all $\text{OH}^- \rightarrow {}^1\text{W}$ ΔQ_{CT}^{3B} contributions associated with fragments in

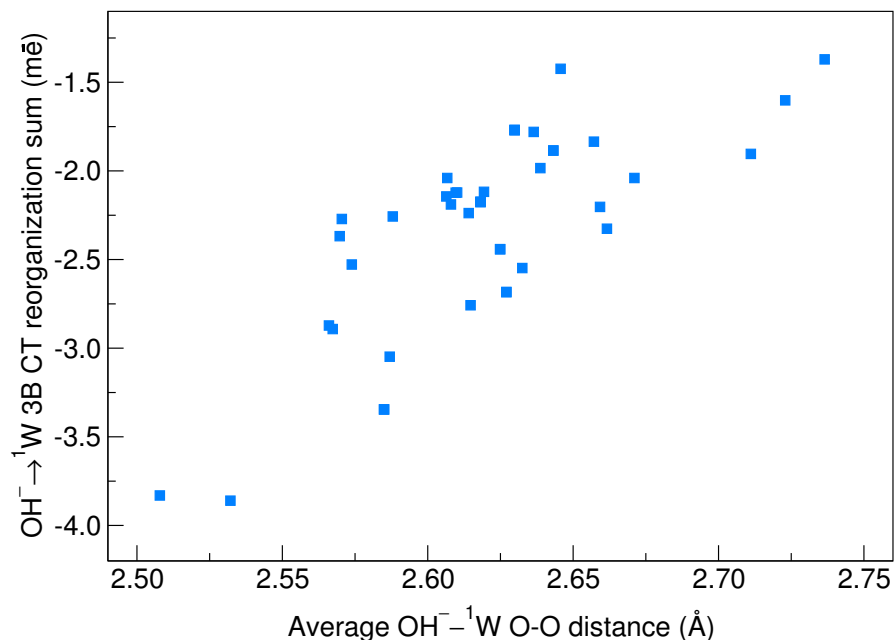


Figure 4.6. Correlation plot between the sum of 3B $\text{OH}^- \rightarrow {}^1\text{W}$ CT “electron reorganization” (in millielectrons, $m\bar{e}$) and average OH^- –water O–O distance in $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments.

$\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ are negative. This is explained by considering the finite extent of electron charge available for hydrogen bond donation from the OH^- lone pairs (see Section 4.2.1 of Ref. 7) which must be shared between both of the ${}^1\text{W}$ molecules in the $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments. This picture is completely consistent with the description of anticooperative hydrogen bonding and, as expected, parallels the occurrence of anticooperativity seen in $\{\text{W}_{\leftarrow BC}^A, A \leftarrow \text{W}^B, A \leftarrow \text{W}^C\}_{FG}$ fragments.[67, 47, 49, 69, 70] The correlation seen in Fig. 4.6 is also similar to the correlation between the total-fragment 3B energy contributions and the average OH^- – ${}^1\text{W}$ O–O distance found in Fig. 4.5 (keeping in mind the inverse relationship between ΔQ_{CT} and E_{CT}^{ALMO} mentioned above). These results are consistent with the SAPT analysis reported in Ref. 65 showing that the interactions in $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments are primarily inductive in nature, since SAPT induction energy includes both POL and CT contributions that are defined separately within the ALMO EDA framework.[137]

The dependence of the total 3B energy as well as the 3B CT and POL decomposition terms on the average OH^- – ${}^1\text{W}$ – ${}^2\text{W}$ O–O–O angle in $\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A \leftarrow {}^2\text{W}^B\}_{FG}$ fragments

is shown in Fig. 4.7. This shows that the total 3B energies in these fragments are mostly due to 3B CT and POL interactions. The total 3B and 3B POL energies are correlated with the $\text{OH}^- - {}^1\text{W} - {}^2\text{W}$ O–O–O angle, becoming more attractive as the angle becomes more linear, while the 3B CT energy shows a smaller correlation with the angle, appearing to be roughly constant for angles greater than 100° . The observed correlation can be explained by the fact that the O–O–O angle is related to the alignment of the ${}^1\text{W}$ and ${}^2\text{W}$ dipole moments.

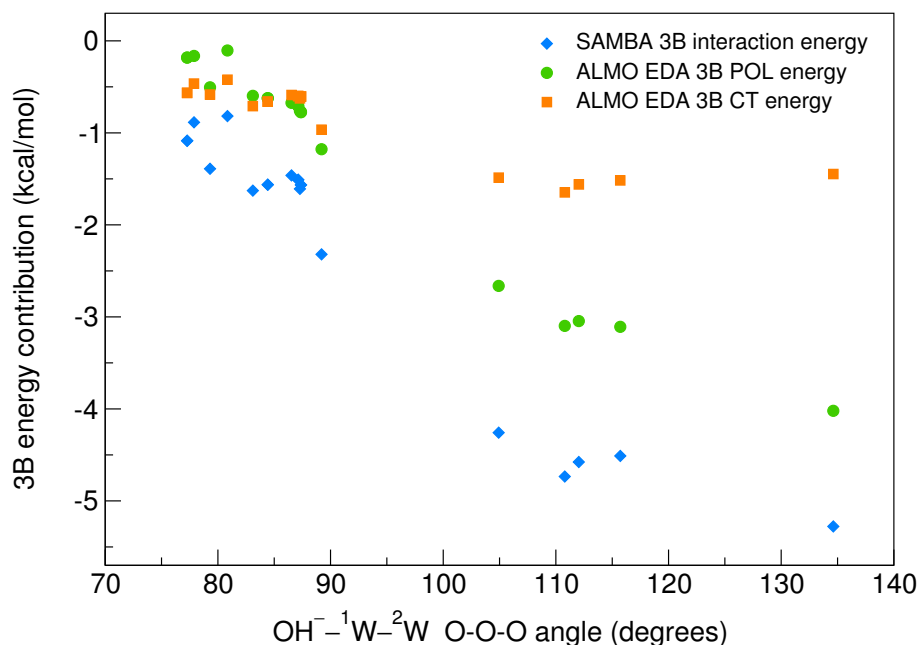


Figure 4.7. Analysis of 3B SAMBA interaction energy contributions and ALMO energy decompositions as a function of $\text{OH}^- - {}^1\text{W} - {}^2\text{W}$ O–O–O angle in $\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A \leftarrow {}^2\text{W}_G^B\}_{FG}$ fragments.

Further insights into the nature of hydrogen bonding in $\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A \leftarrow {}^2\text{W}_G^B\}_G$ are gained from the analysis of ΔQ_{CT}^{3B} shown in Fig. 4.8 as a function of the $\text{OH}^- - {}^1\text{W} - {}^2\text{W}$ O–O–O angle. First, ΔQ_{CT}^{3B} calculated for ${}^1\text{W}_{\leftarrow B}^A \rightarrow A \leftarrow {}^2\text{W}_G^B$ (green symbols) are exclusively positive. This result demonstrates that the presence of OH^- enhances the strength of hydrogen bonding between the two water molecules, which is precisely the definition of cooperative hydrogen bonding. Second, ΔQ_{CT} calculated for $\text{OH}^- \rightarrow A \leftarrow {}^2\text{W}_G^B$ (blue symbols) is diminished by the presence of ${}^1\text{W}_{\leftarrow B}^A$ in each 3-mer, relative to the substituent $\{\text{OH}^-, {}^2\text{W}_G\}_G$ fragment. Finally,

ΔQ_{CT}^{3B} from OH^- to ${}^1\text{W}_{\leftarrow B}^A$ (orange symbols) are exclusively positive, and similar in magnitude to the ${}^1\text{W}_{\leftarrow B}^A \rightarrow A \leftarrow {}^2\text{W}^B \Delta Q_{CT}^{3B}$.

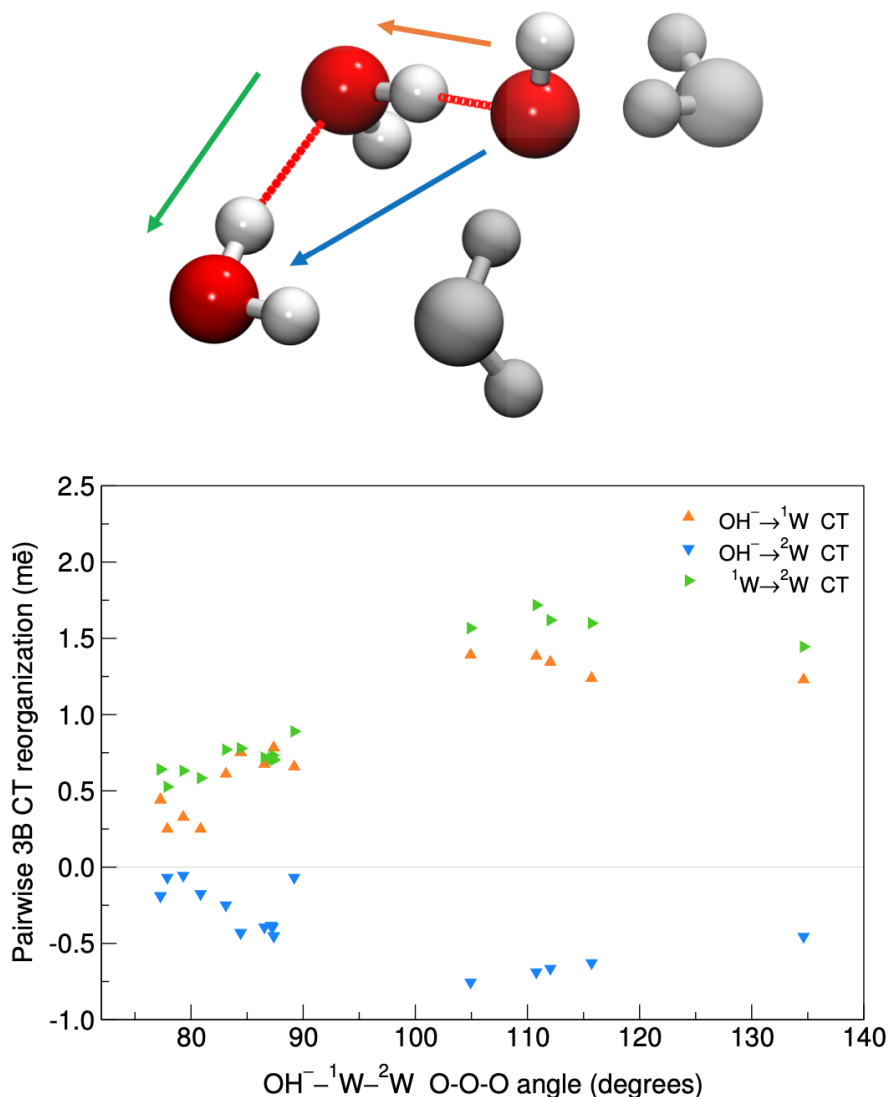


Figure 4.8. 3B ALMO monomer $\rightarrow$ monomer CT reorganization vs $\text{OH}^- - {}^1\text{W} - {}^2\text{W}$ O-O-O angle for $\{\text{OH}^-, {}^1\text{W}_{\leftarrow B}^A, A \leftarrow {}^2\text{W}^B\}_{FG}$ fragments.

A similar energy decomposition analysis and charge transfer analysis can be performed on the cooperative and anticooperative protonated water clusters. 1st generation ALMO EDA calculations[136, 137] at the $\omega\text{B97M-V/def2-TZVPPD}$ level of theory were performed on anticooperative $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$, and cooperative $\{\text{H}_3\text{O}^+, B \leftarrow {}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B\}_{FG}$ fragments (refer to Sec. A.1 for specific details).

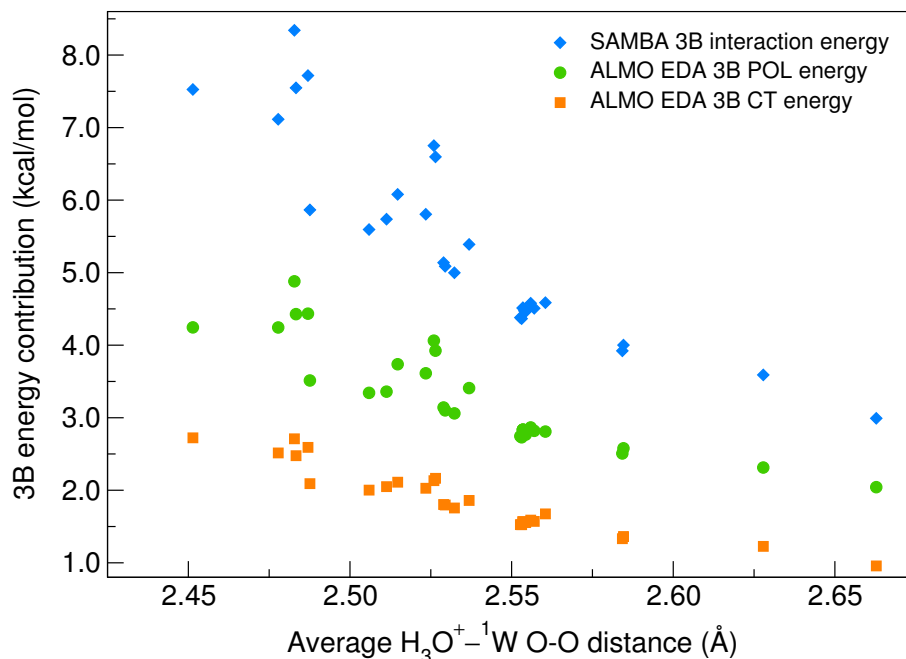


Figure 4.9. Analysis of SAMBA 3B energy contributions and ALMO energy decompositions as a function of the average $\text{H}_3\text{O}^+ - \text{water O-O}$ distance in $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments.

The dependence of $E_{\text{POL}}^{\text{ALMO},3\text{B}}$ and $E_{\text{CT}}^{\text{ALMO},3\text{B}}$ on the average $\text{H}_3\text{O}^+ - \text{W O-O}$ distance is shown in Fig. 4.9. Here, $E_{\text{POL}}^{\text{ALMO},3\text{B}}$ and $E_{\text{CT}}^{\text{ALMO},3\text{B}}$ dominate the total 3B energies of these fragments, with the magnitude of $E_{\text{POL}}^{\text{ALMO},3\text{B}}$ being approximately a factor of two greater than that of $E_{\text{CT}}^{\text{ALMO},3\text{B}}$ for all $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments. Fig. 4.9 also shows that the 3B energy of $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments is negatively-correlated with the average $\text{H}_3\text{O}^+ - \text{W O-O}$ distance, similar to the anticooperative $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments.

Fig. 4.10 shows the correlation between the sum of the pairwise 3B CT “electron reorganization,” $\Delta Q_{\text{CT}}^{3\text{B}}$ (in units of millielectrons, $m\bar{e}$), from ${}^1\text{W}$ to H_3O^+ (${}^1\text{W} \rightarrow \text{H}_3\text{O}^+$ CT) and the average $\text{H}_3\text{O}^+ - \text{W O-O}$ distance. The exclusively-negative values of $\Delta Q_{\text{CT}}^{3\text{B}}$ seen here reflect the exclusively-positive $E_{\text{CT}}^{\text{ALMO},3\text{B}}$ values shown in Fig. 4.9. This reduction of the pairwise ${}^1\text{W} \rightarrow \text{H}_3\text{O}^+$ CT contribution in the presence of the other ${}^1\text{W}$ molecule indicates that anticooperativity in $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments is due in part to the inability of H_3O^+ to accept electron density from each of the two ${}^1\text{W}$ molecules to the same extent that it can accept density from a single ${}^1\text{W}$ in isolation, which is similar to the effect seen in $\{\text{OH}^-, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ fragments. This

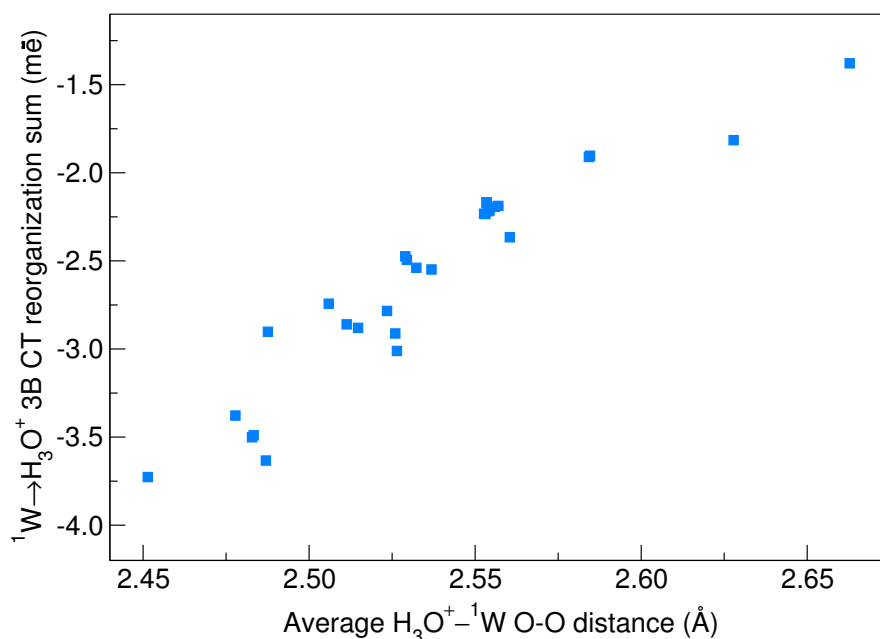


Figure 4.10. Correlation plot between the sum of pairwise 3B ${}^1\text{W} \rightarrow \text{H}_3\text{O}^+$ CT “electron reorganization,” $\Delta Q_{\text{CT}}^{3\text{B}}$ (in millielectrons, $m\bar{e}$) and average $\text{H}_3\text{O}^+ - {}^1\text{W}$ O–O distance in $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{\text{FG}}$ fragments.

is further supported when considering the positive correlation of 2B $\Delta Q_{\text{CT}}^{2\text{B}}$ in $\{\text{H}_3\text{O}^+, {}^1\text{W}\}_{\text{FG}}$ fragments with the $\text{H}_3\text{O}^+ - {}^1\text{W}$ O–O distance (see Supporting Information of Ref. [66]). It should be noted that similar arguments can be made regarding the anticooperative $E_{\text{POL}}^{\text{ALMO},3\text{B}}$ in $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{\text{FG}}$ fragments since polarization of the central H_3O^+ with either ${}^1\text{W}$ molecules will arise in opposition to the polarization due to the other ${}^1\text{W}$ molecule.

The total 3B energies of the cooperative $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^A, {}_{\leftarrow A}{}^2\text{W}^B\}_{\text{FG}}$ fragments were found to correlate with the distance between the ${}^1\text{W} - {}^2\text{W}$ (H-bonded) donor-H and acceptor-O, such that the 3B energy becomes more negative as the H–O distance decreases, as shown in Fig. 4.11. The dependence of $E_{\text{POL}}^{\text{ALMO},3\text{B}}$ and $E_{\text{CT}}^{\text{ALMO},3\text{B}}$ on this coordinate is also shown in Fig. 4.11. As in the case of anticooperative fragments discussed above, $E_{\text{POL}}^{\text{ALMO},3\text{B}}$ and $E_{\text{CT}}^{\text{ALMO},3\text{B}}$ are the dominant contributions to the total 3B energy, resulting in a value for $E_{\text{POL}}^{\text{ALMO},3\text{B}}$ that is approximately twice as large as the corresponding $E_{\text{CT}}^{\text{ALMO},3\text{B}}$ contribution.

Further insight into the nature of H-bonding in $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^A, {}_{\leftarrow A}{}^2\text{W}^B\}_{\text{FG}}$ fragments

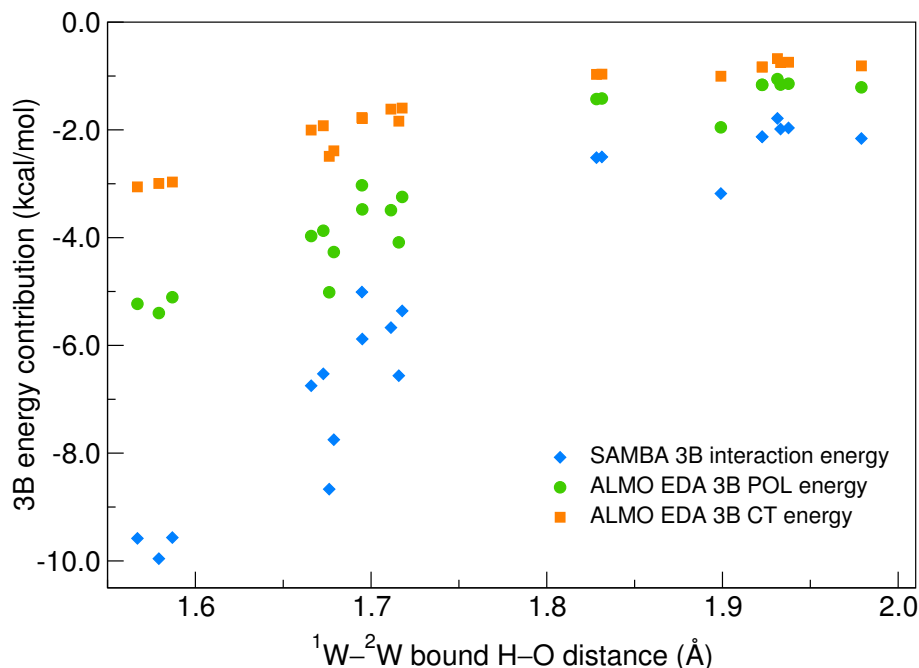


Figure 4.11. Analysis of 3B SAMBA interaction energy contributions and ALMO energy decompositions as a function of the ${}^1\text{W}-{}^2\text{W}$ (H-bonded) donor-H and acceptor-O distance described in the main text in $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^{\text{A}}, {}^2\text{W}_{\leftarrow\text{A}}^{\text{B}}\}_{\text{FG}}$ fragments.

is gained through the analysis of the corresponding $\Delta Q_{\text{CT}}^{3\text{B}}$, as shown in Fig. 4.12. Here, the (directional) pairwise CT reorganization of ${}^1\text{W} \rightarrow \text{H}_3\text{O}^+$, ${}^2\text{W} \rightarrow \text{H}_3\text{O}^+$, and ${}^2\text{W} \rightarrow {}^1\text{W}$ are plotted against the corresponding ${}^1\text{W}-{}^2\text{W}$ H-O distance. This analysis illustrates that cooperativity manifests through an enhancement of both ${}^1\text{W} \rightarrow \text{H}_3\text{O}^+$ and ${}^2\text{W} \rightarrow {}^1\text{W}$ $\Delta Q_{\text{CT}}^{3\text{B}}$ contributions in the 3-mer relative to the $\Delta Q_{\text{CT}}^{2\text{B}}$ term in the corresponding 2-mers. The ${}^2\text{W} \rightarrow \text{H}_3\text{O}^+$ $\Delta Q_{\text{CT}}^{3\text{B}}$ contribution is negative for each fragment, but the magnitudes of these effects are less significant than the other two. While the ΔQ_{CT} analysis of the $\{\text{H}_3\text{O}^+, {}^1\text{W}^{\text{A}}, {}^1\text{W}^{\text{B}}\}_{\text{FG}}$ fragments (Fig. 4.10) suggests that the anticooperativity results from the geometric opposition of the directions of the dominant CT reorganizations (i.e. ${}^1\text{W}^{\text{A}} \rightarrow \text{H}_3\text{O}^+$ and ${}^1\text{W}^{\text{B}} \rightarrow \text{H}_3\text{O}^+$), Fig. 4.12 suggests that the observed cooperativity in $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^{\text{A}}, {}^2\text{W}_{\leftarrow\text{A}}^{\text{B}}\}_{\text{FG}}$ fragments is a consequence of the dominant CT reorganization having similar directionality (i.e. pointing from H_3O^+ toward ${}^2\text{W}$). Rationalization of the $E_{\text{POL}}^{\text{ALMO},3\text{B}}$ term follows from similar arguments.

3B energy decomposition analysis of the protonated water cluster fragments can also

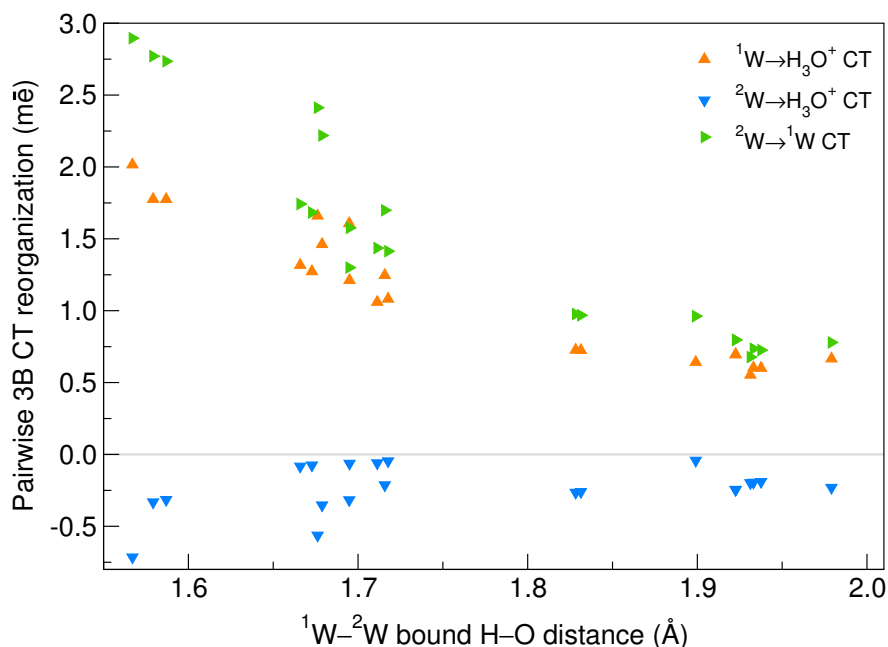


Figure 4.12. Pairwise 3B ALMO monomer $\rightarrow$ monomer CT reorganization vs. the ${}^1\text{W}-{}^2\text{W}$ (H-bonded) donor-H and acceptor-O distance described in the main text for $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B\}_{FG}$ fragments.

aid in the analysis of the accuracy of XC-functionals in reproducing the reference total 3B energies discussed in Sec. 3.3.4. Given the strong 3B charge transfer effects inherent to the $\{\text{H}_3\text{O}^+, {}^1\text{W}^A, {}^1\text{W}^B\}_{FG}$ and $\{\text{H}_3\text{O}^+, {}_{B\leftarrow}{}^1\text{W}^A, {}^2\text{W}_{\leftarrow A}^B\}_{FG}$ fragments, and that the inclusion of Hartree-Fock exchange decreases the deviation from the 3B reference energies (Sec. 3.3.4), one may hypothesize that the delocalization error[12, 13, 14] is the primary culprit of the trends seen in Fig. 3.25. There is evidence[120] that effects due to the delocalization error manifest in the charge transfer term from ALMO EDA calculations, which would imply that for a given XC-functional, the deviation relative to CCSD(T) 3B energy should correlate with its corresponding $E_{CT}^{\text{ALMO},3B}$ term. This hypothesis is supported by the analysis shown in Fig. 4.13 that demonstrates that there is a nearly linear relation between the mean 3B error and $E_{CT}^{\text{ALMO},3B}$ for each XC-functional. Furthermore, the trends seen here agree with trends found in earlier work investigating the delocalization error for various XC-functionals (see Fig. 6 of Ref. 14). Note that 2B energies calculated with the XC-functionals were found to mostly correlate with

their respective ALMO-EDA frozen energies (see Supporting Information of Ref. 66).

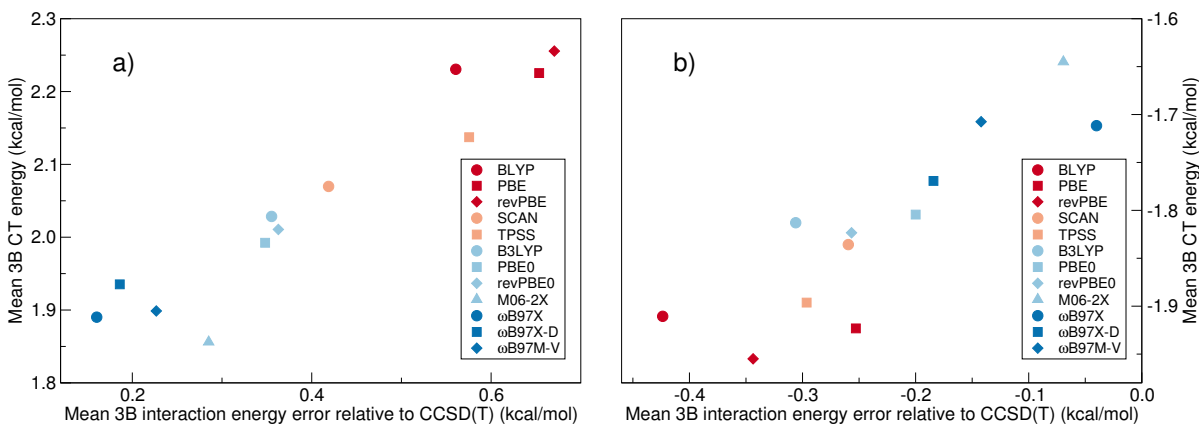


Figure 4.13. Correlation between the mean 3B energy error and the mean ALMO EDA 3B CT energy of the 3-mer fragment groups: a) $\{\text{H}_3\text{O}^+, {}^1\text{W}^{\text{A}}, {}^1\text{W}^{\text{B}}\}_{\text{FG}}$ and b) $\{\text{H}_3\text{O}^+, {}_{\text{B} \leftarrow} {}^1\text{W}^{\text{A}}, {}^2\text{W}^{\text{B}}_{\leftarrow \text{A}}\}_{\text{FG}}$, calculated with each XC-functional. Each marker corresponds to one functional, and the color of the marker specifies the level of DFT approximation (GGAs are red, meta-GGAs are orange, hybrids are light blue, and range-separated hybrids are blue).

4.5 Acknowledgements

Section 4.4 is almost entirely a reprint of the material as they appear in “Assessing Many-Body Effects of Water Self-Ions. I: $\text{OH}^-(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2018**, 14 (4), 1982–1997, and “Assessing Many-Body Effects of Water Self-Ions. II: $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2019**, 15 (9), 4816–4833.

Sections 4.2, and 4.3 are almost entirely reprints of the material as they appear in “Nature of Halide–Water Interactions: Insights from Many-Body Representations and Density Functional Theory” *J. Chem. Theory Comput.*, **2019**, 15 (5), 2983–2995 and “Nature of Alkali–Water Interactions: Insights from Many-Body Representations and Density Functional Theory II” *J. Chem. Theory Comput.*, **2020**, 16 (5), 3055–3072.

Chapter 5

Model Potential Energy Functions

Molecular dynamics (MD) simulations have become an invaluable tool, potentially allowing for detailed investigation of chemical phenomena that are experimentally inaccessible.[138, 139, 140, 141, 142, 143, 144, 145, 146] Presently, the major computational bottleneck associated with MD simulations is handling the electronic degrees of freedom[147] that determine the potential energy landscape of the molecular system of interest.[148] Electronic wave function methods and, in particular, configuration interaction and its approximations, such as the coupled cluster hierarchy,[148, 149] offer “chemical” accuracy but scale very poorly with system size,[150] and are therefore computationally intractable in MD simulations.[151, 152]

One of the most efficient alternatives to wave function methods is the use of analytical potential energy functions (PEFs),[151, 152] commonly known as force fields (FFs), which effectively integrate out the electronic degrees of freedom. The ideal construction of a FF entails generating a lossless representation of the Born-Oppenheimer PES associated with the molecular system of interest, which is then used to evolve in time the dynamics of the remaining nuclear degrees of freedom. Unfortunately, the construction of accurate PESs suffers from numerous complications due to the difficulties in describing many-electron systems. For example, eliminating the electronic degrees of freedom often requires that FFs be specifically tailored to particular molecular systems to achieve sufficient accuracy, and that each atom be assigned a unique identity that is preserved during the MD simulation.

Beyond these complications, PEF construction is not trivial since there does not exist a unique functional representation that accurately describes the target PES.[151] The most common approach to the development of PEFs is to express the total potential energy of a molecular system as a sum of contributions arising from different types of interactions. Intramolecular distortion energies (i.e. 1B energies) are often partitioned into terms involving stretching, bending, and dihedral coordinates. Intermolecular interactions tend to be much more difficult to represent, especially in configurations involving large inter-monomer orbital overlap (IMOO) at close range. Intermolecular interactions are usually represented as a sum of individual terms describing the different types of physical interactions, such as those discussed in Chapter 4, for example permanent electrostatics (ELEC), polarization (POL), charge transfer (CT), London dispersion (DISP), and Pauli repulsion (PAULI).

Some of the most successful PEFs are constructed to reproduce the individual terms in the MBE (Eq. 3.13) calculated with high-level electronic structure theory methods such as CCSD(T).[153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168] Model PEFs that accurately reproduce many-body energies at all orders are desirable because of their applicability to systems beyond the training data used in their development. For example, the MB-pol[160, 161, 162, 52] PEF was developed from energies of small clusters (2B and 3B energies of $(\text{H}_2\text{O})_2$ and $(\text{H}_2\text{O})_3$ clusters, respectively), as well as electrostatic properties of isolated water monomers and 2-mers, but is able to reproduce bulk properties of pure water systems across different phases.[52, 169, 170, 171]

5.1 Development of Many-Body PEFs: The MB-nrg Methodology

As discussed in earlier chapters, for a given molecular system, the Born-Oppenheimer PES for some electronic state (usually the ground state) can be thought of as the potential energy between the nuclei after integrating out the electronic degrees of freedom, assuming

the Born-Oppenheimer approximation is valid. The fact that we integrated out the electronic coordinates means the PES is just a function of the positions of the nuclei (which we may call “atomic sites,” or “atoms”). In Chapter 3, we were referring to the PES as the total electronic energy, E_{tot} . Recall that for a given assignment of monomer identities, we may also use E_{bind} as the PES, which differs from E_{tot} by a constant (the sum of the total electronic energies of the optimized monomers). E_{bind} is a more convenient quantity to use when modeling the PES so that the (intramolecular) 1B distortion energies are 0 for optimized monomers.

When the MBE (Eqs. 3.13 and 3.14) for a given system converges rapidly, such that a classical representation of the inductive electrostatics (polarization) is sufficient to describe the high order many-body terms, then it is possible to construct a model PEF that accurately, and efficiently describes the true PES of the system by applying the MB-nrg methodology. An MB-nrg PEF consists of two parts: first is the “underlying” long-range potential which includes terms representing ELEC, POL, and DISP interactions described in Chapter 4, which collectively reproduce the interaction energies between monomer units in the limit of large inter-monomer separations. The second part of an MB-nrg PEF can be seen as a close-range correction term, which corrects for deficiencies in the long-range terms for small inter-monomer separations. The key idea of the MB-nrg methodology is that explicit close-range corrections can be applied at the low-order terms of the MBE, usually just the 2B and 3B terms which can be obtained at a relatively low computational cost with the accurate electronic structure methods such as CCSD(T), while higher-order many-body energies (which require more expensive electronic structure calculations to obtain) are implicitly reproduced from the classical POL interactions in the long-range potential.

5.1.1 TTM-nrg PEFs: Obtaining the Long-Range Potential

In order to develop an MB-nrg PEF for some system, we must first obtain the long-range potential for that system. These long-range interaction terms can be used to construct a kind of intermediate many-body PEF which we call the TTM-nrg[172, 165] potential (TTM refers to the

Thole-type model electrostatics used in the potential). We first describe the development of a TTM-nrg PEF, and then extending the PEF into MB-nrg will only require adding the close-range correction term.

The intermolecular interactions in a TTM-nrg PEF are described by the following terms:

$$V_{INT}^{TTM-nrg} = V_{ELEC}^{TTM-nrg} + V_{POL}^{TTM-nrg} + V_{PAULI}^{TTM-nrg} + V_{DISP}^{TTM-nrg} \quad (5.1)$$

The TTM-nrg model employs an extended Thole-type model[173] describing ion–water permanent electrostatic ($V_{ELEC}^{TTM-nrg}$) and polarization ($V_{POL}^{TTM-nrg}$) interactions.[174] The TTM-nrg functional form employs “smeared” induced dipoles (centered at the atom sites) and smeared point charges within the exponential-4 smearing scheme introduced in Ref. 175). A set of point charges are assigned to each monomer unit, chosen to reproduce the electrostatic potential surrounding the isolated monomer, according to methods such as CHELPG[176] or CM5.[177] In general, the point charges can be fixed, or geometry-dependent (for geometry-dependent charges, forces must be computed via the chain rule). These partial charges are often centered on each atomic site, although one notable exception to this is the use of the M-site[178, 160] in many water models (placing the negative partial charge on the M-site rather than the water oxygen improves the description of the quadrupole moments on a water monomer). Point dipole polarizabilities are assigned to atomic sites[163, 166] in order to describe POL interactions,[120, 119] including high-order terms in the MBE.

Effective isotropic atomic dipole polarizabilities, $\alpha_a^{\text{eff},M}$ for each atom, a , in a given monomer, M , were determined from:

$$\alpha_a^{\text{eff},M} = \alpha_a^{\text{free}} \frac{V_a^{\text{eff},M}}{V_a^{\text{free}}} \quad (5.2)$$

where α_a^{free} and V_a^{free} are the dipole polarizability and atomic volume of atom a in isolation, and $V_a^{\text{eff},M}$ is the atomic volume of atom a within monomer, M . Both V_a^{free} and $V_a^{\text{eff},M}$ can be

calculated with XDM6.

The TTM-nrg pairwise atom-atom P+D interactions (from the sum of $V_{PAULI}^{TTM-nrg}$ and $V_{DISP}^{TTM-nrg}$) between two monomers, M_1 and M_2 , take the form of a Buckingham potential:

$$V_{PAULI}^{TTM-nrg} + V_{DISP}^{TTM-nrg} = \sum_{i \in M_1} \sum_{j \in M_2} A_{ij} e^{-b_{ij} R_{ij}} - f_6^{TT}(\delta_{ij} R_{ij}) \frac{C_{ij}^{(6)}}{R_{ij}^6} \quad (5.3)$$

where indices i and j refer to the atomic sites in each of the two monomers. The first term on the right-hand side represents close-range repulsive PAULI interactions. A_{ij} and b_{ij} are fitted parameters specific to each pair of atom types, and R_{ij} is the distance between atoms i and j . The second term on the right-hand side represents DISP interactions which are damped with the Tang-Toennies function $f_6^{TT}(\delta_{ij})$.^[179] $C_{ij}^{(6)}$ is the dispersion coefficient specific to each pair of atom types, and δ_{ij} is a damping parameter, which is set to be equal to the corresponding parameter, b_{ij} in the PAULI term.^[172] All $C_{ij}^{(6)}$ coefficients were calculated using the exchange-hole dipole model (XDM6),^[180, 181, 182] as implemented in the Postg software.^[183, 184]

The exponential function representing PAULI (often called a Born-Mayer potential) has a physical justification based on the overlap between closed shell orbitals.^[185] Additionally, the long-range behavior of DISP interactions can be shown to have R^{-6} dependence with second order perturbation theory.^[7]

An explicit representation of CT is in general omitted, although attempts have been made to developing effective functional forms representing CT effects.^[186, 187, 188, 189, 190, 191] Terms specific to CT interactions are not included in TTM-nrg PEFs, although CT interactions are implicitly included due to the fact that the model is fitted to reference energies which do include CT effects.

5.1.2 MB-nrg PEFs

After developing the TTM-nrg PEF for the system of interest, all of the terms needed in the long-range potential are determined. Next, we must replace the pairwise Born-Mayer

potential, $V_{PAULI}^{TTM-nrg}$ (the first term on the right-hand side of Eq. 5.3) with 2B (and if necessary, 3B) correction terms in order to obtain an MB-nrg PEF. As stated above, the correction terms are needed to correct for deficiencies in the long-range potential. For example, while the atom-centered point charges are able to reproduce the electrostatic potential surrounding a monomer, it lacks the inhomogeneity of the true charge distribution in the monomer, namely a positively charged nuclei enclosed in a diffuse electron density, and therefore lacks (attractive) charge penetration (CP) effects which occur at small inter-monomer separations when the electron densities overlap. Additionally, terms representing CT interactions are not explicitly represented in TTM-nrg PEFs, despite being important in many systems (see Chapter 4).

$$V_{INT}^{MB-nrg} = s^{(2B)}(x)V_{PIP}^{(2B)} + [s^{(3B)}(x)V_{PIP}^{(3B)}] + V_{ELEC}^{TTM-nrg} + V_{POL}^{TTM-nrg} + V_{DISP}^{TTM-nrg} \quad (5.4)$$

In order to correct for all of these effects, MB-nrg PEFs include 2B (and 3B) correction terms represented by permutationally invariant polynomials[192] (PIPs), V_{PIP} , although the methodology would apply to any sufficiently flexible function, such as neural nets and Gaussian approximation potentials.[193] These correction terms are multiplied by a sigmoid-shaped function, $s(x)$, (which we call a *switching function*) dependent on some geometric coordinate, x , (for example the distance between the first atom in each monomer in the 2-mer for a 2B correction term) that is equal to 1 within some inner-cutoff value, and 0 beyond some outer-cutoff value, smoothly turning off the correction term between the two cutoffs. The switching function is important because the correction terms are relatively expensive to calculate, but only important for small inter-monomer separations while the long-range potential has the proper behavior at large inter-monomer separations.

The PIPs used in MB-nrg PEFs are discussed in the next few sections, and some considerations regarding the switching functions are discussed afterward.

5.1.3 Brief Motivation for Permutationally Invariant Polynomials

Permutationally Invariant Polynomials[192] (PIPs) are flexible functions which are able to approximately represent the PES for a given system, or the 2B or 3B correction term in an MB-nrg PEF. PIPs are constructed such that identical atomic sites in a given molecular system are treated identically in the PES representation. For example, in a molecule like ethanol ($\text{CH}_3\text{CH}_2\text{OH}$), the three methyl-hydrogens (in CH_3) are identical between themselves, but are not identical with the other three hydrogens since they have different bonding-environments (furthermore, the alcohol-hydrogen in OH, $\text{H}''(1)$, is not identical with any other hydrogens, $\text{H}^{(1)}$ and $\text{H}^{(2)}$, and the two carbons are not identical). A PIP for ethanol will be “permutationally invariant” because it will be invariant to how you label identical atoms, so, for any particular conformation, labeling the methyl-hydrogens as $\text{H}^{(1)}$, $\text{H}^{(2)}$, $\text{H}^{(3)}$, or $\text{H}^{(2)}$, $\text{H}^{(1)}$, $\text{H}^{(3)}$ will yield the same energies, gradients, etc..

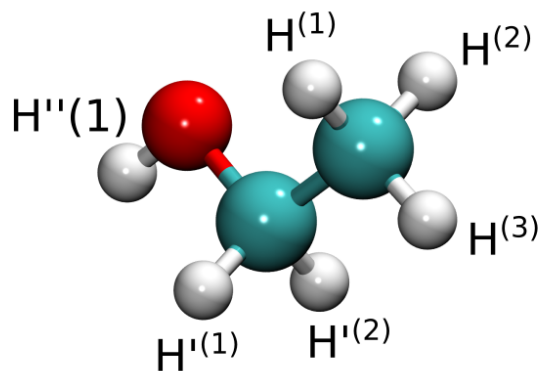


Figure 5.1. An ethanol monomer with hydrogen (atomic sites) labeled based on their chemical identities. $\text{H}^{(1)}$, $\text{H}^{(2)}$, and $\text{H}^{(3)}$ are all identical, $\text{H}^{(1)}$ and $\text{H}^{(2)}$ are identical with each other.

There are three major reasons why permutational invariance is desirable:

1. The true PES will be permutationally invariant because of the indistinguishability of identical particles.[194]

2. It is cheaper to fit and evaluate PIPs, since the permutational invariance is enforced by forcing all permutationally invariant parameters to be equal (e.g. in the ethane PIP, all fitted parameters between methyl-hydrogen $H^{(1)}$ and the methyl-carbon will be equal to the corresponding parameters between $H^{(2)}$ and the methyl-carbon). Fewer distinct parameters means there is a smaller search space during the fitting.
3. There is increased sampling in the training set due to the symmetry (in a CH_4 PIP, for example, the training set would need to be four times larger to get the same sampling for fitting a non-permutationally invariant function).

Note that permutational invariance is enforced in almost any force field, since all equivalent bonds will have the same force constants, etc..

5.1.4 Permutationally Invariant Polynomials For N Identical Particles

For simplicity, let's assume that we have N identical atomic sites, A , in our system, A_N , and label each site with an integer from 1 to N . PIPs are generally functions of the distances between every pair of site, so in A_N , the PIP will depend on distances $R_{1,2}, R_{1,3}, \dots, R_{1,N}, R_{2,3}, R_{2,4}, \dots, R_{N-1,N}$. In order to capture the correct shape of the PES with as few terms as possible, PIPs are generally polynomials of (let's call them) "secondary variables," $\chi_{1,2}, \chi_{1,3}, \dots, \chi_{N-1,N}$. Typically, we use simple exponential-type variables (but other types of secondary variables have also been used) defined as:

$$\chi_{i,j} \equiv \exp[-b_{i,j}R_{i,j}] \quad (5.5)$$

where $b_{i,j}$ is a fitted (nonlinear) parameter specific to the type of pair of sites i and j (note that these parameters must be permutationally invariant, so in the case of our A_N system, all $b_{i,j}$ parameters will be equal).

In general, a degree- D polynomial of variables, $\chi_{1,2}, \chi_{1,3}, \dots, \chi_{N-1,N}$, can be written as:

$$V_{\text{PIP}} = \sum_{d=1}^D \sum_{\mathbf{p} \in \mathcal{P}(d|g)} c_{\mathbf{p}} \mathcal{S}[\chi_{1,2}^{p_{1,2}} \chi_{1,3}^{p_{1,3}} \dots \chi_{\alpha}^{p_{\alpha}} \dots \chi_{N-1,N}^{p_{N-1,N}}] \quad (5.6)$$

$$\mathcal{P}(d|m) \equiv \{\mathbf{p} = (p_{1,2}, p_{1,3}, \dots, p_{\alpha}, \dots, p_{N-1,N}), p_{\alpha} \in \mathbb{Z}_0 \mid \sum_{\alpha} p_{\alpha} = d\} \quad (5.7)$$

where D is the degree of the polynomial, $\mathcal{P}(d|g)$ is the set of integer partitions of d ($1 \leq d \leq D$) into $g = \binom{N}{2}$ non-negative, ordered summands ($\mathbb{Z}_0$ being the set of non-negative integers) with each summand corresponding to one interatomic pair, $\alpha = (i, j)$, $c_{\mathbf{p}}$ is the coefficient of the monomial labeled by partition $\mathbf{p}$, $\mathcal{S}$ is the monomial symmetrization operator which forces the coefficients of monomials similar up to permutations of identical nuclei to be equal,[192] and χ_{α} are exponential-type variables of the interatomic distances between pairs labeled by α .

For example, for $N = 3 \Rightarrow g = \binom{3}{2} = 3$ and $d = 1$:

$$\mathcal{P}(1|3) = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\} \quad (5.8)$$

and for $N = 3 \Rightarrow g = 3$ and $d = 2$:

$$\mathcal{P}(2|3) = \{(2, 0, 0), (0, 2, 0), (0, 0, 2), (1, 1, 0), (1, 0, 1), (0, 1, 1)\} \quad (5.9)$$

so a degree $D = 2$ PIP for A_3 can be written as:

$$\begin{aligned} V_{\text{PIP}} = & c_{(1,0,0)} \mathcal{S}[\chi_{1,2}] + c_{(0,1,0)} \mathcal{S}[\chi_{1,3}] + c_{(0,0,1)} \mathcal{S}[\chi_{2,3}] \\ & + c_{(2,0,0)} \mathcal{S}[\chi_{1,2}^2] + c_{(0,2,0)} \mathcal{S}[\chi_{1,3}^2] + c_{(0,0,2)} \mathcal{S}[\chi_{2,3}^2] \\ & + c_{(1,1,0)} \mathcal{S}[\chi_{1,2}\chi_{1,3}] + c_{(1,0,1)} \mathcal{S}[\chi_{1,2}\chi_{2,3}] + c_{(0,1,1)} \mathcal{S}[\chi_{1,3}\chi_{2,3}] \end{aligned} \quad (5.10)$$

which after symmetrization becomes:

$$\begin{aligned} V_{\text{PIP}} = & c_{(1,0,0)} (\chi_{1,2} + \chi_{1,3} + \chi_{2,3}) \\ & + c_{(2,0,0)} (\chi_{1,2}^2 + \chi_{1,3}^2 + \chi_{2,3}^2) \\ & + c_{(1,1,0)} (\chi_{1,2}\chi_{1,3} + \chi_{1,2}\chi_{2,3} + \chi_{1,3}\chi_{2,3}) \end{aligned} \quad (5.11)$$

Notice that symmetrization makes the PIP invariant to permutations of identical sites, since all monomials that are equivalent up to these permutations are forced to have the same coefficient (i.e. linear parameter, $c_{\mathbf{p}}$). We usually call these sums of symmetric monomials “terms.”

The $N = 3$ example was given for simplicity, but it doesn’t give a complete picture of the symmetrization. For example, if we were to write out the general form of a PIP for A_4 ($N = 4$), you should convince yourself that the monomials, $\chi_{1,2}\chi_{1,3}$ and $\chi_{1,2}\chi_{3,4}$ are not equivalent up to permutations of identical sites, since the first monomial depends on one site twice (site 1), but the second monomial depends on each site exactly once ($\chi_{1,2}\chi_{1,3}$ is symmetric with $\chi_{1,2}\chi_{2,3}$, for example, while $\chi_{1,2}\chi_{3,4}$ is symmetric with $\chi_{1,3}\chi_{2,4}$). This complication isn’t seen in the $N = 3$ case because there are only three sites. This idea is examined further in our discussion of monomial connectivity graphs in Sec. 5.1.10.

5.1.5 “MB-Fitting” Permutationally Invariant Polynomials

Since the PIPs depend on both linear and nonlinear parameters, the polynomial coefficients ($c_{\mathbf{p}}$) and exponential parameters ($b_{i,j}$), respectively, we use a combination of (weighted) linear least squares (LLS), including Tikhonov regularization,[195] and some nonlinear solver (we use the Nelder-Meade “simplex” optimizer). This is accomplished by initializing a guess for the nonlinear parameters, solving the LLS with the nonlinear parameters fixed, and then passing the (weighted) error from LLS into the cost function for the nonlinear solver, and iterating (in other words, LLS is coded directly into the Nelder-Meade cost function). Technically, for K nonlinear parameters, the Nelder-Meade algorithm uses K distinct initial guesses for the K parameters (totaling K^2 initial parameters), but we do not discuss the nonlinear solver in detail here.

For simplicity, let’s assume that we doing ordinary (unweighted) LLS without any regularization. Additionally, let’s start with LLS for a general polynomial (which readily extends to PIPs). A general g -dimensional D -degree polynomial with linear coefficients, $\mathbf{c} = [c_{\mathbf{p}}]$ (we are

considering the ordered list of coefficients to be the vector, $\mathbf{c}$) can be written as:

$$f(\mathbf{x}; \mathbf{c}) = \sum_{d=1}^D \sum_{\mathbf{p} \in \mathcal{P}(d|g)} c_{\mathbf{p}} x_1^{p_1} x_2^{p_2} \cdots x_{\alpha}^{p_{\alpha}} \cdots x_g^{p_g} \quad (5.12)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_g)$ and $\mathcal{P}(d|g)$ is given by Eq. 5.7. Given T samples, $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}, \dots, \mathbf{x}^{(T)}$ (i.e. T configurations in the training set), each corresponding to its own known value, $y^{(1)}, y^{(2)}, \dots, y^{(n)}, \dots, y^{(T)}$ (i.e. reference energies to fit), we want to find the unique vector of linear coefficients, $\mathbf{c} = [c_{\mathbf{p}}]$, that minimizes the following cost function:

$$\varepsilon(\mathbf{c}) = \sum_n (f(\mathbf{x}^{(n)}; \mathbf{c}) - y^{(n)})^2 \quad (5.13)$$

Taking our T samples, and the general form of our polynomial model, $f(\mathbf{x}; \mathbf{c})$, we can construct the rectangular design matrix, where each column corresponds to one monomial of $f(\mathbf{x}; \mathbf{c})$, and each row corresponds to one sample. In one dimension (writing x_n instead of $x^{(n)}$ for the n^{th} sample), the rectangular design matrix, $\mathbf{M}$ for a D -degree polynomial is:

$$\mathbf{M} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^D \\ 1 & x_2 & x_2^2 & \cdots & x_2^D \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_T & x_T^2 & \cdots & x_T^D \end{bmatrix}$$

If $\mathbf{M}$ is full-rank, LLS has a solution if $T \geq D$, so $\mathbf{M}$ will generally have more rows than columns ($D = T$ corresponds to a polynomial interpolation).

Applying the matrix, $\mathbf{M}$, to a given vector of polynomial coefficients, $\mathbf{c}$, yields the vector whose components are the polynomial, $f(x; \mathbf{c})$, evaluated at each sampled point:

$$\mathbf{M}\mathbf{c} = [f(x_1; \mathbf{c}), f(x_2; \mathbf{c}), \dots, f(x_T; \mathbf{c})] \quad (5.14)$$

Writing the reference values of the training set, y_n , as a vector, $\mathbf{y} = [y_1, y_2, \dots, y_T]$, the LLS cost function can be written as:

$$\varepsilon(\mathbf{c}) = \|\mathbf{M}\mathbf{c} - \mathbf{y}\|_2^2 \quad (5.15)$$

where $\|\cdot\|_2$ denotes the l_2 (Euclidian) norm in a T -dimensional vector space. As a side note, the corresponding cost function for weighted LLS with Tikhonov regularization would be:

$$\varepsilon(\mathbf{c}) = \|\mathbf{W}^{\frac{1}{2}}(\mathbf{M}\mathbf{c} - \mathbf{y})\|_2^2 + \|\Gamma\mathbf{c}\|_2^2 \quad (5.16)$$

where $\mathbf{W}$ is the diagonal matrix of weights, and Γ is the Tikhonov matrix which we take to be a constant, α , multiplied by the identity matrix, $\Gamma = \alpha\mathbf{1}$.

In (rare) cases where the cost function can be minimized to 0, we are simply solving the typical “ $\mathbf{A}\mathbf{x} = \mathbf{b}$,” or using the variables that we have been working with: $\mathbf{M}\mathbf{c} = \mathbf{y}$. In the general case, however, there will be some error, which amounts to a remainder-vector called the residual, $\mathbf{r}$ defined such that: $\mathbf{M}\mathbf{c} + \mathbf{r} = \mathbf{y}$.

It helps to consider the geometric interpretation of what we are discussing: The statement that $\mathbf{M}\mathbf{c} = \mathbf{y} - \mathbf{r}$ is saying that the vector, $\mathbf{y} - \mathbf{r}$ lies within the range of matrix $\mathbf{M}$ (also know as the column space). Specifically, $\mathbf{y} - \mathbf{r}$ is the linear combination of the columns of $\mathbf{M}$ whose expansion coefficients are the elements of $\mathbf{c}$ (the coefficients of our polynomial). The fact that the cost function (Eq. 5.15) cannot be minimized to 0 (in general) means that our vector of reference values, $\mathbf{y}$, must lie outside the column space of $\mathbf{M}$. Furthermore, if $\mathbf{c}$ minimizes the cost function, then the residual vector, $\mathbf{r}$ must be orthogonal to the column space of $\mathbf{M}$, and the vector, $\mathbf{y} - \mathbf{r}$, must be the orthogonal projection of $\mathbf{y}$ onto the column space of $\mathbf{M}$. By the fundamental theorem of linear algebra, $\mathbf{M}^*\mathbf{r} = 0$ (since the subspace orthogonal to the column space of $\mathbf{M}$ is the null

space of the conjugate transpose of $\mathbf{M}$, $\mathbf{M}^*$) so we have:

$$\mathbf{M}^* \mathbf{M} \mathbf{c} = \mathbf{M}^* (\mathbf{y} - \mathbf{r}) \quad (5.17)$$

$$\mathbf{M}^* \mathbf{M} \mathbf{c} = \mathbf{M}^* \mathbf{y} \quad (5.18)$$

which means we may construct the so-called psuedoinverse of $\mathbf{M}$, $\mathbf{M}^+ = (\mathbf{M}^* \mathbf{M})^{-1} \mathbf{M}^*$, and simply apply it to $\mathbf{y}$ to find the optimal $\mathbf{c}$, but this is less efficient than the standard approaches (but follows along similar lines).

One such approach is to construct the QR factorization of $\mathbf{M}$ into the product of a unitary matrix, $\mathbf{Q}$ (an orthogonal matrix in real vector spaces), and an upper-triangular matrix, $\mathbf{R}$. Since the columns of $\mathbf{Q}$ form an orthogonal basis of the column space of $\mathbf{M}$, the matrix $\mathbf{Q} \mathbf{Q}^*$ is the orthogonal projector onto the column space of $\mathbf{M}$. Therefore, we can project $\mathbf{y}$ into the vector, $\mathbf{Q} \mathbf{Q}^* \mathbf{y}$ (which lies in the column space of $\mathbf{M}$), and find the vector $\mathbf{c}$ which is mapped to $\mathbf{Q} \mathbf{Q}^* \mathbf{y}$ due to application of $\mathbf{M} = \mathbf{Q} \mathbf{R}$:

$$\mathbf{Q} \mathbf{R} \mathbf{c} = \mathbf{Q} \mathbf{Q}^* \mathbf{y} \quad (5.19)$$

$$\mathbf{R} \mathbf{c} = \mathbf{Q}^* \mathbf{y} \quad (5.20)$$

where we get the latter equation after left-multiplying both sides of the former equation by the inverse of $\mathbf{Q}$ ($\mathbf{Q}^{-1} = \mathbf{Q}^*$, since $\mathbf{Q}$ is unitary). The solution, $\mathbf{c}$, for this equation is easily determined after obtaining the vector, $\mathbf{Q}^* \mathbf{y}$, since $\mathbf{R}$ is upper-triangular.

Notice that all we are doing is determining the optimal linear parameters (the vector, $\mathbf{c}$), given the matrix, $\mathbf{M}$ and the vector of reference values, $\mathbf{y}$. Therefore, to extend this procedure to fitting linear coefficients of the PIPs (fitting to energies), all we need to do is insert each sum of symmetric monomials (i.e. terms of monomials sharing the same linear coefficient) into one of the columns of $\mathbf{M}$ with each row corresponding to the terms evaluated for one training set configuration. If we label the n^{th} term, evaluated for the i^{th} training set configuration, $\mathbf{x}^{(i)}$, as

$m_n(\mathbf{x}^{(i)})$, then we have (assuming there are G terms):

$$\mathbf{M}_{PIP} = \begin{bmatrix} m_1(\mathbf{x}^{(1)}) & m_2(\mathbf{x}^{(1)}) & \cdots & m_G(\mathbf{x}^{(1)}) \\ m_1(\mathbf{x}^{(2)}) & m_2(\mathbf{x}^{(2)}) & \cdots & m_G(\mathbf{x}^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ m_1(\mathbf{x}^{(T)}) & m_2(\mathbf{x}^{(T)}) & \cdots & m_G(\mathbf{x}^{(T)}) \end{bmatrix}$$

and we follow the same procedure outlined above. It should also be noted that when fitting a PIP for an MB-nrg potential (Eq. 5.4), we need to subtract the sum of the long-range terms, $V_{ELEC}^{TTM-nrg} + V_{POL}^{TTM-nrg} + V_{DISP}^{TTM-nrg}$, from the corresponding reference energy for each training set configuration when constructing the $\mathbf{y}$ vector.

5.1.6 Fitting Permutationally Invariant Polynomial Gradients

Since the (partial) derivative of a polynomial is just another polynomial, the same procedure can be adapted to fit gradients, except now we will need more rows in the design matrix: one for each partial derivative of each configuration (for N sites and T configurations, we need $3NT$ rows), and the reference values will be a vector of all gradient components of all training set configurations, $[\partial_{x_1}E^{(1)}, \partial_{y_1}E^{(1)}, \dots, \partial_{z_N}E^{(1)}, \partial_{x_1}E^{(2)}, \dots, \partial_{z_N}E^{(T)}]$. Let's consider the secondary variable for sites 1 and 2, $\chi_{1,2}$, with the sites located at positions, $\mathbf{r}_1 = [x_1, y_1, z_1]$ and $\mathbf{r}_2 = [x_2, y_2, z_2]$, and $R_{12} = \|\mathbf{r}_1 - \mathbf{r}_2\|_2$:

$$\begin{aligned} \chi_{1,2}(R_{1,2}) &= \exp[-b_{1,2}R_{1,2}] \\ &= \exp\left[-b_{1,2}((x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2)^{\frac{1}{2}}\right] \end{aligned} \tag{5.21}$$

If we take the partial derivative of $\chi_{1,2}^p$ with respect to the z coordinate of site 1, we get:

$$\begin{aligned}\frac{\partial}{\partial z_1} \chi_{1,2}^p(R_{1,2}) &= \frac{-pb_{1,2}(z_1 - z_2)}{((x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2)^{\frac{1}{2}}} \chi_{1,2}^p(R_{1,2}) \\ &= \frac{-pb_{1,2}(z_1 - z_2)}{R_{1,2}} \chi_{1,2}^p(R_{1,2})\end{aligned}\tag{5.22}$$

We will need to build the design matrix from terms like these. If we take the z_1 partial derivative of a monomial where only one variable depends on the position of site 1, then we get the same result, with the other variables in the product unchanged, for example:

$$\frac{\partial}{\partial z_1} \chi_{1,2}^p \chi_{2,3} \chi_{3,4}^2 \chi_{3,6}^4 = \frac{-pb_{1,2}(z_1 - z_2)}{R_{1,2}} \chi_{1,2}^p \chi_{2,3} \chi_{3,4}^2 \chi_{3,6}^4\tag{5.23}$$

If we take the z_1 partial derivative of a term where multiple variables depend on the position of site 1, we need to use the product rule, for example:

$$\frac{\partial}{\partial z_1} \chi_{1,2}^{p_1} \chi_{1,3}^{p_2} \chi_{4,8}^5 = \left(\frac{-p_1 b_{1,2}(z_1 - z_2)}{R_{1,2}} + \frac{-p_2 b_{1,3}(z_1 - z_3)}{R_{1,3}} \right) \chi_{1,2}^{p_1} \chi_{1,3}^{p_2} \chi_{4,8}^5\tag{5.24}$$

All terms from the partial derivative of the same parent term will have the same linear coefficient, and will need to be summed into the same matrix element. Additionally, different components of the gradient (the $3N$ rows of the design matrix for each training set configuration) will need to be arranged so that each column of the matrix corresponds to one linear coefficient (again, this is determined by the parent terms). Once the design matrix has been constructed, and we have built the reference vector of gradient components that we are fitting to, the fitting procedure should be otherwise unchanged. It should also be noted that when fitting a PIP for an MB-nrg potential (Eq. 5.4), we need to subtract the gradient of the long-range terms, $\nabla(V_{ELEC}^{TTM-nrg} + V_{POL}^{TTM-nrg} + V_{DISP}^{TTM-nrg})$, from the corresponding vector of reference partial derivatives for each training set configuration.

5.1.7 Considerations Regarding Switching Functions

All PIPs in an MB-nrg PEF are multiplied by a switching function, dependent on some geometric coordinate, which smoothly turns off the PIP outside of some threshold value of the coordinate. For example, in the MB-pol PEF, the switching function is equal to 1 for (water-water) O–O distances less than 4.5 Å, and 0 for O–O distances greater than 6.5 Å. However, for PEFs involving monomers such as the Zundel ion (O_2H_5^+) or ethane (C_2H_6), there is no single atom-atom distance that will be acceptable for the switching function. For example, for the

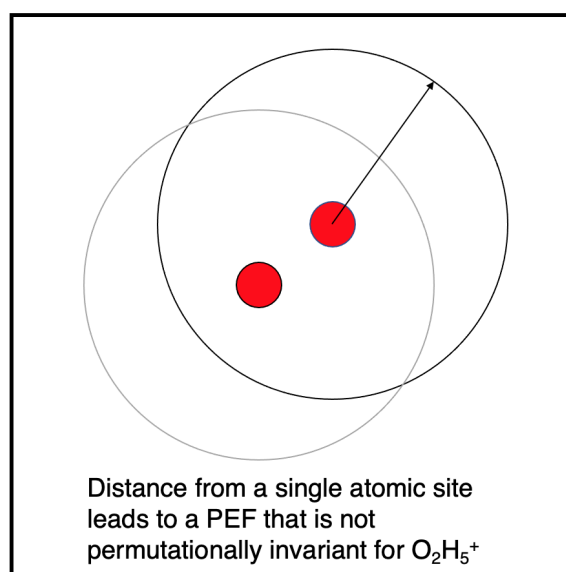


Figure 5.2. The boundary of the outer-cutoff of a switching function dependent on the distance from a single atom. This leads to a PEF that is not permutationally invariant in monomers such as ethane and the Zundel ion.

ethane dimer, if one were to use a single C–C distance for the switching function, then the PEF would not be permutationally invariant, because in regions between the inner- and outer-cutoff distances, the switching function would have different values depending on the permutation of the carbon atoms (Fig. 5.2). Additionally, it is not advisable to use the average of multiple distances in the switching function. For example, if one wanted to develop an ethane-water MB-nrg PEF, use of the average (ethane-water) C–O distance results in the switching function cutoff boundary to be an ellipsoid, rather than the sphere resulting from dependence on a single

distance (Fig. 5.3). Furthermore, the dimensions of the ellipsoid boundary would actually depend on the C–C distance. The best solution is to instead use distances between the geometric centers

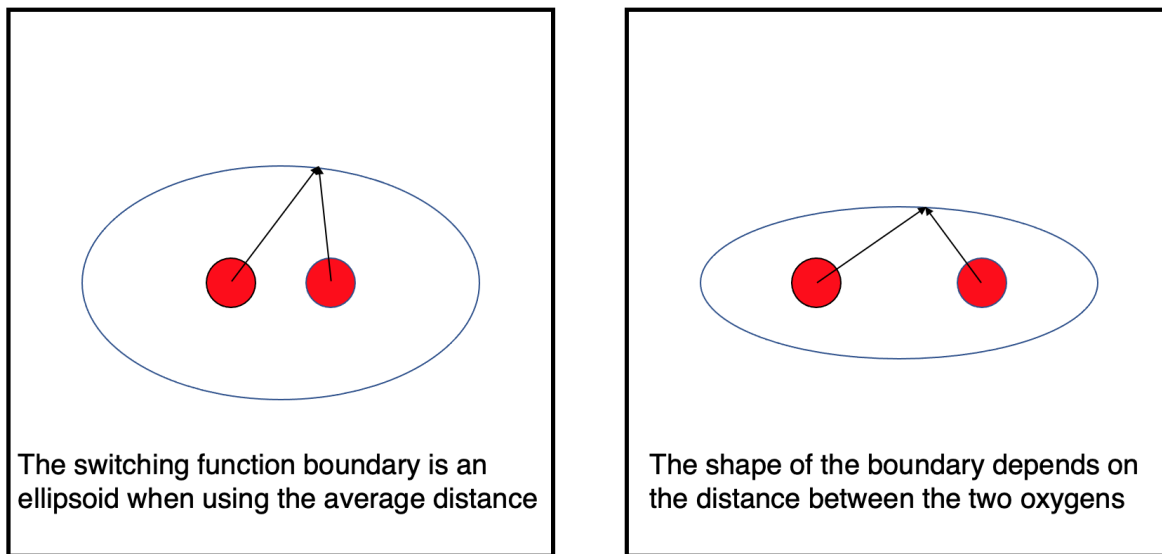


Figure 5.3. The boundary of the outer-cutoff of a switching function dependent on the average distance from two atoms. The average distance results in an ellipsoid-shaped boundary whose dimensions depend on the internal coordinates of the monomer.

of atomic sites in each monomer (Fig. 5.4). In the ethane-water PEF, this would be the distance between the water-O and the midpoint between the ethane carbons. It should be noted that the geometric center is not suited for monomers which can dissociate, since the dissociating fragments will approach, and eventually surpass the switching function outer-cutoff boundary.

5.1.8 Developing MB-nrg PEFs

The typical workflow for PEF development is shown in Fig. 5.5. The two major components needed in PEF development are the model (i.e. the functional form of the PEF) and the training set (the data used to fit the model). We discuss both components in the next two sections.

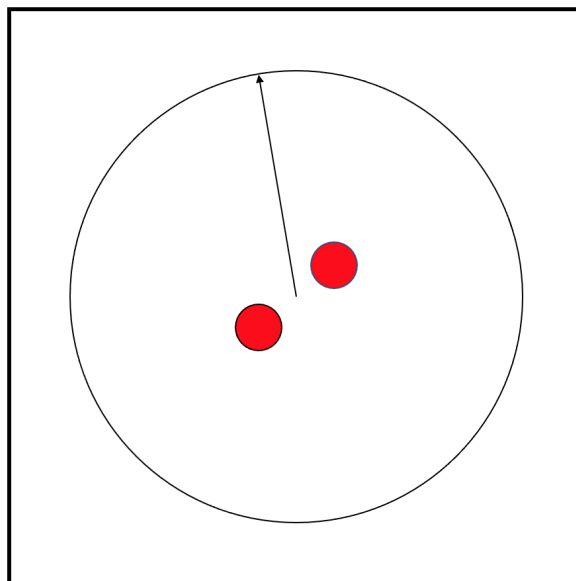


Figure 5.4. The boundary of the outer-cutoff of a switching function dependent on the distance from the geometric center of two atoms. This leads to a PEF that is permutationally invariant (assuming a reasonable choice of atomic sites for the geometric center). The boundary is spherical, and does not depend on internal coordinates of the monomer.

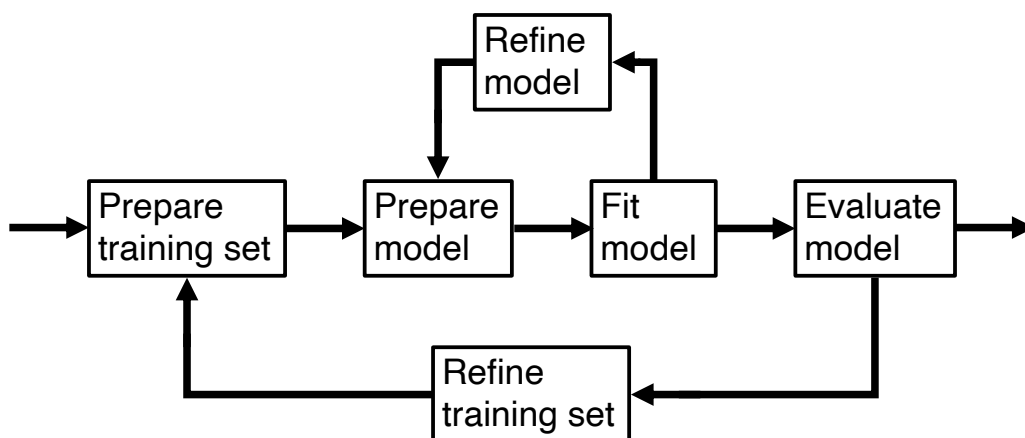


Figure 5.5. PEF development flow chart.

5.1.9 Training Set Preparation and Refinement

In general, a fitting algorithm will attempt to minimize a cost function with respect to the model parameters (for example, see Sec. 5.1.5). At the most basic level, the cost function is the error of the (parameterized) model with respect to the training set data (although the cost function may have additional dependences, for example on the magnitudes of the parameters, as is the case in Tikhonov regression).

When fitting TTM-nrg and MB-nrg PEFs, the training set consists of nuclear configurations and their corresponding reference energies. For example, when fitting a 3B water-water-water PEF, the training set will include some number of $(\text{H}_2\text{O})_3$ configurations paired with their 3B energies, each calculated with the same electronic structure method. The electronic structure method chosen must be cheap enough that a sufficiently large number of energies may be calculated with the available resources, while also accurate enough that MD simulations using the PEF will be a good representation of reality. The electronic structure method must be uniform within a training set, but different methods can be used for complementary PEFs. For example, one might use a more expensive method for the 2B PEF training set, and a cheaper method for the 3B training set (Sec. 3.1.4).

Unfortunately, the full configuration space of a many-body system is enormous. For N atomic sites (in three spatial dimensions), there are $3N$ nuclear degrees of freedom. A uniform, cubic grid of M points in each dimension will have M^{3N} total configurations, so we need to be more selective with the configurations included in the training set. A few examples of ways to generate training set configurations include:

1. extracting configurations from MD simulations
2. normal mode sampling
3. regular or random scans (including rotations)

Usually, a combination of different training set generation methods is the safest choice, since it avoids biasing the training set through the use of a single source of configurations. While

3N-dimensional configuration space can be a bit intimidating, there are some general principles that can guide us through the training set generation process.

The most important consideration when generating a training set is to obtain configurations that are representative of the subset of configuration space that will be sampled in our MD simulations. For example, a PEF used in high pressure simulations must be fitted to a training set which includes configurations with sufficiently small intermolecular separations (we will return to this point when discussing Fig. 5.7). Extracting configurations from simulations is the most direct way to satisfy this consideration, but a combination of normal mode sampling and scans can be used as well.

In addition to including configurations within the subset of configuration space that is sampled during simulations, it is equally important to include configurations just-beyond this subset, within the boundary between the configurations that will be sampled and the (higher-energy) configurations that should never be sampled. This is because we never want our PEF to extrapolate beyond the training data (we only want interpolation). This is especially important when fitting MB-nrg PEFs because the PIPs are flexible enough that their behavior beyond the boundaries of the training set data will likely be completely unphysical. Given that these configurations exist beyond the subset of configuration space sampled in simulations, we will not be able to generate them from extracting configurations from MD, without increasing the temperature or pressure, or adding some external bias potential. Alternatively, we can perform scans which include close-range configurations.

Since an MB-nrg PEF can have hundreds to thousands of flexible parameters, a large number of training set configurations will be needed, so a combination of different training set generation methods is likely the best option. TTM-nrg PEFs, on the other hand, have relatively few flexible parameters (Sec. 5.1.1), so a few physically-motivated scans of rigid monomer units will often be sufficient. It is important to note, however, that a training set containing rigid monomer units will not be representative of the important subset of configuration space when the 1B PES of the relevant monomers have slow degrees of freedom (e.g. important dihedral

angles) with small potential energy barriers, for example the Zundel ion (see Sec. 6.1).

The importance of a training set that is representative of the configurations sampled in simulations can be seen through a comparison between two TTM-nrg PEFs for acetonitrile-acetonitrile interactions which were fitted to different training sets. The first PEF was fitted to 16 radial scans with rigid monomers. Eight of the 16 scans included configurations with negative (attractive) 2B energies, while the other eight scans included exclusively repulsive configurations. The second PEF was fitted to the eight attractive scans, alone. Two examples of attractive and repulsive scans are shown in Fig. 5.6. While the PEF fitted to all 16 scans will be more accurate across the entire configuration space, the PEF fitted only to the attractive scans will be more accurate for configurations sampled in condensed phase simulations in ambient conditions.

This point is illustrated in Fig. 5.7, which shows a comparison between three PEFs for acetonitrile-acetonitrile interactions showing the density over time in NPT simulations of pure acetonitrile (panels a, c, and e) and the test set correlation plots for each PEF (the test set was generated with radial scans of rigid monomers with random relative orientations). Panels a and b correspond to an MB-nrg PEF, panels c and d correspond to the TTM-nrg PEF fitted only to the attractive scans, and panels e and f correspond to the TTM-nrg PEF fitted to both attractive and repulsive scans. A comparison of panels d and f shows that fitting to all 16 scans results in a more balanced PEF which is relatively accurate in the high-energy configurations. However, panels c and e show that the PEF fitted only to attractive scans is more effective at reproducing the experimental density of liquid acetonitrile, despite being less balanced than the other PEF. It should also be noted that the MB-nrg density (Fig. 5.7a) is slightly overestimated, likely due to missing repulsive 3B interactions, which might require a 3B PIP correction term (Sec. 5.1.2).

The model evaluation and training set refinement steps are critical to ensuring that we have developed a model PEF that is sufficiently accurate. Once the model has been fitted, we can attempt to use it in an MD simulation under the thermodynamic conditions (temperature, pressure, etc.) that we are interested in studying, and extract test set configurations which we will use to evaluate the model. The test set configurations must correspond to the PEF that we are

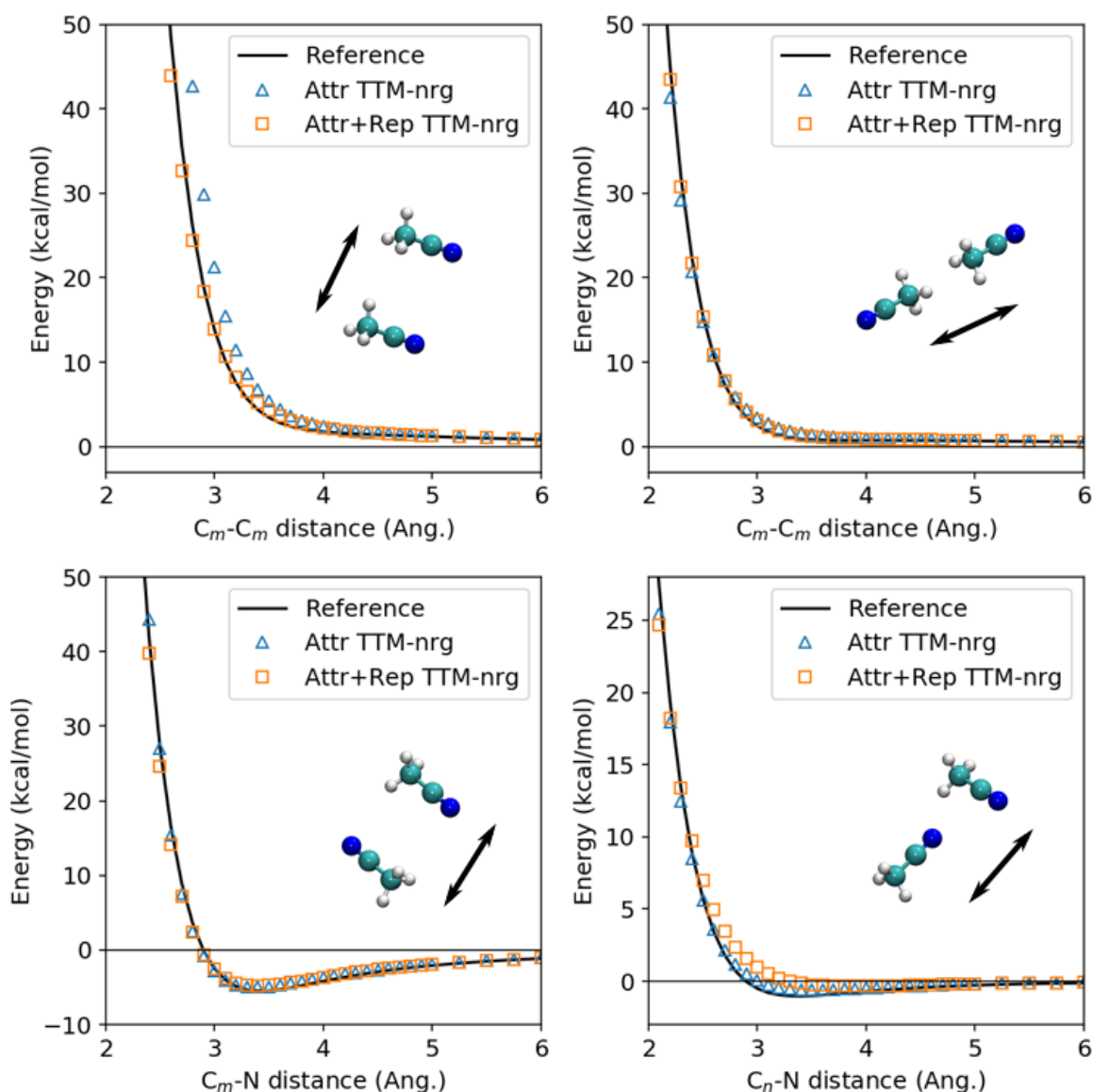


Figure 5.6. Repulsive (panels a, b) and attractive (panels c, d) acetonitrile-acetonitrile scans used in TTM-nrg training sets.

evaluating, i.e. 2-mer configurations for a 2B PEF, 3-mers for a 3B PEF, etc.. If the simulation is unstable, indicating flaws in the model PEF, we must obtain test set configurations leading up to the instability, otherwise, we may obtain test set configurations from random sampling after some equilibration time. In any case, after obtaining the new test set, we need to calculate reference energies, and compare them to the energies predicted by the model. If the model does

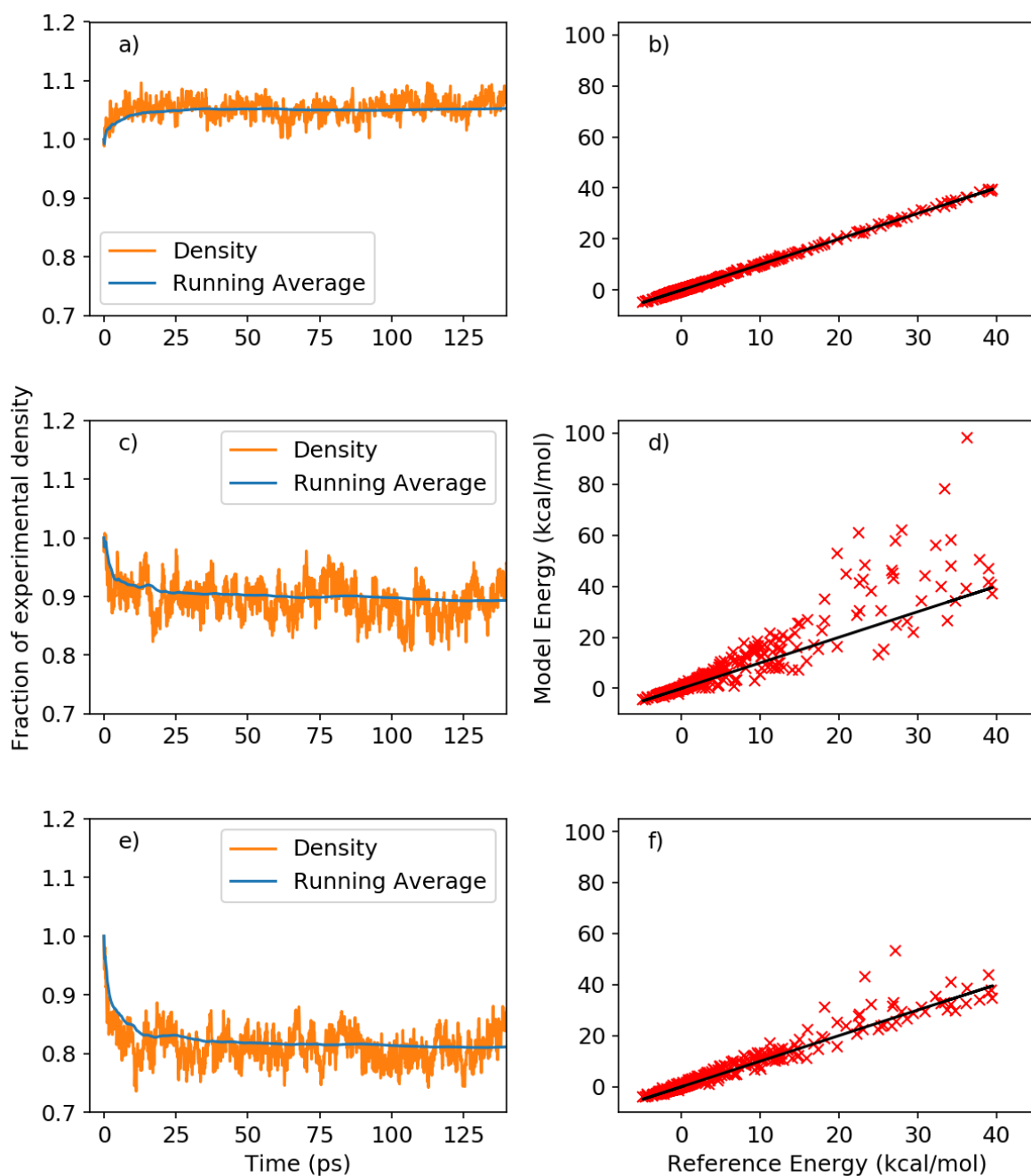


Figure 5.7. Comparison of NPT densities over time and test set correlation plots between three model PEFs for acetonitrile-acetonitrile interactions. MB-nrg (panels a, b), TTM-nrg fitted to attractive scans (panels c, d), and TTM-nrg fitted to attractive and repulsive scans (panels e, f).

not achieve the desired accuracy, we can add the test set into the original training set, and refit the model. If the PEF does not improve after several iterations, we may need to refine the model.

5.1.10 Model Preparation and Refinement: Monomial Connectivity Graphs

Since the functional form of a TTM-nrg PEF is fixed (only dependent on the system), the model cannot be refined, so the corresponding path in the flow chart (Fig. 5.5) is never used. For PEFs involving PIPs (e.g. 1B PEFs and MB-nrg PEFs), one may increase the complexity of the model (for example by increasing the PIP degree) to achieve higher accuracy, or decrease the complexity to avoid overfitting. Additionally one can use monomial filters, which allow the model developers to increase the polynomial degree, while pruning the higher-degree monomials that are unnecessary or superfluous.

The various (symmetrized) monomial terms can be categorized based on their *connectivity graphs*. Since the PIPs used here depend on all atom-atom pairwise distances, the nodes of the connectivity graphs correspond to the atomic sites, and the edges correspond to the variables (atom-atom distances) included in a given monomial term. The connectivity graph is a representation of the potential energy coupling between configurational degrees of freedom due to a given monomial. Therefore, the idea of monomial connectivity graphs can be used to design PIP filters that eliminate monomials involving degrees of freedom that are weakly coupled or not coupled at all.

An example application of the filters based on monomial connectivity graphs is in the development of the acetonitrile-acetonitrile 2B PIP. Each acetonitrile is treated as one methyl group fragment bound to one nitrile group fragment.¹ From this perspective, it seems at least plausible that, for example, the intrafragment H–H distance within the methyl group of molecule 1 is not strongly coupled to the intermolecular interactions between the molecule 1 nitrile group and molecule 2, unless the monomial also contains intermolecular variables between the molecule

¹This fragment-based filter approach (without monomial connectivity graphs) was initially developed and implemented by Ethan Bull-Vulpe, Chanin Tangtartharakul, Marc Riera, and Francesco Paesani.

1 methyl group and molecule 2 (Fig. 5.8). This conditional statement can be summarized in terms of the connectivity of the relevant monomial connectivity graphs, such that monomials whose graphs have no path between the molecule 1 H sites and the molecule 1 N site, for example.² The connectivity is more clear (pictorially) after eliminating nodes that are not included in the monomial graph under question (Fig. 5.9)

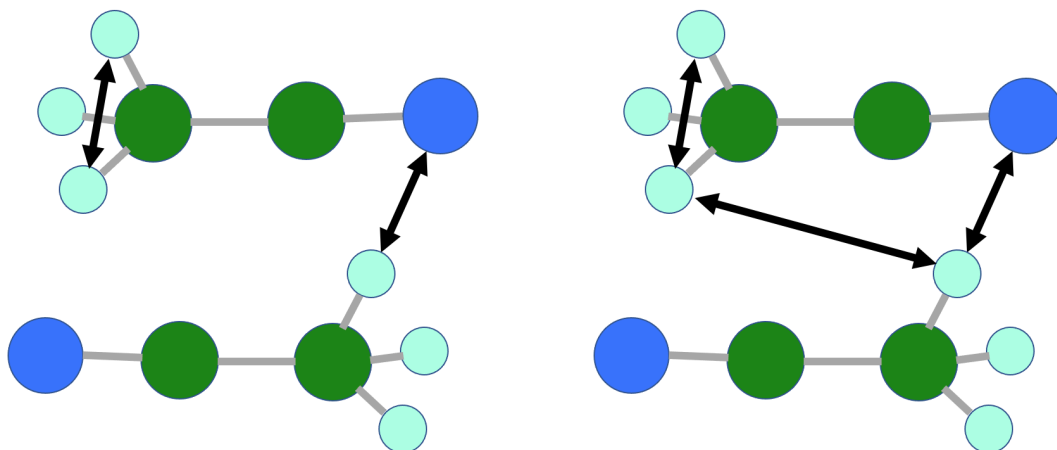


Figure 5.8. Examples of acetonitrile-acetonitrile 2B PIP monomials represented by their monomial connectivity graphs. The monomial represented by the graph on the left-hand side includes an intrafragment (methyl group) H–H variable on molecule 1, and an intermolecular variable between the molecule 1 nitrogen and a molecule 2 hydrogen. The monomial represented by the graph on the right-hand side includes the same two types of variables, but also includes a variable between the molecule 1 methyl group hydrogen and the molecule 2 hydrogen. The left graph is disconnected, and is therefore likely to be less important than the connected graph on the right.

The filter depicted in Figs. 5.8 and 5.9 can be easily defined without any notion of a monomial connectivity graph. However, Figs. 5.10 and 5.11 show an example of a pair of monomials that would be difficult to distinguish without appealing to the idea of monomial connectivity. All of the methyl H sites on either molecule are identical, so filters that only distinguish monomials based on atomic site types (e.g. methyl H, methyl C, nitrile C, etc.), or

²Our implementation generalizes the notion of connectivity through the use of shortest paths. Rather than simply categorizing monomials based on whether there is or is not a path between two chosen nodes, we can further categorize monomials based on the smallest number of edges (PIP variables) that separate the two nodes. We use the Floyd-Warshall algorithm to obtain the lengths of the shortest paths between each pair of nodes for each PIP monomial term.

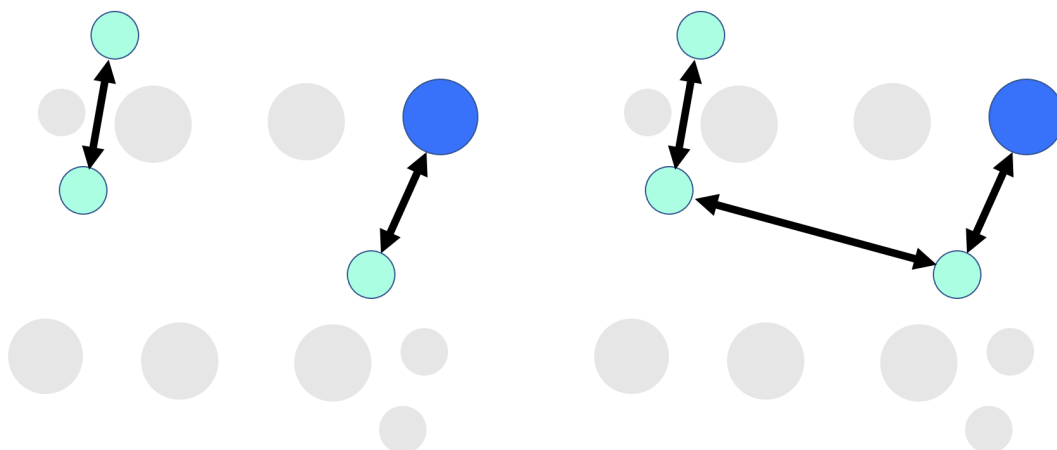


Figure 5.9. Acetonitrile-acetonitrile 2B PIP monomial connectivity graphs from Fig. 5.8 with all non-relevant atomic sites grayed out.

even variable types (intermolecular C–C, intramolecular C–C, intramolecular C–H, etc.) will be insufficient to distinguish the monomials in Figs. 5.10 and 5.11. When applying the concept of monomial connectivity, this pair of monomials is easily distinguished. This type of filter is useful in filtering polynomials involving the Zundel ion, O_2H_5 , since all five Zundel H sites are identical, which leads to many monomial terms that are difficult to distinguish by other means.

Table 5.1 shows an example comparison between MB-nrg PIP filters for the 2B $\text{H}_3\text{O}^+ - \text{H}_2\text{O}$ (non-reactive) diabatic state PEF described in Sec. 7.2. Here, “wRMSD” refers to the weighted (training set) RMSD, “Low RMSD” and “Low Max” refer to the RMSD and maximum error in training set configurations with binding energies 25 kcal/mol above the minimum, “Test RMSD” refers to the test set RMSD, and “Number of Terms” refers to the number of monomials defining a given PIP. “3^o” and “4^o” refer to 3rd and 4th degree PIPs, respectively. Each entry in the table represents the most accurate of 100 fits for a given filter, with accuracy defined by the test (set) RMSD. The “Intra-2” PIPs correspond to filtering 4th degree monomials containing 2+ degree intramolecular variables (4^o Intra-2 is a good proxy for the accuracy limit of 4^o PIPs). The “MB-pol” filtered PIP corresponds to filtering 4th degree monomials containing intramolecular variables whose degrees do not add up to 2. The “Graph”

filtered PIP corresponds to filtering 4th degree monomials by the maximum path length within their monomial connectivity graphs, with a maximum path length of 3 connecting intramolecular variables, and a maximum path length of 4 connecting intramolecular variables to intermolecular variables, or connecting intermolecular variables with each other. Comparing the test set RMSD and the number of monomials between each PIP, it can be seen that the use of filters defined in terms of monomial connectivity graphs can allow one to tune the accuracy of the PIP in a controlled manner without introducing too many superfluous monomials.

In the previous examples (Figs. 5.8 and 5.9 and Figs. 5.10 and 5.11), it was implicit that the disconnected graphs were less relevant in describing intermolecular interactions, and were therefore targeted for removal by graph connectivity filters. However, one can imagine situations in which the disconnected graphs might be more important. For example, when describing the interactions between two or more large polymers (for example proteins), disconnected monomials can be helpful in describing the large-scale spatial relationships between the polymers, since they are more efficient with the number of involved atomic sites with respect to the monomial degree. For example, for a fourth degree polynomial, the most ‘inefficient’ types of monomials are fourth degree in a single variable (a single atom-atom pair), while the most ‘efficient’ types of monomials are fully disconnected, and involve four pairs of distinct atomic sites.

Table 5.1. Example comparison between PIPs defined by different filters.

PIP	wRMSD	Low RMSD	Low Max	Test RMSD	Number of Terms
3 ^o Intra-2	0.25	0.18	1.59	0.41	583
4 ^o Intra-2	0.16	0.11	1.02	0.26	3372
4 ^o MB-pol Filters	0.22	0.15	1.46	0.34	1496
4 ^o Graph Filters	0.18	0.12	1.20	0.29	1724

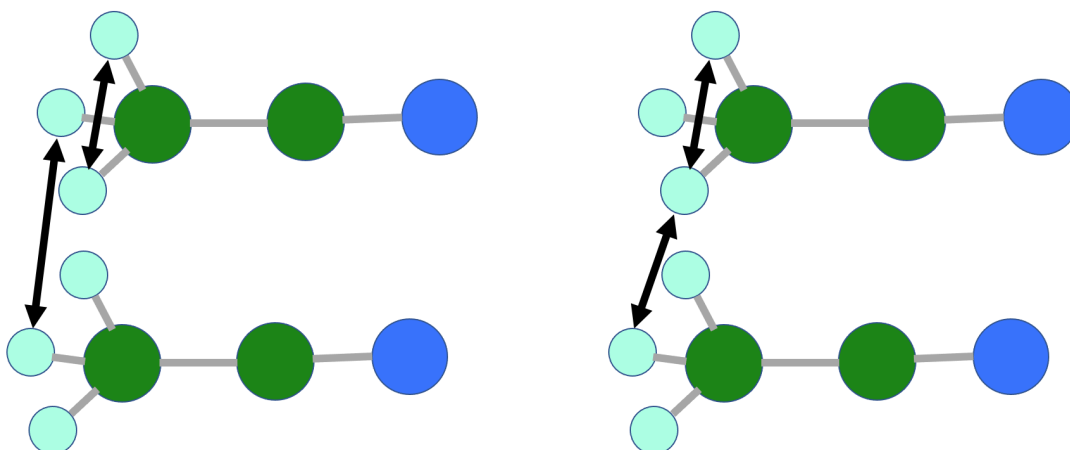


Figure 5.10. Examples of two acetonitrile-acetonitrile 2B PIP monomials, represented by their monomial connectivity graphs, that cannot be distinguished without the use of monomial connectivity graphs due to the permutational invariance of methyl hydrogens on the same fragment. The monomial shown on the left-hand side is disconnected, and is therefore likely to be less important than the monomial shown on the right-hand side whose monomial connectivity graph is connected.

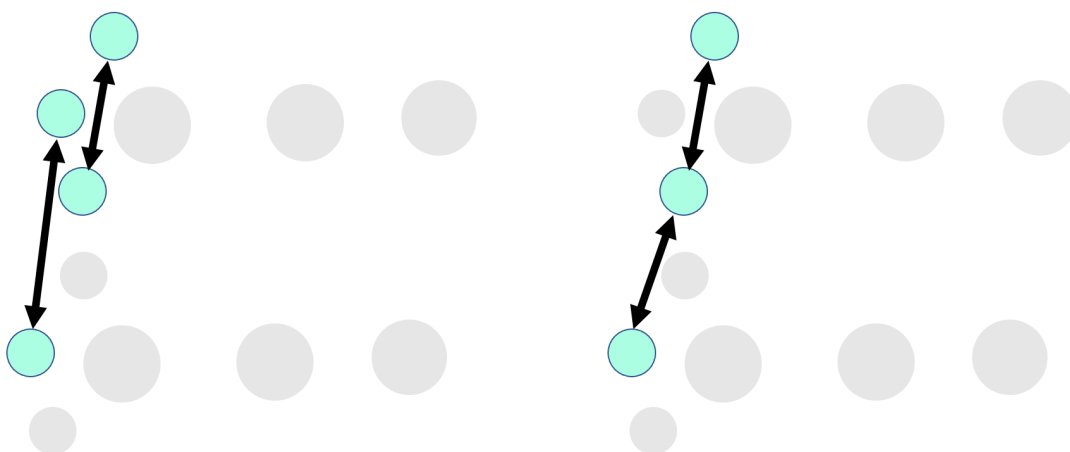


Figure 5.11. Acetonitrile-acetonitrile 2B PIP monomial connectivity graphs from Fig. 5.10 with all non-relevant atomic sites grayed out.

5.2 Evaluating Alkali Ion–Water PEFs

An example of a system for which one can apply the MB-nrg methodology is an alkali ion dissolved in water. In the next section, we evaluate the accuracy of each PEF in optimized 2-mer

and 3-mer $M^+(H_2O)$ and $M^+(H_2O)_2$ ($M = Li, Na, K, Rb, Cs$) clusters, as well as simple radial scans. The three PEFs for alkali ion–water interactions, AMOEBA, TTM-nrg, and MB-nrg are described in detail in Sec. C.1.

We first evaluate the accuracy of each of the PEFs discussed here in reproducing the 2B energies of the $M^+(H_2O)$ clusters and the 3B energies of the $M^+(H_2O)_2$ clusters discussed in Sec. 3.4. Deviations from the CCSD(T) reference 2B and 3B energies for all tested PEFs are shown in Figs. 3.26 and 3.27, respectively.

Additionally, we evaluated the ability of the PEFs to reproduce CCSD(T) reference 2B energies along simple rigid scans. 2B scans of each alkali ion along the water H–O–H bisector were performed by optimizing the water monomer at the RI-MP2[196, 197] level of theory with the aug-cc-pVTZ basis set, initially placing each ion, except for Li^+ , 1.5 Å away from the water-oxygen (Li^+ was initially placed 1.25 Å away from the oxygen), coplanar with the water molecule, on the opposite side of the water’s hydrogen atoms (Fig. 5.12). Each ion was displaced along the water bisector (in the plane) up to 5.0 Å from the oxygen in 0.1 Å increments.

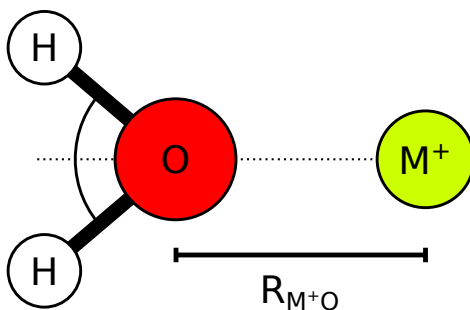


Figure 5.12. Schematic representation of the geometry used in the scans along the R_{M+O} distance carried out for each of the four $M^+(H_2O)$ dimers, with $M = Li, Na, K, Rb, Cs$. The ion remains in the plane of the water molecule.

5.2.1 2B Energies of Alkali Ion–Water Dimers

AMOEBA underestimates the strength of the interactions in all alkali ion–water dimers, deviating from the CCSD(T) 2B energies by at most 2.3 kcal/mol in $Rb^+(H_2O)$, and by as little as 0.3 kcal/mol in $K^+(H_2O)$. AMOEBA is among the least accurate models for $Li^+(H_2O)$,

$\text{Rb}^+(\text{H}_2\text{O})$, and $\text{Cs}^+(\text{H}_2\text{O})$, although it outperforms revPBE for all dimers. We point out that AMOEBA was parameterized to reproduce bulk properties of aqueous solutions, so inaccuracy in gas phase clusters is not unexpected. TTM-nrg also underestimates the interaction strength in $\text{Rb}^+(\text{H}_2\text{O})$, and $\text{Cs}^+(\text{H}_2\text{O})$, having errors of 1.4 and 1.6 kcal/mol, respectively. TTM-nrg and AMOEBA exhibits similar accuracy for $\text{Cs}^+(\text{H}_2\text{O})$ and $\text{Na}^+(\text{H}_2\text{O})$, but AMOEBA is more accurate than TTM-nrg for $\text{K}^+(\text{H}_2\text{O})$. Interestingly, TTM-nrg shows negative errors in $\text{Li}^+(\text{H}_2\text{O})$ and $\text{Na}^+(\text{H}_2\text{O})$, but positive errors for all other dimers. Among all tested analytic PEFs and XC-functionals, MB-nrg is, on average, the most accurate model for all alkali ion–water dimers, displaying the largest deviations (-0.1 kcal/mol) from the reference CCSD(T) values for $\text{Li}^+(\text{H}_2\text{O})$ and $\text{Na}^+(\text{H}_2\text{O})$.

5.2.2 3B Energies of Alkali Ion–Water Trimers

Among the PEFs, TTM-nrg shows very large, positive 3B energy errors for the $\text{Li}^+(\text{H}_2\text{O})_2$ and $\text{Na}^+(\text{H}_2\text{O})_2$ trimers in their minimum-energy geometries, deviating from the CCSD(T) reference by 2.88 and 0.92 kcal/mol, respectively. Given that the TTM-nrg ion–water–water 3B energy is entirely due to classical POL, these large errors may suggest that close-range quantum effects such as CT may be important for reproducing these 3B interactions. However, the AMOEBA PEFs yields 3B energies for $\text{Li}^+(\text{H}_2\text{O})_2$ and $\text{Na}^+(\text{H}_2\text{O})_2$ that deviate from the CCSD(T) 3B energy by -0.34 and 0.01 kcal/mol, respectively, despite being composed solely of classical POL, themselves. Looking at the energy decomposition analysis of 2B interaction energies of $\text{Li}^+(\text{H}_2\text{O})$ and $\text{Na}^+(\text{H}_2\text{O})$ dimers, as well as decompositions of 3B energies of $\text{Li}^+(\text{H}_2\text{O})_2$ and $\text{Na}^+(\text{H}_2\text{O})_2$ trimers, it appears that AMOEBA tends to predict POL energies that are similar to ALMO-EDA POL energies of XC-functionals, while TTM-nrg and MB-nrg (classical) POL energies tend to be larger in magnitude than the corresponding ALMO-EDA values. Note that there is widespread agreement between the tested XC-functionals with respect to their ALMO-EDA POL energies. This difference in POL energies between AMOEBA and TTM-nrg for Li^+ and Na^+ dimers and trimers could be due to a combination of factors, including

differences in atom-centered dipole polarizabilities, differences in permanent charges, the use of permanent dipoles and quadrupoles in AMOEBA, differences in Thole damping parameters, or in the different damping schemes used by each model (see Sec. C.1). These differences in POL energies between the two sets of models may be accentuated in the Li^+ and Na^+ dimers and trimers due to the small distances between the ion and the water molecule(s) in these structures, compared to those in the K^+ , Rb^+ , and Cs^+ dimers and trimers. We note, however, that TTM-nrg was found to predict, on average, better 3B energies for halide–water trimers than AMOEBA,[120] indicating that there may be some amount of error cancellation, for example in the non-negligible 3B ALMO-EDA CT energy found in the global minimum energy $\text{Li}^+(\text{H}_2\text{O})_2$ trimer, as shown in the Supporting Information of Ref. [119] (note that non-negligible 3B ALMO-EDA CT energies were also found in the halide–water trimers[120]).

Irrespective of the specific underlying cause leading to the large deviations in the TTM-nrg 3B interaction energies of the minimum energy $\text{Li}^+(\text{H}_2\text{O})_2$ and $\text{Na}^+(\text{H}_2\text{O})_2$ trimers, Fig. 3.27 demonstrates that the 3B MB-nrg PIPs (Secs. 5.1.2 and C.1.3) provide sufficiently accurate close-range corrections to the underlying classical model, which for 3B and higher-body energy terms, is identical to those of TTM-nrg.

5.2.3 2B Rigid Scans

To assess the ability of AMOEBA, TTM-nrg, and MB-nrg to describe alkali ion–water interactions, Fig. 5.13 shows comparisons with CCSD(T) interaction energies calculated for radial scans obtained by moving each alkali ion along the direction of the water H–O–H bisector, as described above and depicted in Fig. 5.12. The most apparent feature seen here is the tendency of AMOEBA to predict interaction energies that are too repulsive at close range for all dimers except for $\text{K}^+(\text{H}_2\text{O})$, for which AMOEBA is slightly too repulsive near the minimum-energy distance (by less than 0.5 kcal/mol), and becomes slightly too attractive at close range (by around -2 kcal/mol at an $\text{K}^+ - \text{O}$ distance of 2 Å. For all other dimers, but most significantly for $\text{Rb}^+(\text{H}_2\text{O})$ and $\text{Cs}^+(\text{H}_2\text{O})$, at close range, AMOEBA becomes too repulsive at too large

of an M^+-O distance. This artificially-steep repulsive wall results in overestimation of the minimum-energy distances by around 0.2 Å in both $Rb^+(H_2O)$ and $Cs^+(H_2O)$. The AMOEBA 2B energy was also found to diverge from CCSD(T)-F12b at close range in halide–water dimers scanned along the water O–H axis, producing 2B curves that were too repulsive (and too steep) at too large of an $X^- - O$ distance, leading to overestimated minimum-energy distances.[120] However, while AMOEBA was additionally found to predict minimum energies that are too attractive in the halide–water dimer scans,[120] we find here that AMOEBA predicts minimum energies that are too repulsive in the alkali ion–water scans. For all dimers, the AMOEBA interaction energy converges to the reference CCSD(T) interaction energies at large $M^+ - O$ distances where the permanent electrostatics are the dominant interaction.[151] We again point out that AMOEBA was parameterized to reproduce bulk properties of aqueous solutions, so inaccuracy in these gas phase scans is not unexpected.

TTM-nrg performs better than AMOEBA on average in all dimer scans except for parts of the $Na^+(H_2O)$ and $K^+(H_2O)$ scans. It should be noted, however, that despite performing better than AMOEBA on average, TTM-nrg deviates from the CCSD(T) reference energies by around 2 kcal/mol or more in some configurations at close range in each dimer scan. As is the case for AMOEBA, TTM-nrg appears to converge to the CCSD(T) curves at long range. This agreement of AMOEBA and TTM-nrg with CCSD(T) at long range is consistent with the fact that both models use similar representations of the ELEC energy, which is the dominant interaction at large ion–water separations.[151] While TTM-nrg tends to find the correct minimum-energy distances along each scan, the TTM-nrg 2B interaction energies at the minima differ from the corresponding CCSD(T) values by more than 1 kcal/mol in $K^+(H_2O)$, $Rb^+(H_2O)$ and $Cs^+(H_2O)$. Interestingly, this behavior was not apparent in the the halide–water TTM-nrg scans, which were slightly too repulsive in the close range repulsive region, but were otherwise quite accurate at all other ion–water separations except for $F^-(H_2O)$, for which TTM-nrg predicts that the minimum energy is ~ 1.3 kcal/mol lower than the CCSD(T) value.[120]

MB-nrg reproduces nearly quantitatively the CCSD(T) 2B energies along each scan.

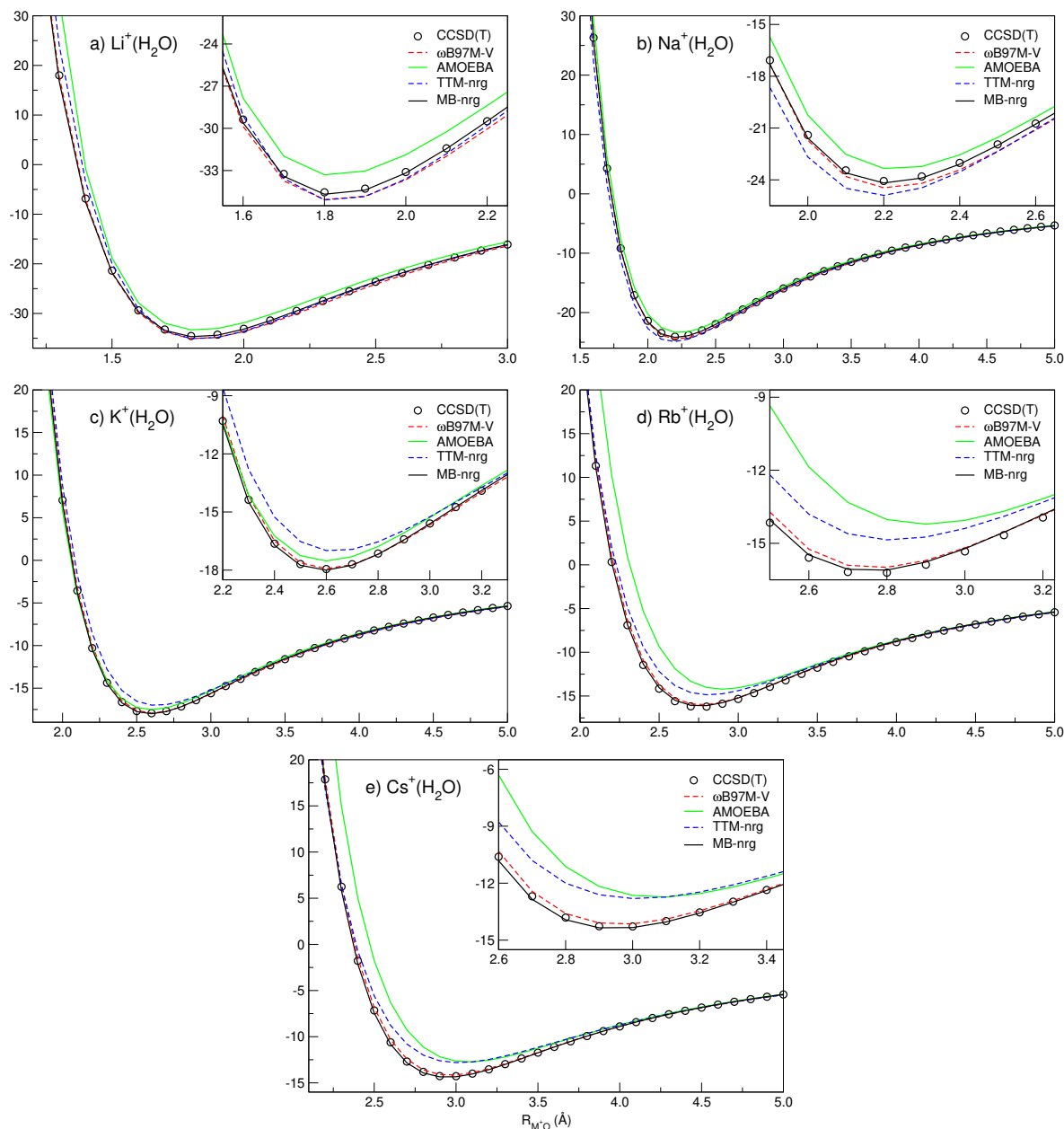


Figure 5.13. Comparisons between CCSD(T) (open circles) and $\omega B97M-V$ (dotted red), AMOEBA (solid green), TTM-nrg (dotted blue), and MB-nrg (solid black) interaction energies calculated for radial scans of $M^+(H_2O)$ dimers, with $M = Li$ (a), Na (b), K (c), Rb (d), Cs (e), in which each alkali ion is displaced along the water H–O–H bisector, on the opposite side of the water-hydrogens, in the water-plane, as described above.

The largest deviations of MB-nrg relative to CCSD(T) occur within the repulsive regions in each scan but the deviations remain within ~ 0.25 kcal/mol at medium and long ranges. Given

that the MB-nrg differs from TTM-nrg only by the replacement of the TTM-nrg Born-Mayer repulsive potential with PIPs (Section C.1.3), the higher accuracy of MB-nrg than TTM-nrg in all configurations along each scan illustrates the importance of including effective quantum corrections to the TTM-nrg classical interactions at close range, keeping in mind that TTM-nrg reproduces the CCSD(T) 2B energies in the long range regime of each scan. Additional dimer scans are reported in the SI, which show that, independently of the water orientation, MB-nrg successfully reproduces CCSD(T)/CBS interaction energies of alkali ion–water dimers, even in close range configurations where either TTM-nrg or AMOEBA deviate from CCSD(T), further demonstrating the importance of the MB-nrg PIPs in correcting the deficiencies of classical descriptions of molecular interactions.

Fig. B.2 shows that the ELEC energies of the minimum energy alkali ion–water dimers calculated with each of the tested PEFs are similar to those calculated with ALMO-EDA. This is in contrast with the ELEC energies of halide–water dimers calculated with the same PEFs, which were consistently less attractive than those calculated with ALMO-EDA. Panels a, c, and e of Fig. 5.14 shows that the PEF ELEC energies converge to the ALMO-EDA (ω B97M-V) ELEC energies at long range, but remain less attractive than the ALMO-EDA ELEC energies as the water–oxygen–ion distance decreases. This divergence between the two sets of ELEC energies is due to the lack of attractive charge penetration interactions in the PEFs, which employ (damped) partial atomic charges. The fact that PEF ELEC energies agree with ALMO-EDA ELEC energies in the minimum energy alkali ion–water dimers indicates that these structures have oxygen–ion distances that lie outside the onset of charge penetration effects. Additionally, it should be noted that a divergence in the ELEC energies are not seen in the $\text{Li}^+(\text{H}_2\text{O})$ scans because charge penetration effects occur at smaller Li^+ –oxygen distances due to the small radius of Li^+ .

The only difference between decompositions of the TTM-nrg and MB-nrg 2B interaction energies is in their Pauli+Disp terms, for which TTM-nrg uses pairwise Born-Mayer functions, while MB-nrg uses the 2B PIPs (Secs. 5.1.2 and C.1.3). As seen in Fig. B.2, MB-nrg Pauli+Disp

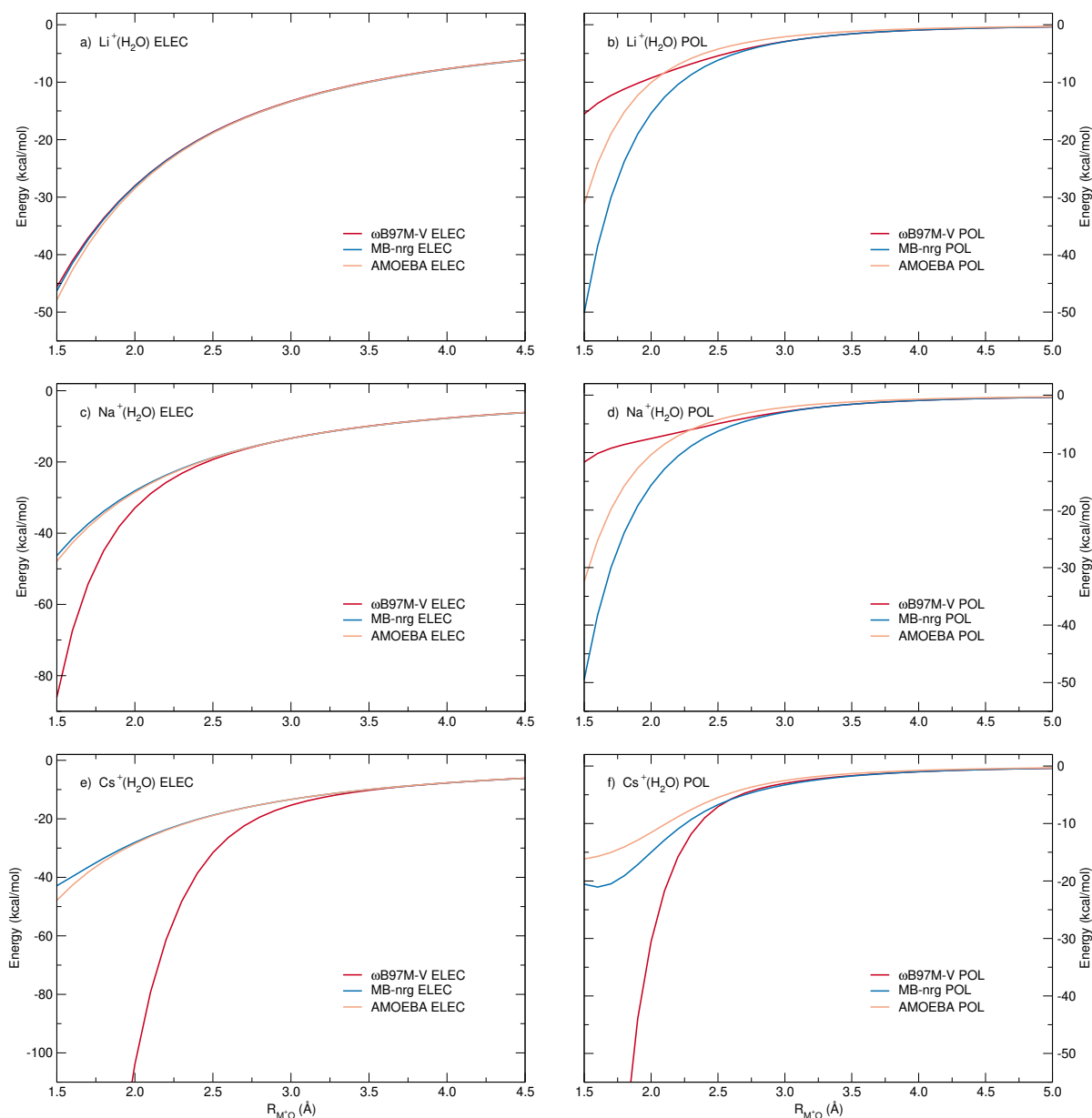


Figure 5.14. ELEC and POL energies calculated with each of the three sets of PEFs analyzed in this study, and the ω B97M-V functional (from ALMO-EDA) along the scans described above and depicted in Fig. 5.12 for the $\text{Li}^+(\text{H}_2\text{O})$ (panels a and b), $\text{Na}^+(\text{H}_2\text{O})$ (panels c and d), and $\text{Cs}^+(\text{H}_2\text{O})$ (panels e and f) dimers.

energies are more repulsive than the TTM-nrg Pauli+Disp energies in $\text{Li}^+(\text{H}_2\text{O})$ and $\text{Na}^+(\text{H}_2\text{O})$, but are less repulsive in the other three dimers. Given that MB-nrg predicts 2B interaction energies that are on average the most consistent with the CCSD(T) reference 2B energies,

and considering the agreement in high-resolution vibrational spectra of small alkali ion–water clusters[167] and in the EXAFS spectra of Cs^+ in water[168] calculated from MB-nrg simulations compared with those measured in experiments, it is clear that the PIPs provide accurate quantum corrections to the underlying long-range classical (TTM-nrg) potential.

5.3 Acknowledgements

Section 5.2 is almost entirely a reprint of the material as they appear in “Nature of Alkali–Water Interactions: Insights from Many-Body Representations and Density Functional Theory II” *J. Chem. Theory Comput.*, **2020**, 16 (5), 3055–3072.

Chapter 6

Molecular Dynamics: Simulations of The Zundel Ion in Acetonitrile

Aqueous acid/base chemistry plays a key role in many biological processes and technological applications.[198, 199, 200, 201, 202, 203, 204, 205] Accordingly, the excess proton in water has been widely studied both experimentally,[206, 207, 208, 209, 210, 211, 212, 213, 214, 115, 215, 116, 117, 216] and theoretically.[217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241] Protonated water clusters in the gas phase offer another perspective on the excess proton in water, and have been studied both experimentally,[242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253] and theoretically.[254, 255, 256, 257, 258, 259, 210, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 75, 4, 270] Gas phase clusters have fewer degrees of freedom than bulk systems, which usually makes gas phase measurements easier interpret, and calculations less costly, however, the finite size of clusters raises questions about extending results to bulk phase solutions. Despite these efforts, the structure of the excess proton in water and the mechanisms of aqueous proton transport are still debated.[217, 218, 219, 221, 207, 230, 209, 210, 208, 212, 232, 213, 234, 235, 237, 215, 115, 238, 116, 117, 216]

There are three widely-accepted fundamental protonated water units, namely the hydronium ion, H_3O^+ , in which an excess proton is localized on a single water molecule, the Eigen ion,[109, 110, 111] O_4H_9^+ , in which a hydronium ion donates one hydrogen

bond to three surrounding water molecules, and the Zundel ion,[106, 107, 108] O_2H_5^+ , in which the excess proton is shared between two water molecules. There is currently no consensus in the literature regarding the dominant structure of the excess proton in liquid water at room temperature, with some results suggesting that an Eigen-type structure is dominant,[224, 207, 230, 210, 232, 215] and other results suggesting that a Zundel-type structure is dominant,[212, 213, 115, 116, 117, 216, 227] although it is generally accepted that the true structure is some dynamically fluctuating combination of these two limiting motifs.[217, 221, 222, 224, 226, 209, 212, 232, 233, 115, 116, 117, 216, 227, 239]

In either case, Zundel-type structures are important to consider given that aqueous proton transfer (PT) must involve the transfer of the excess proton between two water molecules, although recent spectroscopic experiments provide strong evidence that the (distorted) Zundel ion exists as a long-lived species in aqueous solution,[115, 117, 216] rather than simply the transition state during PT events. Given its importance in aqueous PT, and possibly in the solvation structure of the excess proton in water, several studies used ultrafast pump-probe vibrational spectroscopy to study the Zundel ion solvated in liquid acetonitrile.[271, 2, 272, 273] Acetonitrile was chosen as a solvent because it is polar and aprotic, allowing for the dissociation of strong acids without participating in PT, in principle yielding vibrational spectra that are not complicated with ultrafast fluctuations of the hydrogen bonding structures and long-range proton transport through the solvent, as occurs in aqueous solutions.[271] These experiments, supplemented with QM/MM simulations,[2, 273] investigated the role of solvent fluctuations on modulating the Zundel ion vibrational spectra,[271, 2, 273] the relaxation dynamics of the Zundel ion and the surrounding solvent after vibrational excitation,[271, 272] and the effect of adding additional water molecules to the solution.[273]

Here, we employ many-body molecular dynamics (MD) simulations in an attempt to model the Zundel ion solvated in liquid acetonitrile, and provide an analysis of the structure of the Zundel ion in acetonitrile, the structure of the acetonitrile solvent, and the vibrational properties of the solvated Zundel ion. Unfortunately, practical difficulties complicated our study,

but we discuss our methodology and results for pedagogical purposes.

6.1 The Failure of TTM-nrg For The Flexible Zundel Ion, and The Importance of Evaluating PEFs with Test Sets Taken From Simulations

The importance of evaluating a model PEF with a test set taken from MD simulations (Sec. 5.1.9) is illustrated in the case of modeling Zundel ion-acetonitrile interactions with a TTM-nrg PEF. The Zundel ion, being the protonated water dimer (Sec. 3.3), is very flexible, in particular in the relative orientations of its flanking-OH bonds, i.e. in the relative orientations of the flanking water molecules solvating the shared (excess) proton. This conformational flexibility of the Zundel ion makes its interactions with acetonitrile too complicated to describe with a simple TTM-nrg PEF, which, apart from the classical polarization, is constructed entirely from pairwise atom-atom interaction terms (Sec. 5.1.1). Additionally, the use of a training set consisting of rigid monomer units obscured the inaccuracy of the model during the development stage, and we had not evaluated the PEF with a test set extracted from the MD simulations prior to analyzing the results.

The Zundel-acetonitrile TTM-nrg training set consisted of physically motivated scans of the Zundel-acetonitrile dimer using rigid monomers in their global minimum energy configurations, similar to those shown in Fig. 5.6 for the acetonitrile-acetonitrile TTM-nrg PEF (which was also used in the simulations). The 2B training set correlation plot for the TTM-nrg fit is shown in Fig. 6.1. The RMSD for training set configurations with reference energies below 15 kcal/mol is 1.04 kcal/mol.

The original test set was constructed from radial scans and random rotations of rigid monomers in their global minimum energy configurations. 2B test set correlation plots for the whole training set, and for the part of the training set with reference energies below 15 kcal/mol are shown in Figs. 6.2 and 6.3, respectively. The RMSD for test set configurations with reference energies below 15 kcal/mol is 1.215 kcal/mol.

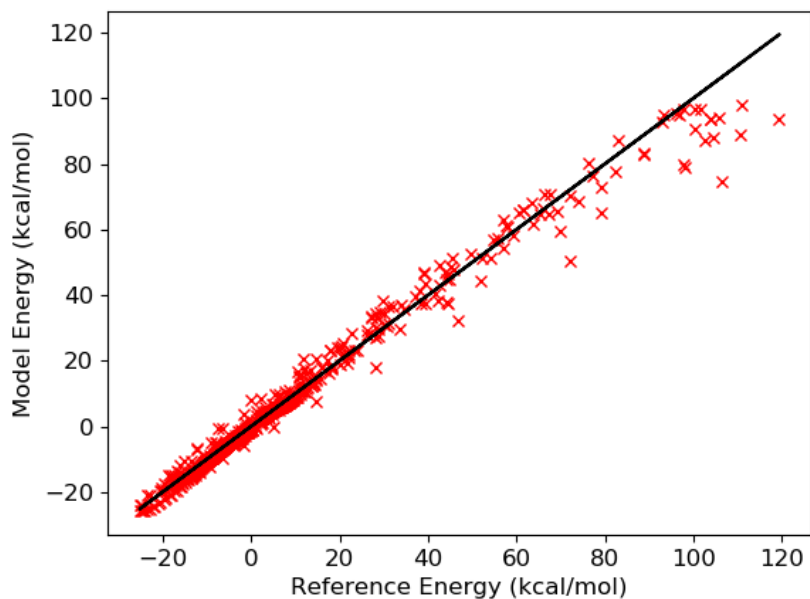


Figure 6.1. Training set correlation plot for the Zundel-acetonitrile TTM-nrg PEF. Reference 2B energies were calculated with ω B97M-V/aug-cc-pVTZ.

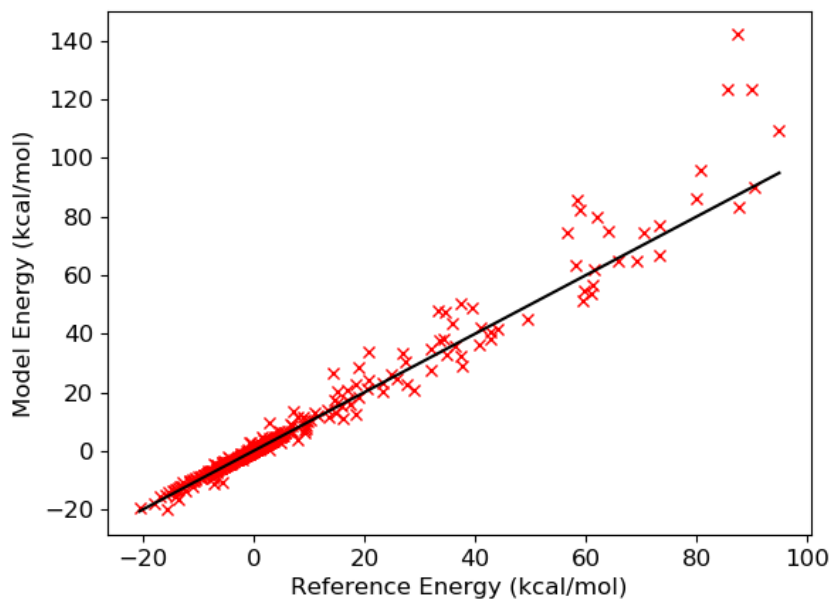


Figure 6.2. Test set correlation plot for the Zundel-acetonitrile TTM-nrg PEF. Reference 2B energies were calculated with ω B97M-V/aug-cc-pVTZ.

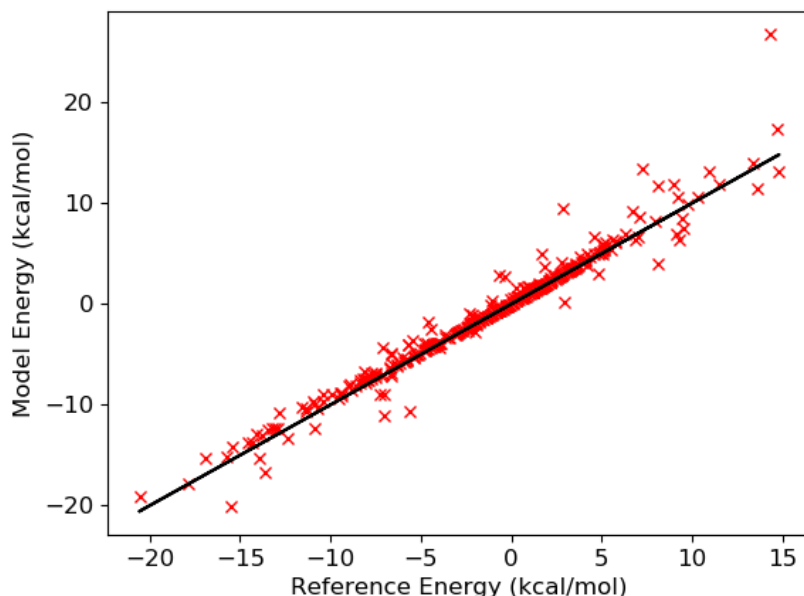


Figure 6.3. Test set (the same as in Fig. 6.2) correlation plot for the Zundel-acetonitrile TTM-nrg PEF for configurations with reference energies below 15 kcal/mol. Reference 2B energies were calculated with ω B97M-V/aug-cc-pVTZ.

However, when evaluating the TTM-nrg model for Zundel-acetonitrile interactions with a test set was taken from the TTM-nrg simulations, it became clear that the major structural and dynamic properties of the solvated Zundel ion seen in our simulations were actually an artifact of deficiencies of the representation of the relevant interactions. The 2B test set correlation plot for dimer configurations extracted from simulations are shown in Fig. 6.4.

The difference in accuracy seen in the original test set (Fig. 6.3) compared to the test set extracted from simulations (Fig. 6.4), particularly in the low energy region of the plot (corresponding to configurations in which the acetonitrile molecules are directly coordinated to the Zundel ion in Fig. 6.4), can be understood by comparing the Zundel monomer conformation used in the training and test sets (Fig. 6.5) to a typical Zundel conformation seen in our MD simulations of the Zundel ion dissolved in liquid acetonitrile using the TTM-nrg PEF.

Notice that in the simulations, the flanking protons in the Zundel ion orient themselves in similar directions in order to maximize the pairwise atom-atom interactions (for example the

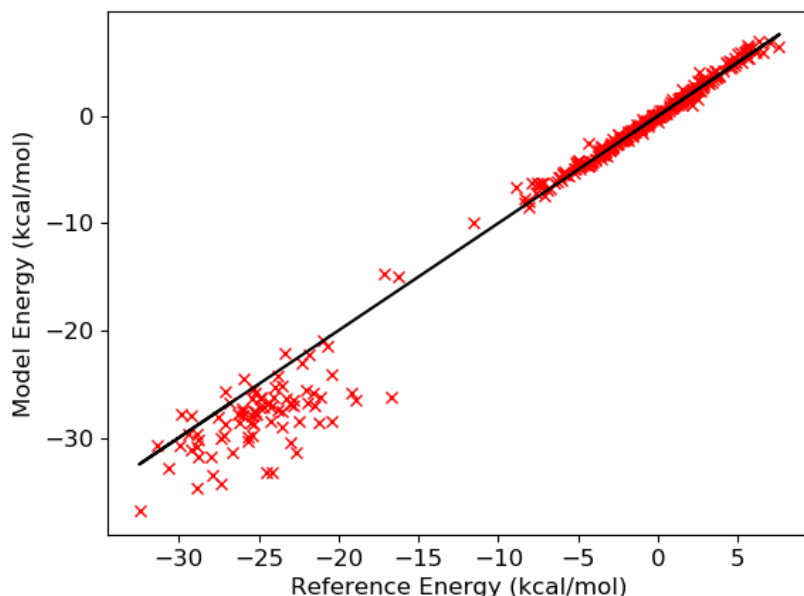


Figure 6.4. Test set correlation plot for the Zundel-acetonitrile TTM-nrg PEF with configurations taken from the TTM-nrg MD simulations. Reference 2B energies were calculated with ω B97M-V/aug-cc-pVTZ.

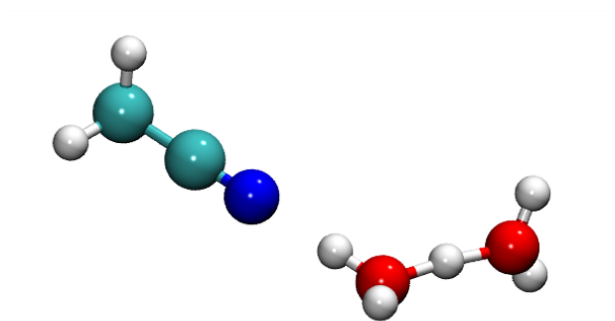


Figure 6.5. An example Zundel-acetonitrile dimer configuration from the TTM-nrg training set where both the Zundel ion and the acetonitrile molecule are in their (isolated monomer) global minimum energy conformations.

electrostatic interactions between the positively charged Zundel-H and the negatively charged acetonitrile-N, as well as the attractive dispersion interactions) between the Zundel ion and the acetonitrile molecules directly coordinated to it. It should additionally be noted that this is not due to deficiencies in our model for 1B distortion energies of the Zundel monomer, which was evaluated to be sufficiently accurate in reproducing the reference energies for configurations

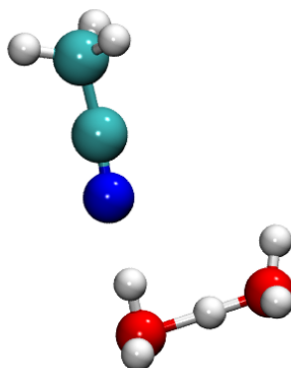


Figure 6.6. A typical Zundel-acetonitrile dimer configuration in our TTM-nrg simulations of the Zundel ion dissolved in liquid acetonitrile. The Zundel flanking protons are oriented in similar directions, which maximizes the pairwise atom-atom interactions (in the TTM-nrg model PEF) with the acetonitrile molecules directly coordinated to the Zundel ion.

taken from our TTM-nrg simulations, as can be seen in Fig. 6.7.

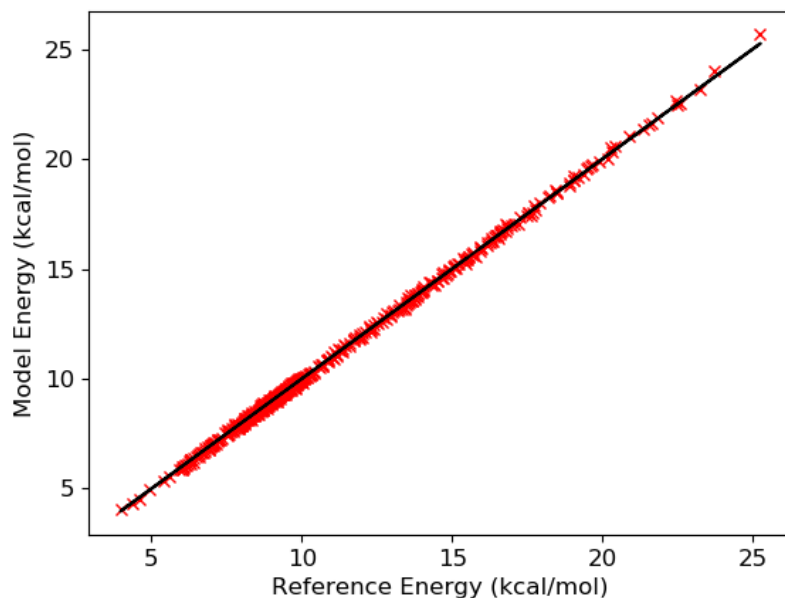


Figure 6.7. 1B Zundel test set correlation plot for configurations taken from our TTM-nrg MD simulations.

Although the simulations are not a reliable representation of the experiments, we discuss the analysis of TTM-nrg simulations of the Zundel ion in acetonitrile in the interest of demonstrating some techniques for analyzing MD simulations.

6.2 Analysis of The TTM-nrg Simulations of The Zundel Ion Dissolved in Acetonitrile

6.2.1 Simulation Details

All classical molecular dynamics (MD) simulations of the Zundel ion in acetonitrile solvent were carried out using the I-Pi[274] software with the in-house MBX driver. The systems simulated all consisted of 128 acetonitrile molecules and one Zundel ion. Each simulation was carried out at constant volume in a cubic box (with periodic boundary conditions) at the experimental density of neat acetonitrile,[275] of around 0.7759 g/cm^3 .¹

All free energy (FE) profiles, radial distribution functions (RDFs), radial-angular distribution functions (RADFs), orientation correlation functions (OCFs), and electric field autocorrelation functions were calculated from a 2.25 ns trajectory simulated at constant temperature (NVT) at 298.15 Kelvin continued from a 500 ps equilibration simulation. All infrared (IR) spectra, and velocity power spectra were calculated from 17 trajectories simulated at constant energy (NVE) for 100 ps each, with initial states (positions and velocities) taken from the NVT simulation in 25 ps increments after equilibration.

Gas phase MD simulations of the Zundel ion were carried out using the LAMMPS software[277] with the in-house MBX driver.

All simulations used a timestep of 0.2 fs. The NVT simulations of the Zundel ion in acetonitrile used a Langevin thermostat[278, 279] with a time constant of 1 ps. NVT simulations of the gas phase Zundel ion used a Nosé-Hoover thermostat.

¹The authors published a better estimate of the density (0.77669 g/cm^3 at 298.15 K) in Ref. 276.

6.2.2 Calculating Vibrational Spectra and Power Spectrum Decompositions

The IR absorption spectrum lineshape function, $I(\omega)$, can be expressed as:

$$I(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \langle \hat{\mathbf{M}}(0) \hat{\mathbf{M}}(\tau) \rangle e^{-i\omega\tau} d\tau \quad (6.1)$$

where angle brackets, $\langle \rangle$, denote the ensemble-averaged expectation value, $\langle \hat{\mathbf{M}}(0) \hat{\mathbf{M}}(\tau) \rangle$ is the autocorrelation function, at time delay, τ , of the total system dipole moment operator, $\hat{\mathbf{M}}$, in the Heisenberg picture.[280] Using the classical system dipole moment, $\mathbf{M}(t)$, (instead of the quantum operator $\hat{\mathbf{M}}(t)$) the IR absorption coefficient is computed within the harmonic approximation[280] as:

$$\alpha_{\text{HA}}(\omega) = \frac{2\pi\beta\omega^2}{3cVn(\omega)} \int_{-\infty}^{\infty} \langle \mathbf{M}(0) \cdot \mathbf{M}(\tau) \rangle e^{-i\omega\tau} d\tau \quad (6.2)$$

where $\beta = 1/k_B T$, c is the speed of light, V is the total volume, and $n(\omega)$ is the frequency-dependent index of refraction of the system at temperature, T .

All IR spectra reported here were calculated according to Eq. 6.2. The total system dipole moment, $\mathbf{M}(t)$, is calculated at several levels of approximation. The crudest approximation (referred to as “1B point charge”) calculates the dipole moment from the position trajectory, using the fixed point charge associated with each atomic site (section 5.1.1). The next approximation (referred to as “1B point charge + induction”) adds the contributions from the induced point dipole moments associated with each atomic site to the 1B point charge total dipole moment. The final two approximations replace the Zundel ion fixed point charges with the 1B Zundel dipole moment surface (DMS), taken from reference 257, in both the 1B, and the 1B + induction approximations, yielding the “1B DMS” and the “1B DMS + induction” approximations, respectively.

The IR spectrum reports on the vibrations of a system indirectly, since the intensity is weighted by the changes in the dipole moment. A more direct picture of the vibrations in

a system can be obtained through the velocity power spectrum (VPS). The power spectrum, $S_x(\omega)$, of some (real-valued) time-dependent signal, $x(t)$, is defined as the expectation value of the squared-magnitude of the Fourier transform of that signal, $S_x(\omega) \equiv \langle |\mathcal{F}[x(t)](\omega)|^2 \rangle$. If the signal results from a stationary process, the Wiener-Khinchin theorem[281] states that the power spectrum can be calculated from the Fourier transform of the autocorrelation function, $C_x(\tau)$, of the signal:

$$\begin{aligned} S_x(\omega) &= \langle |\mathcal{F}[x(t)](\omega)|^2 \rangle \\ &= \mathcal{F}[C_x(t)](\omega) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \langle x(0)x(\tau) \rangle e^{-i\omega\tau} d\tau \end{aligned} \quad (6.3)$$

Note that the autocorrelation function of a signal resulting from a stationary process depends only on the time delay: $C_x(t_1, t_2) \equiv \langle x(t_1)x(t_2) \rangle \xrightarrow{\text{stationary}} \langle x(0)x(t_2 - t_1) \rangle \equiv C_x(t_2 - t_1)$, which is the case when the system is equilibrated.

When the signal is vector-valued (e.g. the velocity or the dipole moment), the autocorrelation function is calculated from the ensemble average of the dot product of the signal with itself after time delay, τ : $C_{\mathbf{x}}(\tau) = \langle \mathbf{x}(0) \cdot \mathbf{x}(\tau) \rangle$. If the signal at time, t , is expanded in terms of some basis, $\mathbf{x}(t) = \sum_n c_n(t) \mathbf{b}'_n(t) = \sum_n \mathbf{b}_n(t)$ (with the expansion coefficients absorbed into the time-dependent basis functions, $\mathbf{b}_n(t)$), the power spectrum can be written as:

$$\begin{aligned} S_{\mathbf{x}}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\langle \left(\sum_n \mathbf{b}_n(0) \right) \cdot \left(\sum_m \mathbf{b}_m(\tau) \right) \right\rangle e^{-i\omega\tau} d\tau \\ &= \sum_n \sum_m \frac{1}{2\pi} \int_{-\infty}^{\infty} \langle \mathbf{b}_n(0) \cdot \mathbf{b}_m(\tau) \rangle e^{-i\omega\tau} d\tau \\ &= \sum_n S_{\mathbf{b}_n}(\omega) + \sum_n \sum_{m \neq n} \frac{1}{2\pi} \int_{-\infty}^{\infty} \langle \mathbf{b}_n(0) \cdot \mathbf{b}_m(\tau) \rangle e^{-i\omega\tau} d\tau \end{aligned} \quad (6.4)$$

The sesquilinearity of the autocorrelation function (or equivalently, the linearity of the Fourier transform) allows the total power spectrum to be decomposed into the sum of power spectra of each basis component, $S_{\mathbf{b}_n}(\omega)$, and the Fourier transforms of the cross-correlations between

different basis vectors. If the signal were to be expanded into an orthogonal basis consisting of vectors that do not depend on time (i.e. $\mathbf{x}(t) = \sum_n c_n(t) \mathbf{b}_n$), then all of the cross-correlations would be 0. In the interest of decomposing the Zundel ion VPS into physically-meaningful components, we expand the velocity vectors into internal coordinates which depend on the geometry, and therefore must change with time.

The Zundel (in acetonitrile) VPS is decomposed into components due to the flanking protons, H_F , the shared proton, H^* , and the oxygens. The collective motions of the four H_F were decomposed into radial (stretching), latitudinal (bending and wagging), and longitudinal (twisting and rocking) components. Fig. 6.8 shows displacement vectors for these three kinds of motions (each arrow should be understood to correspond to the three non-zero components of its own 12-dimensional vector, $[x_1, y_1, z_1, x_2, \dots, z_4]$). Linear combinations of the displacement vectors shown form a basis of the 12 collective motions of the Zundel H_F . We reiterate that these basis vectors depend on the geometry of the molecule, so that for each H_F , the radial vector lies parallel with the corresponding O– H_F axis, latitudinal vectors lie within the corresponding flanking- H_F –O– H_F plane, perpendicular to the radial vector, and longitudinal vectors are perpendicular to their H_F –O– H_F planes.

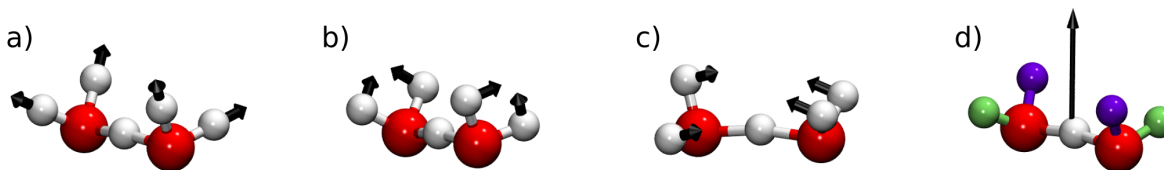


Figure 6.8. Internal coordinate basis vectors for Zundel ion H_F velocities and the approximate C_2 axis shown on a typical Zundel ion configuration from our simulations. The basis vectors are separated into: a) radial, b) latitudinal and c) longitudinal components. Panel d shows the approximate C_2 axis as a black arrow, and also highlights the two pairs of H_F that are interchanged upon C_2 rotation (in lime and violet).

As discussed in Secs. 6.2.3 and 6.2.4, the Zundel ion (in acetonitrile) maintains a fairly rigid shape (due to the effects discussed in Sec. 6.1), which makes it reasonable to assume that it has approximate $\mathcal{C}_2$ symmetry, allowing for a further decomposition of the VPS into components

belonging to the A and B irreducible representations of the $\mathcal{C}_2$ point group. The $\mathcal{C}_2$ -symmetric Zundel ion has two pairs of equivalent H_F that are interchanged upon C_2 rotation (shown in lime and violet in Fig. 6.8d). If we label the two H_F bonded to one oxygen, 1 and 2, and label the two H_F bonded to the other oxygen, 3 and 4, such that H_F 1 and 3 are equivalent, and 2 and 4 are equivalent, then the linear combinations of displacement vectors (in Fig. 6.8) summing to modes in the A irreducible representation have coefficients: $[1, 1, 1, 1]/2$, and $[1, -1, 1, -1]/2$, while the linear combinations summing to modes in the B irreducible representation have coefficients: $[1, 1, -1, -1]/2$ and $[1, -1, -1, 1]/2$.

The H^* motions are also decomposed into modes according to their irreducible representations, with motion along the approximate C_2 axis belonging to the A irreducible representation, and the motion along the proton transfer coordinate (PTC), defined as the difference of the distances between each Zundel oxygen and H^* , and along the remaining perpendicular direction both belonging to the B irreducible representation.

6.2.3 The Structure of the Zundel Ion in Acetonitrile in TTM-nrg Simulations

In the interest of clarity, we refer to the Zundel ion solvated in acetonitrile (not including the solvent) as Z_{acn}^+ , and refer to the gas phase Zundel ion as Z_{gas}^+ .

Fig. 6.9 shows the free energy (FE) of Z_{acn}^+ and Z_{gas}^+ with respect to the O—O distance and PTC (at 298.15 K). It shows that H^* is equidistant from each oxygen (PTC = 0) in most probable configuration for either Zundel ion. However, Z_{acn}^+ tends to have a slightly larger O—O distance than Z_{gas}^+ , having FE minima at around 2.43 and 2.40 Å, respectively. For Z_{acn}^+ and Z_{gas}^+ with O—O distances smaller than around 2.47 and 2.44 Å, respectively, the free energy along the PTC has a single minimum, while the free energy has two minima for larger O—O distances. As can be seen from comparing the curvature of the contours between the two plots, the FE barrier (centered at PTC = 0) of Z_{gas}^+ grows faster, with increasing O—O distance than that of Z_{acn}^+ .

Fig. 6.10 shows the FE of Z_{acn}^+ and Z_{gas}^+ with respect to their O—O distances (averaged

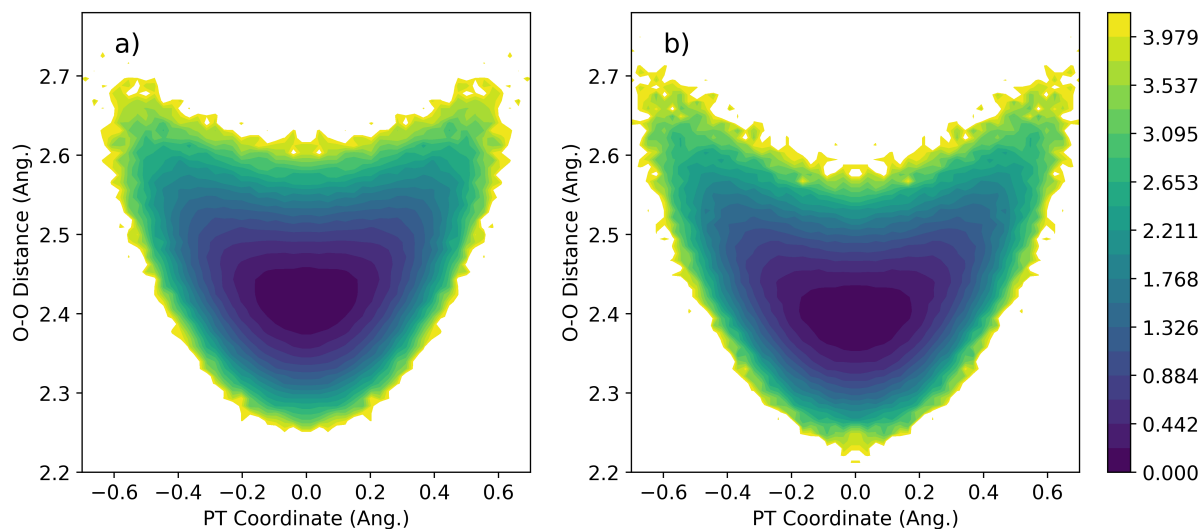


Figure 6.9. Free energy with respect to O–O distance and PTC for a) Z_{acn}^+ , b) Z_{gas}^+ (both symmetrized around PTC).

over their PTC). This shows more clearly that Z_{acn}^+ tends to have a slightly larger O–O distance than Z_{gas}^+ , and also shows that the FE of the two systems along their O–O distances have similar curvatures, and are mostly related by a constant shift of around 0.03 Å.

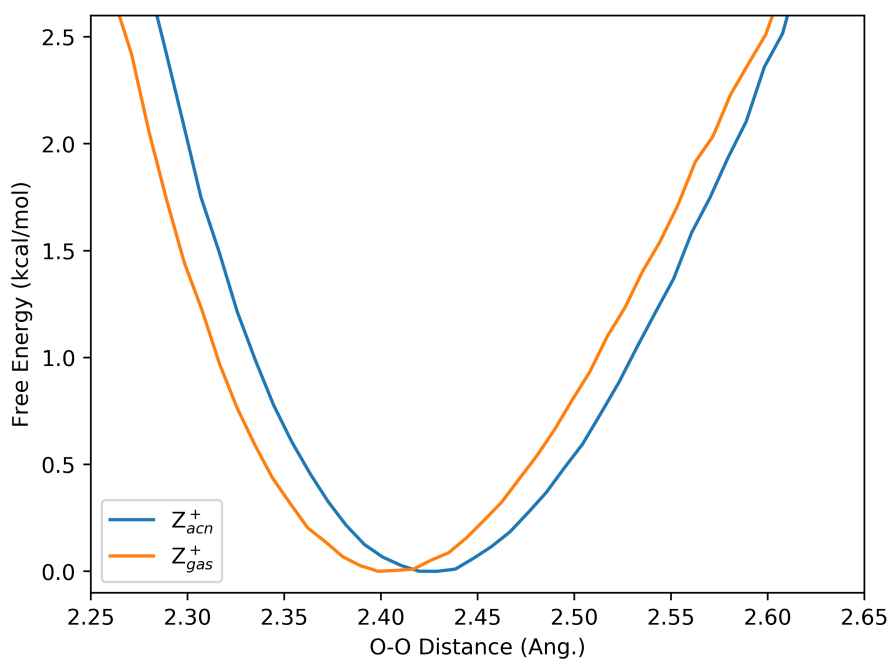


Figure 6.10. Z^+ O–O distance free energy in acetonitrile vs in gas phase.

Fig. 6.11 shows the FE of Z_{acn}^+ and Z_{gas}^+ with respect to their PTC. This shows that both Z_{acn}^+ and Z_{gas}^+ have a single FE minimum at $PTC = 0$, as was seen in Fig. 6.9. However, the FE along the Z_{acn}^+ PTC has a slightly larger curvature than Z_{gas}^+ , which has the effect of redshifting vibrational frequencies along this coordinate in Z_{acn}^+ compared to Z_{gas}^+ (Sec. 6.2.4).

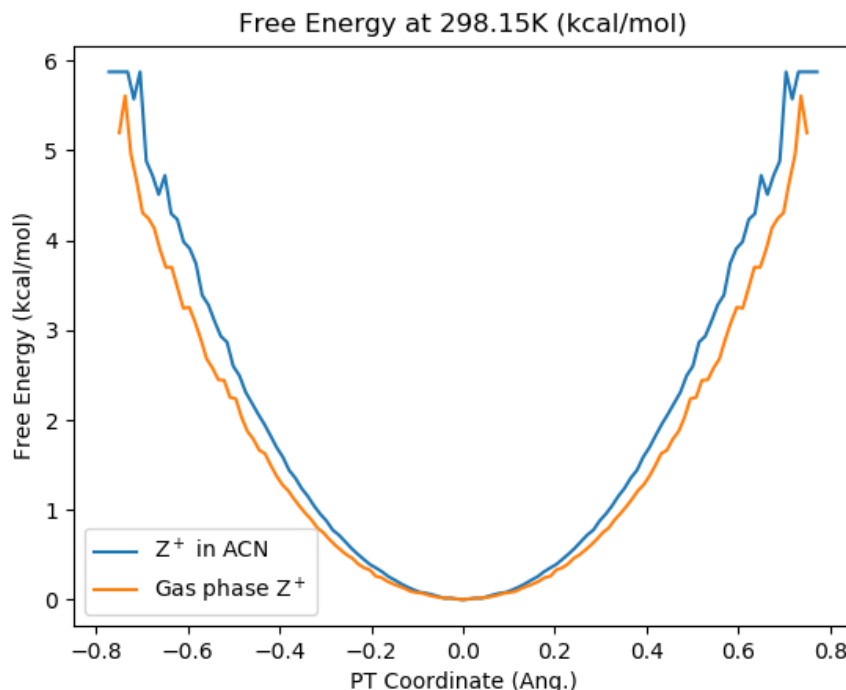


Figure 6.11. Z_{acn}^+ and Z_{gas}^+ FE with respect to the PTC.

Fig. 6.12a shows the FE of Z_{acn}^+ and Z_{gas}^+ along the H_F-O-O angle (with the first O bound to H_F), as shown in Fig. 6.12c (Z_{acn}^+) and Fig. 6.12d (Z_{gas}^+). These curves are averaged over the four distinct (but equivalent) H_F-O-O angles in each Zundel ion. Fig. 6.12a shows that the H_F-O-O angle tends to be around 19° smaller in Z_{acn}^+ than in Z_{gas}^+ . Additionally, the Z_{acn}^+ FE along the H_F-O-O angle has significantly more curvature than that of Z_{gas}^+ . Additionally, the free energy along this angle has slightly greater curvature in acetonitrile than in the gas phase, which is reflected in the vibrational spectra of the longitudinal H_F modes (Sec. 6.2.4). The shift in the Z_{acn}^+ FE relative to the Z_{gas}^+ FE results in angles less than 80° being sampled in solution but not in the gas phase, while angles greater than 150° are sampled in the gas phase, but are

not sampled in solution. Low FE Z_{acn}^+ and Z_{gas}^+ configurations in Fig. 6.12c and Fig. 6.12d show that the smaller H_F-O-O angles in solution make the Zundel ion more compact/compressed compared to the gas phase. This is likely due to the pairwise atom-atom representation of the Zundel-acetonitrile interactions in the TTM-nrg PEF, as discussed in Sec. 6.1.

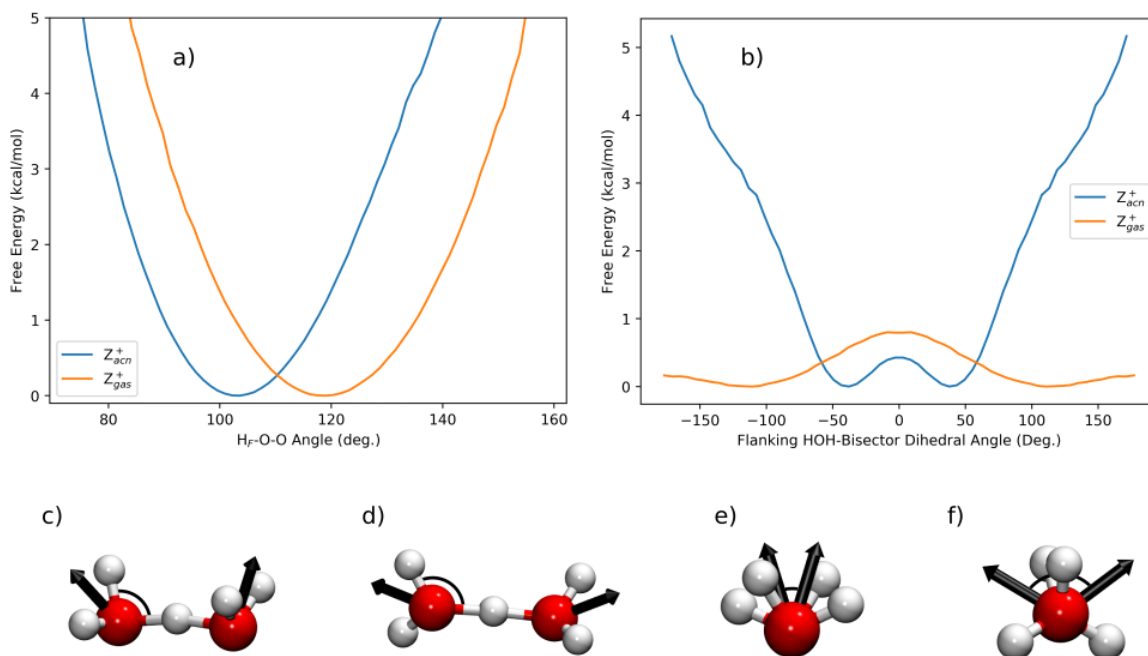


Figure 6.12. a) Free energy along the Z^+ H_F-O-O angle for Z_{acn}^+ and Z_{gas}^+ , b) free energy along the dihedral angle of the Z^+ flanking-H-O-H bisectors for Z_{acn}^+ and Z_{gas}^+ , c) a typical Z_{acn}^+ structure shown with flanking-HOH bisectors (black arrows) and one H_F-O-O angle, d) a typical Z_{gas}^+ structure shown with flanking-HOH bisectors and one H_F-O-O angle, e) the same Z_{acn}^+ structure as in panel c showing its flanking-HOH bisector-oxygen-oxygen-bisector angle, f) the same Z_{gas}^+ structure as in panel d showing its flanking-HOH bisector-oxygen-oxygen-bisector angle.

Fig. 6.12b shows the FE along the Zundel flanking- H_F-O-H_F bisector-oxygen-oxygen-bisector dihedral angle (which we refer to as “the bisector dihedral”) in Z_{acn}^+ and Z_{gas}^+ . This angle is defined such that a dihedral angle of 0° corresponds to both flanking-water units oriented in the same way (bisectors parallel), while an angle of $\pm 180^\circ$ corresponds to the two flanking-water units oriented away from each other (bisectors antiparallel). Note that this dihedral angle can interconvert between 0° and 180° (or any two

angles, $\phi \rightarrow \phi + 180^\circ$) via the “umbrella” (inversion) motion[237] (as occurs in H_3O^+) without traversing any intermediate angles. Due to the symmetry of the Zundel ion, this coordinate is (on average) symmetric around 0° and $\pm 180^\circ$. The same Z_{acn}^+ and Z_{gas}^+ configurations shown in Fig. 6.12c and Fig. 6.12d are reoriented in Fig. 6.12e and Fig. 6.12f, showing their bisector dihedral angles. Note that the black arrows in Fig. 6.12e and Fig. 6.12f show the components of the bisectors perpendicular to the O—O axis for clarity.

Fig. 6.12b shows that the FE of Z_{gas}^+ along this coordinate has minima at around $\pm 111^\circ$, and maxima at $\pm 180^\circ$ and 0° having energies of around 0.2 and 0.8 kcal/mol greater than the minima, respectively. For Z_{acn}^+ , there are maxima at $\pm 180^\circ$ and 0° as well, but the maximum at 0° is smaller than that of Z_{acn}^+ , being around 0.4 kcal/mol above the minima, and the maximum at $\pm 180^\circ$ is substantially larger, having FE around 5 kcal/mol above the minima (although there was very little sampling of large Z_{acn}^+ bisector dihedral angles in our simulations). The FE of the Z_{acn}^+ bisector dihedral angle has minima around $\pm 37.8^\circ$. While the Z_{gas}^+ bisector dihedral FE is relatively flat, particularly for angles $> 100^\circ$ and $< -100^\circ$, for which the two bisectors are oriented away from each other (Fig. 6.12f), the presence of acetonitrile solvent leads to greater FE at larger angles, effectively restricting Z_{acn}^+ to have bisector dihedral angles close to the minima, with the bisectors oriented in similar directions (Fig. 6.12e). Like the effect of the $\text{H}_F\text{—O—O}$ FE, the bisector dihedral FE compresses Z_{acn}^+ (again due to the effects described in Sec. 6.1). This difference also results in larger FE curvature along the Z_{acn}^+ bisector dihedral angle, which is reflected in the vibrational spectra of the H_F (wagging) latitudinal modes (Sec. 6.2.4).

Fig. 6.13 shows the acetonitrile-Zundel N—O, and N—H* radial distribution functions (RDFs) from our simulations.

The narrow peak centered at 2.5 Å and the broad peak centered around 4 Å in the N—O RDF are both predominantly due to acetonitrile molecules bound to the Zundel ion H_F via their nitrogens, with the former peak corresponding to N bound to the O in the N—O distance calculation, while the latter peak corresponding to N bound to the other O. This assignment can be made from considering the N—O—O and N—O— H_F radial-angular distribution functions

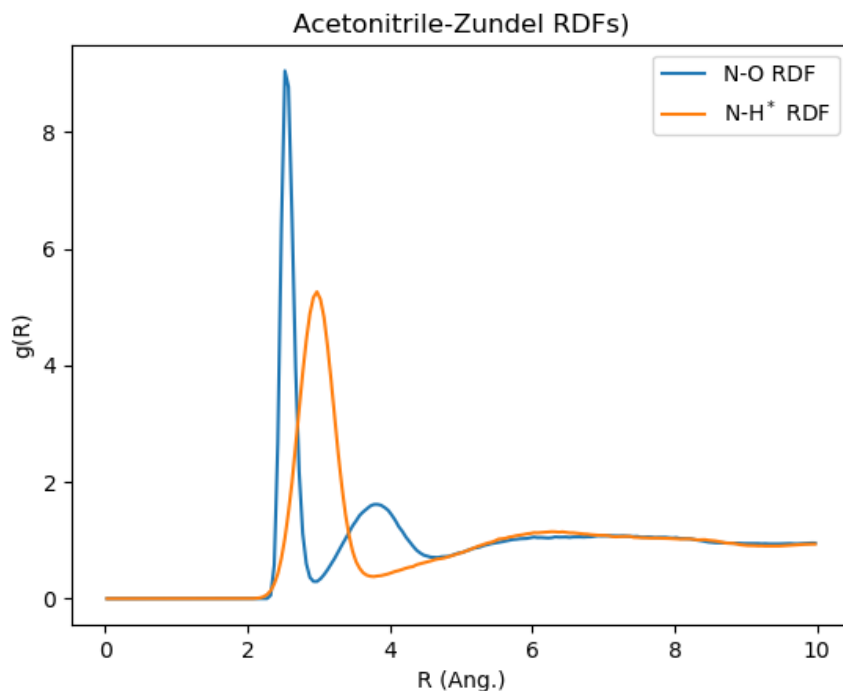


Figure 6.13. N–O and N–H* RDFs.

(RADF), shown in Figs. 6.15a and 6.15c.

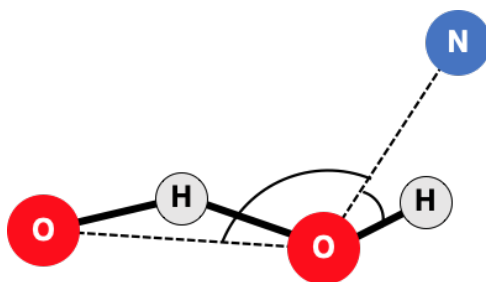


Figure 6.14. Example RADF schematic.

Note that the RADFs in Fig 6.15 are shown in a logarithmic scale so that the outer solvation structure can be visible. The RADFs in Figs. 6.15a and 6.15c correlate the acetonitrile-Zundel N–O distance with the acetonitrile-Zundel-Zundel N–O–O and N–O–H_F angles, respectively. In other words, these plots correlate the position of the acetonitrile-N relative to the Zundel ion with the orientation of the Zundel ion relative to the N. For example, N–O–O angles of 0° or 180° indicate that the N lies on the O–O axis, with an angle of 180° indicating

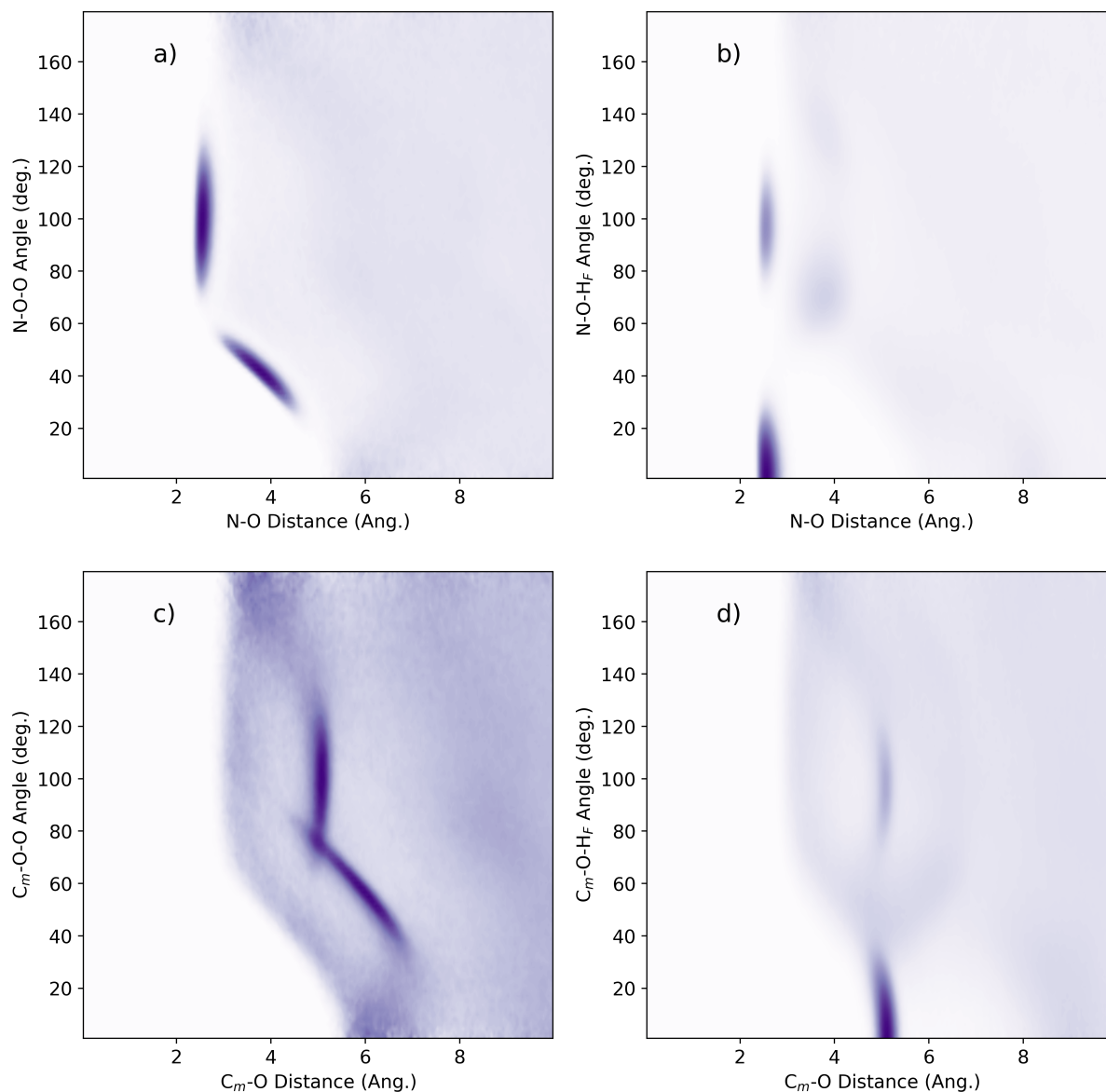


Figure 6.15. Logarithmic scale radial-angular distribution functions: a) N–O–O, b) C_m–O–O, c) N–O–H_F, d) C_m–O–H_F RADFs.

that the N is closer to the Zundel-O used in the N–O distance calculation, while an angle of 0° indicates that N is closer to the other Zundel-O.

Fig. 6.15a shows that the peak at 2.5 Å in the N–O RDF corresponds to N–O–O angles between 80–120° with a maximum around 100°, which corresponds to the most probable Zundel O–O–H_F angle (Fig 6.12a). The N–O–H_F RADF in Fig. 6.15c shows that this peak appears

at two different $\text{N}-\text{O}-\text{H}_F$ angles, one between $0-30^\circ$, and the other between $80-120^\circ$. The peak around 0° confirms that the acetonitrile-N binds to the Zundel- H_F , lying colinear with the corresponding Zundel $\text{O}-\text{H}_F$ axes. The second peak corresponds to N bound to the other Zundel- H_F on the same Zundel-O (hence they have the same $\text{N}-\text{O}$ distance). An example of what this coordination might look like can be seen in Fig. 6.16, which shows an image taken from our simulations.

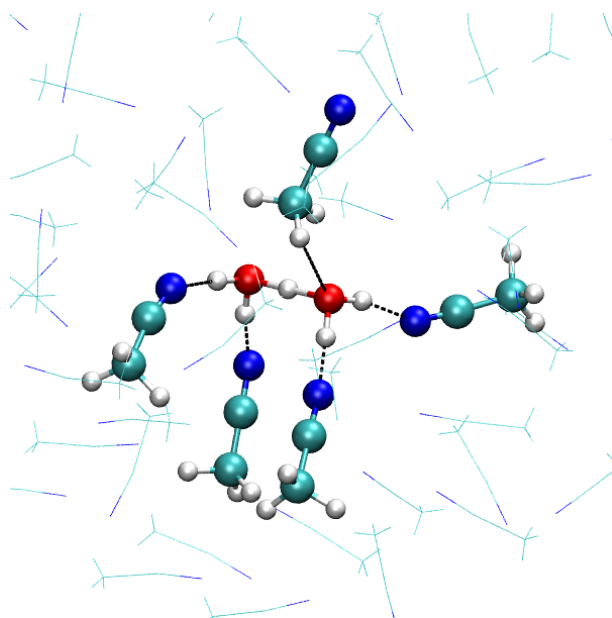


Figure 6.16. Example coordination of acetonitrile molecules to the Zundel ion from our simulations. Highlighted molecules shown in stick and ball representation, surrounding molecules shown as lines. Black dotted lines representing coordination shown for clarity.

The high-density feature between $3-5 \text{ \AA}$ and between $20-60^\circ$ in Fig. 6.15a shows that the broad peak in the $\text{N}-\text{O}$ RDF centered around $4 \text{ \AA}$ is predominantly due to acetonitrile molecules bound to H_F on the (other) Zundel-O which is not used in the $\text{N}-\text{O}$ distance calculation (this can be confirmed from the law of cosines, assuming an $\text{O}-\text{O}$ distance of around $2.5 \text{ \AA}$). Again, Fig. 6.15c shows that this feature is split into two peaks (between $60-90^\circ$, and between $120-150^\circ$) due to the two H_F bonded to that oxygen. Note that this correspondence between the first two peaks in the $\text{N}-\text{O}$ RDF is consistent with the single broad peak in the $\text{N}-\text{H}^*$ RDF (Fig. 6.13) at around $3 \text{ \AA}$, which is not split due to the Zundel- H^* lying between the two Zundel-O.

While this second peak in the N–O RDF is predominantly due to acetonitrile molecules bound to the other Zundel-O, a closer inspection of the regions around N–O distances of 4 Å in the RADFs reveals that there is some acetonitrile-N density which corresponds to acetonitrile molecules that are not directly bound to Zundel- H_F , in particular, at N–O–O angles greater than 160° , and at N–O– H_F angles greater than 150° . This density corresponds to acetonitrile molecules near, but not strongly-bound to the Zundel ion. Density with N–O–O angles greater than 160° is due to acetonitrile molecules roughly colinear with the O–O axis (note that the density around 6 Å with N–O–O angles less than 20° correspond to acetonitrile in the same region on the other side of the O–O axis). Density with N–O– H_F angles greater than 150° correspond to acetonitrile molecules on the oxygen-side of the Zundel O- H_F bonds, which is better understood when considering the acetonitrile-Zundel C_m –O, and C_m – H^* RDFs shown in Fig. 6.17, as well as the C_m –O–O and C_m –O– H_F RADFs shown in Fig. 6.15b and 6.15d.

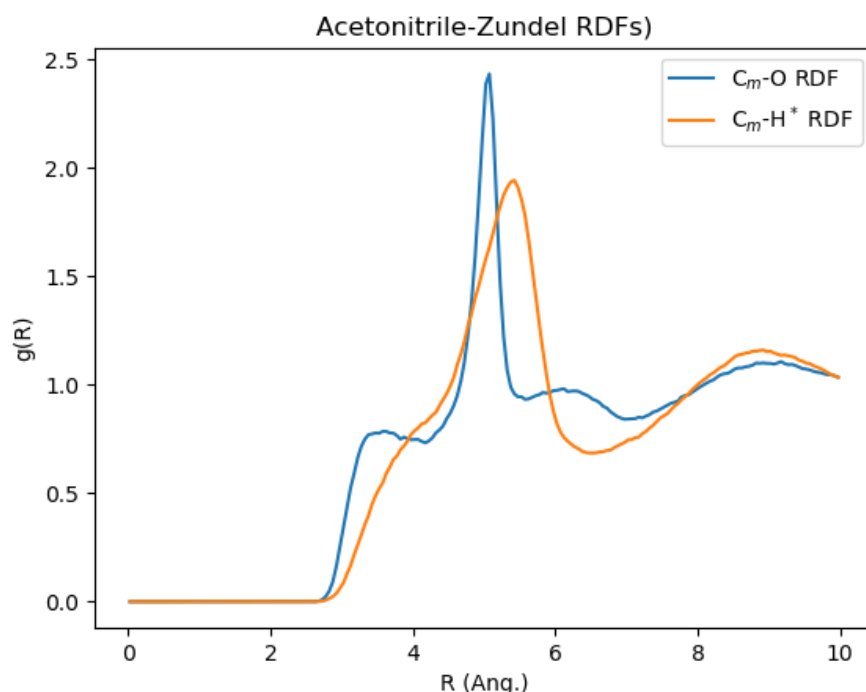


Figure 6.17. C_m –O and C_m – H^* RDFs.

The C_m –O RDF has a broad feature between 3–4.5 Å and a sharp peak at around 5

Å, and two more broad features from 5.5-7 Å and between 7-10 Å. The C_m-O-H_F RADF (Fig. 6.15d) shows that the feature between 3–4.5 Å has a large contribution from C_m with C_m-O-H_F angles greater than 150° (with a maximum around 170°), corresponding to C_m residing on the oxygen-side of the Zundel $O-H_F$ bonds. As is shown in Fig. 6.12b, Z_{acn}^+ tends to have a (staggered) cis-like conformation, with both flanking-HOH bisectors oriented in similar directions. This, in combination with the fact that the Zundel ion in solution tends to have H_F-O-O angles around 100° results in a net Zundel dipole moment perpendicular to the $O-O$ axis (and parallel with the approximate Zundel C_2 axis), which allows the acetonitrile methyl-H to weakly bind to the Zundel-O, opposite the corresponding Zundel- H_F , with C_m-O-H_F angles greater than 150° , and C_m-O distances around 4 Å. An example of what this coordination might look like is shown in Fig. 6.16. It should be noted that this is likely an artifact of the TTM-nrg model, as discussed in Sec. 6.1.

Furthermore, the $O-N-C_m$ RADF shown in Fig. 6.18 correlates the orientation of the acetonitrile molecules relative to the Zundel with the $O-N$ distance, with an $O-N-C_m$ angle of 0° corresponding to the C_m oriented toward the O, and an angle of 180° corresponding to the N oriented toward the O. This RADF shows that virtually all of the density in the first two peaks in the $N-O$ RDF is due to acetonitrile with the N oriented toward the O (having $O-N-C_m$ angles greater than 100°), but density at greater $N-O$ distances do include contributions from acetonitrile molecules with C_m oriented toward the Zundel-O. Note that coordination of the methyl-H to the Zundel-O will usually not have $O-N-C_m$ angles close to 0° (or equivalently, $O-C_m-N$ angles close to 180°), since the acetonitrile $H-C_m-N$ angle tends to be around 110° .

The $N-H^*$ and C_m-H^* RDFs offer a slightly different perspective than the $N-O$ and C_m-O RDFs, but are difficult to interpret due to the dipolar structure of Z_{acn}^+ . These RDFs are better interpreted when correlated with the $N-H^*-C_2$ and the $C_m-H^*-C_2$ angles, where C_2 is taken to be the approximate C_2 axis of the Zundel ion oriented toward the H_F , as shown in Fig. 6.8d, such that angles of 0° correspond to N or C_m lying on the H_F -side of the Zundel ion, and angles of 180° correspond to N or C_m lying on the oxygen-side of the Zundel ion. These

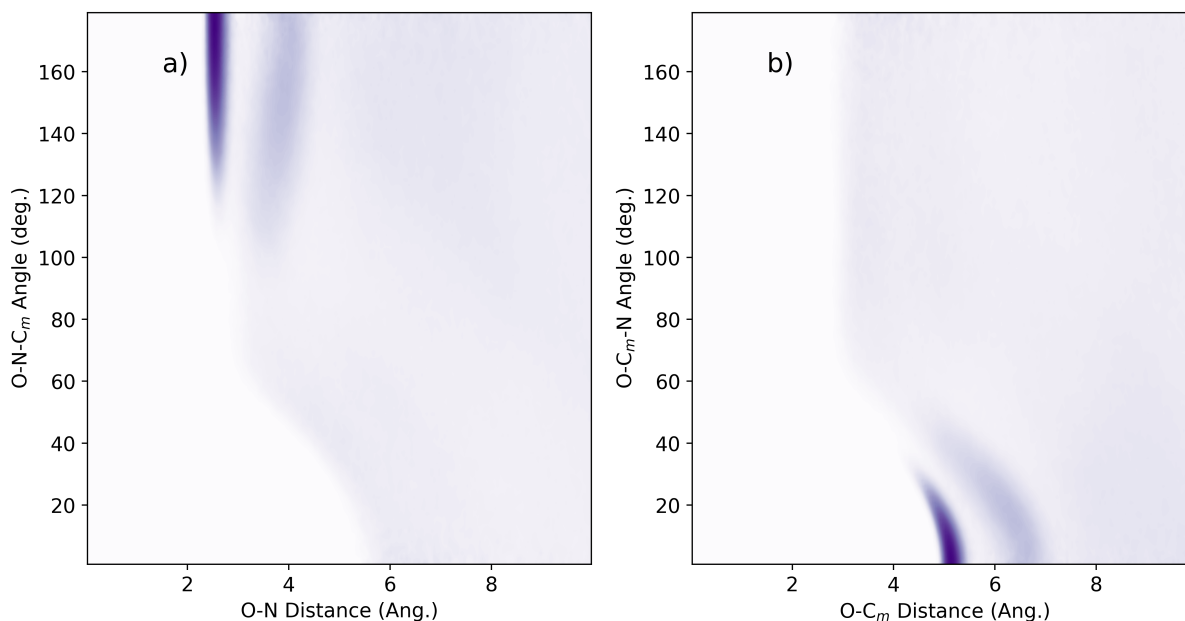


Figure 6.18. Logarithmic scale radial-angular distribution functions: a) O–N– C_m RADF; b) O– C_m –N RADF.

RADFs show the acetonitrile molecules coordinated to the Zundel H_F through their N (the two maxima correspond to the two Zundel-O) at N–H* distances of around 3 Å and N–H*–C₂ angles below 100°, as well as the outer-solvation structure beyond N–H* distances of 5 Å. Additionally, the acetonitrile molecules lying on the oxygen-side of the Zundel appear as density between 3.5-5 Å and angles greater than 120°. Larger density of C_m than N in this region agree with the assessment that acetonitrile molecules weakly bind to the Zundel-O through their methyl groups.

6.2.4 The Dynamics of the Zundel Ion in Acetonitrile in TTM-nrg Simulations

Fig. 6.20 shows a comparison of the IR spectrum of Z_{acn}^+ (but not including contributions to the spectrum due to the dipole moment of the solvent) calculated with four different representations of the Zundel ion dipole moment from the same trajectories. All peaks in the two 1B spectra (the dotted curves in the figure) are centered at around the same frequencies apart from the region from around 800–1400 cm^{−1}. This part of the spectrum includes parts of the signal

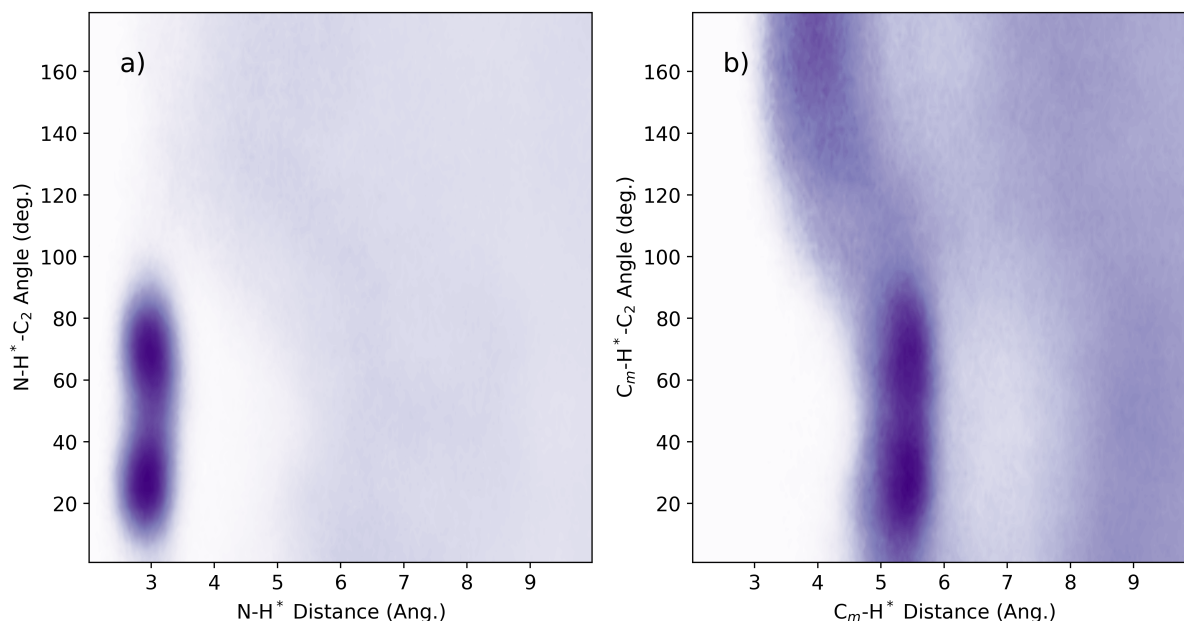


Figure 6.19. Logarithmic scale radial-angular distribution functions: a) $\text{N}-\text{H}^*-\text{C}_2$ -axis; b) $\text{C}_m-\text{H}^*-\text{C}_2$ -axis RADFs.

due to the motion of the shared proton. The differences between the 1B point charge and 1B DMS spectra result from changes in the electronic structure (and hence the charge distribution) of the Zundel ion due mostly to motions of the shared proton which are not captured by fixed point charges.

The 1B + induction spectra (the solid curves in Fig. 6.20) are nearly identical to the corresponding 1B spectra with two exceptions: (1) reduced intensity and splitting of the flanking-OH stretch peaks above 3400 cm^{-1} in the 1B + induction spectra compared to in the 1B spectra; (2) the appearance of a broad feature between $3000\text{--}3400\text{ cm}^{-1}$ in the 1B + induction spectra due to the dipole-polarizabilities of the Zundel ion “reporting” on the C-H stretching vibrations of the acetonitrile methyl groups. Fig. 6.21 shows that this feature in the 1B + induction Zundel spectrum has a similar frequency to the acetonitrile C-H stretch vibrations.

Fig. 6.22 shows a comparison between the IR spectrum of Z_{acn}^+ and Z_{gas}^+ . It should be noted that nuclear quantum effects have been found to be important in simulating the Z_{gas}^+ IR spectrum, even at room temperature[268], but we show it here for the sake of comparison. One

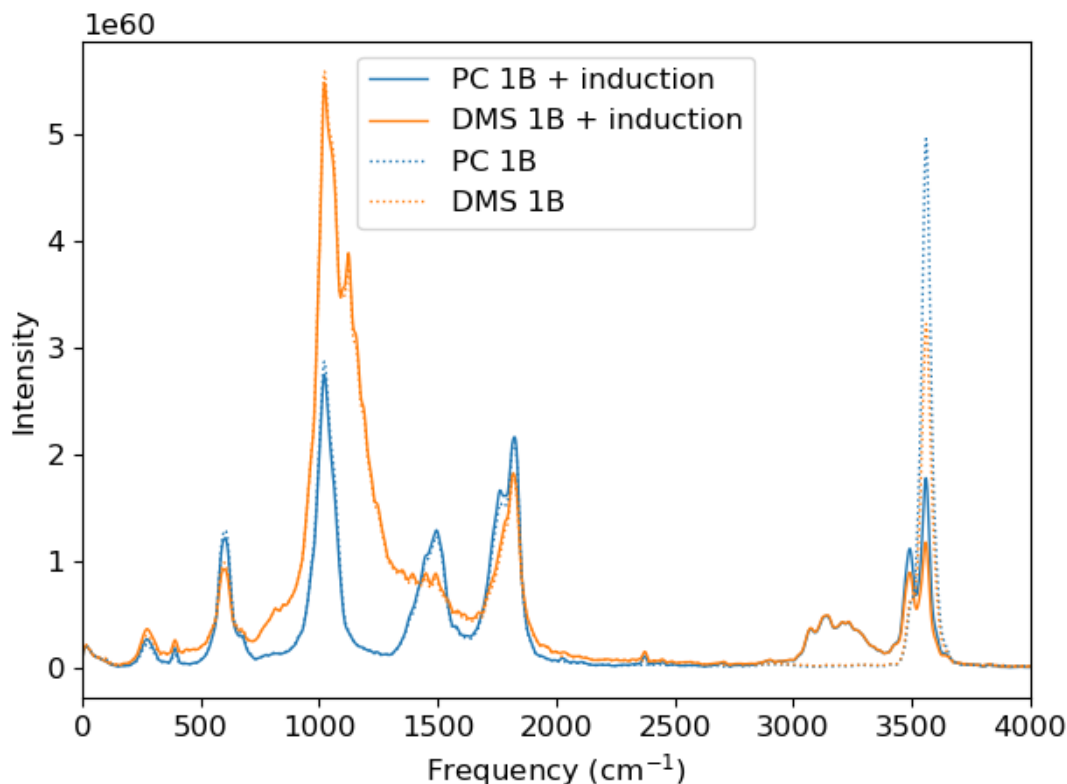


Figure 6.20. Z^+ (in acetonitrile) IR spectra: comparing dipole moment models.

notable difference between the two spectra is the redshift of the flanking-OH stretch frequencies of Z_{acn}^+ compared to those of Z_{gas}^+ , due to the interactions of the acetonitrile molecules bound to the Zundel H_F (as described in Sec 6.2.3), which effectively weaken the Zundel O– H_F bonds. A similar effect can be seen in the redshifting of the OH stretch frequencies in liquid water compared to gas phase water due to strong hydrogen bonding interactions between water molecules. Additionally, the broad feature between 3000-3400 cm^{-1} in the Z_{acn}^+ IR spectrum (described above) does not appear in the Z_{gas}^+ spectrum since there are no acetonitrile molecules in the gas phase simulations. Another difference between the two spectra is the larger IR absorption of the flanking-HOH bending mode (at around 1800 cm^{-1}) for Z_{gas}^+ compared to Z_{acn}^+ . Interestingly, the (unnormalized) velocity power spectra shows that the Z_{acn}^+ asymmetric flanking-HOH bending intensity is larger than that of Z_{gas}^+ , indicating that the larger IR absorption in the gas phase is due to larger changes in the dipole moment due to this vibrational mode in

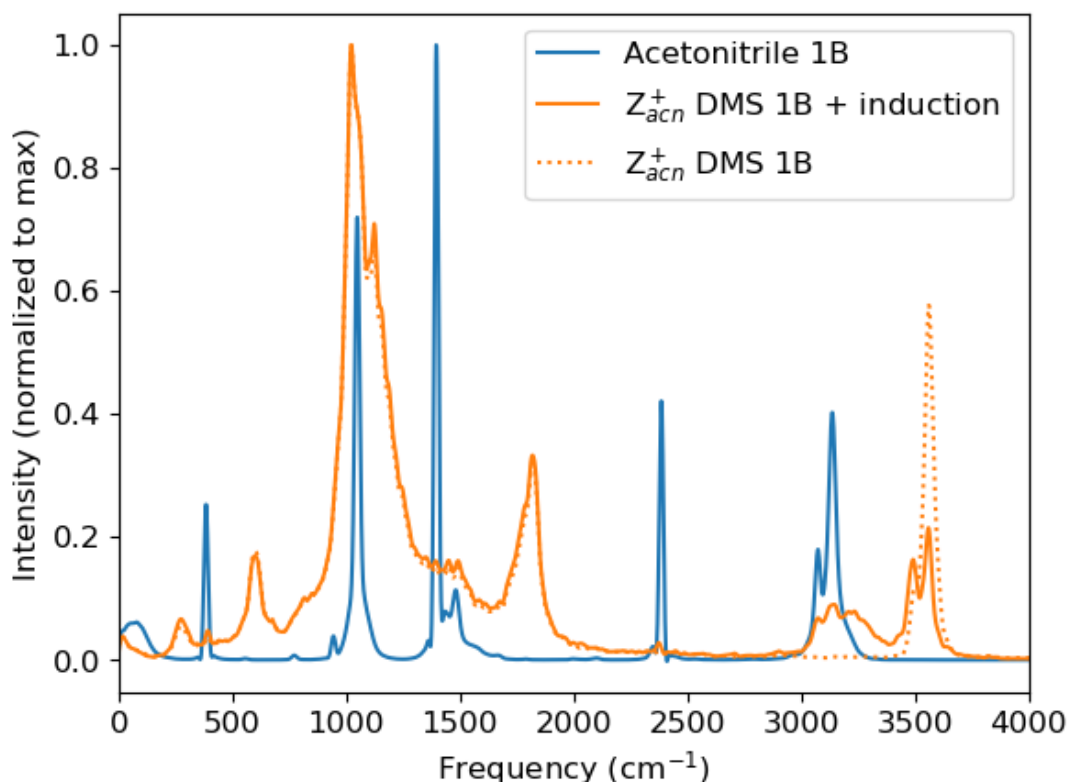


Figure 6.21. Comparing Z^+ (in acetonitrile) and acetonitrile IR spectra to identify peak from induced part of Z^+ spectrum.

Z_{gas}^+ compared to Z_{acn}^+ , which is likely due to differences in the H_F-O-O angle (and possibly the flanking-HOH-bisector dihedral angles) between the two systems.

Fig. 6.23 shows the comparison between the experimental IR spectrum[2] and our calculated IR spectrum. As can be seen, many of the experimentally observed band centers are recovered in our calculations, although the intensities differ substantially. Additionally, the calculated spectrum fails to reproduce the widths of many features. In addition to the lack of nuclear quantum effects in our simulations, the large differences between the experimental and calculated spectra are likely in large part due to inadequacies of the TTM-nrg potential as described in Sec. 6.1. Additionally, the presence of additional water molecules in the experimental solution may also lead to differences between the two spectra.

Fig. 6.24 shows the velocity power spectra of Z_{acn}^+ and Z_{gas}^+ . Comparison with the

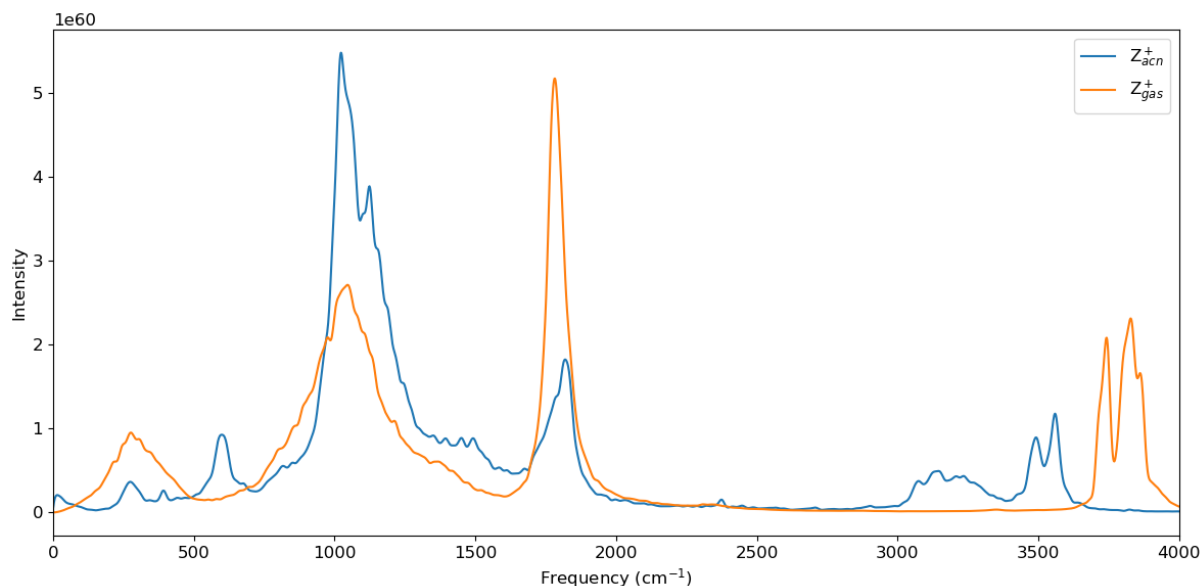


Figure 6.22. Comparison of the DMS + induction IR spectra of Z_{acn}^+ with DMS spectrum of Z_{gas}^+ .

corresponding IR spectra (Fig. 6.22) gives a rough picture of the relative importance of the vibrational power compared to the change in the dipole moment on the IR spectra. For example, this comparison shows that the large signal in the Z_{gas}^+ IR spectrum centered around 1100 cm^{-1} (due to the PTC mode) involves a large change in the Z_{gas}^+ dipole moment, since there is only a weak signal at this frequency in the Z_{gas}^+ power spectrum.

As was discussed in section 6.2.3, Figs. 6.12a and 6.12b show that the Zundel ion conformation is constrained in the presence of the acetonitrile solvent (an artifact of the TTM-nrg model, as discussed in Sec. 6.1). In particular, Fig. 6.12b shows that the flanking HOH-bisector dihedral angle tends to be localized within $\pm 100^\circ$, resulting in a (staggered) cis-like conformation, while Z_{gas}^+ (at 298.15K) is far more flexible (since the FE profile is relatively flat, with the largest barrier being around 1 kcal/mol at an angle of 0°), but tends to assume a trans-like conformation with dihedral angles greater than 100° , or less than -100° . Additionally, Fig. 6.12a shows that the H_F-O-O angle of Z_{acn}^+ tends to be smaller than in the gas phase, making the umbrella inversion motion[237] much less likely. Because of these effective constraints on the flexibility of the Zundel ion due to the solvent, it is reasonable approximation to assume that the Zundel ion

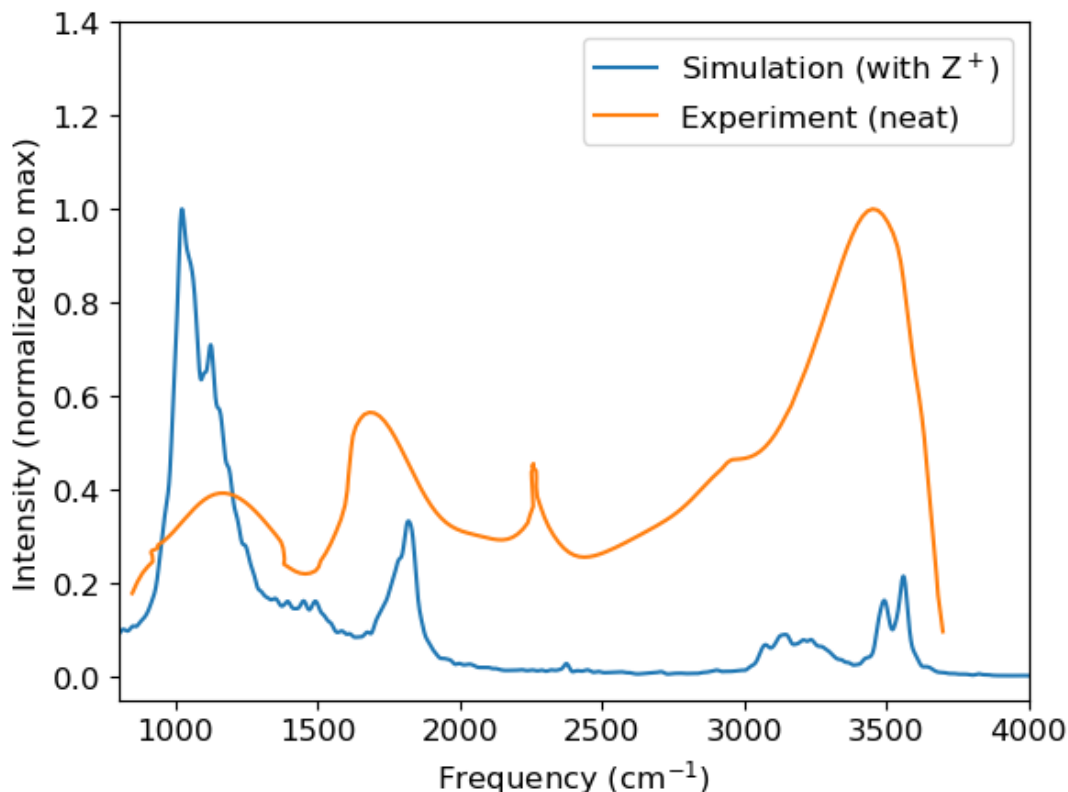


Figure 6.23. Comparison of the DMS + induction IR spectra of Z_{acn}^+ with experimental IR spectrum. The experimental IR spectrum was taken from Fig. 1c in Ref. 2, reproduced with the WebPlotDigitizer.[3]

(in acetonitrile) belongs to the $\mathcal{C}_2$ point group, which allows us to (approximately) decompose its vibrational modes (Sec. 6.2.2) by their irreducible representations.

Fig. 6.25 shows the decomposition of the Zundel (in acetonitrile) velocity power spectrum into the A (blue lines) and B (orange and green lines) irreducible representations of its flanking hydrogen radial (stretching), latitudinal (bending and rocking), and longitudinal (twisting and wagging) modes (in panels a, b, and c, respectively), and its shared proton modes (Fig. 6.25d) along the PTC, and perpendicular to the PTC. Note that the two directions perpendicular to the PTC are decomposed into the direction parallel to the (approximate) C_2 axis (belonging to the A irreducible representation), and perpendicular to both the PTC, and to the C_2 axis (belonging to the B irreducible representation). Motion along the PTC belongs to the B irreducible representation because it has a character of -1 upon C_2 rotation. It should also be noted that each

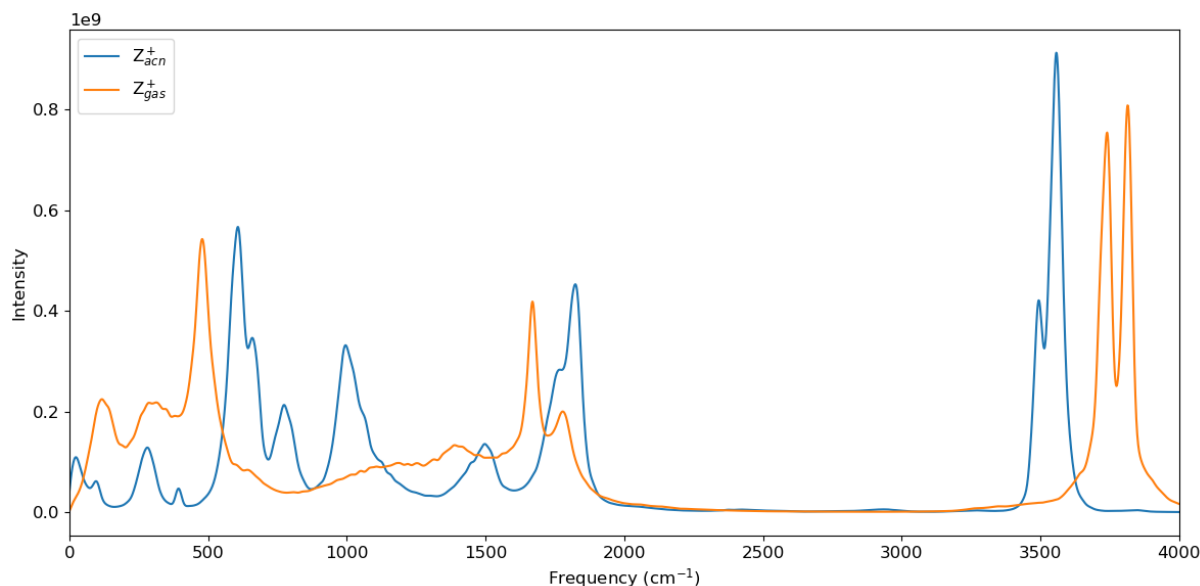


Figure 6.24. Comparison of the total velocity power spectra of Z_{acn}^+ with Z_{gas}^+ .

of the six flanking hydrogen mode types (symmetric stretching, asymmetric stretching, bending, rocking, twisting and wagging) has one mode for each irreducible representation. Fig. 6.25 also shows decompositions of the velocity power spectrum of Z_{gas}^+ at 298.15 K in dotted lines. Since the Zundel ion is not constrained in the gas phase as it is in solution, we do not decompose the $H_F Z_{gas}^+$ spectra into irreducible representations, but we do decompose the H^* modes, shown in blue, orange, and green dotted lines (Fig. 6.25d). The power spectra do not show any significant features between 2000 and 3400 cm^{-1} , so this region is not shown in Fig. 6.25.

Comparing the power spectra of the shared proton modes to those of the flanking proton modes for Z_{acn}^+ , it is clear that both shared proton modes perpendicular to the PTC (solid blue and orange lines) are coupled to longitudinal flanking hydrogen modes (those which have spectral bands at the same frequencies, and belong to the same irreducible representation), namely, the spectral bands with frequencies around 800 and 1500 cm^{-1} having A symmetry, and bands with frequencies around 700 and 1750 cm^{-1} having B symmetry.

The higher-frequency modes (around 1500 and 1750 cm^{-1}) appear weakly coupled, with relatively small H_F longitudinal components, but large H^* components, while the lower-frequency

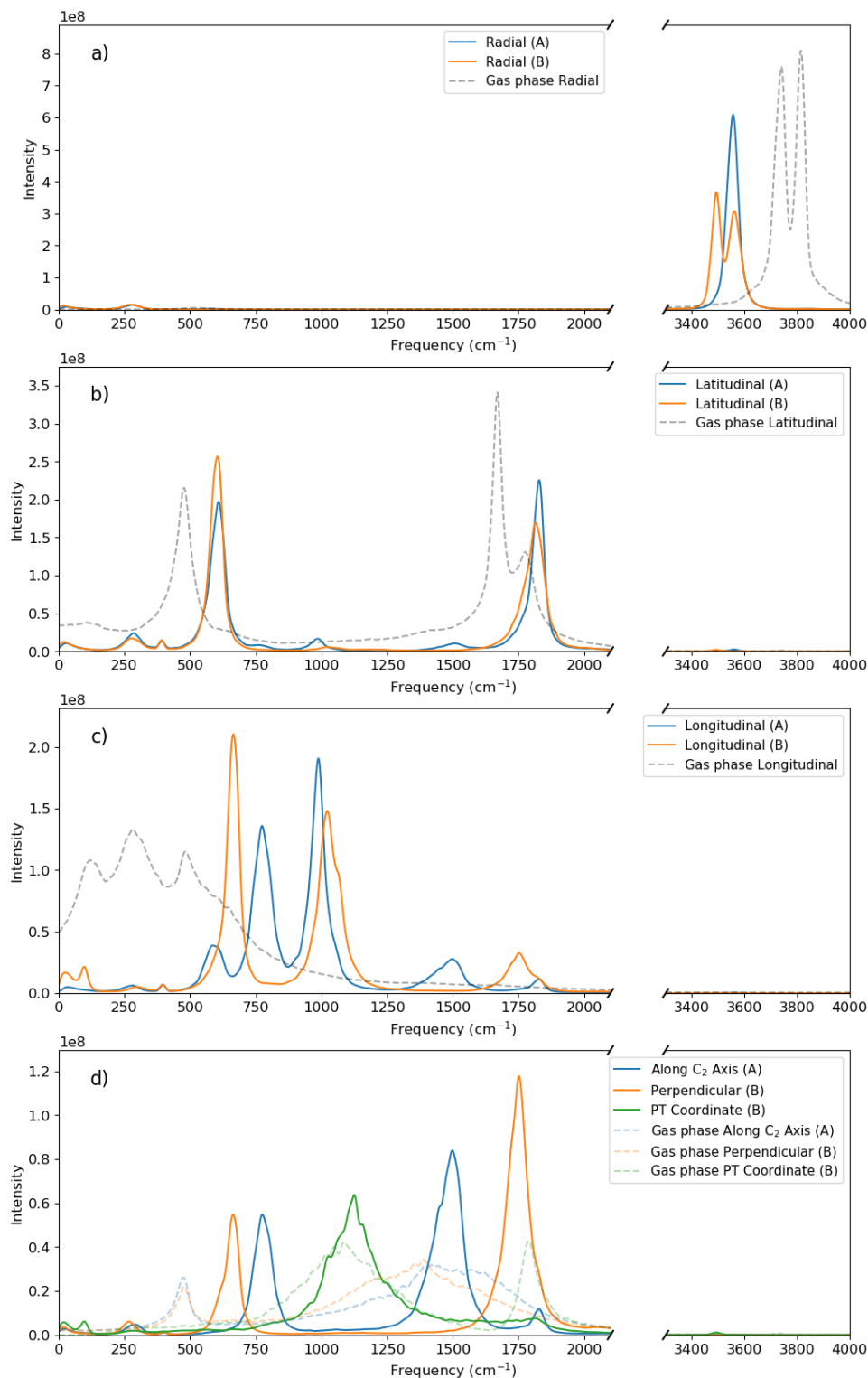


Figure 6.25. Decomposition of the Z_{acn}^+ and Z_{gas}^+ vibrational power spectra into a) H_F radial, b) H_F latitudinal, c) H_F longitudinal, and d) H^* modes. Each H_F mode type is further decomposed, assuming $\mathcal{C}_2$ point group symmetry of the Zundel ion, into A and B irreducible representations.

modes (around 700 and 800 cm^{-1}) appear more strongly coupled, but have slightly more H_F longitudinal than H^* character. With the $\text{H}_F\text{--O--O}$ angles of the Z_{acn}^+ tending to lie around 100° , the H_F longitudinal modes (twisting and wagging) translate the H_F in directions almost parallel with the O--O axis. Thinking of the Zundel H^* , one H_F , and the covalently bonded O as a water-like unit, the coupling of the longitudinal H_F modes to H^* modes perpendicular to the PTC result in (mixed) bending, twisting, and rocking motions of the $\text{H}^*\text{--O--H}_F$ water-like unit. Note that the Z_{acn}^+ longitudinal modes are blueshifted relative to the Z_{gas}^+ longitudinal modes due to the larger FE curvature along the $\text{H}_F\text{--O--O}$ angle (Fig. 6.12a). Furthermore, the Z_{gas}^+ higher-frequency H^* modes perpendicular to the PTC mode (dotted blue and orange lines in Fig. 6.25d) appear completely decoupled from the Z_{gas}^+ H_F longitudinal modes, which are apparently limited to frequencies below 1000 cm^{-1} . We also point out that the low-frequency H_F latitudinal (rocking) modes are blueshifted in Z_{acn}^+ relative to Z_{gas}^+ due to larger FE curvature along the bisector-dihedral angle in Z_{acn}^+ compared to Z_{gas}^+ .

The lower-frequency H^* modes perpendicular to the PTC that appeared at around 700 and 800 cm^{-1} for Z_{acn}^+ appear coalesced into one band, and redshifted to around 500 cm^{-1} in the Z_{gas}^+ spectrum (and actually form a continuum with the higher-frequency bands with nonzero intensity in between). Again, the relative spectral shift of these longitudinal modes in Z_{acn}^+ compared to Z_{gas}^+ is due to differences in the FE curvature along the $\text{H}_F\text{--O--O}$ angle between the two systems (Fig. 6.12a). These low-frequency Z_{gas}^+ H^* modes appear coupled to longitudinal H_F motions, as well as latitudinal H_F modes, specifically, the rocking mode, all lying within the same frequency range.

The H^* PTC spectrum for Z_{acn}^+ appears predominantly as a broad feature from around 1000 to 1300 cm^{-1} , centered at around 1100 cm^{-1} , with a small tail extending to around 700 cm^{-1} , and a more prominent tail extending to around 1900 cm^{-1} . In the Z_{gas}^+ spectrum, this major part of the PTC signal is broader, extending from 750 to 1300 cm^{-1} , and the peak maximum is slightly redshifted compared to the Z_{acn}^+ maximum, reflecting the larger curvature in the O--O FE (Fig. 6.10). The center of the major Z_{acn}^+ PTC band around 1100 cm^{-1} doesn't have

significant spectral overlap with the flanking hydrogen modes, although there is some overlap between the shoulder on the red side of the PTC band and the peak in the longitudinal mode with B symmetry just above 1000 cm^{-1} . The high-frequency tail of the Z_{acn}^+ PTC spectrum appears to overlap with features in the spectra for both the A and B components of the H^* motion perpendicular to the PTC, possibly indicating small couplings between these degrees of freedom. Shared proton motions with opposing symmetries may be coupled despite the symmetry mismatch simply because they correspond to motions of the same, single atom. The high-frequency tail of the Z_{acn}^+ PTC band also extends to 1800 cm^{-1} , possibly suggesting small coupling to the asymmetric H_F (B) bending mode. In the Z_{gas}^+ spectrum, the coupling between H^* PTC motions (dotted green line) and H_F bending motions are far more prominent, indicating that the constraints imposed on the Zundel ion due to the solvent suppress this coupling. As discussed in reference 262, motion of the H^* along the PTC distorts the electronic structure of the Zundel ion from having $O_2H_5^+$ character to $H_3O^+ \cdot H_2O$, modulating the potential along the flanking HOH bending (and wagging) motions, shifting the vibrational frequencies, which results in the observed coupling.[262, 253, 269, 282]

Chapter 7

Chemical Reactions: First Steps Toward A Many-Body Valence Bond Model For Aqueous Proton Transfer

7.1 Background

In MD simulations with MB-nrg PEFs, the monomer identities remain fixed throughout the simulations. In terms of the electronic degrees of freedom, any assignment of monomer identities (i.e. an assignment of covalent bonds between atomic sites belonging to the same monomers) corresponds to some electronic state. Furthermore, for a given (fixed) configuration of atomic sites, one can imagine many ways to assign covalent bonds, each one corresponding to its own electronic state (most of which would be excited states). From this perspective, chemical reactions involving the breaking and forming of covalent bonds can be thought of as the coupling between two or more of these electronic (covalent bonding) states, with the reaction intermediates corresponding to superpositions between the states. This is the general idea behind the empirical valence bond (VB) models.[283, 284, 220, 222, 224, 225, 231, 238] The empirical VB approach was pioneered by Mulliken,[283] applied to modeling enzyme reactions,[284] and later extended to a multi-state model for describing aqueous PT.[220] We set out to adapt the VB approach into the MB-nrg methodology in order to model aqueous PT with the accuracy of the MB-pol PEF, and is correspondingly referred to as MB-VB.

The first component needed to construct any VB model is a non-reactive PEF describing any possible covalent bonding state for the system of interest. This non-reactive PEF in the MB-VB model is represented by a typical MB-nrg PEF, although the fitting procedure will need to be adapted to overcome some complications due to the fact that the PEF must describe high energy electronic (covalent bonding) states, as we will see below. The coupling of these bonding states will allow for the rearrangement of covalent bonds in an MD simulation. Due to the parallels of the VB approach to the treatment of non-adiabatic dynamics within the diabatic representation,[8] the bonding states in a VB model are referred to as *diabatic states* (DS). The DSs collectively form a basis (in practice, a basis approximation) of the VB Hilbert space, from which we may construct a matrix representation of the VB Hamiltonian for our system, the eigenstates of which (formed from superpositions of the DSs) are referred to as *adiabatic states*, with the lowest energy adiabatic state corresponding to the electronic ground state, which is used as the Born-Oppenheimer PES in molecular dynamics simulations. For a given configuration of atomic sites, the binding energies of each DS, obtained from the non-reactive PEF, will correspond to the diagonal elements of the matrix representation of the VB Hamiltonian in the basis of the DSs.

The next component needed is a state search algorithm which is able to identify all important (low energy) DSs in a fixed configuration of atomic sites for each MD time step. In (acidic) simulations of one excess proton dissolved in an N water molecule system, each DS will contain one H_3O^+ monomer, and $N - 1$ water monomers (disregarding auto-ionization of two water molecules into OH^- and H_3O^+). In (basic) simulations of one absent proton (sometimes called a proton “defect”) in an N water molecule system, each DS will contain one OH^- monomer, and $N - 1$ water monomers. In principle, all possible DSs (covalent bonding connectivities) may be obtained to construct the VB Hamiltonian, but in practice a relatively small subset of DSs are needed.

In particular, for each MD time step, the state search algorithm first identifies one (or more) candidate(s) for the lowest energy DS, referred to as “pivot” states. For N_w water

molecules, N_+ H_3O^+ ions, and N_- OH^- ions (and $N = N_w + N_+ + N_-$ oxygens), there are $\binom{N}{N_w} \binom{N-N_w}{N_+} (3N_+ + 2N_w + N_-)! / ((3!)^{N_+} (2!)^{N_w} (1!)^{N_-})$ possible DSs (i.e. ways to assign hydrogens to oxygens), so the pivot states are usually identified by geometric criteria (e.g. minimizing OH bond lengths), rather than evaluating the PEF on all possible DSs. After obtaining the pivot states, the algorithm identifies all DSs which must be coupled to the pivots in the VB Hamiltonian, and subsequently, all DSs which must be coupled to the DSs coupled to the pivots, etc.. Within the VB model for aqueous PT, a pair of states are coupled when they differ by exactly two covalent bonds (i.e. the two states can be seen as the reactant and product, or vice versa, due to a single PT event). The DSs which are directly coupled to a pivot can be seen as belonging to the “first solvation shell” around the pivot, and DSs coupled to those DSs belonging to the second solvation shell around the pivot, etc.. It has been found that the VB Hamiltonian for one excess proton in water can be constructed from DSs in the first three solvation shells around the pivot in order to conserve the energy in MD simulations.[231]

The final component needed for a VB model is the coupling potential between pairs of DSs, which correspond to the off-diagonal matrix elements of the VB Hamiltonian in the basis of the DSs. As a first approximation, the coupling potential can be fitted to reproduce the ground electronic state (binding) energy of the relevant dimer (i.e. the H_2O and H_3O^+ which differ by two covalent bonds between the two coupled DSs in the case of the excess proton). Additionally, one can include effects due to the solvent in the parameterization of the coupling potential.[220]

Once the PEFs for the DS energy and the coupling potential and the state search algorithm are obtained, the Hellman-Feynman theorem is used to determine the forces used to propagate the molecular dynamics.[220]

7.2 MB-VB

In the context of the VB method, each DS for a given nuclear configuration corresponds to an electronic state with well-defined monomer units. For example, in a symmetric Zundel ion

configuration (Sec. 3.3), the two lowest energy DSs correspond to the electronic states having well-defined H_3O^+ and H_2O units, with the shared proton being covalently-bonded to exactly one oxygen in each DS (Fig. 7.1), while the (adiabatic) ground electronic state can be seen as having two equivalent covalent-like bonds between the shared proton and each oxygen (i.e. a 3-center 4-electron bond).[112, 113, 114] From this perspective, the DSs can be thought of as electronic states with limited (electronic) charge transfer (CT) between the defined monomer units. For example, each Zundel DS can be seen as an electronic state that lacks the CT from the water-monomer oxygen to the shared proton.

Although there is no single, unique way to define VB DSs,[285] the development of model DS PEFs should follow the following ideas:

1. For all possible nuclear configurations of the system, all DSs must have larger binding energies than the adiabatic ground state.
2. For configurations characterized by ambiguity in monomer unit definitions (for example, the symmetric Zundel ion), the adiabatic ground state wave function must be well-approximated as a linear combination of the relevant DS wave functions.
3. For configurations characterized by non-ambiguous monomer unit definitions (for example, H_3O^+ and H_2O separated by a large distance), the adiabatic ground state should be well-approximated by the lowest energy DS alone (i.e. the coupling should close to zero).

The first two points follow directly from the basic idea of the VB method, which defines the adiabatic ground state energy as the lowest eigenvalue of the VB Hamiltonian in the basis of the DSs for any given nuclear configuration. The third point follows from the interpretation of the DSs as being defined by well-defined covalent bonding connectivities, and hence well-defined monomer units. This third point is illustrated in the cartoon plot in Fig. 7.1 in which the adiabatic ground state energy (shown in the black dotted curve) converges to each of the DS energies (the solid red and blue curves) for values of the reaction coordinate far from the transition state (where the DS energies cross).

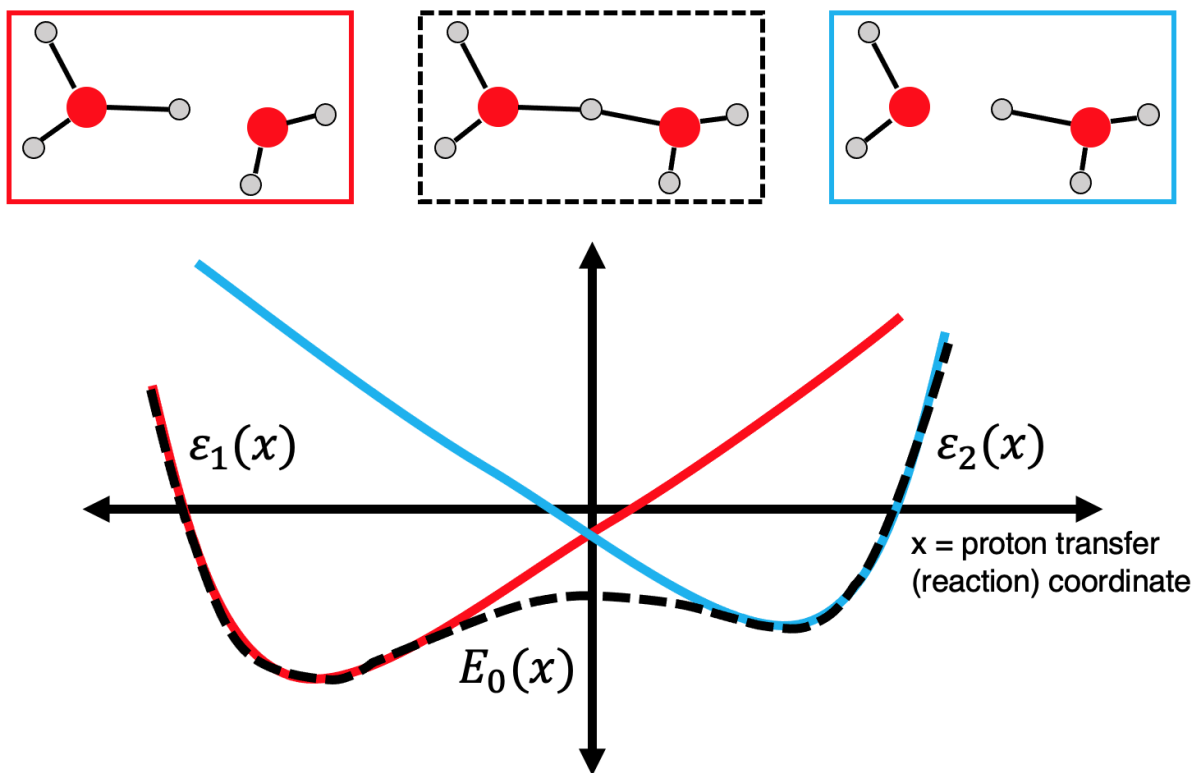


Figure 7.1. Cartoon depicting the bonding connectivity of the two lowest energy diabatic states (enclosed in the red and blue boxes), and the adiabatic ground state (enclosed in the black box) of the Zundel ion (shown to be slightly asymmetric), and a cartoon plot of the diabatic state energies, $\epsilon_1(x)$ and $\epsilon_2(x)$, along the PTC (used as a reaction coordinate) shown as solid red and blue curves. The ground adiabatic state energy, $E_0(x)$, is shown as the black dotted curve on the same axes.

In our MB-VB approach, we model the DS energy with a typical MB-nrg PEF (Sec. 5.1). In particular, the 1B monomer distortion potentials and long-range (TTM-nrg-like) potentials are obtained in exactly the same manner as in any other non-reactive MB-nrg PEF, while the diabatic character of the PEF is contained in the PIP component of the 2B PEF. It is important to note, however, that high energy DSs identified by the VB state search algorithm will have highly distorted monomers, so the 1B training sets must include sufficiently distorted monomers (for example, H_3O^+ and H_2O monomers with large O–H distances) so that the 1B PEFs do not need to extrapolate. Therefore, the biggest uncertainty in how to develop an MB-VB model is in the 2B DS PEF fitting procedure.

With the intent of developing a DS PEF that remains close to the adiabatic ground state PES, we initially attempted to fit the model directly to the adiabatic (reference) energies. Using the Zundel 1B training set used in Chapter 6, we permute the atomic coordinates of each configuration from the O_2H_5 ordering into an OH_3OH_2 ordering, such that each hydrogen is paired with the closest oxygen (i.e. minimizing the sum of the O–H distances). In terms of the cartoon plot in Fig. 7.1, this OH_3OH_2 training set only includes DS 1 (red) configurations with negative x values (on the left-hand side of the x -axis), or equivalently, DS 2 (blue) configurations with positive x values. In other words, this initial training set only includes configurations representing the low energy DS. The test set was permuted in the same way.

A 2B MB-nrg PEF, based on monomial connectivity graphs (Sec. 5.1.10), was fitted to the reference (adiabatic) 2B energies for the training set, resulting in a model that reproduced the test set reference (adiabatic) 2B energies with an RMSD of 0.29 kcal/mol (see Table 5.1). However, despite the apparent accuracy of the model for the low energy DS configurations, this fitting procedure ignores the high energy DS region of configuration space, leading to a serious PIP extrapolation problem (Sec. 5.1.9). High energy DS configurations are characterized by a shared proton that is closer to the water oxygen than the H_3O^+ oxygen, which results in overly attractive electrostatic and dispersion energies from the long range potential. Without any high energy DS representation in the training set, the behavior of the 2B PIP is uncontrolled (i.e. extrapolating) in these configurations, and can be overpowered by the electrostatics and dispersion. As a result, many of these configurations are evaluated to be lower in binding energy than the adiabatic ground state. One example of this extrapolation problem is shown in Fig. 7.2 for a simple rigid scan along the PTC, with the DS 2 curve (the orange dotted line) crossing the adiabatic ground state curve (the solid black line) for several configurations with negative PTC. More extreme examples were also found in the training set. It should be noted that the convergence of each DS curve to the adiabatic ground state curve on either side of the plot illustrates the accuracy of the DS PEF in the low energy DS regions of configuration space.

Based on these results, it is clear that the high energy DS region of configuration space

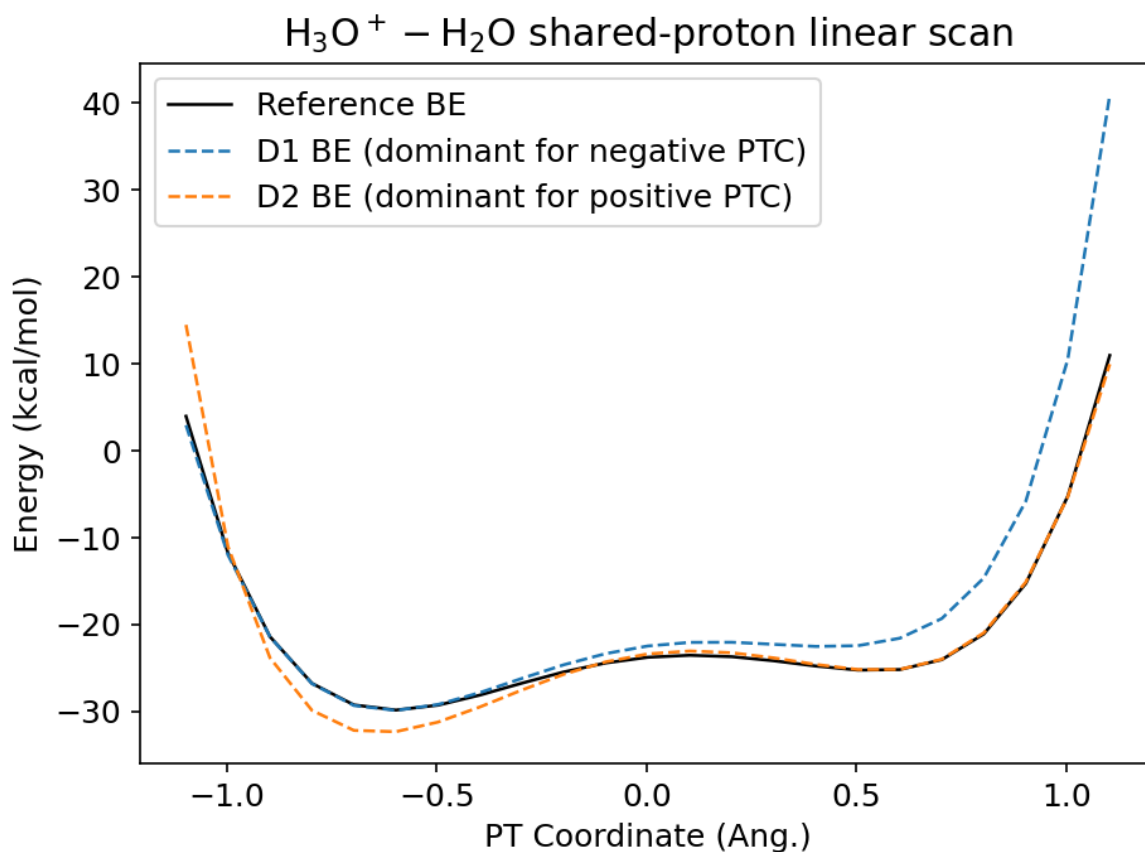


Figure 7.2. Example of the DS PEF extrapolation problem for the protonated water dimer on a rigid scan along the PTC. Blue and orange dotted lines show the binding energies of DS 1 and DS 2, respectively, for the model PEF described here, while the solid black line shows the reference adiabatic ground state binding energy. Due to the extrapolation of the DS PEF PIP, the orange curve crosses the black curve, indicating that the DS 2 binding energy is more attractive than the adiabatic ground state energy, violating the requirement that the ground state energy is the lowest possible energy.

must be represented in the training set in order to force the DS PEF to satisfy the requirement that all DS binding energies must be higher than the adiabatic ground state binding energies for any given nuclear configuration. In order to obtain reference energies for the entire configuration space (for high energy and low energy DS configurations), we return to the idea that DSs are defined by some limitation in the CT between monomer units.

7.2.1 Model DS PEFs Based on ALMO-EDA Charge Transfer Energies

One approach to defining VB DSs in terms of CT-free electronic states is in the use of constrained DFT (CDFT).[285] Given a set of user-defined monomer units and associated monomer charges, CDFT partitions space into regions for each monomer, and solves the Kohn-Sham equations for the system under the constraint that the electron density within each region corresponds to the user-defined monomer charge.[127] As discussed in Chapter 4, another approach to defining CT-free electronic states is through the use of the ALMO-EDA polarized state. Based on the analysis in Ref. 127 we choose to define DSs in terms of the ALMO-EDA definition of CT.

We took the Zundel 1B training set, identified the shared proton, H^* , in each configuration by minimizing the sum of the $O-H^*$ distances, $i = \arg \min_i ||\mathbf{X}_{O^{(1)}} - \mathbf{X}_{H_i}||_2 + ||\mathbf{X}_{O^{(2)}} - \mathbf{X}_{H_i}||_2$ (where $\mathbf{X}$ denotes the position vector of a given atom, and i labels the five Zundel hydrogens). We then created two DS training sets with the atoms permuted for the low energy and high energy DSs, i.e. atom orders: $O^{(1)}H_2^{(O^{(1)})}H^*O^{(2)}H_2^{(O^{(2)})}$ and $O^{(2)}H_2^{(O^{(2)})}H^*O^{(1)}H_2^{(O^{(1)})}$ (where $H_2^{(O^{(1)})}$ and $H_2^{(O^{(2)})}$ denote the two flanking hydrogens bound to $O^{(1)}$ and $O^{(2)}$, respectively). Note that these two training sets (for DS 1 and DS 2) correspond to the same nuclear configurations (and therefore the same adiabatic ground state), but have different assignments of monomer units (the first ordering defining $O^{(1)}$ to be the H_3O^+ oxygen, and the second ordering defining $O^{(2)}$ to be the H_3O^+ oxygen), resulting in different values of the associated ALMO-EDA CT energies. The two DS training sets correspond to negative and positive PTC values, respectively, with the intent of separating the low energy DS configurations from the high energy DS configurations into their own training sets, although this assumes that the sign of the PTC determines the DS binding energy, which is not strictly true, in general.¹ Logarithmic scale histograms of the ALMO-EDA CT energies ($\omega B97M-V/aug-cc-pVTZ$) with respect to the reference (adiabatic ground state) binding energies and the PTC for these two DS training sets are shown in Fig. 7.3. Note that

¹The use of the difference in total 1B energy, or the difference in CT energy, between two DSs may be a more useful reaction “coordinate,” however we simply use the geometric PTC as a reaction coordinate here for simplicity.

negative PTC (roughly) corresponds to the low energy DS, hence the CT energies tends to 0 with decreasing PTC as the monomer unit definitions become less ambiguous, with the low energy DS binding energy converging to the adiabatic ground state binding energy.

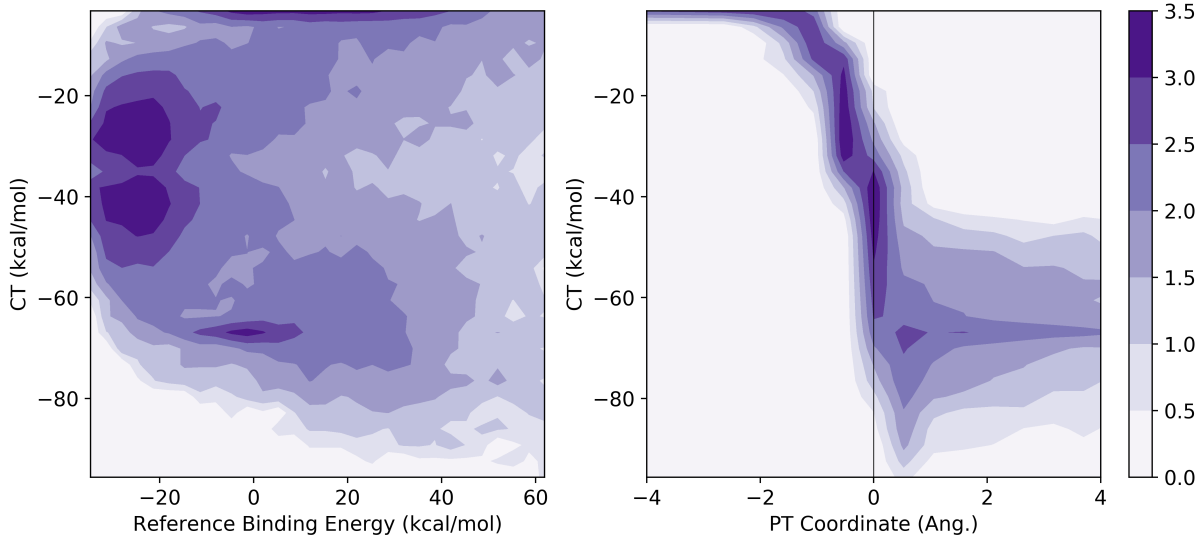


Figure 7.3. Logarithmic scale histogram (calculated as: $z(x,y) = \log_{10}(h(x,y) + 1)$) of the ALMO-EDA CT energy with respect to the reference (adiabatic ground state) binding energy and the PTC for all configurations in the Zundel 1B training set. Note that in PTC histogram, negative PTC values represent the low energy DS, hence the CT energies converge to 0 as the PTC becomes more negative.

In the interest of making the coupling potential close-range and small in magnitude, we introduce a model DS binding energy potential, $V_{\text{BE, model}}(\mathbf{X}, x)$, defined by subtracting the ALMO-EDA CT energy, $E_{\text{CT}}(\mathbf{X}, x)$, scaled by a switching function, $s(x)$, from the adiabatic ground state binding energy, $V_{\text{BE, adiabatic}}(\mathbf{X})$, where $V_{\text{BE, adiabatic}}(\mathbf{X})$ only depends on the nuclear configuration, X , but not the monomer definitions, the switching function, $s(x)$, depends on the PTC, x , but not X , and $E_{\text{CT}}(\mathbf{X}, x)$ depends on the nuclear configuration, X , as well as the monomer definitions which we can imagine being parameterized by the PTC, x (although the PTC is not directly used in the calculation of the ALMO-EDA CT energy). This model DS binding energy is defined as:

$$V_{\text{BE, model}}(\mathbf{X}, x) = V_{\text{BE, adiabatic}}(\mathbf{X}) - (s(x)E_{\text{CT}}(\mathbf{X}, x)) \quad (7.1)$$

with the switching function, $s(x)$ defined as:

$$s(x) = \frac{c}{2} \tanh(ax - b) + \frac{c}{2} \quad (7.2)$$

Here, the switching parameter, a (taken to be non-negative), corresponds to the steepness of the curve, parameter, b , corresponds to the x -value of the inflection point (the center of the curve), and c corresponds to the asymptote in the limit of large x , while the asymptote in the limit of small x remains 0 (switched off) for positive values of a . Note that parameters $a = 0$, $b = 0$, $c = 2$ correspond to $s(x) = 1$ for all x , and makes the model binding energy simply equal to $V_{\text{BE, adiabatic}}(\mathbf{X}) - E_{\text{CT}}(\mathbf{X}, x)$. Examples of the shape of the switching function for a few values of switch parameters a and c are shown in Fig. 7.4 (panels a, c, e, g, and i).

Having computed the reference adiabatic ground state binding energies as well as the DS 1 and DS 2 CT energies for each configuration in the Zundel 1B training set, we can compute the exact coupling potential values, $v_{12}(\mathbf{X}, x)$, for the 2-state VB model for each parameterization of our model DS binding energy potential (Eq. 7.1) through the secular equation for 2-dimensional real symmetric (Hermitian) matrices:

$$v_{12}(\mathbf{X}, x) = [(\varepsilon_1(\mathbf{X}, -x) - E_0(\mathbf{X}))(\varepsilon_2(\mathbf{X}, x) - E_0(\mathbf{X}))]^{\frac{1}{2}} \quad (7.3)$$

where $\varepsilon_1(\mathbf{X}, -x)$ and $\varepsilon_2(\mathbf{X}, x)$ correspond to the DS binding energies for DS 1 and DS 2, respectively, (i.e. the diagonal matrix elements of the 2-state VB Hamiltonian in the basis of the two DSs), and are calculated as $V_{\text{BE, model}}(\mathbf{X}, x)$ from Eq. 7.1 (with the dependence on $-x$ and x emphasizing that the two DSs have the same PTC magnitudes, but with opposite signs), and $E_0(\mathbf{X}) = V_{\text{BE, adiabatic}}(\mathbf{X}, x)$ is the reference adiabatic ground state binding energy. Note that substituting equivalent terms, and using the definition of $\varepsilon(\mathbf{X}, x) = V_{\text{BE, model}}(\mathbf{X}, x)$ (Eq. 7.1), we

can simplify Eq. 7.3 as:

$$v_{12}(\mathbf{X}, x) = [s(-x)E_{\text{CT}}(\mathbf{X}, -x)s(x)E_{\text{CT}}(\mathbf{X}, x)]^{\frac{1}{2}} \quad (7.4)$$

with $E_{\text{CT}}(\mathbf{X}, x)$ and $E_{\text{CT}}(\mathbf{X}, -x)$ corresponding to the ALMO-EDA CT energies for the low energy and high energy DSs, respectively. Note that Eq. 7.4 doesn't depend directly on $E_0(\mathbf{X})$, due to the fact that the dependence of the reference adiabatic ground state energy on the eigenvalues of the VB Hamiltonian are localized in the diagonal matrix elements, defined by Eq. 7.1, while the dependence on v_{12} is implicit.

Logarithmic scale histograms of $v_{12}(\mathbf{X}, x)$ with respect to the PTC for a few different switching function parameterizations are shown in Fig. 7.4 (panels b, d, f, h, and j) along with plots of the corresponding switching functions. Note that panels a and b correspond to the model where no switching function is used at all (with switch parameters $a = 0$, $b = 0$, $c = 2$). From these histograms, it can be seen that increasing switch parameter, a (the steepness), reduces the tail of $v_{12}(\mathbf{X}, x)$ in the limit of large magnitude PTC (x), while switch parameter, c (the large PTC asymptote), reduces the magnitude of $v_{12}(\mathbf{X}, x)$.

From these histograms, it seems that the switching function with large a and small c result in a model DS PEF that is close-range and small in magnitude. However, preliminary applications of our VB model to the protonated water clusters studied in Sec 3.3 seem to indicate that higher accuracy is achieved through small a and large c .

7.2.2 Preliminary Tests of MB-VB on Protonated Water Clusters

Fitting an MB-nrg PEF to model DS energies in the form of Eq. 7.1 is challenging, but appears possible.² However, in the interest of understanding the accuracy of the approach under ideal settings, we do not attempt to fit PEFs to the DS energies, nor to the coupling potentials, and

²To ensure accuracy in the low energy DS region of configuration space, one might need to apply a weighting function to the LLS cost function based on the PTC or some other reaction coordinate (possibly taking a form similar to Eq. 7.2, but with a negative a parameter).

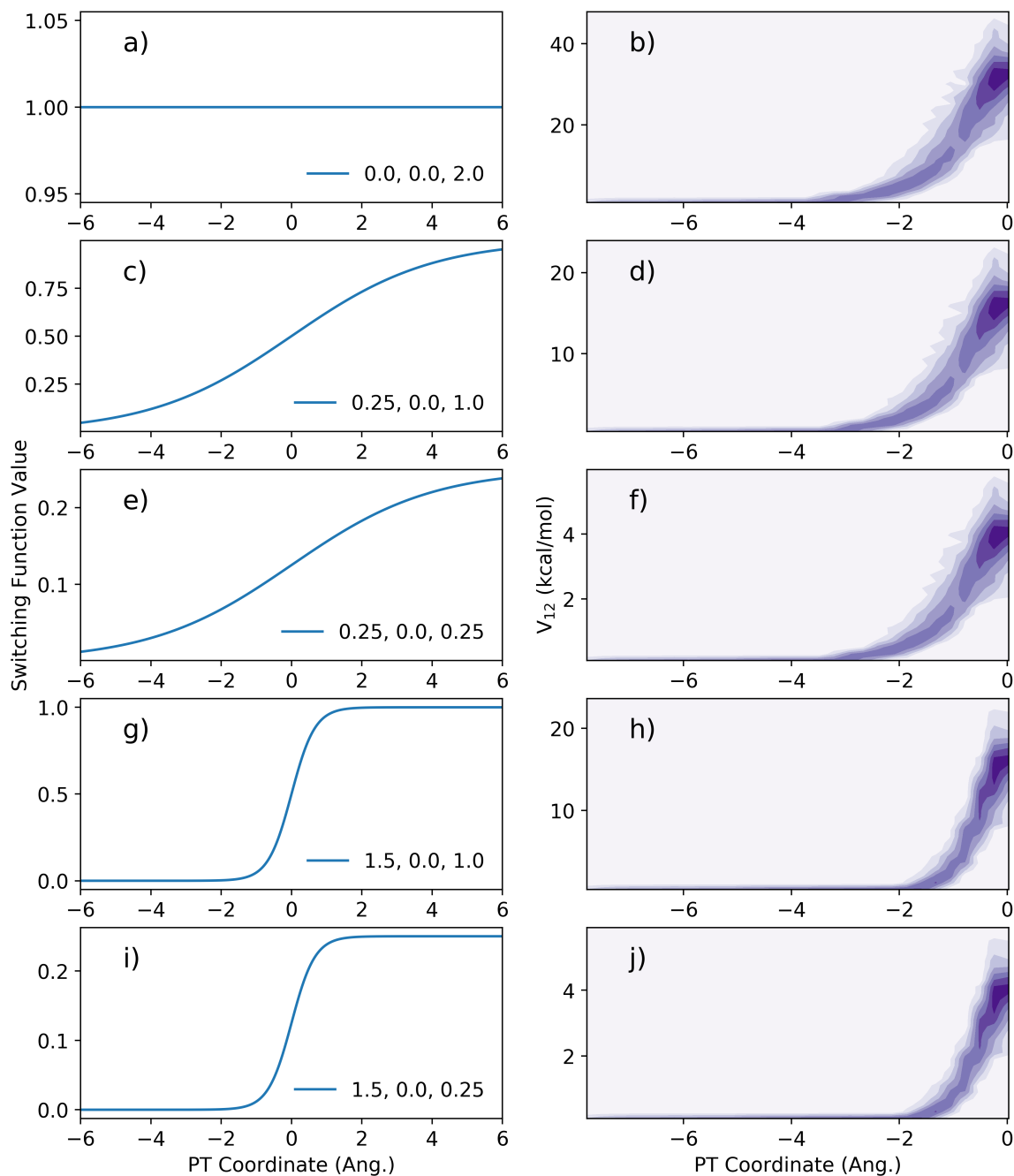


Figure 7.4. Switching function plots (Eq. 7.2) for different values of a and c parameters, and logarithmic scale histograms (calculated as: $z(x,y) = \log_{10}(h(x,y) + 1)$) of the coupling potential with respect to the PTC for each nuclear configuration in the Zundel 1B training set using the model binding energy potential from Eq. 7.1 for each set of switching function parameters. Panels (a) and (b) correspond to the model where no switching function is used at all. Note that increasing switch parameter, a (the steepness), reduces the tail of v_{12} in the limit of large magnitude PTC, while switch parameter, c (the large PTC asymptote), reduces the magnitude of v_{12} .

instead construct the VB Hamiltonians from 2B energies taken directly from electronic structure calculations supplemented with higher-body electrostatics from the long range H_3O^+ –water potential, and with the full MB-pol water potential (1B + 2B + 3B + higher-body electrostatics).

The procedure used is as follows. First, each cluster configuration is processed by the VB state search algorithm, only considering states formed by PT between H_3O^+ hydrogens and water oxygens at a distance of at most 1.8 Å. Each DS found by the algorithm has a single well-defined covalent bonding connectivity (Fig. 7.5), and is represented by an atom permutation into an $\text{OH}_3\text{OH}_2\text{OH}_2\dots$ ordering. The total system 1B, total water-water 2B, and total system higher-body (3B+) energy is evaluated for each DS (i.e. the total system binding energy without any 2B H_3O^+ –water energy contribution). Next, for each DS, all H_3O^+ –water dimer binding energies and ALMO-EDA CT energies are calculated with $\omega\text{B97M-V/aug-cc-pVTZ}$, and the 2B interaction energy is computed with Eq. 7.1 (subtracting the total 1B distortion energy from the resulting binding energy). The sum of the 2B H_3O^+ –water interaction energies for each DS (from Eq. 7.1) are added to the total system 1B, water-water 2B, and higher body energy for the corresponding DS, resulting in the VB Hamiltonian diagonal matrix elements (Fig. 7.5). Finally, for each pair of DSs differing by two covalent bonds, the corresponding coupling (off-diagonal) matrix element is computed by taking the relevant O_2H_5^+ (Fig. 7.5), and evaluating Eq. 7.4 with the two DS 2B H_3O^+ –water ALMO-EDA CT energies.

As mentioned above, we believe that the ideal MB-VB model would have coupling potentials that are close-range and small in magnitude, with the low energy DS being a good approximation to the adiabatic ground state without any coupling to the other DSs. In order to test whether meeting these conditions yields the highest accuracy, we applied MB-VB models with various sets of a and c switching function parameters (with $b = 0$) to the protonated water clusters listed in Table 7.1. This comparison also tests the sensitivity of the results with respect to the model parameterization.

In more comprehensive future studies, it will be important to evaluate the accuracy of the energy gradients (with respect to nuclear coordinates) of the MB-VB models, but in this

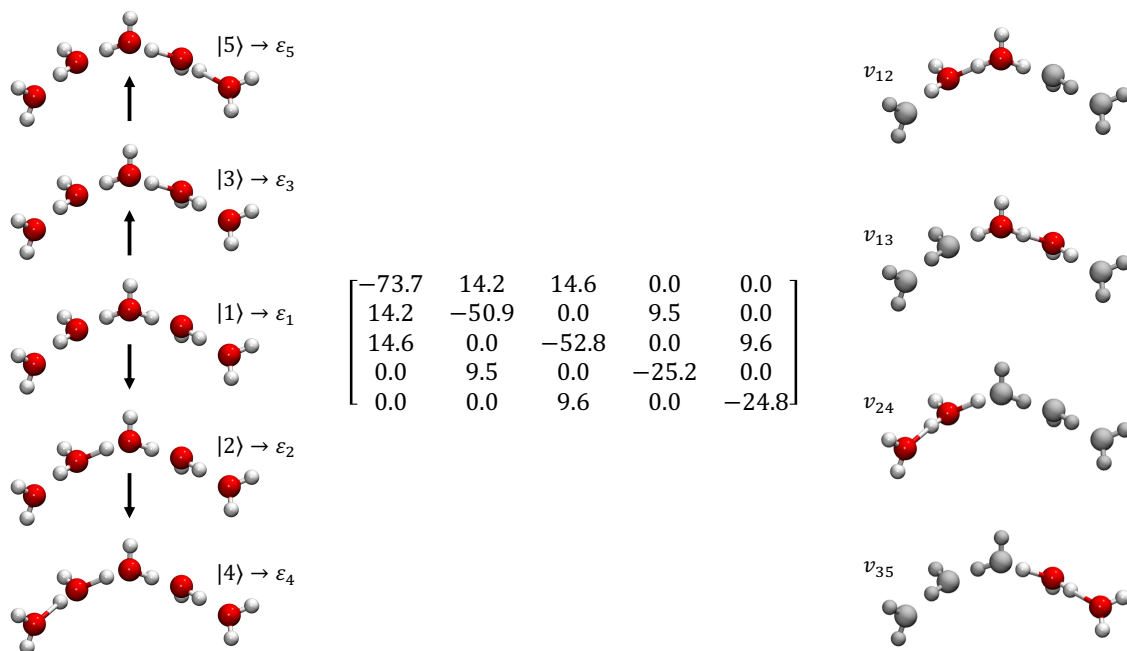


Figure 7.5. Example VB Hamiltonian for $\text{H}_3\text{O}^+(\text{H}_2\text{O})_4$ cluster 5β (Fig. 3.15) in kcal/mol using Eq. 7.1 with switch parameters $a = 0.25$, $b = 0.0$, $c = 1.0$. Covalent bond connectivities for each DS are shown on the left, and dimers used in each coupling matrix element calculation are shown on the right (gray molecules are not used in the calculations, and are only shown for context).

preliminary test, we simply evaluate the accuracy of the models in reproducing binding energy differences between pairs of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ conformers. Specifically: for each cluster size, $n + 1$ (the number of monomers including H_3O^+), we evaluated the MB-VB binding energy on each $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ conformer (conformers are labeled as $C_i^{(n)}$, indexed by i). We then computed the difference in MB-VB energies between each pair of conformers, and subtracted the difference in reference binding energies for the same pair. The absolute values of these differences of binding energy differences were summed, and the sums were normalized by dividing by the number of pairs of $(n + 1)$ -monomer conformers, and by dividing by the number of monomers, $n + 1$, resulting in an error per monomer (per conformer pair). The errors for each cluster size were

summed, yielding the total error shown in Eq. 7.5.

$$\begin{aligned} \text{Error} &= \sum_{n=3}^5 \frac{1}{n+1} \sum_{C_i^{(n)}} \sum_{C_j^{(n)} \neq C_i^{(n)}} \binom{N_n}{2}^{-1} |\Delta_{ij}^{\text{ref}} - \Delta_{ij}^{\text{MB-VB}}| \\ \Delta_{ij}^{\text{ref}} &= V_{\text{BE, ref}}(C_i^{(n)}) - V_{\text{BE, ref}}(C_j^{(n)}) \\ \Delta_{ij}^{\text{MB-VB}} &= V_{\text{BE, MB-VB}}(C_i^{(n)}) - V_{\text{BE, MB-VB}}(C_j^{(n)}) \end{aligned} \quad (7.5)$$

The relative energy errors with respect to $\omega\text{B97M-V/aug-cc-pVTZ}$ reference binding energies (calculated with Eq. 7.5) for several MB-VB model parameterizations are shown in the heat map in Fig. 7.6. Based on these results, it appears that, on average, the highest accuracy is achieved through a small a (steepness) parameter, and a large c (asymptote) parameter. Deviations of the MB-VB models from the reference binding energies for the tested sets of switching function parameters for each individual protonated water cluster are shown in Figs. D.1 – D.10. Deviations from the reference binding energies for the MB-VB model defined by switch parameters $a = 0.25$, $b = 0$, and $c = 1$ are listed in Table 7.1.

It is important to note that these preliminary tests use the MB-pol water potential for all water 1B distortion energies, and 2B water–water and 3B water–water–water interaction energies, which is fitted to CCSD(T) reference energies, and therefore may deviate slightly from the $\omega\text{B97M-V/aug-cc-pVTZ}$ binding energies used here as reference values. In particular, the error of 0.64 kcal/mol in the $\text{H}_3\text{O}(\text{H}_2\text{O})$ 2-mer shown in Table 7.1 is due entirely to errors in the (total) 1B distortion energy, since we use $\omega\text{B97M-V/aug-cc-pVTZ}$ directly in our preliminary MB-VB test calculations, specifically for the $\text{H}_3\text{O}-\text{H}_2\text{O}$ 2B energies and v_{12} coupling potentials.

Overall, this initial model shows that there is some potential in generalizing the MB-nrg methodology towards VB models for simulating aqueous PT reactions with molecular dynamics.

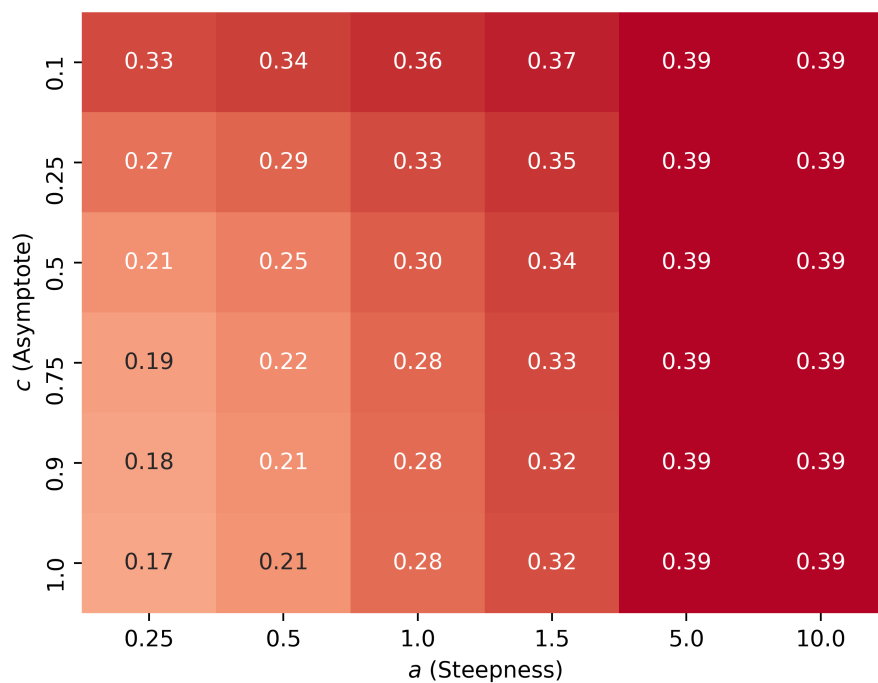


Figure 7.6. Heat map showing relative accuracy of MB-VB models compared to ω B97M-V/aug-cc-pVTZ reference binding energies with various sets of switching function parameters evaluated on the protonated water clusters listed in Table 7.1. Errors are calculated with Eq. 7.5, and have units of kcal/mol/cluster pair/monomer.

Table 7.1. MB-VB binding energies (in kcal/mol) for some of the low-lying isomers of $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters shown in Figs. 3.20 (panels a and b) and 3.15, and their deviations relative to $\omega\text{B97M-V/aug-cc-pVTZ}$ binding energies. 2B H_3O^+ –water energies evaluated directly from ALMO-EDA using Eq. 7.1 with switch parameters $a = 0.25$, $b = 0.0$, $c = 1.0$.

Cluster	MB-VB binding energy	Deviation from $\omega\text{B97M-V/aug-cc-pVTZ}$ reference
2-mer	−34.06	0.64
3-mer	−57.64	0.51
4-mer β	−72.84	1.32
4-mer δ	−73.10	0.68
4-mer γ	−78.08	−0.33
5-mer β	−86.32	1.51
5-mer δ	−92.19	−0.20
5-mer γ	−89.29	0.36
5-mer ε	−91.03	0.21
6-mer α	−102.64	0.57
6-mer β	−104.29	0.06
6-mer γ	−104.39	0.55

Chapter 8

Conclusions

Here, we have demonstrated a number of decomposition schemes and related tools for studying the fundamental components of many-body chemical systems, and applied them to study systems such as gas phase aqueous clusters, lattice systems with many-body interactions, the Zundel ion dissolved in acetonitrile, and protonated water. In particular, we have shown how many-body energy decomposition analysis can illuminate the physical cause of cooperative interactions in protonated and deprotonated water clusters, and introduced the concept of fragment group decompositions to facilitate the analysis. Additionally, charge transfer energies obtained from energy decomposition analysis were a critical component in our valence bond model for aqueous proton transfer. Scaling relations for many-body energies in lattice systems were developed through the extension of the fragment group decomposition. Design principles for developing model potential energy functions within the TTM-nrg and MB-nrg frameworks were outlined. Decompositions based on graph connectivity were shown to increase the computational efficiency of permutationally invariant polynomials. Power spectrum decompositions were used to analyze the vibrational properties of a model for the Zundel ion dissolved in acetonitrile. The preceding work lays out many guiding steps for the development of improved decomposition schemes for the use in many-body chemical applications.

Appendix A

Electronic Structure Calculation Details

A.1 Protonated and Deprotonated Water Clusters

Each cluster was optimized at the RI-MP2[196, 197] level of theory with the aug-cc-pVTZ[286] basis set using a gradient optimization-threshold of 1.0×10^{-6} a.u., an optimization step-size of 1.0×10^{-6} a.u., and a gradient precision of 1.0×10^{-8} a.u.

When calculating many-body energies, in order to guarantee chemical accuracy at all orders, reference energies for the individual MBE terms used in the analysis reported in Sec. 3.2 were calculated using the SAMBA method[46] described in Sec. 3.1.

The reference 2B interaction energies were computed at the coupled cluster with single, double, and perturbative triple excitations, i.e., CCSD(T), level of theory[287, 288] extrapolated to the complete basis set limit (CBS). The energies obtained with the augmented, correlation-consistent triple (aug-cc-pVTZ) and quadruple (aug-cc-pVQZ) basis sets[286] were extrapolated via the two-point extrapolation recommended in Ref. 46 for CCSD(T) calculations,

$$E_{HF}^{CBS} = E_{HF}^X - \left(\frac{E_{HF}^X - E_{HF}^Y}{e^{-\alpha\sqrt{X}} - e^{-\alpha\sqrt{Y}}} \right) e^{-\alpha\sqrt{X}} \quad (\text{A.1})$$

$$E_{corr}^{CBS} = \frac{Y^\beta E_{corr}^Y - X^\beta E_{corr}^X}{Y^\beta - X^\beta} \quad (\text{A.2})$$

Here, the Hartree Fock (Eq. A.1) and correlation (Eq. A.2) contributions to the total CCSD(T)

energy are extrapolated separately. In both equations, X denotes the lower-basis set cardinal number (e.g., $X = 3$ for aug-cc-pVTZ), and Y denotes the upper-basis set cardinal, while α and β are parameters that must be optimized for the given basis sets used in the extrapolation. Following Ref. 289, the two extrapolations were carried out with $\alpha = 5.790$ and $\beta = 3.050$, respectively.

Additionally, it was found necessary to correct for the basis set superposition error (BSSE) in computing all CCSD(T) reference energies. As discussed in Section 3.1, when using the SAMBA method, multiple BSSE correction schemes can be used for different terms in the MBE. Here, the reference 2B interaction energies were computed using K=2-centered CP corrections, and the 3B interaction energies were computed at the CCSD(T)/aug-cc-pVTZ level of theory using K=3-centered CP corrections. All reference energies for 4B and higher-order terms were computed within the SSCP scheme at the explicitly correlated, CCSD(T)-f12b/aug-cc-pVTZ level of theory.[290, 291]

To investigate the accuracy of different DFT exchange-correlation models in describing many-body effects in OH^- – water interactions, calculations with the aug-pc-3[292] basis set were performed with: BLYP,[103] PBE,[293] revPBE,[26] SCAN,[294] TPSS,[30] B3LYP,[295] PBE0,[34] revPBE0, M06-2X,[35] ω B97X-D,[296] and ω B97M-V.[31] Excluding ω B97X-D and ω B97M-V, all DFT energies were calculated with and without Grimme’s D3(0) empirical dispersion correction (GD3),[297] so that the effect of the GD3 corrections could be evaluated along with the corresponding functionals.

In the analysis of 3B interactions, we made use of the (1st generation) absolutely-localized molecular orbital energy decomposition analysis (ALMO EDA) method,[136] including the associated charge transfer analysis (CTA),[137] to dissect 3B interactions into individual, physical contributions. All ALMO EDA calculations were performed with ω B97M-V in combination with the Karlsruhe def2-TZVPPD[298, 299] basis set. ω B97M-V was used because it yields, on average, energies closer to the SAMBA reference data than any other functional tested (see Secs. 3.2.4 and 3.3.4), and the def2-TZVPPD basis set because it is large enough to fully capture

charge transfer (CT) effects, but still small to avoid inter-monomer-orbital linear-dependence, which can lead to unwanted muddling between polarization (E_{POL}^{ALMO}) and CT (E_{CT}^{ALMO}) ALMO EDA energy terms. The ALMO EDA method was devised in such a way that pairwise, directional (monomer $\rightarrow$ monomer) E_{CT}^{ALMO} (ΔE_{CT}^{ALMO}), and CT “electron reorganization” (ΔQ_{CT}), are well-defined, even in clusters with more than two monomers.[136, 137]

All CCSD(T) calculations and RI-MP2 optimizations were carried out using MOLPRO, version 2015.1.[300] DFT calculations were carried out with Gaussian 09[301] for all functionals except revPBE, revPBE0, SCAN, and ω B97M-V, for which Q-Chem 4.4[302] was used.

A.2 Halide–Water Clusters

All $X^-(H_2O)$ dimers were optimized at the RI-MP2[196, 197] level of theory with the aug-cc-pVTZ[286] basis set. Calculations involving Br^- and I^- used effective core potentials with 10 and 28 electrons, respectively.[303] A threshold of 1.0×10^{-6} a.u., a step size of 1.0×10^{-6} a.u., and a gradient precision of 1.0×10^{-8} a.u. were used in all optimizations.

Reference interaction energies for all $X^-(H_2O)$ dimers were calculated at the explicitly correlated coupled cluster level of theory, CCSD(T)-F12b,[290, 291] in the complete basis set (CBS) limit that was achieved via a two-point extrapolation[304] between the values obtained with the aug-cc-pVTZ and aug-cc-pVQZ basis sets,[286] using effective core potentials with 10 and 28 electrons[303] for calculations of clusters containing Br^- and I^- , respectively. All CCSD(T)-F12b and RI-MP2 calculations were performed using MOLPRO, version 2015.1.[300]

See the next section for a list of the XC-functionals used.

A.3 Alkali Ion–Water Clusters

$M^+(H_2O)$ and $M^+(H_2O)_2$ ($M = Li, Na, K, Rb, Cs$) cluster structures were obtained from MP2[305, 306] geometry optimizations using the aug-cc-pVTZ basis set[286, 307] for water, the cc-pwCVTZ basis set[308] for Li and Na, and the cc-pwCVTZ-PP basis set[309] for

K, Rb, and Cs using the ECP10MDF, ECP28MDF, and ECP46MDF effective core potentials, respectively.[310] A threshold of 1.0×10^{-6} a.u., a step size of 1.0×10^{-6} a.u., and a gradient precision of 1.0×10^{-8} a.u. were used in all optimizations.

Reference interaction energies were computed at the CCSD(T)[311, 288] level of theory in the complete basis set (CBS) limit using a two-point extrapolation[46] with the aug-cc-pVTZ and aug-cc-pVQZ basis sets for O and H, the cc-pwCVTZ and cc-pwCVQZ basis sets for Li and Na, and the cc-pwCVTZ-PP and cc-pwCVQZ-PP basis sets for K, Rb, and Cs, using the ECP10MDF, ECP28MDF, and ECP46MDF effective core potentials, respectively. In all of the CCSD(T) calculations, the valence electrons were correlated for O and H, while for the metals, the $n - 1$ valence shell was also correlated. All optimizations and energy calculations were performed using MOLPRO 2015.1.[300]

The ALMO-EDA 2 method[123] was used in all energy decompositions that were carried out using the following XC functionals: BLYP,[23, 24] PBE,[25] revPBE,[26] SCAN,[29] TPSS,[30] B97M-rV,[31, 32] B3LYP,[33] PBE0,[34] revPBE0, M06-2X,[35] ω B97X,[39] and ω B97M-V.[31, 41] Excluding B97M-rV and ω B97M-V, all DFT energies were calculated with and without the D3(0) empirical dispersion corrections.[297] Since ω B97X-D uses D2 empirical dispersion corrections[40] which are unavailable for Cs, we do not include ω B97X-D in our analysis, but do briefly discuss its accuracy in 2B energies of $\text{K}^+(\text{H}_2\text{O})$ and $\text{Rb}^+(\text{H}_2\text{O})$ dimers in Section 3.4.1. It should also be noted that in addition to using D2 corrections, the ω B97X-D functional itself has a different parameterization than that of ω B97X, which means that ALMO-EDA terms other than the DISP energy will differ between the ω B97X-D and ω B97X-D3. The effects of different dispersion corrections on alkali metal ion–water interactions are also investigated through the analysis of the D3(BJ)[118] corrections which is reported in the Supporting Information of Ref. [119]. All ALMO-EDA 2 calculations were carried out with Q-Chem 5.1[312] and used the same basis sets as those used in the reference CCSD(T) calculations.

Appendix B

Energy Decomposition Analysis Comparisons

B.1 Halide—Water Clusters

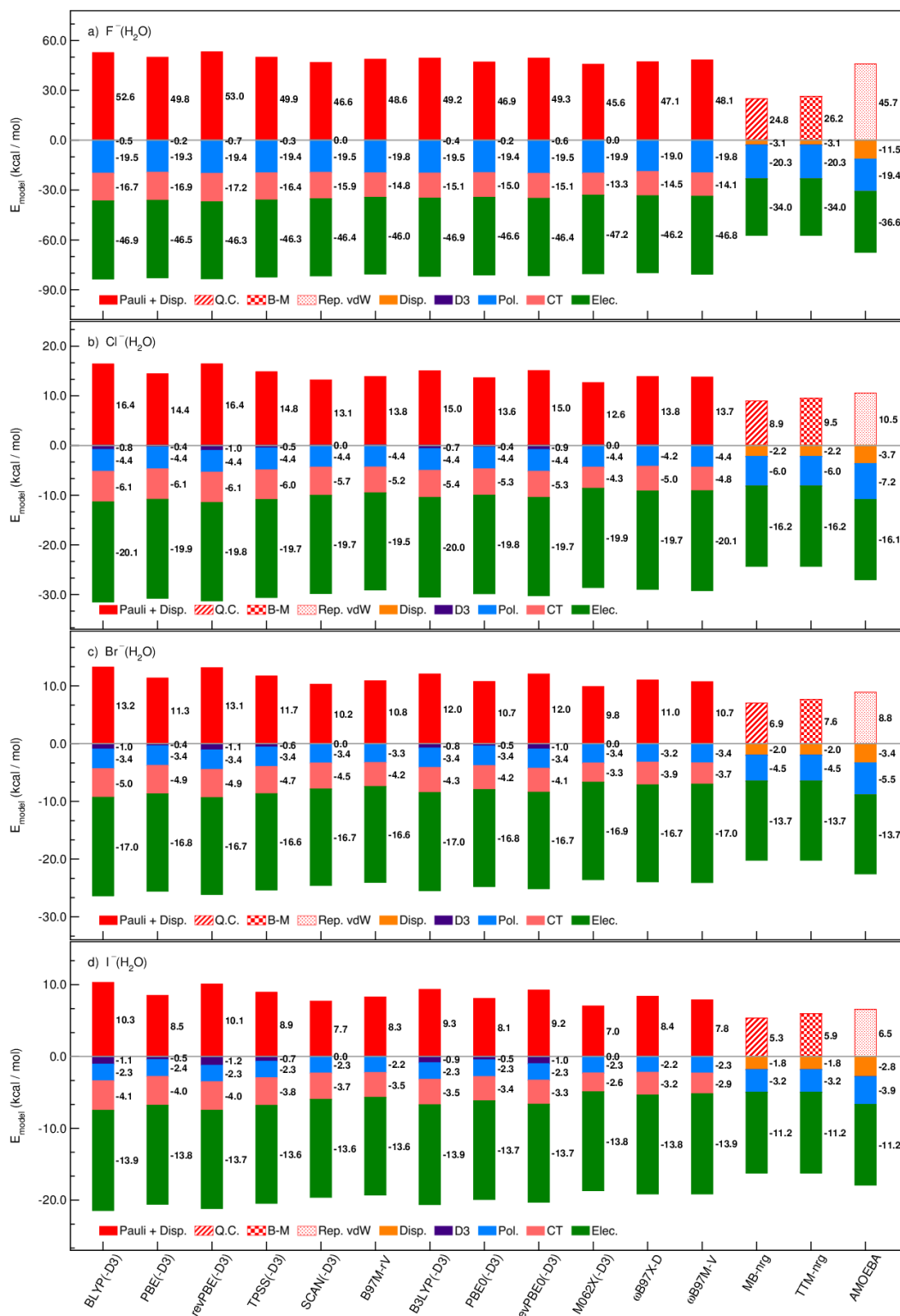


Figure B.1. Energy decompositions for each minimum energy X⁻(H₂O) dimer, with X = F (a), Cl (b), Br (c), and I (d), with numerical energy values (in kcal/mol) shown on the right-hand side of each bar.

B.2 Alkali Ion–Water Clusters

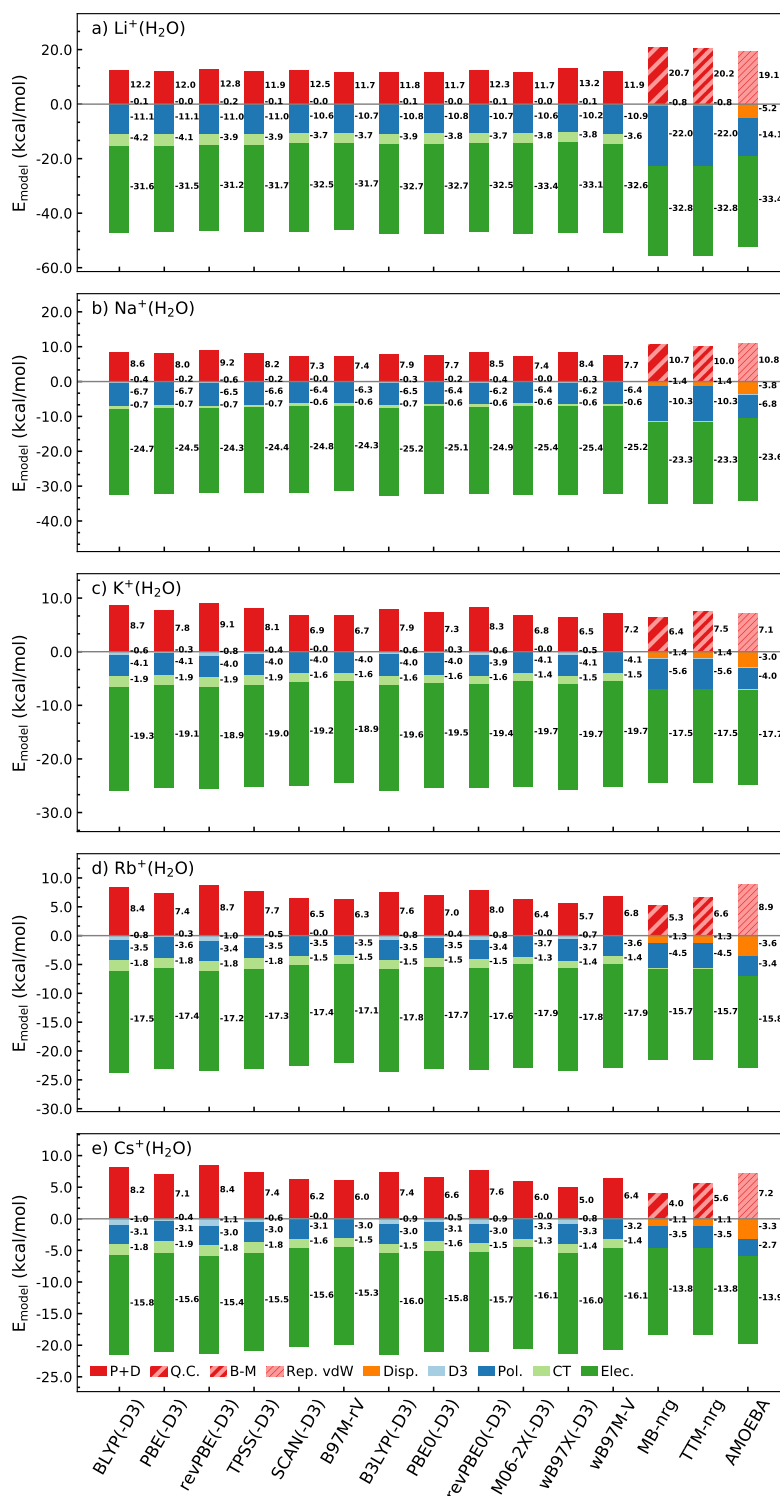


Figure B.2. Energy decompositions for each minimum energy $M^+(H_2O)$, with M = Li (a), Na (b), K (c), Rb (d), Cs (e), with numerical energy values (in kcal/mol) shown on the right-hand side of each bar.

Appendix C

Potential Energy Functions

C.1 Alkali Ion–Water PEF Descriptions

C.1.1 The AMOEBA force field

The Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) is a polarizable FF commonly used in MD simulations.[313, 314, 315, 316, 317, 318] A thorough description of the AMOEBA FF can be found in the original references.[313, 151] The AMOEBA functional form for the total interaction energy is expressed as:

$$V_{\text{INT}}^{\text{AMOEBA}} = V_{\text{ELEC}}^{\text{AMOEBA}} + V_{\text{POL}}^{\text{AMOEBA}} + V_{\text{Pauli+Disp}}^{\text{AMOEBA}} \quad (\text{C.1})$$

The AMOEBA FF employs classical Coulombic equations to describe the permanent electrostatic potential energy, $V_{\text{ELEC}}^{\text{AMOEBA}}$, using atom-centered point multipoles (monopoles through the quadrupole) calculated via distributed multipole analysis (DMA) from MP2/aug-cc-pVTZ.[313] The polarization energy, $V_{\text{POL}}^{\text{AMOEBA}}$, is treated explicitly, including self-consistent atom-centered induced point dipoles. Atomic isotropic polarizabilities are derived from fits to experimental data.[313, 151] To avoid overpolarization at close range, a Thole-type electrostatic damping scheme, with exponential-3 damping is used in $V_{\text{POL}}^{\text{AMOEBA}}$. [313, 173] None of the terms in $V_{\text{ELEC}}^{\text{AMOEBA}}$ are damped.

The remaining energy contribution, $V_{\text{Pauli+Disp}}^{\text{AMOEBA}}$, is modeled with the buffered 14-7

potential[319] to capture close-range pairwise additive interactions and can be decomposed into terms which mainly correspond to Pauli-repulsion and London dispersion:

$$\begin{aligned} V_{\text{Pauli+Disp}}^{\text{AMOEBA}} &= \sum_{i < j} \epsilon_{ij} \left(\frac{1 + \delta}{\rho_{ij} + \delta} \right)^7 \left(\frac{1 + \gamma}{\rho_{ij}^7 + \gamma} - 2 \right) \\ &= \sum_{i < j} \epsilon_{ij} \left(\frac{1 + \delta}{\rho_{ij} + \delta} \right)^7 \left(\frac{1 + \gamma}{\rho_{ij}^7 + \gamma} \right) - 2 \epsilon_{ij} \left(\frac{1 + \delta}{\rho_{ij} + \delta} \right)^7 \end{aligned} \quad (\text{C.2})$$

where $\rho = |\mathbf{R}_i - \mathbf{R}_j|/R_{ij}^0$. $|\mathbf{R}_i - \mathbf{R}_j|$ is the interatomic separation distance between atoms i and j , and R_{ij}^0 and ϵ_{ij} are flexible parameters, specific to the atomic species of atoms i and j , representing the minimum energy distance and energy well depth respectively. The values of the buffering constants are set to $\delta = 0.07$ and $\gamma = 0.12$, and were originally determined from fits to noble gas data.[319] The values for R_{ii}^0/R_{jj}^0 and $\epsilon_{ii}/\epsilon_{jj}$ for homonuclear dimers were obtained from fits to MP2/aug-cc-pVTZ energies and experimental data, while values for heteronuclear dimers are calculated from the combining rules:[313]

$$R_{ij} = \frac{(R_{ii}^0)^3 + (R_{jj}^0)^3}{(R_{ii}^0)^2 + (R_{jj}^0)^2} \quad (\text{C.3})$$

$$\epsilon_{ij} = \frac{4\epsilon_{ii}\epsilon_{jj}}{(\epsilon_{ii}^{1/2} + \epsilon_{jj}^{1/2})^2} \quad (\text{C.4})$$

C.1.2 The TTM-nrg PEF

Within the TTM-nrg PEF,[172, 165] water–water interactions are represented by the MB-pol PEF,[160, 159, 50, 161, 162] while ion–water interactions are represented by analytical functions representing ELEC, POL, DISP, and PAULI:

$$V_{\text{INT}}^{\text{TTM-nrg}} = V_{\text{ELEC}}^{\text{TTM-nrg}} + V_{\text{POL}}^{\text{TTM-nrg}} + V_{\text{PAULI}}^{\text{TTM-nrg}} + V_{\text{DISP}}^{\text{TTM-nrg}} \quad (\text{C.5})$$

TTM-nrg $V_{\text{ELEC}}^{\text{TTM-nrg}}$ and $V_{\text{POL}}^{\text{TTM-nrg}}$ terms use an extended Thole-type model[173] with exponential-4 damping to represent permanent and induced electrostatic interactions, respectively, as de-

scribed in detail in refs 172 and 165. The ion dipole polarizabilities were calculated at the CCSD(T)/cc-pwCV5Z level of theory.[165]

The TTM-nrg Pauli repulsion energy, $V_{\text{PAULI}}^{\text{TTM-nrg}}$, is represented by a pairwise sum of Born-Mayer functions

$$V_{\text{PAULI}}^{\text{TTM-nrg}} = A_{OX} e^{-b_{OX} R_{OX}} + A_{H_1X} e^{-b_{H_1X} R_{H_1X}} + A_{H_2X} e^{-b_{H_2X} R_{H_2X}} \quad (\text{C.6})$$

A_{ij} and b_{ij} in Eq. C.6 are flexible parameters, specific to the atomic species of atoms i and j , determined by fitting to reference CCSD(T) interaction energies (after subtracting $V_{\text{ELEC}}^{\text{TTM-nrg}}$ and $V_{\text{POL}}^{\text{TTM-nrg}}$). The TTM-nrg dispersion energy ($V_{\text{DISP}}^{\text{TTM-nrg}}$) is represented by the following function:

$$V_{\text{DISP}}^{\text{TTM-nrg}} = -f_6^{TT}(\delta_{XO} R_{XO}) \frac{C_{6,XO}}{R_{XO}^6} - f_6^{TT}(\delta_{XH_1} R_{XH_1}) \frac{C_{6,XH_1}}{R_{XH_1}^6} - f_6^{TT}(\delta_{XH_2} R_{XH_2}) \frac{C_{6,XH_2}}{R_{XH_2}^6} \quad (\text{C.7})$$

$C_{6,ij}$ in Eq. C.7, specific to atoms i and j , are fitted independently of the other terms using the exchange dipole method (XDM)[320, 321, 322] calculated with the Postg software.[183, 323] Damping is achieved using the Tang-Toennies ($f_6^{TT}(\delta_{\alpha\beta})$) damping scheme.[179] Additionally, the damping parameters, δ_{ij} , are constrained to be equal to the corresponding b_{ij} parameters in Eq. C.6.

C.1.3 The MB-nrg PEF

The MB-nrg PEFs for alkali ion–water interactions[166] are rigorously derived from the many-body expansion.[324] The MB-nrg PEFs adopt the same functional form and parameterizations as the TTM-nrg ELEC, POL, and DISP terms, $V_{\text{ELEC}}^{\text{TTM-nrg}}$, $V_{\text{POL}}^{\text{TTM-nrg}}$, and $V_{\text{DISP}}^{\text{TTM-nrg}}$, respectively, but replace the TTM-nrg Born-Mayer functions, $V_{\text{PAULI}}^{\text{TTM-nrg}}$, with PIPs[192] for ion–water 2B and ion–water–water 3B interactions, $V_{2B}^{\text{MB-nrg}}$ and $V_{3B}^{\text{MB-nrg}}$, respectively.

The PIPs can be thought of as quantum corrections to the underlying classical TTM-nrg-like potential, accounting for close-range PAULI and CT interactions, as well as correcting

deficiencies in the TTM-nrg ELEC and DISP terms.[120] The PIPs are turned on via sigmoidal “switching functions” at small intermonomer separations, leaving the remaining TTM-nrg terms to capture long range interactions, and many-body polarization.[163, 166] The PIPs are “permutationally invariant” with respect to interchangeable atoms in the system.[192] The PIPs are functions of the distances between the ion and the six sites of an MB-pol water molecule. The coefficients of both 2B and 3B PIPs were optimized using Tikhonov regression[195] (also known as ridge regression) to reproduce reference 2B and 3B energies calculated at the CCSD(T)-F12b[290, 291]/CBS level of theory for Li^+ –water and Na^+ –water interactions, and at the CCSD(T)/CBS level of theory for K^+ –water, Rb^+ –water, and Cs^+ –water interactions.[166]

Appendix D

MB-VB

D.1 Application of MB-VB Models With Various Switching Parameters on Protonated Water Clusters

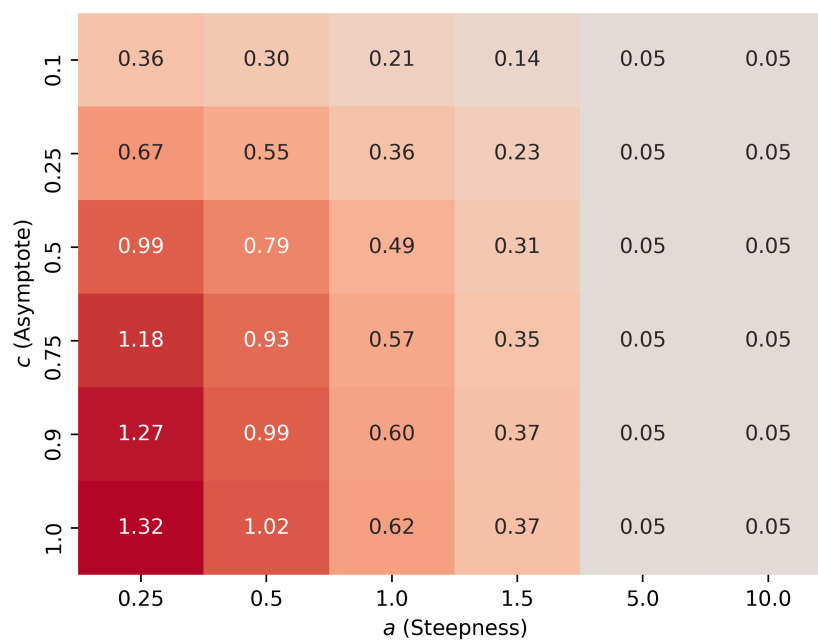


Figure D.1. Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ reference binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_4$ cluster, 4β .

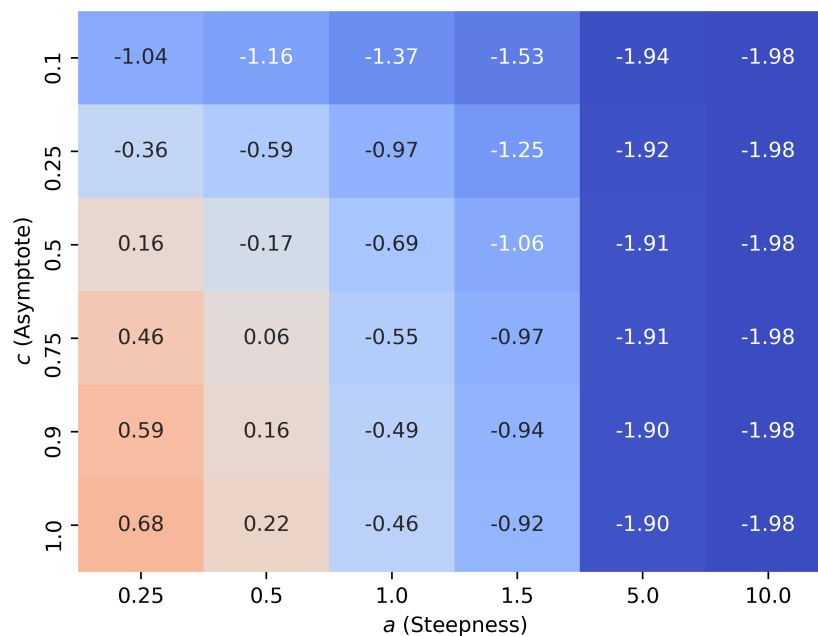


Figure D.2. Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_4$ cluster, 4δ .

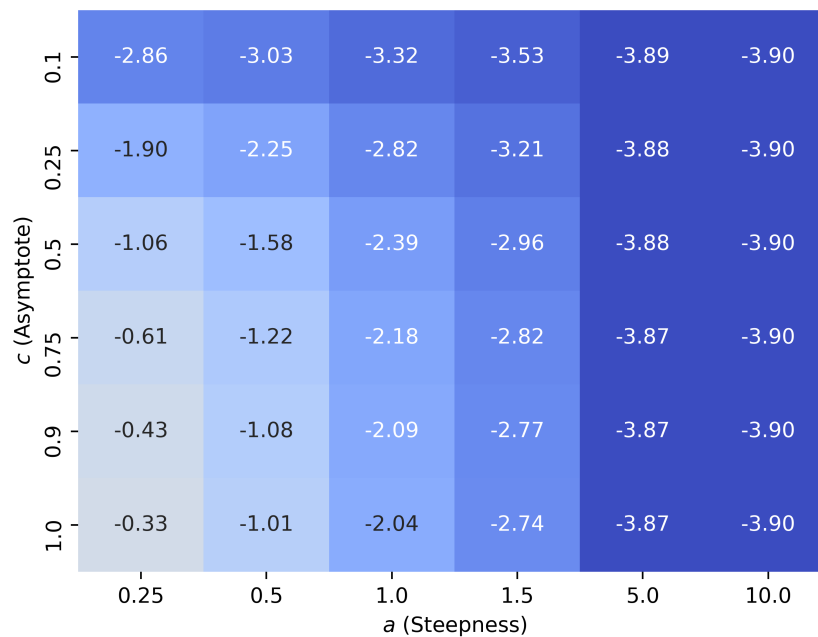


Figure D.3. Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_4$ cluster, 4γ .

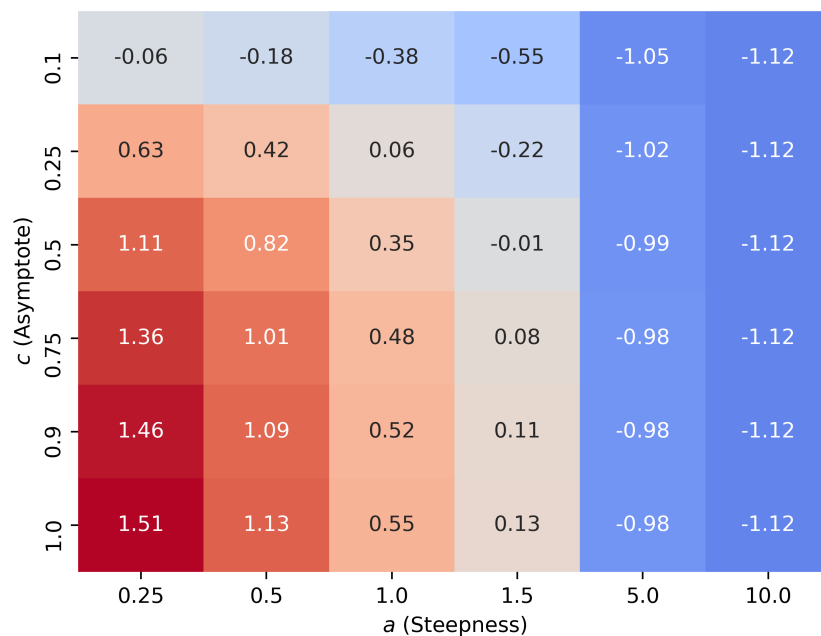


Figure D.4. Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ reference binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_5$ cluster, 5β .

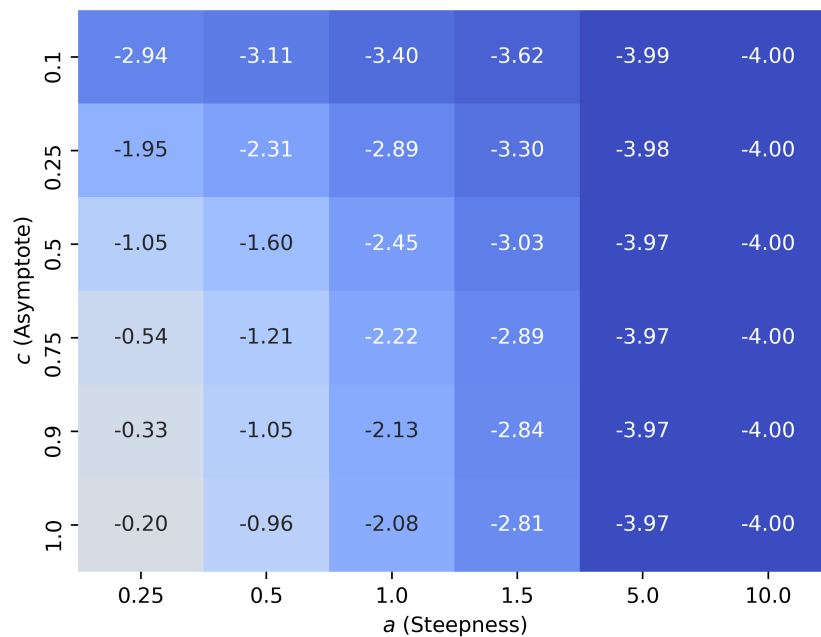


Figure D.5. Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ reference binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_5$ cluster, 5δ .

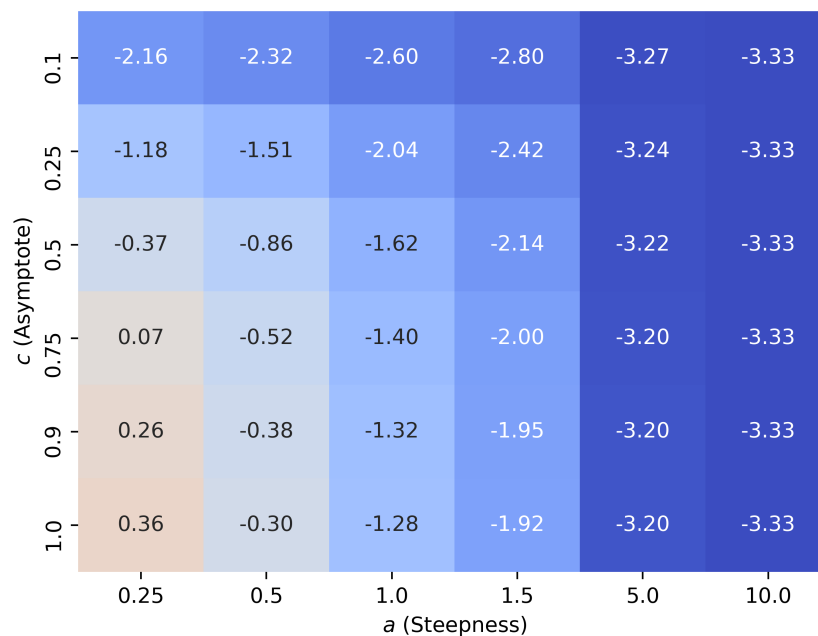


Figure D.6. Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_5$ cluster, 5γ .

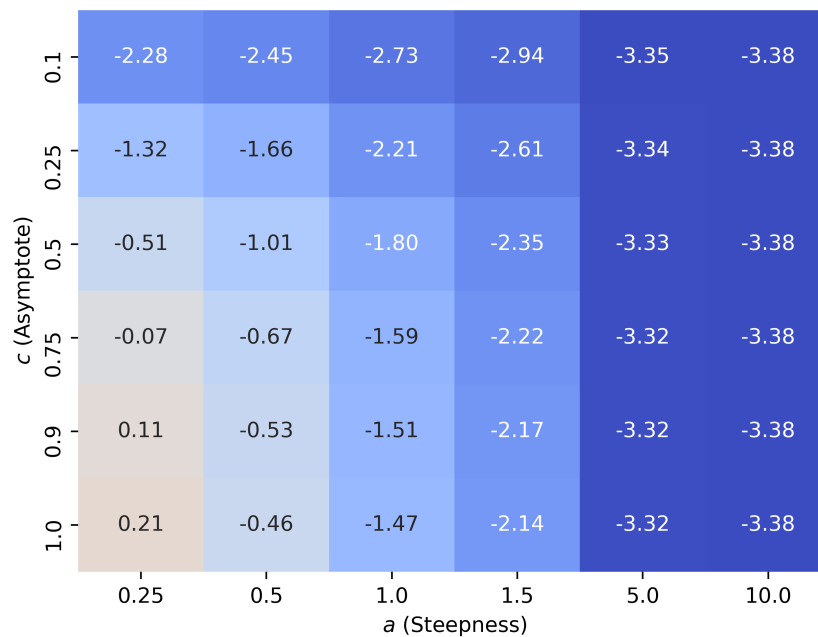


Figure D.7. Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_5$ cluster, 5ϵ .

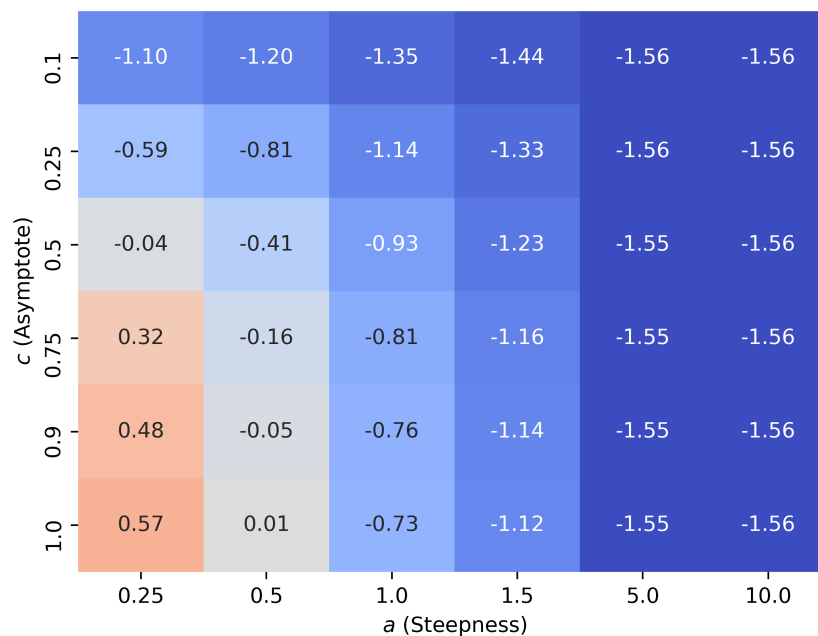


Figure D.8. Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_6$ cluster, 6α .

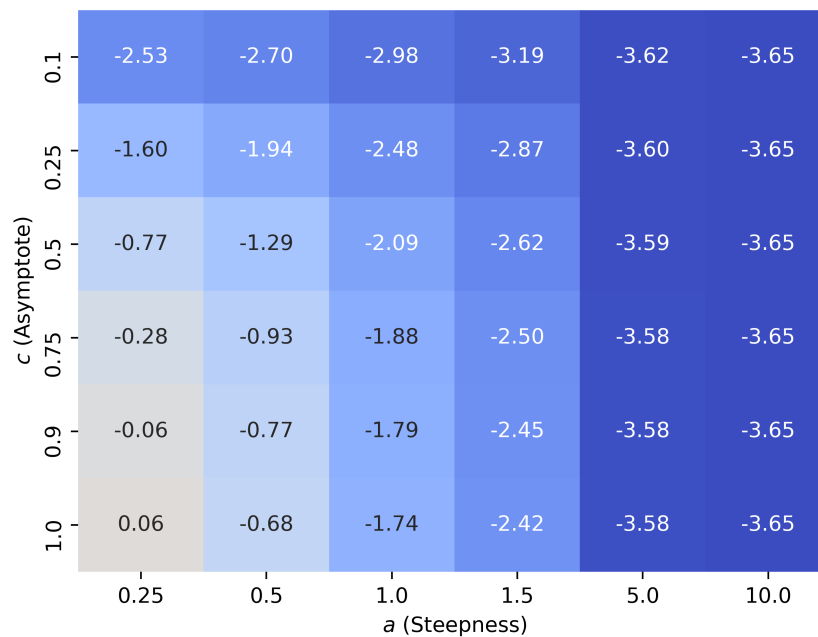


Figure D.9. Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_6$ cluster, 6β .

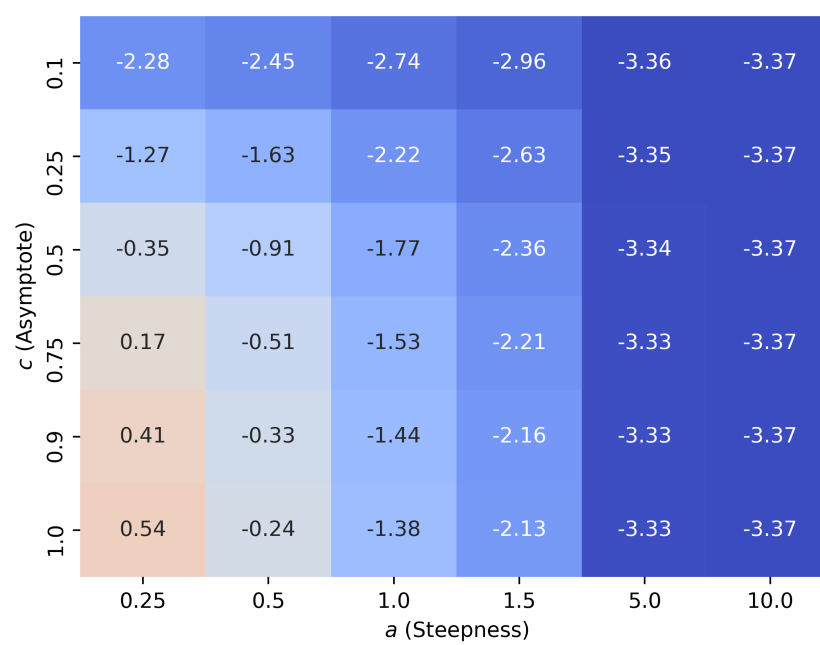


Figure D.10. Deviations of MB-VB models with various a and b switch parameters with respect to ω B97M-V/aug-cc-pVTZ references binding energies (kcal/mol) for $\text{H}_3\text{O}(\text{H}_2\text{O})_6$ cluster, 6γ .

Appendix E

Acknowledgements

Appendices A, B, and C are almost entirely reprints of the material as they appear in “Assessing Many-Body Effects of Water Self-Ions. I: $\text{OH}^-(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2018**, 14 (4), 1982–1997, “Assessing Many-Body Effects of Water Self-Ions. II: $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ Clusters” *J. Chem. Theory Comput.*, **2019**, 15 (9), 4816–4833, “Nature of Halide–Water Interactions: Insights from Many-Body Representations and Density Functional Theory” *J. Chem. Theory Comput.*, **2019**, 15 (5), 2983–2995 and “Nature of Alkali–Water Interactions: Insights from Many-Body Representations and Density Functional Theory II” *J. Chem. Theory Comput.*, **2020**, 16 (5), 3055–3072.

Bibliography

- [1] Felix Bloch. Heisenberg and the Early Days of Quantum Mechanics. *Phys. Today*, 29(12): 23–27, 1976.
- [2] Fabian Dahms, Benjamin P Fingerhut, Erik TJ Nibbering, Ehud Pines, and Thomas Elsaesser. Large-Amplitude Transfer Motion of Hydrated Excess Protons Mapped by Ultrafast 2D IR Spectroscopy. *Science*, 357(6350):491–495, 2017.
- [3] Ankit Rohatgi. Webplotdigitizer: Version 4.4. <https://automeris.io/WebPlotDigitizer>, 2020.
- [4] Colin K Egan and Francesco Paesani. Assessing Many-Body Effects of Water Self-Ions. I: $\text{OH}^-(\text{H}_2\text{O})_n$ Clusters. *J. Chem. Theory Comput.*, 14(4):1982–1997, 2018.
- [5] Memoirs of the Life, Writings, and Discoveries of Sir Isaac Newton, vol. 2 (Edinburgh: 1855) Chapter XVIII. <http://www.newtonproject.ox.ac.uk/view/texts/normalized/OTHE00079>. Accessed: 2021-05-30.
- [6] Attila Szabo and Neil S Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Courier Corporation, 2012.
- [7] Anthony Stone. *The Theory of Intermolecular Forces*. OUP Oxford, 2013.
- [8] Volkhard May and Oliver Kühn. *Charge and Energy Transfer Dynamics in Molecular Systems*. John Wiley & Sons, 2008.
- [9] Ivar Stakgold and Michael J Holst. *Green’s Functions and Boundary Value Problems*, volume 99. John Wiley & Sons, 2011.
- [10] Robert G Parr and Weitao Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press and Clarendon Press, 1989.
- [11] John P Perdew and Alex Zunger. Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems. *Phys. Rev. B*, 23(10):5048, 1981.

- [12] John P Perdew, Robert G Parr, Mel Levy, and Jose L Balduz Jr. Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy. *Phys. Rev. Lett.*, 49(23):1691, 1982.
- [13] Paula Mori-Sánchez, Aron J. Cohen, and Weitao Yang. Many-Electron Self-Interaction Error in Approximate Density Functionals. *J. Chem. Phys.*, 125(20):201102, 2006.
- [14] Diptarka Hait and Martin Head-Gordon. Delocalization Errors in Density Functional Theory Are Essentially Quadratic in Fractional Occupation Number. *J. Phys. Chem. Lett.*, 9(21):6280–6288, 2018.
- [15] Rodney J Bartlett and George D Purvis. Many-Body Perturbation theory, Coupled-Pair Many-Electron theory, and the Importance of Quadruple Excitations for the Correlation Problem. *Int. J. Quantum Chem.*, 14(5):561–581, 1978.
- [16] Rodney J Bartlett. Many-body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Ann. Rev. Phys. Chem.*, 32(1):359–401, 1981.
- [17] Oktay Sinanoğlu. Many-Electron Theory of Atoms and Molecules. I. Shells, Electron Pairs vs Many-Electron Correlations. *J. Chem. Phys.*, 36(3):706–717, 1962.
- [18] Oktay Sinanoğlu. Many-Electron Theory of Atoms and Molecules. *Proc. Natl. Acad. Sci. U. S. A.*, 47(8):1217, 1961.
- [19] Rodney J Bartlett, JD Watts, SA Kucharski, and J Noga. Non-Iterative Fifth-Order Triple and Quadruple Excitation Energy Corrections in Correlated Methods. *Chem. Phys. Lett.*, 165(6):513–522, 1990.
- [20] Pierre Hohenberg and Walter Kohn. Inhomogeneous Electron Gas. *Phys. Rev.*, 136(3B):B864, 1964.
- [21] Walter Kohn and Lu Jeu Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, 140(4A):A1133, 1965.
- [22] John P Perdew and Karla Schmidt. Jacob’s Ladder of Density Functional Approximations for the Exchange-Correlation Energy. In *AIP Conf. Proc.*, volume 577, pages 1–20. AIP, 2001.
- [23] Axel D Becke. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A*, 38(6):3098, 1988.
- [24] Chengteh Lee, Weitao Yang, and Robert G Parr. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B*, 37(2):785, 1988.

- [25] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77(18):3865, 1996.
- [26] Yingkai Zhang and Weitao Yang. Comment on “Generalized Gradient Approximation Made Simple”. *Phys. Rev. Lett.*, 80(4):890, 1998.
- [27] Andreas Görling and Mel Levy. Requirements for Correlation Energy Density Functionals from Coordinate Transformations. *Phys. Rev. A*, 45(3):1509, 1992.
- [28] Jianwei Sun, Richard C Remsing, Yubo Zhang, Zhaoru Sun, Adrienn Ruzsinszky, Haowei Peng, Zenghui Yang, Arpita Paul, Umesh Waghmare, Xifan Wu, Michael L. Klein Klein, and John P. Perdew. Accurate First-Principles Structures and Energies of Diversely Bonded Systems from an Efficient Density Functional. *Nat. Chem*, 8(9):831, 2016.
- [29] Jianwei Sun, Adrienn Ruzsinszky, and John P Perdew. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett*, 115(3):036402, 2015.
- [30] Jianmin Tao, John P. Perdew, Viktor N. Staroverov, and Gustavo E. Scuseria. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.*, 91(14):146401, 2003.
- [31] Narbe Mardirossian and Martin Head-Gordon. Mapping the Genome of Meta-Generalized Gradient Approximation Density Functionals: The Search for B97M-V. *J. Chem. Phys.*, 142(7):074111, 2015.
- [32] Narbe Mardirossian, Luis Ruiz Pestana, James C Womack, Chris-Kriton Skylaris, Teresa Head-Gordon, and Martin Head-Gordon. Use of the rVV10 Nonlocal Correlation Functional in the B97M-V Density Functional: Defining B97M-rV and Related Functionals. *J. Phys. Chem. Lett*, 8(1):35–40, 2016.
- [33] Axel D Becke. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.
- [34] Carlo Adamo and Vincenzo Barone. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.*, 110(13):6158–6170, 1999.
- [35] Yan Zhao and Donald G Truhlar. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.*, 120(1):215–241, 2008.
- [36] Hisayoshi Iikura, Takao Tsuneda, Takeshi Yanai, and Kimihiko Hirao. A Long-Range Correction Scheme for Generalized-Gradient-Approximation Exchange Functionals. *J. Chem. Phys.*, 115(8):3540–3544, 2001.

- [37] Takeshi Yanai, David P Tew, and Nicholas C Handy. A New Hybrid Exchange–Correlation Functional Using The Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.*, 393(1-3):51–57, 2004.
- [38] Narbe Mardirossian and Martin Head-Gordon. ω B97X-V: A 10-Parameter, Range-Separated Hybrid, Generalized Gradient Approximation Density Functional with Nonlocal Correlation, Designed by a Survival-of-the-Fittest Strategy. *Phys. Chem. Chem. Phys.*, 16(21):9904–9924, 2014.
- [39] Jeng-Da Chai and Martin Head-Gordon. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.*, 128(8):084106, 2008.
- [40] Jeng-Da Chai and Martin Head-Gordon. Long-Range Corrected Hybrid Density Functionals with Damped Atom–Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.*, 10(44):6615–6620, 2008.
- [41] Narbe Mardirossian and Martin Head-Gordon. ω B97M-V: A Combinatorially Optimized, Range-Separated Hybrid, Meta-GGA Density Functional with VV10 Nonlocal Correlation. *J. Chem. Phys.*, 144(21):214110, 2016.
- [42] Richard FW Bader. Atoms in Molecules. *Acc. Chem. Res.*, 18(1):9–15, 1985.
- [43] Richard FW Bader. A Quantum Theory of Molecular Structure and its Applications. *Chem. Rev.*, 91(5):893–928, 1991.
- [44] Ilya G Kaplan, Ruben Santamaria, and Octavio Novaro. Non-Additive Forces in Atomic Clusters: The Case of Ag_n . *Mol. Phys.*, 84(1):105–114, 1995.
- [45] Andreas Hermann, Robert P Krawczyk, Matthias Lein, Peter Schwerdtfeger, Ian P Hamilton, and James JP Stewart. Convergence of the Many-Body Expansion of Interaction Potentials: From van der Waals to Covalent and Metallic Systems. *Phys. Rev. A*, 76(1):013202, 2007.
- [46] Urszula Góra, Rafał Podeszwa, Wojciech Cencek, and Krzysztof Szalewicz. Interaction Energies of Large Clusters from Many-Body Expansion. *J. Chem. Phys.*, 135(22):224102, 2011.
- [47] Lars Ojamäe and Kersti Hermansson. *Ab Initio* Study of Cooperativity in Water Chains: Binding Energies and Anharmonic Frequencies. *J. Phys. Chem.*, 98(16):4271–4282, 1994.
- [48] Sotiris S Xantheas. *Ab Initio* Studies of Cyclic Water Clusters $(\text{H}_2\text{O})_n$, $n=1-6$. II. Analysis of Many-Body Interactions. *J. Chem. Phys.*, 100(10):7523–7534, 1994.
- [49] Sotiris S. Xantheas. Cooperativity and Hydrogen Bonding Network in Water Clusters.

Chem. Phys., 258(2):225–231, 2000.

- [50] Gregory R. Medders and Francesco Paesani. Many-Body Convergence of the Electrostatic Properties of Water. *J. Chem. Theory Comput.*, 9(11):4844–4852, 2013.
- [51] Gregory R. Medders, Volodymyr Babin, and Francesco Paesani. Development of a “First-Principles” Water Potential with Flexible Monomers. III. Liquid Phase Properties. *J. Chem. Theory Comput.*, 10(8):2906–2910, 2014.
- [52] Sandeep K Reddy, Shelby C. Straight, Pushp Bajaj, C. Huy Pham, Marc Riera, Daniel R. Moberg, Miguel A. Morales, Chris Knight, Andreas W. Götz, and Francesco Paesani. On the Accuracy of the MB-Pol Many-Body Potential for Water: Interaction Energies, Vibrational Frequencies, and Classical Thermodynamic and Dynamical Properties from Clusters to Liquid Water and Ice. *J. Chem. Phys.*, 145(19):194504, 2016.
- [53] Samuel F Boys and FJMP Bernardi. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.*, 19(4):553–566, 1970.
- [54] Bryan H. Wells and Stephen Wilson. van der Waals Interaction Potentials: Many-Body Basis Set Superposition Effects. *Chem. Phys. Lett.*, 101(4-5):429–434, 1983.
- [55] Julia C. White and Ernest R. Davidson. An Analysis of the Hydrogen Bond in Ice. *J. Chem. Phys.*, 93(11):8029–8035, 1990.
- [56] Pierre Valiron and István Mayer. Hierarchy of Counterpoise Corrections for N-Body Clusters: Generalization of the Boys-Bernardi Scheme. *Chem. Phys. Lett.*, 275(1-2):46–55, 1997.
- [57] Ren-Jie Lin, Quoc Chinh Nguyen, Yew-Soon Ong, Kaito Takahashi, and Jer-Lai Kuo. Temperature Dependent Structural Variations of $\text{OH}^-(\text{H}_2\text{O})_n$, $n=4-7$: Effects on Vibrational and Photoelectron Spectra. *Phys. Chem. Chem. Phys.*, 17(29):19162–19172, 2015.
- [58] Masato Morita and Kaito Takahashi. Multidimensional oh local mode calculations for $\text{oh}^-(\text{h}_2\text{o})_3$ —importance of intermode anharmonicity. *Phys. Chem. Chem. Phys.*, 15(1):114–124, 2013.
- [59] Elangannan Arunan, Gautam R. Desiraju, Roger A. Klein, Joanna Sadlej, Steve Scheiner, Ibon Alkorta, David C. Clary, Robert H. Crabtree, Joseph J. Dannenberg, Pavel Hobza, Henrik G. Kjaergaard, Anthony C. Legon, Benedetta Mennucci, and David J. Nesbitt. Definition of the Hydrogen Bond (IUPAC Recommendations 2011). *Pure Appl. Chem.*, 83(8):1637–1641, 2011.

- [60] Mark E Tuckerman, Dominik Marx, and Michele Parrinello. The Nature and Transport Mechanism of Hydrated Hydroxide Ions in Aqueous Solution. *Nature*, 417(6892):925, 2002.
- [61] Noam Agmon, Huib J Bakker, R Kramer Campen, Richard H Henchman, Peter Pohl, Sylvie Roke, Martin Thmer, and Ali Hassanali. Protons and Hydroxide Ions in Aqueous Systems. *Chem. Rev*, 116(13):7642–7672, 2016.
- [62] Mark E Tuckerman, Dominik Marx, Michael L Klein, and Michele Parrinello. On the Quantum Nature of the Shared Proton in Hydrogen Bonds. *Science*, 275(5301):817–820, 1997.
- [63] Claire C M Samson and Wim Klopper. *Ab Initio* Calculation of Proton Barrier and Binding Energy of the $(\text{H}_2\text{O})\text{OH}^-$ Complex. *J. Mol. Struc. Theochem.*, 586(1):201–208, 2002.
- [64] Joseph R Roscioli, Eric G Diken, Mark A Johnson, Samantha Horvath, and Anne B McCoy. Prying apart a water molecule with anionic h-bonding: A comparative spectroscopic study of the $x^- \cdot \text{H}_2\text{O}$ ($x = \text{oh}, \text{o}, \text{f}, \text{cl}, \text{and br}$) binary complexes in the $600\text{--}3800\text{ cm}^{-1}$ region. *J. Phys. Chem. A*, 110(15):4943–4952, 2006.
- [65] Naima Turki, Anne Milet, Ali Rahmouni, Ourida Ouamerali, Robert Moszynski, Elise Kochanski, and Paul E S Wormer. Theoretical study of the $\text{oh}^-(\text{h}_2\text{o})_2$ system: Nature and importance of three-body interactions. *J. Chem. Phys.*, 109(17):7157–7168, 1998.
- [66] Colin K Egan and Francesco Paesani. Assessing Many-Body Effects of Water Self-Ions. II: $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ Clusters. *J. Chem. Theory Comput.*, 15(9):4816–4833, 2019.
- [67] Henry S. Frank and Wen-Yang Wen. Ion-Solvent Interaction. Structural Aspects of Ion-Solvent Interaction in Aqueous Solutions: A Suggested Picture of Water Structure. *Discuss. Faraday Soc.*, 24:133–140, 1957.
- [68] Alan E. Reed and Frank Weinhold. Natural Bond Orbital Analysis of Near-Hartree–Fock Water Dimer. *J. Chem. Phys.*, 78(6):4066–4073, 1983.
- [69] Keiichi Ohno, Mari Okimura, Nobuyuki Akai, and Yukiteru Katsumoto. The Effect of Ccooperative Hydrogen Bonding on the OH Stretching-Band Shift for Water Clusters Studied by Matrix-Isolation Infrared Spectroscopy and Density Functional Theory. *Phys. Chem. Chem. Phys.*, 7(16):3005–3014, 2005.
- [70] C. J. Tainter and J. L. Skinner. The Water Hexamer: Three-Body Interactions, Structures, Energetics, and OH-Stretch Spectroscopy at Finite Temperature. *J. Chem. Phys.*, 137(10):104304, 2012.
- [71] Erika A. Cobar, Paul R. Horn, Robert G. Bergman, and Martin Head-Gordon. Examination

- of the Hydrogen-Bonding Networks in Small Water Clusters ($n = 2-5, 13, 17$) Using Absolutely Localized Molecular Orbital Energy Decomposition Analysis. *Phys. Chem. Chem. Phys.*, 14(44):15328–15339, 2012.
- [72] Shannon McDonald, Lars Ojamäe, and Sherwin J. Singer. Graph Theoretical Generation and Analysis of Hydrogen-Bonded Structures with Applications to the Neutral and Protonated Water Cube and Dodecahedral Clusters. *J. Phys. Chem. A*, 102(17):2824–2832, 1998.
- [73] Maihemutijiang Jieli and Misako Aida. Classification of OH Bonds and Infrared Spectra of the Topology-Distinct Protonated Water Clusters $\text{H}_3\text{O}^+(\text{H}_2\text{O})_{n-1}$ ($n \leq 7$). *J. Phys. Chem. A*, 113(8):1586–1594, 2009.
- [74] Jun-Ho Choi and Minhaeng Cho. Ion Aggregation in High Salt Solutions. II. Spectral Graph Analysis of Water Hydrogen-Bonding Network and Ion Aggregate Structures. *J. Chem. Phys.*, 141(15):154502, 2014.
- [75] Joseph P. Heindel, Qi Yu, Joel M. Bowman, and Sotiris S. Xantheas. Benchmark Electronic Structure Calculations for $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$, $n = 0-5$, Clusters and Tests of an Existing 1, 2, 3-Body Potential Energy Surface with a New 4-Body Correction. *J. Chem. Theory Comput.*, 14(9):4553–4566, 2018.
- [76] Jean Jules Fifen and Noam Agmon. Structure and Spectroscopy of Hydrated Sodium Ions at Different Temperatures and the Cluster Stability Rules. *J. Chem. Theory Comput.*, 12(4):1656–1673, 2016.
- [77] Revati Kumar, J. R. Schmidt, and James L. Skinner. Hydrogen Bonding Definitions and Dynamics in Liquid Water. *J. Chem. Phys.*, 126(20):05B611, 2007.
- [78] James A Platts and Keith E Laidig. Proton Transfer in Ionic Hydrogen Bonds. *J. Phys. Chem.*, 100(32):13455–13461, 1996.
- [79] Erica A Price, William H Robertson, Eric G Diken, Gary H Weddle, and Mark A Johnson. Argon predissociation infrared spectroscopy of the hydroxide–water complex ($\text{oh}^- \cdot \text{h}_2\text{o}$). *Chem. Phys. Lett.*, 366(3):412–416, 2002.
- [80] Gaétan Weck, Anne Milet, Robert Moszynski, and Elise Kochanski. Role of cancellation of errors in *ab initio* calculations: Structure and energetics of the $\text{oh}^-(\text{h}_2\text{o})$ system and electric dipole properties of the subsystems. *J. Phys. Chem. A*, 106(50):12084–12094, 2002.
- [81] Xinchuan Huang, Bastiaan J Braams, Stuart Carter, and Joel M Bowman. Quantum Calculations of Vibrational Energies of H_3O_2^- on an *Ab Initio* potential. *J. Am. Chem. Soc.*, 126(16):5042–5043, 2004.

- [82] Eric G Diken, Jeffrey M Headrick, Joseph R Roscioli, Joseph C Bopp, Mark A Johnson, and Anne B McCoy. Fundamental Excitations of the Shared Proton in the H_3O_2^- and H_5O_2^+ Complexes. *J. Phys. Chem. A*, 109(8):1487–1490, 2005.
- [83] Eric G Diken, Jeffrey M Headrick, Joseph R Roscioli, Joseph C Bopp, Mark A Johnson, Anne B McCoy, Xinchuan Huang, Stuart Carter, and Joel M Bowman. Argon Predissociation Spectroscopy of the $\text{OH}^- \text{H}_2\text{O}$ and $\text{Cl}^- \text{H}_2\text{O}$ Complexes in the 1000–1900 cm^{-1} Region: Intramolecular Bending Transitions and the Search for the Shared-Proton Fundamental in the Hydroxide Monohydrate. *J. Phys. Chem. A*, 109(4):571–575, 2005.
- [84] Anne B McCoy, Xinchuan Huang, Stuart Carter, and Joel M Bowman. Quantum studies of the vibrations in H_3O_2^- and D_3O_2^- . *J. Chem. Phys.*, 123(6):064317, 2005.
- [85] Hua-Gen Yu. A rigorous full-dimensional quantum dynamics calculation of the vibrational energies of h_3o_2^- . *J. Chem. Phys.*, 125(20):204306, 2006.
- [86] Kimichi Suzuki, Motoyuki Shiga, and Masanori Tachikawa. Temperature and Isotope Effects on Water Cluster Ions with Path Integral Molecular Dynamics Based on the Fourth Order Trotter Expansion. *J. Chem. Phys.*, 129(14):144310, 2008.
- [87] Martina Kaledin, John M Moffitt, Craig R Clark, and Fareeha Rizvi. *Ab Initio* Molecular Dynamics Simulations of the Infrared Spectra of H_3O_2^- and D_3O_2^- . *J. Chem. Theory Comput.*, 5(5):1328–1336, 2009.
- [88] Anne Milet, Tatiana Korona, Robert Moszynski, and Elise Kochanski. Anisotropic Intermolecular Interactions in Van der Waals and Hydrogen-Bonded Complexes: What Can We Get from Density Functional Calculations? *J. Chem. Phys.*, 111(17):7727–7735, 1999.
- [89] Julie A Anderson and Gregory S Tschumper. Characterizing the Potential Energy Surface of the Water Dimer with DFT: Failures of Some Popular Functionals for Hydrogen Bonding. *J. Phys. Chem. A*, 110(22):7268–7271, 2006.
- [90] G Avila, G J Kroes, and M C van Hemert. The Photodissociation of the Water Dimer in the $\tilde{\text{A}}$ Band: A Twelve-Dimensional Quasiclassical Study. *J. Chem. Phys.*, 128(14):144313, 2008.
- [91] Rustam Z Khaliullin, Alexis T Bell, and Martin Head-Gordon. Electron Donation in the Water–Water Hydrogen Bond. *Chem. Eur. J.*, 15(4):851–855, 2009.
- [92] Kirill Kuyanov-Prozument, Myong Yong Choi, and Andrey F Vilesov. Spectrum and Infrared Intensities of OH-Stretching Bands of Water Dimers. *J. Chem. Phys.*, 132(1):014304, 2010.

- [93] Thomas D Kühne and Rustam Z Khaliullin. Electronic Signature of the Instantaneous Asymmetry in the first Coordination Shell of Liquid Water. *Nat. Commun.*, 4:1450, 2013.
- [94] Branko Ruscic. Active Thermochemical Tables: Water and Water Dimer. *J. Phys. Chem. A*, 117(46):11940–11953, 2013.
- [95] Volodymyr Babin, Claude Leforestier, and Francesco Paesani. Development of a “First Principles” Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *J. Chem. Theory Comput.*, 9(12):5395–5403, 2013.
- [96] Claude Leforestier. Water Dimer Equilibrium Constant Calculation: A Quantum Formulation Including Metastable States. *J. Chem. Phys.*, 140(7):074106, 2014.
- [97] J Coleman Howard, Jordan D Enyard, and Gregory S Tschumper. Assessing the Accuracy of Some Popular DFT Methods for Computing Harmonic Vibrational Frequencies of Water Clusters. *J. Chem. Phys.*, 143(21):214103, 2015.
- [98] William T S Cole, Ray S Fellers, Mark R Viant, Claude Leforestier, and Richard J Saykally. Far-infrared vrt spectroscopy of the water dimer: Characterization of the 20 μm out-of-plane librational vibration. *J. Chem. Phys.*, 143(15):154306, 2015.
- [99] Maxim Tafipolsky and Kay Ansorg. Toward a physically motivated force field: Hydrogen bond directionality from a symmetry-adapted perturbation theory perspective. *J. Chem. Theory Comput.*, 12(3):1267–1279, 2016.
- [100] Piotr Jankowski, Garold Murdachaew, Robert Bukowski, Omololu Akin-Ojo, Claude Leforestier, and Krzysztof Szalewicz. *Ab Initio* Water Pair Potential with Flexible Monomers. *J. Phys. Chem. A*, 119(12):2940–2964, 2015.
- [101] Frank Jensen. *Introduction to Computational Chemistry*. John Wiley & Sons, 2 edition, 2007.
- [102] David J Tozer and Nicholas C Handy. The Importance of the Asymptotic Exchange-Correlation Potential in Density Functional Theory. *Mol. Phys.*, 101(17):2669–2675, 2003.
- [103] Axel D. Becke. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A*, 38(6):3098, 1988.
- [104] J P Perdew, Kieron Burke, and M Ernzerhof. A Reply to the Comment by Yingkai Zhang and Weitao Yang. *Phys. Rev. Lett.*, 80(4):891, 1998.
- [105] Michael J Gillan, Dario Alfè, and Angelos Michaelides. Perspective: How Good is DFT

- for Water? *J. Chem. Phys.*, 144(13):130901, 2016.
- [106] G Zundel and H Metzger. Energiebänder der Tunnelnden Überschuß-Protonen in Flüssigen Säuren. Eine IR-Spektroskopische Untersuchung der Natur der Gruppierungen H_5O_2^+ . *Z. Phys. Chem.*, 58(5_6):225–245, 1968.
 - [107] R Janoschek, EG Weidemann, H Pfeiffer, and G Zundel. Extremely High Polarizability of Hydrogen Bonds. *J. Am. Chem. Soc.*, 94(7):2387–2396, 1972.
 - [108] R Janoschek, Erich G Weidemann, and Georg Zundel. Calculated Frequencies and Intensities Associated with Coupling of the Proton Motion with the Hydrogen Bond Stretching Vibration in a Double Minimum Potential Surface. *J. Chem. Soc., Faraday Trans. 2*, 69:505–520, 1973.
 - [109] Eigen Wicke, M Eigen, and Th Ackermann. Über den Zustand des Protons (Hydroniumions) in Wässriger Lösung. *Z. Phys. Chem.*, 1(340):194, 1954.
 - [110] Manfred Eigen and L De Maeyer. Self-Dissociation and Protonic Charge Transport in Water and Ice. *Proc. R. Soc. London, Ser. A*, 247(1251):505–533, 1958.
 - [111] M Eigen. Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. Part I: Elementary Processes. *Angew. Chem. Int. Ed.*, 3(1):1–19, 1964.
 - [112] R. Parthasarathi, V. Subramanian, and N. Sathyamurthy. Hydrogen Bonding Without Borders: An Atoms-In-Molecules Perspective. *J. Phys. Chem. A*, 110(10):3349–3351, 2006.
 - [113] R. Parthasarathi, V. Subramanian, and N. Sathyamurthy. Hydrogen Bonding in Protonated Water Clusters: An Atoms-In-Molecules Perspective. *J. Phys. Chem. A*, 111(51):13287–13290, 2007.
 - [114] Krešimir Molčanov, Christian Jelsch, Emmanuel Wenger, Jernej Stare, Anders Ø Madsen, and Biserka Kojić-Prodić. Experimental Evidence of a 3-Centre, 2-Electron Covalent Bond Character of the Central O–H–O Fragment on the Zundel Cation in Crystals of Zundel Nitrilic Tetrahydrate. *CrystEngComm*, 19(28):3898–3901, 2017.
 - [115] Martin Thämer, Luigi De Marco, Krupa Ramasesha, Aritra Mandal, and Andrei Tokmakoff. Ultrafast 2D IR Spectroscopy of the Excess Proton in Liquid Water. *Science*, 350(6256):78–82, 2015.
 - [116] Clyde A. Daly Jr, Louis M. Streacker, Yuchen Sun, Shannon R. Pattenaude, Ali A. Hassanali, Poul B. Petersen, Steven A. Corcelli, and Dor Ben-Amotz. Decomposition of the Experimental Raman and Infrared Spectra of Acidic Water into Proton, Special Pair, and Counterion Contributions. *J. Phys. Chem. Lett.*, 8(21):5246–5252, 2017.

- [117] Joseph A. Fournier, William B. Carpenter, Nicholas H. C. Lewis, and Andrei Tokmakoff. Broadband 2D IR Spectroscopy Reveals Dominant Asymmetric H_5O_2^+ Proton Hydration Structures in Acid Solutions. *Nat. Chem.*, 10(9):932, 2018.
- [118] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem*, 32(7):1456–1465, 2011.
- [119] Colin K Egan, Brandon B Bizzarro, Marc Riera, and Francesco Paesani. Nature of Alkali Ion–Water Interactions: Insights from Many-Body Representations and Density Functional Theory. II. *J. Chem. Theory Comput.*, 16(5):3055–3072, 2020.
- [120] Brandon B Bizzarro, Colin K Egan, and Francesco Paesani. Nature of Halide–Water Interactions: Insights from Many-Body Representations and Density Functional Theory. *J. Chem. Theory Comput.*, 15(5):2983–2995, 2019.
- [121] Joseph P Heindel and Sotiris S Xantheas. The Many-Body Expansion for Aqueous Systems Revisited: I. Water–Water Interactions. *J. Chem. Theory Comput.*, 16(11):6843–6855, 2020.
- [122] Aaron Meurer, Christopher P. Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B. Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K. Moore, Sartaj Singh, Thilina Rathnayake, Sean Vig, Brian E. Granger, Richard P. Muller, Francesco Bonazzi, Harsh Gupta, Shivam Vats, Fredrik Johansson, Fabian Pedregosa, Matthew J. Curry, Andy R. Terrel, Štěpán Roučka, Ashutosh Saboo, Isuru Fernando, Sumith Kulal, Robert Cimrman, and Anthony Scopatz. Sympy: symbolic computing in python. *PeerJ Computer Science*, 3:e103, January 2017.
- [123] Paul R Horn, Yuezhi Mao, and Martin Head-Gordon. Probing Non-Covalent Interactions with a Second Generation Energy Decomposition Analysis Using Absolutely Localized Molecular Orbitals. *Phys. Chem. Chem. Phys.*, 18(33):23067–23079, 2016.
- [124] Paul R Horn, Yuezhi Mao, and Martin Head-Gordon. Defining the Contributions of Permanent Electrostatics, Pauli Repulsion, and Dispersion in Density Functional Theory Calculations of Intermolecular Interaction Energies. *J. Chem. Phys.*, 144(11):114107, 2016.
- [125] Rustam Z Khaliullin, Erika A Cobar, Rohini C Lochan, Alexis T Bell, and Martin Head-Gordon. Unravelling the Origin of Intermolecular Interactions Using Absolutely Localized Molecular Orbitals. *The Journal of Physical Chemistry A*, 111(36):8753–8765, 2007.
- [126] Paul R Horn and Martin Head-Gordon. Polarization Contributions to Intermolecular Interactions Revisited with Fragment Electric-Field Response Functions. *J. Chem. Phys.*, 143(11):114111, 2015.

- [127] Yuezhi Mao, Qinghui Ge, Paul R Horn, and Martin Head-Gordon. On the Computational Characterization of Charge-Transfer Effects in Noncovalently Bound Molecular Complexes. *J. Chem. Theory Comput.*, 14(5):2401–2417, 2018.
- [128] John P Perdew, K Burke, and M Ernzerhof. Perdew, Burke, and Ernzerhof Reply. *Phys. Rev. Lett.*, 80(4):891, 1998.
- [129] Jean-Philip Piquemal, Antonio Marquez, Olivier Parisel, and Claude Giessner-Prettre. A CSOV Study of the Difference Between HF and DFT Intermolecular Interaction Energy Values: The Importance of the Charge Transfer Contribution. *J. Comput. Chem.*, 26(10):1052–1062, 2005.
- [130] Qin Wu, Paul W Ayers, and Yingkai Zhang. Density-Based Energy Decomposition Analysis for Intermolecular Interactions with Variationally Determined Intermediate State Energies. *J. Chem. Phys.*, 131(16):164112, 2009.
- [131] F-F Wang, Glen Jenness, WA Al-Saidi, and KD Jordan. Assessment of the Performance of Common Density Functional Methods for Describing the Interaction Energies of (H₂O)₆ Clusters. *J. Chem. Phys.*, 132(13):134303, 2010.
- [132] Frank Jensen. Describing Anions by Density Functional Theory: Fractional Electron Affinity. *J. Chem. Theory Comput.*, 6(9):2726–2735, 2010.
- [133] Sándor Kristyán and Peter Pulay. Can (Semi) Local Density Functional Theory Account for the London Dispersion Forces? *Chem. Phys. Lett.*, 229(3):175–180, 1994.
- [134] Riccardo Sabatini, Tommaso Gorni, and Stefano De Gironcoli. Nonlocal van der Waals Density Functional Made Simple and Efficient. *Phys. Rev. B*, 87(4):041108, 2013.
- [135] Stefan Grimme. Density Functional Theory with London Dispersion Corrections. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 1(2):211–228, 2011.
- [136] Rustam Z. Khaliullin, Erika A. Cobar, Rohini C. Lochan, Alexis T. Bell, and Martin Head-Gordon. Unravelling the Origin of Intermolecular Interactions Using Absolutely Localized Molecular Orbitals. *J. Phys. Chem. A*, 111(36):8753–8765, 2007.
- [137] Rustam Z. Khaliullin, Alexis T. Bell, and Martin Head-Gordon. Analysis of Charge Transfer Effects in Molecular Complexes Based on Absolutely Localized Molecular Orbitals. *J. Chem. Phys.*, 128(18):184112, 2008.
- [138] Wilfred F van Gunsteren and Herman J C Berendsen. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angew. Chem., Int. Ed. Engl.*, 29(9):992–1023, 1990.

- [139] Arieh Warshel. Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights. *Annu. Rev. Biophys. Biomol. Struct.*, 32(1):425–443, 2003.
- [140] Mireia Garcia-Viloca, Jiali Gao, Martin Karplus, and Donald G Truhlar. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science*, 303(5655):186–195, 2004.
- [141] V Yamakov, D Wolf, S R Phillpot, A K Mukherjee, and H Gleiter. Deformation-Mechanism Map for Nanocrystalline Metals by Molecular-Dynamics Simulation. *Nat. Mater.*, 3(1):43, 2004.
- [142] Steve O Nielsen, Carlos F Lopez, Goundla Srinivas, and Michael L Klein. Coarse Grain Models and the Computer Simulation of Soft Materials. *J. Phys.: Condens. Matter*, 16(15):R481, 2004.
- [143] Albert J Markvoort, P A J Hilbers, and S V Nedeau. Molecular Dynamics Study of the Influence of Wall-Gas Interactions on Heat Flow in Nanochannels. *Phys. Rev. E*, 71(6):066702, 2005.
- [144] Martin Karplus and John Kuriyan. Molecular Dynamics and Protein Function. *Proc. Natl. Acad. Sci. U. S. A.*, 102(19):6679–6685, 2005.
- [145] Gregory A Voth. Computer Simulation of Proton Solvation and Transport in Aqueous and Biomolecular Systems. *Acc. Chem. Res.*, 39(2):143–150, 2006.
- [146] Marco De Vivo, Matteo Masetti, Giovanni Bottegoni, and Andrea Cavalli. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.*, 59(9):4035–4061, 2016.
- [147] Mark Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford university press, 2010.
- [148] Attila Szabo and Neil S Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Courier Corporation, 2012.
- [149] Rodney J Bartlett and Monika Musiał. Coupled-Cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.*, 79(1):291, 2007.
- [150] Martin Head-Gordon and Emilio Artacho. Chemistry on the Computer. *Phys. Today*, 61(4):58, 2008.
- [151] Yuezhi Mao, Omar Demerdash, Martin Head-Gordon, and Teresa Head-Gordon. Assessing Ion–Water Interactions in the AMOEBA Force Field Using Energy Decomposition Analysis of Electronic Structure Calculations. *J. Chem. Theory Comput.*, 12(11):5422–

5437, 2016.

- [152] George C Schatz. The Analytical Representation of Electronic Potential-Energy Surfaces. *Rev. Mod. Phys.*, 61(3):669, 1989.
- [153] Robert Bukowski, Krzysztof Szalewicz, Gerrit C. Groenenboom, and Ad van der Avoird. Predictions of the Properties of Water from First Principles. *Science*, 315(5816):1249–1252, 2007.
- [154] Robert Bukowski, Krzysztof Szalewicz, Gerrit C. Groenenboom, and Ad van der Avoird. Polarizable interaction potential for water from coupled cluster calculations. II. Applications to dimer spectra, virial coefficients, and simulations of liquid water. *J. Chem. Phys.*, 128(9):094314, 2008.
- [155] Yimin Wang, Benjamin C. Shepler, Bastiaan J. Braams, and Joel M. Bowman. Full-Dimensional, *Ab Initio* Potential Energy and Dipole Moment Surfaces for Water. *J. Chem. Phys.*, 131(5):054511, 2009.
- [156] Yimin Wang, Xinchuan Huang, Benjamin C. Shepler, Bastiaan J. Braams, and Joel M. Bowman. Flexible, *Ab Initio* Potential, and Dipole Moment Surfaces for Water. I. Tests and Applications for Clusters up to the 22-mer. *J. Chem. Phys.*, 134(9):094509, 2011.
- [157] Yimin Wang and Joel M. Bowman. *Ab Initio* Potential and Dipole Moment Surfaces for Water. II. Local-Monomer Calculations of the Infrared Spectra of Water Clusters. *J. Chem. Phys.*, 134(15):154510, 2011.
- [158] Volodymyr Babin, Gregory R. Medders, and Francesco Paesani. Toward a Universal Water Model: First Principles Simulations from the Dimer to the Liquid Phase. *J. Phys. Chem. Lett.*, 3(24):3765–3769, 2012.
- [159] Gregory R. Medders, Volodymyr Babin, and Francesco Paesani. A Critical Assessment of Two-Body and Three-Body Interactions in Water. *J. Chem. Theory Comput.*, 9(2):1103–1114, feb 2013.
- [160] Volodymyr Babin, Claude Leforestier, and Francesco Paesani. Development of a “first principles” water potential with flexible monomers: Dimer potential energy surface, vrt spectrum, and second virial coefficient. *J. Chem. Theory Comput.*, 9(12):5395–5403, 2013.
- [161] Volodymyr Babin, Gregory R. Medders, and Francesco Paesani. Development of a “First Principles” Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. *J. Chem. Theory Comput.*, 10(4):1599–1607, 2014.

- [162] Gregory R. Medders, Volodymyr Babin, and Francesco Paesani. Development of a “First-Principles” Water Potential with Flexible Monomers. III. Liquid Phase Properties. *J. Chem. Theory Comput.*, 10(8):2906–2910, 2014.
- [163] Pushp Bajaj, Andreas W. Götz, and Francesco Paesani. Toward Chemical Accuracy in the Description of Ion–Water Interactions Through Many-Body Representations. I. Halide–Water Dimer Potential Energy Surfaces. *J. Chem. Theory Comput.*, 12(6):2698–2705, 2016.
- [164] Pushp Bajaj, Xiao-Gang Wang, Tucker Carrington Jr., and Francesco Paesani. Vibrational Spectra of Halide-Water Dimers: Insights on Ion Hydration from Full-Dimensional Quantum Calculations on Many-Body Potential Energy Surfaces. *J. Chem. Phys.*, 148(10):102321, 2018.
- [165] Marc Riera, Andreas W Götz, and Francesco Paesani. The i-TTM model for *Ab Initio*-Based Ion–Water Interaction Potentials. II. Alkali Metal Ion–Water Potential Energy Functions. *Phys. Chem. Chem. Phys.*, 18(44):30334–30343, 2016.
- [166] Marc Riera, Narbe Mardirossian, Pushp Bajaj, Andreas W. Götz, and Francesco Paesani. Toward Chemical Accuracy in the Description of Ion–Water Interactions Through Many-Body Representations. Alkali-Water Dimer Potential Energy Surfaces. *J. Chem. Phys.*, 147(16):161715, 2017.
- [167] Marc Riera, Sandra E Brown, and Francesco Paesani. Isomeric Equilibria, Nuclear Quantum Effects, and Vibrational Spectra of $M^+(H_2O)_{n=1-3}$ Clusters, with $M = Li, Na, K, Rb,$ and Cs , Through Many-Body Representations. *J. Phys. Chem. A*, 122(27):5811–5821, 2018.
- [168] Debbie Zhuang, Marc Riera, Gregory K Schenter, John L Fulton, and Francesco Paesani. Many-Body Effects Determine the Local Hydration Structure of Cs^+ in Solution. *J. Phys. Chem. Lett.*, 10:406–412, 2019.
- [169] C. Huy Pham, Sandeep K. Reddy, Karl Chen, Chris Knight, and Francesco Paesani. Many-Body Interactions in Ice. *J. Chem. Theory Comput.*, 13(4):1778–1784, 2017.
- [170] Sandra E. Brown, Andreas W. Götz, Xiaolu Cheng, Ryan P. Steele, Vladimir A. Mandelshtam, and Francesco Paesani. Monitoring Water Clusters “Melt” Through Vibrational Spectroscopy. *J. Am. Chem. Soc.*, 139(20):7082–7088, 2017.
- [171] Daniel R. Moberg, Shelby C. Straight, Christopher Knight, and Francesco Paesani. Molecular Origin of the Vibrational Structure of Ice I_h . *J. Phys. Chem. Lett.*, 2017.
- [172] Daniel J Arismendi-Arrieta, Marc Riera, Pushp Bajaj, Rita Prosmiti, and Francesco Paesani. i-TTM Model for *Ab Initio*-Based Ion–Water Interaction Potentials. 1. Halide–

- Water Potential Energy Functions. *J. Phys. Chem. B*, 120(8):1822–1832, 2015.
- [173] B T Thole. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem. Phys.*, 59(3):341–350, 1981.
- [174] Jonas Sala, Elvira Guardia, and Marco Masia. The Polarizable Point Dipoles Method with Electrostatic Damping: Implementation on a Model System. *J. Chem. Phys.*, 133(23):234101, 2010.
- [175] C J Burnham, D J Anick, P K Mankoo, and G F Reiter. The Vibrational Proton Potential in Bulk Liquid Water and Ice. *J. Chem. Phys.*, 128(15):154519, 2008.
- [176] Curt M Breneman and Kenneth B Wiberg. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J. Comput. Chem.*, 11(3):361–373, 1990.
- [177] Aleksandr V Marenich, Steven V Jerome, Christopher J Cramer, and Donald G Truhlar. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.*, 8(2):527–541, 2012.
- [178] George S Fanourgakis and Sotiris S Xantheas. The Flexible, Polarizable, Thole-Type Interaction Potential for Water (TTM2-F) Revisited. *J. Phys. Chem. A*, 110(11):4100–4106, 2006.
- [179] K T Tang and J Peter Toennies. An Improved Simple Model for the van der Waals Potential Based on Universal Damping Functions for the Dispersion Coefficients. *J. Chem. Phys.*, 80(8):3726–3741, 1984.
- [180] Axel D. Becke and Erin R. Johnson. A Density-Functional Model of the Dispersion Interaction. *J. Chem. Phys.*, 123(15):154101, 2005.
- [181] Axel D Becke and Erin R Johnson. Exchange-Hole Dipole Moment and the Dispersion Interaction Revisited. *J. Chem. Phys.*, 127(15):154108, 2007.
- [182] Jing Kong, Zhengting Gan, Emil Proynov, Marek Freindorf, and Thomas R Furlani. Efficient Computation of the Dispersion Interaction with Density-Functional Theory. *Phys. Rev. A*, 79(4):042510, 2009.
- [183] Felix O. Kannemann and Axel D. Becke. van der Waals Interactions in Density-Functional Theory: Intermolecular Complexes. *J. Chem. Theory Comput.*, 6(4):1081–1088, 2010.
- [184] A Otero-De-La-Roza and Erin R Johnson. Non-Covalent Interactions and Thermochemistry Using XDM-Corrected Hybrid and Range-Separated Hybrid Density Functionals. *J.*

- Chem. Phys.*, 138(20):204109, 2013.
- [185] Mary J Van Vleet, Alston J Misquitta, Anthony J Stone, and Jordan R Schmidt. Beyond Born–Mayer: Improved Models for Short-Range Repulsion in *Ab Initio* Force Fields. *J. Chem. Theory Comput.*, 12(8):3851–3870, 2016.
 - [186] Claude Millot, Jean-Christophe Soetens, Marília TC Martins Costa, Matthew P Hodges, and Anthony J Stone. Revised Anisotropic Site Potentials for the Water Dimer and Calculated Properties. *J. Phys. Chem. A*, 102(4):754–770, 1998.
 - [187] Adri C T Van Duin, Siddharth Dasgupta, Francois Lorant, and William A Goddard. ReaxFF: a Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A*, 105(41):9396–9409, 2001.
 - [188] Riccardo Chelli, Marco Pagliai, Piero Procacci, Gianni Cardini, and Vincenzo Schettino. Polarization Response of Water and Methanol Investigated by a Polarizable Force Field and Density Functional Theory Calculations: Implications for Charge Transfer. *J. Chem. Phys.*, 122(7):074504, 2005.
 - [189] Tristan G A Youngs and Christopher Hardacre. Application of Static Charge Transfer Within an Ionic-Liquid Force Field and Its Effect on Structure and Dynamics. *ChemPhysChem*, 9(11):1548–1558, 2008.
 - [190] Alexis J Lee and Steven W Rick. The Effects of Charge Transfer on the Properties of Liquid Water. *J. Chem. Phys.*, 134(18):184507, 2011.
 - [191] Marielle Soniat and Steven W Rick. The Effects of Charge Transfer on the Aqueous Solvation of Ions. *J. Chem. Phys.*, 137(4):044511, 2012.
 - [192] Bastiaan J Braams and Joel M Bowman. Permutationally Invariant Potential Energy Surfaces in High Dimensionality. *Int. Rev. Phys. Chem.*, 28(4):577–606, 2009.
 - [193] Thuong T Nguyen, Eszter Székely, Giulio Imbalzano, Jörg Behler, Gábor Csányi, Michele Ceriotti, Andreas W Götz, and Francesco Paesani. Comparison of Permutationally Invariant Polynomials, Neural Networks, and Gaussian Approximation Potentials in Representing Water Interactions Through Many-Body Expansions. *J. Chem. Phys.*, 148(24):241725, 2018.
 - [194] Hugh C Longuet-Higgins. The Symmetry Groups of Non-Rigid Molecules. *Mol. Phys.*, 6(5):445–460, 1963.
 - [195] Andrey N. Tikhonov. Solution of Incorrectly Formulated Problems and the Regularization Method. In *Soviet Math. Dokl.*, volume 5, pages 1035–1038, 1963.

- [196] Martin Feyereisen, George Fitzgerald, and Andrew Komornicki. Use of Approximate Integrals in *Ab Initio* Theory. An Application in MP2 Energy Calculations. *Chem. Phys. Lett.*, 208(5-6):359–363, 1993.
- [197] Florian Weigend and Marco Häser. RI-MP2: First Derivatives and Global Consistency. *Theor. Chem. Acc.*, 97(1-4):331–340, 1997.
- [198] An-Suei Yang and Barry Honig. On the pH Dependence of Protein Stability. *J. Mol. Biol.*, 231(2):459–474, 1993.
- [199] Jessica MJ Swanson, C Mark Maupin, Hanning Chen, Matt K Petersen, Jiancong Xu, Yujie Wu, and Gregory A Voth. Proton Solvation and Transport in Aqueous and Biomolecular Systems: Insights from Computer Simulations. *J. Phys. Chem. B*, 111(17):4300–4314, 2007.
- [200] Robin L Hayes, Stephen J Paddison, and Mark E Tuckerman. Proton Transport in Triflic Acid Hydrates Studied via Path Integral Car- Parrinello Molecular Dynamics. *J. Phys. Chem. B*, 113(52):16574–16589, 2009.
- [201] Ivan M Krivobokov, Evgeniy N Gribov, Alexey G Okunev, Giuseppe Spoto, and Valentin N Parmon. Transport Properties of Modified Nafion Membranes: Effect of Zeolite and Precursors. *Solid State Ionics*, 180(40):1694–1701, 2010.
- [202] Guillermo A Luduena, Thomas D Ku?hne, and Daniel Sebastiani. Mixed Grotthuss and Vehicle Transport Mechanism in Proton Conducting Polymers from *Ab Initio* Molecular Dynamics Simulations. *Chem. Mater.*, 23(6):1424–1429, 2011.
- [203] Francesco Paesani. Molecular Mechanisms of Water-Mediated Proton Transport in MIL-53 Metal–Organic Frameworks. *J. Phys. Chem. C*, 117(38):19508–19516, 2013.
- [204] Daiane Damasceno Borges, Sabine Devautour-Vinot, Hervé Jobic, Jacques Ollivier, Farid Nouar, Rocio Semino, Thomas Devic, Christian Serre, Francesco Paesani, and Guillaume Maurin. Proton Transport in a Highly Conductive Porous Zirconium-Based Metal–Organic Framework: Molecular Insight. *Angew. Chem.*, 128(12):3987–3992, 2016.
- [205] Thais Grancha, Jess Ferrando-Soria, Joan Cano, Pedro Amors, Beatriz Seoane, Jorge Gascon, Montse Bazaga-García, Enrique R Losilla, Aurelio Cabeza, Donatella Armentano, et al. Insights into the Dynamics of Grotthuss Mechanism in a Proton-Conducting Chiral bio MOF. *Chem. Mater.*, 28(13):4608–4615, 2016.
- [206] NB Librovič, VP Sakun, and ND Sokolov. H^+ and OH^- Ions in Aqueous Solutions Vibrational Spectra of Hydrates. *Chem. Phys.*, 39(3):351–366, 1979.
- [207] Omar F Mohammed, Dina Pines, Jens Dreyer, Ehud Pines, and Erik TJ Nibbering.

- Sequential Proton Transfer Through Water Bridges in Acid-Base Reactions. *Science*, 310(5745):83–86, 2005.
- [208] Sander Woutersen and Huib J Bakker. Ultrafast Vibrational and Structural Dynamics of the Proton in Liquid Water. *Phys. Rev. Lett.*, 96(13):138305, 2006.
- [209] A Botti, F Bruni, MA Ricci, and AK Soper. Eigen Versus Zundel Complexes in HCl-Water Mixtures. *J. Chem. Phys.*, 125(1):014508, 2006.
- [210] Bernd Winter, Manfred Faubel, Ingolf V Hertel, Christian Pettenkofer, Stephen E Bradforth, Barbara Jagoda-Cwiklik, Lukasz Cwiklik, and Pavel Jungwirth. Electron Binding Energies of Hydrated H_3O^+ and OH^- : Photoelectron Spectroscopy of Aqueous Acid and Base Solutions Combined with Electronic Structure Calculations. *J. Am. Chem. Soc.*, 128(12):3864–3865, 2006.
- [211] BJ Siwick, MJ Cox, and HJ Bakker. Long-Range Proton Transfer in Aqueous Acid-Base Reactions. *J. Phys. Chem. B*, 112(2):378–389, 2008.
- [212] Maciej Śmiechowski and Janusz Stangret. ATR FT-IR H_2O Spectra of Acidic Aqueous Solutions. Insights About Proton Hydration. *J. Mol. Struct.*, 878(1-3):104–115, 2008.
- [213] Evgenii S Stoyanov, Irina V Stoyanova, and Christopher A Reed. The Structure of the Hydrogen Ion (H_{aq}^+) in Water. *J. Am. Chem. Soc.*, 132(5):1484–1485, 2010.
- [214] John L Fulton and Mahalingam Balasubramanian. Structure of Hydronium (H_3O^+)/Chloride (Cl^-) Contact Ion Pairs in Aqueous Hydrochloric Acid Solution: A Zundel-Like Local Configuration. *J. Am. Chem. Soc.*, 132(36):12597–12604, 2010.
- [215] Dominique Decka, Gerhard Schwaab, and Martina Havenith. A THz/FTIR Fingerprint of the Solvated Proton: Evidence for Eigen Structure and Zundel Dynamics. *Phys. Chem. Chem. Phys.*, 17(17):11898–11907, 2015.
- [216] William B Carpenter, Joseph A Fournier, Nicholas HC Lewis, and Andrei Tokmakoff. Picosecond Proton Transfer Kinetics in Water Revealed With Ultrafast IR Spectroscopy. *J. Phys. Chem. B*, 122(10):2792–2802, 2018.
- [217] Mark Tuckerman, Kari Laasonen, Michiel Sprik, and Michele Parrinello. *Ab Initio* Molecular Dynamics Simulation of the Solvation and Transport of Hydronium and Hydroxyl Ions in Water. *J. Chem. Phys.*, 103(1):150–161, 1995.
- [218] Noam Agmon. The Grotthuss Mechanism. *Chem. Phys. Lett.*, 244(5-6):456–462, 1995.
- [219] Noam Agmon. Hydrogen Bonds, Water Rotation and Proton Mobility. *J. Chim. Phys. Phys.-Chim. Biol.*, 93:1714–1736, 1996.

- [220] Udo W. Schmitt and Gregory A. Voth. Multistate Empirical Valence Bond Model for Proton Transport in Water. *J. Phys. Chem. B*, 102(29):5547–5551, 1998.
- [221] Dominik Marx, Mark E Tuckerman, Jürg Hutter, and Michele Parrinello. The Nature of the Hydrated Excess Proton in Water. *Nature*, 397(6720):601–604, 1999.
- [222] Rodolphe Vuilleumier and Daniel Borgis. An Extended Empirical Valence Bond Model for Describing Proton Mobility in Water. *Isr. J. Chem.*, 39(3-4):457–467, 1999.
- [223] Rodolphe Vuilleumier and Daniel Borgis. Transport and Spectroscopy of the Hydrated Proton: a Molecular Dynamics Study. *J. Chem. Phys.*, 111(9):4251–4266, 1999.
- [224] Tyler JF Day, Alexander V Soudackov, Martin Čuma, Udo W Schmitt, and Gregory A Voth. A Second Generation Multistate Empirical Valence Bond Model for Proton Transport in Aqueous Systems. *J. Chem. Phys.*, 117(12):5839–5849, 2002.
- [225] Giuseppe Brancato and Mark E Tuckerman. A polarizable Multistate Empirical Valence Bond Model for Proton Transport in Aqueous Solution. *J. Chem. Phys.*, 122(22):224507, 2005.
- [226] Sergei Izvekov and Gregory A Voth. *Ab Initio* Molecular-Dynamics Simulation of Aqueous Proton Solvation and Transport Revisited. *J. Chem. Phys.*, 123(4):044505, 2005.
- [227] D Asthagiri, LR Pratt, and JD Kress. *Ab Initio* Molecular Dynamics and Quasichemical Study of H^+ (aq). *Proc. Natl. Acad. Sci.*, 102(19):6704–6708, 2005.
- [228] Hadas Lapid, Noam Agmon, Matt K Petersen, and Gregory A Voth. A Bond-Order Analysis of the Mechanism for Hydrated Proton Mobility in Liquid Water. *J. Chem. Phys.*, 122(1):014506, 2005.
- [229] Dominik Marx. Proton Transfer 200 Years After von Grotthuss: Insights from *Ab Initio* Simulations. *ChemPhysChem*, 7(9):1848–1870, 2006.
- [230] Tatiana Murakhtina, Jasper Heuft, Evert Jan Meijer, and Daniel Sebastiani. First Principles and Experimental 1H NMR Signatures of Solvated Ions: the Case of HCl (aq). *ChemPhysChem*, 7(12):2578–2584, 2006.
- [231] Yujie Wu, Hanning Chen, Feng Wang, Francesco Paesani, and Gregory A Voth. An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transport. *J. Phys. Chem. B*, 112(2):467–482, 2008.
- [232] Omer Markovitch, Hanning Chen, Sergei Izvekov, Francesco Paesani, Gregory A Voth, and Noam Agmon. Special Pair Dance and Partner Selection: Elementary Steps in Proton Transport in Liquid Water. *J. Phys. Chem. B*, 112(31):9456–9466, 2008.

- [233] Jessica MJ Swanson and Jack Simons. Role of Charge Transfer in the Structure and Dynamics of the Hydrated Proton. *J. Phys. Chem. B*, 113(15):5149–5161, 2009.
- [234] Chris Knight and Gregory A Voth. The Curious Case of the Hydrated Proton. *Acc. Chem. Res.*, 45(1):101–109, 2012.
- [235] Ali Hassanali, Federico Giberti, Jérôme Cuny, Thomas D. Kühne, and Michele Parrinello. Proton Transfer Through the Water Gossamer. *Proc. Natl. Acad. Sci. USA*, 110(34):13723–13728, 2013.
- [236] Waldemar Kulig and Noam Agmon. A ‘Clusters-in-Liquid’ Method for Calculating Infrared Spectra Identifies the Proton-Transfer Mode in Acidic Aqueous Solutions. *Nat. Chem.*, 5(1):29–35, 2013.
- [237] Ali A Hassanali, Federico Giberti, Gabriele C Sosso, and Michele Parrinello. The Role of the Umbrella Inversion Mode in Proton Diffusion. *Chem. Phys. Lett.*, 599:133–138, 2014.
- [238] Takuya Mabuchi, Akinori Fukushima, and Takashi Tokumasu. A Modified Two-State Empirical Valence Bond Model for Proton Transport in Aqueous Solutions. *J. Chem. Phys.*, 143(1):014501, 2015.
- [239] Rajib Biswas, William Carpenter, Joseph A. Fournier, Gregory A. Voth, and Andrei Tokmakoff. IR Spectral Assignments for the Hydrated Excess Proton in Liquid Water. *J. Chem. Phys.*, 146(15):154507, 2017.
- [240] William B Carpenter, Qi Yu, John H Hack, Bogdan Dereka, Joel M Bowman, and Andrei Tokmakoff. Decoding the 2D IR Spectrum of the Aqueous Proton with High-Level VSCF/VCI Calculations. *J. Chem. Phys.*, 153(12):124506, 2020.
- [241] Giuseppe Cassone. Nuclear Quantum Effects Largely Influence Molecular Dissociation and Proton Transfer in Liquid Water Under an Electric Field. *J. Phys. Chem. Lett.*, 11:8983–8988, 2020.
- [242] AJ Cunningham, JD Payzant, and Paul Kebarle. Kinetic Study of the Proton Hydrate $\text{H}^+(\text{H}_2\text{O})_n$ Equilibria in the Gas Phase. *J. Am. Chem. Soc.*, 94(22):7627–7632, 1972.
- [243] LI Yeh, M Okumura, JD Myers, JM Price, and YT Lee. Vibrational Spectroscopy of the Hydrated Hydronium Cluster Ions $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ ($n=1,2,3$). *J. Chem. Phys.*, 91(12):7319–7330, 1989.
- [244] Jeffrey M Headrick, Eric G Diken, Richard S Walters, Nathan I Hammer, Richard A Christie, Jun Cui, Evgeniy M Myshakin, Michael A Duncan, Mark A Johnson, and Kenneth D Jordan. Spectral Signatures of Hydrated Proton Vibrations in Water Clusters. *Science*, 308(5729):1765–1769, 2005.

- [245] Nathan I Hammer, Eric G Diken, Joseph R Roscioli, Mark A Johnson, Evgeniy M Myshakin, Kenneth D Jordan, Anne B McCoy, Xinchuan Huang, Joel M Bowman, and Stuart Carter. The Vibrational Predissociation Spectra of the $\text{H}_5\text{O}_2^+ \cdot \text{RG}_n$ ($\text{RG} = \text{Ar}, \text{Ne}$) Clusters: Correlation of the Solvent Perturbations in the Free OH and Shared Proton Transitions of the Zundel Ion. *J. Chem. Phys.*, 122(24):244301, 2005.
- [246] Huan-Cheng Chang, Chih-Che Wu, and Jer-Lai Kuo. Recent Advances in Understanding the Structures of Medium-Sized Protonated Water Clusters. *Int. Rev. Phys. Chem.*, 24(3-4):553–578, 2005.
- [247] Anna Gutberlet, Gerhard Schwaab, Özgür Birer, Marco Masia, Anna Kaczmarek, Harald Forbert, Martina Havenith, and Dominik Marx. Aggregation-Induced Dissociation of HCl (H_2O)₄ Below 1 K: The Smallest Droplet of Acid. *Science*, 324(5934):1545–1548, 2009.
- [248] Kenta Mizuse, Naohiko Mikami, and Asuka Fujii. Infrared Spectra and Hydrogen-Bonded Network Structures of Large Protonated Water Clusters $\text{H}^+(\text{H}_2\text{O})_n$ ($n=20-200$). *Angew. Chem. Int. Ed.*, 49(52):10119–10122, 2010.
- [249] GE Douberly, RS Walters, J Cui, KD Jordan, and MA Duncan. Infrared Spectroscopy of Small Protonated Water Clusters, $\text{H}^+(\text{H}_2\text{O})_n$ ($n=2-5$): Isomers, Argon Tagging, and Deuteration. *J. Phys. Chem. A*, 114(13):4570–4579, 2010.
- [250] Nadja Heine, Matias R Fagiani, Mariana Rossi, Torsten Wende, Giel Berden, Volker Blum, and Knut R Asmis. Isomer-Selective Detection of Hydrogen-Bond Vibrations in the Protonated Water Hexamer. *J. Am. Chem. Soc.*, 135(22):8266–8273, 2013.
- [251] Joseph A Fournier, Conrad T Wolke, Mark A Johnson, Tuguldur T Odbadrakh, Kenneth D Jordan, Shawn M Kathmann, and Sotiris S Xantheas. Snapshots of Proton Accommodation at a Microscopic Water Surface: Understanding the Vibrational Spectral Signatures of the Charge Defect in Cryogenically Cooled $\text{H}^+(\text{H}_2\text{O})_{n=2-28}$ Clusters. *J. Phys. Chem. A*, 119(36):9425–9440, 2015.
- [252] Conrad T Wolke, Joseph A Fournier, Laura C Dzugan, Matias R Fagiani, Tuguldur T Odbadrakh, Harald Knorke, Kenneth D Jordan, Anne B McCoy, Knut R Asmis, and Mark A Johnson. Spectroscopic Snapshots of the Proton-Transfer Mechanism in Water. *Science*, 354(6316):1131–1135, 2016.
- [253] Chinh H Duong, Olga Gorlova, Nan Yang, Patrick J Kelleher, Mark A Johnson, Anne B McCoy, Qi Yu, and Joel M Bowman. Disentangling the Complex Vibrational Spectrum of the Protonated Water Trimer, $\text{H}^+(\text{H}_2\text{O})_3$, With Two-Color IR-IR Photodissociation of the Bare Ion and Anharmonic VSCF/VCI Theory. *J. Phys. Chem. Lett.*, 8(16):3782–3789, 2017.
- [254] Mark E Tuckerman, Dominik Marx, Michael L Klein, and Michele Parrinello. On the

- Quantum Nature of the Shared Proton in Hydrogen Bonds. *Science*, 275(5301):817–820, 1997.
- [255] David J Wales. Rearrangements and Tunneling Splittings of Protonated Water Dimer. *J. Chem. Phys.*, 110(21):10403–10409, 1999.
- [256] Philip L Geissler, Christoph Dellago, and David Chandler. Chemical Dynamics of the Protonated Water Trimer Analyzed by Transition Path Sampling. *Phys. Chem. Chem. Phys.*, 1(6):1317–1322, 1999.
- [257] Xinchuan Huang, Bastiaan J Braams, and Joel M Bowman. *Ab Initio* Potential Energy and Dipole Moment Surfaces for H_5O_2^+ . *J. Chem. Phys.*, 122(4):044308, 2005.
- [258] Anne B McCoy, Xinchuan Huang, Stuart Carter, Marc Y Landeweer, and Joel M Bowman. Full-Dimensional Vibrational Calculations for H_5O_2^+ Using an *Ab Initio* Potential Energy Surface. *J. Chem. Phys.*, 122(6):061101, 2005.
- [259] N Jiten Singh, Mina Park, Seung Kyu Min, Seung Bum Suh, and Kwang S Kim. Magic and Antimagic Protonated Water Clusters: Exotic Structures with Unusual Dynamic Effects. *Angew. Chem. Int. Ed.*, 45(23):3795–3800, 2006.
- [260] Ilgyou Shin, Mina Park, Seung Kyu Min, Eun Cheol Lee, Seung Bum Suh, and Kwang S Kim. Structure and Spectral Features of $\text{H}^+(\text{H}_2\text{O})_7$: Eigen Versus Zundel Forms. *J. Chem. Phys.*, 125(23):234305, 2006.
- [261] Oriol Vendrell, Fabien Gatti, David Lauvergnat, and Hans-Dieter Meyer. Full-Dimensional (15-Dimensional) Quantum-Dynamical Simulation of the Protonated Water Dimer. I. Hamiltonian Setup and Analysis of the Ground Vibrational State. *J. Chem. Phys.*, 127(18):184302, 2007.
- [262] Oriol Vendrell, Fabien Gatti, and Hans-Dieter Meyer. Full Dimensional (15-Dimensional) Quantum-Dynamical Simulation of the Protonated Water Dimer. II. Infrared Spectrum and Vibrational Dynamics. *J. Chem. Phys.*, 127(18):184303, 2007.
- [263] Oriol Vendrell, Fabien Gatti, and Hans-Dieter Meyer. Dynamics and Infrared Spectroscopy of the Protonated Water Dimer. *Angew. Chem. Int. Ed.*, 46(36):6918–6921, 2007.
- [264] Mina Park, Ilgyou Shin, N Jiten Singh, and Kwang S Kim. Eigen and Zundel Forms of Small Protonated Water Clusters: Structures and Infrared Spectra. *J. Phys. Chem. A*, 111(42):10692–10702, 2007.
- [265] Gulafroz Meraj and Ajay Chaudhari. Hydrogen Bonded Hydrated Hydronium and Zündel Ion Complexes. *J. Mol. Liq.*, 190:1–5, 2014.

- [266] Waldemar Kulig and Noam Agmon. Both Zundel and Eigen Isomers Contribute to the IR Spectrum of the Gas-Phase H_9O_4^+ Cluster. *J. Phys. Chem. B*, 118(1):278–286, 2014.
- [267] Qi Yu and Joel M Bowman. How the Zundel (H_5O_2^+) Potential Can be Used to Predict the Proton Stretch and Bend Frequencies of Larger Protonated Water Clusters. *J. Phys. Chem. Lett.*, 7(24):5259–5265, 2016.
- [268] Félix Mouhat, Sandro Sorella, Rodolphe Vuilleumier, Antonino Marco Saitta, and Michele Casula. Fully Quantum Description of the Zundel Ion: Combining Variational Quantum Monte Carlo with Path Integral Langevin Dynamics. *J. Chem. Theory Comput.*, 13(6):2400–2417, 2017.
- [269] Qi Yu and Joel M Bowman. High-Level Quantum Calculations of the IR Spectra of the Eigen, Zundel, and Ring Isomers of $\text{H}^+(\text{H}_2\text{O})_4$ Find a Single Match to Experiment. *J. Am. Chem. Soc.*, 139(32):10984–10987, 2017.
- [270] Gianluca Bertaina, Giovanni Di Liberto, and Michele Ceotto. Reduced Rovibrational Coupling Cartesian Dynamics for Semiclassical Calculations: Application to the Spectrum of the Zundel Cation. *J. Chem. Phys.*, 151(11):114307, 2019.
- [271] Fabian Dahms, Rene Costard, Ehud Pines, Benjamin P Fingerhut, Erik TJ Nibbering, and Thomas Elsaesser. The Hydrated Excess Proton in the Zundel Cation H_5O_2^+ : The Role of Ultrafast Solvent Fluctuations. *Angew. Chem.*, 128(36):10758–10763, 2016.
- [272] Achintya Kundu, Fabian Dahms, Benjamin P Fingerhut, Erik TJ Nibbering, Ehud Pines, and Thomas Elsaesser. Ultrafast Vibrational Relaxation and Energy Dissipation of Hydrated Excess Protons in Polar Solvents. *Chem. Phys. Lett*, 713:111–116, 2018.
- [273] Achintya Kundu, Fabian Dahms, Benjamin P Fingerhut, Erik TJ Nibbering, Ehud Pines, and Thomas Elsaesser. Hydrated Excess Protons in Acetonitrile/Water Mixtures: Solvation Species and Ultrafast Proton Motions. *J. Phys. Chem. Lett.*, 10(9):2287–2294, 2019.
- [274] Michele Ceriotti, Joshua More, and David E Manolopoulos. i-PI: A Python Interface for *Ab Initio* Path Integral Molecular Dynamics Simulations. *Comput. Phys. Commun.*, 185(3):1019–1026, 2014.
- [275] M del Carmen Grande, HL Bianchi, and CM Marschoff. On the Density of Pure Acetonitrile. In *J. Argent. Chem. Soc.*, volume 92, pages 109–114. SciELO Argentina, 2004.
- [276] María del Carmen Grande, Jorge Alvarez Juliá, Carmen R Barrero, Carlos M Marschoff, and Hugo L Bianchi. The (Water + Acetonitrile) Mixture Revisited: A New Approach for Calculating Partial Molar Volumes. *J. Chem. Thermodyn.*, 38(6):760–768, 2006.
- [277] Steve Plimpton. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J.*

- Comput. Phys.*, 117(1):1–19, 1995.
- [278] Michele Ceriotti, Giovanni Bussi, and Michele Parrinello. Langevin Equation with Colored Noise for Constant-Temperature Molecular Dynamics Simulations. *Phys. Rev. Lett.*, 102(2):020601, 2009.
 - [279] Michele Ceriotti, Giovanni Bussi, and Michele Parrinello. Colored-Noise Thermostats À La Carte. *J. Chem. Theory Comput.*, 6(4):1170–1180, 2010.
 - [280] Rafael Ramirez, Telesforo López-Ciudad, Padma Kumar P, and Dominik Marx. Quantum Corrections to Classical Time-Correlation Functions: Hydrogen Bonding and Anharmonic Floppy Modes. *J. Chem. Phys.*, 121(9):3973–3983, 2004.
 - [281] Abraham Nitzan. *Chemical Dynamics in Condensed Phases: Relaxation, Transfer and Reactions in Condensed Molecular Systems*. Oxford university press, 2006.
 - [282] Qi Yu, William B Carpenter, Nicholas HC Lewis, Andrei Tokmakoff, and Joel M Bowman. High-Level VSCF/VCI Calculations Decode the Vibrational Spectrum of the Aqueous Proton. *J. Phys. Chem. B*, 123(33):7214–7224, 2019.
 - [283] Robert S Mulliken. The Interaction of Electron Donors and Acceptors. *J. Chim. Phys.*, 61:20–38, 1964.
 - [284] Johan Åqvist and Arie Warshel. Simulation of Enzyme Reactions Using Valence Bond Force Fields and Other Hybrid Quantum/Classical Approaches. *Chem. Rev.*, 93(7):2523–2544, 1993.
 - [285] Gongyi Hong, Edina Rosta, and Arie Warshel. Using the Constrained DFT Approach in Generating Diabatic Surfaces and Off Diagonal Empirical Valence Bond Terms for Modeling Reactions in Condensed Phases. *J. Phys. Chem. B*, 110(39):19570–19574, 2006.
 - [286] Thom H. Dunning. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.*, 90(2):1007–1023, 1989.
 - [287] George D. Purvis III and Rodney J. Bartlett. A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.*, 76(4):1910–1918, 1982.
 - [288] Krishnan Raghavachari, Gary W Trucks, John A Pople, and Martin Head-Gordon. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.*, 157(6):479–483, 1989.
 - [289] Frank Neese and Edward F. Valeev. Revisiting the Atomic Natural Orbital Approach for Basis Sets: Robust Systematic Basis Sets for Explicitly Correlated and Conventional

- Correlated *Ab Initio* Methods? *J. Chem. Theory Comput.*, 7(1):33–43, 2010.
- [290] Thomas B. Adler, Gerald Knizia, and Hans Joachim Werner. A Simple and Efficient CCSD(T)-F12 Approximation. *J. Chem. Phys.*, 127:221106, 2007.
- [291] Gerald Knizia, Thomas B. Adler, and Hans Joachim Werner. Simplified CCSD(T)-F12 Methods: Theory and Benchmarks. *J. Chem. Phys.*, 130:054104, 2009.
- [292] Frank Jensen. Polarization Consistent Basis Sets: Principles. *J. Chem. Phys.*, 115(20):9113–9125, 2001.
- [293] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77(18):3865, 1996.
- [294] Jianwei Sun, Adrienn Ruzsinszky, and John P. Perdew. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.*, 115:036402, Jul 2015.
- [295] Axel D. Becke. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.
- [296] Jeng-Da Chai and Martin Head-Gordon. Long-Range Corrected Hybrid Density Functionals with Damped Atom–Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.*, 10(44):6615–6620, 2008.
- [297] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A Consistent and Accurate *Ab Initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.*, 132(15):154104, 2010.
- [298] Florian Weigend and Reinhart Ahlrichs. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.*, 7(18):3297–3305, 2005.
- [299] Dmitrij Rappoport and Filipp Furche. Property-Optimized Gaussian Basis Sets for Molecular Response Calculations. *J. Chem. Phys.*, 133(13):134105, 2010.
- [300] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györfly, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O’Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and M. Wang. *MOLPRO, a Package of Ab Initio Programs*. Cardiff, UK, version 2015.1 edition, 2015.

see <http://www.molpro.net>.

- [301] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. *Gaussian 09*. Gaussian Inc., Wallingford, CT, Revision d.01 edition, 2009.
- [302] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kús, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio Jr., H. Dop, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyayev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, P. A. Pieniazek, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, N. Sergueev, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, V. Vanovschi, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhou, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xua, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon. *Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. Mol. Phys.*, 113:184–215, 2015.
- [303] Kirk A. Peterson, Detlev Figgen, Erich Goll, Hermann Stoll, and Michael Dolg. Systematically Convergent Basis Sets with Relativistic Pseudopotentials. II. Small-Core

Pseudopotentials and Correlation Consistent Basis Sets for the Post-d Group 16–18 Elements. *J. Chem. Phys.*, 119(21):11113–11123, 2003.

- [304] J. Grant Hill, Kirk A. Peterson, Gerald Knizia, and Hans-Joachim Werner. Extrapolating MP2 and CCSD Explicitly Correlated Correlation Energies to the Complete Basis Set Limit with First and Second Row Correlation Consistent Basis Sets. *J. Chem. Phys.*, 131(19):194105, 2009.
- [305] Chr. Møller and Milton S Plesset. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.*, 46(7):618, 1934.
- [306] Adel El Azhary, Guntram Rauhut, Peter Pulay, and Hans-Joachim Werner. Analytical Energy Gradients for Local Second-Order Møller–Plesset Perturbation Theory. *J. Chem. Phys.*, 108(13):5185–5193, 1998.
- [307] Rick A Kendall, Thom H Dunning Jr, and Robert J Harrison. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.*, 96(9):6796–6806, 1992.
- [308] Brian P Prascher, David E Woon, Kirk A Peterson, Thom H Dunning, and Angela K Wilson. Gaussian Basis Sets for Use in Correlated Molecular Calculations. VII. Valence, Core-Valence, and Scalar Relativistic Basis Sets for Li, Be, Na, and Mg. *Theor. Chem. Acc.*, 128(1):69–82, 2011.
- [309] J Grant Hill and Kirk A Peterson. Gaussian Basis Sets for Use in Correlated Molecular Calculations. XI. Pseudopotential-Based and All-Electron Relativistic Basis Sets for Alkali Metal (K–Fr) and Alkaline Earth (Ca–Ra) Elements. *J. Chem. Phys.*, 147(24):244106, 2017.
- [310] Ivan S Lim, Peter Schwerdtfeger, Bernhard Metz, and Hermann Stoll. All-Electron and Relativistic Pseudopotential Studies for the Group 1 Element Polarizabilities from K to Element 119. *J. Chem. Phys.*, 122(10):104103, 2005.
- [311] George D Purvis III and Rodney J Bartlett. A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.*, 76(4):1910–1918, 1982.
- [312] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kús, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio Jr., H. Dop, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya,

- J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyayev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, P. A. Pieniazek, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, N. Sergueev, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, V. Vanovschi, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhou, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xua, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.*, 113:184–215, 2015.
- [313] Pengyu Ren and Jay W Ponder. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B*, 107(24):5933–5947, 2003.
- [314] Alan Grossfield, Pengyu Ren, and Jay W Ponder. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *J. Am. Chem. Soc.*, 125(50):15671–15682, 2003.
- [315] Pengyu Ren and Jay W Ponder. Temperature and Pressure Dependence of the AMOEBA Water Model. *J. Phys. Chem. B*, 108(35):13427–13437, 2004.
- [316] Jay W Ponder, Chuanjie Wu, Pengyu Ren, Vijay S Pande, John D Chodera, Michael J Schnieders, Imran Haque, David L Mobley, Daniel S Lambrecht, Robert A DiStasio Jr, , Martin Head-Gordon, Gary N. I. Clark, Margaret E. Johnson, and Teresa Head-Gordon. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B*, 114(8):2549–2564, 2010.
- [317] Johnny C Wu, Gaurav Chattree, and Pengyu Ren. Automation of AMOEBA Polarizable Force Field Parameterization for Small Molecules. *Theor. Chem. Acc.*, 131(3):1138, 2012.
- [318] Lee-Ping Wang, Teresa Head-Gordon, Jay W Ponder, Pengyu Ren, John D Chodera, Peter K Eastman, Todd J Martinez, and Vijay S Pande. Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B*, 117(34):9956–9972, 2013.
- [319] Thomas A Halgren. The Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters. *J. Am.*

- Chem. Soc.*, 114(20):7827–7843, 1992.
- [320] Axel D Becke and Erin R Johnson. Exchange-Hole Dipole Moment and the Dispersion Interaction. *J. Chem. Phys.*, 122(15):154104, 2005.
- [321] Erin R Johnson and Axel D Becke. A Post-Hartree–Fock Model of Intermolecular Interactions. *J. Chem. Phys.*, 123(2):024101, 2005.
- [322] Erin R Johnson and Axel D Becke. A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections. *J. Chem. Phys.*, 124(17):174104, 2006.
- [323] A Otero-De-La-Roza and Erin R Johnson. Non-Covalent Interactions and Thermochemistry Using XDM-Corrected Hybrid and Range-Separated Hybrid Density Functionals. *J. Chem. Phys.*, 138(20):204109, 2013.
- [324] D Hankins, J W Moskowitz, and F H Stillinger. Water Molecule Interactions. *J. Chem. Phys.*, 53(12):4544–4554, 1970.