Machine Learning for Multiscale Simulation of Complex Molecular Systems: Phosphorylation Reactions and Excited States Energy Surfaces

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat)

von der KIT-Fakultät für Chemie und Biowissenschaften des

Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

M. Sc. Christian Carl Schmidt

aus Ludwigsburg (Baden-Württemberg)

Tag der mündlichen Prüfung: 15. Dezember 2025
1. Referent: Prof. Dr. Marcus Elstner
2. Referent: Prof. Dr. Pascal Friederich

„Everything around you that you call life was made up by people that were no smarter than you.“

— Steve Jobs

Danksagung

Ich möchte mich zuallererst bei Prof. Dr. Marcus Elstner für die Möglichkeit zur Promotion bedanken. Danke für die Betreuung, die Hilfestellungen und die Wegweisung – sowohl im beruflichen als auch im persönlichen Kontext.

Besonderer Dank geht vor allem an Lukas Petersen: mein Mitstreiter im Kampf gegen die bösen neuronalen Netzwerke, die sich uns entgegenstellten; gegen den Code, der uns jedes Weiterkommen verweigerte; und gegen jede Rechnung, die uns wie eine feindliche Armee entgegentrat.

Danke geht an Dr. Tomáš Kubař, der bei technischen Problemen es nie müde war, Soft- oder Hardware-Anfragen zu bearbeiten. Danke an alle Kolleginnen und Kollegen, mit denen ich im fachlichen Austausch stand und die mir die Möglichkeit gegeben haben, mit ihnen zusammenzuarbeiten. Mein Dank gilt auch meinen Studentinnen und Studenten, die zu diesem Werk beigetragen und dabei meine Psyche in Mitleidenschaft gezogen haben.

Ein letzter Dank geht zum Schluss an Julian Böser, ohne den mein Verstand vermutlich nicht mehr in einem brauchbaren Zustand wäre. Ein allerletzter Dank geht an einen weiteren meiner besten Freunde, Christoph Bickmann – für alles. Wir hatten im ersten Studiensemester gemeinsam den Blick auf eine Dissertation geworfen, und du meintest: „Das ist genau das, was ich einmal machen will." Wir haben es geschafft!

Ein aller-allerletzter Dank geht an meine Eltern, die mich mein Leben lang unterstützt haben – es gibt keine Worte, die diesem Dank gerecht würden, also unterlasse ich jeden Versuch. Und ein aller-aller-allerletzter Dank geht an meinen Bruder, der mein Leben stets mit seiner positiven Art bereichert und mit seinem Humor belustigt.

Danke, danke, danke, danke, danke, danke. Danke, danke, danke, danke, danke. Danke, danke, danke, danke. Vielen Dank. Noch mehr danke. Danke an alle! Und bedenke stets:

> *„Mach dich niemals auf den Acker ohne deinen Tacker."*
>
> *— Alexander F. Schmidt*

Abstract

Molecular physics faces a blunt trade-off: the most accurate quantum methods work only for small systems, while scalable models handle bigger ones at the cost of accuracy. Complex environments, such as explicit solvents and proteins, push atom counts into the thousands to hundreds of thousands and demand long trajectories, so state dynamics and free-energy mapping remain prohibitively expensive even with methods combining quantum mechanics and molecular mechanics (QM/MM). Over the last decade, machine learning has emerged as a credible path to address broad classes of problems. Applied here, it offers the prospect of turning costly high-level calculations into fast predictors for complex systems – bridging the gap between accuracy and scalability.

This thesis develops ML tools for multiscale simulation of complex molecular systems, built around two pillars that target different sources of complexity. First, phosphorylation chemistry via Δ-QM/MM: a low-level QM baseline runs with an ML correction that recovers hybrid-DFT reaction energies and barriers. Deployed on-the-fly, the correction reshapes free-energy surfaces and removes known transition-state biases. Generalization is demonstrated by testing trained models on previously unseen phosphorylations: across a 12-system cross-validation, most held-out reactions are predicted with $\lesssim 1\,\mathrm{kcal\,mol^{-1}}$ RMSE in energy. Two interchangeable neural network backends are used as the production Δ-corrector: an HDNNP that learns energies and derives forces, and a compact MACE model that learns energies and forces jointly. Second, fluorescence with learned-QM: the QM/MM force calculation is completely replaced by a neural model trained on bright-state labels, while a property head maps structure to vertical emission energies and oscillator strengths for spectrum assembly. The on-the-fly NN/MM simulation follows the bright state through frequent reorderings and reproduces solvatochromic trends. A 1 ns trajectory at 0.5 fs resolution completes in only a few hours rather than months, enabling ensemble statistics at practical cost. A charge equilibrated MACE-EQ confirms very low force errors.

Two compact studies support these pillars. Retinal absorption reuses the fluorescence property network to predict spectra and recover qualitative protein shifts. A short sampling study surveys data generation strategies and compares their coverage and stability.

Overall, machine learning moves multiscale simulation from impractical to practical, and from seemingly impossible to possible: Δ-learning delivers transferable phosphorylation reaction free energies in solution, and learned-QM forces plus property heads enable excited-state spectra at a fraction of the cost. Since its rise over the last decade, machine learning has promised a broad toolkit for hard scientific problems; this work shows that its potential in molecular physics is far from exhausted and points toward data-driven models that accelerate materials prediction and design.

Zusammenfassung

Die Molekülphysik steht vor einem Konflikt: Die genauesten quantenchemischen Methoden sind auf kleine Systeme beschränkt, während skalierbare Methoden größere Systeme nur mit Genauigkeitseinbußen behandeln. Komplexe Umgebungen wie explizite Lösungsmittel und Proteine treiben die Atomzahlen in die Tausende bis Hunderttausende und erfordern lange Trajektorien; dadurch bleiben Zustandsdynamik und Freie-Energie-Abbildungen selbst mit kombinierten Quantenmechanik/Molekularmechanik-Ansätzen (QM/MM) prohibitiv teuer. In den letzten zehn Jahren hat sich maschinelles Lernen als glaubwürdiger Weg erwiesen, breite Klassen von Problemen anzugehen. Hier angewandt, bietet es die Aussicht, teure hochgenaue Rechnungen in schnelle Prädiktoren für komplexe Systeme zu verwandeln — und so die Lücke zwischen Genauigkeit und Skalierbarkeit zu schließen.

Diese Arbeit entwickelt ML-Werkzeuge für Multiskalen-Simulationen komplexer molekularer Systeme, aufgebaut auf zwei Säulen, die unterschiedliche Quellen der Komplexität adressieren. Erstens, Phosphorylierungschemie via Δ-QM/MM: Eine niedrigstufige QM-Baseline läuft mit einer ML-Korrektur, die Reaktionsenergien und -barrieren auf Hybrid-DFT-Niveau zurückgewinnt. Im laufenden Betrieb eingesetzt, formt die Δ-Korrektur Freie-Energie-Flächen um und entfernt bekannte Verzerrungen an Übergangszuständen. Generalisierung wird gezeigt, indem trainierte Modelle auf zuvor ungesehene Phosphorylierungen geprüft werden: Über eine Kreuzvalidierung mit 12 Systemen werden die meisten zurückgehaltenen Reaktionen mit $\lesssim 1\,\mathrm{kcal\,mol^{-1}}$ RMSE$_E$ vorhergesagt. Zwei austauschbare neuronale Backends dienen als produktiver Δ-Korrektor: ein HDNNP, das Energien lernt und Kräfte ableitet, sowie ein kompaktes MACE-Modell, das Energien und Kräfte gemeinsam lernt. Zweitens, Fluoreszenz mit gelernten-QM-Kräften: Die QM/MM-Kraftberechnung wird vollständig durch ein neuronales Modell ersetzt, das auf Labels des hellen Zustands trainiert ist, während ein Property-Head die Struktur auf vertikale Emissionsenergien und Oszillatorstärken für die Spektrenassemblierung abbildet. Die laufende NN/MM-Simulation verfolgt den hellen Zustand trotz häufiger Neuordnungen und reproduziert solvatochromische Trends. Eine 1 ns-Trajektorie mit 0,5 fs Auflösung wird pro Lösungsmittel in wenigen Stunden statt in Monaten berechnet und ermöglicht so Ensemblestatistiken zu praktischen Kosten. Ein ladungsausgeglichenes MACE-EQ bestätigt sehr geringe Kraftfehler.

Zwei kompakte Studien stützen diese Säulen. Die Retinal-Absorption nutzt das Fluoreszenz-Property-Netz erneut, um Spektren vorherzusagen und qualitative Proteinschiebungen zurückzugewinnen. Eine kurze Sampling-Studie sichtet Strategien zur Datengenerierung und vergleicht deren Abdeckung und Stabilität.

In der Summe verschiebt maschinelles Lernen Multiskalen-Simulationen vom Unpraktischen ins Praktische — und vom scheinbar Unmöglichen ins Mögliche: Δ-Learning liefert übertragbare freie Reaktionsenergien für Phosphorylierungen in Lösung, und gelernte

QM-Kräfte plus Property-Heads ermöglichen Spektren angeregter Zustände zu einem Bruchteil der Kosten. Seit seinem Aufstieg im letzten Jahrzehnt verspricht maschinelles Lernen einen breiten Werkzeugkasten für harte wissenschaftliche Probleme; diese Arbeit zeigt, dass sein Potential in der Molekülphysik bei Weitem nicht ausgeschöpft ist und weist auf datengetriebene Modelle, die die Vorhersage und das Design von Materialien beschleunigen.

Contents

Danksagung . . . . **i**

Abstract . . . . **ii**

Zusammenfassung . . . . **iii**

List of Figures . . . . **ix**

List of Tables . . . . **xii**

I. Introduction **1**
- 1.1. Scope & Contributions at a Glance 2
- 1.2. Motivation 3
- 1.3. Biomolecular Simulations at the Physics-Chemistry-Biology Interface . . 4
- 1.4. Machine-Learned Quantum Chemistry 5
- 1.5. Phosphorylation in Biological Environments 6
- 1.6. Fluorescence and Protein Sensors 7
- 1.7. Scope, Contributions, and Roadmap 7

II. Methods I: Physical Foundations & Simulation Methods **9**
- 2.1. Overview 10
- 2.2. Phosphorylation 10
- 2.3. Molecular Photophysics 11
  - 2.3.1. Ground-State Absorption 11
  - 2.3.2. Fluorescence and Solvatochromism 11
  - 2.3.3. Solvent: Dielectric Response 13
  - 2.3.4. Emissive-State Identity and State Tracking 14
  - 2.3.5. Spectrum Construction 15
- 2.4. Quantum Mechanical Methods 20
  - 2.4.1. Ab Initio Methods 20
  - 2.4.2. Density Functional Theory 21
  - 2.4.3. Density Functional Tight Binding 25
  - 2.4.4. Excited States TD-LC-DFTB 28
- 2.5. Molecular Dynamics 29
  - 2.5.1. Classical Force Field 30

2.5.2. System Propagation 31
2.5.3. Temperature and Pressure 31
2.6. Quantum Mechanics / Molecular Mechanics 32
2.7. Minimum Free Energy Path with Nudged Elastic Band 36

III. Methods II: Neural-Network Strategies 41
3.1. Overview 42
3.2. Machine Learning 42
3.2.1. Artificial Neural Networks 43
3.2.2. Backpropagation Training 44
3.2.3. Hyperparameters & Optimization 46
3.2.4. Label Z-Normalization 47
3.3. High-Dimensional Neural Network Potentials 49
3.4. MACE: Equivariant Message Passing 55
3.4.1. Graph Neural Networks 55
3.4.2. Core MACE 56
3.4.3. MACE-EQ 57
3.5. Phosphorylation: Δ-Learning Networks 59
3.5.1. Δ Network: HDNNP 60
3.5.2. Δ Network: MACE 61
3.5.3. Dataset Construction 61
3.5.4. Δ-QM/MM Runtime Coupling 63
3.6. Fluorescence Networks: QM Replacement 63
3.6.1. Force Network 64
3.6.2. Benchmark: MACE-EQ 65
3.6.3. Property Network 66
3.6.4. Training Protocols 70
3.6.5. Dataset Construction 70
3.6.6. NN/MM Runtime Coupling 70
3.7. Sampling Strategies 72
3.7.1. Data Sampling 72
3.7.2. Free-Simulation Sampling 73
3.7.3. NMA Off-Equilibrium Generator 73
3.7.4. PCA-Biased Metadynamics 74
3.7.5. RMSD-Biased Metadynamics 75
3.7.6. Uncertainty-Guided Sampling 75
3.7.7. Sampling Diagnostics and Visualization 77

IV. Computational Workflow & System Integration 81
4.1. Overview 82
4.2. Phosphorylation 83
4.3. Fluorescence 85

V. Neural Network-Driven Phosphorylation Predictions **87**
5.1. Overview 88
5.2. Introduction 88
5.3. Dataset Construction 90
  5.3.1. Reaction Candidates 91
  5.3.2. Selected Reactions 101
  5.3.3. Reaction Diagnostics 102
5.4. Δ Network: HDNNP 107
  5.4.1. Training 108
  5.4.2. Cross-Validation 110
  5.4.3. Δ-QM/MM Application 116
5.5. Δ Network: MACE 119
  5.5.1. Training 119
  5.5.2. Cross-Validation 122
  5.5.3. Δ-QM/MM Application 127
5.6. Discussion 129
5.7. Summary 131

VI. Machine-Learned Bright-State Dynamics for Fluorescence in Explicit Solvent **133**
6.1. Overview 134
6.2. Introduction 134
6.3. Dataset Construction 136
  6.3.1. Solvent: Dielectric Permittivity 136
  6.3.2. Dataset Diagnostics 138
  6.3.3. State Identity and Charge Choice 146
6.4. Force Model 149
  6.4.1. Training 149
  6.4.2. Cross-Validation 152
  6.4.3. Benchmark: MACE–EQ 154
  6.4.4. Application in NN/MM 156
6.5. Property Model 161
  6.5.1. Training 161
  6.5.2. Spectral Predictions 163
6.6. Discussion 170
6.7. Summary 174

VII. Neural Network Prediction of Retinal Absorption Spectra **175**
7.1. Overview 176
7.2. Introduction 176
7.3. Dataset Construction 178
  7.3.1. Excited-State Targets 178
  7.3.2. Diagnostics 179

7.4. Property Network 184
  7.4.1. Training 184
  7.4.2. Protein Application 189
7.5. Discussion 189
7.6. Summary 190

VIII. Sampling Strategies for ML Potentials **191**
8.1. Overview 192
8.2. Introduction 192
8.3. Case Studies 194
  8.3.1. Case: FR0 195
  8.3.2. Case: 4AP 201
  8.3.3. Case: ACDAN 204
  8.3.4. Case: Retinal 207
8.4. Discussion 211
8.5. Summary 212

IX. Summary and Outlook **213**
9.1. Scope and Context 214
9.2. Summary of Main Findings 214
9.3. Integration 215
9.4. Limitations 216
9.5. Outlook 217
9.6. Concluding Remark 218

Bibliography **219**

A. Appendix **247**

List of Figures

1. Machine Learning bridges Simulation Method Landscape. 4
2. Schematic N-Phosphorylation Reaction. 10
3. Schematic Jablonski Diagram for Absorption, Relaxation, and Emission. . . . 12
4. Passive vs. Active Environment. 13
5. Machine Learning for State Crossings. 15
6. Subtractive QM/MM Bookkeeping. 35
7. Schematic Second-Generation HDNNP. 50
8. Atom-centered Symmetry Functions. 52
9. Schematic Convolutional Layer. 52
10. Schematic Excited States Neural Network Models: Forces and Properties. . . 67
11. Excited States Force Network Architecture. 68
12. Excited States Property Network Architecture. 69
13. Phosphorylation Workflow. 84
14. Fluorescence Workflow. 86
15. Phosphorylation with Total Charge −1. 92
16. Phosphorylation with Total Charge −2. 92
17. Phosphorylation with Total Charge −2 and alternative Nucleophiles. 93
18. Phosphorylation with Total Charge −3. 93
19. Accepted for ML: Representative FES. 96
20. Excluded FES: Representative FES. 97
21. Retained Phosphorylation Reactions: Fractional-Occupation Diagnostics. . . 98
22. Discarded Phosphorylation Reactions: Fractional-Occupation Diagnostics. . 99
23. DFTB non-converged Snapshots across Reactions. 100
24. Normalized ΔE Label Distributions per Reaction. 103
25. Distributions of ΔE Labels used for Δ-learning. 104
26. Residual Phosphorylation Force Magnitudes per Reaction. 106
27. t-SNE Embedding of Phosphor Geometry Descriptors by Reaction Family. . . 106
28. Δ-learning for Total Charge −1 and −2. 108
29. Δ-learning Parity by Total Charge Combinations. 109
30. Reaction 4: 4-MeI + Ph–PO_4; LOOCV Parity Plots by Training Pool 112
31. Reaction 5: 4-MeI + *t*Bu–PO_4; LOOCV Parity Plots by Training Pool 113
32. Reaction 6: 4-MeI + Me–PO_4; LOOCV Parity Plots by Training Pool 114
33. Reaction 7: 4-MeI + Me–PP–PO_4; LOOCV Parity Plots for two Training Pools 115
34. Reaction 12: 4-MeI + Me–PP–PO_4 (HK); LOOCV Parity Plot, Trained on −2 . 115
35. On-the-fly Δ-QM/MM: FES Comparisons. 117
36. Forces along the FES: 4–MeI + Ph–PO_4Et. 118
37. Forces along the FES: 4–MeI + Me–PP–PO_4. 118

38. Forces along the FES: ImH + Me–PO_4. 118
39. Core MACE on Phosphorylation Δ-Learning. 120
40. MACE-EQ on Phosphorylation Δ-Learning. 121
41. Core MACE Δ-model: LOOCV (1–3). 123
42. Core MACE Δ-model: LOOCV (4–6). 124
43. Core MACE Δ-model: LOOCV (7–9). 125
44. Core MACE Δ-model: LOOCV (10–12). 126
45. On-the-fly Δ–QM/MM with core MACE: FES and path (2/2). 128
46. Chemical Structure of FR0. 135
47. FR0 Reference and Labeled Dihedrals. 139
48. FR0 in Solution: 2D Dihedral Occupancies of (D_1, D_2). 140
49. FR0 in Solution: 1D Formyl Dihedral. 141
50. FR0 in Solution: Planarity Diagnostics. 142
51. FR0 in Solution: Energies and Force Magnitudes. 143
52. FR0: ESP Distributions at QM Atoms by Solvent. 144
53. FR0: Vertical Excitation Energies and Oscillator Strengths by Solvent. 145
54. FR0 Partial Charges. 147
55. FR0 ESP Sensitivity by Atom and Solvent. 148
56. FR0 Parity Plots: Forces. 150
57. Force Model: Parity Plots with Density Masking. 151
58. Force Model State Energy LOSO CV: Outliers in two Solvents. 153
59. MACE-EQ Force Model. 154
60. Live NN/MM forces vs. TD-LC-DFTB (Water, DMSO). 158
61. Live NN/MM forces vs. TD-LC-DFTB (DMF, ACN). 159
62. Live NN/MM forces vs. TD-LC-DFTB (MeOH, Acetone). 159
63. Live NN/MM forces vs. TD-LC-DFTB (DCM, EtOAc). 159
64. Live NN/MM forces vs. TD-LC-DFTB (Dioxane, Hexane). 160
65. Property Network Parity (all solvents). 162
66. Fluorescence Spectra from weighted Histograms (NN vs. DFTB). 166
67. NEA vs. E–ZTFC Emission Spectra (NN/MM). 167
68. Fluorescence Spectra from three Estimators. 168
69. EGCF-Kubo Emission Spectra (Energy & Wavelength). 169
70. NN/MM Emission-Energy Trace with TD-LC-DFTB Validation (DMSO). . . . 171
71. Photoisomerization of Retinal. 177
72. Retinal Backbone and Bond Indexing. 179
73. Retinal: BLA Sampling. 180
74. Retinal: ESP Distributions. 180
75. Retinal: Excitation Energies and Oscillator Strengths. 180
76. OM2 vs. Δ. 181
77. BLA vs. E: 11-cis. 182
78. BLA vs. E: all-trans. 183
79. Property Model: Parity Plots on full Dataset. 188
80. Chemical Structure of FR0, Retinal, ACDAN, and 4AP. 193
81. FR0: Dihedrals and Reference Plane/Atoms. 195
82. FR0: 2D Dihedral Coverage for the Diethylamino Tail. 196

83. FR0: NMA Off-Equilibrium Generator. 196
84. FR0: Formyl Oxygen Dihedral across Methods. 197
85. FR0: Out-Of-Plane Angles for Selected Atoms. 198
86. FR0: Energy/Force Boxplots across Sampling Methods. 199
87. 4AP: Atoms and Reference Plane used for Planarity. 201
88. 4AP: Out-Of-Plane Planarity across Sampling Methods. 202
89. 4AP: Energy/Force Boxplots across Sampling Methods. 202
90. ACDAN: Dihedrals and Planarity Probes. 204
91. ACDAN: Planarity by Sampling Method. 205
92. ACDAN: Energy/Force Distributions across Sampling Methods. 205
93. Retinal: Numbered Polyene Backbone used for Diagnostics. 207
94. Retinal: Backbone Planarity across Sampling Methods. 208
95. Retinal: Backbone Dihedral Densities across Sampling Methods. 209
96. Retinal: Energy/Force Boxplots across Sampling Methods. 209

A.1. NN/MM Emission Traces: Water and DMSO. 260
A.2. NN/MM Emission Traces: DMF and ACN. 261
A.3. NN/MM Emission Traces: Methanol and Acetone. 262
A.4. NN/MM Emission Traces: DCM and Ethyl acetate. 263
A.5. NN/MM Emission Traces: Dioxane and Hexane. 264

List of Tables

1. Ab initio Methods Overview 21
2. Parameters for Free Energy Nudged Elastic Band Method. 39
3. Δ-ML Architecture: HDNNP. 60
4. Excited States Force Network Architecture. 64
5. Force Model: Ablation of Loss Terms. 65
6. Excited States Property Network Architecture. 66
7. Successfully Sampled Phosphorylation Reaction Systems. 91
8. Discarded Phosphorylation Reaction Systems. 91
9. Reaction-path Energetics from 2D FES (P–O vs. P–N). 95
10. ML-selected Phosphorylation Reactions. 101
11. Dataset Summary per Reaction (ΔE only; raw and z-normalized). 102
12. Phosphorylation: Reaction- and Component Atomization Energies. 105
13. Training performance by charge set and per reaction. 107
14. Phosphorylation: Cross-Validation. 111
15. FES: Δ-QM/MM with HDNNP (DFTB vs. DFTB+NN). 116
16. Core MACE Loss Schedules and Validation Errors. 119
17. Core MACE Δ-model: leave-one-reaction-out cross-validation. 122
18. FES: Δ-QM/MM with MACE. 127
19. Δ-QM/MM vs. DFTB: Runtime. 127
20. Solvents: Dielectric Permittivity – Simulation vs. Literature. 137
21. FR0 Data-Generation and Bright-State Selection. 138
22. Sampling State Occupancies and Switching Statistics. 146
23. HP Search Space and Selected Configuration. 149
24. Force Model Performance per Solvent Combination. 150
25. FR0 Force Model: Leave-One-Solvent-Out Cross-Validation. 152
26. Wall-Clock Cost: TD-LC-DFTB/MM vs. NN/MM (1 ns). 156
27. NN/MM Force Network Stability for FR0 across Solvents. 157
28. NN/MM Bright-State Statistics per Solvent (TD-LC-DFTB Assignment). . . . 157
29. NN/MM vs. TD-LC-DFTB Force Parity. 158
30. Property Network: HP Search Space vs. Selected (All 10). 162
31. Property Network: Test Performance per Solvent. 162
32. Spectrum Energy-Domain Peaks and Widths. 163
33. Spectrum Wavelength Peaks and Widths before/after Calibration. 165
34. Retinal: Training Systems Overview. 178
35. Property Network: HP Search Space for Retinal. 184
36. Retinal: cis+trans Ion Curriculum. 186
37. Retinal: Solvent Curricula. 187

38. Property Network: Test Performance across Scopes. 187
39. Property Model: Architectures. 188
40. Retinal: Protein Absorption vs. bR. 189
41. FR0: MACE MAE Grids across Sampling Strategies. 200
42. 4AP: MACE MAE Grids across Sampling Strategies. 203
43. ACDAN: MACE MAE Grids across Sampling Strategies. 206
44. Retinal: MACE MAE Grids across Sampling Strategies. 210

A.1. Settings for Phosphorylation Sampling. 248
A.2. MD Run Control. 248
A.3. Nonbonded Settings and PBC. 249
A.4. Thermostat 249
A.5. QM/MM: Engine-Level. 249
A.6. DFTB+ during Metadynamics. 249
A.7. Metadynamics (PLUMED). 250
A.8. Reference DFT for Training. 250
A.9. DFTB Single-Points for Δ-Learning. 250
A.10. Phosphorylation: Molecular and Atomization Energies. 251
A.11. Atom Reference Energies: DFT vs DFTB. 252
A.12. Vanilla MACE Training Configuration. 252
A.13. MACE-EQ Training Configuration. 252
A.14. Atomic Reference Energies for Δ-Atomization (MACE Views). 252
A.15. GROMACS QM/MM Settings: ImH + Me–PO_4. 253
A.16. GROMACS QM/MM Settings: Phenyl (Ethyl) and His Reactions. 254
A.17. DFTB(+ML) Input Settings used in On-the-fly QM/MM. 255
A.18. MM Parameter Provenance. 255
A.19. Solvent Parameter Provenance. 256
A.20. GROMACS QM/MM Settings: Excited-State Data Generation. 257
A.21. TD–LC–DFTB Single-Point Settings per Geometry. 258
A.22. MACE–EQ Hyperparameters. 259
A.23. Hyperparameters per Solvent Combination. 259
A.24. Property Network Hyperparameters per Solvent. 265
A.25. Calibration Settings for Spectrum Estimators. 265
A.26. Retinal Training Data Overview. 266
A.27. Enhanced Sampling Parameters. 267
A.28. Uncertainty and Calibration Parameters. 267
A.29. Metadynamics Configuration. 268
A.30. Solvent Models. 268
A.31. GROMACS QM/MM Settings. 269
A.32. DFTB2 Parameters. 270

Part I

Introduction

Scope & Contributions at a Glance

Goal. Implement and operationally deploy ML methods for atomistic simulation – selecting and adapting architectures, building reproducible data/training pipelines, and coupling them to DFTB+ and GROMACS – *as tools to extend or replace quantum-chemical simulations*: either by Δ-learning that augments a fast QM baseline or by NN-based forces that substitute the QM step (NN/MD). Ensemble-scale predictions (e.g., FES and fluorescence spectra) then serve to validate the engineered workflows.

Implemented toolkits. Two reusable stacks were built and used throughout this thesis: *(1) Δ-QM/MM toolkit.* Two interchangeable backends – multi-molecule HDNNP Δ-models and a MACE variant – each coupled on-the-fly to DFTB+ and GROMACS to augment the QM step during MD. The models learn residuals to a higher-level reference, yielding higher-accuracy energies/forces at near-baseline cost. *(2) Learned-QM fluorescence toolkit (NN/MM).* Two independent networks: a *force model* that replaces the per-frame excited-state QM step, and a *property model* that predicts vertical excitation/emission energies and oscillator strengths from structure+ESP. Inputs include per-atom electrostatic potentials from the MM field, making predictions *explicitly environment-sensitive*. Both are integrated on-the-fly with GROMACS; a MACE-EQ was used for benchmarking.

From data to live simulation (end-to-end). Two independent workflows are established. *Δ-QM/MM track*: from *data generation & labeling* (QM/MM sampling, QM cutout, DFT/DFTB single points, atomization-based ΔE and optional $\Delta \mathbf{F}$ targets), through *training & selection* (standardized features, HDNNP or MACE backends, hyperband model selection, multi-molecule training, validation/uncertainty controls), to *runtime coupling* via exported models with DFTB+/GROMACS hooks – at run time the learned correction augments the DFTB step each MD step. *Learned-QM fluorescence track*: from *data generation & labeling* in explicit solvent (trajectory snapshots, bright-state selection, ESP features, excitation/emission energies and oscillator strengths), through *training & selection* of two networks (a force model for QM forces; a property model for vertical excitation/emission and f; hyperband search), to *runtime coupling* where NN–MM provides forces and property predictions on-the-fly in GROMACS so spectra can be assembled from large ensembles. Datasets, descriptor libraries, and coupling APIs are provided as reusable assets.

Applications and outcomes. *Phosphorylation:* Multi-molecule Δ-learning learns transferable residuals for phosphorylation and yields sensible predictions on *held-out* reaction variants (unseen nucleophiles/leaving groups), with on-the-fly Δ-QM/MM coupling exercised. *Fluorescence (FR0):* An environment-sensitive *force model* plus a separate *property model* enables following the bright state and assembling ensemble spectra from large configuration sets at tractable cost.

1.2. Motivation

Quantum-mechanical (QM) simulations have reshaped how we understand structure, reactivity, and spectroscopy in complex molecular systems.[1,2] At the same time, computational scaling remains the dominant bottleneck when biological realism demands large systems, long trajectories, and ensemble averages.[3,4] (Fig. 1) Ground-state density-functional theory (DFT) – as formalized by Hohenberg-Kohn and Kohn-Sham[5,6] – is the workhorse for electronic-structure calculations, yet diagonalization-based implementations typically scale cubically with system size, and higher-level *ab initio* methods scale even more steeply.[3,4] For electronically excited states, time-dependent DFT (TDDFT) is widely used, but per-snapshot calculations across trajectory-level ensembles are often too costly for routine prediction of inhomogeneously broadened spectra.[2,7,8]

This thesis addresses that gap by developing *physically informed* neural-network (NN) models that either add Δ-corrections to a fast quantum baseline (e.g., DFTB3 with 3OB) or, where feasible, replace the expensive per-frame QM step with learned models, so that ensemble-scale predictions become tractable without sacrificing essential electronic-structure physics.[9–16] Recent advances in neural network potentials and Δ-learning have demonstrated near-DFT fidelity for energies, forces, and selected properties at orders-of-magnitude lower cost, providing the methodological leverage required for realistic ensembles.[9,12–14]

Two scientific drivers motivate the applications considered here. First, **phosphorylation** is a cornerstone of cellular regulation and energy transduction – kinase signaling, ATP-coupled metabolism, and on/off modulation of protein function hinge on forming and breaking P-O bonds – so predictive models in realistic environments directly impact mechanistic insight and inhibitor design.[17,18] Second, **fluorescence** with solvatochromic dyes converts nanoscale changes in local environment into optical signals and underpins a vision of minimally invasive biosensing and potentially continuous biochemical monitoring, where robust molecular readouts could inform diagnostics and closed-loop therapeutics.[19,20] A compact supporting component demonstrates portability of the fluorescence workflow to a second chromophore (retinal) and sketches how configuration-space coverage can be extended beyond the setups used here via data-driven collective variables and enhanced-sampling strategies.

The unifying aim is to make *routine, ensemble-scale* predictions of free energies and spectra in chemically diverse, explicitly solvated settings feasible at biomolecular scale, thereby shortening design cycles in areas ranging from catalysis to biosensing.[7,8,21] Recent reviews synthesize the emerging ML/MM landscape, emphasizing two practical routes – Δ-ML on top of a semiempirical/DFT baseline and learned intramolecular potentials under an external MM field – together with charge-aware and increasingly polarizable couplings for transferability in condensed phase.[22]

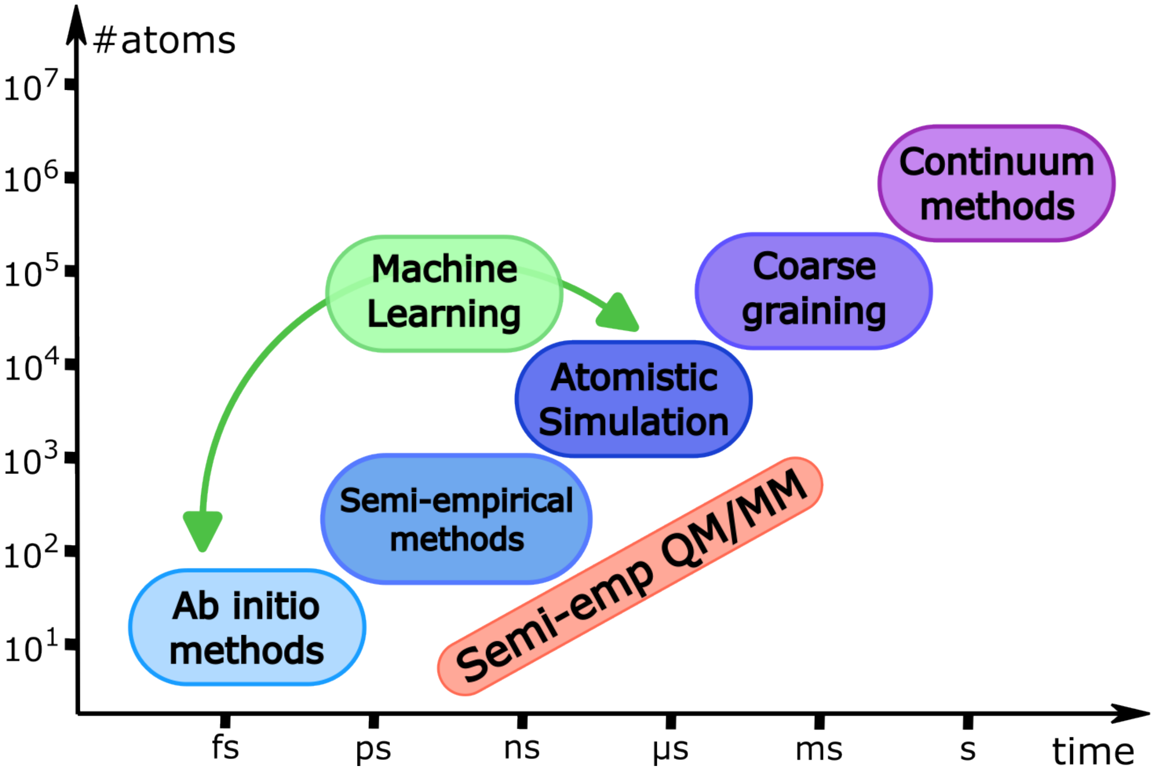


Figure 1.: Method landscape (schematic): accessible system sizes (#atoms) vs. time. Ab initio and semiempirical QM anchor the small/short end; atomistic MD, coarse graining, and continuum models extend reach. The green arrows indicate how machine learning can *bridge the gap* by elevating a fast baseline (Δ-learning) or replacing the per-frame QM step, pushing atomistic simulations toward larger systems and longer times at practical cost. Positions are illustrative. Adapted from Keith (2021).[23]

1.3. Biomolecular Simulations at the Physics-Chemistry-Biology Interface

Biomolecular simulations integrate physical laws with chemical specificity to interrogate biological function, combining statistical mechanics with atomistic models to connect structure, dynamics, and observables.[24,25] Classical molecular dynamics (MD) provides access to long-time trajectories and thermally relevant configurational ensembles for proteins, nucleic acids, and small molecules, and has advanced our understanding of enzyme mechanisms and ligand recognition in biophysics and biochemistry.[24,26] Quantum-mechanical (QM) methods are employed when explicit electronic responses – bond-making/breaking, polarization, charge transfer – are essential to the observable of interest.[1,27]

For ground states, density-functional theory (DFT) – as formalized by Hohenberg-Kohn and Kohn-Sham[5,6] – is the workhorse of electronic-structure calculations in chemistry and materials science, yet standard diagonalization typically leads to cubic scaling with system size and higher-level *ab initio* methods scale even more steeply in practice.[3,4,28,29] Hybrid multiscale approaches such as QM/MM localize the expensive QM treatment while embedding it in a classical environment to balance accuracy and cost, enabling reactive centers to be treated realistically within proteins and enzymes.[1,27]

For electronically excited states, time-dependent DFT (TDDFT) is widely used for organic chromophores and biomolecular probes in solution, but evaluating excited-state properties across large thermal ensembles – required for realistic, inhomogeneously broadened spectra – is often computationally prohibitive in routine workflows, leading to practical restrictions that can affect fidelity.[2,7,8,30]

The observables central to this work – reaction free energies and barrier statistics for reactivity, and broadened spectra for spectroscopy – are ensemble properties that require averaging over thermally accessible configurations (with spectra assembled from many configurations for quantitative comparison to experiment).[7,8,21,31] Accurate free-energy profiles are obtained by coupling biased sampling to proper reweighting, for example via WHAM or MBAR.[32,33]

Semiempirical quantum methods such as DFTB3 with the 3OB parametrization provide a convenient, MD-compatible baseline that captures broad chemical trends at substantially reduced cost.[10,11] These scaling and ensemble demands motivate machine-learning acceleration that preserves essential electronic-structure physics while making routine, statistically robust predictions feasible at biomolecular scale.[9,12,14]

1.4. Machine-Learned Quantum Chemistry

The first wave of machine-learned quantum chemistry established that neural network potentials (NNPs) can approximate high-dimensional potential-energy surfaces for small organic molecules in the gas phase with near-DFT accuracy at orders-of-magnitude lower cost.[12–15] This progress was catalyzed by standardized datasets (e.g., QM7/QM9, MD17, ANI) and by architectures that encode locality and symmetries.[13,34–36] In parallel, Δ-learning showed that correcting a fast quantum baseline toward a higher-level reference is a pragmatic route to accuracy that retains the baseline's MD compatibility and robustness.[9–11]

In the past five years, the scope has expanded from gas-phase benchmarks to condensed-phase and multiscale settings. Equivariant message-passing networks (e.g., PaiNN, NequIP, MACE) improved data efficiency and the smoothness of learned forces, enabling longer, more stable MD and broader chemical coverage; at the same time, ML has been integrated with QM/MM and enhanced sampling – either as a Δ-correction that augments a semiempirical or DFT baseline or, where feasible, as a direct learned-QM surrogate within NN/MD.[23,37–39] Active/uncertainty-driven selection further reduced labeling cost and helped prevent out-of-distribution drift during dynamics.[40] A recent overview consolidates these developments for condensed-phase ML/MM, highlighting explicit environment signals (ESP), flexible charges, and polarizable embeddings as key ingredients for robust transfer across microenvironments.[22]

For excited states, recent reviews document rapid progress from property prediction (E_{vert}, oscillator strengths) to learned excited-state PESs and non-adiabatic couplings for small-/medium systems, with first steps toward solvent-aware modeling.[16] Nevertheless, many established pipelines either target gas-phase molecules or avoid explicit environments. In explicitly solvated simulations two challenges dominate: (i) **scalability** (large systems, long trajectories, ensemble statistics) and (ii) **transferability** across chemistries, charge states, and microenvironments. Direct "learned-QM" models trained on curated gas-phase data often struggle here – particularly along reactive coordinates or when long-range polarization matters – so hybrid strategies that retain a quantum baseline (via Δ-learning) or make environment signals explicit in the input tend to be more reliable for condensed phase.[16,23]

The contributions of this thesis, developed during the last three to four years, sit squarely within this shift toward realistic condensed-phase simulation. For reactivity, a Δ-QM/MM toolkit targets **phosphorylation** as an archetype of complex P-centered chemistry – with highly polarized, multi-center bonding and strong environment dependence – and learns transferable corrections across many reaction variants (unseen nucleophiles/leaving groups), enabling potentials of mean force at biomolecular scale while preserving the robustness and embedding of the DFTB baseline.[9–11] For spectroscopy, a **learned-QM (NN–MM)** workflow for solvatochromic fluorescence in explicit solvent incorporates environment signals directly into the model to deliver stable dynamics and ensemble-scale spectra, and its portability is illustrated on retinal. Across both settings, the unifying theme is scalable, transferable ML for explicit-solvent, multiscale simulations, supported by coverage diagnostics and sampling strategies designed to keep models within their domain of reliability.[15,16,38,39]

1.5. Phosphorylation in Biological Environments

Phosphorylation is a cornerstone of cellular regulation and energy transduction – kinase signaling, metabolic switching via ATP and related phosphates, and controlled on/off modulation of protein function all hinge on forming and breaking P–O bonds – making predictive models of these reactions directly relevant to biology and medicine.[17,18] From a technology perspective, reliable computation of phosphoryl chemistry in realistic environments would inform inhibitor design, enzyme engineering, and the interpretation of phospho-proteomics, advancing both basic mechanistic insight and translational applications.[18]

Accurate modeling is challenging because phosphorylation traverses strongly polarized, charge-transfer, and occasionally hypervalent bonding regions around phosphorus; barriers and reaction energies are highly environment dependent and must be evaluated as ensemble free energies rather than single-geometry numbers.[1,21,31] While first-principles electronic-structure methods provide the necessary physics, their cost scales steeply with system size and sampling length, so straightforward QM or QM/MM evaluations along full reaction ensembles are often prohibitive.[1,3,4,27]

In this thesis we address the bottleneck by elevating a fast quantum baseline (DFTB3 with appropriate organic parameters) toward hybrid-DFT fidelity via Δ-learning, yielding near-DFT forces and energetics at DFTB-like speed for the computation of potentials of mean force and barrier statistics.[9–11] The ML models are trained and validated on chemically diverse microenvironments and coupled to enhanced-sampling with proper reweighting to produce statistically robust free energies at biomolecular scale, with uncertainty-aware selection and active learning to mitigate out-of-distribution configurations.[12,21,33]

1.6. Fluorescence and Protein Sensors

Fluorescence spectroscopy converts nanoscale changes in polarity, hydrogen bonding, and packing into measurable shifts of absorption and emission, making solvatochromic dyes powerful reporters of local environment.[19] Protein-dye constructs exploit this principle for biosensing by translating molecular recognition into optical signals; in the glucose/galactose-binding protein (GGBP/GBP), strategically placed environment-sensitive dyes yield pronounced glucose-dependent responses, and simulations illuminate how binding reshapes hydrogen-bond networks and chromophore conformations.[20,41] Looking ahead, such constructs underpin a vision of minimally invasive, potentially continuous biochemical monitoring, where robust molecular readouts could inform diagnostics and closed-loop therapeutics; strongly solvatochromic probes like FR0 (a fluorene analogue of Prodan) are attractive building blocks for this purpose.[19,42] From a modeling perspective, realistic spectra are ensemble properties that require extensive sampling and per-snapshot excited-state evaluations to reconstruct inhomogeneously broadened line shapes, and explicit-solvent treatments are often necessary to capture specific solute-solvent interactions that drive solvatochromism.[8,43–45] This computational burden motivates accelerated approaches that preserve the relevant physics while enabling ensemble-scale predictions.[43] Related ML/MM studies on excited-state reactivity and spectroscopy in solution likewise point to solvent-aware descriptors, charge flexibility, and multi-state formulations as the most reliable path forward.[22]

In this thesis we take a step toward protein embedding by establishing and validating a solution-phase workflow for FR0: learned NNPs provide forces to propagate Born–Oppenheimer dynamics on a learned surface, and auxiliary models predict emission energies and oscillator strengths so that spectra can be assembled from many configurations at tractable cost.[8,16,23,43,46]

1.7. Scope, Contributions, and Roadmap

This thesis focuses on two core scientific settings in which electronic-structure cost constrains biological realism: (i) phosphorylation in realistic environments (reactivity) and (ii) fluorescence of the strongly solvatochromic dye FR0 in explicit solvent (spectroscopy). The aim is to *implement and operationalize* machine-learning toolkits that either *extend* a fast quantum baseline via Δ-learning or *replace* the per-frame QM step with learned forces, so that ensemble-scale simulations become feasible at practical computational cost. The emphasis is on methods *as used and implemented in this thesis*: architectural templates, dataset conventions, training protocols, and tight runtime coupling to DFTB+ and GROMACS.

Guided by an implementation-first perspective, the contributions are as follows. A Δ-QM/MM toolkit with two interchangeable backends – multi-molecule HDNNPs and a MACE/MACE-EQ variant – is implemented, trained on atomization-based Δ targets, and integrated on-the-fly so that learned corrections augment the DFTB step during QM/MM. Models are trained across multiple reaction families and evaluated on held-out variants (unseen nucleophiles/leaving groups), indicating transferable residuals within phosphoryl chemistry. For fluorescence, a learned-QM (NN–MM) toolkit is established comprising two independent networks: a force model that provides Born–Oppenheimer forces on a learned surface, and a property model that predicts vertical excitation/emission energies and oscillator strengths from structure together with per-atom ESP, yielding *explicitly environment-sensitive* predictions. This enables following the bright state and assembling ensemble spectra in solution at tractable cost; training is performed at the (TD-)DFTB level, and direct experimental benchmarking lies outside the present scope. Across both applications, reusable assets—standardized datasets, descriptor handling, training/selection protocols, and runtime hooks – are provided for end-to-end use.

The remainder of the thesis is organized to reflect this flow from background to implementation and application. *Methods I* (Chap. II) reviews the physics background (QM, MM, QM/MM, ensembles). *Methods II* (Chap. III) sets out the machine-learning methods as actually employed here, including network specifications, dataset conventions, training protocols, and coupling patterns, recording what is implemented and used in later chapters. *Integration* (Chap. IV) composes these building blocks into coherent, end-to-end workflows with schematics and runtime interfaces. *Phosphorylation* (Chap. V) applies and tests the Δ-QM/MM toolkit; *Fluorescence* (Chap. VI) establishes the solution-phase learned-QM workflow for FR0; *Retinal* (Chap. VII) demonstrates portability of the fluorescence pipeline to a second chromophore. Finally, *Sampling* (Chap. VIII) summarizes the production choices (unbiased QM/MM, well-tempered metadynamics) and *goes beyond* them by benchmarking exploratory generators, before a concluding *Summary/Outlook* (Chap. IX) synthesizes lessons and limitations.

Part II

Methods I: Physical Foundations & Simulation Methods

2.1. Overview

This chapter provides two components. *First*, concise physical background needed for the later results: condensed-phase phosphorylation (Sec. 2.2) and molecular photophysics of solvatochromic dyes (Sec. 2.3). *Second*, the computational toolbox used throughout this work: quantum-chemical models (DFT/DFTB and TD-LC-DFTB), classical molecular dynamics (MD), and hybrid QM/MM simulations. Project-specific machine-learning architectures and dataset workflows are deferred to the companion Methods II chapter.

Units and conventions. Unless stated otherwise, all energies are reported in electronvolts (eV) or Hartree (Ha), lengths in ångström (Å), times in femtoseconds (fs), temperatures in Kelvin (K), and oscillator strengths are dimensionless. For dielectric properties, **SI units** are used: dipole moments in Cm and volumes in m^3. Where internal atomic units are used in formulae (e.g., oscillator strengths in the length form), the corresponding physical units are stated alongside the equations if needed.

2.2. Phosphorylation

Phosphoryl-transfer reactions proceed by nucleophilic attack on a phosphorus center with concerted reorganization of P–O and P–N bonds in a condensed-phase environment. Depending on substituents and solvation, mechanisms span a continuum between associative (pentacoordinate, S_N2(P)-like) and dissociative (metaphosphate-like) limits, with trigonal-bipyramidal transition structures commonly found in quantum-chemical calculations. Local electrostatics and protonation states shape the potential-energy surface in condensed phase; where present, specific metal coordination (e.g., Mg^{2+}) can stabilize anionic intermediates and transition-state–like arrangements. A representative reaction is illustrated in Fig. 2.

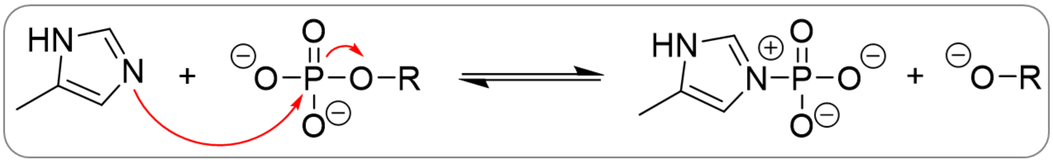


Figure 2.: Representative N-phosphorylation. An imidazole nucleophile attacks a phosphate monoester (left), forming a phosphoramidate and an alkoxide leaving group (right). Red arrows indicate electron flow; charges are shown for one resonance form. The diagram serves as a concrete reference example for the condensed-phase mechanisms discussed in Sec. 2.2.

2.3. Molecular Photophysics

Molecular photophysics concerns how molecular chromophores interact with electromagnetic radiation in condensed environments. Electronic excitation is followed by vibronic and solvent relaxation, which together control intensities, peak positions, and line shapes. Throughout, we compute spectroscopically relevant quantities – oscillator strengths and vertical emission energies – that serve as inputs to the spectral simulations, while accounting for how the environment can stabilize, mix, or reorder nearby singlet states.

2.3.1. Ground-State Absorption

The probability of absorption or emission is quantified by the dimensionless oscillator strength f.[47–49] For a transition $|1\rangle \longrightarrow |2\rangle$ with energies E_1, E_2 one has

$$f_{12} = \frac{|E_2 - E_1|}{3a_0^2 R_\mathrm{y}} \sum_{\alpha=x,y,z} \left|\langle 1 \,|\, \hat{R}_\alpha \,|\, 2\rangle\right|^2, \tag{2.1}$$

where the many-electron position operator along direction α is

$$\hat{R}_\alpha = \sum_{i=1}^{N_e} \hat{r}_{i,\alpha}. \tag{2.2}$$

Here $a_0 \approx 0.529\,\text{Å}$ is the Bohr radius and $R_\mathrm{y} \approx 13.61\,\mathrm{eV}$ the Rydberg energy. Large f corresponds to a *bright* transition observable with high intensity in absorption or emission; small f defines a *dark* transition.[47,48] A schematic of absorption and fluorescence steps is shown in Fig. 3.

2.3.2. Fluorescence and Solvatochromism

Fluorescence in molecular chromophores proceeds in three steps.[48,50] See also Fig. 3. First, the fluorophore absorbs light and is promoted to an electronically and vibrationally excited state. As indicated by the Jablonski diagram, the system then relaxes nonradiatively to a lower vibrational level, contributing to the Stokes shift.[50] Finally, radiative decay to the electronic ground state occurs – typically on the nanosecond timescale – and often ends in an excited vibrational level, further contributing to the Stokes shift.[48,50] The instability of the excited state reflects that an allowed electronic transition connects the excited and ground states.[47]

For the FR0 molecule analyzed here, only singlet states are considered,[42] because triplet lifetimes are typically longer and phosphorescence quantum yields are low under the present conditions.[51] In the Jablonski picture, additional singlet states (e.g., S_2) lie above S_1.[52] Because of the separation of timescales between internal conversion and fluorescence, population generally relaxes to S_1 prior to emission in accordance with Kasha's rule.[53]

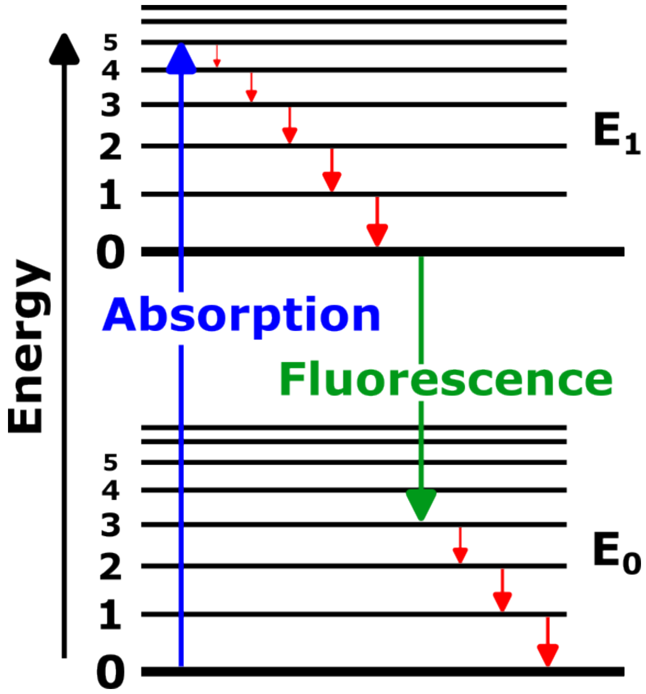


Figure 3.: Schematic Jablonski diagram illustrating absorption (blue), nonradiative vibrational relaxation (red), and fluorescence (green) between the ground (E_0) and an excited (E_1) electronic state.

For clarity, Fig. 3 depicts only S_1. Anti-Kasha emission from higher singlets has been reported in some systems.[54] In our calculations, emission properties are evaluated for multiple low-lying singlet states, and spectral construction uses the state with substantial oscillator strength at a given geometry; thus, contributions that formally correspond to higher-index singlets can be represented. However, we do not model nonadiabatic lifetimes or population kinetics, so establishing mechanistic anti-Kasha behavior for FR0 lies beyond the present scope.

A key observable in this chapter is the environment dependence of emission (solvatochromism).[55,56] States with larger dipole moments are stabilized in polar environments; if emission proceeds from such a state, the spectrum red-shifts.[56–58]

Standard workflows draw many geometries from MD and evaluate excited states per geometry (ensemble/NEA), often augmented by FC vibronic structure and by QM/MM embedding for explicit environments.[43,45,59,60] For large biochromophores, LR-LC-TD-DFTB (linear-response time-dependent long-range-corrected DFTB) offers cost-effectiveness with competitive trends relative to LR-TD-DFT,[61] and analytical LC-TD-DFTB gradients enable excited-state MD and QM/MM sampling.[62] Multiscale simulations provide the framework to combine a quantum region with a large explicit environment.[27,45]

2.3.3. Solvent: Dielectric Response

For solvent model diagnostics the relative permittivity is computed from total dipole fluctuations in SI units under metallic boundary conditions,[25,63,64]

$$\varepsilon_r = 1 + \frac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{3\,\varepsilon_0\,V\,k_\mathrm{B} T}, \tag{2.3}$$

with total dipole moment $\mathbf{M}$ (C m), simulation box volume V (m^3), and Boltzmann constant k_B ($\mathrm{J\,K^{-1}}$). Sufficiently large systems and long sampling ensure convergence of Eq. (2.3).

Passive vs. active environment. Following Toldo *et al.*, we distinguish a *passive environment*, in which the surroundings shift excited-state energies without changing their order, from an *active environment*, in which solvent or protein reorganization can invert the ordering of the lowest singlets along nuclear motion.[65] Figure 4 visualizes this distinction: the left panel shows a passive environment (level shifts without reordering), whereas the right panel illustrates an active environment with solvent-driven S_1/S_2 reordering; labels (1)–(3) indicate vertical excitation, relaxation, and solvent-driven crossing, respectively. For Prodan-type push–pull dyes (including FR0), two low-lying singlets of different character – a locally excited state (LE, often ^1L_b) and an intramolecular charge-transfer state (ICT, often ^1L_a) – dominate the oscillator-strength budget; their relative order is strongly environment-sensitive, and solvent polarity and specific interactions can drive frequent S_1/S_2 reorderings and, more rarely, transient brightening of a higher singlet (e.g., S_3).[19,42,55] This motivates identity-aware analysis and our oscillator-strength–based bright-state labeling (Sec. 2.3.4).

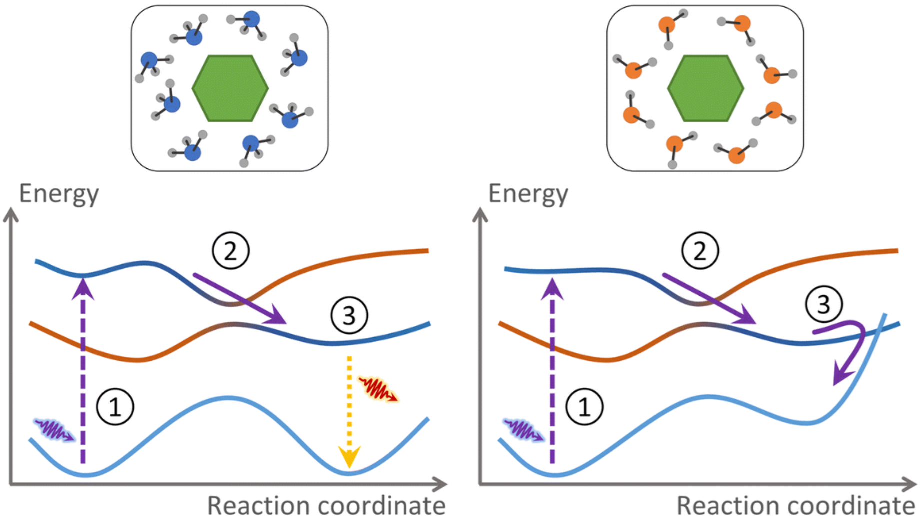


Figure 4.: Passive vs. active environment. Left: a passive environment shifts levels but does not reorder the lowest singlets. Right: an active environment (e.g., specific solvent motions) drives reordering within the S_1/S_2 manifold; (1) vertical excitation, (2) relaxation, (3) solvent-driven crossing. Adapted from Toldo Ref.[65]. Licensed under CC BY 3.0.

Push–pull dyes. Push-pull dyes are donor-acceptor chromophores with a conjugated bridge (D-π-A): an electron-donating group and an electron-withdrawing group are connected through π-conjugation, so that photoexcitation promotes an intramolecular charge-transfer (ICT) state.[19,50] The large excited-state dipole ($\Delta\mu \gg 0$) underlies strong solvatochromism and environmental sensitivity.[56,58] In this thesis we use FR0 – a fluorene analogue of Prodan – as the primary model dye; its structure is shown in Fig. 46 in the Fluorescence chapter (Sec. VI).[42]

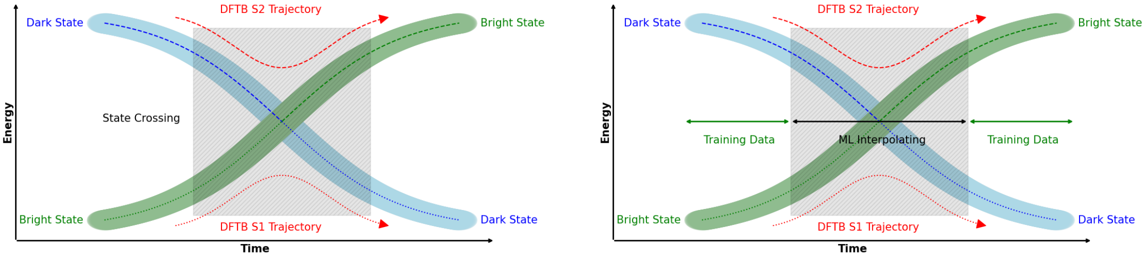
2.3.4. Emissive-State Identity and State Tracking

In condensed environments, several low-lying singlets (typically LE and ICT) can lie close in energy and *reorder* along nuclear motion. Adiabatic indices ("S_1", "S_2") therefore do not label unique *state identities*; index-based analysis induces *root flipping* and can steer dynamics onto the wrong electronic state.[66,6768–70] For fluorescence, the emissive (bright) surface is decisive; we thus identify it operationally via the oscillator strength f (see Eq. 2.1) and follow the *bright state* across the S_1/S_2 manifold.

Figure 5 summarizes the strategy. Panel 5a illustrates that tracking by adiabatic index can follow the *wrong* trajectory near an S_1/S_2 crossing, whereas identity tracking follows the physical bright surface. Panel 5b shows our per-geometry labeling: for each snapshot we keep a single *bright* state (large f) and *discard* geometries with ambiguous identity (e.g., small maximum f or near the crossing). This deliberately avoids providing labels in the immediate crossing region. As a result, the ML model is tasked to *interpolate* smoothly between well-identified regimes on either side of the crossing, rather than to extrapolate from mislabeled or contradictory data; see Panel 5c.

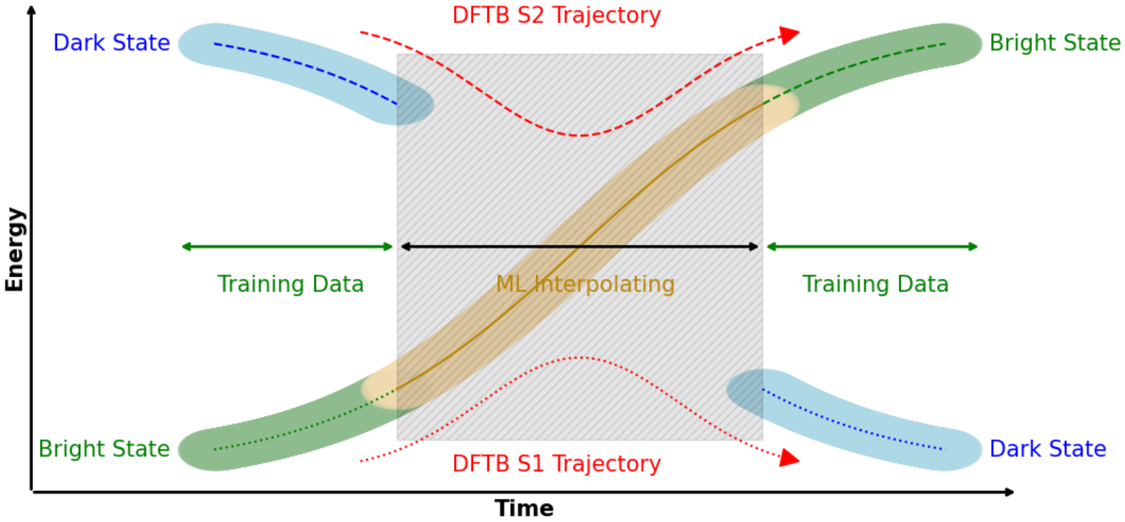
Nonadiabatic dynamics (FSSH, AIMS, (ML-)MCTDH) provide principled state tracking,[71–73] but explicit-solvent, nanosecond trajectories remain impractical in routine workflows.[74] We therefore use f as an identity cue and explicitly sample the bright surface; numerical thresholds and exact selection parameters are given later in the dedicated machine-learning methods chapter.

Scope note: protein-embedded chromophores. The condensed-phase photophysics and QM/MM machinery summarized above are used unchanged for protein-bound retinal. Colour tuning in rhodopsins is governed by conjugation (bond-length alternation, BLA) and local electrostatics around the protonated Schiff base; the corresponding application-specific background and data design are treated in the retinal chapter (Chapter VII). Machine-learning architecture and loss functions are identical to those introduced for fluorescence and are detailed in chapter III.



(a) Adiabatic S_1/S_2 crossing: index tracking (red) fails to follow the physical bright surface.

(b) Per-geometry bright/dark labeling; ambiguous frames near the crossing are excluded.



(c) The NN interpolates a continuous bright surface across the excluded crossing region.

Figure 5.: Machine Learning for States Crossings: State identity vs. adiabatic indices near S_1/S_2 crossings. Identity tracking via the oscillator strength f avoids root flipping and focuses sampling on the emissive surface.

2.3.5. Spectrum Construction

This section formulates two practical estimators for absorption/emission spectra from trajectory snapshots: (i) the *Nuclear Ensemble Approach* (NEA), and (ii) the combined *Ensemble-Zero-Temperature Franck–Condon* (E-ZTFC) method. Formal background and reviews are given in Refs.[59,75] (NEA) and Refs.[60,76] (E-ZTFC).

Inputs. From a trajectory, a set of snapshots $\{\mathbf{R}_i\}_{i=1}^{N}$ is sampled. For each snapshot i the vertical transition energy ΔE_i (absorption or emission) and an associated intensity factor S_i are evaluated. For electric–dipole allowed transitions, S_i is commonly taken proportional to the oscillator strength f_i.[59,60] Angular frequencies are defined by $\omega_i = \Delta E_i/\hbar$ (used below only when working on a frequency axis). A spectral kernel $K_\sigma(\omega)$ (typically Gaussian) with width parameter σ models homogeneous/instrumental broadening.[59,76]

2.3.5.1. Nuclear Ensemble Approach (NEA)

NEA accounts for *inhomogeneous* (temperature- and environment-induced) broadening by averaging vertical gaps and intensities over a snapshot ensemble; explicit vibronic progressions are not included beyond what is effectively mimicked by the ensemble.[59,75]

Estimator. Given $\{(\omega_i, S_i)\}_{i=1}^{N}$, the (unnormalized) spectrum reads

$$I_{\mathrm{NEA}}(\omega) \;=\; \sum_{i=1}^{N} w_i\, S_i\, K_\sigma\big(\omega - \omega_i\big), \tag{2.4}$$

with optional weights w_i (default $w_i{=}1/N$). The kernel K_σ can be chosen Gaussian,

$$K_\sigma(\Delta) = \exp\Big[-\frac{\Delta^2}{2\sigma^2}\Big], \tag{2.5}$$

or Lorentzian; normalization is applied after assembly.[59] If snapshots were generated from a distribution different from the target ensemble (e.g. biased sampling), w_i incorporate the usual reweighting factors (Boltzmann or bias-removal).[77,78]

Practical recipe (energy domain).

1. **Axis & grid.** Work in *energy* (eV). Define a fine grid $\{E_k\}_{k=1}^{K}$ that covers your snapshot energies with margin (e.g., $E_{\min} - 3\sigma$ to $E_{\max} + 3\sigma$, step 1–5 meV). A safe rule is $\Delta E \lesssim \sigma/10$.[59]

2. **Per-snapshot quantities.** For each snapshot $i = 1, \dots, N$

$$E_i \equiv \Delta E_i \quad \text{(in eV)}, \qquad S_i = \begin{cases} f_i, & \text{absorption or relative emission shape,} \\ E_i^3\, f_i, & \text{emission with } \omega^3 \text{ weighting (shape).} \end{cases} \tag{2.6}$$

   The ω^3 (E^3) emission weighting reflects the frequency dependence of the spontaneous emission rate and is standard in snapshot-based estimators.[60,76] If sampling is unbiased and all frames are taken at equal time stride, set $w_i{=}1$; for biased sampling use the reweighting w_i (Boltzmann/bias removal).[77]

3. **Kernel K_σ & width σ.** Use a Gaussian lineshape (homogeneous/instrumental broadening) with width σ in *eV*

$$K_\sigma(\Delta) \;=\; \exp\Big[-\tfrac{\Delta^2}{2\sigma^2}\Big], \qquad \Delta_{ik} \equiv E_k - E_i. \tag{2.7}$$

   Choose σ either from an experimental FWHM (*Full Width at Half Maximum*) via

$$\sigma = \frac{\mathrm{FWHM}}{2\sqrt{2\ln 2}} \;\approx\; \frac{\mathrm{FWHM}}{2.355}, \tag{2.8}$$

   or pick a practical range 0.03–0.10 eV and calibrate against a reference.[59,76] (Lorentz/Voigt are possible; Gaussian is standard.)

4. **Assemble the spectrum (unnormalized) & normalize.**

$$I_E(E_k) \;=\; \sum_{i=1}^{N} w_i\, S_i\; \exp\Big[-\tfrac{(E_k-E_i)^2}{2\sigma^2}\Big]. \tag{2.9}$$

   There are two main *normalization options*

$$\text{(area-normalized)}\quad \tilde{I}_E(E_k) = \frac{I_E(E_k)}{\sum_{k=1}^{K} I_E(E_k)\,\Delta E}, \tag{2.10}$$

$$\text{(peak-normalized)}\quad \tilde{I}_E(E_k) = \frac{I_E(E_k)}{\max_{1\le k\le K} I_E(E_k)}. \tag{2.11}$$

   Area normalization is common for comparing spectral shapes across conditions.[59]

5. **Plot in wavelength.** To plot on a wavelength axis λ (nm), transform *after* building I_E. With $E(\lambda) = hc/\lambda$ and $hc \approx 1239.84$ eV nm, the spectral density transforms with the Jacobian

$$I_\lambda(\lambda) \;=\; I_E\big(E(\lambda)\big)\left|\tfrac{dE}{d\lambda}\right| \;=\; I_E\big(\tfrac{hc}{\lambda}\big)\,\frac{hc}{\lambda^2}. \tag{2.12}$$

   (First interpolate I_E at $E{=}hc/\lambda$, then apply the Jacobian; renormalize if desired.) This post-construction transform avoids shape distortions from a nonlinear axis change.[59]

2.3.5.2. Ensemble–Zero-Temperature Franck–Condon (E-ZTFC)

Contrast to NEA. NEA captures *inhomogeneous* broadening by averaging vertical transition energies over a snapshot ensemble and smoothing with a simple lineshape kernel; explicit vibronic progressions are not included beyond what is implicitly mimicked by the ensemble.[59,75] E-ZTFC augments this ensemble distribution with a precomputed *zero-temperature Franck–Condon* vibronic shape $J_{\mathrm{ZTFC}}(\omega)$ and combines both by convolution, thereby adding systematic high-frequency vibronic structure to the inhomogeneously broadened profile.[60,76]

What changes in practice?

1. Compute once the vibronic 0 K shape $J_{\mathrm{ZTFC}}(\omega)$ (harmonic FC analysis) for a reference geometry $\mathbf{R}_0$ (gas phase or small cluster).[76]

2. Build the snapshot ensemble $\{(\omega_i, S_i, w_i)\}$ exactly as in NEA (environment/temperature enter here).[59]

3. Replace the NEA kernel $K_\sigma(\omega-\omega_i)$ by the shifted vibronic shape $J_{\mathrm{ZTFC}}\big(\omega-(\omega_i-\omega_0)\big)$, which is equivalent to the convolution form.[60,76]

4. Optionally add a modest homogeneous broadening; normalize as for NEA (area or peak).

Construction. Let $J_{\text{ZTFC}}(\omega)$ be the zero-temperature vibronic lineshape (dimensionless; normalized), centered at a reference vertical transition frequency ω_0 of a chosen geometry $\mathbf{R}_0$. Define the ensemble distribution $P(\epsilon) = \sum_i w_i\, S_i\, \delta(\epsilon - \omega_i + \omega_0)$, i.e., vertical energies relative to ω_0 weighted by snapshot intensities. E-ZTFC then approximates the spectrum by the *convolution*

$$I_{\text{E-ZTFC}}(\omega) \;=\; \int d\epsilon\; P(\epsilon)\; J_{\text{ZTFC}}(\omega - \epsilon), \tag{2.13}$$

optionally followed by a small homogeneous broadening. In discrete practice this is the sum of shifted copies of J_{ZTFC}:

$$I_{\text{E-ZTFC}}(\omega) \;\approx\; \sum_{i=1}^{N} w_i\, S_i\; J_{\text{ZTFC}}\big(\omega - (\omega_i - \omega_0)\big). \tag{2.14}$$

Practical recipe.

1. Compute $J_{\text{ZTFC}}(\omega)$ (harmonic FC, T=0) for the chromophore at $\mathbf{R}_0$.[76]
2. Generate $\{(\omega_i, S_i, w_i)\}$ as in NEA.[59]
3. Assemble $I_{\text{E-ZTFC}}$ via Eq. (2.14) (or Eq. (2.13)); apply optional homogeneous broadening and normalization.[60]

Remarks and caveats. (i) Choose ω_0 consistently (e.g., the vertical frequency at $\mathbf{R}_0$). (ii) Avoid double counting: all temperature/solvent broadening is carried by the *ensemble*; the ZTFC part remains at T=0.[76] (iii) For very strong, nonlocal chromophore–solvent coupling, fully dynamical approaches (e.g., cumulant expansion) may be preferable.[7]

2.3.5.3. Energy-gap correlation and cumulant lineshape (EGCF–Kubo)

A complementary, dynamics-based route constructs spectra from the *time correlation* of the vertical energy gap along a trajectory.[7,59,75] In this *energy-gap correlation* (EGCF) framework the instantaneous emission (or absorption) gap $\Delta E(t)$ is sampled along a trajectory. After removing the mean, $\delta\epsilon(t) = \Delta E(t) - \langle \Delta E \rangle$, the (stationary) autocorrelation function is

$$C(t) \;=\; \big\langle\, \delta\epsilon(0)\, \delta\epsilon(t)\, \big\rangle. \tag{2.15}$$

Practically, $C(t)$ is estimated by windowed time averages, often followed by gentle damping (e.g., cosine^2 or Gaussian) and zero padding to reduce high-frequency ringing in subsequent transforms.

From correlation to spectral density. Under classical linear-response assumptions, the *spectral density* $J(\omega)$ associated with the gap fluctuations is obtained from the cosine transform of $C(t)$ with the usual fluctuation–dissipation prefactor,

$$J(\omega) \;=\; \frac{\omega}{\pi k_{\mathrm{B}} T} \int_0^{\infty} C(t)\,\cos(\omega t)\,dt, \qquad \omega \geq 0, \tag{2.16}$$

which is the relation implemented in many energy-gap workflows and reviews.[59,75] Smoothing (e.g., spline interpolation over a low-frequency cutoff) may be applied to suppress numerical oscillations before subsequent use. Negative lobes in $J(\omega)$ caused by noise are commonly clipped to zero to enforce positivity.

Cumulant lineshape (second order). Assuming Gaussian gap fluctuations (second-order cumulant), a lineshape function $g(t)$ is built from either $C(t)$ (time domain) or $J(\omega)$ (frequency domain). In the time domain,

$$g(t) \;=\; \frac{1}{\hbar^2} \int_0^{t} d\tau \int_0^{\tau} d\tau'\, C(\tau'), \tag{2.17}$$

while the equivalent frequency-domain form uses $J(\omega)$ (with the high-/low-temperature quantum correction via $\coth(\beta\hbar\omega/2)$ if desired).[7] For an absorption transition with reference frequency ω_0 the cumulant expression reads

$$I(\omega) \;\propto\; \mathrm{Re} \int_0^{\infty} dt\; \exp\Big[i\big(\omega - \omega_0\big)t - g(t)\Big], \tag{2.18}$$

optionally followed by a modest homogeneous broadening (e.g., convolution with a narrow Gaussian/Lorentzian) and normalization (area or peak). For emission, an ω^3 factor can be included to reflect the spontaneous-emission rate dependence when comparing *shapes.*[60,76]

Practical recipe.

1. **Input series.** Provide a uniformly sampled time series $\{\Delta E(t_n)\}_{n=0}^{N-1}$ (absorption or emission gap; $t_{n+1} - t_n = \Delta t$). Subtract the mean to form $\delta\epsilon(t_n)$.
2. **Correlation estimate.** Compute $C(t)$ by windowed averaging of $\delta\epsilon(0)\delta\epsilon(t)$ up to a finite lag. Apply gentle damping of the long-t tail and zero padding to stabilize the transform.
3. **Spectral density.** Evaluate $J(\omega)$ from Eq. (2.16) by a cosine transform with the $\omega/(\pi k_{\mathrm{B}} T)$ prefactor. Smooth/clamp to enforce a nonnegative $J(\omega)$.
4. **Lineshape and spectrum.** Build $g(t)$ via Eq. (2.17) (or from $J(\omega)$), then assemble $I(\omega)$ by Eq. (2.18). Apply optional homogeneous broadening and normalize (area or max). For emission-shape comparisons, multiply by ω^3 before normalization.

Scope and caveats. The EGCF-cumulant construction captures *dynamical* broadening from environment-driven energy-gap fluctuations, complementing NEA/E–ZTFC which emphasize inhomogeneous ensembles and vibronic structure.[7,59,75,76] Accuracy relies on (i) near-Gaussian statistics of $\delta\epsilon(t)$ (second-order truncation), (ii) stationarity and adequate sampling (uniform Δt; sufficient correlation-time coverage), and (iii) consistent unit handling in Eqs. (2.16)–(2.17) (the $\hbar^{-2}$ factor is required if $C(t)$ is in energy^2). When strong non-Gaussian effects or pronounced vibronic progressions dominate, higher-order/cumulant or vibronic treatments may be preferable.[7,60]

2.4. Quantum Mechanical Methods

This chapter provides an overview of the quantum-mechanical approaches applied in this dissertation. Density Functional Theory (DFT) is employed as a reference for reaction energies and benchmarking, while Density Functional Tight Binding (DFTB) serves as a computationally efficient method for large data generation and sampling. The time-dependent long-range corrected DFTB (TD-LC-DFTB) method is further introduced for describing excited-state properties relevant to fluorescence studies. Together, these methods enable accurate yet efficient modeling of the systems investigated in this work.

Each subsection begins with a brief note on the relevance of the respective method in the context of this work, followed by a summary of its theoretical foundations.

2.4.1. Ab Initio Methods

Wavefunction-based *ab initio* methods solve the electronic Schrödinger equation using antisymmetrized products of spin-orbitals (Slater determinants) and provide a systematic route to electron correlation at the expense of steeper scaling.[4,79,80] A compact overview of correlation treatment, formal scaling, and size-extensivity is given in Table 1.

Hartree-Fock (HF). Hartree–Fock provides a mean-field reference in which the N-electron wavefunction is approximated by a single Slater determinant that minimizes the expectation value of the Hamiltonian.[79] HF includes exchange exactly but *omits dynamical correlation*; the missing part $E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}$ is recovered by post-HF methods.

Many-body perturbation theory (MPn). Møller-Plesset perturbation theory treats correlation as a perturbation to the HF reference.[81] MP2 (formal $\mathcal{O}(N^5)$) is widely used for thermochemistry and noncovalent interactions; higher orders (MP3/MP4) are costlier and may converge unreliably.[79]

Configuration Interaction (CI). CI expands the wavefunction in a linear combination of excited determinants relative to HF.[80] Truncated CI (e.g., CIS, CISD) is variational but *not size-extensive*; full configuration interaction (FCI) is exact within a given one-electron basis but grows combinatorially and is restricted to very small molecules.

Coupled Cluster (CC). Coupled cluster employs the exponential ansatz $|\Psi\rangle = e^{\hat{T}}|\Phi_0\rangle$ with connected excitations.[4] CCSD is size-extensive and systematically improvable; perturbative triples, CCSD(T), deliver a reliable single-reference benchmark (formal $\mathcal{O}(N^7)$).[82,83] Genuin multi-reference cases require specialized methods (e.g., CASSCF/MRCI), which are beyond this work's scope.

Table 1.: Ab initio methods at a glance. N denotes a generic size measure (e.g., basis functions).

Method	Correlation	Formal scaling	Size extensivity
HF	none (mean field)	$\mathcal{O}(N^4)$	yes
MP2	2nd-order (perturbative)	$\mathcal{O}(N^5)$	yes (linked)
CIS	singles (excited states)	$\mathcal{O}(N^5)$	–
CISD	singles+doubles (variational)	$\mathcal{O}(N^6)$	**no**
CCSD	singles+doubles (connected)	$\mathcal{O}(N^6)$	yes
CCSD(T)	+ perturbative triples	$\mathcal{O}(N^7)$	yes (approx.)
FCI	exact in given basis	combinatorial	yes

Relevance for this work. High-level *ab initio* methods provide gold-standard references for small systems and for benchmarking density functionals.[4] In this dissertation, range-separated hybrid DFT is sufficient for ground-state phosphorus chemistry at the target scale, while DFTB/LC-TD-DFTB enable efficient sampling and excited-state screening. Ab initio methods are summarized for context and positioning; production workflows rely on DFT/DFTB, as detailed in the following sections.

2.4.2. Density Functional Theory

Density Functional Theory (DFT) is one of the most widely applied methods in computational chemistry due to its favorable balance between accuracy and computational efficiency[5,6,84]. Unlike wavefunction-based methods such as Hartree-Fock (HF) or post-Hartree–Fock approaches (e.g., MP2, CCSD(T))[79,84], which treat the many-electron wavefunction explicitly, DFT reformulates the quantum-mechanical many-body problem in terms of the electron density $\rho(\mathbf{r})$[85]. Since the density is a three-dimensional function independent of the number of electrons, this approach significantly reduces the complexity of electronic structure calculations, while retaining predictive accuracy for a broad range of chemical systems. Applications include reaction energy calculations, structure optimizations, and spectroscopic property predictions for molecules and solids.[84,86]

The theoretical foundations of DFT were established in two seminal papers: the Hohenberg-Kohn theorems[5] and the Kohn–Sham equations[6]. These works reformulated quantum mechanics of interacting electrons using universal functionals of the electron density and paved the way for modern density functional approximations.

Relevance in this work. DFT provides the highest level of accuracy used in this dissertation and serves primarily as a reference for benchmarking and for generating correction data in the Δ-learning framework. Its predictive power for reaction energies and optimized structures makes it the baseline for evaluating approximate methods such as DFTB, which are used for large-scale sampling and machine-learning potentials.

Hohenberg-Kohn Theorems and Energy Functional. The Hohenberg-Kohn theorems provide the formal foundation of density functional theory.[5,84,85] The first theorem states[5] that the external potential $V_{\text{ext}}(\mathbf{r})$ (up to an additive constant) and therefore the full Hamiltonian and ground-state wavefunction are unique functionals of the ground-state electron density $\rho(\mathbf{r})$. Consequently, all ground-state observables can be expressed as functionals of $\rho(\mathbf{r})$. The second theorem[5] establishes a variational principle: the ground-state energy E_0 is obtained by minimizing the energy functional

$$E[\rho] = \int V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r})\,d\mathbf{r} + F[\rho], \tag{2.19}$$

where $F[\rho]$ is a universal functional including kinetic energy and electron-electron interactions. The only constraint is the conservation of particle number

$$\int \rho(\mathbf{r})\,d\mathbf{r} = N. \tag{2.20}$$

In the commonly used decomposition, the classical Coulomb interaction is separated explicitly

$$F[\rho] = \frac{1}{2}\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}d\mathbf{r}' + g[\rho], \tag{2.21}$$

where $g[\rho]$ contains all exchange and correlation contributions.[85] For systems with nearly uniform or slowly varying densities, $g[\rho]$ can be approximated by a gradient expansion of the energy density

$$g[\rho] = g_0(\rho) + g_2(\rho)\,(\nabla\rho)^2 + g_4(\rho)\,(\nabla^2\rho)^2 + \dots \tag{2.22}$$

This gradient expansion connects to generalized Thomas-Fermi models and forms the conceptual basis for Generalized Gradient Approximations (GGAs).[86] Gradient expansions alone, however, are insufficient to describe features such as shell structure and Friedel oscillations, which require orbital-dependent or higher-level functionals for accurate treatment.[84]

Kohn–Sham Formalism. Although the Hohenberg-Kohn theorems guarantee the existence of a universal functional, they do not provide its explicit form.[5] Kohn and Sham introduced a practical reformulation by mapping the interacting electron system onto a fictitious system of non-interacting electrons that reproduces the exact ground-state density.[6,84,87] The total energy functional is expressed as

$$E[\rho] = T_s[\rho] + V_{\text{ext}}[\rho] + J[\rho] + E_{\text{xc}}[\rho], \tag{2.23}$$

where $T_s[\rho]$ is the kinetic energy of the non-interacting reference system, $V_{\text{ext}}[\rho]$ is the external potential energy, $J[\rho]$ is the classical Coulomb (Hartree) term, and $E_{\text{xc}}[\rho]$ is the exchange-correlation functional.[28]

Variation of this functional with respect to the single-particle orbitals $\psi_i(\mathbf{r})$ yields the Kohn-Sham equations

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{ext}}(\mathbf{r}) + V_H(\mathbf{r}) + V_{\text{xc}}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}), \tag{2.24}$$

where $V_H(\mathbf{r})$ is the Hartree potential and $V_{\text{xc}}(\mathbf{r}) = \delta E_{\text{xc}}/\delta\rho(\mathbf{r})$ is the exchange-correlation potential. These equations are solved self-consistently, analogous to the Hartree-Fock method but using local potentials.[79]

A key property of the Kohn-Sham construction is that it provides the exact kinetic energy of a non-interacting reference system while isolating all many-body effects into the exchange-correlation term $E_{\text{xc}}[\rho]$. This separation enables systematic improvements of the functional's accuracy by refining approximations to $E_{\text{xc}}[\rho]$ without altering the underlying equations.[28,86] In addition, the Kohn-Sham approach clarifies differences to earlier approximations such as the Slater exchange potential, which underestimates the exact Hartree-Fock exchange by a factor of $2/3$ and motivated the development of modern hybrid functionals.[84]

Historical Context. Early gradient expansion techniques introduced by Hohenberg and Kohn provided one of the first systematic corrections beyond the uniform electron gas approximation.[5,85] These expansions explicitly linked energy corrections to electronic polarizabilities and higher-order response functions and inspired later generalized gradient approximations (GGAs).[88] Although modern GGAs and meta-GGAs now use empirically or semi-empirically parameterized forms,[89] understanding this historical step highlights the conceptual evolution from local density models to today's hybrid and range-separated functionals.

Exchange-Correlation Functionals and Jacob's Ladder. The exchange-correlation functional $E_{\text{xc}}[\rho]$ is the only unknown term in the Kohn-Sham formulation and determines the overall accuracy of density functional theory.[28,87] Various approximations form a hierarchy commonly described as Perdew's "Jacob's Ladder",[86] progressing from the simplest local models to highly accurate hybrid and double-hybrid functionals:

- **Local Density Approximation (LDA):** Relies solely on the local electron density, derived from the uniform electron gas model.[6]
- **Generalized Gradient Approximation (GGA):** Incorporates density gradients to improve upon LDA; typical examples include PBE and BLYP.[88,90,91]
- **Meta-GGA:** Adds higher-order ingredients such as the kinetic energy density (e.g., TPSS).[89]
- **Hybrid Functionals:** Combine exact Hartree-Fock exchange with GGA components; prominent examples are B3LYP and PBE0.[92,93]
- **Range-Separated Hybrids:** Separate short- and long-range exchange interactions, such as ωB97X or CAM-B3LYP.[94,95]
- **Double Hybrids:** Include perturbative correlation corrections (e.g., B2PLYP).[96]

The functional used in this work, ωB97M-V, is a range-separated hybrid with nonlocal correlation (VV10) that balances short-range GGA exchange-correlation with long-range exact exchange.[94,97] Its accuracy for charged and polar systems makes it well suited for the phosphorylation reactions studied in this work.

Basis Sets. Practical DFT calculations rely on the choice of an appropriate basis set to represent molecular orbitals. Gaussian-type orbital (GTO) basis sets are widely employed due to their computational efficiency.[84] Common families include Pople's split-valence sets such as 6-31G,[98] the correlation-consistent Dunning sets (cc-pVXZ) designed for systematic convergence in correlated calculations,[99] and the Karlsruhe def2 family optimized for balanced accuracy across the periodic table.[100] For this work, the minimally augmented ma-def2-TZVP basis set[101] is employed in combination with the ωB97M-V functional, providing diffuse functions that are essential for accurately describing charged phosphorylated species. Energies from quantum-chemical calculations are reported in Hartree (Ha) unless specified; spectral transition energies are given in electronvolts (eV).

ωB97M-V as a reliable reference. Range-separated hybrids with nonlocal correlation such as ωB97M-V combine three ingredients that are crucial for charged and highly polar phosphoryl chemistry: (i) *range separation* mitigates self-interaction and improves barriers and charge-separated states; (ii) the *VV10* nonlocal correlation captures dispersion without ad hoc D terms; and (iii) the meta-GGA ingredient improves thermochemistry and conformer energetics.[94,102] Across broad benchmarks (GMTKN55 and related sets), ωB97M-V ranks among the top tier for main-group thermochemistry, barrier heights, and noncovalent interactions – exactly the error channels that matter for P–O bond reorganization and anionic transition structures in phosphorylation.[102] In this thesis we pair ωB97M-V with a diffuse, triple-ζ basis (ma-def2-TZVP) to stabilize anions/metaphosphate-like motifs and obtain robust reference energetics for Δ-learning and validation.

Despite the broad applicability of DFT, intrinsic limitations remain, including self-interaction errors, underestimated band gaps, and missing long-range dispersion interactions.[28] Dispersion can be accounted for either via semi-empirical DFT-D corrections[96] or nonlocal correlation functionals such as VV10.[97]

In summary, density functional theory provides the most accurate quantum-mechanical reference in this dissertation and forms the basis for benchmarking approximate methods. Its combination of the Kohn–Sham formalism and modern range-separated hybrid functionals such as ωB97M-V enables reliable treatment of charged phosphorylated species. The resulting DFT data serve as a cornerstone for both the Δ-learning correction to DFTB and the validation of excited-state approaches described in the following sections.

2.4.3. Density Functional Tight Binding

Density Functional Tight Binding (DFTB) is an approximate quantum-chemical method derived from Density Functional Theory (DFT).[5,6] By restricting the description to valence electrons and employing a minimal Linear Combination of Atomic Orbitals (LCAO) basis, it reduces computational cost compared to full DFT calculations.[103] The method expands the DFT total energy around a reference density of neutral atoms, resulting in a simplified Hamiltonian and parametrized integrals that retain key electronic structure features while remaining computationally efficient.

Relevance in this work. DFTB is employed as an approximate quantum-mechanical method that balances efficiency and accuracy. It is used for large-scale sampling in molecular dynamics and as the baseline in the Δ-learning scheme, where neural networks recover DFT-level accuracy. This dual role allows explicit treatment of thousands of configurations that would be prohibitive at the DFT level.

Historical development. DFTB originated from early work on density-functional-based tight-binding methods by Porezag and co-workers, who formulated a minimal-basis approach using DFT-derived matrix elements and short-range repulsive potentials.[104] This first-generation scheme (DFTB1) approximated the total energy as the sum of the band-structure energy and a parametrized repulsive term. A major advance was achieved by Elstner et al., who introduced self-consistent charge corrections (SCC-DFTB), resulting in the second-generation formalism (DFTB2).[105] DFTB2 incorporates charge redistribution effects through a second-order Taylor expansion of the energy functional, enabling accurate treatment of polar and charge-transfer systems. Subsequent developments led to the third generation (DFTB3), which added third-order corrections to improve the description of hydrogen bonding, proton transfer, and polarization phenomena.[106,107] These advancements provided the foundation for later extensions, such as long-range corrected (LC) formulations and time-dependent variants relevant to this work.[108,109]

Theoretical formulation. The total energy functional of Density Functional Theory (DFT) can be systematically expanded as a Taylor series around a reference density ρ_0 composed of neutral atomic densities,[103–105]

$$
\begin{aligned}
E[\rho(\mathbf{r})] = E[\rho_0(\mathbf{r})] &+ \int \left.\frac{\delta E[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})}\right|_{\rho_0} \delta\rho(\mathbf{r}) \\
&+ \frac{1}{2}\iint \left.\frac{\delta^2 E[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right|_{\rho_0} \delta\rho(\mathbf{r})\delta\rho(\mathbf{r}') \\
&+ \frac{1}{6}\iiint \left.\frac{\delta^3 E[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right|_{\rho_0} \delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')\delta\rho(\mathbf{r}'') + \dots
\end{aligned}
\tag{2.25}
$$

Truncating this series after the second-order term leads to the second-generation self-consistent-charge formalism (DFTB2), while inclusion of the third-order term yields the DFTB3 approach with improved treatment of hydrogen bonding and polarization.[106,107]

The energy expression for DFTB2 is commonly written as,[103,105]

$$
E = E_0 + \sum_i n_i \langle \varphi_i | H_0 | \varphi_i \rangle + \frac{1}{2}\sum_{a,b} \gamma_{ab}\Delta q_a \Delta q_b + E_{\text{rep}}, \tag{2.26}
$$

where E_0 is the reference energy, H_0 is the Kohn-Sham-like Hamiltonian for the reference density, and Δq_a are Mulliken charge fluctuations on atom a. The Coulomb-like interaction is captured by the γ_{ab} term, and E_{rep} is a short-range repulsive contribution parametrized against full DFT energies.[103,104]

The Mulliken charge q_a on atom a is computed from the density matrix $P_{\mu\nu}$ and the overlap matrix $S_{\mu\nu}$ as,[105,110]

$$
q_a = \frac{1}{2}\sum_{\mu\in a}\sum_{\nu}(P_{\mu\nu} + P_{\nu\mu})S_{\mu\nu}. \tag{2.27}
$$

These charges are iteratively updated until self-consistency is achieved, which is the hallmark of the SCC-DFTB approach.

The density matrix $P_{\mu\nu}$ is constructed from the occupied molecular orbitals φ_i and their occupation numbers n_i according to

$$
P_{\mu\nu} = \sum_i n_i \, c_{\mu i} \, c_{\nu i}, \tag{2.28}
$$

where $c_{\mu i}$ are the molecular orbital coefficients in the atomic orbital basis. The overlap matrix $S_{\mu\nu}$ is defined as

$$
S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle, \tag{2.29}
$$

where ϕ_μ are the atomic basis functions used in the minimal Linear Combination of Atomic Orbitals (LCAO) representation employed by DFTB.

Parametrization and extensions. All two-center Hamiltonian and overlap integrals in DFTB are precomputed using full DFT calculations and stored in Slater-Koster tables.[104,105] These integrals, together with repulsive energy terms, are parameterized to reproduce DFT reference data for specific element combinations. Widely used parameter sets include `mio` (minimal organics),[105] `3ob` (third-order organic-biochemical),[106] and `ob2`,[108] the latter optimized for organic and biological systems including phosphorus-containing compounds. DFTB3 incorporates additional parameters for the third-order expansion term, improving the description of hydrogen bonding, proton transfer, and highly polarizable environments.[106,107] Beyond the core formalism, empirical dispersion corrections (e.g., DFTB3-D3)[111] and long-range corrected (LC) functionals[109] extend the applicability of DFTB to weakly bound systems, charge-transfer excitations, and excited-state calculations.

Advantages and limitations. The primary advantage of DFTB lies in its computational efficiency: it achieves orders-of-magnitude speedups compared to full DFT while retaining key quantum-chemical features such as bond formation, polarization, and charge transfer.[103] This efficiency enables molecular dynamics simulations on nanosecond timescales for systems containing hundreds to thousands of atoms, which would be prohibitive with conventional DFT. However, the method's accuracy strongly depends on the quality and transferability of its parameter sets. Limitations arise for systems outside the training domain of the parametrization, such as strongly correlated transition-metal complexes, high-spin states, or systems with significant multi-reference character. Furthermore, dispersion interactions and long-range charge-transfer states require either empirical corrections (e.g., D3) or explicitly range-separated functionals to achieve quantitative accuracy.[109,111]

Phosphorus-specific notes. For phosphoryl chemistry, DFTB3 is a practical baseline *if* one employs parameter sets that explicitly address P-bond systematics. The 3OB line and its P/S extensions introduce targeted corrections that rebalance P=O vs. P–O bonding and reduce overpolarization around hypervalent phosphorus;[106–108] more recently, P–N energetics relevant to phosphoramidates were tightened, improving the description of PN formation/cleavage windows.[112] These fixes materially lower the baseline error for phosphoryl transfer. Nevertheless, bare DFTB still struggles in strongly charge-separated/metaphosphate-like regions and along pentacoordinate transition structures, where long-range polarization and multi-reference character loom large.[10] Accordingly, in this thesis DFTB3 (with the above P/O and P–N updates) is used as an *efficient sampling engine and fast QM baseline* rather than as a final predictor: Δ-QM/MM models trained to hybrid-DFT reference data correct the remaining P=O/P–O/P–N imbalances at near-DFTB cost while preserving MD stability and the existing QM/MM plumbing.

In summary, DFTB provides a computationally efficient approximation to DFT that retains essential quantum-mechanical features such as bond formation and charge transfer. Its speed enables extensive sampling for molecular dynamics and serves as a practical baseline for the Δ-learning correction, which recovers near-DFT accuracy for the phosphorylation reactions studied in this work.

2.4.4. Excited States TD-LC-DFTB

Time-dependent long-range corrected density functional tight binding (TD-LC-DFTB) combines the efficiency of the DFTB framework (see previous subsection) with extensions that enable the description of excited electronic states and charge-transfer processes. The ground-state formalism, including the Taylor expansion of the total energy and the definition of Mulliken charges, follows the formulation introduced in Section 2.4.3 (see Eq. 2.25 and Eq. 2.27). In this section, only the modifications required for long-range correction (LC) and time-dependent (TD) response are discussed.

Relevance in this work TD-LC-DFTB extends the DFTB framework to excited states and is used in this dissertation for predicting fluorescence spectra and excited-state dynamics. The inclusion of long-range exchange correction is crucial for describing charge-transfer excitations, enabling accurate yet efficient simulations of solvatochromic dyes over extended timescales.

Long-range correction. Standard DFTB inherits the local exchange-correlation description of its parent DFT functional, which often leads to failures for charge-transfer excitations and Rydberg states. Long-range corrected DFTB (LC-DFTB) addresses this by partitioning the exchange into short- and long-range components using a Yukawa-type separation.[109] The short-range part is treated with the chosen local exchange-correlation functional, while the long-range part is replaced by exact Hartree–Fock exchange. The range-separation parameter ω controls the partitioning and is typically chosen as $\omega = 0.3$ for organic and biomolecular systems.[108] This correction significantly improves the description of long-range excitations and charge-separated states, which are relevant in both photophysics and bioinorganic catalysis.

Linear-response formalism. Excited-state properties are obtained via a linear-response formalism analogous to time-dependent density functional theory (TD-DFT)[113]. In TD-DFTB, this leads to a generalized eigenvalue problem of Casida form

$$\begin{pmatrix} A & K \\ K & A \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \Omega \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}, \tag{2.30}$$

where X and Y are excitation and de-excitation amplitudes, respectively, and Ω contains the excitation energies. The matrices A and K are constructed from orbital energy differences and Coulomb-like coupling terms

$$A_{ij,kl} = (\epsilon_j - \epsilon_i)\delta_{ik}\delta_{jl} + K_{ij,kl} \tag{2.31a}$$

$$K_{ij,kl} = 2\sum_{a,b} q_a^{ij} q_b^{kl} \left[\gamma_{ab} + m_{ab}\right], \tag{2.31b}$$

where q_a^{ij} are Mulliken transition charges defined via the density and overlap matrices (see Eq. 2.28).[49] The γ_{ab} terms represent Coulomb interactions (as in ground-state DFTB), and m_{ab} accounts for spin contributions; spin-flip terms are neglected in this work, as triplet-state dynamics are outside the timescale of interest.

Oscillator strengths. Transition dipole moments and oscillator strengths f are obtained from the eigenvectors of Eq. 2.30, enabling direct computation of absorption spectra. The oscillator strength of an excitation with energy Ω is given by

$$f = \frac{2}{3}\Omega \sum_{\alpha=x,y,z} \left| \sum_{ij,a} R_{\alpha,a} q_a^{ij} \sqrt{\frac{\omega_{ij}}{\Omega}} (F_{ij\uparrow} + F_{ij\downarrow}) \right|^2, \tag{2.32}$$

where $R_{\alpha,a}$ are Cartesian components of the atomic positions and ω_{ij} are orbital energy differences.[49,114]

Application. Compared to conventional TD-DFTB, the LC extension is particularly important for capturing charge-transfer and Rydberg excitations, which are otherwise underestimated.[109] While higher-order methods like DFTB3[106] enhance ground-state energetics, the LC correction provides the decisive improvement for excited-state calculations. Leveraging this efficiency-accuracy trade-off, TD-LC-DFTB is used to compute excitation energies for thousands of geometries along molecular dynamics trajectories, providing both converged spectral profiles and training data for the machine-learning spectral models developed in later chapters.

Overall, TD-LC-DFTB bridges the gap between efficient ground-state methods and the accurate treatment of excited-state properties. The combination of long-range exchange correction and linear-response formalism enables the simulation of charge-transfer excitations and fluorescence spectra over extended timescales, making it particularly well suited for the solvatochromic dyes analyzed in this dissertation.

2.5. Molecular Dynamics

Molecular Dynamics (MD) [115–117] simulations are based on classical mechanics and describe the motion of nuclei under the Born–Oppenheimer approximation. All particles are treated in their electronic ground state, and the potential energy $V(\{\mathbf{r}\})$ is expressed as a function of atomic coordinates only.

Relevance in this work In this work, molecular dynamics is primarily applied to describe the non-reactive environment in hybrid QM/MM simulations, most notably the solvent (e.g., water). Since the electronic structure of the solvent does not change during the reactions studied here, its electrons can be assumed to remain in the ground state. A classical force-field treatment is therefore sufficient and avoids the computational cost of a full quantum-mechanical description.[25,117] Using MD for the solvent allows efficient sampling of solvent configurations and dynamics, while the reactive region is treated at the quantum-mechanical level.[1] This hybrid approach combines accuracy and computational efficiency and is widely used for reactions in condensed-phase systems.

2.5.1. Classical Force Field

A classical force field defines the potential energy surface used to calculate the forces acting on the atoms.[115] The total potential energy $V(\{\mathbf{r}\})$ is separated into bonded and non-bonded contributions

$$V = V_{\text{bonded}} + V_{\text{nonbonded}}. \tag{2.33}$$

Bonded interactions include three types of terms: bond stretching, angle bending, and torsional dihedrals. They are typically described by harmonic potentials.[115]

$$V_{ij}^{\text{b}} = \frac{1}{2}k_{ij}^{\text{b}}\left(r_{ij} - r_{ij}^{0}\right)^2 \tag{2.34a}$$

$$V_{ijk}^{\text{a}} = \frac{1}{2}k_{ijk}^{\text{a}}\left(\theta_{ijk} - \theta_{ijk}^{0}\right)^2 \tag{2.34b}$$

$$V_{ijkl}^{\text{d}} = \frac{1}{2}k_{ijkl}^{\text{d}}\left(\xi_{ijkl} - \xi_{ijkl}^{0}\right)^2 . \tag{2.34c}$$

Non-bonded interactions consist of electrostatics and van der Waals contributions. Electrostatics are modeled with a Coulomb potential

$$V_{ij}^{\text{C}} = \frac{q_i q_j}{\epsilon_r r_{ij}}, \tag{2.35}$$

while van der Waals interactions are usually represented by a Lennard–Jones potential

$$V_{ij}^{\text{LJ}} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]. \tag{2.36}$$

For periodic systems, long-range electrostatics are evaluated using an Ewald-based method, most commonly Particle Mesh Ewald (PME).[118,119] The force field parameters used in this work originate from the GROMOS force field family,[120] which is implemented in GROMACS and well-suited for biomolecular systems.[121]

2.5.2. System Propagation

The time evolution of the system is obtained by integrating classical mechanics' equations of motion

$$\mathbf{F}_i = m_i \mathbf{a}_i = -\frac{\partial V}{\partial \mathbf{r}_i} \tag{2.37}$$

following standard MD formalisms.[115,118] Positions and velocities are updated in discrete time steps Δt. In this work a time step of $\Delta t = 1\,\mathrm{fs}$ or $\Delta t = 0.5\,\mathrm{fs}$ is used, which is small enough to resolve high-frequency bond vibrations. The Leap–Frog algorithm,[115,118,122] is applied for the integration of the equations of motion

$$\mathbf{v}\left(t + \frac{1}{2}\Delta t\right) = \mathbf{v}\left(t - \frac{1}{2}\Delta t\right) + \frac{\mathbf{F}(t)}{m}\Delta t \tag{2.38a}$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}\left(t + \frac{1}{2}\Delta t\right)\Delta t. \tag{2.38b}$$

Initial velocities are generated according to a Maxwell–Boltzmann distribution at the reference temperature.[117] The probability density for the speed v of a particle with mass m at temperature T is given by

$$f(v) = 4\pi \left(\frac{m}{2\pi k_\mathrm{B} T}\right)^{3/2} v^2 \exp\left(-\frac{mv^2}{2k_\mathrm{B} T}\right), \tag{2.39}$$

where k_B is the Boltzmann constant. This distribution ensures that the initial velocities reflect the correct thermal energy corresponding to the target temperature.

The time step is constrained by the fastest motions in the system; in the present simulations, bond length constraints (LINCS algorithm) [123] are employed to allow a stable integration at femtosecond resolution.

2.5.3. Temperature and Pressure

Temperature and pressure are evaluated from the instantaneous kinetic energy and the virial of the system.[25,124] The temperature follows directly from the equipartition theorem

$$T = \frac{2}{f k_\mathrm{B}} E_\mathrm{kin} = \frac{1}{f k_\mathrm{B}} \sum_i m_i v_i^2, \tag{2.40}$$

where f is the number of degrees of freedom.

The pressure is obtained using the virial expression

$$P = \frac{N k_\mathrm{B} T}{V} + \frac{1}{3V} \sum_{i<j} \mathbf{r}_{ij} \cdot \mathbf{F}_{ij}, \tag{2.41}$$

where the second term accounts for the contribution of intermolecular forces to the mechanical pressure.

Thermostats. To maintain reference conditions during a simulation, coupling algorithms are employed. Thermostats control the kinetic energy of the system to generate a desired ensemble, most commonly the canonical (NVT) ensemble. A simple and widely used scheme is the Berendsen thermostat,[125] which rescales velocities smoothly towards the reference temperature. The scaling factor λ applied to all velocities is derived from the deviation of the current temperature T from the reference value T_{ref} and follows

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_{\text{ref}}}{T} - 1\right)}, \tag{2.42}$$

where τ is the coupling constant controlling how quickly the system approaches T_{ref}. While numerically stable and useful for equilibration, this method does not yield correct canonical fluctuations.

For production runs, the Nosé–Hoover thermostat,[126,127] is typically preferred, which extends the system by an additional degree of freedom acting as a heat reservoir. The corresponding equations of motion introduce a friction term proportional to $\gamma = \dot{s}/s$, where s is the extended variable of the thermostat

$$\ddot{\mathbf{r}}_i = \frac{\mathbf{F}_i}{m_i} - \gamma \dot{\mathbf{r}}_i. \tag{2.43}$$

This approach generates a proper canonical ensemble and preserves realistic temperature fluctuations.

Barostats Analogous coupling schemes exist for pressure control in the isothermal-isobaric (NPT) ensemble. The Berendsen barostat rescales the box dimensions to drive the instantaneous pressure P towards a reference value P_{ref},

$$\mu = 1 - \frac{\beta \Delta t (P_{\text{ref}} - P)}{3\tau}, \tag{2.44}$$

where β is the isothermal compressibility. Although efficient for equilibration, this method suppresses volume fluctuations. The Parrinello-Rahman barostat,[128] instead introduces box vectors as dynamic variables, allowing anisotropic fluctuations and correct sampling of the NPT ensemble.

2.6. Quantum Mechanics / Molecular Mechanics

The QM/MD approach combines a quantum-mechanical description of a reactive subsystem with a classical molecular dynamics treatment of the surrounding environment.[1,27,129,130] This hybrid strategy allows for the accurate modeling of chemical reactivity within the QM region while maintaining computational efficiency by treating the solvent or matrix at the classical level.[131,132]

Relevance in this work. The combined QM/MD framework is employed throughout this work to describe chemically active molecules at the quantum-mechanical level while capturing their interaction with a surrounding environment treated classically.[1,130] In all systems studied, the QM region is fully contained within one or several molecules, and no covalent bonds cross the QM-MM boundary.[1] As a result, no link atoms are required, and the coupling between the regions arises solely from non-bonded interactions. This setup enables accurate modeling of reactivity and excited-state properties in solution or condensed phases at tractable computational cost.[131]

Energy decomposition. The total energy of the hybrid system can be written as

$$E = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QMMM}}. \tag{2.45}$$

Here, E_{QM} and E_{MM} are the energies of the isolated QM and MM subsystems, respectively, and E_{QMMM} describes their interaction[130,131]

$$E_{\mathrm{QMMM}} = -\sum_{m\in MM} q_m \int \frac{n(\mathbf{r})}{|\mathbf{r}-\mathbf{r}_m|}\,\mathrm{d}\mathbf{r} + \sum_{m\in MM}\sum_{n\in QM} \frac{q_m Z_n}{|\mathbf{r}_n-\mathbf{r}_m|} + E_{\mathrm{LJ}}. \tag{2.46}$$

q_m are partial charges of MM atoms at positions $\mathbf{r}_m$, $n(\mathbf{r})$ is the QM electron density, Z_n are nuclear charges of QM atoms, and E_{LJ} is the non-electrostatic Lennard–Jones term.[131]

$$E_{\mathrm{LJ}} = \sum_{m\in MM}\sum_{n\in QM} 4\epsilon_{mn}\left[\left(\frac{\sigma_{mn}}{|\mathbf{r}_n-\mathbf{r}_m|}\right)^{12} - \left(\frac{\sigma_{mn}}{|\mathbf{r}_n-\mathbf{r}_m|}\right)^{6}\right], \tag{2.47}$$

where ϵ_{mn} and σ_{mn} are Lennard-Jones parameters.

Mechanical vs. electrostatic embedding. Two main strategies exist to couple the QM and MM regions in hybrid simulations.[1,27,132] In **mechanical embedding**, the QM subsystem is solved in vacuum and the interaction with the MM region is added only at the classical level, neglecting polarization of the QM density by the environment. In **electrostatic embedding**, the MM point charges are included directly in the QM Hamiltonian as an external potential, allowing the QM electron density to polarize in response to the solvent field.[132] The following sections present the Hamiltonian formulations for each embedding scheme.

Mechanical embedding. In mechanical embedding, the QM calculation is carried out in vacuum and the interaction with the MM region is added only at the classical level. The Hamiltonian can be written as

$$\hat{H}_{\mathrm{ME}} = \hat{H}^{0}_{\mathrm{QM}} + \hat{H}_{\mathrm{MM}} + \sum_{i\in \mathrm{QM}}\sum_{j\in \mathrm{MM}} \frac{q_i q_j}{|\mathbf{R}_i-\mathbf{R}_j|} + V^{\mathrm{LJ}}_{\mathrm{QMMM}}, \tag{2.48}$$

where $\hat{H}^0_{\mathrm{QM}}$ is the gas-phase QM Hamiltonian, $\hat{H}_{\mathrm{MM}}$ is the classical Hamiltonian of the MM subsystem, the third term is the Coulomb interaction between QM and MM partial charges q_i and q_j, and the fourth term $V^{\mathrm{LJ}}_{\mathrm{QMMM}}$ is the Lennard-Jones non-electrostatic interaction.[130]

The total energy in this scheme is then given by

$$E^{\mathrm{ME}}_{\mathrm{tot}} = \left\langle \Psi_{\mathrm{QM}} \middle| \hat{H}^0_{\mathrm{QM}} \middle| \Psi_{\mathrm{QM}} \right\rangle + E_{\mathrm{MM}} + E_{\mathrm{classical}}(\mathrm{QM}, \mathrm{MM}), \tag{2.49}$$

where the classical interaction energy $E_{\mathrm{classical}}(\mathrm{QM}, \mathrm{MM})$ includes the Coulomb and Lennard-Jones terms shown above but is not part of the electronic structure calculation of the QM region.

Electrostatic embedding. In electrostatic embedding, the MM environment enters directly into the QM Hamiltonian as an external electrostatic potential. The Hamiltonian becomes

$$\hat{H}_{\mathrm{EE}} = \hat{H}^0_{\mathrm{QM}} + \hat{H}_{\mathrm{MM}} - \sum_{a=1}^{N_e} \sum_{j\in\mathrm{MM}} \frac{q_j}{|\mathbf{r}_a - \mathbf{R}_j|} + \sum_{i\in\mathrm{QM}} \sum_{j\in\mathrm{MM}} \frac{Z_i q_j}{|\mathbf{R}_i - \mathbf{R}_j|} + V^{\mathrm{LJ}}_{\mathrm{QMMM}}, \tag{2.50}$$

where $\hat{H}^0_{\mathrm{QM}}$ is the gas-phase QM Hamiltonian, $\hat{H}_{\mathrm{MM}}$ is the classical Hamiltonian of the MM subsystem, the third term is the Coulomb interaction of the MM point charges q_j with all N_e of the QM electrons (included self-consistently), the fourth term is the interaction of the MM point charges q_j with the QM nuclei Z_i, and $V^{\mathrm{LJ}}_{\mathrm{QMMM}}$ is the non-electrostatic Lennard-Jones interaction.[132]

The total energy in this scheme is then

$$E^{\mathrm{EE}}_{\mathrm{tot}} = \left\langle \Psi_{\mathrm{QM}} \middle| \hat{H}^0_{\mathrm{QM}} - \sum_{a=1}^{N_e} \sum_{j\in\mathrm{MM}} \frac{q_j}{|\mathbf{r}_a - \mathbf{R}_j|} + \sum_{i\in\mathrm{QM}} \sum_{j\in\mathrm{MM}} \frac{Z_i q_j}{|\mathbf{R}_i - \mathbf{R}_j|} \middle| \Psi_{\mathrm{QM}} \right\rangle + E_{\mathrm{MM}} + E_{\mathrm{LJ}}(QM, MM), \tag{2.51}$$

where the Coulomb potential of the MM charges is included directly in the electronic structure calculation, leading to polarization of the QM density by the solvent field.

Meaning of the MM charges q_j. The charges q_j are not bare nuclear charges nor explicit electronic charges of the MM region. Instead, they are *effective point charges* defined by the classical force field (e.g., TIP3P for water) and are parametrized to reproduce the electrostatics of the MM molecules.[133] These charges approximate the combined effect of nuclei and electrons in the MM subsystem and act as fixed external sources in the QM calculation. Their positions $\mathbf{R}_j$ evolve dynamically with the classical MD trajectory, so the external potential changes along the simulation.

Additive vs. subtractive bookkeeping. Two equivalent energy bookkeepings are common in QM/MM. We adopt the *subtractive* (ONIOM-like) convention to avoid double counting of the QM region: the total energy is assembled as

$$E_{\mathrm{tot}} = E_{\mathrm{MM}}(\mathrm{QM} \cup \mathrm{MM}) + E_{\mathrm{QM}}(\mathrm{QM\,in\,field}) - E_{\mathrm{MM}}(\mathrm{QM})\,,$$

with MM point charges entering the QM Hamiltonian (electrostatic embedding).

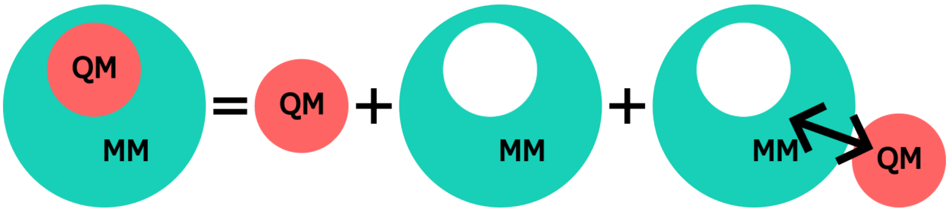


Figure 6.: Subtractive (ONIOM-like) QM/MM bookkeeping: $E_{\mathrm{tot}} = E_{\mathrm{MM}}(\mathrm{QM} \cup \mathrm{MM}) + E_{\mathrm{QM}}(\mathrm{QM\,in\,field}) - E_{\mathrm{MM}}(\mathrm{QM})$. This avoids double counting of the QM region; MM point charges act as an external field in the QM Hamiltonian.

Force evaluation in QM/MM dynamics. In the Born-Oppenheimer approximation, the electronic wavefunction is solved for fixed nuclear coordinates, and the resulting potential energy surface is used to propagate the nuclei using classical mechanics.[1,130] The quantities referred to as "forces" in this context are therefore the negative gradients of the total energy with respect to the nuclear positions. For the QM atoms, the energy gradient consists of two contributions: the gradient of the QM Hamiltonian (which may or may not include the external MM potential, depending on the embedding scheme) and the gradient of the classical non-bonded QM–MM interaction terms. The quantum-mechanical contribution is given by

$$\tilde{\mathbf{F}}_{\mathrm{QM}} = -\nabla_{\mathrm{QM}} \langle \Psi_{\mathrm{QM}} | \hat{H}_{\mathrm{QM}} | \Psi_{\mathrm{QM}} \rangle. \tag{2.52}$$

The additional classical contribution originates from electrostatic and van-der-Waals interactions between QM and MM atoms,

$$\mathbf{F}'_{\mathrm{QM}} = -\nabla_{\mathrm{QM}} \left(E_{\mathrm{Coulomb}}(QM, MM) + E_{\mathrm{LJ}}(QM, MM) \right). \tag{2.53}$$

The total force acting on the QM atoms is then

$$\mathbf{F}_{\mathrm{QM,tot}} = \tilde{\mathbf{F}}_{\mathrm{QM}} + \mathbf{F}'_{\mathrm{QM}}. \tag{2.54}$$

Analogously, the forces on the MM atoms are obtained by taking the gradient with respect to their coordinates. These forces enter the classical equations of motion used for the molecular-dynamics propagation described in Section 2.5.

The QM/MD framework provides quantum accuracy for the reactive molecular subsystem and classical efficiency for the surrounding environment. The explicit inclusion of environmental charges in the QM Hamiltonian (electrostatic embedding) ensures that polarization effects are captured, which is critical for modeling fluorescence shifts, charge-transfer excitations, and chemical reactions investigated in this work.[130–132]

2.7. Minimum Free Energy Path with Nudged Elastic Band

For a general chemical reaction at finite temperature, the central object is the minimum free energy path (MFEP) on a precomputed free energy surface (FES) defined over suitable collective variables (CVs). The MFEP connects basins through the lowest pass and captures the thermodynamically preferred sequence across the transition-state region.[134,135] In this work, the FES is treated as fixed; the method optimizes a path on this surface rather than sampling or updating the surface.[136] Chain-of-states relaxations such as nudged elastic band (NEB) and its free-energy variants (FENEB) are well suited to this task, and climbing-image modifications refine the highest image near the saddle.[137,138] Related string formulations provide an alternative update with explicit reparametrization.[134,135]

Scope. On a fixed, precomputed FES, this work employs a standard chain-of-states relaxation (FENEB). For robustness in early steps, a two-phase "Waterfall" initializer biases progress first in one CV and then in the other, and a weak form-preserving regularizer maintains this staging until the band approaches the pass. While similar heuristics may exist under different names, this specific pairing is not commonly reported in FENEB applications.[136,137,139] The term "Waterfall" is used descriptively; in scientific software and visualization the name "waterfall" commonly denotes staged or stepwise representations.[140]

Core idea. *The band itself is the object being optimized: a discrete chain anchored at the two endpoints that is iteratively deformed until the band coincides with an MFEP on the fixed FES.* The deformation uses two complementary effects: (i) perpendicular free-energy forces push each inner image sideways downhill to eliminate any gradient component normal to the current band; (ii) gentle springs act along the band to keep neighboring images evenly spaced, avoiding clustering in basins and gaps near the pass.[137] Initialization matters because the early shape decides which channel the band explores; a naive straight line can cut ridges, whereas a chemically informed staging can place the initial band nearer a relevant valley before the surface forces refine the route.[135] In this work, a two-phase "Waterfall" initializer is used and, optionally, a weak form-preserving regularizer is added so that the staged character is not immediately lost; to our knowledge, this specific combination is not standard in FENEB reports.[136] Because the chain remains continuously connected between basins, it must cross the inter-basin divide; repeated sideways downhill "denting" lowers the maximum height along the chain until no further perpendicular descent is possible, which occurs at a pass.[134,138]

Algorithm. Inputs: fixed FES F on a rectangular CV box $\Omega = [x_1^{\min}, x_1^{\max}] \cdot [x_2^{\min}, x_2^{\max}]$; endpoints $\mathbf{x}_0, \mathbf{x}_{N-1}$ in different basins; number of images N; step size $\alpha > 0$; spring constant $k > 0$; optional Waterfall reference $\{\mathbf{x}_i^{\mathrm{WF}}\}$ and weight $\lambda \geq 0$.

1. **Initialization.** Build an initial band $\{\mathbf{x}_i\}_{i=0}^{N-1}$ between $\mathbf{x}_0$ and $\mathbf{x}_{N-1}$ (linear or Waterfall), then equalize spacing by arc length. Keep endpoints fixed.

2. **Per iteration: tangents.** For each inner image $i = 1, \dots, N-2$, compute the central-difference unit tangent $\boldsymbol{\tau}_i$ and the parallel/perpendicular projectors $\mathbf{P}_i^{||}$ and $\mathbf{P}_i^{\perp}$ (definitions in Eq. (2.56)).
3. **Per iteration: forces from FES and springs.** Evaluate the free-energy gradient $g_i = \nabla F(\mathbf{x}_i)$. Form the perpendicular (path-shaping) force $\mathbf{F}_i^{\perp} = -\mathbf{P}_i^{\perp} g_i$ and the along-band spring force $\mathbf{F}_i^{||} = \mathbf{P}_i^{||}\,\mathbf{F}_i^{\text{spring}}$ with $\mathbf{F}_i^{\text{spring}} = k[(\mathbf{x}_{i+1} - \mathbf{x}_i) + (\mathbf{x}_{i-1} - \mathbf{x}_i)]$ (Eqs. (2.57)–(2.58)).
4. **Optional: Waterfall regularizer.** Add $\mathbf{F}_i^{\text{WF}} = \lambda(\mathbf{x}_i^{\text{WF}} - \mathbf{x}_i)$ to stably preserve the staged form in early steps (Eq. (2.60)).
5. **Update all inner images synchronously.**

$$\mathbf{x}_i \longleftarrow \Pi_\Omega\Big(\mathbf{x}_i + \alpha\,[\mathbf{F}_i^{\perp} + \mathbf{F}_i^{||} + \mathbf{F}_i^{\text{WF}}]\Big), \qquad i = 1, \dots, N-2, \tag{2.55}$$

with the componentwise projection Π_Ω that clips to the CV box (Eq. (2.59)).

6. **Equalize spacing (optional each macro-step).** Reparametrize by equal arc length to avoid clustering and holes along the band.
7. **Convergence test.** Stop when $\max_{i=1,\dots,N-2} \|\mathbf{F}_i^{\perp} + \mathbf{F}_i^{||}\| < \varepsilon$ for a small tolerance ε. The highest-energy image then marks the pass region; an optional climbing-image refinement can be applied late to sharpen the saddle.[138]

Definitions used by the algorithm. For inner images $i = 1, \dots, N-2$, the central tangent and projectors are

$$\boldsymbol{\tau}_i = \frac{\mathbf{x}_{i+1} - \mathbf{x}_{i-1}}{\|\mathbf{x}_{i+1} - \mathbf{x}_{i-1}\|}, \qquad \mathbf{P}_i^{||} = \boldsymbol{\tau}_i\boldsymbol{\tau}_i^{\top}, \qquad \mathbf{P}_i^{\perp} = \mathbf{I} - \boldsymbol{\tau}_i\boldsymbol{\tau}_i^{\top}. \tag{2.56}$$

The spring and perpendicular forces are

$$\mathbf{F}_i^{\text{spring}} = k\big[(\mathbf{x}_{i+1} - \mathbf{x}_i) + (\mathbf{x}_{i-1} - \mathbf{x}_i)\big], \tag{2.57}$$

$$\mathbf{F}_i^{||} = \mathbf{P}_i^{||}\mathbf{F}_i^{\text{spring}}, \qquad \mathbf{F}_i^{\perp} = -\mathbf{P}_i^{\perp}\nabla F(\mathbf{x}_i). \tag{2.58}$$

The CV-box projection used for clipping is

$$\Pi_\Omega(x_1, x_2) = \big(\min\{\max\{x_1, x_1^{\min}\}, x_1^{\max}\},\ \min\{\max\{x_2, x_2^{\min}\}, x_2^{\max}\}\big). \tag{2.59}$$

If a Waterfall reference path is employed, the regularizer is

$$U_{\text{WF}}(\{\mathbf{x}_i\}) = \frac{\lambda}{2}\sum_{i=1}^{N-2} \|\mathbf{x}_i - \mathbf{x}_i^{\text{WF}}\|^2, \qquad \mathbf{F}_i^{\text{WF}} = \lambda(\mathbf{x}_i^{\text{WF}} - \mathbf{x}_i). \tag{2.60}$$

MFEP condition and why NEB enforces it. Along an MFEP, the free-energy gradient has no component perpendicular to the path,

$$\mathbf{P}^{\perp}\nabla F(\mathbf{x}(s)) = \mathbf{0} \quad \text{for all } s \in [0,1]. \tag{2.61}$$

Intuitively, if a perpendicular component remained at some point, one could “dent” the path sideways to lower its local height; repeating this would reduce the path’s maximum height until no such perpendicular descent exists.[134] NEB implements exactly this logic: the update keeps only the perpendicular downhill component from the surface to shape the path and only the parallel spring component to redistribute images.[137] At the saddle itself, $\nabla F = \mathbf{0}$, which trivially satisfies Eq. (2.61); climbing-image NEB flips the parallel component of the highest image so it ascends to the saddle more sharply.[138]

Gradients on a fixed FES: simple smoothing and differences. The update requires a stable gradient ∇F at image locations. If tabulated F is noisy, apply a mild Gaussian local average on a regular grid and then centered finite differences:

$$\widetilde{F}[m,n] = \sum_{i=-r}^{r}\sum_{j=-r}^{r} w_\sigma(i,j)\,\widehat{F}[m-i,n-j], \qquad \sum_{i,j} w_\sigma(i,j) = 1, \tag{2.62}$$

$$\nabla\widetilde{F}(x_1,x_2) = \left(\frac{\widetilde{F}(x_1+h,x_2)-\widetilde{F}(x_1-h,x_2)}{2h},\ \frac{\widetilde{F}(x_1,x_2+h)-\widetilde{F}(x_1,x_2-h)}{2h}\right). \tag{2.63}$$

If the FES is already smooth, set $\sigma = 0$ and skip Eq. (2.62).[136]

Initializations: linear vs. Waterfall. Let the band endpoints be $\mathbf{x}_0 = (a_1, a_2)$ (reactant) and $\mathbf{x}_{N-1} = (b_1, b_2)$ (product), where a_j and b_j are the coordinates of CV_j (with $\mathrm{CV}_1 \equiv x_1$, $\mathrm{CV}_2 \equiv x_2$) at the endpoints. Denote $\mathbf{a} = (a_1, a_2)$, $\mathbf{b} = (b_1, b_2)$, and the spans $\Delta_1 = b_1 - a_1$, $\Delta_2 = b_2 - a_2$. A simple reference start path is the straight line between endpoints,

$$\mathbf{x}^{\mathrm{lin}}(s) = \mathbf{a} + s(\mathbf{b}-\mathbf{a}), \quad s \in [0,1]. \tag{2.64}$$

This changes both CVs proportionally for all s and can cut ridges; it is *not* used in this work.[135]

Waterfall (smooth C^1 variant). Let $s_0 \in (0,1)$ be the phase split and $f_1, f_2 \in (0,1)$ the fractional progresses reached at s_0 for CV1 and CV2, respectively. Define the segment coordinates $t_1(s) = s/s_0$ and $t_2(s) = (s-s_0)/(1-s_0)$, and choose a C^1 “smoothstep” ramp $\phi(t)$ that satisfies $\phi(0) = 0$, $\phi(1) = 1$, and $\phi'(0) = \phi'(1) = 0$ so that the path has flat ends and continuous first derivatives at the phase split; a convenient closed-form choice is the cubic Hermite $\phi(t) = 3t^2 - 2t^3$. Then the staged path is given by Eq. (2.65)

$$\mathbf{x}^{\mathrm{WF}}(s) = \begin{cases} \begin{pmatrix} a_1 + f_1\,\phi\big(t_1(s)\big)\,\Delta_1 \\ a_2 + f_2\,\phi\big(t_1(s)\big)\,\Delta_2 \end{pmatrix}, & 0 \le s \le s_0, \\ \begin{pmatrix} a_1 + \big[f_1 + (1-f_1)\,\phi\big(t_2(s)\big)\big]\,\Delta_1 \\ a_2 + \big[f_2 + (1-f_2)\,\phi\big(t_2(s)\big)\big]\,\Delta_2 \end{pmatrix}, & s_0 < s \le 1\ . \end{cases} \tag{2.65}$$

Waterfall (as implemented, piecewise linear). For N images set $s_i = i/(N-1)$ and choose $s_0 \in (0,1)$. The phase-local coordinates are $t_1(s_i) = s_i/s_0$ for $s_i < s_0$ and $t_2(s_i) = (s_i - s_0)/(1 - s_0)$ for $s_i \geq s_0$. Using early-phase fractions $f_1, f_2 \in (0,1)$ and late-phase baselines $m_1, m_2 \in [0,1)$, the initialized images are given by Eq. (2.66)

$$\mathbf{x}^{\mathrm{WF}}(s_i) = \begin{cases} \begin{pmatrix} a_1 + f_1\, t_1(s_i)\, \Delta_1 \\ a_2 + f_2\, t_1(s_i)\, \Delta_2 \end{pmatrix}, & 0 \leq s_i \leq s_0, \\ \begin{pmatrix} a_1 + \left[m_1 + (1-m_1) t_2(s_i)\right] \Delta_1 \\ a_2 + \left[m_2 + (1-m_2)\, t_2(s_i)\right] \Delta_2 \end{pmatrix}, & s_0 < s_i \leq 1\ . \end{cases} \tag{2.66}$$

Continuity at s_0 holds for CV_j *if and only if* $m_j = f_j$; otherwise a small designed jump appears at s_0 (acceptable here because this is only an initializer). All practical parameter values used in this work are summarized in Table 2.

Table 2.: Practical choices used in this work for the Minimum Free Energy Path with Nudged Elastic Band method.

Category	Quantity (symbol)	Value / rule
Waterfall init	Phase split (s_0)	0.7
	Early fractions (f_1, f_2)	$f_1 = 0.10$ (CV1), $f_2 = 0.90$ (CV2)
	Late baselines (m_1, m_2)	$m_1 = 0.07$, $m_2 = 0.90$
Images	Number of images (N)	30
Smoothing	Gaussian std. (σ)	1.5 (grid units)
Gradient	FD step (h)	$5 \cdot 10^{-3}$ (CV units)
Springs	Spring constant (k)	0.2
Integrator	Step size (α)	initial $3 \cdot 10^{-3}$; then 10^{-3} if $\max_i \lVert \mathbf{F}_i^{\mathrm{tot}} \rVert > 5 \cdot 10^{-2}$; $3 \cdot 10^{-3}$ if $2 \cdot 10^{-2} < \max_i \lVert \mathbf{F}_i^{\mathrm{tot}} \rVert \leq 5 \cdot 10^{-2}$; otherwise $5 \cdot 10^{-3}$
Regularizer	Waterfall weight (λ)	0.3 (early iterations)
Convergence	Tolerance (ε)	stop when $\max_i \lVert \mathbf{F}_i^{\perp} + \mathbf{F}_i^{\parallel} \rVert < 5 \cdot 10^{-4}$

Part III

Methods II: Neural-Network Strategies

3.1. Overview

This chapter sets out the machine-learning *methods as used and implemented in this thesis*. It fixes the architectural templates, dataset conventions, and the runtime coupling patterns that underpin the application chapters, bridging the literature to the concrete setups actually employed here. The focus is on *component-level* descriptions across the full pipeline: how datasets are generated and labeled; which network architectures are employed (inputs/features, layer stacks, activations, readouts, losses, training protocols); and how these models are coupled at run time to simulation engines (GROMACS[121] and DFTB+[129]). The elements are presented as modular building blocks; composition into end-to-end workflows is deferred to the Integration chapter (Chap. IV). System-specific hyperparameters and thresholds are reported in the respective result chapters.

Whereas Methods I summarizes physical background, Methods II *selects, adapts, and specifies* the methods actually used in the applications. I first introduce essentials of neural networks and training (Secs. 3.2.1–3.2.3). I then detail the two methods employed – second-generation HDNNPs with ACSF descriptors and $SO(3)$-equivariant message passing (MACE/MACE-EQ, Sec. 5.5) – in the roles they play here. For phosphorylation, I use Δ-learning against DFTB3 on the isolated QM region (Sec. 3.5), including dataset conventions and splits (Sec. 3.5.3) and a Δ-QM/MM correction path coupled on the fly (Sec. 3.5.4). For fluorescence, I deploy direct force/property models driven by geometry and per-atom ESP from the MM field (Sec. 3.6); forces are energy gradients and bright-state identity follows Methods I (Sec. 2.3.4), while *system-specific* thresholds are stated in the respective result chapters. Data-generation choices are summarized in Sec. 3.7: production datasets rely on unbiased QM/MM and well-tempered metadynamics, whereas additional strategies are introduced here for context and comparatively assessed – beyond the application use cases – in the open-ended Sampling chapter (Chap. VIII). The application chapters – Phosphorylation (Chap. V), Fluorescence (Chap. VI), and Retinal (Chap. VII) – refer back to this chapter rather than repeating method details.

3.2. Machine Learning

This chapter introduces the machine learning (ML) methods employed throughout this work. ML models provide an efficient way to approximate quantum-mechanical properties such as energies and forces, which are otherwise computationally expensive to evaluate. By learning from precomputed reference data, these models can reproduce high-level accuracy at a fraction of the cost, enabling large-scale sampling and long molecular dynamics trajectories that would be infeasible with direct quantum calculations.

The use of ML is central to this dissertation in two ways: first, as an additive correction to low-cost methods via delta-learning, and second, as a direct replacement of quantum-mechanical contributions in hybrid QM/MM simulations. Both strategies significantly accelerate simulations while retaining quantum-level fidelity in the relevant reactive regions.

3.2.1. Artificial Neural Networks

Machine Learning (ML) models predict the output of a given set of input data, based on data that is not directly stored.[141] The underlying assumption is that there exists a functional relation f between the input $\mathbf{x}$ and output data $\mathbf{y}$, which can be approximated by a sufficiently designed model.[142] The ML model of choice for this work are deep neural networks (NN).[143,144] These models are structured in layers and consist of an input layer, multiple hidden layers followed by activation functions, and an output layer with a final activation function.[145] The concept of artificial neurons originates from the early work of McCulloch and Pitts, who first proposed a mathematical model of neuronal activity using a threshold function as activation.[146] Each linear layer i uses the output $\mathbf{y}_{i-1}$ of the previous layer as its input $\mathbf{x}_i = \mathbf{y}_{i-1}$ and computes

$$\mathbf{y}_i = f(\mathbf{x}_i) = \mathbf{W}^{(i)}\mathbf{x}_i. \tag{3.1}$$

Here, $\mathbf{x}_i \in \mathbb{R}^{n+1}$ and $\mathbf{y}_i \in \mathbb{R}^m$, with $\mathbf{W}^{(i)} \in \mathbb{R}^{m\times(n+1)}$ being the weight matrix connecting input and output. The dimension $n+1$ accounts for a bias component.

Role of activation functions. If only linear transformations were used across all layers, the overall network would reduce to a single linear mapping, regardless of its depth. This severely limits its expressive power: it would be incapable of modeling nonlinear relationships that are ubiquitous in physical and chemical systems.[144] To overcome this, nonlinear activation functions σ are applied after each linear layer

$$\mathbf{h}_i = \sigma\left(\mathbf{W}^{(i)}\mathbf{x}_i\right). \tag{3.2}$$

These functions enable the composition of nonlinear mappings and form the basis of the **universal approximation theorem**, which states that even a single hidden layer with suitable nonlinearities can approximate any continuous function on a compact domain to arbitrary precision.[147]

Historical and modern activation functions. Early neural networks employed sigmoid or hyperbolic tangent (tanh) activations, which map inputs to bounded ranges

$$\sigma_{\text{sigmoid}}(x) = \frac{1}{1+e^{-x}}, \qquad \sigma_{\tanh}(x) = \tanh(x). \tag{3.3}$$

These functions are smooth and differentiable, properties that are crucial for gradient-based optimization such as backpropagation.[148] However, their saturation for large $|x|$ leads to vanishing gradients, slowing training in deep architectures.

A major breakthrough came with the Rectified Linear Unit (ReLU), defined as

$$\sigma_{\text{ReLU}}(x) = \max(0, x), \tag{3.4}$$

which mitigates vanishing gradients and accelerates convergence, especially in very deep networks.[149] Variants such as the Exponential Linear Unit (ELU)[150] and Softplus smooth the discontinuity at zero, offering improved stability

$$\sigma_{\mathrm{ELU}}(x) = \begin{cases} x & x > 0 \\ \alpha(e^x - 1) & x \leq 0 \end{cases}, \qquad \sigma_{\mathrm{Softplus}}(x) = \log(1 + e^x). \tag{3.5}$$

Implications for molecular modeling. In the context of molecular property prediction, activation functions allow networks to capture strongly nonlinear relations between atomic coordinates and target quantities such as energies or spectra. The choice of activation impacts both expressiveness and training dynamics: ReLU and ELU are common for hidden layers due to their stability and efficiency, while the final output layer may use linear activations (e.g., for energy prediction) or bounded ones (e.g., for probabilities).

In this work, feedforward architectures with nonlinear activations are employed for supervised prediction tasks, with weights initialized randomly and optimized using backpropagation.[151]

3.2.2. Backpropagation Training

Training is performed by minimizing a loss function $L(\mathbf{x})$ that quantifies the deviation between the network prediction $\mathbf{y}$ and the target value $\hat{\mathbf{y}}$. As an example, we illustrate the *quadratic loss*

$$L(\mathbf{x}) = \frac{1}{2}|\mathbf{y} - \hat{\mathbf{y}}|^2 = \frac{1}{2}|f(\mathbf{x}) - \hat{\mathbf{y}}|^2. \tag{3.6}$$

The entire network is represented by f. The gradient of L with respect to each weight matrix is computed by the backpropagation algorithm, originally introduced to efficiently propagate gradients in multilayer perceptrons.[142,148,152,153]

Purely linear case. If no activation functions are present, each layer performs a linear mapping. The gradient then simplifies to a product of weight matrices

$$\mathbf{W}^{(i)}_{\mathrm{new}} = \mathbf{W}^{(i)}_{\mathrm{old}} - \eta \Delta \mathbf{W}^{(i)} \tag{3.7}$$

$$\Delta \mathbf{W}^{(i)} = \frac{\mathrm{d}L}{\mathrm{d}\mathbf{W}^{(i)}} = \left[\prod_{l=i+1}^{L} \mathbf{W}^{(l)T} \right] \left(\frac{\partial L}{\partial \mathbf{y}} \right) \mathbf{x}^{(i)T}, \tag{3.8}$$

where η is the learning rate and L is the number of layers.[144]

Including activation functions. In practical networks, nonlinear activation functions σ are applied after each linear transformation (see Sec. 3.2.1) to introduce nonlinearity and increase representational power.[144,149] Their derivatives enter the gradient computation via the chain rule. For *elementwise* activations at layer i with pre-activations $z^{(i)} \in \mathbb{R}^{N_i}$, we can write the Jacobian as a diagonal matrix

$$\Sigma^{(i)\prime} = \begin{pmatrix} \sigma'(z_1) & 0 & \cdots & 0 \\ 0 & \sigma'(z_2) & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma'(z_{N_i}) \end{pmatrix}, \tag{3.9}$$

where z_j are the pre-activation values for the neurons in that layer.[144] This matrix form explicitly encodes the elementwise derivatives while allowing a compact vector-matrix formulation for backpropagation. Typical activations for which this diagonal representation applies include ReLU, tanh, ELU, and GELU;[144,149,154] see also the practical overviews in Goodfellow et al. and Glorot & Bengio for usage and initialization guidance.[144,149]

In general, let

$$J^{(i)} = \frac{\partial\, \sigma^{(i)}\big(z^{(i)}\big)}{\partial z^{(i)}}$$

denote the Jacobian at layer i. For *coupled* activations or normalizations (e.g., softmax; layer/batch normalization during training; attention score normalization), $J^{(i)}$ has off-diagonal terms and should not be replaced by a diagonal matrix.[144] For softmax $s = \mathrm{softmax}(z)$ one has

$$\frac{\partial s_j}{\partial z_k} = s_j\,(\delta_{jk} - s_k) \quad \Longrightarrow \quad J = \mathrm{diag}(s) - s\,s^\top,$$

which is dense.[144] In practice, backpropagation uses Jacobian–vector products without forming J explicitly (e.g., with cross-entropy, $\nabla_z L = s - y$).[142,144,148,152,153] Batch/LayerNorm are not activations in the strict sense. During training their Jacobians contain off-diagonal terms due to coupling via mean/variance; in inference they reduce to an affine per-feature transform (effectively diagonal).[144]

The weight gradient in presence of activations is then modified to

$$\Delta \mathbf{W}^{(i)} = \frac{\mathrm{d}L}{\mathrm{d}\mathbf{W}^{(i)}} = \left[\prod_{l=i+1}^{L+S} \mathbf{W}^{(l)T} \Sigma^{(l)\prime}\right] \left(\frac{\partial L}{\partial \mathbf{y}}\right) \mathbf{x}^{(i)T}, \tag{3.10}$$

where S is the number of activation layers interleaved with the linear layers.

Interpretation. This formulation shows that each activation layer contributes a diagonal Jacobian Σ' to the gradient chain. For saturating activations such as sigmoid or tanh, entries of Σ' become small for large $|z|$, which can lead to vanishing gradients in deep networks.[148] In contrast, ReLU-like activations[149] have derivatives of either 0 or 1, which mitigates this effect and is one reason for their widespread adoption in modern architectures.[144]

Practical training. Training is typically performed using stochastic gradient descent (SGD) with mini-batches, providing a compromise between convergence speed and noise in gradient estimates.[144] The learning rate remains a critical hyperparameter: excessively high values lead to divergence, while overly low values slow convergence.

Performance metrics. To evaluate model performance, the coefficient of determination R^2[155], the mean absolute error (MAE)[156], and the root mean squared error (RMSE)[157] are employed

$$\begin{aligned} R^2 &= 1 - \frac{|\mathbf{y} - \hat{\mathbf{y}}|^2}{|\mathbf{y} - \bar{\mathbf{y}}|^2} \\ \mathrm{MAE} &= \frac{1}{N}|\mathbf{y} - \hat{\mathbf{y}}| \\ \mathrm{RMSE} &= \sqrt{\frac{1}{N}|\mathbf{y} - \hat{\mathbf{y}}|^2}. \end{aligned} \tag{3.11}$$

Here, $\mathbf{y} \in \mathbb{R}^N$ denotes the prediction and $\bar{\mathbf{y}} \in \mathbb{R}^N$ the mean value of the targets $\hat{\mathbf{y}}$, with components

$$\bar{y}_i = \frac{1}{N} \sum_{i}^{N} \mathbf{y}_i. \tag{3.12}$$

3.2.3. Hyperparameters & Optimization

The predictive performance of neural networks depends sensitively on the choice of hyperparameters, such as the learning rate, number of layers, and number of neurons per layer. Since these parameters are not learned during training, systematic search strategies are required to identify suitable configurations. Classical approaches like grid or random search sample hyperparameter combinations uniformly and train each for a fixed number of epochs.[158] While effective for small search spaces, these methods become inefficient for modern deep networks due to the high cost of full training runs.

To address this, this work employs the *Hyperband* algorithm, which reformulates hyperparameter tuning as a resource allocation problem.[159] Hyperband combines two ideas: (i) random sampling of many configurations (exploration) and (ii) successive halving, where poorly performing candidates are discarded early and remaining candidates receive more resources (exploitation). This balances coverage of the search space with computational efficiency.

Algorithmic principle. Hyperband defines a maximum resource budget R (e.g., maximum training epochs) and a reduction factor η (commonly $\eta = 3$). The algorithm computes the maximum number of halving stages

$$s_{\max} = \lfloor \log_\eta R \rfloor, \tag{3.13}$$

and an overall budget

$$B = (s_{\max} + 1)R. \tag{3.14}$$

For each bracket $s = s_{\max}, s_{\max} - 1, \dots, 0$, Hyperband initializes a number of configurations n and an initial resource per configuration r as

$$n = \left\lceil \frac{B}{R} \frac{\eta^s}{s+1} \right\rceil, \tag{3.15}$$

$$r = R\,\eta^{-s}. \tag{3.16}$$

Each bracket runs a *successive halving* procedure: all n configurations are trained for r epochs, the worst $1 - 1/\eta$ fraction is discarded, and the survivors are retrained with η times more epochs. This continues until only one configuration remains or the maximum budget is reached. Hyperband repeats this for all brackets, covering both aggressive and conservative trade-offs between breadth (many short runs) and depth (few long runs).

Advantages. Compared to pure random search, Hyperband achieves orders of magnitude speedup by discarding poor candidates early, yet maintains broad coverage of the search space. Unlike Bayesian optimization, it does not rely on surrogate models of the loss landscape, making it scalable to high-dimensional spaces frequently encountered in molecular machine learning.[23]

3.2.4. Label Z-Normalization

When training the Δ-learning models, we standardize targets $x_1, \dots, x_n$ within each reaction family by a per-family z-score[144,151]

$$z_i \;=\; \frac{x_i - \hat{\mu}}{\hat{\sigma}}, \tag{3.17}$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the sample mean and sample standard deviation computed *on the training split* of that family. By construction, the transformed *training* set then has mean 0 and standard deviation 1 (up to rounding and the n vs. $n{-}1$ convention). On validation/test splits, using the *training* $(\hat{\mu}, \hat{\sigma})$ generally yields mean $\neq 0$ and std. $\neq 1$; this is intended to avoid leakage[144]. This normalization *does not* fix the range. If $a = \min_i x_i$ and $b = \max_i x_i$, then

$$\mathrm{Range}_z \;=\; \max_i z_i - \min_i z_i \;=\; \frac{b - \hat{\mu}}{\hat{\sigma}} - \frac{a - \hat{\mu}}{\hat{\sigma}} \;=\; \frac{b - a}{\hat{\sigma}}. \tag{3.18}$$

CDF and "upper tail" probability. Let $Z \sim \mathcal{N}(0, 1)$ denote a standard normal variable. Its cumulative distribution function (CDF) is[160]

$$\Phi(x) \;=\; \Pr(Z \leq x) \;=\; \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt, \tag{3.19}$$

and the probability to lie to the *right* of x (the "upper tail") is $1 - \Phi(x) = \Pr(Z > x)$.

Maxima from i.i.d. samples. Idealizing the per–family z-scores as i.i.d. standard normals $Z_1, \ldots, Z_n$, the sample maximum

$$M_n \;=\; \max(Z_1, \ldots, Z_n) \tag{3.20}$$

obeys[161]

$$\Pr(M_n \le x) \;=\; \Pr(Z_1 \le x, \ldots, Z_n \le x) \;=\; \big[\Phi(x)\big]^n, \tag{3.21}$$

because all n values must fall below x. A convenient way to locate a *typical* level of M_n is to consider how many points exceed a threshold x. With indicators $I_i = \mathbf{1}\{Z_i > x\}$ and the exceedance count $N_x = \sum_{i=1}^{n} I_i$, first note that each $I_i \in \{0, 1\}$ is Bernoulli with success probability

$$p_x \;=\; \Pr(Z > x) \;=\; 1 - \Phi(x). \tag{3.22}$$

Hence

$$\mathbb{E}[I_i] \;=\; 0 \cdot \Pr(I_i = 0) + 1 \cdot \Pr(I_i = 1) \;=\; \Pr(Z_i > x) \;=\; 1 - \Phi(x). \tag{3.23}$$

By linearity of expectation (independence not required),

$$\mathbb{E}[N_x] \;=\; \sum_{i=1}^{n} \mathbb{E}[I_i] \;=\; n \Pr(Z > x) \;=\; n\big(1 - \Phi(x)\big). \tag{3.24}$$

Choosing x so that $\mathbb{E}[N_x] \approx 1$ places x near the typical maximum,[161] which yields

$$1 - \Phi(x) \;\approx\; \frac{1}{n} \quad\Longrightarrow\quad x \;\approx\; \Phi^{-1}\!\Big(1 - \frac{1}{n}\Big). \tag{3.25}$$

Remark. $\mathbb{E}[I_i] = p_x$ is a probability, whereas $\mathbb{E}[N_x] = \sum_{i=1}^{n} \mathbb{E}[I_i] = np_x$ is an *expected count*, not a probability. For i.i.d. Z_i, one also has $N_x \sim \text{Binomial}(n, p_x)$ and $\Pr(M_n \le x) = \Phi(x)^n$, consistent with the same scale.

By symmetry of $\mathcal{N}(0, 1)$, the minimum typically sits near $-x$, so the *typical* z-range is

$$\text{Range}_z \;\approx\; x - (-x) \;=\; 2x \;=\; 2\,\Phi^{-1}\!\Big(1 - \frac{1}{n}\Big). \tag{3.26}$$

A closely related choice is the *median* of M_n, defined by $\Pr(M_n \le x) = \frac{1}{2}$, i.e. $\Phi(x)^n = \frac{1}{2}$, which yields

$$x \;\approx\; \Phi^{-1}\!\Big(1 - \frac{\ln 2}{n}\Big), \tag{3.27}$$

agreeing with (3.25) to leading order in 1/n.[161]

Asymptotics and numerical scale. For large n, the normal upper quantile admits the classical extreme-value approximation[162]

$$\Phi^{-1}\!\left(1-\frac{1}{n}\right) \;\approx\; \sqrt{2\ln n} \;-\; \frac{\ln\ln n + \ln(4\pi)}{2\sqrt{2\ln n}}, \tag{3.28}$$

so for $n \approx 9{\cdot}10^3$ to $1.2{\cdot}10^4$ one gets $x \approx 3.7$ and hence a typical $\mathrm{Range}_z \approx 7.4$ via (3.26). This explains why, after z-normalization, different reaction families with similar sizes n show very similar observed ranges: it is a property of normal extremes at fixed n, not an intrinsic consequence of the normalization.

Diagnostics for non-normal tails. For large z, the standard tail (Mills) approximation gives[163]

$$\Pr(Z > z) \;\approx\; \frac{1}{z\sqrt{2\pi}}\, e^{-z^2/2}. \tag{3.29}$$

Values like $z \approx 25$ are astronomically unlikely under $\mathcal{N}(0,1)$ (far beyond any practical n), and thus indicate outliers, a mixture/heavy tail, or QC issues rather than ordinary fluctuation.

Practical takeaways. (i) z-normalization enforces mean 0 and std. 1 on the training split but does *not* enforce equal ranges across families; equal-looking ranges arise naturally when families are roughly normal and of similar size n (Eqs. (3.25)–(3.26)). (ii) For $n \sim 10^4$, expect maxima near +3.7 and minima near −3.7, hence ranges near 7.4 (Eqs. (3.28)–(3.26)). (iii) Investigate extremes that significantly exceed this scale; robust options include light outlier clipping (e.g., $|z| \leq 5$) and robust scaling based on median and IQR/MAD[164].

3.3. High-Dimensional Neural Network Potentials

High-dimensional neural network potentials (HDNNPs) provide accurate and efficient representations of potential energy surfaces (PES) for systems with thousands of atoms. Since the seminal work by Behler and Parrinello, *"Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces"* (2007)[12], HDNNPs have evolved into a versatile framework that embeds key physical symmetries into machine-learning models.[165] A comprehensive taxonomy is given in Behler's review, *"Four Generations of High-Dimensional Neural Network Potentials"* (2021).[165]

Generations and scope. First-generation models used a single feed-forward network for the total system energy and suffered from scalability and symmetry limitations.[165] Second-generation HDNNPs (Fig. 7) introduced an additive atom-centered decomposition that is invariant to translation, rotation, and permutation of identical atoms, as reviewed in Behler, *"Machine Learning Potentials for Atomistic Simulations"* (2021).[166] Third- and fourth-generation models incorporate long-range electrostatics and nonlocal effects (e.g., charge transfer) via explicit charge networks or global equilibration.[165] Which generation is appropriate depends on the target property and the degree of nonlocality involved.

Second-generation HDNNP. Within the HDNNP family, only *second-generation* models are employed (Fig. 7); accordingly, only this generation is developed in detail here. The application-specific motivation is discussed in Sec. 3.5.1. For a second-generation HDNNP, the system's total energy is written as a sum of atomic contributions,

$$E_{\text{total}} = \sum_{i=1}^{N} E_i = \sum_{i=1}^{N} f(\{G_i\}), \tag{3.30}$$

where each atomic term E_i depends only on the local environment of atom i within a finite cutoff R_c, encoded by atom-centered symmetry functions (ACSFs).[165,167] The ACSF design goals and explicit forms are summarized next.

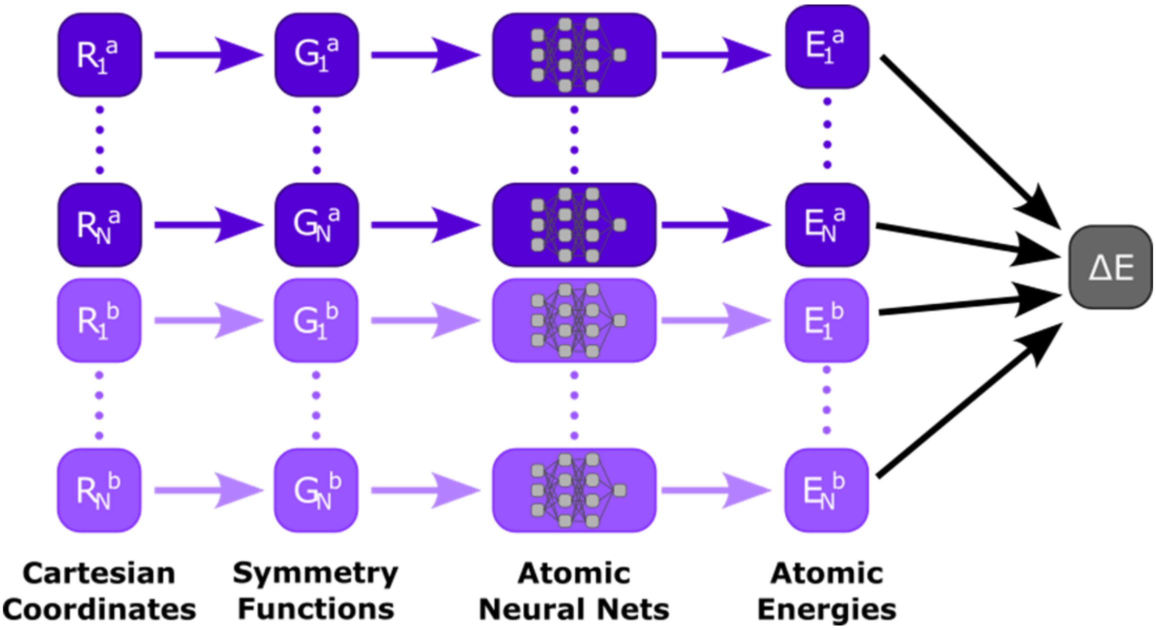


Figure 7.: Schematic architecture of a second-generation HDNNP. Each atom is represented by a local atomic environment described by symmetry functions G_i^{ν}, which serve as input to atom-centered neural networks of identical architecture but independent weights; the atomic contributions are summed to yield the total energy.

Symmetry functions. A smooth cutoff ensures locality and differentiability,

$$f_c(R_{ij}) = \begin{cases} \frac{1}{2}\left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1\right] & \text{if } R_{ij} \le R_c, \\ 0 & \text{if } R_{ij} > R_c, \end{cases} \tag{3.31}$$

where R_{ij} is the interatomic distance. Radial ACSFs collect pairwise information,

$$G_i^{\mathrm{rad}} = \sum_{j \neq i} \exp\big[-\eta (R_{ij} - R_s)^2 \big] \, f_c(R_{ij}) \,, \tag{3.32}$$

with Gaussian width η and center R_s, while angular ACSFs encode three-body correlations,

$$G_i^{\mathrm{ang}} = 2^{1-\zeta} \sum_{j,\, k \neq i;\, j \neq k} \big(1 + \lambda \cos\theta_{ijk}\big)^{\zeta} \exp\big[-\eta (R_{ij}^2 + R_{ik}^2 + R_{jk}^2) \big] f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}) \,, \tag{3.33}$$

where $\lambda \in \{-1, +1\}$ selects the angular center and ζ controls angular sharpness. Several (η, R_s) and (η, λ, ζ) settings per element provide structural resolution.[166] Radial angular ACSFs are shown for different parameters in Fig. 8.

Design goals and invariances. ACSFs act as a *local fingerprint* of the neighborhood of atom i within a finite receptive field R_c.[12,165,166] They are constructed to (i) be invariant to translation/rotation, (ii) be permutation-invariant for identical atoms, and (iii) be smooth/differentiable. The radial terms (Eq. 3.32) resolve coordination shells; the angular terms (Eq. 3.33) encode three-body geometry. Higher-order effects enter via the NN nonlinearity.[165]

Fingerprint/CNN analogy and practical choices. Conceptually, R_c plays the role of a CNN *receptive field* (Fig. 9):[144] the network "sees" an atom together with its neighborhood. Banks of radial (η, R_s) and angular (η, λ, ζ) functions act like fixed filters; the network learns their combinations to detect patterns (coordination, angles, polarization). A modest, overlapping grid balances resolution and cost.[166] Because ACSFs decay smoothly to zero (Figure 8a), forces remain stable in MD/QM/MM. In this work, ACSFs provide *local* features for Δ-learning on top of DFTB, whose long-range electrostatics already capture nonlocal interactions.[165]

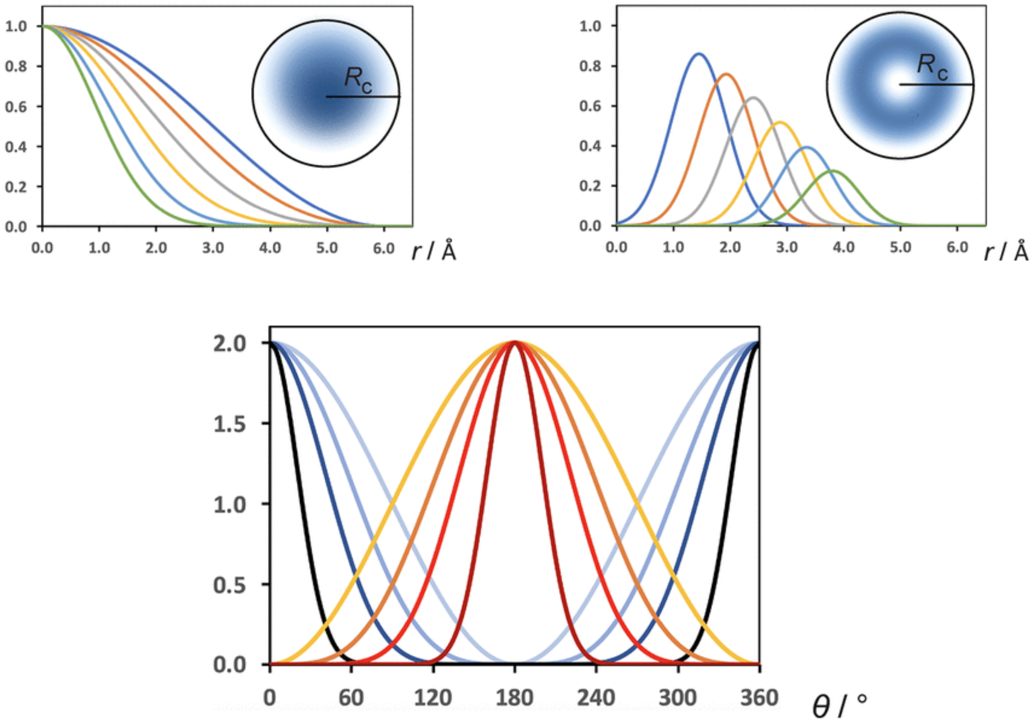


Figure 8.: Atom-centered symmetry functions (ACSFs) used in second-generation HDNNPs. Top-left: radial ACSFs with $R_s = 0$ and $\eta \in \{0.00,\ 0.03,\ 0.08,\ 0.16,\ 0.30,\ 0.50\}\ \text{Å}^{-2}$; parameters chosen for $R_c = 6$ Å and a minimum interatomic distance of 1 Å to yield approximately equidistant turning points. Top-right: radial ACSFs with $\eta = 0$ and $R_s \in \{1.5,\ 2.0,\ 2.5,\ 3.0,\ 3.5,\ 4.0\}$ Å. Bottom: angular prefactor $2^{1-\zeta}(1 + \lambda \cos\theta)^{\zeta}$ with $\lambda = \pm 1$ and $\zeta \in \{1, 2, 4, 16\}$ (blue/black for $\lambda = +1$, orange/brown for $\lambda = -1$). Adapted from Figs. 7 (panels a,b) and 9 of Behler, *Chem. Rev.* **2021**, 121, 10037–10111.[165] Reprinted (adapted) with permission from J. Behler, "Four Generations of High-Dimensional Neural Network Potentials," *Chemical Reviews* 2021, 121, 10037–10111. Copyright 2021 American Chemical Society.

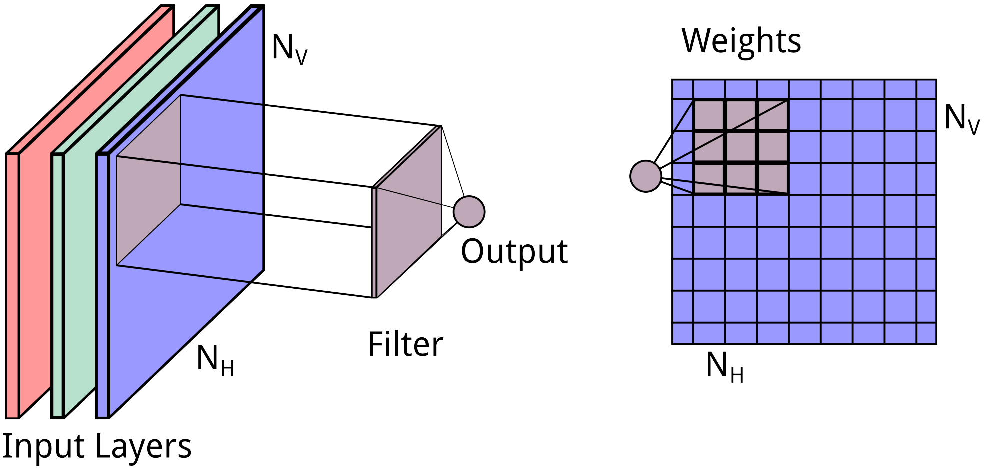


Figure 9.: Convolutional layer (schematic). **Left:** a 3-channel input image (RGB) of size $N_V \times N_H$ is convolved with a $k_V \times k_H$ kernel that spans all channels; the shaded patch marks the local receptive field whose weighted sum produces one activation in the next feature map (one output neuron shown). **Right:** front view of the same operation: the kernel slides across the grid; at each position the shaded window is multiplied–summed with the underlying pixels to yield the corresponding output value. This notion of a finite receptive field motivates the analogy to atom-centered descriptors limited by a cutoff R_c used later.

Forces and analytic derivatives (second generation). For the additive form in Eq. (3.30), forces follow by the chain rule,

$$
\begin{aligned}
\frac{\partial E_{\text{total}}}{\partial \vec{r}_l} &= \sum_{i=1}^{N} \frac{\partial E_{\text{total}}}{\partial E_i} \frac{\partial E_i}{\partial \vec{r}_l}, \\
&= \sum_{i=1}^{N} \left[\sum_{\tilde{m}=1}^{m} \frac{\partial E_i}{\partial G^{\text{rad}}_{i,\tilde{m}}} \frac{\partial G^{\text{rad}}_{i,\tilde{m}}}{\partial \vec{r}_l} + \sum_{\tilde{m}'=1}^{m'} \frac{\partial E_i}{\partial G^{\text{ang}}_{i,\tilde{m}'}} \frac{\partial G^{\text{ang}}_{i,\tilde{m}'}}{\partial \vec{r}_l} \right], \\
&= \sum_{i=1}^{N} \left[\sum_{\tilde{m}=1}^{m} \frac{\partial E_i}{\partial G^{\text{rad}}_{i,\tilde{m}}} \sum_{j\neq i} \frac{\partial G^{\text{rad}}_{i,\tilde{m}}}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \vec{r}_l} + \sum_{\tilde{m}'=1}^{m'} \frac{\partial E_i}{\partial G^{\text{ang}}_{i,\tilde{m}'}} \sum_{j\neq i} \sum_{k\neq i,j} \sum_{ab\in\{ij,ik,jk\}} \frac{\partial G^{\text{ang}}_{i,\tilde{m}'}}{\partial r_{ab}} \frac{\partial r_{ab}}{\partial \vec{r}_l} \right],
\end{aligned}
\tag{3.34}
$$

where the distance derivatives are

$$
r_{ij} = \left\| \vec{r}_i - \vec{r}_j \right\|, \qquad \frac{\partial r_{ij}}{\partial \vec{r}_l} = \begin{cases} \dfrac{\vec{r}_i - \vec{r}_j}{r_{ij}} & \text{for } l = i, \\ \dfrac{\vec{r}_j - \vec{r}_i}{r_{ij}} & \text{for } l = j, \\ \vec{0} & \text{for } l \notin \{i, j\}, \end{cases} \tag{3.35}
$$

and for the cutoff (3.31),

$$
\frac{\partial f_c(r_{ij})}{\partial r_{ij}} = \begin{cases} -\dfrac{\pi}{2R_c} \sin\left(\pi r_{ij}/R_c\right) & \text{if } r_{ij} \leq R_c, \\ 0 & \text{if } r_{ij} > R_c, \end{cases} \tag{3.36}
$$

which yields for the radial term (**??**),

$$
\frac{\partial G^{\text{rad}}_i}{\partial r_{ij}} = \sum_{j\neq i} \exp\left[-\eta (r_{ij} - R_s)^2 \right] \left[-2\eta (r_{ij} - R_s)\, f_c(r_{ij}) + \frac{\partial f_c(r_{ij})}{\partial r_{ij}} \right], \tag{3.37}
$$

and auxiliary angular derivatives (law of cosines and product rule) are

$$\begin{aligned}
\frac{\partial}{\partial r_{ij}} \cos\theta_{ijk} &= \frac{1}{r_{ik}} - \frac{\cos\theta_{ijk}}{r_{ij}}\,, \\
\frac{\partial}{\partial r_{ik}} \cos\theta_{ijk} &= \frac{1}{r_{ij}} - \frac{\cos\theta_{ijk}}{r_{ik}}\,, \\
\frac{\partial}{\partial r_{jk}} \cos\theta_{ijk} &= -\frac{r_{jk}}{r_{ij} r_{ik}}\,, \\
\frac{\partial}{\partial r_{ij}} e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)} &= -2\eta r_{ij} e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)}\,, \\
\frac{\partial}{\partial r_{ik}} e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)} &= -2\eta r_{ik} e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)}\,, \\
\frac{\partial}{\partial r_{jk}} e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)} &= -2\eta r_{jk} e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)}\,, \\
\frac{\partial}{\partial r_{ij}}\Big[f_c(r_{ij}) f_c(r_{ik}) f_c(r_{jk})\Big] &= \frac{\partial f_c(r_{ij})}{\partial r_{ij}} f_c(r_{ik}) f_c(r_{jk})\,, \\
\frac{\partial}{\partial r_{ik}}\Big[f_c(r_{ij}) f_c(r_{ik}) f_c(r_{jk})\Big] &= f_c(r_{ij}) \frac{\partial f_c(r_{ik})}{\partial r_{ik}} f_c(r_{jk})\,, \\
\frac{\partial}{\partial r_{jk}}\Big[f_c(r_{ij}) f_c(r_{ik}) f_c(r_{jk})\Big] &= f_c(r_{ij}) f_c(r_{ik}) \frac{\partial f_c(r_{jk})}{\partial r_{jk}}\,,
\end{aligned} \tag{3.38}$$

so that the angular descriptor derivatives read

$$\begin{aligned}
\frac{\partial G_i^{\text{ang}}}{\partial r_{ij}} &= \sum_{j\neq i} \sum_{k\neq i,j} 2^{1-\zeta} (1+\lambda\cos\theta_{ijk})^{\zeta} e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)} f_c(r_{ik}) f_c(r_{jk}) \\
&\quad \times \left[\zeta\lambda \frac{\frac{1}{r_{ik}} - \frac{\cos\theta_{ijk}}{r_{ij}}}{1+\lambda\cos\theta_{ijk}} f_c(r_{ij}) \;-\; 2\eta r_{ij} f_c(r_{ij}) \;+\; \frac{\partial f_c(r_{ij})}{\partial r_{ij}} \right], \\
\frac{\partial G_i^{\text{ang}}}{\partial r_{ik}} &= \sum_{j\neq i} \sum_{k\neq i,j} 2^{1-\zeta} (1+\lambda\cos\theta_{ijk})^{\zeta} e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)} f_c(r_{ij}) f_c(r_{jk}) \\
&\quad \times \left[\zeta\lambda \frac{\frac{1}{r_{ij}} - \frac{\cos\theta_{ijk}}{r_{ik}}}{1+\lambda\cos\theta_{ijk}} f_c(r_{ik}) \;-\; 2\eta r_{ik} f_c(r_{ik}) \;+\; \frac{\partial f_c(r_{ik})}{\partial r_{ik}} \right], \\
\frac{\partial G_i^{\text{ang}}}{\partial r_{jk}} &= \sum_{j\neq i} \sum_{k\neq i,j} 2^{1-\zeta} (1+\lambda\cos\theta_{ijk})^{\zeta} e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)} f_c(r_{ij}) f_c(r_{ik}) \\
&\quad \times \left[\zeta\lambda \frac{\frac{r_{jk}}{r_{ij} r_{ik}}}{1+\lambda\cos\theta_{ijk}} f_c(r_{jk}) \;-\; 2\eta r_{jk} f_c(r_{jk}) \;+\; \frac{\partial f_c(r_{jk})}{\partial r_{jk}} \right].
\end{aligned}$$

3.4. MACE: Equivariant Message Passing

Positioning. While second-generation HDNNPs encode locality via fixed atom-centered symmetry functions (ACSFs), MACE learns many-body, $SO(3)$-equivariant features on radius graphs. In this thesis, second-generation HDNNPs are used for local Δ-corrections to DFTB3 in phosphorylation (Sec. 3.5), whereas *MACE-EQ* is employed (i) in the same phosphoryl-transfer setting to enable Δ-learning with an explicit total-charge constraint, and (ii) in the fluorescence project to incorporate the environment via per-atom electrostatic potentials (ESP) as inputs (Sec. 3.6).

Equivariant graph neural networks learn energies and forces on atomistic graphs while respecting translation, rotation, and permutation symmetries.[168–173] This section recalls the essentials of message passing (Sec. 3.4.1), then presents the MACE architecture (Sec. 3.4.2), and finally its charge–equilibration extension MACE-EQ (Sec. 3.4.3).

3.4.1. Graph Neural Networks

Graph Neural Networks (GNNs) operate on graph-structured data with information distributed over nodes and edges.[168,174] In molecular ML, nodes are atoms and edges represent pairwise interactions; this naturally encodes local and global environments.

Graph representation. A molecular graph is $G = (V, E)$ with nodes $v \in V$ and edges $(v, w) \in E$. Node features $\mathbf{x}_v \in \mathbb{R}^{d_x}$ (e.g., element) and edge features $\mathbf{e}_{vw} \in \mathbb{R}^{d_e}$ (e.g., distance) initialize hidden states

$$\mathbf{h}_v^0 = \mathbf{x}_v. \tag{3.39}$$

Message passing framework. Following Gilmer *et al.*,[174] updates iterate messages and state transforms

$$\mathbf{m}_v^{t+1} = \sum_{w \in \mathcal{N}(v)} M_t\big(\mathbf{h}_v^t,\ \mathbf{h}_w^t,\ \mathbf{e}_{vw}\big)\,, \tag{3.40}$$

$$\mathbf{h}_v^{t+1} = U_t\big(\mathbf{h}_v^t,\ \mathbf{m}_v^{t+1}\big)\,, \tag{3.41}$$

with learnable M_t and U_t; gated units (e.g., GRUs) are common:[175] $\mathbf{h}_v^{t+1} = \mathrm{GRU}(\mathbf{h}_v^t, \mathbf{m}_v^{t+1})$.

Readout. After T steps a permutation-invariant readout R produces graph outputs, $\mathbf{y}_G = R\big(\{\mathbf{h}_v^T\}_{v \in V}\big)$, typically a sum/mean plus an MLP.

Edge features and continuous interactions. Continuous edge features (distances, angles) are embedded via small networks to obtain interaction-dependent weights, $A(\mathbf{e}_{vw}) = W_2\,\sigma(W_1\,\mathbf{e}_{vw})$, allowing smooth geometric dependence.[173]

Symmetry considerations. Physical targets obey translation/rotation/permutation constraints; energies are invariant and forces are rotation-equivariant.[170] One route to encode these constraints uses irreducible representations and tensor products; the concrete realization for MACE follows below (Sec. 3.4.2).[171,172]

Beyond pairwise interactions. Higher-order effects can be captured by many-body message constructions,[39] instantiated explicitly in MACE.

3.4.2. Core MACE

The *MACE* model maps atomic positions and elements to energy via $SO(3)$-equivariant message passing on a radius-limited graph.[39,176] Each node i (element Z_i, position $\mathbf{r}_i$) carries features in irreducible $SO(3)$ representations, indexed by (ℓ, m) and channel k; (s) denotes the message-passing layer.

Initialization and channel mixing. Scalar (ℓ=0) features initialize from a learnable element embedding

$$h^{(0)}_{i,k00} = \sum_{z} W_{kz}\, \delta_{zZ_i}, \tag{3.42}$$

then mix linearly within each (ℓ, m) block:

$$\bar{h}^{(s)}_{i,k\ell m} = \sum_{\tilde{k}} W^{(s)}_{k\tilde{k}\ell}\, h^{(s)}_{i,\tilde{k}\ell m}. \tag{3.43}$$

Radial basis and learnable radial weights. Distances $r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$ are expanded in a smooth radial basis (e.g., spherical Bessel with polynomial cutoff f_{cut}),

$$j^0_n(r_{ij}) = \sqrt{\frac{2}{r_{\text{cut}}}}\, \frac{\sin\!\left(\frac{n\pi r_{ij}}{r_{\text{cut}}}\right)}{r_{ij}}\, f_{\text{cut}}(r_{ij}), \tag{3.44}$$

processed by small MLPs into distance-dependent weights:

$$R^{(s)}_{k\eta\ell_1\ell_2\ell_3}(r_{ij}) = \mathrm{MLP}\big(\{j^0_n(r_{ij})\}_n\big)\,. \tag{3.45}$$

Equivariant messages via tensor products. With $\hat{\mathbf{r}}_{ij} = \mathbf{r}_{ij}/r_{ij}$ and real spherical harmonics $Y^m_\ell(\hat{\mathbf{r}}_{ij})$, messages use Clebsch–Gordan tensor products,[170–172]

$$\phi^{(s)}_{ij,k\eta\ell_3 m_3} = \sum_{\ell_1,\ell_2} \sum_{m_1,m_2} C^{\ell_3 m_3,\eta}_{\ell_1 m_1,\ell_2 m_2}\, R^{(s)}_{k\eta\ell_1\ell_2\ell_3}(r_{ij})\; Y^{m_1}_{\ell_1}(\hat{\mathbf{r}}_{ij})\; \bar{h}^{(s)}_{j,k\ell_2 m_2}. \tag{3.46}$$

Aggregation with learnable weights gives

$$A^{(s)}_{i,k\ell_3 m_3} = \sum_{\tilde{k},\eta} W^{(s)}_{k\tilde{k}\eta\ell_3} \sum_{j\in\mathcal{N}(i)} \phi^{(s)}_{ij,\tilde{k}\eta\ell_3 m_3}. \tag{3.47}$$

Higher effective body order and updates. Stacking layers raises effective body order and extends the receptive field. Products of aggregated tensors are organized into higher-order tensors and recoupled via CG maps (schematically),

$$B^{(s),\nu}_{i,\eta_\nu kLM} = \sum_{\{(\ell,m)\}} C^{LM,\eta_\nu}_{\{\ell m\}} \prod_{\xi=1}^{\nu} A^{(s)}_{i,k\ell_\xi m_\xi}. \tag{3.48}$$

Residual updates then read

$$m^{(s)}_{i,kLM} = \sum_{\nu,\eta_\nu} W^{(s),\nu}_{Z_i,\eta_\nu kL}\, B^{(s),\nu}_{i,\eta_\nu kLM}, \qquad h^{(s+1)}_{i,kLM} = \sum_{\tilde{k}} W^{(s)}_{kL,\tilde{k}}\, m^{(s)}_{i,\tilde{k}LM} + \sum_{\tilde{k}} W^{(s)}_{kZ_iL,\tilde{k}}\, h^{(s)}_{i,\tilde{k}LM}. \tag{3.49}$$

Equivariance and readout. Scalar targets are invariant and vector targets must be rotation-equivariant. MACE enforces this by carrying features in irreducible $SO(3)$ representations and coupling them via CG products; only ℓ=0 channels feed the energy readouts. After S layers,

$$E_i = \sum_{s=1}^{S} E_i^{(s)} = \sum_{s=1}^{S} R^{(s)}\big(\mathbf{h}_i^{(s)}\big), \qquad R^{(s)}\big(\mathbf{h}_i^{(s)}\big) = \begin{cases} \sum_k W_k^{(s)}\, h^{(s)}_{i,k00} & s < S, \\ \mathrm{MLP}\Big(\{h^{(s)}_{i,k00}\}_k\Big) & s = S. \end{cases} \tag{3.50}$$

The total energy and forces are

$$E_{\text{total}} = \sum_{i=1}^{N} E_i, \qquad \mathbf{F}_i = -\nabla_{\mathbf{r}_i} E_{\text{total}}, \tag{3.51}$$

with forces from automatic differentiation.

Practical knobs. Three hyperparameters dominate accuracy/cost: (i) channels k (width), (ii) maximum angular momentum $L_{\max}$ (anisotropy order), and (iii) depth S (effective body order; receptive field $\sim S \cdot r_{\text{cut}}$). Invariant variants use $L_{\max}$=0 (e.g., "MACE 64-0"); modest depths like S=2 already capture up to ~three-body correlations within $2r_{\text{cut}}$ and often reach sub-kcal/mol MAEs on standard molecular benchmarks.[177,178]

3.4.3. MACE-EQ

Implementation note. The MACE-EQ formulation follows the differentiable charge-equilibration paradigm. The integration into our QM/MM runtime (GROMACS[121] hook, charge-constraint solve, and data pipeline) was implemented in this work; the core MACE-EQ method was originally developed by Lukas Petersen.

Use cases in this work. MACE-EQ is used in two roles. First, for phosphorylation it enables Δ-learning against DFTB3 while enforcing the prescribed *total charge* of the QM region via a differentiable charge-equilibration solve (Sec. 3.5). Second, for fluorescence property models it serves as a graph backbone that can *ingest environment information* through per-atom electrostatic potentials (ESP) from the MM embedding (Sec. 3.6); in this setting no Δ-target is used.

To absorb residual polarization/charge-transfer differences between DFT and DFTB within a Δ-learning setup, MACE is augmented with a differentiable *charge–equilibration* (Eq) head in the spirit of electronegativity equalization[179] and charge-equilibration models,[180] including modern variants such as QTPIE.[181] The Eq term adds a physically motivated long-range component while keeping the short-range energy in the invariant MACE readout.

Electronegativities, hardness, and kernel. Each MACE block t contributes a scalar *electronegativity* $\chi_i^{(t)}$ from the ℓ=0 channels; the final per-atom value is

$$\chi_i = \sum_{t=0}^{T} \chi_i^{(t)}. \tag{3.52}$$

Optional diagonal hardness terms η_i may be predicted. A screened Coulomb kernel $\gamma_{ij}(r_{ij})$ (finite cutoff; smooth near field) defines the interaction matrix

$$J_{ij} = \eta_i \delta_{ij} + \left(1 - \delta_{ij}\right)\gamma_{ij}(r_{ij}). \tag{3.53}$$

Note on explicit environment features (ablated). Early prototypes added an extra per-atom environment feature Φ_i and used an effective RHS $(\chi_i + \Phi_i)$. In the *final model used here*, no additional Φ_i node feature is used; the Eq system is solved with RHS $-\chi_i$ and only the prescribed total charge as a constraint. Ablations showed no robust gain from Φ_i but increased variance and complexity. Consequently, the environment feature Φ_i is only directly added to the scalar electronegativity χ_i.

Constrained Eq solve (prescribed total charge). Given the system's total charge q_{tot}, charges $\mathbf{q} = (q_1, \dots, q_N)$ minimize

$$E_{\text{Eq}}(\mathbf{q}) = \sum_i \chi_i q_i + \tfrac{1}{2} \sum_{i,j} q_i J_{ij} q_j, \quad \text{s.t.} \quad \sum_i q_i = q_{\text{tot}}. \tag{3.54}$$

Introducing a Lagrange multiplier λ yields the augmented linear system

$$\left(\begin{array}{cccc|c} J_{11} & J_{12} & \cdots & J_{1N} & 1 \\ J_{21} & J_{22} & \cdots & J_{2N} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ J_{N1} & J_{N2} & \cdots & J_{NN} & 1 \\ \hline 1 & 1 & \cdots & 1 & 0 \end{array}\right) \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \\ \lambda \end{pmatrix} = \begin{pmatrix} -\chi_1 - \Phi_1 \\ -\chi_2 - \Phi_2 \\ \vdots \\ -\chi_N - \Phi_N \\ q_{\text{tot}} \end{pmatrix}, \tag{3.55}$$

solved once per geometry with a stable, differentiable routine (implicit differentiation in backpropagation). The charges are not fed back into later MACE blocks.

Total energy and training. The model energy for Δ-learning is

$$E_{\text{tot}} = E_{\text{MACE}} + E_{\text{Eq}}, \tag{3.56}$$

with forces $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} E_{\text{tot}}$ from automatic differentiation. Supervision follows the same energy/force-Δ objective as in Sec. 3.4.2. Core-MACE baselines are trained identically but without the Eq head and without supplying q_{tot}.

Context. Related strategies – learning a short-range local energy and augmenting it with a physically motivated electrostatic term – appear in fourth-generation HDNNPs as well.[182]

3.5. Phosphorylation: Δ-Learning Networks

The concept of Δ-machine learning (Δ-ML), first formalized by Ramakrishnan et al., offers an efficient route to elevate inexpensive quantum methods by training on the energy difference between low- and high-fidelity levels.[9] This concept has since been extended to modern neural network architectures; notably, Smith et al. demonstrated with the ANI-1ccx model that transfer learning from DFT to CCSD(T) data can achieve near gold-standard accuracy while retaining efficiency suitable for large datasets.[183] This strategy has been shown to systematically improve semi-empirical methods toward DFT or even CCSD(T) accuracy and has recently been extended to condensed-phase simulations.[183,184]

Rather than predicting absolute DFT energies, the model learns the residual,

$$\Delta E = E_{\text{DFT}} - E_{\text{DFTB}}\,, \tag{3.57}$$

consistent with the Δ-ML strategy.[9,34,183,185]

Two realizations of this Δ-ML idea are used in the phosphorylation project: (i) a second-generation HDNNP that maps fixed ACSF fingerprints to per-atom corrections and constitutes the main results of this work; and (ii) an $SO(3)$-equivariant graph model (MACE) that supplies the same local correction with learned many-body features and an explicit total-charge constraint. In both cases the supervised targets are $\Delta E = E_{\text{DFT}} - E_{\text{DFTB}}$ (and, where available, $\Delta\mathbf{F} = \mathbf{F}_{\text{DFT}} - \mathbf{F}_{\text{DFTB}}$) computed on the isolated QM region; architecture-specific details follow in Secs. 3.5.1 and 3.5.2.

Conventions (atomization Δ). All ΔE labels use *atomization energies* at each level (DFT, DFTB3). When DFTB uses a finite Fermi temperature, the molecular energy equals the Mermin free energy. Atomic references are computed consistently at the same level/smearing.

3.5.1. Δ Network: HDNNP

Why second generation for this work. Third- and fourth-generation HDNNPs become essential for properties dominated by nonlocal physics (e.g., redox-active materials, strongly polarized interfaces), but they increase data requirements and model complexity and may duplicate physics already present in DFTB.[165] Because DFTB3 already supplies long-range electrostatics and approximate charge transfer,[10] we hypothesize the remaining discrepancy to DFT for phosphorylation is primarily local; a second-generation HDNNP matches this error structure, preserves linear scaling, maintains interpretability, and minimizes runtime. Accordingly, the phosphorylation application presented later in this thesis employs a *second-generation* HDNNP (Fig. 7) as a local Δ-correction to DFTB3; the general methodology is summarized below.[10,165]

Δ-learning with second-generation HDNNPs. For a second-generation HDNNP, the (correction) energy is written as a sum of atomic contributions,

$$\Delta E = \sum_{i=1}^{N} \Delta E_i = \sum_{i=1}^{N} f(\{G_i\}), \tag{3.58}$$

where each atomic term E_i depends only on the local environment of atom i within a finite cutoff R_c, encoded by atom-centered symmetry functions (ACSFs).[165] In the Δ-learning context, the HDNNP predicts

$$E_{\mathrm{NN}} = \sum_{i} \Delta E_i \approx \Delta E, \tag{3.59}$$

with ΔE_i derived solely from local descriptors, aligning with the expected locality of the DFTB residual.

Loss. Unless stated otherwise, models are trained on energy deltas only, $\mathcal{L}_{\mathrm{HDNNP}} = \frac{1}{2}\,|\Delta\hat{E} - \Delta E|^2$, with forces obtained by automatic differentiation during dynamics. Variants with joint energy+force supervision are possible but not used by default.

Table 3.: HDNNP Δ-model components.

Stage	Summary
Atomic descriptors	Radial/Angular ACSFs; cutoff R_c; element channels
Per-atom MLPs	One MLP per element (H, C, N, O, P, Mg); shared architecture, independent weights
Pooling	Sum over atoms to total ΔE
Output	$\Delta E = \sum_A \varepsilon_A(\mathrm{ACSF}_A)$
Training target	$\mathrm{MSE}(\Delta E)$; forces via $-\nabla E_{\mathrm{NN}}$ (no force labels by default)
Optimizer/HPs	*(placeholder: optimizer, LR schedule, batch size, epochs)*

3.5.2. Δ Network: MACE

MACE/MACE-EQ is evaluated as an alternative to the HDNNP Δ-learning approach to learn Δ-forces and MACE-EQ specifically to satisfy a hard total-charge constraint. MACE implements a radius-limited, $SO(3)$-equivariant message-passing network with atomic energy readouts ΔE_i; the Δ correction is

$$\Delta E_{\mathrm{NN}}(\mathbf{R}, \mathbf{Z}; q_{\mathrm{tot}}) = \sum_i \Delta E_i, \tag{3.60}$$

where q_{tot} is the prescribed total charge of the QM region. The Eq head solves a differentiable charge-equilibration system to enforce $\sum_i q_i = q_{\mathrm{tot}}$; its energy contribution is added to the invariant readout (see Sec. 3.4.3).

Targets and loss. Training uses the same labels as for HDNNP but with a graph backbone

$$\mathcal{L}_{\mathrm{MACE}} = \tfrac{1}{2}\left|\Delta\hat{E} - \Delta E\right|^2 + \lambda_F \left\|\Delta\hat{\mathbf{F}} - \Delta\mathbf{F}\right\|^2, \qquad \Delta\hat{\mathbf{F}} = -\nabla_{\mathbf{R}} E_{\mathrm{NN}}, \tag{3.61}$$

with $\lambda_F = 0$ if force labels are not available. As for HDNNP, atomization conventions are used for cross-system consistency (Sec. 3.5.3).

3.5.3. Dataset Construction

Targets and conventions (Δ-learning). In all phosphorylation experiments the supervised targets are energy and (where available) force *deltas* between a high-level DFT reference and the DFTB3 baseline on the *isolated QM region*. To stabilize scales across systems we primarily use *atomization* energies at each level,

$$E^{\mathrm{level}}_{\mathrm{atom}} = E^{\mathrm{mol}}_{\mathrm{level}} - \sum_A E^{\mathrm{atom}}_{A,\mathrm{level}}, \tag{3.62}$$

$$\Delta E = E^{\mathrm{DFT}}_{\mathrm{atom}} - E^{\mathrm{DFTB}}_{\mathrm{atom}}. \tag{3.63}$$

When DFTB employs a finite Fermi temperature, $E^{\mathrm{mol}}_{\mathrm{DFTB}}$ is taken as the Mermin electronic free energy $F^{\mathrm{Mermin}}_{\mathrm{DFTB}}$; atomic references are treated consistently at both levels. For force labels (if present), $\Delta\mathbf{F} = \mathbf{F}_{\mathrm{DFT}} - \mathbf{F}_{\mathrm{DFTB}}$.

Sampling by QM/MM well-tempered metadynamics. Reactive configurations are generated with QM/MM well-tempered metadynamics in explicit solvent. All systems except the histidine kinase (HK) are sampled in pure water boxes; HK is embedded in a protein environment. The MM environment is TIP3P water[133] (AMBER99SB-ILDN for proteins[186]); the QM region is propagated with DFTB3 (3ob[11,107]) plus phosphorus SRPs (SRP-OPhyd, SRP-PNmod).[106,187] Electrostatic embedding couples QM and MM.[130] After minimization and NVT equilibration (Nosé–Hoover) to calibrate CV fluctuations, biases are placed on two CVs that follow the reaction: the *forming* P–N and *breaking* P–O distances. This captures both concerted and stepwise mechanisms. Bias parameters (Gaussian height 1.2 kJ mol^{-1}, width 0.004 nm, deposition every 0.5 ps, bias factor 50) promote exploration without unphysical excursions.[78,188]

Uniform extraction along the reaction path. From converged (or near-converged) free-energy surfaces, configurations are drawn *uniformly along the minimum-free-energy path (MFEP)* with small plateaus included to cover high-curvature neighborhoods. We explicitly *avoid* overweighting the TS region: reactant, TS-like, and product-like segments are sampled with equal density along the MFEP arc length.

QM cutout and sanitization. Each stored snapshot is reduced to a gas-phase QM subsystem comprising the nucleophile, the phosphorus donor/leaving group, and (if present) coordinated Mg^{2+}. A linker atom is used *only* for the histidine-kinase (HK) system to cap the QM/MM boundary; aqueous systems require no linkers. Waters and distant MM atoms are removed; ring integrity (imidazole) and connectivity of the P–O scaffold are preserved. Sanitization enforces (i) topology consistency with the CV state, (ii) distance checks removing frames with unphysical close contacts (P–N, P–O, Mg–O), and (iii) an SCF pre-check (see below).

Single-point calculations (reference and baseline). Gas-phase single points are performed on the sanitized QM geometries at two levels: *(DFT)* ωB97M-V[94,102] with ma-def2-TZVP[101] (0 K occupations; no Fermi smearing), consistent with best-practice guidance for anionic phosphorus chemistry at moderate cost[189] and used in related phosphate-transfer QM/MM studies.[190] *(DFTB3)* same settings/parameters as for sampling (3ob + SRPs) with fractional occupations via a Fermi temperature of 300 K (500 K tested but not required); the Mermin free energy is used as E_{DFTB} in all formulas. Frames are retained if both levels converge numerically; no additional filtering by HOMO–LUMO gap is applied.

Label construction and variants. Primary labels are ΔE from atomization energies (above) and, where available, $\Delta\mathbf{F}$. For completeness we also export a *raw* total-energy variant $\Delta E_{\mathrm{raw}} = E_{\mathrm{mol}}^{\mathrm{DFT}} - E_{\mathrm{mol}}^{\mathrm{DFTB}}$. In this work, HDNNP models are trained with ΔE (atomization-based), while MACE(-Eq) is evaluated with both ΔE and ΔE_{raw}; in practice we observe negligible differences in downstream metrics, but atomization-based deltas are preferred as the canonical target.

Splits, views, and parity. To ensure fair comparison between architectures, the same source geometries feed two export “views”: (i) HDNNP (ACSFs computed on the fly) and (ii) MACE (radius graphs). We use balanced, group-aware splits by reaction family and charge state; geometry membership is shared across views so that HDNNP and MACE train/validate on identical splits.

Quality control, provenance, and reproducibility. The automated pipeline performs: uniform MFEP subsampling; QM isolation/sanitization; batched DFT/DFTB single points with retry policies; label assembly; and split construction. Detailed provenance tables, manifests, and outlier diagnostics are reported in the phosphorylation chapter V.

3.5.4. Δ-QM/MM Runtime Coupling

In all Δ-models, the QM/MM step augments DFTB3 with a learned correction

$$\begin{aligned} E^{\mathrm{tot}}(\mathbf{R}) &= E_{\mathrm{DFTB}}(\mathbf{R}) + E_{\mathrm{NN}}(\mathbf{R}), \\ \mathbf{F}_i^{\mathrm{tot}}(\mathbf{R}) &= \mathbf{F}_i^{\mathrm{DFTB}}(\mathbf{R}) + \mathbf{F}_{\mathrm{NN}}(\mathbf{R}), \end{aligned} \tag{3.64}$$

so that corrections act only on QM atoms while MM forces and long-range terms remain unchanged. Forces are obtained by automatic differentiation

$$F_{\mathrm{NN}}(\mathbf{R}) = -\nabla_{\mathbf{R}_i} E_{\mathrm{NN}}(\mathbf{R}). \tag{3.65}$$

HDNNP path. Training uses ΔE only; at runtime forces are obtained by automatic differentiation of E_{NN} (no separate force head).

MACE(-Eq) path. Energy+force multitask, where force labels stabilize training; at runtime forces are still taken as $-\nabla E_{\mathrm{NN}}$ to preserve energy conservation.

Advantages and challenges. Augmenting QM calculations with ML models dramatically accelerates simulations, enabling access to longer timescales and larger systems. Delta-learning reduces data requirements and improves stability.[182] However, this approaches relies on training data that comprehensively cover the relevant configurational space. Transferability to new geometries or chemical environments remains a challenge, and special care must be taken to prevent energy drift or instabilities during extended simulations.[38,191]

Notes. Implementation details (per-step hooks, energy bookkeeping, Eq solve) are documented in the Phosphorylation chapter (Chap. V); the charge-equilibration formulation is summarized in Sec. 3.4.3.

3.6. Fluorescence Networks: QM Replacement

Note on backbone. For fluorescence, MACE-EQ is used as the backbone to pass environment information into the model via per-atom ESP features from the MM field (Sec. 3.4.3); in this application the targets are direct properties (energies/oscillator strengths), not Δ-labels.

Conventions. For a system with N atoms, coordinates are $\mathbf{R} \in \mathbb{R}^{N\times 3}$, per-atom MM electrostatic potentials (ESP) are $\boldsymbol{\phi} \in \mathbb{R}^{N}$. Flattening concatenates all Cartesian components across all atoms. Absolute values and powers act componentwise; means are taken over components. Let $P = \frac{N(N-1)}{2}$ and $P' \leq P$ when pair masking is used.

3.6.1. Force Network

The purpose of this network is to predict a scalar bright-state energy $\hat{E}(\mathbf{R}, \boldsymbol{\phi})$ and obtain forces $\hat{\mathbf{F}} = -\nabla_{\mathbf{R}}\hat{E}$ via automatic differentiation. The layer stack and data flow are summarised in Table 4, and a schematic flow diagram is shown in Fig. 11. A schematic representation of the force network is shown in Fig. 10a.

Input per sample: $(N \times 4)$ tensor with columns $[x, y, z, \mathrm{ESP}]$.

Table 4.: Excited states force-network architecture: layer stack and data flow from inputs to energy and forces.

Stage	Output shape	Summary
Inverse-distance + ESP feature layer	$\mathbb{R}^{P'+N}$	Builds inverse distances $1/r_{ij}$ (all $i < j$); an optional dataset-wide mask keeps pairs with $1/r_{ij}$ above a cutoff, then concatenates per-atom ESP $\boldsymbol{\phi}$.
Normalization (fixed)	$\mathbb{R}^{P'+N}$	Constant standardization with separate mean/variance for the inverse-distance block and for the ESP block (estimated from training data).
Dense(H, ELU) $\times L$	$\mathbb{R}^{H}$	Fully connected trunk with L hidden layers of width H.
Dense(1) (linear)	$\mathbb{R}$	Scalar energy $\hat{E}$.
Energy→Force wrapper	$\mathbb{R} \oplus \mathbb{R}^{3N}$	Forces $\hat{\mathbf{F}} = -\nabla_{\mathbf{R}}\hat{E}$ (ESP is treated as an external feature and is not differentiated).

Force-model objective (generic form). Using a compact notation in which $|\cdot|$ denotes the mean absolute value over the flattened force vector (all atoms, all Cartesian components) and powers act componentwise before averaging, the loss reads

$$\mathcal{L}_{\mathrm{force}}(\lambda_E) \;=\; \lambda_E \,|E - \hat{E}|^2 \;+\; |\mathbf{F} - \hat{\mathbf{F}}| \;+\; 10^4\,|\mathbf{F} - \hat{\mathbf{F}}|^4. \tag{3.66}$$

Force-model objective (fluorescence-specific settings). Building on Eq. 3.66, only the energy weight λ_E is tuned; the L_1 force term carries unit weight, and the quartic force penalty uses a fixed prefactor 10^4 to damp outliers in regions with state reordering. The ESP channel is treated as an external, non-differentiated input during training and deployment.

Loss-function ablation and NN/MM stability. The primary objective is *accurate forces*; the energy term serves as a weak, diagnostic anchor. Using the pooled 10-solvent dataset, loss variants were compared that combine energy MSE with absolute, quadratic, and quartic force penalties while keeping $\lambda_E = 10^{-3}$ in Eq. 3.66. Table 5 summarizes the outcomes. The adopted composite loss (energy MSE + $|\mathbf{F}|$ + $10^4\,|\mathbf{F}|^4$) yields the best *force* accuracy together with NN/MM trajectories that remain stable up to nanoseconds (water is a known outlier). In contrast, replacing the absolute force term with a purely quadratic penalty (energy MSE + $|\mathbf{F}|^2$ + $|\mathbf{F}|^4$) can appear competitive on validation metrics (including R^2) yet leads to *unstable* live NN/MM runs that often crash within picoseconds. This indicates that the $|\mathbf{F}|$ term controls heavy-tailed force errors, the quartic term suppresses rare outliers, and the energy term acts only as a mild scale anchor.

Table 5.: Ablation of loss terms for the force model (training on the pooled 10-solvent dataset). A checkmark (x) indicates inclusion; a dash (–) indicates omission. Metrics reported as R^2 and mean absolute error (MAE). *Units:* MAE (forces) in kcal mol^{-1} Å^{-1} and MAE (energies) in kcal mol^{-1}. For all experiments the energy-term weight in Eq. 3.66 was fixed to $\lambda_E = 10^{-3}$. The top row (bold) is the loss function used in this chapter.

E	$\|\mathbf{F}\|$	$\|\mathbf{F}\|^2$	$\|\mathbf{F}\|^4$	R^2_F	MAE$_F$	R^2_E	MAE$_E$
x	**x**	–	**x**	**0.990**	**1.82**	**0.967**	**3.35**
–	x	–	x	0.969	3.65	< 0	489.18
x	–	x	x	0.979	2.93	0.977	2.64
–	–	x	x	0.991	1.65	0.972	3.37
x	–	x	–	0.963	3.96	0.907	5.23
–	–	x	–	0.954	4.42	< 0	2825.41
x	–	–	x	0.955	4.42	0.936	4.08
–	–	–	x	0.964	3.86	< 0	1087.39
x	x	–	–	0.990	1.80	0.914	4.98
–	x	–	–	0.989	1.95	< 0	3462.02

3.6.2. Benchmark: MACE-EQ

A MACE-EQ force network is included as a late-stage *reference baseline*, not as the primary production model. After the feedforward force network (Sec. 3.6.1) had been used throughout the study, MACE-EQ was added at the end to obtain a benchmark for performance and NN/MM stability. The choice of MACE-*EQ* (rather than an invariant graph model) is motivated by the need to expose the solvent environment to the model: the environmental potential V is represented by per-atom MM electrostatic potentials (ESP) $\boldsymbol{\phi}$ and fed to the network, yielding an energy functional that depends on geometry and environment,

$$E_{\text{ref}} = E_{\text{MACE-Eq}}(\mathbf{R}, \boldsymbol{\phi}) \quad \Longrightarrow \quad \mathbf{F} = -\nabla_{\mathbf{R}} E_{\text{ref}}, \tag{3.67}$$

with $\boldsymbol{\phi}_i \equiv V_{\text{MM}}(\mathbf{r}_i)$. Training uses the same splits and objective as in Eq. 3.66 to enable a like-for-like comparison. Methodological details of MACE and the Eq extension are given in Secs. 5.5 and 3.4.3.

3.6.3. Property Network

The aim is to predict bright-state vertical emission energy $\Delta\hat{E}_{\mathrm{em}}$ and oscillator strength $\hat{f}$ from structure and ESP, which are then used to estimate the absorption/emission spectrum (Sec. 2.3.5). The layer stack and data flow are summarised in Table 6, and a schematic flow diagram is shown in Fig. 12. A schematic representation of the property network is shown in Fig. 10b.

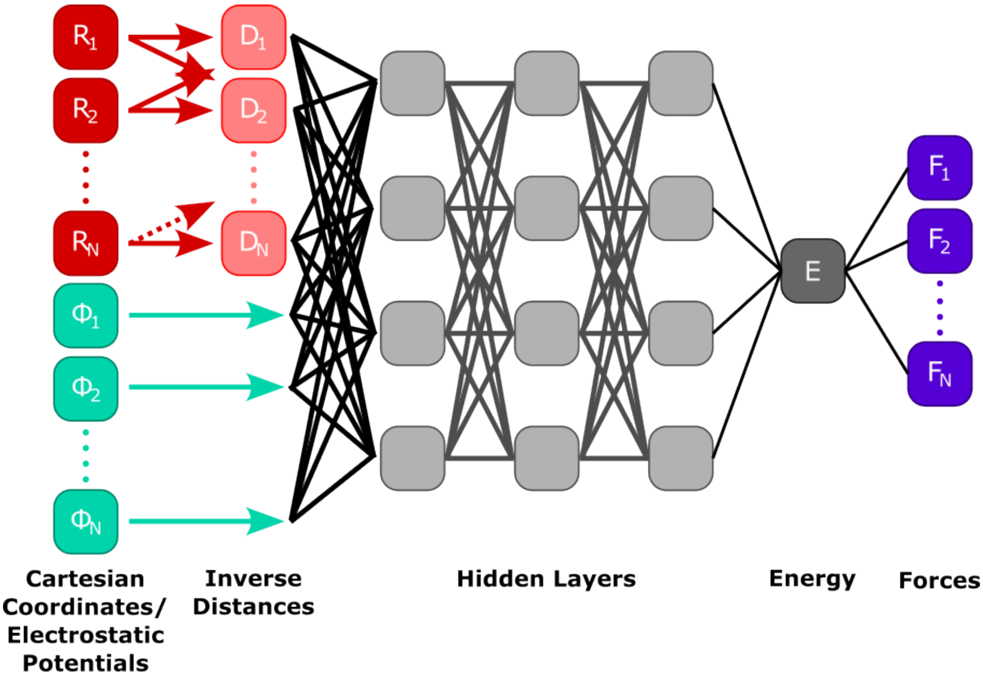
Property-model objective. A two-target loss penalizes errors in ΔE_{em} and f,

$$\mathcal{L}_{\mathrm{prop}} \;=\; |\Delta E_{\mathrm{em}} - \widehat{\Delta E}_{\mathrm{em}}|^2 \;+\; |f - \hat{f}|^2 \;+\; \lambda'_{\mathrm{L2}}\, \|W\|_2^2. \tag{3.68}$$

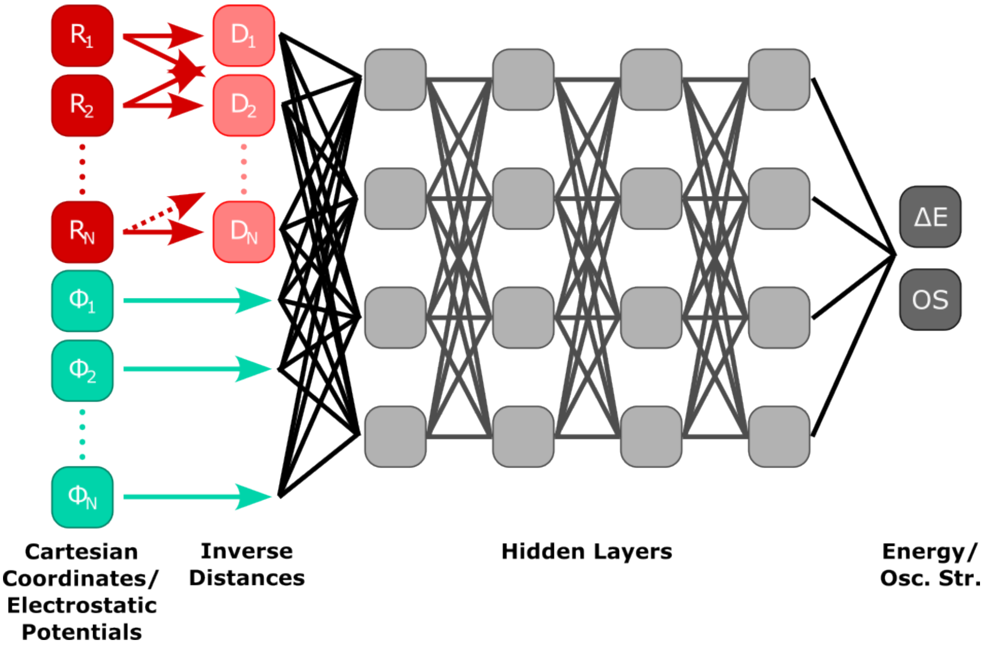
Inputs per sample: geometry $\mathbf{R} \in \mathbb{R}^{N\times 3}$, ESP $\boldsymbol{\phi} \in \mathbb{R}^{N}$.

Table 6.: Excited states property-network architecture: layer stack and data flow from geometry and ESP to $\widehat{\Delta E}_{\mathrm{em}}$ and $\hat{f}$.

Stage	Output shape	Summary
Coordinate scaling (fixed)	$\mathbb{R}^{N\times 3}$	Standardization of $\mathbf{R}$ (per channel).
Geometric features (inverse distances)	$\mathbb{R}^{P}$	All inverse distances $1/r_{ij}$ (flattened).
Constant normalization	$\mathbb{R}^{P}$	Fixed standardization of geometric features (training statistics).
Concatenate with ESP	$\mathbb{R}^{P+N}$	Concatenate normalized geometric features with per-atom ESP $\boldsymbol{\phi}$.
Head: $\mathrm{MLP}_{\mathrm{energy}} \rightarrow$ Dense(1)	$\mathbb{R}$	$\Delta\widehat{E}_{\mathrm{em}}$ (hidden activation configured; width/depth tuned; linear output).
Head: $\mathrm{MLP}_{\mathrm{osc}} \rightarrow$ Dense(1)	$\mathbb{R}$	$\hat{f}$ (hidden activation configured; width/depth tuned; linear output).
Outputs (concatenated)	$\mathbb{R}^{2}$	$[\Delta\widehat{E}_{\mathrm{em}}, \hat{f}]$.



(a) Force Model replacing QM (*NN/MM*). Inputs: inverse distances within the QM region and MM electrostatic potentials at QM atoms; outputs: $\hat{E}$ and $\hat{\mathbf{F}}$.



(b) Property model for spectrum prediction. Outputs: vertical emission energy $\Delta\hat{E}_{\text{em}}$ and oscillator strength $\hat{f}$.

Figure 10.: Schematic representation of environment sensitive neural network models within the fluorescence project: the force network and property network. Architectural details are in Fig. 11 and 12.

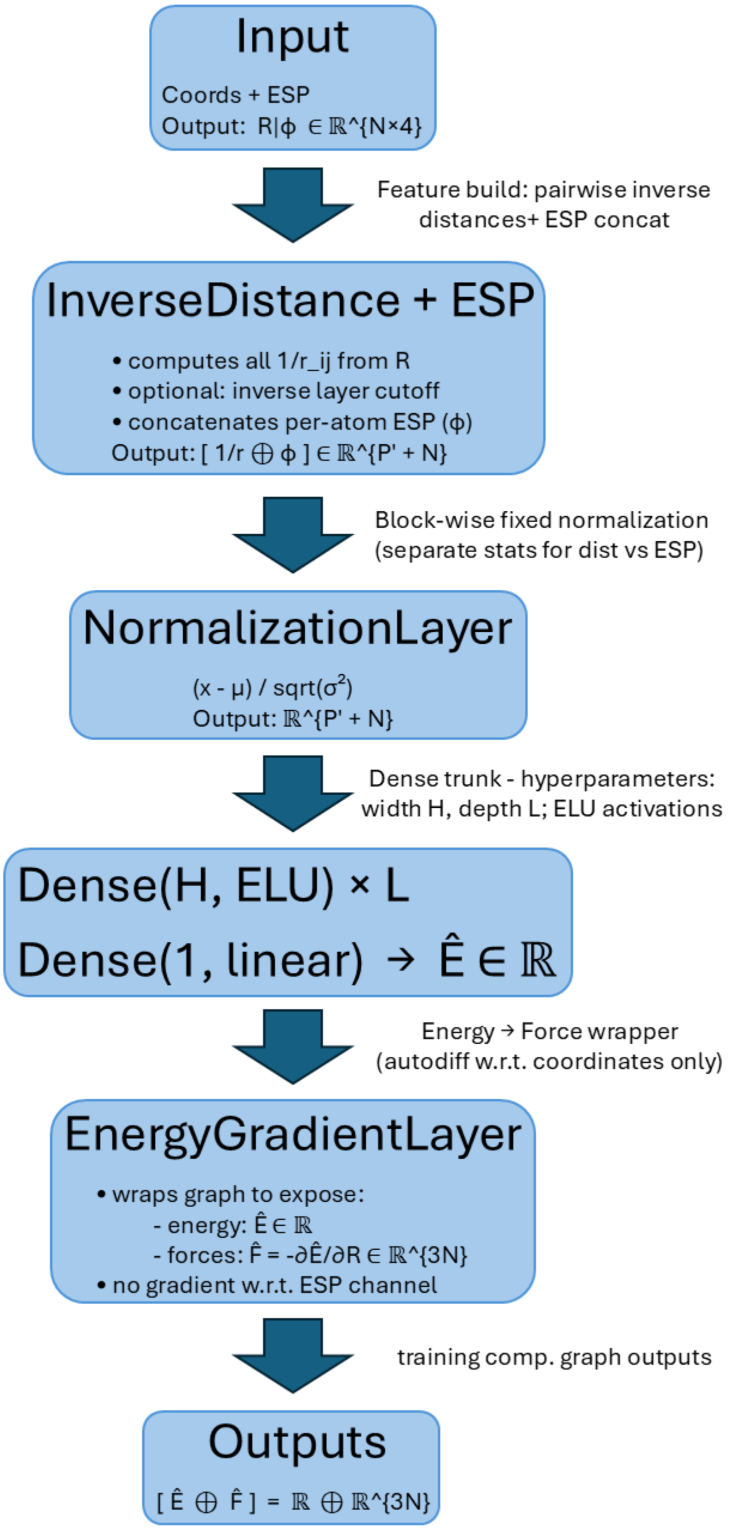


Figure 11.: Excited states force-network schematic (energy ⟶ forces by automatic differentiation). Inputs are Cartesian coordinates $\mathbf{R} \in \mathbb{R}^{N\times3}$ and per-atom ESP $\boldsymbol{\phi} \in \mathbb{R}^{N}$. A feature layer computes all pairwise inverse distances $1/r_{ij}$ (optionally masked dataset-wide to keep pairs above a cutoff) and concatenates $\boldsymbol{\phi}$ to yield a vector of length $P' + N$ with $P' \leq N(N-1)/2$. A fixed normalisation (separate statistics for distance vs. ESP block) is followed by an MLP trunk with L hidden layers of width H (ELU activations) and a linear output for the bright-state energy $\hat{E}$. The *EnergyGradientLayer* exposes forces $\hat{\mathbf{F}} = -\nabla_{\mathbf{R}}\hat{E}$ (no gradients w.r.t. the ESP channel), so the model outputs $[\hat{E} \ \oplus \ \hat{\mathbf{F}}] \in \mathbb{R} \oplus \mathbb{R}^{3N}$.

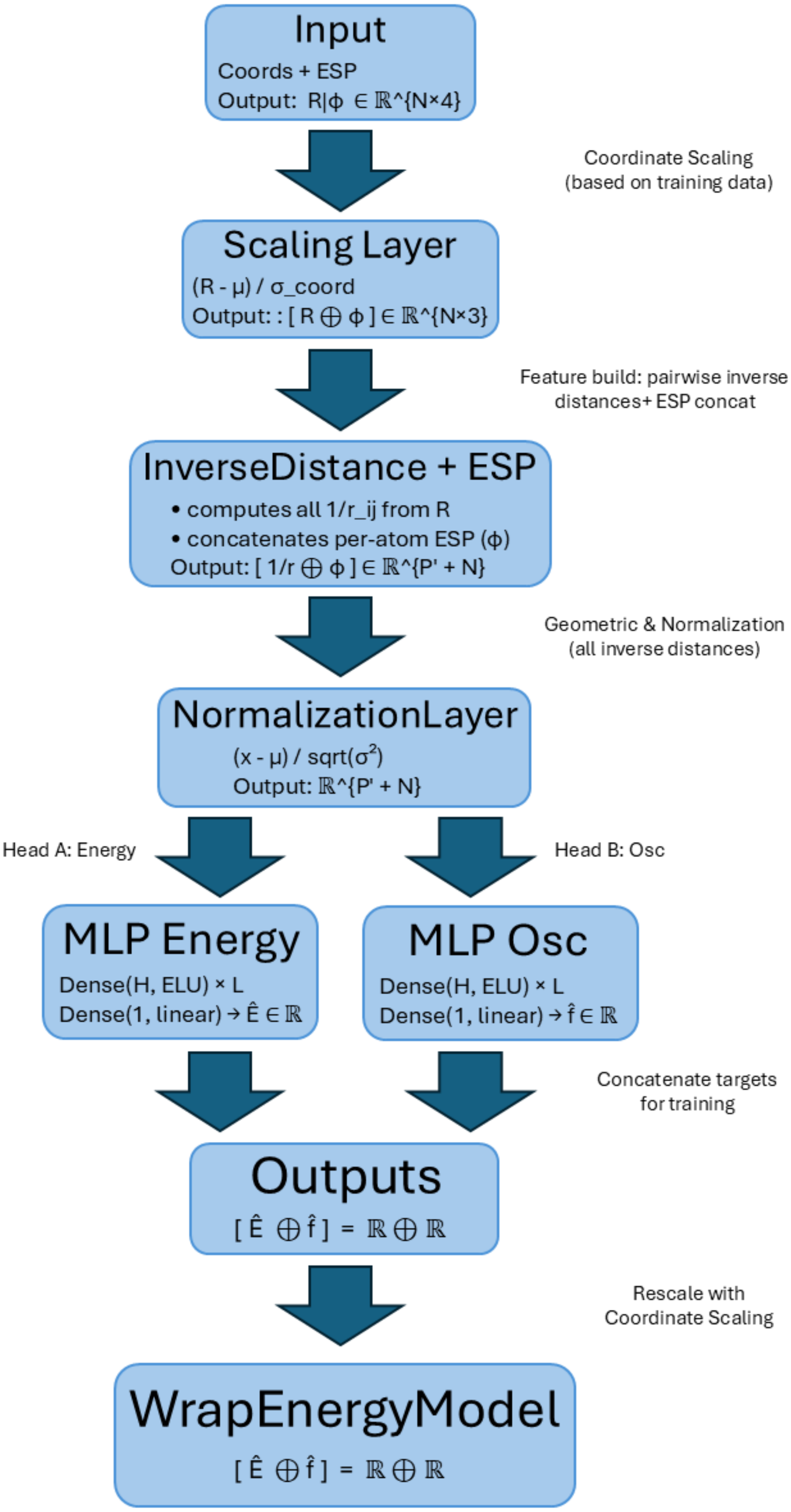


Figure 12.: Excited states property-network schematic for predicting bright-state vertical emission energy $\Delta\widehat{E}_{\mathrm{em}}$ and oscillator strength $\hat{f}$ from structure and ESP. Inputs are coordinates $\mathbf{R} \in \mathbb{R}^{N\times 3}$ and per-atom ESP $\boldsymbol{\phi} \in \mathbb{R}^{N}$. A coordinate *scaling layer* (fixed from training statistics) precedes a feature layer that computes all pairwise inverse distances $1/r_{ij}$ and concatenates $\boldsymbol{\phi}$ to yield a vector of length $P + N$ $(P = N(N-1)/2)$. A constant *normalization layer* standardizes the features. Two MLP heads (hidden ELU, linear output) map to $\Delta\widehat{E}_{\mathrm{em}}$ (Head A) and $\hat{f}$ (Head B); their outputs are concatenated to $[\Delta\widehat{E}_{\mathrm{em}}, \hat{f}]$. The *WrapEnergyModel* applies the inverse target scaling so that reported predictions are in physical units.

3.6.4. Training Protocols

Optimization and model selection. Training uses Adam with a cosine-decay restarts schedule. Hyperparameters (hidden width/depth, learning-rate schedule, batch size, and λ_E) are selected with Hyperband as described in the global Methods subsection *Hyperparameters and Optimization Strategies* (Sec. 3.2.3).[23,158,159] Final choices are made by a Pareto compromise between validation errors (energies, forces/properties) and NN/MM stability metrics (energy drift, temperature control), and documented in the SI.

3.6.5. Dataset Construction

Explicit-solvent configurations are generated by propagating TD-LC-DFTB/MM trajectories on the low-lying singlet surfaces S_1 and S_2, seeded from equilibrated MM frames; multiple short replicas broaden coverage near S_1/S_2 reorderings. For each stored geometry, vertical singlet excitations S_1–S_3 and their oscillator strengths $\{f_k\}$ are computed in the presence of the MM field. A stepwise, graphical workflow is shown in the Integration chapter (Sec. 4.3, Fig. 14).

A snapshot is tagged *bright* by selecting $k^\star = \arg\max_{k\in\{1,2,3\}} f_k$ and keeping it if $f_{k^\star} \geq f_{\min}$ (threshold specified in the fluorescence chapter). For retained frames we store state energy $E_{k^\star}$, forces $\mathbf{F}_{k^\star}$, vertical emission energy $\Delta E_{\mathrm{em}} = E_{k^\star} - E_{S_0}$, oscillator strength $f_{k^\star}$, and the per-atom MM electrostatic potential $\boldsymbol{\phi}$. For training this yields two task-specific tuples with identical inputs: *Force network:* $(\mathbf{R}, \boldsymbol{\phi};\ E_{k^\star}, \mathbf{F}_{k^\star})$. *Property network:* $(\mathbf{R}, \boldsymbol{\phi};\ \Delta E_{\mathrm{em}}, f_{k^\star})$. Otherwise the snapshot is discarded. This state-agnostic rule captures the LE/ICT manifold and avoids root-index artefacts.

Optional absorption set: from S_0 trajectories all snapshots are kept; the strongest vertical excitation (typically within S_1–S_3) is assigned and $(\mathbf{R}, \boldsymbol{\phi};\ \Delta E_{\mathrm{abs}}, f)$ recorded for the property model.

Datasets are stratified by solvent to preserve ESP statistics across polarities; inputs $(\mathbf{R}, \boldsymbol{\phi})$ are standardized per channel using training-set moments. Basic QC (geometry sanity, SCF convergence) is applied. *No enhanced/bias sampling is used in this work*; thus all frames carry unit statistical weight $w_i = 1$. If enhanced or biased sampling were employed, per-frame reweighting must be recorded (store w_i directly, or the bias potential and metadata to reconstruct it) and applied consistently in ensemble averages (e.g., spectrum construction), dataset statistics/standardization, and per-solvent distribution reports. As a run-level diagnostic, the dielectric constant ε_r of pure-solvent boxes is computed from dipole fluctuations (Eq. 2.3) and reported in Sec. 6.3.1.

3.6.6. NN/MM Runtime Coupling

During fluorescence simulations a neural network provides the *QM-region forces* directly, while the MM environment is treated classically. Energy conservation for the QM part follows from using the gradient of a scalar $\hat{E}$; ESP features are treated as external, non-differentiated inputs originating from the MM field.

Full replacement of QM. In the second strategy, the entire QM contribution to the energy and forces is replaced by a neural network potential. The ML model learns a direct mapping

$$E_{\mathrm{ML}}(\mathbf{R}, \mathbf{\Phi}) \approx E_{\mathrm{QM}}(\mathbf{R}, \mathbf{\Phi}), \qquad \mathbf{F}_{\mathrm{ML}}(\mathbf{R}, \mathbf{\Phi}) = -\nabla E_{\mathrm{ML}}(\mathbf{R}, \mathbf{\Phi}), \tag{3.69}$$

where $\mathbf{R}$ are the atomic coordinates of the QM region and $\mathbf{\Phi}$ the ESP acting on each QM atom. This approach eliminates QM calculations during simulation, enabling nanosecond trajectories at a cost comparable to classical MD.[173,191]

Integration into molecular dynamics. Both strategies provide forces at each MD time step to propagate the nuclei using standard integrators such as Velocity-Verlet

$$m_i \frac{\mathrm{d}^2 \mathbf{r}_i}{\mathrm{d}t^2} = \mathbf{F}_{i,\mathrm{ML}} + \mathbf{F}_{i,\mathrm{MM}}, \tag{3.70}$$

where $\mathbf{F}_{i,\mathrm{MM}}$ are classical contributions from the surrounding MM environment. Electrostatic interactions can be included via the electrostatic potential (ESP) from MM point charges as additional inputs to the ML model.[16,192]

Advantages and challenges. Replacing QM calculations with ML models dramatically accelerates simulations, enabling access to longer timescales and larger systems. Full replacement avoids any QM calls during propagation.[182] However, this approaches relies on training data that comprehensively cover the relevant configurational space. Transferability to new geometries or chemical environments remains a challenge, and special care must be taken to prevent energy drift or instabilities during extended simulations.[38,191]

Applications and generality. This ML/MD framework is not limited to specific chemical reactions or properties. Separate models can be trained for forces, energies, or excited-state observables (e.g., emission energies or oscillator strengths). The modular structure makes it broadly applicable to photochemistry, catalysis, and biomolecular systems.[16,193]

Notes. Implementation details (per-step hooks, energy bookkeeping, Eq solve) are documented in the Integration chapter IV; the charge-equilibration formulation is summarized in Sec. 3.4.3.

3.7. Sampling Strategies

This section summarizes how we generate reference configurations. Two points up front:

(i) Production use in the phosphoryl-transfer project. All quantitative results reported for the phosphorus study rely on *(a)* unbiased QM/MM ("free simulation") and *(b)* standard well-tempered metadynamics with *hand-picked* distance collective variables, namely the P–O (leaving group) and P–N (nucleophile) distances, which are conventional for SN2(P)-like pathways. These choices keep the setup transparent and targeted at the reaction coordinate of interest.

(ii) Exploratory methods for benchmarking only. Additional strategies – normal-mode displacements (NMA), PCA-biased metadynamics, RMSD-biased metadynamics, and uncertainty – guided sampling-are *explored* later in the dedicated Sampling chapter to assess coverage and efficiency. Unless stated otherwise, data generated by these exploratory variants are *not* used in the production models for the phosphoryl-transfer results.

The remainder of this section briefly documents each generator; detailed performance comparisons appear in the Sampling chapter.

3.7.1. Data Sampling

Training data must cover the configurational and energetic regions that control the target observables and the model's generalization in condensed phase.[25,117] We employ complementary generators:

- **Unbiased QM/MM ("free simulation"):** thermally accessible configurations from the Boltzmann distribution (Sec. 3.7.2).
- **Normal-mode displacements (NMA):** cheap, controlled off-equilibrium distortions around minima (Sec. 3.7.3).[84]
- **Metadynamics along 1D CVs:** history-dependent bias to accelerate slow coordinates. Here we *explore* two generic CVs—PCA projection (Sec. 3.7.4) and RMSD (Sec. 3.7.5)—while the *production* phosphoryl-transfer runs use hand-crafted P–O and P–N distance CVs.
- **Uncertainty-guided sampling:** active-learning loop that steers toward out-of-distribution regions (Sec. 3.7.6).

Metadynamics in one line. Metadynamics augments the potential by depositing Gaussians along a chosen CV $s(\mathbf{r})$, with well-tempered tempering to ensure convergence and permit free-energy reconstruction.[77,78,188]

3.7.2. Free-Simulation Sampling

In this chapter, *free simulation* denotes an **unbiased** QM/MM trajectory (no added bias potential) in explicit solvent; the theoretical background and force evaluation follow Sections 2.5 and 2.6. Concretely, the chromophore is treated quantum-mechanically and embedded in a classical environment via electrostatic embedding, while the solvent is propagated with a standard force field in periodic boundary conditions.[1,130] For efficiency, the QM region is described with self-consistent-charge DFTB3 using an appropriate parameterization, whereas the MM part is integrated with GROMACS in the target ensemble (NVT or NPT) using the algorithms specified in Section 2.5; the resulting snapshots (geometries, QM energies, and forces) constitute the baseline dataset.[121,129] For the nonreactive chromophores considered here (absorption/emission observables), the adequacy of free simulation is governed less by exceeding chemical barriers than by the correlation times of relevant degrees of freedom (torsions, ring motions, solvent reorganization) and by total trajectory length; enhanced-sampling methods below are therefore used not to model reactions, but to broaden configurational and environmental coverage and to generate diverse off-equilibrium geometries for training and stress-testing ML models.[25,117]

In the phosphoryl-transfer project, this unbiased baseline is complemented by well-tempered metadynamics with hand-picked distance CVs ($d_{\mathrm{P-O}}$, $d_{\mathrm{P-N}}$) to target barrier crossing along the reaction coordinate, following standard metadynamics practice.[77,78,188]

3.7.3. NMA Off-Equilibrium Generator

Near a local minimum $\mathbf{r}_0$, the potential is approximated harmonically as

$$V(\mathbf{r}) \;\approx\; V(\mathbf{r}_0) + \tfrac{1}{2}\,(\mathbf{r}-\mathbf{r}_0)^{\top}\mathbf{H}^{(\mathrm{cart})}(\mathbf{r}-\mathbf{r}_0), \tag{3.71}$$

with the Cartesian Hessian $\mathbf{H}^{(\mathrm{cart})}$ whose elements are $\partial^2 V/\partial r_i \partial r_j$ at $\mathbf{r}_0$.[84] Introducing mass-weighted coordinates $\mathbf{q} = \mathbf{M}^{1/2}(\mathbf{r}-\mathbf{r}_0)$ and the *mass-weighted Hessian*

$$\mathbf{H} \;=\; \mathbf{M}^{-1/2}\,\mathbf{H}^{(\mathrm{cart})}\,\mathbf{M}^{-1/2}, \tag{3.72}$$

Newton's equations linearize to $\ddot{\mathbf{q}} = -\,\mathbf{H}\,\mathbf{q}$, which is the standard starting point for normal-mode analysis.[84,194] Diagonalization solves the (symmetric) eigenproblem

$$\mathbf{H}\,\mathbf{L} \;=\; \mathbf{L}\,\boldsymbol{\Omega}^2, \tag{3.73}$$

where the columns $\mathbf{l}_k$ of $\mathbf{L}$ are the *mass-weighted orthonormal eigenvectors* (normal-mode directions) and $\boldsymbol{\Omega}^2 = \mathrm{diag}(\omega_1^2, \ldots, \omega_{3N}^2)$ collects the squared angular frequencies.[84] With the standard normalization

$$\mathbf{L}^{\top}\mathbf{L} = \mathbf{I}, \qquad \mathbf{L}^{\top}\mathbf{H}\,\mathbf{L} = \boldsymbol{\Omega}^2, \tag{3.74}$$

the *normal-mode coordinates* $\mathbf{Q} = \mathbf{L}^{\top}\mathbf{q}$ obey the decoupled equations of motion

$$\ddot{\mathbf{Q}} + \boldsymbol{\Omega}^2\mathbf{Q} = \mathbf{0} \iff \ddot{Q}_k + \omega_k^2 Q_k = 0 \quad (k = 1, \ldots, 3N), \tag{3.75}$$

i.e., each mode behaves as a harmonic oscillator with frequency ω_k.[84,194,195] For a nonlinear molecule, six eigenvalues correspond to overall translations/rotations and are (ideally) zero; the remaining $3N - 6$ positive eigenvalues define the vibrational modes.[84]

For sampling, a displacement along a single mode k with mass-weighted amplitude α is mapped back to Cartesian coordinates as

$$\mathbf{r}(\alpha; k) \;=\; \mathbf{r}_0 \;+\; \alpha\, \mathbf{M}^{-1/2}\, \mathbf{l}_k, \tag{3.76}$$

and superpositions follow by linearity.[84] In the classical harmonic limit, the thermal variance of Q_k is

$$\langle Q_k^2 \rangle \;=\; \frac{k_\mathrm{B} T}{\omega_k^2}, \tag{3.77}$$

while the quantum expression is $\langle Q_k^2 \rangle = \frac{\hbar}{2\omega_k} \coth\big(\frac{\hbar\omega_k}{2k_\mathrm{B}T}\big)$.[84] Low-frequency modes typically encode collective motions; high-frequency modes capture localized distortions, which motivates using bounded amplitudes to remain within the harmonic basin before recomputing energies/forces for displaced structures at the chosen electronic level.[194,195]

Scope. This subsection documents an *exploratory* generator used for coverage/benchmarking only; it is *not* used for the production phosphoryl-transfer or fluorescence properties models reported in this thesis.

3.7.4. PCA-Biased Metadynamics

Principal component analysis (PCA) identifies orthogonal directions of maximal positional variance (“essential dynamics”) from an unbiased reference trajectory.[196,197] Let $\mathbf{r}(t) \in \mathbb{R}^{3N}$ be the Cartesian coordinates of the solute at time t. To remove overall translation and rotation, each frame is optimally superposed to a reference structure $\mathbf{r}_\mathrm{ref}$ (optimal superposition).[198] With centered displacements $\mathbf{u}(t) = \mathbf{r}(t) - \langle \mathbf{r} \rangle$, where $\langle \cdot \rangle$ denotes the time average over the aligned trajectory, the covariance matrix

$$\mathbf{C} \;=\; \left\langle \mathbf{u}\,\mathbf{u}^\top \right\rangle \;\in\; \mathbb{R}^{3N\times 3N} \tag{3.78}$$

is diagonalized as

$$\mathbf{C}\,\mathbf{P} \;=\; \mathbf{P}\,\boldsymbol{\Lambda}, \qquad \mathbf{P}^\top\mathbf{P} = \mathbf{I}, \tag{3.79}$$

yielding principal axes (columns $\mathbf{p}_k$ of $\mathbf{P}$) and variances $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1 \geq \lambda_2 \geq \dots)$. The projection onto PC1 defines a one–dimensional collective variable (CV)

$$s(\mathbf{r}) \;=\; \mathbf{p}_1^\top \big(\mathbf{r} - \langle \mathbf{r} \rangle\big), \tag{3.80}$$

which we bias in metadynamics. (Optionally, a mass-weighted PCA uses $\tilde{\mathbf{u}} = \mathbf{M}^{1/2}\mathbf{u}$, leading to $\mathbf{C}_\mathrm{mw} = \langle \tilde{\mathbf{u}}\tilde{\mathbf{u}}^\top \rangle$ and $s(\mathbf{r}) = \mathbf{p}_1^\top \mathbf{M}^{1/2}(\mathbf{r} - \langle \mathbf{r} \rangle)$.[196,197])

Metadynamics augments the potential with a history-dependent bias in CV space.[188] Depositing Gaussian kernels along the visited s values gives

$$V_{\text{bias}}(s,t) \;=\; \sum_{t'\le t} w(t')\,\exp\!\left[-\frac{\big(s-s(\mathbf{r}(t'))\big)^2}{2\sigma^2}\right], \tag{3.81}$$

with kernel width σ and height $w(t')$. In well-tempered metadynamics the kernel height is tempered by the accumulated bias with factor $\gamma > 1$, ensuring smooth convergence; the free energy along s follows asymptotically from the bias as[78]

$$F(s) \;\approx\; -\frac{\gamma}{\gamma-1}\,V_{\text{bias}}(s,t\!\to\!\infty) \;+\; \text{const.} \tag{3.82}$$

All simulations in this work use the standard implementation in PLUMED.[77]

Scope. PCA-biased metadynamics is evaluated in the Sampling chapter as an exploratory variant to broaden coverage; it is not used for production data in the phosphoryl-transfer study, where hand-picked distance CVs are employed.

3.7.5. RMSD-Biased Metadynamics

We use the root–mean–square deviation (RMSD) to a single, pre-optimized reference geometry as a one–dimensional collective variable. For Cartesian coordinates $\mathbf{r}(t) = (\mathbf{r}_1,\ldots,\mathbf{r}_N) \in \mathbb{R}^{3N}$ and a fixed reference $\mathbf{r}^{\text{ref}}$, the RMSD after optimal rigid-body superposition is

$$\text{RMSD}(\mathbf{r}) \;=\; \sqrt{\frac{1}{N}\sum_{i=1}^{N}\big|\,\mathbf{r}_i-\mathbf{R}\,\mathbf{r}_i^{\text{ref}}-\mathbf{t}\,\big|^2}, \tag{3.83}$$

where the rotation $\mathbf{R}$ and translation $\mathbf{t}$ minimize the least-squares fit (closed-form quaternion solution).[198] Here $|\cdot|$ denotes the Euclidean norm. Setting $s(\mathbf{r}) = \text{RMSD}(\mathbf{r})$, metadynamics augments the potential with a history-dependent bias in s-space to discourage revisits of explored regions,[188] and its well-tempered variant ensures smooth convergence and permits free-energy reconstruction from the asymptotic bias,[78] as implemented in PLUMED.[77]

Usage in this thesis. RMSD-biased metadynamics is included for exploration/benchmarking only; production phosphoryl-transfer results use distance CVs (P–O, P–N).

3.7.6. Uncertainty-Guided Sampling

Uncertainty-guided sampling integrates an interatomic neural-network potential with an on-the-fly uncertainty estimator and an enhanced-sampling driver in a closed active-learning loop. The aim is to turn the model's *epistemic* (data-driven ignorance, as opposed to aleatoric noise) uncertainty into a scalar collective variable (CV; a smooth function

$\xi(\mathbf{x})$ of atomic coordinates that can be biased) which steers dynamics toward out-of-distribution (OOD) regions, from which new reference labels are harvested to improve the model iteratively.[199–202] A typical cycle begins from a small *seed* of labeled configurations (energies and forces) with which an initial equivariant message-passing potential (here: MACE) is trained.[39] Seed structures are curated from short free trajectories, labeled at the target electronic level (DFTB2), and used to train MACE with forces and energies; DFTB details follow the standard SCC-DFTB/DFTB3 literature.[10,11,105,203] The trained network then drives enhanced sampling, while an uncertainty score $\xi(\mathbf{x})$ is computed online for each configuration $\mathbf{x}$ to decide which frames to label and when to update the model.[204–206]

Uncertainty score. The estimator is defined in the network's latent space: for each configuration $\mathbf{x}$, a fixed embedding $\mathbf{z}(\mathbf{x})$ is extracted from a designated MACE layer. A K-component Gaussian mixture model (GMM; weights w_k, means $\boldsymbol{\mu}_k$, full covariances Σ_k, with $\sum_k w_k = 1$, $w_k \geq 0$) is fitted via EM to embeddings of the current training pool.[199] Uncertainty is the negative log-likelihood

$$\xi(\mathbf{x}) \;=\; -\,\log\!\left(\sum_{k=1}^{K} w_k\,\mathcal{N}\!\big(\mathbf{z}(\mathbf{x})\,\big|\,\boldsymbol{\mu}_k,\,\Sigma_k\big)\right). \tag{3.84}$$

Embeddings that lie in low-density regions under the training distribution receive large ξ and are considered novel. Writing the soft responsibilities

$$\gamma_k(\mathbf{z}) = \frac{w_k\,\mathcal{N}(\mathbf{z}\mid\boldsymbol{\mu}_k,\Sigma_k)}{\sum_j w_j\,\mathcal{N}(\mathbf{z}\mid\boldsymbol{\mu}_j,\Sigma_j)},$$

one obtains

$$\nabla_{\mathbf{x}}\xi(\mathbf{x}) \;=\; \Big(\nabla_{\mathbf{x}}\mathbf{z}(\mathbf{x})\Big)^{\top}\sum_{k=1}^{K}\gamma_k(\mathbf{z})\,\Sigma_k^{-1}\big(\mathbf{z}-\boldsymbol{\mu}_k\big), \tag{3.85}$$

computed by automatic differentiation (AD) through $\mathbf{z}(\mathbf{x})$; a finite-difference fallback is possible when AD is unavailable.[199] Intuitively, the negative gradient of Eq. (3.85) points from the embedding toward nearby mixture centers, weighted by their soft memberships, and is mapped back to coordinate space by $\nabla_{\mathbf{x}}\mathbf{z}$.

Bias potential used here (eABF + GaMD, implemented in ASE). To *use* uncertainty as a steering coordinate, $\xi(\mathbf{x})$ is treated as a 1D CV coupled to an extended-variable adaptive-biasing scheme and augmented by a GaMD boost; all bias terms and bookkeeping are implemented with *ASE* hooks (no PLUMED) to keep the CV, its gradient, and the bias *in-Python* next to the NN, enabling end-to-end autodiff for Eq. (3.85) and minimizing external dependencies.[207] The total potential during sampling is

$$V_{\mathrm{tot}}(\mathbf{x},\lambda,t) = V(\mathbf{x}) + \tfrac{1}{2}\kappa\big(\xi(\mathbf{x})-\lambda\big)^2 + B_{\mathrm{ABF}}(\lambda,t) + \Delta V_{\mathrm{GaMD}}(\mathbf{x}), \tag{3.86}$$

where λ is the extended variable and $\kappa = 1/(\beta\sigma^2)$ sets the eABF coupling width.[204,205] The GaMD boost has the standard piecewise-harmonic form[206]

$$\Delta V_{\mathrm{GaMD}}(\mathbf{x}) \;=\; \begin{cases} \frac{1}{2}\,k\,\big(E - V(\mathbf{x})\big)^2, & V(\mathbf{x}) < E, \\ 0, & V(\mathbf{x}) \geq E, \end{cases} \tag{3.87}$$

with (E, k) determined from short unbiased energy statistics (min/mean/max) to keep the boost smooth and reweightable.[206] The ABF contribution $B_{\mathrm{ABF}}(\lambda, t)$ integrates the running mean force along the CV; its derivative satisfies

$$\frac{\partial B_{\mathrm{ABF}}}{\partial \lambda}(\lambda, t) \;=\; \big\langle F_\xi \big\rangle_{\lambda,t} \;=\; \frac{1}{N_\lambda(t)} \sum_{i:\,\xi(\mathbf{x}_i)\approx\lambda} \mathbf{F}(\mathbf{x}_i)\cdot \frac{\nabla_{\mathbf{x}}\xi(\mathbf{x}_i)}{\big\|\nabla_{\mathbf{x}}\xi(\mathbf{x}_i)\big\|^2}, \tag{3.88}$$

so that eABF cancels the mean free-energy gradient along ξ, while GaMD helps cross barriers in orthogonal directions.[204,206]

Online screening and retraining. Frames are screened by two criteria: (i) an *adaptive* uncertainty gate $\xi_{\mathrm{threshold}}$ (e.g., a running upper quantile) that tightens generation-by-generation to keep exploration focused on genuinely novel regions and follow the model's improving coverage,[199] and (ii) an adaptive cosine-distance filter in latent space to enforce diversity among candidates.[199] Minimal physical sanity checks (distance bounds) and a short FIRE relaxation discard broken structures.[208] Accepted candidates are *labeled* at the target level (self-consistent DFTB2 with `ob2-1-1`),[10,11,105,203] added to the pool, and the surrogate is retrained. The active-learning loop (biased sampling → screening → labeling → retraining) repeats until uncertainty and validation errors plateau or a budget is reached.[199]

Scope and limitations. The uncertainty-guided setup is included to broaden configuration-space coverage; it is *not* used in the production phosphoryl-transfer or fluorescence results. Limitations include the cost of repeated retraining/labeling, possible calibration drift of the GMM as the embedding changes (mitigated by refitting after each retrain), sensitivity to the choice of embedding layer and K, and the need for consistent reweighting if ensemble observables are estimated from biased trajectories.[205,206]

3.7.7. Sampling Diagnostics and Visualization

This subsection defines the geometric diagnostics we use across chapters to assess how different sampling strategies populate configuration space. We focus on (i) torsional coverage in 1D/2D and (ii) planarity for quasi-planar chromophores.

Planarity for quasi-planar molecules. Several systems (FR0, 4AP, ACDAN, Fig. 80) are expected to remain largely planar in their chromophoric cores. We therefore monitor *out-of-plane deviations* of selected atoms relative to a reference plane on the aromatic ring. Choose a *reference atom* C^* on the benzenoid ring and two second-nearest ring carbons B and C (two bonds away). Define the unit bond vectors $\hat{\mathbf{u}}_1 = (\mathbf{r}_B - \mathbf{r}_{C^*})/\|\mathbf{r}_B - \mathbf{r}_{C^*}\|$ and $\hat{\mathbf{u}}_2 = (\mathbf{r}_C - \mathbf{r}_{C^*})/\|\mathbf{r}_C - \mathbf{r}_{C^*}\|$, and the *unit* plane normal

$$\hat{\mathbf{n}} \;=\; \frac{\hat{\mathbf{u}}_1 \times \hat{\mathbf{u}}_2}{\|\hat{\mathbf{u}}_1 \times \hat{\mathbf{u}}_2\|}. \tag{3.89}$$

For any monitored atom X (e.g., heteroatoms adjacent to the ring), form the *unit* vector $\hat{\mathbf{v}}_X = (\mathbf{r}_X - \mathbf{r}_{C^*})/\|\mathbf{r}_X - \mathbf{r}_{C^*}\|$. The out-of-plane angle is the complement of the angle between $\hat{\mathbf{n}}$ and $\hat{\mathbf{v}}_X$,

$$\alpha_X \;=\; 90^\circ - \arccos\big(|\hat{\mathbf{n}}\cdot\hat{\mathbf{v}}_X|\big), \tag{3.90}$$

so $\alpha_X = 0^\circ$ means that X lies in the plane, and larger values quantify buckling/pyramidalisation. All vectors are normalized, hence $|\hat{\mathbf{n}} \cdot \hat{\mathbf{v}}_X| \in [0, 1]$ and the sign of $\hat{\mathbf{n}}$ is immaterial.

Backbone planarity for polyenes (Retinal). For long, nearly planar backbones it is impractical to track many individual atoms. We instead build *local plane normals* along the chain and aggregate them into a single planarity score.

Let $\mathbf{r}_1, \dots, \mathbf{r}_M$ be the ordered heavy-atom coordinates of the polyene backbone and

$$\hat{\mathbf{b}}_i = \frac{\mathbf{r}_{i+1} - \mathbf{r}_i}{\|\mathbf{r}_{i+1} - \mathbf{r}_i\|} \qquad (i = 1, \dots, M-1) \tag{3.91}$$

the unit bond directions. For each inner index $i = 2, \dots, M-1$ define

$$\gamma_i = \big\|\hat{\mathbf{b}}_{i-1} \times \hat{\mathbf{b}}_i\big\| = \sin\angle(\hat{\mathbf{b}}_{i-1}, \hat{\mathbf{b}}_i), \qquad \hat{\mathbf{n}}_i = \frac{\hat{\mathbf{b}}_{i-1} \times \hat{\mathbf{b}}_i}{\gamma_i}. \tag{3.92}$$

Sites with nearly collinear bonds ($\gamma_i < \tau$, e.g. $\tau \in [0.05, 0.1]$) have ill-defined planes and are omitted. If K normals remain, a *global planarity score* in $[0, 1]$ is

$$S_{\text{glob}} \;=\; \frac{2}{K(K-1)} \sum_{\substack{i,j\in\mathcal{I}\\ i<j}} \big|\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_j\big|, \tag{3.93}$$

where $\mathcal{I}$ is the index set of valid normals. The normalisation $2/[K(K-1)]$ is the reciprocal of the number of unordered pairs $\binom{K}{2}$, so S_{glob} is the *average* pairwise alignment. The absolute value removes the up/down ambiguity of plane normals.

For an angle-type readout we report the *effective misalignment*

$$\theta_{\text{glob}} = \arccos\big(S_{\text{glob}}\big) \in [0^\circ, 90^\circ], \tag{3.94}$$

so $\theta_{\text{glob}} = 0^\circ$ denotes a highly planar backbone. For small spreads (in radians) θ_{glob} behaves like an RMS of the pairwise misalignment angles, since $\cos\delta \approx 1 - \frac{1}{2}\delta^2$ implies $\arccos(\langle\cos\delta\rangle) \approx \sqrt{\langle\delta^2\rangle}$.

2D torsional coverage (Ramachandran-style maps). To visualise coupled soft modes we plot hex-binned occupancies in the (ϕ, ψ) plane, where each dihedral is computed from four atoms and wrapped to $[-180°, 180°]$. Binning is periodic in both directions; color encodes either counts or density (counts divided by the total number of frames). These maps quickly reveal basin connectivity, bridging between basins, and sparse excursions.

1D torsion distributions. When a single dihedral is diagnostic (e.g., a formyl twist), we report the 1D distribution as a density-normalised histogram on $(-180°, 180°]$ with periodic binning. Kernel density estimates may be overlaid for smoother trends but are not required.

Energy and force summaries. For each dataset we summarise per-frame total energies and force magnitudes by boxplots. Energies are in $\mathrm{kcal\,mol^{-1}}$; force magnitudes are $\|\mathbf{F}\|$ in $\mathrm{kcal\,mol^{-1}\,\AA^{-1}}$. These plots highlight heavy tails and strained outliers (useful for spotting unstable sampling such as extreme NMA perturbations).

Implementation notes. Angles are computed from Cartesian coordinates in the simulation frame; periodic wrapping is applied before binning. When comparing methods we keep bin sizes and normalisation (counts vs. density) identical across panels. For the polyene planarity scores we ignore near-linear sites using a small threshold τ and use $w_i = \gamma_i$ (or γ_i^2) to stabilize the average; absolute dot products or one-time sign alignment make the metrics orientation-invariant.

Part IV

Computational Workflow & System Integration

4.1. Overview

This chapter gives a schematic, end-to-end view of the two pipelines used throughout the dissertation – phosphorylation (Sec. 4.2) and fluorescence (Sec. 4.3). A substantial part of the work consisted in *engineering, implementing, and stabilizing* these workflows so that complex QM/MM sampling, dataset construction, model training, and deployment can be run reproducibly and with minimal manual intervention. The pipelines are modular (swappable backends such as HDNNP/MACE, DFT/DFTB), scriptable (batch execution with consistent manifests and quality control), and shared in a form that collaborators could use without modification.

At a high level, the phosphorylation pipeline converts QM/MM metadynamics into Δ-learning datasets (difference of atomization energies) and couples the trained model back into Δ-QM/MM dynamics (Fig. 13). The fluorescence pipeline prepares solvent/chromophore systems, launches parallel QM/MM trajectories, assembles bright-state emission/absorption datasets, trains a force model (for NN/MM) and a property model, and constructs spectra via NEA/E-ZTFC (Fig. 14). Details of electronic-structure settings and network architectures are deferred to Methods I/II; here the focus is the integration logic and data flow.

4.2. Phosphorylation

QM/MM metadynamics trajectories are converted into Δ-learning datasets and the trained model is coupled back into dynamics. Parameters are deferred; this section only describes the flow that corresponds to Fig. 13.

1. **QM/MM metadynamics (GROMACS[121] + PLUMED[77]).** Well-tempered metadynamics in explicit solvent is run for several nanoseconds (typically ∼5–10 ns) using the forming P–N and breaking P–O distances as collective variables. *Outputs:* trajectory frames, CV traces, and a free-energy surface (FES).

2. **Generate geometries (cut-out & uniform sampling).** The minimum-free-energy path (MFEP) is cut out from the FES (with small plateaus around it), and frames are drawn *uniformly* along this polyline. In practice ∼12 000 frames are drawn to end up with ∼10 000 converged labels per reaction after quality control. (Examples of MFEP cut-outs and FES are shown in the Phosphorylation chapter.)

3. **DFT & DFTB single points.** For each geometry, gas-phase single points are computed at the reference DFT level and at DFTB3 (same QM cut as during sampling). For DFTB, a finite Fermi temperature (300 K) is used to aid SCF convergence (fractional occupations); 500 K was tested but not required. *Outputs:* per-frame energies (and forces if available) at both levels and convergence diagnostics.

4. **Training data.** For every retained geometry, Δ-labels are formed: $\Delta E = E_{\mathrm{DFT}} - E_{\mathrm{DFTB}}$ and, where available, $\Delta \mathbf{F} = \mathbf{F}_{\mathrm{DFT}} - \mathbf{F}_{\mathrm{DFTB}}$. *Outputs:* tables of ($\mathbf{R}$; ΔE, $\Delta \mathbf{F}$) plus quality control flags.

5. **Convert to atomization energies.** Both DFT and DFTB totals are converted to atomization energies by subtracting level-consistent atomic references; the supervised target is the *difference of atomization energies* $\Delta E = E^{\mathrm{DFT}}_{\mathrm{atom}} - E^{\mathrm{DFTB}}_{\mathrm{atom}}$ (cf. Methods II). *Outputs:* standardized inputs/targets for NN training.

6. **Neural-network training.** Charge-stratified Δ-models are trained; HDNNP is used as the primary backend, with a MACE/Δ view available for cross-checks (same splits; see Methods II for architecture and splits). *Outputs:* final checkpoints selected by a Pareto trade-off between validation error and NN/MM stability.

7. **Δ-QM/MM simulation.** During production dynamics the learned correction augments the baseline: $E^{\mathrm{tot}} = E_{\mathrm{DFTB}} + E_{\mathrm{NN}}$ and $\mathbf{F}^{\mathrm{tot}} = \mathbf{F}_{\mathrm{DFTB}} + \mathbf{F}_{\mathrm{NN}}$, yielding on-the-fly corrected trajectories (see Phosphorylation chapter for results).

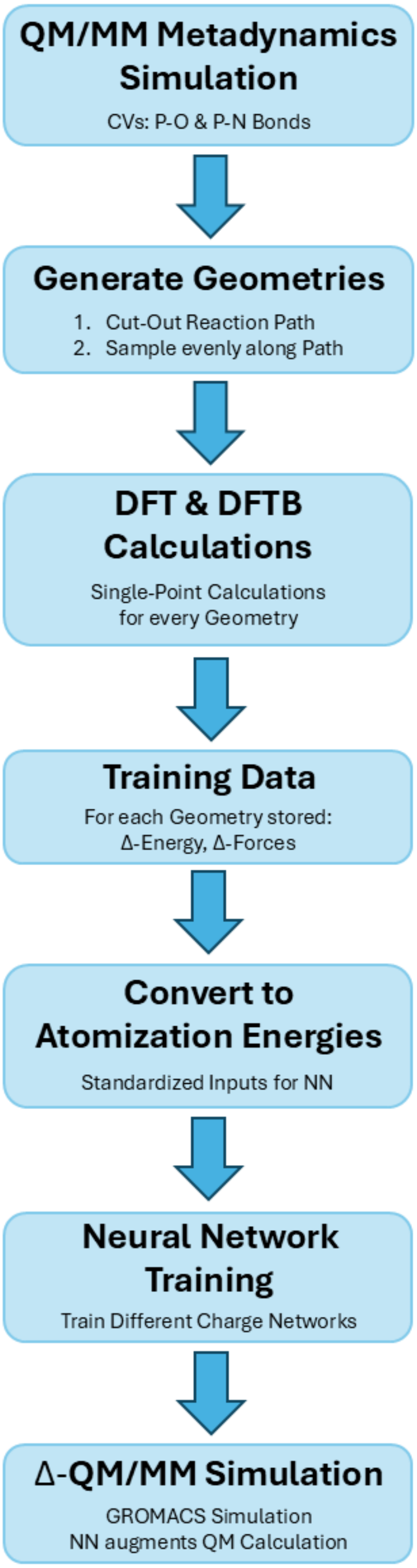


Figure 13.: Schematic phosphorylation pipeline. (1) QM/MM metadynamics in explicit solvent (CVs: P–O and P–N); (2) MFEP cut-out and uniform sampling; (3) DFT and DFTB single points; (4) assemble Δ-energy/force training data; (5) convert to atomization energies; (6) train Δ-models (HDNNP; MACE view); (7) on-the-fly corrected Δ-QM/MM dynamics.

4.3. Fluorescence

This section summarizes the end-to-end workflow (Fig. 14) used in the fluorescence project. It is schematic by design; electronic-structure specifics (TD-LC-DFTB/Casida, state tracking) are given in Methods I, and model/label details in Methods II (Secs. 2.4.4, 3.6).

1. **Force-Field Setup (Solvent & Chromophore).** Obtain or build MM parameters for each solvent and the chromophore, adapting or deriving missing terms as needed. *Outputs:* solvated topologies/parameters for all solvent systems.
2. **Simulation Preparation (GROMACS).** Place the chromophore in the chosen solvent box and run minimization, NVT and NPT equilibration (electrostatic embedding for later QM/MM). *Outputs:* equilibrated MM systems ready for production.
3. **MM Production & Seed Selection.** Run a MM production trajectory and draw N random seed geometries. *Purpose:* spawn many independent QM/MM runs in parallel.
4. **Parallel QM/MM Free Dynamics.** For each seed, launch a free QM/MM trajectory in **(i) ground state** S_0 for absorption datasets, or **(ii) excited states** for fluorescence, typically propagating both S_1 and S_2. TD-LC-DFTB (Casida) provides state energies and identities along the runs. *Outputs:* N trajectories per solvent (for S_0 and/or $S_{1,2}$).
5. **Concatenation & Single-Point Evaluation.** Concatenate the N runs into one ensemble per solvent/state family. For *every* stored geometry, compute single-point properties for S_0–S_3 (state energies, vertical gaps, oscillator strengths, and forces) at the same TD-LC-DFTB level. *Outputs:* per-geometry tables for S_0–S_3.
6. **Bright-state Selection & Dataset Assembly.** *Fluorescence:* From the S_1/S_2 sampling, for each snapshot choose the brightest among S_1–S_3; keep it if $f \geq f_{\text{min}}$ and record $(E, \mathbf{F}, \Delta E_{\text{em}}, f, \boldsymbol{\phi})$. *Absorption (optional):* From S_0 runs keep all snapshots, assign the strongest vertical excitation (typically within S_1–S_3), and record $(\Delta E_{\text{abs}}, f, \boldsymbol{\phi})$. *Output:* two datasets with identical features (coordinates + per-atom ESP $\boldsymbol{\phi}$).
7. **Model Training.** *Force Model (QM Replacement):* learn a scalar bright-state energy $\hat{E}(\mathbf{R}, \boldsymbol{\phi})$; forces are obtained as $\hat{\mathbf{F}} = -\nabla_{\mathbf{R}}\hat{E}$. *Property Model (Spectrum):* learn $(\Delta\hat{E}_{\text{em}}, \hat{f})$ on the emission dataset (and analogously $(\Delta\hat{E}_{\text{abs}}, \hat{f})$ on the absorption dataset). *Outputs:* trained checkpoints selected on validation and NN/MM stability diagnostics.
8. **NN/MM Deployment.** In production MD the force model replaces the QM calculation and provides the QM-region forces on-the-fly; the ESP from the MM field enters as an input. The property model is queried along the trajectory for $(\Delta E_{\text{em}}, f)$.
9. **Spectrum Construction.** From the MD snapshots and property predictions, assemble spectra using the Nuclear Ensemble Approach (NEA) and/or E-ZTFC as described in Sec. 2.3.5. For absorption, the ground-state dataset and the property model suffice without NN/MM deployment.

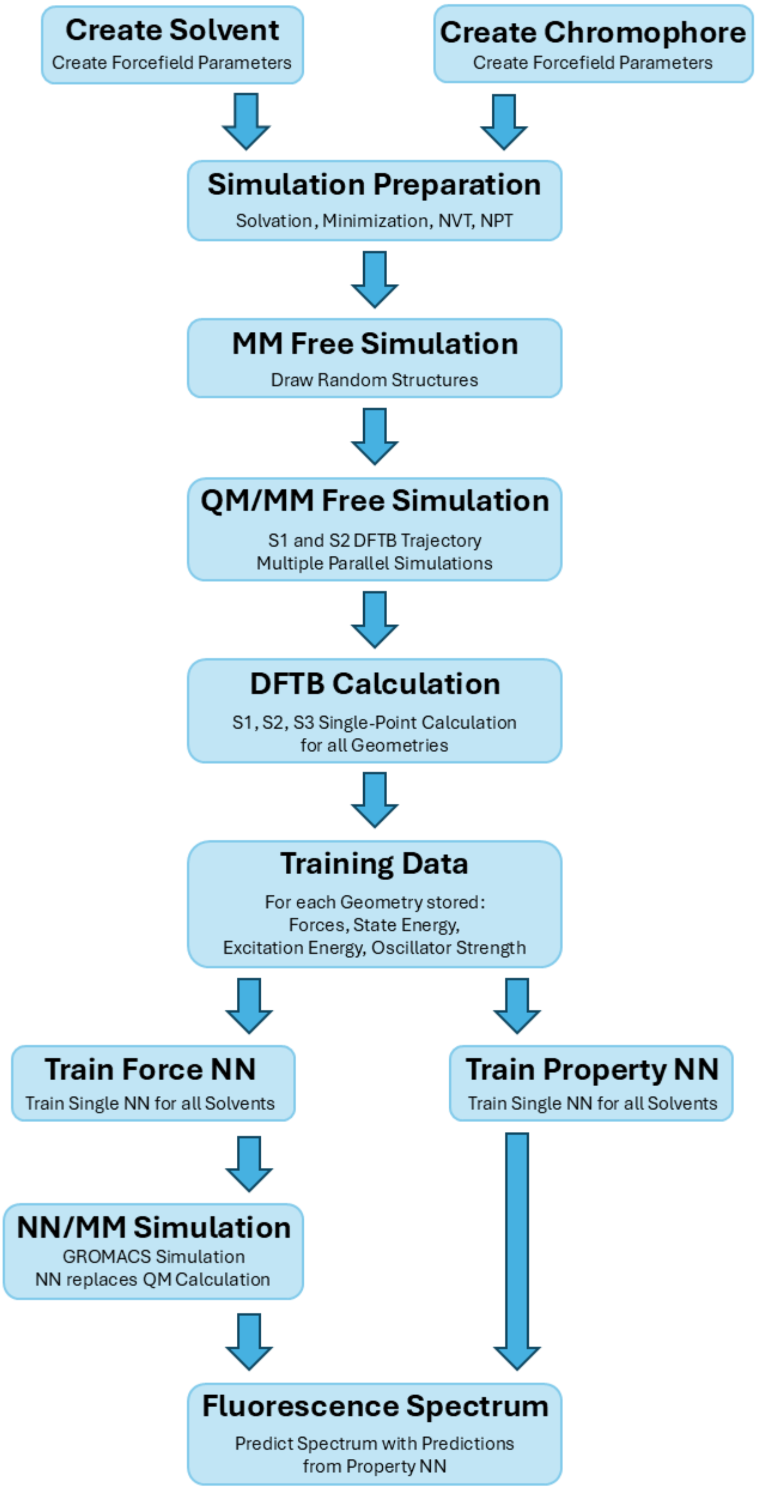


Figure 14.: Schematic fluorescence pipeline: (i) build solvent/chromophore force fields; (ii) prepare solvated systems (minimization, NVT, NPT); (iii) MM production and selection of N random seeds; (iv) parallel free QM/MM dynamics in S_0 (absorption) and/or S_1/S_2 (fluorescence); (v) concatenate runs and evaluate single-point properties for S_0–S_3; (vi) assemble emission and (optional) absorption datasets via bright-state selection; (vii) train the force and property NNs; (viii) run NN/MM with the force NN and construct spectra from property NN predictions.

Part V

Neural Network-Driven Phosphorylation Predictions

5.1. Overview

This chapter develops and validates a local Δ-machine-learning correction to a semi-empirical quantum baseline for phosphoryl-transfer chemistry with nitrogen nucleophiles. The technical core is a second-generation HDNNP trained on $\Delta E = E_{\mathrm{DFT}} - E_{\mathrm{DFTB}}$ labels; datasets are derived from QM/MM well-tempered metadynamics and processed by a reproducible pipeline. The chapter reports how chemically diverse reaction snapshots are generated, how reference/baseline single points are computed, and how atomization-based Δ labels are constructed (Methods I–II; workflow in Chap. IV). It then presents training procedures, on-the-fly integration into GROMACS for corrected energies and forces, generalization tests to unseen phosphate motifs, and a backend swap that reuses the same pipeline with a graph-based MACE/MACE-EQ baseline for comparison. Limitations and implications for long-timescale simulation are discussed at the end.

Author Contributions. This chapter was carried out in collaboration with Lena Eichinger. The present author implemented the HDNNP framework end-to-end-descriptor code, training and inference workflows, and a versioned data pipeline – and equipped the same pipeline with a MACE backend. Discussions on design choices, implementation details, and practical troubleshooting were conducted with Lukas Petersen. Training data were generated jointly by the present author, Lena Eichinger, and Dominik Hachenthal, covering metadynamics simulations, structure extraction, and single-point QM calculations; the division of labor is as follows: datasets 1–3 (present author and Eichinger), 4–6 and 13–16 (Hachenthal and present author), and 7–12 (Eichinger), building on Eichinger's earlier histidine-kinase work during her master's thesis and subsequent contributions as a doctoral researcher. Model training was performed by the present author and Eichinger across multiple cycles with cross-validation. The Δ-QM/MM metadynamics protocol and inputs were prepared within the above pipeline by the present author. MACE production was performed by the present author. Production simulations using the trained HDNNP were executed by the present author jointly with Eichinger.

5.2. Introduction

Phosphorylation reactions involving nitrogen species play essential roles in a wide range of biological processes.[209,210] In bacterial signal transduction, for instance, histidine kinases mediate phosphotransfer between histidine and aspartate residues to regulate adaptive responses.[112,211,212] Similarly, in energy metabolism, creatine and arginine kinases rely on rapid phosphorylation to buffer ATP levels during periods of high energy demand.[213–217] Beyond their physiological importance, phosphoryl-transfer reactions are also central to enzyme catalysis and the design of therapeutic inhibitors or artificial catalysts[18] that target or exploit these pathways. Understanding the energetics and mechanisms of these reactions at an atomistic level is therefore crucial for elucidating enzyme function and for guiding rational design efforts.

From a computational perspective, accurate quantum chemical simulations are required to capture the formation and cleavage of phosphorus–nitrogen (P–N) and phosphorus–oxygen (P–O) bonds, including the subtle changes in coordination and charge distribution that accompany phosphorylation.[106,187] Density Functional Theory (DFT) provides sufficient accuracy but is computationally prohibitive for extensive sampling of reaction pathways in condensed-phase environments.[1,218] Density Functional Tight Binding (DFTB3) offers a cost-efficient alternative, two to three orders of magnitude faster than DFT, but at the expense of reduced accuracy – particularly for phosphorus chemistry, where standard parameterizations tend to overbind P–O bonds and misrepresent local polarization.[10,106]

The concept of Δ-machine learning (Δ-ML), first formalized by Ramakrishnan et al., offers an efficient route to elevate inexpensive quantum methods by training on the energy difference between low- and high-fidelity levels.[9] This concept has since been extended to modern neural network architectures; notably, Smith et al. demonstrated with the ANI-1ccx model that transfer learning from DFT to CCSD(T) data can achieve near gold-standard accuracy while retaining efficiency suitable for large datasets.[183] This strategy has been shown to systematically improve semi-empirical methods toward DFT or even CCSD(T) accuracy and has recently been extended to condensed-phase simulations.[183,184]

Recent machine-learning approaches have begun to bridge the gap for phosphorus chemistry specifically.[192] The DPRc model introduced by Zeng et al. (2021) demonstrated that neural-network corrections can improve semi-empirical QM/MM simulations of RNA phosphorylation by correcting MNDO/d and DFTB2 baselines toward higher fidelity.[219] More recently, Giese et al. (2025) employed a MACE-based[39] Δ-ML potential to elevate AM1/d QM/MM simulations to DFT accuracy for a series of RNA transesterification reactions.[220] While these studies represent important milestones, architectures relying on global graph representations and explicit nonlocal corrections can overlap with DFTB3's existing long-range treatment and increase computational overhead; in this chapter I therefore emphasize local, atom-centered corrections as a complementary path.

In this chapter, DFTB3 is augmented with a second-generation HDNNP Δ-learning correction (atom-centered symmetry functions) and benchmarked against an $SO(3)$-equivariant MACE/MACE-EQ baseline (Methods II, Secs. 3.5, 5.5). Background on phosphoryl transfer is summarized in Methods I, and the end-to-end workflow is documented in the Integration chapter (Chap. IV). The focus here is purely evaluative: assembling and analyzing the phosphorylation datasets; constructing atomization-based residual labels (ΔE, and $\Delta \mathbf{F}$ where available); applying group-aware splits by reaction family with charge states pooled; per-family z-scaling; and reporting diagnostics and held-out accuracy. Live behavior is assessed via on-the-fly Δ-QM/MM corrections (HDNNP, ΔE only) and NN/MM runs (MACE/MACE-EQ, ΔE+$\Delta \mathbf{F}$).

5.3. Dataset Construction

This section documents the composition and labeling of the phosphorylation training data. Generation protocols are given in Methods II (Sec. 3.5.3) and the end-to-end phosphorylation workflow is summarized in the Integration chapter (Sec. 4.2). Here the settings that matter for reproducibility of the datasets and labels are recorded, together with the reaction coverage actually used for ML.

All datasets originate from enhanced sampling of phosphoryl-transfer events in explicit solvent. The reactive core (QM region) contains the nucleophile and the phosphorus moiety and, where relevant, coordinated Mg^{2+}. Except for the histidine-kinase system (protein embedding), sampling was performed in water with a classical MM environment; key settings are summarized in Table A.1. Labels follow the *atomization-based* Δ-definition used throughout Methods II. Minimal reproducibility settings (MD engine/nonbonded/PME, thermostat, QM/MM and DFTB+, metadynamics, and the quantum reference levels for labels and Δ-learning) are listed in the Appendix: MD run control (Table A.2); Nonbonded and PBC (Table A.3); Thermostat / ensemble (Table A.4); QM/MM engine-level settings (Table A.5); DFTB+ Hamiltonian during metadynamics (Table A.6); PLUMED metadynamics parameters (Table A.7); DFT reference level for labels (Table A.8); and DFTB+ single-point setup for Δ-learning (Table A.9). All systems were simulated in explicit water (TIP3P), as listed in the solvent provenance table (Appendix, Table A.19).

5.3.1. Reaction Candidates

A broad pool of nucleophile-donor combinations was attempted; Tables 7 and 8 list, respectively, systems that could be sampled to completion and systems discarded after quality control. Figures 15–18 depict the successfully sampled chemistries, grouped by formal QM-region charge; numbered panels map to Table 7 (grey: used for ML; orange: excluded despite successful sampling).

Table 7.: Successfully sampled reaction systems (numbers map to figure panels); a subset was retained for ML training. "±" indicates that both ethyl/no-ethyl variants were sampled. Rows with *ML use = No* were excluded from training despite successful sampling due to ill-conditioned PES/FES (see text). Where a number appears twice (e.g., 7), it denotes environment variants of the same chemistry (water vs. HK embedding).

Nr.	Nucleophile	Phosphorus donor motif	Ethyl	Charge	ML use
1, 4	4-MeI	Phenyl phosphate	±	−1, −2	Yes
2, 5	4-MeI	*tert*-butyl phosphate	±	−1, −2	Yes
3, 6	4-MeI	Methyl phosphate	±	−1, −2	Yes
13, 14	4-MeI	Anhydride phosphate	±	−1, −2	**No**†
15, 16	4-MeI	Diphosphate (PP)	±	−2, −3	**No**†
8	Imidazole	Methyl phosphate	—	−2	Yes
7	4-MeI	Triphosphate (Mg^{2+})	—	−2	Yes
7	4-MeI	Triphosphate (Mg^{2+}/HK)	—	−2	Yes
9	Arginine	Triphosphate (Mg^{2+})	—	−2	Yes
10	N^G-Methylarg.	Triphosphate (Mg^{2+})	—	−2	Yes
11	Creatine (deprot.)	Triphosphate (Mg^{2+})	—	−3	Yes

† Excluded due to ill-conditioned potential/free-energy surfaces (poor basin connectivity, noisy barriers).

Table 8.: Discarded Phosphorylation Reaction Systems: Attempted systems excluded after quality control; no figure numbers assigned. Reasons include problematic free-energy profiles (FES), unstable coordination, or lack of a stable reaction pathway under the chosen embedding.

Nucleophile	Phosphorus donor motif	Ethyl sub.
4-MeI	Triphosphate–Me	±
4-MeI	Triphosphate–H	±
4-MeI	ATP	±
Arginine (prot.)	Methyl phosphate	—
Arginine (prot.)	Triphosphate	—
N^G-Methylarg. (prot.)	Methyl phosphate	—
N^G-Methylarg. (prot.)	Triphosphate	—
Creatine (prot.)	Triphosphate	—

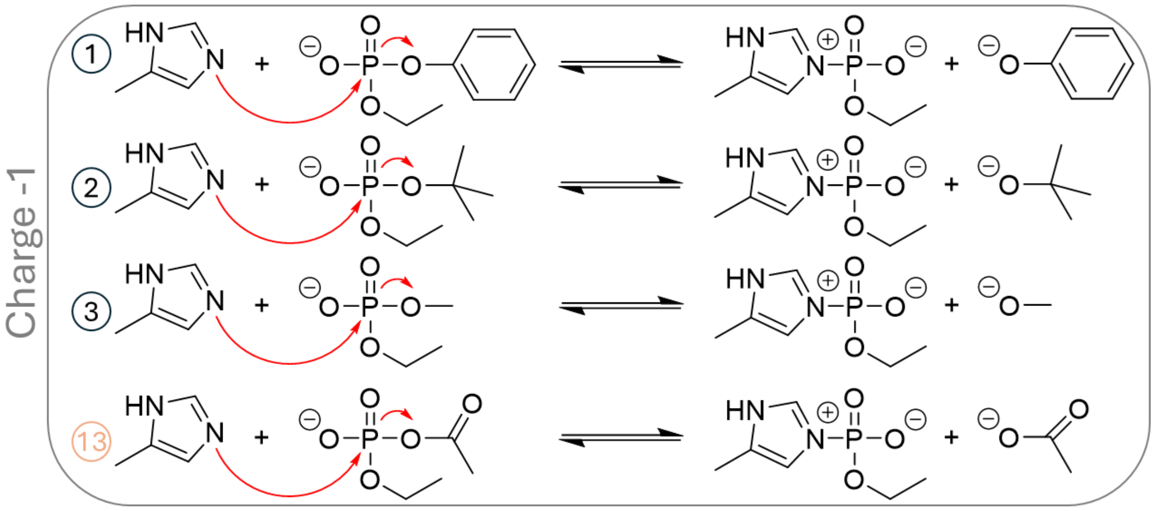


Figure 15.: Reaction schemes (Charge −1) for 4-methylimidazole with phenyl-, *tert*-butyl-, methyl-, and anhydride–phosphate donors. Panel indices ①–③, ⑬ match Table 7. Orange indices denote systems excluded from ML despite successful sampling (Nr. 13), consistent with Table 7.

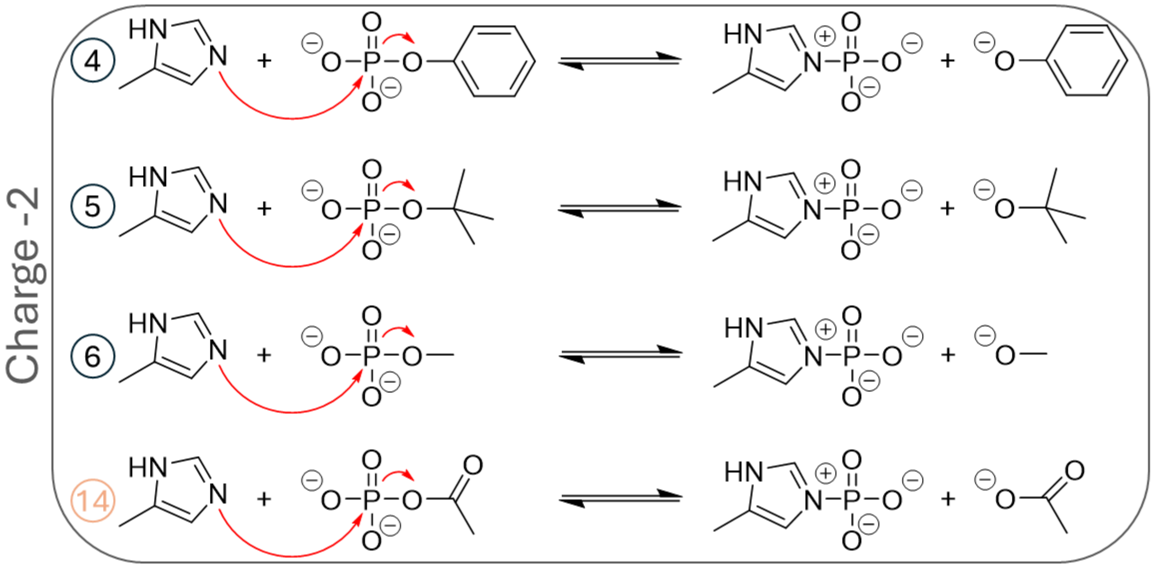


Figure 16.: Reaction schemes (Charge −2) for the same donor series as in Fig. 15. Panel indices ④–⑥, ⑭ match Table 7. Orange index marks the anhydride system (Nr. 14) excluded from ML.

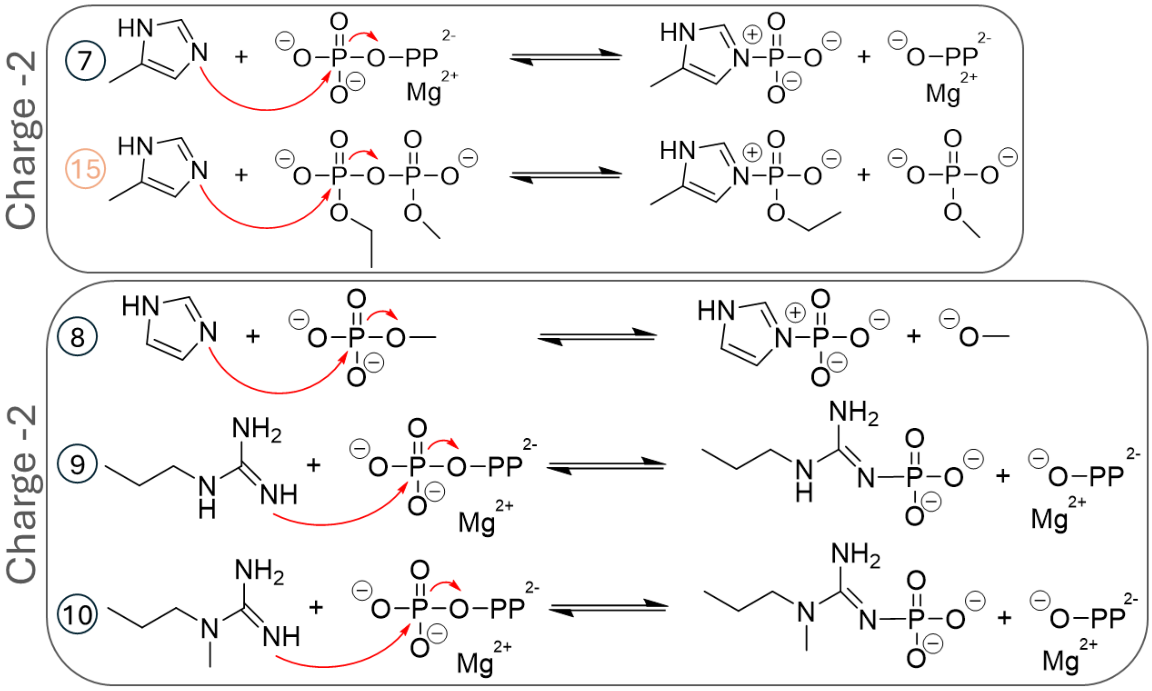


Figure 17.: Reaction schemes (Charge −2) for triphosphate donors coordinated by Mg^{2+}: 4-methylimidazole (bulk water and HK-embedded variant, both indexed ⑦), imidazole (⑧), arginine (deprot., ⑨), N^G-methylarginine (deprot., ⑩) – and the diphosphate control (⑮). ⑮ is excluded from ML (orange index), see Table 7.

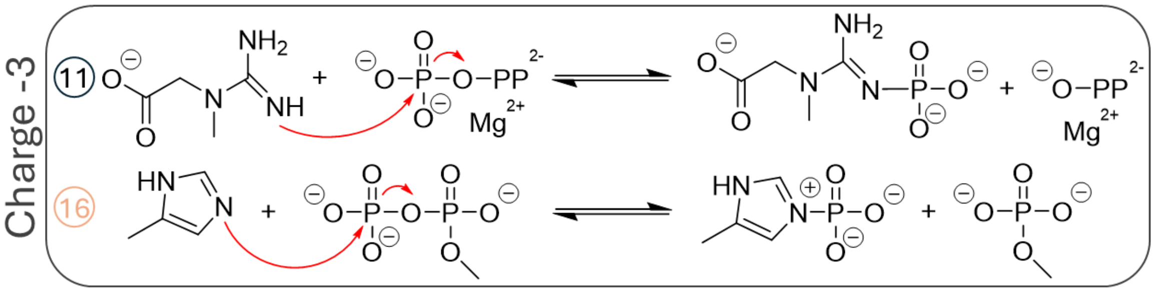


Figure 18.: Reaction schemes (Charge −3): creatine (deprot.) + triphosphate + Mg^{2+} (⑪) and the 4-MeI diphosphate case (⑯). Nr. 16 is excluded from ML (orange index), see Table 7.

The sampled set comprises (i) a *model series in water (MM)* based on imidazole scaffolds—primarily 4-methylimidazole (4-MeI) and imidazole—combined with donors spanning Me–PO_4, Ph–PO_4, *tert*-butyl–PO_4, anhydride (PO_3–PO_4), and diphosphate motifs, each with/without an ethyl substituent ("±"); and (ii) a *biochemical series* with Mg^{2+}-coordinated triphosphate donors sampled both in bulk water and within the histidine-kinase (HK) protein environment, paired with guanidine nucleophiles (arginine, N^G-methylarginine) and creatine. The reaction schematics are shown in Figs. 15–18: Charge −1 (Nr. ①–③, ⑬), Charge −2 model series (Nr. ④–⑥, ⑭), Charge −2 triphosphates with Mg^{2+} (Nr. ⑦, ⑨, ⑩), and Charge −3 (Nr. ⑪, ⑯).

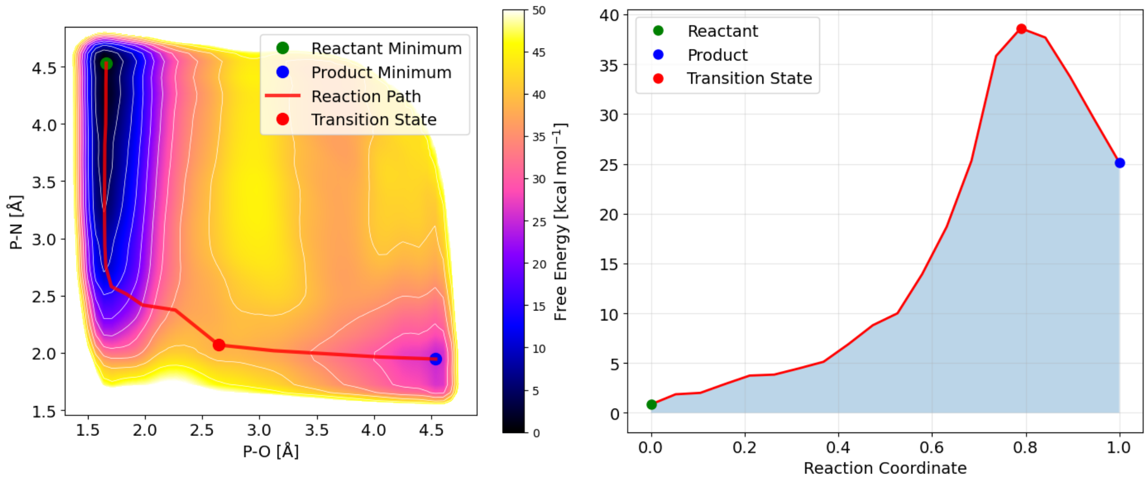
In the tables, *deprot.* denotes the *neutral guanidine form* for arginine/N^G-methylarginine (deprotonated guanidinium, formal charge 0), whereas for *creatine* it denotes the *carboxylate anion* (formal charge −1). Across the series, formal QM-region charges span −1 to −3, governed by phosphate stoichiometry, Mg^{2+} coordination, and nucleophile protonation state. **All entries in Table 7 were successfully sampled**, but *not all* were retained for ML: the *anhydride* (Nr. 13, 14) and *diphosphate* (Nr. 15, 16) cases were excluded due to ill-conditioned PES/FES (poor basin connectivity, noisy barriers). Table 8 lists additional attempted systems that failed quality control and therefore have no figure indices.

Electronic-structure sanity check and reaction selection. For illustration we analyse *seven* representative families from the larger pool—three ultimately retained and four discarded. Selection combines three diagnostics: (i) inspection of 2D free-energy surfaces (FES) in the $(\mathrm{P-O}, \mathrm{P-N})$ plane together with the minimum free-energy paths (MFEP) cut-out to check basin connectivity and barrier placement (retained examples in Fig. 19, excluded ones in Fig. 20); (ii) HOMO/LUMO fractional-occupation histograms from Mermin–DFTB3 single points at $T_\mathrm{F} = 300$ K, where tight peaks (HOMO≈ 2, LUMO≈ 0) indicate a healthy gap (Fig. 21) and broad partial occupations signal gap collapse (Fig. 22); and (iii) maps of SCF non-convergences over the $(\mathrm{P-O}, \mathrm{P-N})$ grid, used only to *localize* numerically fragile regions (typically stretched/dissociative corners; Fig. 23). Energies and barriers reported below are obtained by tracing the MFEP with the waterfall FENEB variant described in Sec. 2.7, which provides a smooth path and a TS estimate on the FES. Quantitatively, the MFEP-based energetics in Table 9 show that the three retained 4-MeI families (Me–PO_4, *t*Bu–PO_4, Ph–PO_4) and their ethyl variants have coherent paths with moderate-to-high yet chemically plausible barriers ($\Delta E^\ddagger \sim 20$–$41\ \mathrm{kcal\,mol^{-1}}$) and transition states near $(\mathrm{P-O}, \mathrm{P-N}) \approx (3.0\pm0.3,\ 1.9$–$2.3)$ Å; the imidazole control (ImH+Me–PO_4) is similar ($\Delta E^\ddagger = 37.37\ \mathrm{kcal\,mol^{-1}}$). By contrast, the four excluded anhydride/diphosphate families exhibit ill-conditioned FES (broad plateaus or late/high barriers; Fig. 20) and pathological occupations (Fig. 22); although their apparent MFEP barriers can be small ($\sim$3–19 $\mathrm{kcal\,mol^{-1}}$) and ΔE_rxn even negative, these cases correlate with clustered SCF failures (Fig. 23) and would yield noisy labels. Based on (i)–(iii), we retain the Me/*t*Bu/Ph phosphate series (with and without ethyl) and de-prioritize the anhydride and diphosphate families.

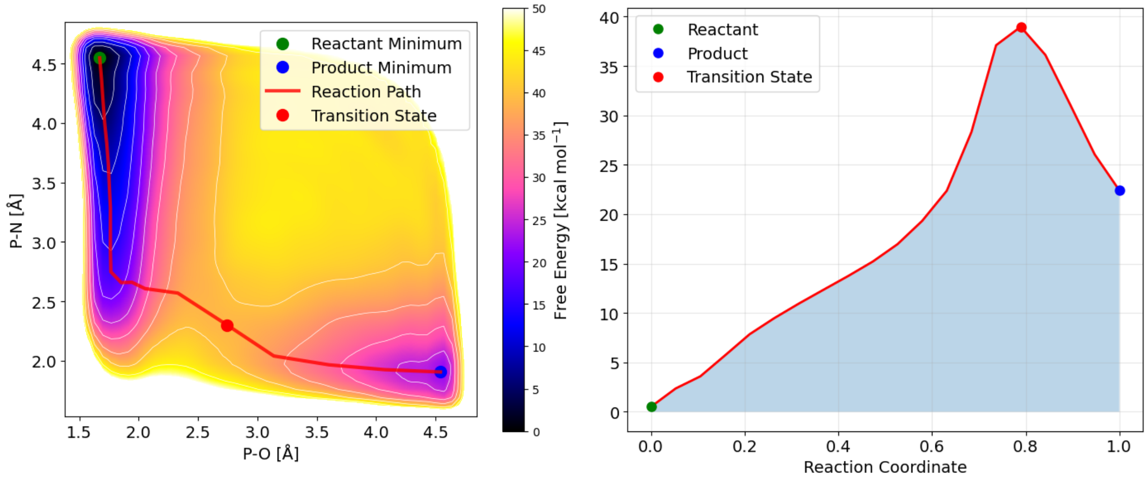
Table 9.: Training Data: Reaction-path energetics extracted from the 2D free-energy surfaces (FES; CVs: P–O and P–N); DFTB3 baseline *without diffusion*. Reported are the reactant and product energies (E_R, E_P), the transition-state energy (E_TS), the activation barrier $\Delta E^\ddagger = E_\mathrm{TS} - E_\mathrm{R}$, the reaction energy $\Delta E_\mathrm{rxn} = E_\mathrm{P} - E_\mathrm{R}$, and the TS coordinates along the FES. Units: energies in kcal mol^{-1}, distances in Å. Energetics are obtained by tracing the MFEP with the waterfall FENEB variant (Sec. 2.7). For the HK-embedded case (row marked [†]), the TS coordinates use alternative CVs; see note below. Numbers map to reaction schemes in Figs. 15–18.

Nr.	Reaction	E_R	E_P	E_TS	$\Delta E^\ddagger$	ΔE_rxn	TS (P–O, P–N)
1	4-MeI + Ph–PO$_4$Et	0.22	15.16	23.62	23.40	14.94	(2.17, 2.06)
2	4-MeI + *t*Bu–PO$_4$Et	0.32	32.16	41.55	41.23	31.84	(2.99, 1.95)
3	4-MeI + Me–PO$_4$Et	0.27	32.95	39.30	39.03	32.68	(2.99, 1.87)
4	4-MeI + Ph–PO$_4$	0.35	9.41	20.98	20.62	9.06	(2.27, 2.33)
5	4-MeI + *t*Bu–PO$_4$	0.52	22.42	38.97	38.45	21.90	(2.74, 2.30)
6	4-MeI + Me–PO$_4$	0.86	25.13	38.61	37.75	24.27	(2.64, 2.07)
7	4-MeI + Me–PP–PO$_4$	25.01	0.18	33.92	8.91	−24.82	(1.84, 2.25)
8	ImH + Me–PO$_4$	0.42	27.73	37.80	37.37	27.31	(3.01, 2.02)
9	Arg + Me–PP–PO$_4$	19.93	0.24	38.67	18.74	−19.69	(1.88, 2.23)
10	MeArg + Me–PP–PO$_4$	20.22	0.36	40.14	19.91	−19.86	(1.91, 2.24)
11	Crn + Me–PP–PO$_4$	21.27	0.14	38.15	16.88	−21.13	(1.99, 2.19)
12	4-MeI + Me–PP–PO$_4$ (from HK)	23.05	9.30	26.92	3.86	−13.75	(−2.21, 1.62)[†]
13	~~4-MeI~~ + ~~CO$_2$Me–PO$_4$Et~~	6.95	0.29	16.29	9.34	−6.66	(2.08, 2.38)
14	~~4-MeI~~ + ~~CO$_2$Me–PO$_4$~~	11.41	0.27	15.23	3.82	−11.14	(2.28, 2.64)
15	~~4-MeI~~ + ~~MePO$_3$–PO$_4$Et~~	5.26	0.40	24.39	19.13	−4.86	(2.24, 2.31)
16	~~4-MeI~~ + ~~MePO$_3$–PO$_4$~~	14.43	0.42	24.43	10.00	−14.00	(2.16, 3.31)

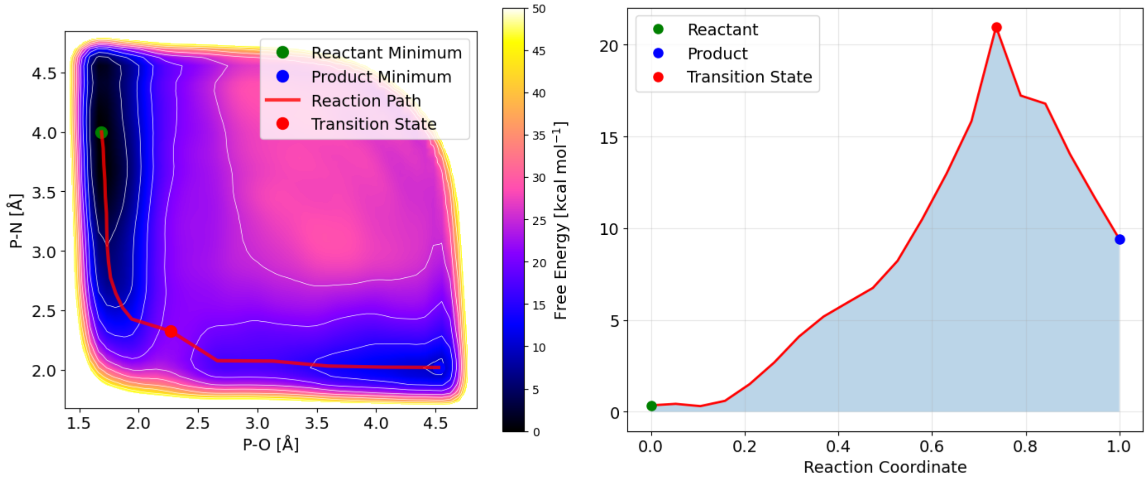
[†] Entry taken from the histidine-kinase (HK) embedded simulation. For this row, the TS coordinate pair does *not* use the distances (P–O, P–N); instead it uses the alternative progress variables (NHO, OPN), where NHO tracks the proton-transfer coordinate and OPN tracks the phosphoryl-transfer coordinate (units Å).



(a) 4-MeI + Me−PO$_4$

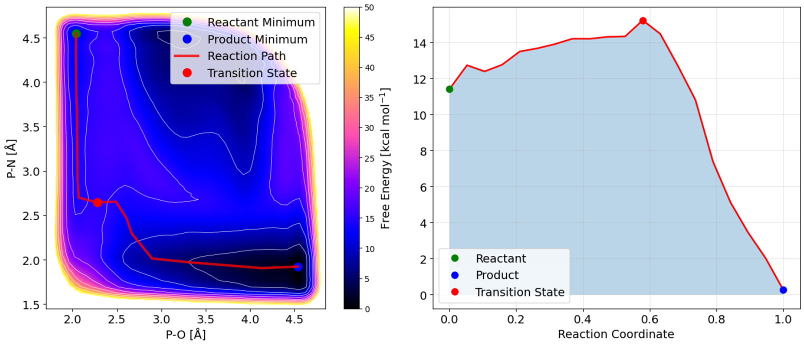


(b) 4-MeI + tBu−PO$_4$

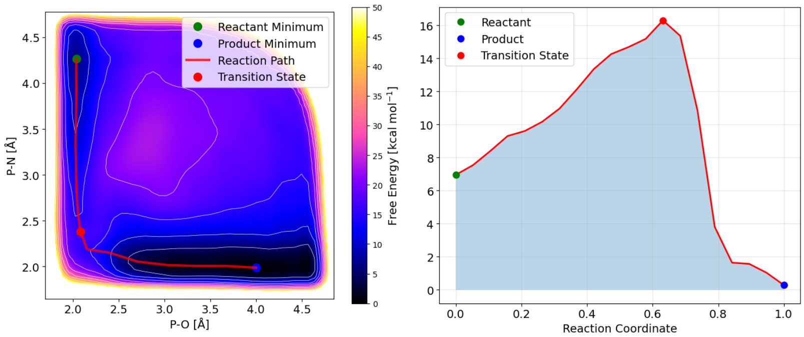


(c) 4-MeI + Ph−PO$_4$

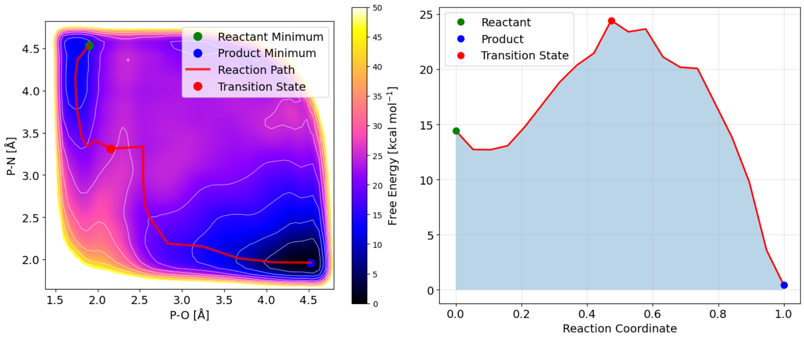
Figure 19.: Accepted data for ML. Three phosphorylation families that were *retained* for dataset generation and ML training: 4-MeI + Me−PO$_4$, 4-MeI + tBu−PO$_4$, and 4-MeI + Ph−PO$_4$. FES in the (P−O, P−N) plane; the red polyline is the MFEP cut-out, green/blue dots mark reactant/product minima, and the red dot as TS estimate. All three systems are used as training families.



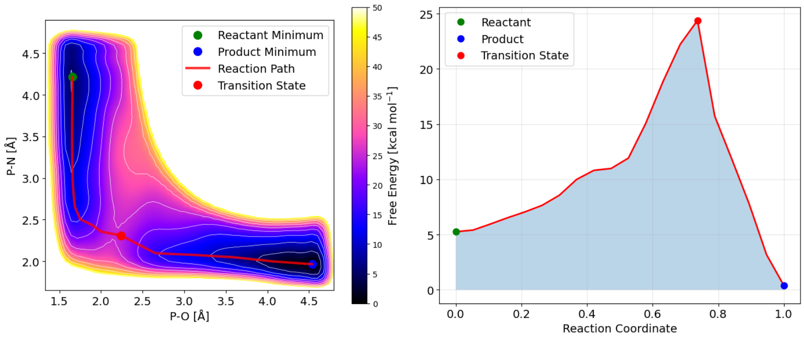
(a) 4-MeI + $CO_2Me{-}PO_4$



(b) 4-MeI + $CO_2Me{-}PO_4Et$

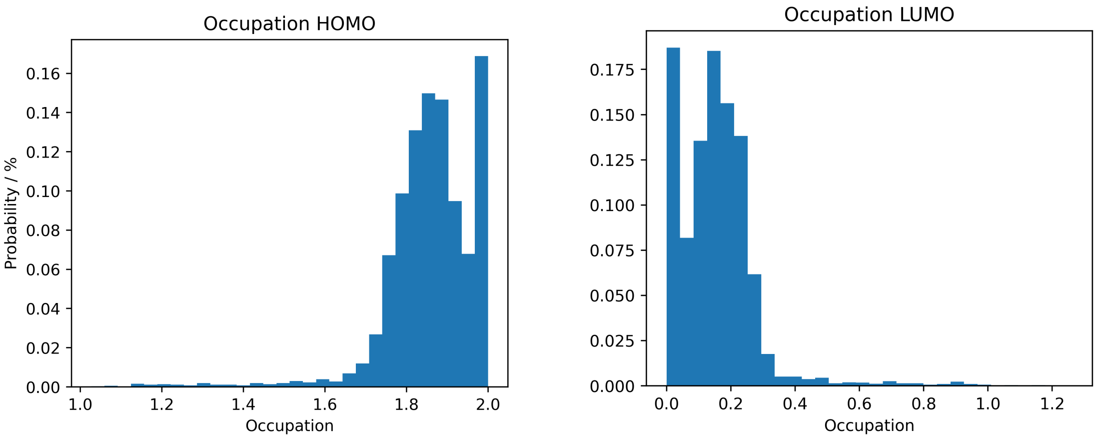


(c) 4-MeI + $MePO_3{-}PO_4$

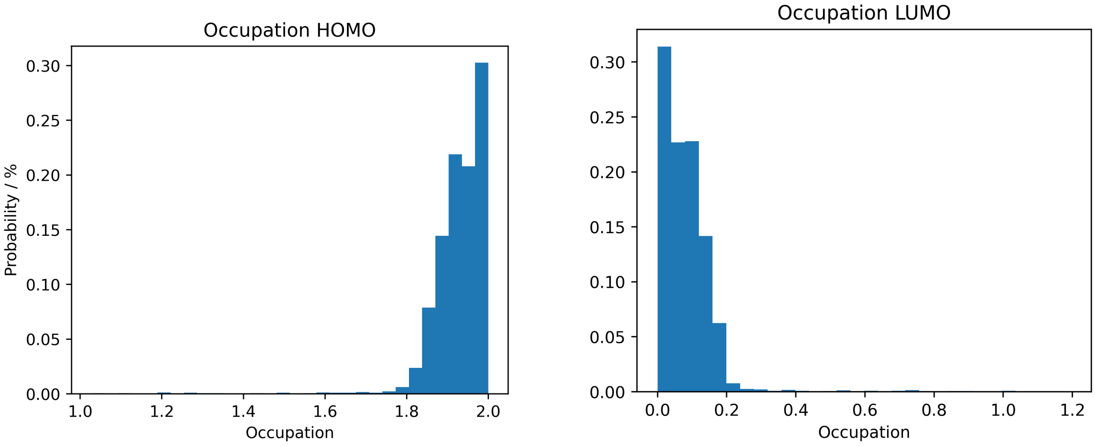


(d) 4-MeI + $MePO_3{-}PO_4Et$

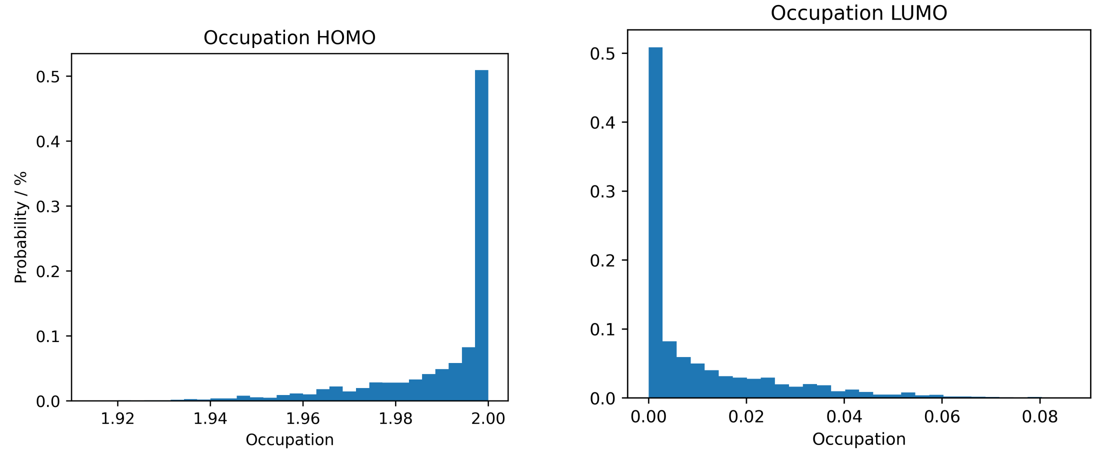
Figure 20.: Excluded data from ML. Free-energy surfaces (FES) in the (P−O, P−N) plane with extracted minimum free-energy paths (MFEP) for four reaction families that were *excluded* from ML training. MFEP (red), reactant/product minima (green/blue), and the highest point along the path (red dot). Broad plateaus, ill-defined basins, or high/late barriers along the MFEP correlate with unstable sampling and noisy labels; therefore these families were not taken forward.



(a) 4-MeI + Me–PO_4



(b) 4-MeI + *t*Bu–PO_4



(c) 4-MeI + Ph–PO_4

Figure 21.: Fractional-occupation diagnostics for retained reactions. Histograms of HOMO and LUMO occupations from Mermin–DFTB3 single points (Fermi temperature 300 K) over uniformly drawn geometries of each candidate reaction. Well-behaved sets show a tight HOMO peak near 2 and LUMO near 0, indicating a robust HOMO–LUMO gap across sampled configurations. Panels correspond to three retained families (*4-MeI + Me–*PO_4, *4-MeI + tBu–*PO_4, *4-MeI + Ph–*PO_4 as labeled).

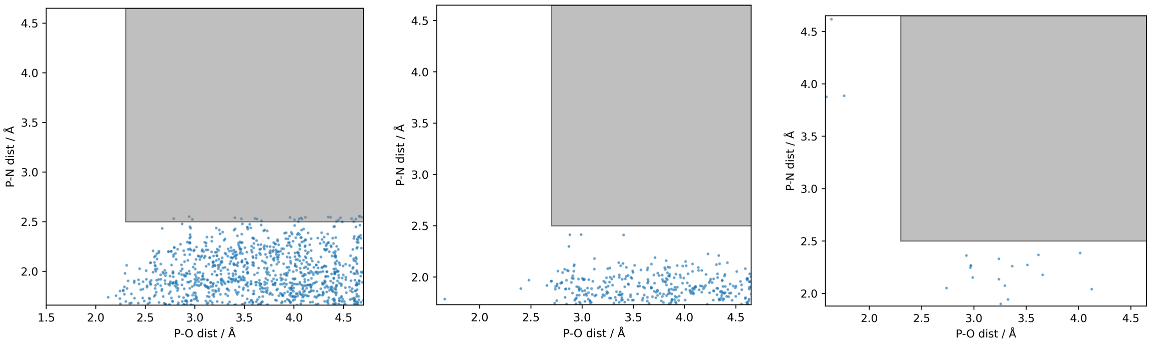
(a) 4-MeI + $CO_2Me–PO_4$

(b) 4-MeI + $CO_2Me–PO_4Et$

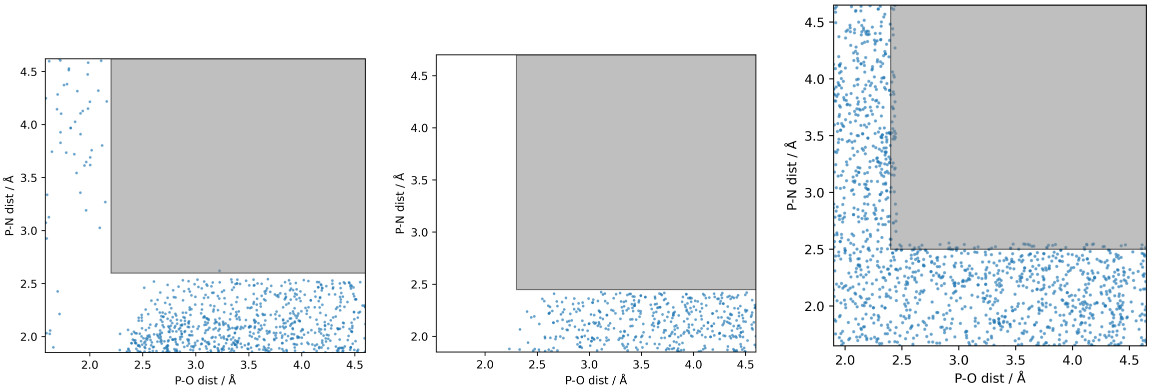
(c) 4-MeI + $MePO_3–PO_4$

(d) 4-MeI + $MePO_3–PO_4Et$

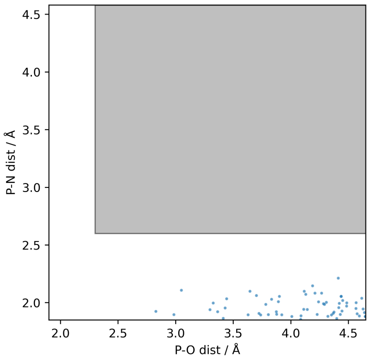
Figure 22.: Fractional-occupation diagnostics for discarded reactions. Same analysis as Fig. 21, now for the four excluded families (*4-MeI + $CO_2Me–PO_4$*, *4-MeI + $CO_2Me–PO_4Et$*, *4-MeI + $MePO_3–PO_4$*, *4-MeI + $MePO_3–PO_4Et$*). Broad HOMO depletion and/or substantial partial LUMO population indicate frequent gap collapse and near-metallic (or strongly charge-transfer) situations. Such heavy-tailed occupation patterns correlate with SCF instabilities and noisy labels, and these families were therefore excluded from training.



(a) 4-MeI + Me–PO_4 **(b)** 4-MeI + *t*Bu–PO_4 **(c)** 4-MeI + Ph–PO_4



(d) 4-MeI + $MePO_3$–PO_4 **(e)** 4-MeI + $MePO_3$–PO_4Et **(f)** 4-MeI + CO_2Me–PO_4



(g) 4-MeI + CO_2Me–PO_4Et

Figure 23.: Where DFTB single points fail. For each reaction family, points indicate sampled geometries in the plane of forming P–N distance (vertical) vs. breaking P–O distance (horizontal). Symbols mark frames where Mermin-DFTB3 single points did *not* converge; the shaded region highlights the stretched/dissociative corner where failures are most frequent. The clustering of non-convergences at long P–O and/or large P–N is consistent with the fractional-occupation pathologies seen in Fig. 22, and supports our choice to exclude those families in the final dataset curation. All non-converged frames are filtered by the QC pipeline before label construction.

5.3.2. Selected Reactions

From the successfully sampled pool (Table 7), a subset was retained for ML to ensure well-behaved potential/free-energy landscapes and balanced coverage across charge states, donor motifs, and environments. Systems flagged as ill-conditioned (anhydrides and diphosphates) were excluded to avoid contaminating the residual with sampling pathologies.

Table 10 lists the *final* ML set, reporting for each entry the combined *Reaction* (single column), the total QM-region charge, whether Mg^{2+} is part of the QM cutout, and the simulation environment.

To contextualize label magnitudes used in Δ-learning, the compact atomization summaries in Table 12 provide (i) component-wise values for optimized fragments and (ii) combined nucleophile+donor totals for consistency checks. As indicated in the captions, these tables include a few systems beyond the final ML subset for completeness; struck-through rows were *not* used for ML training. Reaction indices (Nr.) map to the scheme panels in Figs. 15–18, and MFEP-based barriers/coordinates are summarized in Table 9.

Table 10.: Reactions retained for machine learning (ML), grouped by total charge and annotated with the presence of Mg^{2+} in the QM cutout and the simulation environment.

Nr.	**Reaction**		Total charge	Mg^{2+}	Environment
1	4-MeI	+ Ph–PO_4Et	−1	—	bulk water
2	4-MeI	+ *t*Bu–PO_4Et	−1	—	bulk water
3	4-MeI	+ Me–PO_4Et	−1	—	bulk water
4	4-MeI	+ Ph–PO_4	−2	—	bulk water
5	4-MeI	+ *t*Bu–PO_4	−2	—	bulk water
6	4-MeI	+ Me–PO_4	−2	—	bulk water
7	4-MeI	+ Me–PP–PO_4	−2	Yes	bulk water
8	ImH	+ Me–PO_4	−2	—	bulk water
9	Arg	+ Me–PP–PO_4	−2	Yes	bulk water
10	N^G-MeArg	+ Me–PP–PO_4	−2	Yes	bulk water
11	Crn (deprot.)	+ Me–PP–PO_4	−3	Yes	bulk water
12†	4-MeI	+ Me–PP–PO_4	−2	Yes	HK-embedded

† HK = histidine-kinase protein embedding (same QM cutout; different environment).

5.3.3. Reaction Diagnostics

This subsection reports diagnostics for both *energy* and *force* residual labels collected for the phosphorylation datasets. Datasets use *group-aware* splits by reaction family (charge states pooled), require convergence of both DFT and DFTB single points, and balance MFEP coverage across reactant, transition, and product regions including plateau segments.

Table 11.: Dataset summary per reaction (ΔE **only**): count and raw label statistics together with per-reaction z-normalized minima and maxima (computed on the training split of each reaction and applied unchanged to validation/test). Energies in kcal mol^{-1}.

	Reaction	n	Mean	Std	Min	Max	z_{min}	z_{max}	z-Range
1	4-MeI + Ph-PO$_4$Et	11846	−13.75	3.93	−27.66	−1.17	−3.54	3.21	6.75
2	4-MeI + tBu-PO$_4$Et	11522	7.79	4.53	−8.77	21.34	−3.66	2.99	6.65
3	4-MeI + Me-PO$_4$Et	10873	5.79	4.65	−9.19	29.01	−3.22	4.99	8.21
4	4-MeI + Ph-PO$_4$	9980	−8.39	3.63	−20.16	5.14	−3.25	3.73	6.98
5	4-MeI + tBu-PO$_4$	9664	10.74	3.84	−5.32	25.41	−4.18	3.82	8.00
6	4-MeI + CH$_3$-PO$_4$	8964	11.54	4.89	−2.28	135.83	−2.82	25.40	28.22
7	4-MeI + Me-PP-PO$_4$	11940	50.39	7.18	25.44	76.53	−3.47	3.64	7.11
8	ImH + Me-PO$_4$	10406	8.41	3.52	−4.24	23.53	−3.59	4.30	7.89
9	Arg + Me-PP-PO$_4$	11855	59.24	8.01	30.50	83.94	−3.59	3.09	6.68
10	MeArg + Me-PP-PO$_4$	11993	57.15	7.92	27.41	82.66	−3.76	3.22	6.98
11	Crn + Me-PP-PO$_4$	11996	42.71	8.31	14.04	68.36	−3.45	3.08	6.53
12	4-MeI + Me-PP-PO$_4$ (HK)	11965	41.04	7.41	15.48	68.08	−3.45	3.65	7.10

Energies (ΔE). Table 11 summarizes, for each reaction, the number of frames and descriptive statistics of the **energy** residuals, together with the per-reaction z-normalized extrema (training statistics applied unchanged to validation/test). Reaction sizes cluster around 9–12k frames; raw spreads differ across donors/charge states as expected. The raw distributions and per-reaction boxplots are shown in Fig. 25; after per-reaction standardization the histograms collapse to similar bell shapes with comparable tail reach (Fig. 24). One case (4-MeI + CH$_3$-PO$_4$) exhibits a larger normalized range driven by a small number of anomalous frames; the bulk aligns with the other families. Reference atomization energies at optimized geometries are listed in Table 12.

Expected z-range. As outlined in Methods II (Sec. 3.2.4), sample sizes of order 10^4 imply typical standardized maxima/minima at a few standard deviations from zero, yielding an expected overall z-range of ~7–8. The observed per-reaction z-ranges in Table 11 fall between ~6.6 and 8.2 (median ≈7.1), in line with this prediction; the sole clear deviation is the CH$_3$-PO$_4$ outlier case noted above.

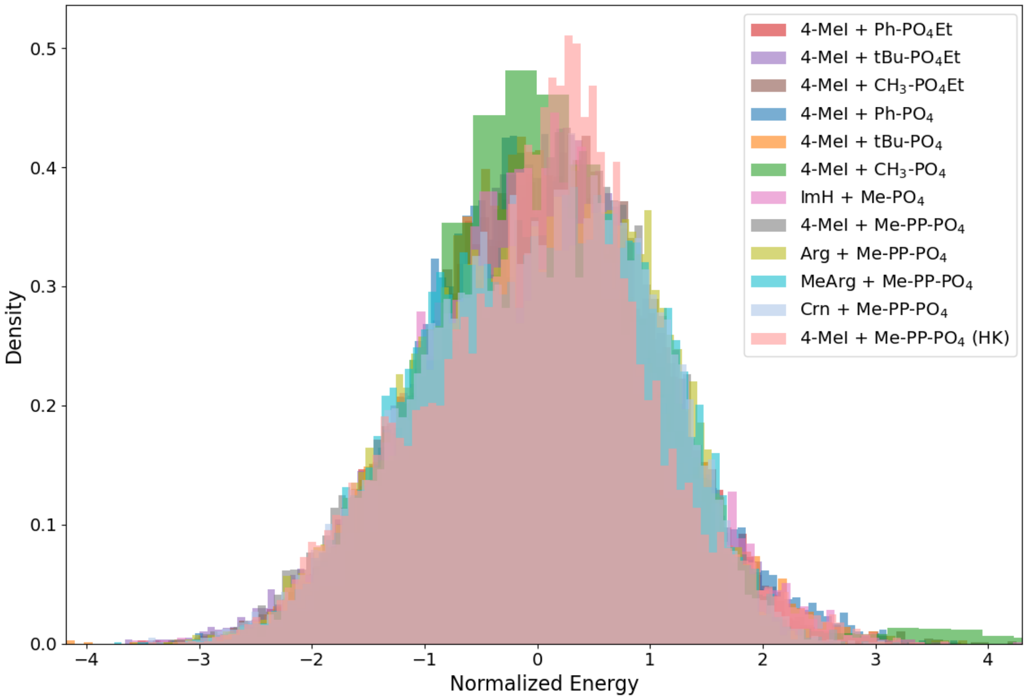


Figure 24.: Per-reaction histograms of standardized ΔE labels (one overlay per reaction). As expected after per-reaction standardization, the bulk collapses well; small differences in tail reach reflect residual outliers and slight variations in sample size.

Forces (ΔF). Residual **force** labels are recorded for all frames. Figure 26 compiles per-reaction box–whisker summaries of the *per-configuration force magnitude* on a logarithmic scale (kcal mol^{-1} Å^{-1}). Medians and IQRs are comparable across reactions; long right tails stem from a small number of strained/outlier frames. No z-scaling is applied to forces here. These diagnostics are used downstream by the MACE/MACE-EQ models, whereas the HDNNP setup trained next uses *energies only*.

Geometry embedding. A t-SNE embedding of the geometry descriptors (Fig. 27) shows compact, well-separated clusters by reaction family, with intra-family spread reflecting MFEP coverage. The HK variant appears more diffuse, consistent with its broader conformational context.

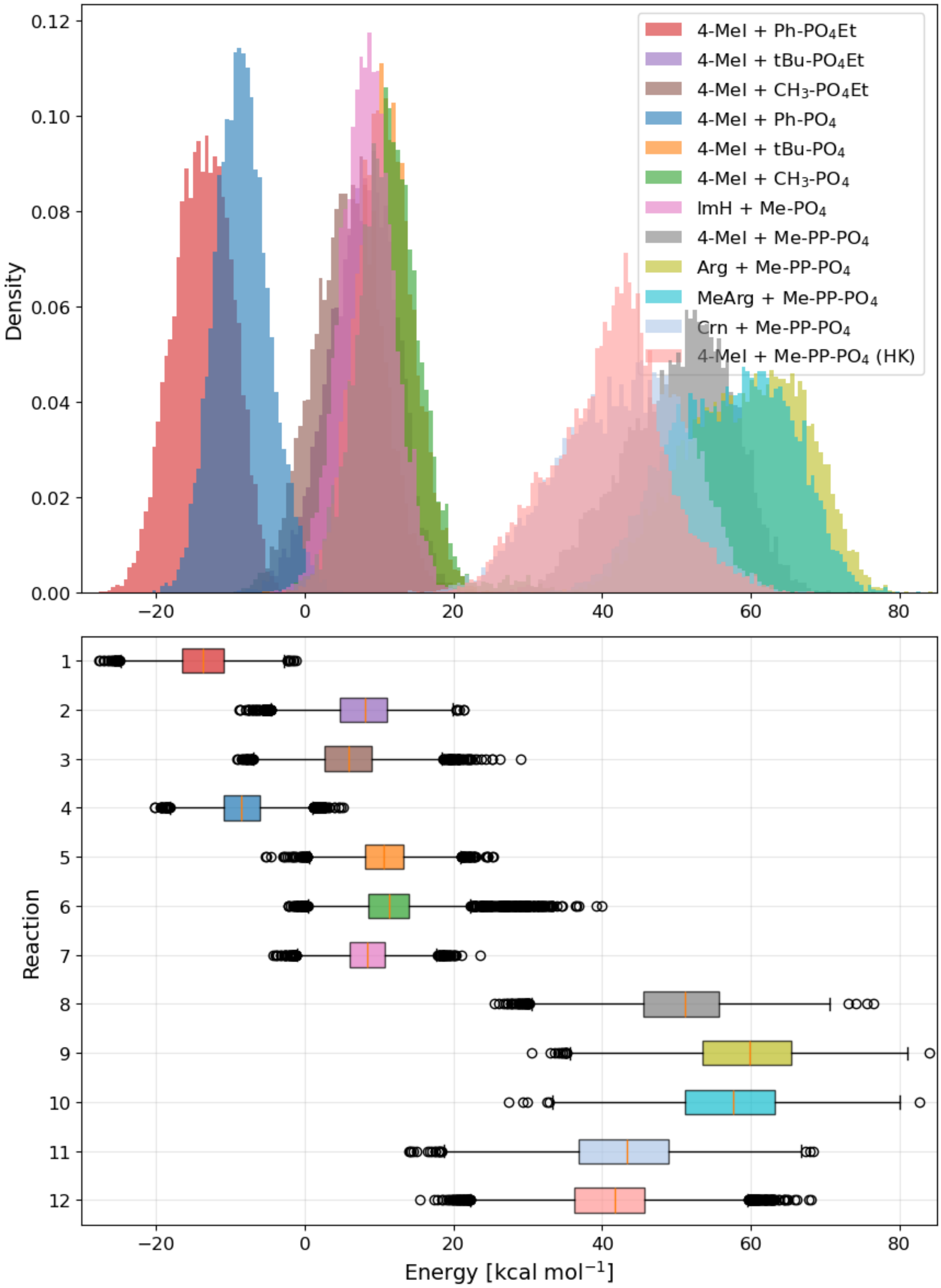


Figure 25.: Distributions of atomization-residual labels $\Delta E = E^{\mathrm{DFT}}_{\mathrm{atom}} - E^{\mathrm{DFTB}}_{\mathrm{atom}}$ used as targets in the Δ-learning models (see Sec. 5.3.2). **Top:** stacked probability histograms for each ML-selected reaction reaction (legend). **Bottom:** per-reaction box-and-whisker summary (median, IQR box, whiskers at $1.5 \times$ IQR; circles denote outliers). The panels visualize location and spread of the training labels across families. Energies in kcal mol^{-1}.

Table 12.: Atomization energies $E_{\text{atom}}^{\text{DFT}}$ and $E_{\text{atom}}^{\text{DFTB}}$ for (a) full reaction systems (nucleophile + phosphate leaving group) and (b) the corresponding optimized components. Energies in kcal mol^{-1}; numbers map to reaction schemes in Figs. 15–18. *Struck-through entries: excluded from ML.*

(a) Reaction systems: nucleophile + leaving group. Includes number of covalent bonds (double bonds counted as 1) and error per bond.

Nr.	**Reaction**		$E_{\text{atom}}^{\text{DFT}}$	$E_{\text{atom}}^{\text{DFTB}}$	ΔE	Bonds	ΔE/Bond
1	4-MeI	+ Ph–PO_4Et	−3900.5	−3871.0	−29.5	35	−0.84
2	4-MeI	+ tBu–PO_4Et	−3806.5	−3795.2	−11.3	36	−0.31
3	4-MeI	+ Me–PO_4Et	−2914.5	−2903.7	−10.8	27	−0.40
4	4-MeI	+ Ph–PO_4	−3193.5	−3175.1	−18.4	28	−0.66
5	4-MeI	+ tBu–PO_4	−3086.7	−3085.4	−1.3	29	−0.04
6	4-MeI	+ Me–PO_4	−2190.1	−2190.8	0.7	20	0.04
7	4-MeI	+ Me–PP–PO_4	−3382.9	−3376.8	−6.1	28	−0.22
8	ImH	+ Me–PO_4	−1892.6	−1892.1	−0.5	17	−0.03
9	Arg	+ Me–PP–PO_4	−3808.6	−3812.5	3.9	33	0.12
10	MeArg	+ Me–PP–PO_4	−4097.3	−4099.5	2.2	36	0.06
11	Crn	+ Me–PP–PO_4	−4182.9	−4170.4	−12.5	32	−0.39
12	4-MeI	+ Me–PP–$PO_4^{\dagger}$	−3382.9	−3376.8	−6.1	28	−0.22
13	~~4-MeI~~	+ ~~CO_2Me–PO_4Et~~	−3327.2	−3310.1	−17.1	29	−0.59
14	~~4-MeI~~	+ ~~CO_2Me–PO_4~~	−2657.1	−2652.1	−5.0	22	−0.23
15	~~4-MeI~~	+ ~~$MePO_3$–PO_4Et~~	−3396.8	−3386.2	−10.6	31	−0.34
16	~~4-MeI~~	+ ~~$MePO_3$–PO_4~~	−2602.5	−2606.3	3.8	24	0.16

† By construction of our atomization protocol, the HK-embedded variant is identical (in E_{atom}) to the bulk-water triphosphate case.

(b) Components used in the reaction systems (optimized fragments). Includes number of covalent bonds and error per bond.

Group	$E_{\text{atom}}^{\text{DFT}}$	$E_{\text{atom}}^{\text{DFTB}}$	ΔE	Bonds	ΔE/Bond
Ph–PO_4Et	−2623.61	−2614.37	−9.24	23	−0.40
tBu–PO_4Et	−2529.58	−2538.62	9.04	24	0.38
Me–PO_4Et	−1637.57	−1647.11	9.54	15	0.64
Ph–PO_4	−1916.65	−1918.53	1.88	16	0.12
tBu–PO_4	−1809.81	−1828.78	18.97	17	1.12
Me–PO_4	−913.19	−934.26	21.07	8	2.63
Me-PP-PO_4	−2106.01	−2120.20	14.19	16	0.89
4-MeI	−1276.89	−1256.59	−20.30	12	−1.69
ImH	−979.40	−957.84	−21.56	9	−2.40
Arg	−1702.60	−1692.31	−10.29	17	−0.61
N^G-MeArg	−1991.29	−1979.33	−11.96	20	−0.60
Crn (deprot.)	−2076.88	−2050.20	−26.68	16	−1.67
CO_2Me–PO_4Et	−2050.33	−2053.50	3.17	17	0.19
CO_2Me–PO_4	−1380.20	−1395.52	15.32	10	1.53
$MePO_3$–PO_4Et	−2119.90	−2129.59	9.69	19	0.51
$MePO_3$–PO_4	−1325.61	−1349.76	24.15	12	2.01

Abbrev.: Me = methyl; Et = ethyl; tBu = tert-butyl; Ph = phenyl; CO_2Me = methoxycarbonyl; PO_4 = (mono)phosphate; PP = diphosphate motif; ImH = imidazole; 4-MeI = 4-methylimidazole; Arg = arginine; MeArg = methylarginine; Crn = creatinine. A trailing “Et” denotes an ethyl ester at the terminal phosphate.

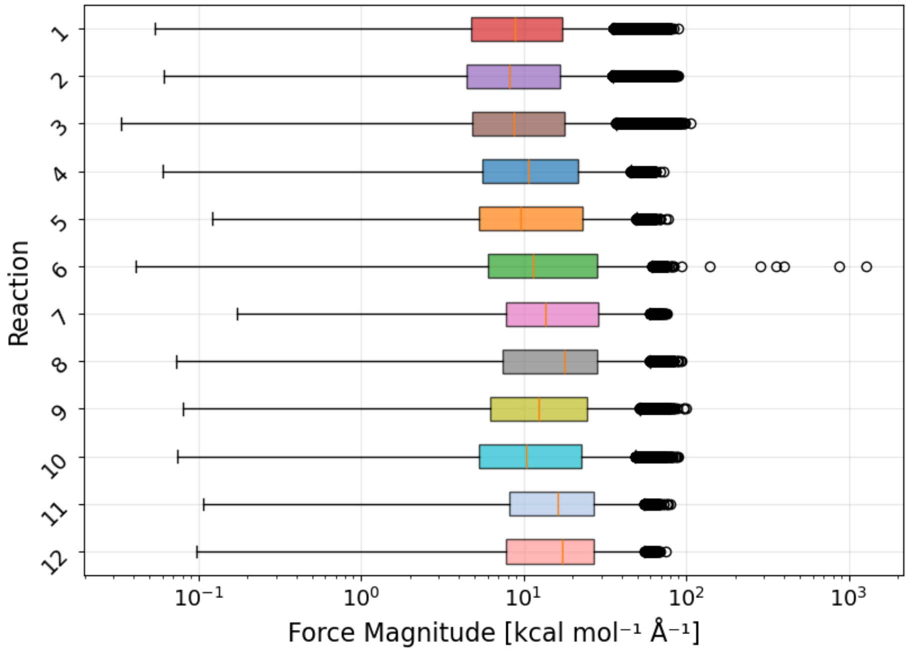


Figure 26.: Distribution of *residual force magnitudes* per configuration, grouped by reaction (indices 1–12; numbering as in Table 12a). Box-whisker summaries (median, IQR box; whiskers at 1.5 × IQR; markers denote outliers) are shown on a logarithmic x-axis (units: $\mathrm{kcal\,mol^{-1}\,Å^{-1}}$). These statistics refer to the **force** labels used later with the MACE/MACE-EQ models; the HDNNP setup in this chapter is trained on **energy** residuals only. No z-scaling is applied to forces here; the very long right tails stem from a small number of strained/outlier frames, while the bulk lies in a narrow band across reactions.

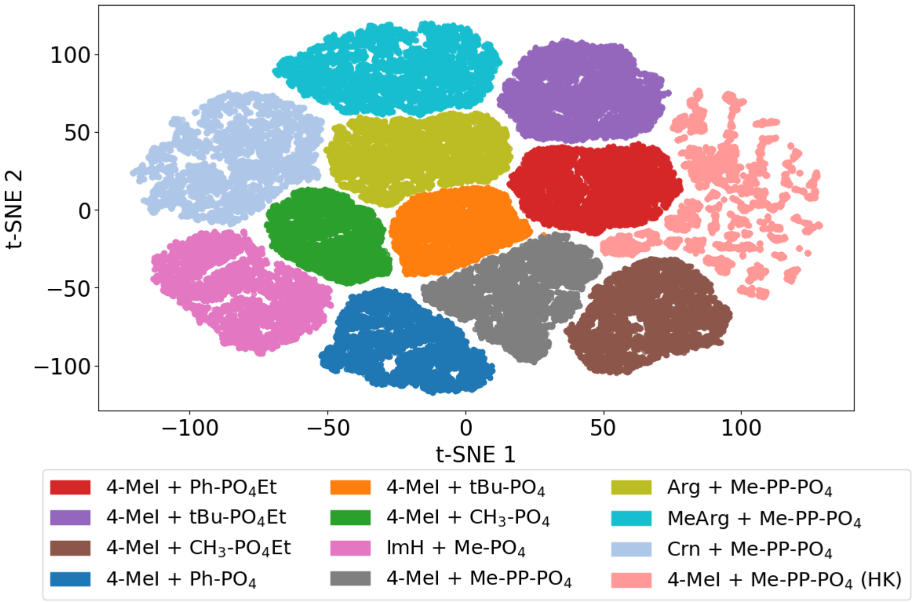


Figure 27.: Two-dimensional t-SNE embedding of per-frame *geometry descriptors* (the same inputs used by the HDNNP) coloured by reaction family. Compact, well-separated clouds indicate that snapshots cluster by chemical motif; within-family spread reflects the coverage along the MFEP. The histidine-kinase variant (HK) appears more diffuse, consistent with its broader conformational context. Axes are t-SNE coordinates (arbitrary units); the plot is illustrative and not used for training.

5.4. Δ Network: HDNNP

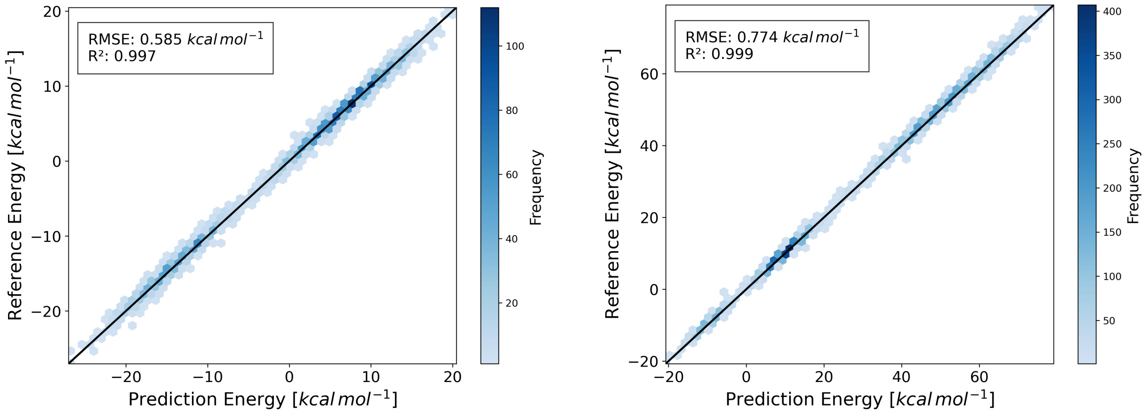
This section assembles and evaluates Δ-learning high-dimensional neural network potentials (HDNNP) for the phosphorylation reaction set defined in Sec. 3.5.3. The model design and featurization follow the neural-network methodology outlined in Part III, and the end-to-end data/compute flow is described in Sec. 4.2. Three components are covered: *(i) Training:*single-reaction models and charge-pooled models (total charge −1, −2, and their combinations, including −1/−2/−3); *(ii) Cross-validation:* stratified splits with parity and error statistics (reported as R^2 and RMSE in kcal mol^{-1}); and *(iii) Δ-QM/MM application:* deployment of the trained HDNNP as a Δ-corrector within a QM/MM workflow and assessment of stability and accuracy. A particular point of interest is charge pooling: although the network architecture is not explicitly charge-aware, aggregated training by total charge and across charges is tested for transferability and robustness.

Table 13.: Training and test performance of Δ-learning HDNNP models for pooled *charge sets* (top, bold) and *individual reactions* (bottom, numbered). Metrics: coefficient of determination R^2 and RMSE in kcal mol^{-1}.

	Reaction		Total charge	R^2	RMSE
	All -1, -2, -3		−1/−2/−3	**0.999**	**0.72**
	All -1, -2		−1/−2	0.999	0.69
	All -1		−1	0.997	0.59
	All -2		−2	0.999	0.77
1	4-MeI	+ Ph–PO_4Et	−1	0.969	0.68
2	4-MeI	+ *t*Bu–PO_4Et	−1	0.976	0.70
3	4-MeI	+ Me–PO_4Et	−1	0.981	0.63
4	4-MeI	+ Ph–PO_4	−2	0.987	0.41
5	4-MeI	+ *t*Bu–PO_4	−2	0.986	0.47
6	4-MeI	+ Me–PO_4	−2	0.991	0.47
7	4-MeI	+ Me–PP–PO_4	−2	0.990	0.34
8	ImH	+ Me–PO_4	−2	0.987	0.81
9	Arg	+ Me–PP–PO_4	−2	0.983	1.03
10	N^G-MeArg	+ Me–PP–PO_4	−2	0.985	0.94
11	Crn	+ Me–PP–PO_4	−3	0.980	1.19
12	4-MeI	+ Me–PP–PO_4 (HK)	−2	0.991	0.67

5.4.1. Training

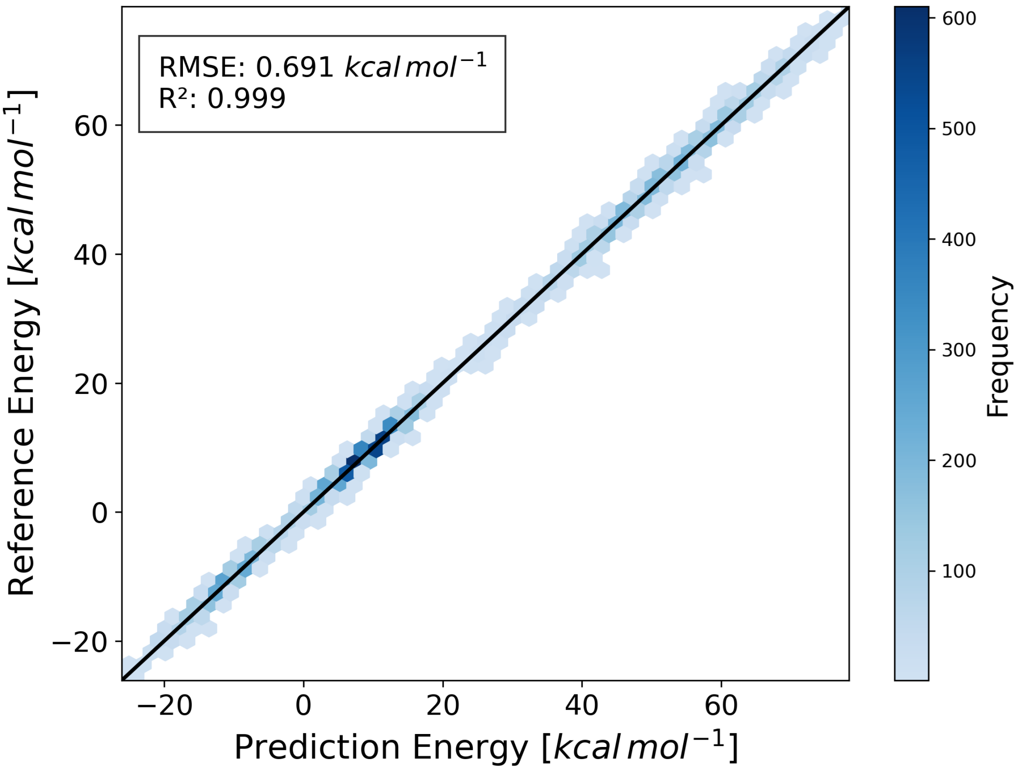
Δ-learning HDNNP models were trained on (i) individual reaction datasets and (ii) pooled datasets grouped by the *total charge* of the QM region (all −1, all −2), as well as their combinations (−1 & −2; and −1/−2/−3). Despite the network architecture not being explicitly charge-aware, pooling by charge and across charges yields consistently high accuracy. Models trained on all charge −1 data, all charge −2 data, and their union show excellent parity; adding the (small) charge −3 set (creatine) to the pooled training preserves performance. Representative parity plots are provided in Fig. 28 (separate models, total charge −1 and −2) and Fig. 29 (mixed models, −1/−2 and −1/−2/−3).



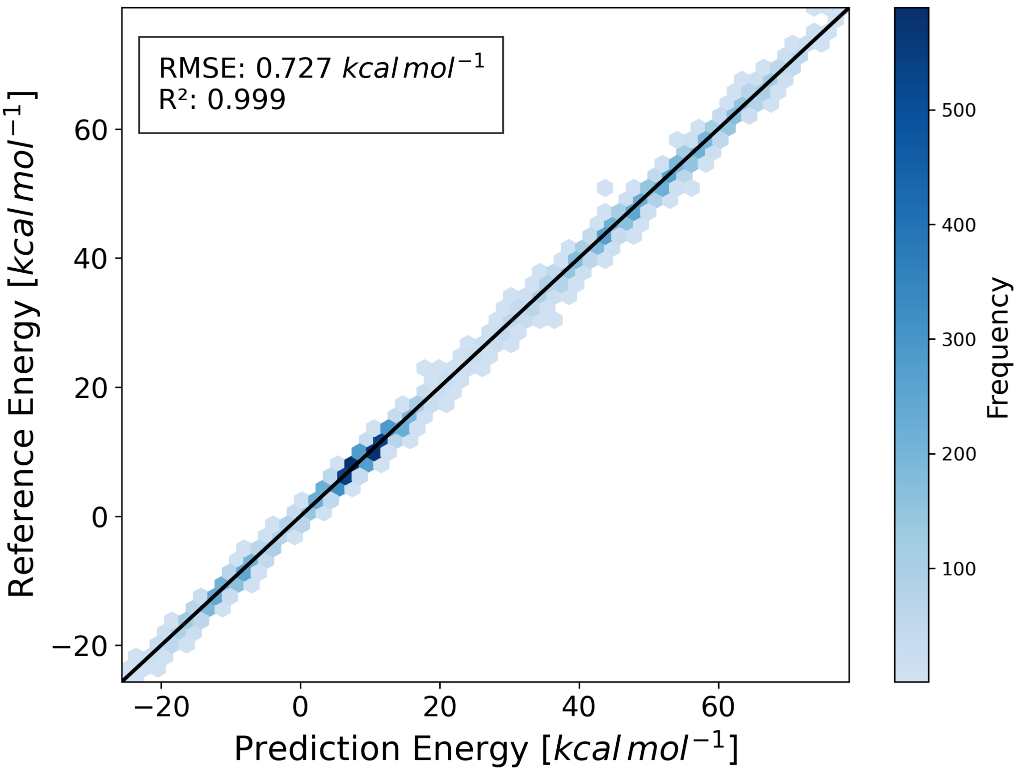
(a) Total charge −1

(b) Total charge −2

Figure 28.: HDNNP Δ-learning parity plots for the HDNNP Δ-learning models grouped by *total charge* of the QM region. Separate models for ML data of total charge −1 (left) and −2 (right).



(a) Mixed model (total charge −1 and −2 combined)



(b) Mixed model (total charge −1, −2, and −3 combined)

Figure 29.: HDNNP Δ-learning parity plots for the HDNNP Δ-learning models grouped by *total charge* of the QM region. Top: model for ML data of total charge −1 and −2. Bottom: model trained on the union with total charge −3.

5.4.2. Cross-Validation

Table 14 summarizes leave-one-out cross-validation (LOOCV), where one reaction is held out and the model is trained on the remaining reactions grouped by total charge(s) (-1, -2, or $-1/-2$). The left-out reaction is always excluded from the corresponding training pool. A dagger (†) after -2 indicates the -2 subset excluding Mg-counterion systems (reactions 7, 9, 10, and 12). Metrics are the coefficient of determination R^2 and RMSE.

Selected Ph, *t*Bu, Me reactions (4–6). For Reaction 4 (Ph leaving group), performance depends strongly on the training pool: using the mixed pool $-1/-2^\dagger$ yields $R^2_{\mathrm{CV}} = 0.582$ with RMSE = 2.345 kcal mol^{-1}, whereas training on $-2^\dagger$ only or -1 only leads to $R^2_{\mathrm{CV}} < 0$ with larger RMSE. The parity plots in Fig. 30 show a clear linear trend with a pronounced constant offset whenever the pool is mismatched. Reaction 5 (*t*Bu leaving group) follows the same pattern (Fig. 31): the mixed pool $-1/-2^\dagger$ provides the best generalization ($R^2_{\mathrm{CV}} = 0.816$, RMSE = 1.64 kcal mol^{-1}), while $-2^\dagger$ alone is near-random, and -1 alone is inadequate ($R^2_{\mathrm{CV}} < 0$). For Reaction 6 (Me leaving group), the same tendency is observed once the corresponding plots are added (Fig. 32): the mixed pool $-1/-2^\dagger$ performs best, followed by $-2^\dagger$, whereas -1 alone fails to generalize ($R^2_{\mathrm{CV}} < 0$).

His and histidine kinase. For the His system (Reaction 7; isolated Triphosphate), both pools generalize, with reasonable performance from the -2 pool compared to $-1/-2$; see Fig. 33. The histidine kinase system (Reaction 12) also achieves good CV on the -2 pool; see Fig. 34. These two cases indicate that restricting training to the closest charge class can be beneficial when the target shares specific physicochemical features (e.g., counterions or coordination environments) with the training data.

Systematic offset in cross-application. Across several folds with low or negative R^2_{CV}, the parity plots nevertheless show a nearly linear relation accompanied by an approximately constant energy shift relative to the diagonal. This suggests a bias dominated by an additive offset rather than a change in slope. While such a shift is critical for absolute energies, its impact on forces is expected to be smaller because forces are energy gradients; a detailed force analysis is presented separately.

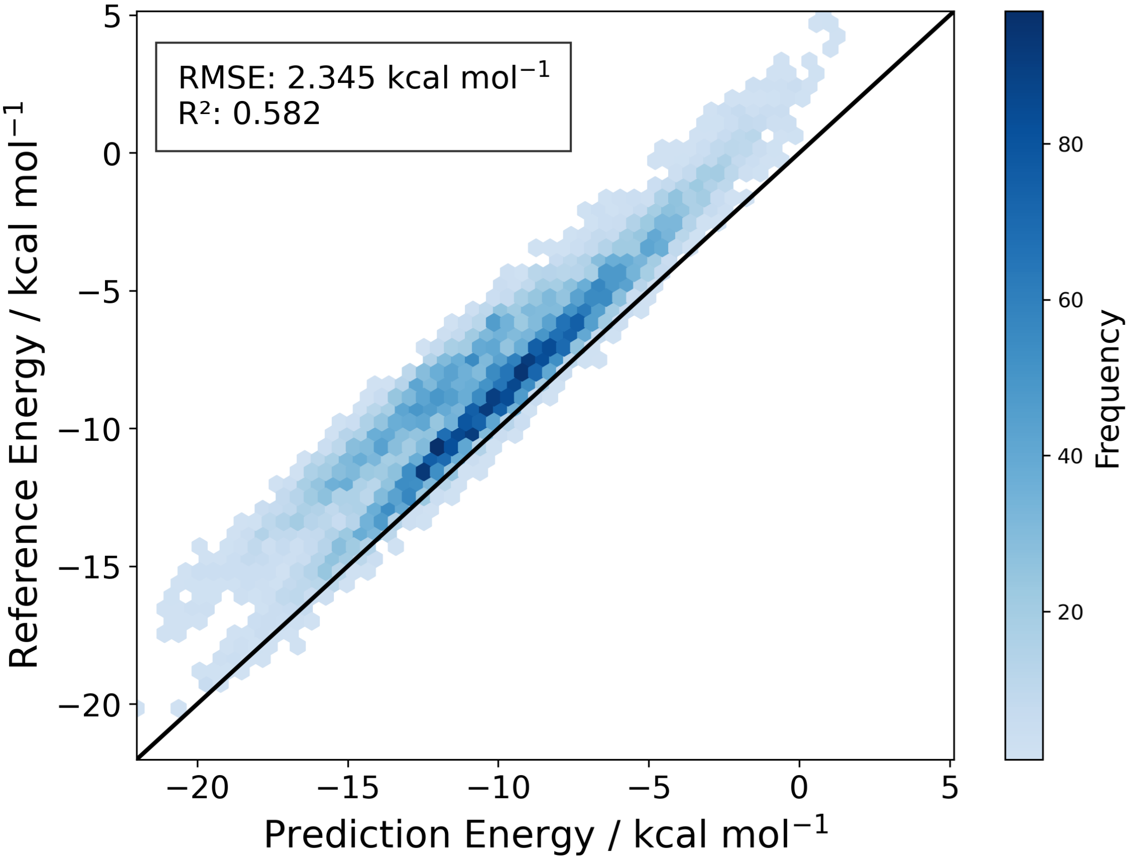
Implications for pooling strategy.

1. For -2 noEt reactions (4–6), mixing -1 and $-2^\dagger$ in the training pool consistently improves CV relative to using $-2^\dagger$ alone, indicating useful transfer between neighboring charge states when Mg-counterion systems are excluded.
2. For His/HK (7, 12), training on -2 alone yields the strongest CV, consistent with a closer mechanistic match to the target systems.

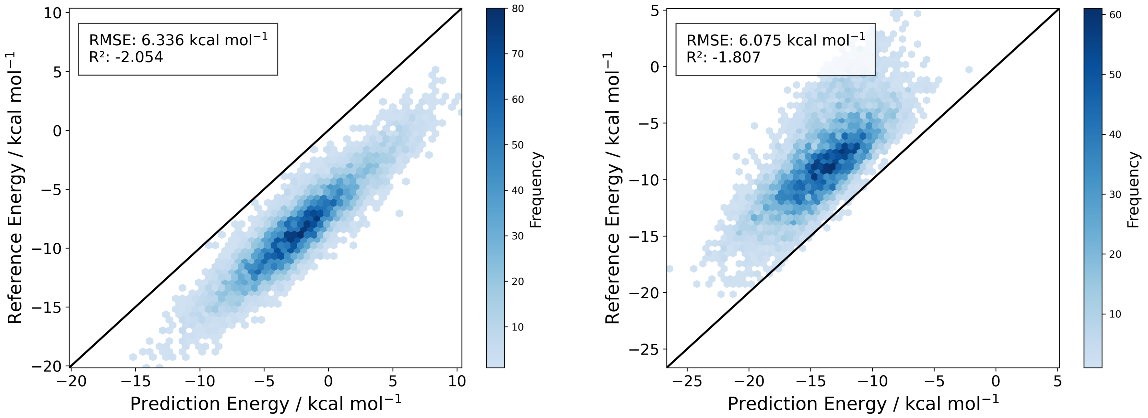
3. Negative R^2_{CV} values indicate clear extrapolation failure; however, the associated plots commonly reveal a near-constant bias that could be corrected by a simple offset calibration if desired.

Table 14.: Leave-one-out cross-validation (one reaction held out per fold). The *Train pool* indicates the set of molecules grouped by their total charge(s) (e.g., −1, −2, or −1/−2) used for training; in each fold the left-out molecule itself is excluded from that pool, even if it belongs to the same charge group. A dagger † after −2 denotes the −2 subset *excluding* reactions that require a Mg counterion (reactions 7, 9, 10, and 12). Metrics: coefficient of determination R^2 and RMSE in kcal mol^{-1}.

	Left-out reaction	**Charge**	**Train pool**	R^2 (train)	RMSE (train)	R^2 (CV)	RMSE (CV)
1	4-MeI + Ph–PO_4Et	−1	-1 / -2†	0.995	0.52	0.644	2.34
			-2†	0.996	0.56	< 0	17.57
			-1	0.982	0.62	< 0	11.12
2	4-MeI + *t*Bu–PO_4Et	−1	-1 / -2†	0.998	0.50	0.754	2.25
			-2†	0.996	0.56	< 0	15.98
			-1	0.996	0.65	< 0	9.97
3	4-MeI + Me–PO_4Et	−1	-1 / -2†	0.977	1.64	0.835	1.89
			-2†	0.996	0.56	< 0	16.67
			-1	0.997	0.68	< 0	11.90
4	4-MeI + Ph–PO_4	−2	-1 / -2†	0.973	1.62	0.582	2.35
			-2†	0.984	0.54	< 0	6.34
			-1	0.996	0.67	< 0	6.08
5	4-MeI + *t*Bu–PO_4	−2	-1 / -2†	0.997	0.54	0.816	1.65
			-2†	0.948	2.25	0.048	3.75
			-1	0.996	0.67	< 0	3.90
6	4-MeI + Me–PO_4	−2	-1 / -2†	0.997	0.53	0.839	1.60
			-2†	0.998	0.39	0.387	3.14
			-1	0.996	0.67	< 0	5.51
7	4-MeI + Me–PP–PO_4	−2	-1 / -2	0.999	0.75	0.740	3.66
			-2	0.999	0.82	0.896	2.31
8	ImH + Me–PO_4	−2	-1 / -2†	0.978	1.62	0.883	1.21
			-2†	0.995	0.74	< 0	2.23
			-1	0.996	0.67	0.153	3.24
9	Arg + Me–PP–PO_4	−2	-1 / -2	0.999	0.66	0.912	2.38
			-2	0.999	0.75	0.911	2.39
10	MeArg + Me–PP–PO_4	−2	-1 / -2	0.999	0.65	< 0	14.43
			-2	0.999	0.66	< 0	9.42
11	Crn + Me–PP–PO_4	−3	-1 / -2	0.999	0.65	< 0	38.76
			-2	0.999	0.72	< 0	51.16
12	4-MeI + Me–PP–PO_4	−2	-1 / -2	0.999	0.67	0.829	3.07
			-2	0.999	0.67	0.850	2.87



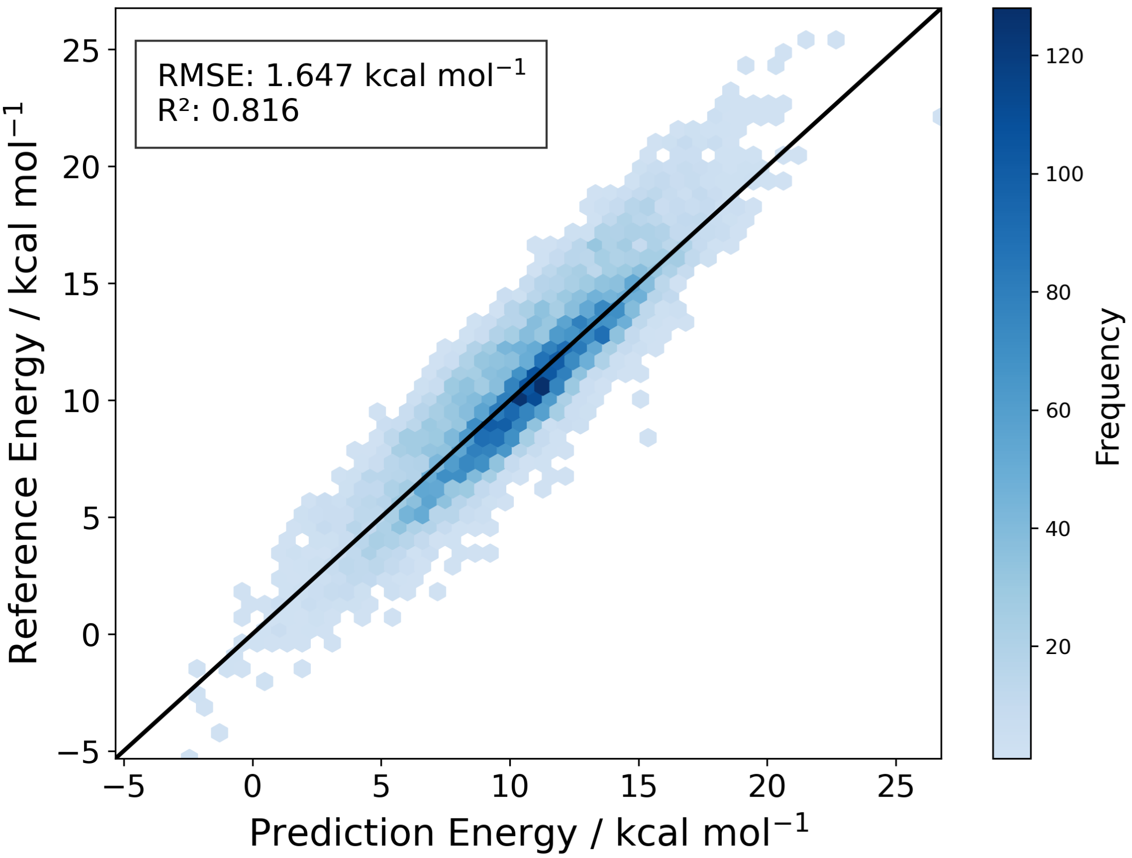
(a) Train pool (total charge): −1/−2†.



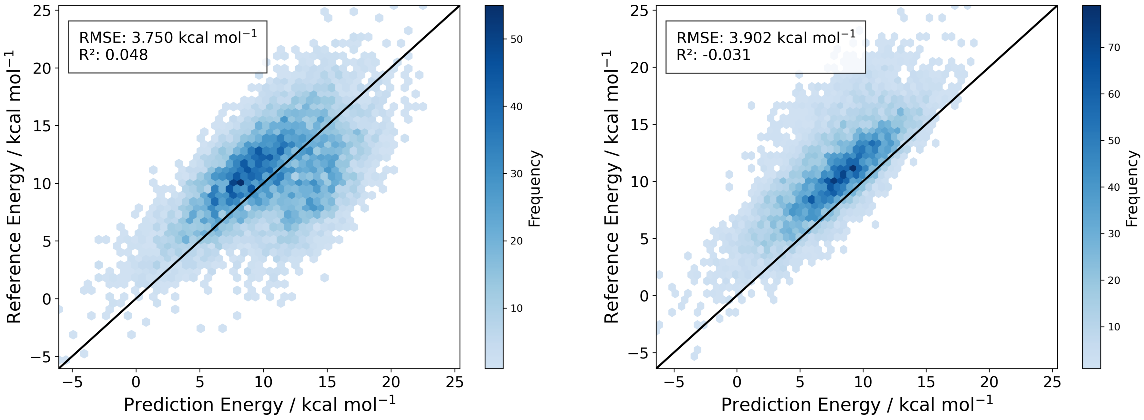
(b) Train pool (total charge): −2† only.

(c) Train pool (total charge): −1 only.

Figure 30.: Leave-one-out CV for **Reaction 4: 4-MeI + Ph–PO$_4$**. Each panel shows a parity plot for the left-out reaction after training on molecules grouped by their total charge: (a) −1/−2†, (b) −2† only, (c) −1 only. The left-out reaction is excluded from the respective pool. Here, −2† denotes the −2 subset excluding Mg-counterion systems (reactions 7, 9, 10, and 12).



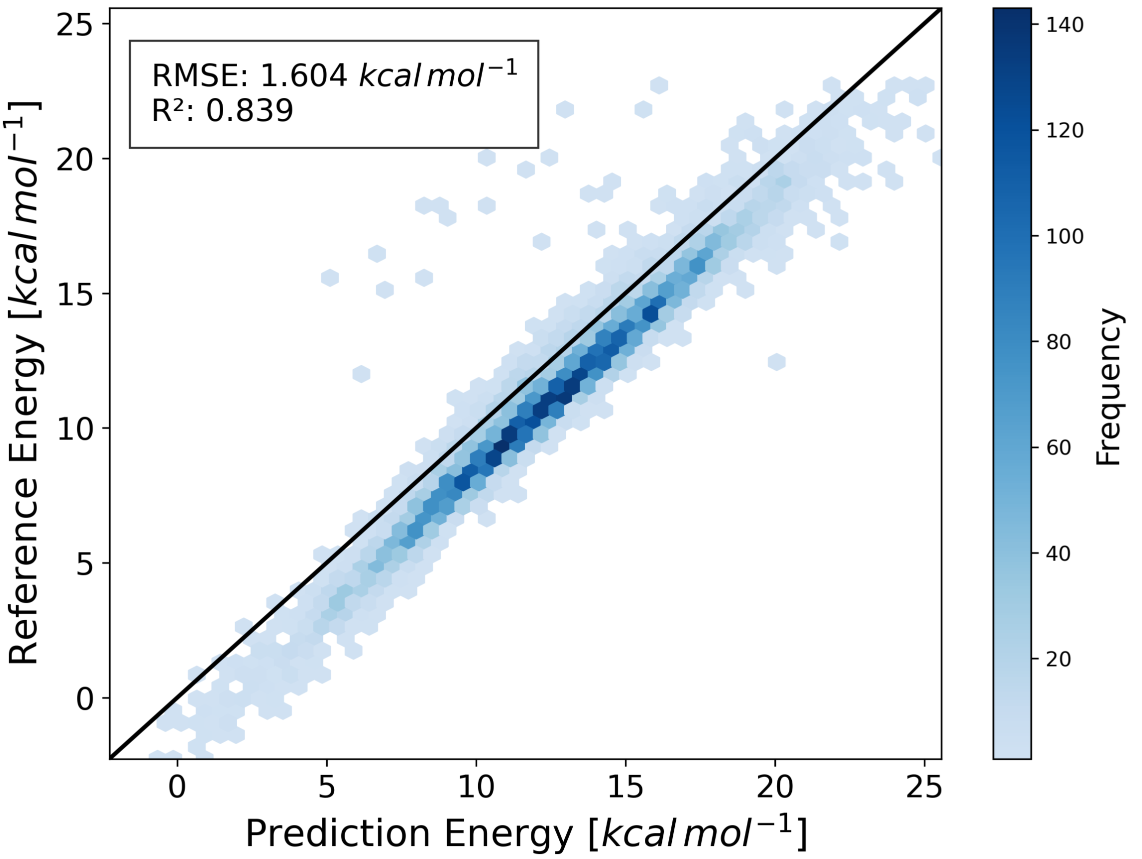
(a) Train pool (total charge): $-1/-2^{\dagger}$.



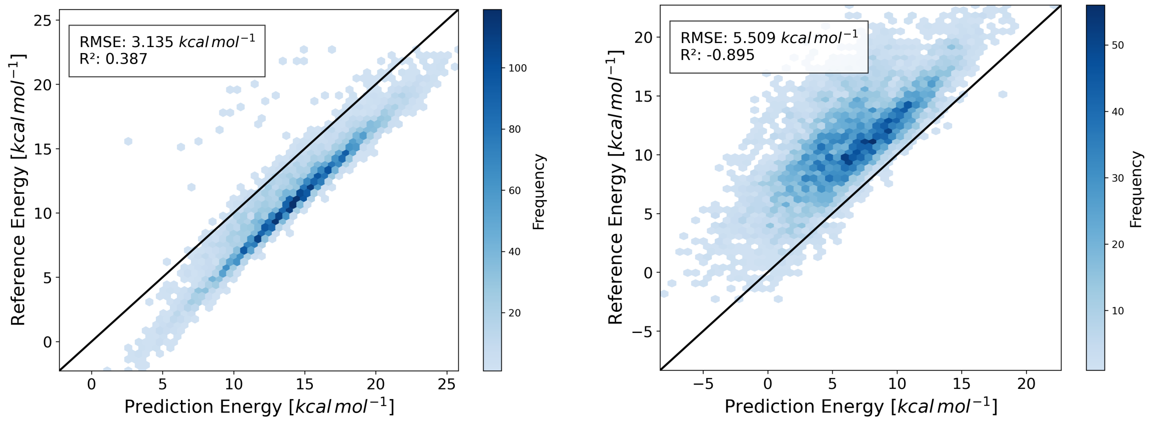
(b) Train pool (total charge): $-2^{\dagger}$ only.

(c) Train pool (total charge): -1 only.

Figure 31.: Leave-one-out CV for **Reaction 5: 4-MeI + *t*Bu–PO$_4$**. Panels correspond to training on (a) $-1/-2^{\dagger}$, (b) $-2^{\dagger}$ only, and (c) -1 only (left-out reaction excluded in each case). $-2^{\dagger}$ excludes Mg-counterion systems (reactions 7, 9, 10, 12).



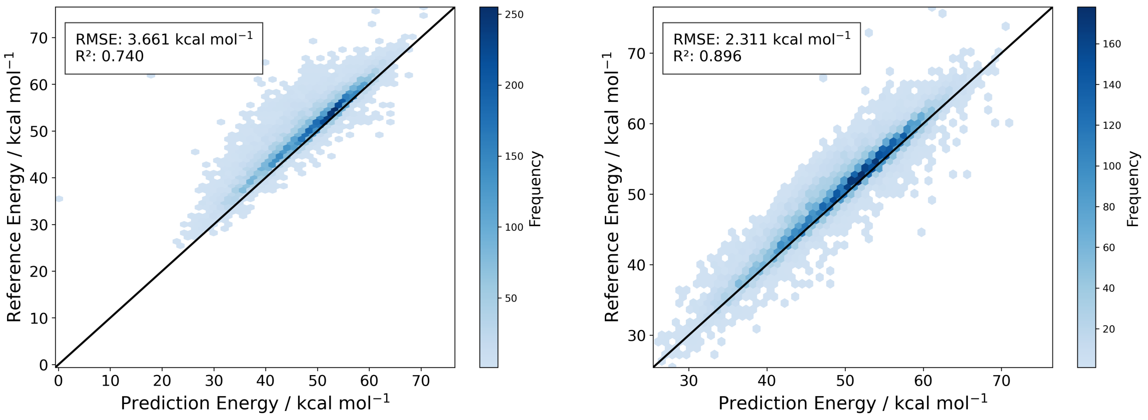
(a) Train pool (total charge): $-1/-2^{\dagger}$.



(b) Train pool (total charge): $-2^{\dagger}$ only.

(c) Train pool (total charge): -1 only.

Figure 32.: Leave-one-out CV for **Reaction 6: 4-MeI + Me–PO$_4$**. Planned panels mirror Reactions 4 and 5: (a) $-1/-2^{\dagger}$, (b) $-2^{\dagger}$ only, (c) -1 only; the left-out reaction is excluded from the respective pool. $-2^{\dagger}$ omits Mg-counterion systems (reactions 7, 9, 10, 12).



(a) Train pool (total charge): −1/−2.

(b) Train pool (total charge): −2 only.

Figure 33.: Leave-one-out CV for **Reaction 7: 4-MeI + Me–PP–PO$_4$**. Parity plots are shown after training on (a) −1/−2 and (b) −2 only. In both cases the left-out reaction is excluded from the training pool.

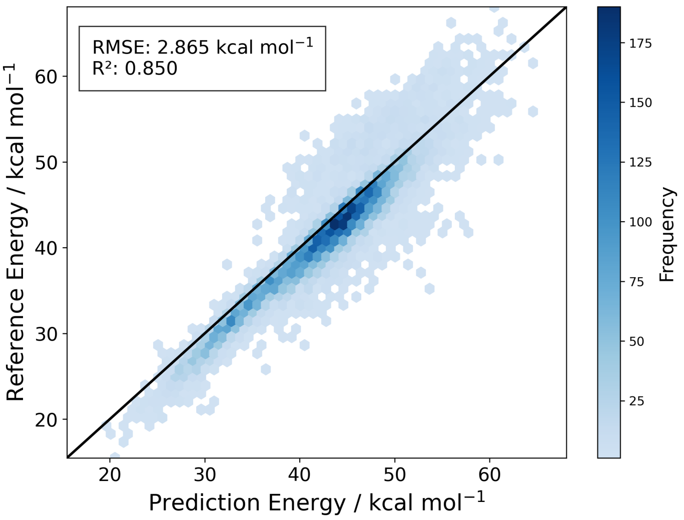


Figure 34.: Leave-one-out CV for **Reaction 12: 4-MeI + Me–PP–PO$_4$ (HK)**. Parity plot evaluated on the left-out reaction after training on molecules with total charge −2 (left-out reaction excluded from the pool).

5.4.3. Δ-QM/MM Application

This section applies on-the-fly Δ-QM/MM corrections for three representative systems using the trained HDNNP Δ-model. For each case, FES are obtained under the same metadynamics protocol, once with DFTB and once with DFTB+NN (Fig. 35). We report shifts in reaction energy (ΔE_{rxn}), activation barrier ($\Delta E^{\ddagger}$), and TS location; in parallel, force consistency over the FES grids is assessed (Figs. 36–38). Runtimes are reported in Table 19. All simulation parameters are compiled in Tables A.15, A.16, and A.17.

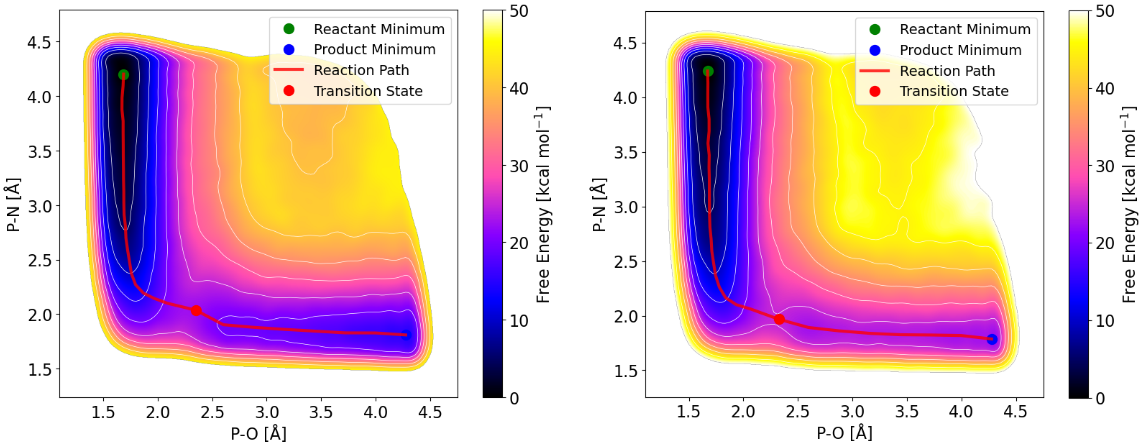
Ethyl–phenyl phosphate leaving group (4-MeI + Ph–PO_4Et). ΔE_{rxn} increases from 15.1 to 22.4 kcal mol^{-1} and the barrier from 22.0 to 24.5 kcal mol^{-1} (DFTB → DFTB+NN). The TS shifts toward the associative corner, (2.35, 2.04) Å → (2.32, 1.97) Å. Forces tighten markedly (RMSE 9.4→2.7 kcal mol^{-1} Å^{-1}; R^2 0.89→0.99).

Isolated triphosphate (4-MeI + Me–PP–PO_4). ΔE_{rxn} becomes less exergonic (−26.5 → −20.3 kcal mol^{-1}) and the barrier rises (8.0→13.5 kcal mol^{-1}). The TS moves away from the compact region, (1.88, 2.18) Å → (1.84, 2.28) Å. Force metrics improve in parallel (RMSE 12.78→4.45 kcal mol^{-1} Å^{-1}; R^2 0.80→0.98).

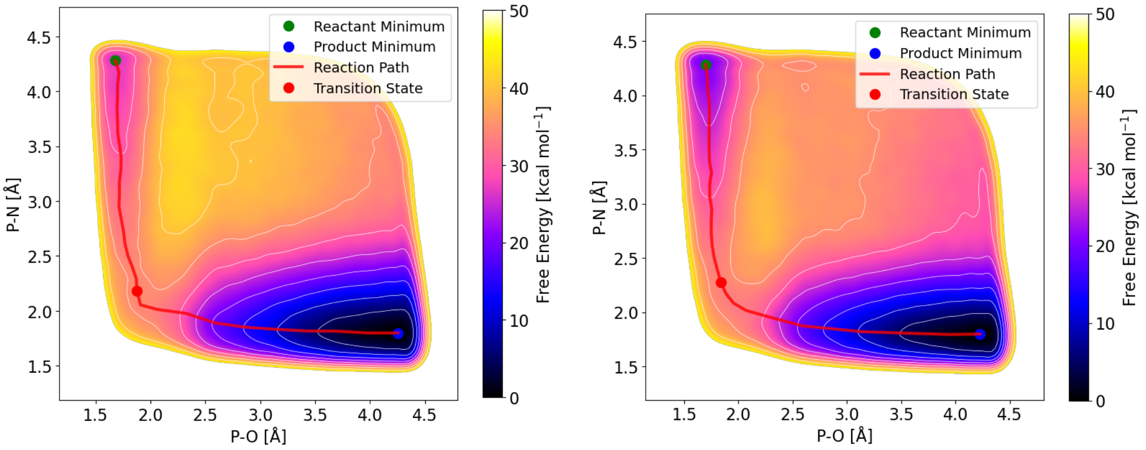
Imidazole nucleophile (ImH + Me–PO_4). ΔE_{rxn} increases from 26.3 to 30.1 kcal mol^{-1} and the barrier from 36.6 to 40.7 kcal mol^{-1}. The TS shifts to longer P–O and shorter P–N, (2.67, 2.18) Å → (2.91, 2.10) Å. Forces likewise improve (RMSE 12.74→3.21 kcal mol^{-1} Å^{-1}; R^2 0.90→0.99).

Table 15.: FES: Δ-QM/MM with HDNNP (DFTB vs. DFTB+NN); CVs: P–O and P–N. Two levels: the DFTB baseline *(with diffusion)* and the on-the-fly Δ–corrected surface (DFTB+HDNNP). Reactant and product energies (E_R, E_P), the transition-state energy (E_{TS}), the activation barrier $\Delta E^{\ddagger} = E_{TS} - E_R$, the reaction energy $\Delta E_{\mathrm{rxn}} = E_P - E_R$, and the TS coordinates. Energies in kcal mol^{-1}, distances in Å.

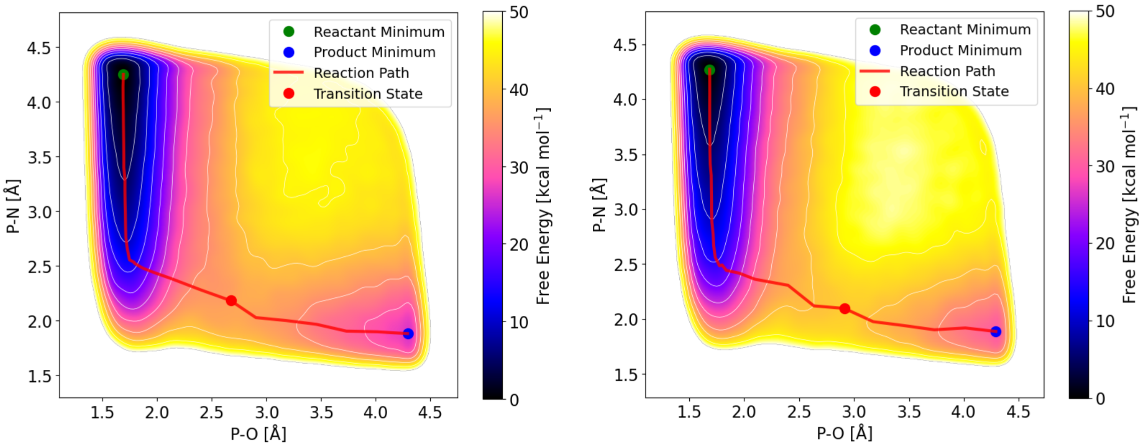
Nr.	Reaction	Level	E_R	E_P	E_{TS}	$\Delta E^{\ddagger}$	ΔE_{rxn}	TS (P–O, P–N)
1	4–MeI + Ph–PO_4Et	DFTB	0.3	15.4	22.3	22.0	15.1	(2.35, 2.04)
		HDNNP	0.3	22.6	24.8	24.5	22.4	(2.32, 1.97)
7	4–MeI + Me–PP–PO_4	DFTB	26.7	0.2	34.6	8.0	−26.5	(1.88, 2.18)
		HDNNP	20.5	0.2	34.0	13.5	−20.3	(1.84, 2.28)
8	ImH + Me–PO_4	DFTB	0.3	26.5	36.9	36.6	26.3	(2.67, 2.18)
		HDNNP	0.3	30.4	41.0	40.7	30.1	(2.91, 2.10)



(a) 4-MeI + Ph–PO$_4$Et, DFTB **(b)** 4-MeI + Ph–PO$_4$Et, HDNNP



(c) 4-MeI + Me–PP–PO$_4$, DFTB **(d)** 4-MeI + Me–PP–PO$_4$, HDNNP



(e) ImH + Me–PO$_4$, DFTB **(f)** ImH + Me–PO$_4$, HDNNP

Figure 35.: Free-energy surfaces (FES) in the (P–O, P–N) plane for three systems: Ethyl-phenyl phosphate with 4-MeI (top), ImH + Me–PO$_4$ (middle), and 4-MeI + Me–PP–PO$_4$ (bottom). Left: DFTB3 baseline; right: DFTB+NN corrections applied on-the-fly. Quantitative metrics are summarized in Table 15.

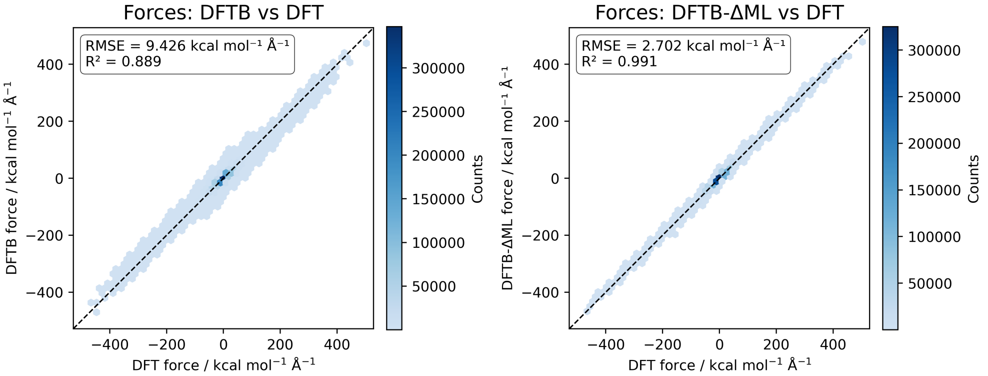


Figure 36.: Force comparison for **4–MeI + Ph–PO$_4$Et**. Left: DFTB3 vs. DFT forces over all FES geometries. Right: on–the–fly DFTB+ΔML vs. DFT forces for the same set. Insets report RMSE (kcal mol^{-1} Å^{-1}) and R^2.

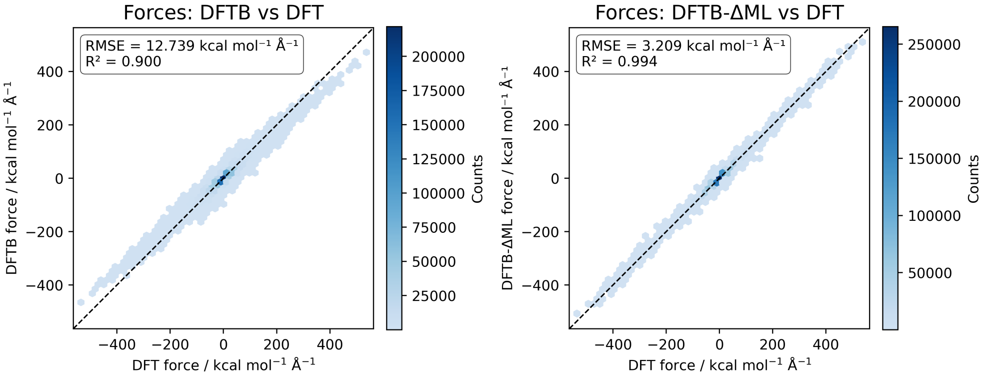


Figure 37.: Force comparison for **4–MeI + Me–PP–PO$_4$**. Left: DFTB3 vs. DFT; right: DFTB+ΔML vs. DFT. Improvements in force RMSE/R^2 mirror the corrected FES topography.

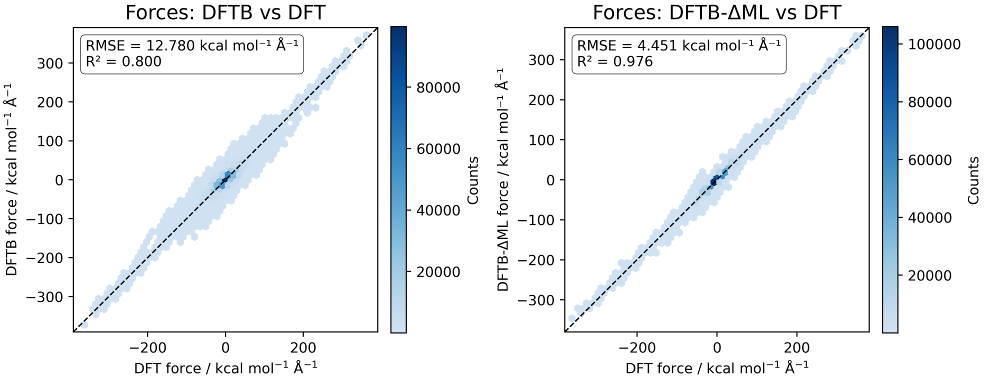


Figure 38.: Force comparison for **ImH + Me–PO$_4$**. Left: DFTB3 vs. DFT; right: DFTB+ΔML vs. DFT. The Δ–correction markedly reduces RMSE and tightens the correlation.

5.5. Δ Network: MACE

In parallel to the local HDNNP, a graph-based, $SO(3)$-equivariant Δ-model based on MACE was evaluated.[39] Two variants were considered: a core (ScaleShift) MACE with local message passing, and a QEq-augmented *MACE-EQ* that couples a charge-equilibration layer to the network and exposes electrostatic observables (charges/ESP). Both models were trained on the atomization-based residual targets used in Sec. 5.4, $\Delta E = E^{\mathrm{DFT}}_{\mathrm{atom}} - E^{\mathrm{DFTB}}_{\mathrm{atom}}$ and, where available, ΔF, with identical splits and quality control. This section reports training behaviour and accuracy; on-the-fly application is deferred to Sec. 5.4.3.

5.5.1. Training

The core model is a compact MACE with two interaction blocks ($l_{\max} = 1$, 128 channels), a 5 Å message cutoff, SiLU nonlinearity, and SWA/EMA for the final weights. Training minimizes a combined energy/force loss under two complementary schedules that trade off energy and force accuracy: (i) an *energy-focused* phase followed by an SWA phase emphasizing forces, and (ii) a *force-focused* phase followed by an SWA phase emphasizing energies. Hyperparameters are summarized in Table A.22; the two loss schedules and their validation errors are listed in Table 16. Atomic reference energies used for Δ-atomization are given in Table A.14.

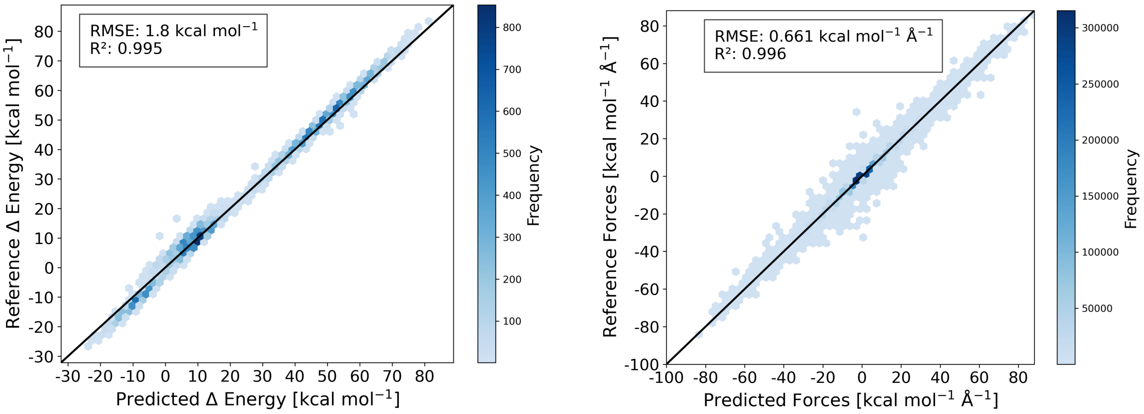
For MACE-EQ, the same data and splits are employed, with a smaller equivariant core (64 channels, $r_{\mathrm{cut}} = 8$ Å) augmented by a QEq/ESP head and a loss that jointly penalizes energies, forces, and electrostatics. Key differences relative to the core setup are summarized in Table A.13.

Table 16.: Core MACE: loss weights and validation errors. Columns report energy:force weights w_E! :!w_F before stochastic weight averaging (pre) and during SWA. RMSE_E in meV/atom; RMSE_F in meV/Å. "Best" denotes the best validation checkpoint *before* SWA; "SWA" denotes the metrics of the SWA(+EMA) averaged final model.

Sched.	$w_E : w_F$ (pre)	$w_E : w_F$ (SWA)	RMSE_E (Best)	RMSE_F (Best)	$\mathrm{RMSE}_E/\mathrm{RMSE}_F$ (SWA)
E→F	1000 : 1	1 : 100	0.7	43.4	2.5 / 28.6
F→E	1 : 1000	100 : 1	6.4	13.5	0.5 / 14.6

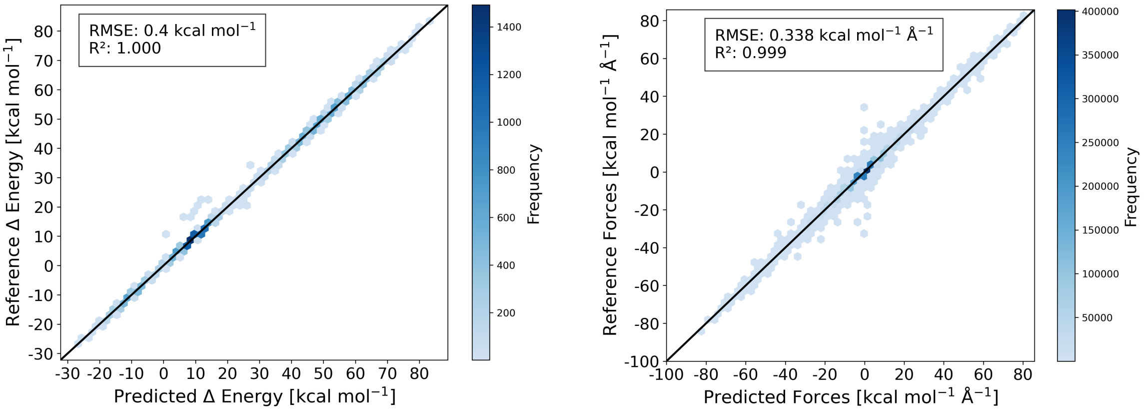
Results (core MACE). Across both schedules we observe the expected trade-off: when energies dominate the pre-SWA phase, force accuracy lags and vice versa. Figures 39a–b (Energy-focus) and 39c–d (Force-focus) visualize the test parity: forces are learned reasonably well (tight diagonal bands), while *energy* residuals show large scatter and a pronounced systematic offset from the diagonal. This indicates that the local message-passing model captures *gradients* of the residual potential reliably, but struggles to place the absolute ΔE consistently across motifs and charge states – a behaviour that mirrors the HDNNP/MACE contrast discussed in Sec. 5.4.

Results (MACE-EQ). Adding charge equilibration (Fig. 40) does *not* rescue the ΔE target in our phosphorylation set: the energy parity collapses to two off-diagonal clusters, and forces degrade compared to the best core setting. Given the smaller backbone, tighter scheduler, and the extra QEq optimization, we interpret this as under-parameterization plus loss competition with electrostatic terms. As a consequence, in the phosphorylation chapter we treat MACE primarily as a *force* learner (for NN/MM tests later), while ΔE–based on-the-fly corrections are provided by the HDNNP.



(a) Energy-focus: ΔE

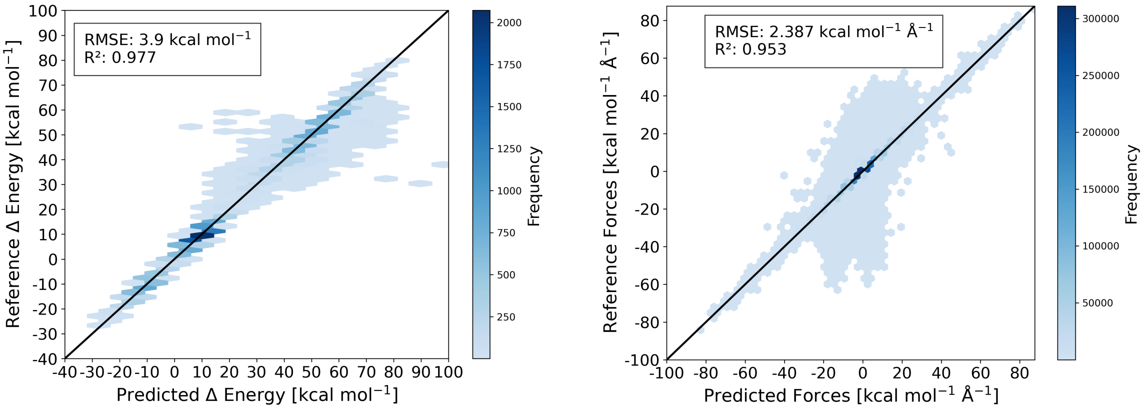
(b) Energy-focus: $\Delta \mathrm{F}$



(c) Force-focus: ΔE

(d) Force-focus: $\Delta \mathrm{F}$

Figure 39.: Core MACE on phosphorylation Δ-learning: parity for two loss schedules. *Top:* pre-SWA energy-focused training (E:F=1000:1) followed by SWA with force emphasis (E:F=1:100). *Bottom:* pre-SWA force-focused (E:F=1:1000) followed by SWA with energy emphasis (E:F=100:1). Values of RMSE/R^2 for each panel are annotated inside the plots.



(a) MACE-EQ: ΔE

(b) MACE-EQ: ΔF

Figure 40.: MACE-EQ on phosphorylation Δ-learning. Force parity remains reasonable but weaker than the best core setting; energy residuals collapse into off-diagonal clusters, indicating a systematic offset and poor calibration for ΔE.

Take-away. For phosphorylation Δ-learning, the equivariant message-passing model provides useful *forces* but unreliable *absolute* ΔE. Charge equilibration (MACE-EQ) does not cure this behaviour in our current configuration. Accordingly, subsequent live tests (Sec. 5.4.3) use MACE primarily as an NN/MM force model, while the energy-only Δ-correction (for Δ-QM/MM) is provided by the HDNNP.

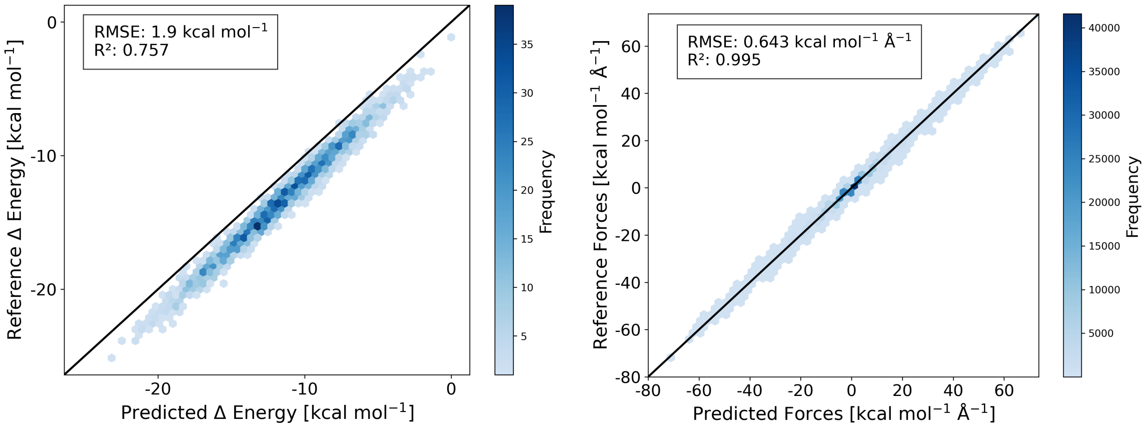
5.5.2. Cross-Validation

Cross-validation is performed in a leave-one-reaction-out (LOOCV) scheme over the twelve reactions: in each fold one reaction is withheld, a single MACE Δ–model is trained on the remaining eleven, and performance is recorded both on the training remainder and on the held–out reaction. Numerical summaries are compiled in Table 17, and the corresponding parity plots for each held-out reaction are shown in Figs. 41–44.

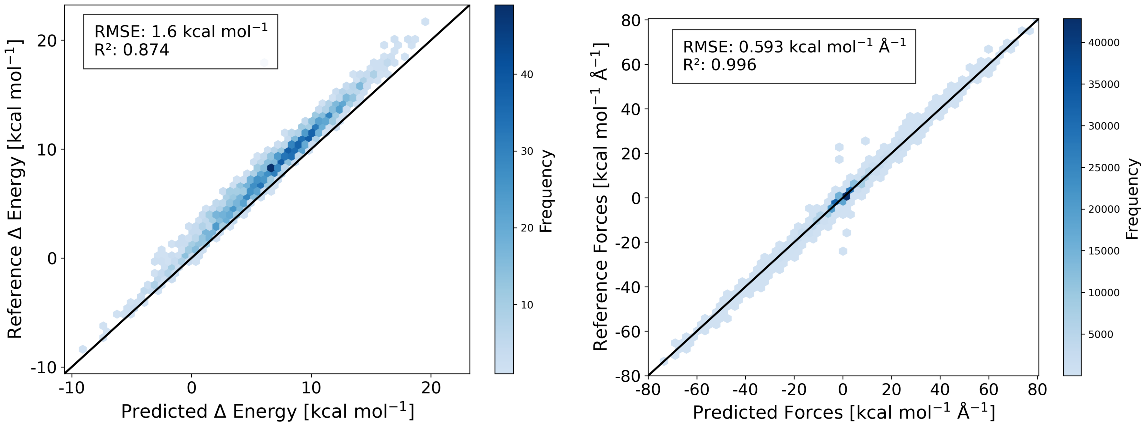
Across reactions, training (on 11/12 reactions) fits show uniformly small energy errors ($\mathrm{RMSE}_E \approx 0.34$–$0.86\,\mathrm{kcal\,mol^{-1}}$) and tight force fits ($\mathrm{RMSE}_F \approx 0.33$–$1.17\,\mathrm{kcal\,mol^{-1}}$), generally lower than the HDNNP training errors reported earlier; the sole training outlier is *Arg + Me–PP–PO*$_4$. In the held–out tests (Table 17; Figs. 41–44) energies vary by reaction: most remain $\lesssim 1\,\mathrm{kcal\,mol^{-1}}$ RMSE_E, while *Ph–PO*$_4$*Et* and *t*Bu–PO$_4$Et show higher RMSE_E (1.91/1.62). *Crn + Me–PP–PO*$_4$ is a clear outlier (negative R^2_E, $\mathrm{RMSE}_E >$ $11\,\mathrm{kcal\,mol^{-1}}$). Despite these energy offsets, forces remain robust in almost all cases ($\mathrm{RMSE}_F \leq 1.76\,\mathrm{kcal\,mol^{-1}}$), with *Crn + Me–PP–PO*$_4$ again the exception ($R^2_F = 0.855$, $\mathrm{RMSE}_F = 4.61$). Visual parity plots verify that most energy discrepancies arise from near–constant shifts relative to the diagonal – a pattern also seen for HDNNP – which does not affect forces, consistent with forces being energy gradients.

Table 17.: Core MACE Δ-model: Leave-one-reaction-out (LOOCV). Left block: model **trained on 11 reactions** (the remainder) – report RMSE only; note that for these training fits $R^2_E \equiv 1.000$ **for all** and $R^2_F = 0.999$ **for all except Arg + Me–PP–PO$_4$ (**0.989**)**. Right block: **held–out reaction (LOOCV)** with R^2 followed by RMSE. RMSE_E in $\mathrm{kcal\,mol^{-1}}$; RMSE_F in $\mathrm{kcal\,mol^{-1}\,Å^{-1}}$.

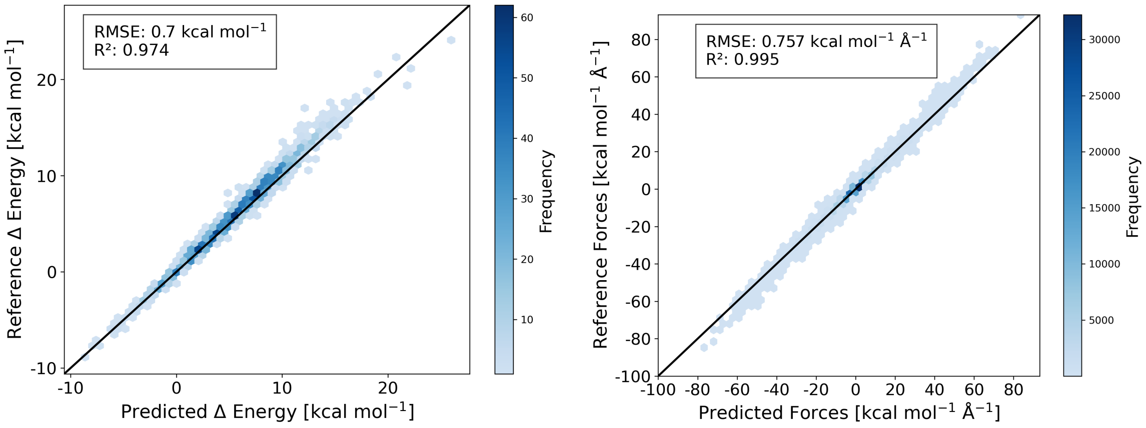
		Trained on 11 reactions		**Held–out (LOOCV)**			
Nr.	**Reaction**	RMSE_E	RMSE_F	R^2_E	RMSE_E	R^2_F	RMSE_F
1	4–MeI + Ph–PO$_4$Et	0.34	0.34	0.757	1.91	0.995	0.64
2	4–MeI + *t*Bu–PO$_4$Et	0.36	0.35	0.874	1.62	0.996	0.59
3	4–MeI + Me–PO$_4$Et	0.37	0.35	0.974	0.74	0.995	0.76
4	4–MeI + Ph–PO$_4$	0.40	0.34	0.956	0.76	0.994	0.79
5	4–MeI + *t*Bu–PO$_4$	0.39	0.35	0.940	0.93	0.997	0.60
6	4–MeI + Me–PO$_4$	0.34	0.35	0.955	1.00	0.994	0.98
7	4–MeI + Me–PP–PO$_4$	0.35	0.34	0.952	1.57	0.997	0.65
8	ImH + Me–PO$_4$	0.40	0.34	0.988	0.40	0.984	1.64
9	Arg + Me–PP–PO$_4$	0.86	1.17	0.936	2.03	0.973	1.76
10	MeArg + Me–PP–PO$_4$	0.36	0.34	0.976	1.21	0.988	1.11
11	Crn + Me–PP–PO$_4$	0.41	0.33	< 0	11.68	0.855	4.61
12	4–MeI + Me–PP–PO$_4$	0.40	0.34	0.979	1.09	0.985	1.47



(a) (1) 4–MeI + Ph–PO_4Et

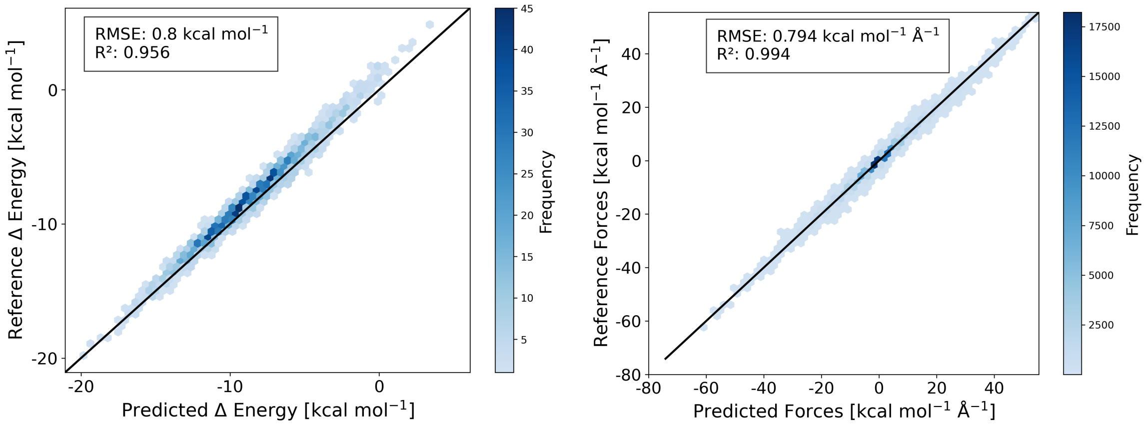


(b) (2) 4–MeI + *t*Bu–PO_4Et

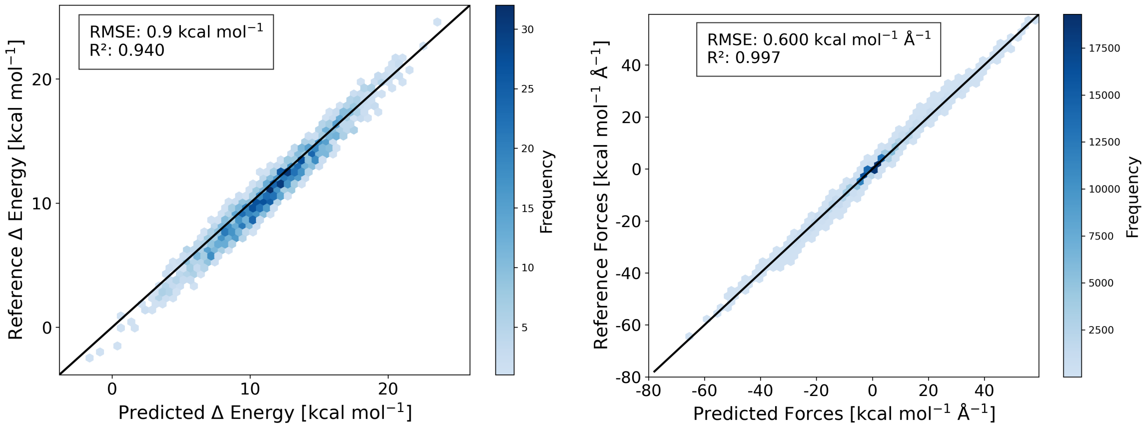


(c) (3) 4–MeI + Me–PO_4Et

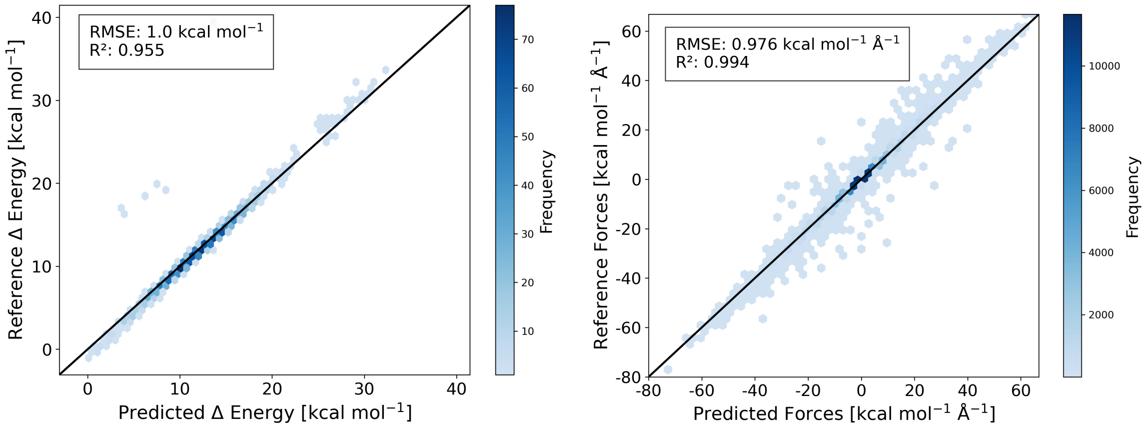
Figure 41.: Core MACE Δ–model: Leave-one-reaction-out cross validation of *(reactions 1 – 3)*. Models are trained on the remaining 11 reactions and validated on the held–out reaction. Each row shows one reaction (left: ΔE, right: forces).



(a) (4) 4–MeI + Ph–PO$_4$

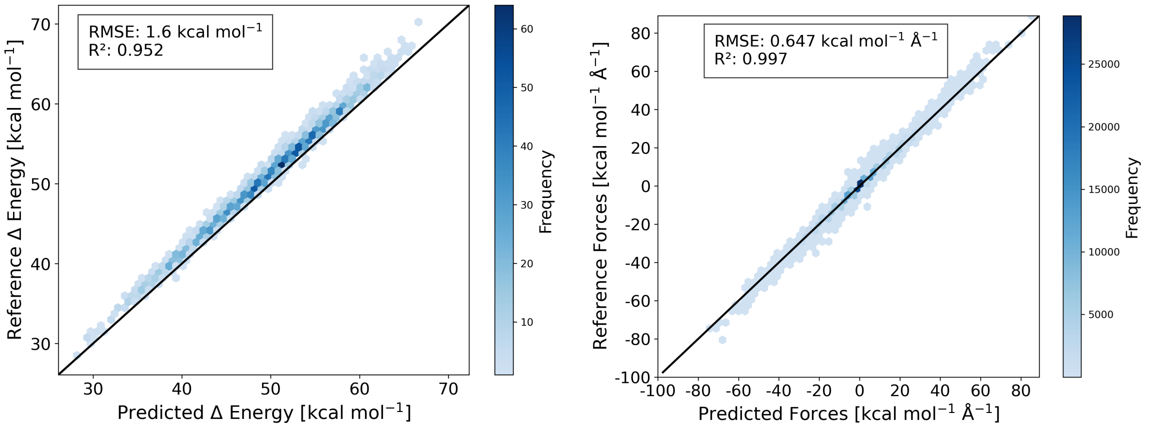


(b) (5) 4–MeI + *t*Bu–PO$_4$

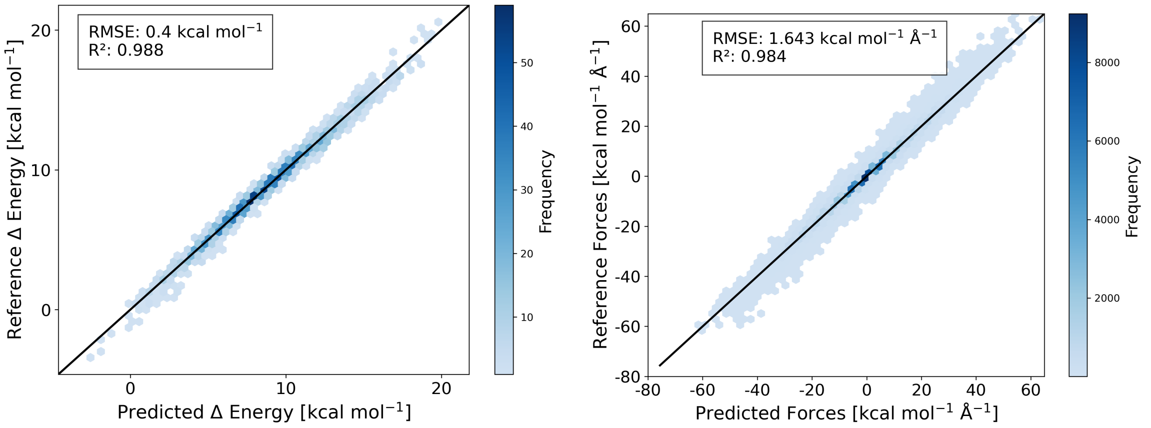


(c) (6) 4–MeI + Me–PO$_4$

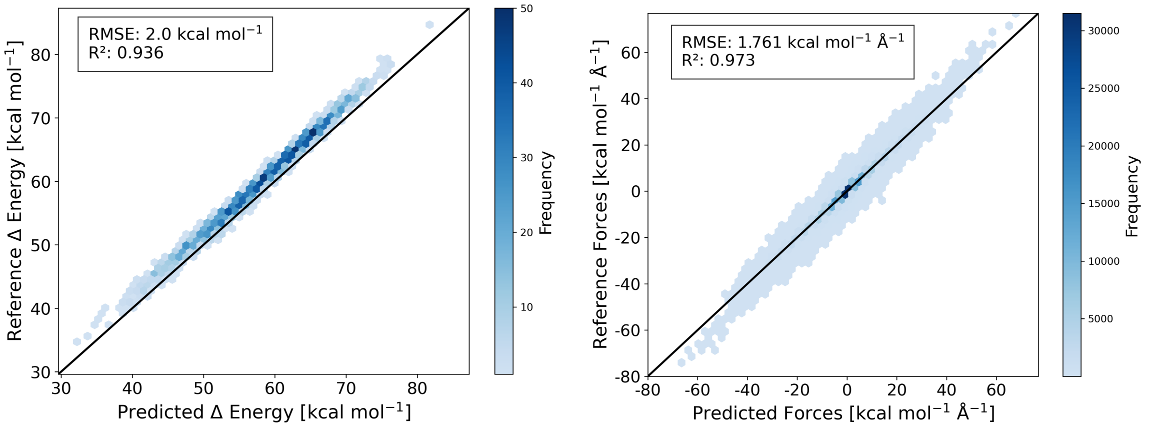
Figure 42.: Core MACE Δ-model: Leave-one-reaction-out cross validation of *(reactions 4 – 6)*. Same setup and plotting conventions as Fig. 41.



(a) (7) 4–MeI + Me–PP–PO$_4$

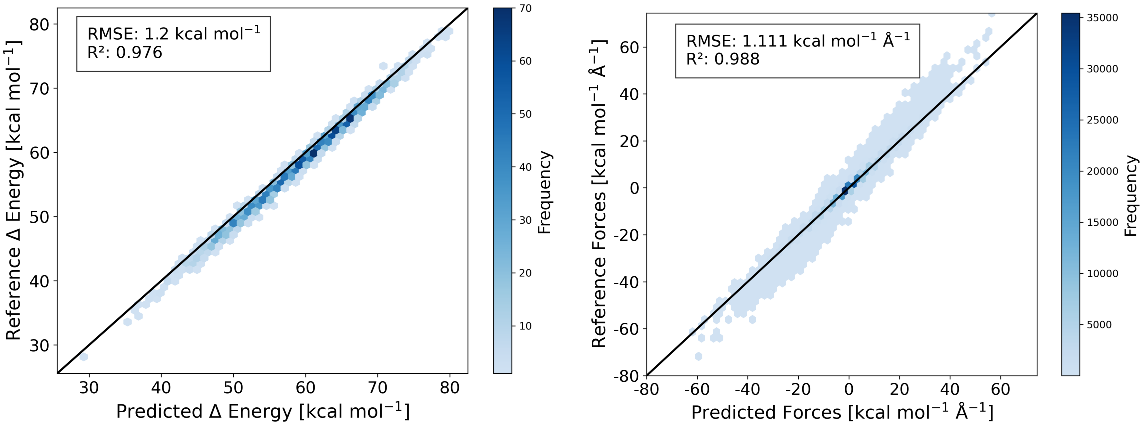


(b) (8) ImH + Me–PO$_4$

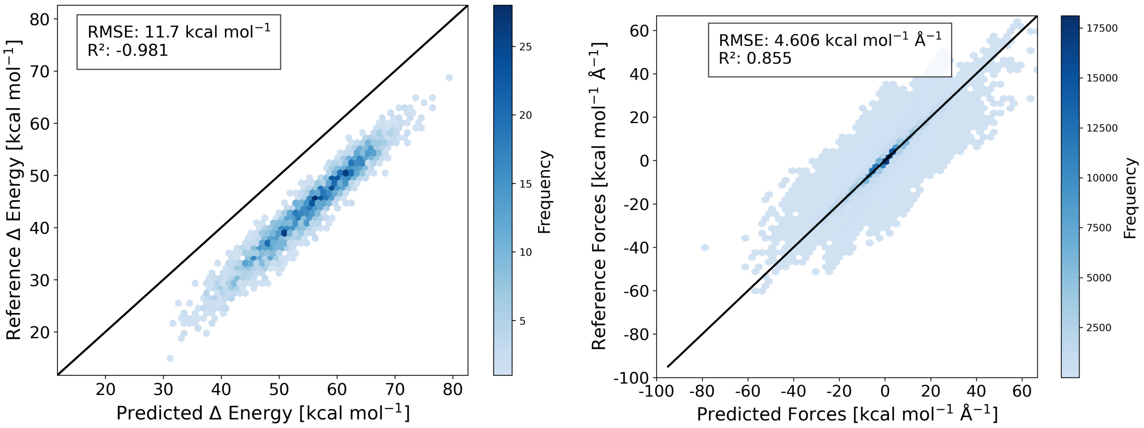


(c) (9) Arg + Me–PP–PO$_4$

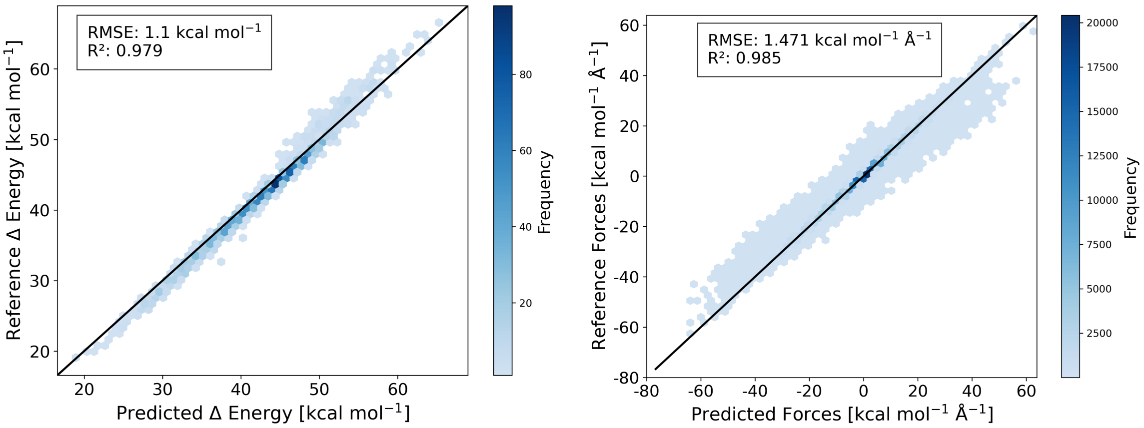
Figure 43.: Core MACE Δ-model: Leave-one-reaction-out cross validation of *(reactions 7 – 9)*. Same setup and plotting conventions as Fig. 41.



(a) (10) MeArg + Me–PP–PO_4



(b) (11) Crn + Me–PP–PO_4



(c) (12) 4–MeI + Me–PP–PO_4

Figure 44.: Core MACE Δ-model: Leave-one-reaction-out cross validation of *(reactions 10 – 12)*. Same setup and plotting conventions as Fig. 41.

5.5.3. Δ-QM/MM Application

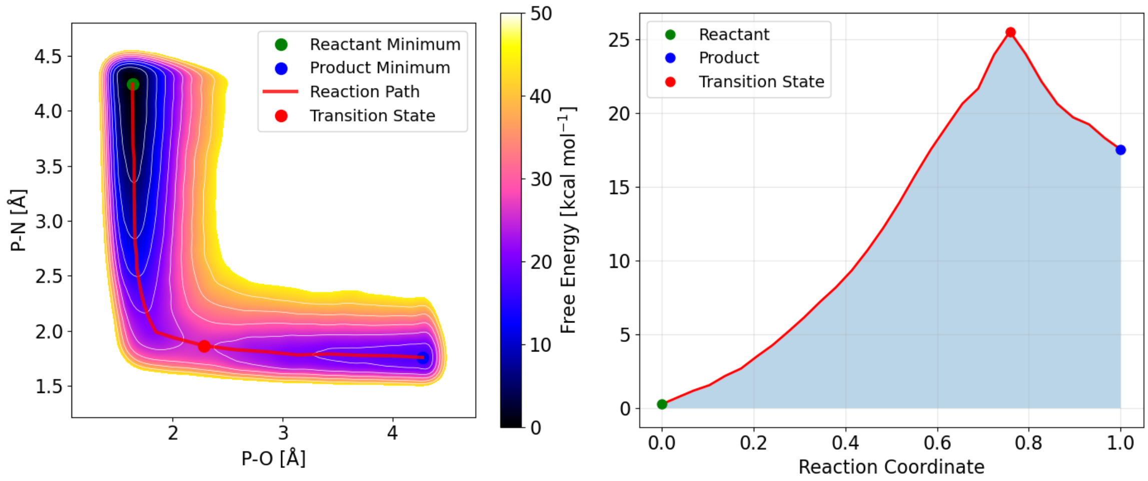
This section mirrors Sec. 5.4.3: all collective variables, metadynamics parameters, and post-processing are identical for strict comparability; the only change is the learned head used for the Δ correction. Using the core MACE Δ model in the QM step, Fig. 45 gathers the three benchmark reactions in one figure; each row shows a two-panel composite with the 2D FES in (P−O, P−N) on the left and the corresponding 1D profile along that path on the right. In all cases the landscapes exhibit the characteristic L-shaped associative S_N2 channel with a reactant basin at short P–O and long P–N, a product basin at long P–O and short P–N, and an associative corner along the pathway. Energies in Table 18 report reactant/product energies, the transition-state energy, activation barrier $\Delta E^{\ddagger} = E_{TS} - E_R$, reaction energy $\Delta E_{rxn} = E_P - E_R$, and the TS coordinates in the FES. Runtimes are reported in Table 19. Simulation parameters are compiled in Tables A.15, A.16, and A.17.

Table 18.: FES: Δ-QM/MM with MACE: reaction-path energetics extracted from the core MACE Δ-QM/MM free-energy surfaces (CVs: P–O and P–N). Reactant and product energies (E_R, E_P), transition-state energy (E_{TS}), activation barrier $\Delta E^{\ddagger}$, reaction energy ΔE_{rxn}, and TS coordinates. Energies in kcal mol^{-1}, distances in Å. DFTB baseline (with diffusion); HDNNP and MACE on-the-fly Δ-corrected surfaces.

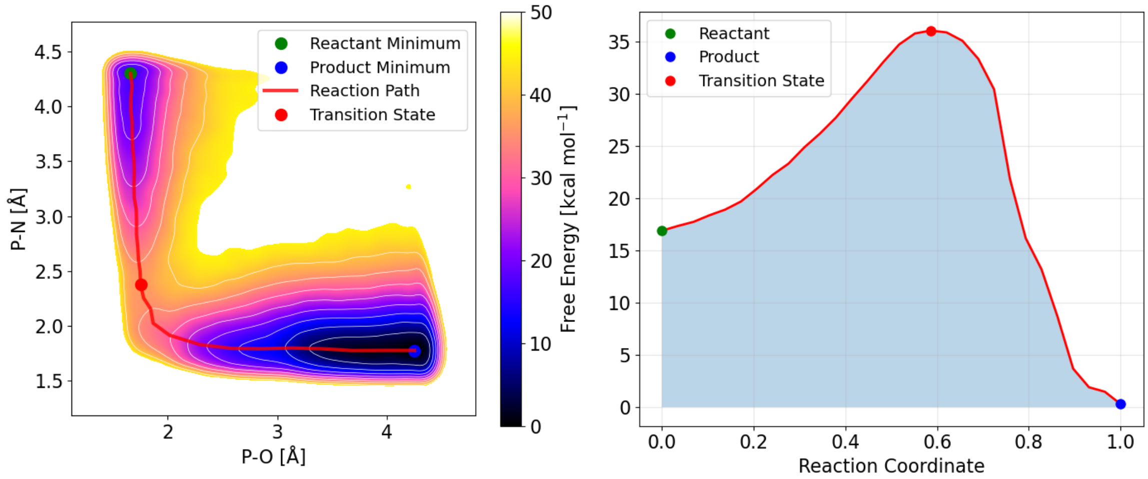
Nr.	Reaction	Level	E_R	E_P	E_{TS}	$\Delta E^{\ddagger}$	ΔE_{rxn}	TS (P–O, P–N)
1	4–MeI + Ph–PO$_4$Et	DFTB	0.3	15.4	22.3	22.0	15.1	(2.35, 2.04)
		HDNNP	0.3	22.6	24.8	24.5	22.4	(2.32, 1.97)
		MACE	0.3	17.5	25.5	25.2	17.3	(2.29, 1.86)
7	4–MeI + Me–PP–PO$_4$	DFTB	26.7	0.2	34.6	8.0	−26.5	(1.88, 2.18)
		HDNNP	20.5	0.2	34.0	13.5	−20.3	(1.84, 2.28)
		MACE	16.9	0.3	36.0	19.2	−16.6	(1.75, 2.38)
8	ImH + Me–PO$_4$	DFTB	0.3	26.5	36.9	36.6	26.3	(2.67, 2.18)
		HDNNP	0.3	30.4	41.0	40.7	30.1	(2.91, 2.10)
		MACE	0.4	35.2	45.6	45.2	34.8	(2.60, 1.96)

Table 19.: Δ-QM/MM Application: runtime per walker for the benchmark reactions. Wall time per nanosecond for DFTB (with diffusion) benchmark; “HDNNP” and “MACE” are on-the-fly Δ–corrected runs.

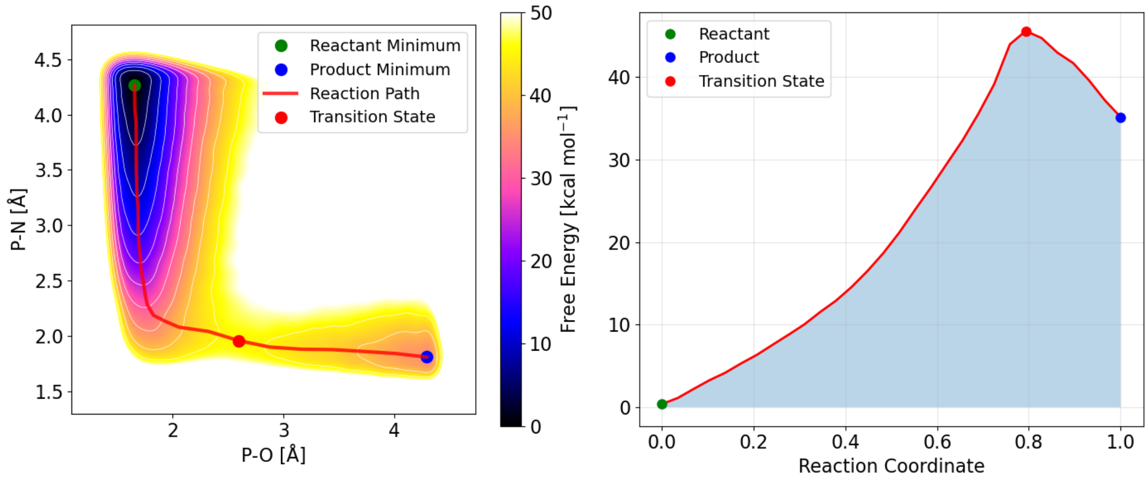
Nr.	Reaction	Level	Runtime (h ns^{-1})	Walkers
1	4–MeI + Ph–PO$_4$Et	DFTB	26.0	1/CPU node
		HDNNP	270.0	1/CPU node
		MACE	90.0	2/GPU
7	4–MeI + Me–PP–PO$_4$	DFTB	22.0	1/CPU node
		HDNNP	190.0	1/CPU node
		MACE	82.0	2/GPU
8	ImH + Me–PO$_4$	DFTB	27.0	1/CPU node
		HDNNP	47.0	1/CPU node
		MACE	72.0	2/GPU



(a) 4–MeI + Ph–PO_4Et



(b) 4–MeI + Me–PP–PO_4



(c) ImH + Me–PO_4

Figure 45.: On-the-fly Δ–QM/MM metadynamics with the core MACE Δ head. Each subfigure contains two panels: left, the 2D free-energy surface in (P–O, P–N) with the reaction path and stationary points; right, the 1D free-energy profile along that path. Simulation parameters match Sec. 5.4.3. The landscapes exhibit the characteristic L-shaped associative S_N2 channel.

5.6. Discussion

The results in this chapter support the feasibility of elevating semi-empirical QM baselines for phosphoryl transfer by a local Δ−ML correction trained on atomization residuals. The curated datasets (Sec. 5.3) combine enhanced sampling with strict single-point quality control and per-family z-standardization. Histogram diagnostics and expected extreme-value ranges (Fig. 24) indicate that label scaling behaves as anticipated for sample sizes of $O(10^4)$.

Target 1: DFTB with on-the-fly HDNNP Δ correction approaching DFT quality. For the HDNNP Δ−network, test errors are uniformly low across reaction families and charge-pooled settings (Table 13), indicating that the learned correction is precise enough to place per-frame energies close to the ωB97M-V target. When deployed on-the-fly in Δ-QM/MM (Sec. 5.4.3), the corrected FES systematically correct characteristic DFTB3 biases: for 4−MeI + Ph−PO_4Et, the product over-stabilization is mitigated and the barrier rises modestly; for ImH + Me−PO_4, both ΔG and the barrier increase by ~4 kcal mol^{-1} with a TS shift toward the associative corner; for the isolated triphosphate, the strong exergonicity is tempered and the barrier increases by ~6 kcal mol^{-1}. In all cases, the relocation of the TS away from compact, hypervalent geometries is chemically consistent with the high-level reference and with the offsets observed in cross-validation. Taken together, these outcomes show that DFTB3 + HDNNP Δ corrections can track the ωB97M-V landscape closely enough to change mechanistic conclusions in a chemically sensible way.

Target 2: HDNNP generalization across reaction families.. Pooling by total charge is effective despite the model not being explicitly charge aware: separate −1 and −2 pools and their union yield near-identical parity, and inclusion of the small −3 subset does not degrade performance (Table 13). Leave-one-out cross-validation (Table 14) clarifies transfer limits and the pooling strategy. For the −2 set without ethyl group (Reactions 4–6), the mixed pool $-1/-2^{\dagger}$ generalizes best; training on $-2^{\dagger}$ only or −1 only often produces negative R^2_{CV}, yet the parity remains nearly linear with an additive shift – pointing to a constant offset bias rather than slope errors. Crucially, the triphosphate systems (His/HK; Reactions 7 and 12) benefit from −2-only training, consistent with a closer physicochemical match (counterions, coordination). The HK case reaches strong CV scores without HK-specific training data in the pool, indicating that the network *learns transferable phosphoryl-transfer physics* rather than memorizing donor/nucleophile identities.

MACE: strong core model training and underperforming MACE–EQ. Beyond the HDNNP, the core MACE Δ-model trains very well on the pooled phosphorylation data: on the training remainder of 11/12 reactions (leave-one-reaction-out cross-validation) it attains uniformly small errors, $\mathrm{RMSE}_E \approx 0.34$–$0.86$ kcal mol^{-1} and $\mathrm{RMSE}_F \approx 0.33$–$1.17$ kcal mol^{-1} (Table 17), typically lower than the per-reaction errors reported for the HDNNP fits earlier in this chapter. Cross-validation on the held-out reaction performs also competitive: most reactions remain at $\mathrm{RMSE}_E \lesssim 1\,\mathrm{kcal\,mol^{-1}}$ with robust forces ($\mathrm{RMSE}_F \leq 1.76\,\mathrm{kcal\,mol^{-1}}$), while *Ph–PO$_4$Et* and *t*Bu–PO$_4$Et exhibit higher energy RMSE (1.906/1.620) and *Crn + Me–PP–PO$_4$* is the clear outlier (negative R_E^2, $\mathrm{RMSE}_E > 11$; see Fig. 41–44). Where discrepancies appear, the parity plots are dominated by near-constant energy shifts rather than slope errors, which explains why force agreement remains strong. By contrast, the QEq–augmented MACE–EQ variant underperforms on energies in our configuration and slightly degrades forces. Practically, core MACE is therefore a solid force learner with acceptable energy calibration in most held-out cases. In line with this, preliminary NN/MM production tests indicate that free/metadynamics simulations of the phosphorylations are stable when driven by the Δ–MACE model.

Limitations. (i) All corrections inherit systematic differences between the target level (ωB97M–V) and the baseline DFTB3 with special reaction parameters; changing either level will shift absolute numbers. (ii) Defining labels via atomization improves additivity and size consistency but can introduce small biases if fragments and full complexes are not treated exactly the same (geometry, charge, boundary choices). (iii) Reaction families with ill-conditioned free-energy surfaces or frequent SCF failures were excluded. The model is trained on this curated chemical window – transfer outside it is not guaranteed. (iv) The HDNNP used for Δ-QM/MM was trained on energies only. Including forces would likely further stabilize metadynamics and sharpen transition-state localization. (v) Pooling by total charge is only a proxy for true charge awareness. The charge-aware MACE-Eq is not yet ideally integrated for Δ-learning; its electrostatic head currently competes with the $\Delta E/\Delta\mathbf{F}$ objectives and can hurt energy calibration. Better coupling and weighting of electrostatics, or explicit long-range terms, should reduce the near-constant offsets seen in some folds of the leave-one-reaction-out cross-validation.

Simulation time. On-the-fly Δ-corrections raise the cost from ∼22–27 $\mathrm{h\,ns^{-1}}$ for the DFTB baseline to ∼47–272 $\mathrm{h\,ns^{-1}}$ with HDNNP, whereas MACE delivers the correction at ∼72–90 $\mathrm{h\,ns^{-1}}$—i.e., about 2–3× lower overhead than HDNNP (Table 19). Note that the HDNNP runtimes reflect a *CPU-only* implementation (use of TensorFlow ragged tensors forced CPU execution), while the MACE metadynamics simulation employed walker distributed across *four GPUs*; therefore the wall-time comparison likely overestimates the HDNNP overhead relative to a GPU-optimized variant.

Outlook. Future improvements include (i) joint energy/force training for the HDNNP Δ-corrector to improve stability and transition-state localization; (ii) lightweight treatments of long range and charge – either explicit far-field terms or compact QEq heads; (iii) fragment benchmarks at higher levels to calibrate absolute energies; (iv) incorporating environment electrostatics (ESP) into the local HDNNP, e.g., via MM-derived fields/potentials or learned embedding features, since the current model only "sees" the cut QM region. For the equivariant line, a main priority is improving the charge-aware MACE-EQ for Δ-learning: better electrostatic treatment and alternative electrostatic couplings that act as corrections rather than absolute contributions.

5.7. Summary

This chapter demonstrates two central points. First, an on-the-fly Δ-QM/MM scheme that augments DFTB3 with a local, atomization-based HDNNP correction can approach the ωB97M-V energy landscape for phosphoryl transfer: sub-kcal test errors on held-out frames, consistent transition-state shifts, and systematic reduction of known DFTB3 biases translate into tangible changes of free-energy profiles and barriers in three representative reactions. Second, the correction generalizes across reaction families when data are pooled by total charge, with clear guidance from leave-one-reaction-out cross-validation on when mixing charges helps and when specialization is preferable. The histidine-kinase system attains sufficient cross-validated accuracy from training alone, indicating that the network learns transferable phosphoryl-transfer physics. In parallel, a compact MACE model trained on the same Δ labels attains sub–kcal mol^{-1} energy errors on the training fits and, in leave–one–reaction–out tests, remains at or below ~1 kcal mol^{-1} RMSE$_E$ for most reactions, generally outperforming the HDNNP energy fits; notable exceptions are Ph–PO$_4$Et, *t*Bu–PO$_4$Et, and Crn + Me–PP–PO$_4$, where RMSE$_E$ increases to ~1.6–2.0 and ~11.7 kcal mol^{-1}, respectively. Parity plots indicate that many discrepancies arise from an approximately constant energy shift relative to the diagonal, which leaves force learning largely unaffected. Taken together, both MACE and HDNNP constitute viable Δ-ML correctors for on-the-fly Δ-QM/MM free-energy mapping; they provide robust forces with competitive energy accuracy across most families, and the choice between them can be made within the existing workflow.

Part VI

Machine-Learned Bright-State Dynamics for Fluorescence in Explicit Solvent

6.1. Overview

This chapter *reports and analyzes results* for predicting fluorescence spectra of solvatochromic dyes in explicit solvent by *learning and following the emissive bright state* within the S_1–S_3 manifold, remaining robust to state reordering. The learned-QM (NN/MM) setup used here – force model for dynamics on the bright surface and property model mapping structure + MM-field ESP to vertical emission/excitation energies and oscillator strengths – follows the architectural/coupling choices defined in Methods II (Chap. III) and Integration (Chap. IV).

Author Contributions. The present author developed and implemented the complete workflow for fluorescence learning and simulation: explicit-solvent LR-TD-LC-DFTB/QM/MM sampling; generation and curation of bright-state labels with S_1–S_3 identity checks; a force model trained on the bright surface and validated for nanosecond-scale stability; a property model using structure + ESP features to predict emission/excitation energies and oscillator strengths; assembly of ensemble emission spectra linked to solvent-response diagnostics; and an end-to-end, versioned data pipeline for training and inference. An internal master's thesis by Manuel Enns, co-advised by the present author, provided early exploratory ideas that motivated this direction;[221] all methods, data, and models reported here were (re)built from first principles by the present author. The study is established on FR0[42] (Fig. 46) with a brief portability check to retinal (Chap. VII); all training uses TD-LC-DFTB (Secs. 2.3, 2.3.4).

6.2. Introduction

Fluorescence spectroscopy provides a sensitive window into local structure and dynamics in complex molecular environments, with solvatochromic dyes transducing nanoscale changes in polarity, hydrogen bonding, and packing into shifts of absorption and emission.[19] Protein–dye constructs leverage this principle to convert molecular recognition into optical signals for biosensing; a prominent example is the glucose/galactose-binding protein (GGBP), where placement of an environment-sensitive dye at strategic sites yields large glucose-induced fluorescence changes,[20] and simulations have clarified how binding reshapes hydrogen-bond networks and chromophore conformations to guide sensor optimization.[41] A general method is developed and validated to determine fluorescence spectra from explicit-solvent sampling that is *robust to state reordering* among the lowest singlets by directly targeting the emissive bright state; the method is established and tested on the strongly solvatochromic FR0 fluorophore (Fig. 46), a fluorene analogue of Prodan.[42] See Sec. 2.3 for general photophysics and Sec. 2.3.4 for state-identity considerations.

Quantitative prediction of emission energies and line shapes must reconcile (i) environmental inhomogeneity and dynamics, (ii) vibronic structure, and (iii) the identity of the *emissive bright state* along nuclear motion. Ensemble approaches such as the *nuclear*

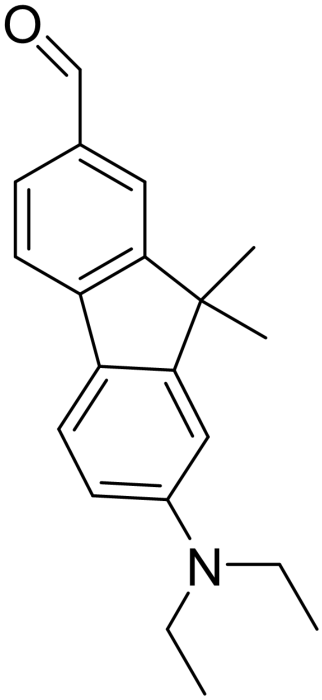


Figure 46.: Chemical structure of FR0.[42]

ensemble approach (NEA) capture inhomogeneous broadening, while *Franck–Condon* (FC) methods capture vibrational progressions; combined ensemble–FC strategies improve fidelity but remain computationally demanding.[59,60] A second challenge is solvent treatment: implicit models often miss specific interactions in heterogeneous environments, whereas explicit-solvent sampling (e.g., QM/MM) is more reliable for charge-transfer states that drive solvatochromism.[43,44] For Prodan-like push-pull dyes (including FR0), two low-lying singlets of different character – locally excited (LE, often ^1L_b) and intramolecular charge transfer (ICT, ^1L_a) – dominate oscillator strength and govern solvatochromism; higher singlets typically relax ultrafast and are thus irrelevant on the ns emission timescale, though occasional bright S_3 episodes can occur in certain environments.[19,42]

The identity-tracking bottleneck is addressed by *learning* the bright state. Using explicit-solvent QM/MM sampling with LR-TD-LC-DFTB on FR0 as training data, a force model is trained to provide forces for dynamics *on the bright surface* across reordering within the S_1–S_3 manifold, and a property model is trained to predict emission energies and oscillator strengths from structure+ESP for spectral simulation – forming an ML–MM scheme that preserves state identity and accelerates spectrum prediction.[16,23,46] All training in this chapter is performed at the (TD-)DFTB level. Details of the learning-based workflow are given in Chap. III.

6.3. Dataset Construction

Two coupled data streams underpin this chapter while details are deferred to Chap. III and Secs. 6.3.1–6.3.2. (i) Pure-solvent controls quantify the MM models via the static permittivity ε_r from dipole-moment fluctuations at 300 K under conducting boundaries (Tab. 20), providing a compact proxy for the ESP feature space used by the networks. (ii) Explicit-solvent LR-TD-LC-DFTB/QM/MM sampling of *FR0* in ten solvents supplies the learning set for the force and property models: 1 ns trajectories with a 100 fs snapshot stride; per snapshot, the lowest N_{ex}=10 singlets are evaluated in the MM field and a frame is retained if $\max_{k\in\{1,2,3\}} f_k \geq 0.7$ (Tab. 21). Feature definitions, targets, units, charge embedding (fixed S_2), data splits, and cross-validation protocols follow Chap. III. The next subsections report ε_r (bias diagnostic; Sec. 6.3.1) and inspect the excited-state sets via structural/electrostatic coverage (Sec. 6.3.2). Detailed settings and parameter provenance are collected in the Appendix: MM parameters and charge models for *FR0* (Table A.18); solvent models and sources (Table A.19); QM/MM MD settings for the free excited–state data-generation runs (Table A.20); and TD–LC–DFTB/Casida single-point settings used to re-evaluate each snapshot (Table A.21).

6.3.1. Solvent: Dielectric Permittivity

Because our environment descriptor is the electrostatic potential (ESP) on the QM atoms (see Methods II, Sec. 3.6), the bulk dipolar response of the solvent model matters. As a compact global diagnostic we compute the static relative permittivity ε_r from total dipole-moment fluctuations using Eq. (2.3) under conducting ("tin-foil") Ewald boundary conditions; MD settings follow Methods I (Sec. 2.5). All simulations reported here were performed at $T = 300$ K with production windows of 1 ns.

Table 20 compares simulated ε_r with literature values at 25°C (298 K) taken from Reichardt and Welton (Appendix A, Table A-1).[56] Small temperature differences between 298 and 300 K can shift absolute numbers slightly, but the dominant trends are systematic.

Table 20.: Solvents: Relative permittivity ε_r from 1 ns trajectories at 300 K (this work) versus literature values at 25 °C (298 K) from Reichardt & Welton.[56] $\Delta = \varepsilon_r^{\text{Sim}} - \varepsilon_r^{\text{Lit}}$.

Solvent	ε_r(Sim)	ε_r(Lit)	Δ	Δ [%]
Water	97.1	78.4	18.7	23.9
DMSO	44.3	46.5	−2.2	−4.6
DMF	22.7	36.7	−14.0	−38.2
Acetonitrile	23.6	35.9	−12.3	−34.3
Methanol	20.0	32.7	−12.7	−38.8
Acetone	10.4	20.7	−10.3	−49.8
Dichloromethane	3.9	8.9	−5.0	−56.3
Ethyl acetate	4.5	6.0	−1.5	−25.2
1,4-Dioxane	1.1	2.2	−1.1	−50.2
n-Hexane	1.0	1.9	−0.9	−46.8

The overall pattern matches expectations for non-polarizable, fixed-charge models: water (tip3p) is overpolarized in our setup (here $\varepsilon_r = 97.1$ vs. 78.36 in the literature), whereas many aprotic organics – especially acetone, DCM, dioxane, and hexane – and several moderately polar solvents (DMF, ACN, MeOH) underestimate ε_r by roughly 25–55 %.[25,56] This reflects known limitations of fixed-charge force fields (missing electronic polarizability and model-dependent dipole correlations).

These deviations matter for spectra because ε_r correlates with the mean and spread of the ESP distribution entering the NN inputs (Methods II, Sec. 3.6). Underestimated ε_r compresses solvent response and tends to underpredict solvatochromic red shifts, whereas overestimated ε_r (as in water) has the opposite tendency. The NN/MM workflow reproduces trends across solvents whose ESP statistics match those seen during training but cannot compensate for systematic bias in the underlying solvent physics. Absolute experiment–theory offsets are therefore interpreted in light of Table 20, and ε_r is used as a control metric for the solvent models. Potential remedies include polarizable force fields, reparametrization of partial charges for key solvents, or simple ESP rescaling calibrated to experimental ε_r.

6.3.2. Dataset Diagnostics

This subsection inspects the *training data* used for the fluorescence models. Everything here concerns the **excited state**: we only retain *bright-state* configurations and analyse their structural statistics. Data are generated with the FR0 workflow in Methods II (Sec. 3.6.5; cf. Sec. 4.3): explicit-solvent TD-LC-DFTB/MM trajectories, Tab. 21, with snapshots saved every 100 fs. For each snapshot the lowest $N_{\mathrm{ex}} = 10$ singlet excitations are evaluated in the MM field and a frame is marked *bright* if

$$\max_{k \in \{1,2,3\}} f_k \;\geq\; 0.7,$$

with S_1–S_3 considered for brightness. Only such bright-state frames contribute to the data set; feature definitions (geometry + per-atom ESP) and partitioning follow Methods II.

Table 21.: Fixed settings for FR0 data generation and bright-state selection.

Quantity	Value
Temperature	300 K
Pressure	1.013 bar
Snapshot stride	100 fs
QM level	TD-LC-DFTB
Singlet excitations evaluated, N_{ex}	10
States considered for brightness	S_1, S_2, S_3
Brightness threshold, f_{thresh}	0.7

To understand what our bright-state training sets look like, we now summarise **free** simulations of *FR0* in ten solvents. We proceed in four steps: (i) Ramachandran-style 2D occupancies for the dialkylamino dihedrals (D_1, D_2), (ii) the 1D distribution of the formyl dihedral D_{O}, (iii) planarity diagnostics relative to the aromatic ring (metric defined in Methods II, Sec. 3.7.7), and (iv) boxplots of total energies and force magnitudes. Figure 47 provides the reference labelling used throughout.

2D torsional coverage across solvents. Figure 48 shows hex-binned occupancies of (D_1, D_2) for ten solvents. Across media the topology is similar – distinct basins in the four quadrant – while the *relative* weights of these basins vary with solvent, reflecting different steric/electrostatic environments around the dialkylamino group.

Formyl torsion D_{O}. Figure 49 shows that the formyl dihedral does not undergo free rotation during the free simulations: distributions are sharply localized – typically near anti (±180°) in polar solvents, with some aprotic/apolar media (e.g., ethyl acetate, hexane) favoring a planar (∼ 0°) basin. The qualitative pattern is reproducible across repeats.

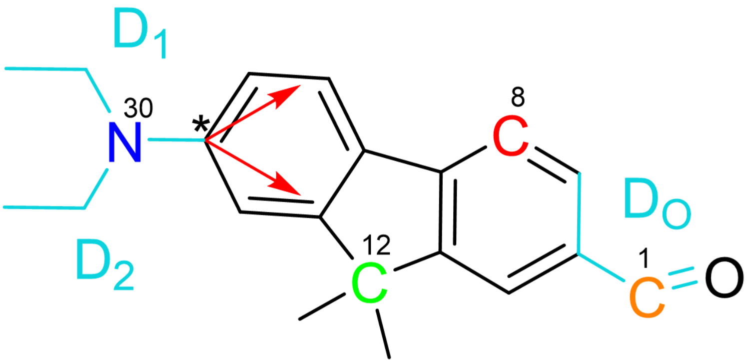


Figure 47.: Reference sketch of *FR0* with the two dialkylamino dihedrals D_1, D_2 (cyan) and the formyl dihedral D_O. The red arrows indicate the ring-plane basis used for planarity diagnostics (Methods II, Sec. 3.7.7). Planarity is monitored for four atoms: the dimethylamino nitrogen N_{30} (blue), a bridgehead carbon C_{12} (green), the ring carbon C_8 (red), and the formyl carbon C_1 (orange).

Planarity around the chromophore. Using the angle-to-plane metric defined in Methods II (Sec. 3.7.7), Fig. 50 compares out-of-plane deviations for four diagnostic atoms (donor N; bridgehead C12; ring C8; formyl C1). Ring carbons remain near-planar, while the donor nitrogen and the formyl carbon show the largest and most solvent-sensitive excursions.

Energy and force summaries. Finally, Fig. 51 summarizes per-frame total energies and force magnitudes for the same free simulations. Distributions are broadly comparable across solvents; differences appear mainly in spread and tails, which mirrors the breadth seen in the torsional/planarity diagnostics.

Electrostatic environment (ESP). The MM electrostatic potential (ESP) sampled at the QM atoms (Methods II, Sec. 3.6.1), we sanity-check its distribution across solvents. Figure 52 shows per-solvent boxplots of the per-atom ESP (kcal mol^{-1}). Polar media (Water, MeOH, DMSO, DMF, ACN) exhibit broader distributions. Moderately polar aprotics (Acetone, DCM, Ethyl Acetate, Dioxane) are narrower, and the non-polar reference (Hexane) is very tight and nearly constant.

Vertical excitation energies and oscillator strengths. For each stored snapshot we evaluate the TD-LC-DFTB vertical excitation energies and oscillator strengths. Figure 53 summarizes the distributions per solvent: the *top* panel shows boxplots of the bright-state vertical excitation energy E_{vert}, the bottom panel shows the corresponding oscillator strengths f. Within each solvent the energy distributions are fairly compact with occasional long tails, indicative of transient state reordering or local structural distortions captured by the QM/MM simulation.

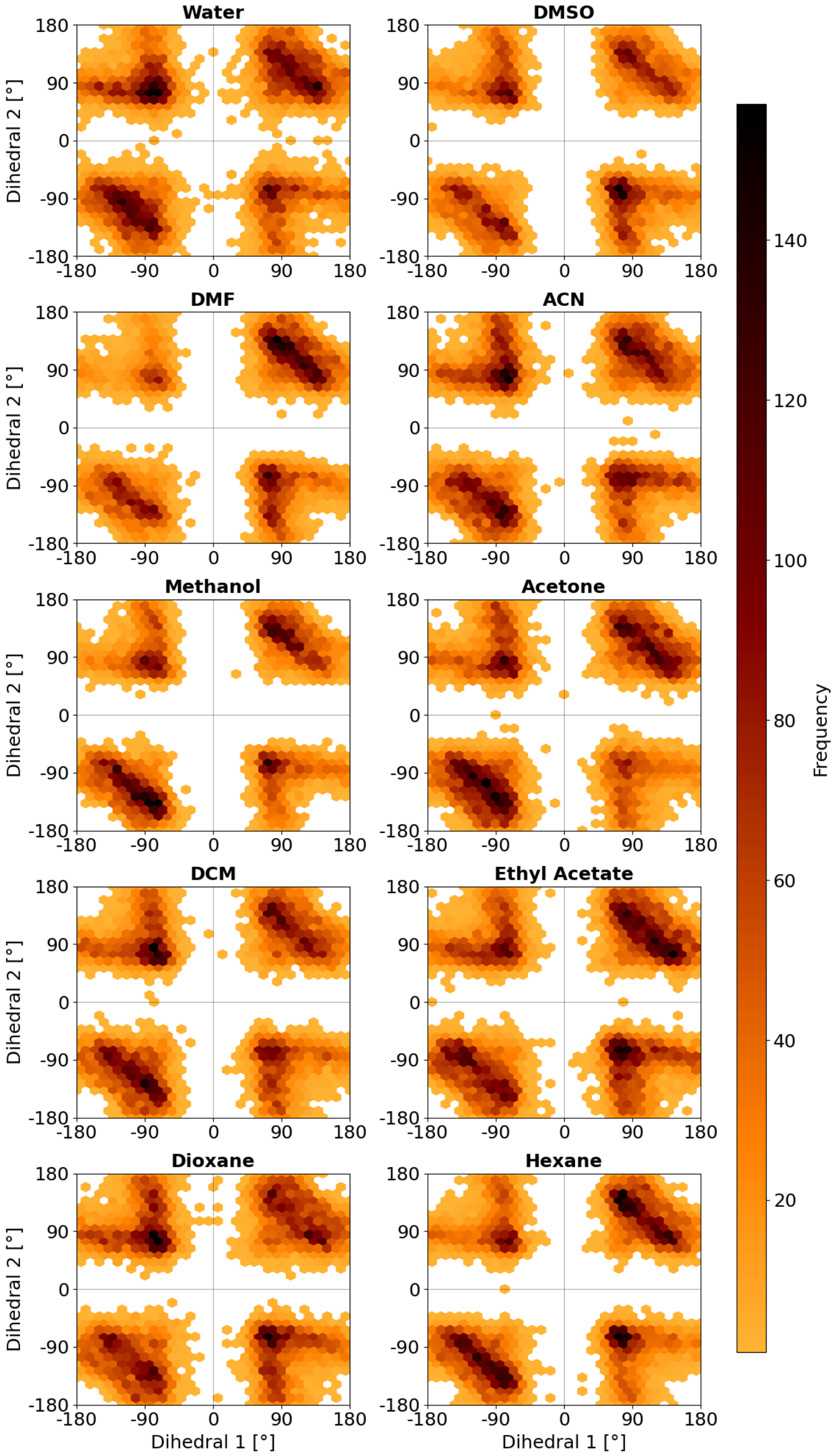


Figure 48.: *FR0* dialkylamino dihedrals (D_1, D_2) in ten solvents (free simulation; hex-binned densities). The overall basin structure is conserved, while solvent changes the population balance between basins.

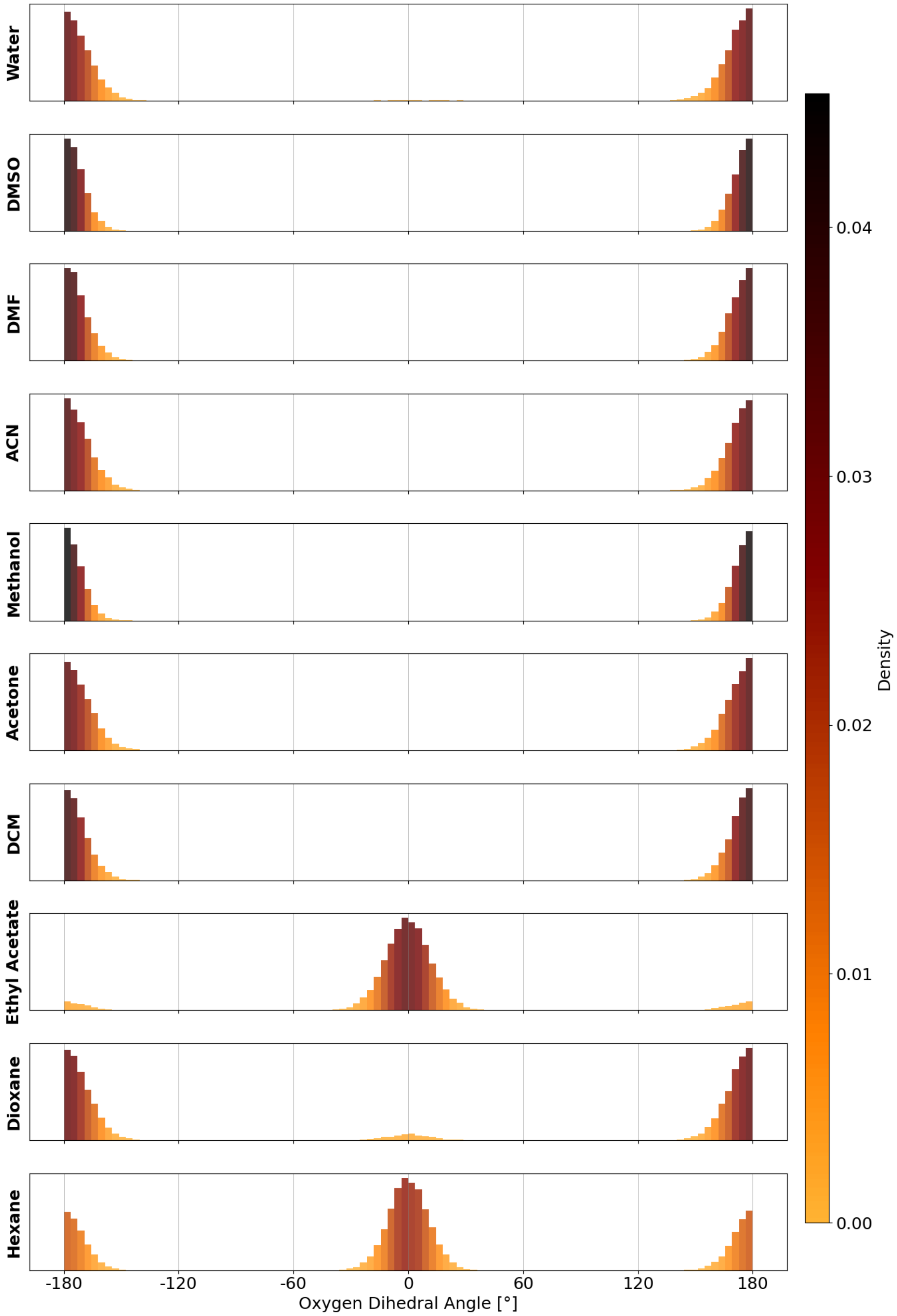


Figure 49.: One-dimensional distributions of the formyl dihedral D_O by solvent. Densities are normalized per panel to make relative peak weights comparable.

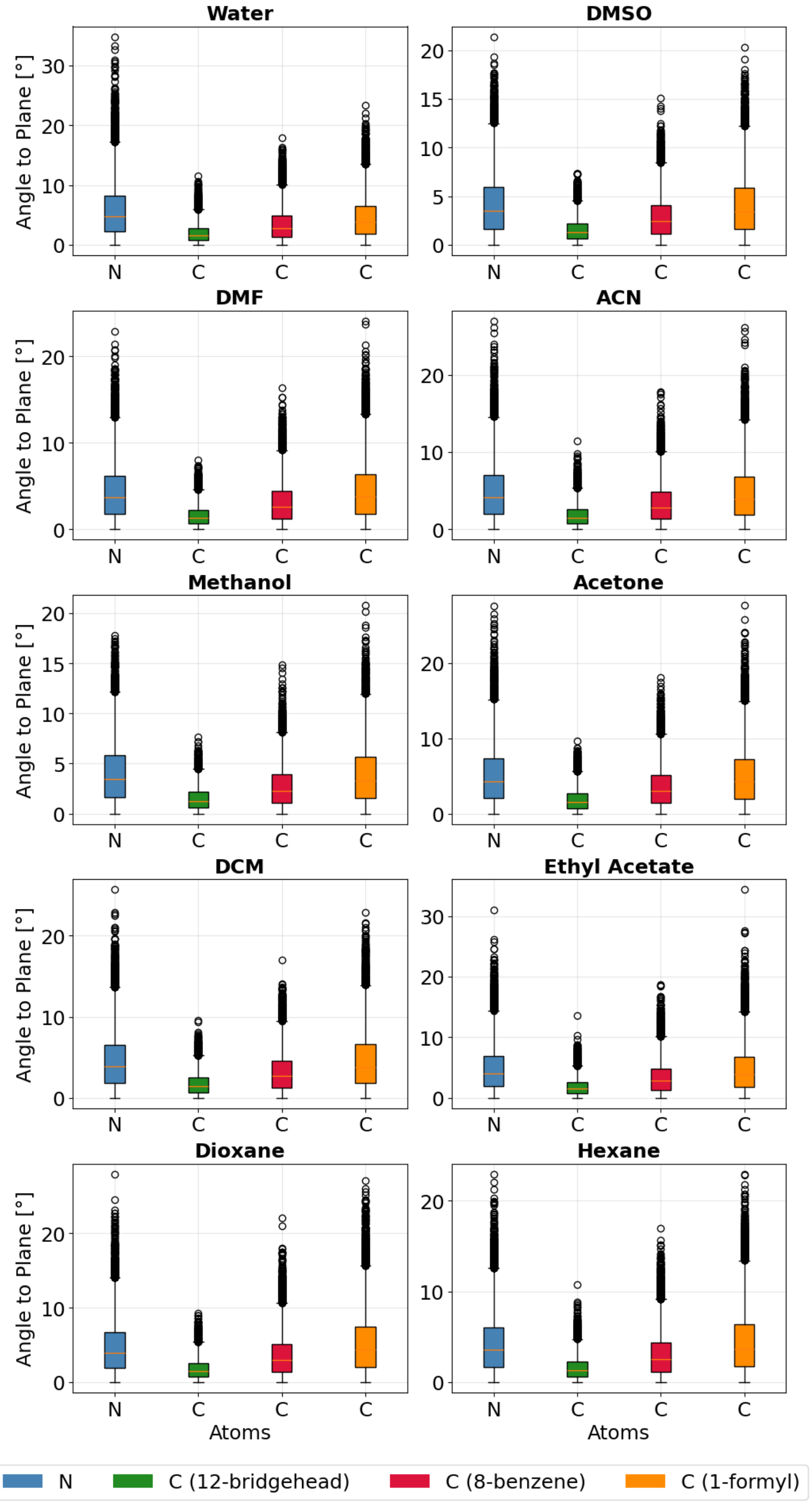


Figure 50.: Angle between each diagnostic atom and the ring plane (degrees; lower is more planar). Labels follow Fig. 47. The bridgehead C12 is the most rigid; N and C1 (formyl) exhibit the widest spreads and solvent dependence.

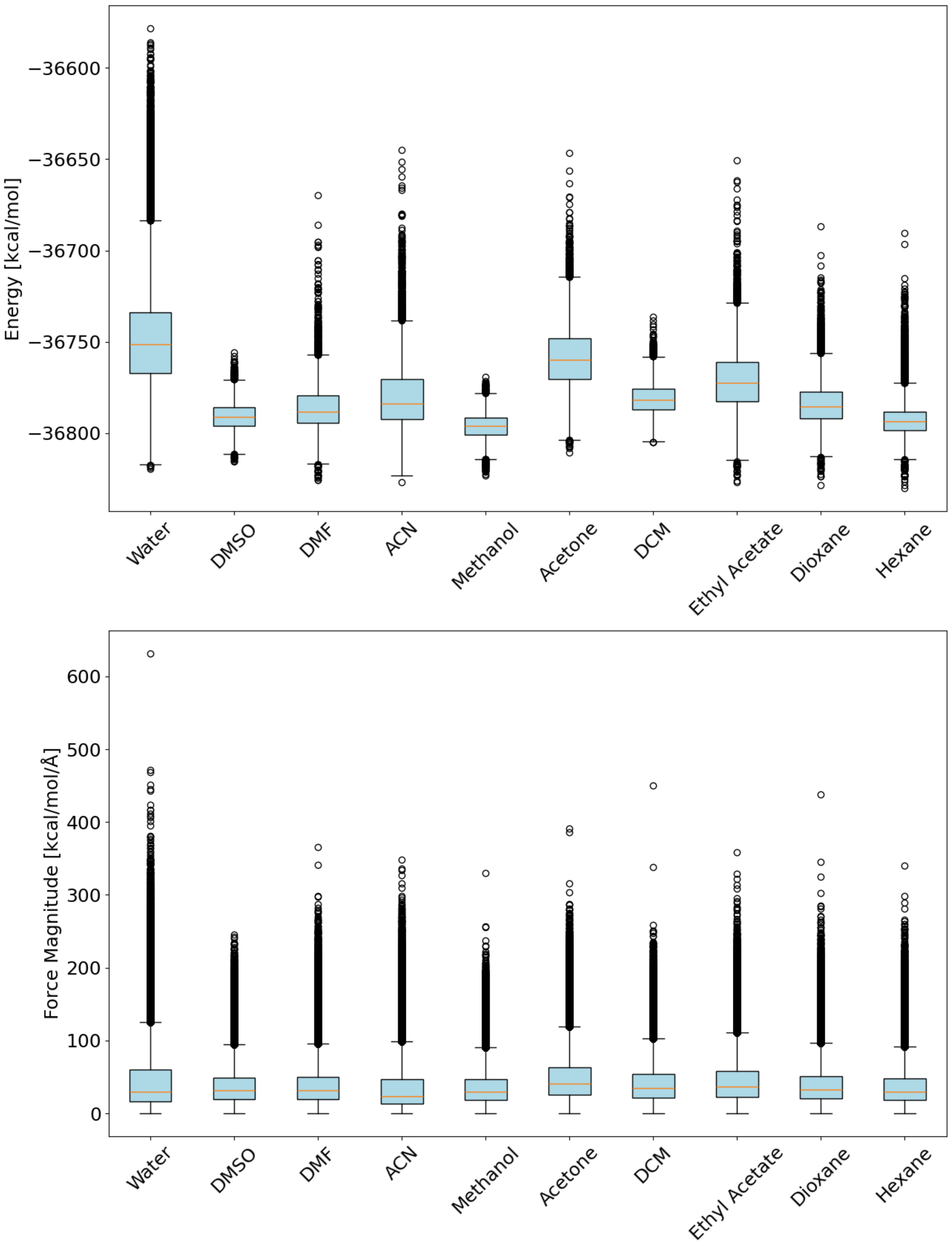


Figure 51.: State energies (top) and force magnitudes (bottom) for *FR0* in *all ten solvents*. These summaries complement the structural diagnostics by indicating the prevalence of strained outliers.

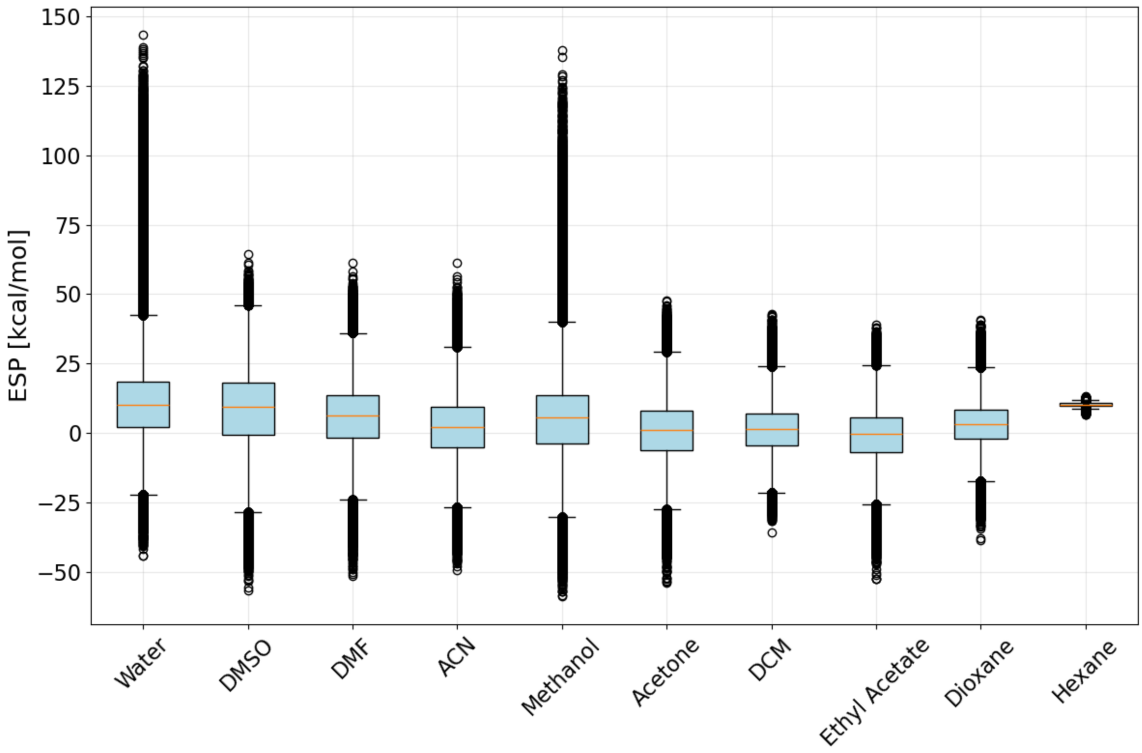


Figure 52.: Electrostatic potential (ESP) of the MM atoms evaluated at QM atomic positions for *all ten solvents*. Polar solvents yield broader, heavy-tailed ESPs; non-polar Hexane is narrow and nearly constant.

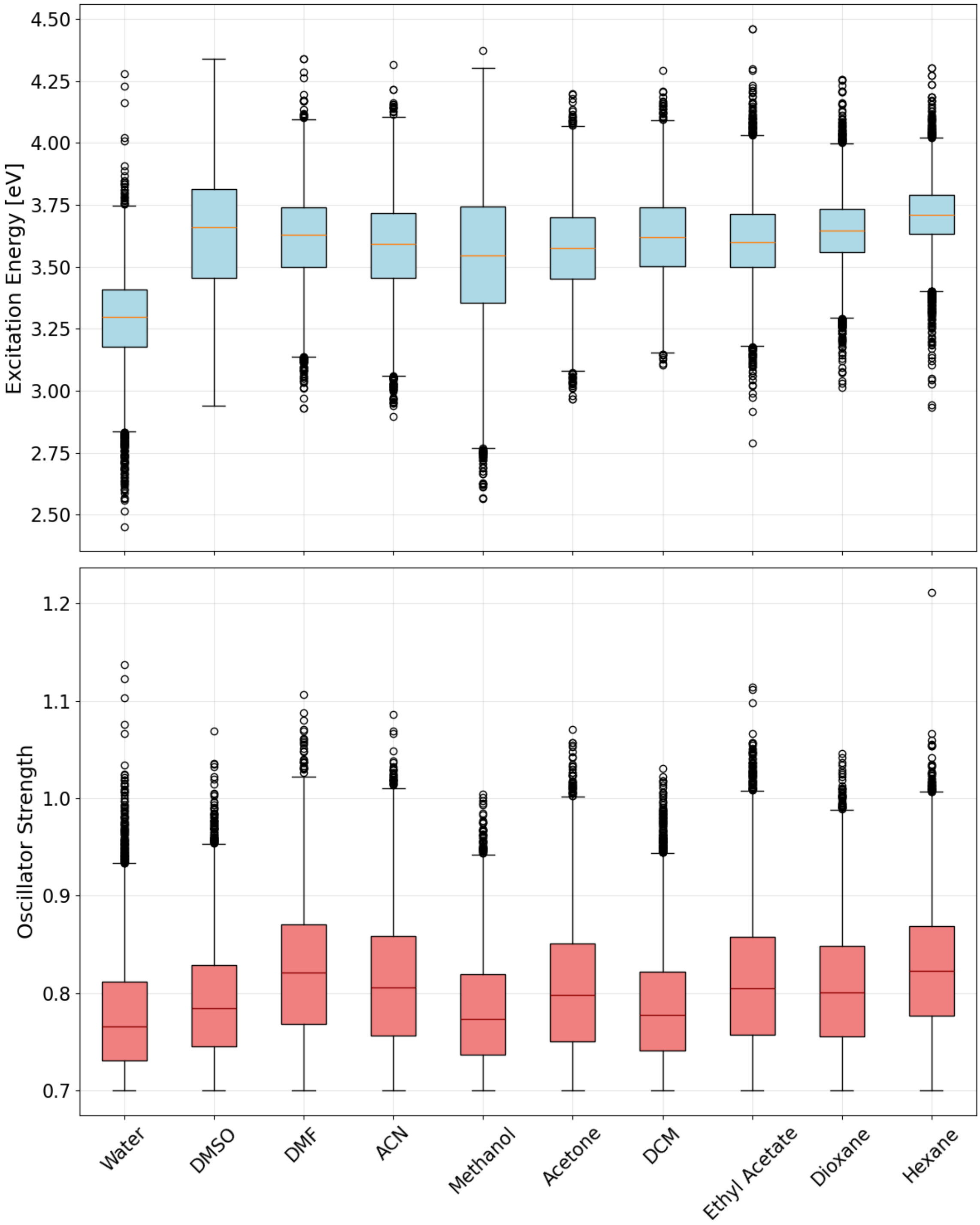


Figure 53.: Vertical excitation energies E_{vert} (eV) for the bright state (top); corresponding oscillator strengths f (bottom). Only frames passing the brightness criterion $\max_{k\in\{1,2,3\}} f_k \geq 0.7$ are included.

6.3.3. State Identity and Charge Choice

For *FR0* the identity of the bright state is strongly solvent-dependent. Gas-phase control trajectories show virtually no S_1/S_2 exchanges (Table 22), whereas explicit solvent frequently induces swaps. The per-atom ESP distributions in Fig. 55 differ markedly between all solvents and four selected sites, with polar media showing broad, shifted ranges and hexane exhibiting a very narrow band. Taken together, the pronounced, site-resolved ESP sensitivity indicates that fluorescence spectra can be predicted as a function of the local ESP field sampled at the QM atoms, making per-atom ESP a natural control variable for spectrum construction.

Bright-state populations and switching. Table 22 quantifies the bright-state assignment after recomputing the electronic states for each geometry from the two DFTB trajectories (**DFTB S_1** and **DFTB S_2** denote the *source trajectory*). Across solvents the recomputed bright state is most often S_2. In several cases – e.g. ethyl acetate and 1,4-dioxane – geometries taken from the DFTB S_1 trajectory are assigned S_2 in $\gtrsim$ 99% of frames ($S_1/S_2/S_3$ = 1.0/99.0/0.0 and 0.3/99.6/0.0, respectively), while vacuum shows essentially no switching (dashed row), underscoring that swapping is solvent-induced.

Table 22.: Sampling state occupancies and switching statistics for FR0. The blocks **DFTB S_1** and **DFTB S_2** indicate the *source trajectory* from which geometries were taken. For each geometry the bright state is re-evaluated; the columns $S_1/S_2/S_3$ report the fraction of frames whose (recomputed) bright state has character of State 1/2/3 (in %). "Switches" is the fraction of frames with a change of that assignment relative to the previous frame.

	DFTB S_1				**DFTB S_2**			
Solvent	S_1 [%]	S_2 [%]	S_3 [%]	Switches [%]	S_1 [%]	S_2 [%]	S_3 [%]	Switches [%]
Water	42.3	55.5	2.3	50.8	38.7	60.9	0.3	26.8
DMSO	0.0	100.0	0.0	0.0	41.8	58.2	0.0	42.1
DMF	9.1	90.8	0.0	14.6	1.3	98.7	0.0	2.4
Acetonitrile	13.1	86.9	0.0	23.1	4.0	95.9	0.1	6.9
Methanol	40.5	59.5	0.0	0.1	52.0	48.0	0.0	43.6
Acetone	7.7	92.3	0.0	12.9	4.7	95.3	0.0	8.2
DCM	0.0	100.0	0.0	0.0	16.9	83.1	0.0	26.3
Ethyl acetate	1.0	99.0	0.0	1.8	4.9	94.9	0.2	8.8
Dioxane	0.3	99.6	0.0	0.7	1.7	98.2	0.0	3.4
Hexane	0.0	100.0	0.0	0.0	1.0	99.0	0.0	1.9
Vacuum	0.0	100.0	0.0	0.0	0.6	99.4	0.0	1.3

ESP sensitivity at four representative sites. To demonstrate that the MM electrostatic environment varies both *by solvent* and *within the molecule*, Fig. 55 reports ESP distributions for *four illustrative atoms*: the formyl oxygen O(0), a benzylic ring carbon C(1), the bridgehead carbon C(12), and the donor nitrogen N(30). The selection spans acceptor, aromatic core, linker, and donor and is meant as a *minimal diagnostic*, not an exhaustive per-atom survey. For a given site, different solvents produce clearly distinct ESP histograms; within one solvent, the four sites also separate. Such solvent- and site-specific ESP shifts are consistent with Stark-type modulation of transition energies and oscillator strengths, and motivate using per-atom ESP as an input feature for predicting fluorescence spectra.

Fixed charges for NN/MM. State-resolved partial charges from ωB97X/def2-TZVP (Fig. 54) differ mainly at the donor/acceptor termini and are otherwise similar. Given that the recomputed bright state in solution is predominantly S_2 (Table 22), the NN/MM simulations employ a *single, fixed* set of S_2 charges. This aligns the embedding with the most likely excited-state character while avoiding state-dependent charge switching during dynamics.

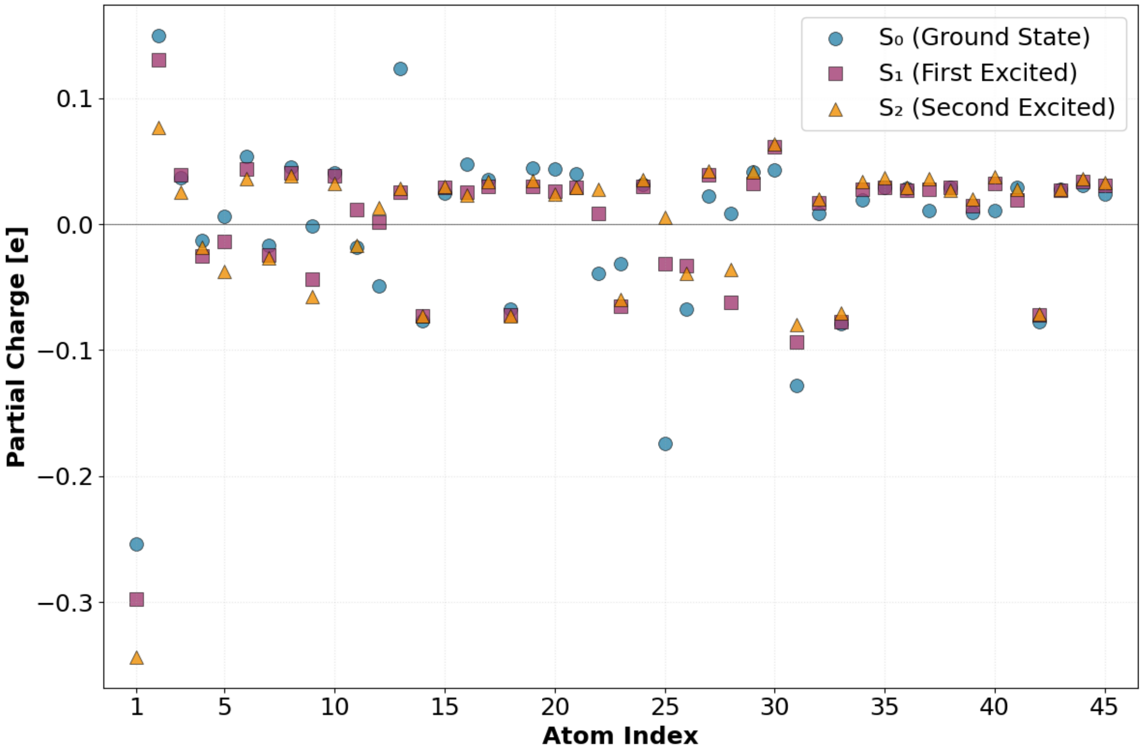
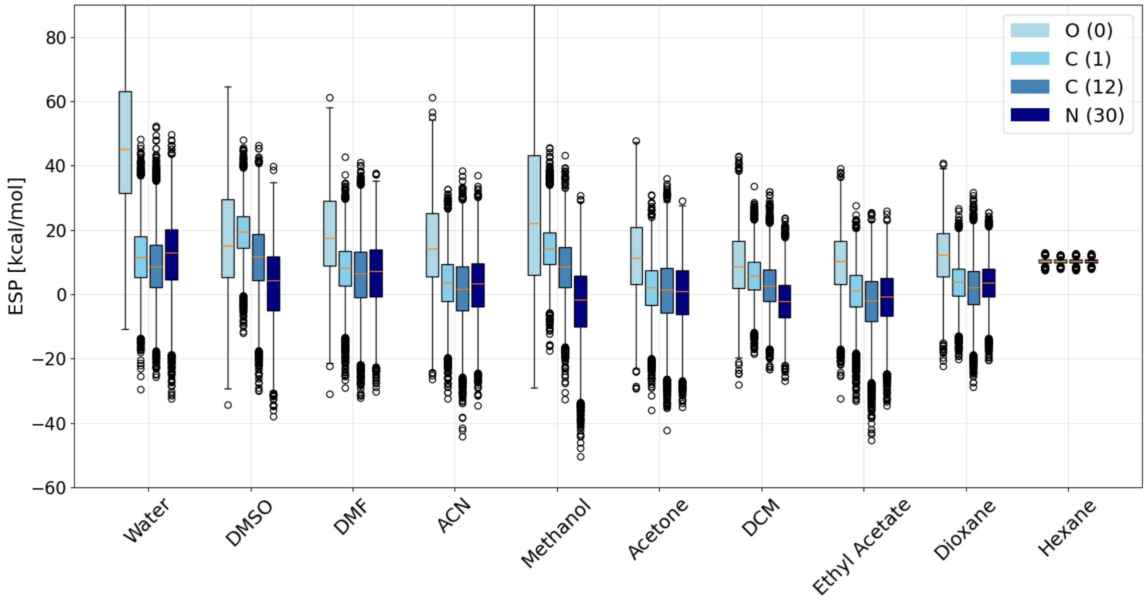
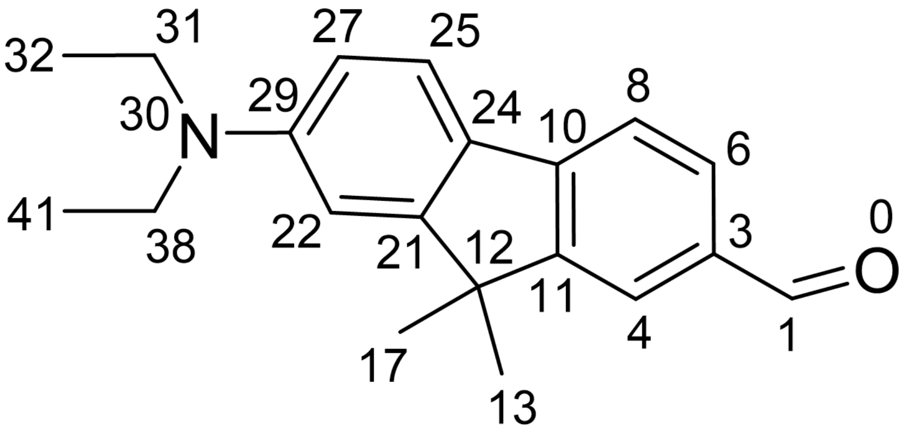


Figure 54.: FR0 atomic partial charges (vacuum) for S_0, S_1, and S_2 from ωB97X/def2-TZVP. Differences are concentrated near the donor/acceptor termini; elsewhere the patterns are similar. Given the predominance of S_2 in Table 22, S_2 charges are adopted for NN/MM.



(a) ESP at O(0), C(1), C(12), and N(30) across solvents.



(b) FR0 with atom indices used in this section.

Figure 55.: MM electrostatic potential (ESP): solvent-dependent at four FR0 sites. Polar solvents broaden and shift the distributions in an atom-specific manner, whereas hexane shows a narrow band. The reference sketch (bottom) defines the atom indices.

6.4. Force Model

This section trains and evaluates neural force fields that replace the QM gradient in explicit solvent (NN/MM). We first focus on our *inverse-distance + ESP* model (Methods II, Sec. 3.6.1), then benchmark against a *MACE-EQ* baseline, and finally demonstrate NN/MM applications.

6.4.1. Training

The inverse-distance + ESP force model introduced in Methods II (Sec. 3.6.1) is trained using as inputs (i) inverse interatomic distances within the QM region and (ii) the MM electrostatic potential sampled at the QM atomic positions. This allows solvent reorganisation to modulate the energy landscape and state ordering while keeping the descriptor compact.

Hyperparameter search. Model capacities and regularisation are selected by a Hyperband-style search with early stopping. The search space and the configuration chosen for training on *all ten solvents* are summarised in Tab. 23; per-solvent training outcomes and the corresponding selected hyperparameters are listed in Tab. 24.

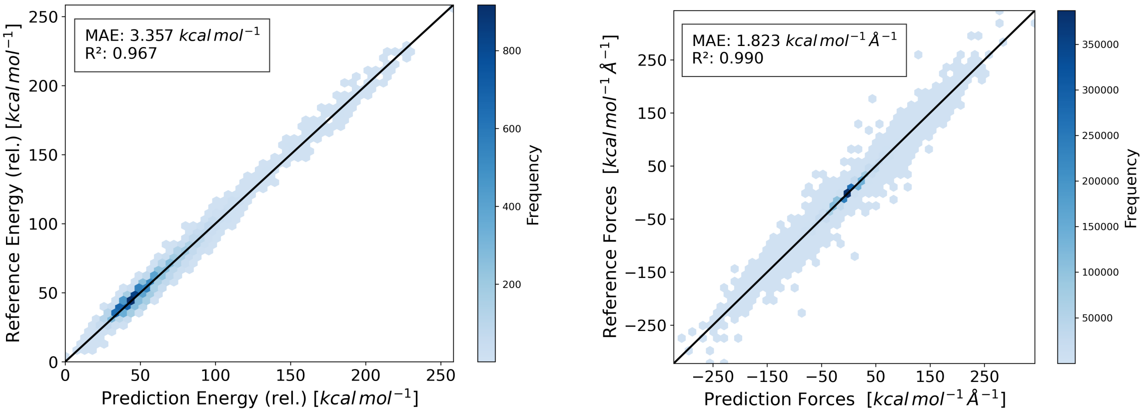
Table 23.: Hyperparameter search space and selected configuration for the model trained on *all 10 solvents*.

Hyperparameter	Search space	Best (all)
Hidden neurons / layer	200–1000 (stepsize: 50)	850
Hidden layers	2–4	2
Initial learning rate	$\{10^{-3}, 5{\cdot}10^{-4}, 10^{-4}\}$	$5{\cdot}10^{-4}$
L2 penalty	$\{10^{-3}, 5{\cdot}10^{-4}, 10^{-4}, 5{\cdot}10^{-5}\}$	10^{-4}
Energy/force loss ratio λ	$\{10^{-2}, 5{\cdot}10^{-3}, 10^{-3}, 5{\cdot}10^{-4}\}$	$5{\cdot}10^{-3}$
Batch size	128	128
Fit epochs	2000	2000
HP budget / trial	300 epochs	300
Hyperband factor	18	18

Parity on the pooled dataset. Figure 56 compares predictions to references for the model trained on the pooled set of *all ten* solvents. The forces – used as the primary learning target for NN/MM – lie tightly on the diagonal (MAE $\approx$ 1.8 kcal mol^{-1} Å^{-1}, $R^2 \approx 0.99$).

Robustness to sparse outliers. To demonstrate that near-diagonal force parity is not driven by a few rare bins, hexagons with occupancy below a threshold τ are masked (Fig.57). Increasing τ from 10 to 100 removes only very sparse high-force wings; the dense core remains unchanged and the summary metrics stay essentially constant (MAE $\approx 1.8\,\mathrm{kcal\,mol^{-1}\,Å^{-1}}$, $R^2 \approx 0.99$).

Energy diagnostics. Forces remain the primary learning target and show reasonable agreement throughout the study (cf. Fig. 56 and Table 25). Energies are reported only as a diagnostic. Further discussion of energy diagnostics in cross-evaluation section (Sec. 6.4.2).



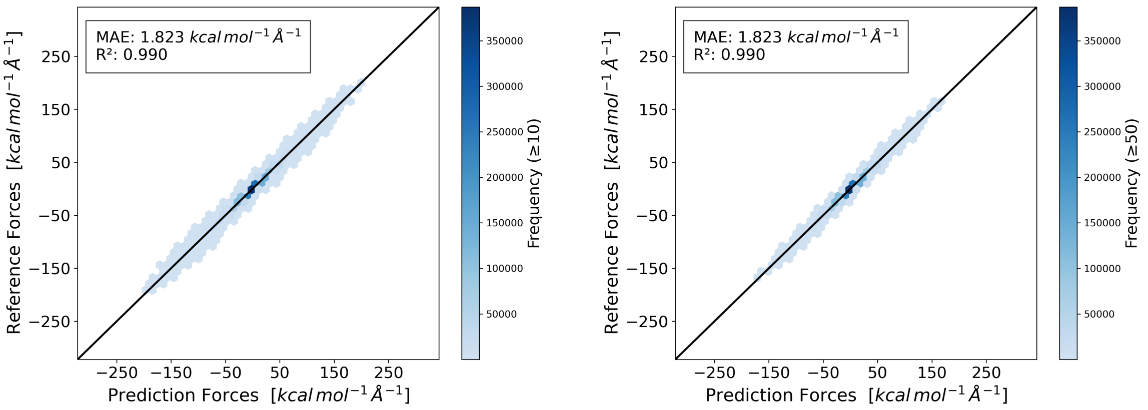
(a) Energies

(b) Forces

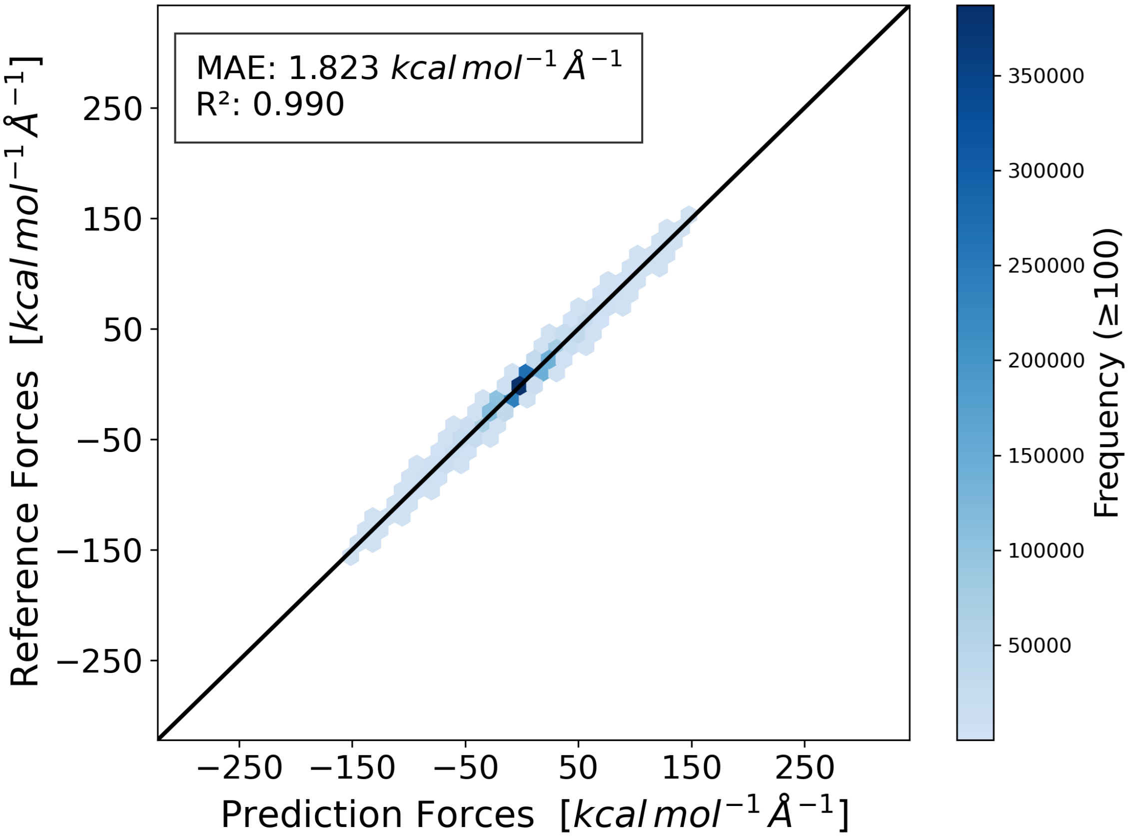
Figure 56.: FR0 ML potential: Prediction-reference parity across *all ten solvents*. Left: total electronic energies in (*reference diagnostic*). Right: per-component forces (*learning target*).

Table 24.: Force model training performance: trained per *single* solvent and for the pooled set of *all ten solvents*. MAE_F in $\mathrm{kcal\,mol^{-1}\,Å^{-1}}$; MAE_E in $\mathrm{kcal\,mol^{-1}}$. Hyperparameter details are provided in Appx. Table A.23.

Training combination	R^2_F	MAE_F	R^2_E	MAE_E
All 10 solvents	0.990	1.82	0.967	3.36
Water	0.967	4.00	0.978	4.19
DMSO	0.978	2.81	0.827	2.59
DMF	0.982	2.48	0.956	2.01
Acetonitrile	0.975	2.76	0.963	2.88
Methanol	0.980	2.54	0.879	1.94
Acetone	0.977	3.49	0.949	2.89
DCM	0.979	2.91	0.878	2.40
Ethyl acetate	0.976	3.30	0.931	3.34
1,4-Dioxane	0.980	2.68	0.705	4.08
n-Hexane	0.981	2.46	0.920	2.21



(a) $\tau = 10$ **(b)** $\tau = 50$



(c) $\tau = 100$

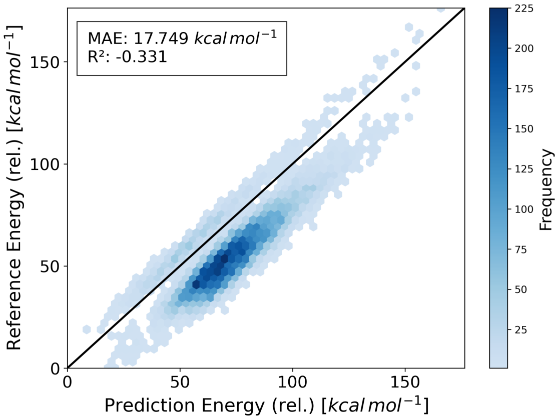
Figure 57.: Force model trained on *all ten solvents*: parity with density mask (bins with count $< \tau$ hidden).

6.4.2. Cross-Validation

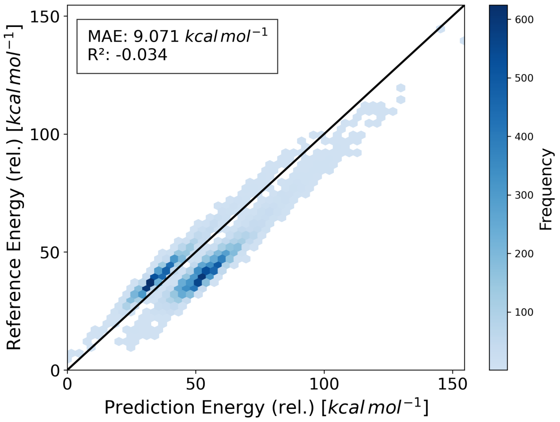
Forces remain the primary learning target and show reasonable agreement throughout the study (cf. Fig. 56 and Table 25). Energies are reported only as a diagnostic. Two solvents stand out in the leave-one-solvent-out (LOSO) evaluation: for **n-hexane**, the held-out energy metrics are poor with $R_E^2 < 0$ and $\mathrm{MAE}_E = \mathbf{9.07}\,\mathrm{kcal\,mol^{-1}}$ (Table 25); for **ethyl acetate**, LOSO yields $R_E^2 < 0$ and $\mathrm{MAE}_E = \mathbf{17.75}\,\mathrm{kcal\,mol^{-1}}$. Including the respective solvent in training improves ethyl acetate substantially but leaves hexane problematic: with *all ten* solvents in training, evaluation on *hexane* gives $R_E^2 = 0.557$ and $\mathrm{MAE}_E = \mathbf{6.07}\,\mathrm{kcal\,mol^{-1}}$, whereas evaluation on *ethyl acetate* reaches $R_E^2 = 0.853$ and $\mathrm{MAE}_E = \mathbf{5.61}\,\mathrm{kcal\,mol^{-1}}$. By contrast, *single-solvent* models trained only on the target solvent perform well for both; *ethyl acetate:* $R_E^2 = 0.931$, $\mathrm{MAE}_E = \mathbf{3.35}\,\mathrm{kcal\,mol^{-1}}$; see Table 24). The parity maps in Fig. 58 show that the pathology manifests as two nearly parallel ridges (constant offsets across subsets) for ethyl acetate and hexane, while DMF displays a single tight diagonal. A plausible contributor for hexane is the extremely narrow ESP range in the training data (Fig. 52), which weakens the energy signal when hexane is mixed with high-polarity solvents; importantly, the force quality remains high in all cases.

Table 25.: FR0 Force Model: Leave-one-solvent-out cross-validation; the network is trained on 9/10 solvents and evaluated on the held-out solvent. Reported are R^2 and MAE_F is in $\mathrm{kcal\,mol^{-1}\,Å^{-1}}$; MAE_E is in $\mathrm{kcal\,mol^{-1}}$.

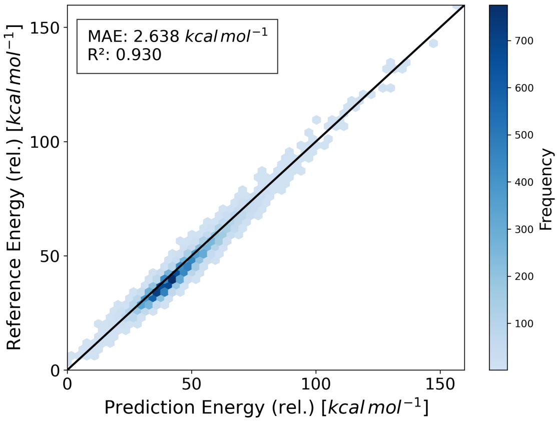
Held-out solvent	R_F^2	MAE_F	R_E^2	MAE_E
Water	0.977	2.99	0.973	5.33
DMSO	0.992	1.67	0.707	3.50
DMF	0.994	1.46	0.930	2.64
Acetonitrile	0.987	1.96	0.829	6.07
Methanol	0.993	1.44	0.851	2.18
Acetone	0.991	2.11	0.945	3.54
Dichloromethane	0.992	1.70	0.762	3.56
Ethyl acetate	0.988	2.29	−0.331	17.75
Dioxane	0.991	1.75	0.897	3.01
Hexane	0.993	1.45	−0.034	9.07



(a) LOSO: Ethyl acetate



(b) LOSO: n-Hexane

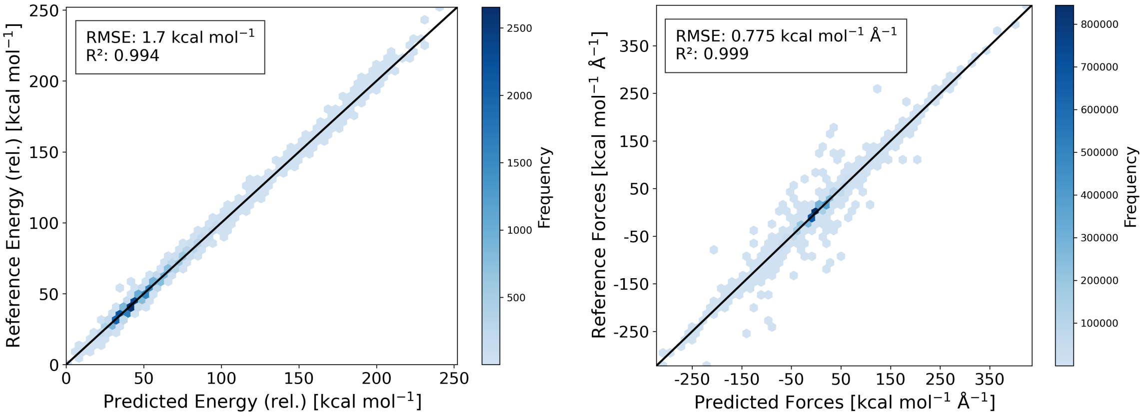


(c) LOSO: DMF

Figure 58.: Force model state energy outliers: Leave-one-solvent-out cross-validation – prediction-reference parity (kcal mol^{-1}; hexbin density). Ethyl acetate **(a)** and n-Hexane **(b)** exhibit two nearly parallel ridges offset from the y=x line, consistent with a constant energy shift in subsets of the data (e.g., per-batch zeroing or state reindexing), whereas DMF **(c)** shows a single tight ridge. These panels serve as a *diagnostic only*: throughout the chapter the **forces** are the learning target used in NN/MM (energies are shown for sanity-checks).

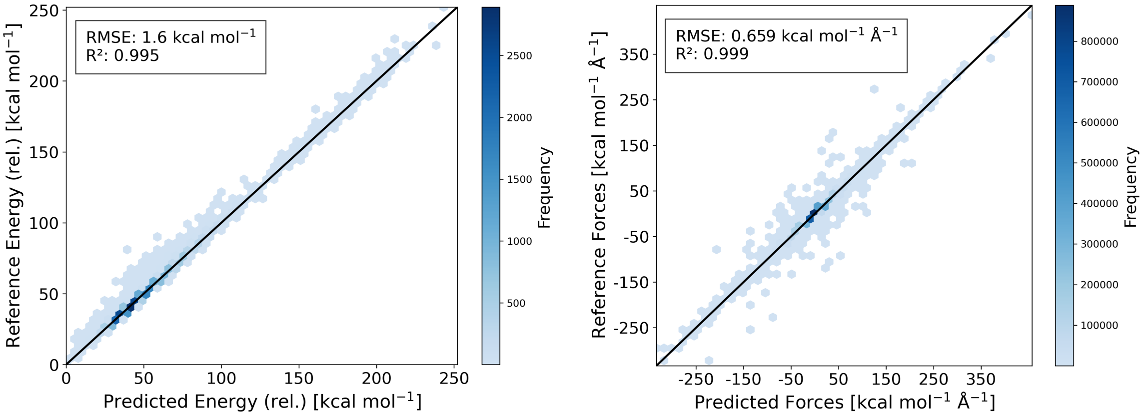
6.4.3. Benchmark: MACE–EQ

This subsection benchmarks MACE–EQ (Methods II, Sec. 3.4.3) against the inverse-distance + ESP baseline used throughout this chapter (Methods II, Sec. 3.6.1). MACE-EQ augments MACE with charge-equilibration (QEq) and electrostatic–potential (ESP) terms and is trained to predict bright-state energies and forces for the emissive surface employed in the NN/MM workflows. Two variants differ only in the loss weighting between energy and forces: an *energy-focused* run and a *force-focused* run. In both cases the weighting is complemented by a swap during *stochastic weight averaging (SWA)*: the main phase emphasizes one target (energy or forces), while the SWA phase inverts the emphasis to balance the final model (Table A.22). Trainings results are shown in Fig. 59.



(a) Energy–focused: energies

(b) Energy–focused: forces



(c) Force–focused: energies

(d) Force–focused: forces

Figure 59.: MACE-EQ force model: trained on all solvents. Panels (a,b) are energy-focused; (c,d) are force-focused. In-panel text reports RMSE and R^2.

Comparison to the baseline force model. Relative to Fig. 56b (inverse-distance + ESP model), the MACE-EQ parity plots span a broader force range ($F > 350$ kcal mol^{-1} Å^{-1} are well covered) while delivering *markedly lower* force errors: the MACE–EQ force RMSE lies in the $0.66–0.78$ kcal mol^{-1} Å^{-1} band, i.e. approximately half the baseline's level. For context, the baseline used a **10%** test split whereas the MACE-EQ runs used **20%**; the larger held–out portion increases the chance that very rare high-force outliers appear in the test set and thus extends the visible range beyond ~ 350. These tail points are sparse and do not alter the aggregate conclusions – $R^2 \geq 0.999$ together with stable NN/MM performance supports MACE–EQ as a high–fidelity reference or drop–in replacement in the fluorescence workflow.

6.4.4. Application in NN/MM

We deploy the trained inverse-distance + ESP force model in live NN/MM simulations across ten solvents and assess its performance under production conditions. In this section we proceed in stages: first we check *stability*; later we examine how well trajectories adhere to the intended bright-state surface and how force/energy predictions compare to independent reference windows.

Stability. Here, "stable" means a 1 ns NN/MM run completes without a crash for the chosen model and time step. Tables 27 a and b summarize this as a binary pass/fail at **1 fs** and **0.5 fs**; Table 27b complements this with the *survival percentage* at **1 fs**. At **1 fs**, joint training on *All 10* solvents runs stably in **10/10** cases (Table 27 a). Halving the step to **0.5 fs** boosts robustness: with *First 6* solvents the model is already stable in **7/10** systems (Table 27 b). Per-solvent (*Same*) models benefit most from 0.5 fs (from **1/10** stable at 1 fs to **5/10** at 0.5 fs), underscoring the advantage of both a smaller time step and broader training data. For unstable 1 fs cases, Table 27b shows how far runs progressed before the first crash.

Computational cost. A single explicit-solvent TD-LC-DFTB/MM trajectory of 1 ns (1 fs time-steps) with excited-state evaluation (Casida) required **90–120 days** wall time – depending on the solvent – on the tested setup. Replacing the QM gradient with the trained NN/MM force model reduces this to **hours** while preserving bright-state identity and force fidelity. Table 26 summarizes wall times and the implied speedup factors.

Table 26.: Wall times for 1 ns trajectories per solvent (0.5 fs time-steps). TD-LC-DFTB/MM refers to explicit-solvent runs with excited-state evaluation; NN/MM uses the learned force model. Speedup computed as $(90\text{–}120\ \text{days})/(t_{\text{NN/MM}} \cdot 0.5\,\text{fs})$.

Solvent	**NN/MM** [h]	**Speedup** [×]
Water	9	480–640
DMSO	8	540–720
DMF	9	480–640
ACN	7	618–822
Methanol	8	540–720
Acetone	8	540–720
DCM	5	864–1152
Ethyl acetate	8	540–720
Dioxane	10	432–576
Hexane	9	480–640

Table 27.: NN/MM stability of the force network for FR0 across solvents: (a) **1 fs** vs. (b) **0.5 fs**. A mark “x” indicates a 1 ns simulation finished without crash; “–” indicates early failure. NN/MM trajectory survival percentage of the 1 ns target duration completed before the first crash, of a model trained *on all ten solvents* at a **0.5 fs**. Columns: *Same* = trained and evaluated on the *same* solvent; *First 6* = joint training on H_2O, DMSO, DMF, ACN, MeOH, Acetone; *All 10* = joint training on all ten solvents (only used at 1 fs). Unless noted otherwise, the force loss is Eq. 3.66 with energy weight $\lambda_E = 10^{-3}$.

(a) 1 fs

Solvent	Same	First 6	All 10
Water	–	–	–
DMSO	–	x	x
DMF	x	x	x
ACN	–	x	x
Methanol	–	x	x
Acetone	–	–	x
DCM	–	–	x
Ethyl acetate	–	x	x
Dioxane	–	–	x
Hexane	–	x	x

(b) 0.5 fs

Solvent	Same	First 6	All 10
Water	–	–	x
DMSO	–	–	x
DMF	x	x	x
ACN	–	x	x
Methanol	–	x	x
Acetone	–	–	x
DCM	x	x	x
Ethyl acetate	x	x	x
Dioxane	x	x	x
Hexane	x	x	x

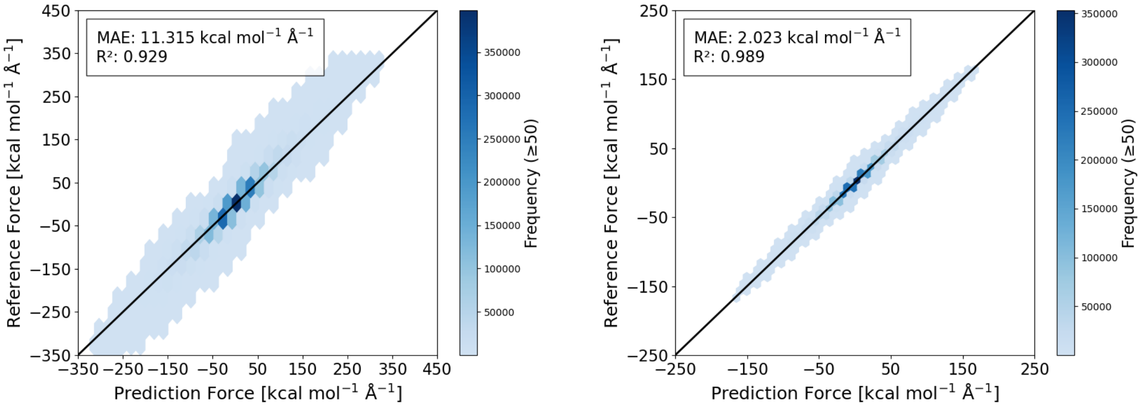
Table 28.: NN/MM trajectories: bright-state statistics of *FR0* with a force-model in place of the QM gradient. For every NN/MM geometry, the bright state ($S_1/S_2/S_3$) is assigned *a posteriori* by a TD-LC-DFTB calculation. **Crossings** is the percentage of frames in which the assigned bright state differs from the previous frame. Trajectories here were propagated using fixed S_2 charges.

Solvent	Crossings [%]	Time in S_1 [%]	Time in S_2 [%]	Time in S_3 [%]
Water	17.4	90.2	8.7	1.0
DMSO	20.1	87.4	12.6	0.0
DMF	23.2	85.4	14.6	0.0
ACN	24.2	85.6	14.4	0.0
Methanol	20.4	87.4	12.6	0.0
Acetone	32.3	79.4	20.6	0.0
DCM	36.0	66.0	34.0	0.0
Ethyl acetate	43.5	69.8	30.2	0.0
Dioxane	43.0	63.3	36.7	0.0
Hexane	3.9	2.0	98.0	0.0

Live-force cross-validation against TD-LC-DFTB. During 1 ns NN/MM production we logged the *forces actually used* by the integrator. For each saved geometry we then recomputed TD-LC-DFTB and extracted the forces of the *brightest* state (highest oscillator strength) as an approximate reference. Because frequent state crossings can misassign the "true" emissive surface, this procedure introduces label noise (especially in highly solvating media); hence the reported MAE/R^2 should be read as an estimator of force fidelity. Despite this caveat, parity maps remain relatively tightly clustered around y=x for all solvents.

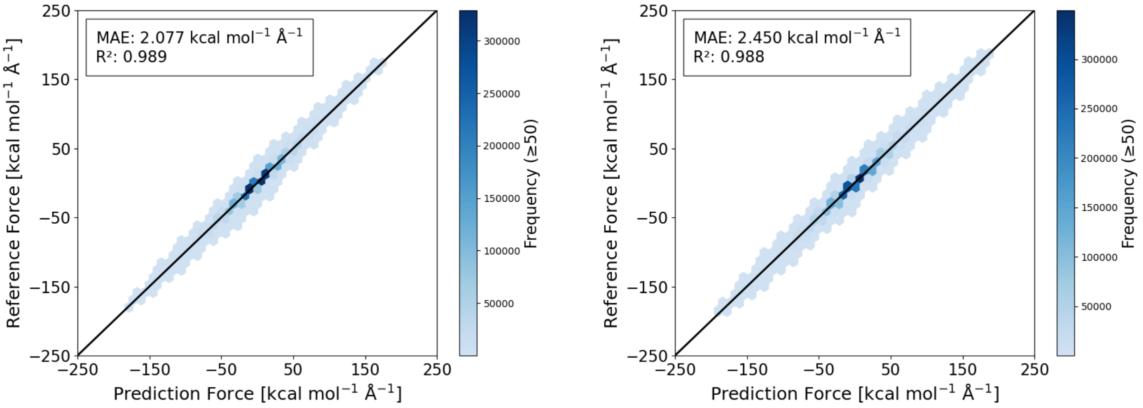
Table 29.: NN/MM forces recalculated with TD-LC-DFTB. Summary of R^2 and MAE (in kcal mol^{-1} Å^{-1}) from the parity plots above. Reference forces are taken from the TD-LC-DFTB state with the highest oscillator strength at each snapshot.

Solvent	**R^2**	**MAE**
Water	0.929	11.32
DMSO	0.989	2.02
DMF	0.989	2.08
Acetonitrile	0.988	2.45
Methanol	0.988	2.65
Acetone	0.987	2.64
Dichloromethane	0.986	1.64
Ethyl acetate	0.986	2.71
Dioxane	0.987	1.80
Hexane	0.990	1.40



(a) Water. **(b)** DMSO.

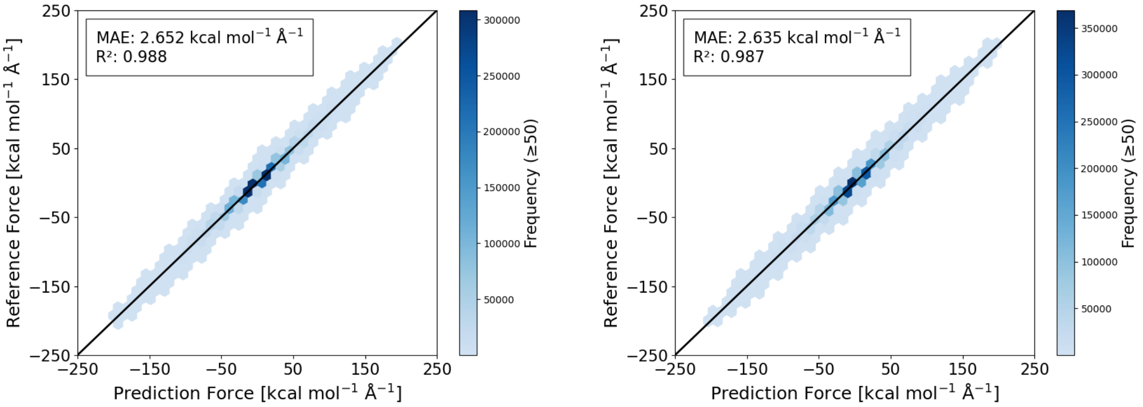
Figure 60.: Parity maps of **Water** and **DMSO**.



(a) DMF.

(b) Acetonitrile.

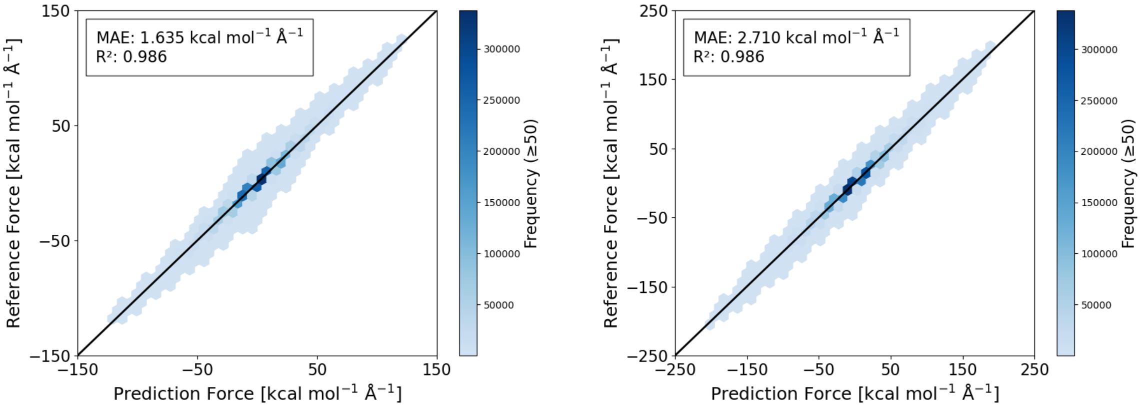
Figure 61.: Parity maps for **DMF** and **ACN**.



(a) Methanol.

(b) Acetone.

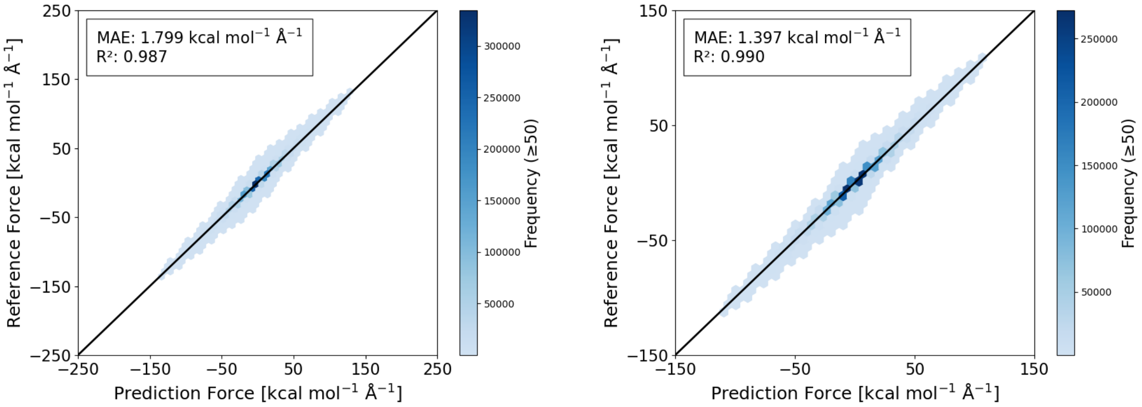
Figure 62.: Parity maps for **MeOH** and **Acetone**.



(a) Dichloromethane.

(b) Ethyl acetate.

Figure 63.: Parity maps for **DCM** and **Ethyl acetate**.



(a) 1,4-Dioxane.

(b) *n*-Hexane.

Figure 64.: Parity maps for **Dioxane** and **Hexane**.

6.5. Property Model

Emission spectra are constructed from NN/MM snapshots using a nuclear-ensemble (ensemble-based) estimator with Gaussian homogeneous broadening σ, weighting each configuration by its predicted oscillator strength. Peak positions and bandwidths across solvents are compared with experiment where available, and the dependence on σ is analyzed. Systematic offsets are interpreted in light of solvent ε_r diagnostics from pure-solvent simulations (SI), which modulate the ESP distributions entering the model inputs.

The *property model* (Sec. 3.6.3) maps structures to the bright-state vertical emission energy ΔE_{em} and oscillator strength f for spectrum construction.

6.5.1. Training

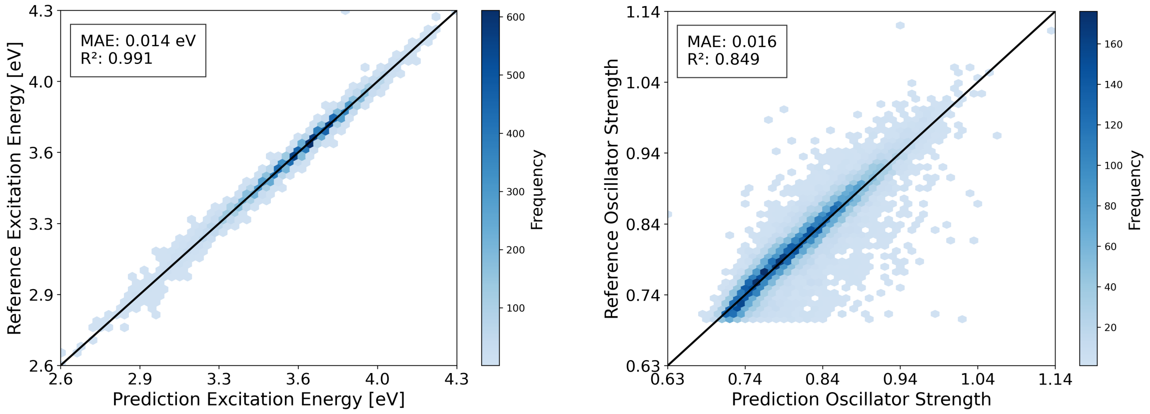
The property network (Sec. 3.6.3) uses as inputs the QM geometry (inverse interatomic distances within the QM region) together with the MM electrostatic potential (ESP) evaluated at the QM atoms, and predicts the bright-state vertical emission energy ΔE_{em} and the oscillator strength f. The objective is the unweighted sum of mean-squared errors across the two output heads. Hyperparameters are selected via a Hyperband search (Table 30).

The **primary model** is trained on *all ten solvents* (FR0 in H_2O, DMSO, DMF, ACN, MeOH, acetone, DCM, ethyl acetate, dioxane, hexane) and is used for all spectral predictions in Sec. 6.5.2. Its test performance (Fig. 65, Table 31) is R^2=0.991 (MAE = 0.01 eV) for energies and R^2=0.849 (MAE = 0.02) for oscillator strengths.

For **diagnostics**, single-solvent baselines (one model per solvent) were trained with the same protocol. Contrary to the intuitive expectation that a per-solvent model should perform best on its own domain, the pooled-all model generally *matches or outperforms* single-solvent models in energy accuracy (Table 31). A likely reason is that pooling exposes the network to a broader range of ESP/geometry patterns, acting as data-driven regularization and improving generalization. The f head is intrinsically noisier than the energy head but remains adequate for spectral weighting; spectrum shapes are dominated by ΔE_{em} together with the chosen homogeneous broadening.

All spectra reported in this chapter therefore use the *pooled-all* property network; single-solvent results are retained only to document comparative performance.

Fixed (non-tuned) settings: 90/10 train-validation split; mean-squared-error loss applied to both heads; leaky-softplus hidden activation (α=0.03) and linear output layers; early stopping with patience 250 within a 2000-epoch training budget.



(a) Excitation energy

(b) Oscillator strength

Figure 65.: Prediction-reference parity of the *property network* trained on *all ten solvents.* (a) Vertical excitation energies (eV): MAE = 0.014 eV, $R^2 = 0.991$. (b) Oscillator strengths (dimensionless): MAE = 0.016, $R^2 = 0.849$. Points are shown as hexbin densities; the black diagonal indicates $y{=}x$.

Table 30.: Property network: Hyperparameter search space and the settings selected for the best model trained on *all ten solvents.*

Hyperparameter	**Search space**	**Selected (All 10)**
Hidden neurons / layer (per head)	20–100 (step 5)	Energy 40; Oscillator 25
Hidden layers (per head)	2–8 (step 1)	Energy 6; Oscillator 5
Learning rate	$\{10^{-3}, 5{\cdot}10^{-4}, 10^{-4}\}$	$5{\cdot}10^{-4}$
Regulariser	L2	L2
Hyperband budget / trial (epochs)	20	20
Hyperband factor	2	2

Table 31.: Test performance of the *property network* predicting bright-state vertical emission energy and oscillator strength. Rows list single-solvent training/testing and the pooled *All 10* model. Energy MAE in eV; oscillator-strength MAE is dimensionless. Detailed model settings are given in App. Table A.24.

Solvent	R^2_{Energy}	$\text{MAE}_{\text{Energy}}$ [eV]	R^2_f	MAE_f
Water	0.855	0.06	0.529	0.04
DMSO	0.988	0.02	0.826	0.02
DMF	0.980	0.02	0.831	0.02
ACN	0.974	0.02	0.751	0.02
Methanol	0.988	0.02	0.739	0.02
Acetone	0.967	0.03	0.782	0.02
DCM	0.983	0.02	0.736	0.02
Ethyl acetate	0.964	0.02	0.756	0.02
Dioxane	0.969	0.02	0.821	0.02
Hexane	0.966	0.02	0.904	0.01
All 10	0.991	0.01	0.849	0.02

6.5.2. Spectral Predictions

Fluorescence spectra are constructed from NN/MM data using four estimators: (i) weighted histograms of vertical emission energies (*Hist*), (ii) the Nuclear Ensemble Approach (*NEA*), (iii) the vibronically augmented *E–ZTFC*, and (iv) the dynamics–based *EGCF* (energy–gap correlation with second–order cumulant; Methods I, Sec. 2.3.5.3). All spectra use oscillator–strength weighting with the ω^3 factor for emission and are area–normalized. Spectra are assembled on the energy axis and, where shown, mapped to wavelength via $I_\lambda(\lambda) = I_E(E(\lambda))\, hc/\lambda^2$. For the *ensemble* estimators (Hist/NEA/E–ZTFC) we used a 1 ns trajectory propagated at 0.5 fs but *stored every 100th frame* ($\Delta t_{\text{out}} = 50\,\text{fs}$). In contrast, EGCF requires a uniformly sampled, high-resolution time series; we therefore re–ran NN/MM for 1 ns at 0.5 fs and *stored every frame*.

Table 32.: Spectrum energy-domain peak positions and widths across methods and solvents (Hist, NEA, E–ZTFC, EGCF, DFTB). $\boldsymbol{\mu}$: peak energy (at maximum). $\boldsymbol{\sigma}$: Gaussian standard deviation from the histogram fit. $\boldsymbol{\Gamma}$: full width at half maximum (FWHM) measured from the NEA, E–ZTFC, and EGCF profiles (not assumed Gaussian). All quantities are in eV. *Right-most column: simulated relative permittivity* $\varepsilon_r^{\text{Sim}}$.

	Hist		**NEA**		**E–ZTFC**		**EGCF**		**DFTB**		
Solvent	μ	σ	μ	Γ	μ	Γ	μ	Γ	μ	σ	$\varepsilon_r^{\text{Sim}}$
Water	2.93	0.31	3.00	0.73	3.11	0.77	2.74	0.49	2.92	0.33	97.1
DMSO	3.38	0.14	3.37	0.34	3.49	0.45	3.28	0.08	3.38	0.14	44.3
DMF	3.39	0.15	3.41	0.34	3.50	0.46	3.29	0.08	3.39	0.15	22.7
ACN	3.36	0.16	3.37	0.36	3.45	0.48	3.23	0.09	3.36	0.15	23.6
Methanol	3.33	0.17	3.35	0.41	3.45	0.50	3.20	0.08	3.34	0.16	20.0
Acetone	3.37	0.17	3.39	0.40	3.47	0.49	3.23	0.09	3.37	0.16	10.4
DCM	3.50	0.11	3.51	0.26	3.58	0.40	3.44	0.07	3.51	0.12	3.9
Ethyl ac.	3.41	0.17	3.43	0.40	3.54	0.49	3.26	0.07	3.41	0.17	4.5
Dioxane	3.51	0.12	3.51	0.29	3.60	0.42	3.44	0.05	3.51	0.13	1.1
Hexane	3.68	0.11	3.67	0.24	3.76	0.38	3.61	0.02	3.68	0.11	1.0

EGCF–Kubo. Complementing the ensemble/vibronic estimators above, we also construct emission spectra via the energy-gap correlation and second-order cumulant (Kubo) formalism (Methods I, Sec. 2.3.5.3). From the 0.5 fs time series of the vertical emission gap $\Delta E(t)$ (about 2×10^6 values per solvent), the stationary autocorrelation $C(t)$ was computed, gentle long-t damping and zero padding were applied, and a cosine transform was performed. Spectra were then assembled from the second-order cumulant lineshape (including the standard ω^3 factor for emission) and area-normalized. Figure 69 shows the resulting EGCF–Kubo emission spectra on the *energy* (top) and *wavelength* (bottom) axes; vertical dotted lines mark the per-solvent maxima. Peak positions follow the expected solvatochromic ordering (water → red/broad; hexane → blue/narrow), consistent with NEA/E–ZTFC. Quantitatively, EGCF peaks are typically red–shifted by ~0.08–0.12 eV relative to NEA/E–ZTFC for several aprotic solvents (see Tables 32 and 33). This behavior is reasonable: the cumulant spectrum is anchored to the *dynamics* of $\Delta E(t)$ and, under finite–window damping and mildly non-Gaussian fluctuations, its maximum can lie closer to the mean gap than the ensemble histogram maximum. Widths behave as anticipated: water shows the largest FWHM, while weakly polar/aprotic media are comparatively narrow. Practically, EGCF complements NEA/E–ZTFC: it adds explicit dynamical broadening and a transparent spectral–density $J(\omega)$ at the cost of dense storage and stricter sampling (uniform Δt, sufficient correlation–time coverage).

Cross–method consistency. Across solvents, all four estimators preserve the same ordering of emission maxima (Figs. 66, 67, 68, 69). E–ZTFC peaks are systematically blue–shifted relative to NEA by ~ 0.07–0.12 eV, and EGCF peaks are typically modestly red–shifted relative to NEA/E–ZTFC for several aprotic solvents (energy–domain values in Table 32).

Peak positions and widths (energy domain). Histogram fits report $\boldsymbol{\sigma}$ (Gaussian SD); NEA/E–ZTFC/EGCF report $\boldsymbol{\Gamma}$ (FWHM, no Gaussian assumption). Trends are solvent dependent: water shows the lowest μ (red–shift) and the broadest bands (e.g., NEA $\Gamma = 0.73$ eV, E–ZTFC $\Gamma = 0.77$ eV, EGCF $\Gamma = 0.49$ eV), whereas non–polar hexane is blue–shifted and very narrow (EGCF $\Gamma = 0.02$ eV).

Relation to dielectric environment. Including the simulated relative permittivity $\varepsilon_r^{\mathrm{Sim}}$ in Table 32 links spectral shifts to polarity: larger ε_r (e.g., water 97.1) correlates with red–shifted and broader emission, while small ε_r (hexane 1.0) correlates with blue–shifted, narrower bands (cf. Table 20).

Wavelength domain and calibration. Peak wavelengths $\boldsymbol{\mu}^{\lambda}$ and widths $\boldsymbol{\Gamma}^{\lambda}$ are listed before and after a single rigid calibration that aligns the methanol spectrum to 570 nm (Table 33). Because the calibration is a constant shift per method, Γ^{λ} is unchanged between panels (a) and (b).

Experiment vs. prediction and dielectric trends. Using the calibrated NEA peaks (Table 33b), the experimental ordering from Kucherak *et al.*[42] for the solvents available in our set is MeOH (570) > ACN (518) > DMF (512) > DCM (497) > Dioxane (445) nm. The NEA ordering is MeOH (569) > ACN (566) > DMF (563) > Dioxane (552) $\gtrsim$ DCM (552) nm, i.e., correct except for a nearly tied (slightly inverted) DCM/Dioxane pair. Overall, the predicted solvent spread is strongly *compressed* relative to experiment, consistent with the fixed-charge solvent models: literature ε_r are underestimated for several aprotic solvents and overestimated for water (Table 20). Across all ten solvents, longer wavelengths broadly increase with ε_r (red-shift in more polar media); the DCM vs. Dioxane reversal reflects both the small $\Delta\varepsilon_r$ at the low end and compressed Stark response in DFTB/MM. These trends support the interpretation that dielectric bias maps into the ESP feature space and narrows the solvatochromic lever arm. Upgrading the QM level and/or using polarizable embeddings should widen the dynamic range.

NN vs. TD–LC–DFTB at identical geometries. Peak energies from NN histograms closely match TD–LC–DFTB recomputations at the same snapshots; see Fig. 66. This supports the use of NN predictions for large-scale spectral assembly.

Take–home. (i) Solvent polarity controls both peak position and width; (ii) E–ZTFC introduces modest blue shifts and broader profiles relative to NEA while preserving solvent ordering; (iii) EGCF adds dynamical broadening and a route to $J(\omega)$ with stricter sampling requirements; (iv) NN-based spectra reproduce TD–LC–DFTB trends with near-quantitative agreement. Method settings are listed in App. Table A.25.

Table 33.: Spectrum wavelength-domain peak positions μ^λ and FWHM Γ^λ (both in nm) for four estimators (Hist, NEA, E–ZTFC, EGCF). (a) lists uncalibrated peaks; (b) shows peaks after applying a single offset so that the methanol spectrum aligns to 570 nm. *Since calibration is a rigid shift, Γ^λ is identical in (a) and (b).*

(a) Peak wavelengths μ^λ and widths Γ^λ (uncalibrated).

	Hist		**NEA**		**E–ZTFC**		**EGCF**	
Solvent	μ^λ	Γ^λ	μ^λ	Γ^λ	μ^λ	Γ^λ	μ^λ	Γ^λ
Water	415	100	406	95	395	98	453	82
DMSO	366	36	367	37	354	46	379	10
DMF	365	37	363	37	352	46	377	9
ACN	368	40	366	40	356	50	384	11
Methanol	370	43	369	45	357	52	387	10
Acetone	367	42	365	43	355	51	384	11
DCM	353	27	353	27	345	39	360	7
Ethyl acetate	362	42	361	42	349	49	380	8
Dioxane	352	29	353	29	341	40	360	6
Hexane	337	23	337	22	328	34	344	2

(b) Peak wavelengths μ^λ and widths Γ^λ (calibrated to MeOH = 570 nm). Experimental[42] maxima; dielectric constants ε_r (literature[56] at 298 K; simulation at 300 K) listed in Tab. 20; rounded to integer.

	Hist		**NEA**		**E–ZTFC**		**EGCF**		**Exp.[42]**	ε_r	
Solvent	μ^λ	Γ^λ	μ^λ	Γ^λ	μ^λ	Γ^λ	μ^λ	Γ^λ		**Lit.[56]**	**Sim.**
Water	613	100	606	95	606	98	636	82	—	78	97
DMSO	564	36	567	37	565	46	561	10	—	46	44
DMF	563	37	563	37	563	46	560	9	512	37	23
ACN	566	40	566	40	566	50	567	11	518	36	24
Methanol	568	43	569	45	568	52	570	10	570	33	20
Acetone	564	42	565	43	566	51	567	11	—	21	10
DCM	551	27	552	27	556	39	543	7	497	9	4
Ethyl acetate	560	42	561	42	560	49	563	8	—	6	5
Dioxane	550	29	552	29	551	40	543	6	445	2	1
Hexane	535	23	537	22	539	34	527	2	—	2	1

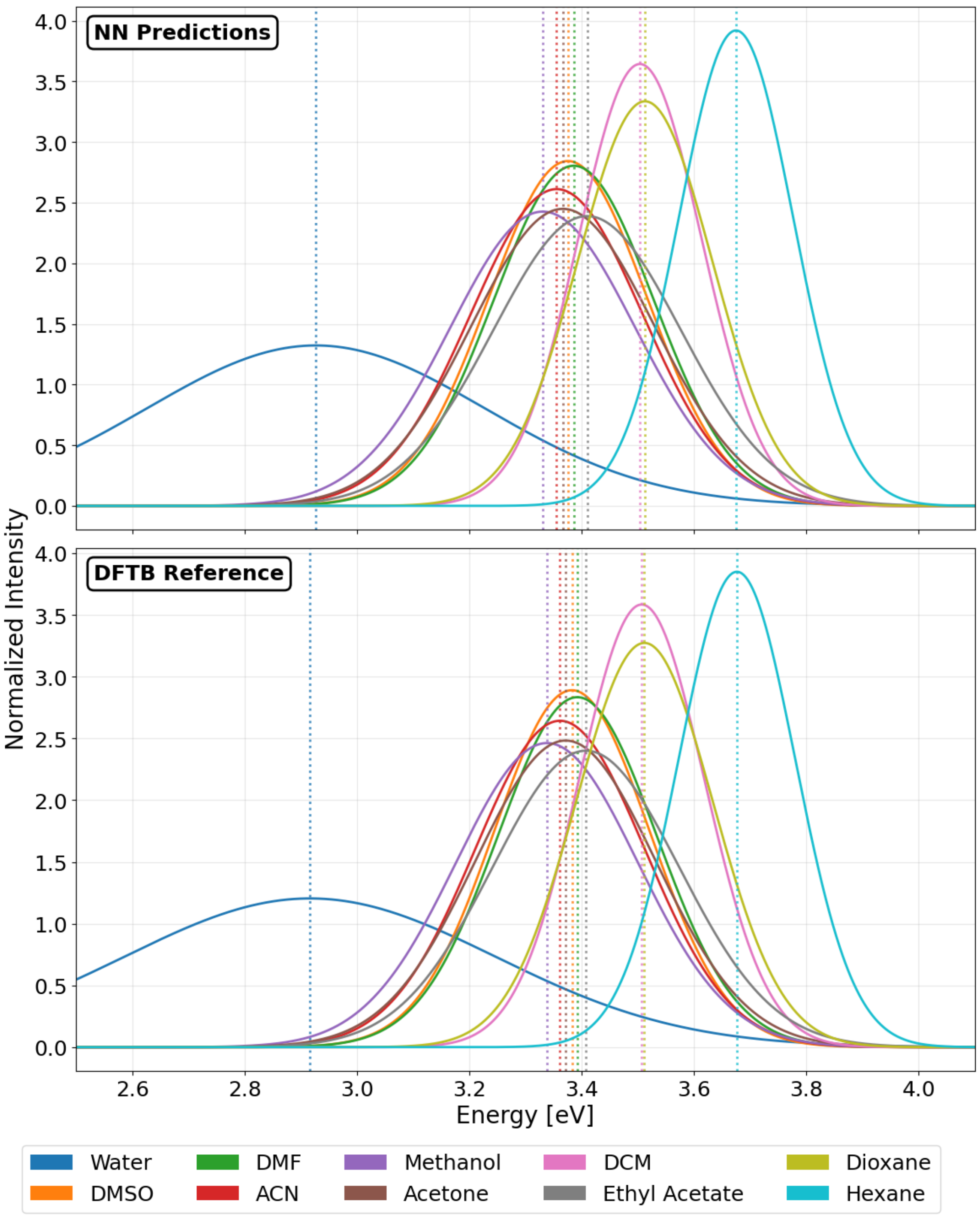
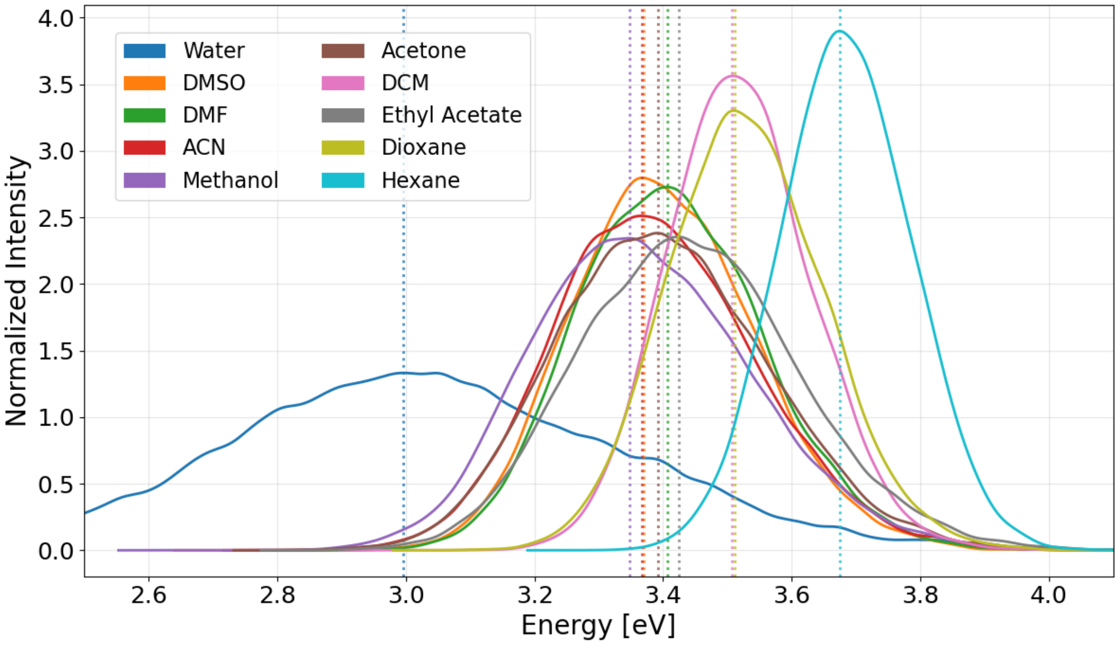
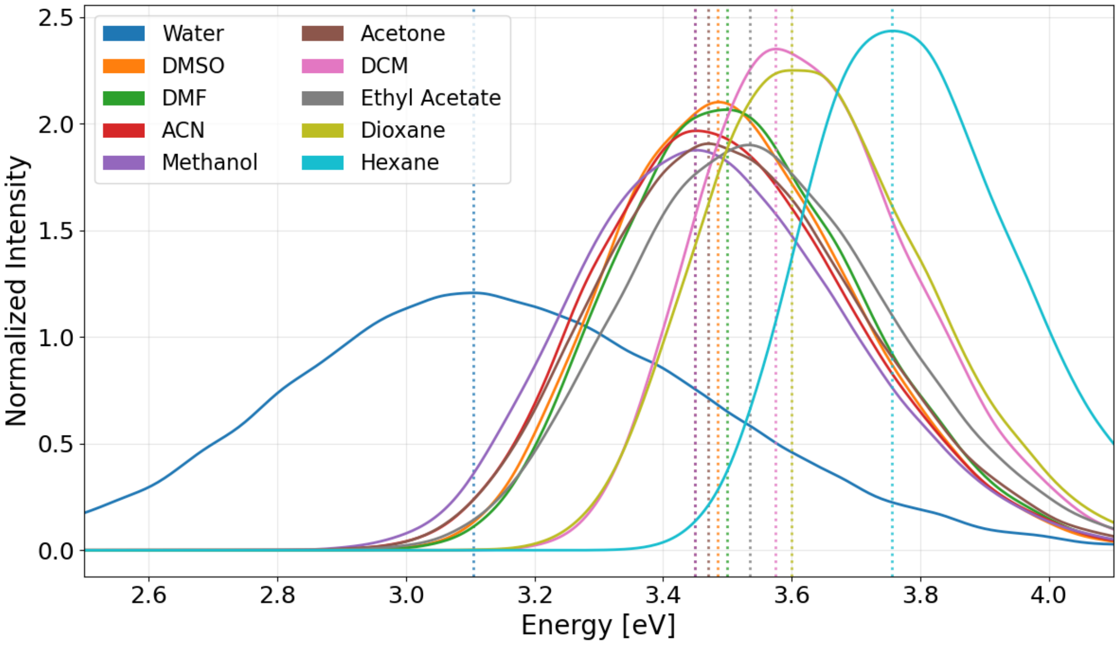


Figure 66.: Fluorescence spectra obtained from weighted histograms of vertical emission energies (method overview in Sec. 2.3.5). **Top:** NN predictions; **bottom:** TD-LC-DFTB recomputations at the same NN/MM geometries (reference). For each solvent, histograms of E_i are weighted by f_i and ω^3, smoothed/fitted, and area-normalized; the **dotted vertical lines** mark peak positions. NN and DFTB show consistent solvent ordering of the maxima, with water displaying the broadest, red-shifted band and non-polar solvents yielding narrower, blue-shifted bands.

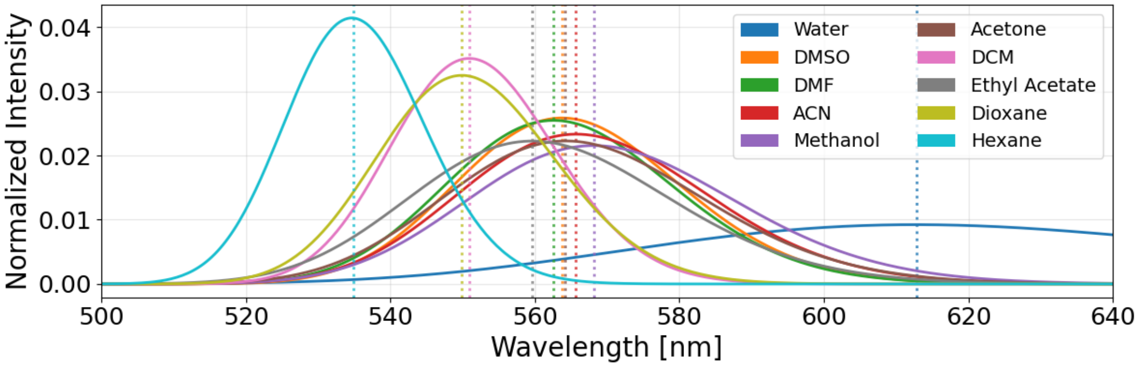


(a) Nuclear Ensemble Approach (NEA) emission spectra from NN/MM snapshots; area-normalized.

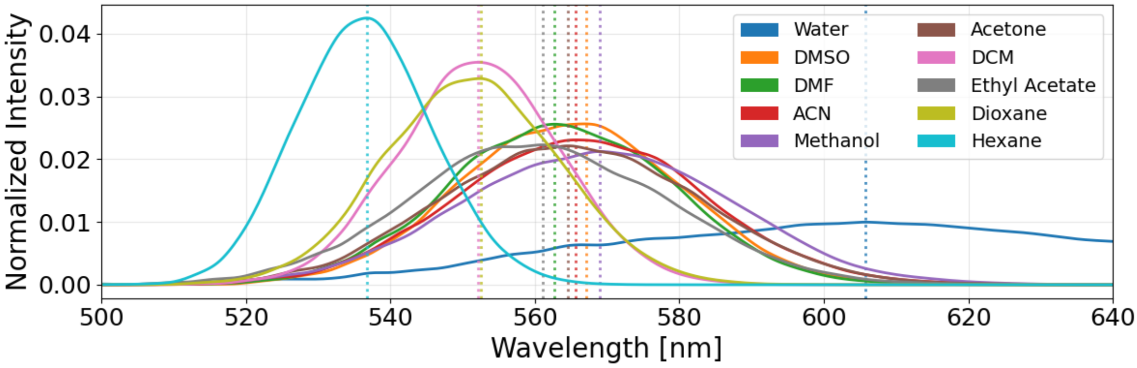


(b) E–ZTFC emission spectra built from the same snapshots using a single 0 K vibronic shape J_{ZTFC}.

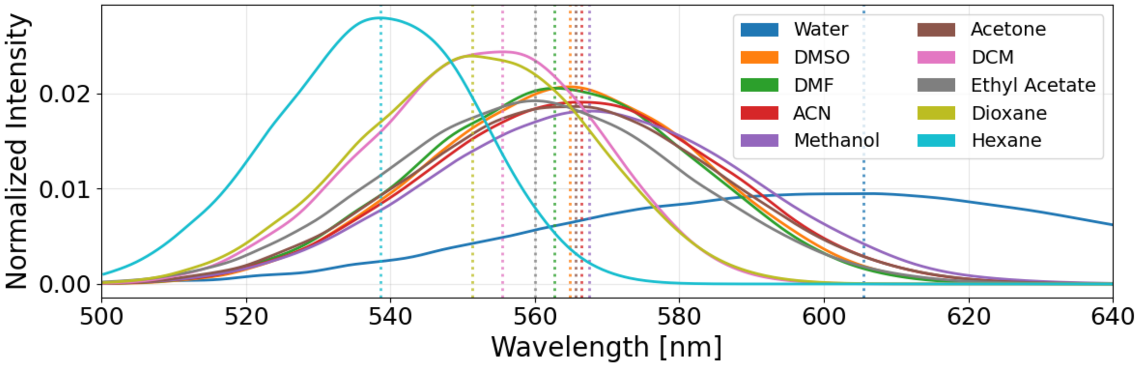
Figure 67.: Comparison of **NEA** (top; Sec. 2.3.5.1) and **E–ZTFC** (bottom; Sec. 2.3.5.2) emission spectra across solvents. Both panels use the identical NN/MM snapshot set and oscillator-strength weighting (including the ω^3 factor for emission) with area normalization. Dotted vertical lines indicate per-solvent peak positions. NEA applies a homogeneous-broadening kernel to the ensemble of vertical gaps, whereas E–ZTFC augments the ensemble by convolving with a precomputed 0 K vibronic shape, adding systematic vibronic structure while preserving the solvent ordering of the maxima.



(a) Histogram method.

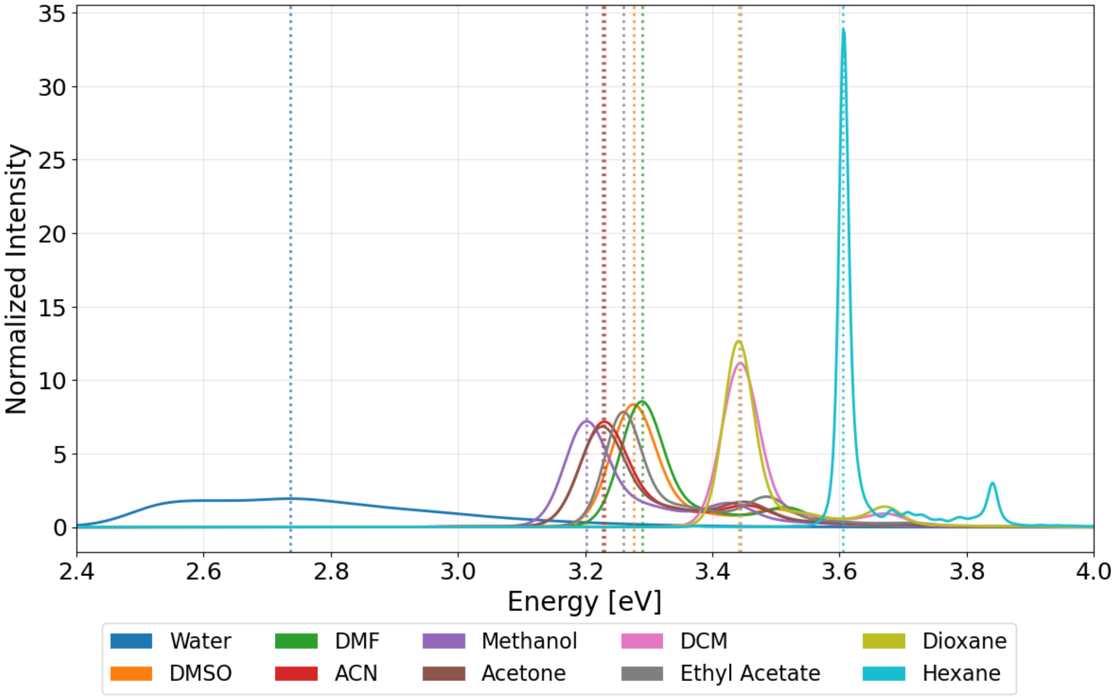


(b) Nuclear Ensemble Approach (NEA).

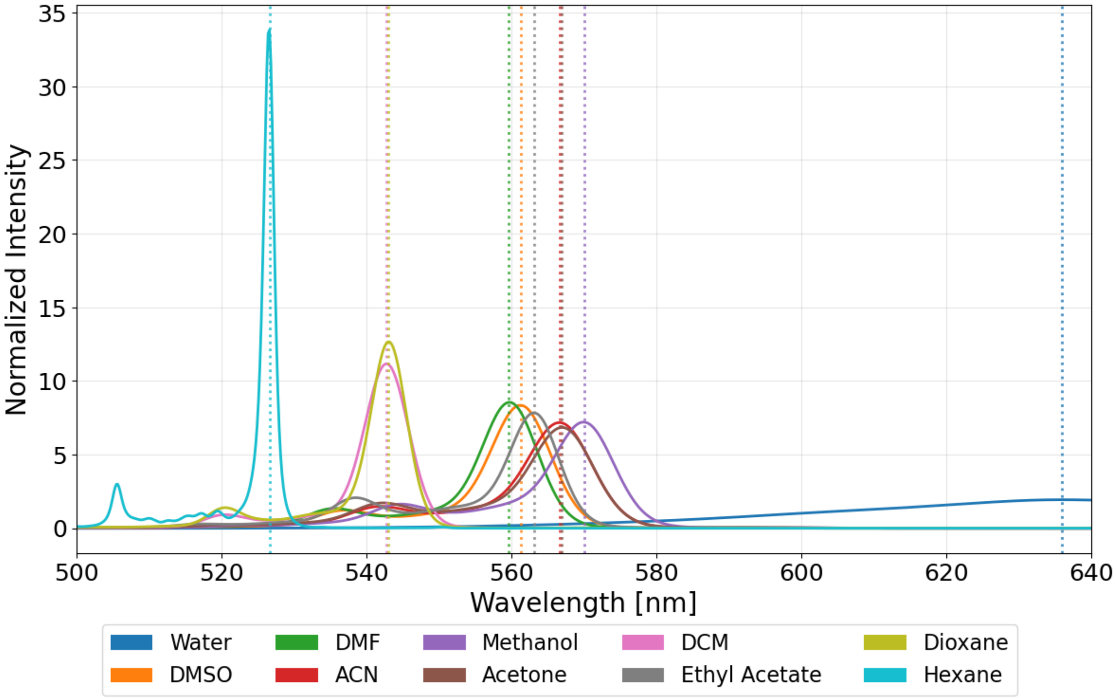


(c) Ensemble–Zero-Temperature Franck–Condon (E–ZTFC).

Figure 68.: Fluorescence emission spectra from three estimators. Panels (A)–(C) show normalized intensity vs. wavelength for the *Histogram*, *NEA*, and *E–ZTFC* constructions, respectively. Colored curves correspond to different solvents (legend); vertical dotted lines mark the corresponding emission maxima for each solvent. All spectra are *predicted by a neural-network model.* Spectra were assembled on the energy axis and transformed to wavelength using $I_\lambda(\lambda) = I_E(E(\lambda))\,(hc/\lambda^2)$ with $hc = 1239.84$ eV nm, then normalized for shape comparison. A *single rigid wavelength offset* was applied (per method) so that the mean emission wavelength of the methanol spectrum matches the experimental value of 570 nm; the same offset is used for all solvents shown.



(a) Energy domain: EGCF–Kubo emission spectra.



(b) Wavelength domain: EGCF–Kubo emission spectra; λ obtained from E with the Jacobian, cf. Methods I.

Figure 69.: EGCF-Kubo emission spectra from 1 ns NN/MM runs with 0.5 ns frame spacing (every frame saved; 2×10^6 samples per solvent).

6.6. Discussion

The central aim of this chapter was to learn and follow the emissive bright state through frequent reorderings in the S_1–S_3 manifold and to use this capability to assemble solvent-dependent fluorescence spectra efficiently. On the force side, both the inverse-distance + ESP model (Methods II, Sec. 3.6.1) and the MACE–EQ variant (Sec. 6.4.3) deliver high-fidelity forces on the bright surface, enabling stable NN/MM dynamics across ten solvents (Table 27). On the property side, the network maps structure + ESP to bright-state vertical energies and oscillator strengths with low errors (Fig. 65, Table 31), allowing construction of emission spectra whose peak orderings and widths follow expected solvatochromic trends (Figs. 66–68, 69; Tables 32, 33).

Bright–state identity and switching. The observations are consistent with the *hypothesis* that solvent–specific electrostatics are a principal driver of state reordering. Vacuum controls show essentially no $S_1 \leftrightarrow S_2$ exchanges (Table 22), whereas per–atom ESP distributions vary strongly across solvents and sites (Fig. 55), providing a plausible mechanism for solvent–induced reordering. Using fixed S_2 charges for NN/MM aligns with the predominance of S_2 and avoids embedding discontinuities (Fig. 54). A posteriori assignment along NN/MM trajectories (Table 28) indicates that dynamics remain close to the intended bright surface. As an illustration, Fig. 70 overlays TD–LC–DFTB S_1–S_3 recomputations for each NN/MM geometry in DMSO and compares the NN emission energy with the TD–LC–DFTB energy of the brightest state at each frame, showing close agreement; analogous traces for the other solvents are provided in the Appendix.

Example trajectory (DMSO). Figure 70 illustrates an NN/MM trajectory of *FR0* in **DMSO**. The top panel overlays TD-LC-DFTB recomputations for S_1–S_3 at each NN/MM geometry and marks the instantaneous bright state; the bottom panel compares the NN emission energy with the TD-LC-DFTB energy of the *brightest* state at each frame, showing close agreement along the trajectory. NN/MM itself carries no state label – state identity is assigned *a posteriori* from the TD-LC-DFTB evaluation. Analogous plots for the remaining nine solvents are provided in the Appendix (Figs. A.1, A.2, A.3, A.4, A.5).

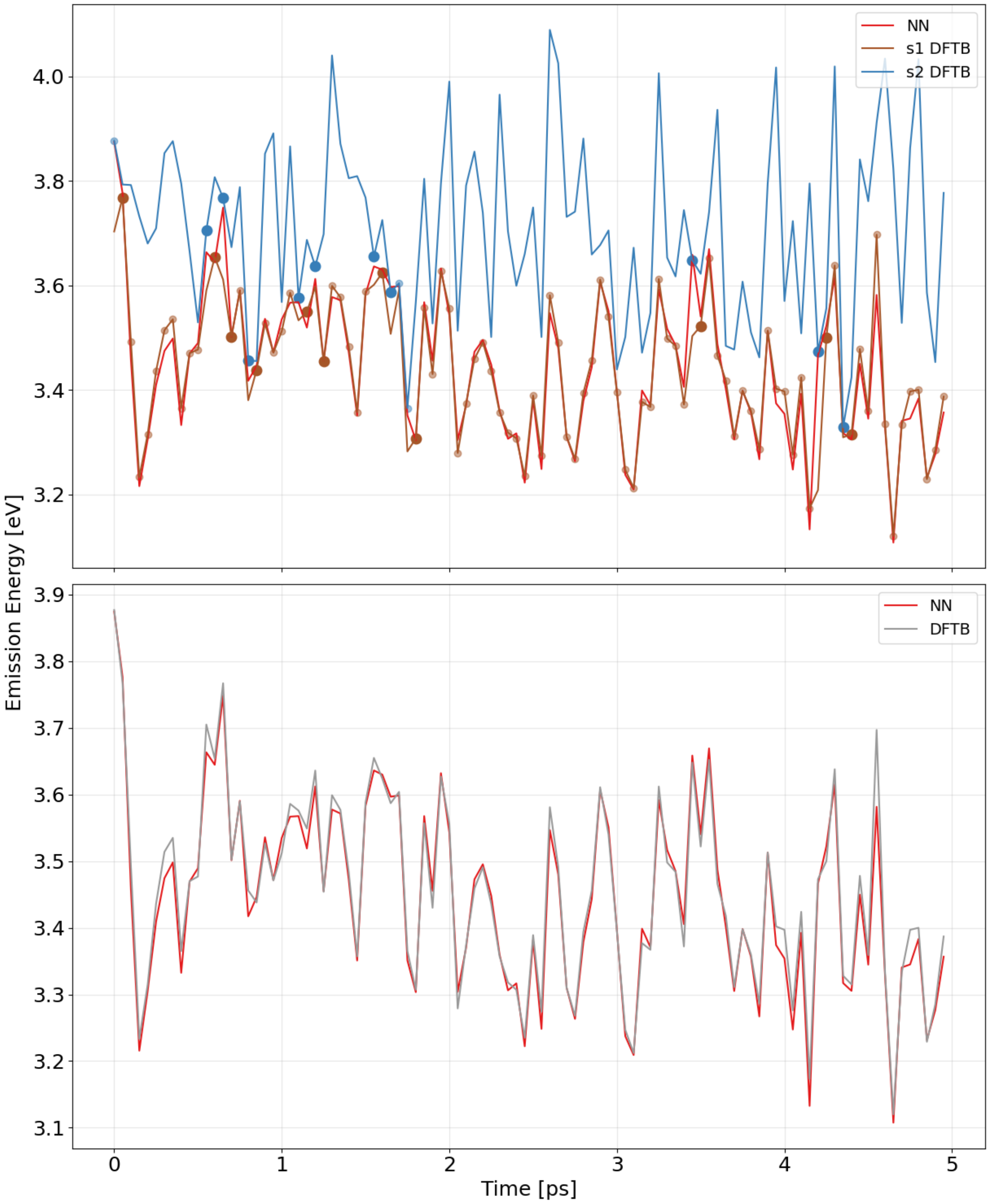


Figure 70.: Emission-energy trace for an NN/MM trajectory of **FR0** in **DMSO**. **Top:** Each point corresponds to one NN/MM geometry. The red curve is the NN-predicted vertical emission energy; the overlaid TD-LC-DFTB recomputations for S_1 and S_2 at the *same* geometries are shown in their respective colors. Marker colors indicate which of S_1–S_2 is the bright state for that frame based on the DFTB recomputation. **Bottom:** Direct comparison of the NN energy (red) with the DFTB energy of the *brightest* of S_1–S_2 at each frame, illustrating close agreement along the trajectory. NN/MM itself does not carry a state label; state identity is inferred *a posteriori* from the TD-LC-DFTB evaluation. See Appendix Figs. A.1, A.2, A.3, A.4, and A.5 for the other solvents.

Force models: accuracy, stability, and dynamic range. The pooled inverse-distance + ESP model reaches $R_F^2 \approx 0.99$ (Fig. 56); stability improves with smaller steps and broader training (Table 27). MACE–EQ further halves the force RMSE relative to the baseline and exhibits excellent $R^2 \geq 0.999$ (Fig. 59).

Energies vs. forces and cross–domain generalization. While forces are the learning target used in NN/MM, energy heads reveal useful diagnostics. In LOSO tests, hexane and ethyl acetate show energy pathologies when held out (Table 25, Fig. 58), manifesting as near-parallel ridges consistent with small, solvent-specific offsets. A plausible driver is the *mismatch of ESP support* when very narrow (hexane) and broad (polar) distributions are mixed (cf. Fig. 52). This does not undermine NN/MM because forces remain accurate; two straightforward mitigations are recommended in pooled training: (i) make input statistics comparable across solvents (e.g., per-solvent z-normalization of per-atom ESPs or inclusion of a learned solvent identifier), and (ii) increase the representation of rare, high-force frames (e.g., oversample those frames or add a few targeted reference windows) so that pooled training sees the tails that single-solvent models naturally learn.

Computational feasibility and speed. Following the physical bright-state trajectory is not practically attainable with direct TD-LC-DFTB/MM: standard TD-DFTB provides states in a fixed ordering per step, so one can run separate $S_1/S_2/S_3$ trajectories, but this *does not* reproduce the true bright-state path across reorderings and thus cannot yield the correct emission statistics. The NN/MM force model *does* enable bright-state following and reduces wall time dramatically: a 1 ns at 1 fs time-steps explicit-solvent run that would require ∼90–120 days per solvent with TD-LC-DFTB/MM (reference-only S_1/S_2 trajectories) completes in ∼7–10 hours at 0.5 fs time-steps with NN/MM, i.e., well below 24 h (see Table 26 for representative timings and speedups). This combination of *feasibility* and *orders-of-magnitude speedup* is a principal advantage of the approach.

Spectral estimators and consistency. Across Hist/NEA/E–ZTFC/EGCF, the ordering of emission maxima is consistent and tracks polarity (water red/broad → hexane blue/narrow), with modest, method–characteristic shifts (NEA → E–ZTFC: blue by ∼0.07–0.12 eV; EGCF: mildly red for several aprotics), see Table 32. EGCF adds dynamical broadening tied to the gap autocorrelation and provides access to a spectral density at the cost of dense sampling (Fig. 69).

Comparison to experiment (Kucherak 2010). For the solvents common to our study and Ref. [42], the experimental ordering of fluorescence maxima (blue → red; FR0, $\lambda_{\max}$ in nm) is: *1,4-dioxane (445) < DCM (497) < DMF (512) < ACN (518) < MeOH (570).* Using the calibrated wavelength peaks (Tab. 33b; aligned to MeOH), the NEA estimator gives: *DCM (552)* $\lesssim$ *1,4-dioxane (552) < DMF (563) < ACN (566) < MeOH (569).* Thus, the relative order among *DMF–ACN–MeOH* matches experiment (ACN redder than DMF; MeOH the reddest), while the *DCM/dioxane* pair is essentially degenerate in NEA and hence numerically ambiguous. The other estimators behave similarly: *Histogram* and *E–ZTFC* reproduce the experimental relation *dioxane* < *DCM* (Tab. 33b), whereas *EGCF* yields another near tie (DCM 543 nm vs. dioxane 543 nm). Given the tiny separations between DCM and dioxane in our predictions, this local inversion should be viewed as within method/fit uncertainty rather than a systematic trend.

Absolute offsets are expected because the training/reference level is TD-LC-DFTB. Here a single-reference alignment (MeOH → 570 nm) was used to remove the dominant rigid shift and focus on *relative solvatochromic ordering*. Residual differences originate chiefly from (i) the dielectric response of the fixed-charge MM models (Tab. 20), which compresses/expands the ESP feature range that drives solvatochromic shifts, and (ii) the TD-DFTB electronic structure itself. For quantitative agreement to experiment, the same workflow could be retrained/validated at a higher level (e.g. ωB97X/def2-TZVP in TDA-TDDFT) or by recalibrating the MM embedding (polarizable FF or ESP rescaling). The present results should be read as a *proof of principle*: the NN/MM force model reproduces the bright surface well enough to generate physically ordered spectra, and the property model tracks relative shifts across solvents; the absolute placement then follows from the chosen electronic-structure level.

Role of the solvent dielectric response. Simulated permittivities deviate from literature by -25 to -55% for several organics and are overestimated for water (Table 20). Because ESP enters as an input feature, such dielectric biases compress or inflate the solvent–response space seen by the networks and help explain near-rigid offsets in absolute peak positions and widths. A constant wavelength shift used for visualization (Table 33b) improves apparent agreement but does *not* correct the underlying physics; the offsets are attributable to the fixed-charge embeddings. Embedding improvements – polarizable force fields, reparametrized partial charges for key solvents, or calibrated ESP rescaling – are therefore recommended.

Limitations and extensions. (i) Rare bright S_3 episodes are largely filtered by the brightness threshold and remain underrepresented; targeted sampling or multi-state training could reduce residual label noise. (ii) The fixed S_2 charge embedding is physically motivated but approximate; state-consistent polarizable embeddings or learned charge equilibration during dynamics (QEq-style) are promising. (iii) For proteins, strongly heterogeneous and anisotropic fields will stretch the ESP distribution; environment-aware embeddings and periodic active learning should help maintain coverage. (iv) FR0 is only moderately flexible—apart from two terminal ethyl rotors and a formyl twist the surface is relatively flat; in free MD the formyl torsion *does not* freely rotate but remains localized. This keeps the learning task challenging yet tractable; more fluxional chromophores (e.g., retinal) are expected to be harder. (v) The cutoff $f \geq 0.7$ is a pragmatic choice that works for FR0, where a dominant bright state is usually present; systems without a clear bright state (e.g., 4AP) call for a multi-state formulation – learn simultaneous S_1–S_3 energies/forces with state hopping. (vi) Extension to dye–protein complexes is plausible but remains to be validated; heterogeneous fields and broader conformational space will likely require modest additional training (fine-tuning/active learning on protein snapshots) and systematic testing before routine application.

Practical guidance. For routine NN/MM sampling of solvatochromic dyes, the inverse-distance + ESP model with 0.5 fs steps and pooled training offers a robust default; use MACE–EQ as a high–fidelity reference or drop-in replacement when very low force errors are required. When transferring across distinct polarity regimes, prefer solvent-aware normalization and periodically augment training with tail frames. For spectra, NEA is a reliable baseline; E–ZTFC adds vibronic texture, and EGCF is recommended when time-correlation information is of interest.

6.7. Summary

An explicit-solvent workflow was developed that *learns the emissive bright state* and preserves its identity across frequent S_1–S_3 reorderings. A force model (inverse-distance + ESP) enables stable, ns-scale NN/MM at a tiny fraction of TD–LC–DFTB cost, and a property model maps structure + ESP to vertical emission energies and oscillator strengths for spectrum assembly. Across ten solvents, spectra from Hist/NEA/E–ZTFC/EGCF reproduce the expected solvatochromic ordering and bandwidth trends; residual absolute offsets are largely attributable to dielectric biases of fixed-charge solvent models that modulate the ESP feature space. Crucially, direct TD-LC-DFTB/MM cannot follow the *physical* bright-state trajectory across state reorderings; at best, separate S_1/S_2/S_3 reference trajectories can be run, which require ~90–120 days at 1 fs time-steps per solvent and still do not yield the correct bright-state statistics. In contrast, NN/MM completes 1 ns at 0.5 fs time-steps per solvent in ~7–10 hours (well below 24 h), see Table 26, providing both feasibility and $\mathcal{O}(10^4-10^6)\times$ speedups. A higher-capacity benchmark (MACE–EQ) halves force RMSE while retaining excellent correlation and broader dynamic-range coverage, supporting its use as a reference or drop-in replacement. Overall, the results establish a practical, identity-consistent ML–MM pipeline for fluorescence prediction in complex environments and provide a template for extension to protein–dye systems.

Part VII

Neural Network Prediction of Retinal Absorption Spectra

7.1. Overview

This chapter integrates retinal absorption as a second pillar in the thesis-wide theme of *machine-learning surrogates for spectroscopic observables.* Specifically, excitation energies (E) and oscillator strengths (f) for retinal in solution and in protein environments are targeted. The aims are twofold: (i) to consolidate and contextualize prior results obtained with a neural-network (NN) framework when applied to retinal, and (ii) to extend these results with additional validation steps and quantitative coverage metrics. In contrast to the fluorescence chapter (Chapter VI), which focuses on emission and may include force learning for dynamical propagation, the present chapter deals solely with *absorption*; consequently, the data design differs in detail, while the core NN architecture remains reusable.

Author Contributions. This chapter is complementary to the retinal-absorption chapter in Katharina Spies' dissertation and shares the same foundational building blocks: the neural-network architecture developed by the present author and the fundamental datasets generated by Katharina Spies together with supervised students Béryl Naëmi Greb, Lara Nagel, and Jemina Enkelmann (solvent/ion ensembles for 11-cis and all-trans; protein preparations; excited-state labels and structural analyses).[222–224] Building on this basis, the present author performed independent data curation, retrained the networks, and added quantitative comparison and coverage metrics. For the protein application, the identical trained networks from the Spies project were used to independently reproduce the reported trends; the numerical values listed here follow Spies' dissertation for consistency.

7.2. Introduction

Retinylidene proteins span a broad color range because the retinal chromophore's $S_0 \rightarrow S_1$ gap is highly sensitive to its local electrostatic environment and to conjugation along the polyene into the β-ionone ring.[225,226] In practice, counterions near the protonated Schiff base (RSBH$^+$), hydrogen-bond networks, and ring planarity cooperate to tune absorption via differential stabilization of ground vs. excited states.[225,227] This manifests from blue-shifted channelrhodopsins to red-absorbing variants such as Chrimson.[227–229] These color-tuning principles are well established across microbial and animal rhodopsins and are consistent with structural and spectroscopic data.[225,226,230] See Fig. 71 for the 11-cis → all-trans photoisomerization and the structural landmarks discussed above.[225,226]

From a physics perspective, two diagnostic levers connect structure to spectra: (i) bond–length alternation (BLA) along the polyene and (ii) the local electric field at the chromophore, represented here by per-atom electrostatic potentials (ESP).[225,230–233] The BLA–excitation-energy relation is well documented across rhodopsins and rationalizes blue/red shifts via changes in conjugation.[230,231] Electrostatic color tuning – most prominently the effect of counterions near the protonated Schiff base and charges near the β-ionone ring –

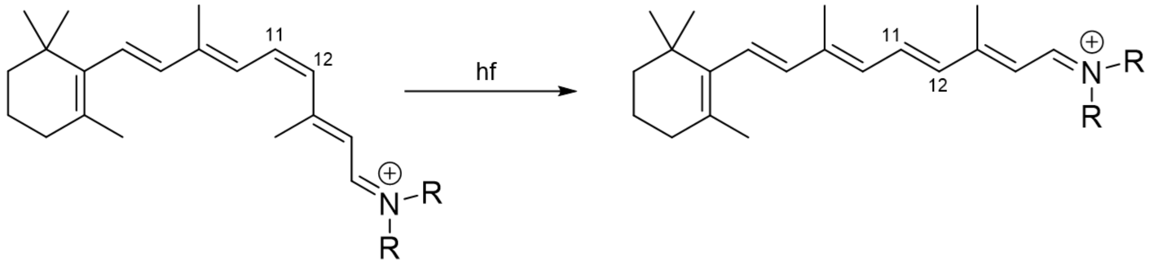


Figure 71.: Photoisomerization of retinal from 11-cis to all-trans upon photon absorption ($h\nu$). The C_{11}=C_{12} double bond (labels 11 and 12) changes configuration; the β-ionone ring and the protonated Schiff base ($RSBH^+$) are indicated. This elementary event initiates the rhodopsin photocycle.[225,226]

has been established in structural/spectroscopic reviews and explicit modeling studies, including polarization-aware treatments of the opsin shift.[225,232,233]

Accurate modeling of these effects with excited-state QM/MM is computationally demanding because spectral observables converge only with broad conformational sampling (ion placements, side-chain motions, torsions), and because technical details such as periodic-boundary handling in electrostatic embeddings can bias targets if not treated carefully, as discussed in Section 7.3.[226] This motivates machine-learning surrogates trained on QM/MM snapshots: once clean reference targets are available, neural networks can map geometry and ESP to excitation energies and oscillator strengths at negligible runtime.[16] Recent work surveys ML for electronically excited states and photodynamics and demonstrates protein-relevant use cases, including SchNarc and SPAINN frameworks, Δ-style corrections and solvatochromism, as well as exciton-transfer/electrostatics-aware applications.[16,234–239]

Within the broader thesis narrative, retinal absorption provides a compact but stringent test case for the machine-learning strategy pursued here. The network architecture introduced in the fluorescence chapter (Chapter VI) is deliberately reused, but the task is changed: here, absorption properties ($S_0 \rightarrow S_1$ excitation energies and oscillator strengths) are predicted, whereas the fluorescence chapter focuses on emission. This separation keeps the learning machinery constant while highlighting absorption-specific data design and validation.

7.3. Dataset Construction

Two retinal isomers (*all-trans*, *11-cis*) were simulated in vacuum and multiple solvents (water, alcohols, anisole, toluene, DMSO, pyridine, THF). To emulate counterion scenarios, ions were placed near $RSBH^+$ or the β-ionone ring and restrained at 3 Å. Classical MD was run in the NVT ensemble at moderately elevated temperature (~500 K) to enhance solvent/ion mixing, with positional restraints on the retinal chromophore to avoid known artifacts of generic force fields for RPSB systems.[229,240–242] QM/MM MD employed DFTB3 with the *3OB* parameterization[10] and its vibrationally optimized extension *3OB-f*.[11] Snapshots provided (i) QM-region geometries in a fixed definition (retinal covalently linked to lysine; consistent capping throughout) and (ii) electrostatic potentials (ESP) on QM atoms derived from the surrounding environment.

7.3.1. Excited-State Targets

Excitation energies and oscillator strengths were computed for the retinal+lysine QM region using two semiempirical levels: OM2/MRCI and long-range-corrected TD-DFTB. Only $S_0 \rightarrow S_1$ transitions surpassing a minimal oscillator-strength threshold were retained.

Target hygiene: periodic boundaries. All excited-state labels used below were computed with PBC-consistent electrostatics. Box-broken point-charge exports (no PBC unwrapping) bias LC–TD–DFTB energies, whereas PBC-aware point charges or on-the-fly ESP fields yield mutually consistent values (deviations below 0.01 eV). All targets in this chapter follow the PBC-safe route.

Table 34.: Overview of retinal training systems (solvents and ion placements) for *11-cis* and *all-trans* retinal. Entries in parentheses indicate systems *excluded from the final training set* after per-system validation because they degraded generalization of the property network (poor predictive R^2, typically < 0.85, often linked to chromophore distortions in nonpolar media with strong local fields). This pruning follows the performance-based curation described in Section 7.3 (Diagnostics). Counts per system are given in Appendix Table A.26.

	11-cis retinal						all-trans retinal					
	pure	Cl^-	2 Cl^-	Na^+	Na^+ (r)	NaCl	pure	Cl^-	2 Cl^-	Na^+	Na^+ (r)	NaCl
Vacuum	X						X					
Water	X	X	(X)	X	(X)	(X)	X	X		(X)	(X)	
Anisole							X	(X)	(X)		(X)	(X)
DMSO							X	(X)	X	X	X	(X)
Methanol	(X)	X	X	X	X	(X)	X	(X)	X			
Ethanol							X	X	(X)	X	X	X
Pyridine							X	X	(X)	(X)	(X)	X
Toluene	X	X	(X)		(X)	X	X	(X)	(X)		X	
Chloroform	X	(X)	X		(X)							
Isobutanol	X	(X)	(X)		(X)	(X)						
THF	X	X	X	(X)	X							

7.3.2. Diagnostics

Retinal is treated as a fixed QM region consisting of the chromophore and the covalent lysine linkage (protonated Schiff base, RSBH$^+$). The backbone used for all structural diagnostics follows the numbered path in Fig. 72 from the β-ionone ring into the polyene and ends at the iminium C=N of the Schiff base.

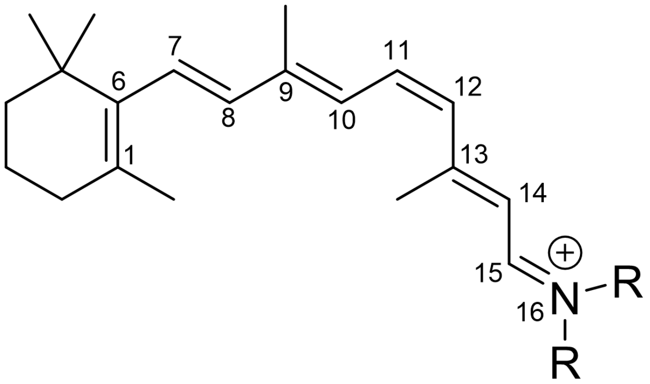


Figure 72.: Retinal backbone and bond indexing used for BLA. Atom indices 1–16 run from the β-ionone ring (1–6) through the polyene (6–15) to the protonated Schiff base (15–16).

Definition of BLA. The bond-length alternation (BLA) is evaluated along the retinal backbone from atom 1 to 16 (including the ring double bond and the iminium C=N). Define the ordered path and consecutive bonds

$$P = (1, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16)$$

$$B = \{(1, 6), (6, 7), (7, 8), (8, 9), (9, 10), (10, 11), (11, 12), (12, 13), (13, 14), (14, 15), (15, 16)\}$$

Partition B into formal double and single bonds,

$$D = \{(1, 6), (7, 8), (9, 10), (11, 12), (13, 14), (15, 16)\}$$

$$S = \{(6, 7), (8, 9), (10, 11), (12, 13), (14, 15)\}$$

With $r_{ij} = |\mathbf{R}_i - \mathbf{R}_j|$ and $|S|$, $|D|$ denoting the numbers of single and double bonds, the BLA is

$$\mathrm{BLA} = \overline{r_S} - \overline{r_D} = \frac{1}{|S|} \sum_{(i,j)\in S} r_{ij} - \frac{1}{|D|} \sum_{(i,j)\in D} r_{ij}. \tag{7.1}$$

Positive BLA indicates pronounced alternation (longer single vs. shorter double bonds); reduced or negative values reflect increased delocalization or distortions.

BLA and ESP coverage. Box plots in Figs. 73 and 74 summarize, respectively, the distributions of BLA and the electrostatic potentials (ESP) *across all QM atoms* of the retinal+lysine region and across all snapshots. The solution ensembles span—and typically exceed—the ranges encountered in the protein trajectories, indicating that protein inference largely remains within training support. Outliers originate from high-field configurations in weakly polar media.

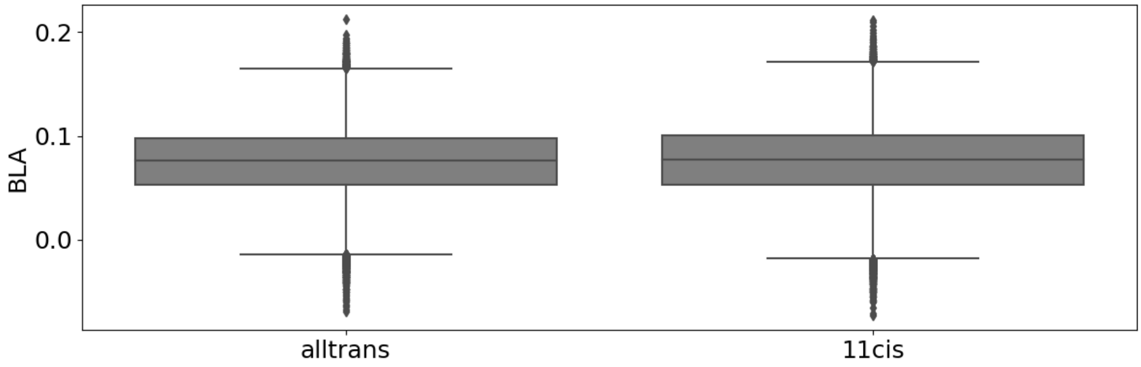


Figure 73.: Sampling of bond-length alternations (BLA) for *all-trans* and *11-cis* retinal. Boxes show the interquartile range, whiskers the 1.5 IQR; points indicate outliers.

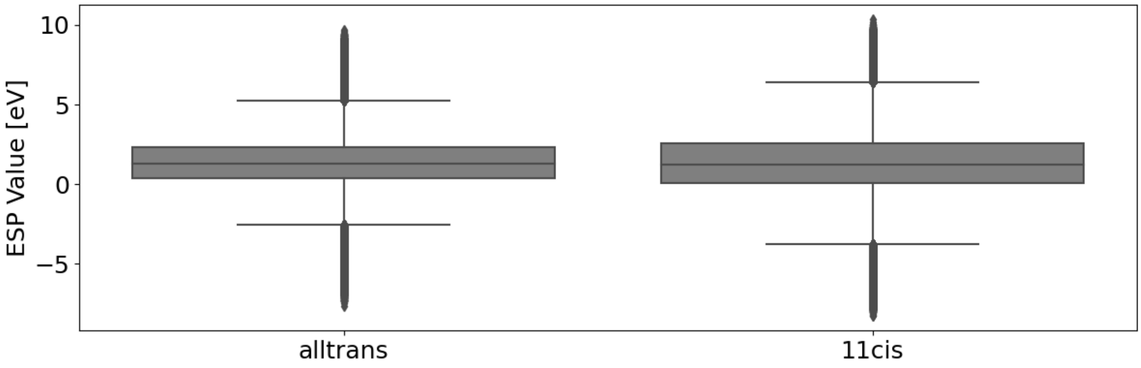


Figure 74.: Electrostatic potential (ESP) distributions sampled around the retinal chromophore for *all-trans* and *11-cis*.

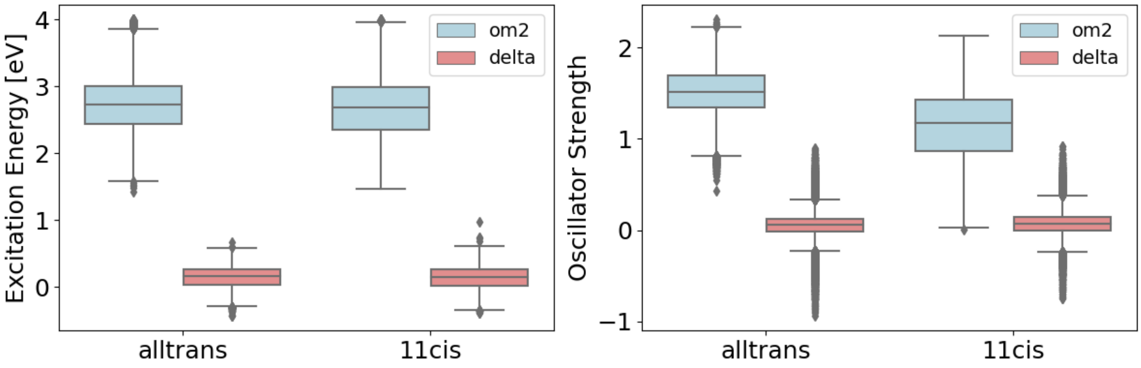


Figure 75.: Excitation energies and oscillator strengths for *all-trans* and *11-cis* retinal. Left panels: OM2/MRCI labels; right panels: Δ-learning labels.

Excitation-energy and f distributions. Aggregate statistics for excitation energies and oscillator strengths of both isomers are shown in Fig. 75. The isomers exhibit overlapping ranges with solvent-dependent shifts consistent with BLA trends; Δ-labels track OM2/MRCI with a mild energy-dependent residual and environment-specific spread (cf. Fig. 76).

Deformations and curation. Because BLA is computed as BLA $= \overline{r_S} - \overline{r_D}$ (no absolute value), a small number of snapshots in high-field, nonpolar setups show BLA < 0 and appear as outliers in Fig. 73. These cases reflect local chromophore distortions, e.g., a cation pinned near RSBH$^+$.

Motivation for Δ-learning. Define the residual $\Delta = E_{\mathrm{OM2}} - E_{\mathrm{DFTB}}$. Two baselines that exploit E_{DFTB} at inference time are a global shift $\widehat{E}_{\mathrm{OM2}} = E_{\mathrm{DFTB}} + c$ with $c = \overline{\Delta}$ (train), and an affine correction $\widehat{E}_{\mathrm{OM2}} = a + b\,E_{\mathrm{DFTB}}$. Figure 76 shows $|\Delta| \ll E_{\mathrm{OM2}}$ yet a solvent/ion-dependent spread at fixed E_{OM2}, which limits both baselines and motivates learning $\hat{\Delta} = f(\text{geometry}, \text{ESP})$.

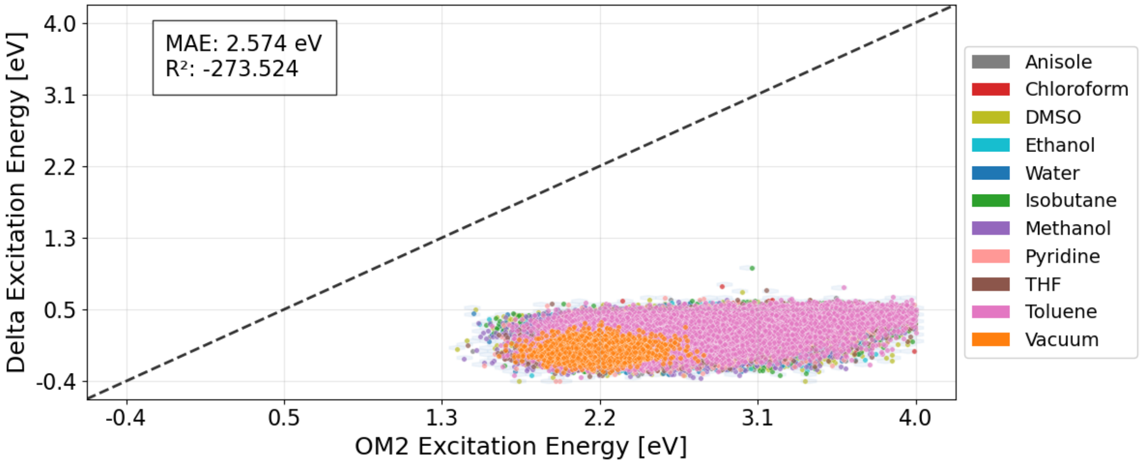


Figure 76.: OM2 excitation energy vs. $\Delta = E_{\mathrm{OM2}} - E_{\mathrm{DFTB}}$. The residual spread across solvents/ions motivates Δ-learning with geometry/ESP.

Relation of BLA and E. Across solvents and ion scenarios, the mean excitation energy $\langle E \rangle$ scales approximately linearly with the mean BLA (Figs. 77, 78): larger BLA correlates with blue shifts (e.g., anions near RSBH$^+$), smaller BLA with red shifts (e.g., cations near the β-ionone ring).

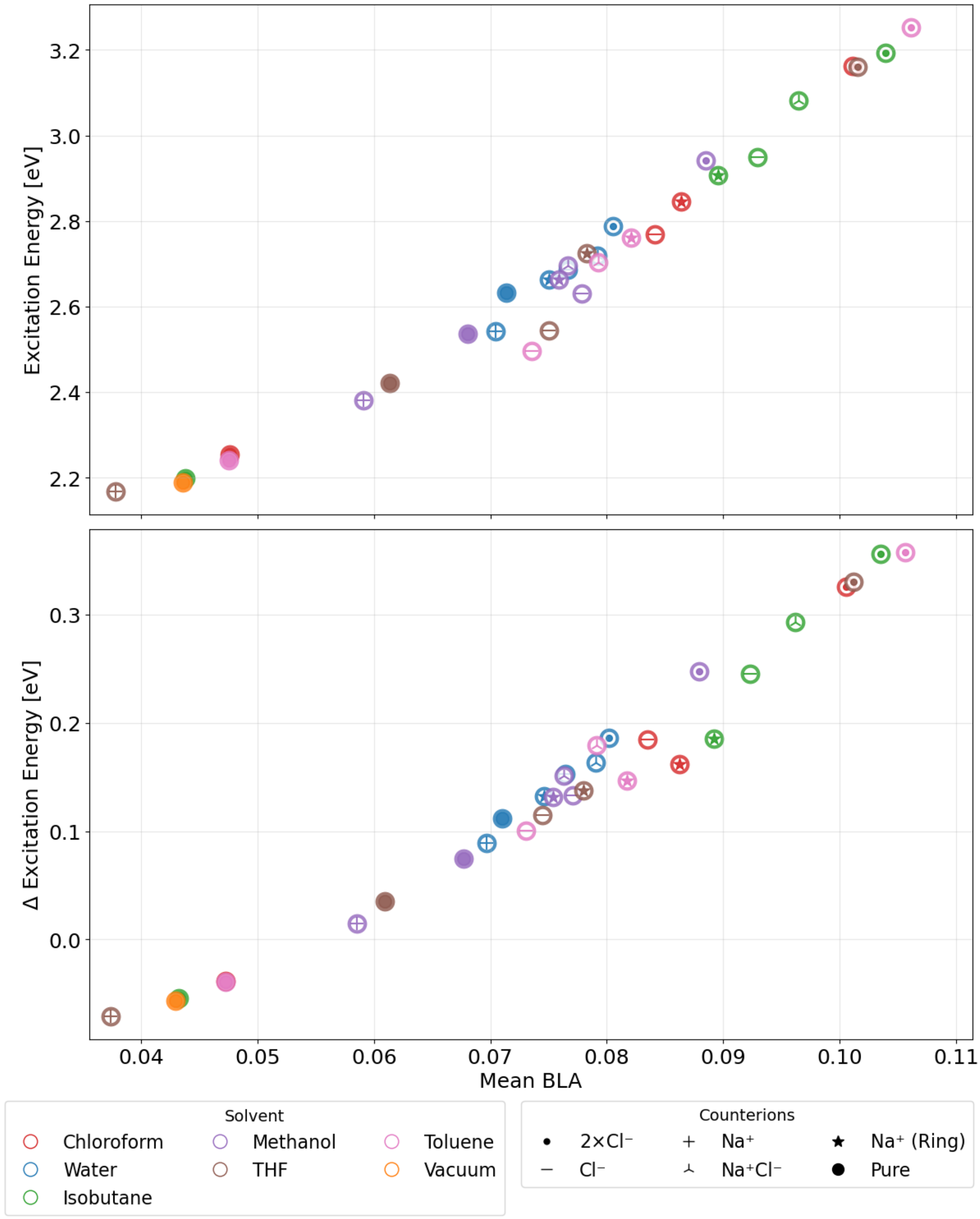


Figure 77.: Mean BLA vs. excitation energy: *11-cis* retinal (top: OM2/MRCI; bottom: Δ). Colors indicate solvents; symbols encode counterions. The data follow an approximately linear relation.

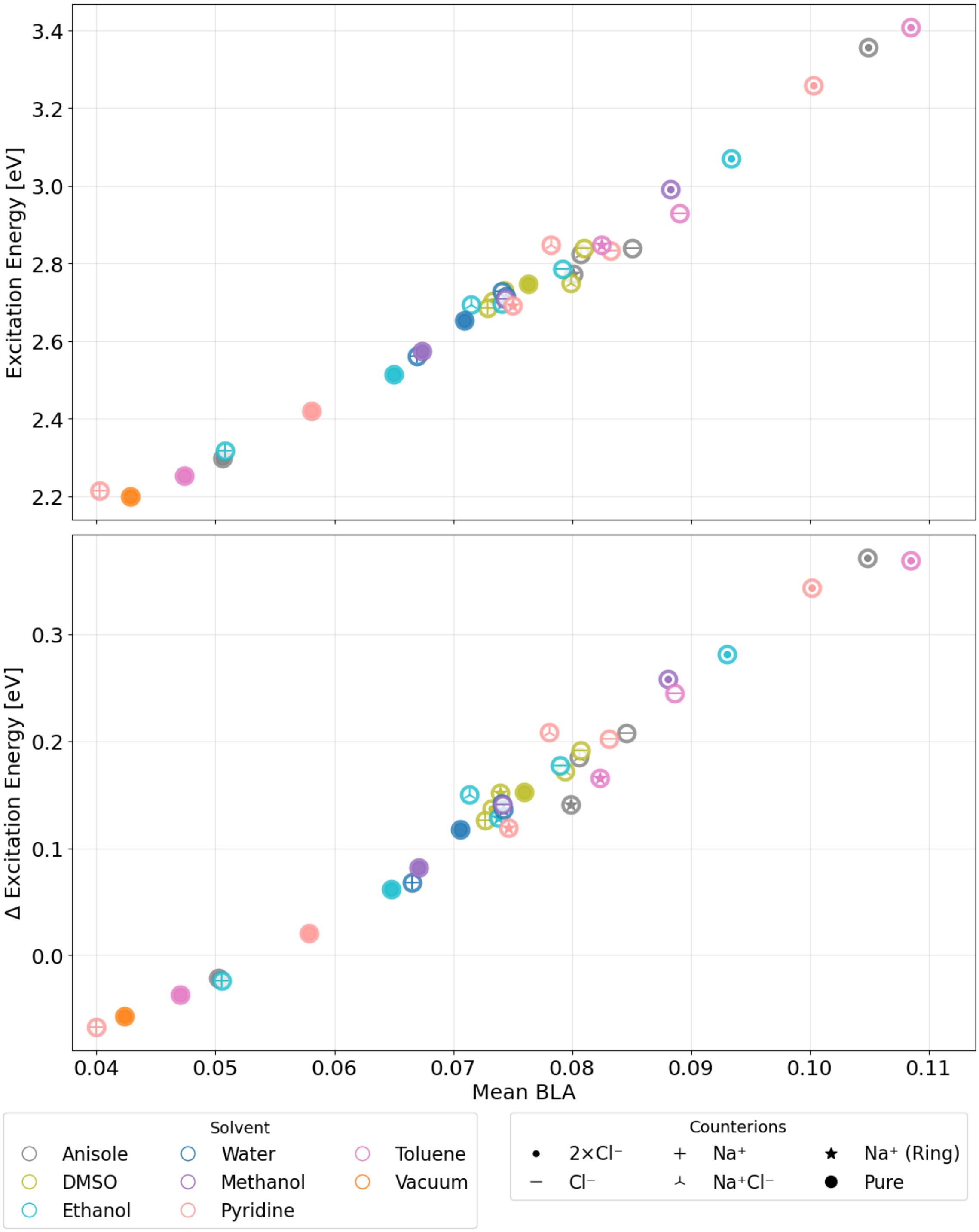


Figure 78.: Mean BLA vs. excitation energy: *all-trans* retinal (top: OM2/MRCI; bottom: Δ). Colors indicate solvents; symbols encode counterions. The points align closely along a straight line.

7.4. Property Network

This section first summarizes the learning setup for retinal+lysine QM cutouts embedded in solvent-like electrostatics, then reports training outcomes on curriculum-style datasets and finally applies the trained models to proteins. The graph network architecture, featurization, and losses are identical to Chapter VI – see also Methods II (Chapter III) and the integration chapter (Sec. 4.3); only the task mapping differs.

Learning setup and dataflow. Inputs are fixed-size QM coordinates and atom types for the retinal+lysine region, augmented with per-atom ESP sampled from the environment. Outputs are the $S_0 \rightarrow S_1$ excitation energy E and oscillator strength f. Two heads (separate regressors for E and f) are trained. Two labeling strategies are considered: direct regression to OM2/MRCI and Δ-learning that lifts LC–TD–DFTB to OM2/MRCI. For absorption, spectra are constructed from the predicted (E, f) distributions; forces and nonadiabatic couplings are out of scope.

7.4.1. Training

This subsection evaluates two labeling strategies – direct OM2/MRCI and Δ-learning with Δ = OM2 – DFTB – on three data scopes (CIS+TRANS, TRANS, CIS), all assessed on a fixed held-out panel. Results are presented in three parts: (i) curriculum experiments in which solvents (per isomer) and ions (in toluene) are added stepwise; (ii) a head-to-head comparison of OM2/MRCI vs. Δ on the full scopes, including parity/density plots; and (iii) the selected network capacities from the tuner (Table 35). This structure makes it explicit what is being compared (labeling strategy) and over which datasets (scope and curriculum).

Table 35.: Hyperparameter search space for the property network.

Hyperparameter	Search space
Hidden neurons / layer (per head)	20–100 (step 5)
Hidden layers (per head)	2–8 (step 1)
Learning rate	$\{10^{-3}, 5\cdot10^{-4}, 10^{-4}\}$
Regulariser	L2
Hyperband budget / trial (epochs)	20
Hyperband factor	2

Curriculum-style data aggregation. To quantify how additional environments improve generalization, stepwise training curricula are assembled. For solvents (per isomer): Tol → Tol+MeOH → Tol+MeOH+H_2O. For ions (combined isomers in toluene): no counterion → +Cl^- → +2Cl^- → +Na^+ (ring) → +NaCl. At each step, the property network is retrained and test R^2/MAE are recorded on a fixed held-out panel.

Results at a glance. Across the three solvent curricula in Table 37 and the ion curriculum in Table 36, test $\mathrm{MAE_E}$ sits in the 0.06–0.10 eV range, while R^2_E increases as environments are added (typically from 0.71–0.77 for Tol to 0.81–0.85 for Tol+MeOH+H_2O, and up to ~0.91–0.95 when ions are present in toluene). For oscillator strengths, MAE_f is 0.07–0.08 with R^2_f between 0.78 and 0.95. On the full-scope models without solvent filtering (Table 38), OM2/MRCI attains $R^2_\mathrm{E} = 0.90$–0.91 and $R^2_f = 0.94$–0.95 with $\mathrm{MAE_E} = 0.07$–0.08 eV and $\mathrm{MAE}_f = 0.06$–0.06. The Δ head yields lower $\mathrm{MAE_E}$ (0.04–0.04 eV) but lower R^2_E (0.85–0.87) and weaker calibration for f ($R^2_f = 0.71$–0.78, $\mathrm{MAE}_f = 0.04$–0.05). With the filtered training set (starred rows), OM2/MRCI improves to $R^2_\mathrm{E} = 0.93$ and $R^2_f = 0.95$, while the Δ energy head reaches $\mathrm{MAE_E} = 0.03$ eV.

OM2 vs. Δ. On the full CIS+TRANS dataset, the Δ model halves $\mathrm{MAE_E}$ relative to OM2/MRCI (0.04 vs. 0.08 eV) but attains a lower R^2_E (0.87 vs. 0.91), see Table 38 and Fig. 79. The parity plots reveal a mild regression-to-the-mean behavior for Δ (compressed spread along the identity), which improves MAE but penalizes R^2 once the target range widens. For oscillator strengths, OM2/MRCI is better calibrated ($R^2_f = 0.95$ vs. 0.76), even though Δ attains a slightly lower MAE_f (0.04 vs. 0.06). The same pattern holds for the TRANS-only and CIS-only scopes: Δ excels in $\mathrm{MAE_E}$ (0.04–0.04 eV) while OM2/MRCI retains higher R^2 for both E and f.

Effect of the solvent curricula. For **11-cis**, R^2_E increases from 0.71 (Tol) to 0.81 (Tol + MeOH + H_2O), whereas $\mathrm{MAE_E}$ increases from 0.069 to 0.098 eV; R^2_f decreases slightly with additional solvents (0.87 → 0.82). **All-trans** shows a similar trend: R^2_E peaks with Tol + MeOH (0.85) while $\mathrm{MAE_E}$ climbs to 0.091 eV at Tol + MeOH + H_2O. **Combined isomers** benefit most in R^2_E (0.77 → 0.85) with modest changes in f. Overall, adding environments improves generalization (higher R^2) at the cost of slightly larger MAE on the fixed panel – consistent with increased diversity and target variance.

Effect of ions in toluene. Introducing counterions strongly raises R^2 for both E and f (to $\gtrsim 0.91$ and $\gtrsim 0.93$, respectively), while $\mathrm{MAE_E}$ increases (0.06 → 0.09–0.10 eV; Table 36). The elevated R^2 indicates that the network captures the systematic electrostatic trend imposed by the ions, whereas the larger MAE reflects sharper local field variations and a wider energetic spread.

Model capacity and training stability. Architectures chosen by the tuner follow an intuitive pattern (Table 39): Δ models employ deeper/larger heads than OM2/MRCI, consistent with learning residual structure across diverse environments. Learning rates remain in a narrow stable band (10^{-4}–10^{-3}).

Filtered training (Spies, “Network 3”). When solvents that degraded generalization are excluded, performance improves markedly for OM2/MRCI (R^2_E=0.93, R^2_f=0.95) and yields the lowest MAE for the Δ energy head (0.03 eV), see the starred rows in Table 38. This mirrors the curation rationale from Diagnostics and is the setup used for the protein application.

Table 36.: cis+trans-Retinal: Ion curriculum in toluene.

Curriculum step	R^2_E	MAE_E [eV]	R^2_f	MAE_f
Tol (no ion)	0.774	0.06	0.877	0.07
+ Cl^-	0.935	0.07	0.948	0.07
+ 2 Cl^-	0.924	0.09	0.938	0.08
+ Na^+ (ring)	0.913	0.10	0.933	0.08
+ NaCl	0.916	0.09	0.927	0.08

Table 37.: Solvent curriculum. Test metrics on a fixed held-out panel.

(a) 11-cis-Retinal

Curriculum step	R^2_E	MAE_E [eV]	R^2_f	MAE_f
Tol	0.713	0.07	0.870	0.07
Tol + MeOH	0.781	0.10	0.833	0.08
Tol + MeOH + H_2O	0.807	0.10	0.816	0.08

(b) all-trans-Retinal

Curriculum step	R^2_E	MAE_E [eV]	R^2_f	MAE_f
Tol	0.738	0.07	0.827	0.08
Tol + MeOH	0.850	0.08	0.805	0.08
Tol + MeOH + H_2O	0.833	0.09	0.784	0.08

(c) cis+trans-Retinal

Curriculum step	R^2_E	MAE_E [eV]	R^2_f	MAE_f
Tol	0.774	0.06	0.877	0.07
Tol + MeOH	0.838	0.08	0.865	0.08
Tol + MeOH + H_2O	0.851	0.08	0.871	0.08

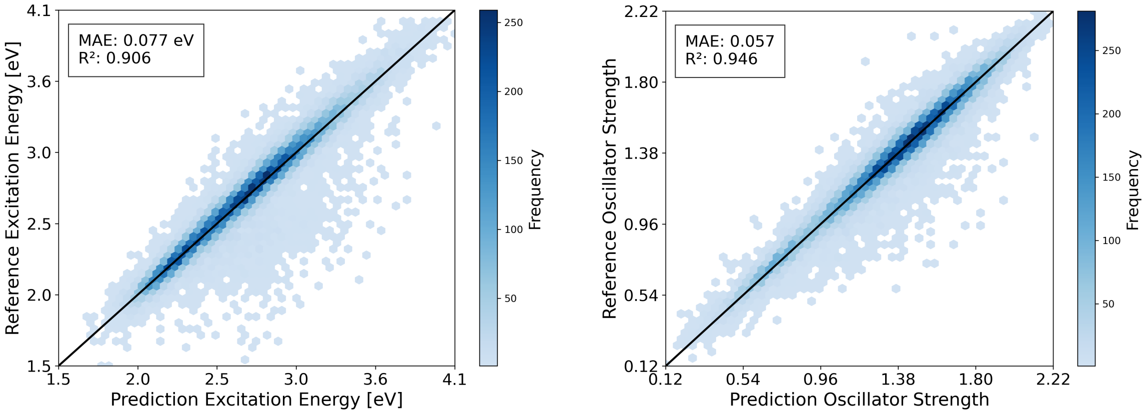
Table 38.: Property network: Test-set performance across data scopes (all solvents). **Reference method:** OM2/MRCI and Δ = OM2 − DFTB. f is dimensionless. Rows marked with * use the filtered training set in which selected solvents were excluded to boost performance (cf. Spies, *Networks 3*).

Reference method	Data scope	R^2_E	MAE_E [eV]	R^2_f	MAE_f
OM2/MRCI	CIS + TRANS	0.91	0.08	0.95	0.06
Δ	CIS + TRANS	0.87	0.04	0.76	0.04
OM2/MRCI	TRANS	0.91	0.07	0.89	0.06
Δ	TRANS	0.85	0.04	0.71	0.05
OM2/MRCI	CIS	0.90	0.08	0.94	0.06
Δ	CIS	0.87	0.04	0.78	0.04
OM2/MRCI*	CIS + TRANS	**0.93**	**0.07**	**0.95**	**0.06**
Δ^*	CIS + TRANS	**0.90**	**0.03**	**0.78**	**0.04**

Table 39.: Property model: architectures and key hyperparameters across data scopes (all solvents). **Reference method:** OM2/MRCI and Δ = OM2 – DFTB. Columns list depths before neuron counts for E and f.

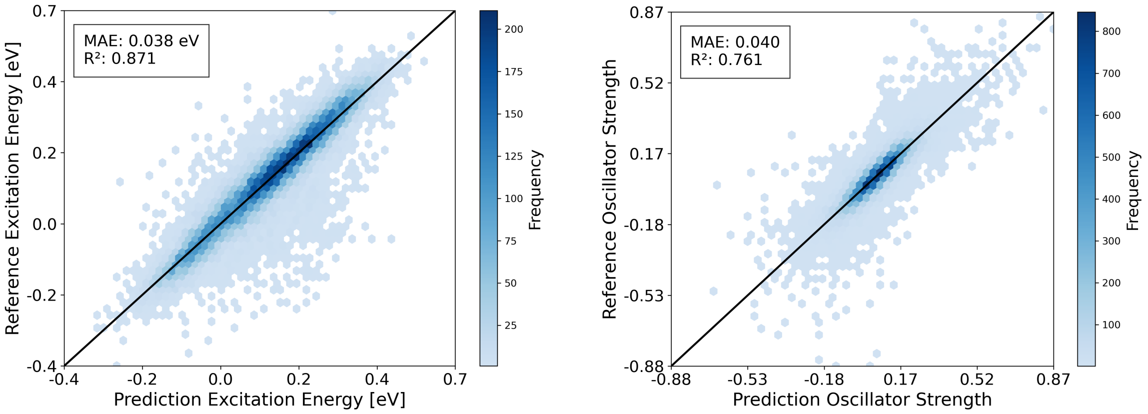
Reference method	Data scope	depth$_\mathrm{E}$	n_E	depth$_f$	n_f	lr
OM2/MRCI	CIS + TRANS	3	30	4	50	1e−4
Δ	CIS + TRANS	4	70	8	40	1e−3
OM2/MRCI	TRANS	3	20	8	30	5e−4
Δ	TRANS	4	100	5	90	1e−4
OM2/MRCI	CIS	2	20	8	40	1e−4
Δ	CIS	7	60	6	40	1e−4

Figure 79.: Property Model: OM2/MRCI and Δ. CIS + TRANS – all solvents. Parity plots for models trained on the full dataset. Left column: excitation energy E; right column: oscillator strength f (dimensionless). Top row: OM2/MRCI; bottom row: Δ-learning (Δ = OM2 – DFTB). Insets report MAE and R^2.



(a) OM2/MRCI – Excitation Energy

(b) OM2/MRCI – Oscillator Strength



(c) Δ – Excitation Energy

(d) Δ – Oscillator Strength

7.4.2. Protein Application

The trained networks are applied to six proteins (ChR2, bR, Chrimson and its S169A mutant, rhodopsin and E181Q). Only geometries and ESP of the retinal+lysine QM region are available; no excited-state labels exist for these trajectories. Accordingly, *relative* shifts with bR as reference are reported and compared qualitatively to experimental $\lambda_{\max}$ (Table 40). Both model variants recover the expected ordering consistent with electrostatic color tuning: ChR2 is blue-shifted relative to bR; Chrimson (and its S169A mutant) form the red-tuned set; rhodopsin is blue-shifted relative to bR. For the key pair bR vs. ChR2, both models predict a negative shift (blue) relative to bR, in line with experiment (see Table 40).

Table 40.: Retinal: protein absorption comparison using bR as reference (Δ vs bR in eV). Experimental $\lambda_{\max}$ are indicative. Predicted mean excitation energies E (eV) are shown for an OM2/MRCI network and a Δ-model. *Note:* Both ML columns use the solvent-filtered training setup ("Network 3") from Spies, where selected solvents were excluded to improve performance; the results are reproduced here in collaboration with K. Spies.

Protein	Experiment		OM2/MRCI NN		Δ-model	
	$\lambda_{\max}$ / nm	eV	E / eV	Δ vs bR / eV	E / eV	Δ vs bR / eV
Chrimson S169A	608	2.04	2.56	−0.14	2.57	−0.13
Chrimson	590	2.10	2.66	−0.04	2.68	−0.02
bR	570	2.18	2.70	0.00	2.70	0.00
Rhodopsin E181Q	508	2.44	2.77	+0.07	2.71	+0.01
Rhodopsin	500	2.48	2.82	+0.12	2.75	+0.05
ChR2	470	2.62	3.01	+0.31	3.09	+0.39

7.5. Discussion

The training diagnostics indicate that physically motivated descriptors capture the main levers that tune retinal absorption. Mean E increases approximately linearly with mean BLA, reproducing the classic trend across rhodopsins. ESP distributions are reported for completeness to document the range sampled in solution. Nonpolar, high-field scenarios can distort the chromophore and degrade generalization – performance-based pruning of such cases improves stability, consistent with the solvent-filtered "Network 3" strategy.

Across curricula, adding MeOH and subsequently H_2O to toluene generally improves R^2_{E} for the combined isomer set while leaving R^2_f high and stable; for single-isomer curricula the benefits depend on the specific holdout and can trade a small MAE increase for better R^2. The ion curriculum reveals that strong local fields increase distributional breadth and can raise MAE for E, even though R^2 remains high – an expected outcome when the label variance increases with field strength.

Direct OM2/MRCI models are consistently strong for f and attain $R_f^2 \approx 0.94$ on the full combined set, whereas Δ-learning achieves notably lower MAE in E but weaker R_f^2. This pattern is consistent with the view that semiempirical excited-state methods encode oscillator-strength trends with fewer systematic biases than energies, while Δ-corrections are effective primarily on E. Solvent filtering (Spies "Network 3") increases combined-set R_E^2 to ~0.93 (OM2/MRCI) and ~0.90 (Δ), indicating that targeted curation can remove distributional tails that otherwise harm generalization.

On proteins, the expected color ordering is recovered for both models when shifts are reported relative to bR, which reduces sensitivity to absolute offsets. Nevertheless, quantitative shifts remain imperfect: residual systematic errors in E transfer and differences between solution-like and protein-embedded electrostatics (e.g., polarization, specific H-bond networks) are not fully captured by the present semiempirical labels. Increasing label fidelity (e.g., higher-level QM/MM for selected snapshots) or incorporating explicit polarization-aware descriptors would likely improve the absolute shifts but at substantial computational cost. Uncertainty quantification and domain-adaptation schemes could further qualify predictions in outlying electrostatic regimes.

7.6. Summary

This chapter reuses a unified property-network architecture to learn retinal absorption from QM/MM snapshots of a fixed retinal+lysine QM region embedded in diverse electrostatic environments. Clean targets were ensured via PBC-consistent electrostatics. Diagnostics show physically consistent trends (BLA–E correlation, ESP coverage, torsional sampling) and motivate curation of distorted, high-field cases. Curriculum experiments quantify how solvents and ions affect generalization; direct OM2/MRCI labeling is advantageous for f, whereas Δ-learning reduces MAE in E. On proteins, qualitative color ordering relative to bR is reproduced using solvent-filtered models derived from Spies' "Network 3". Remaining discrepancies in absolute shifts point to the need for higher-fidelity excited-state labels and/or polarization-aware features. Overall, the results establish a robust proof of principle for ML surrogates of retinal absorption and delineate a clear path toward quantitative protein spectrochemistry with improved labeling and targeted curation.

Part VIII

Sampling Strategies for ML Potentials

8.1. Overview

This chapter presents and extends a set of sampling methods for training machine-learning potentials and for ML/MM simulations. I apply five strategies – free molecular dynamics, normal-mode perturbations, PCA-biased metadynamics, RMSD – biased metadynamics, and uncertainty-guided sampling – to four representative systems: *FR0*, *Retinal*, *4-Aminophthalimide* (4AP), and *ACDAN*. The objective is to broaden our sampling toolbox beyond what we used in earlier projects and to assess how these methods capture challenging motions – large-amplitude torsions and cis-trans isomerisation – so that simulations of demanding molecules become more informative and reliable.

Author Contributions. As part of the fluorescence project (Chapter VI), the initial FR0 study underpinning this chapter was carried out by Dominik Hachenthal under my supervision.[243] Pascal Schmidt developed a complementary uncertainty-guided sampling strategy for FR0, which I co-advised.[244] Initial unbiased MD simulations of *4-Aminophthalimide* (4AP) and *ACDAN* were first performed by Anna Ulrich in her Master's thesis, which I supervised.[245] The FR0 results reported here are adopted from these works. Building on these contributions, this chapter consolidates the approaches and extends them to *Retinal*, 4AP, and *ACDAN* under a unified protocol; for these molecules, Pascal Schmidt provided simulation support using methods implemented by Hachenthal.

8.2. Introduction

Sampling governs which configurations a model will ever see and, hence, which regions of configurational space can be learned and reliably simulated. In this chapter we compare general-purpose strategies – free simulation, normal-mode analysis (NMA), PCA-biased metadynamics, RMSD-biased metadynamics, and uncertainty-guided sampling – purely from a *sampling* perspective. In practice, two goals must be balanced: (i) *coverage* of relevant configurational space, including rare or off-equilibrium motifs; and (ii) *force reliability in dynamics*, i.e., producing data that yield predictors stable and accurate enough to be embedded back into the QM/MM loop. Our prior studies illustrate the spectrum of approaches: the phosphorus project used hand-crafted CVs with metadynamics; the FR0 work relied mainly on free QM/MM; Retinal highlights systems where large, conjugated-dihedral rearrangements make CV choice important. Here we place these philosophies side by side and examine their behavior across molecules with distinct dominant degrees of freedom.

Systems. Four molecules are considered to span common soft modes relevant to sampling (Figure 80): *FR0* (7-(diethylamino)-9,9-dimethyl-9H-fluorene-2-carbaldehyde), characterized by flexible aldehyde and dialkylamino torsions; *Retinal*, a polyene chromophore undergoing 11-*cis* ↔ all-*trans* isomerization dominated by coupled C=C/C−C dihedrals; *4-Aminophthalimide (4AP)*, in which ring planarity and the imide-adjacent torsion define the soft coordinates; and *ACDAN* (6-acetyl-2-(dimethylamino)naphthalene), where the acetyl twist and dimethylamino planarity/pyramidalization govern the principal low-frequency motions.

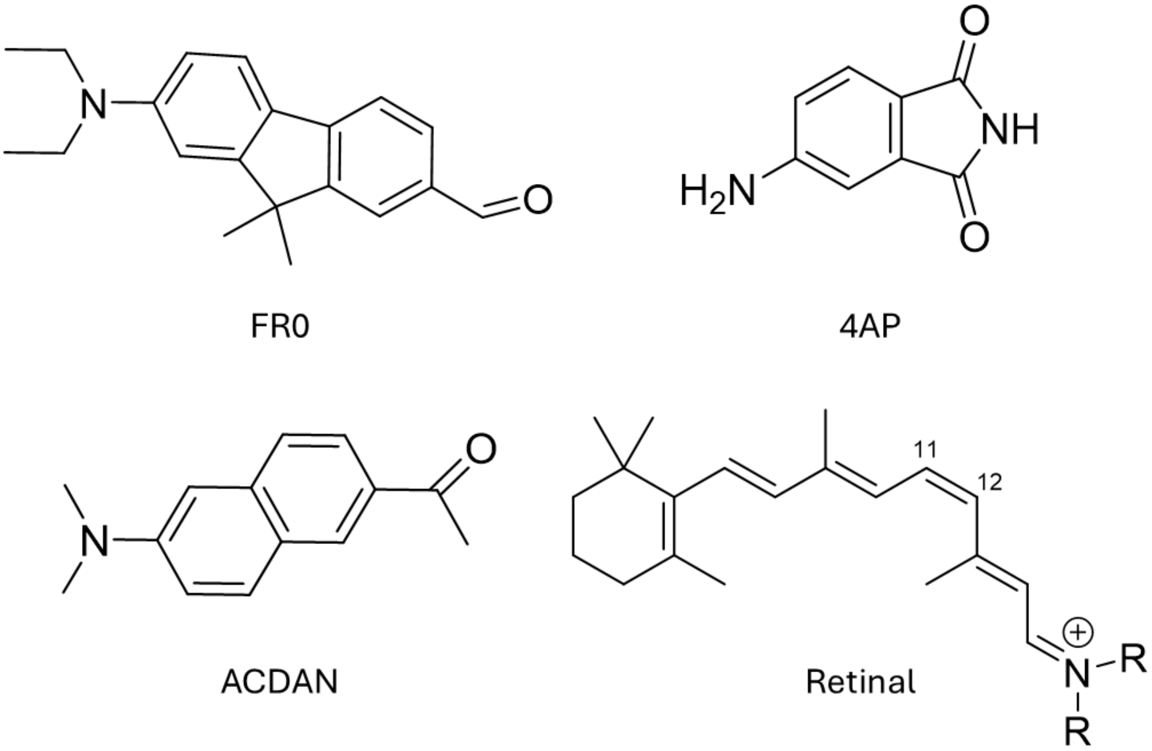


Figure 80.: Chemical structures of the four systems considered in this chapter: FR0, Retinal, ACDAN, and 4AP.

8.3. Case Studies

To keep the comparison meaningful, core settings are harmonised across methods and molecules: identical solvent models; the same timestep, write-out frequency, and thermostat/barostat; a common QM/MM level for labelling energies and forces; a shared network architecture and optimiser schedule; and identical train/validation splits by trajectory index. Method-specific knobs (e.g., metadynamics hill width/height and walker count; uncertainty thresholds; normal-mode amplitudes) are varied only within narrow, molecule-agnostic ranges and are listed in the Appendix.

Parameter tables. Complete parameter listings for the sampling setups are provided in the Appendix: enhanced sampling (uncertainty-guided workflow; Table A.27), uncertainty quantification and calibration (Table A.28), metadynamics configuration (RMSD- and PCA-biased; Table A.29), solvent models per system (Table A.30), MM force-field provenance for the classical region (Table A.18), generic GROMACS QM/MM ground-state MD settings (Table A.31), and DFTB2 settings used for re-labelling (ground state; Table A.32). Unless otherwise noted in the case-specific text, these settings apply across molecules.

Scope. All data in this chapter stem from *ground-state* (S_0) simulations. Ground-state energies are reported as diagnostics; *forces* are the primary target because they govern MD stability and provide the supervised signal for ML force fields.

Solvents. Multiple solvents were simulated per molecule; for the case studies one solvent per system is fixed to avoid conflating solvent and method effects: *FR0*, *4AP*, and *ACDAN* in DMSO; *Retinal* in water.

MACE cross-evaluation. For transfer comparisons between sampling strategies, MACE models are trained on the *QM region only* (without explicit environment) to isolate configurational transfer. The environment is present during QM/MM labelling and MD but not in the MACE training input.

8.3.1. Case: FR0

This section compares five sampling strategies for *FR0* strictly on configurational coverage. All diagnostics and cross-evaluations use the MACE architecture (energy and force), trained per sampling strategy and tested across all datasets. For the flexible diethylamino tail we analyse the coupled dihedrals (D_1, D_2); for the formyl group we consider the single dihedral D_0 (Fig. 81).

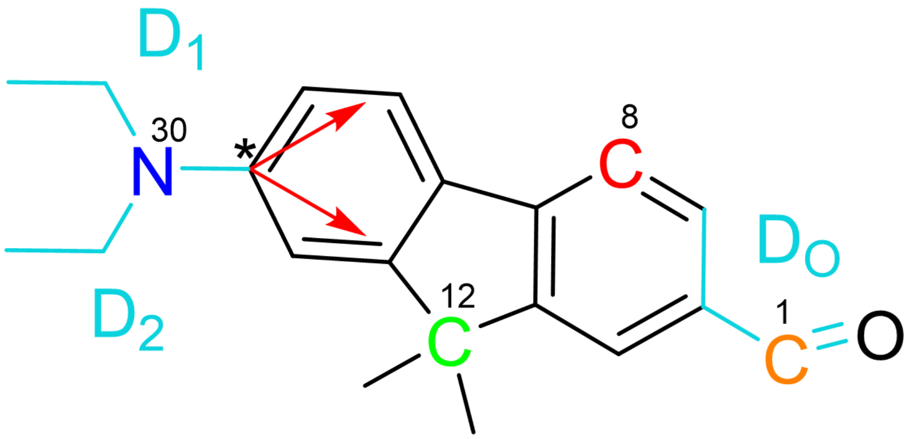
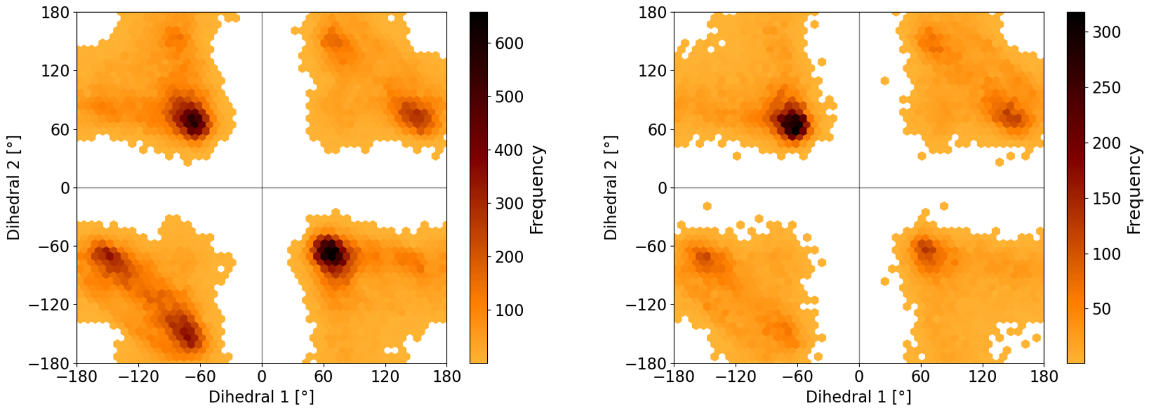
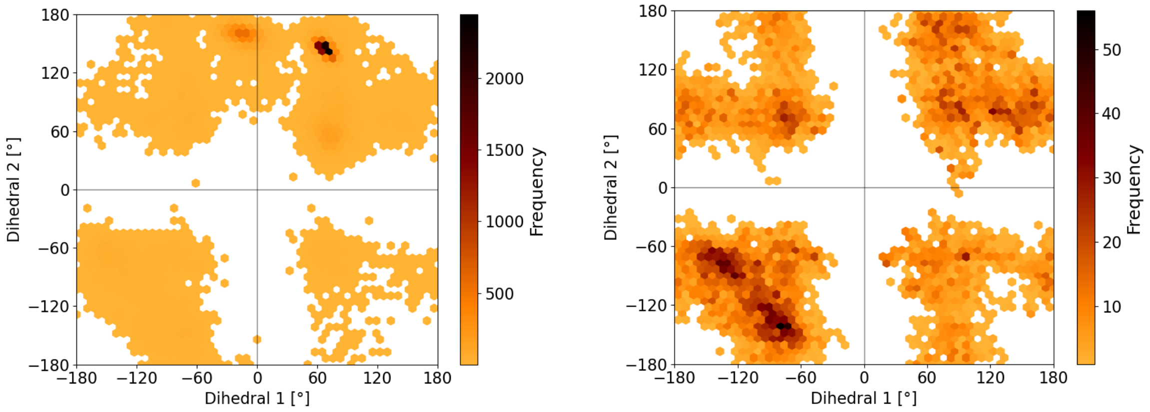


Figure 81.: FR0 with dihedrals and planarity markers. D_1 and D_2 are the two *dialkylamino* ethyl-arm torsions, and D_0 is the *formyl* torsion about the aryl–CH=O linkage. The starred aryl carbon (C^*) is the reference atom; the two red arrows indicate second-nearest ring-carbon directions from C^* that span the reference plane (see Implementation Notes on planarity). Out-of-plane angles are evaluated for the four *numbered* atoms N_{30} (blue), C_8 (green), C_1 (red), C_{12} (orange) relative to this plane.

Coverage in (D_1, D_2). The 2D maps in Fig. 82 show complementary behaviors: *Free MD* yields compact, contiguous basins; *PCA-biased* sampling broadens these basins and bridges nearby regions; *RMSD-biased* sampling reaches the widest spread, populating remote pockets; *Uncertainty-guided* sampling fills fringe regions while remaining sparser in the cores. The separate *NMA* panel (Fig. 83) produces sparse island clusters around local mode displacements with little overlap to the other strategies.



(a) Free simulation **(b)** PCA-biased metadynamics



(c) RMSD-biased metadynamics **(d)** Uncertainty-guided sampling

Figure 82.: Two-dimensional dihedral occupancy for the diethylamino tail of *FR0* (hex-binned frequencies). Top: Free MD and PCA-biased metadynamics. Bottom: RMSD-biased metadynamics and uncertainty-guided sampling.

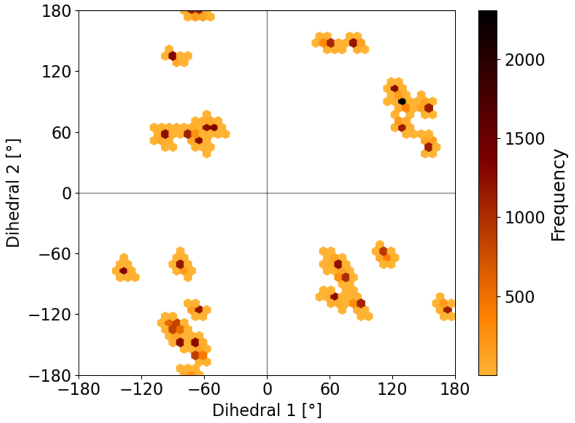


Figure 83.: Normal-mode *off-equilibrium generator* (NMA): 2D dihedral density for the diethylamino tail. Mode-following perturbations generate sparse islands around local displacements.

Formyl torsion D_0**.** The 1D distributions (Fig. 84) mirror the 2D picture: free and PCA concentrate near the planar conformer; RMSD spreads into secondary lobes; uncertainty extends the tails; NMA gives sharp, isolated peaks characteristic of small-amplitude normal-mode moves.

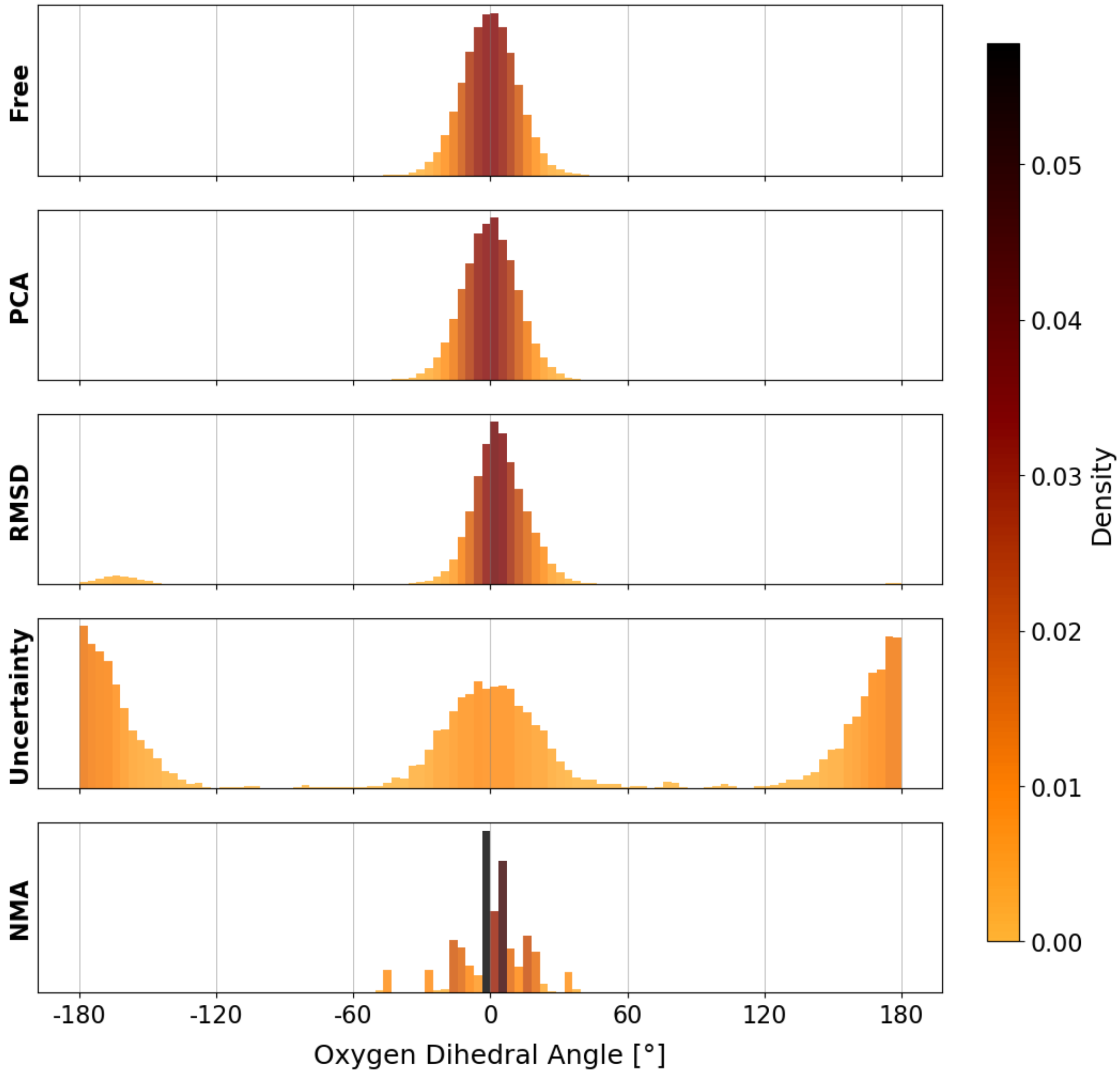


Figure 84.: FR0: One-dimensional distribution of the formyl oxygen dihedral across sampling methods (density-normalized histograms).

Planarity at selected sites. Using the planarity definition from Sec. 3.7.7 (reference plane at the starred aryl carbon C^*; Fig. 81), we monitor out-of-plane angles α_X for four diagnostic atoms: the dialkylamino $\mathbf{N}_{30}$, the fluorene **bridgehead** carbon $\mathbf{C}_{12}$, a benzenoid carbon $\mathbf{C}_8$, and the **formyl** carbon $\mathbf{C}_1$. Figure 85 summarizes the distributions by sampling method: free and PCA keep all sites close to the plane (small medians, narrow spreads), uncertainty as well as NMA is similar with slightly broader tails, RMSD occasionally pushes benzenoid/formyl carbons further out of plane.

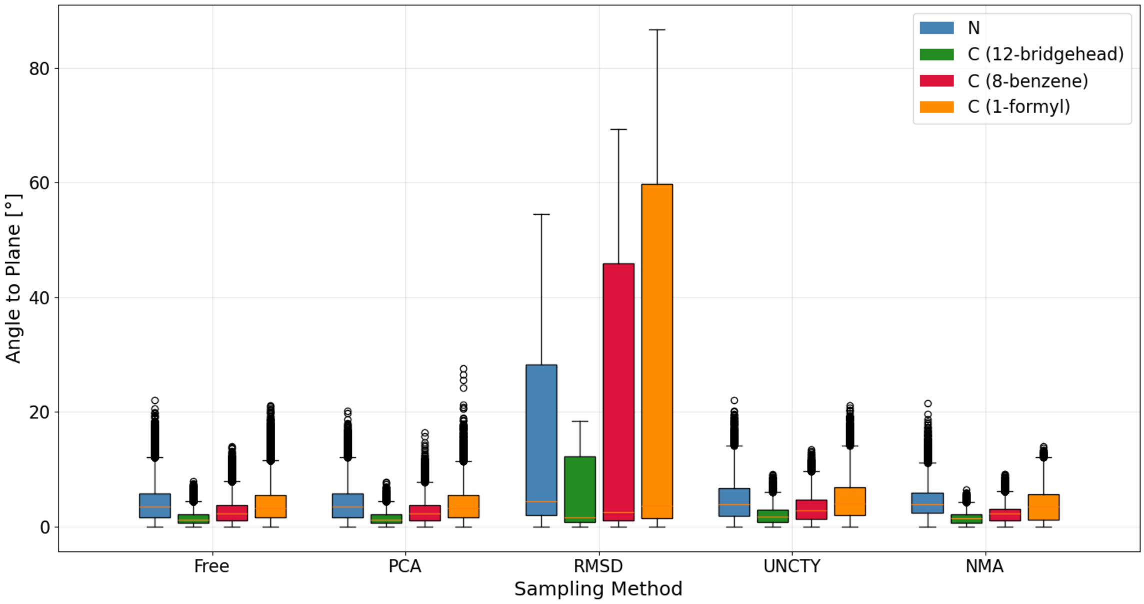
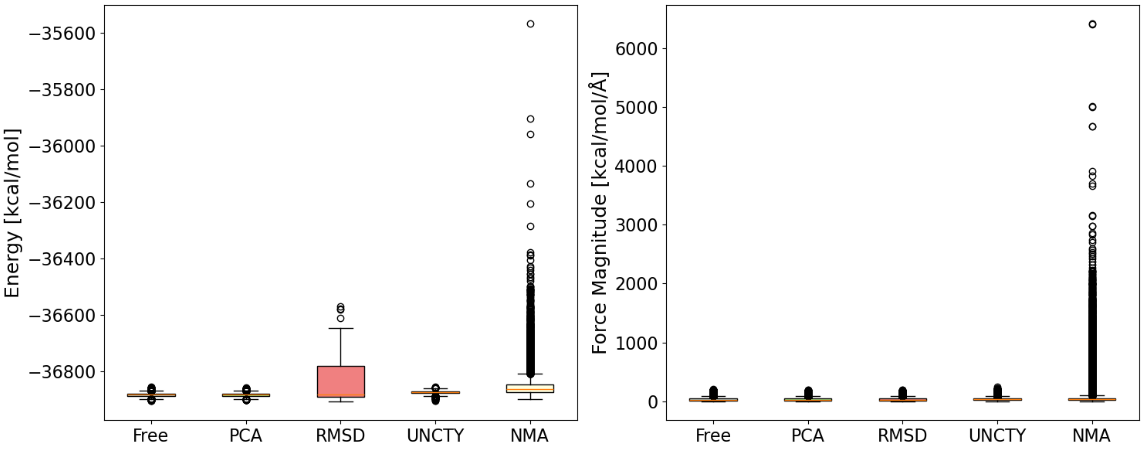
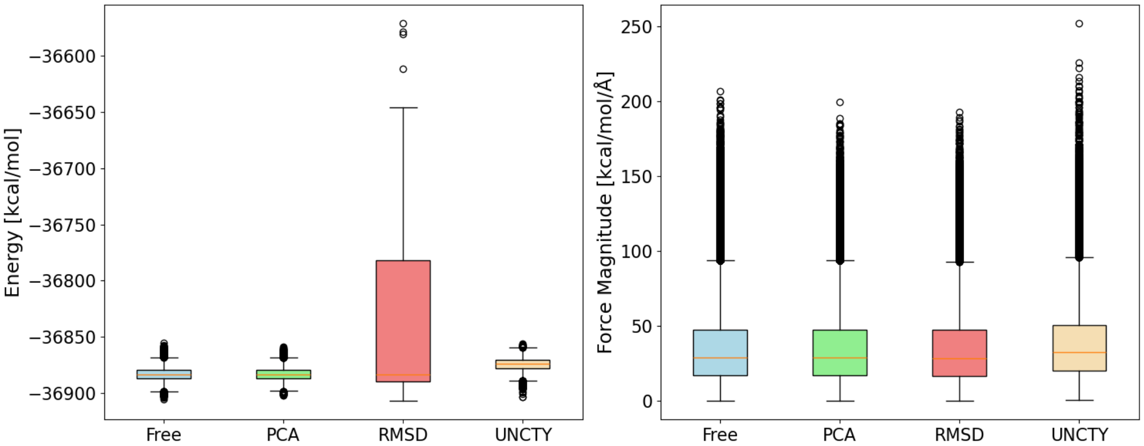


Figure 85.: FR0 planarity by method. Boxplots of the out-of-plane angle α_X (degrees) relative to the reference plane at C^* for four atoms: **N** (dialkylamino), $\mathbf{C}_{12}$ (bridgehead), $\mathbf{C}_8$ (benzenoid), and $\mathbf{C}_1$ (formyl). Methods on the x-axis. The planarity construction in Sec. 3.7.7; the reference atom and IDs are shown in Fig. 81.

Energy and force distributions. Figure 86 summarizes per-frame *energies* and *force magnitudes* for each strategy. Free and PCA concentrate in narrow energy bands with moderate forces; RMSD widens the spread as intended; uncertainty sampling extends the support while keeping the cores comparable to free/PCA. Including **NMA** (top panel) reveals heavy-tailed force outliers and occasional large energy excursions; removing NMA (bottom panel) improves comparability of the remaining strategies. These diagnostics validate that NMA generates atypical, high-gradient configurations that are of limited value for model training and transfer on *FR0*.



(a) Methods (with NMA): Free, PCA, RMSD, Uncertainty, NMA.



(b) Methods (without NMA): Free, PCA, RMSD, Uncertainty.

Figure 86.: FR0: boxplots of per-frame *energy* (left) and *force magnitude* (right) for each sampling strategy. Units: energies in kcal mol^{-1}; forces in kcal mol^{-1} $\mathrm{Å}^{-1}$. Including NMA (top) highlights its heavy-tailed, high-force outliers; zooming in without NMA (bottom) makes the remaining strategies comparable. This motivates excluding NMA from the subsequent molecule-wise analyses in this chapter.

MACE cross-evaluation. The MAE grids in Table 41 (single **MACE** architecture; energy and force) show pronounced diagonal minima (in-domain) and the expected transfer structure: free and PCA models generalize well to each other, and RMSD remains competitive on those tests. Averaged over the *non-NMA* test sets, however, the **UNCTY-trained** model attains the lowest mean error for both targets, making it the best overall generalizer on FR0. By contrast, NMA behaves as an outlier with poor mutual transfer and is therefore omitted from the remaining systems (4-AP, ACDAN and Retinal).

		Model [Energy]				
		Free	**PCA**	**RMSD**	**UNCTY**	**NMA**
Test Dataset	**Free**	0.04	0.05	0.06	0.05	0.13
	PCA	0.04	0.05	0.06	0.05	0.13
	RMSD	0.03	0.05	0.04	0.04	0.09
	UNCTY	0.16	0.18	0.18	0.06	0.28
	NMA	$> 10^3$	$> 10^2$	$> 10^1$	$> 10^5$	0.09

(a) Energy MAE $[\mathrm{kcal\,mol^{-1}}]$.

		Model [Force]				
		Free	**PCA**	**RMSD**	**UNCTY**	**NMA**
Test Dataset	**Free**	0.08	0.08	0.10	0.08	0.21
	PCA	0.08	0.09	0.10	0.08	0.21
	RMSD	0.07	0.08	0.08	0.08	0.21
	UNCTY	0.26	0.28	0.24	0.11	0.44
	NMA	$> 10^3$	$> 10^2$	$> 10^1$	$> 10^5$	0.09

(b) Force MAE $[\mathrm{kcal\,mol^{-1}\,\AA^{-1}}]$.

Table 41.: Cross-evaluation on **FR0** (without environment) with a single **MACE** architecture. Each column is a MACE model trained on data from the indicated sampling strategy; each row reports the MAE on the corresponding *test* dataset. The grids capture in-domain performance (diagonal) and transfer across strategies (off-diagonal). "UNCTY" denotes uncertainty-guided sampling. Cross-evaluation values reproduced from Pascal Schmidt's FR0 MACE analysis.

8.3.2. Case: 4AP

Compared to *FR0*, whose soft coordinates are large-amplitude torsions, *4-aminophthalimide* (4AP) is essentially planar. Accordingly, sampling strategies are compared via *planarity* rather than torsional coverage.

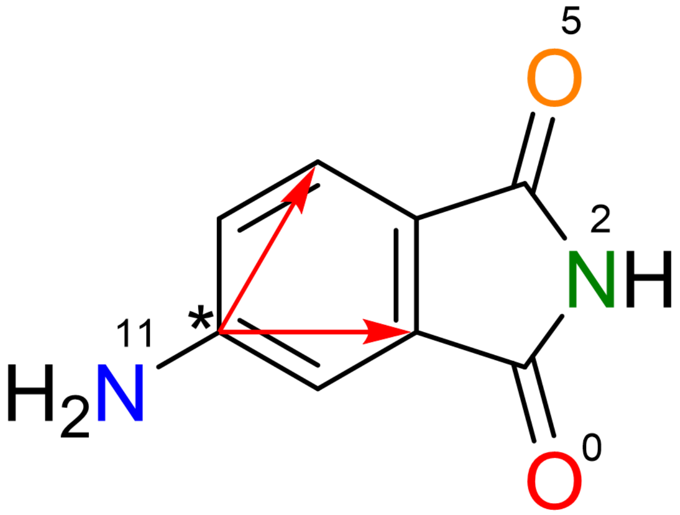


Figure 87.: 4AP with the two red arrows indicating the bond directions that span the reference plane at the highlighted carbon. Monitored atoms: aniline N_{11} (blue), imide N_2 (green), and the two carbonyl O_0 and O_5 atoms (red and orange). Angles are defined as out-of-plane deviations relative to this plane.

Planarity metric and setup. Planarity is evaluated exactly as defined in the Methods II chapter (Sec. 3.7.7): a reference plane on the benzenoid core and per-atom out-of-plane angles α_X for the four highlighted atoms in Fig. 87. We compare *Free*, *PCA-biased*, *RMSD-biased*, and *Uncertainty-guided* sampling; *NMA* is omitted here due to poor performance on *FR0* (cf. Sec. 8.3.1).

Planarity and selected sites. Across all strategies, 4AP remains close to planar, as expected for this π-conjugated system. Figure 88 shows narrow angle distributions for the aniline N_{11}, imide N_2, and both carbonyl oxygens O_0/O_5. *Free* and *PCA*-biased sampling yield the tightest interquartile ranges with small medians at all probe sites. *RMSD*-biased sampling increases spread and produces occasional high-angle outliers – useful for stress-testing the model on less planar motifs. *Uncertainty*-guided sampling sits between Free/PCA and RMSD: coverage extends beyond the cores while keeping the bulk of frames near the reference plane. Carbonyl oxygens show slightly broader tails than nitrogens, consistent with transient tightening at the imide/amide linkage.

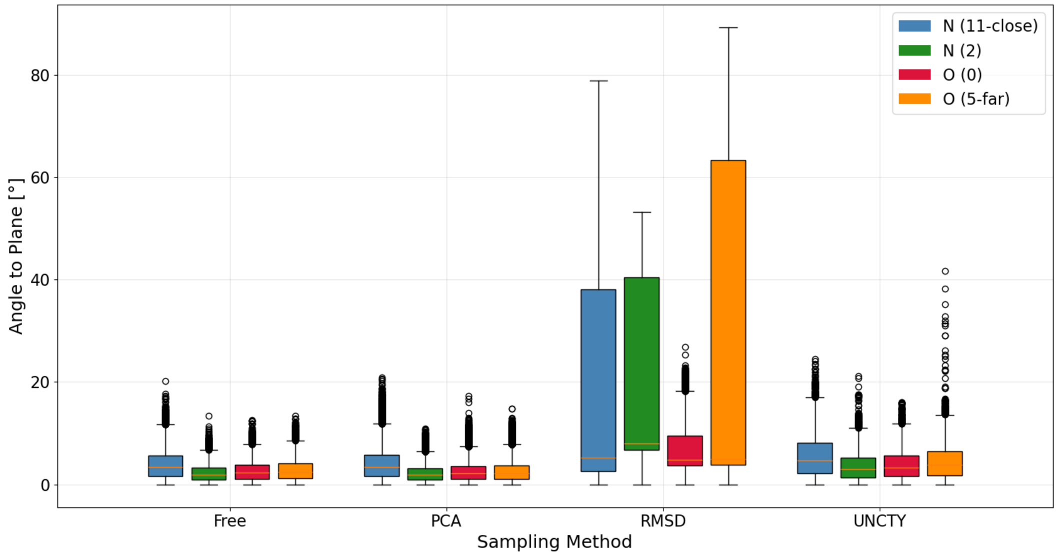


Figure 88.: 4AP: boxplots of out-of-plane angles α_X (degrees; 0° is in-plane) for the aniline $\mathbf{N}_{11}$, imide $\mathbf{N}_2$, and the two carbonyl $\mathbf{O}_0$ and $\mathbf{O}_5$ atoms across sampling methods (Free, PCA, RMSD, UNCTY). The planarity construction in Sec. 3.7.7; the reference atom and IDs are shown in Fig. 87.

Energy and force distributions. Figure 89 summarises per-frame energies and force magnitudes. *Free* and *PCA* concentrate in narrow energy bands with moderate forces, matching their near-planar geometry. *RMSD* broadens both distributions, as expected from deliberately distorted configurations. *Uncertainty* extends support relative to Free/PCA but avoids the heaviest tails, suggesting it adds informative diversity without flooding training with extreme, high-gradient frames. These diagnostics align with the planarity analysis: methods that push the molecule further from the aromatic plane also populate higher-energy, higher-force regions.

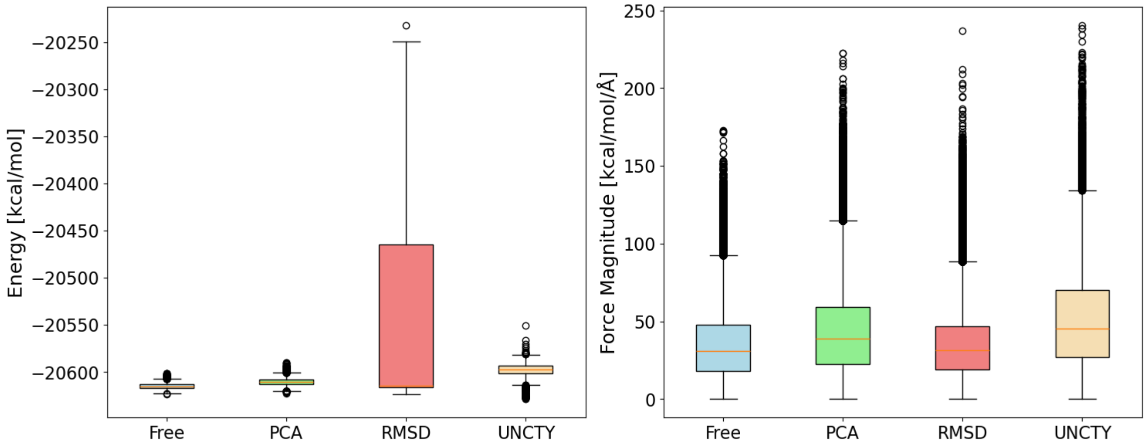


Figure 89.: 4-Aminophthalimide (4AP): boxplots of per-frame *energy* (left, kcal mol^{-1}) and *force magnitude* (right, kcal mol^{-1} Å^{-1}) for the available sampling strategies. Each box shows the interquartile range with median; whiskers span 1.5 IQR; points mark outliers. As in the FR0 plots, the panel summarises how each sampling method populates energy/force regions.

MACE cross-evaluation. The MAE grids for *4AP* (Table 42) show clear diagonal minima. For *forces*, *Free* and *PCA* transfer well onto each other (Free→PCA / PCA→Free = 0.09 / 0.09 kcal mol^{-1} Å^{-1}), yet both perform poorly on *RMSD*-biased tests ($\geq$ 7 kcal mol^{-1} Å^{-1}), consistent with deliberately off-basin geometries. The *UNCTY*-trained model is best in-domain (0.15 kcal mol^{-1} Å^{-1}) and remains moderate on Free/PCA tests (0.07–0.08 kcal mol^{-1} Å^{-1}) but not on RMSD (4.80 kcal mol^{-1} Å^{-1}). For *energies*, asymmetry is stronger: the *RMSD*-trained model transfers back to Free/PCA with small errors (0.02 / 0.02 kcal mol^{-1}), whereas models trained on Free, PCA, or UNCTY fail on RMSD tests (17–22 kcal mol^{-1}).

		Model [Energy]			
		Free	**PCA**	**RMSD**	**UNCTY**
Test Dataset	**Free**	0.01	0.02	0.02	0.33
	PCA	0.01	0.02	0.02	0.33
	RMSD	17.25	20.69	0.03	22.23
	UNCTY	0.31	0.30	0.33	0.04

(a) Energy MAE [kcal mol^{-1}].

		Model [Force]			
		Free	**PCA**	**RMSD**	**UNCTY**
Test Dataset	**Free**	0.06	0.09	0.11	0.07
	PCA	0.08	0.05	0.13	0.08
	RMSD	7.69	7.20	0.13	4.80
	UNCTY	0.27	0.29	0.33	0.15

(b) Force MAE [kcal mol^{-1} Å^{-1}].

Table 42.: Cross-evaluation on **4AP** (QM region only) with a single **MACE** architecture. Each column is a MACE model trained on data from the indicated sampling strategy; each row reports the MAE on the corresponding *test* dataset. The grids capture in-domain performance (diagonal) and transfer across strategies (off-diagonal).

8.3.3. Case: ACDAN

Given ACDAN's kinship to FR0, the analysis focuses on planarity and on the acetyl/formyl dihedral. Three sampling strategies are considered: *Free* MD, *PCA*-biased metadynamics, and *RMSD*-biased metadynamics. NMA is excluded based on the FR0 results; *uncertainty*-guided sampling could not be run for ACDAN due to technical issues.

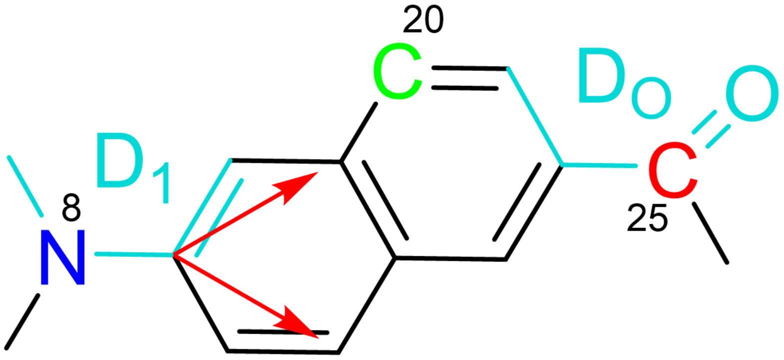


Figure 90.: **ACDAN** with the two analysed dihedrals: D_1 (dialkylamino twist about the N-aryl bond) and D_O (acetyl/formyl dihedral along the aryl-CO linkage). The two red arrows indicate the anchor vectors that span the aromatic reference plane (planarity metric defined in Sec. 3.7.7). Planarity is monitored for three atoms: the dimethylamino nitrogen N_8 (blue), a ring carbon C_{20} (green), and the carbonyl carbon C_{25} (red).

Planarity. Free MD keeps all three probes close to the plane – small medians, tight IQRs – (Fig. 91). *PCA*-biased metadynamics behaves very similarly to Free but with slightly tighter interquartile ranges, i.e., excursions are steered along soft principal modes. *RMSD*-biased sampling broadens the distributions substantially and creates pronounced tails, reflecting deliberate off-basin geometries.

Energy and force distributions. Per-frame energies and force magnitudes (Fig. 92) reflect the structural picture from the planarity analysis. *Free* MD occupies a narrow energy band with moderate forces. *RMSD*-biased sampling broadens both distributions and generates heavy tails, with a noticeable upward shift of the median force magnitude. This behaviour is consistent with the larger out-of-plane excursions at N_8 and C_{25} (Fig. 91), i.e., frames driven further from the aromatic reference plane also populate higher-energy, high-gradient regions.

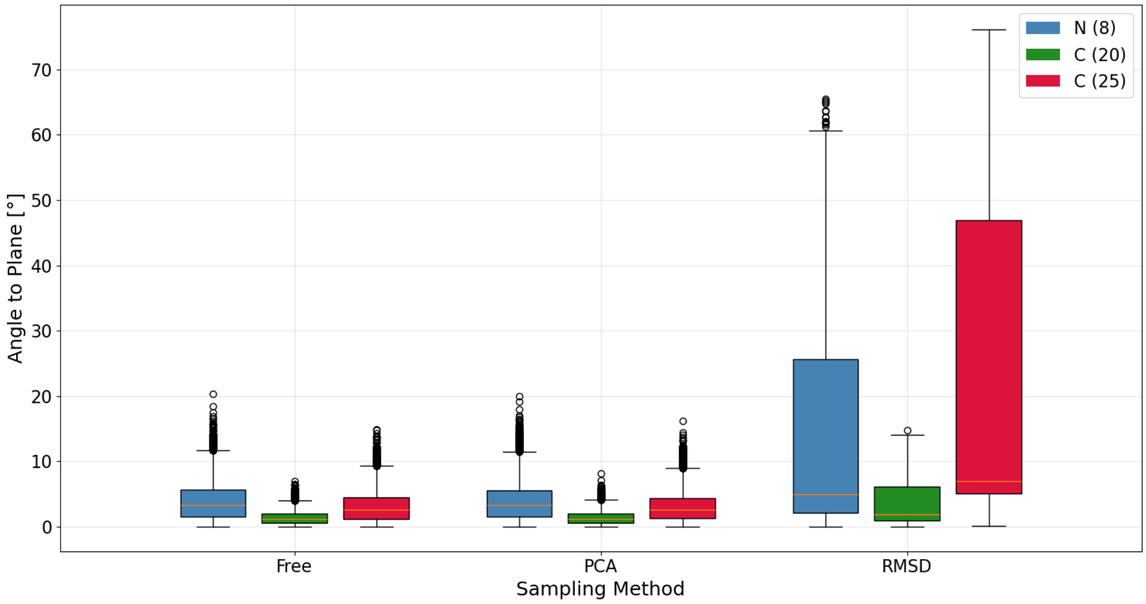


Figure 91.: ACDAN: out-of-plane angle distributions α_X (degrees) for N_8, C_{20}, and C_{25} under *Free*, *PCA*, and *RMSD* sampling. Angles are measured relative to the aromatic reference plane (Sec. 3.7.7).

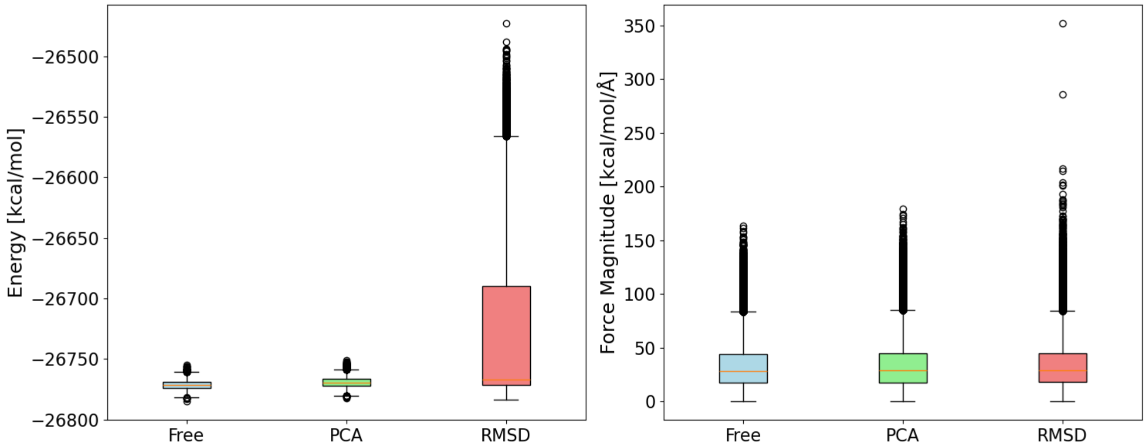


Figure 92.: ACDAN: per-frame *energy* (left, kcal mol^{-1}) and *force magnitude* (right, $\text{kcal mol}^{-1}\,\text{Å}^{-1}$) for *Free*, *PCA*, and *RMSD*. PCA tightens the distributions relative to Free; RMSD generates the broadest support and the heaviest tails. Uncertainty-guided sampling was not available for this system due to technical constraints.

MACE cross-evaluation. For *ACDAN* (Table 43), cross-regime transfer is weak. In this paragraph, energies are presented in kcal mol^{-1} and forces in kcal mol^{-1} Å^{-1}. Diagonals are lowest (Energy / Force MAE: Free 0.022 / 0.073, PCA 0.59 / 0.77, RMSD 0.82 / 0.86). Off-diagonal errors rise markedly, especially when applying Free- or PCA-trained models to *RMSD* tests (Energy 9.17 / 2.49; Force 3.92 / 4.62 for Free / PCA). Notably, the *RMSD*-trained model tolerates PCA tests comparatively well (Energy 0.60; Force 0.78), whereas the reverse direction is poor (RMSD tests with PCA model: Force 4.62). These trends indicate that models should be trained within the intended sampling regime – or on an explicitly mixed dataset – since near-equilibrium training does not generalize to off-basin configurations for this system.

Model [Energy]

Test Dataset	Free	PCA	RMSD
Free	0.02	1.54	1.56
PCA	1.47	0.59	0.60
RMSD	9.17	2.49	0.82

(a) Energy MAE [kcal mol^{-1}].

Model [Force]

Test Dataset	Free	PCA	RMSD
Free	0.07	1.18	1.14
PCA	1.36	0.77	0.78
RMSD	3.92	4.62	0.86

(b) Force MAE [kcal mol^{-1} Å^{-1}].

Table 43.: Cross-evaluation on **ACDAN** (without environment) with a single **MACE** architecture. Each column is a MACE model trained on data from the indicated sampling strategy; each row reports the MAE on the corresponding *test* dataset. The grids capture in-domain performance (diagonal) and transfer across strategies (off-diagonal).

8.3.4. Case: Retinal

Retinal presents a distinct challenge because 11-cis ↔ all-trans isomerisation involves coupled C=C/C–C dihedrals along a polyene backbone. Three sampling strategies are considered – Free MD, PCA-biased metadynamics, and uncertainty-guided sampling (UNCTY) – and the analysis assesses (i) backbone planarity via the global angle Θ_{glob}, (ii) per-bond dihedral angle distributions along C_6–C_{16}, (iii) per-frame energy and force statistics, and (iv) cross-evaluation of MACE models trained within each regime. RMSD-biased metadynamics is omitted due to technical constraints.

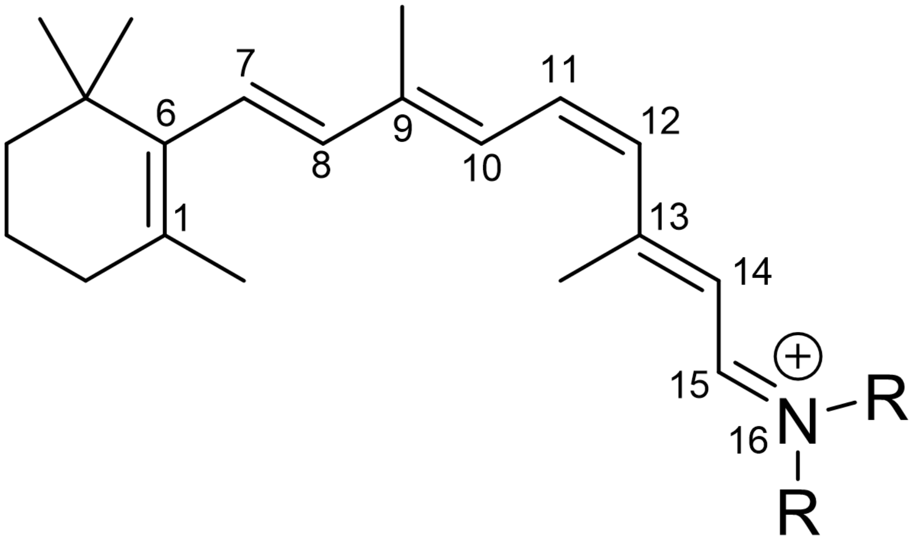


Figure 93.: Retinal with the conjugated backbone numbered (C_6–C_{16}). This ordering is used throughout the Retinal analysis: it defines the per–bond torsions ϕ_i and the backbone–planarity metric of Sec. 3.7.7, where local plane normals $\hat{\mathbf{n}}_i$ from bonds $(i{-}1, i)$ and $(i, i{+}1)$ are aggregated into the global coherence S_{glob} (and angle θ_{glob}). The terminal iminium nitrogen is shown schematically and is not part of the numbered backbone.

Backbone planarity. Backbone planarity is quantified with the global angle $\Theta_{\mathrm{glob}} \in [0°, 90°]$ from Sec. 3.7.7. Across methods, *Free* MD exhibits the broadest distribution with a higher median, indicating frequent moderate corrugations along the polyene chain. *PCA*-biased sampling reduces both median and interquartile range relative to Free MD, consistent with excursions that are steered along principal, low-curvature modes of the training data. *UNCTY* (uncertainty-guided) is tightest overall, yielding the smallest median and the narrowest IQR among the three. All methods retain occasional high-angle outliers; these are rarer under PCA and UNCTY and are attributable to transient local kinks or frames with near-linear segments where local-plane estimation becomes less robust. Because the metric averages absolute dot products of local normals, values are strictly non-negative and the empirical distribution is not centred at 0° even for small thermal corrugations (Fig. 94).

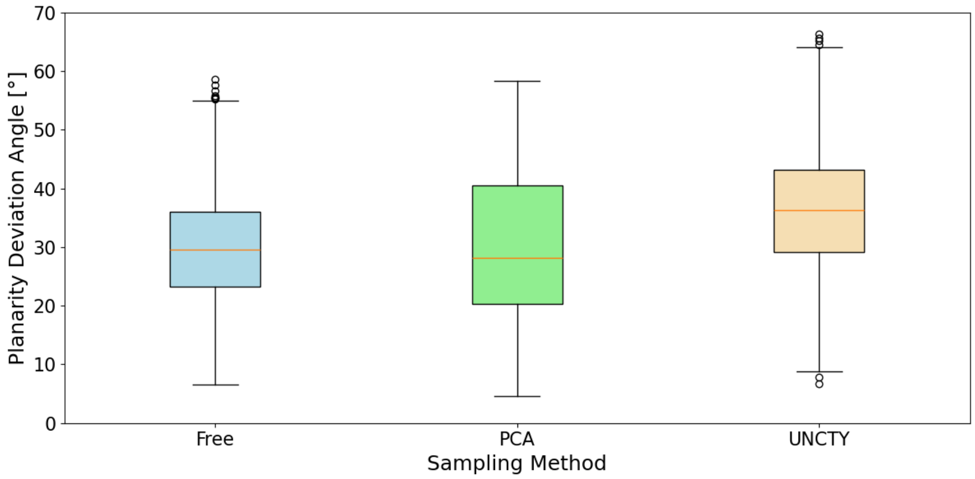


Figure 94.: Retinal: boxplots of the global backbone planarity angle Θ_{glob} (degrees) for *Free*, *PCA*, and *UNCTY*. Lower values indicate a more nearly planar polyene chain; PCA tightens the distribution relative to Free, UNCTY tightens it further, i.e., a backbone closer to planarity with fewer moderate corrugations. In contrast to FR0, 4AP, and ACDAN – where atom-wise out-of-plane angles relative to a fixed aromatic plane are used – retinal is assessed with this backbone-wide planarity metric.

Backbone dihedral coverage. Column-wise angular densities for consecutive backbone dihedrals (Fig. 95) show the expected polyene pattern: most bonds localise near *s-trans* (±180°), while the central junction (e.g., 6–7–8–9) peaks close to 0° (*s-cis*). Relative to *Free* MD, *PCA* sharpens several columns and reduces diffuse background, indicating sampling that remains closer to the dominant conformers while still visiting the characteristic *s-cis* pocket at the centre. *UNCTY* further concentrates density around the same modes and suppresses low-probability off-angles, yielding the narrowest ridges across the backbone. Towards the chain ends a mild broadening remains visible under all methods, consistent with the increased conformational softness of terminal segments; PCA and UNCTY nonetheless curtail leakage into higher-energy quadrants. Overall, PCA and UNCTY preserve the qualitative conformer ordering from Free MD while tightening the distributions in a way that aligns with the planarity trends.

Energy and force distributions. Energy and force statistics (Fig. 96) follow the structural trends with one caveat. The *Free* trajectories show a consistent absolute energy offset relative to *PCA* and *UNCTY*; repeated internal checks reproduce the same shift, so cross-method *levels* of the energy should not be interpreted. Comparison is therefore based on (i) the within-method spread and (ii) the force distributions. In that view, *Free* exhibits the widest energy band and the broadest force tails, *PCA* narrows both and lowers the median force magnitude, and *UNCTY* tightens the distributions further and suppresses extreme outliers, yielding the smallest interquartile ranges. These trends track the decrease in Θ_{glob} and the sharpening of the dihedral densities, i.e., methods that keep the backbone closer to the conjugated reference geometry populate less extreme – and more learnable – force regimes.

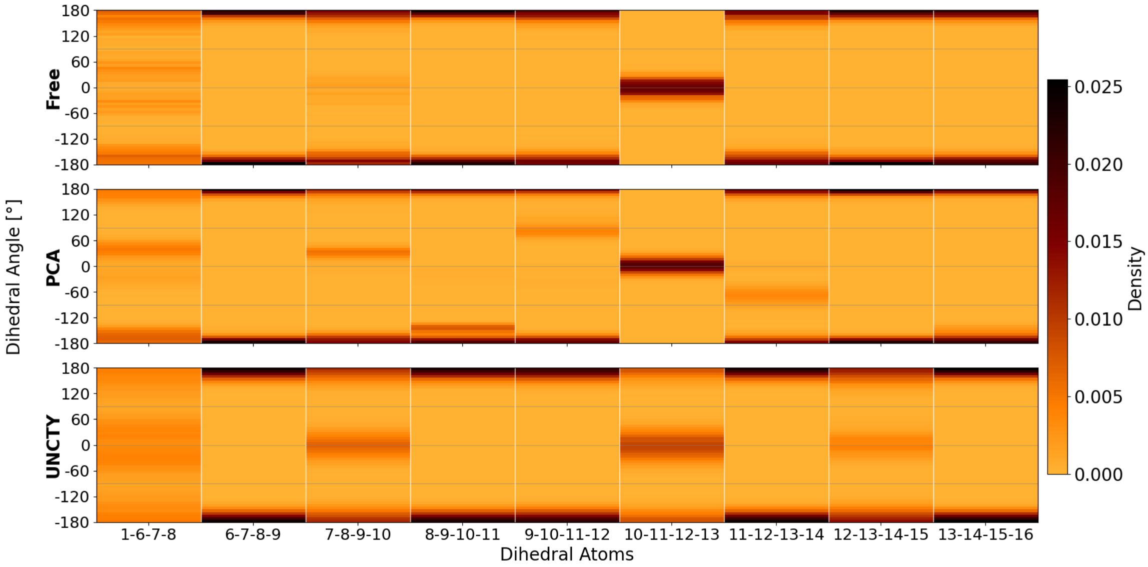


Figure 95.: Retinal: heatmap of per-dihedral angular densities along the numbered backbone from Fig. 93. Columns list consecutive four-atom dihedrals; rows give the dihedral angle $(-180°, 180°]$. *Free* MD, *PCA*, and *UNCTY* (uncertainty-guided sampling). Most dihedrals occupy *s-trans* (±180°).

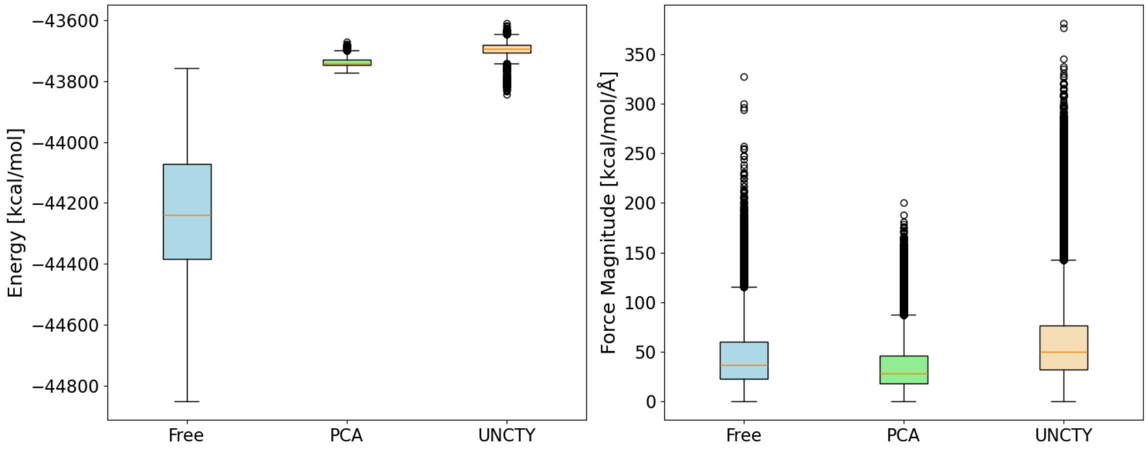


Figure 96.: Retinal: per-frame *energy* (left, $\mathrm{kcal\,mol^{-1}}$) and *force magnitude* (right, $\mathrm{kcal\,mol^{-1}\,Å^{-1}}$) for *Free*, *PCA*, and *UNCTY*. RMSD-biased metadynamics could not be included here due to technical problems.

MACE cross-evaluation. For *Retinal* (Table 44), *energies* reveal strong distributional mismatch: only the *UNCTY* model attains a small in-domain Energy MAE (0.31 kcal mol^{-1}), whereas *PCA* is already sizable in-domain (8.46 kcal mol^{-1}) and *Free* is large (159.67 kcal mol^{-1}). Off-domain energy errors for all pairs are equal or larger than $\gtrsim$ 100 kcal mol^{-1}. *Forces* are more discriminative for MD: *UNCTY* achieves the best in-domain Force MAE (0.30 kcal mol^{-1} Å^{-1}), *PCA* is second (0.83 kcal mol^{-1} Å^{-1}), and *Free* is worst (4.80 kcal mol^{-1} Å^{-1}). Cross-transfer remains limited: *PCA* and *Free* are roughly interchangeable on each other (4.195 kcal mol^{-1} Å^{-1} on Free tests with PCA model, 4.02 kcal mol^{-1} Å^{-1} on PCA tests with Free model), neither generalizes well to *UNCTY* tests (3.29 and 5.91 kcal mol^{-1} Å^{-1}), and the *UNCTY*-trained model does not yield low error on Free/PCA tests (5.93 / 4.26 kcal mol^{-1} Å^{-1}). Consistent with the absolute-energy offset observed for the Free trajectories (Fig. 96), the poor performance involving the Free model is attributed to this systematic shift, which inflates energy losses and perturbs the joint energy-force objective; consequently, force-based training and transfer suffer whenever the off-energy Free distribution is involved.

Model [Energy]

Test Dataset	Free	PCA	UNCTY
Free	159.67	510.60	422.46
PCA	521.10	8.46	92.68
UNCTY	423.42	102.17	0.31

(a) Energy MAE [kcal mol^{-1}].

Model [Force]

Test Dataset	Free	PCA	UNCTY
Free	4.80	4.19	5.93
PCA	4.02	0.83	4.26
UNCTY	5.91	3.29	0.30

(b) Force MAE [kcal mol^{-1} Å^{-1}].

Table 44.: Cross-evaluation on **Retinal** (without environment) with a single **MACE** architecture. Each column is a MACE model trained on data from the indicated sampling strategy; each row reports the MAE on the corresponding *test* dataset. The grids capture in-domain performance (diagonal) and transfer across strategies (off-diagonal).

8.4. Discussion

Across molecules, a coherent pattern emerges from the structural diagnostics and the MACE cross-evaluations (FR0: Table 41; 4AP: Table 42; ACDAN: Table 43; Retinal: Table 44). *Forces* – not energies – govern MD stability and therefore remain the primary learning target; energy statistics are still informative as sanity checks and to expose distributional mismatch.

What each strategy contributes. *Free* MD supplies robust, near-equilibrium cores. *PCA*-biased metadynamics enriches local diversity with minimal heavy tails, effectively tightening distributions around dominant modes. *RMSD*-biased metadynamics deliberately generates off-basin geometries and hence broad force and energy tails – valuable for stress testing, but typically difficult to transfer to from near-equilibrium training. *UNCTY* (uncertainty-guided sampling) targets under-represented regions while keeping tails comparatively controlled, often producing the best in-domain force accuracy. *NMA* produced heavy-tailed outliers and poor cross-transfer on FR0 and was therefore not pursued for the remaining systems – for the purposes of this chapter it is considered non-competitive and omitted.

Transferability by system. For **FR0**, Free and PCA generalize well to each other, RMSD remains competitive on those tests, and the UNCTY-trained model attains the lowest mean error across non-NMA test sets (Table 41). For **4AP**, Free ↔ PCA transfer is strong on forces, UNCTY is best in-domain and acceptable on Free/PCA, whereas RMSD as a *test* distribution breaks transfer for models trained on Free, PCA, or UNCTY; notably, the RMSD-trained *energy* model transfers back to Free/PCA with small errors (Table 42). For **ACDAN**, cross-regime transfer is weak: diagonal minima dominate and applying Free or PCA models to RMSD tests yields large errors, while the RMSD-trained model tolerates PCA tests somewhat better than vice versa (Table 43). For **Retinal**, energies manifest severe distributional mismatch across regimes, whereas forces are more diagnostic: UNCTY achieves the best in-domain force MAE, PCA is second, and Free is worst; cross-transfer remains limited (Table 44).

Practical guidance. Mixed training sets should be considered: build a Free+PCA backbone, then *subsample* UNCTY and RMSD to harvest their benefits without importing heavy tails. Keep the combined UNCTY+RMSD share modest (about 10–30%), trim high-force outliers, and thin by diversity in CV space. Where distributions mismatch (e.g., energy offsets in Retinal Free), center energies per source or emphasize a force-dominant loss. UNCTY is effective for patching genuine gaps; RMSD is best for robustness checks or as a small dedicated slice.

Limitations and outlook. All comparisons use a single network family (MACE) and fixed hyperparameters; alternative architectures or active-learning criteria may shift absolute errors. CV definitions influence PCA- and RMSD-biased protocols; systematic CV tuning could further improve transfer. For Retinal – with its coupled polyene dihedrals and narrow near-planar basin under UNCTY – per-regime models or mixture training with explicit weighting appear preferable to a single mixed model.

8.5. Summary

Free MD provides a reliable baseline across systems. PCA-biasing concentrates sampling around dominant conformers and improves local coverage with limited tails. Uncertainty-guided selection extends support into under-sampled regions while maintaining learnable force distributions, often yielding the best in-domain force accuracy. RMSD-biased sampling efficiently generates challenging off-basin configurations that are useful for robustness and for training specialized components, but they transfer poorly if used in isolation against near-equilibrium tests. Normal-mode sampling produced heavy tails and weak transfer on FR0 and was excluded thereafter.

In practice, committing to a single sampling strategy entails trade-offs. Free MD can be entirely adequate (cf. Chapter VI) but may require long trajectories to reach coverage. RMSD-biased sampling rapidly generates off-basin geometries – useful for stress testing, but it can under-sample typical states; PCA tightens local coverage but explores less; uncertainty-guided sampling often offers a middle ground by extending support with comparatively controlled tails. As an alternative to a single choice, it can be helpful to *subsample and mix* – e.g., keep a Free/PCA core, add a light uncertainty-guided slice to patch gaps, and optionally a trimmed RMSD fraction for robustness.

Part IX

Summary and Outlook

9.1. Scope and Context

This thesis develops machine-learning models for *multiscale* simulation of *complex molecular systems* in *complex environments*. The work has two main pillars: (i) a **phosphorylation** line that uses a QM+ML/MM approach based on DFTB with Δ-learning to reach DFT accuracy on reaction energies in solution, and (ii) a **fluorescence** line that couples an NN/MM force model for excited states to a property network for spectroscopic observables. Two compact additions support these pillars: a **retinal absorption** study that reuses the fluorescence property network for $S_0 \longrightarrow S_1$ and a **sampling** comparison that systematises data generation. Across chapters the emphasis is on clean targets, consistent electrostatics handling, coverage diagnostics, and reusability of the same geometry+ESP description across tasks. The overall goal is practical multiscale modelling in a complex environment – not gas-phase problems.

9.2. Summary of Main Findings

Phosphorus chemistry is a representative case of *complex chemistry* in solution. This chapter tests whether Δ–learning can lift DFTB to DFT (ωB97X–like) quality for *reaction energies and generalizes across reaction families* in a QM+ML/MM setting.
HDNNP (central result). Training the Δ–head on reaction-energy labels makes the combined DFTB+Δ model match DFT profiles on in–domain and held–out variants, and transfers across related phosphorylations within the family. Forces are then added to stabilize dynamics, but accuracy conclusions are driven by ΔE. With this ordering, barrier regions and products are reproduced under explicit solvent and remain predictive for previously unseen but chemically related cases. In practical terms, the Δ–correction increases wall time over pure DFTB by a factor of ~ 2 to ~ 10 in the tested size range, while retaining the key advantage of a QM+ML/MM workflow: DFT-grade energetics at costs compatible with routine trajectory work in complex environments.
A compact MACE model trained on the same Δ labels learns *energies and forces jointly*, achieving sub-kcal mol^{-1} energy fits on training data and $\lesssim 1$ kcal mol^{-1} RMSE$_E$ for most held–out reactions, with robust force learning. Thus, both HDNNP and MACE constitute viable Δ-ML routes for free-energy mapping. In this chapter, *both* HDNNP and MACE were deployed on-the-fly as Δ correctors. On-the-fly Δ raises cost from 22–27 to 47–272 h ns^{-1} with HDNNP (CPU-only), while MACE delivers 72–90 h ns^{-1} (~2–3× lower); with four GPU walkers, Table 19 likely overestimates the HDNNP overhead.

The **fluorescence** chapter asks whether a 1 ns excited-state dynamics for a dye in a complex environment – whose adiabatic excited states undergo frequent state crossings – can be run and its spectrum predicted within a single coherent pipeline. The answer is yes. An NN/MM *force model* trained on QM/MM labels across a solvent panel produces nanosecond-stable trajectories with a 0.5 fs step – at wall times of $\sim$ 7–10 hours for runs that would otherwise require $\sim$ 90–120 days *with* 1.0 *fs time steps* as direct TD-LC-DFTB/MM production. Because the dye's adiabatic excited states frequently reorder (“root flipping”), framewise single-point TD-LC-DFTB cannot reliably track bright-state identity over time; the NN/MM dynamics therefore provide the sampling vehicle. On the same geometry+ESP cutouts, a *property network* with two heads (emission energy and oscillator strength) predicts spectra that reproduce experimental trends across solvents. The architecture is kept fixed and the gains come from data design and labels. A MACE benchmark under the same protocol verifies the force accuracy.

Retinal was used only to test whether the fluorescence property network can be *repurposed for absorption*. With the same retinal-lysine QM region and electrostatics, diagnostics recover the expected BLA–energy relation, and solution-phase ESP magnitudes overlap the protein-like range. Two label strategies were compared: (i) direct OM2/MRCI and (ii) Δ-learning (OM2 − DFTB). For absorption, the direct OM2/MRCI training provided the more stable calibration, whereas Δ-learning did tend to underperform in the protein application (attenuated shift magnitudes and increased scatter after transfer). Consequently, the result here is methodological: the network can be applied to absorption without architectural changes; further performance gains are expected to come primarily from improved electronic-structure labels (i.e., more reliable reference methods/levels), which lies outside the scope of this thesis.

Sampling is compared as a support module for force learning and ML/MM stability: free MD, PCA-biased metadynamics, RMSD-biased metadynamics, normal-mode perturbations, and uncertainty-guided selection. On *FR0* (with cross-evaluations using MACE) free and PCA give robust cores, RMSD and uncertainty add reach, and normal modes create heavy-tailed forces with poor transfer – hence NMA is not used for production datasets. The same planarity and dihedral diagnostics explain energy/force statistics for *4AP* and *ACDAN*. For *retinal*, backbone planarity and dihedral maps show that PCA and uncertainty tighten distributions relative to free MD while preserving the characteristic *s*-trans/*s*-cis pattern. Cross-evaluations for non-FR0 systems are planned but are not required for the conclusions in the main pillars.

9.3. Integration

The two pillars share one multiscale pattern: fixed QM cutouts, consistent treatment of electrostatics via per-atom ESP on the QM region, coverage-aware curation, and reuse of the same input description across tasks. In the *reactive* line (phosphorylation) this becomes DFTB + Δ in QM+ML/MM for reaction energies with family-level transfer; in the *spectroscopic* line (fluorescence → retinal) the same cutouts feed an NN/MM force model and a property network. The common thread is that compact models plus disciplined data design make multiscale simulation in a complex environment feasible.

9.4. Limitations

Several constraints are shared across chapters. Absolute levels remain sensitive to label fidelity (DFTB3 with SRP, TD-LC-DFTB, OM2/MRCI) and to how electrostatics are represented; per-atom ESP features approximate polarization but do not encode full many-body response. Fixed-charge MM embeddings introduce solvent-dependent dielectric biases that can compress or inflate the ESP feature range, yielding near-rigid energy or wavelength shifts. Δ-learning can reduce variance and MAE while lowering R^2 when target ranges widen. Coverage is finite by design – curated windows stabilize training but limit out-of-domain guarantees. Finally, Δ adds wall-time overhead relative to pure baselines (DFTB or TD-DFTB), even though costs remain far below higher-level QM.

Phosphorylation. Δ-correctors inherit systematic parts of the DFTB3 ↔ DFT gap; changing either level will shift absolute numbers. Absolute ΔE anchoring across reaction families is weak: per-family standardization and fragment-based residual references can induce near-constant family-specific offsets in ΔE (visible as a rigid shift in parity plots), which depress R^2_E but leave forces unaffected. Mitigation – add a per-family bias term or calibrate each family on a small set of frames. Pooling by total charge is a proxy for true charge awareness; the Δ correction is assumed local – information on long-range interactions is not part of the present model. The deployed HDNNP Δ-QM/MM used energy-only training, which can limit metadynamics robustness. Core MACE learns forces strongly but shows energy outliers for some held-out reactions; the charge-augmented MACE-EQ variant should improve on core MACE, but underperforms, likely due to the present integration of electrostatics.

Fluorescence. Bright-state identity is recovered from TD-LC-DFTB single points along NN/MM trajectories; frequent state reorderings mean that raw TD-DFTB trajectories cannot follow the physical bright state. The fixed S_2 charge embedding – chosen to avoid discontinuities – is an approximation; state-consistent polarizable embeddings or learned charge equilibration would be preferable. Absolute spectral positions carry solvent-specific offsets attributable to fixed-charge dielectric biases and to the TD-DFTB reference; a single-solvent alignment removes the dominant rigid shift but does not fix the underlying physics. Rare S_3 episodes remain under-represented after brightness filtering. Extension to protein-dye systems is plausible but remains to be validated under heterogeneous, anisotropic fields.

Retinal. Calibration quality is limited by semiempirical labels; direct OM2/MRCI stabilizes oscillator-strength learning, whereas Δ-to-DFTB reduces energy MAE but transfers imperfectly to proteins. Absolute color shifts in proteins remain sensitive to label level and to missing polarization in the ESP description.

Sampling. Comparisons use a single architecture (MACE – without environment) and fixed hyperparameters; other choices may shift absolute errors. RMSD-biased and normal-mode protocols generate heavy-tailed force distributions that are valuable for stress tests but transfer poorly if over-represented in training. CV definitions for PCA influence coverage; mixture curricula mitigate but do not remove distributional mismatch. Uncertainty sampling improves in-domain force accuracy, yet its benefit depends on how uncertainty is estimated and on the diversity of candidate pools.

9.5. Outlook

Near-term work aims to raise the fidelity of training labels, improve how electrostatics and charge flow are represented, and make data coverage and diagnostics more systematic across all case studies.

Phosphorylation. Priorities are (i) joint energy+force training for the HDNNP Δ corrector to further stabilize metadynamics, (ii) charge awareness or explicit long-range terms in otherwise local models. On the equivariant line, the immediate goal is to make MACE-EQ robust in the Δ setting: rework how electrostatics couple to the $\Delta E/\Delta \mathbf{F}$ objectives. Core MACE remains a strong force learner but is charge agnostic; MACE-EQ should provide charge awareness once the electrostatic head is better integrated.

Fluorescence. The immediate bottleneck is the electronic-structure reference: TD-LC-DFTB does not reliably calibrate bright-state energies or orderings. The priority is a level upgrade – e.g., ωB97X – and a two-step relabelling: first retrain the property heads at the higher level to recalibrate emission energies and oscillator strengths; then generate excited-state forces at the same level on representative solvent snapshots and retrain the NN/MM force model, comparing dynamics and spectra side by side. Next, upgrade the embedding from fixed-charge to *mutually polarizable*: use polarizable MM so that the solvent polarizes the QM region and the QM back-polarizes the solvent; within the QM cutout replace fixed S_2 charges with *flexible charges* updated on the fly to capture charge redistribution across state crossings. A second thrust is multi-state learning: predict S_1-S_3 energies, forces, and oscillator strengths jointly, expose state identity, and enable simple surface hopping – essential for systems without a dominant bright state (e.g., 4AP). Extension to protein-dye complexes then becomes a direct next application.

Retinal. Improve label fidelity on selected snapshots with higher–level QM/MM to calibrate absolute shifts, and add polarization–aware descriptors to better capture protein fields. Where transfer errors point to under–sampled regions, expand beyond free MD by adding targeted sampling – e.g., uncertainty–driven selection and modest PCA– or RMSD–biased windows around problematic coordinates – to cover rare but relevant configurations without importing heavy tails.

Sampling. The full sampling matrix was not exercised across systems: for *ACDAN*, uncertainty-guided selection could not be integrated due to code errors; for *retinal*, RMSD-biased metadynamics also failed to run. All cross-evaluations used core MACE only (no explicit environment terms), and retinal cross-validation was comparatively weak – causes need targeted follow-up on label quality, distribution mismatch, and descriptor support, but this was out of scope time-wise. Priority items are a robust implementation of uncertainty sampling for ACDAN and RMSD-biased windows for retinal, plus deeper diagnostics of the retinal case. Going forward, continue systematic method comparisons and include new strategies such as active learning that interleave training with acquisition to benchmark against the current fixed setups.

9.6. Concluding Remark

This thesis set out to test whether compact machine learning models can enable *multiscale* simulation of *complex molecular systems* in *complex environments* – concretely, phosphoryl transfer in solution and excited state energy surfaces of dyes. The two pillars deliver on this goal. For phosphorylation, on-the-fly DFTB+Δ corrections approach DFT quality and transfer across related reaction families, yielding free-energy maps that correct baseline biases while remaining practical for routine trajectories. For fluorescence, an NN/MM force model coupled to a property network sustains nanosecond dynamics on the emissive surface, preserves bright-state identity through frequent reorderings, and reproduces solvent trends in emission spectra – making bright-state following feasible where direct TD-LC-DFTB/MM is not. The retinal case shows that the same property architecture can be repurposed for absorption without changing the network design.

Taken together, the results support a simple message: with PBC-consistent electrostatics, stable force labels, and coverage-aware curation, *compact* models are sufficient to deliver reliable ML/MM dynamics and transferable spectroscopic predictions in complex environments. Building on the systems and the foundations developed here, the same framework is well positioned to extend into biomolecular settings – enabling quantitative protein spectrochemistry and simulations of reactive events within structured macromolecular environments.

Bibliography

[1] Hans Martin Senn and Walter Thiel. "QM/MM Methods for Biomolecular Systems". en. In: *Angewandte Chemie International Edition* 48.7 (2009). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.200802019, pp. 1198–1229. ISSN: 1521-3773. DOI: `10.1002/anie.200802019`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.200802019` (visited on 07/27/2025).

[2] Andreas Dreuw and Martin Head-Gordon. "Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules". In: *Chemical Reviews* 105.11 (Nov. 2005). Publisher: American Chemical Society, pp. 4009–4037. ISSN: 0009-2665. DOI: `10.1021/cr0505627`. URL: `https://doi.org/10.1021/cr0505627` (visited on 08/20/2025).

[3] Stefan Goedecker. "Linear scaling electronic structure methods". In: *Reviews of Modern Physics* 71.4 (July 1999). Publisher: American Physical Society, pp. 1085–1123. DOI: `10.1103/RevModPhys.71.1085`. URL: `https://link.aps.org/doi/10.1103/RevModPhys.71.1085` (visited on 08/20/2025).

[4] Rodney J. Bartlett and Monika Musiał. "Coupled-cluster theory in quantum chemistry". In: *Reviews of Modern Physics* 79.1 (Feb. 2007). Publisher: American Physical Society, pp. 291–352. DOI: `10.1103/RevModPhys.79.291`. URL: `https://link.aps.org/doi/10.1103/RevModPhys.79.291` (visited on 08/20/2025).

[5] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". en. In: *Physical Review* 136.3B (Nov. 1964), B864–B871. ISSN: 0031-899X. DOI: `10.1103/PhysRev.136.B864`. URL: `https://link.aps.org/doi/10.1103/PhysRev.136.B864` (visited on 01/10/2024).

[6] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". en. In: *Physical Review* 140.4A (Nov. 1965), A1133–A1138. ISSN: 0031-899X. DOI: `10.1103/PhysRev.140.A1133`. URL: `https://link.aps.org/doi/10.1103/PhysRev.140.A1133` (visited on 01/12/2024).

[7] Tim J. Zuehlsdorff and Christine M. Isborn. "Modeling absorption spectra of molecules in solution". en. In: *International Journal of Quantum Chemistry* 119.1 (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.25719, e25719. ISSN: 1097-461X. DOI: `10.1002/qua.25719`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.25719` (visited on 08/20/2025).

[8] Christine M. Isborn et al. “The Charge Transfer Problem in Density Functional Theory Calculations of Aqueously Solvated Molecules”. In: *The Journal of Physical Chemistry B* 117.40 (Oct. 2013). Publisher: American Chemical Society, pp. 12189–12201. ISSN: 1520-6106. DOI: `10.1021/jp4058274`. URL: `https://doi.org/10.1021/jp4058274` (visited on 08/20/2025).

[9] Raghunathan Ramakrishnan et al. “Big Data Meets Quantum Chemistry Approximations: The -Machine Learning Approach”. en. In: *Journal of Chemical Theory and Computation* 11.5 (May 2015), pp. 2087–2096. ISSN: 1549-9618, 1549-9626. DOI: `10.1021/acs.jctc.5b00099`. URL: `https://pubs.acs.org/doi/10.1021/acs.jctc.5b00099` (visited on 07/30/2025).

[10] Michael Gaus, Qiang Cui, and Marcus Elstner. “DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB)”. In: *Journal of Chemical Theory and Computation* 7.4 (Apr. 2011). Publisher: American Chemical Society, pp. 931–948. ISSN: 1549-9618. DOI: `10.1021/ct100684s`. URL: `https://doi.org/10.1021/ct100684s` (visited on 07/30/2025).

[11] Michael Gaus, Albrecht Goez, and Marcus Elstner. “Parametrization and Benchmark of DFTB3 for Organic Molecules”. In: *Journal of Chemical Theory and Computation* 9.1 (Jan. 2013). Publisher: American Chemical Society, pp. 338–354. ISSN: 1549-9618. DOI: `10.1021/ct300849w`. URL: `https://doi.org/10.1021/ct300849w` (visited on 07/24/2025).

[12] Jörg Behler and Michele Parrinello. “Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces”. In: *Physical Review Letters* 98.14 (Apr. 2007). Publisher: American Physical Society, p. 146401. DOI: `10.1103/PhysRevLett.98.146401`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.98.146401` (visited on 07/24/2025).

[13] J. S. Smith, O. Isayev, and A. E. Roitberg. “ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost”. en. In: *Chemical Science* 8.4 (2017). Publisher: Royal Society of Chemistry, pp. 3192–3203. DOI: `10.1039/C6SC05720A`. URL: `https://pubs.rsc.org/en/content/articlelanding/2017/sc/c6sc05720a` (visited on 08/20/2025).

[14] Oliver T. Unke and Markus Meuwly. “PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges”. In: *Journal of Chemical Theory and Computation* 15.6 (June 2019). Publisher: American Chemical Society, pp. 3678–3693. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.9b00181`. URL: `https://doi.org/10.1021/acs.jctc.9b00181` (visited on 08/20/2025).

[15] K. T. Schütt et al. “SchNetPack: A Deep Learning Toolbox For Atomistic Systems”. In: *Journal of Chemical Theory and Computation* 15.1 (Jan. 2019). Publisher: American Chemical Society, pp. 448–455. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.8b00908`. URL: `https://doi.org/10.1021/acs.jctc.8b00908` (visited on 08/20/2025).

[16] Julia Westermayr and Philipp Marquetand. “Machine Learning for Electronically Excited States of Molecules”. In: *Chemical Reviews* 121.16 (Aug. 2021). Publisher: American Chemical Society, pp. 9873–9926. ISSN: 0009-2665. DOI: `10.1021/acs.chemrev.0c00749`. URL: `https://doi.org/10.1021/acs.chemrev.0c00749` (visited on 07/28/2025).

[17] F. H. Westheimer. “Why Nature Chose Phosphates”. In: *Science* 235.4793 (Mar. 1987). Publisher: American Association for the Advancement of Science, pp. 1173–1178. DOI: `10.1126/science.2434996`. URL: `https://www.science.org/doi/abs/10.1126/science.2434996` (visited on 08/20/2025).

[18] Philip Cohen. “Protein kinases — the major drug targets of the twenty-first century?” en. In: *Nature Reviews Drug Discovery* 1.4 (Apr. 2002). Publisher: Nature Publishing Group, pp. 309–315. ISSN: 1474-1784. DOI: `10.1038/nrd773`. URL: `https://www.nature.com/articles/nrd773` (visited on 07/31/2025).

[19] Andrey S. Klymchenko. “Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications”. In: *Accounts of Chemical Research* 50.2 (Feb. 2017). Publisher: American Chemical Society, pp. 366–375. ISSN: 0001-4842. DOI: `10.1021/acs.accounts.6b00517`. URL: `https://doi.org/10.1021/acs.accounts.6b00517` (visited on 08/17/2025).

[20] Faaizah Khan, Luigi Gnudi, and John C. Pickup. “Fluorescence-based sensing of glucose using engineered glucose/galactose-binding protein: A comparison of fluorescence resonance energy transfer and environmentally sensitive dye labelling strategies”. In: *Biochemical and Biophysical Research Communications* 365.1 (Jan. 2008), pp. 102–106. ISSN: 0006-291X. DOI: `10.1016/j.bbrc.2007.10.129`. URL: `https://www.sciencedirect.com/science/article/pii/S0006291X07023091` (visited on 08/17/2025).

[21] Christophe Chipot. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*. en. Springer Series in Chemical Physics Ser v.86. Berlin, Heidelberg: Springer Berlin / Heidelberg, 2007. ISBN: 978-3-540-38447-2 978-3-540-38448-9.

[22] Juan Santiago Grassano et al. *From QM/MM to ML/MM: A new era in multiscale modeling*. en. Aug. 2025. DOI: `10.26434/chemrxiv-2025-lhwt7`. URL: `https://chemrxiv.org/engage/chemrxiv/article-details/6894368d728bf9025e421b0d` (visited on 10/16/2025).

[23] John A. Keith et al. “Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems”. In: *Chemical Reviews* 121.16 (Aug. 2021). Publisher: American Chemical Society, pp. 9816–9872. ISSN: 0009-2665. DOI: `10.1021/acs.chemrev.1c00107`. URL: `https://doi.org/10.1021/acs.chemrev.1c00107` (visited on 08/17/2025).

[24] Martin Karplus and J. Andrew McCammon. “Molecular dynamics simulations of biomolecules”. en. In: *Nature Structural Biology* 9.9 (Sept. 2002). Publisher: Nature Publishing Group, pp. 646–652. ISSN: 1545-9985. DOI: `10.1038/nsb0902-646`. URL: `https://www.nature.com/articles/nsb0902-646` (visited on 08/20/2025).

[25] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids.* en. Google-Books-ID: WFExDwAAQBAJ. Oxford University Press, Aug. 2017. ISBN: 978-0-19-252470-6.

[26] Andrew R. Leach. *Molecular Modelling: Principles and Applications.* en. Pearson Education, 2001. ISBN: 978-0-582-38210-7.

[27] A. Warshel and M. Levitt. "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme". In: *Journal of Molecular Biology* 103.2 (May 1976), pp. 227–249. ISSN: 0022-2836. DOI: `10.1016/0022-2836(76)90311-9`. URL: `https://www.sciencedirect.com/science/article/pii/0022283676903119` (visited on 07/28/2025).

[28] Kieron Burke. "Perspective on density functional theory". In: *The Journal of Chemical Physics* 136.15 (Apr. 2012), p. 150901. ISSN: 0021-9606. DOI: `10.1063/1.4704546`. URL: `https://doi.org/10.1063/1.4704546` (visited on 07/25/2025).

[29] D R Bowler and T Miyazaki. "methods in electronic structure calculations". en. In: *Reports on Progress in Physics* 75.3 (Feb. 2012), p. 036503. ISSN: 0034-4885. DOI: `10.1088/0034-4885/75/3/036503`. URL: `https://dx.doi.org/10.1088/0034-4885/75/3/036503` (visited on 08/20/2025).

[30] Carlo Adamo and Denis Jacquemin. "The calculations of excited-state properties with Time-Dependent Density Functional Theory". en. In: *Chemical Society Reviews* 42.3 (2013). Publisher: Royal Society of Chemistry, pp. 845–856. DOI: `10.1039/C2CS35394F`. URL: `https://pubs.rsc.org/en/content/articlelanding/2013/cs/c2cs35394f` (visited on 08/20/2025).

[31] Mark E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation.* en. Google-Books-ID: eLXLEAAAQBAJ. Oxford University Press, Aug. 2023. ISBN: 978-0-19-255961-6.

[32] Shankar Kumar et al. "THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method". en. In: *Journal of Computational Chemistry* 13.8 (1992). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.540130812, pp. 1011–1021. ISSN: 1096-987X. DOI: `10.1002/jcc.540130812`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540130812` (visited on 08/20/2025).

[33] Michael R. Shirts and John D. Chodera. "Statistically optimal analysis of samples from multiple equilibrium states". In: *The Journal of Chemical Physics* 129.12 (Sept. 2008), p. 124105. ISSN: 0021-9606. DOI: `10.1063/1.2978177`. URL: `https://doi.org/10.1063/1.2978177` (visited on 08/20/2025).

[34] Matthias Rupp et al. "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning". In: *Physical Review Letters* 108.5 (Jan. 2012). Publisher: American Physical Society, p. 058301. DOI: `10.1103/PhysRevLett.108.058301`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.108.058301` (visited on 07/28/2025).

[35] Raghunathan Ramakrishnan et al. “Quantum chemistry structures and properties of 134 kilo molecules”. en. In: *Scientific Data* 1.1 (Aug. 2014). Publisher: Nature Publishing Group, p. 140022. ISSN: 2052-4463. DOI: `10.1038/sdata.2014.22`. URL: `https://www.nature.com/articles/sdata201422` (visited on 09/29/2025).

[36] Stefan Chmiela et al. “Machine learning of accurate energy-conserving molecular force fields”. In: *Science Advances* 3.5 (May 2017). Publisher: American Association for the Advancement of Science, e1603015. DOI: `10.1126/sciadv.1603015`. URL: `https://www.science.org/doi/full/10.1126/sciadv.1603015` (visited on 09/29/2025).

[37] Kristof Schütt, Oliver Unke, and Michael Gastegger. “Equivariant message passing for the prediction of tensorial properties and molecular spectra”. en. In: *Proceedings of the 38th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, July 2021, pp. 9377–9388. URL: `https://proceedings.mlr.press/v139/schutt21a.html` (visited on 09/29/2025).

[38] Simon Batzner et al. “E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials”. en. In: *Nature Communications* 13.1 (May 2022). Publisher: Nature Publishing Group, p. 2453. ISSN: 2041-1723. DOI: `10.1038/s41467-022-29939-5`. URL: `https://www.nature.com/articles/s41467-022-29939-5` (visited on 07/28/2025).

[39] Ilyes Batatia et al. “MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields”. en. In: ().

[40] Justin S. Smith et al. “Less is more: Sampling chemical space with active learning”. In: *The Journal of Chemical Physics* 148.24 (May 2018), p. 241733. ISSN: 0021-9606. DOI: `10.1063/1.5023802`. URL: `https://doi.org/10.1063/1.5023802` (visited on 09/29/2025).

[41] Ziwei Pang et al. “Unravelling the mechanism of glucose binding in a protein-based fluorescence probe: molecular dynamics simulation with a tailor-made charge model”. en. In: *Physical Chemistry Chemical Physics* 24.4 (2022), pp. 2441–2453. ISSN: 1463-9076, 1463-9084. DOI: `10.1039/D1CP03733A`. URL: `http://xlink.rsc.org/?DOI=D1CP03733A` (visited on 01/12/2024).

[42] Oleksandr A. Kucherak et al. “Fluorene Analogues of Prodan with Superior Fluorescence Brightness and Solvatochromism”. In: *The Journal of Physical Chemistry Letters* 1.3 (Feb. 2010). Publisher: American Chemical Society, pp. 616–620. DOI: `10.1021/jz9003685`. URL: `https://doi.org/10.1021/jz9003685` (visited on 01/10/2024).

[43] Tim J. Zuehlsdorff et al. “Optical spectra in the condensed phase: Capturing anharmonic and vibronic features using dynamic and static approaches”. en. In: *The Journal of Chemical Physics* 151.7 (Aug. 2019), p. 074111. ISSN: 0021-9606, 1089-7690. DOI: `10.1063/1.5114818`. URL: `https://pubs.aip.org/jcp/article/151/7/074111/197927/Optical-spectra-in-the-condensed-phase-Capturing` (visited on 02/15/2024).

[44] Sapana V. Shedge et al. “Explicit environmental and vibronic effects in simulations of linear and nonlinear optical spectroscopy”. In: *The Journal of Chemical Physics* 154.8 (Feb. 2021), p. 084116. ISSN: 0021-9606. DOI: `10.1063/5.0038196`. URL: `https://doi.org/10.1063/5.0038196` (visited on 08/17/2025).

[45] Uriel N. Morzan et al. “Spectroscopy in Complex Environments from QM–MM Simulations”. en. In: *Chemical Reviews* 118.7 (Apr. 2018), pp. 4071–4113. ISSN: 0009-2665, 1520-6890. DOI: `10.1021/acs.chemrev.8b00026`. URL: `https://pubs.acs.org/doi/10.1021/acs.chemrev.8b00026` (visited on 08/17/2025).

[46] Bao-Xin Xue, Mario Barbatti, and Pavlo O. Dral. “Machine Learning for Absorption Cross Sections”. en. In: *The Journal of Physical Chemistry A* 124.35 (Sept. 2020), pp. 7199–7210. ISSN: 1089-5639, 1520-5215. DOI: `10.1021/acs.jpca.0c05310`. URL: `https://pubs.acs.org/doi/10.1021/acs.jpca.0c05310` (visited on 08/17/2025).

[47] Peter Atkins, Julio De Paula, and James Keeler. *Atkins’ Physical Chemistry*. en. 12th ed. Oxford University Press, Dec. 2022. ISBN: 978-0-19-884781-6 978-0-19-199216-2. DOI: `10.1093/hesc/9780198847816.001.0001`. URL: `https://www.oxfordsciencetrove.com/view/10.1093/hesc/9780198847816.001.0001/isbn-9780198847816` (visited on 01/10/2024).

[48] Marcus Elstner. *Physikalische Chemie II: Quantenmechanik und Spektroskopie*. de. Berlin, Heidelberg: Springer, 2021. ISBN: 978-3-662-61461-7 978-3-662-61462-4. DOI: `10.1007/978-3-662-61462-4`. URL: `https://link.springer.com/10.1007/978-3-662-61462-4` (visited on 01/09/2024).

[49] T. A. Niehaus et al. “Tight-binding approach to time-dependent density-functional response theory”. In: *Physical Review B* 63.8 (Feb. 2001). Publisher: American Physical Society, p. 085108. DOI: `10.1103/PhysRevB.63.085108`. URL: `https://link.aps.org/doi/10.1103/PhysRevB.63.085108` (visited on 01/10/2024).

[50] “Instrumentation for Fluorescence Spectroscopy”. en. In: *Principles of Fluorescence Spectroscopy*. Ed. by Joseph R. Lakowicz. Boston, MA: Springer US, 2006, pp. 27–61. ISBN: 978-0-387-46312-4. DOI: `10.1007/978-0-387-46312-4_2`. URL: `https://doi.org/10.1007/978-0-387-46312-4_2` (visited on 08/16/2025).

[51] Jerker Widengren, Uelo Mets, and Rudolf Rigler. “Fluorescence correlation spectroscopy of triplet states in solution: a theoretical and experimental study”. In: *The Journal of Physical Chemistry* 99.36 (Sept. 1995). Publisher: American Chemical Society, pp. 13368–13379. ISSN: 0022-3654. DOI: `10.1021/j100036a009`. URL: `https://doi.org/10.1021/j100036a009` (visited on 01/10/2024).

[52] A. Jablonski. “Efficiency of Anti-Stokes Fluorescence in Dyes”. en. In: *Nature* 131.3319 (June 1933). Number: 3319 Publisher: Nature Publishing Group, pp. 839–840. ISSN: 1476-4687. DOI: `10.1038/131839b0`. URL: `https://www.nature.com/articles/131839b0` (visited on 01/10/2024).

[53] Michael Kasha. “Characterization of electronic transitions in complex molecules”. en. In: *Discussions of the Faraday Society* 9 (1950), p. 14. ISSN: 0366-9033. DOI: `10.1039/df9500900014`. URL: `http://xlink.rsc.org/?DOI=df9500900014` (visited on 01/10/2024).

[54] Alexander P. Demchenko, Vladimir I. Tomin, and Pi-Tai Chou. “Breaking the Kasha Rule for More Efficient Photochemistry”. In: *Chemical Reviews* 117.21 (Nov. 2017). Publisher: American Chemical Society, pp. 13353–13381. ISSN: 0009-2665. DOI: `10.1021/acs.chemrev.7b00110`. URL: `https://doi.org/10.1021/acs.chemrev.7b00110` (visited on 08/16/2025).

[55] Franziska E. Wolff et al. “Origin of the Solvatochromism in Organic Fluorophores with Flexible Side Chains: A Case Study of Flugi-2”. In: *The Journal of Physical Chemistry A* 123.21 (May 2019). Publisher: American Chemical Society, pp. 4581–4587. ISSN: 1089-5639. DOI: `10.1021/acs.jpca.9b02474`. URL: `https://doi.org/10.1021/acs.jpca.9b02474` (visited on 01/09/2024).

[56] Christian Reichardt and Thomas Welton. *Solvents and Solvent Effects in Organic Chemistry*. en. John Wiley & Sons, Dec. 2010. ISBN: 978-3-527-32473-6.

[57] Christian Reichardt. “Solvatochromic Dyes as Solvent Polarity Indicators”. en. In: *Chemical Reviews* 94.8 (Dec. 1994), pp. 2319–2358. ISSN: 0009-2665, 1520-6890. DOI: `10.1021/cr00032a005`. URL: `https://pubs.acs.org/doi/abs/10.1021/cr00032a005` (visited on 08/16/2025).

[58] Noboru Mataga, Yozo Kaifu, and Masao Koizumi. “Solvent Effects upon Fluorescence Spectra and the Dipolemoments of Excited Molecules”. In: *Bulletin of the Chemical Society of Japan* 29.4 (Apr. 1956), pp. 465–470. ISSN: 0009-2673. DOI: `10.1246/bcsj.29.465`. URL: `https://doi.org/10.1246/bcsj.29.465` (visited on 08/16/2025).

[59] Rachel Crespo-Otero and Mario Barbatti. “Spectrum simulation and decomposition with nuclear ensemble: formal derivation and application to benzene, furan and 2-phenylfuran”. en. In: *Marco Antonio Chaer Nascimento: A Festschrift from Theoretical Chemistry Accounts*. Ed. by Fernando R. Ornellas and Maria João Ramos. Berlin, Heidelberg: Springer, 2014, pp. 89–102. ISBN: 978-3-642-41163-2. DOI: `10.1007/978-3-642-41163-2_9`. URL: `https://doi.org/10.1007/978-3-642-41163-2_9` (visited on 08/17/2025).

[60] Ajay Khanna et al. “Calculating absorption and fluorescence spectra for chromophores in solution with ensemble Franck–Condon methods”. In: *The Journal of Chemical Physics* 161.4 (July 2024), p. 044121. ISSN: 0021-9606. DOI: `10.1063/5.0217080`. URL: `https://doi.org/10.1063/5.0217080` (visited on 08/17/2025).

[61] Beatrix M. Bold et al. “Benchmark and performance of long-range corrected time-dependent density functional tight binding (LC-TD-DFTB) on rhodopsins and light-harvesting complexes”. en. In: *Physical Chemistry Chemical Physics* 22.19 (2020), pp. 10500–10518. ISSN: 1463-9076, 1463-9084. DOI: `10.1039/C9CP05753F`. URL: `http://xlink.rsc.org/?DOI=C9CP05753F` (visited on 01/12/2024).

[62] Monja Sokolov et al. “Analytical Time-Dependent Long-Range Corrected Density Functional Tight Binding (TD-LC-DFTB) Gradients in DFTB+: Implementation and Benchmark for Excited-State Geometries and Transition Energies”. In: *Journal of Chemical Theory and Computation* 17.4 (Apr. 2021). Publisher: American Chemical

Society, pp. 2266–2282. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.1c00095`. URL: `https://doi.org/10.1021/acs.jctc.1c00095` (visited on 08/17/2025).

[63] Martin Neumann. "Dipole moment fluctuation formulas in computer simulations of polar systems". In: *Molecular Physics* 50.4 (Nov. 1983). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00268978300102721, pp. 841–858. ISSN: 0026-8976. DOI: `10.1080/00268978300102721`. URL: `https://doi.org/10.1080/00268978300102721` (visited on 08/16/2025).

[64] Martin Neumann. "Computer simulation and the dielectric constant at finite wavelength". In: *Molecular Physics* 57.1 (Jan. 1986). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00268978600100081, pp. 97–121. ISSN: 0026-8976. DOI: `10.1080/00268978600100081`. URL: `https://doi.org/10.1080/00268978600100081` (visited on 08/16/2025).

[65] Josene M. Toldo et al. "Surface hopping modeling of charge and energy transfer in active environments". en. In: *Physical Chemistry Chemical Physics* 25.12 (2023). Publisher: Royal Society of Chemistry, pp. 8293–8316. DOI: `10.1039/D3CP00247K`. URL: `https://pubs.rsc.org/en/content/articlelanding/2023/cp/d3cp00247k` (visited on 08/26/2025).

[66] *6.6. Excited States Calculations - ORCA 6.0 Manual.* URL: `https://www.faccts.de/docs/orca/6.0/manual/contents/typical/excitedstates.html?utm_source=chatgpt.com` (visited on 08/27/2025).

[67] Lan Nguyen Tran, Jacqueline A. R. Shea, and Eric Neuscamman. "Tracking Excited States in Wave Function Optimization Using Density Matrices and Variational Principles". en. In: *Journal of Chemical Theory and Computation* 15.9 (Sept. 2019), pp. 4790–4803. ISSN: 1549-9618, 1549-9626. DOI: `10.1021/acs.jctc.9b00351`. URL: `https://pubs.acs.org/doi/10.1021/acs.jctc.9b00351` (visited on 08/27/2025).

[68] Felix Plasser and Leticia González. "Communication: Unambiguous comparison of many-electron wavefunctions through their overlaps". In: *The Journal of Chemical Physics* 145.2 (July 2016), p. 021103. ISSN: 0021-9606. DOI: `10.1063/1.4958462`. URL: `https://doi.org/10.1063/1.4958462` (visited on 08/17/2025).

[69] F. Plasser. "TheoDORE: A toolbox for a detailed and automated analysis of electronic excited state computations". In: *The Journal of Chemical Physics* 152.8 (Feb. 2020), p. 084108. ISSN: 0021-9606. DOI: `10.1063/1.5143076`. URL: `https://doi.org/10.1063/1.5143076` (visited on 08/17/2025).

[70] Felix Plasser, Anna I. Krylov, and Andreas Dreuw. "libwfa: Wavefunction analysis tools for excited and open-shell electronic states". en. In: *WIREs Computational Molecular Science* 12.4 (2022). _eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1595, e1595. ISSN: 1759-0884. DOI: `10.1002/wcms.1595`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1595` (visited on 08/17/2025).

[71] John C. Tully. "Molecular dynamics with electronic transitions". In: *The Journal of Chemical Physics* 93.2 (July 1990), pp. 1061–1071. ISSN: 0021-9606. DOI: `10.1063/1.459170`. URL: `https://doi.org/10.1063/1.459170` (visited on 08/17/2025).

[72] Aaron M. Virshup et al. “Photodynamics in Complex Environments: Ab Initio Multiple Spawning Quantum Mechanical/Molecular Mechanical Dynamics”. In: *The Journal of Physical Chemistry B* 113.11 (Mar. 2009). Publisher: American Chemical Society, pp. 3280–3291. ISSN: 1520-6106. DOI: `10.1021/jp8073464`. URL: `https://doi.org/10.1021/jp8073464` (visited on 08/17/2025).

[73] Hans-Dieter Meyer. “Studying molecular quantum dynamics with the multiconfiguration time-dependent Hartree method”. en. In: *WIREs Computational Molecular Science* 2.2 (2012). _eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.87, pp. 351–374. ISSN: 1759-0884. DOI: `10.1002/wcms.87`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.87` (visited on 08/17/2025).

[74] Saikat Mukherjee et al. “Simulations of molecular photodynamics in long timescales”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 380.2223 (Mar. 2022). Publisher: Royal Society, p. 20200382. DOI: `10.1098/rsta.2020.0382`. URL: `https://royalsocietypublishing.org/doi/full/10.1098/rsta.2020.0382` (visited on 08/17/2025).

[75] Péter P. Fehér, Ádám Madarász, and András Stirling. “Multiscale Modeling of Electronic Spectra Including Nuclear Quantum Effects”. In: *Journal of Chemical Theory and Computation* 17.10 (Oct. 2021). Publisher: American Chemical Society, pp. 6340–6352. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.1c00531`. URL: `https://doi.org/10.1021/acs.jctc.1c00531` (visited on 08/30/2025).

[76] T. J. Zuehlsdorff and C. M. Isborn. “Combining the ensemble and Franck-Condon approaches for calculating spectral shapes of molecules in solution”. In: *The Journal of Chemical Physics* 148.2 (Jan. 2018), p. 024110. ISSN: 0021-9606. DOI: `10.1063/1.5006043`. URL: `https://doi.org/10.1063/1.5006043` (visited on 08/30/2025).

[77] Gareth A. Tribello et al. “PLUMED 2: New feathers for an old bird”. In: *Computer Physics Communications* 185.2 (Feb. 2014), pp. 604–613. ISSN: 0010-4655. DOI: `10.1016/j.cpc.2013.09.018`. URL: `https://www.sciencedirect.com/science/article/pii/S0010465513003196` (visited on 08/24/2025).

[78] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. “Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method”. In: *Physical Review Letters* 100.2 (Jan. 2008). Publisher: American Physical Society, p. 020603. DOI: `10.1103/PhysRevLett.100.020603`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.100.020603` (visited on 08/15/2025).

[79] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory.* en. Courier Corporation, July 1996. ISBN: 978-0-486-69186-2.

[80] C. David Sherrill and Henry F. Schaefer. “The Configuration Interaction Method: Advances in Highly Correlated Approaches”. In: *Advances in Quantum Chemistry.* Ed. by Per-Olov Löwdin et al. Vol. 34. Academic Press, Jan. 1999, pp. 143–269. DOI: `10.1016/S0065-3276(08)60532-8`. URL: `https://www.sciencedirect.com/science/article/pii/S0065327608605328` (visited on 09/29/2025).

[81] Chr. Møller and M. S. Plesset. “Note on an Approximation Treatment for Many-Electron Systems”. In: *Physical Review* 46.7 (Oct. 1934). Publisher: American Physical Society, pp. 618–622. DOI: `10.1103/PhysRev.46.618`. URL: `https://link.aps.org/doi/10.1103/PhysRev.46.618` (visited on 09/29/2025).

[82] George D. Purvis III and Rodney J. Bartlett. “A full coupled-cluster singles and doubles model: The inclusion of disconnected triples”. In: *The Journal of Chemical Physics* 76.4 (Feb. 1982), pp. 1910–1918. ISSN: 0021-9606. DOI: `10.1063/1.443164`. URL: `https://doi.org/10.1063/1.443164` (visited on 09/29/2025).

[83] Krishnan Raghavachari et al. “A fifth-order perturbation comparison of electron correlation theories”. In: *Chemical Physics Letters* 157.6 (May 1989), pp. 479–483. ISSN: 0009-2614. DOI: `10.1016/S0009-2614(89)87395-6`. URL: `https://www.sciencedirect.com/science/article/pii/S0009261489873956` (visited on 09/29/2025).

[84] Frank Jensen. *Introduction to Computational Chemistry*. en. Google-Books-ID: UZOVDQAAQBAJ. John Wiley & Sons, Feb. 2017. ISBN: 978-1-118-82599-0.

[85] Robert G. Parr. “Density Functional Theory of Atoms and Molecules”. en. In: *Horizons of Quantum Chemistry*. Ed. by Kenichi Fukui and Bernard Pullman. Dordrecht: Springer Netherlands, 1980, pp. 5–15. ISBN: 978-94-009-9027-2. DOI: `10.1007/978-94-009-9027-2_2`.

[86] John P. Perdew and Karla Schmidt. “Jacob’s ladder of density functional approximations for the exchange-correlation energy”. In: *AIP Conference Proceedings* 577.1 (July 2001), pp. 1–20. ISSN: 0094-243X. DOI: `10.1063/1.1390175`. URL: `https://doi.org/10.1063/1.1390175` (visited on 07/24/2025).

[87] Klaus Capelle. “A bird’s-eye view of density-functional theory”. en. In: *Brazilian Journal of Physics* 36 (Dec. 2006). Publisher: Sociedade Brasileira de Física, pp. 1318–1343. ISSN: 0103-9733, 1678-4448. DOI: `https://doi.org/10.1590/S0103-97332006000700035`. URL: `https://www.scielo.br/j/bjp/a/dWXh7yLvsDSpZwzVv4wYcxF/` (visited on 07/25/2025).

[88] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized Gradient Approximation Made Simple”. In: *Physical Review Letters* 77.18 (Oct. 1996). Publisher: American Physical Society, pp. 3865–3868. DOI: `10.1103/PhysRevLett.77.3865`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.77.3865` (visited on 07/25/2025).

[89] Jianmin Tao et al. “Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids”. In: *Physical Review Letters* 91.14 (Sept. 2003). Publisher: American Physical Society, p. 146401. DOI: `10.1103/PhysRevLett.91.146401`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.91.146401` (visited on 07/25/2025).

[90] A. D. Becke. “Density-functional exchange-energy approximation with correct asymptotic behavior”. In: *Physical Review A* 38.6 (Sept. 1988). Publisher: American Physical Society, pp. 3098–3100. DOI: `10.1103/PhysRevA.38.3098`. URL: `https://link.aps.org/doi/10.1103/PhysRevA.38.3098` (visited on 07/25/2025).

[91] Chengteh Lee, Weitao Yang, and Robert G. Parr. “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density”. In: *Physical Review B* 37.2 (Jan. 1988). Publisher: American Physical Society, pp. 785–789. DOI: `10.1103/PhysRevB.37.785`. URL: `https://link.aps.org/doi/10.1103/PhysRevB.37.785` (visited on 07/25/2025).

[92] Axel D. Becke. “Density-functional thermochemistry. III. The role of exact exchange”. In: *The Journal of Chemical Physics* 98.7 (Apr. 1993), pp. 5648–5652. ISSN: 0021-9606. DOI: `10.1063/1.464913`. URL: `https://doi.org/10.1063/1.464913` (visited on 07/25/2025).

[93] Carlo Adamo and Vincenzo Barone. “Toward reliable density functional methods without adjustable parameters: The PBE0 model”. In: *The Journal of Chemical Physics* 110.13 (Apr. 1999), pp. 6158–6170. ISSN: 0021-9606. DOI: `10.1063/1.478522`. URL: `https://doi.org/10.1063/1.478522` (visited on 07/25/2025).

[94] Narbe Mardirossian and Martin Head-Gordon. “B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation”. In: *The Journal of Chemical Physics* 144.21 (June 2016), p. 214110. ISSN: 0021-9606. DOI: `10.1063/1.4952647`. URL: `https://doi.org/10.1063/1.4952647` (visited on 08/12/2025).

[95] Takeshi Yanai, David P Tew, and Nicholas C Handy. “A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP)”. In: *Chemical Physics Letters* 393.1 (July 2004), pp. 51–57. ISSN: 0009-2614. DOI: `10.1016/j.cplett.2004.06.011`. URL: `https://www.sciencedirect.com/science/article/pii/S0009261404008620` (visited on 07/25/2025).

[96] Stefan Grimme. “Semiempirical GGA-type density functional constructed with a long-range dispersion correction”. de. In: *Journal of Computational Chemistry* 27.15 (2006). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20495, pp. 1787–1799. ISSN: 1096-987X. DOI: `10.1002/jcc.20495`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20495` (visited on 07/25/2025).

[97] Oleg A. Vydrov and Troy Van Voorhis. “Nonlocal van der Waals density functional: The simpler the better”. In: *The Journal of Chemical Physics* 133.24 (Dec. 2010), p. 244103. ISSN: 0021-9606. DOI: `10.1063/1.3521275`. URL: `https://doi.org/10.1063/1.3521275` (visited on 07/25/2025).

[98] W. J. Hehre, R. F. Stewart, and J. A. Pople. “Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals”. In: *The Journal of Chemical Physics* 51.6 (Sept. 1969), pp. 2657–2664. ISSN: 0021-9606. DOI: `10.1063/1.1672392`. URL: `https://doi.org/10.1063/1.1672392` (visited on 07/25/2025).

[99] Thom H. Dunning Jr. “Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen”. In: *The Journal of Chemical Physics* 90.2 (Jan. 1989), pp. 1007–1023. ISSN: 0021-9606. DOI: `10.1063/1.456153`. URL: `https://doi.org/10.1063/1.456153` (visited on 07/25/2025).

[100] 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". en. In: *Physical Chemistry Chemical Physics* 7.18 (Aug. 2005). Publisher: The Royal Society of Chemistry, pp. 3297–3305. ISSN: 1463-9084. DOI: `10.1039/B508541A`. URL: `https://pubs.rsc.org/en/content/articlelanding/2005/cp/b508541a` (visited on 07/25/2025).

[101] Jingjing Zheng, Xuefei Xu, and Donald G. Truhlar. "Minimally augmented Karlsruhe basis sets". en. In: *Theoretical Chemistry Accounts* 128.3 (Feb. 2011), pp. 295–305. ISSN: 1432-2234. DOI: `10.1007/s00214-010-0846-z`. URL: `https://doi.org/10.1007/s00214-010-0846-z` (visited on 07/25/2025).

[102] Narbe Mardirossian and Martin Head-Gordon. "Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals". In: *Molecular Physics* 115.19 (Oct. 2017). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00268976.2017.1333644, pp. 2315–2372. ISSN: 0026-8976. DOI: `10.1080/00268976.2017.1333644`. URL: `https://doi.org/10.1080/00268976.2017.1333644` (visited on 07/30/2025).

[103] Fernand Spiegelman et al. "Density-functional tight-binding: basic concepts and applications to molecules and clusters". In: *Advances in Physics: X* 5.1 (Jan. 2020). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/23746149.2019.1710252, p. 1710252. ISSN: null. DOI: `10.1080/23746149.2019.1710252`. URL: `https://doi.org/10.1080/23746149.2019.1710252` (visited on 01/10/2024).

[104] D. Porezag et al. "Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon". In: *Physical Review B* 51.19 (May 1995). Publisher: American Physical Society, pp. 12947–12957. DOI: `10.1103/PhysRevB.51.12947`. URL: `https://link.aps.org/doi/10.1103/PhysRevB.51.12947` (visited on 07/25/2025).

[105] M. Elstner et al. "Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties". In: *Physical Review B* 58.11 (Sept. 1998). Publisher: American Physical Society, pp. 7260–7268. DOI: `10.1103/PhysRevB.58.7260`. URL: `https://link.aps.org/doi/10.1103/PhysRevB.58.7260` (visited on 07/24/2025).

[106] Michael Gaus et al. "Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications". In: *Journal of Chemical Theory and Computation* 10.4 (Apr. 2014). Publisher: American Chemical Society, pp. 1518–1537. ISSN: 1549-9618. DOI: `10.1021/ct401002w`. URL: `https://doi.org/10.1021/ct401002w` (visited on 07/24/2025).

[107] Xiya Lu et al. "Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications". In: *The Journal of Physical Chemistry B* 119.3 (Jan. 2015). Publisher: American Chemical Society, pp. 1062–1082. ISSN: 1520-6106. DOI: `10.1021/jp506557r`. URL: `https://doi.org/10.1021/jp506557r` (visited on 07/24/2025).

[108] Van Quan Vuong et al. “Parametrization and Benchmark of Long-Range Corrected DFTB2 for Organic Molecules”. In: *Journal of Chemical Theory and Computation* 14.1 (Jan. 2018). Publisher: American Chemical Society, pp. 115–125. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.7b00947`. URL: `https://doi.org/10.1021/acs.jctc.7b00947` (visited on 01/10/2024).

[109] Takao Tsuneda and Kimihiko Hirao. “Long-range correction for density functional theory”. en. In: *WIREs Computational Molecular Science* 4.4 (2014). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1178, pp. 375–390. ISSN: 1759-0884. DOI: `10.1002/wcms.1178`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1178` (visited on 01/10/2024).

[110] R. S. Mulliken. “Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I”. In: *The Journal of Chemical Physics* 23.10 (Oct. 1955), pp. 1833–1840. ISSN: 0021-9606. DOI: `10.1063/1.1740588`. URL: `https://doi.org/10.1063/1.1740588` (visited on 07/24/2025).

[111] Stefan Grimme et al. “A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu”. In: *The Journal of Chemical Physics* 132.15 (Apr. 2010), p. 154104. ISSN: 0021-9606. DOI: `10.1063/1.3382344`. URL: `https://doi.org/10.1063/1.3382344` (visited on 07/25/2025).

[112] Mayukh Kansari et al. “Mechanism of activation and autophosphorylation of a histidine kinase”. en. In: *Communications Chemistry* 7.1 (Sept. 2024). Publisher: Nature Publishing Group, p. 196. ISSN: 2399-3669. DOI: `10.1038/s42004-024-01272-6`. URL: `https://www.nature.com/articles/s42004-024-01272-6` (visited on 07/31/2025).

[113] Mark E. Casida. *Time-Dependent Density Functional Response Theory for Molecules | Recent Advances in Density Functional Methods*. en. Archive Location: world. DOI: `10.1142/9789812830586_0005`. URL: `https://www.worldscientific.com/doi/epdf/10.1142/9789812830586_0005` (visited on 07/24/2025).

[114] Carsten A. Ullrich and Zeng-hui Yang. “A brief compendium of time-dependent density-functional theory”. In: *Brazilian Journal of Physics* 44.1 (Feb. 2014). arXiv:1305.1388 [cond-mat, physics:physics], pp. 154–188. ISSN: 0103-9733, 1678-4448. DOI: `10.1007/s13538-013-0141-2`. URL: `http://arxiv.org/abs/1305.1388` (visited on 01/10/2024).

[115] Mark Abraham et al. “GROMACS 2023.2 Manual”. In: (July 2023). Publisher: Zenodo. DOI: `10.5281/zenodo.8134388`. URL: `https://zenodo.org/records/8134388` (visited on 01/10/2024).

[116] Wilfred F. van Gunsteren and Herman J. C. Berendsen. “Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry”. en. In: *Angewandte Chemie International Edition in English* 29.9 (1990). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.199009921, pp. 992–1023. ISSN: 1521-3773. DOI: `10.1002/anie.199009921`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.199009921` (visited on 07/27/2025).

[117] "Front Matter". In: *Understanding Molecular Simulation (Third Edition)*. Ed. by Daan Frenkel and Berend Smit. Academic Press, Jan. 2023, pp. i–iii. ISBN: 978-0-323-90292-2. DOI: `10.1016/B978-0-32-390292-2.00002-7`. URL: `https://www.sciencedirect.com/science/article/pii/B9780323902922000027` (visited on 07/25/2025).

[118] R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles*. Boca Raton: CRC Press, Mar. 2021. ISBN: 978-0-367-80693-4. DOI: `10.1201/9780367806934`.

[119] Ulrich Essmann et al. "A smooth particle mesh Ewald method". In: *The Journal of Chemical Physics* 103.19 (Nov. 1995), pp. 8577–8593. ISSN: 0021-9606. DOI: `10.1063/1.470117`. URL: `https://doi.org/10.1063/1.470117` (visited on 07/27/2025).

[120] Walter R. P. Scott et al. "The GROMOS Biomolecular Simulation Program Package". In: *The Journal of Physical Chemistry A* 103.19 (May 1999). Publisher: American Chemical Society, pp. 3596–3607. ISSN: 1089-5639. DOI: `10.1021/jp984217f`. URL: `https://doi.org/10.1021/jp984217f` (visited on 07/27/2025).

[121] Mark James Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 1-2 (Sept. 2015), pp. 19–25. ISSN: 2352-7110. DOI: `10.1016/j.softx.2015.06.001`. URL: `https://www.sciencedirect.com/science/article/pii/S2352711015000059` (visited on 08/16/2025).

[122] R. W Hockney, S. P Goel, and J. W Eastwood. "Quiet high-resolution computer models of a plasma". In: *Journal of Computational Physics* 14.2 (Feb. 1974), pp. 148–158. ISSN: 0021-9991. DOI: `10.1016/0021-9991(74)90010-2`. URL: `https://www.sciencedirect.com/science/article/pii/0021999174900102` (visited on 07/27/2025).

[123] Berk Hess et al. "LINCS: A linear constraint solver for molecular simulations". en. In: *Journal of Computational Chemistry* 18.12 (1997). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291096-987X%28199709%2918%3A12%3C1463%3A%3AAID-JCC4%3E3.0.CO%3B2-H, pp. 1463–1472. ISSN: 1096-987X. DOI: `10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-987X%28199709%2918%3A12%3C1463%3A%3AAID-JCC4%3E3.0.CO%3B2-H` (visited on 07/27/2025).

[124] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. en. Google-Books-ID: jyipEAAAQBAJ. Elsevier, July 2023. ISBN: 978-0-323-91318-8.

[125] H. J. C. Berendsen et al. "Molecular dynamics with coupling to an external bath". In: *The Journal of Chemical Physics* 81.8 (Oct. 1984), pp. 3684–3690. ISSN: 0021-9606. DOI: `10.1063/1.448118`. URL: `https://doi.org/10.1063/1.448118` (visited on 07/27/2025).

[126] Shuichi Nosé. “A unified formulation of the constant temperature molecular dynamics methods”. In: *The Journal of Chemical Physics* 81.1 (July 1984), pp. 511–519. ISSN: 0021-9606. DOI: `10.1063/1.447334`. URL: `https://doi.org/10.1063/1.447334` (visited on 07/27/2025).

[127] William G. Hoover. “Canonical dynamics: Equilibrium phase-space distributions”. In: *Physical Review A* 31.3 (Mar. 1985). Publisher: American Physical Society, pp. 1695–1697. DOI: `10.1103/PhysRevA.31.1695`. URL: `https://link.aps.org/doi/10.1103/PhysRevA.31.1695` (visited on 07/27/2025).

[128] M. Parrinello and A. Rahman. “Polymorphic transitions in single crystals: A new molecular dynamics method”. In: *Journal of Applied Physics* 52.12 (Dec. 1981), pp. 7182–7190. ISSN: 0021-8979. DOI: `10.1063/1.328693`. URL: `https://doi.org/10.1063/1.328693` (visited on 01/16/2024).

[129] B. Hourahine et al. “DFTB+, a software package for efficient approximate density functional theory based atomistic simulations”. In: *The Journal of Chemical Physics* 152.12 (Mar. 2020), p. 124101. ISSN: 0021-9606. DOI: `10.1063/1.5143190`. URL: `https://doi.org/10.1063/1.5143190` (visited on 07/27/2025).

[130] Gerrit Groenhof. “Introduction to QM/MM Simulations”. en. In: *Biomolecular Simulations: Methods and Protocols.* Ed. by Luca Monticelli and Emppu Salonen. Totowa, NJ: Humana Press, 2013, pp. 43–66. ISBN: 978-1-62703-017-5. DOI: `10.1007/978-1-62703-017-5_3`. URL: `https://doi.org/10.1007/978-1-62703-017-5_3` (visited on 07/27/2025).

[131] Sven Roßbach and Christian Ochsenfeld. “Influence of Coupling and Embedding Schemes on QM Size Convergence in QM/MM Approaches for the Example of a Proton Transfer in DNA”. In: *Journal of Chemical Theory and Computation* 13.3 (Mar. 2017). Publisher: American Chemical Society, pp. 1102–1107. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.6b00727`. URL: `https://doi.org/10.1021/acs.jctc.6b00727` (visited on 07/30/2025).

[132] Miquel Huix-Rotllant and Nicolas Ferré. “Analytic Energy, Gradient, and Hessian of Electrostatic Embedding QM/MM Based on Electrostatic Potential-Fitted Atomic Charges Scaling Linearly with the MM Subsystem Size”. In: *Journal of Chemical Theory and Computation* 17.1 (Jan. 2021). Publisher: American Chemical Society, pp. 538–548. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.0c01075`. URL: `https://doi.org/10.1021/acs.jctc.0c01075` (visited on 07/27/2025).

[133] William L. Jorgensen et al. “Comparison of simple potential functions for simulating liquid water”. In: *The Journal of Chemical Physics* 79.2 (July 1983), pp. 926–935. ISSN: 0021-9606. DOI: `10.1063/1.445869`. URL: `https://doi.org/10.1063/1.445869` (visited on 08/15/2025).

[134] Luca Maragliano et al. “String method in collective variables: Minimum free energy paths and isocommittor surfaces”. In: *The Journal of Chemical Physics* 125.2 (July 2006), p. 024106. ISSN: 0021-9606. DOI: `10.1063/1.2212942`. URL: `https://doi.org/10.1063/1.2212942` (visited on 09/08/2025).

[135] Davide Branduardi, Francesco Luigi Gervasio, and Michele Parrinello. “From A to B in free energy space”. In: *The Journal of Chemical Physics* 126.5 (Feb. 2007), p. 054103. ISSN: 0021-9606. DOI: `10.1063/1.2432340`. URL: `https://doi.org/10.1063/1.2432340` (visited on 09/08/2025).

[136] Jonathan A. Semelak et al. “Minimum Free Energy Pathways of Reactive Processes with Nudged Elastic Bands”. In: *Journal of Chemical Theory and Computation* 19.18 (Sept. 2023). Publisher: American Chemical Society, pp. 6273–6293. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.3c00366`. URL: `https://doi.org/10.1021/acs.jctc.3c00366` (visited on 09/08/2025).

[137] Graeme Henkelman and Hannes Jónsson. “Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points”. In: *The Journal of Chemical Physics* 113.22 (Dec. 2000), pp. 9978–9985. ISSN: 0021-9606. DOI: `10.1063/1.1323224`. URL: `https://doi.org/10.1063/1.1323224` (visited on 09/08/2025).

[138] Graeme Henkelman, Blas P. Uberuaga, and Hannes Jónsson. “A climbing image nudged elastic band method for finding saddle points and minimum energy paths”. In: *The Journal of Chemical Physics* 113.22 (Dec. 2000), pp. 9901–9904. ISSN: 0021-9606. DOI: `10.1063/1.1329672`. URL: `https://doi.org/10.1063/1.1329672` (visited on 09/08/2025).

[139] Matthias U. Bohner et al. “Nudged-elastic band used to find reaction coordinates based on the free energy”. In: *The Journal of Chemical Physics* 140.7 (Feb. 2014), p. 074109. ISSN: 0021-9606. DOI: `10.1063/1.4865220`. URL: `https://doi.org/10.1063/1.4865220` (visited on 09/08/2025).

[140] *waterfall - Waterfall plot - MATLAB*. de. URL: `https://www.mathworks.com/help/matlab/ref/waterfall.html` (visited on 09/08/2025).

[141] S Agatonovic-Kustrin and R Beresford. “Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research”. In: *Journal of Pharmaceutical and Biomedical Analysis* 22.5 (June 2000), pp. 717–727. ISSN: 0731-7085. DOI: `10.1016/S0731-7085(99)00272-1`. URL: `https://www.sciencedirect.com/science/article/pii/S0731708599002721` (visited on 01/11/2024).

[142] A.K. Jain, Jianchang Mao, and K.M. Mohiuddin. “Artificial neural networks: a tutorial”. In: *Computer* 29.3 (Mar. 1996). Conference Name: Computer, pp. 31–44. ISSN: 1558-0814. DOI: `10.1109/2.485891`. URL: `https://ieeexplore.ieee.org/abstract/document/485891` (visited on 01/11/2024).

[143] Vivienne Sze et al. “Efficient Processing of Deep Neural Networks: A Tutorial and Survey”. In: *Proceedings of the IEEE* 105.12 (Dec. 2017). Conference Name: Proceedings of the IEEE, pp. 2295–2329. ISSN: 1558-2256. DOI: `10.1109/JPROC.2017.2761740`. URL: `https://ieeexplore.ieee.org/abstract/document/8114708` (visited on 01/11/2024).

[144] I. Goodfellow. *Deep Learning. I. Goodfellow, Y. Bengio, A. Courville.* 2016.

[145] Barry J. Wythoff. "Backpropagation neural networks: A tutorial". In: *Chemometrics and Intelligent Laboratory Systems* 18.2 (Feb. 1993), pp. 115–155. ISSN: 0169-7439. DOI: `10.1016/0169-7439(93)80052-J`. URL: `https://www.sciencedirect.com/science/article/pii/016974399380052J` (visited on 01/11/2024).

[146] Warren S. McCulloch and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity". en. In: *The bulletin of mathematical biophysics* 5.4 (Dec. 1943), pp. 115–133. ISSN: 1522-9602. DOI: `10.1007/BF02478259`. URL: `https://doi.org/10.1007/BF02478259` (visited on 07/28/2025).

[147] G. Cybenko. "Approximation by superpositions of a sigmoidal function". en. In: *Mathematics of Control, Signals and Systems* 2.4 (Dec. 1989), pp. 303–314. ISSN: 1435-568X. DOI: `10.1007/BF02551274`. URL: `https://doi.org/10.1007/BF02551274` (visited on 07/28/2025).

[148] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". en. In: *Nature* 323.6088 (Oct. 1986). Publisher: Nature Publishing Group, pp. 533–536. ISSN: 1476-4687. DOI: `10.1038/323533a0`. URL: `https://www.nature.com/articles/323533a0` (visited on 07/28/2025).

[149] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. "Deep Sparse Rectifier Neural Networks". en. In: ().

[150] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs).* arXiv:1511.07289 [cs]. Feb. 2016. DOI: `10.48550/arXiv.1511.07289`. URL: `http://arxiv.org/abs/1511.07289` (visited on 01/10/2024).

[151] Christopher M. Bishop. *Pattern recognition and machine learning.* en. Information science and statistics. New York: Springer, 2006. ISBN: 978-0-387-31073-2.

[152] ROBERT Hecht-nielsen. "III.3 - Theory of the Backpropagation Neural Network**Based on "nonindent" by Robert Hecht-Nielsen, which appeared in Proceedings of the International Joint Conference on Neural Networks 1, 593–611, June 1989. © 1989 IEEE." In: *Neural Networks for Perception.* Ed. by Harry Wechsler. Academic Press, Jan. 1992, pp. 65–93. ISBN: 978-0-12-741252-8. DOI: `10.1016/B978-0-12-741252-8.50010-8`. URL: `https://www.sciencedirect.com/science/article/pii/B9780127412528500108` (visited on 01/11/2024).

[153] P.J. Werbos. "Backpropagation through time: what it does and how to do it". en. In: *Proceedings of the IEEE* 78.10 (Oct. 1990), pp. 1550–1560. ISSN: 00189219. DOI: `10.1109/5.58337`. URL: `http://ieeexplore.ieee.org/document/58337/` (visited on 01/11/2024).

[154] Dan Hendrycks and Kevin Gimpel. *Gaussian Error Linear Units (GELUs).* arXiv:1606.08415 [cs]. June 2023. DOI: `10.48550/arXiv.1606.08415`. URL: `http://arxiv.org/abs/1606.08415` (visited on 08/27/2025).

[155] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis.* en. Google-Books-ID: tCIgEAAAQBAJ. John Wiley & Sons, Feb. 2021. ISBN: 978-1-119-57875-8.

[156] Cj Willmott and K Matsuura. “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance”. en. In: *Climate Research* 30 (2005), pp. 79–82. ISSN: 0936-577X, 1616-1572. DOI: `10.3354/cr030079`. URL: `http://www.int-res.com/abstracts/cr/v30/n1/p79-82/` (visited on 08/23/2025).

[157] Rob J. Hyndman and Anne B. Koehler. “Another look at measures of forecast accuracy”. In: *International Journal of Forecasting* 22.4 (Oct. 2006), pp. 679–688. ISSN: 0169-2070. DOI: `10.1016/j.ijforecast.2006.03.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0169207006000239` (visited on 08/23/2025).

[158] James Bergstra and Yoshua Bengio. “Random search for hyper-parameter optimization”. In: *J. Mach. Learn. Res.* 13.null (Feb. 2012), pp. 281–305. ISSN: 1532-4435.

[159] Lisha Li et al. *Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization.* en. arXiv:1603.06560 [cs, stat]. June 2018. URL: `http://arxiv.org/abs/1603.06560` (visited on 01/11/2024).

[160] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables.* en. Google-Books-ID: MtU8uP7XMvoC. Courier Corporation, Jan. 1965. ISBN: 978-0-486-61272-0.

[161] Herbert A. David and Haikady N. Nagaraja. *Order Statistics.* en. Google-Books-ID: bdhzFXg6xFkC. John Wiley & Sons, Mar. 2004. ISBN: 978-0-471-65401-8.

[162] M. R. Leadbetter, G. Lindgren, and H. Rootzen. *Extremes and Related Properties of Random Sequences and Processes.* en. Springer Science & Business Media, Dec. 2012. ISBN: 978-1-4612-5449-2.

[163] William Feller. *An introduction to probability theory and its applications.* Vol. 963. Wiley New York, 1971. URL: `https://www.cambridge.org/core/journals/mathematical-gazette/article/mathematical-association/DF2D61900F68DD3EE37B6D4F7B9B5ABA` (visited on 09/10/2025).

[164] Peter J. Rousseeuw and Christophe Croux. “Alternatives to the Median Absolute Deviation”. In: *Journal of the American Statistical Association* 88.424 (Dec. 1993). Publisher: ASA Website _eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1993.10476408, pp. 1273–1283. ISSN: 0162-1459. DOI: `10.1080/01621459.1993.10476408`. URL: `https://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476408` (visited on 09/10/2025).

[165] Jörg Behler. “Four Generations of High-Dimensional Neural Network Potentials”. In: *Chemical Reviews* 121.16 (Aug. 2021). Publisher: American Chemical Society, pp. 10037–10072. ISSN: 0009-2665. DOI: `10.1021/acs.chemrev.0c00868`. URL: `https://doi.org/10.1021/acs.chemrev.0c00868` (visited on 01/18/2024).

[166] Jörg Behler and Gábor Csányi. “Machine learning potentials for extended systems: a perspective”. en. In: *The European Physical Journal B* 94.7 (July 2021), p. 142. ISSN: 1434-6036. DOI: `10.1140/epjb/s10051-021-00156-1`. URL: `https://doi.org/10.1140/epjb/s10051-021-00156-1` (visited on 07/30/2025).

[167] Jörg Behler. "Atom-centered symmetry functions for constructing high-dimensional neural network potentials". In: *The Journal of Chemical Physics* 134.7 (Feb. 2011), p. 074106. ISSN: 0021-9606. DOI: `10.1063/1.3553717`. URL: `https://doi.org/10.1063/1.3553717` (visited on 08/27/2025).

[168] Franco Scarselli et al. "The Graph Neural Network Model". In: *IEEE Transactions on Neural Networks* 20.1 (Jan. 2009), pp. 61–80. ISSN: 1941-0093. DOI: `10.1109/TNN.2008.2005605`. URL: `https://ieeexplore.ieee.org/abstract/document/4700287` (visited on 07/28/2025).

[169] Justin Gilmer et al. "Message Passing Neural Networks". en. In: *Machine Learning Meets Quantum Physics*. Ed. by Kristof T. Schütt et al. Cham: Springer International Publishing, 2020, pp. 199–214. ISBN: 978-3-030-40245-7. DOI: `10.1007/978-3-030-40245-7_10`. URL: `https://doi.org/10.1007/978-3-030-40245-7_10` (visited on 07/28/2025).

[170] Nathaniel Thomas et al. *Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds*. en. arXiv:1802.08219 [cs]. May 2018. DOI: `10.48550/arXiv.1802.08219`. URL: `http://arxiv.org/abs/1802.08219` (visited on 07/28/2025).

[171] Brandon Anderson, Truong Son Hy, and Risi Kondor. "Cormorant: Covariant Molecular Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc., 2019. URL: `https://proceedings.neurips.cc/paper/2019/hash/03573b32b2746e6e8ca98b9123f2249b-Abstract.html` (visited on 08/16/2025).

[172] Fabian Fuchs et al. "SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 1970–1981. URL: `https://proceedings.neurips.cc/paper/2020/hash/15231a7ce4ba789d13b722cc5c955834-Abstract.html` (visited on 08/16/2025).

[173] K. T. Schütt et al. "SchNet – A deep learning architecture for molecules and materials". In: *The Journal of Chemical Physics* 148.24 (Mar. 2018), p. 241722. ISSN: 0021-9606. DOI: `10.1063/1.5019779`. URL: `https://doi.org/10.1063/1.5019779` (visited on 07/28/2025).

[174] Justin Gilmer et al. "Neural Message Passing for Quantum Chemistry". en. In: *Proceedings of the 34th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, July 2017, pp. 1263–1272. URL: `https://proceedings.mlr.press/v70/gilmer17a.html` (visited on 08/26/2025).

[175] Kyunghyun Cho et al. *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*. en. arXiv:1406.1078 [cs]. Sept. 2014. DOI: `10.48550/arXiv.1406.1078`. URL: `http://arxiv.org/abs/1406.1078` (visited on 07/28/2025).

[176] James P. Darby et al. “Tensor-Reduced Atomic Density Representations”. In: *Physical Review Letters* 131.2 (July 2023). Publisher: American Physical Society, p. 028001. DOI: `10.1103/PhysRevLett.131.028001`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.131.028001` (visited on 08/26/2025).

[177] Philipp Pracht et al. “Efficient Composite Infrared Spectroscopy: Combining the Double-Harmonic Approximation with Machine Learning Potentials”. In: *Journal of Chemical Theory and Computation* 20.24 (Dec. 2024). Publisher: American Chemical Society, pp. 10986–11004. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.4c01157`. URL: `https://doi.org/10.1021/acs.jctc.4c01157` (visited on 08/26/2025).

[178] Elena Gelžinytė et al. “Transferable Machine Learning Interatomic Potential for Bond Dissociation Energy Prediction of Drug-like Molecules”. In: *Journal of Chemical Theory and Computation* 20.1 (Jan. 2024). Publisher: American Chemical Society, pp. 164–177. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.3c00710`. URL: `https://doi.org/10.1021/acs.jctc.3c00710` (visited on 08/26/2025).

[179] Wilfried J. Mortier, Swapan K. Ghosh, and S. Shankar. “Electronegativity-equalization method for the calculation of atomic charges in molecules”. en. In: *Journal of the American Chemical Society* 108.15 (July 1986), pp. 4315–4320. ISSN: 0002-7863, 1520-5126. DOI: `10.1021/ja00275a013`. URL: `https://pubs.acs.org/doi/abs/10.1021/ja00275a013` (visited on 08/16/2025).

[180] Anthony K. Rappe and William A. Goddard. “Charge equilibration for molecular dynamics simulations”. en. In: *The Journal of Physical Chemistry* 95.8 (Apr. 1991), pp. 3358–3363. ISSN: 0022-3654, 1541-5740. DOI: `10.1021/j100161a070`. URL: `https://pubs.acs.org/doi/abs/10.1021/j100161a070` (visited on 08/16/2025).

[181] Jiahao Chen and Todd J. Martínez. “QTPIE: Charge transfer with polarization current equalization. A fluctuating charge model with correct asymptotics”. In: *Chemical Physics Letters* 438.4 (Apr. 2007), pp. 315–320. ISSN: 0009-2614. DOI: `10.1016/j.cplett.2007.02.065`. URL: `https://www.sciencedirect.com/science/article/pii/S0009261407002618` (visited on 08/16/2025).

[182] Tsz Wai Ko et al. “A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer”. en. In: *Nature Communications* 12.1 (Jan. 2021). Publisher: Nature Publishing Group, p. 398. ISSN: 2041-1723. DOI: `10.1038/s41467-020-20427-2`. URL: `https://www.nature.com/articles/s41467-020-20427-2` (visited on 07/28/2025).

[183] Justin S. Smith et al. “Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning”. en. In: *Nature Communications* 10.1 (July 2019). Publisher: Nature Publishing Group, p. 2903. ISSN: 2041-1723. DOI: `10.1038/s41467-019-10827-4`. URL: `https://www.nature.com/articles/s41467-019-10827-4` (visited on 07/28/2025).

[184] K. T. Schütt et al. “Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions”. en. In: *Nature Communications* 10.1 (Nov. 2019). Publisher: Nature Publishing Group, p. 5024. ISSN: 2041-1723. DOI:

10.1038/s41467-019-12875-2. URL: https://www.nature.com/articles/s41467-019-12875-2 (visited on 07/31/2025).

[185] Peter Zaspel et al. "Boosting Quantum Machine Learning Models with a Multilevel Combination Technique: Pople Diagrams Revisited". In: *Journal of Chemical Theory and Computation* 15.3 (Mar. 2019). Publisher: American Chemical Society, pp. 1546–1559. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.8b00832. URL: https://doi.org/10.1021/acs.jctc.8b00832 (visited on 07/28/2025).

[186] Kresten Lindorff-Larsen et al. "Improved side-chain torsion potentials for the Amber ff99SB protein force field". en. In: *Proteins: Structure, Function, and Bioinformatics* 78.8 (2010). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.22711, pp. 1950–1958. ISSN: 1097-0134. DOI: 10.1002/prot.22711. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.22711 (visited on 08/15/2025).

[187] Mayukh Kansari, Lena Eichinger, and Tomáš Kubař. "Extended-sampling QM/MM simulation of biochemical reactions involving P–N bonds". en. In: *Physical Chemistry Chemical Physics* 25.14 (2023). Publisher: Royal Society of Chemistry (RSC), pp. 9824–9836. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/d2cp05890a. URL: https://xlink.rsc.org/?DOI=D2CP05890A (visited on 07/25/2025).

[188] Alessandro Laio and Michele Parrinello. "Escaping free-energy minima". In: *Proceedings of the National Academy of Sciences* 99.20 (Oct. 2002). Publisher: Proceedings of the National Academy of Sciences, pp. 12562–12566. DOI: 10.1073/pnas.202427399. URL: https://www.pnas.org/doi/abs/10.1073/pnas.202427399 (visited on 08/15/2025).

[189] Markus Bursch et al. "Best-Practice DFT Protocols for Basic Molecular Computational Chemistry". en. In: *Angewandte Chemie* 134.42 (2022). https://onlinelibrary.wiley.com/doi/pdf/10.1002/ange.202205735, e202205735. ISSN: 1521-3757. DOI: 10.1002/ange.202205735. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ange.202205735 (visited on 08/15/2025).

[190] Dénes Berta et al. "Mechanism-Based Redesign of GAP to Activate Oncogenic Ras". In: *Journal of the American Chemical Society* 145.37 (Sept. 2023). Publisher: American Chemical Society, pp. 20302–20310. ISSN: 0002-7863. DOI: 10.1021/jacs.3c04330. URL: https://doi.org/10.1021/jacs.3c04330 (visited on 08/15/2025).

[191] Jörg Behler. "Perspective: Machine learning potentials for atomistic simulations". In: *The Journal of Chemical Physics* 145.17 (Nov. 2016), p. 170901. ISSN: 0021-9606. DOI: 10.1063/1.4966192. URL: https://doi.org/10.1063/1.4966192 (visited on 07/28/2025).

[192] Claudia L. Gómez-Flores et al. "Accurate Free Energies for Complex Condensed-Phase Reactions Using an Artificial Neural Network Corrected DFTB/MM Methodology". In: *Journal of Chemical Theory and Computation* 18.2 (Feb. 2022). Publisher: American Chemical Society, pp. 1213–1226. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.1c00811. URL: https://doi.org/10.1021/acs.jctc.1c00811 (visited on 07/25/2025).

[193] Kristof T. Schütt et al. “Quantum-chemical insights from deep tensor neural networks”. en. In: *Nature Communications* 8.1 (Jan. 2017). Publisher: Nature Publishing Group, p. 13890. ISSN: 2041-1723. DOI: `10.1038/ncomms13890`. URL: `https://www.nature.com/articles/ncomms13890` (visited on 07/28/2025).

[194] B Brooks and M Karplus. “Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor.” In: *Proceedings of the National Academy of Sciences* 80.21 (Nov. 1983). Publisher: Proceedings of the National Academy of Sciences, pp. 6571–6575. DOI: `10.1073/pnas.80.21.6571`. URL: `https://www.pnas.org/doi/abs/10.1073/pnas.80.21.6571` (visited on 08/24/2025).

[195] Jianpeng Ma. “Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes”. English. In: *Structure* 13.3 (Mar. 2005). Publisher: Elsevier, pp. 373–380. ISSN: 0969-2126. DOI: `10.1016/j.str.2005.02.002`. URL: `https://www.cell.com/structure/abstract/S0969-2126(05)00065-1` (visited on 08/24/2025).

[196] Andrea Amadei, Antonius B. M. Linssen, and Herman J. C. Berendsen. “Essential dynamics of proteins”. en. In: *Proteins: Structure, Function, and Bioinformatics* 17.4 (1993). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.340170408, pp. 412–425. ISSN: 1097-0134. DOI: `10.1002/prot.340170408`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.340170408` (visited on 08/24/2025).

[197] Charles C. David and Donald J. Jacobs. “Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins”. en. In: *Protein Dynamics: Methods and Protocols*. Ed. by Dennis R. Livesay. Totowa, NJ: Humana Press, 2014, pp. 193–226. ISBN: 978-1-62703-658-0. DOI: `10.1007/978-1-62703-658-0_11`. URL: `https://doi.org/10.1007/978-1-62703-658-0_11` (visited on 08/24/2025).

[198] Berthold K. P. Horn. “Closed-form solution of absolute orientation using unit quaternions”. EN. In: *JOSA A* 4.4 (Apr. 1987). Publisher: Optica Publishing Group, pp. 629–642. ISSN: 1520-8532. DOI: `10.1364/JOSAA.4.000629`. URL: `https://opg.optica.org/josaa/abstract.cfm?uri=josaa-4-4-629` (visited on 08/24/2025).

[199] Aik Rui Tan, Johannes C. B. Dietschreit, and Rafael Gómez-Bombarelli. “Enhanced sampling of robust molecular datasets with uncertainty-based collective variables”. In: *The Journal of Chemical Physics* 162.3 (Jan. 2025), p. 034114. ISSN: 0021-9606. DOI: `10.1063/5.0246178`. URL: `https://doi.org/10.1063/5.0246178` (visited on 08/25/2025).

[200] Burr Settles. “From Theories to Queries: Active Learning in Practice”. en. In: *Active Learning and Experimental Design workshop In conjunction with AISTATS 2010*. ISSN: 1938-7228. JMLR Workshop and Conference Proceedings, Apr. 2011, pp. 1–18. URL: `https://proceedings.mlr.press/v16/settles11a.html` (visited on 09/04/2025).

[201] David J. C. MacKay. “Information-Based Objective Functions for Active Data Selection”. In: *Neural Computation* 4.4 (July 1992), pp. 590–604. ISSN: 0899-7667. DOI: `10.1162/neco.1992.4.4.590`. URL: `https://doi.org/10.1162/neco.1992.4.4.590` (visited on 09/04/2025).

[202] D. A. Cohn, Z. Ghahramani, and M. I. Jordan. “Active Learning with Statistical Models”. en. In: *Journal of Artificial Intelligence Research* 4 (Mar. 1996), pp. 129–145. ISSN: 1076-9757. DOI: `10.1613/jair.295`. URL: `https://www.jair.org/index.php/jair/article/view/10158` (visited on 09/04/2025).

[203] Maximilian Kubillus et al. “Parameterization of the DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems”. In: *Journal of Chemical Theory and Computation* 11.1 (Jan. 2015). Publisher: American Chemical Society, pp. 332–342. ISSN: 1549-9618. DOI: `10.1021/ct5009137`. URL: `https://doi.org/10.1021/ct5009137` (visited on 08/24/2025).

[204] Adrien Lesage et al. “Smoothed Biasing Forces Yield Unbiased Free Energies with the Extended-System Adaptive Biasing Force Method”. In: *The Journal of Physical Chemistry B* 121.15 (Apr. 2017). Publisher: American Chemical Society, pp. 3676–3685. ISSN: 1520-6106. DOI: `10.1021/acs.jpcb.6b10055`. URL: `https://doi.org/10.1021/acs.jpcb.6b10055` (visited on 08/25/2025).

[205] Jeffrey Comer et al. “The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask”. In: *The Journal of Physical Chemistry B* 119.3 (Jan. 2015). Publisher: American Chemical Society, pp. 1129–1151. ISSN: 1520-6106. DOI: `10.1021/jp506633n`. URL: `https://doi.org/10.1021/jp506633n` (visited on 08/25/2025).

[206] Yinglong Miao, Victoria A. Feher, and J. Andrew McCammon. “Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation”. In: *Journal of Chemical Theory and Computation* 11.8 (Aug. 2015). Publisher: American Chemical Society, pp. 3584–3595. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.5b00436`. URL: `https://doi.org/10.1021/acs.jctc.5b00436` (visited on 08/25/2025).

[207] Ask Hjorth Larsen et al. “The atomic simulation environment—a Python library for working with atoms”. en. In: *Journal of Physics: Condensed Matter* 29.27 (June 2017). Publisher: IOP Publishing, p. 273002. ISSN: 0953-8984. DOI: `10.1088/1361-648X/aa680e`. URL: `https://dx.doi.org/10.1088/1361-648X/aa680e` (visited on 09/04/2025).

[208] Erik Bitzek et al. “Structural Relaxation Made Simple”. In: *Physical Review Letters* 97.17 (Oct. 2006). Publisher: American Physical Society, p. 170201. DOI: `10.1103/PhysRevLett.97.170201`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.97.170201` (visited on 08/25/2025).

[209] Tony Hunter. “Signaling—2000 and Beyond”. English. In: *Cell* 100.1 (Jan. 2000). Publisher: Elsevier, pp. 113–127. ISSN: 0092-8674, 1097-4172. DOI: `10.1016/S0092-8674(00)81688-8`. URL: `https://www.cell.com/cell/abstract/S0092-8674(00)81688-8` (visited on 07/31/2025).

[210] Peter J Kennelly. “Protein kinases and protein phosphatases in prokaryotes: a genomic perspective”. In: *FEMS Microbiology Letters* 206.1 (Jan. 2002), pp. 1–8. ISSN: 0378-1097. DOI: `10.1111/j.1574-6968.2002.tb10978.x`. URL: `https://doi.org/10.1111/j.1574-6968.2002.tb10978.x` (visited on 07/31/2025).

[211] Ann M. Stock, Victoria L. Robinson, and Paul N. Goudreau. “Two-Component Signal Transduction”. en. In: *Annual Review of Biochemistry* 69.Volume 69, 2000 (July 2000). Publisher: Annual Reviews, pp. 183–215. ISSN: 0066-4154, 1545-4509. DOI: `10.1146/annurev.biochem.69.1.183`. URL: `https://www.annualreviews.org/content/journals/10.1146/annurev.biochem.69.1.183` (visited on 07/31/2025).

[212] John S. Parkinson and Eric C. Kofoid. “COMMUNICATION MODULES IN BACTERIAL SIGNALING PROTEINS”. en. In: *Annual Review of Genetics* 26.Volume 26, 1992 (Dec. 1992). Publisher: Annual Reviews, pp. 71–112. ISSN: 0066-4197, 1545-2948. DOI: `10.1146/annurev.ge.26.120192.000443`. URL: `https://www.annualreviews.org/content/journals/10.1146/annurev.ge.26.120192.000443#` (visited on 07/31/2025).

[213] W. Ross Ellington. “Evolution and Physiological Roles of Phosphagen Systems”. en. In: *Annual Review of Physiology* 63.Volume 63, 2001 (Mar. 2001). Publisher: Annual Reviews, pp. 289–325. ISSN: 0066-4278, 1545-1585. DOI: `10.1146/annurev.physiol.63.1.289`. URL: `https://www.annualreviews.org/content/journals/10.1146/annurev.physiol.63.1.289` (visited on 07/31/2025).

[214] Brenda Martins Vasconcellos et al. “A comprehensive review of arginine kinase proteins: What we need to know?” In: *Biochemistry and Biophysics Reports* 40 (Dec. 2024), p. 101837. ISSN: 2405-5808. DOI: `10.1016/j.bbrep.2024.101837`. URL: `https://www.sciencedirect.com/science/article/pii/S2405580824002012` (visited on 07/31/2025).

[215] Markus Wyss and Rima Kaddurah-Daouk. “Creatine and Creatinine Metabolism”. In: *Physiological Reviews* 80.3 (July 2000). Publisher: American Physiological Society, pp. 1107–1213. ISSN: 0031-9333. DOI: `10.1152/physrev.2000.80.3.1107`. URL: `https://journals.physiology.org/doi/full/10.1152/physrev.2000.80.3.1107` (visited on 07/31/2025).

[216] T Wallimann et al. “Intracellular compartmentation, structure and function of creatine kinase isoenzymes in tissues with high and fluctuating energy demands: the ’phosphocreatine circuit’ for cellular energy homeostasis.” In: *Biochemical Journal* 281.Pt 1 (Jan. 1992), pp. 21–40. ISSN: 0264-6021. DOI: `10.1042/bj2810021`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1130636/` (visited on 07/31/2025).

[217] Kouji Uda et al. “Evolution of the arginine kinase gene family”. In: *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 1.2 (June 2006), pp. 209–218. ISSN: 1744-117X. DOI: `10.1016/j.cbd.2005.10.007`. URL: `https://www.sciencedirect.com/science/article/pii/S1744117X05000250` (visited on 07/31/2025).

[218] Orlando Acevedo and William L. Jorgensen. “Advances in Quantum and Molecular Mechanical (QM/MM) Simulations for Organic and Enzymatic Reactions”. In: *Accounts of Chemical Research* 43.1 (Jan. 2010). Publisher: American Chemical Society, pp. 142–151. ISSN: 0001-4842. DOI: `10.1021/ar900171c`. URL: `https://doi.org/10.1021/ar900171c` (visited on 07/31/2025).

[219] Jinzhe Zeng et al. "Development of Range-Corrected Deep Learning Potentials for Fast, Accurate Quantum Mechanical/Molecular Mechanical Simulations of Chemical Reactions in Solution". en. In: *Journal of Chemical Theory and Computation* 17.11 (Nov. 2021), pp. 6993–7009. ISSN: 1549-9618, 1549-9626. DOI: `10.1021/acs.jctc.1c00201`. URL: `https://pubs.acs.org/doi/10.1021/acs.jctc.1c00201` (visited on 07/30/2025).

[220] Timothy J. Giese, Jinzhe Zeng, and Darrin M. York. "Transferability of MACE Graph Neural Network for Range Corrected -Machine Learning Potential QM/MM Applications". en. In: *The Journal of Physical Chemistry B* 129.22 (June 2025), pp. 5477–5490. ISSN: 1520-6106, 1520-5207. DOI: `10.1021/acs.jpcb.5c02006`. URL: `https://pubs.acs.org/doi/10.1021/acs.jpcb.5c02006` (visited on 07/30/2025).

[221] Manuel Enns. "Fluorescence Analysis of FR0 in Complex Environments with Quantum Mechanical and Machine Learning Approaches". English. MA thesis. Karlsruhe Institute of Technology, Oct. 2023.

[222] Béryl Na{\"e}mi Greb. "Training of {NN} for the Prediction of 11-cis-Retinylidene Protein Absorption Spectra". English. Bachelor's thesis. Karlsruhe Institute of Technology, July 2024.

[223] Lara Maria Nagel. "Trainingsdatengenerierung des neuronalen Netzes von 11-cis-Retinal basierend auf der Verwendung von {QM/MM}-Simulationen". German. Bachelor's thesis. Karlsruher Institut für Technologie, Sept. 2024.

[224] Jemina Enkelmann. "Training Data Generation for Neural Networks describing all-trans Retinal using QM/MM simulations". English. Bachelor's thesis. Karlsruhe Institute of Technology, Sept. 2024.

[225] Oliver P. Ernst et al. "Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms". In: *Chemical Reviews* 114.1 (Jan. 2014). Publisher: American Chemical Society, pp. 126–163. ISSN: 0009-2665. DOI: `10.1021/cr4003769`. URL: `https://doi.org/10.1021/cr4003769` (visited on 08/18/2025).

[226] Franziska Schneider, Christiane Grimm, and Peter Hegemann. "Biophysics of Channelrhodopsin". en. In: *Annual Review of Biophysics* 44.Volume 44, 2015 (June 2015). Publisher: Annual Reviews, pp. 167–186. ISSN: 1936-122X, 1936-1238. DOI: `10.1146/annurev-biophys-060414-034014`. URL: `https://www.annualreviews.org/content/journals/10.1146/annurev-biophys-060414-034014` (visited on 08/18/2025).

[227] Kazumasa Oda et al. "Crystal structure of the red light-activated channelrhodopsin Chrimson". en. In: *Nature Communications* 9.1 (Sept. 2018). Publisher: Nature Publishing Group, p. 3949. ISSN: 2041-1723. DOI: `10.1038/s41467-018-06421-9`. URL: `https://www.nature.com/articles/s41467-018-06421-9` (visited on 08/18/2025).

[228] Katharina Spies, Beatrix M. Bold, and Marcus Elstner. “Complex active site structures influence absorption spectrum of Chrimson wild type and mutants”. en. In: *Physical Chemistry Chemical Physics* 27.25 (2025). Publisher: Royal Society of Chemistry, pp. 13360–13370. DOI: `10.1039/D5CP00762C`. URL: `https://pubs.rsc.org/en/content/articlelanding/2025/cp/d5cp00762c` (visited on 10/09/2025).

[229] Yanan Guo et al. “Active site structure and absorption spectrum of channelrhodopsin-2 wild-type and C128T mutant”. en. In: *Chemical Science* 7.6 (2016). Publisher: Royal Society of Chemistry, pp. 3879–3891. DOI: `10.1039/C6SC00468G`. URL: `https://pubs.rsc.org/en/content/articlelanding/2016/sc/c6sc00468g` (visited on 08/18/2025).

[230] Michael Hoffmann et al. “Color Tuning in Rhodopsins: The Mechanism for the Spectral Shift between Bacteriorhodopsin and Sensory Rhodopsin II”. In: *Journal of the American Chemical Society* 128.33 (Aug. 2006). Publisher: American Chemical Society, pp. 10808–10818. ISSN: 0002-7863. DOI: `10.1021/ja062082i`. URL: `https://doi.org/10.1021/ja062082i` (visited on 08/18/2025).

[231] Gerd G. Kochendoerfer et al. “Resonance Raman Examination of the Wavelength Regulation Mechanism in Human Visual Pigments”. In: *Biochemistry* 36.22 (June 1997). Publisher: American Chemical Society, pp. 6577–6587. ISSN: 0006-2960. DOI: `10.1021/bi970322o`. URL: `https://doi.org/10.1021/bi970322o` (visited on 08/18/2025).

[232] Marius Wanko et al. “Effect of Polarization on the Opsin Shift in Rhodopsins. 2. Empirical Polarization Models for Proteins”. In: *The Journal of Physical Chemistry B* 112.37 (Sept. 2008). Publisher: American Chemical Society, pp. 11468–11478. ISSN: 1520-6106. DOI: `10.1021/jp802409k`. URL: `https://doi.org/10.1021/jp802409k` (visited on 08/18/2025).

[233] Marius Wanko et al. “Effect of Polarization on the Opsin Shift in Rhodopsins. 1. A Combined QM/QM/MM Model for Bacteriorhodopsin and Pharaonis Sensory Rhodopsin II”. In: *The Journal of Physical Chemistry B* 112.37 (Sept. 2008). Publisher: American Chemical Society, pp. 11462–11467. ISSN: 1520-6106. DOI: `10.1021/jp802408g`. URL: `https://doi.org/10.1021/jp802408g` (visited on 08/18/2025).

[234] Julia Westermayr, Michael Gastegger, and Philipp Marquetand. “Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics”. In: *The Journal of Physical Chemistry Letters* 11.10 (May 2020). Publisher: American Chemical Society, pp. 3828–3834. DOI: `10.1021/acs.jpclett.0c00527`. URL: `https://doi.org/10.1021/acs.jpclett.0c00527` (visited on 08/18/2025).

[235] Sascha Mausenberger et al. “S pai NN: equivariant message passing for excited-state nonadiabatic molecular dynamics”. en. In: *Chemical Science* 15.38 (2024). Publisher: Royal Society of Chemistry, pp. 15880–15890. DOI: `10.1039/D4SC04164J`. URL: `https://pubs.rsc.org/en/content/articlelanding/2024/sc/d4sc04164j` (visited on 08/18/2025).

[236] Amanda Arcidiacono et al. “Predicting Solvatochromism of Chromophores in Proteins through QM/MM and Machine Learning”. In: *The Journal of Physical Chemistry A* 128.18 (May 2024). Publisher: American Chemical Society, pp. 3646–3658. ISSN: 1089-5639. DOI: `10.1021/acs.jpca.4c00249`. URL: `https://doi.org/10.1021/acs.jpca.4c00249` (visited on 08/18/2025).

[237] Monja Sokolov et al. “Non-adiabatic molecular dynamics simulations provide new insights into the exciton transfer in the Fenna–Matthews–Olson complex”. en. In: *Physical Chemistry Chemical Physics* 26.28 (2024). Publisher: Royal Society of Chemistry, pp. 19469–19496. DOI: `10.1039/D4CP02116A`. URL: `https://pubs.rsc.org/en/content/articlelanding/2024/cp/d4cp02116a` (visited on 08/18/2025).

[238] Patrizia Mazzeo et al. “Electrostatic embedding machine learning for ground and excited state molecular dynamics of solvated molecules”. en. In: *Digital Discovery* 3.12 (2024). Publisher: Royal Society of Chemistry, pp. 2560–2571. DOI: `10.1039/D4DD00295D`. URL: `https://pubs.rsc.org/en/content/articlelanding/2024/dd/d4dd00295d` (visited on 08/18/2025).

[239] Elena Betti et al. “Insights into Energy Transfer in Light-Harvesting Complex II Through Machine-Learning Assisted Simulations”. In: *The Journal of Physical Chemistry B* 128.21 (May 2024). Publisher: American Chemical Society, pp. 5188–5200. ISSN: 1520-6106. DOI: `10.1021/acs.jpcb.4c01494`. URL: `https://doi.org/10.1021/acs.jpcb.4c01494` (visited on 08/18/2025).

[240] M. Eichinger et al. “A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields”. In: *The Journal of Chemical Physics* 110.21 (June 1999), pp. 10452–10467. ISSN: 0021-9606. DOI: `10.1063/1.479049`. URL: `https://doi.org/10.1063/1.479049` (visited on 10/09/2025).

[241] Hiroshi C. Watanabe et al. “Towards an Understanding of Channelrhodopsin Function: Simulations Lead to Novel Insights of the Channel Mechanism”. In: *Journal of Molecular Biology* 425.10 (May 2013), pp. 1795–1814. ISSN: 0022-2836. DOI: `10.1016/j.jmb.2013.01.033`. URL: `https://www.sciencedirect.com/science/article/pii/S0022283613000648` (visited on 10/09/2025).

[242] *Mechanism by which Untwisting of Retinal Leads to Productive Bacteriorhodopsin Photocycle States | The Journal of Physical Chemistry B*. URL: `https://pubs.acs.org/doi/full/10.1021/jp505818r` (visited on 10/09/2025).

[243] Dominik Hachenthal. “Enhanced Molecular Sampling Techniques for Neural Network Models: Application to Fluorescent Molecules”. English. MA thesis. Karlsruhe Institute of Technology, Feb. 2025.

[244] Pascal Schmidt. “Uncertainty Sampling as an Enhanced Molecular Sampling Technique for Neural Network Models”. English. MA thesis. Karlsruhe Institute of Technology, July 2025.

[245] Anna Ulrich. “Comparison of Adiabatic and Non-Adiabatic Computational Approaches to Organic Fluorophores”. MA thesis. Karlsruhe Institute of Technology, Oct. 2024.

[246] Junmei Wang et al. “Development and testing of a general amber force field”. In: *Journal of Computational Chemistry* 25.9 (July 2004). Publisher: John Wiley & Sons, Ltd, pp. 1157–1174. ISSN: 0192-8651. DOI: `10.1002/jcc.20035`. URL: `https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.20035` (visited on 10/10/2025).

[247] Junmei Wang et al. “Automatic atom type and bond type perception in molecular mechanical calculations”. In: *Journal of Molecular Graphics and Modelling* 25.2 (Oct. 2006), pp. 247–260. ISSN: 1093-3263. DOI: `10.1016/j.jmgm.2005.12.005`. URL: `https://www.sciencedirect.com/science/article/pii/S1093326305001737` (visited on 10/10/2025).

[248] Christopher I. Bayly et al. “A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model”. en. In: *The Journal of Physical Chemistry* 97.40 (Oct. 1993), pp. 10269–10280. ISSN: 0022-3654, 1541-5740. DOI: `10.1021/j100142a004`. URL: `https://pubs.acs.org/doi/abs/10.1021/j100142a004` (visited on 10/10/2025).

[249] Carl Caleman et al. “Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant”. In: *Journal of Chemical Theory and Computation* 8.1 (Jan. 2012). Publisher: American Chemical Society, pp. 61–74. ISSN: 1549-9618. DOI: `10.1021/ct200731v`. URL: `https://doi.org/10.1021/ct200731v` (visited on 10/10/2025).

A. Appendix

Table A.1.: Key settings for sampling phosphorylation datasets.

Component	Setting
MM model	TIP3P water; AMBER99SB-ILDN for HK protein embedding
QM (sampling)	DFTB3 (3ob) + SRP-OPhyd / SRP-PNmod; electrostatic embedding
Metadynamics CVs	Forming P–N and breaking P–O distances (concerted & stepwise covered)
WT-meta params	Height 1.2 kJ mol^{-1}; width 0.004 nm; stride 0.5 ps; bias factor 50
Equilibration	Minimization + NVT (Nosé–Hoover) to calibrate CV fluctuations
MFEP sampling	Uniform along MFEP (reactant/TS/product balanced); plateaus included
QM cutout	Nucleophile + phosphorus donor/leaver + Mg^{2+} (if present); linker only in HK
DFT level	ωB97M-V / ma-def2-TZVP; 0 K occupations (no smearing)
DFTB level	3ob + SRP-OPhyd/PNmod; Fermi $T = 300$ K (500 K tested); Mermin free energy used
Convergence policy	Keep frames only if both DFT and DFTB converge; no HOMO–LUMO gap filtering
Label(s)	Primary: $\Delta E = E_{\text{atom}}^{\text{DFT}} - E_{\text{atom}}^{\text{DFTB}}$; optional: $\Delta\mathbf{F}$ and ΔE_{raw}
Splitting	Balanced, group-aware by reaction family and charge state; identical splits across views

Table A.2.: MD run control and output (GROMACS .mdp).

Parameter	Value
Integrator	`md`
Time step	1 fs (`dt = 0.001`)
No. of steps	10,000,000 (≈ 10 ns)
COM motion removal	`comm-mode = Linear`, `nstcomm = 20`
Coordinates (compressed)	`nstxout-compressed = 20`
Energies / log frequency	`nstenergy = 20`, `nstlog = 20`

Table A.3.: Nonbonded settings and periodic boundary conditions.

Parameter	Value
Cutoff scheme	`Verlet`
Neighbor list update	`nstlist = 10`
Real-space cutoff list	`rlist = 1.0 nm`
Electrostatics	`PME`, `rcoulomb = 1.0 nm`
van der Waals	`Cut-off`, `rvdw = 1.0 nm`
Dispersion correction	`No`
PBC	`pbc = xyz`

Table A.4.: Thermostat and ensemble.

Parameter	Value
Thermostat	`Nosé-Hoover` (`Tcoupl = Nose-Hoover`)
Temperature	300 K (`ref-t = 300`)
Coupling time	0.5 ps (`tau-t = 0.5`)
Groups / stride	`tc-grps = System`, `nsttcouple = 1`
Barostat	– (NVT ensemble)

Table A.5.: QM/MM settings in the MD engine.

Parameter	Value
QM/MM	enabled (`QMMM = yes`)
QM region	phosphoryl moiety (generic subset)
Total QM charge	−2 (`QMcharge = -2`)
Spin multiplicity	set per state (left unspecified in `.mdp`)
MM charge scaling	`MMChargeScaleFactor = 1`
Note	External DFTB+ used for QM (see Table A.6)

Table A.6.: DFTB+ settings for the QM region during metadynamics.

Setting	Value
SCC / order	SCC-DFTB3 (`SCC = Yes`, `ThirdOrderFull = Yes`)
Parameter set	3OB with P/N extension (H, C, N, O, P)
Charge	−2 (`Charge = -2`)
Max SCC iterations	250
Temperature (Fermi filling)	300 K
H-correction	Damping, exponent 4.0
Max angular momentum	H:s; C/N/O:p; P:d
Hubbard derivatives	C −0.1492, N −0.1535, O −0.1575, H −0.1857, P −0.14
Analysis	`CalculateForces = Yes`
Slater–Koster mapping	`Type2FileNames` scheme (paths omitted)

Table A.7.: Well-tempered-style metadynamics keywords (PLUMED).

Keyword	Value
Collective variables	`ARG = dist_P-N, dist_P-O`
Gaussian pace	`PACE = 200` (MD steps)
Gaussian height	`HEIGHT = 0.5` (engine energy units)
Gaussian widths	`SIGMA = 0.02, 0.02` (distance units)
HILLS output	`FILE = HILLS`
Multi-walkers	`WALKERS_N = 16`
Restart stride	`WALKERS_RSTRIDE = 500`

Table A.8.: Reference DFT level for training data (ORCA).

Item	Specification
Functional / dispersion	ωB97M-V (VV10 nonlocal correlation)
Basis set	ma-def2-TZVP
Quantity	analytic gradients (`Engrad`)
Population analysis	Hirshfeld

Table A.9.: DFTB+ single-point setup for Δ-learning (DFT − DFTB).

Setting	Value
Hamiltonian / SCC	DFTB3, `SCC = Yes`, `ThirdOrderFull = Yes`
Dispersion	D3(BJ): $a_1 = 0.746$, $a_2 = 4.191$, $s_6 = 1.0$, $s_8 = 3.209$
Param set	3OB with P/N extension (H, C, N, O, P)
Charge	-2
SCC tolerance	10^{-10}
Fermi temperature	300 K
H-correction	Damping, exponent 4.0
Max angular momentum	H:s; C/N/O/Mg:p; P:d
Hubbard derivatives	C −0.1492, N −0.1535, O −0.1575, H −0.1857, P −0.14, Mg −0.02
Outputs	forces, band info, results tag (paths omitted)

Table A.10.: Molecular total energies E^{mol}, sums of isolated atomic reference energies $\sum_A E_A^{\mathrm{atom}}$, and atomization energies E_{atom} for optimized geometries (ωB97M-V/ma-def2-TZVP). Columns list DFT and DFTB values (energies in E_{h}; atomization energies and ΔE in kcal mol^{-1}). We use $E_{\mathrm{atom}}^{\mathrm{level}} = E_{\mathrm{level}}^{\mathrm{mol}} - \sum_A E_{A,\mathrm{level}}^{\mathrm{atom}}$ and $\Delta E = E_{\mathrm{atom}}^{\mathrm{DFT}} - E_{\mathrm{atom}}^{\mathrm{DFTB}}$. Atomic reference energies are listed in Table A.11.

(a) Reaction assemblies: *nucleophile + phosphate leaving group.*

Reaction			DFT			DFTB			ΔE
Nr.	Nucleophile	Leaving group	E^{mol}	$\sum_A E_A^{\mathrm{atom}}$	$E_{\mathrm{atom}}^{\mathrm{DFT}}$	E^{mol}	$\sum_A E_A^{\mathrm{atom}}$	$E_{\mathrm{atom}}^{\mathrm{DFTB}}$	[kcal/mol]
1	4-MeI	+ Ph–PO$_4$Et	−1218.8954	−1212.6795	−3900.5	−46.6711	−40.5023	−3871.0	−29.5
2	4-MeI	+ tBu–PO$_4$Et	−1145.0655	−1138.9995	−3806.5	−44.7811	−38.7330	−3795.2	−11.3
3	4-MeI	+ Me–PO$_4$Et	−1027.1376	−1022.4931	−2914.5	−37.3504	−32.7231	−2903.7	−10.8
4	4-MeI	+ Ph–PO$_4$	−1139.5989	−1134.5097	−3193.5	−41.2759	−36.2160	−3175.1	−18.4
5	4-MeI	+ tBu–PO$_4$	−1065.7487	−1060.8297	−3086.7	−39.3636	−34.4467	−3085.4	−1.3
6	4-MeI	+ Me–PO$_4$	−947.8133	−944.3232	−2190.1	−31.9282	−28.4369	−2190.8	0.7
7	4-MeI	+ Me–PP–PO$_4$	−2282.6146	−2277.2236	−3382.9	−56.4118	−51.0305	−3376.8	−6.1
8	ImH	+ Me–PO$_4$	−908.5038	−905.4877	−1892.6	−29.4488	−26.4336	−1892.1	−0.5
9	Arg	+ Me–PP–PO$_4$	−2340.3653	−2334.2959	−3808.6	−60.7214	−54.6458	−3812.5	3.9
10	MeArg	+ Me–PP–PO$_4$	−2379.6608	−2373.1313	−4097.3	−63.1821	−56.6491	−4099.5	2.2
11	Crn	+ Me–PP–PO$_4$	−2528.4302	−2521.7643	−4182.9	−68.7400	−62.0940	−4170.4	−12.5
12	4-MeI	+ CO$_2$Me–PO$_4$Et	−1215.7627	−1210.4604	−3327.2	−45.7260	−40.4510	−3310.1	−17.1
13	4-MeI	+ CO$_2$Me–PO$_4$	−1136.5249	−1132.2906	−2657.1	−40.3912	−36.1648	−2652.1	−5.0
14	4-MeI	+ MePO$_3$–PO$_4$Et	−1594.3302	−1588.9171	−3396.8	−49.2435	−43.8473	−3386.2	−10.6
15	4-MeI	+ MePO$_3$–PO$_4$	−1514.8945	−1510.7472	−2602.5	−43.7145	−39.5610	−2606.3	3.8

Note: numbering in this appendix subtable follows the order listed here and is not intended to replicate numbering in other tables.

(b) Components (optimized fragments).

Component	E^{mol}	$\sum_A E_A^{\mathrm{atom}}$	$E_{\mathrm{atom}}^{\mathrm{DFT}}$	E^{mol}	$\sum_A E_A^{\mathrm{atom}}$	$E_{\mathrm{atom}}^{\mathrm{DFTB}}$	ΔE
	[E_{h}]		[kcal/mol]	[E_{h}]		[kcal/mol]	[kcal/mol]
Ph–PO$_4$Et	−953.360 446	−949.179 455	−2623.61	−32.7808	−28.6145	−2614.37	−9.24
tBu–PO$_4$Et	−879.530 610	−875.499 468	−2529.58	−30.8908	−26.8452	−2538.63	9.05
Me–PO$_4$Et	−761.602 646	−758.993 006	−1637.57	−23.4602	−20.8353	−1647.11	9.54
Ph–PO$_4$	−874.063 985	−871.009 608	−1916.65	−27.3857	−24.3283	−1918.53	1.88
tBu–PO$_4$	−800.213 743	−797.329 620	−1809.81	−25.4733	−22.5590	−1828.78	18.97
Me–PO$_4$	−682.278 423	−680.823 158	−913.19	−18.0379	−16.5491	−934.26	21.07
Me-PP-PO$_4$	−2017.079 648	−2013.723 498	−2106.01	−42.5215	−39.1428	−2120.20	14.19
4-MeI	−265.534 908	−263.500 058	−1276.89	−13.8903	−11.8878	−1256.59	−20.30
ImH	−226.225 345	−224.664 571	−979.40	−11.4109	−9.8845	−957.84	−21.56
Arg	−323.285 615	−320.572 353	−1702.60	−18.1999	−15.5030	−1692.31	−10.29
N^G-MeArg	−362.581 169	−359.407 840	−1991.29	−20.6606	−17.5063	−1979.33	−11.96
Crn (deprot.)	−511.350 547	−508.040 824	−2076.88	−26.2185	−22.9513	−2050.20	−26.68
CO$_2$Me–PO$_4$Et	−950.227 766	−946.960 351	−2050.33	−31.8357	−28.5633	−2053.50	3.17
CO$_2$Me–PO$_4$	−870.989 988	−868.790 503	−1380.20	−26.5009	−24.2770	−1395.53	15.33
MePO$_3$–PO$_4$Et	−1328.795 282	−1325.417 002	−2119.90	−35.3532	−31.9595	−2129.59	9.69
MePO$_3$–PO$_4$	−1249.359 642	−1247.247 154	−1325.61	−29.8242	−27.6733	−1349.76	24.15

Table A.11.: Atom reference energies used to compute atomization energies. Values in Hartree (E_h).

Atom	**DFT** [E_h]	**DFTB** [E_h]
H	−0.498 873	−0.279 662
C	−37.837 740	−1.443 973
N	−54.577 928	−2.216 950
O	−75.064 802	−3.141 968
P	−341.229 589	−1.698 262
Mg	−200.052 348	−0.345 355

Table A.12.: Vanilla MACE (ScaleShift) architecture and training configuration used for phosphorylation.

Component	**Setting**
Model	ScaleShiftMACE; 2 RealAgnosticResidualInteractionBlocks; $l_\mathrm{max} = 1$
Channels / irreps	128 channels; hidden irreps $128 \times 0e + 128 \times 1o$; MLP irreps $16 \times 0e$
Correlation order	3; radial basis: 8 Bessel; radial MLP [64, 64, 64]
Cutoff	5.0 Å (internal r_max 8.0 Å)
Activation	SiLU; parameters $\approx 7.47 \times 10^5$
Batch / epochs	batch 16 (train/val), max 50 epochs
Optimizer	Adam ($\beta = 0.9$, amsgrad=True), lr 0.01, weight decay 5×10^{-7}
Scheduler	ReduceLROnPlateau (factor 0.8, patience 50)
Stabilization	grad clip 10.0; SWA start epoch 36, lr_SWA=0.001; EMA with decay 0.99
Targets	ΔE (atomization residual), $\Delta \mathbf{F}$; loss = weighted RMSE (per-atom energy)

Table A.13.: MACE-Eq (QEq+ESP) configuration. Differences vs. vanilla are highlighted.

Component	**Setting**
Model	`maceQEqESP` (charges + ESP head); hidden irreps $64 \times 0e + 64 \times 1o$
Cutoff / basis	$r_\mathrm{cut} = 8.0$ Å; 8 Bessel; cutoff basis 5; SiLU
Batch / epochs	batch 10 (train/val), 50 epochs; workers 4
Optimizer / sched.	Adam, lr 0.01; ReduceLROnPlateau (factor 0.8, patience 3), $\gamma = 0.9993$
Stabilization	grad clip 10.0; SWA lr 0.001; EMA 0.99
Targets / loss	energy + forces primary (weights: pre-SWA 1:1000, SWA 100:1); EFQRMSE; additionally uses total charge, ESP, ESP gradients
Atom refs	identical to Table A.14

Table A.14.: Δ-Atom reference energies (eV) used to convert totals to atomization energies in the MACE views.

Element	H	C	N	O	Mg	P
E_atom	-5.965	-990.325	-1424.816	-1957.121	-5434.307	-9239.123

Table A.15.: GROMACS QM/MM MD settings used for the **ImH + Me–PO$_4$** reaction in all three levels (DFT, HDNNP, MACE).

Parameter	**Value/Setting**
Integrator and sampling	
Integrator	MD
Initial time	0 ps
Timestep	1 fs
Steps	10^6 (total time 1 ns)
COM motion removal	mode: Linear; groups: System; frequency (`nstcomm`) = 200
Cutoffs and neighbor lists	
Cutoff scheme	Verlet
Neighbor update	every 10 steps
Periodic boundary cond.	xyz
Coulomb	PME, cutoff 1.0 nm
van der Waals	Cut-off, cutoff 1.0 nm
Dispersion correction	No
Thermostat	
Thermostat	Nosé–Hoover
T-groups	`IPh`, `SOL`
τ_T	0.5 ps, 0.5 ps
T_{ref}	300 K, 300 K
Couple every	1 step
QM/MM coupling	
QM/MM	Enabled
QM method / basis	RHF / STO-3G *(required by interface; ignored for DFTB runs)*
QM charge / multiplicity	−2 / (not specified)
MM charge scaling	1.0
Output / miscellaneous	
Compressed coordinates	every 200 steps (`nstxout-compressed=200`), group: System
Energies / log	every 200 steps (`nstenergy=nstlog=200`)
Coordinates / velocities	`nstxout=0`, `nstvout=0`; forces `nstfout=0`

Table A.16.: GROMACS QM/MM MD settings used for **4–MeI + Ph–PO$_4$Et** (“phenyl”) and **4–MeI + Me–PP–PO$_4$** (“his”) in all three levels (DFT, HDNNP, MACE).

Parameter	**Value/Setting**
Integrator and sampling	
Integrator	MD
Initial time	0 ps
Timestep	1 fs
Steps	10^6 (total time 1 ns)
COM motion removal	mode: Linear; groups: System; frequency (`nstcomm`) = 20
Cutoffs and neighbor lists	
Cutoff scheme	Verlet
Neighbor update	every 10 steps
Periodic boundary cond.	xyz
Coulomb	PME, cutoff 1.0 nm
van der Waals	Cut-off, cutoff 1.0 nm
Dispersion correction	No
Thermostat	
Thermostat	Nosé–Hoover
T-groups	`IPh`, `SOL`
τ_T	0.5 ps, 0.5 ps
T_{ref}	300 K, 300 K
Couple every	1 step
QM/MM coupling	
QM/MM	Enabled
QM group	`IPh`
QM method / basis	RHF / STO-3G *(required by interface; ignored for DFTB runs)*
QM charge / multiplicity	−1 / (not specified)
MM charge scaling	1.0
Output / miscellaneous	
Compressed coordinates	every 500 steps (`nstxout-compressed=500`), group: System
Energies / log	every 500 steps (`nstenergy=nstlog=500`)
Coordinates / velocities	`nstxout=0`, `nstvout=0`; forces `nstfout=0`

Table A.17.: Core `dftb_in.hsd` settings common to the DFTB baseline and the Δ-corrected runs (HDNNP and MACE). The HDNNP variant *adds* a neural-network block with symmetry functions; DFTB and MACE use only the basic block shown here.

Parameter	Value/Setting
Electronic structure (DFTB core)	
Theory level	DFTB (self-consistent, SCC=Yes; third-order: `ThirdOrderFull=Yes`)
Max. SCC iterations	250
Slater–Koster set	3ob–OP–PN (C, H, N, O, P)
Max. angular momentum	H: s; C/N/O: p; P: d
Molecular charge	reaction-dependent (-2 for ImH+Me–PO_4; -1 for Ph/His)
Occupation / smearing	Fermi; $T_F = 300$ K
Dispersion	D3(BJ): $a_1 = 0.746$, $a_2 = 4.191$, $s_6 = 1.0$, $s_8 = 3.209$
Hubbard derivatives	C: -0.1492, N: -0.1535, O: -0.1575, H: -0.1857, P: -0.14
H correction	Damping, exponent = 4.0
Analysis / output	
Forces	calculated (`CalculateForces=Yes`)
Detailed output	off (`WriteDetailedOut=No`)
HDNNP-specific augmentation (only for HDNNP Δ)	
ML block	`MachineLearning = NeuralNet {...}`
Symmetry functions	present; radial: 35, angular: 50, sum: 925
Cutoffs (SF)	radial cutoff ≈ 10 Å (18.8973 bohr); angular cutoff ≈ 6 Å (11.3384 bohr)
Atomic species	defined for H, C, N, O, P
Energy scaling / units	`ScaleEnergy=Yes`; `UnitIsKcalMol=No` (i.e., internal DFTB units)
Network params	supplied via `NeuralNetworkFiles` (paths omitted)

Table A.18.: Parameter provenance for the classical MM configuration used in the Fluorescence chapter and in the Sampling chapter.

System	Provenance / Setup
FR0, 4AP, ACDAN	GAFF/GAFF2[246,247]; charges: RESP (HF/6-31G*)[248] (AM1-BCC alternative).
Retinal–Lys	GAFF/GAFF2; charges: RESP (HF/6-31G*); QM/MM: DFTB3/3OB-f[10,11,129].
Water	TIP3P[133].
Other solvents	GROMACS-compatible parameters (VirtualChemistry)[115,249].

Table A.19.: Parameter provenance for the solvents used in the fluorescence simulations.

Solvent	Model / FF	Provenance (short)
Water	TIP3P	Jorgensen et al. (1983)
Acetonitrile	OPLS-AA	VirtualChemistry DB
Acetone	OPLS-AA	VirtualChemistry DB
DMF	OPLS-AA	VirtualChemistry DB
DMSO	OPLS-AA	VirtualChemistry DB
Methanol	OPLS-AA	VirtualChemistry DB
Dichloromethane	OPLS-AA	VirtualChemistry DB
Ethyl acetate	OPLS-AA	VirtualChemistry DB
1,4-Dioxane	OPLS-AA	VirtualChemistry DB
n-Hexane	OPLS-AA	VirtualChemistry DB

Note: Organic solvents: OPLS-AA topologies from the *GROMACS Molecule & Liquid Database* (VirtualChemistry; based on Caleman 2012). For use in GROMACS see Abraham et al. (2023). Water: TIP3P from Jorgensen et al. (1983).[115,133,249]

Table A.20.: GROMACS QM/MM MD settings used for the free simulations that generate the excited-state training/evaluation geometries in the fluorescence chapter.

Parameter	**Value/Setting**
Integrator and sampling	
Integrator	MD
Steps	10^6
Timestep	1 fs
Continuation	Yes
Constraints	Hydrogen bonds
Cutoffs and neighbor lists	
Cutoff scheme	Verlet
Neighbor update	every 10 steps
Coulomb	PME, cutoff 1.0 nm
van der Waals	Cut-off, cutoff 1.0 nm
Dispersion correction	Energy and pressure
Thermostat / Barostat	
Thermostat	Nosé–Hoover, $\tau_T = 0.2$ ps, $T = 300$ K
Barostat	Parrinello–Rahman (NPT), $\tau_P = 2.0$ ps, $P = 1.013$ bar
Compressibility	4.5×10^{-5} bar^{-1}
QM/MM coupling	
QM/MM	Enabled
QM method	RHF
QM basis	STO-3G
QM charge / multiplicity	0 / 1
MM charge scaling	1.0
Output / miscellaneous	
Periodic boundary conditions	xyz
Velocity generation	Off
Coordinates (compressed)	every 100 steps; precision 10^6
Energies / log	energies every step; log every 1000 steps
Coordinates / velocities	`nstxout=0`, `nstvout=0`

Table A.21.: TD–LC–DFTB (Casida) single-point settings used to re-evaluate each geometry from the free simulations (bright-state analysis, forces, and transition properties).

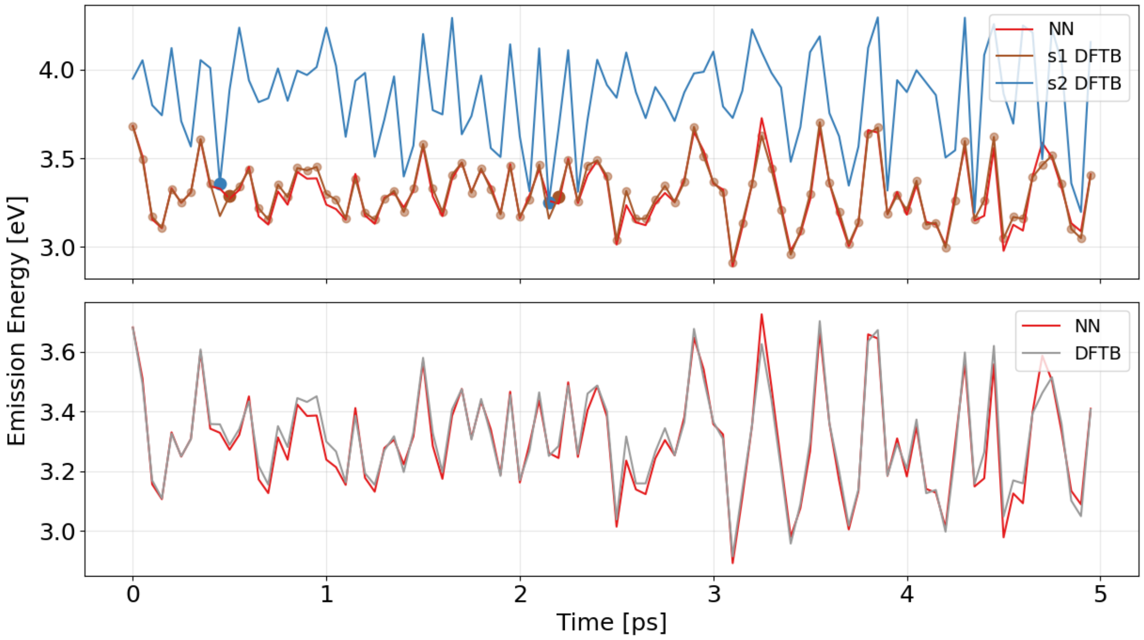
Parameter	**Value/Setting**
Electronic structure	
Theory level	DFTB (self-consistent, SCC=Yes)
SCC tolerance	1×10^{-10}
Max. SCC iterations	100
Slater–Koster set	`ob2-1-1` (C,H,N,O)
Max. angular momentum	H: s; C/N/O: p
Molecular charge	0
Range separation	
Long-range correction	Enabled (`RangeSeparated = LC`)
Screening	Thresholded, threshold = 1.0×10^{-6}
Excited-state (Casida) block	
Number of excitations	10 (vertical)
Symmetry	singlet
State of interest	1
Transition dipole	written (`WriteTransitionDipole = Yes`)
Mulliken (excited state)	written (`WriteMulliken = Yes`)
Other detailed writes	transitions / coeffs / eigenvectors: off
Analysis and output	
Forces	calculated (`CalculateForces = Yes`)
Mulliken analysis	enabled
Detailed output	written (`WriteDetailedOut = Yes`)
Restart frequency	20
Random seed	0
Parser / bookkeeping	
Parser version	7
Write processed input	`dftb_pin.hsd` (`WriteHSDInput = Yes`)

Table A.22.: Training configuration for the two MACE–Eq runs. Shared settings apply to both; only the loss weights differ between the *energy–focused* and *force–focused* variants. SWA inverts the main-phase weights; EMA decay = 0.99.

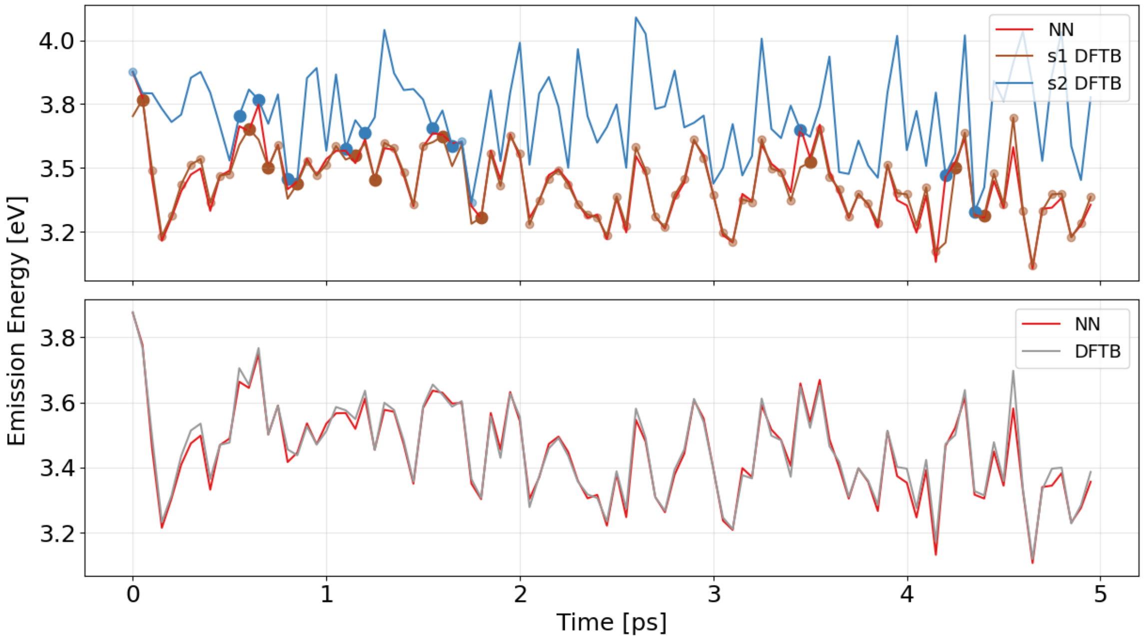
Category	**Parameter**	**Shared**	**Values (E | F focus)**
Architecture	Core	MACE+QEq+ESP	—
	Hidden irreps	`64x0e+64x1o`	—
	Cutoff r	8.0 Å	—
	Radial basis	8 Bessel, 5 cutoff	—
	Activation	SiLU	—
Optimisation	Opt./LR	Adam, 1×10^{-2}	—
	Batch/Epochs	10 / 50	—
	Scheduler	ReduceLROnPlateau (0.8,3)	—
	Grad. clip	10.0	—
	SWA/LR	on, 1×10^{-3}	—
	EMA	on, decay 0.99	—
Targets & loss	Main loss	$E + F$ (+ESP terms)	—
	Main weights	—	E:1000, F:1 | E:1, F:1000
	SWA weights	—	E:10, F:1000 | E:1000, F:10
Electrostatics	E_0 refs	H/C/N/O	—
	Formal charges	from data	—
	ESP supervision	`esp`, `esp_gradient`	—

Table A.23.: Selected hyperparameters corresponding to the performance in Table 24.

Training combination	Layers × Neurons	lr_0	L2	λ
All 10 solvents	2×850	$5{\cdot}10^{-4}$	10^{-4}	$5{\cdot}10^{-3}$
Water	4×800	$1{\cdot}10^{-4}$	$1{\cdot}10^{-3}$	$5{\cdot}10^{-3}$
DMSO	3×850	$1{\cdot}10^{-4}$	$1{\cdot}10^{-3}$	$5{\cdot}10^{-4}$
DMF	2×950	$1{\cdot}10^{-4}$	$1{\cdot}10^{-4}$	$1{\cdot}10^{-3}$
Acetonitrile	3×550	$5{\cdot}10^{-4}$	$1{\cdot}10^{-3}$	$5{\cdot}10^{-3}$
Methanol	4×900	$1{\cdot}10^{-4}$	$5{\cdot}10^{-4}$	$1{\cdot}10^{-3}$
Acetone	3×950	$1{\cdot}10^{-4}$	$5{\cdot}10^{-4}$	$1{\cdot}10^{-2}$
DCM	4×700	$1{\cdot}10^{-4}$	$5{\cdot}10^{-4}$	$5{\cdot}10^{-4}$
Ethyl acetate	4×950	$1{\cdot}10^{-4}$	$1{\cdot}10^{-3}$	$5{\cdot}10^{-3}$
1,4-Dioxane	2×850	$1{\cdot}10^{-4}$	$1{\cdot}10^{-4}$	$1{\cdot}10^{-3}$
n-Hexane	3×800	$1{\cdot}10^{-4}$	$5{\cdot}10^{-5}$	$1{\cdot}10^{-2}$

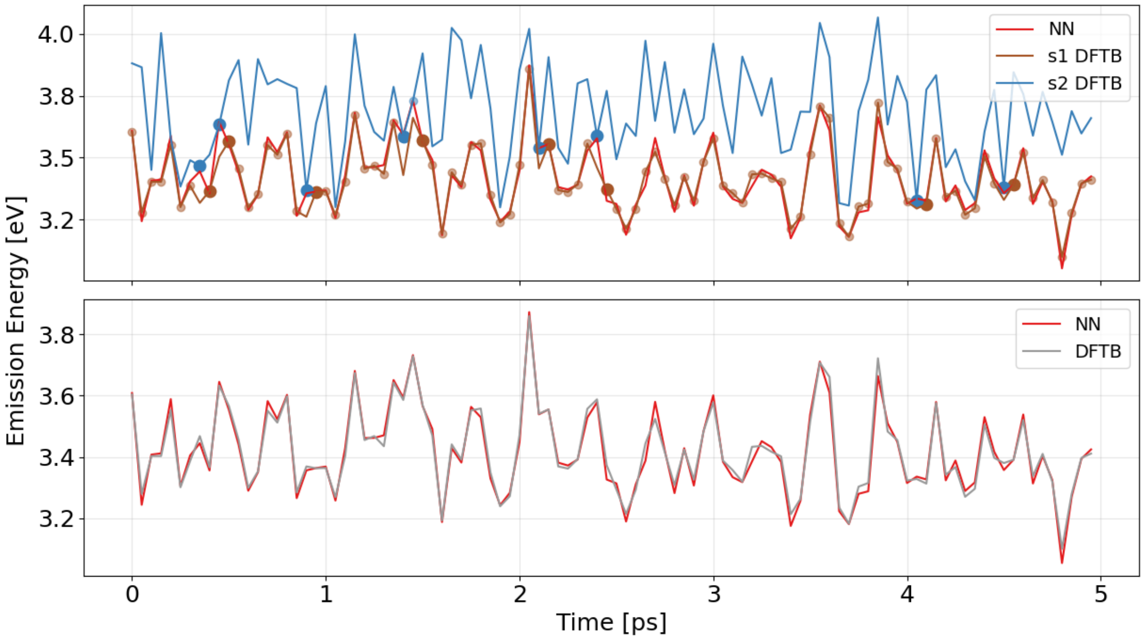


(a) Water

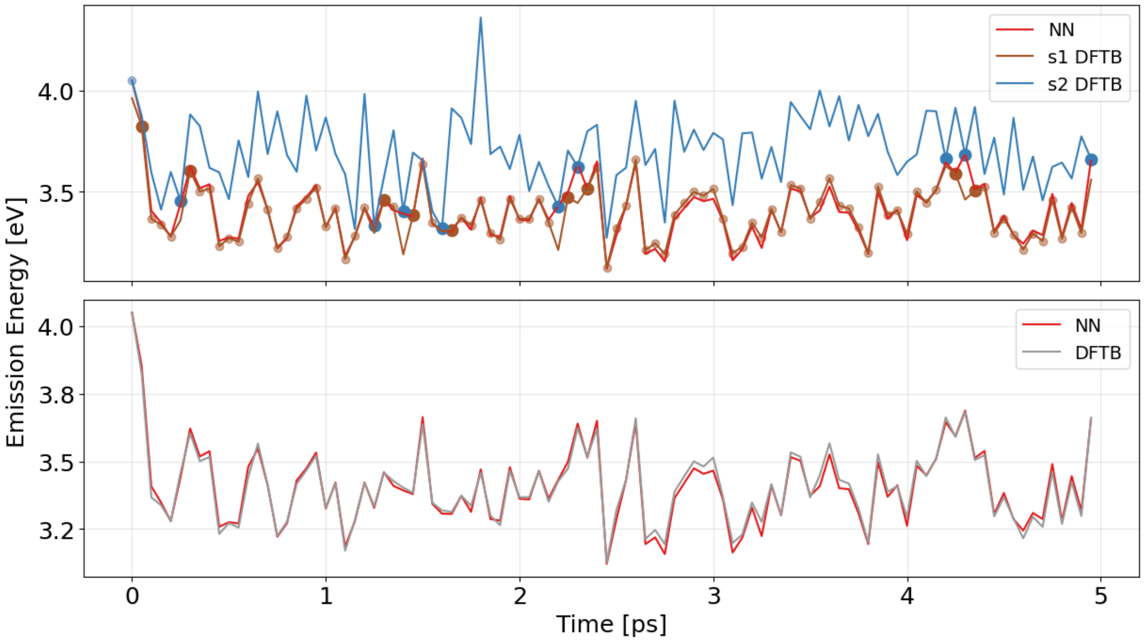


(b) DMSO

Figure A.1.: Bright-state emission along NN/MM trajectories. Each panel shows (top subplot) NN emission energy together with single-point DFTB $S_1/S_2/S_3$ values for the same geometries, and (bottom subplot) NN emission compared to the DFTB emission of the recomputed bright state. Top: Water; bottom: DMSO.

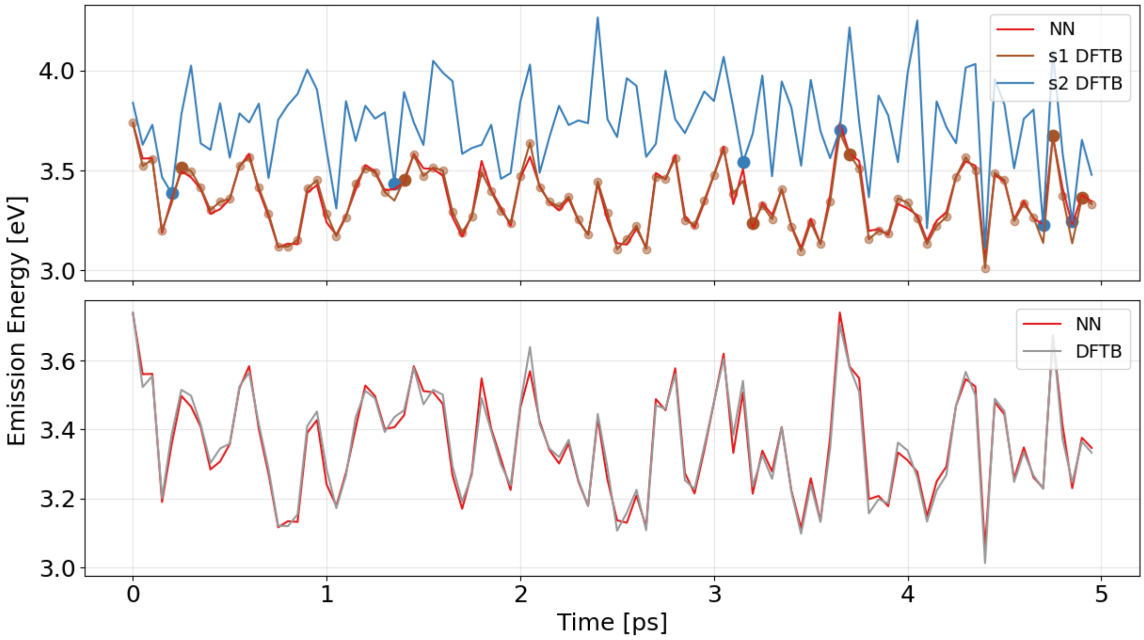


(a) DMF

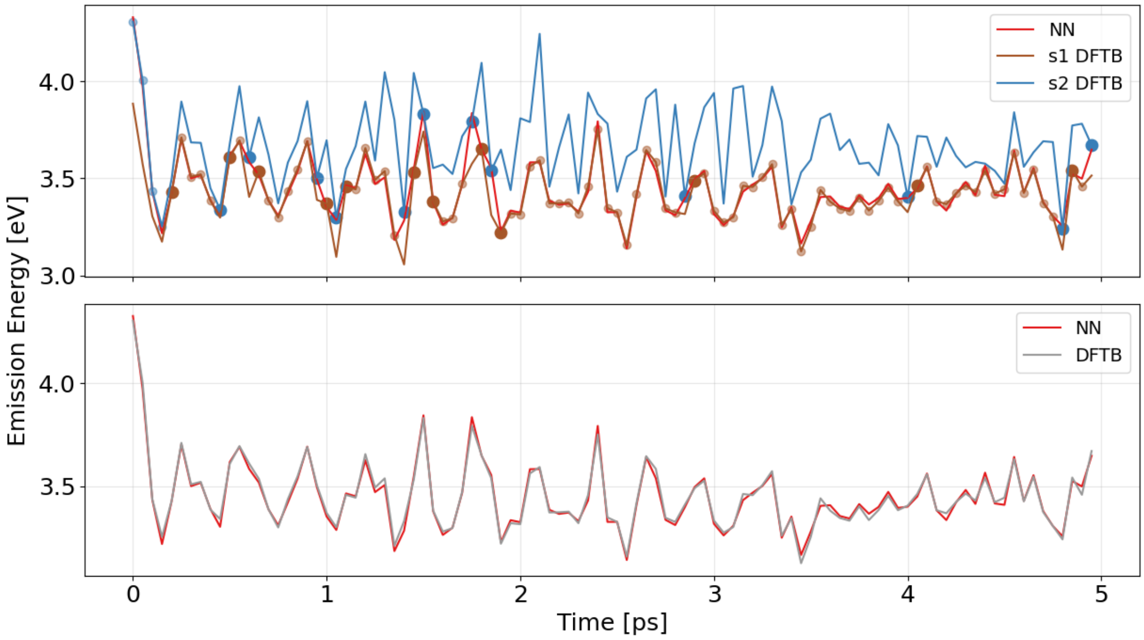


(b) Acetonitrile (ACN)

Figure A.2.: Bright-state emission along NN/MM trajectories. Top: DMF; bottom: ACN. Subplots as in Fig. A.1.

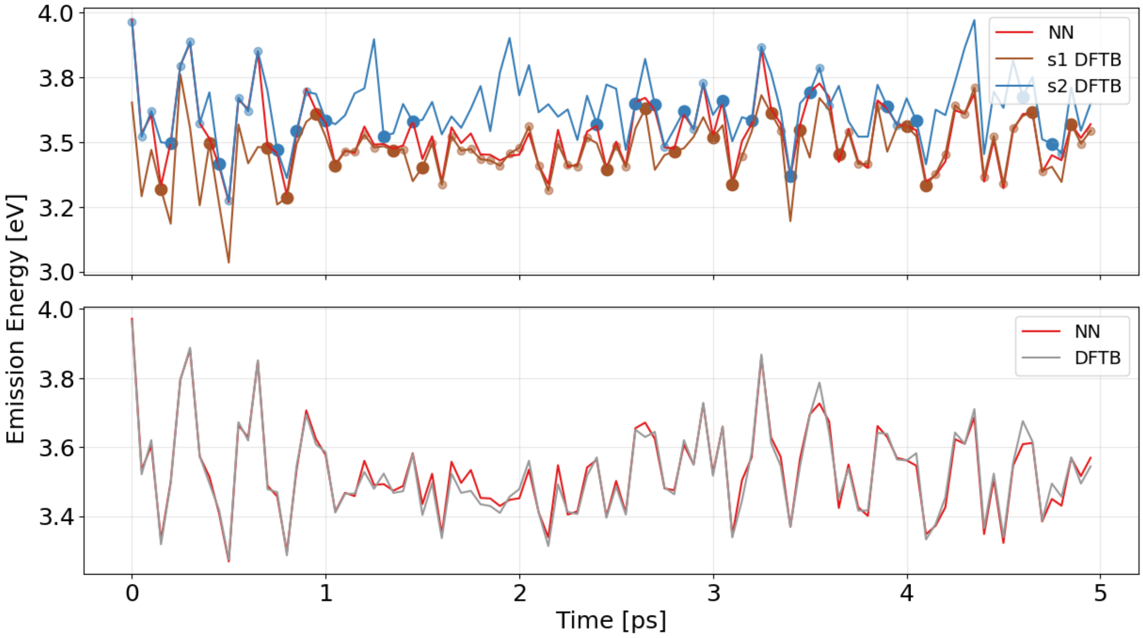


(a) Methanol

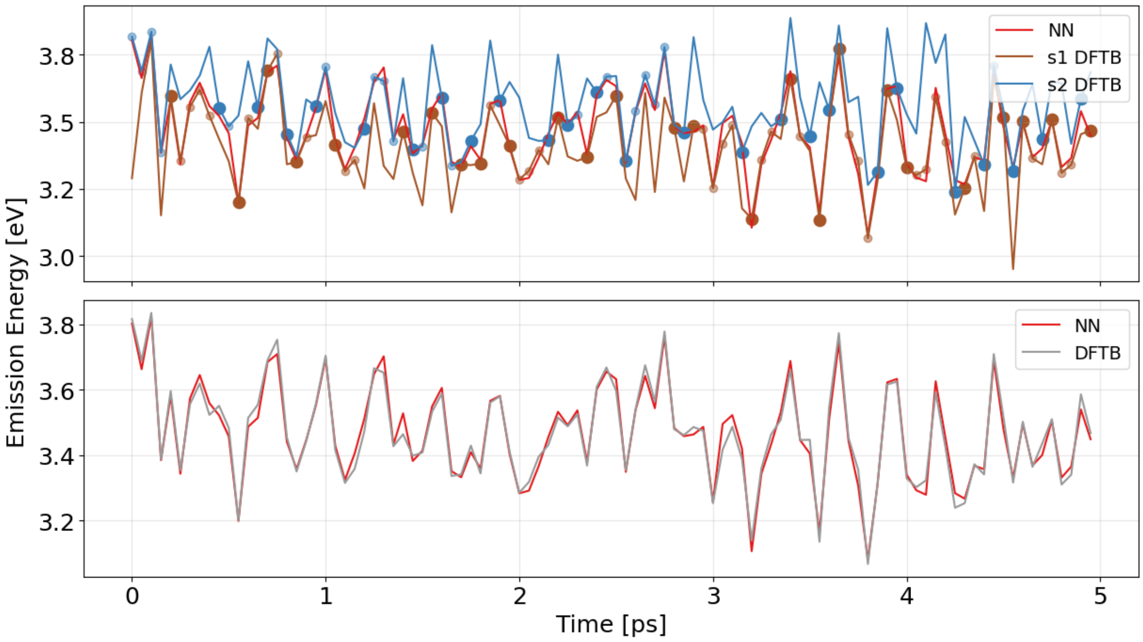


(b) Acetone

Figure A.3.: Bright-state emission along NN/MM trajectories. Top: Methanol; bottom: Acetone. Subplots as in Fig. A.1.

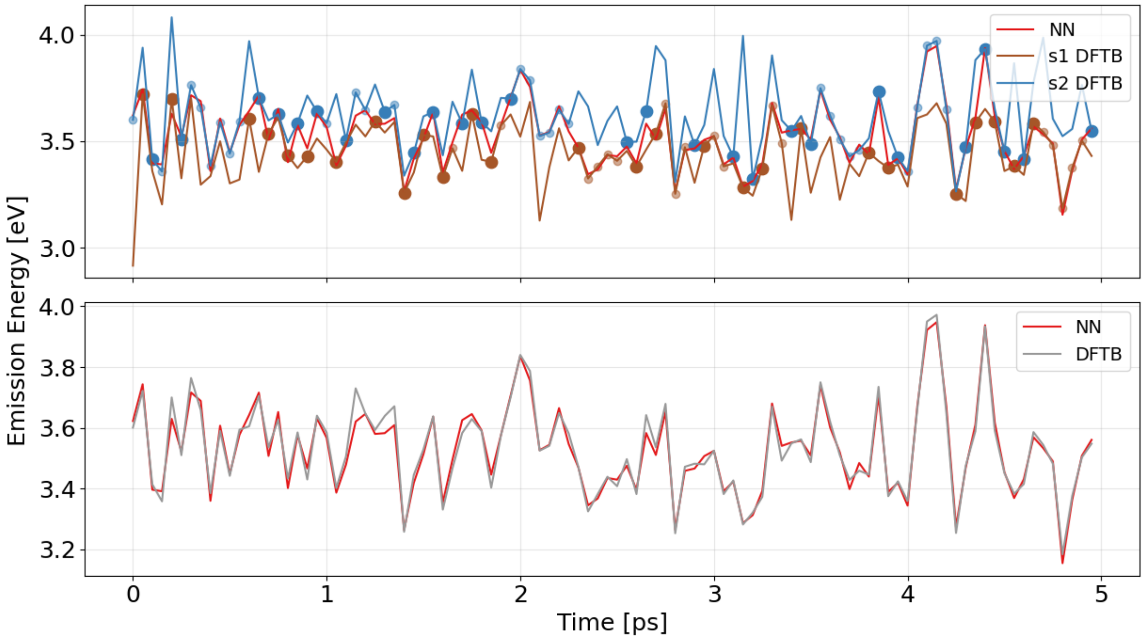


(a) Dichloromethane (DCM)

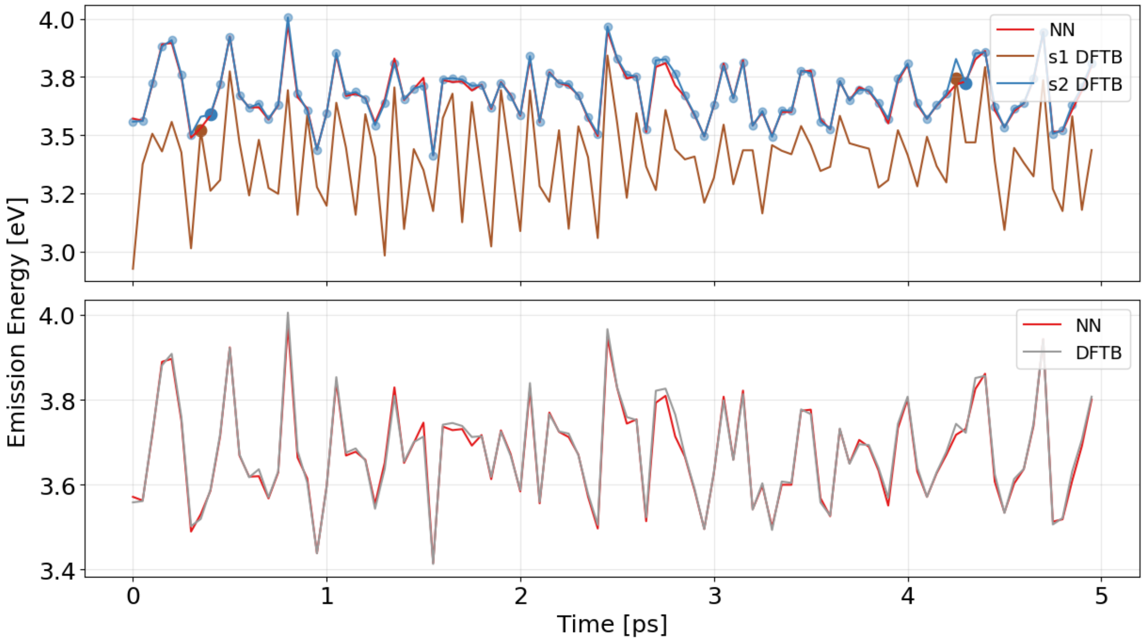


(b) Ethyl acetate

Figure A.4.: Bright-state emission along NN/MM trajectories. Top: DCM; bottom: Ethyl acetate. Subplots as in Fig. A.1.



(a) 1,4-Dioxane



(b) n-Hexane

Figure A.5.: Bright-state emission along NN/MM trajectories. Top: 1,4-Dioxane; bottom: n-Hexane. Subplots as in Fig. A.1.

Table A.24.: Model configuration of the property network used for each solvent and the pooled *All 10* set. n_E/depth$_E$: hidden width / depth of the energy head; n_f/depth$_f$: width / depth of the oscillator-strength head; lr_0: initial learning rate; “Best epoch” from the final fit.

Solvent	n_E	depth$_E$	n_f	depth$_f$	lr_0	Best epoch
Water	50	6	70	7	0.0005	15
DMSO	90	8	90	4	0.0001	68
DMF	80	8	20	5	0.0005	59
ACN	90	6	40	8	0.0005	53
Methanol	50	7	20	3	0.0005	53
Acetone	70	6	20	4	0.0005	63
DCM	60	6	60	8	0.0010	83
Ethyl acetate	20	5	20	7	0.0005	55
Dioxane	20	2	20	3	0.0005	71
Hexane	60	8	80	7	0.0005	58
All 10	40	6	25	5	0.0005	86

Table A.25.: Calibration settings used for constructing fluorescence spectra from energy-space data, prior to wavelength transform and MeOH alignment.

Method	**Parameter**	**Value**	**Unit / Setting**
Histogram	bins	50	
	energy_range		[2.5, 4.1] eV
	sigma_guess	0.2	eV
	maxfev	5000	function evaluations
	bin_width	0.032	eV (derived)
	x_fine	1000	points (fine grid)
NEA	sigma_eV	0.02	eV (Gaussian kernel)
	kernel		`gauss`[a]
	grid_step_eV	0.005	eV
	emission_weight		E^3 (emission)
	normalize		`area`
	energy_range		auto (data $\pm 5\sigma$)
E–ZTFC	vibronic_freq_eV	0.15	eV
	huang_rhys	0.8	–
	n_modes	3	(0–0, 0–1, 0–2, 0–3)
	vib_range	1.5	eV (template, ±1.5 eV)
	vib_kernel_width	0.02	eV
	grid_step_eV	0.005	eV
	extra_gauss_eV	0.02	eV
	emission_weight		E^3 (emission)
	normalize		`area`
	energy_range		[2.5, 5.0] eV

Table A.26.: Overview of the training snapshots used for the retinal models (split into two subtables): numbers of OM2/MRCI labels (OM2), Δ-learning labels (Delta), and totals per solvent/counterion combination. Abbreviations – Solvents: ANI = anisole, CHL = chloroform, DMSO = dimethyl sulfoxide, ETH = ethanol, H_2O = water, IBU = isobutanol, MOH = methanol, PYR = pyridine, THF = tetrahydrofuran, TOL = toluene, *vac* = vacuum. Counterion configs: Cl^- (one chloride near $RSBH^+$), $2Cl^-$ (two chlorides), Na^+ (near $RSBH^+$), Na^+ (r) (near β-ionone ring), NaCl (Na^+ and Cl^- present), Pure (no targeted counterion).

Solvent	Counterion	OM2	Delta
ANI	Cl^-	3963	3903
ANI	$2Cl^-$	3762	3746
ANI	NaCl	3878	3858
ANI	Na^+ (ring)	3807	3767
ANI	Pure	3740	3685
CHL	Cl^-	3875	3786
CHL	$2Cl^-$	3814	3665
CHL	Na^+ (ring)	3848	3776
CHL	Pure	3947	3898
DMSO	Cl^-	3855	3814
DMSO	$2Cl^-$	3740	3716
DMSO	NaCl	3805	3692
DMSO	Na^+	3714	3684
DMSO	Na^+ (ring)	3727	3688
DMSO	Pure	3732	3694
ETH	Cl^-	3868	3835
ETH	$2Cl^-$	3914	3872
ETH	NaCl	3784	3767
ETH	Na^+	3773	3724
ETH	Na^+ (ring)	3635	3592
ETH	Pure	3523	3491
H2O	Cl^-	1874	1856
H2O	$2Cl^-$	3792	3735
H2O	NaCl	939	924
H2O	Na^+	1830	1795
H2O	Na^+ (ring)	7453	7365
H2O	Pure	1844	1822
IBU	Cl^-	3929	3855
IBU	$2Cl^-$	3911	3832
IBU	NaCl	3883	3836
IBU	Na^+ (ring)	3828	3760
IBU	Pure	3994	3903
MOH	Cl^-	7459	7304
MOH	$2Cl^-$	7758	7648
MOH	NaCl	3764	3730
MOH	Na^+	3765	3692
MOH	Na^+ (ring)	3701	3622
MOH	Pure	7233	7143
PYR	Cl^-	3919	3895
PYR	$2Cl^-$	3835	3811
PYR	NaCl	3833	3816
PYR	Na^+	3904	3849
PYR	Na^+ (ring)	3657	3607
PYR	Pure	3551	3524
THF	Cl^-	3581	3507
THF	$2Cl^-$	3822	3722
THF	Na^+	3976	3897
THF	Na^+ (ring)	3707	3668
THF	Pure	3685	3611
TOL	Cl^-	7662	7528
TOL	$2Cl^-$	7438	7268
TOL	NaCl	3850	3828
TOL	Na^+ (ring)	7825	7723
TOL	Pure	7838	7703
GRAND TOTAL (incl. vac)		**230744**	**227325**

Note: Vacuum (*vac*, Pure) contains OM2 = 4,000, Δ = 3,893

Table A.27.: Enhanced sampling simulation parameters for the Sampling chapter.

Parameter	Value/Setting
Simulation settings	
Timestep	$\Delta t = 0.5$ fs
Simulation length	50,000 steps (+20% each generation)
Temperature	300 K
Thermostat	Langevin (friction = 1.0 ps^{-1})
Auxiliary mass	$m_\lambda = 50.0$ amu (extended mass)
CV range	-30.0 to 30.0
Grid spacing	$\Delta\xi = 0.1$
Enhanced sampling configuration	
Sampling method	latent embedding
Sampling lower limit	$\xi_{\text{threshold}} = 2.0$ (+20% each generation)
Accelerated MD integration	
aMD scheme	GaMD (upper bound approach)
aMD parameter	0.01
Application	Enabled for enhanced exploration
Physical constraints and QC	
Distance constraints	$r_{\text{min}} = 0.75$ Å to $r_{\text{max}} = 2.0$ Å
Relaxation algorithm	FIRE, 10 steps
Confidence margin	0.1 (conformational assessment)
Confidence parameter	$k_{\text{conf}} = 100.0$
Statistical analysis	
Linkage method	Complete linkage (clustering)
Full sampling threshold	$n_{\text{full}} = 100.0$
SAMD parameter	$c_0 = 0.0001$ (self-adaptive MD)

Table A.28.: Uncertainty quantification and calibration parameters.

Parameter	Value
Gaussian mixture model (latent space)	
Uncertainty metric	$\xi(\mathbf{x}) = -\log\big(\sum_{k=1}^{15} w_k\, \mathcal{N}\big(\zeta(\mathbf{x}) \mid \boldsymbol{\mu}_k, \Sigma_k\big)\big)$
Number of components	$K = 15$
Covariance type	Full matrices
Initialization	K-means++
Convergence tolerance	10^{-6} (log-likelihood change)
Maximum iterations	1000
Ordering method	System mean
Conformal prediction calibration	
Calibration	$\text{CI}_{95\%}(\mathbf{x}) = [\hat{y}(\mathbf{x}) - \hat{q}_{0.95},\ \hat{y}(\mathbf{x}) + \hat{q}_{0.95}]$
Significance level	$\alpha = 0.05$ (95% confidence)

Table A.29.: Metadynamics configuration for the Sampling chapter (shared settings and CV definitions).

Parameter	Value/Setting
Shared deposition / multi-walker	
Gaussian height	0.5
Deposition pace	500 steps
Walkers (count)	16
Walker stride	500 steps
HILLS output	Enabled
RMSD-biased metadynamics	
Collective variable	PCA / RMSD
Gaussian width	0.1
Bias type	Standard metadynamics
PCA-biased metadynamics	
Collective variables	Top 10 principal components (PC1–PC10) / Reference Structure for RMSD
Note	PCs from trajectory PCA; biased along largest-variance modes / Geometry Optimize

Table A.30.: Solvent models used in ground-state simulations.

System	Solvent / Model	Note / Citation
FR0	DMSO	parameters as in[55]
4AP	DMSO	parameters as in[55]
ACDAN	DMSO	parameters as in[55]
Retinal	Water (TIP3P)	standard TIP3P model

Table A.31.: GROMACS QM/MM ground-state MD settings.

Parameter	**Value/Setting**
Integrator and sampling	
Integrator	MD
Steps	10^6
Timestep	1 fs
Continuation	Yes
Constraints	Hydrogen bonds
Cutoffs and neighbor lists	
Cutoff scheme	Verlet
Neighbor update	every 10 steps
Coulomb	PME, cutoff 1.0 nm
van der Waals	Cut-off, cutoff 1.0 nm
Dispersion correction	Energy and pressure
Thermostat / Barostat	
Thermostat	Nosé–Hoover, $\tau_T = 0.2$ ps, $T = 300$ K
Barostat	Parrinello–Rahman (NPT), $\tau_P = 2.0$ ps, $P = 1.013$ bar
Compressibility	4.5×10^{-5} bar^{-1}
QM/MM coupling	
QM/MM	Enabled
QM method	RHF
QM basis	STO-3G
QM charge / multiplicity	0 / 1
MM charge scaling	1.0
Output / miscellaneous	
Periodic boundary conditions	xyz
Velocity generation	Off
Trajectory compression	every 100 steps, precision 10^6
Energies / log	energies every step; log every 1000 steps

Table A.32.: DFTB2 Computational Parameters for Quantum-Chemical Ground Truth.

Parameter	**Value/Setting**
Electronic structure	
Theory level	DFTB2 (SCC enabled)
Slater–Koster set	ob2-1-1
SCC tolerance	10^{-7} Hartree
Maximum SCC iterations	500
Molecular charge	0
Range separation	Long-range corrected; screening threshold 10^{-16}
Basis / angular momentum	
Hydrogen	$l_{\text{max}} = s$
Carbon, nitrogen, oxygen	$l_{\text{max}} = p$
Analysis and output	
Force calculation	Enabled
Random seed	0
Notes	
Excited-state section present	Casida block specified, but only ground state used in this chapter

Declaration of Authorship and Tool Disclosure

I hereby declare that I have written this dissertation independently and have not used any sources or aids other than those indicated. All passages taken from other sources – whether literally or in substance – are cited accordingly. This dissertation has not been submitted, in whole or in part, for any previous degree or qualification. I have adhered to the rules of good scientific practice throughout.

Disclosure of software and AI-assisted tools used. In preparing, editing, and typesetting this dissertation I used the following tools as *aids only*: GitHub Copilot (boilerplate code completions and minor LaTeX snippets), DeepL/DeepL Write (translation and style suggestions), OpenAI ChatGPT (language polishing and occasional LaTeX troubleshooting), Elicit (literature discovery and scoping), and Grammarly (grammar and style suggestions). All scientific ideas, analyses, results, and conclusions are my own; all suggestions from these tools were critically reviewed and revised by me; no passages were adopted verbatim without proper citation; no data, results, or references were fabricated or manipulated by these tools.