

Washington University in St. Louis

Washington University Open Scholarship

Arts & Sciences Electronic Theses and
Dissertations

Arts & Sciences

12-19-2023

A Quantum Physics-Based Force Field for Biomolecular Simulation

Roseane dos Reis Silva
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

 Part of the [Computational Chemistry Commons](#)

Recommended Citation

dos Reis Silva, Roseane, "A Quantum Physics-Based Force Field for Biomolecular Simulation" (2023). *Arts & Sciences Electronic Theses and Dissertations*. 3209.
https://openscholarship.wustl.edu/art_sci_etds/3209

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology & Biomedical Sciences
Computational and Molecular Biophysics

Dissertation Examination Committee:

Jay Ponder, Chair

Kathleen Hall

Alex Holehouse

Janice Robertson

Andrea Soranno

A Quantum Physics-Based Force Field for Biomolecular Simulation

by

Roseane dos Reis Silva

A dissertation presented to
Washington University in St. Louis
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2023
St. Louis, Missouri

© 2023, Roseane dos Reis Silva

Table of Contents

List of Figures	iv
List of Tables	vi
Acknowledgments.....	viii
Abstract.....	xi
Chapter 1. Introduction	1
1.1. Molecular dynamic simulations and force fields for biomolecules	3
1.1.1. Point Charge Models.....	4
1.1.2. Polarizable models	6
1.2. <i>Ab initio</i> methods and the parametrization of classical force fields.....	8
1.2.1. Symmetry Adapted Perturbation Theory (SAPT).....	9
1.2.2. Coupled Cluster methods for total intermolecular energy	11
1.3. Quantum Chemistry databases	11
1.3.1. The S101x7 database	12
1.3.2. Non-Covalent Interactions Atlas (NCIA)	13
1.3.3. DES370k database	14
1.4. A database for organic liquids.....	15
1.5. HIPPO: Hydrogen-like Intermolecular Polarizable Potential	15
1.6. Structure of this dissertation.....	16
1.7. References	18
Chapter 2. The HIPPO Force Field and Water Model.....	20
2.1. Introduction	20
2.2. Theory	23
2.3. Methods.....	32
2.4. Results	37
2.5. Discussion	62
2.6. Conclusions	72
2.7. References	78
Chapter 3. The Benzene Model	86

3.1. Introduction	86
3.2. Methods	90
3.3. Results	92
3.3.1. Dimer potential energy surface	93
3.3.2. Liquid	98
3.4. Conclusion	100
3.5. References	100
Chapter 4. Computational tools for force field parametrization	104
4.1. Building the target dataset for parametrization	104
4.1.1. Structure of the Data Set	105
4.2. PolFit: a python tool for parametrization of the HIPPO force field	110
4.2.1. An example of fitting a molecule using this tool	110
4.2.2. Analyzing and visualizing	118
4.3. Future work	122
4.4. References	123
Chapter 5. Modeling Organic Molecules with the HIPPO Force Field	124
5.1. Introduction	124
5.2. Methods	125
5.2.1. Selection of molecules	125
5.2.2. Computational details	126
5.3. Results	127
5.4. References	136
Chapter 6. Future directions for the HIPPO model	138
6.1. Bridging Classical Physics and Advanced Machine Learning	138
6.2. Development of the HIPPO Force Field for Proteins	139
6.3. References	142
Appendix A	143
Appendix B	146
Appendix C	148
Appendix D	150
Appendix E	162

List of Figures

Figure 1.1.	The Point Charge Force Field model	4
Figure 1.2.	Predicted structures of the UUCG RNA tetraloop.....	6
Figure 1.3.	SAPT approach to intermolecular energy decomposition for two monomers A and B.	9
Figure 1.4.	Schematic representation of the dimers in the S101 Database	12
Figure 2.1.	Schematic of a HIPPO atom. The blue shaded area represents the valence electron density, and the red point represents the point core charge	24
Figure 2.2.	Error in the electrostatic potential: HIPPO vs. point multipoles neglecting charge penetration.....	39
Figure 2.3.	Energy components for water dimer dissociation curve (A) and “flap angle” degree of freedom (B)..	41
Figure 2.4.	Structures of the ten Smith water dimers obtained from full geometry optimization at the MP2/aug-cc-pV5Z level.....	44
Figure 2.5.	Total energies and components for ten water dimer stationary points.	47
Figure 2.6.	Three-body energies for water trimer as a function of intermolecular distance. ..	49
Figure 2.7.	Three-body energies for water clusters tetramer through octamer	51
Figure 2.8.	Water radial distribution at 298 K and 1 atm.....	54
Figure 2.9.	Thermodynamic water properties for the HIPPO model (green) compared to experiment (black) for temperatures from 248 to 373 K at atmospheric pressure (1 atm).....	59
Figure 2.10.	Dynamical water properties for the HIPPO model (green) compared to experiment (black) for temperatures from 248 to 373 K at atmospheric pressure (1 atm).....	60
Figure 2.11.	Self-diffusion coefficient vs. cubic box size at 298 K.....	60
Figure 2.12.	Surface Tension for the HIPPO model (green) compared to experiment (black) for temperatures from 248 K to 458 K.....	61
Figure 2.13.	Density of the HIPPO model (green) compared to experiment (black) for pressures from 1 to 9,375 atm at room temperature (298 K).....	61
Figure 2.14.	Second virial coefficient of the HIPPO water model (green) compared to experiment (black) for temperatures from 298 to 575 K	62

Figure 2.15.	Water radial distribution function for the HIPPO-SAPT water model before optimization with ForceBalance at 298 K and 1 atm.....	69
Figure 2.16.	A non-exhaustive taxonomy of classical water models.....	73
Figure 3.1.	Total intermolecular energies for benzene-benzene and benzene-water dimers using the HIPPO, AMOEBA and OPLS-AA force fields	94
Figure 3.2.	HIPPO energy components for dimers in the S101x7 and NCIA R739x5 databases, and randomly generated conformations	95
Figure 3.3.	AMOEBA energy components for dimers in the S101x7 and NCIA R739x5 databases, and randomly generated conformations	96
Figure 3.4.	OPLS-AA energy components for dimers in the S101x7 and NCIA R739x5 databases, and randomly generated conformations	97
Figure 3.5.	Temperature dependence of liquid benzene at 1 atm	99
Figure 3.6.	Center to center radial distribution function of HIPPO, AMOEBA and OPLS-AA compared to experimental neutron scattering data.....	99
Figure 4.1.	Sample output for a finalized parametrization process for analysis and visualization	121
Figure 5.1.	Comparison between experimental density and HIPPO results for molecules in data set 1.	128
Figure 5.2.	Comparison between experimental Enthalpy of Vaporization and HIPPO results for molecules in data set 1.	128

List of Tables

Table 2.1.	HIPPO Water Monomer Properties. All calculations performed on experimental, gas phase geometry where the Z-axis is the C2 axis, the molecule lies in the XZ-plane, and the O atom is along the negative Z-axis.	38
Table 2.2.	Water dimer binding energies for HIPPO compared to ab Initio reference calculations. Dimer geometries were taken from the Supporting Information of reference 69; Ref 1 energies are from reference 66, and Ref 2 values are from reference 68. Dimer stabilization energies ⁶⁷ and total deformation energies at the CCSD(T)/pV5Z level are shown, as are complete basis set (CBS) extrapolated values. ^{70, 71} HIPPO dimer energies are provided for single point calculations at the CCSD(T)/pV5Z geometry, and for fully optimized HIPPO structures. Also shown are the QM and HIPPO RO-O dimer distances, the HIPPO structure RMS vs. MP2/aug-cc-pV5Z optima, and the number of negative frequencies (n) for CCSD(T)/pV5Z and HIPPO optima. All energies are in kcal/mol, and the RO-O distance and HIPPO RMS values are in Angstroms.	45
Table 2.3.	Water cluster binding energies (kcal/mol) with HIPPO compared to ab initio calculations. a Structures and reference values from reference 73. b Hexamer structures and reference energies from reference 74.	49
Table 2.4.	Water properties at room temperature (298 K and 1 atm). a Reference 75. b Reference 76. c Reference 77. d Reference 78. e Reference 79. f Reference 80..	52
Table 2.5.	Ice properties from HIPPO model compared with experimental density (kg/m ³) and lattice energy (kcal/mol). a Experimental values from reference 89. b Values from the ICE10 data set. ⁹⁰	58
Table 2.6.	HIPPO-SAPT water properties at room temperature (298 K and 1 atm).	70
Table 2.7.	Selected properties of some water models used in MD simulation. HIPPO values are from the current work. Parameterization and data for other models are taken from: AMOEBA+, ⁵³ AMOEBA03, ⁹ AMOEBA14, ¹¹⁰ TTM3-F, ¹¹¹ SWM4-NDP and SWM6, ¹¹² MB-POL, ⁶² MB-UCB, ¹¹³ TIP3P, ¹¹⁴ TIP4P-Ew ¹¹⁵ and TIP5P. ¹¹⁶ Dimer energy and heat of vaporization are in kcal/mol, density in g/cm ³ , and diffusion coefficient as 10 ⁻⁵ cm ² /s.	72
Table 3.1.	Molecular polarizability of benzene. Experimental value from reference ⁴⁴ . Ab initio calculation performed in PSI4 software.	93
Table 3.2.	Benzene liquid properties at 298 K. a Reference 47, b Reference 48, c Reference 49, d Reference 50, e Reference 51.	98

Table 3.3.	Hydration free energy of benzene for different polarizable models.	100
Table 5.1.	35 molecules from Data set 1 (Appendix D) selected to undergo the first round of parameterization. These molecules had parameters validated with QM and experimental liquid data.	129
Table 5.2.	101 molecules from Data set 2 (Appendix D) that have finalized parameters, after fitting to available ab initio references from NCIA7-10 and DES370k11 databases.	130
Table 5.3.	HIPPO density compared to experimental values. Density is presented in kg/m ³ . Experimental values from references 12, 13.	133
Table 5.4.	HIPPO enthalpy of vaporization (H _{vap}) compared to experimental values. H _{vap} is presented in kJ/mol. Experimental values from references 12, 13.	134
Table 5.5.	HIPPO dielectric constant (ϵ) compared to experimental values. Dielectric constants (unitless) from references 12, 13.	135

Acknowledgments

My journey in science has been anything but linear, and numerous individuals have played important roles along its course. I owe a deep debt of gratitude to the community at Washington University, who embraced me as a fellow scientist. Their belief allowed me to fuel my lifelong passion for learning and to embark on the complexities of a fundamental scientific challenge.

Additionally, I would like to acknowledge the research grants that have facilitated the development of novel force fields, including the one detailed in my dissertation. Special mention goes to the National Institute of Health for their Grants Nos. R01 GM106137 and R01 GM114237.

I want to express my gratitude to the professors and mentors I had throughout my educational career. You were invaluable to my growth as a scientist.

To my advisor, Jay. You are a kind and understanding person who made my time as a graduate student smoother than I thought it would be. Your trust and respect for my opinions and ideas gave me the confidence to pursue my research.

Andrea and Melissa, your warmth and support as I integrated into the lab and the university were invaluable. You made the transition to a new country and environment infinitely smoother. Melissa, our conversations were crucial during some of my most challenging times.

To my high school teachers, especially Paulo Vicente and Azly. Paulo Vicente, you introduced me to Physics and are one of the reasons I decided to pursue the subject as a career. I will always be grateful for all you taught me. Azly, witnessing a math teacher who mirrored my reflection was an inspiration. You believed in me from the start and instigated my curiosity about the subject you taught.

To my friends and neighbors for over half my lifetime, Lena, Mada, Leo, Layse, and Jeovana. Your unwavering belief in my possibilities of success as a scientist was crucial. You have been there for me throughout all my years as a graduate student, encouraging and reminding me of all I had overcome to get where I am today.

To the friends in my entering class. I am eternally grateful and lucky I have met you at the beginning of my journey. You are all so kind and genuinely caring. Having you around helped me through many hardships over the last few years, especially during the Covid-19 pandemic.

Family is the compass that guides us, and mine has been my bedrock. Through adversities and challenges, their resilience has been awe-inspiring. I would not be here today if it weren't for my hard-working family. You all have faced many adversities and worked through impossible situations to ensure our family thrived.

To my parents. You have supported my endeavors in science even when you did not fully understand them. Through the most challenging years in high school and college, you were there for me through early mornings, getting home from college past midnight, and doing what you could to make my life easier. You worked double to support me, and I would not be here today if it weren't for your efforts in giving me and my siblings all you did not have.

To my siblings, Marilia and Emerson. I am grateful to have you both in my corner. You made it easier for me to be away and still be present in critical moments.

Lastly, to my nephews, Manoel and Mateus. You are the light in my life. Manoel, watching you grow, albeit from a distance, during my graduate studies has been a source of boundless joy. You always make me laugh, and I love seeing your face light up whenever I call you.

Roseane dos Reis Silva

Washington University in St. Louis

December 2023

ABSTRACT OF DISSERTATION

A Quantum Physics-Based Force Field for Biomolecular Simulation

by

Roseane dos Reis Silva

Doctor of Philosophy in Biology & Biomedical Sciences

Computational and Molecular Biophysics

Washington University in St. Louis, 2023

Professor Jay Ponder, Chair

This dissertation explores the development of a physics-based force field meticulously tailored to guide molecular dynamics simulations of biomolecules. At its core, the Hydrogen-like Intermolecular Polarizable Potential (HIPPO) force field emerges as an apex of achievement, devised to refine the precision of short-range intermolecular interactions. HIPPO's inception marks a profound stride towards heightened realism in molecular simulations, anchored in Quantum Physics theory and fortified by state-of-the-art Quantum Chemistry calculations. This force field's efficacy is attested by its systematic application in constructing models featuring diverse organic molecules, from water to benzene, and those with motifs resembling proteins and nucleic acids. HIPPO achieves an accuracy of 1 kcal/mol for each of its energy components of electrostatic, induction, repulsion, and dispersion when compared against *ab initio* Symmetry Adapted Perturbation Theory calculations while exhibiting striking conformity with an array of experimental bulk phase properties. HIPPO performs this without imposing a significant computational burden, thus positioning it comparably to the widely used AMOEBA force field.

Chapter 1. Introduction

Throughout the past three decades, the world has experienced a true revolution in research and using computers and computational tools for many aspects of our lives. Amongst the many promises the advances in computational power have brought, the one where it can aid the cure of many long-standing human diseases is of utmost interest and importance to society.

Many possibilities exist of how such a goal can be achieved. From large-scale studies of human genetics and the application of tools to help understand inner patterns to the prediction of protein structure from sequence to the development of new chemical compounds and beyond. Those are all areas where the computational revolution has had a profound impact. And it continues to allow us to explore science and help us tackle complex problems we have yet to answer, like how to better and faster create new drugs to treat human diseases.

Molecular simulations are at the core of solving this problem and many others. It combines our centuries of Mathematics, Physics, Chemistry, and Biology knowledge into computer code. It allows us to ask questions about the behavior of biological systems at an atomic level. Although its early development started almost a century ago, the research into adequately performing such simulations is still ongoing. And the recent advances in computational power have allowed us to explore more complex models that were almost impossible a few decades ago.

In the pursuit of predicting the dynamic behaviors and interactions of biological molecules, the field of biophysics faces the formidable challenge of applying foundational physical principles to complex, large systems. The Hydrogen-like Intermolecular Polarizable Potential (HIPPO) force field emerges as a pivotal atomistic approach to address this challenge. HIPPO dissects the intermolecular potential energy within biomolecular interactions into discrete and meaningful

components—electrostatics, polarization, dispersion, and exchange-repulsion. Each of these components was derived using Quantum Physics principles, bridging the gap there exists in classical force field potentials when it comes to the description of atomic interactions.

The present dissertation unfolds the development of a physics-driven force field tailored for the simulation of biomolecules. The HIPPO force field emerges as an achievement, meticulously designed to enhance the precision of short-range intermolecular interactions in large systems. Anchored in Quantum Physics theory and underpinned by cutting-edge Quantum Chemistry calculations, HIPPO's creation represents a profound step towards realism in molecular modeling. As a testament to its efficacy, I have methodically employed this force field to construct models encompassing diverse organic molecules, including water, benzene, and numerous counterparts adorned with protein and nucleic acid-like motifs. By integrating its functional form into molecular dynamics simulations, I will show that HIPPO achieves an impressive accuracy within 1 kcal/mol for each energy component, benchmarked against quantum calculations. Moreover, it shows close agreement with various experimental bulk phase properties while having comparable computational costs to the widely used AMOEBA force field.

This dissertation chronicles the unveiling of these molecular models, encompassing the intricate dance of intermolecular interactions, the characterization of condensed phase attributes, and the evolution of computational methodologies that empower such pursuits. Within this framework, my work resonates as an emblem of harnessing the capacities of advanced computing to invigorate the fidelity of protein simulations. Among the myriad possibilities, an outcome of profound consequence emerges—the potential to predict drug binding with uncanny precision, thereby edging closer to the expeditious realization of transformative treatments for human diseases.

1.1. Molecular dynamic simulations and force fields for biomolecules

The field of molecular dynamics simulations in Computational biophysics encompasses many areas. It starts with our Quantum Mechanics (QM) knowledge, which has described the atomic inner structure and how atoms interact through the Schrödinger equation. Its results ultimately lead to a thorough description of the forces acting on the system. From there, Newton's second law will tell us how the atoms move, given the forces acting on them, throughout time. The forces and the atomic and electronic structure are dynamic and constantly change with atoms' rearrangement. The result is a time-lapse snapshot showing the trajectory of the atoms in a molecular system. With that information, many questions can be asked, like how tightly small molecules bind to a protein target, what is the most likely protein conformation, how stable a particular crystal is, and so on.

If we could simply use QM for arbitrarily large systems, such as the coronavirus SARS-CoV-2 spike protein bound to the human receptor with over 30 thousand atoms, we would be done and have most of the answers we need. However, *ab initio* calculations (first-principles methods to solve Schrodinger's equation) are extremely expensive and scale poorly with the system's size. That means we can only use high-level *ab initio* methods for very small systems, having at most a few dozen molecules.

The solution to the problem is to approximate QM to the required level of accuracy while being feasible for large-scale simulations. This is achieved in Molecular Mechanics (MM) approaches, where atomic interactions are described by classical empirical potentials often referred to as force fields. Such models will have the same goal to explain how atoms behave and to yield the forces acting on each of them, which can then reveal how molecules move through time. Many different force fields exist at varying approximations, accuracy, and computational cost levels.¹

The most straightforward and least expensive is the Point Charge force field, introduced almost half a century ago and remains the most widely used model for biomolecular simulations.²

1.1.1. Point Charge Models

In Quantum Mechanics, atoms are described as having a positively charged nucleus surrounded by orbitals that hold electrons. In Point charge force fields, each atom is approximated as having a fixed-point charge at the atom center, and they interact electrostatically through Coulomb's law. Besides that, a van der Waals empirical term compares QM's equivalent of repulsion and dispersion. Simple harmonic potentials represent the bonded valence interactions.

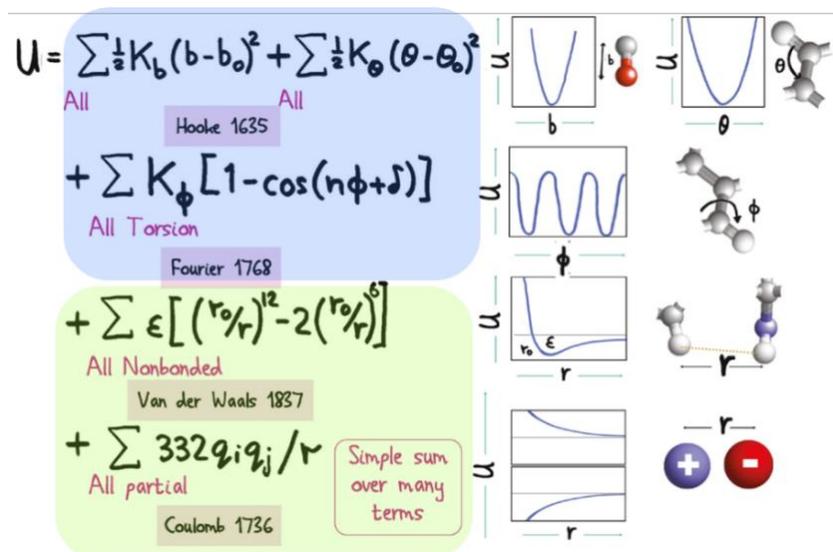


Figure 1.1. The Point Charge Force Field model.² A classical, empirical model with intramolecular (shaded blue) and intermolecular (shaded green) energy terms. (Reproduced from Michael Levitt's Nobel lecture)

Point charge models were crucial for the advancement of computational biophysics as they allowed many scientific discoveries in biomolecular systems and have helped establish the field.³ Numerous protein conformational studies using these models showed how valuable molecular mechanics can be for biology and how much we could gain if the model were expanded. Thanks to the exponential increase in computational resources since the model was first introduced,

simulation accuracy can be significantly improved by switching the atomic description to ones closer to the correct physical representation provided by QM.⁴

One of the main shortcomings of the simple, point charge description is its purely pairwise nature, far from the quantum description. It misses capturing the responses atoms can have to one another, commonly referred to as the many-body effect. Another apparent problem is the poor description of electrostatics. In classical electrostatics, a charge distribution can be expanded in multipole contributions of charge, dipole, quadrupole, etc. Although the charge is usually the most significant contributor to the electrostatics energy, many molecules have important higher-order moments. Although the inclusion of more complex terms can reduce the performance of computer simulations, the inclusion of many-body effects and polarization can significantly improve the calculated potential energy surface of biomolecules.^{1, 4, 5} But, with increased computational capabilities, these more complex models can be used in biology to improve accuracy when point charge potentials are incomplete and fail to reproduce some molecular properties.

One practical example of how the incorrect potential can influence the course of a simulation is shown in Figure 1.2. This figure shows the performance of current force fields in predicting the fold of the UUCG RNA tetraloop.⁶ The experimental structure of this RNA has been elucidated by NMR (Nuclear Magnetic Resonance). It is known to spend more than 90% of its time in the conformation represented by cluster 5.

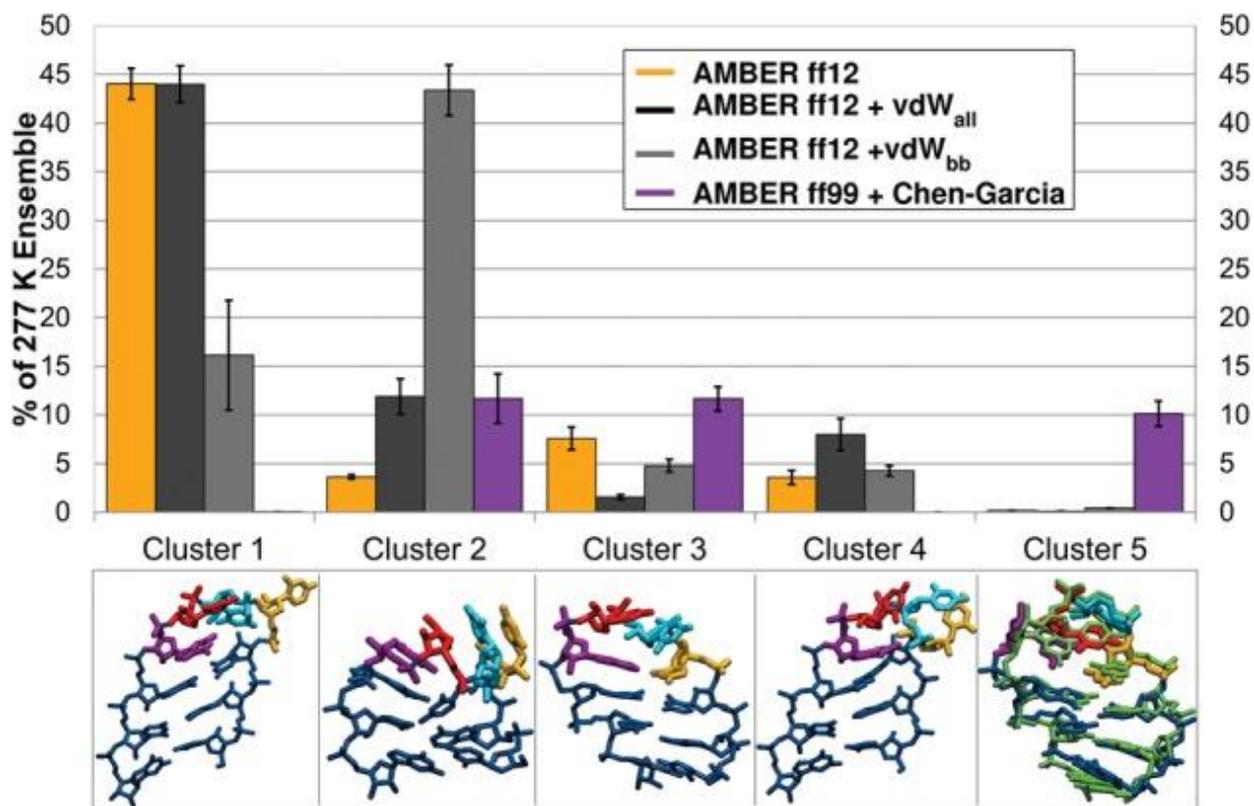


Figure 1.2. Predicted structures of the UUCG RNA tetraloop. The experimental structure is Cluster 5. No version of the AMBER force field can reproduce the correct conformation. Reproduced from reference⁶.

1.1.2. Polarizable models

In molecular mechanics, a polarizable site has the ability to respond to its electrostatic environment. In other words, classical polarization is the effect that other charged atoms have on a given atomic site. This effect can be included in classical models through different routes, and several polarizable force fields exist, either by including fluctuating charges, inducible dipoles, or putting charges on a spring, like the Drude Oscillators.⁴

In this group, the AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) model provides a straightforward approach to including some of the terms lacking in the simpler model.⁷ In AMOEBA, the electrostatic potential was expanded to include permanent

atomic monopole (charge), dipole, and quadrupole moments at each atomic center. Moreover, an explicit polarization term is included via the mutual induction of dipoles at the atomic center. Various models exist for this force field, including water⁷, organic molecules⁸, and nucleic acids⁹.

This force field has been implemented using fast GPUs (Graphics Processing Units), dramatically improving the model's performance, despite its greater complexity compared to point charge models. The advances of GPU and massive parallelization have allowed AMOEBA to be applied to large biological systems, including the SARS-CoV-2 protease.¹⁰

The present dissertation will present the development of molecular models for a new polarizable potential called HIPPO (Hydrogen-like Intermolecular Polarizable Potential), which was built upon the backbone of the AMOEBA force field. HIPPO, however, abandons the longstanding classical description of atoms as point charges (or points dipoles and quadrupoles.) Instead, it describes the atoms as charge distributions, and almost every term is derived using this notion.

HIPPO falls into the new category of physics-based force fields, where every term has a stronger connection to the rigorous QM equivalent. Moreover, the parametrization procedure involves calibration to high-level *ab initio* calculations before fitting against experimental condensed-phased data. Throughout this dissertation, I will show the advantages of having a force field model that is physically grounded yet capable of reproducing experimental data. This model has a new level of complexity compared to AMOEBA but has suffered no significant reduction in performance, presenting itself as a clear next-generation to the AMOEBA model. Because Quantum Chemistry and *ab initio* methods are at the core of HIPPO's model and inception, the next session will be dedicated to such topics. Then, a review of HIPPO's theoretical development will be provided.

1.2. *Ab initio* methods and the parametrization of classical force fields

All classical force fields have parameters to determine for every atom or classes of atoms. In the past, force fields parameters were determined based solely on experimental, condensed-phase data. This approach can lead to highly empirical parameters that are not transferable between chemically similar molecules. Another issue there is great reliance in cancellation of errors during fitting, leading to difficult to parametrize models.

Computational hardware advances and ongoing research in Quantum Chemistry have made *ab initio* calculations more accessible¹¹, allowing their use for the parametrization of classical models. An example that has been successful in AMOEBA is using QM methods for computing the multipoles on each atom⁸, which HIPPO has inherited. QM intermolecular energies computed between small cluster of molecules are also quantities that can be valuable to parametrization and validation, which can now be used in large scale parametrization of force fields.

However, the fact that classical models must split the total energy of interaction into components, such as electrostatics and van der Waals, is not ideal, since those are not QM observables. A solution to the issue is to assign these quantities theoretically using *ab initio* Energy Decomposition Analysis (EDA) schemes. The most powerful method for determining force field functional form is Symmetry Adapted Perturbation Theory (SAPT), which uses perturbation theory to decompose total intermolecular energy into electrostatics, induction, dispersion, and exchange-repulsion components¹²⁻¹⁴. SAPT was chosen as the basis for developing and parameterizing the HIPPO model^{5, 15}, because it provided the ideal blueprint for a first principles physics-based functional form to replace the empirical Point Charge function (Rackers, 2019.)¹⁶

The largest part of my work in the HIPPO force field was coming up with a way to develop robust, automated, and parallel methods for parametrization. Although the functional form of HIPPO is physically more accurate, it comes with more parameters that needs to be determined. Hence, making this parametrization method and obtaining good parameters are just as essential for the success and usage of the force field in practical applications.

During the parametrization, I have used thousands of SAPT calculations, computing a few of them myself, and they are essential for this dissertation. However, a thorough review of the method is beyond the goal. Important work and reviews of the method are referenced.¹²⁻¹⁴

1.2.1. Symmetry Adapted Perturbation Theory (SAPT)

Intermolecular energies are considered very small when compared to the energy of intramolecular bonds. For this reason, it can be treated as a perturbation to the isolated monomers. To compute and decompose the interaction energy of a dimer, SAPT applies perturbation theory to perform the quantum calculations. As shown in Figure 1.3, the idea is that the Hamiltonian (the quantum operator representing the total energy of a system) of the dimer complex is going to be represented as the sum of the Hamiltonian of the monomers plus a perturbation term representing the intermolecular interaction operator, V . The calculation proceeds in orders of perturbation, full derivation in ¹⁶.

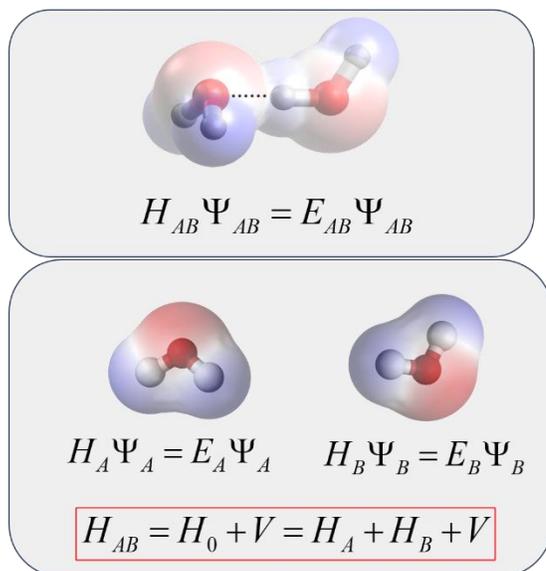


Figure 1.3. SAPT approach to intermolecular energy decomposition for two monomers A and B.

Internal energy: solving the equation using only the H_{AB} Hamiltonian yields the internal energy of the monomers, isolated. The zeroth order perturbation is not usually used in SAPT.

Electrostatics: the first order perturbation yields Coulomb's law. It is the electrostatics interaction between the electronic charge densities of monomers A and B, in their original state.

$$E_1 = \iint \rho_A \frac{1}{r} \rho_B dV^2 \quad (1.1)$$

Induction: the second order and higher perturbations gives the induction, which is the deformation of one monomer charge density in response to the other one.

$$E_2 (ind) = E_2(ind, B \rightarrow A) + E_2(ind, A \rightarrow B) \quad (1.2)$$

Dispersion: the second order perturbation has a “left-over” part, which is due to the interaction between the mutually perturbed charge density of the monomers. It is the instantaneous response of the charge density to the correlated changes in the other monomer density.

$$E_2 (disp) = E_2 + E_2(ind) \quad (1.3)$$

Repulsion: the last term on the SAPT decomposition is less straightforward than the other ones. It is called exchange-repulsion and it arises from the application of Pauli's exclusion principle requirement that all electronic wavefunctions are antisymmetric. The wavefunctions coming from the perturbations were not antisymmetric, and SAPT applies a method to symmetrize those wave functions and then recalculate each of the previous components. The repulsion term is then the difference between the energies computed using the antisymmetric versus originally symmetric wavefunctions.

$$E_{exch} = \sum_{i=0}^2 E_i^{SAPT} - E_i \quad (1.4)$$

The SAPT calculations can be performed with different numbers of perturbations¹⁷, starting with the zeroth order, which provides the sum of the internal energy of the monomers; the first order perturbation yields electrostatic, referred to as SAPT0; higher orders will provide dispersion and induction. All the calculations I performed used SAPT2+, with the aug-cc-pVTZ bases set.

1.2.2. Coupled Cluster methods for total intermolecular energy

SAPT was the main *ab initio* method for parametrization of the components of HIPPO force field of electrostatics, induction, dispersion and repulsion. However, the total energy of the interaction was often fitted against the highest-level calculation available. In many occasions, the highest-level calculation was the SAPT total energy (the sum of its components). But, for many cases, a total intermolecular energy computed with the *ab initio* method called CCSD(T) (Coupled Cluster Singles Doubles with perturbative (Triples))¹⁸ excitations is available. In these cases, the energy components were fit to SAPT references, whereas the total energy was fit to the CCSD(T) total, considered one of the gold standard method for non-covalent interactions. These references usually came from the published databases of intermolecular energy of dimer systems.

1.3. Quantum Chemistry databases

The calibration of parameters within the HIPPO force field involves aligning them with high-level quantum data, a process that demands computationally intensive calculations. Consequently, leveraging published databases to access required data for parameterization is essential. In the construction of the models showcased in this dissertation, three key databases were used: S101x7¹⁹, the Non-Covalent Interactions Atlas²⁰⁻²³ (NCIA), and the DES370k²⁴ databases.

1.3.1. The S101x7 database

This database contains 101 dimers involving hydrogen, carbon, nitrogen, oxygen, phosphorus, sulfur, fluorine, chlorine, and bromine. The molecules selected covers common chemical space found in organic and biological systems. It was specifically designed to capture the so-called charge penetration effect missing in force fields modeling atoms as points, including the AMOEBA force field, which uses point multipoles.¹⁹ An illustrative representation of this set is shown in **Figure 1.4**.

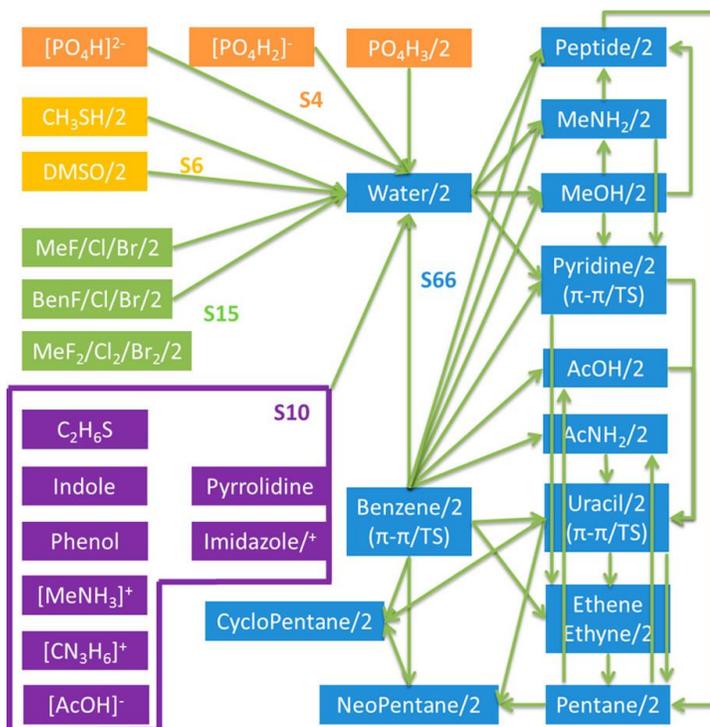


Figure 1.4. Schematic representation of the dimers in the S101 Database. The arrows connecting the molecules indicates a dimer; the “/2” designation indicates a homodimer; the “/+(-)” notation indicates the presence of both neutral and charged species to the molecule. Reproduced from reference¹⁹.

This was the initial set used on HIPPO's parametrization and allowed for validation and theoretical calibration of the potential. In this dissertation the molecules included in databases are treated aren't treated separately. They were added to a larger dataset of molecules, and all quantum information available for each molecule across different databases were condensed in one reference set. This is also true for the other QM databases. The notation 'S101x7' shall be used to refer to data coming from this particular set.

1.3.2. Non-Covalent Interactions Atlas (NCIA)

The Non-Covalent Interactions Atlas²⁰⁻²³ project provides a wide range of non-covalent intermolecular interactions with benchmark energy computed using advanced quantum chemical methods. This is a large dataset encompassing common organic elements, and investigating particular kinds of interactions, like sigma-holes, dispersion-drive systems, hydrogen bonding, etc. This project is even more remarkable considering it provides 10 or 5 points of the dissociation curve for the dimers at the highest level of quantum calculation available. The total intermolecular energies are computed using the coupled-cluster singles and doubles with perturbative triples [CCSD(T)] method

The NCIA data is available at their webpage (nciatlas.org) where it is subdivided into different categories of interactions. The sets I have used for my project are listed below.

- a. *SH250x10*, sigma-hole interactions. This dataset includes molecules involving bonds of the elements chlorine, bromine, iodine, sulfur, selenium, phosphorus and arsenic. These specific non-covalent interactions result from the existence of a sigma-hole, a region of positive electrostatic potential on an otherwise electronegative atom.²³ For this work, I have excluded the molecules where Selenium and Arsenic were present.

This dataset is being used for validation of the HIPPO parametrization procedures, since it involves various heterodimers not used during fitting of parameters.

- b. *R739x5*, repulsive contacts in an extended chemical space. This data set focuses on repulsive contacts in molecular complexes, covering organic molecules, sulfur, phosphorus, and halogens.²⁰ This set was crucial for the development of HIPPO, since data on conformations far from the global minima are not usually available.
- c. *HB375x10*, hydrogen-bonding in organic molecules. This dataset offers a collection of dimers with hydrogen bonds of OH, NH and CH groups with oxygen and nitrogen, plus a control group of complexes of the same molecules without H-bonds.²² Perhaps the most important interaction in biomolecular systems, hydrogen bonds were extensively evaluated in HIPPO using this database, proving its natural ability to capture such interactions in many configurations.
- d. *HB300SPXx10*, hydrogen bonding extended to sulfur, phosphorus and halogens. As the name suggests, this database is an extension of the original hydrogen-bonding *HB375x10* database.²¹

These four databases will be often presented as the target for the intermolecular energies HIPPO. The legend of plots will use the names just provided, *SH250x10*, *R739x5*, *HB375x10*, and *HB300SPXx10*.

1.3.3. DES370k database

The DES370k dataset contains interaction energies for more than 370 thousand of dimer geometries computed using CCSD(T).²⁴ It contains almost four thousand distinct dimer pairs, covering a wide chemical space and types of non-covalent interactions. Besides providing total

interaction energies, this database also provides SAPT0¹⁷ calculations for several of the conformations.

1.4. A database for organic liquids

The parametrization of HIPPO using organic liquid experimental data drew inspiration from Caleman et al.'s publication²⁵ that assesses the efficacy of classical models in replicating condensed phase properties. This publication, along with its accompanying database, enables a direct comparison between HIPPO models and prevalent point charge models.

In this study, a comprehensive evaluation of classical force fields with respect to their ability to replicate condensed phase properties of organic liquids is presented. Several key properties are considered, including density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. The aim was to provide a benchmark for assessing the performance and accuracy of force fields used in molecular simulations of organic compounds. By doing so, this work provided valuable insights into the strengths and limitations of the evaluated models, which was useful for guiding HIPPO in the direction of addressing shortcomings of current models. All of the 146 molecules included in this database were used.

1.5. HIPPO: Hydrogen-like Intermolecular Polarizable Potential

The standard biomolecular simulation model for over three decades has been the point charge force field. This model has been successful in folding proteins and reproducing enzyme-inhibitor binding interactions. However, it has been found that this standard model is missing some key physics. The HIPPO model, on the other hand, is a new class of force field that includes the most relevant and important physics from the start. It is derived and parameterized to explicitly

reproduce the *ab initio* energy components from Symmetry Adapted Perturbation Theory (SAPT). The HIPPO model introduces a model electron density around every atom, departing from the atoms-as-points model of the standard force field. The HIPPO model offers several improvements over conventional models. It is able to reproduce each separate component of the intermolecular energy relative to SAPT within chemical accuracy. The inclusion of a charge density model in the HIPPO model solves the longstanding "charge penetration problem" in molecular modeling. The polarization model in the HIPPO model yields better molecular polarizabilities than leading polarizable force fields. The dispersion functional form produces a damping function with true physical meaning. The exchange-repulsion describes the anisotropy of halogen bonding with drug molecules more accurately than any alternative force field. Overall, our HIPPO model works naturally for simulating water and a variety of other organic molecules, as the results of my work will show.

The HIPPO model is a new class of force field that includes important physics from the start and is derived and parameterized to reproduce *ab initio* energy components. It offers improvements over conventional models in terms of reproducing intermolecular energy components, solving the charge penetration problem, and accurately describing polarization, dispersion, and exchange-repulsion effects.

The first chapter will have a deeper introduction to the model and its parameters.

1.6. Structure of this dissertation

This opening chapter has provided an overview of the critical importance and challenges associated with molecular simulations. Within this context, a brief background on force fields and

quantum data was presented. Central to the discussion was an introduction to the HIPPO model, emphasizing its significance and objectives in addressing current challenges in the field.

Drawing its core content from our inaugural publication on the subject¹⁵, Chapter 2 offers a detailed introduction to the HIPPO force field. A specific focus is given to the development and validation of the water model nested within the HIPPO framework. The outcomes and implications of these findings are critically discussed in light of existing knowledge in the broader scientific community.

Chapter 3 delves deeply into the benzene model, spotlighting its pivotal role as a foundational organic molecule. It will provide insights into the model's development, its validation process, and potential applications. Additionally, this chapter provides a sneak peek into our soon-to-be-released publication centered on this topic.

In Chapter 4, the challenges associated with parametrizing force fields are introduced. To address these challenges, I detail the software tools I developed, designed to automate the HIPPO parametrization process. Beyond parametrization, the analytical utilities crafted for improved simulation analysis are expounded upon. The chapter concludes with a reflection on potential enhancements and the future trajectory of the software.

Chapter 5 ventures into the diverse range of organic molecules that I have parameterized using the HIPPO model. A meticulous presentation of results is provided, highlighting the challenges faced, the solutions derived, and how these findings compare with pre-existing models and force fields.

Taking a forward-looking stance, the final Chapter 6 embarks on an exploration of the forthcoming projects envisioned for the HIPPO model. Noteworthy discussions encompass the

envisaged expansion of the parametrization tool, especially its symbiosis with machine learning-driven force fields, and preliminary insights into the creation of a protein force field under the HIPPO umbrella. The chapter wraps up by pondering the potential challenges ahead and strategies earmarked to navigate them.

1.7. References

1. Dauber-Osguthorpe P, Hagler AT. Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there? *J. Comput. Aid. Mol. Des.*, *33*, 133-203 (2019).
2. Mccammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature*, *267*, 585-90 (1977).
3. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. *Neuron*, *99*, 1129-43 (2018).
4. Baker CM. Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.*, *5*, 241-54 (2015).
5. Rackers JA, Wang Q, Liu C, Piquemal J-P, Ren P, Ponder JW. An Optimized Charge Penetration Model for Use with the AMOEBA Force Field. *Phys. Chem. Chem. Phys.*, *19*, 276-91 (2017).
6. Bergonzo C, Henriksen NM, Roe DR, Cheatham TE. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA*, *21*, 1578-90 (2015).
7. Ren P, Ponder JW. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *The Journal of Physical Chemistry B*, *107*, 5933-47 (2003).
8. Ren P, Wu C, Ponder JW. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.*, *7*, 3143-61 (2011).
9. Zhang C, Lu C, Jing Z, Wu C, Piquemal J-P, Ponder JW, Ren P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.*, *14*, 2084-108 (2018).
10. El Khoury L, Jing Z, Cuzzolin A, Deplano A, Loco D, Sattarov B, Hédin F, Wendeborn S, Ho C, El Ahdab D, Jaffrelot Inizan T, Sturlese M, Sosic A, Volpiana M, Lugato A, Barone M, Gatto B, Macchia ML, Bellanda M, Battistutta R, Salata C, Kondratov I, Iminov R, Khairulin A, Mykhalonok Y, Pochevko A, Chashka-Ratushnyi V, Kos I, Moro S, Montes M, Ren P, Ponder JW, Lagardère L, Piquemal J-P, Sabbadin D. Computationally driven discovery of SARS-CoV-2 M^{pro} inhibitors: from design to experimental validation. *Chemical Science*, *13*, 3674-87 (2022).
11. Turney JM, Simmonett AC, Parrish RM, Hohenstein EG, Evangelista FA, Fermann JT, Mintz BJ, Burns LA, Wilke JJ, Abrams ML, Russ NJ, Leininger ML, Janssen CL, Seidl ET, Allen WD, Schaefer HF, King RA, Valeev EF, Sherrill CD, Crawford TD. Psi4: an open-source ab initio electronic structure program. *WIREs Comput. Mol. Sci.*, *2*, 556-65 (2012).
12. Jeziorski B, Moszynski R, Szalewicz K. Perturbation Theory Approach of Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.*, *94*, 1887-930 (1994).
13. Rybak S, Jeziorski B, Szalewicz K. Many-Body Symmetry-Adapted Perturbation Theory of Intermolecular Interactions. H₂O and HF Dimers. *J. Chem. Phys.*, *95*, 6576-601 (1991).

14. Szalewicz K. Symmetry-Adapted Perturbation Theory of Intermolecular Forces. *WIREs Comput. Mol. Sci.*, 2, 254-72 (2012).
15. Rackers JA, Silva RR, Wang Z, Ponder JW. Polarizable Water Potential Derived from a Model Electron Density. *J. Chem. Theory Comput.*, 17, 7056-84 (2021).
16. Rackers JA. A Physics-Based Intermolecular Potential for Biomolecular Simulation: Washington University in St. Louis; 2019.
17. Parker TM, Burns LA, Parrish RM, Ryno AG, Sherrill CD. Levels of Symmetry Adapted Perturbation Theory (SAPT). I. Efficiency and Performance for Interaction Energies. *J. Chem. Phys.*, 140, 094106 (2014).
18. Purvis GD, III, Bartlett RJ. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of Chemical Physics*, 76, 1910-8 (1982).
19. Wang Q, Rackers JA, He C, Qi R, Narth C, Lagardere L, Gresh N, Ponder JW, Piquemal J-P, Ren P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.*, 11, 2609-18 (2015).
20. Kříž K, Nováček M, Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets 3: Repulsive Contacts. *J. Chem. Theory Comput.*, 17, 1548-61 (2021).
21. Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets 2: Hydrogen Bonding in an Extended Chemical Space. *J. Chem. Theory Comput.*, 16, 6305-16 (2020).
22. Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.*, 16, 2355-68 (2020).
23. Kříž K, Řezáč J. Non-covalent interactions atlas benchmark data sets 4: σ -hole interactions. *Phys. Chem. Chem. Phys.*, 24, 14794-804 (2022).
24. Donchev AG, Taube AG, Decolvenaere E, Hargus C, Mcgibbon RT, Law K-H, Gregersen BA, Li J-L, Palmo K, Siva K, Bergdorf M, Klepeis JL, Shaw DE. Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Scientific Data*, 8, (2021).
25. Caleman C, van Maaren PJ, Hong M, Hub JS, Costa LT, van der Spoel D. Force Field Benchmark for Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.*, 8, 61-74 (2012).

Chapter 2. The HIPPO Force Field and Water Model

This chapter will provide a complete description of the HIPPO model and how it stands in the realm of current force fields. The work encapsulates my initial two years as a force field developer, a journey filled with challenges and growth. During this period, I honed my skills in graphics card programming to enhance the model's speed and laid the foundational groundwork for the force field's parametrization procedure. While perfecting the model required numerous iterations and patience, the outcome is a testament to the attainability of a swift and precise biomolecular force field.

2.1. Introduction

Water is perhaps the most studied of all molecules, both experimentally and theoretically. In addition to its obvious importance for life on Earth, water is of interest due to: (1) its unique physical properties, including a density maximum near 4°C with normal ice being less dense than the liquid, (2) its ability to solvate a wide range of disparate chemical species, (3) the great variety of its solid-phase crystal forms and richness of its phase diagram, and (4) its paradigmatic hydrogen bonding interaction and the related hydrophobic effect. The first atom-based water potential available as a quantitative model dates back nearly a century to the work of Bernal and Fowler.¹ The ST2 model of Rahman and Stillinger,² among other models from that period, was suitable for use in some of the initial molecular dynamics (MD) simulations. During the early 1980s, the TIPS³ and SPC⁴ families of water potentials were developed, and they are still used in present day modeling projects. Since that time, a large number of additional water models have been proposed for use in simulation, ranging from coarse-grained empirical functions that represent several molecules by a single-site particle,^{5, 6} to detailed density functional theory-based (DFT) MD calculations,⁷ to massive simulations using machine-learned potentials.⁸

Here we propose a new water model near the intersection of empirical models fit to reproduce macroscopic properties, and *ab initio* models derived entirely from first-principles physics. This new model, referred to as HIPPO for Hydrogen-like Intermolecular Polarizable Potential, is derived directly from a model electron density obtained from *ab initio* results and electronic structure theory, but then parameterized to improve agreement with target experimental data. As such, the HIPPO water model provides a computationally efficient form for use in large-scale simulations, while allowing for analysis and decomposition in terms of physically validated energetic components.

In one sense, this new model is a natural extension of previous polarizable force fields. In particular, it extends the logic that has made the AMOEBA force field successful.^{9, 10} The main advance of AMOEBA was to show that intermolecular interactions at medium range cannot always be handled through cancellation of errors, as they are in point charge force fields.¹¹ This insight motivated the inclusion of dipole polarization and atom-centered multipoles into the model. Much recent work, however, shows that despite AMOEBA's more elaborate functional form, it still relies on significant error cancellation at short range. The archetypal example of this behavior is the p-stacking interaction, exemplified by the benzene dimer.^{12, 13} Studies of this system have shown that despite its atomic multipole and polarization terms, AMOEBA exhibits some of the same short-range problems as simpler force fields. A principal aim of the HIPPO model is to reduce this kind of reliance on error compensation at short range.

The way in which HIPPO achieves this aim, however, makes it more than a simple extension of AMOEBA. Much of the short-range error in force fields is due to their reliance on point approximations and the lack of an explicit charge density. In the p-stacking case, for example, the error in the electrostatic interaction is given a widely adopted name: charge penetration.

Analogous errors occur in other force field components, but they all arise from the same inappropriate density treatment.

HIPPO addresses this problem directly by including a description of the electron density explicitly in the model. It is far from the first empirical potential to include a model for the density; other models, most notably the Gaussian Electrostatic Model (GEM),¹⁴⁻¹⁷ have made use of explicit charge densities. However, HIPPO is the first force field to use an electron density model in constructing each component of the total potential. As we will detail in the Theory section, every non-valence term derives its form from charge densities and the interactions between them. This distinction makes HIPPO a new class of density-based model.

The choice of density-based form is not arbitrary, as HIPPO follows from Symmetry Adapted Perturbation Theory (SAPT) quantum energy decomposition analysis.^{18,19} SAPT divides the total interaction energy of a system into four physically meaningful components: electrostatics, polarization, Pauli repulsion and dispersion. Importantly and as its name implies, SAPT does this through the use of perturbation theory. The base, or unperturbed state, is represented by isolated molecules, and the energy components are computed as perturbations from that state as two molecules are allowed to interact. This perturbation theory logic lends itself well to classical approximation. As is detailed in previous work, each HIPPO term uses the atomic electron density model to construct a classical equivalent of the corresponding SAPT term.²⁰⁻²² In this way, HIPPO is not just parameterized against SAPT components; it can itself be considered a classical approximation of SAPT.

Conceptually, one might be tempted to assume the elaborate functional form of HIPPO would lead to a large increase in computational cost over similar polarizable models. This, however, misjudges the nature of the complexity in the underlying model. Atomic charge densities

only overlap at short range, and the highest cost additions to HIPPO are restricted to the relatively few interactions in that regime. In this way, HIPPO is able to employ a more complex functional form while maintaining a computational cost roughly equal to that of other polarizable force fields.

The following sections provide: (1) a unified summary of the theoretical underpinnings of the portions the HIPPO force field needed for a water model, (2) a description of the computational and simulation methodologies used, (3) HIPPO results compared against quantum mechanical and experimental data for gas phase clusters, liquid water and ice, and (4) discussion of strengths and limitations of the HIPPO model and the suitability of SAPT as a framework for force field development.

2.2. Theory

In the HIPPO force field, every atom is represented by two components: a model valence electron density and a core point charge. The atomic electron density, illustrated in Figure 2.1, emulates that of a hydrogen-like atom,

$$\rho_{HIPPO} = \frac{Q\zeta^3}{8\pi} e^{-\zeta r} + Z \delta(r) \quad (1)$$

where Q is the valence charge of the atom, Z is the core charge, ζ controls the width of the electron density, and δ is the Kronecker delta function. The HIPPO density also includes consistent higher-order atomic dipole and quadrupole terms for describing anisotropy. This model density is used to derive all four intermolecular energy terms that compose the HIPPO force field,

$$U_{HIPPO} = U_{electrostatic} + U_{induction} + U_{dispersion} + U_{Pauli\ repulsion} \cdot \quad (2)$$

The general forms and derivations of these terms have been detailed in several references ^{20, 21, 22} describing the piecewise development of the model. To provide a unified picture, we present here a comprehensive definition of each term.

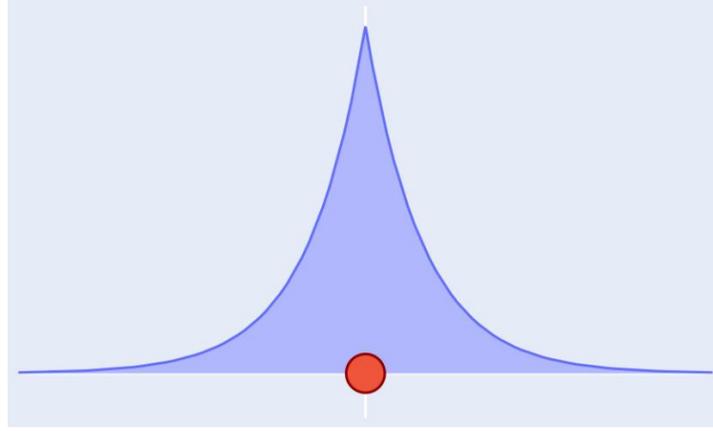


Figure 2.1. Schematic of a HIPPO atom. The blue shaded area represents the valence electron density, and the red point represents the point core charge.

Electrostatic Energy. Like its progenitor, the AMOEBA force field, the HIPPO electrostatic term is anisotropic, utilizing atomic multipole moments through the quadrupole. Since each atom in the model is represented by a core charge and a smeared density, the pairwise Coulomb interaction has four components. The HIPPO electrostatic energy is defined as,

$$U_{electrostatic}^{HIPPO} = \sum_{i>j} Z_i T_{ij} Z_j + Z_i \mathbf{T}_{ij}^* \vec{M}_j + Z_j \mathbf{T}_{ji}^* \vec{M}_i + \vec{M}_i \mathbf{T}_{ij}^{overlap} \vec{M}_j \quad (3a)$$

$$\vec{M} = \left(Q, [\mu_x, \mu_y, \mu_z], \begin{bmatrix} \Theta_{xx} & \Theta_{xy} & \Theta_{xz} \\ \Theta_{yx} & \Theta_{yy} & \Theta_{yz} \\ \Theta_{zx} & \Theta_{zy} & \Theta_{zz} \end{bmatrix} \right) \quad (3b)$$

$$T_{ij} = \frac{1}{r_{ij}} \quad (3c)$$

$$\mathbf{T}_{ij}^* = [1 \quad \nabla \quad \nabla^2] \left(\frac{1}{r_{ij}} f_{ij}^{damp}(r_{ij}) \right) \quad (3d)$$

$$\mathbf{T}_{ij}^{overlap} = \begin{bmatrix} 1 & \nabla & \nabla^2 \\ \nabla & \nabla^2 & \nabla^3 \\ \nabla^2 & \nabla^3 & \nabla^4 \end{bmatrix} \left(\frac{1}{r_{ij}} f_{ij}^{overlap}(r_{ij}) \right) \quad (3e)$$

where the first term represents the core-core repulsion, the second and third terms represent the core–density attractions and the fourth term represents the density–density repulsion. The \mathbf{M} vector contains the multipole moments (charge, dipole and traceless quadrupole) and Q and Z represent the core and density charges constrained to satisfy the relation for the total atomic partial charge $q_i = Z_i + Q_i$. The f^{damp} and $f^{overlap}$ terms in equations 3d and 3e are of critical importance. They result directly from the electrostatic potential generated by the model density,

$$V(r) = \frac{Q}{r} \left[1 - \left(1 + \frac{1}{2} \zeta r \right) e^{-\zeta r} \right] . \quad (4)$$

This gives the core–density attractions,

$$U_{core-density} = Z_i V_j(r_{ij}) = Z_i \left(\frac{1}{r_{ij}} f_{ij}^{damp}(r_{ij}) \right) q_j \quad (5a)$$

$$f_{ij}^{damp}(r_{ij}) = 1 - \left(1 + \frac{1}{2} \zeta_j r_{ij} \right) e^{-\zeta_j r_{ij}} \quad (5b)$$

yielding the “one-center” damping factor that goes into \mathbf{T}^* . The density–density repulsion is given by

$$U_{density-density} = \frac{1}{2} \left[\int \rho_i(\mathbf{r}) V_j(\mathbf{r}) dv + \int \rho_j(\mathbf{r}) V_i(\mathbf{r}) dv \right] = q_i \left(\frac{1}{r_{ij}} f_{ij}^{overlap}(r_{ij}) \right) q_j \quad (6a)$$

$$f_{ij}^{overlap}(r_{ij}) = \begin{cases} 1 - \left(1 + \frac{11}{16}\zeta r_{ij} + \frac{3}{16}(\zeta r_{ij})^2 + \frac{1}{48}(\zeta r_{ij})^3\right) e^{-\zeta r_{ij}}, & \zeta_i = \zeta_j \\ 1 - A^2 \left(1 + 2B + \frac{\zeta_i}{2} r_{ij}\right) e^{-\zeta_i r_{ij}} - B^2 \left(1 + 2A + \frac{\zeta_j}{2} r_{ij}\right) e^{-\zeta_j r_{ij}}, & \zeta_i \neq \zeta_j \end{cases} \quad (6b)$$

$$\text{with } B = \frac{\zeta_i^2}{\zeta_i^2 - \zeta_j^2} \text{ and } A = \frac{\zeta_j^2}{\zeta_j^2 - \zeta_i^2}, \quad (6c)$$

where the integrals are evaluated according to the method of Coulson.²³ The $f^{overlap}$ term is the “two-center” damping factor necessary to compute the fourth term of the HIPPO electrostatic potential energy. The terms necessary for higher-order multipole interactions are obtained by successive gradient operations applied to each of the damping factors as specified in equations 3d and 3e. In the interest of clarity, the explicit equations for all orders of the multipole interaction energy are enumerated in Appendix A. In the limit of large a , both damping factors tend to unity and the undamped point multipole interaction energy is recovered. In practice, the use of finite densities remedies the well-documented charge penetration problem of electrostatics.^{13, 24-29} In total, the HIPPO electrostatic model has five parameters per atom: a core charge Z , a valence charge Q , a dipole moment m , a quadrupole moment Q , and a “charge penetration” damping parameter ζ .

Induction Energy. In addition to the permanent core charge and density-based multipoles, HIPPO includes a point inducible dipole at every atomic site. The induction energy of the model is defined as,

$$U_{induction}^{HIPPO} = \sum_i \frac{1}{2} \vec{\mu}_i^{ind} \vec{F}_i^{perm} - \sum_{i>j} \varepsilon_i e^{-\eta_j r_{ij}} + \varepsilon_j e^{-\eta_i r_{ij}} \quad (7)$$

where the first term represents the polarization energy of the induced dipoles interacting with the permanent electric field and the second term represents a small pairwise exponential charge transfer function. The polarization term is the source of many-body energy in the force field. The induced dipoles are determined by solving the system of linear equations,

$$\boldsymbol{\mu} = \boldsymbol{\alpha}(\mathbf{F}^{perm} + \mathbf{F}^{ind}) \quad (8)$$

where the vectors are defined as $\mathbf{m} = [m_1, m_2, m_3, \dots, m_n]$ and similarly for \mathbf{F}^{perm} (the field due to the permanent multipoles), \mathbf{F}^{ind} (the field due to the induced dipoles) and $\boldsymbol{\alpha}$ (the atomic polarizabilities). The permanent and induced electric fields are calculated in the same manner, with the same parameters, as described in the previous section. In this way, the electric fields for the polarization model are completely consistent with the permanent electrostatics portion of the model. For completeness, the full equations describing polarization are detailed in Appendix B. The only additional parameter necessary for the polarization model is the atomic polarizability of each atom, denoted by a . Finally, the charge transfer function requires two parameters per atom: a prefactor e and an atom-based damping factor h .³⁰

Dispersion. The dispersion interaction between atoms arises from the interaction energy of correlated, instantaneous induced dipole moments. In the point approximation, this gives the canonical $1/r^6$ dependence associated with London dispersion.³¹ Because the HIPPO model represents valence electrons as densities, the functional dependence is somewhat modified. The dispersion energy between two atoms with instantaneous induced dipoles, m_i and m_j , is found by solving Schrödinger's equation,

$$\frac{1}{M} \frac{\delta^2 \Psi}{\delta z_i^2} + \frac{1}{M} \frac{\delta^2 \Psi}{\delta z_j^2} + \frac{2}{\hbar} \left(E - \frac{1}{2} k z_i^2 - \frac{1}{2} k z_j^2 - U_{dipole-dipole} \right) \Psi = 0 \quad (9)$$

where, for the case of correlated, parallel dipoles,

$$\begin{aligned}
U_{dipole-dipole} &= \mu_i \left(\nabla^2 \left(\frac{1}{r_{ij}} f_{ij}^{overlap}(r_{ij}) \right) \right) \mu_j = \frac{\mu_i \mu_j}{r^3} \lambda_3^{overlap} - \frac{3(\mu_i r)(\mu_j r)}{r^5} \lambda_5^{overlap} \\
&= \frac{\mu_i \mu_j}{r^3} (3\lambda_5^{overlap} - \lambda_3^{overlap}) = \frac{\mu_i \mu_j}{r^3} f_{damp}^{dispersion} .
\end{aligned} \tag{10}$$

The damping factors, l_3 and l_5 , that define f_{damp} for dispersion are derived from the action of the gradient operator and are identical to those for the dipole-dipole interaction energy as defined in Appendix A. Solving the Schrödinger equation from equation 9 yields,

$$E = \frac{1}{2} \hbar(\omega_1 + \omega_2) \tag{11}$$

$$\omega_1 = \omega_0 \sqrt{1 - \frac{2Q^2}{r^3 k} f_{damp}^{dispersion}} , \quad \omega_2 = \omega_0 \sqrt{1 + \frac{2Q^2}{r^3 k} f_{damp}^{dispersion}} . \tag{12}$$

This energy expression can be effectively approximated with a binomial expansion,

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \dots \tag{13}$$

and the total energy thus becomes,

$$E = \hbar\omega_0 - \frac{Q^4 \hbar\omega_0}{2r^6 k^2} + \dots . \tag{14}$$

Subtracting the energy of two infinitely separated dipoles ($\hbar\omega_0$) and substituting the parameter C_6

for $\frac{Q^2 \sqrt{\hbar\omega_0}}{\sqrt{2}k}$ gives the pairwise dispersion energy,

$$U_{dispersion}^{HIPPO} = - \sum_{i < j} \frac{C_6^i C_6^j}{r^6} (f_{damp}^{dispersion})_{ij}^2 . \tag{15}$$

It is well known that accurate modeling of the dispersion energy at short range requires the use of a damping function.³²⁻⁴¹ HIPPO provides a non-empirical damping function derived from the

dipole density-dipole density interaction. The model requires only one C_6 parameter per atom since the parameters for the damping function are fixed at their electrostatic model values.

Pauli Repulsion. The final element of the HIPPO model is a density-based, multipolar model for Pauli Repulsion. Pauli repulsion is a consequence of the rearrangement of electron density that occurs when the Pauli exclusion principle is applied to electron densities of two unperturbed interacting molecules.⁴²⁻⁴⁹ In previous work, we show that the primary change in electron density, relative to the unperturbed reference state, is an evacuation of electron density from the internuclear region.²² The energy associated with this accumulation of charge in the internuclear region is proportional to

$$U_{Pauli\ repulsion} \propto \frac{S^2}{R} \quad (16a)$$

$$S = \int \phi_i \phi_j dv \quad (16b)$$

where S is the overlap integral between the atomic orbitals on i and j , and R is the internuclear distance. To obtain suitable quantities to implement this model, we use the ansatz

$$\rho = \phi^* \phi \quad (17)$$

to define real, atomic pseudo-orbitals as:

$$\phi = \sqrt{\rho} = \sqrt{\frac{Q\zeta^3}{8\pi}} e^{-\frac{\zeta r}{2}}. \quad (18)$$

These pseudo-orbitals define the charge-charge portion of the overlap integral,

$$S = \int \phi_i \phi_j dv = \sqrt{Q_i Q_j \zeta_i^3 \zeta_j^3} \left[\frac{1}{2X^3 R} \left(\zeta_i (RX - 2\zeta_j) e^{-\frac{\zeta_j R}{2}} + \zeta_j (RX + 2\zeta_i) e^{-\frac{\zeta_i R}{2}} \right) \right] \quad (19)$$

with

$$X = \left(\frac{\zeta_i}{2}\right)^2 - \left(\frac{\zeta_j}{2}\right)^2 .$$

From the bracket term, we can define

$$f_{exp}^{repulsion}(R) = \begin{cases} \frac{1}{\zeta^3} \left(1 + \frac{\zeta R}{2} + \frac{1}{3} \left(\frac{\zeta R}{2}\right)^2\right) e^{-\frac{\zeta R}{2}}, & \zeta_i = \zeta_j \\ \frac{1}{2X^3 R} \left[\zeta_i (RX - 2\zeta_j) e^{-\frac{\zeta_j R}{2}} + \zeta_j (RX + 2\zeta_i) e^{-\frac{\zeta_i R}{2}} \right], & \zeta_i \neq \zeta_j \end{cases} . \quad (20)$$

This allows writing S^2 in the familiar Coulombic form,

$$\frac{S_{charge-charge}^2}{R} = Q_i T_{pauli} Q_j \quad (21)$$

with

$$T_{pauli} = \frac{\zeta_i^3 \zeta_j^3}{R} (f_{exp}^{repulsion})^2 \quad (22)$$

where T_{pauli} (and, in turn, S^2) is dominated at short range by the exponential $f^{repulsion}$ term.

The anisotropy of the HIPPO repulsion model is obtained through its use of atomic multipole moments. Because S^2 has a clearly Coulombic form, we can include higher-order terms in the same manner as for electrostatics,

$$\frac{S_{total}^2}{R} = \sum_{i>j} \vec{M}_i T_{ij}^{repulsion} \vec{M}_j \quad (23a)$$

where

$$\vec{M} = \left(Q, [\mu_x, \mu_y, \mu_z], \begin{bmatrix} \Theta_{xx} & \Theta_{xy} & \Theta_{xz} \\ \Theta_{yx} & \Theta_{yy} & \Theta_{yz} \\ \Theta_{zx} & \Theta_{zy} & \Theta_{zz} \end{bmatrix} \right) \quad (23b)$$

$$\mathbf{T}_{ij}^{repulsion} = \begin{bmatrix} 1 & \nabla & \nabla^2 \\ \nabla & \nabla^2 & \nabla^3 \\ \nabla^2 & \nabla^3 & \nabla^4 \end{bmatrix} (T_{pauli}). \quad (23c)$$

The multipole moments used are identical to those from the electrostatics calculation and $T^{repulsion}$ is a natural generalization of T_{pauli} . The interpretation here is that just as the charge component of the multipole expansion has a density, so too do the dipole and quadrupole moments. The various multipolar terms described in equation 23 represent the overlaps between the pseudo-orbitals associated with each individual density component. This definition of S^2 allows us to establish an anisotropic repulsion model we call the Multipolar Pauli Repulsion model,

$$U_{Pauli\ repulsion}^{HIPPO} = \sum_{i < j} \frac{K_i K_j}{r_{ij}} S_{total}^2. \quad (24)$$

A complete derivation of this model is detailed in our previous work.²² Full equations defining the model as presented here, with higher-order terms included, are presented in Appendix C. The HIPPO repulsion model introduces three parameters per atom: a proportionality constant K, an exponential parameter a, and a valence charge Q. Note that although analogous to their counterparts in the electrostatics derivation, the parameters ζ and Q are allowed to differ from their adopted values in the electrostatic energy term.

Valence Terms. The HIPPO water model is fully flexible. It includes a bond stretching term and angle bending term, whose functional forms are the same modified harmonic potentials used in AMOEBA⁹ and originally taken from work by Allinger on the MM3 force field.⁵⁰ Stretch-bend and Urey-Bradley coupling terms are not used. HIPPO does include a charge flux term which

couples the atomic partial charges with the H-O stretching motions and the H-O-H angle,⁵¹ and serves to provide a dipole moment derivative surface in better agreement with quantum mechanical calculations.⁵² Previous work with the AMOEBA+ force field has shown that this charge flux term correctly reproduces the average increase in the H-O-H angle, from 104.5° to roughly 106°, that occurs when transferring water from gas to liquid phase.⁵³ The inclusion of this term, with parameters optimized for the HIPPO water model, yields the same correct behavior for the average angle value.

2.3. Methods

Code Implementation. HIPPO calculations in this paper were performed with the Tinker Version 8, Tinker-OpenMM, and Tinker9 packages.⁵⁴⁻⁵⁶ Implementation of HIPPO was undertaken by Josh Rackers and Jay Ponder in Tinker, Joshua Rackers, Zhi Wang and Roseane Silva in Tinker-OpenMM, and Zhi Wang and Roseane Silva in Tinker9. Molecular dynamics simulations data in the paper were performed with Tinker9 on our in-house GPU cluster. All subsequent analysis was performed using Tinker on workstation CPU hardware and Tinker9 on the GPU cluster.

The Tinker9 code is optimized for standard simple partial charge force fields and for the AMOEBA potential, while the HIPPO code is unoptimized. Molecular dynamics benchmarks for three 24051 atoms, 62.23 Å cubic water boxes using current Tinker9 code and an NVIDIA 3070 Ti GPU are as follows: TIP3P, 325.5 ns/day (2.0 fs steps, rigid water via SETTLE); AMOEBA, 29.1 ns/day, and HIPPO 24.6 ns/day (both run with 2.0 fs steps, RESPA multiple time step integrator, SCF induced dipole convergence to 0.00001 Debye RMS). A looser induce dipole convergence of 0.01 D is sufficient for many production calculations, and its use increases the speed of AMOEBA and HIPPO to 43.4 ns/day and 33.6 ns/day, respectively. Based on

comparative timings with CPU code, we estimate that fully optimized Tinker9 HIPPO code will be at least as fast as AMOEBA, and likely about 25% faster.

In order to facilitate model development, our current HIPPO implementation is written for ease of modification instead of for computational speed. First, multipole, polarization and repulsion terms are computed in independent, modular code sections, requiring redundant evaluation of the geometric and interaction terms for dipoles and quadrupoles. Second, the multipole and polarization are directly computed in the global Cartesian coordinate frame, without use of spherical harmonics or prior rotation of pair interactions into quasi-internal frames.⁵⁷ Speed advantages for HIPPO compared to AMOEBA include the use of particle mesh Ewald summation (PME) for dispersion interactions,²¹ and HIPPO's simpler gradient computation due to its use of unified exclusion and scaling rules for induced dipole and energy calculations.

Parameterization Procedure.

Stage One: Fit to SAPT data

The initial multipole and valence parameters were fit to monomer data. The multipole parameters were obtained using a protocol analogous to that for AMOEBA parameterization¹⁰ and initial bond and angle parameters were taken from AMOEBA. The rest of the initial parameters pertaining to the intermolecular potential were fit exclusively to SAPT data. SAPT2+ reference calculations with an aug-cc-pVDZ basis set were performed on 27 water dimer structures. These structures included seven points on the dissociation curve, ten points on the canonical dimer angular surface, and the ten stationary point dimer structures of Smith, *et al.*⁵⁸ Each term of the force field was fit to its corresponding component from the SAPT decomposition. Observations on the quality of the resulting parameters can be found in the discussion section.

Stage Two: Constrained Genetic Algorithm Search

The initial parameter set obtained through fitting to SAPT energy components needed further adjustment to better match condensed phase properties. To improve the liquid water properties while keeping the features of the SAPT fitting, we continued optimizing the model by performing a global search in parameter space centered at the initial values. A differential evolution optimizer from the Scipy 1.8 package was used. The objective function of this optimizer has two main components: the energy decomposition of the Smith dimers and the heat of vaporization of water at room temperature.

While this optimizer was generating liquid data, a second function was simultaneously evaluating the liquid properties from each simulation. This function was merely a tool to select simulations with desired properties. The goal was to find simulations with liquid density, heat of vaporization and self-diffusion coefficient within 1% of their experimental values. The search was ended upon generation of five parameter sets satisfying all requirements. One of these sets was chosen as the best to continue the parametrization.

Stage Three: Parameter Refinement with ForceBalance

Following the global search in parameter space, we used a least square optimizer to fit a wider range of properties and to guarantee we were at a local minimum in parameter space. For this step, we used the ForceBalance (FB) program.⁵⁹ The goal of this final parameterization step was to obtain a model with desired condensed phase properties across a wide range of temperature and pressure.

The distinctive feature of FB is its ability to compute parametric derivatives of condensed phase properties from MD simulations using thermodynamic fluctuation equations. To refine

parameters, we set a minimal objective function including experimental densities, enthalpies of vaporization, and dielectric constants over a range of temperatures from 261 K to 373 K. No other condensed phase properties were considered in the fitting procedure.

Computational Details. All properties and simulations were obtained using the HIPPO force field as implemented in the Tinker and Tinker9 packages. To compute condensed phase properties, MD simulations of liquid water were performed. Unless otherwise noticed, properties were computed based on simulation of a cubic box of dimension ~ 50 Å and containing 4,200 water molecules. The thermodynamics properties listed in Table 2.4 were calculated from simulations at constant pressure and temperature. All simulations were performed using the RESPA (Reversible Reference System Propagator Algorithm) integrator coupled with a Monte Carlo barostat⁶⁰ and the Bussi thermostat.⁶¹ For FB fitting, each MD simulation ran for 2 ns using a 2.0 fs time step, with a 0.5 ns equilibration phase and 1.5 ns production phase. The energy components of water dimers and clusters were calculated using the ANALYZE program in Tinker.

The temperature dependence of water properties was computed from a total of 40 simulations carried out at atmospheric pressure (1 atm), for temperatures ranging from 248 K to 373 K. Each simulation was started at the experimental density for the respective temperature and ran for at least 20 ns using a 2.0 fs time step; the first 2 ns of the simulations were discarded as equilibration. For temperatures less than 300 K, the production MD was extended by 10 ns to guarantee convergence of properties.

The self-diffusion coefficient was computed following the steps described for the MB-Pol model.⁶² We chose to run simulations in ~ 100 Å cubic boxes with 33,500 water molecules. The larger box size was used to reduce known finite size effects in the calculation of self-diffusion coefficients.⁶³ We simulated 26 temperatures in total, ranging from 248 K to 373 K. For each

simulation temperature, a box was built such that its density matched the experimental density for that temperature. Then, each simulation box was equilibrated for 0.5 ns in an NPT ensemble at atmospheric pressure. Following equilibration, we ran an additional 1.5 ns trajectory. From this trajectory, thirty different structures were selected, at 50 ns intervals. From those structures, thirty independent NVT trajectories of 100 ps were obtained. Then, we ran 100 ps simulations in an NVE ensemble. The self-diffusion coefficient was computed from each NVE trajectory and averaged over the 30 independent calculations for each temperature.

In order to evaluate finite size effects in computation of the self-diffusion coefficient, we ran additional simulations with different box sizes at room temperature (298 K). Each simulation was run for 4 ns in NVT ensemble, and the self-diffusion coefficient was computed using the final 3.5 ns of data. Five cubic box simulations were performed: 300 water molecules in ~ 20 Å box, 900 molecules in ~ 30 Å box, 4,200 molecules in a ~ 50 Å box, 17,100 molecules in a ~ 80 Å box, and finally 33,500 molecules in a ~ 100 Å box.

To calculate the surface tension of liquid water, we first selected four starting structures, at least 100 ps apart, from the production phase of our NPT simulations at different temperatures. Each structure was then simulated for 500 ps in an NVT ensemble. Then the Z-axis of each cubic box was expanded to three times the X-axis and Y-axis dimensions.⁶⁴ The final system geometries were slabs with $X = Y = \sim 50$ Å, and $Z = \sim 150$ Å, with a vacuum layer along the Z-axis over each side of the slab. Each system was then simulated in the NVT ensemble for 10 ns. The surface tension was calculated from the last 9.5 ns of data using the pressure tensor,⁶⁵ which was computed every picosecond. The final surface tension value reported for each temperature is the average of the four independent calculations.

The pressure dependence of the liquid water density was computed from a total of 10 simulations at room temperature (298 K), and with target pressure ranging from 1 atm to 9000 atm. Each simulation was started at the experimental density for the respective pressure and run for 10 ns using a 2.0 fs time step, with the first 2 ns as equilibration. The cubic box size for this set of simulations was ~ 30 Å with 900 water molecules.

We selected eight ice crystal structures to compute lattice energy and density. Ice energies were computed after energy minimization of the initial structure using a steepest descent algorithm. Each minimized structure was then simulated for 10 ns in the NPT ensemble. The average density of each ice crystal was computed using the last 8 ns of trajectory data. The target temperature and pressure for each simulation were set to the respective values reported for each polymorph crystal structure.

2.4. Results

Because both the functional form and parameterization of the HIPPO water model are rooted in quantum mechanics, we set out to test the accuracy of the model against both experimental condensed phase data and *ab initio* calculations. In this section we will move from small to large clusters, starting from the properties of the water monomer and dimers, through successively larger clusters and up to condensed phase. By showing this behavior across scales we hope to demonstrate the power of a first-principles derived potential energy function.

Parameters. Full specification of the HIPPO force field water model includes 37 refined parameter values. Explicit values for these parameters with their associated units are provided in the Supporting Information as Table S1. Several of the parameters are highly correlated, such that the effective number of parameter degrees of freedom required for the HIPPO model is lower than

the number of raw parameters. While many of the parameter classes also used in previous AMOEBA-like water models, such as the atomic multipole values on oxygen and hydrogen, adopt similar values in HIPPO, the differences observed are important to the accurate reproduction of many water properties. Finally, where earlier work on individual components of the HIPPO model considered additional molecules,²⁰⁻²² the HIPPO water values reported here are in line with periodic trends across these other molecules and atom types.

Monomer. The foundation of the HIPPO model is the monomer electron density. The fidelity of the rest of the model relies on an accurate representation of the true electron density of the molecule. Table 2.1 shows that HIPPO reproduces the monomer multipole moments and polarizability of an isolated water molecule with a satisfactory level of agreement.

Table 2.1. HIPPO Water Monomer Properties. All calculations performed on experimental, gas phase geometry where the Z-axis is the C_2 axis, the molecule lies in the XZ-plane, and the O atom is along the negative Z-axis.

	Dipole (D)	Quadrupole (B)			Polarizability (\AA^{-3})		
	d_z	Q_{xx}	Q_{yy}	Q_{zz}	a_{xx}	a_{yy}	a_{zz}
HIPPO	1.843	2.48	-2.38	-0.10	1.613	1.289	1.362
Experiment	1.855	2.63	-2.50	-0.13	1.528	1.415	1.468

Additionally, HIPPO accurately reproduces the electrostatic potential around the water monomer as illustrated in Figure 2.2. The “Multipole Only” panel shows the signature of the “charge penetration” effect with a large negative error near the molecular surface. The point multipole model systematically underestimates the electrostatic potential at short range. Previous work has shown that including a simple density model can largely eliminate this charge penetration error, and this is clearly true for the HIPPO model. The “HIPPO” panel in Figure 2.2 shows that

error in the electrostatic potential at short range is greatly reduced relative to the undamped point multipole model.

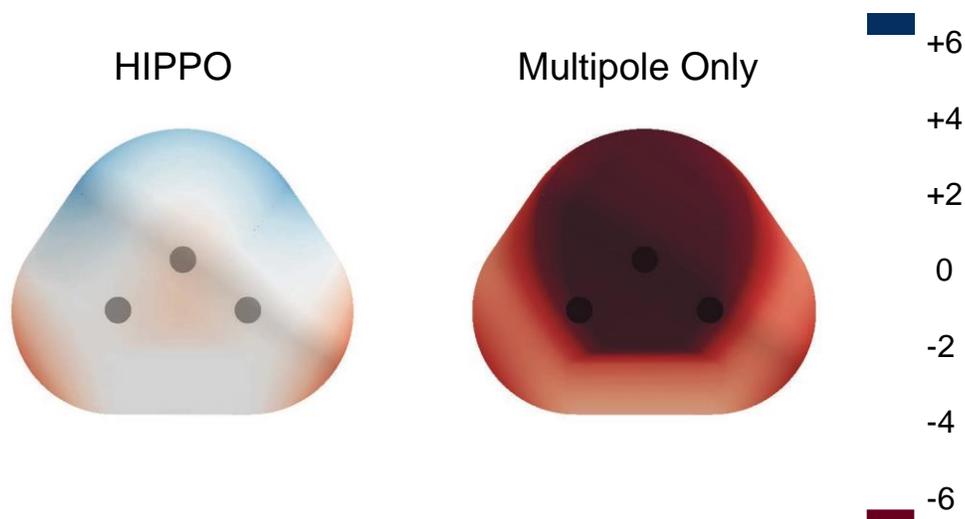


Figure 2.2. Error in the electrostatic potential: HIPPO vs. point multipoles neglecting charge penetration. The plot on the right shows the error in the electrostatic potential at the van der Waals surface for the undamped point multipole model. The plot on the right shows the same for HIPPO. Both use the same set of multipoles through quadrupole. Values are given in kcal/mol/electron.

Dimers. The water dimer potential energy surface is foundational to the overall model because it is the first place where the entire intermolecular energy function comes into play. For HIPPO in particular, this surface is of tremendous importance as the density-based terms of the intermolecular potential energy function are constructed specifically to reproduce dimer intermolecular interactions. Because it has been extensively studied, we have selected three separate “slices” of the dimer potential energy surface on which to evaluate the HIPPO model: the canonical water dimer dissociation curve, the angular dependence of the water dimer hydrogen bond angle, and the ten well-studied stationary points of Smith *et al.*⁵⁸ For each of these slices we evaluate the HIPPO model relative to two references. First, we compare the total energies of HIPPO to the total energies from *ab initio* calculations. Second, we compare the components of the HIPPO intermolecular potential energy function to their corresponding components from a SAPT decomposition.

The dissociation curve of the canonical water dimer is an important piece of the dimer potential energy surface because it contains information about the balance between short-range effects like repulsion and charge penetration, and long-range effects such as dispersion and multipole electrostatics. To generate this curve, we took the water-water interaction structures from the S101x7 database.²⁹ These structures represent the water dimer at points from 0.7 to 1.1 times the equilibrium distance. The results for HIPPO *vs.* the *ab initio* reference data are plotted in Figure 2.3A.

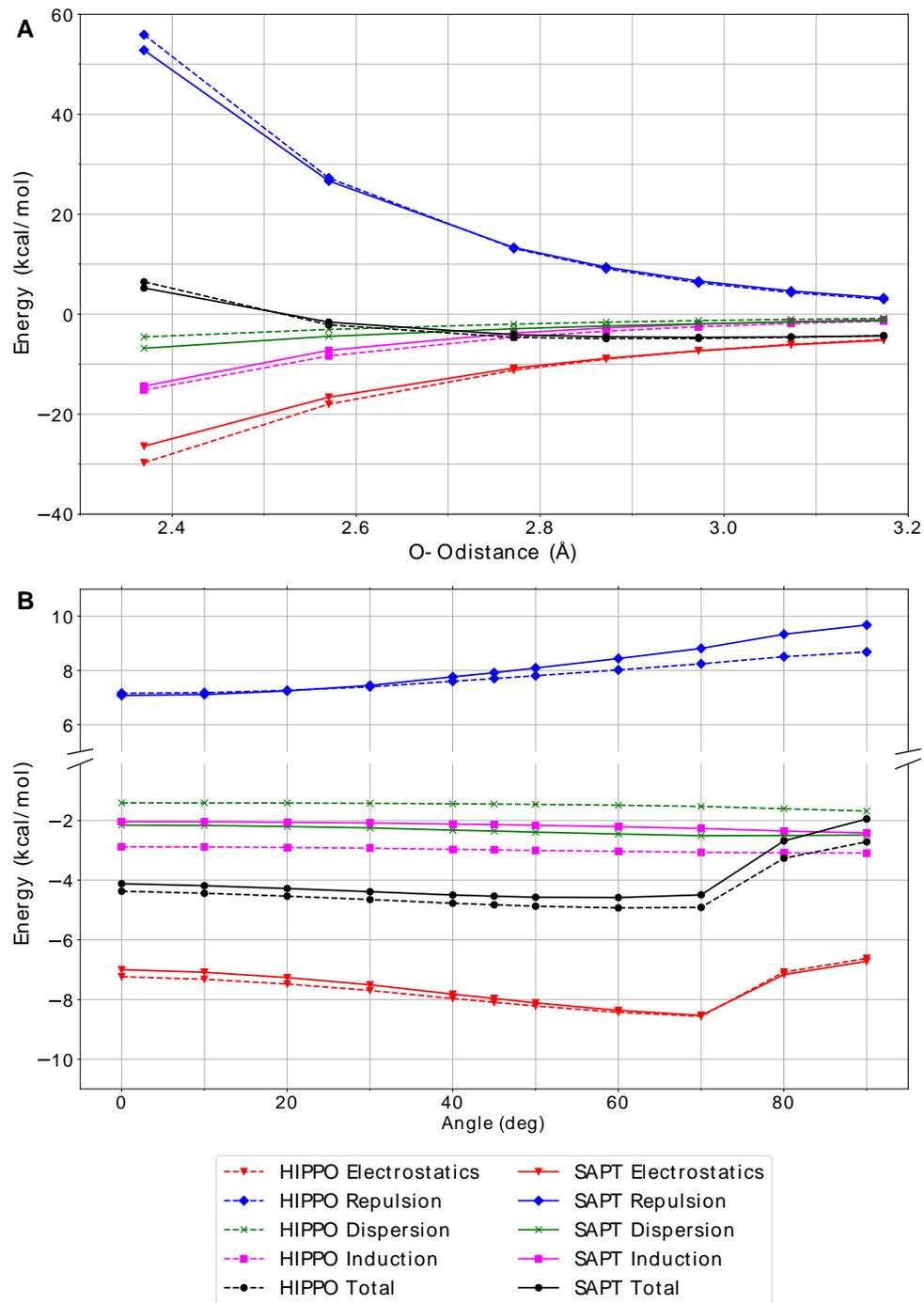


Figure 2.3. Energy components for water dimer dissociation curve (A) and “flap angle” degree of freedom (B). HIPPO components are shown in dashed lines and SAPT reference energies are shown in solid. The extent of dissociation is represented by the O-O distance. The “flap angle” is defined as the angle between the O-O vector and the plane of the acceptor water molecule.

The HIPPO total energy matches the SAPT total energy closely throughout the distance range. Even for the closest points, O-O distances that are rarely sampled in ambient water, the agreement is good. This agreement across the range can be attributed to the fidelity with which HIPPO matches the components of the SAPT energy. In particular, the repulsion and electrostatic curves, which point force fields fail to reproduce at short and long range simultaneously, are in excellent agreement throughout the curve. Importantly, HIPPO is able to capture the short-range physics without compromising the long-range behavior.

Another critical aspect of the water dimer potential energy surface is the hydrogen bond angle. To generate structures for this part of the surface we varied the so-called “flap angle” of the canonical water dimer as illustrated in the inset of Figure 3B. The behavior with respect to this angle is important because it contains information about the anisotropy of the water molecule. Work on the AMOEBA force field has shown that anisotropy in the electrostatics vis-a-vis point multipoles helps reproduce the directionality of hydrogen bonding in water as well as other systems. Here we examine the anisotropy of not just the electrostatics, but the other energy components as well. Plotted in Figure 2.3B is the change in total energy, as well as the change in each of the components, as the flap angle of the water dimer is changed from 0° to 90° . The SAPT curves illustrate an interesting phenomenon. While the dispersion and induction components of the intermolecular energy are largely unchanged across the scan, the electrostatics and repulsion components vary dramatically and in opposite directions. In fact, the trends in these two components counterbalance each other. The minimum of the electrostatic curve lies near 70° . However, the optimal hydrogen angle for the water dimer is known to be slightly smaller, around 60° .

Figure 2.3B shows that this is nearly entirely due to the countervailing angular dependence of the repulsion curve. It also shows that HIPPO matches the angular dependence of both the electrostatics and repulsion curves well. The anisotropy in the repulsion curve is noteworthy since this is the first force field to include multipolar anisotropic repulsion. This gives a flap angle for the minimized HIPPO water dimer of 63° , near the experimental value of 57° . This underscores the importance of including anisotropy, not just in the electrostatics portion of the force field, but the repulsion as well. Without the angular dependence of the multipolar repulsion model, as is the case in the vast majority of isotropic Lennard-Jones van der Waals functions, the flap angle of the water dimer would be incorrect. Curiously, the original multipole-based AMOEBA model corrected this issue empirically by scaling down the quadrupole moments of each atom by a factor of 0.73, but misdiagnosed the problem. The key to capturing the anisotropy of the potential energy surface of the water dimer seems to be in including anisotropy in the repulsion as well.

The final piece of the dimer potential energy surface we examined is the ten Smith water dimers. These dimers are all stationary points on the water dimer potential energy surface and as such, they form a representative sample of the various dominant dimer configurations in the condensed phase. There are a variety of both hydrogen bonded and non-hydrogen bonded structures in the set, making it a good test of the accuracy of the model with relevant contact geometries beyond the canonical configuration.⁶⁶ Fully optimized *ab initio* structures at the MP2/aug-cc-pV5Z level were computed as part of the present study, and are depicted in Figure 2.4. From the geometry of each dimer at the MP2/aug-cc-pV5Z level, we then determined a “gold standard” counterpoise-corrected CCSD(T) total stabilization energy for each dimer compared to the energy of two optimized, separated monomers at the same level of theory.⁶⁷ Note that these energies contain the deformation energy of the water monomers upon dimer formation. The

coordinates of the optimized Smith dimers are provided in Supporting Information. Only dimer 1 is a true minimum on the potential energy surface, while the other dimers have one to three negative Hessian eigenvalues.

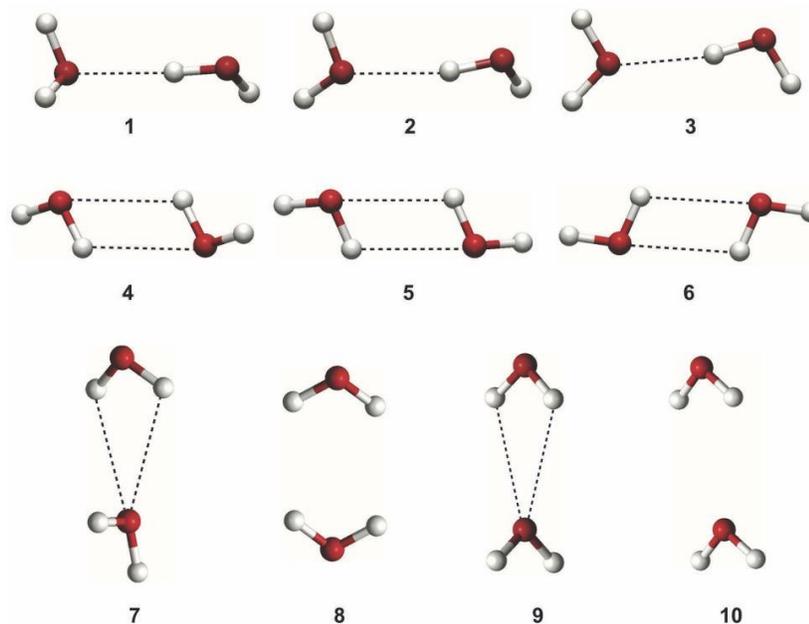


Figure 2.4. Structures of the ten Smith water dimers obtained from full geometry optimization at the MP2/aug-cc-pV5Z level. The dashed lines represent hydrogen-oxygen interactions that are roughly within the distance corresponding to the hydrogen bonding. Dimers 1-3 each contain a single hydrogen bond, and are variations of the global minimum structure 1. Dimers 4-6 contain two hydrogen bonds between a pair of antiparallel O-H bonds. Dimer 7, 9 and 10 have two weaker hydrogen bonds of approximately equal distance provided by a single donor water. Dimer 8 has stacked, displaced molecules with H-H interactions as the closest contacts. Atomic coordinates are provided in Supporting Information.

In Table 2.2, the structures and energetics of the Smith dimers optimized with the HIPPO model are compared against *ab initio* reference data.^{66, 68} In addition to previously reported reference energy values, the MP2/aug-cc-pV5Z structures computed here were used to generate CCSD(T)/CBS energies for all ten dimers. The root mean square energy difference between the CCSD(T)/CBS and HIPPO values is 0.129 kcal/mol, and the average structural RMS with all atoms weighted equally is 0.075 Å. Overall, the structural and energetic agreement is excellent. Dimers 4 and 5 exhibit the largest deviation between QM and HIPPO results. In both cases, the

HIPPO optima have lower energies and smaller intermolecular contact distances, perhaps due to a small error in the interaction between antiparallel O–H bonds. The energies of dimers 7 to 10 differ the most between the earlier rigid monomer interaction energies of Tschumper, *et al.*,⁶⁶ and the fully flexible values reported by Wang and Bowman⁶⁸ or the flexible CCSD(T) values reported here. Unsurprisingly, three of those dimers exhibit the largest deformation energies upon dimer formation. Comparison of HIPPO energies with a limited set of other empirical water models is detailed in Table S2 of the Supporting Information.

Table 2.2. Water dimer binding energies for HIPPO compared to *ab Initio* reference calculations. Dimer geometries were taken from the Supporting Information of reference⁶⁹; Ref 1 energies are from reference⁶⁶, and Ref 2 values are from reference⁶⁸. Dimer stabilization energies⁶⁷ and total deformation energies at the CCSD(T)/pV5Z level are shown, as are complete basis set (CBS) extrapolated values.^{70, 71} HIPPO dimer energies are provided for single point calculations at the CCSD(T)/pV5Z geometry, and for fully optimized HIPPO structures. Also shown are the QM and HIPPO R_{O-O} dimer distances, the HIPPO structure RMS vs. MP2/aug-cc-pV5Z optima, and the number of negative frequencies (n) for CCSD(T)/pV5Z and HIPPO optima. All energies are in kcal/mol, and the R_{O-O} distance and HIPPO RMS values are in Angstroms.

Dimer	Ref 1	Ref 2	CCSD(T) /pV5Z	Deform	CCSD(T) /CBS	CCSD(T) Ro-o	HIPPO (sngl)	HIPPO (opt)	#neg n	HIPPO RMS	HIPPO Ro-o
1	-4.968	-4.98	-4.956	0.041	-4.967	2.895	-4.917	-4.957	0	0.054	2.884
2	-4.453	-4.45	-4.447	0.038	-4.459	2.905	-4.330	-4.339	1	0.104	2.913
3	-4.418	-4.38	-4.398	0.037	-4.410	2.911	-4.232	-4.238	2	0.017	2.916
4	-4.250	-4.23	-4.262	0.029	-4.281	2.800	-4.378	-4.574	1	0.103	2.756
5	-3.998	-3.97	-4.014	0.032	-4.034	2.771	-3.994	-4.193	1	0.161	2.754
6	-3.957	-3.91	-3.969	0.036	-3.991	2.748	-3.823	-3.913	3	0.044	2.729
7	-3.256	-3.15	-3.157	0.092	-3.168	2.952	-3.090	-3.121	2	0.028	2.917
8	-1.300	-1.46	-1.417	0.035	-1.425	3.325	-1.354	-1.377	3	0.046	3.271
9	-3.047	-3.18	-3.197	0.114	-3.208	3.018	-3.169	-3.184	1	0.031	2.971
10	-2.182	-2.28	-2.275	0.096	-2.286	3.168	-2.278	-2.295	2	0.025	3.118

Further structural and energetic results, comparing HIPPO with *ab initio* results on the ten dimers, are plotted in Figure 2.5. The figure shows two levels of comparison. First, it compares the total interaction energy for each dimer. Along with the HIPPO values, two *ab initio* results are shown. The first is the SAPT total energy at the SAPT2+ level. The second *ab initio* values are the CCSD(T)/CBS results obtained in this work. It is interesting to note there is some disagreement between the SAPT and CCSD(T) results. For some dimers, the SAPT value differs by ~0.5 kcal/mol. This shows that although SAPT2+ is useful for determining individual components of the energy function, it is not a replacement for high-level coupled cluster total energy calculations. Optimized HIPPO dimer structures and energies are in good agreement with the CCSD(T) results for all ten dimers. This indicates an accurate balance between the hydrogen bonded and non-hydrogen bonded configurations. The origin of this balance is illustrated by the second level of comparison in Figure 2.5, the components of the interaction energy. The electrostatics, repulsion, dispersion and induction components of the HIPPO model match the SAPT decomposition in a consistent fashion across the dimer configurations. This demonstrates the agreement in total energies is not coming from cancellation of errors, indicating that a similar agreement should hold for other water dimer configurations outside this set of ten structures.

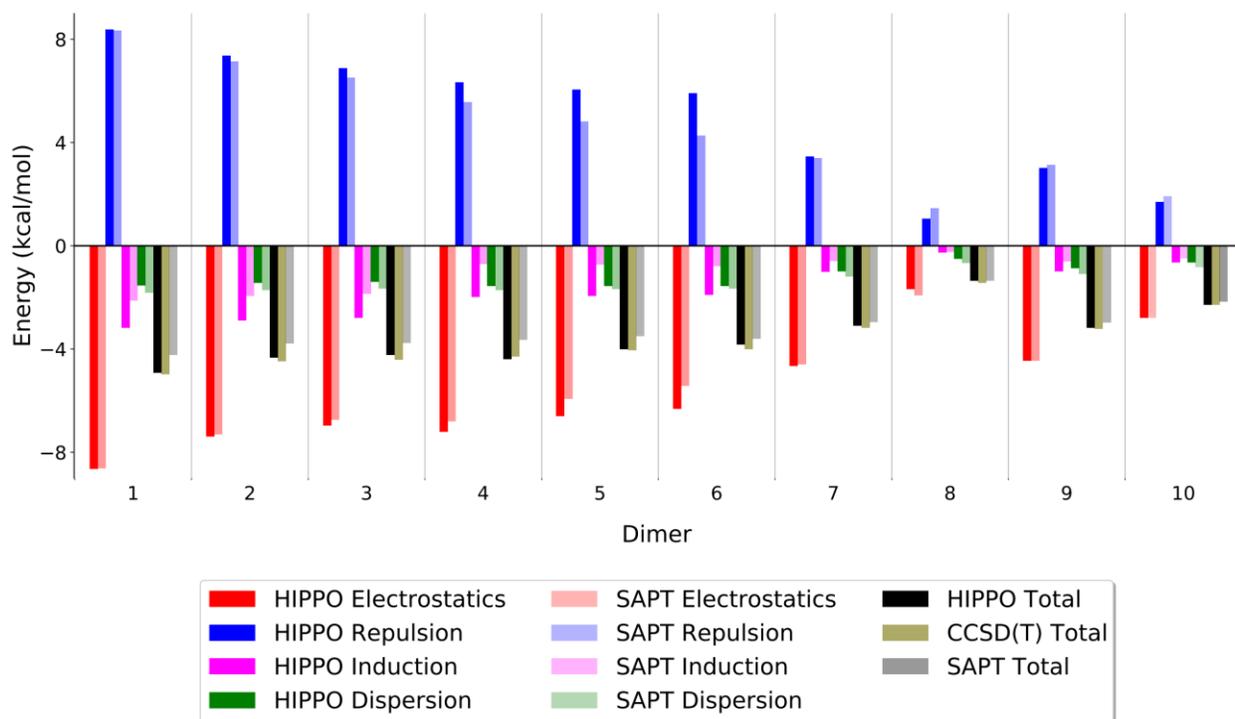


Figure 2.5. Total energies and components for ten water dimer stationary points. HIPPO and SAPT components are shown as colored bars. Although all errors are under 0.5 kcal/mol, these show some compensation on the part of HIPPO between the induction and dispersion components. The black, tan, and grey bars represent the HIPPO, CCSD(T), and SAPT values, respectively. Notably, HIPPO is in better agreement with the CCSD(T) data than SAPT, suggesting that the HIPPO component errors relative to SAPT are within the intrinsic error of the SAPT methodology.

Larger Clusters. We next tested the HIPPO model on larger clusters of water ranging from three to twenty molecules. The goal here is to span as much of the gap as possible between gas phase and condensed phase. For these larger clusters, SAPT data becomes difficult to interpret, but there are two types of data relevant to evaluating the HIPPO model. First, we compare total cluster binding energies. This provides a measure of how well the potential energy function performs as water becomes more liquid-like. Second, we compare the many-body energies. The average dipole moment of a water molecule increases steadily upon moving from monomer to dimer to clusters to condensed phase. This implies that in order for any model to achieve agreement with QM data for both clusters and condensed phase, it must include many-body effects. Thus, we

compare the many-body energies from *ab initio* calculations with the many-body energies from the classical HIPPO polarization function.

HIPPO compares very well with gold standard CCSD(T) benchmark total energy calculations moving from gas phase dimers toward bulk-like clusters. As shown in Table 2.2 and Table 2.3, the agreement for structures through the hexamer is within 0.57 kcal/mol on average. Moreover, the relative ranking of unique structures is also quite accurate. For example, HIPPO ranks the eight reference water hexamer structures in the same order as CCSD(T) calculations. Lastly, the HIPPO minima are structurally very similar to the reference QM-optimized structures, indicating the accuracy of the local potential energy landscape.

Unlike pairwise force fields, where the total energy of a system is simply the sum of the energies of every pair of interactions, HIPPO is polarizable. This means that it is designed to reproduce the non-additive portion of intermolecular interactions. To quantify the amount of non-additivity, we compute the three-body energy of a range of different water clusters. The three-body energy is defined as

$$E_{3B} = E_{total} - \sum_i \sum_j E_{ij} + \sum_i E_i \quad (25)$$

where second and third terms represent the sums of the two-body and one-body energies, respectively. The four-body and higher terms are negligible in the case of water.⁷² The first test set for three-body energies is the water trimer at a range of intermolecular contact distances. Starting from the structure depicted in the inset of Figure 2.6, the distances d_1 and d_2 were varied systematically. The three-body energy was computed at the MP2 level of theory and compared to the HIPPO three-body energy. Figure 2.6 shows that across the range of distances HIPPO agrees well with the *ab initio* result. Particularly at distances near equilibrium the agreement is very good.

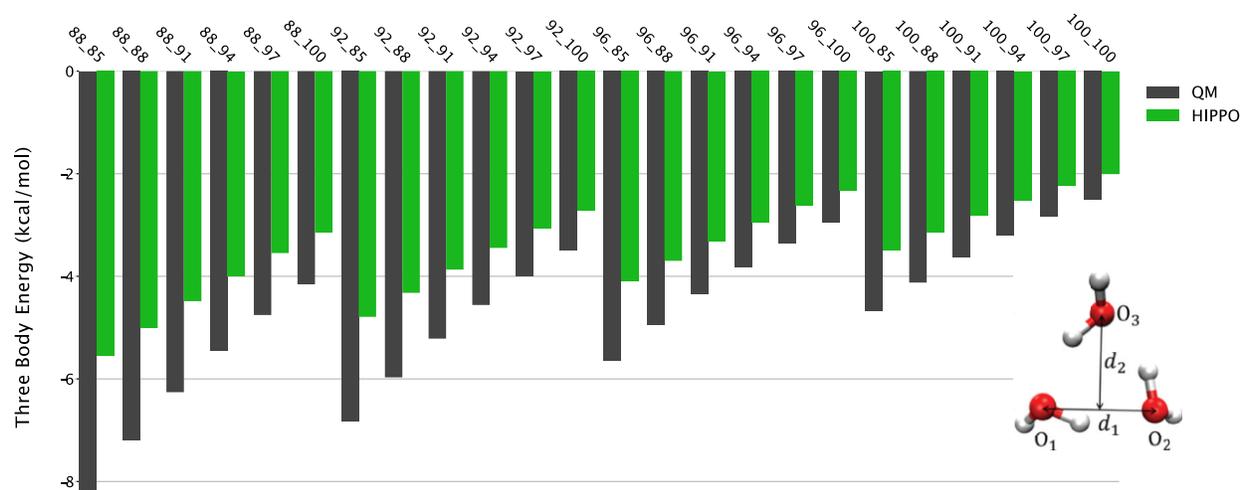


Figure 2.6. Three-body energies for water trimer as a function of intermolecular distance. HIPPO is within 1 kcal/mol for near-equilibrium structures. The X-axis values represent d_1 and d_2 , respectively as shown in the inset as percentages of the equilibrium distances. QM data is generated at the MP2/aug-cc-pVTZ level.

Table 2.3. Water cluster binding energies (kcal/mol) with HIPPO compared to *ab initio* calculations. ^a Structures and reference values from reference ⁷³. ^b Hexamer structures and reference energies from reference ⁷⁴.

Cluster	Structure	Reference	HIPPO	HIPPO Minimum	Difference
Trimer ^a		-15.77	-15.417	-15.767	0.00
Tetramer ^a		-27.39	-25.695	-26.685	0.71
Pentamer ^a		-35.90	-32.994	-34.582	1.32
Hexamer ^b	Prism	-45.92	-44.169	-46.145	-0.23
	Cage	-45.67	-43.635	-45.387	0.28
	Bag	-44.30	-41.106	-43.364	0.94
	Chair	-44.12	-40.484	-42.543	1.58
	Book A	-45.20	-42.359	-44.245	0.96
	Book B	-44.90	-42.103	-43.958	0.94
	Boat A	-43.13	-39.576	-41.548	1.58
	Boat B	-43.07	-39.612	-41.555	1.52
Octamer ^a	D _{2d}	-73.0	-68.309	-71.547	1.5
	S ₄	-72.9	-68.253	-71.559	1.3

11-mer ^a	43'4	-104.6	-94.775	-100.232	4.4
	515a	-1040	-93.635	-99.377	4.6
16-mer ^a	AABB	164.1	-155.457	-161.556	2.5
	ABAB	164.2	-155.875	-161.836	2.4
	Antiboat	164.6	-152.799	-159.634	5.0
	Boat A	164.4	-152.457	-159.357	5.0
	Boat B	164.2	-152.400	-159.425	4.8
17-mer ^a	552'5	-175.7	-161.740	-169.938	5.8
	Sphere	-175.0	-162.549	-170.681	4.3
				MAD:	
Summary:	Dimer – Hexamer			0.57	
	Octamer – 17-mer			3.8	

This level of agreement illustrates two important points. First, it shows that the HIPPO model is effective in capturing the many-body effect, and thus may perform well across the spectrum from gas to condensed phase. Second, it suggests that the majority of the *ab initio* many-body energy can be classified as polarization. It is well known that other categories of intermolecular interaction such as dispersion, charge transfer and repulsion have many-body components. The data in Figure 2.6 shows that for water, however, these appear to be small. The HIPPO three-body energy is systematically smaller than the *ab initio* result, but only by a small amount. It is only at the closest points, where water rarely accesses in the condensed phase, that it appears that higher-order many-body effects start to be significant.

To assess if the agreement with small-scale *ab initio* many-body results translates to liquid water, we also computed the three-body energy of progressively larger clusters. For water clusters of four to eight molecules, we computed the three-body energy at the MP2 level of theory and compared it against HIPPO results. Figure 2.7 shows the trends seen in the trimer test case hold for larger clusters.

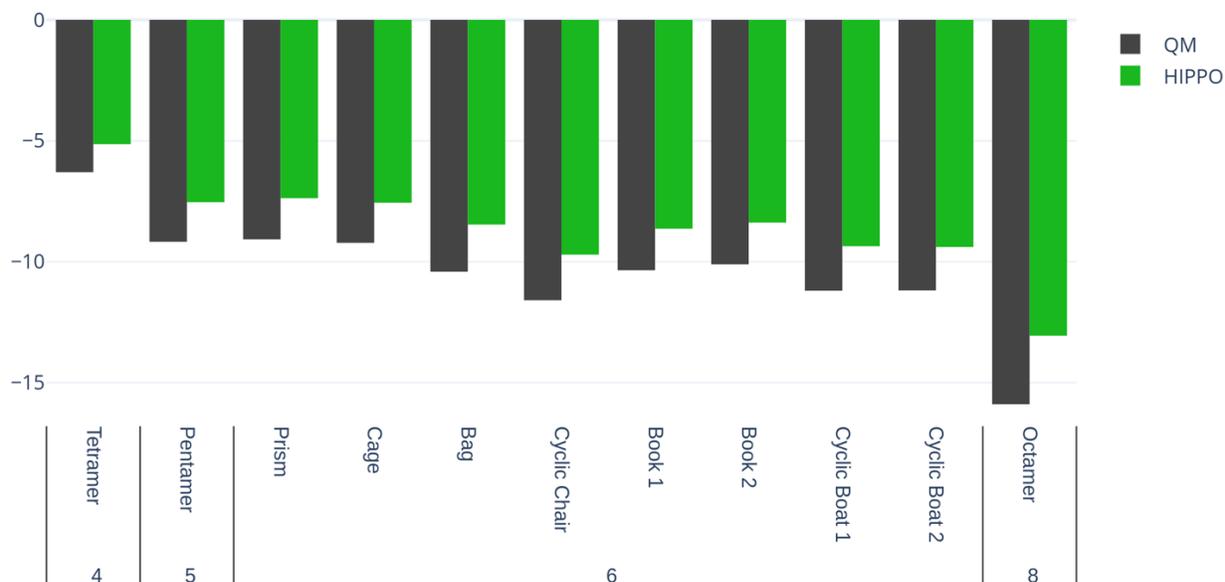


Figure 2.7. Three-body energies for water clusters tetramer through octamer. QM data is generated at the MP2 level of theory and aug-cc-pVTZ basis set.

Just as in the trimer case, the HIPPO result always slightly underestimates the magnitude of the total *ab initio* three-body energy. This validates the observation from the trimers that many-body effects in higher-order terms appear to stay small as cluster size grows. Additionally, the behavior of the three-body energy with geometry appears to be in good agreement with the reference data. HIPPO correctly predicts the ordering of the amount of three-body energy in the eight water hexamer structures. These structures are picked to be representative of fully hydrated water. It is difficult to estimate what the many-body energy of a full, condensed phase system of water is, but this result for the hexamers suggests that HIPPO may give an adequate representation.

Liquid Properties. In addition to accurately modeling *ab initio* data, it is important for a water model to accurately reproduce experimental liquid phase properties as well. We have tested the HIPPO water model on a wide variety of experimental observables at room temperature and ambient pressure and present the results in this section.

Table 2.4. Water properties at room temperature (298 K and 1 atm). ^a Reference ⁷⁵. ^b Reference ⁷⁶. ^c Reference ⁷⁷. ^d Reference ⁷⁸. ^e Reference ⁷⁹. ^f Reference ⁸⁰.

Property	HIPPO	Experimental	Abs. Deviation
Density (kg/m ³)	996.492	997.045 ^a	0.553 (0.06%)
Enthalpy of Vaporization (kJ/mol)	43.806	43.989 ^b	0.183 (0.42%)
Static Dielectric Constant	76.878	78.409 ^c	1.531 (1.95%)
Self-Diffusion Coefficient (10 ⁻⁵ cm ² /s)	2.557	2.299 ^d	0.258 (11.22%)
Surface Tension (mJ/m ²)	74.918	71.99 ^e	2.928 (4.07%)
Second Virial Coefficient (L/mol)	-1.2612	-1.158 ^f	0.103 (8.91%)

The primary thermodynamic and dynamic properties of the HIPPO model are collected in Table 2.4 along with the known experimental values. The density is in excellent agreement with experiment, with an error of less than 0.1%. The heat of vaporization, or the amount of energy required to transfer a water molecule from the liquid phase to the gas phase, is also in excellent agreement with an error relative to experiment of 0.4%. Both of these values were included in the objective function of the parameter refinement step, so good agreement is expected.

Likewise, the dynamic properties of HIPPO water are in close agreement with experiment. The self-diffusion coefficient of water measures how quickly or freely water molecules move in the liquid phase. The predicted diffusion coefficient of HIPPO differs from the experimental value by 11%. This is a reasonable agreement for a quantity that is known to be quite sensitive to details of molecular dynamics simulations. The HIPPO model is also in excellent agreement with the experimental dielectric constant of water. This is also a highly sensitive quantity for molecular dynamics simulations, and the HIPPO prediction is within 2% of the experimental result. The agreement with the experimental dielectric constant indicates that the HIPPO water electrostatic

environment is accurate. This is important not just for the properties of water, but for the future use of the HIPPO model solvating small molecules, ions and ultimately biological macromolecules. Lastly, the surface tension, a stress test for how well a water model handles the balance between bulk solution and interfaces, of the HIPPO model is in excellent agreement with experiment. The accuracy in the surface tension suggests that HIPPO will model solvation of both polar and hydrophobic species equally well.

The structural properties of liquid water are also of great interest for both the study of pure water and water as a solvent. As the canonical example of the hydrogen bond and because of its bent shape, liquid water represents a balance between many orientations of water-water interactions. To probe these structural properties, we compared the experimental radial distribution functions and second virial coefficient of water to those predicted by HIPPO.

Plotted in Figure 2.8 are the O-O, O-H, H-H radial distribution functions of water. Panel A in Figure 2.8 shows that the O-O radial distribution function of HIPPO water is in good agreement with experiment. The position of the first peak at 2.785 Å and height at 3.0 is within the experimental uncertainty of the experimental curve. The entire curve lies within the “family” of O-O $g(r)$ curves described by Brookes and Head-Gordon.⁸¹ The O-O $g(r)$ represents the coarse molecular level of structure in liquid water. The close agreement of HIPPO shows that the force field has the correct number of molecules in each solvation shell. At a finer level of detail, the O-H and H-H curves are also in close agreement with experiment. Panels B and C of Figure 2.8 show that the positions of the peaks in these curves agree with experiment. Moreover, the relative heights of the first and second peaks in the O-H and H-H curves correspond closely to the relative heights from the experimental model. This suggests that not only are the correct number of molecules in each solvation shell, but the average orientations of those molecules are in line with reality as well.

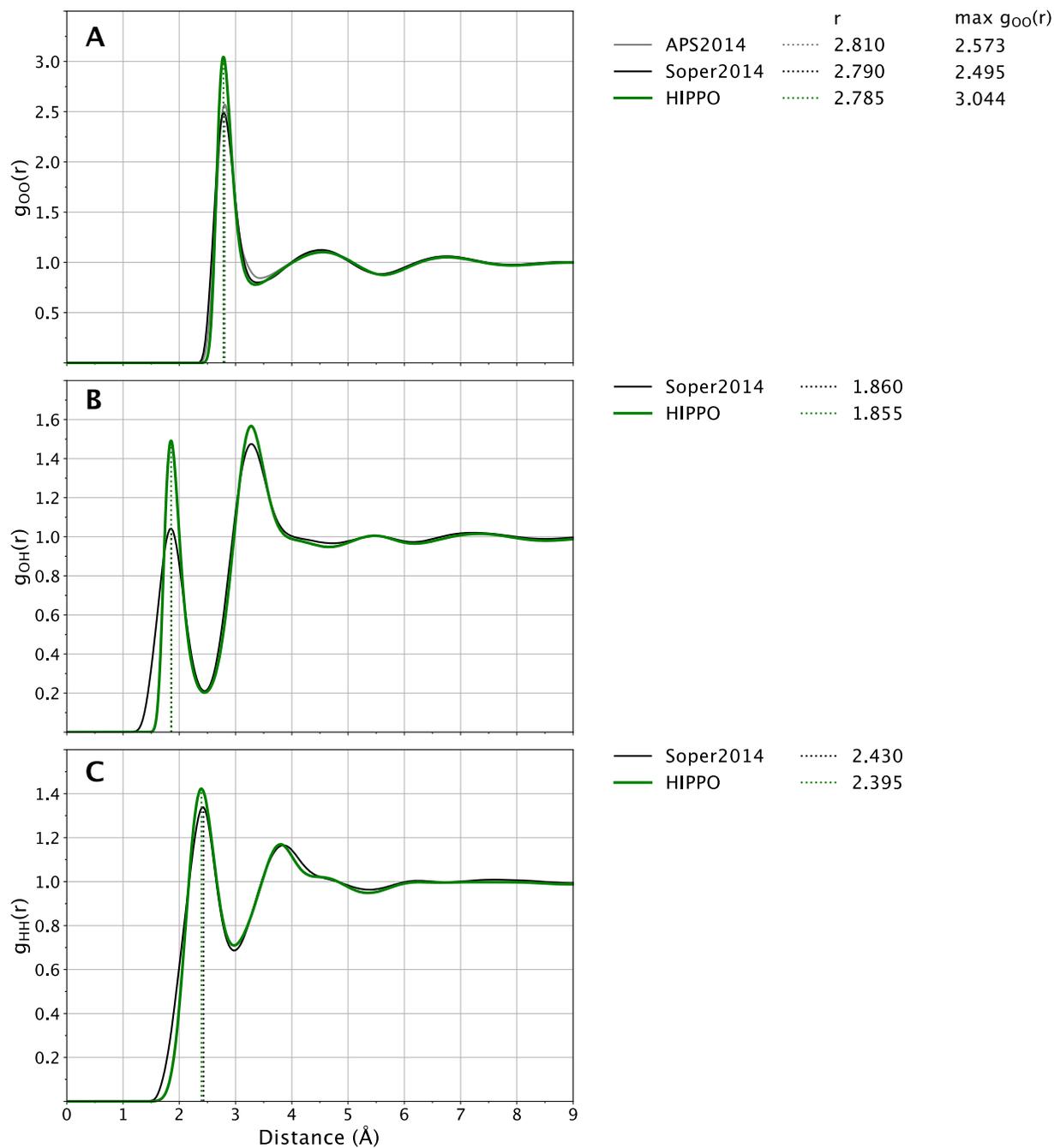


Figure 2.8. Water radial distribution at 298 K and 1 atm. HIPPO results are shown in green and experimental in black. First peaks of the HIPPO distribution are indicated with dotted green vertical lines. Experimental curves from references⁸¹⁻⁸³.

Temperature and Pressure Dependence. To stress the water model, we also tested the HIPPO model at a range of temperatures and pressures. This data is included to evaluate how well the HIPPO model performs in a variety of conditions away from room temperature and ambient pressure. We calculated the density, enthalpy of vaporization, heat capacity, dielectric constant, self-diffusion coefficient, thermal expansion coefficient, and isothermal compressibility at temperatures ranging from super-cool up to the boiling point. The same was done for the second virial coefficient. We also calculated density at a range of pressures, up to 10,000 atm. Of these, only the density, enthalpy of vaporization and dielectric constant were included in the fitting procedure. The results are presented in Figure 2.9 to Figure 2.14, and are discussed in detail below.

The temperature dependence of the density of liquid water is unique. At high temperatures, the dependence is intuitive and straightforward: the higher the temperature, the lower the density. At lower temperatures, near the freezing point, however, the correlation is less intuitive. The curve “turns over” and the density starts to decrease as the temperature decreases. This dual dependence leads to the characteristic “temperature of maximum density” of water. Panel A of Figure 2.9 shows that HIPPO reproduces the entirety of this curve with exceptional accuracy. The error relative to experiment is less than 1% for all points on the curve. The HIPPO temperature of maximum density is 277 ± 2 K, which is in near-perfect agreement with the experimental value of 277 K.

The enthalpy of vaporization temperature dependence is simple. As temperature increases, the DH_{vap} decreases. This matches our intuitive understanding of how much energy it takes to remove a water model from the liquid phase. Panel B of Figure 2.9 shows that HIPPO exhibits this same behavior, but the slope of the curve is slightly steeper than experiment. The result is a near-

perfect enthalpy of vaporization at room temperature with errors of $\pm 3\%$ at the respective ends of the tested temperature spectrum.

The heat capacity of the HIPPO water model is plotted in panel D of Figure 2.9. Heat capacity is closely related to the derivative with respect to temperature of the enthalpy of vaporization. Since the slope of the enthalpy of vaporization shown in panel B is largely unchanged over the temperature range, the heat capacity is nearly a constant with respect to temperature. However, since the slope of the HIPPO model for the enthalpy of vaporization in panel B is too steep, the calculated heat of vaporization is noticeably higher than experiment. This difference is the result of a known shortcoming in all classical models of water: the neglect of nuclear quantum effects (NQE). Rough corrections of 6 cal/mol/K and 2 cal/mol/K have been suggested.^{84, 85} The ForceBalance program also implements an NQE correction for the enthalpy of vaporization,^{86, 87} and these corrected values are plotted in panel B of Figure 2.9. Analysis of the NQE correction and its ramifications are discussed in greater detail in the Discussion section below.

As a model whose intended future use is the solvation of biological macromolecules, the dielectric constant is of great importance. One of the main practical implications of using a polarizable water model in biomolecular simulations is accurately modeling both water in bulk solvent and isolated water molecules in, for instance, a protein binding pocket. We calculated the dielectric constant of the HIPPO model and the results are plotted in panel A of Figure 2.10. The dielectric constant is notoriously sensitive and difficult to converge. However, the HIPPO model shows good agreement across the range of temperatures. This stands in contrast to most fixed charge water models whose dielectric constants change very little with temperature.⁸⁸

Another typically sensitive property of water models is the self-diffusion coefficient. This property is also important to future biomolecular simulations because it is a contributing factor to

producing accurate timescales for macromolecular dynamics. Plotted in panel B of Figure 2.10 is the self-diffusion coefficient of HIPPO water *vs.* temperature. It is clear that the overall shape of the temperature dependence curve is correct, with the HIPPO diffusion slightly higher than experiment. The self-diffusion coefficient is a rough measure of the balance of hydrogen bonding *vs.* other types of intermolecular interactions in water. The agreement of HIPPO with experiment indicates this balance is accurate. Due to the steep rise in diffusion coefficient with temperature, the 11% overestimation by HIPPO at room temperature corresponds to only a small error along the temperature dimension. For example, the computed HIPPO coefficient of $2.557 \pm 0.026 \times 10^{-5}$ cm²/s at 298 K is equivalent to the experimental value at roughly 304 K. Figure 11 shows the dependence of the diffusion coefficient on the reciprocal dimension of the cubic simulation box, 1/L. The variation with box size is in agreement with the well-known correction suggested by Yeh and Hummer.⁶³ Since the Yeh-Hummer correction depends on the shear viscosity of each model, we feel a diffusion *vs.* 1/L plot provides the best diffusion estimate at infinite box size for any specific water model. The HIPPO value obtained from Figure 11 is 2.568×10^{-5} cm²/s at 298 K, which is very close to the average of 2.557×10^{-5} cm²/s from multiple 100 Å cubic box simulations.

Finally, we show how the HIPPO water model performs under extreme conditions. Panels C and E show the temperature dependence of the thermal expansion coefficient and isothermal compressibility, respectively. The HIPPO compressibility is higher than the experimental value. This agrees with the pressure dependence of the density shown in Figure 2.13, where the density is greater than predicted for high pressure simulations. Note, however, the units of compressibility are small. Water is very difficult to compress and the HIPPO model of water is only slightly less so. The agreement of the thermal expansion coefficient with experiment is better. Cold water

expands rapidly as it is heated up, but the rate of expansion slows as the temperature increases. HIPPO reproduces this trend, mirroring the behavior seen in the density vs. temperature curve.

Ice Properties. In addition to liquid properties, we tested the properties of ice crystals. Due to its variety of structures, ice is a stringent test of the intermolecular potential. The intermolecular distances are generally shorter than in liquid water and thus stress the repulsive wall of the model. We computed lattice energies and densities for ten different ice polymorphs across a range of conditions. The HIPPO results are shown against curated experimental data in Table 2.5. Predicted densities are in error by no more than 2.5% and the lattice energies are all within 3% of the experimental values.

Table 2.5. Ice properties from HIPPO model compared with experimental density (kg/m^3) and lattice energy (kcal/mol). ^a Experimental values from reference ⁸⁹. ^b Values from the ICE10 data set.⁹⁰

	Density				Lattice Energy			
	HIPPO	Expt ^a	Diff	% Error	HIPPO	Ref ^b	Diff	% Error
Ice XI	949.9	934	15.9	1.7	-13.804	-14.10	0.3	-2.1
Ice Ih	910.8	920	-9.2	-1.0	-13.699	-14.07	0.4	-2.8
Ice IX	1164.8	1194	-29.2	-2.4	-13.769	-13.97	0.2	-1.4
Ice XIV	1316.7	1294	22.7	1.8	-13.360	-13.74	0.4	-2.9
Ice XV	1355.6	1364	-8.4	-0.6	-13.365	-13.48	0.1	-0.8
Ice Ic	951.8	931	20.8	2.2				
Ice Ica	947.2	931	17.2	1.8				
Ice XIII	1279.1	1251	28.1	2.2				
		MAD:	18.9			MAD:	0.3	

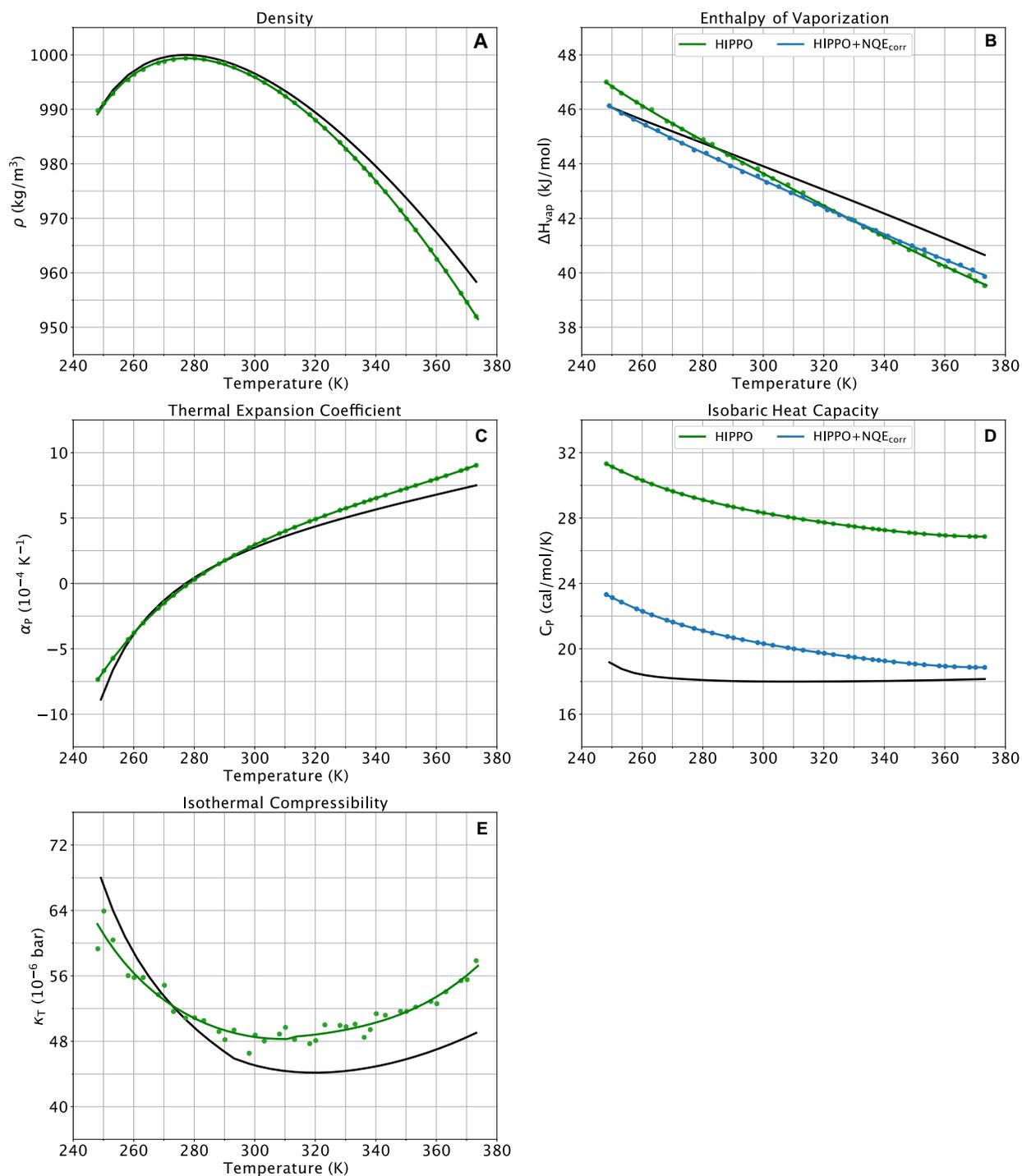


Figure 2.9. Thermodynamic water properties for the HIPPO model (green) compared to experiment (black) for temperatures from 248 to 373 K at atmospheric pressure (1 atm). (A), (C), (D), (E) experimental values from reference ⁷⁵. (B) experimental data from reference ⁷⁶.

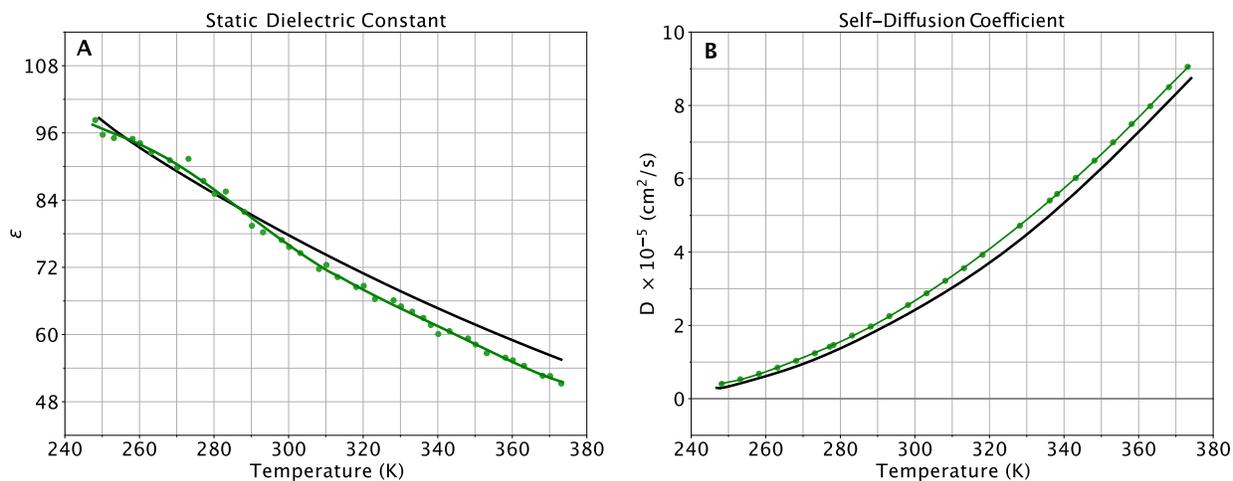


Figure 2.10. Dynamical water properties for the HIPPO model (green) compared to experiment (black) for temperatures from 248 to 373 K at atmospheric pressure (1 atm). (A) experimental values from reference ⁷⁷. (B) experimental values taken from references ^{78, 91, 92}.

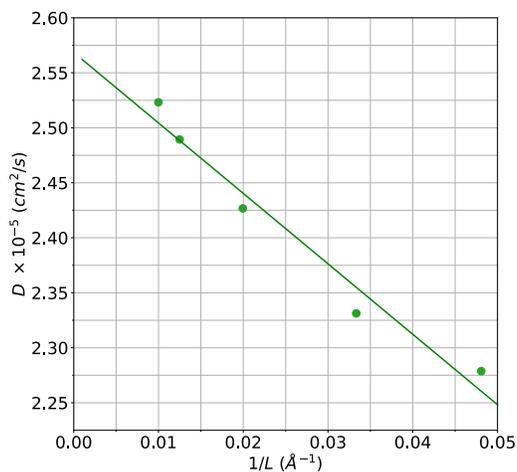


Figure 2.11. Self-diffusion coefficient vs. cubic box size at 298 K. The extrapolated y-axis intercept, corresponding to the estimated diffusion coefficient at infinite box size, is $2.568 \times 10^{-5} \text{ cm}^2/\text{s}$.

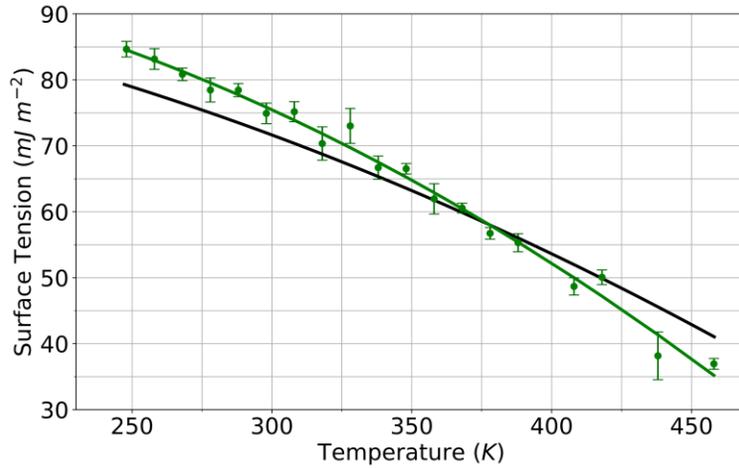


Figure 2.12. Surface Tension for the HIPPO model (green) compared to experiment (black) for temperatures from 248 K to 458 K. Experimental values from references ⁷⁹ and ⁹³.

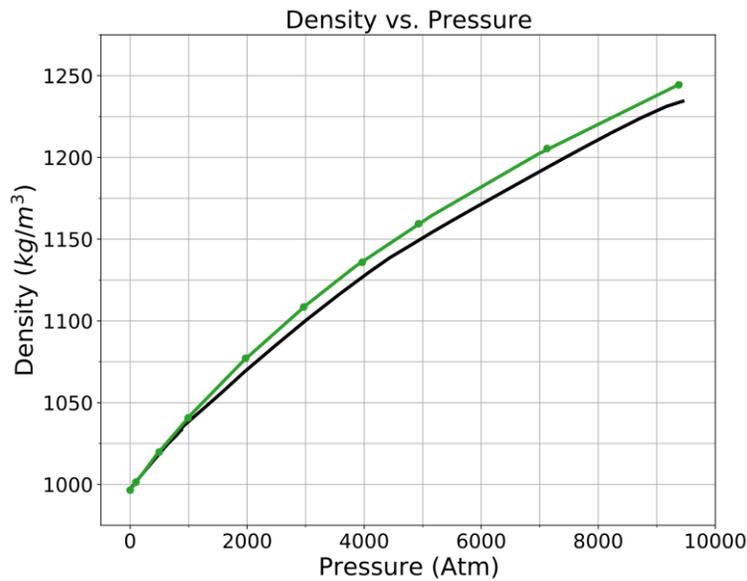


Figure 2.13. Density of the HIPPO model (green) compared to experiment (black) for pressures from 1 to 9,375 atm at room temperature (298 K). Experimental values from reference ⁹⁴.

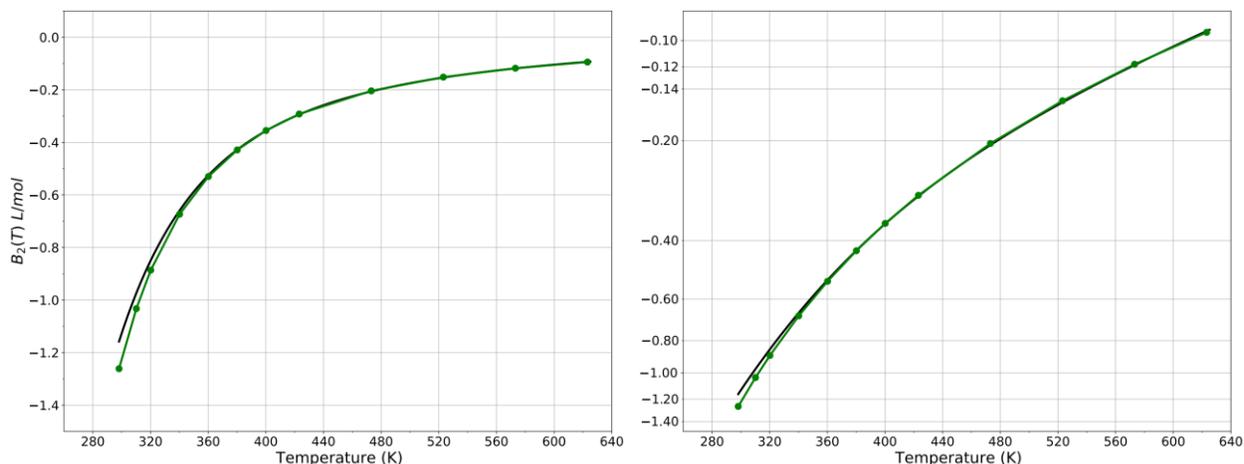


Figure 2.14. Second virial coefficient of the HIPPO water model (green) compared to experiment (black) for temperatures from 298 to 575 K. Experimental values from references ⁸⁰ and ⁹⁵.

2.5. Discussion

Implications of Parameter Space. HIPPO is a force field derived from our understanding of how atoms and molecules interact in short range. For this reason, our first goal in building a water model was to guarantee it could accurately reproduce high-level QM calculations for different configurations of water dimers. This is the reason we fit the initial water parameters using the SAPT energy decomposition as a reference for each energy component.

The second and most relevant goal of HIPPO water model is to be appropriate to a variety of MD applications, including solvation of biomolecules. Therefore, the model needs to agree with experimental data available within a small tolerance. The initial parameter set obtained after SAPT fitting did not meet our requirements for the condensed phase properties and led us to continue improving the model.

We chose to perform a constrained global search in parameter space using the repulsion, charge transfer and dispersion parameters of oxygen and hydrogen, centered at their initial values. With the exception of the charge transfer parameters, each parameter was not allowed to vary by

any more than 5%. Besides providing improvement of liquid properties, we chose this method because it could give us insight into how the features of the potential energy surface of water dimers related to condensed phase properties with respect to parameter space. Our optimizer was set to only compute liquid properties for parameters that kept the energy of water dimers within an average deviation 0.5 kcal per energy component, compared to the SAPT reference. The flexibility of 0.5 kcal in SAPT component is explained by the fact that SAPT calculations have intrinsic errors compared to gold standard CCSD(T) values. With that requirement, we were able to generate hundreds of water models. This showed we were dealing with a rough potential energy surface and the initial SAPT fitting put the model in a shallow minimum well. Upon computing water properties at room temperature for all the models generated, we selected the one with the smallest combined deviation from condensed phase experimental data and SAPT energy components.

Using the large amount of simulation data generated during parameter optimization, correlation analysis was performed between the energy components of low energy water dimer structures and liquid properties computed for the same set of parameters. Beyond a few obvious exceptions involving repulsion, no clear correlation was seen between calculated liquid properties and dimer total energy or components. Although there is some selection bias in the data, the 0.5 kcal/mol variance permitted for SAPT components should have allowed observation of correlation if it existed. The lack of correlation suggests the model parameter space is rugged. This in turn suggests orders of magnitude more QM dimer and cluster data would be needed to build a completely *ab initio* force field. This suspicion is given credence by the experience of the MB-pol and GEM water models, both of which required thousands of structures to produce well-determined models.^{96, 97}

Limitations of a Classical Model. By nature of being a classical model, HIPPO has a set of limitations. As illustrated in the Results section, the agreement between experimental and *ab initio* data, while good, is not perfect. These inconsistencies generally arise because of the classical approximations the HIPPO model employs. In this section, we will briefly enumerate some of the most important limitations of the model. We will also rationalize why, despite these limitations, HIPPO is capable of agreement with experiment as good or better than some of the best published water models.

Nuclear Quantum Effects. One of the most prominent areas for which the HIPPO water model is in disagreement with experiment is the heat capacity. This discrepancy is rooted in a physical effect that the HIPPO model does not directly address: Nuclear Quantum Effects or NQEs. NQEs show up in a variety of physical attributes of water. An instructive comparison is between H₂O and D₂O, where D₂O is meant to represent “classical” water with significantly less impact from nuclear quantum effects. The density of D₂O is 0.3% smaller, the dielectric constant is 0.5% smaller, and the enthalpy of vaporization is 3.3% larger than those of H₂O. These are mostly small effects that have been largely accounted for *via* our parameterization procedure. The heat capacity, however, is different. C_p at room temperature for D₂O is 11% larger than that of H₂O. This difference is too large to be covered by flexibility in parameterization, and furthermore the nature of the difference makes it virtually impossible to do so with a classical model.

The root of all NQEs, but most especially the heat capacity effect, is the treatment of hydrogens as classical oscillators. According to the Born-Oppenheimer approximation, under which conventional molecular dynamics operates, both intra- and intermolecular vibrational modes of hydrogen in the HIPPO model are treated classically. This treatment is essentially incorrect from the standpoint of quantum mechanics, where the vibrations of hydrogen should be

treated at quantum oscillators. The characteristic vibrations of a water molecule lie in the frequency range 1000-4000 cm^{-1} . However, at room temperature the amount of available thermal energy, $k_B T$, corresponds to a frequency of $\sim 200 \text{ cm}^{-1}$. This means that for virtually all of the vibrational modes of hydrogen atoms in water, the spacing between energy levels is much greater than the amount of thermal energy available. At room temperature, corrections of $6 \text{ cal mol}^{-1} \text{ K}^{-1}$ and $2 \text{ cal mol}^{-1} \text{ K}^{-1}$ to account for this difference between quantum oscillators and the classical model have been proposed.^{84, 98} Moreover, these considerations show that the magnitude of the error caused by imposing a classical model on a quantum system is temperature dependent. As temperature is decreased, vibrational excitation becomes more and more difficult as $k_B T$ drops. However, when temperature is increased, $k_B T$ becomes closer to the energy spacing of the hydrogen atom's low-frequency modes, allowing more vibrational excitation. The upshot of this temperature dependent error is that one should expect a classical water model to exhibit a heat of vaporization that is too high at low temperature and too low at high temperature. This is exactly the behavior seen in the HIPPO water model, giving it a heat capacity slightly higher than experiment.

Of course, the solution to fix this error in the heat capacity is to use a method that goes beyond the Born-Oppenheimer approximation to include NQEs. Other classical models have used methods such as path integral molecular dynamics (PIMD) or ring-polymer molecular dynamics (RPMD) with some success. Application of this methodology to the HIPPO model would be of great interest, given the otherwise high fidelity with experiment.

There are two likely reasons why HIPPO still attains good agreement with experiment despite not including nuclear quantum effects. The first is that for properties besides heat capacity, the impact of NQEs is small. The second reason is that while HIPPO is rooted in *ab initio* EDA calculations, it is not strictly an *ab initio* model. This means that there is some flexibility in

parameterization that has allowed HIPPO to fit H₂O experimental data without losing fidelity to the SAPT data from which it was originally derived. This flexibility is the driving force behind the parameterization process described in the methods section. In order to include NQEs implicitly, we optimized the initial *ab initio* derived parameters of the water model to reproduce H₂O liquid properties.

Many-Body Effects. The HIPPO model includes many-body effects through its polarization model. This induced dipole model allows for a linear order, classical electrostatic response of each atom to its environment. The results in the “Larger Clusters” section of the Results show that the model captures a majority of the total three-body energy of water clusters. However, there are other many-body effects which the HIPPO model does not include.

The first set of many-body effects excluded from HIPPO are “classical” electrostatic effects. These arise from terms involving higher-order polarizabilities, hyperpolarizabilities or charge transfer. Various water models such as NEMO and the ASP series of Stone and co-workers have included higher-order and hyperpolarizabilities.^{99,100} Similarly, there exist models for many-body charge transfer in water, such as those of Rick¹⁰¹ as well as the forthcoming SIBFA water model.¹⁰² The distinction between these various terms is not well defined and is presently the subject of intense scrutiny. However, their roots, regardless of nomenclature, are the same. They all describe the response of a molecule’s electron density to its electrostatic environment to infinite order. HIPPO includes just the least computationally expensive leading term of the full expansion. Models that include higher-order or hyperpolarizabilities, or charge transfer are attempting to select those additional terms representing the largest additional portion of the full expansion. While HIPPO does include a pairwise charge-transfer term, the decision to not include any of the higher-order many-body effects derives from a simple observation. As shown in Figure 2.7 the missing

part of the HIPPO three-body energy of water clusters is about 0.1 kcal/mol per molecule. This error is an order of magnitude smaller than other errors in the force field relative to *ab initio* results. The comparison indicates why HIPPO is capable of a high degree of agreement with experiment despite neglecting higher-order effects.

Of course, classical effects are not the only thing at play in intermolecular interactions. There exist many-body components to the dispersion and Pauli repulsion components of the intermolecular potential as well. There is a body of work showing that for some systems these quantum many-body effects, particularly many-body dispersion, can be important.¹⁰³⁻¹⁰⁵ There are also a number of models available for including these effects in classical potentials.¹⁰⁶⁻¹⁰⁸ However, the computational cost to include these effects for the purposes of the HIPPO model is prohibitive. Moreover, work from the Head-Gordon and co-workers has shown that the magnitude of these quantum many-body effects is insignificant for water-water and water-ion interactions.⁷² Because many-body dispersion and repulsion account for less than 1% of the total many-body energy, even for close-contact water clusters, they are neglected by the current HIPPO model.

How Good is SAPT for Water? A question one might ask is, “why not fit the HIPPO water model exclusively to SAPT data?” The suggestion in the question certainly has appeal. Fitting exclusively to SAPT would put HIPPO in the category of “*ab initio*” water models. The goal of the HIPPO project, however, is not to recreate any particular level of QM theory; it is to accurately predict experimental thermodynamic results. This is the reason we chose to refine the HIPPO parameters to reproduce the experimental density and heat of vaporization of liquid water. This strategy, however, leaves the HIPPO model in a middle ground that bears some explanation. Why use SAPT if the end result is ultimately fit to experiment?

To answer this question, it is helpful to look at the quality of an alternative “*ab initio*” version of the HIPPO water model. For the purposes of discussion, we will refer to this model as HIPPO-SAPT. Plotted in Figure 2.15 are the room temperature liquid properties of the stage one HIPPO-SAPT water model, fit to SAPT data, not yet refined for any experimental properties. This model is fit exclusively to SAPT2+ data on ~25 water dimer structures. Each of the components was fit individually, as outlined in the methods section. One can see from Figure 2.15 and Table 2.6 that the condensed phase properties of this model at room temperature are not far from the experimental values.

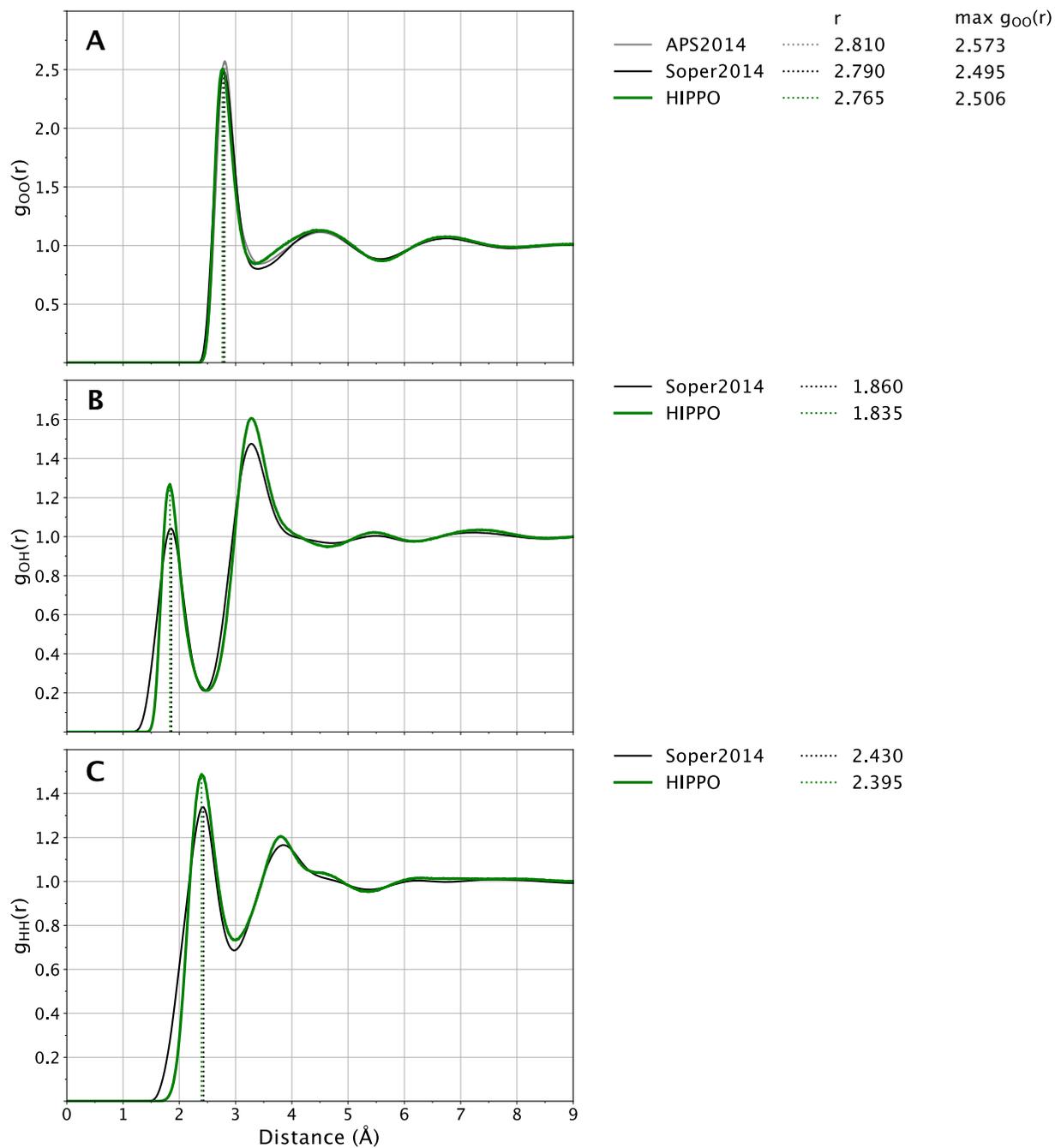


Figure 2.15. Water radial distribution function for the HIPPO-SAPT water model before optimization with ForceBalance at 298 K and 1 atm. HIPPO-SAPT results are shown in green and experimental in black. First peaks of the HIPPO-SAPT distribution are indicated with dotted green vertical lines. Experimental curves from references⁸¹⁻⁸³.

Table 2.6. HIPPO-SAPT water properties at room temperature (298 K and 1 atm).

Property	HIPPO-SAPT	Experiment	Abs. Error (%)
Density (kg/m ³)	974.195	997.045	22.85 (2.3)
Enthalpy of Vaporization (kJ/mol)	41.194	43.989	2.795 (6.4)
Self-Diffusion Coefficient (10 ⁻⁵ cm ² /s)	2.805	2.230	0.575 (25.8)
Static Dielectric Constant	80.050	78.409	1.641 (2.1)

The radial distribution functions for O-O, O-H, and H-H are all in good agreement with experiment. The dielectric constant is also very close to the experimental value. The model is not perfect, however. There are significant discrepancies in the density, enthalpy of vaporization and self-diffusion coefficient, and the density vs. temperature curve for this model exhibits no maximum. These data indicate the quality of SAPT for water. Within the confines of the SAPT2+ level of theory, with an aug-cc-pVDZ basis set, SAPT is capable of producing a “rough” water model, but not one up to the accuracy of empirical polarizable force fields such as AMOEBA.

Because water is the most important component of any biomolecular force field, the level of accuracy of this SAPT “*ab initio*” force field is not sufficient. The general accuracy of the model from these initial parameters, however, tells us about the utility of using SAPT as a reference. HIPPO is built on a series of successive approximations. The model is fit to SAPT, but SAPT has some measurable error relative to CCSD(T), the so-called gold standard of quantum chemistry.

CCSD(T), despite the title, however, is not perfect either. CCSD(T) uses the Born-Oppenheimer approximation, and as such is missing nuclear quantum effects (NQEs). This means that rather than using SAPT as a hard reference, the HIPPO strategy is instead to use SAPT as a guide. The SAPT data serves to solve the biggest problem in non-*ab initio* force fields: overdetermination. Requiring that HIPPO satisfy the SAPT components dramatically limits the parameter space available in the refinement phase of parameterization. This means that while it is not an *ab initio* force field, HIPPO is qualitatively different from empirical force fields because it follows the clearly identifiable series of approximations just described.

Transferability. Within the confines of any particular functional form - density-based, point-charge or otherwise - there are an infinite number of equally good water models. This is a simple consequence of a problem that is overdetermined by its nature. Unlike the simplest fixed partial charge water models, many advanced or polarizable models have several tunable parameters, but a sparse number of experimental observables to fit against. What makes the HIPPO model unique is that it limits itself to a narrow window of parameter space by insisting that SAPT energy decomposition data be satisfied. This is true not just for water-water SAPT calculations. The final parameters of this water model produce an RMS error on the entire S101x7 database of less than 1.0 kcal/mol per component on average. This means that the relaxation of the parameters to fit liquid properties did not disrupt the backbone of the HIPPO framework. These data suggest that the HIPPO water model will not only reproduce pure water properties well, but also perform well as a solvent. Although yet untested, this natural fit between the water model and the rest of the future force field is important. Recent work has shown that various point charge water models can produce dramatically different results for protein simulations.¹⁰⁹ The emphasis on SAPT in the

HIPPO model gives confidence that this water model will work well with the HIPPO small molecule and macromolecule models currently under development.

2.6. Conclusions

The quality of a water model is a subjective quantity. The utility of a particular model depends upon the kinds of scientific questions one wants to answer. The “best” water model for a job will change depending on whether one wants a rough solvation model or a detailed comparison with spectroscopic values. For bulk phase properties, a number of water models traditionally used in molecular dynamics simulation, as well as more recent models, provide generally similar results. Importantly, however, models sufficiently accurate for homogeneous pure water simulation may not be appropriate to account for solvation by water in heterogeneous environments. **Table 2.7** provides a minimal set of pure liquid properties for a subset of available models, including the HIPPO model described in this work.

Table 2.7. Selected properties of some water models used in MD simulation. HIPPO values are from the current work. Parameterization and data for other models are taken from: AMOEBA+,⁵³ AMOEBA03,⁹ AMOEBA14,¹¹⁰ TTM3-F,¹¹¹ SWM4-NDP and SWM6,¹¹² MB-POL,⁶² MB-UCB,¹¹³ TIP3P,¹¹⁴ TIP4P-Ew¹¹⁵ and TIP5P.¹¹⁶ Dimer energy and heat of vaporization are in kcal/mol, density in g/cm³, and diffusion coefficient as 10⁻⁵ cm²/s.

Model	E _{dimer}	Density	DH _{vap}	Diffusion	Dielectric
Reference	-4.97	0.997	10.51	2.30	78.4
HIPPO	-4.96	0.997	10.47	2.56	76.9
AMOEBA+	-4.85	0.998	10.6	2.14	78.8
AMOEBA03	-4.96	1.000	10.48	2.02	81
AMOEBA14	-4.64	0.998	10.63	2.36	79.4
TTM3-F	-5.18	0.994	11.4	2.37	94.4

SWM4-NDP	-5.15	0.994	10.45	2.85	78.0
SWM6	-5.27	0.996	10.52	2.14	78.1
MB-POL	-5.05	1.007	10.93	2.8	68.4
MB-UCB	-5.06	0.999	10.58	–	–
TIP3P	-6.02	0.982	10.45	6.11	82
TIP4P-Ew	-6.18	0.995	10.58	2.44	63.9
TIP5P	-6.78	0.979	10.46	2.78	92

With the above in mind, we conclude by attempting to place the HIPPO water model in context. First, we lay out a general taxonomy of water models and attempt to place HIPPO in that scheme. Second, we present the level of accuracy one can expect when using a water model out of a particular class in the taxonomy. Third, we use these ideas to motivate exactly what the HIPPO model is intended to be useful for. And lastly, we summarize the main scientific points uncovered in the process of developing the HIPPO potential.

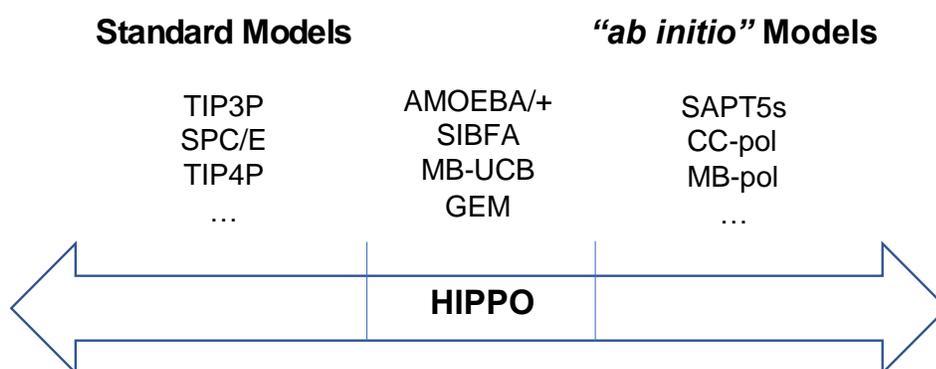


Figure 2.16. A non-exhaustive taxonomy of classical water models.

Despite the staggering number of published classical water models, the existing atom-based models can be roughly grouped into three general categories: empirical, *ab initio* and

physics-based. These three subsets loosely define a spectrum as illustrated in Figure 2.16, with empirical on one side, *ab initio* at the extreme, and physics-based in the middle. On the *ab initio* side are models fit solely to data from quantum mechanical calculations. On the other hand, empirical models are calibrated largely against experimental condensed phase properties. Models in the sparsely populated middle of the spectrum, which we term “physics-based”, attempt to reproduce both bulk phase and quantum mechanical calculation data simultaneously.

Examples of the empirical class of water models are the SPC and TIPS families of potentials. These may vary in the number and placement of interaction sites, but the functional form is essentially fixed: a Lennard-Jones van der Waals function coupled with point charge electrostatics. Because of this limited functional form, such models rely heavily on cancellation of errors. Thus, they are fit primarily to reproduce bulk phase properties of water around room temperature and pressure. The sheer number of published parameterizations of this functional form is a testament to how much flexibility is available during the fitting process. Because of this, most empirical models do give good agreement with the properties of water at room temperature, including a roughly correct description of the radial distribution function. However, these models are typically unable to capture fine-grained details of water structure. They struggle, for example, to correctly rank the ten Smith dimers or accurately predict the 2nd virial coefficient. For this reason, parameterizations of general biomolecular force fields often are calibrated using a specific water model. The main advantage of balancing these model costs is speed. Empirical water models remain the tool of choice when extensive sampling or simulating large systems is of greater importance than quantitative model accuracy.

On the other end of the spectrum are the *ab initio* water models. These can be further subdivided into two camps: (1) unique models intended just for water, such as the ASP-W¹¹⁷ and

TTM¹¹¹ series, which rely on the unique electronic properties of water, and (2) big data-derived water models, such as SAPT-5s,¹¹⁸ CC-pol,¹¹⁹ and MB-pol,⁶² which are based on large amounts of high-quality *ab initio* data. What all *ab initio* models have in common, though, is they are primarily fit to reproduce quantum mechanical data. This gives them a level of accuracy much higher than empirical models. They generally give a high-fidelity description of the Born-Oppenheimer potential energy surface of water, and are able to capture the bulk property temperature dependence (modulo nuclear quantum effects) and detailed structural features of water very accurately. Moreover, many of these models are able to reproduce spectroscopic properties such as vibrational frequencies due to their fidelity to the underlying quantum mechanics. These qualities come with two major tradeoffs. First, because of their complexity, these models are generally much slower than empirical models. They are too slow, for instance, to efficiently sample biomolecular-sized systems. Second, the framework for these models is not easily generalizable to complex, heterogeneous systems. To date, none of the *ab initio* class of models have been successfully extended to produce a complete biomolecular force field.

The final class of water models lies in the space between empirical and *ab initio*. These “physics-based” models attempt to satisfy both quantum mechanical and bulk phase data simultaneously by employing more complex functional forms intended to directly approximate the underlying Born-Oppenheimer potential energy surface. Examples of models in this class are AMOEBA,^{9, 110} AMOEBA+,⁵³ GEM,¹⁶ SIBFA,¹²⁰ MB-UCB,¹¹³ SWM,¹¹² and HIPPO. Because these models are classical approximations, the approximations used mean that there is a slight degradation of the Born-Oppenheimer surface compared to good *ab initio* models. For example, such physics-based models are generally not highly accurate for spectroscopic properties. Several of these models, however, are capable of quantitatively reproducing structural and energetic

properties across a wider range of conditions. For instance, we have shown in this work that HIPPO is capable of quantitatively predicting water dimer properties, cluster energies, and the 2nd virial coefficient. Additionally, physics-based models, unlike the empirical class, can reliably reproduce the temperature dependence of bulk phase liquid properties. A detailed comparison of many-body energetics for the *ab Initio* MB-pol and TTM models against several physics-based polarizable water potentials was recently presented by Lambros and Paesani.¹²¹ We agree with their conclusion that many-body *ab Initio* and physics-based water potentials should continue in parallel and with an eye toward ultimate convergence. HIPPO's use of the SAPT framework and explicit consideration of many-body energies within clusters is an initial step toward such convergence.

Having outlined what purposes best suit each class of model, the question is: If one needs a physics-based model, why consider HIPPO over the alternatives? For predicting many properties of water, HIPPO performs as well or better than the other listed models. However, this does not make HIPPO different in kind from the other models in its class. What makes HIPPO qualitatively different is the systematic, traceable series of approximations upon which it is constructed. HIPPO is based upon a model for charge density, from which every nonbonded term of the force field is derived. This allows the model to provide a direct approximation of Symmetry Adapted Perturbation Theory. SAPT is in turn approximate with respect to the current "gold standard" level of quantum chemistry, CCSD(T), which is in turn an approximation of the exact Born-Oppenheimer surface. Although the errors accumulated across this series of approximations place the derived HIPPO model too far from the exact potential surface to be a true *ab initio* model, this lineage gives HIPPO two properties that make it unique:

1. It dramatically limits the parameter search space for optimization against experimental data.

2. It gives a specific framework from which to build a more general, and complete molecular force field.

These qualities certainly contribute to the fidelity with which HIPPO predicts properties of water, but their primary value will lie in the ability to extend HIPPO to other molecular systems in the future.

An important point to make about the HIPPO force field is that despite its more complete and complex set of equations, the computational cost of the model is roughly equal to other physics-based models. In the Tinker9 and OpenMM 7.4.0 computer codes, both of which implement HIPPO and AMOEBA on GPUs, the difference in cost between the two models is negligible.

Several lessons were learned during the process of developing the HIPPO water model. First, SAPT2+ is insufficient to build an *ab initio* water model. Since the bulk properties of water are sensitive to small changes in the water dimer potential energy surface, we found that fitting only to SAPT2+ data could not produce a satisfactorily precise and accurate model. Second, the use of an underlying charge density is critical to the accurate modeling of both short- and long-range intermolecular interactions. This is obviously true for electrostatics, but it is no less true for other parts of the force field, including polarization, repulsion and dispersion. HIPPO shows that a charge density formulation can produce accurate many-body interactions vis-a-vis a polarization model, and we have demonstrated that a charge density model is also necessary to accurately reproduce van der Waals interactions. Third, atomic anisotropy is essential for a physics-based model, and is necessary to achieve fully correct behavior for water dimers and clusters. Importantly, we show this anisotropy is just as important in the repulsion component of the force field as in the traditional electrostatic portion. Furthermore, the HIPPO functional form illustrates

that the nature of the anisotropy can be effectively captured by an energy model derived from the atomic multipole moments. Finally, and practically, we make the observation that dramatic improvement in the short-range physics of a force field can be incorporated without significant additional computational cost. Because the short-range terms have simple asymptotic behavior, the cost of HIPPO is comparable to or less than many of its physics-based force field peers.

The goal is for the HIPPO water model to become the cornerstone of a general force field for water, ions, organics and biomolecules. The critical importance of water as the solvent in many simulations justifies the high level of attention described in this work. The strength of interactions with monoatomic ions provides a useful stress test for new potentials. We are currently exploring HIPPO water-ion energetics along the lines of prior studies of AMOEBA water with ions.^{122, 123} Continued parameterization for organic molecules and biomolecules will make use of the Caleman, *et. al* database of over 1200 experimental properties and values for 146 organic liquids,¹²⁴ and the S101x7 SAPT data set,²⁹ respectively. From the experience gained with water, the plan is to obtain atomic multipole values and polarizabilities from DMA and potential fitting.¹⁰ Then we will use genetic and least squares optimization methods to fit liquid properties across multiple molecules simultaneously, using SAPT values from S101x7 data as guides via loose restraints. Lessons learned in the development of the HIPPO water model should prove useful as physics-based force fields progress toward maturity.

2.7. References

1. Bernal JD, Fowler RH. A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. *J. Chem. Phys.*, 1, 515-48 (1933).
2. Stillinger FH, Rahman A. Improved Simulation of Liquid Water by Molecular Dynamics. *J. Chem. Phys.*, 60, 1545-57 (1974).
3. Jorgensen WL. Quantum and Statistical Mechanical Studies of Liquids 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.*, 103, 335-40 (1981).

4. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction Models for Water in Relation to Protein Hydration. In: Pullman B, editor. *Intermolecular Forces*: Reidel, Dordrecht; 1981. p. 331-42.
5. Yesylevskyy SO, Schafer LV, Sengupta D, Marrink SJ. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.*, 6, e1000810 (2010).
6. Yesudasan S. Extended MARTINI Water Model for Heat Transfer Studies. *Mol. Phys.*, 118, e1692151 (2020).
7. Chen M, Ko H-Y, Remsing RC, Calegari Andrade MF, Santra B, Z. S, Selloni A, Car R, Klein ML, Perdew JP, Wu X. Ab Initio Theory and Modeling of Water. *Proc. Natl. Acad. Sci. USA*, 114, 10846-51 (2017).
8. Jia W, Wang H, Chen M, Lu D, Lin L, Car R, W. E, Zhang L, editors. Pushing the Limit of Molecular Dynamics with Ab Initio Accuracy to 100 Million Atoms with Machine Learning. ACM Gordon Bell Prizes, SC20; November 2020: ACM.
9. Ren P, Ponder JW. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B*, 107, 5933-47 (2003).
10. Ren P, Wu C, Ponder JW. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.*, 7, 3143-61 (2011).
11. Dudek MJ, Ponder JW. Accurate Modeling of the Intramolecular Electrostatic Energy of Proteins. *J. Comput. Chem.*, 16, 791-816 (1995).
12. Sherrill CD. Energy Component Analysis of Pi Interactions. *Acc. Chem. Res.*, 46, 1020-8 (2013).
13. Slipchenko LV, Gordon MS. Electrostatic Energy in the Effective Fragment Potential Method: Theory and Application to Benzene Dimer. *J. Comput. Chem.*, 28, 276-91 (2007).
14. Piquemal J-P, Cisneros GA, Reinhardt P, Gresh N, Darden TA. Towards a Force Field Based on Density Fitting. *J. Chem. Phys.*, 124, 104101 (2006).
15. Cisneros GA, Piquemal JP, Darden TA. Generalization of the Gaussian Electrostatic Model: Extension to Arbitrary Angular Momentum, Distributed Multipoles, and Speedup with Reciprocal Space Methods. *J. Chem. Phys.*, 125, 184101 (2006).
16. Duke RE, Starovoytov ON, Piquemal J-P, Cisneros GA. GEM*: A Molecular Electronic Density-Based Force Field for Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, 10, 1361-5 (2014).
17. Torabifard H, Starovoytov ON, Ren P, Cisneros GA. Development of an AMOEBA Water Model Using GEM Distributed Multipoles. *Theor. Chem. Acc.*, 134, 101 (2015).
18. Jeziorski B, Moszynski R, Szalewicz K. Perturbation Theory Approach of Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.*, 94, 1887-930 (1994).
19. Szalewicz K. Symmetry-Adapted Perturbation Theory of Intermolecular Forces. *WIREs Comput. Mol. Sci.*, 2, 254-72 (2012).
20. Rackers JA, Wang Q, Liu C, Piquemal J-P, Ren P, Ponder JW. An Optimized Charge Penetration Model for Use with the AMOEBA Force Field. *Phys. Chem. Chem. Phys.*, 19, 276-91 (2017).
21. Rackers JA, Liu C, Ren P, Ponder JW. A Physically Grounded Damped Dispersion Model with Particle Mesh Ewald Summation. *J. Chem. Phys.*, 149, 084115 (2018).
22. Rackers JA, Ponder JW. Classical Pauli Repulsion: An Anisotropic, Atomic Multipole Model. *J. Chem. Phys.*, 150, 084104 (2019).
23. Coulson CA. Two-Centre Integrals Occurring in the Theory of Molecular Structure. *Math. Proc. Cambridge*, 38, 210-23 (1942).

24. Freitag MA, Gordon MS, Jensen JH, Stevens WJ. Evaluation of Charge Penetration between Distributed Multipolar Expansions. *J. Chem. Phys.*, *112*, 7300-6 (2000).
25. Piquemal JP, Gresh N, Giessner-Prettre C. Improved Formulas for the Calculation of the Electrostatic Contribution to the Intermolecular Interaction Energy from Multipolar Expansion of the Electronic Distribution. *J. Phys. Chem. A*, *107*, 10353-9 (2003).
26. Narth C, Lagardere L, Polack E, Gresh N, Wang Q, Bell DR, Rackers JA, Ponder JW, Ren PY, Piquemal JP. Scalable Improvement of SPME Multipolar Electrostatics in Anisotropic Polarizable Molecular Mechanics Using a General Short-Range Penetration Correction Up To Quadrupoles. *J. Comput. Chem.*, *37*, 494-506 (2016).
27. Wang B, Truhlar DG. Including Charge Penetration Effects in Molecular Modeling. *J. Chem. Theory Comput.*, *6*, 3330-42 (2010).
28. Wang B, Truhlar DG. Screened Electrostatic Interactions in Molecular Mechanics. *J. Chem. Theory Comput.*, *10*, 4480-7 (2016).
29. Wang Q, Rackers JA, He C, Qi R, Narth C, Lagardere L, Gresh N, Ponder JW, Piquemal J-P, Ren P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.*, *11*, 2609-18 (2015).
30. Misquitta AJ. Charge Transfer from Regularized Symmetry-Adapted Perturbation Theory. *J. Chem. Theory Comput.*, *9*, 5313-26 (2013).
31. Maitland GC, Rigby M, Smith EB, Wakeham WA. Intermolecular Forces: Their Origin and Determination. Oxford, UK: Oxford University Press; 1981.
32. Tang KT, Toennies JP. An Improved Simple Model for the van der Waals Potential Based on Universal Damping Functions for the Dispersion Coefficients. *J. Chem. Phys.*, *80*, 3726-41 (1984).
33. Brooks FC. Convergence of Intermolecular Force Series. *Phys. Rev.*, *86*, 92-7 (1952).
34. Ahlrichs R, Penco R, Scoles G. Intermolecular Forces in Simple Systems. *Chem. Phys.*, *19*, 119-30 (1977).
35. Grimme S. Density Functional Theory with London Dispersion Corrections. *WIREs Comput. Mol. Sci.*, *1*, 211-28 (2011).
36. Grimme S, Ehrlich S, Georigk L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.*, *32*, 1456-65 (2011).
37. Wu Q, Yang W. Empirical Correction to Density Functional Theory for van der Waals Interactions. *J. Chem. Phys.*, *116*, 515-24 (2002).
38. Chai J-D, Head-Gordon M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.*, *10*, 6615-20 (2008).
39. Johnson ER, Becke AD. A Post-Hartree-Fock Model of Intermolecular Interactions. *J. Chem. Phys.*, *123*, 024101 (2005).
40. Slipchenko LV, Gordon MS. Damping Functions in the Effective Fragment Potential Method. *Mol. Phys.*, *107*, 999-1016 (2010).
41. Verma P, Wang B, Fernandez LE, Truhlar DG. Physical Molecular Mechanics Method for Damped Dispersion. *J. Phys. Chem. A*, *121*, 2855-62 (2017).
42. Salem L. The Forces Between Polyatomic Molecules. II. Short-Range Repulsive Forces. *Proc. R. Soc. Lon. Ser.-A*, *264*, 379-91 (1961).
43. Musher JI, Salem L. Energy of Interaction between Two Molecules. *J. Chem. Phys.*, *44*, 2943-6 (1966).
44. Murrell JN, Randić M, Williams DR. The Theory of Intermolecular Forces in the Region of Small Orbital Overlap. *Proc. R. Soc. Lon. Ser.-A*, *284*, 566-81 (1965).

45. Murrell JN, Shaw G. Intermolecular Forces in the Region of Small Orbital Overlap. *J. Chem. Phys.*, 46, 1768-72 (1967).
46. Gresh N. Energetics of Zn²⁺ Binding to a Series of Biologically Relevant Ligands: A Molecular Mechanics Investigation Grounded on ab Initio Self-Consistent Field Supermolecular Computations. *J. Comput. Chem.*, 16, 856-82 (1995).
47. Piquemal J-P, Chevreau H, Gresh N. Toward a Separate Reproduction of the Contributions to the Hartree– Fock and DFT Intermolecular Interaction Energies by Polarizable Molecular Mechanics with the SIBFA Potential. *J. Chem. Theory Comput.*, 3, 824-37 (2007).
48. Jensen JH, Gordon MS. An Approximate Formula for the Intermolecular Pauli Repulsion between Closed Shell Molecules. *Mol. Phys.*, 89, 1313-25 (1996).
49. Jensen JH, Gordon MS. An Approximate Formula for the Intermolecular Pauli Repulsion between Closed Shell Molecules. II. Application to the Effective Fragment Potential Method. *J. Chem. Phys.*, 108, 4772-82 (1998).
50. Allinger NL, Yuh YH, Lii J-H. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.*, 111, 8551-66 (1989).
51. Dinur U, Hagler AT. Geometry-Dependent Atomic Charges: Methodology and Application to Alkanes, Aldehydes, Ketones, and Amides. *J. Comput. Chem.*, 16, 154-70 (1995).
52. Fanourgakis GS, Xantheas SS. The Bend Angle of Water in Ice Ih and Liquid Water: The Significance of Implementing the Nonlinear Monomer Dipole Moment Surface in Classical Interaction Potentials. *J. Chem. Phys.*, 124, 174504 (2006).
53. Liu C, Piquemal JP, Ren P. Implementation of Geometry-Dependent Charge Flux into the Polarizable AMOEBA+ Potential. *J. Phys. Chem. Lett.*, 11, 419-26 (2020).
54. Rackers JA, Wang Z, Lu C, Laury ML, Lagardere L, Schnieders MJ, Piquemal J-P, Ren P, Ponder JW. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.*, 14, 5273-89 (2018).
55. Harger M, Li D, Wang Z, Dalby K, Lagardere L, Piquemal JP, Ponder JW, Ren P. Tinker-OpenMM: Absolute and Relative Alchemical Free Energies Using AMOEBA on GPUs. *J. Comput. Chem.*, 38, 2047-55 (2017).
56. Wang Z. Tinker9: Next Generation of Tinker with GPU Support. St. Louis: Washington University in St. Louis; 2021.
57. Simmonett AC, Pickard IV FC, Schaefer III HF, Brooks BR. An Efficient Algorithm for Multipole Energies and Derivatives Based on Spherical Harmonics and Extensions to Particle Mesh Ewald. *J. Chem. Phys.*, 140, 184101 (2014).
58. Smith BJ, Swanton DJ, Pople JA, Schaefer III HF. Transition Structures for the Interchange of Hydrogen Atoms Within the Water Dimer. *Chem. Phys.*, 92, 1240-7 (1990).
59. Wang L-P, Martinez TJ, Pande VS. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.*, 5, 1885-91 (2014).
60. Frenkel D, Smit B. Understanding Molecular Simulation: From Algorithms to Applications. Section 5.4, San Diego, CA, USA: Academic Press; 2002.
61. Bussi G, Donadio D, Parrinello M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.*, 126, 014101 (2007).
62. Reddy SK, Straight SC, Bajaj P, Huy Pham C, Riera M, Moberg DR, Morales MA, Knight C, Gotz AW, Paesani F. On the Accuracy of the MB-Pol Many-Body Potential for Water: Interaction Energies, Vibrational Frequencies, and Classical Thermodynamic and Dynamical Properties from Clusters to Liquid Water and Ice. *J. Chem. Phys.*, 145, 194504 (2016).

63. Yeh I-C, Hummer G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B*, *108*, 15873-9 (2004).
64. Gloor GJ, Jackson G, Blas FJ, de Miguel E. Test-Area Simulation Method for the Direct Determination of the Interfacial Tension of Systems with Continuous or Discontinuous Potentials. *J. Chem. Phys.*, *123*, 134703 (2005).
65. Kirkwood JG, Buff FP. The Statistical Mechanical Theory of Surface Tension. *J. Chem. Phys.*, *17*, 338-43 (1949).
66. Tschumper GS, Leininger ML, Hoffman BC, Valeev EF, Schaefer III HF, Quack M. Anchoring the Water Dimer Potential Energy Surface with Explicitly Correlated Computations and Focal Point Analyses. *J. Chem. Phys.*, *116*, 690-701 (2002).
67. Szalewicz K, Jeziorski B. Comment on "On the Importance of the Fragment Relaxation Energy Terms in the Estimation of the Basis Set Superposition Error Correction to the Intermolecular Interaction Energy". *J. Chem. Phys.*, *109*, 1198-200 (1998).
68. Wang Y, Bowman JM. Towards an ab Initio Flexible Potential for Water, and Post-Harmonic Quantum Vibrational Analysis of Water Clusters. *Chem. Phys. Lett.*, *491*, 1-10 (2010).
69. van Duijneveldt-van de Rijdt JGCM, Mooij WTM, van Duijneveldt FB. Testing the quality of some recent water–water potentials. *Phys. Chem. Chem. Phys.*, *5*, 1169-80 (2003).
70. Vasilyev V. Online Complete Basis Set Limit Extrapolation Calculator. *Comput. Theor. Chem.*, *1115*, 1-3 (2017).
71. Halkier A, Helgaker T, Jorgensen P, Klopper W, Olsen J. Basis-Set Convergence of the Energy in Molecular Hartree–Fock Calculations. *Chem. Phys. Lett.*, *302*, 437-46 (1999).
72. Demerdash O, Mao YZ, Liu TY, Head-Gordon M, Head-Gordon T. Assessing Many-Body Contributions to Intermolecular Interactions of the AMOEBA Force Field Using Energy Decomposition Analysis of Electronic Structure Calculations. *J. Chem. Phys.*, *147*, 161721 (2017).
73. Miliordos E, Xantheas SS. An Accurate and Efficient Computational Protocol for Obtaining the Complete Basis Set Limits of the Binding Energies of Water Clusters at the MP2 and CCSD(T) Levels of Theory: Application to (H₂O)_m, m = 2-6, 8, 11, 16, and 17. *J. Chem. Phys.*, *142*, 234303 (2015).
74. Bates DM, Tschumper GS. CCSD(T) Complete Basis Set Limit Relative Energies for Low-Lying Water Hexamer Structures. *J. Phys. Chem. A*, *113*, 3555-9 (2009).
75. Kell GS. Density, Thermal Expansivity, and Compressibility of Liquid Water from 0.deg. to 150.deg.. Correlations and Tables for Atmospheric Pressure and Saturation Reviewed and Expressed on 1968 Temperature Scale. *J. Chem. Eng. Data*, *20*, 97-105 (1975).
76. Wagner W, Pruß A. The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *J. Phys. Chem. Ref. Data*, *31*, 387-535 (2002).
77. Hamelin J, Mehl JB, Moldover MR. The Static Dielectric Constant of Liquid Water Between 274 and 418 K Near the Saturated Vapor Pressure. *Int. J. Thermophys.*, *19*, 1359-80 (1998).
78. Mills R. Self-Diffusion in Normal and Heavy Water in the Range 1-45.deg. *J. Phys. Chem.*, *77*, 685-8 (1973).
79. Vargaftik NB, Volkov BN, Voljak LD. International Tables of the Surface Tension of Water. *J. Phys. Chem. Ref. Data*, *12*, 817-20 (1983).

80. CRC Handbook of Chemistry and Physics. 84th ed. Lide DR, editor. Boca Raton, FL: CRC Press LLC; 2004.
81. Brookes DH, Head-Gordon T. Family of Oxygen–Oxygen Radial Distribution Functions for Water. *J. Phys. Chem. Lett.*, 6, 2938-43 (2015).
82. Soper AK. The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure? *ISRN Phys. Chem.*, 2013, 279463 (2013).
83. Skinner LB, Benmore CJ, Neuefeind JC, Parise JB. The Structure of Water Around the Compressibility Minimum. *J. Chem. Phys.*, 141, 214507 (2014).
84. Levitt M, Hirshberg M, Sharon R, Laidig KE, Daggett V. Calibration and Testing of a Water Model for Simulation of the Molecular Dynamics of Proteins and Nucleic Acids in Solution. *J. Phys. Chem. B*, 101, 5051-61 (1997).
85. Yu H, Hansson T, van Gunsteren WF. Development of a Simple, Self-Consistent Polarizable Model for Liquid Water. *J. Chem. Phys.*, 118, 221-34 (2003).
86. Wang L-P, Chen J, Van Voorhis T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.*, 9, 452-60 (2013).
87. Habershon S, Markland TE, Manolopoulos DE. Competing Quantum Effects in the Dynamics of a Flexible Water Model. *J. Chem. Phys.*, 131, 024501 (2009).
88. Fennell CJ, Li LB, Dill KA. Simple Liquid Models with Corrected Dielectric Constants. *J. Phys. Chem. B*, 116, 6936-44 (2012).
89. Petrenko VF, Whitworth RW. Physics of Ice. Oxford, UK: Oxford University Press; 1999.
90. Brandenburg JG, Maas T, Grimme S. Benchmarking DFT and Semiempirical Methods on Structures and Lattice Energies for Ten Ice Polymorphs. *J. Chem. Phys.*, 142, 124104 (2015).
91. Gillen KT, Douglass DC, Hoch MJR. Self-Diffusion in Liquid Water to -31°C . *J. Chem. Phys.*, 57, 5117-9 (1972).
92. Holz M, Heil SR, Sacco A. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements. *Phys. Chem. Chem. Phys.*, 2, 4740-2 (2000).
93. Vinš V, Fransen M, Hykl J, Hrubý J. Surface Tension of Supercooled Water Determined by Using a Counterpressure Capillary Rise Method. *J. Phys. Chem. B*, 119, 5567-75 (2015).
94. Sato H, Watanabe K, Sengers JMHL, Gallagher JS, Hill PG, Straub J, Wagner W. Sixteen Thousand Evaluated Experimental Thermodynamic Property Data for Water and Steam. *J. Phys. Chem. Ref. Data*, 20, 1023-44 (1991).
95. Kell GS, McLaurin GE, Whalley E, Schneider WG. PVT Properties of Water VII. Vapour Densities of Light and Heavy Water from 150 to 500° C. *P. Roy. Soc. Lond. A Mat.*, 425, 49-71 (1989).
96. Babin V, Leforestier C, Paesani F. Development of a “First Principles” Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *J. Chem. Theory Comput.*, 9, 5395-403 (2013).
97. Gökcan H, Kratz E, Darden TA, Piquemal J-P, Cisneros GA. QM/MM Simulations with the Gaussian Electrostatic Model: a Density-Based Polarizable Potential. *J. Phys. Chem. Lett.*, 9, 3062-7 (2018).
98. Eisenberg DS, Kauzmann W. The Structure and Properties of Water. New York: Oxford University Press; 1969.

99. Wallqvist A, Ahlström P, Karlström G. New Intermolecular Energy Calculation Scheme: Applications to Potential Surface and Liquid Properties of Water. *J. Phys. Chem.*, 94, 1649-56 (1990).
100. Millot C, Stone AJ. Towards an Accurate Intermolecular Potential for Water. *Mol. Phys.*, 77, 439-62 (1992).
101. Rick SW. A Polarizable, Charge Transfer Model of Water Using the Drude Oscillator. *J. Comput. Chem.*, 37, 2060-6 (2016).
102. Lagardere L, El-Khoury L, Naseem-Khan S, Aviat F, Gresh N, Piquemal JP. Towards Scalable and Accurate Molecular Dynamics Using the SIBFA Polarizable Force Field. *AIP Conf. Proc.*, 1906, 030018 (2017).
103. Donchev AG. Many-Body Effects of Dispersion Interaction. *J. Chem. Phys.*, 125, 074713 (2006).
104. Marom N, DiStasio RA, Atalla V, Levchenko S, Reilly AM, Chelikowsky JR, Leiserowitz L, Tkatchenko A. Many-Body Dispersion Interactions in Molecular Crystal Polymorphism. *Angew. Chem. Int. Ed.*, 52, 6629-32 (2013).
105. Elrod MJ, Saykally RJ. Many-Body Effects in Intermolecular Forces. *Chem. Rev.*, 94, 1975-97 (1994).
106. Chaudret R, Gresh N, Parisel O, Piquemal J-P. Many-Body Exchange-Repulsion in Polarizable Molecular Mechanics. I. Orbital-Based Approximations and Applications to Hydrated Metal Cation Complexes. *J. Comput. Chem.*, 32, 2949-57 (2011).
107. Tkatchenko A, DiStasio RA, Car R, Scheffler M. Accurate and Efficient Treatment of Many-Body van der Waals Interactions. *Phys. Rev. Lett.*, 108, 236402 (2012).
108. Tkatchenko A, Scheffler M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.*, 102, 073005 (2009).
109. Liem SY, Popelier PLA. The Influence of Water Potential in Simulation: A Catabolite Activator Protein Case Study. *J. Mol. Model.*, 25, 216 (2019).
110. Laury ML, Wang L-P, Pande VS, Head-Gordon T, Ponder JW. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B*, 119, 9423-37 (2015).
111. Fanourgakis GS, Xantheas SS. Development of Transferable Interaction Potentials for Water. V. Extension of the Flexible, Polarizable, Thole-type Model Potential (TTM3-F, v. 3.0) to Describe the Vibrational Spectra of Water Clusters and Liquid Water. *J. Chem. Phys.*, 128, 074506 (2008).
112. Yu W, Lopes PEM, Roux B, MacKerell JAD. Six-Site Polarizable Model of Water Based on the Classical Drude Oscillator. *J. Chem. Phys.*, 138, 034508 (2013).
113. Das AK, Urban L, Leven I, Loipersberger M, Aldossary A, Head-Gordon M, Head-Gordon T. Development of an Advanced Force Field for Water Using Variational Energy Decomposition Analysis. *J. Chem. Theory Comput.*, 15, 5001-13 (2019).
114. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.*, 79, 926-35 (1983).
115. Horn HW, Swope WC, Pitner JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.*, 120, 9665-78 (2004).
116. Mahoney MW, Jorgensen WL. A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. *J. Chem. Phys.*, 112, 8910-22 (2000).

117. Millot C, Soetens J-C, Martins Costa MTC. Revised Anisotropic Site Potentials for the Water Dimer and Calculated Properties. *J. Phys. Chem. A*, 1998, 754-70 (1998).
118. Groenenboom GC, Mas EM, Bukowski R, Szalewicz K, Wormer PES, van der Avoird A. Water Pair and Three-Body Potential of Spectroscopic Quality from Ab Initio Calculations. *Phys. Rev. Lett.*, 84, 4072-5 (2000).
119. Bukowski R, Szalewicz K, Groenenboom GC, van der Avoird A. Predictions of the Properties of Water from First Principles. *Science*, 315, 1249-52 (2007).
120. Devereux M, Gresh N, Piquemal J-P, Meuwly M. A Supervised Fitting Approach to Force Field Parameterization with Application to the SIBFA Polarizable Force Field. *J. Comput. Chem.*, 35, 1577-91 (2014).
121. Lambros E, Paesani F. How Good are Polarizable and Flexible Models for Water: Insights from a Many-Body Perspective. *J. Chem. Phys.*, 153, 060901 (2020).
122. Grossfield A, Ren P, Ponder JW. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *J. Am. Chem. Soc.*, 125, 15671-82 (2003).
123. Jiao D, King C, Grossfield A, Darden TA, Ren P. Simulation of Ca²⁺ and Mg²⁺ Solvation Using Polarizable Atomic Multipole Potential. *J. Phys. Chem. B*, 110, 18553-9 (2006).
124. Caleman C, van Maaren PJ, Hong M, Hub JS, Costa LT, van der Spoel D. Force Field Benchmark for Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.*, 8, 61-74 (2012).

Chapter 3. The Benzene Model

At the heart of molecular behavior lies quantum mechanics, which offers the most detailed and fundamental description of molecular interactions. For a molecule as significant as benzene, with its delocalized electrons and unique aromatic character, quantum mechanics plays a crucial role in determining its properties. An ideal force field model should, therefore, approximate these quantum mechanical behaviors closely. Doing so ensures that the model captures the essence of the molecule, from its electron distribution to its interaction potentials. By mirroring quantum mechanical properties, the HIPPO force field can provide both qualitative and quantitative accuracy, making simulations involving benzene and other aromatic systems more predictive and reliable for real-world applications.

3.1. Introduction

The benzene molecule occupies a storied position in the history of chemistry, and is arguably the most famous organic molecule. It is certainly of great interest from a theoretical and computational perspective. Benzene was first isolated and characterized in 1825 by a young Michael Faraday¹, who later abandoned experimental chemistry in favor of physics, from compressed illuminating gas obtained by pyrolysis of whale oil. Determination of the molecular structure was a contentious issue considered by many of the best-known chemists of the mid-19th century including Dewar, Ladenburg, Couper, Loschmidt and others. The correct structure with a six-membered ring consisting of alternating single and double carbon-carbon bonds was proposed by August Kekulé in 1865. At an 1890 symposium held in his honor, Kekulé described the idea for the ring structure as appearing to him in a dream of a snake biting its own tail- an ancient symbol known as the ouroboros. In order to explain the number of isomers of various substituted

benzene derivatives, Kekulé proposed in 1872 that his benzene structure must oscillate between two alternative structures such that all the carbon-carbon bonds become equivalent. A theoretical basis for this idea was advanced by Linus Pauling in 1928 as part of his concept of resonance.² Benzene exhibits an anomalous thermodynamic stability in comparison to other similar conjugated hydrocarbon molecules. Rationalization of this extra stabilization was one of the first chemical applications of the new physical theory of quantum mechanics, culminating in 1937 in the simple molecular orbital model devised by Erich Hückel.³ The resulting theory of aromaticity is one of the cornerstones of physical organic chemistry, and helps explain the ubiquity as well as the reactivity of benzene and myriad other aromatic compounds.

Benzene derivatives other aromatic moieties play an important role in biology. The amino acids phenylalanine, tyrosine, histidine and tryptophan contain aromatic side chains. So-called π -cation interactions⁴ involving aromatic amino acids and ions are strongly stabilizing due to a polarization effects that are not well described by traditional biomolecular force fields that do not include explicit polarization.⁵ The nucleic acid bases, both purines and pyrimidines, provide the variability differentiating DNA and RNA sequences. The base “stacking” interactions central to nucleic acid structure result from the same delicate balance of classical electrostatics and dispersion that determines the benzene dimer potential energy surface. For example, the exact vertical stacking of benzene rings lying in horizontal planes is electrostatically unfavorable and is not a minimum energy structure, but rather serves as an intermediate connecting lower energy parallel displaced structures. Finally, we note that phenyl groups and other aromatic and heteroaromatic rings are present in an extraordinarily large number of small molecule drugs. These substructures help to reduce aqueous solubility and provide favorable dispersion interaction and a

hydrophobic thermodynamic signature that enhances relative binding within protein pockets and active sites.

Benzene and the larger polycyclic aromatic hydrocarbons (PAHs) are significant environmental toxins and causative agents in a number of human cancers. The PAHs, found in soot, tar, coal and charred meats, directly induce bladder and scrotal cancer via reactive epoxide intermediates. Benzene itself is uniquely detrimental to human health, distinct from the widely understood effects of PAH molecules. Acute benzene toxicity manifests as pancytopenia and aplastic anemia as well as chromosomal changes.⁶ Repeated, prolonged exposure to benzene in occupational settings is reported to result in as much as a 7-fold increase in chronic myeloid leukemia (CML) and other hematopoietic cancers.⁷ For example, the during 1950s to 1980s drinking water at the U.S. Marine Corps base at Camp Lejeune, North Carolina exposed those present to benzene and other industrial solvents, and imparted a roughly 10% higher chance of death from all cancer, and a 40% higher rate of some subtypes of leukemia.⁸

Many studies have described benzene's interesting combination of experimental properties, such as its large and highly anisotropic molecular polarizability. Polarizability plays a key role in many of the intermolecular interactions of benzene. For example, the experimentally measured gas phase enthalpy of interaction of a benzene molecule with a potassium ion is -19 kcal/mol.⁹ Roughly half of this stabilization is due to induction effects resulting from polarization of the π -cloud of the benzene by the cation. Other properties have only been investigated with the past 15 years, such scattering experiments providing a deeper understanding of the conformations dominating the liquid phase.¹⁰

The potential energy surface for the benzene dimer is extremely rich, with an unusually large number of stationary points.¹¹ High-level quantum calculations have characterized much of

the potential surface, with the parallel-displaced and T-shaped being the lowest energy conformations and almost isoenergetic.¹² The benzene dimer provides a sobering example of the difficulty of determining chemically accurate interaction energies via *ab initio* electronic structure methods, even for seemingly simple systems. Complete basis set Moller-Plesset (MP) correlated calculations for the benzene dimer grossly over-stabilize both the T-shaped and parallel displaced dimer minima, and are particularly poor for treating dispersion-dominated interactions.^{13, 14} Expensive large basis set CCSD(T) calculations are required to achieve chemically accurate results.^{15 16} The crystal structure of benzene interlaces both of these low energy dimer motifs into a herringbone packing of high overall symmetry. A recent study estimated the lattice energy of the benzene crystal using a high-level *ab initio* many-body expansion approach.¹⁷

Numerous classical, empirical force field models of benzene have been developed.¹⁸ A recent comparison of force fields for dimerization and binding of aromatic molecules nicely summarizes a range of currently available models.¹⁹ Simple atomic partial charge models of benzene have only a single free parameter- the amount of charge separation between the carbon and hydrogen atoms. Most such models that attempt to account for the bulk phase properties of liquid benzene place a charge on the carbon of -0.1 to -0.15 electrons.^{20 21} Off-atom charge sites have been used to improve modeling of aromatic p-clouds within a simple partial charge framework, as in the XED force field.²² A common shortcoming of benzene potentials is their inability to capture charge overlap and penetration effects.^{23, 24} A Drude oscillator-based polarizable model for benzene and other aromatic molecules has been developed as part of the CHARMM Drude force field.²⁵ A recent reparameterization of the AMOEBA model for benzene and nucleic acid bases adds induced dipole polarization to atomic multipole electrostatics, but neglects charge penetration and charge transfer effects.²⁶

This study presents a new classical, semi-empirical model of benzene that provides an accurate description of benzene-benzene interactions across its different phases. This model was developed using the Hydrogen-like Polarizable Potential (HIPPO), which includes charge overlap effects by using a charge density to describe the electronic charge on each atom.

3.2. Methods

The parametrization procedure we used for benzene is like the one described in our previous work for water parametrization.²⁷ Multipole and valence parameter were fit to monomer data, and bond and angles were the same as the AMOEBA forcefield.²⁸ Our model also includes a charge flux term, and its parameters were taken from Yang, X. et al²⁹.

For the SAPT fitting procedure, we used 21 benzene dimers from S101x7 database.³⁰ Additional dimer configurations were added from the NCIA database³¹ as a second step in the fitting. These different structures were repulsive contacts of water-benzene and benzene-benzene dimers. Additional conformations from reference³² added high level CCSD(T) benchmarks to the fitting dataset. The finalized model was then evaluated against the full potential energy surfaces of homodimers and water-benzene dimers with the structures in the DES370k database.³³

Computational Details. All properties and simulations were obtained using the HIPPO force field as implemented in the Tinker and Tinker9 packages^{34,35}. To compute condensed phase properties, MD simulations of liquid benzene were performed. Unless otherwise noticed, properties were computed based on simulation of a cubic box of dimension ~ 40 Å and containing 450 benzene molecules. All liquid and solid phase simulations were performed using the RESPA (Reversible Reference System Propagator Algorithm) integrator; a Langevin piston barostat³⁶ and thermostat³⁷ was used for NPT ensembles and a Bussi thermostat was used for NVT ensembles.³⁸

For fitting, each MD simulation ran for 2 ns using a 2.0 fs time step, with a 0.5 ns equilibration phase and 1.5 ns production phase. The energy components of dimers were calculated using the ANALYZE program in Tinker.

The temperature dependence of liquid benzene was computed from three independent NPT simulations at each temperature. Temperatures ranged from 278 K to 348 K at 5 K intervals, totaling 15 temperature points. All 45 simulations ran at atmospheric pressure (1 atm.) Each simulation was started at the experimental density for the respective temperature and ran for 30 ns using a 2.0 fs time step; the first 5 ns of the simulations were discarded as equilibration. The reported results are the averages of the three independent runs. The standard deviation can be found in the supplemental material.

The viscosity was computed using a multi-simulation method and the Green-Kubo approach.^{39, 40} The starting configuration was taken from one of the equilibrated NPT simulations at 298 K (after 5 ns of production) such that the density closely matched the average density at that temperature. From the same initial structure, 40 independent NVT simulations were started each with a different seed velocity. A Bussi thermostat was used, coupled with 2 fs time step RESPA integrator. The simulations ran for 7 ns, and the first 4 ns were discarded as equilibration period. For the remainder 3 ns of production, the pressure tensor was saved every 10 fs. Detailed results are shown in the supplemental material.

The self-diffusion coefficient was computed following a procedure like the one described for the MB-Pol model.⁴¹ Fifty different starting structures were selected from the three equilibrated NPT simulations at 298 K (after at least 5 ns of production and at least 500 ps apart), none with density deviating more than 2% of the average density obtained from the three NPT simulations. Each configuration was then equilibrated for 10 ns in NVT ensemble, with a Bussi thermostat and

2 fs time step RESPA integrator. Following equilibration, each of them ran for another 500 ps in a NVE ensemble. The self-diffusion coefficient was computed from this 500 ps phase. The result presented is the average of the 50 independent calculations. A box-size correction was applied to each calculated self-diffusion coefficient.⁴² The viscosity used in the calculation of the correction was the one computed previously for this model.

To calculate the surface tension at 298 K, we reutilized the 40 NVT simulations previously used to calculate the viscosity. Each simulation box had already been equilibrated for 7 ns in a NVT ensemble. Then the Z-axis of each cubic box was expanded to three times the X-axis and Y-axis dimensions. The final system geometries were slabs with $X = Y = \sim 40 \text{ \AA}$, and $Z = \sim 120 \text{ \AA}$, with a vacuum layer along the Z-axis over each side of the slab. Each system was then further equilibrated for 10 ns in the NVT ensemble followed by a 10 ns production phase. The surface tension was calculated from the last 10 ns of data using the pressure tensor,⁴³ which was computed every picosecond. The final surface tension value is the average of the 40 independent calculations.

3.3. Results

Employing a strategy akin to our prior water model development, we assessed the quality of the benzene model using both ab initio calculations and experimental condensed phase data. In this section, we'll discuss monomer properties, dimer interaction energies, and condensed phase properties.

The complete specification of the HIPPO force field benzene model consists of 37 refined parameter values. Explicit values for these parameters, along with their associated units, are detailed in the Appendix D.

The first property we examined was the molecular polarizability, crucial for accurately reflecting benzene's recognized anisotropic nature. Table 3.1 shows HIPPO is in close agreement with both experimental and quantum calculations.

Table 3.1. *Molecular polarizability of benzene. Experimental value from reference⁴⁴. Ab initio calculation performed in PSI4 software.*

	Polarizability (\AA^{-3})		
	a_{xx}	a_{yy}	a_{zz}
HIPPO	11.782	11.782	6.656
MP2/aug-cc-PVTZ	11.896	11.896	6.614
Experiment	12.26	12.26	6.66

3.3.1. Dimer potential energy surface

Benzene stands as a cornerstone in organic chemistry. A multitude of complex molecules and polymers are derived from benzene. Hence, having a force field model for benzene that can faithfully reproduce its quantum mechanical properties is of paramount importance. This ensures not only a reliable representation of benzene itself but also lays the groundwork for accurately modeling more intricate systems derived from it.

It is well known that point charge models are not capable of reproducing the electrostatics properties of aromatic systems, due to their inability to capture charge penetration, and polarization effects. One of the most important goals of our new model is to be capable of describing π - π interactions at the level of accuracy required to provide predictive power in computing binding free energies. The first step at such task is to reproduce the dimer potential energy surface of benzene. The next figures show the results for HIPPO compared to common models

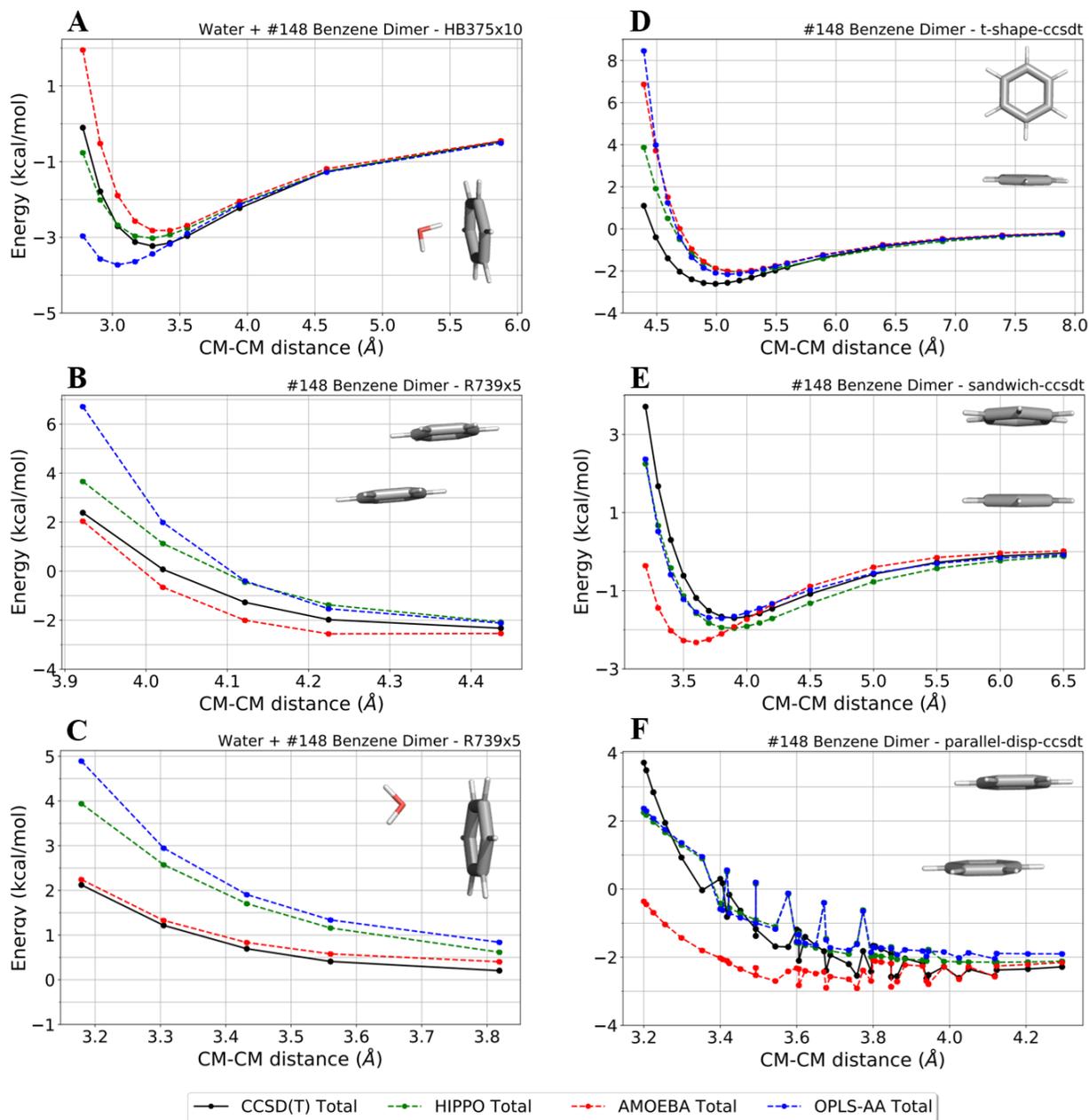


Figure 3.1. Total intermolecular energies for benzene-benzene and benzene-water dimers using the HIPPO, AMOEBA and OPLS-AA force fields. The left side panel (A, B, C) structures and reference values are from the NCIA datasets^{31, 45, 46}, all computed at the CCSD(T)/CBS limit level of theory. The right side structures (D,E,F) and *ab initio* values are from reference³², computed at the CCSD(T)/CBS limit resolution.

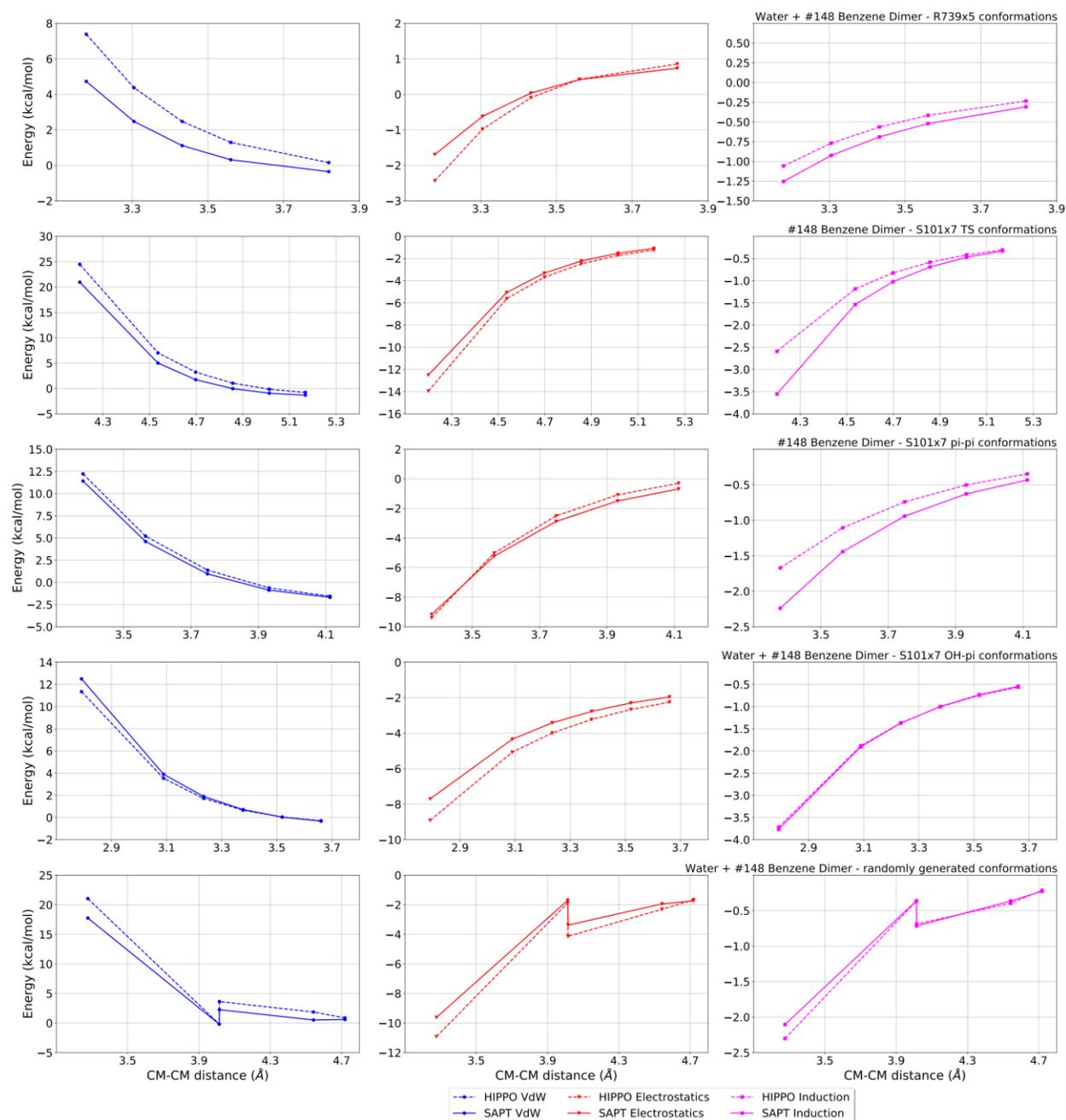


Figure 3.2. HIPPO energy components for dimers in the S101x7 and NCIA R739x5 databases, and randomly generated conformations. The components are Van der Waals, Electrostatics and Induction. Repulsion and Dispersion components were added to yield Van der Waals equivalent and allow comparison to empirical models. Reference values were computed at SAPT2+ level of theory. S101x7 references used the aug-cc-pVDZ basis set, whereas the others were computed in-house using the aug-cc-pVTZ basis set.

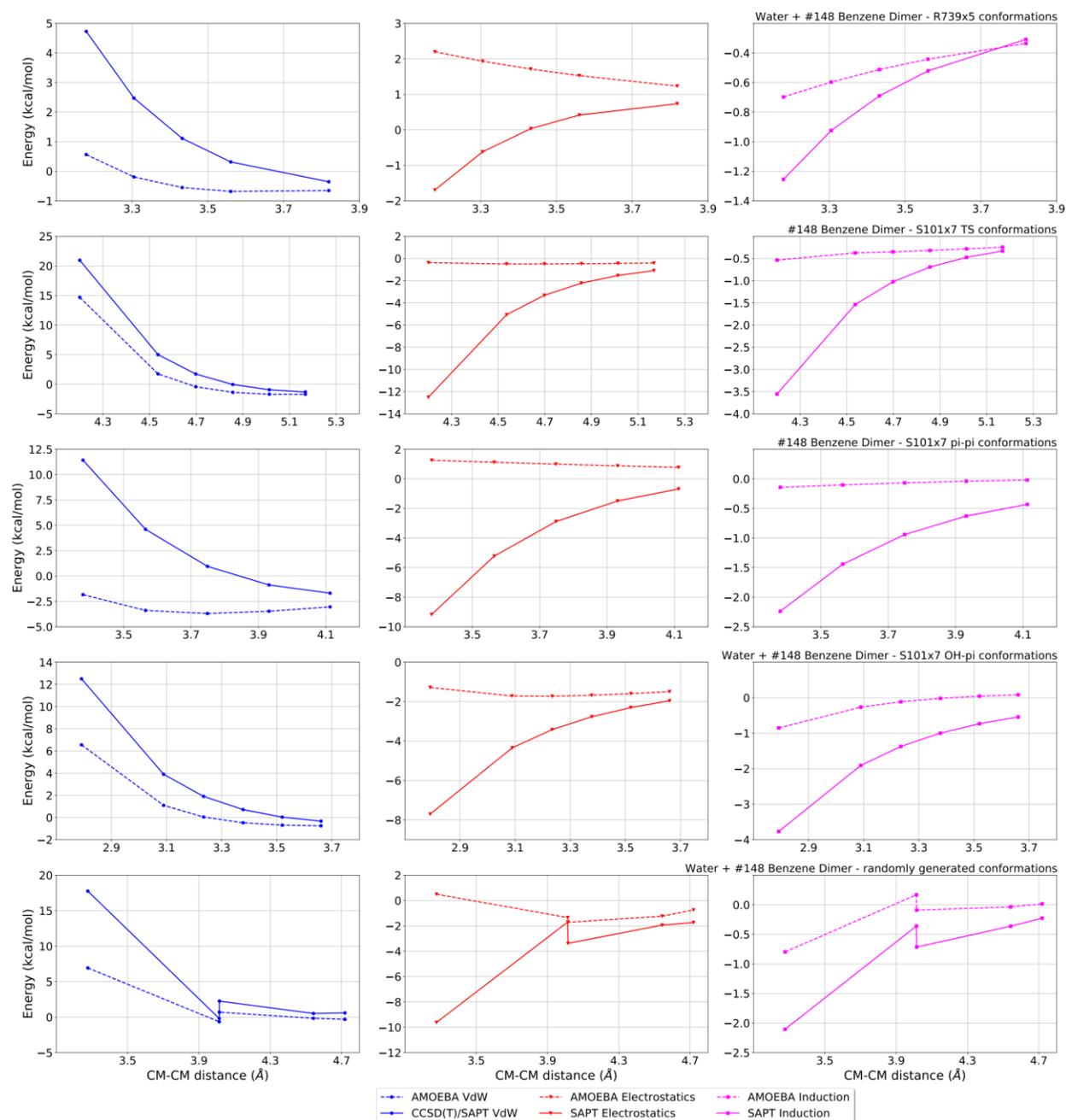


Figure 3.3. AMOEBA energy components for dimers in the S101x7 and NCIA R739x5 databases, and randomly generated conformations. The components are Van der Waals, Electrostatics and Induction. Repulsion and Dispersion components were added to yield Van der Waals equivalent and allow comparison to empirical models. Reference values were computed at SAPT2+ level of theory. S101x7 references used the aug-cc-pVDZ basis set, whereas the others were computed in-house using the aug-cc-pVTZ basis set.

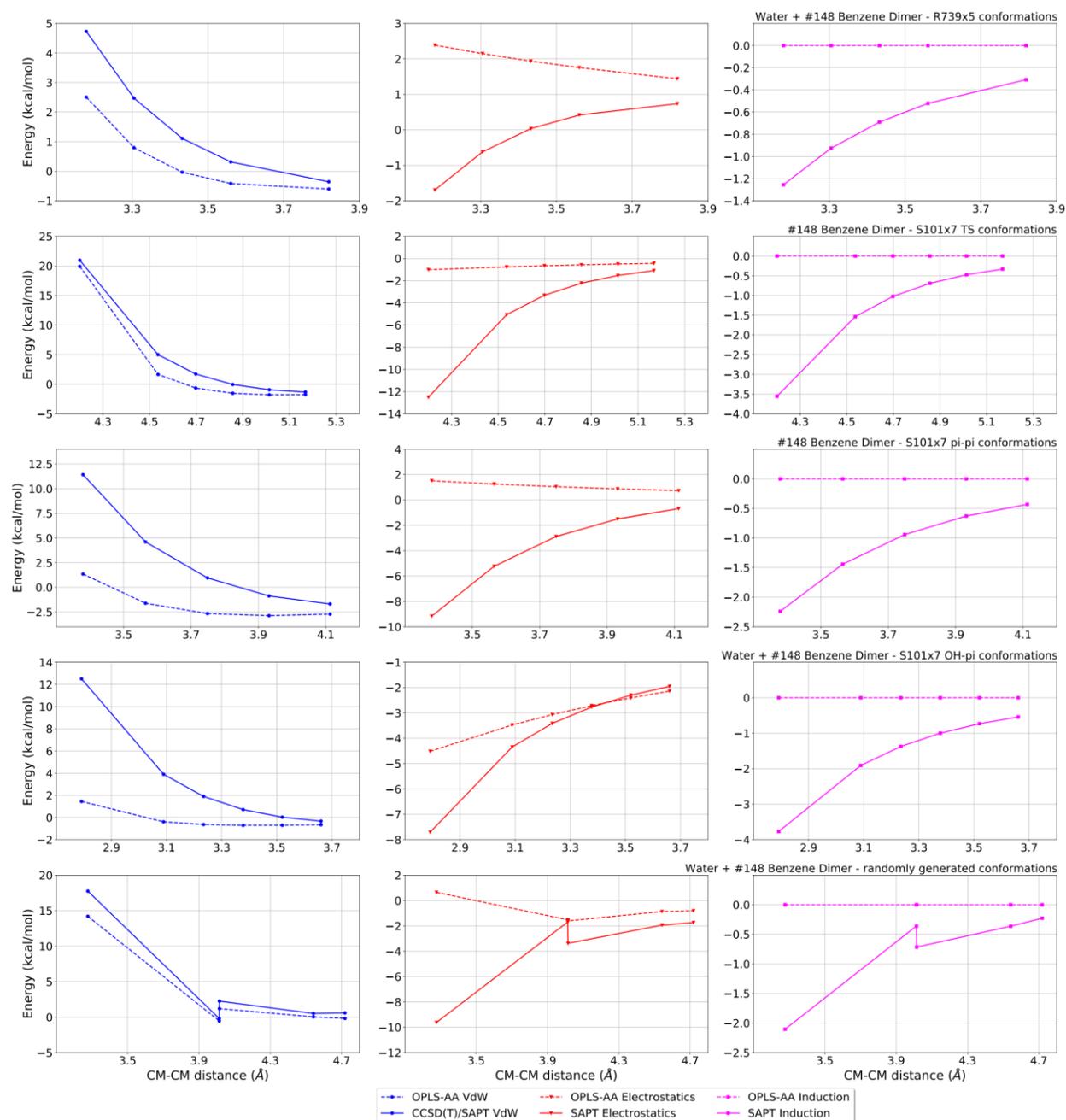


Figure 3.4. OPLS-AA energy components for dimers in the S101x7 and NCIA R739x5 databases, and randomly generated conformations. The components are Van der Waals, Electrostatics and Induction. Repulsion and Dispersion components were added to yield Van der Waals equivalent and allow comparison to empirical models. Reference values were computed at SAPT2+ level of theory. S101x7 references used the aug-cc-pVDZ basis set, whereas the others were computed in-house using the aug-cc-pVTZ basis set.

3.3.2. *Liquid*

Benzene plays a crucial role as an industrial solvent and is a key component in gasoline. Accurately modeling its behavior is vital, given the implications for petrochemical processes and environmental spill modeling, among other areas. Furthermore, benzene's carcinogenic properties necessitate a deep understanding of its interactions in biological systems. A force field that accurately captures its properties is instrumental in simulations that delve into its behavior within biological contexts, its interactions with biomolecules, and potential health ramifications.

Considering the significance of this model within the broader context of the HIPPO force field, meticulous attention was dedicated to ensuring that our model mirrors as many of benzene's liquid properties as possible. As indicated in **Table 3.2**, our model successfully reproduces the density, enthalpy of vaporization, dielectric constant, and viscosity within a 1% margin of their experimental benchmarks. Additionally, the model reasonably captures the surface tension—a notably challenging property to replicate—as well as the diffusion coefficient.

Table 3.2. Benzene liquid properties at 298 K. ^a Reference ⁴⁷, ^b Reference ⁴⁸, ^c Reference ⁴⁹, ^d Reference ⁵⁰, ^e Reference ⁵¹.

Property	HIPPO	Experimental	Abs. Deviation
Density (kg/m ³)	878.524	873.660 ^a	4.864 (0.56%)
Enthalpy of Vaporization (kJ/mol)	33.340	33.827 ^b	0.487 (1.44%)
Static Dielectric Constant	2.298	2.271 ^c	0.027 (1.19%)
Viscosity (10 ⁻⁴ Pa·s)	5.929	6.030 ^a	0.101 (1.67%)
Self-Diffusion Coefficient (10 ⁻⁵ cm ² /s)	1.97	2.19 ^d	0.22 (10.05%)
Surface Tension (mN/m)	31.74	28.22 ^e	3.52 (12.47%)

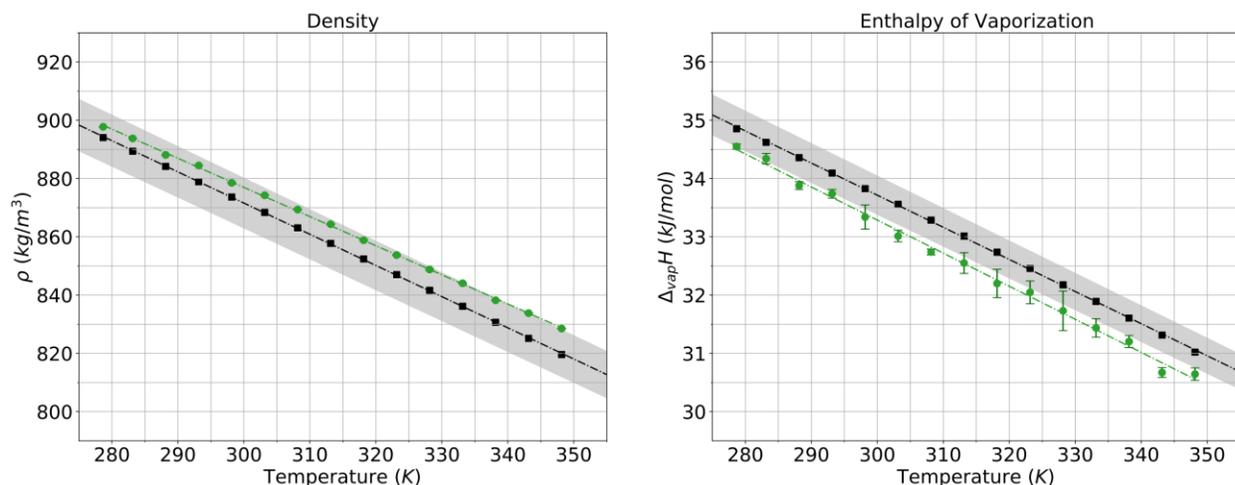


Figure 3.5. Temperature dependence of liquid benzene at 1 atm. Experimental values are represented by black squares, calculated HIPPO values are green circles. (a) Density, shaded area shows results within 1% of experimental data; reference ⁴⁷ (b) Enthalpy of Vaporization, shaded area shows region within 1% of experimental data; reference ⁴⁸ for temperature range 293-350; missing values are extrapolated from available data.

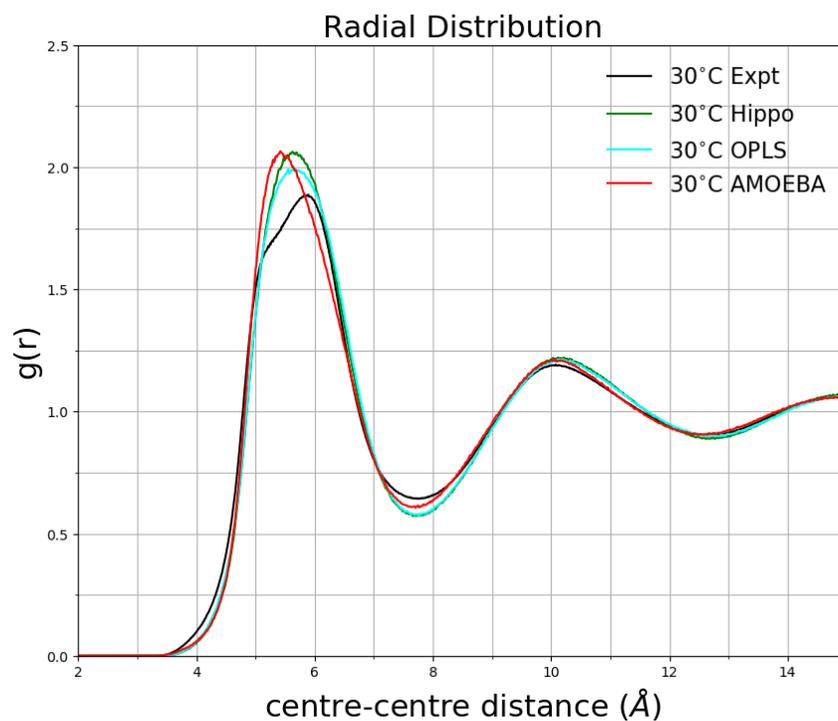


Figure 3.6. Center to center radial distribution function of HIPPO, AMOEBA and OPLS-AA compared to experimental neutron scattering data.¹⁰

Table 3.3. Hydration free energy of benzene for different polarizable models.

Experiment	AMOEBA	HIPPO
-0.87 kcal/mol	-1.0 kcal/mol	-0.5 kcal/mol

3.4. Conclusion

The rigor and precision of the HIPPO model are underscored by its outstanding capability to replicate the ab initio SAPT components and the CCSD(T) total interaction energy within a tight margin of 1 kcal/mol for numerous water-benzene and benzene-benzene dimers. This level of accuracy not only speaks to the model's robustness but also elevates its utility for various applications, especially when considering benzene's crucial role in industry and biology.

Furthermore, the model's compatibility with Tinker9, paired with the efficient utilization of GPUs for simulation, ensures that researchers and professionals can benefit from swift, high-resolution simulations, paving the way for cutting-edge research and applications. Additionally, the introduction of a parametrization tool augments the versatility of the HIPPO model. This tool facilitates the parameterization of diverse molecular models, leveraging the foundational attributes of both benzene and water.

In sum, the HIPPO model stands as a pinnacle of force field development, marrying unparalleled accuracy with functional adaptability. Its capabilities herald a new era of molecular simulations, offering promise for myriad domains, from petrochemical engineering to biomedical research.

3.5. References

1. Tripp EH. The Discovery of Benzene. *Nature*, 115, 909 (1925).

2. Pauling L. The Shared-Electron Chemical Bond. *Proc. Natl. Acad. Sci. USA*, *14*, 359-62 (1928).
3. Hückel E. The Theory of Unsaturated and Aromatic Compounds. *Z. Elektrochem. Angew. Physik. Chem.*, *42*, 752 and 827 (1937).
4. Dougherty DA. The Cation-Pi Interaction. *Acc. Chem. Res.*, *46*, 885-93 (2013).
5. Zheng X, Wu C, Ponder JW, Marshall GR. Molecular Dynamics of beta-Haripin Models of Epigenetic Recognition Motifs. *J. Am. Chem. Soc.*, *134*, 15970-8 (2012).
6. Eastmond DA, Schuler M, Frantz C, Chen H, Parks R, Wang L, Hasegawa L. Characterization and Mechanisms of Chromosomal Alterations Induced by Benzene in Mice and Humans. *Res. Rep. Health Eff. Inst.*, *103*, 1-68 (2001).
7. Rushton L, Schnatter AR, Tang G, Glass DC. Acute Myeloid and Chronic Lymphoid Leukaemias and Exposure of Low-Level Benzene Among Petroleum Workers. *British Journal of Cancer*, *110*, 783-7 (2014).
8. Jung K, Khan A, Mocharnuk R, Olivo-Marston S, McDaniel JT. Clinical Encounter with Three Cancer Patients Affected by Groundwater Contamination at Camp Lejeune: A Case Series and Review of the Literature. *Journal of Medical Case Reports*, *16*, 272 (2022).
9. Sunner J, Nishizawa K, Kebarle P. Ion-Solvent Molecule Interactions in the Gas Phase. *J. Phys. Chem.*, *85*, 1814-20 (1981).
10. Headen TF. Temperature dependent structural changes in liquid benzene studied using neutron diffraction. *Mol. Phys.*, *117*, 3329-36 (2019).
11. Tran F, Weber J, Wesolowski TA. Theoretical Study of the Benzene Dimer by the Density-Functional-Theory Formalism Based on Electron-Density Partitioning. *Helvetica Chimica Acta*, *84*, 1489-503 (2001).
12. Sinnokrot MO, Sherrill CD. Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer: Sandwich, T-Shaped, and Parallel-Displaced Configurations. *J. Phys. Chem. A*, *108*, 10200-7 (2004).
13. Hobza P, Seizle HL, Schlag EW. Potential Energy Surface for the Benzene Dimer. Results of ab Initio CCSD(T) Calculations Show Two Nearly Isoenergetic Structures: T-Shaped and Parallel-Displaced. *J. Phys. Chem.*, *100*, 18790-4 (1996).
14. Sinnokrot MO, Valeev EF, Sherrill CD. Estimates of the Ab Initio Limit of Pi-Pi Interactions The Benzene Dimer. *J. Am. Chem. Soc.*, *124*, 10887-93 (2002).
15. Sherrill CD, Takatani T, Hohenstein EG. An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene–Methane, and Benzene–H₂S. *The Journal of Physical Chemistry A*, *113*, 10146-59 (2009).
16. Pitonak M, Neogrady P, Rezac J, Jurecka P, Urban M, Hobza P. Benzene Dimer: High-Level Wave Function and Density Functional Theory Calculations. *J. Chem. Theory Comput.*, *4*, 1829-34 (2008).
17. Borca CH, Glick ZL, Metcalf DP, Burns LA, Sherrill CD. Benchmark Coupled-Cluster Lattice Energy of Crystalline Benzene and Assessment of Multi-Level Approximations in the Many-Body Expansion. *J. Chem. Phys.*, *158*, 234102 (2023).
18. Fu C-F, Tian SX. A Comparative Study for Molecular Dynamics Simulations of Liquid Benzene. *J. Chem. Theory Comput.*, *7*, 2240-52 (2011).
19. Patmanidis I, Alessandri R, de Vries AH, Marrink SJ. Comparing Dimerization Free Energies and Binding Modes of Small Molecules with Different Force Fields. *Molecules*, *26*, 6069 (2021).

20. Jorgensen WL, Severance DL. Aromatic-Aromatic Interactions: Free Energy Profiles for the Benzene Dimer in Water, Chloroform, and Liquid Benzene. *J. Am. Chem. Soc.*, *112*, 4768-74 (1990).
21. Jorgensen WL, Tirado-Rives J. Potential energy Functions for Atomic-Level simulations of Water and Organic and Biomolecular Systems. *Proc. Natl. Acad. Sci. USA*, *102*, 6665-70 (2005).
22. Chessari G, Hunter CA, Low CMR, Packer MJ, Vinter JG, Zonta C. An Evaluation of Force-Field Treatments of Aromatic Interactions. *Chemistry*, *8*, 2860-7 (2002).
23. Tafipolsky M, B. E. Accurate Intermolecular Potential with Physically Grounded Electrostatics. *J. Chem. Theory Comput.*, *7*, (2011).
24. Rackers JA, Wang Q, Liu C, Piquemal J-P, Ren P, Ponder JW. An Optimized Charge Penetration Model for Use with the AMOEBA Force Field. *Phys. Chem. Chem. Phys.*, *19*, 276-91 (2017).
25. Lopes PEM, Lamoureux G, Roux B, MacKerell JAD. Polarizable Empirical Force Field for Aromatic Compounds Based on the Classical Drude Oscillator. *J. Phys. Chem. B*, *111*, 2873-85 (2007).
26. Zhang C, Bell D, Harger M, Ren P. Polarizable Multipole-Based Force Field for Aromatic Molecules and Nucleobases. *J. Chem. Theory Comput.*, *13*, 666-78 (2017).
27. Rackers JA, Silva RR, Wang Z, Ponder JW. Polarizable Water Potential Derived from a Model Electron Density. *J. Chem. Theory Comput.*, *17*, 7056-84 (2021).
28. Ren P, Ponder JW. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *The Journal of Physical Chemistry B*, *107*, 5933-47 (2003).
29. Yang X, Liu C, Walker BD, Ren P. Accurate description of molecular dipole surface with charge flux implemented for molecular mechanics. *The Journal of Chemical Physics*, *153*, 064103 (2020).
30. Wang Q, Rackers JA, He C, Qi R, Narth C, Lagardere L, Gresh N, Ponder JW, Piquemal J-P, Ren P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.*, *11*, 2609-18 (2015).
31. Kříž K, Nováček M, Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets 3: Repulsive Contacts. *J. Chem. Theory Comput.*, *17*, 1548-61 (2021).
32. Sinnokrot MO, Sherrill CD. High-Accuracy Quantum Mechanical Studies of π - π Interactions in Benzene Dimers. *The Journal of Physical Chemistry A*, *110*, 10656-68 (2006).
33. Donchev AG, Taube AG, Decolvenaere E, Hargus C, McGibbon RT, Law K-H, Gregersen BA, Li J-L, Palmo K, Siva K, Bergdorf M, Klepeis JL, Shaw DE. Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Scientific Data*, *8*, (2021).
34. Rackers JA, Wang Z, Lu C, Laury ML, Lagardère L, Schnieders MJ, Piquemal J-P, Ren P, Ponder JW. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.*, *14*, 5273-89 (2018).
35. Wang Z. Tinker9: Next Generation of Tinker with GPU Support. St. Louis: Washington University in St. Louis; 2021 [updated 2021; cited]; Available from: <https://github.com/TinkerTools/tinker9>.
36. Feller SE, Zhang YH, Pastor RW, Brooks BR. CONSTANT-PRESSURE MOLECULAR-DYNAMICS SIMULATION - THE LANGEVIN PISTON METHOD. *J. Chem. Phys.*, *103*, 4613-21 (1995).
37. Quigley D, Probert MIJ. Langevin dynamics in constant pressure extended systems. *J. Chem. Phys.*, *120*, 11432-41 (2004).

38. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126, 014101 (2007).
39. Zwanzig R. Time-Correlation Functions and Transport Coefficients in Statistical Mechanics. *Annual Review of Physical Chemistry*, 16, 67-102 (1965).
40. Zhang Y, Otani A, Maginn EJ. Reliable Viscosity Calculation from Equilibrium Molecular Dynamics Simulations: A Time Decomposition Method. *J. Chem. Theory Comput.*, 11, 3537-46 (2015).
41. Reddy SK, Straight SC, Bajaj P, Huy Pham C, Riera M, Moberg DR, Morales MA, Knight C, Gotz AW, Paesani F. On the Accuracy of the MB-Pol Many-Body Potential for Water: Interaction Energies, Vibrational Frequencies, and Classical Thermodynamic and Dynamical Properties from Clusters to Liquid Water and Ice. *J. Chem. Phys.*, 145, 194504 (2016).
42. Yeh I-C, Hummer G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B*, 108, 15873-9 (2004).
43. Shi B, Sinha S, Dhir VK. Molecular dynamics simulation of the density and surface tension of water by particle-particle particle-mesh method. *J. Chem. Phys.*, 124, (2006).
44. Applequist J, Carl JR, Fung K-K. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.*, 94, 2952-60 (1972).
45. Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets 2: Hydrogen Bonding in an Extended Chemical Space. *J. Chem. Theory Comput.*, 16, 6305-16 (2020).
46. Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.*, 16, 2355-68 (2020).
47. Lemmon EW, Bell IH, Huber ML, McLinden MO. Thermophysical Properties of Fluid Systems. In: Linstrom PJ, Mallard WG, editors. NIST Chemistry WebBook, NIST Standard Reference Database Number 69. Gaithersburg MD, 20899: National Institute of Standards and Technology.
48. Majer V, Svoboda V. Enthalpies of Vaporization of Organic Compounds: A Critical Review and Data Compilation. Oxford: Blackwell Scientific Publications; 1985.
49. Permittivity (Dielectric Constant) Of Liquids. In: Lide DR, editor. CRC Handbook of Chemistry and Physics. Boca Raton: CRC Press; 2005.
50. Falcone DR, Douglass DC, McCall DW. Self-diffusion in benzene. *The Journal of Physical Chemistry*, 71, 2754-5 (1967).
51. Surface Tension of Common Liquids. In: Lide DR, editor. CRC Handbook of Chemistry and Physics. Boca Raton: CRC Press; 2005.

Chapter 4. Computational tools for force field parametrization

A precise and robust force field is paramount for achieving molecular simulations of the highest fidelity. The pursuit of the optimal force field is a meticulous endeavor, one that necessitates a careful balance between experimental findings and quantum mechanical underpinnings. In this chapter, I present a computational tool I've crafted to optimize the parameterization process for the HIPPO force field, leveraging the capabilities of both Tinker and Tinker9 software. Given HIPPO's semi-empirical nature as a classical model, it's imperative to meticulously calibrate its parameters for each molecule. This ensures that it accurately replicates both quantum mechanics (QM) data and experimental observations. The construction of the method began with a selection of molecules to fit and creation of a reference dataset.

4.1. Building the target dataset for parametrization

To effectively parametrize the HIPPO force field, one needs a thorough set of target data for each molecule. My methodology began with the collation of dimer interaction energy from multiple QM databases. This effort is crucial, primarily because the HIPPO force field aspires to serve as a classical analog to the SAPT energy decomposition method.

The foundational set of molecules was sourced from the S101x7 database¹. Subsequently, the research by Coleman et al². in 2012, which meticulously cataloged a database for organic liquids, became instrumental. Their database, encompassing 146 molecules, not only provided experimental data but also critically assessed more rudimentary classical force fields. This evaluation took into account their proficiency in mimicking condensed phase attributes of organic liquids. Properties such as density, enthalpy of vaporization, heat capacities, surface tension,

isothermal compressibility, volumetric expansion coefficient, and dielectric constant were evaluated. Every single molecule from this database was assimilated into my study.

Upon combining the organic liquid database by Caleman with the S101x7, the count stood at 161 molecules. To this, I integrated data for an additional four molecules that had relevant liquid experimental data, bringing the total to 165 molecules.

HIPPO's ambition isn't limited to reflecting experimental data. It's also tailored to emulate QM calculations. While the S101x7 database had SAPT2+³ calculations for a subset of 30 out of the 165 molecules, I took it upon myself to compute SAPT2+ calculations of water-molecule dimers for the entire set. Each molecule underwent SAPT2+ calculations across five randomly selected dimer conformations. The Psi4 software version 1.6 was used for all calculations.^{4,5}

The pursuit for greater accuracy didn't stop there. My exploration extended to scouring published QM databases to amass even more reference data for fitting and validation purposes. In this endeavor, homodimer CCSD(T) QM calculations from the DES370k⁶ and NCI atlas databases⁷⁻¹⁰ became invaluable additions. Whenever water dimer configurations were accessible within these databases, they too were assimilated into the target reference dataset.

This rigorous approach culminated in a uniquely curated dataset—incorporating 165 molecules poised for HIPPO parametrization—sourced from three distinct QM databases and supplemented with experimental liquid data. I've termed this as 'data set 1', and a comprehensive listing can be found in Appendix D.

4.1.1. Structure of the Data Set

The database is structured to facilitate the parameterization process and is hosted on GitHub (github.com/roseane-reis/analyzetestool). The main components are detailed below:

1. **Base Directory (basedir):**

- This is the root directory housing all data pertinent to the fitting.

2. **Data Directory (datadir):**

- Essential for the fitting procedure, this directory contains all data used during the process. The scripts reference this as **reference-data**.
- Before executing the scripts, the major path needs to be established.
- Inside this directory, five subdirectories provide the necessary data for the parameterization script:

i. **database-info**: Contains database metadata in a Python-specific format (pickle).

ii. **boxes**: This folder contains **liquid.xyz**, a pre-set periodic cubic box of the molecule at an appropriate density for simulation temperature. It's in Tinker XYZ format. Additional files like **monomer.xyz** and Tinker key files are also present, as this location was used for creating the liquid box.

iii. **elec-pot**: Essential for setting the HIPPO-specific parameters. The software **Poltype 2** is employed to obtain bonded valence parameters and the DMA (Distributed Multipole Analysis) from quantum calculations. Two key files are housed here: **monomer.pot** (reference electrostatic potential energy) and **monomer.xyz** (Tinker XYZ file of the monomer post-initial parameter acquisition).

iv. **mol-polarize**: - Houses data on pre-computed or experimental molecular polarizability. All values are derived from quantum calculations. The polarizability components [ax, ay, az] are given, and the units are imperative, being required in inverse angstrom cubed for the fitting process.

v. **prmfiles**: Contains initial Tinker parameter files per molecule. Each molecule is referenced by a sequential number.

vi. **qm-calc**: This is the repository of all reference quantum calculations. The data is structured into three categories: **ccsdt_dimers**, **sapt_dimers**, and **clusters**. Each data type has specific file requirements and formats.

3. **Contents of Pickle Files in database-info:**

- These are dictionaries saved in Python's binary format.
- **database_full**: A list where each molecule is identified by an assigned number. The structure is as follows:

- ID: [name, molecular formula, chemspider CSID, pubchem ID (CID)]
- **molinfo**: This segment houses experimental data on liquid properties, whenever available. The lion's share of data originated from the Virtual Chemistry database (<https://virtualchemistry.org/>). However, additional data points were appended over time from other sources. The structure of this dictionary is delineated as:
 - **ID**: [Temperature, Density, Enthalpy of Vaporization, Dielectric Constant, Isothermal Compressibility Coefficient, Thermal Expansion Coefficient, Surface Tension]
 - **Units**:
 - Temperature: Kelvin (K)
 - Density: kilograms per cubic meter (kg/m³)
 - Enthalpy of Vaporization: kilojoules per mole (kJ/mol)
 - Dielectric Constant: (Unitless)
 - Isothermal Compressibility Coefficient: inverse Gigapascals (1/GPa)
 - Thermal Expansion Coefficient: inverse Kelvin scaled by a factor of 10⁻³ (10⁻³ 1/K)
 - Surface Tension: milliNewtons per meter (mN/m)
 - It's important to note that in instances where specific data points are unavailable, they are indicated with a placeholder value of **-1** in the list.
 - (Note: CIDs are mandatory, whereas CSIDs are optional.)
- **molIDs**: provides a systematized directory of molecules chosen specifically for the fitting procedure. The selection criterion was meticulously framed around molecules possessing quantum computations either with water or as homodimers. Here's a detailed breakdown:
 - Virtual Chemistry Database Inclusions: The initial range, from IDs 1 to 146, encompasses all molecules sourced from the Virtual Chemistry database.
 - S101x7 Database and Experimental Data: Molecules from IDs 147 to 161 are extracted from the S101x7 database. Their experimental liquid data is

primarily curated from reputable resources, such as the National Institute of Standards and Technology (NIST) website (<https://webbook.nist.gov/>) and the Chemistry and Physics Handbook (92nd edition).

- Additional Database Selections: molecules numbered 162 through 165 were handpicked because of their importance, they are: carbon tetrachloride, aniline, methane and ammonia.
- Molecules numbered from 166 to 512 are derived from the DES370k and NCIA databases not included in the initial selection. Their association with experimental condensed phase data varies, with some having this data readily available while others don't. From ID 166 onward, molecules were ordered based on their ascending CID (Compound Identification Number).
- Exclusions: In our pursuit of a refined database, molecules that contain elements such as Selenium, Arsenic, and Boron were purposefully omitted from this inaugural set.
- Total Count: altogether, molIDs boasts an impressive count of 504 distinct molecules.
- **initial_fitting**: represents the molecules chosen for the fitting procedure.
 - Selection criteria: molecules possessing either dimers with water or homodimers. If the necessary SAPT data was absent, five structures were selectively sourced from a water+molecule dimer's dissociation curve present in one of the databases. Following this, SAPT2+ calculations were performed.
 - Total Count: The **initial_fitting** dataset encompasses 473 molecules.
 - Folders **qm-calc**, **prmfiles**, and **boxes** contain files related to this set only.
 - Notable Exclusions: 31 molecules were excluded from the **molIDs** set. The reason being the absence of quantum calculations either for their homodimers and water dimers. Interestingly, these excluded molecules possess quantum calculations for heterodimers that align with molecules from the initial set. They will subsequently be parameterized once the primary set's parameters are finalized and validated.

- **all_dimers_map:** A comprehensive mapping of dimers according to their molecule, pinpointed by their CID.
 - Structure: Each entry begins with another molecule's CID, succeeded by a description of the data's hosting database, a unique ID specific to that database, and the total number of conformations. Example: {CID1: CID2: DESRES_SpecificID_ConformationCount}
- **CIDmolinfo:** A meticulously crafted dictionary holding molecule details, indexed by their CID, pulled from the Pubchem webservice, for every molecule in the dataset. This dictionary proves beneficial when attempting to retrieve specific names or associated data for the molecules.

Data Source:

- The molecules were selected from various databases including the Virtual Chemistry database, the S101x7 database, the NCIA databases (R739x5, HB300SPXx10, HB375x10, and SH250x10), and the DES370k database.
- The primary selection criterion was to choose molecules with available homodimers or heterodimers with water. For molecules from the Virtual Chemistry database lacking quantum data, a SAPT2+ calculation was executed with five randomly generated water+molecule dimers. Data from all these databases were subsequently amalgamated, and reference quantum data was collated from each available source.

Interactive Exploration:

- For a more hands-on and immersive experience, all this data is readily available for interactive exploration via the Jupyter notebook, aptly titled **database_description.ipynb**, hosted on GitHub.

4.2. PolFit: a python tool for parametrization of the HIPPO force field

While developing optimization software for the parametrization of HIPPO, I created several tools to enhance the utilization of the Tinker and Tinker9 software. This toolset primarily facilitates running and analyzing molecular simulations, as well as computing interaction energies for small clusters via the Tinker software, all through a user-friendly Python interface. Additionally, these tools aid in aligning simulation results with experimental and quantum data, offering both statistical analysis and visualization capabilities.

4.2.1. An example of fitting a molecule using this tool

In the course of the fitting process, I developed a script named 'runfit.py'. This script interfaces with my more intricate PolFit program to execute a comprehensive job of fitting a molecule. Its primary objective is to fine-tune and optimize parameters in molecular simulations to align closely with specified energy components derived from reference data. Below is an overview of its core functionality, following the pathway of the parametrization.

The relevant parts of the script are shown below, followed by an explanation of each part.

```
import numpy as np
import os
import sys
import copy
import pickle
from analyzetestool import auxfitting, prmedit
from analyzetestool.process import save_pickle, load_pickle

def save_termfit(termfit, path, n):
    """Drops a file in the directory of fitting with the name of the energy terms
    for which parameters are currently being fitted
    """
    listterm = " ".join(termfit)
    listterm += '\n'
    with open(f"{path}/{n}/termfit.txt", 'w') as file:
        file.write(listterm)

    print(f"Running {listterm}")
```

```

sys.stdout.flush()
return listterm

# List of energy components for which parameters can be fitted.
termfit = ['chgpen', 'dispersion', 'repulsion',
           'polarize', 'chgtrn', 'multipole']

#### UPDATE THESE PATHS IF NEEDED
# ref_data: directory with all the files and target data used for fitting.
ref_data = "/work/roseane/HIPPO/small_molecules/org_molecules/reference-data"

molinfo = load_pickle(f"{ref_data}/database-info/molinfo_dict.pickle")

def runfit(path, n, elfn, cudad):

    os.chdir(path)
    fitpath = f"{path}/{n}"
    molfit = auxfitting.Auxfit(path, n)

    molfit.datadir = ref_data
    molfit.prepare_directories() # Creates directories for carrying out the fit

    molfit.process_prm()          # Reads in the parameter file in the reference folder
    molfit.build_prm_list()       # Create a list of parameters to fit based on termfit
    molfit.make_key()             # Make Tinker software key file, based on given params or
                                ## the ones in the starter parameter file

    molfit.initpotrms = molfit.get_potfit() ## Run a potential fit analysis of the current
parameters
                                ## In key file

    ## PRESERVE SOME PARAMETERS FROM ORIGINAL PRMFILE,
    ## Those parameters are never parametrized with this script
    preserve_terms = ['opbend', 'strbnd', 'torsion', 'bndcflux', 'angcflux']
    files = next(os.walk(fitpath))[2]
    dictfn = np.array([f for f in files if 'newprms' in f])
    modtim = [os.path.getmtime(f"{fitpath}/{f}") for f in dictfn]
    modtim = np.array(modtim)
    inds = np.argsort(modtim)
    if len(dictfn) > 0:
        ndict = f"{fitpath}/{dictfn[inds[-1]]}"

        newdict = load_pickle(ndict)
        print(f>Loading previous prmdict: {dictfn[inds[-1]]}\n")

        for term in preserve_terms:
            newdict[term] = copy.deepcopy(molfit.prmdict[term])

        molfit.prmdict = copy.deepcopy(newdict)

    ## molfit.prmdict is a dictionary of the current parameters.

```

```

## The initial one as made from the reference parameter file
## It gets updated when parameters are changed

testliq = True                ## Keyword to test if a simulation run with the current
parameters                    parameters
molfit.nsteps_test = 10000    ## number of steps in the test run
molfit.fithv = False          ## when running the test, you can fit enthalpy of
vaporization                  vaporization
fitliq = False                ## Keyword to perform a full fitting, including condensed
phased                        phased

## List in variable termfit will pass the energy components to run a fit
## It will allow creation of a parameter list to fit for that specifically
## energy potential

## fit_data() is the function that actually runs an optimization of parameters
## It can use two optimization algorithms: genetic, differential evolution
## Or least-squares (lstsq), both from scipy.
## the arguments passed are (optimizer(genetic,lstsq),fitliq,testliq,diff_step,wide_range)
## wide_range argument sets the rules to make the upper and lower bounds on parameters
## if wide_range=True, it uses the full allowed interval for the parameter
## if wide_range=False, it creates bounds within 10% of parameter initial value,
## for parameter < 5, or 20% for parameters with larger numbers
## diff_step only works with lstsq

## FIT MOLECULAR POLARIZABILITY FIRST
termfit = ['polarize']
listterm = " ".join(termfit)
with open(f"{path}/{n}/termfit.txt",'w') as file:
    file.write(listterm)
molfit.build_prm_list(termfit)

res = molfit.fit_data('lstsq', fitliq,testliq, 0.05, False)
molfit.prmdict = molfit.prmlist_to_dict(res.x)
# Save dictionary of newprms
save_pickle(molfit.prmdict,f"{path}/{n}/newprms.pickle")

print(f"Completed mol. pol. fitting \n")
sys.stdout.flush()

## After finishing parametrizing molecular polarizability,
## change molfit.rungas to allow running gas simulation
## to better fit the enthalpy of vaporization, set with the
## variable molfit.fithv
testliq = True
fitliq = False
molfit.rungas = True
molfit.nsteps_gas = 500000
molfit.fithv = True
os.system(f"touch {path}/{n}/FIT_RUNNING")

```

```

## The next block looks over the reference directory to see
## what types of quantum calculation data is available to fit
## the force field energy to that SAPT component or total energy
if os.path.isfile(f"{ref_data}/qm-calc/{n}/sapt-res-water+mol.npy"):
    molfit.prepare_opt_sapt_dimers()
if os.path.isdir(f"{ref_data}/qm-calc/{n}/sapt_dimers"):
    molfit.prepare_sapt_dimers()
if os.path.isdir(f"{ref_data}/qm-calc/{n}/clusters"):
    molfit.prepare_cluster()
if os.path.isdir(f"{ref_data}/qm-calc/{n}/ccsdt_dimers"):
    molfit.prepare_ccsdt_dimers()

#####
## Function calls to fit energy terms are done bellow
## The optimizer can take one or many energy terms at a time.

## Induction fitting through fit of charge transfer parameters
termfit = ['chgtrn']
listterm = save_termfit(termfit,path,n)
molfit.build_prm_list(termfit)
res = molfit.fit_data('lstsq', fitliq,testliq,0.05,False)

molfit.prmdict = molfit.prmlist_to_dict(res.x)
# Save dictionary of newprms
save_pickle(molfit.prmdict,f"{path}/{n}/newprms8.pickle")

print(f"Completed {listterm}")
sys.stdout.flush()

#####
## Electrostatics fitting through fit of charge penetration parameters
termfit = ['chgpen']
listterm = save_termfit(termfit,path,n)
molfit.build_prm_list(termfit)
res = molfit.fit_data('lstsq', fitliq,testliq,0.05,False)

molfit.prmdict = molfit.prmlist_to_dict(res.x)
# Save dictionary of newprms
save_pickle(molfit.prmdict,f"{path}/{n}/newprms8.pickle")

print(f"Completed {listterm}")
sys.stdout.flush()

molfit.nsteps_test = 25000
molfit.nsteps_gas = 250000
molfit.rungas = True
molfit.fithv = True

## Repulsion fitting through fit of repulsion parameters
termfit = ['repulsion']
listterm = save_termfit(termfit,path,n)
molfit.build_prm_list(termfit)

```

```

res = molfit.fit_data('lstsq', fitliq, testliq, 0.05, False)

molfit.prmdict = molfit.prmlist_to_dict(res.x)
# Save dictionary of newprms
save_pickle(molfit.prmdict, f"{path}/{n}/newprms8.pickle")

print(f"Completed {listterm}")
sys.stdout.flush()

## Dispersion fitting through fit of dispersion parameters
termfit = ['dispersion']
listterm = save_termfit(termfit, path, n)
molfit.build_prm_list(termfit)
res = molfit.fit_data('lstsq', fitliq, testliq)

molfit.prmdict = molfit.prmlist_to_dict(res.x)
# Save dictionary of newprms
save_pickle(molfit.prmdict, f"{path}/{n}/newprms8.pickle")

print(f"Completed {listterm}")
sys.stdout.flush()

## The next blocks fit more than one energy term at a time
#Induction
termfit = ['repulsion', 'chgtrn']
listterm = save_termfit(termfit, path, n)
molfit.build_prm_list(termfit)
res = molfit.fit_data('lstsq', fitliq, testliq, 0.05, False)

molfit.prmdict = molfit.prmlist_to_dict(res.x)
# Save dictionary of newprms
save_pickle(molfit.prmdict, f"{path}/{n}/newprms9.pickle")

print(f"Completed {listterm}")
sys.stdout.flush()

termfit = ['dispersion', 'repulsion', 'chgtrn']
listterm = save_termfit(termfit, path, n)
molfit.build_prm_list(termfit)
res = molfit.fit_data('lstsq', fitliq, testliq, 0.05, False)
molfit.prmdict = molfit.prmlist_to_dict(res.x)
# Save dictionary of newprms
save_pickle(molfit.prmdict, f"{path}/{n}/newprms9.pickle")

print(f"Completed {listterm}")
sys.stdout.flush()

## The next blocks will turn on fitliq to allow the fit to use
## experimental liquid data and run simulation for fitting of the
## parameters to reproduce experimental properties
testliq = False
fitliq = True
molfit.fithv = False

```

```

molfit.gasdcd = f"{path}/{n}/ref_liquid/gas.dcd"
## molinfo has the experimental information per molecule
info = molinfo[n]

## Use a continuation file .dyn in the liquid fitting
## this allows that the box of the simulation is pre-equilibrated
if os.path.isfile(f"{path}/{n}/ref_liquid/liquid.dcd") and not
os.path.isfile(f"{path}/{n}/ref_liquid/liquid.err"):
    molfit.useliqdyn = True

## Change some of the references for two specific cases, not an
## essential part of the fitting, but allows to change the reference
## case by case
if n == 6:      # Methanoic acid
    molfit.liquidref[0][2] = 2*molfit.liquidref[0][2] + 6.76792
if n == 147:   # Acetic acid
    molfit.liquidref[0][2] = 2*molfit.liquidref[0][2] + 5.53816

## info[0] is the temperature of the simulatin. This code block will
## increase the lenght of the simulation if the temperature is lower than
## a threshold because lower temperature simulations converges slower

if info[0] < 275:
    molfit.nsteps = 1500000
    molfit.equil = 1500
elif info[0] < 250:
    molfit.nsteps = 2000000
    molfit.equil = 2000
else:
    molfit.nsteps = 500000
    molfit.equil = 500
    molfit.useliqdyn = False

molfit.rungas = True
molfit.nsteps_gas = 5000000
molfit.gasdcd = f"{path}/{n}/ref_liquid/gas.dcd"

## Use dispersion, repulsion and chgtrn parameters to fit to
## experimental data targets
termfit = ['dispersion', 'repulsion', 'chgtrn']
listterm = save_termfit(termfit, path, n)
molfit.build_prm_list(termfit)
res = molfit.fit_data('lstsq', fitliq, testliq)
molfit.prmdict = molfit.prmlist_to_dict(res.x)
# Save dictionary of newprms
save_pickle(molfit.prmdict, f"{path}/{n}/newprms10.pickle")

print(f"Completed {listterm}")
sys.stdout.flush()

## Save results
if not os.path.isdir(f"{path}/{n}/fit_results"):

```

```
os.system(f"mkdir -p {path}/{n}/fit_results")

resdir = f"{path}/{n}/fit_results"
fname = f"{resdir}/{n}-latest.prm"
```

Initialization: It sets up lists and paths for the terms of energy that will be optimized.

A list `termfit` contains the energy components for which parameters can be optimized.

The `ref_data` path points to the reference data directory.

Preparing the Fitting Environment:

Given a path and a molecule index `n`, the function `runfit` sets up the fitting environment.

It loads molecule-related information from pickled data.

Preparations include creating directories, processing parameters, and checking if there are any previous fitting results to use as a starting point.

Fitting Steps:

The script then begins the fitting process. Each fitting step optimizes the parameters of one or more energy components.

The optimization is done using a function `fit_data` which can employ different optimization algorithms.

For each step:

- Set the energy terms to fit.
- Prepare the fitting environment for those terms.
- Optimize the parameters using reference data.
- Store the optimized parameters in a dictionary and save them.

Different energy terms or components include molecular polarizability, charge transfer, charge penetration, repulsion, and dispersion. They are fitted in various combinations as well.

Fitting with Experimental Data:

After the initial fitting steps using the reference data, the script can perform fitting using experimental liquid data. The number of simulation steps and other parameters may be adjusted depending on the temperature of the experimental data.

The results from this step can be closer to real-world observations.

Finalizing and Saving Results:

The optimized parameters from the latest fitting steps are saved in a results directory with a specific naming convention.

The full script can be found online, under my private GitHub page (<https://github.com/roseane-reis/analyzetestool/blob/main/runfit.py>).

It shows the general strategy of the fitting for the HIPPO force field. The highlight steps are below.

1. Fit the molecular polarizability; as shown in the data setup process, every molecule has a reference QM calculation of the molecular polarizability. The first step then is to fit atomic polarizability to reproduce the molecular one. For this step, the quantum data for dimers has not been loaded. It is optional to do a trial bulk phase calculation at every step, and there is a keyword to turn off this feature. However, this option was always on for the HIPPO parametrization. It ensured parameters would not cause polarization catastrophe during bulk phase simulation.
2. Next, fit the charge transfer parameters, followed by charge penetration. For those components, a test of the stability of the bulk phase simulation is performed, by running a short simulation. The data of the simulation is discarded. It is important to notice that only the component energy is used in the fit, not the total. If fitting only charge penetration, only the electrostatic energy is fit to the reference SAPT target. The total energy is only fitted when more than two parameter sets from different energy terms are included.
3. Fit repulsion parameters, followed by dispersion. Before running this fit, I turn on the run of gas phase simulation at every step along with liquid phase simulation. The heat of vaporization is computed and used as a fitting target.

4. Another cycle of fitting for the repulsion and charge transfer terms is performed, this is to further adjust to the right heat of vaporization and to possibly fix the total interaction energy. The same is done by adding dispersion parameters.
5. The last step in the parametrization is to fit to liquid data. The length of the simulation can be adjusted, and there is a variation with the temperature of the simulation, since lower temperature simulations take longer to converge.
6. Results are saved. A strategy is in place to avoid overwriting files.

4.2.2. Analyzing and visualizing

The analyze tool is a script designed specifically for scrutinizing the outcomes of molecular simulations and parametrization processes. It's tailored for those who want to go beyond basic analyses and delve deep into the intricate details of parameters, per molecule. This is also hosted in my personal GitHub page. Below are its core functionalities:

1. Visualization and Inspection:

The tool isn't just limited to those engaged in an ongoing fit cycle. Even without an active fitting process, it enables users to:

- Visualize Parameters: Easily view and understand results for a given parameter file.
- Track a Fitting Cycle: Whether a fit cycle is still in progress or has been halted, the script allows users to stay updated with the latest outcomes.
- This script leverages the capabilities of the previously built package, ensuring users can visualize the results seamlessly.

- Analyzing Without Fitting: Sometimes, all one needs is to inspect the potential outcomes a parameter file might produce without diving into an actual fit cycle.
- Test Parameters: Just point the tool to a directory (testfit) where you want to test the parameters. RefPath: This is the directory for an actual fit cycle, but if you're just testing, you can set 'refpath' identical to 'testfit'.
- Analyzing Progress in Real-Time, for those knee-deep in a fitting project, real-time analysis can be a boon:
 - Accessing Logs: By navigating to the 'dumpdata' inside the fit directory, the tool retrieves the pickle file logging the ongoing fit.
 - Use_dict Option: This option is designed to provide insights into intermediate parameters during the fit. It's particularly useful if the fit has been carried out in stages. For instance, if 'chgpen' finishes before 'repulsion', the tool will store interim parameters distinct from the final ones saved in 'fit_results'.
 - Peek into the Optimizer's Progress: The true magic lies in how you can peek into the optimizer's ongoing work. As an example, if 'repulsion' hasn't concluded its fitting, you won't see its latest parameters until it's done. However, with the log.pickle file, parameters and error arrays are saved at every iteration. This means users can fetch parameters from any iteration, especially the one with the lowest error thus far.
- Post-Fitting Analysis:

Once a fit process concludes, you might want a comprehensive look at the outcomes. Accessing final results is done by simply activating the `fit_prm=True` option, and the tool will retrieve the latest parameter data from 'fit_results'.

- A sample output: below, I am going to show what the fitting produces as a final output for understanding the results.

```
#29 Dimethyl disulfide C2H6S2  CID: 12232

ref molpol  12.64    8.95    9.46, avg  10.35
   molpol  12.60    10.13   9.37, avg  10.70
rms molpol   0.04    1.18    0.09, avg   0.35

Monomer potential fitting RMS: 0.72

##Dimer results - Fitting to QM datasets##

DESRES_29-water, energy values in kcal/mol

      MAE      Std error  max error  #points  #count[err > 1]
0.269      0.545      5.8014     564      31

DESRES_29-29, energy values in kcal/mol

      MAE      Std error  max error  #points  #count[err > 1]
0.247      0.365      3.5495     528      24

Liquid Dimethyl disulfide @ 298.15 K
Density Ref-Dens %err   HV Ref-HV %err Dielec Ref-D %err
1051.01 1057.32 -0.6  39.44 38.32 2.9  7.70  9.60 -19.8

kappa Ref-k %err alphaT Ref-aT %err #nFrm
1.21 -1.00  0.0  1.71  1.10  55.4 32601
```

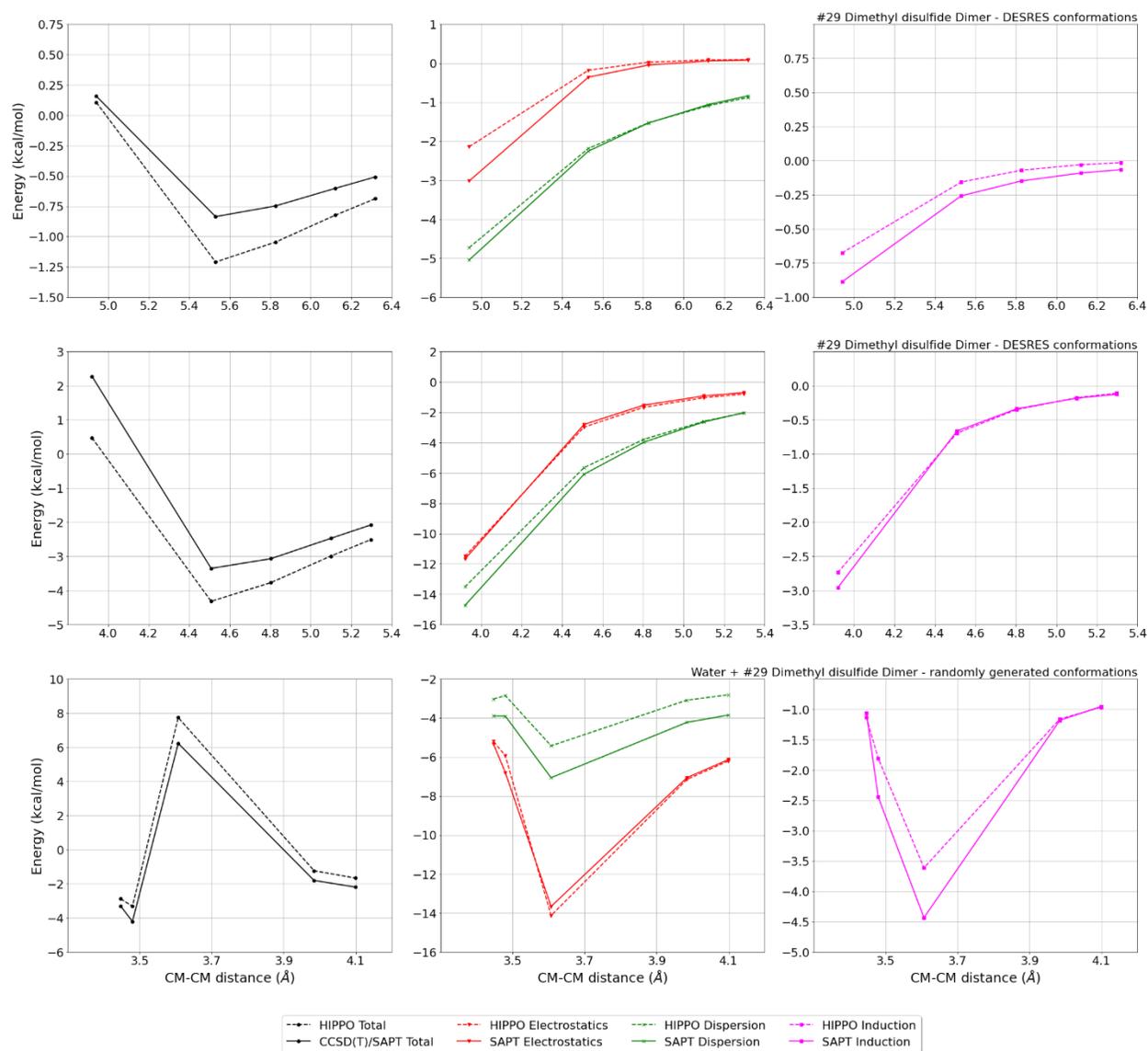


Figure 4.1. Sample output for a finalized parametrization process for analysis and visualization. This figure compares the fit of intermolecular interaction energy for Dimethyl Disulfide homodimers from a variety of databases. The components are electrostatics (red), dispersion (black) and induction (pink). Repulsion term is omitted, but its information is packed into the total energy.

4.3. Future work

The development of this analysis and parametrization tool is crucial for the future viability of HIPPO. This ensures that as new molecules emerge that aren't part of the current database, researchers and experts can still utilize HIPPO efficiently.

A significant avenue for progression lies in seamlessly integrating this tool with other components of the parametrization software, particularly those associated with valence terms. Moreover, enhancing the documentation will not only elevate the user experience but will also make the tool more intuitive and user-friendly.

The subsequent chapter will demonstrate the robustness of this software. It has already been leveraged for the parametrization of several hundred molecules in HIPPO's initial parametrization phase. While there's undeniable room for refining both the code and the accompanying documentation, the current parametrization pipeline has proven its mettle, yielding exemplary model results.

For broader accessibility and to empower users, there's an imminent need for comprehensive and clearer documentation. Furthermore, to bolster its capabilities in statistical analysis, integrating this tool with prevalent software in the domain would be a judicious step forward.

4.4. References

1. Wang Q, Rackers JA, He C, Qi R, Narth C, Lagardere L, Gresh N, Ponder JW, Piquemal J-P, Ren P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.*, *11*, 2609-18 (2015).
2. Caleman C, van Maaren PJ, Hong M, Hub JS, Costa LT, van der Spoel D. Force Field Benchmark for Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.*, *8*, 61-74 (2012).
3. Parker TM, Burns LA, Parrish RM, Ryno AG, Sherrill CD. Levels of Symmetry Adapted Perturbation Theory (SAPT). I. Efficiency and Performance for Interaction Energies. *J. Chem. Phys.*, *140*, 094106 (2014).
4. Smith DGA, Burns LA, Sirianni DA, Nascimento DR, Kumar A, James AM, Schriber JB, Zhang T, Zhang B, Abbott AS. Psi4NumPy: An Interactive Quantum Chemistry Programming Environment for Reference Implementations and Rapid Development. *J. Chem. Theory Comput.*, *14*, 3504-11 (2018).
5. Turney JM, Simmonett AC, Parrish RM, Hohenstein EG, Evangelista FA, Fermann JT, Mintz BJ, Burns LA, Wilke JJ, Abrams ML, Russ NJ, Leininger ML, Janssen CL, Seidl ET, Allen WD, Schaefer HF, King RA, Valeev EF, Sherrill CD, Crawford TD. Psi4: an open-source ab initio electronic structure program. *WIREs Comput. Mol. Sci.*, *2*, 556-65 (2012).
6. Donchev AG, Taube AG, Decolvenaere E, Hargus C, McGibbon RT, Law K-H, Gregersen BA, Li J-L, Palmo K, Siva K, Bergdorf M, Klepeis JL, Shaw DE. Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Scientific Data*, *8*, (2021).
7. Kříž K, Nováček M, Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets 3: Repulsive Contacts. *J. Chem. Theory Comput.*, *17*, 1548-61 (2021).
8. Kříž K, Řezáč J. Non-covalent interactions atlas benchmark data sets 4: σ -hole interactions. *Phys. Chem. Chem. Phys.*, *24*, 14794-804 (2022).
9. Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets 2: Hydrogen Bonding in an Extended Chemical Space. *J. Chem. Theory Comput.*, *16*, 6305-16 (2020).
10. Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.*, *16*, 2355-68 (2020).

Chapter 5. Modeling Organic Molecules with the HIPPO Force

Field

In this chapter, we delve into the practical application of the parametrization tool discussed in Chapter 4. This exercise serves a dual purpose: firstly, to evaluate the robustness of our tool, and secondly, to examine the broader utility and stability of the HIPPO force field when employed across a vast spectrum of molecules. Our findings underline that HIPPO accurately captures the underlying physics of each molecule under scrutiny.

Furthermore, the progression of our parametrization showcases the transferability of our model. We embarked on this journey with smaller molecules and gradually scaled up, ensuring the parameters from previous iterations were carried over to initiate subsequent fitting cycles. Impressively, there was consistent agreement with SAPT calculations, and most results aligned with experimental properties, deviating by a mere 2% from experimental benchmarks.

In further validation, I utilized hydration free energy calculations on a curated set of models. It's heartening to note that HIPPO's forecasts remained unwaveringly within a tight 1 kcal/mol range for every molecule assessed. This endeavor stands as a significant chapter in my journey as a graduate student. It was a path marked by numerous iterations, setbacks, and invaluable lessons. Yet, when I reflect upon the end result – its success and precision – I am filled with a profound sense of accomplishment and pride, recognizing it as the pinnacle of my academic journey.

5.1. Introduction

The age of computational drug discovery is upon us, and with it comes the increasing need for accurate force fields for organic molecules. Small organic molecules lie at the heart of drug

discovery, serving as the building blocks and initial leads for potential therapeutics. The correct simulation and representation of their interactions, geometries, and behaviors in various environments can often decide the fate of a drug candidate in the virtual world long before it undergoes rigorous testing in a laboratory setting.

To simulate the behavior of these molecules, it's imperative to have accurate force fields. A force field describes the forces between atoms in a molecule, allowing researchers to predict how that molecule will behave under various conditions. Getting these force fields right is crucial, not just for small molecules, but also as the building blocks for more complex systems like proteins and nucleic acids. Specifically, the development of these force fields for organic molecules will lay the foundation for a more extensive protein and nucleic acid force field in HIPPO.

This chapter elaborates on the rigorous process of parameterizing 137 distinct organic molecules using our computational tool. Divided into two main datasets, we've ensured that each molecule undergoes validation against quantum mechanics (QM) to guarantee accuracy. The first set contains 36 molecules, and those were further parametrized using experimental liquid data. The second only using *ab initio* results as targets for parametrization.

5.2. Methods

5.2.1. Selection of molecules

*The molecules on **Table 5.1** were selected sequentially from the Data Set 1 (Appendix D, Table D.1) based on size (number of atoms), functional groups presents and amount of reference data available. For the initial parametrization, it was important to have as much QM references as possible so the smallest chemical specie of a given functional group would have an accurate description of its intermolecular potential terms. The molecules on*

Table 5.2 were selected from Data Set 2 (Appendix D, Table D.2) based on availability of initial parameters for a given molecule and again on the amount of Quantum data available. The number of molecules was set based on the availability of computational resources in the lab cluster.

5.2.2. Computational details

Molecular Polarizability. This property was computed in-house using Psi4 software version 1.6.¹ For most molecules, this was computed at using MP2 method with the aug-cc-pVTZ basis set. The calculation of the dipole polarizability was done by a perturbative method, where the dipole was systematically perturbed by a small amount and the energy response was calculated. Then, the polarizability was computed, given induced dipole is proportional to the molecular polarizability times the total electric field of the system. This approach failed in some cases where the symmetry or the lack of a dipole moment in a particular direction caused the calculation to fail. For these cases, the polarizability is calculated with the CCSD method as implemented in Psi4. The later calculation is much more expensive than the former, which means the problem was not solved in a timely manner for this dissertation for a number of molecules. This, however, does not represent a major problem as the fit to the molecular polarizability is an inexpensive calculation, and the deviation from the two methods to compute molecular polarizability are not too significant.

Using the Polfit tool developed and explained in Chapter 4, the optimization of the atomic polarizability parameters was carried out as a first step to the optimization. The initial value was set by the Poltype 2 software², and it was the same as that assigned for the AMOEBA force field.³ A least squares algorithm was applied, as implemented in SciPy⁴ version 1.8. The fit was done against every component of the molecular polarizability eigen vector (output of Psi4). for these properties usually converged within less than a hundred iterations.

Liquid and dimer calculations. Dimer interaction energies were computed using HIPPO through of the ANALYZE suite of the Tinker8 software.⁵ Liquid phase simulations were performed using similar protocols as in Chapters 1 and 3. For every molecule, a 30 Å cubic box was created such that the density matched that of the molecule for the temperature used in fitting. All bulk phase calculations were performed using the Tinker9 software⁶, with the implemented version implemented by me and Dr. Zhi Wang. All the simulations used a Langevin barostat and thermostat and RESPA integrator and the length of the simulation varied per molecule during fitting, dependent on the temperature. Throughout the cycles of fitting the simulation length was also adjusted, as was shown in Chapter 4. At every step, the box was minimized and simulated for at least 50 ps; to fit liquid properties, the minimum duration was 700 ps per iteration. Gas phase simulations used a stochastic Verlet integrator, and those simulations usually ran for 500 ps during fitting.

The results shown in the next session are for the production dynamics, which was a simulated using the same setup as in the fitting stage, but were ran for at least 10 ns. The first 3 ns was discarded as equilibration.

5.3. Results

The next pages and figures are going to show the results after every molecule listed in **Table 5.1** and **Table 5.2** were parameterized. Given the number of interactions per molecule, the majority of the individual results are listed in Appendix E. This chapter will discuss such results and give attention to the global outcome and average results.

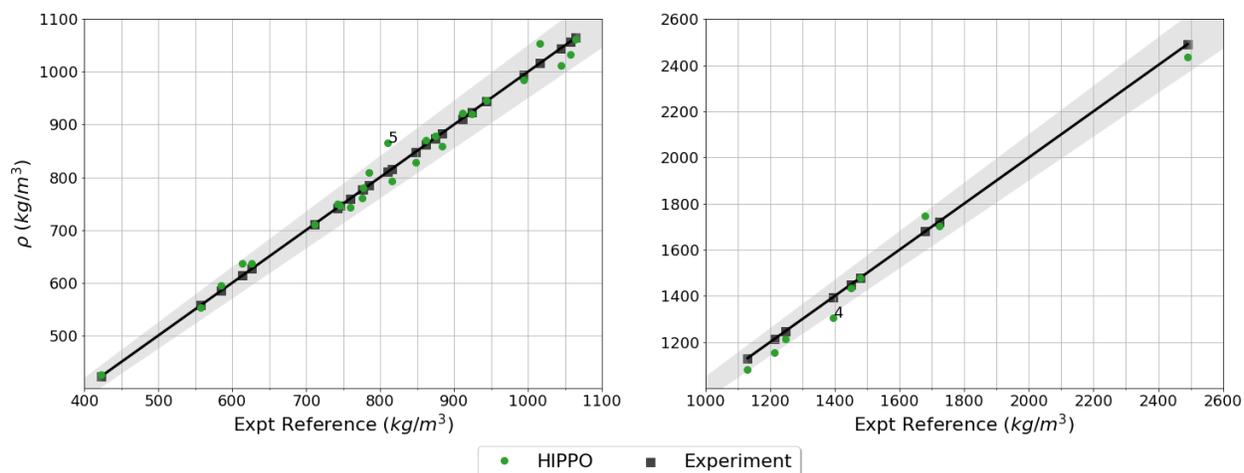


Figure 5.1. Comparison between experimental density and HIPPO results for molecules in data set 1.

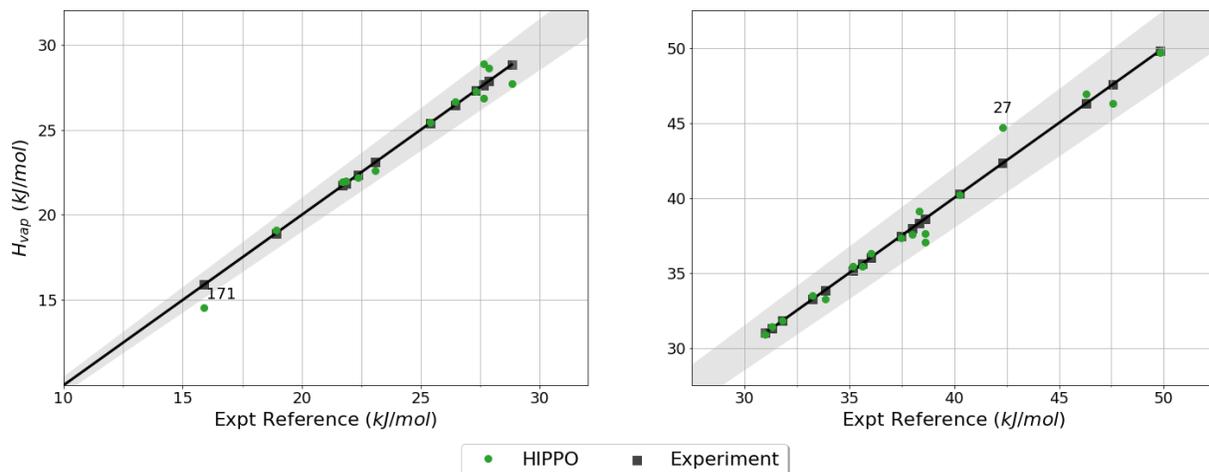


Figure 5.2. Comparison between experimental Enthalpy of Vaporization and HIPPO results for molecules in data set 1.

Table 5.1. 35 molecules from Data set 1 (Appendix D) selected to undergo the first round of parameterization. These molecules had parameters validated with QM and experimental liquid data.

	#ID	CID	Name	Formula	Functional Group
1	1	6212	chloroform	CHCl ₃	chlorine
2	3	3024	dibromomethane	CH ₂ Br ₂	bromine
3	4	6344	dichloromethane	CH ₂ Cl ₂	chlorine
4	5	712	formaldehyde	CH ₂ O	aldehydes
5	7	6323	bromomethane	CH ₃ Br	bromine
6	9	6375	nitromethane	CH ₃ NO ₂	nitros
7	11	6419	1,1,1,2,2-pentachloroethane	C ₂ HCl ₅	chlorine
8	15	6342	methyl cyanide	C ₂ H ₃ N	nitriles
9	18	11	1,2-dichloroethane	C ₂ H ₄ Cl ₂	chlorine
10	20	6332	bromoethane	C ₂ H ₅ Br	bromine
11	21	6337	chloroethane	C ₂ H ₅ Cl	chlorine
12	25	6587	nitroethane	C ₂ H ₅ NO ₂	nitros
13	26	8254	dimethyl ether	C ₂ H ₆ O	ethers
14	27	702	ethanol	C ₂ H ₆ O	alcohols
15	29	12232	dimethyl disulfide	C ₂ H ₆ S ₂	disulfides
16	36	7854	ethyl cyanide	C ₃ H ₅ N	nitriles
17	42	12586	1,3-dioxolane	C ₃ H ₆ O ₂	acetal
18	45	6228	N,N-dimethylformamide	C ₃ H ₇ NO	amides
19	51	7852	propan-1-amine	C ₃ H ₉ N	amines
20	53	6360	isobutane	C ₄ H ₁₀	hydrocarbons
21	64	9260	pyrimidine	C ₄ H ₄ N ₂	pyrimidine; aromatics
22	72	8028	oxolane	C ₄ H ₈ O	ethers; cyclics
23	73	8023	ethoxyethene	C ₄ H ₈ O	ethers
24	76	1127	tetrahydrothiophene	C ₄ H ₈ S	thiols; cyclics
25	120	1140	toluene	C ₇ H ₈	hydrocarbons; benzenes; aromatics
26	148	241	benzene	C ₆ H ₆	benzenes
27	151	9253	cyclopentane	C ₅ H ₁₀	hydrocarbons
28	152	1068	dimethyl sulfide	C ₂ H ₆ S	thiols; sulfides
29	157	6327	methyl chloride	CH ₃ Cl	chlorine
30	158	11638	methyl fluoride	CH ₃ F	fluorine
31	159	6345	methylene fluoride	CH ₂ F ₂	fluorine
32	160	10041	neopentane	C ₅ H ₁₂	hydrocarbons
33	161	8003	pentane	C ₅ H ₁₂	hydrocarbons
34	164	297	methane	CH ₄	hydrocarbons
35	171	402	hydrogen sulfide	H ₂ S	sulfides

Table 5.2. 101 molecules from Data set 2 (Appendix D) that have finalized parameters, after fitting to available *ab initio* references from NCIA⁷⁻¹⁰ and DES370k¹¹ databases.

#	#ID	CID	Name	Formula	Functional Group
1	169	260	bromane	BrH	bromine
2	173	527	propanal	C3H6O	aldehydes
3	178	768	hydrogen cyanide	CHN	nitriles
4	181	878	methanethiol	CH4S	thiols
5	190	1567	2-mercaptoethanol	C2H6OS	thiols
6	194	4685	1,4-dichlorobenzene	C6H4Cl2	chlorine; benzenes; aromatics
7	196	6058	2-aminoethanethiol	C2H7NS	amines; thiols
8	200	6324	ethane	C2H6	hydrocarbons
9	203	6334	propane	C3H8	hydrocarbons
10	204	6335	prop-1-yne	C3H4	hydrocarbons
11	205	6340	iodoethane	C2H5I	iodine
12	207	6343	ethanethiol	C2H6S	sulfides
13	209	6368	1,1-difluoroethane	C2H4F2	fluorine
14	210	6373	fluoroform	CHF3	fluorine
15	214	6403	2,2-dimethylbutane	C6H14	hydrocarbons
16	215	6431	1,1,1,2,2,2-hexafluoroethane	C2F6	fluorine
17	217	6556	2-methylbutane	C5H12	hydrocarbons
18	219	6569	butan-2-one	C4H8O	ketones
19	221	6578	propanamide	C3H7NO	amides
20	223	7239	1,2-dichlorobenzene	C6H4Cl2	chlorine; benzenes
21	224	7288	pentan-3-one	C5H10O	ketones
22	225	7296	methylcyclopentane	C6H12	hydrocarbons; cyclics
23	228	7843	butane	C4H10	hydrocarbons
24	233	7950	1,3,5-trichlorobenzene	C6H3Cl3	chlorine; benzenes
25	236	8008	butanenitrile	C4H7N	nitriles
26	240	8018	2-methoxyethanamine	C3H9NO	ethers; amines
27	242	8025	ethyl formate	C3H6O2	esters
28	244	8070	N,N'-dimethylethane-1,2-diamine	C4H12N2	amines
29	245	8071	1,2-dimethoxyethane	C4H10O2	ethers
30	248	8078	cyclohexane	C6H12	hydrocarbons; cyclics
31	251	8252	prop-1-ene	C3H6	hydrocarbons
32	255	8894	tetrahydropyran	C5H10O	ethers; cyclics
33	256	9086	3-aminopropan-1-ol	C3H9NO	alcohols; amines
34	257	9261	pyrazine	C4H4N2	aromatics
35	259	9620	fluoroethane	C2H5F	fluorine
36	260	9633	1,1,1,2,2-pentafluoroethane	C2HF5	fluorine
37	261	9696	1,2,3,4,5-pentafluorobenzene	C6HF5	fluorine; benzenes
38	262	9745	1,3,5-trifluorobenzene	C6H3F3	fluorine; benzenes
39	263	9805	1,2,3,4,5,6-hexafluorobenzene	C6F6	fluorine; benzenes
40	264	9868	1,1,1-trifluoroethane	C2H3F3	fluorine
41	265	9890	1,1,2-trifluoroethane	C2H3F3	fluorine

42	266	9998	1-fluoropropane	C3H7F	fluorine
43	279	10553	2-methylbut-2-ene	C5H10	hydrocarbons
44	281	10899	1-chloropropane	C3H7Cl	chlorine
45	284	10943	1,3-dichlorobenzene	C6H4Cl2	chlorine; benzenes
46	287	11182	1-methoxypropane	C4H10O	ethers
47	290	11250	2,3-dimethylbut-2-ene	C6H12	hydrocarbons
48	294	11855	1,2,3,4,5-pentachlorobenzene	C6HCl5	chlorine; benzenes
49	298	12051	N,N-diethylformamide	C5H11NO	amides
50	301	12223	1,2-difluoroethane	C2H4F2	fluorine
51	304	12243	2-methylpent-2-ene	C6H12	hydrocarbons
52	305	12251	2-methoxyacetic acid	C3H6O3	carboxylic_acids; ethers
53	306	12253	N-ethylacetamide	C4H9NO	amides
54	307	12309	pent-1-yne	C5H8	hydrocarbons
55	308	12310	pent-2-yne	C5H8	hydrocarbons
56	310	12319	N-ethylformamide	C3H7NO	amides
57	313	12418	1,1,1,2-tetrachloroethane	C2H2Cl4	chlorine
58	315	12703	N,N-diethylacetamide	C6H13NO	amides
59	317	12961	2-sulfanylacetamide	C2H5NOS	amides; thiols
60	319	13129	1,1,1,2-tetrafluoroethane	C2H2F4	fluorine
61	321	13134	dimethyl hydrogen phosphate	C2H7O4P	phosphates
62	322	13195	5-methyl-1H-imidazole	C4H6N2	aromatics
63	328	16592	1-(methylsulfanyl)propane	C4H10S2	disulfides
64	330	16908	1-fluorobutane	C4H9F	fluorine
65	333	22686	N-propylformamide	C4H9NO	amides
66	334	23110	1,2-bis(methylsulfanyl)ethane	C4H10S2	sulfides
67	336	24387	trichlorophosphane	Cl3P	chlorine; phosphines
68	341	24622	1,1,1-trichloropropane	C3H5Cl3	chlorine
69	346	31242	4-ethylphenol	C8H10O	phenols
70	347	31275	1,4-dioxane	C4H8O2	ethers; cyclics
71	350	62540	methoxymethanol	C2H6O2	alcohols; ethers
72	351	66978	N,N'-dimethylpropane-1,3-diamine	C5H14N2	amines
73	353	67899	1,1,1-trifluoropropane	C3H5F3	fluorine
74	354	68152	3-hydroxypropanoic acid	C3H6O3	carboxylic_acids; alcohols
75	357	69020	2-aminoacetamide	C2H6N2O	amides; amines
76	358	69021	2-hydroxyacetamide	C2H5NO2	alcohols; amides
77	360	69657	N-(2-hydroxyethyl)formamide	C3H7NO2	alcohols; amides
78	363	74116	3-methoxypropan-1-ol	C4H10O2	ethers
79	364	75367	3-amino-3-oxopropanoic acid	C3H5NO3	carboxylic_acids; amides
80	365	75551	2-methylsulfanylacetic acid	C3H6O2S	carboxylic_acids; sulfides
81	366	75606	2-formamidoacetic acid	C3H5NO3	carboxylic_acids; amides

82	367	75891	3-(methylamino)propanoic acid	C4H9NO2	carboxylic_acids; amines
83	368	77743	3-methylsulfanylpropan-1-amine	C4H11NS	amines; sulfides
84	369	78925	2-methylsulfanylethanol	C3H8OS	alcohols; sulfides
85	375	87697	2-methylsulfanylethanamine	C3H9NS	amines; sulfides
86	378	94671	disulfanylethane	C2H6S2	thiols; disulfides
87	379	97436	3-aminopropane-1-thiol	C3H9NS	amines; thiols
88	381	122370	methylsulfanylmethanethiol	C2H6S2	thiols; sulfides
89	399	140060	2-methoxyacetamide	C3H7NO2	ethers; amides
90	400	140180	1,3-dimethoxypropane	C5H12O2	ethers
91	402	141892	2-acetamido-N-methylpropanamide	C6H12N2O2	amides
92	406	192802	3-aminopropanamide	C3H8N2O	amides; amines
93	409	226108	N-(2-formamidoethyl)formamide	C4H8N2O2	amides
94	410	232267	2-formamidoacetamide	C3H6N2O2	amides
95	411	263087	3-methylsulfanylpropanamide	C4H9NOS	amides; sulfides
96	412	300977	2-methoxy-N-methylethanamine	C4H11NO	ethers; amines
97	414	350667	N-ethyl-N-methylformamide	C4H9NO	amides
98	419	524894	methoxymethoxyethane	C4H10O2	ethers
99	420	525376	1-(methylsulfanylmethylsulfanyl)propane	C5H12S2	sulfides
100	421	525377	methylsulfanylmethylsulfanylethane	C4H10S2	sulfides
101	427	641811	thioacetone	C3H6S	thiones

Table 5.3. HIPPO density compared to experimental values. Density is presented in kg/m³. Experimental values from references ^{12, 13}.

	#ID	Name	Temp. (K)	Expt.	HIPPO	% Error
1	1	chloroform	298.15	1479.30	1477.95	-0.1
2	3	dibromomethane	298.15	2490.70	2435.11	-2.2
3	4	dichloromethane	298.15	1394.30	1306.43	-6.3
4	5	formaldehyde	253.65	810.53	865.46	6.8
5	7	bromomethane	276.65	1721.95	1702.45	-1.1
6	9	nitromethane	298.15	1130.40	1081.02	-4.4
7	11	1,1,1,2,2-pentachloroethane	293.15	1679.60	1748.66	4.1
8	15	methyl cyanide	298.15	776.00	760.42	-2.0
9	18	1,2-dichloroethane	298.15	1246.30	1215.05	-2.5
10	20	bromoethane	298.15	1449.30	1432.48	-1.2
11	21	chloroethane	273.15	923.90	919.78	-0.4
12	25	nitroethane	298.15	1044.10	1012.28	-3.0
13	26	dimethyl ether	240.00	742.08	748.97	0.9
14	27	ethanol	298.15	784.80	809.29	3.1
15	29	dimethyl disulfide	298.15	1057.32	1032.00	-2.4
16	36	ethyl cyanide	298.15	776.40	780.55	0.5
17	42	1,3-dioxolane	298.15	1064.40	1062.13	-0.2
18	45	N,N-dimethylformamide	298.15	943.30	945.19	0.2
19	51	propan-1-amine	298.15	711.47	710.38	-0.2
20	53	isobutane	243.65	613.53	637.29	3.9
21	64	pyrimidine	298.15	1016.40	1054.13	3.7
22	72	oxolane	298.15	883.70	859.53	-2.7
23	73	ethoxyethene	293.15	758.90	742.97	-2.1
24	76	tetrahydrothiophene	298.15	994.00	984.76	-0.9
25	120	toluene	298.15	861.90	869.85	0.9
26	148	benzene	298.15	874.00	878.09	0.5
27	151	cyclopentane	293.15	745.70	745.89	0.0
28	152	dimethyl sulfide	293.15	848.30	827.53	-2.4
29	157	methyl chloride	298.15	911.00	922.04	1.2
30	158	methyl fluoride	298.15	557.00	553.01	-0.7
31	159	methylene fluoride	221.00	1213.90	1155.52	-4.8
32	160	neopentane	298.15	585.20	595.17	1.7
33	161	pentane	293.15	626.20	637.34	1.8
34	164	methane	111.15	423.11	425.28	0.5
35	171	hydrogen sulfide	281.20	816.00	792.74	-2.9

Table 5.4. HIPPO enthalpy of vaporization (H_{vap}) compared to experimental values. H_{vap} is presented in kJ/mol. Experimental values from references ^{12, 13}.

	#ID	Name	T (K)	Expt.	HIPPO	% Error
1	1	chloroform	298.15	31.28	31.41	0.42
2	3	dibromomethane	298.15	37.45	37.31	-0.38
3	4	dichloromethane	298.15	28.82	27.72	-3.82
4	5	formaldehyde	253.65	23.08	22.60	-2.07
5	7	bromomethane	276.65	-	14.06	-
6	9	nitromethane	298.15	38.62	37.06	-4.04
7	11	1,1,1,2,2-pentachloroethane	293.15	46.29	46.93	1.38
8	15	methyl cyanide	298.15	33.23	33.49	0.77
9	18	1,2-dichloroethane	298.15	35.16	35.43	0.77
10	20	bromoethane	298.15	27.62	26.87	-2.71
11	21	chloroethane	273.15	25.39	25.43	0.17
12	25	nitroethane	298.15	40.24	40.20	-0.10
13	26	dimethyl ether	240.00	21.72	21.92	0.94
14	27	ethanol	298.15	42.32	44.66	5.53
15	29	dimethyl disulfide	298.15	38.32	39.09	2.00
16	36	ethyl cyanide	298.15	36.03	36.30	0.74
17	42	1,3-dioxolane	298.15	35.60	35.44	-0.44
18	45	N,N-dimethylformamide	298.15	47.57	46.29	-2.69
19	51	propan-1-amine	298.15	30.98	30.88	-0.31
20	53	isobutane	243.65	22.35	22.18	-0.78
21	64	pyrimidine	298.15	49.81	49.67	-0.28
22	72	oxolane	298.15	31.80	31.84	0.13
23	73	ethoxyethene	293.15	27.84	28.61	2.78
24	76	tetrahydrothiophene	298.15	38.62	37.60	-2.65
25	120	toluene	298.15	37.99	37.54	-1.17
26	148	benzene	298.15	33.83	33.27	-1.64
27	151	cyclopentane	293.15	27.30	27.27	-0.12
28	152	dimethyl sulfide	293.15	27.65	28.85	4.34
29	157	methyl chloride	298.15	18.92	19.09	0.92
30	158	methyl fluoride	298.15	-	8.81	-
31	159	methylene fluoride	221.00	-	19.25	-
32	160	neopentane	298.15	21.84	22.00	0.73
33	161	pentane	293.15	26.43	26.62	0.73
34	164	methane	111.15	8.17	8.54	4.59
35	171	hydrogen sulfide	281.20	15.90	14.54	-8.56

Table 5.5. HIPPO dielectric constant (ϵ) compared to experimental values. Dielectric constants (unitless) from references ^{12, 13}.

	#ID	Name	T (K)	ϵ (expt)	HIPPO	% error
1	1	chloroform	298.15	4.71	4.13	-12.25
2	3	dibromomethane	298.15	7.23	6.80	-5.93
3	4	dichloromethane	298.15	8.82	8.04	-8.79
4	5	formaldehyde	253.65	-	44.57	-
5	7	bromomethane	276.65	9.71	7.96	-18.03
6	9	nitromethane	298.15	36.56	22.68	-37.96
7	11	1,1,1,2,2-pentachloroethane	293.15	-	4.01	-
8	15	methyl cyanide	298.15	35.69	30.12	-15.60
9	18	1,2-dichloroethane	298.15	10.13	10.60	4.62
10	20	bromoethane	298.15	9.01	9.09	0.93
11	21	chloroethane	273.15	10.41	10.13	-2.74
12	25	nitroethane	298.15	28.29	18.82	-33.48
13	26	dimethyl ether	240.00	6.88	7.98	15.97
14	27	ethanol	298.15	24.85	24.13	-2.90
15	29	dimethyl disulfide	298.15	9.60	6.22	-35.23
16	36	ethyl cyanide	298.15	29.32	23.98	-18.22
17	42	1,3-dioxolane	298.15	-	4.21	-
18	45	N,N-dimethylformamide	298.15	37.22	37.82	1.60
19	51	propan-1-amine	298.15	5.11	4.60	-10.00
20	53	isobutane	243.65	1.85	1.84	-0.36
21	64	pyrimidine	298.15	-	22.92	
22	72	oxolane	298.15	8.04	9.12	13.42
23	73	ethoxyethene	293.15	-	6.36	-
24	76	tetrahydrothiophene	298.15	-	8.00	-
25	120	toluene	298.15	2.37	2.42	1.94
26	148	benzene	298.15	2.27	2.31	1.73
27	151	cyclopentane	293.15	1.97	1.91	-2.83
28	152	dimethyl sulfide	293.15	6.70	5.91	-11.84
29	157	methyl chloride	298.15	9.76	9.70	-0.60
30	158	methyl fluoride	298.15	8.84	8.40	-4.94
31	159	methylene fluoride	221.00	26.91	26.14	-2.85
32	160	neopentane	298.15	1.77	1.75	-1.20
33	161	pentane	293.15	1.84	1.85	0.64
34	164	methane	111.15	1.63	1.67	2.25
35	171	hydrogen sulfide	281.20	6.03	5.46	-9.43

5.4. References

1. Turney JM, Simmonett AC, Parrish RM, Hohenstein EG, Evangelista FA, Fermann JT, Mintz BJ, Burns LA, Wilke JJ, Abrams ML, Russ NJ, Leininger ML, Janssen CL, Seidl ET, Allen WD, Schaefer HF, King RA, Valeev EF, Sherrill CD, Crawford TD. Psi4: an open-source ab initio electronic structure program. *WIREs Comput. Mol. Sci.*, 2, 556-65 (2012).
2. Walker B, Liu C, Wait E, Ren P. Automation of
<scp>AMOEBA</scp>
polarizable force field for small molecules: Poltype 2. *J. Comput. Chem.*, 43, 1530-42 (2022).
3. Ren P, Ponder JW. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *The Journal of Physical Chemistry B*, 107, 5933-47 (2003).
4. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, Van Der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, Vanderplas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, Van Mulbregt P, Vijaykumar A, Bardelli AP, Rothberg A, Hilboll A, Kloeckner A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C, Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik DV, Olivetti E, Martin E, Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold G-L, Allen GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J, Nothman J, Buchner J, Kulick J, Schönberger JL, De Miranda Cardoso JV, Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K, Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ, Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT, Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T, Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T, Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17, 261-72 (2020).
5. Rackers JA, Wang Z, Lu C, Laury ML, Lagardere L, Schnieders MJ, Piquemal J-P, Ren P, Ponder JW. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.*, 14, 5273-89 (2018).
6. Wang Z. Tinker9: Next Generation of Tinker with GPU Support. St. Louis: Washington University in St. Louis; 2021.
7. Kříž K, Nováček M, Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets 3: Repulsive Contacts. *J. Chem. Theory Comput.*, 17, 1548-61 (2021).
8. Kříž K, Řezáč J. Non-covalent interactions atlas benchmark data sets 4: σ -hole interactions. *Phys. Chem. Chem. Phys.*, 24, 14794-804 (2022).
9. Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets 2: Hydrogen Bonding in an Extended Chemical Space. *J. Chem. Theory Comput.*, 16, 6305-16 (2020).
10. Řezáč J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.*, 16, 2355-68 (2020).
11. Donchev AG, Taube AG, Decolvenaere E, Hargus C, McGibbon RT, Law K-H, Gregersen BA, Li J-L, Palmo K, Siva K, Bergdorf M, Klepeis JL, Shaw DE. Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Scientific Data*, 8, (2021).

12. CRC Handbook of Chemistry and Physics. 84th ed. Lide DR, editor. Boca Raton, FL: CRC Press LLC; 2004.
13. Caleman C, van Maaren PJ, Hong M, Hub JS, Costa LT, van der Spoel D. Force Field Benchmark for Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.*, 8, 61-74 (2012).

Chapter 6. Future directions for the HIPPO model

6.1. Bridging Classical Physics and Advanced Machine Learning.

The crux of contemporary force fields, such as HIPPO, lies in their ability to robustly represent intermolecular interactions in molecular systems. While HIPPO has taken great strides in accurately depicting these interactions, certain challenges, like the incomplete classical treatment of bond, angle, and torsion terms, persist. The conventional models, which employ simple harmonic or Morse bond potentials, overlook quantum behaviors like zero-point energy and don't account for the electronic structure's nuances, especially short-range through-space and through-bond interactions. In molecular dynamics, specific angle representations, like the N-Ca-C "tau" angle in proteins, play a pivotal role in defining behavior, yet classical methods have struggled to capture these intricacies.

Recent advancements in machine learning (ML) and neural networks present promising avenues to tackle these challenges. While ML has been pervasively adopted across various scientific fields, its integration into force field development, specifically with HIPPO, presents an exciting frontier. However, it's crucial to emphasize that not all force field components demand an ML-driven approach. HIPPO's treatment of intermolecular physics is commendably precise, interpretable, and extensible. Thus, the primary area ripe for ML enhancement within HIPPO lies in the short-range bonded quantum mechanical interactions.

The proposed framework will aim to harness the power of ML to develop highly accurate valence terms that encapsulate these interactions. By training a neural network on the discrepancies between quantum mechanical (QM) electronic structure values and HIPPO's existing nonbonded

and intermolecular potentials, we aim to derive a more complete model. As this dissertation has shown, we already possess extensive data with numerous QM configuration results, either by using existing databases or computing necessary calculations. As an expansion of this endeavor, and in collaboration with developers of the AMOEBA model, we can generate a comprehensive QM dataset encompassing various configurations of capped amino acids, nucleotides, and selected trimers. In our approach, we will draw inspiration from the recent ML-force field models¹⁻³, with an emphasis on integrating short-range information to reduce computational loads.

The final step would see the amalgamation of these ML-derived forces with HIPPO's nonbonded potentials. This framework would be implemented in our GPU software, Tinker9. The overarching objective remains clear: seamlessly integrate ML models to capture the nuances of short-range intramolecular interactions while retaining HIPPO's strengths in long-range intermolecular domains.

In summation, the future of HIPPO lies at the crossroads of classical physics and advanced machine learning. By amalgamating the strengths of both worlds, we envision a force field model that stands unparalleled in accuracy and efficiency.

6.2. Development of the HIPPO Force Field for Proteins

In our quest to advance the capabilities of molecular dynamics simulations for biological molecules, we are focusing on the development of the HIPPO force field for proteins, inspired by the groundwork laid by the AMOEBA protein model.⁴ Given the vast complexity inherent in protein structures and dynamics, our development roadmap is layered and systematic. Given the experience and success of the AMOEBA protein model, HIPPO will follow on its footsteps. Below are the steps and methodologies we plan to employ:

Electrostatic Properties in the Gas Phase

A hallmark feature of a robust force field is its ability to seamlessly transition from gas phase to solution. To assess this, we will examine the electrostatic properties of amino acids in the gas phase and rigorously compare these to QM ab initio results. Ensuring accurate representation of peptide electrostatic properties across different conformations with a unified set of electrostatic parameters will be our primary goal.

Polyalanine Conformational Free Energy in Solution

Relying on oligopeptide conformational properties in solution, as used in AMOEBA, we will calibrate HIPPO's torsional parameters. By simulating peptides like Ala/Gly/Pro-based sequences and comparing the results to experimental NMR data, we aim to achieve a comprehensive understanding of the force field's capability in accurately representing peptide conformations in solution.

Proline and Glycine Conformational Free Energies in Solution

Proline and glycine, due to their unique structures and roles in proteins, warrant separate examinations. We will validate the ϕ/ψ torsional angle distributions for these amino acids through REMD simulations of representative peptides, subsequently comparing the results to PDB statistical PMF maps.

Secondary Structure Distribution for the Ac- (AAQAA)₃-NH₂ Peptide

An effective protein force field should correctly predict the secondary structure of peptides and strike a balance in peptide conformations. Using REMD simulations, we will simulate longer

helix-forming peptides, like Ac-(AAQAA)₃-NH₂, to investigate helix–coil transitions and compare the results to available experimental data.

Molecular Dynamics Simulations of Protein Systems

To ensure our force field performs well with larger systems, we will run molecular dynamics simulations on representative protein systems. These simulations will be compared against known X-ray or NMR structures.

Calculation of NMR Order Parameters

To further refine and validate our force field, we will compute NMR order parameters from our simulations and compare these values to experimental NMR measurements.

Calculation of Side Chain J-Couplings

As a final validation step, side chain J-couplings will be calculated and benchmarked against experimental values. This will offer insights into the accuracy of our force field in capturing side chain dynamics.

For larger molecules like proteins, a comparison of detailed simulations with limited data such as NMR J-coupling constants, database-derived statistical populations, and atomic structures from X-ray or NMR experiments is essential. The final validation and optimization of the HIPPO force field for proteins will be an iterative process, involving continuous feedback from both simulations and experimental data. In-depth investigations across a variety of protein-related research areas will be required to validate different components of our proposed force field thoroughly. By harnessing the strengths and learning from the limitations of the AMOEBA protein

model, we are optimistic about developing a state-of-the-art HIPPO force field for proteins, promising enhanced accuracy and utility in biological simulations.

6.3. References

1. Glick ZL, Koutsoukas A, Cheney DL, Sherrill CD. Cartesian message passing neural networks for directional properties: Fast and transferable atomic multipoles. *The Journal of Chemical Physics*, 154, (2021).
2. Takaba K, Pulido Ia, Henry M, MacDermott-Opeskin H, Chodera JD, Wang Y. Espaloma-0.3.0: Machine-learned molecular mechanics force field for the simulation of protein-ligand systems and beyond. (2023).
3. Smith JS, Isayev O, Roitberg AE. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical Science*, 8, 3192-203 (2017).
4. Shi Y, Xia Z, Zhang J, Best R, Wu C, Ponder JW, Ren P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.*, 9, 4046-63 (2013).

Appendix A

Electrostatic Energy:

Core-Core:

$$U_{core-core} = Z_i T_{ij} Z_j$$

$$T_{ij} = \frac{1}{R}$$

Core-Density:

$$U_{core-density} = Z_i \mathbf{T}_{ij}^* \vec{M}_j$$

$$\mathbf{T}_{ij}^* = [1 \quad \nabla \quad \nabla^2] T^*$$

$$T^* = \frac{1}{R} f_1^{damp}$$

$$\nabla T^* = -f_3^{damp} \frac{R_\alpha}{R^3}$$

$$\nabla^2 T^* = f_5^{damp} \frac{3R_\alpha R_\beta}{R^5} - f_3^{damp} \frac{\delta_{\alpha\beta}}{R^3}$$

$$f_1^{damp} = 1 - \left(1 + \frac{1}{2} \zeta_j R\right) e^{-\zeta_j R}$$

$$f_3^{damp} = 1 - \left(1 + \zeta_j R + \frac{1}{2} (\zeta_j R)^2\right) e^{-\zeta_j R}$$

$$f_5^{damp} = 1 - \left(1 + \zeta_j R + \frac{1}{2} (\zeta_j R)^2 + \frac{1}{6} (\zeta_j R)^3\right) e^{-\zeta_j R}$$

Density-Density:

$$U_{density-density} = \vec{M}_i \mathbf{T}_{ij}^{overlap} \vec{M}_j$$

$$\mathbf{T}_{ij}^{overlap} = \begin{bmatrix} 1 & \nabla & \nabla^2 \\ \nabla & \nabla^2 & \nabla^3 \\ \nabla^2 & \nabla^3 & \nabla^4 \end{bmatrix} (T^{overlap})$$

$$T^{overlap} = \frac{1}{R} f_1^{overlap}$$

$$\nabla T^{overlap} = -f_3^{overlap} \frac{R_\alpha}{R^3}$$

$$\nabla^2 T^{overlap} = f_5^{overlap} \frac{3R_\alpha R_\beta}{R^5} - f_3^{overlap} \frac{\delta_{\alpha\beta}}{R^3}$$

$$\nabla^3 T^{overlap} = -f_7^{overlap} \frac{15R_\alpha R_\beta R_\gamma}{R^7} + f_5^{overlap} \frac{3(R_\alpha \delta_{\beta\gamma} + R_\beta \delta_{\alpha\gamma} + R_\gamma \delta_{\alpha\beta})}{R^5}$$

$$\nabla^4 T^{overlap}$$

$$= f_9^{overlap} \frac{105R_\alpha R_\beta R_\gamma R_\eta}{R^9}$$

$$- f_7^{overlap} \frac{15(R_\alpha R_\beta \delta_{\gamma\eta} + R_\alpha R_\gamma \delta_{\beta\eta} + R_\alpha R_\eta \delta_{\beta\gamma} + R_\beta R_\gamma \delta_{\alpha\eta} + R_\beta R_\eta \delta_{\alpha\gamma} + R_\gamma R_\eta \delta_{\alpha\beta})}{R^7}$$

$$+ f_5^{overlap} \frac{3(\delta_{\alpha\beta} \delta_{\gamma\eta} + \delta_{\alpha\gamma} \delta_{\beta\eta} + \delta_{\alpha\eta} \delta_{\beta\gamma})}{R^5}$$

$$f_1^{overlap} = \begin{cases} 1 - \left(1 + \frac{11}{16} \zeta R + \frac{3}{16} (\zeta R)^2 + \frac{1}{48} (\zeta R)^3\right) e^{-\zeta R}, & \zeta_i = \zeta_j \\ 1 - A^2 \left(1 + 2B + \frac{\zeta_i}{2} R\right) e^{-\zeta_i R} - B^2 \left(1 + 2A + \frac{\zeta_j}{2} R\right) e^{-\zeta_j R}, & \zeta_i \neq \zeta_j \end{cases}$$

$$f_3^{overlap} = \begin{cases} 1 - \left(1 + \zeta R + \frac{1}{2} (\zeta R)^2 + \frac{7}{48} (\zeta R)^3 + \frac{1}{48} (\zeta R)^4\right) e^{-\zeta R}, & \zeta_i = \zeta_j \\ 1 - A^2 \left(1 + \zeta_i R + \frac{1}{2} (\zeta_i R)^2\right) e^{-\zeta_i R} - B^2 \left(1 + \zeta_j R + \frac{1}{2} (\zeta_j R)^2\right) e^{-\zeta_j R} - \\ 2A^2 B (1 + \zeta_i R) e^{-\zeta_i R} - 2B^2 A (1 + \zeta_j R) e^{-\zeta_j R}, & \zeta_i \neq \zeta_j \end{cases}$$

$$f_5^{overlap} = \begin{cases} 1 - \left(1 + \zeta R + \frac{1}{2}(\zeta R)^2 + \frac{1}{6}(\zeta R)^3 + \frac{1}{24}(\zeta R)^4 + \frac{1}{144}(\zeta R)^5\right) e^{-\zeta R}, & \zeta_i = \zeta_j \\ 1 - A^2 \left(1 + \zeta_i R + \frac{1}{2}(\zeta_i R)^2 + \frac{1}{6}(\zeta_i R)^3\right) e^{-\zeta_i R} - \\ B^2 \left(1 + \zeta_j R + \frac{1}{2}(\zeta_j R)^2 + \frac{1}{6}(\zeta_j R)^3\right) e^{-\zeta_j R} - \\ 2A^2 B \left(1 + \zeta_i R + \frac{1}{3}(\zeta_i R)^2\right) e^{-\zeta_i R} - \\ 2B^2 A \left(1 + \zeta_j R + \frac{1}{3}(\zeta_j R)^2\right) e^{-\zeta_j R}, & \zeta_i \neq \zeta_j \end{cases}$$

$$f_7^{overlap} = \begin{cases} 1 - \left(1 + \zeta R + \frac{1}{2}(\zeta R)^2 + \frac{1}{6}(\zeta R)^3 + \frac{1}{24}(\zeta R)^4 + \frac{1}{120}(\zeta R)^5 + \frac{1}{720}(\zeta R)^6\right) e^{-\zeta R}, & \zeta_i = \zeta_j \\ 1 - A^2 \left(1 + \zeta_i R + \frac{1}{2}(\zeta_i R)^2 + \frac{1}{6}(\zeta_i R)^3 + \frac{1}{30}(\zeta_i R)^4\right) e^{-\zeta_i R} - \\ B^2 \left(1 + \zeta_j R + \frac{1}{2}(\zeta_j R)^2 + \frac{1}{6}(\zeta_j R)^3 + \frac{1}{30}(\zeta_j R)^4\right) e^{-\zeta_j R} - \\ 2A^2 B \left(1 + \zeta_i R + \frac{2}{5}(\zeta_i R)^2 + \frac{1}{15}(\zeta_i R)^3\right) e^{-\zeta_i R} - \\ 2B^2 A \left(1 + \zeta_j R + \frac{2}{5}(\zeta_j R)^2 + \frac{1}{15}(\zeta_j R)^3\right) e^{-\zeta_j R}, & \zeta_i \neq \zeta_j \end{cases}$$

$$f_9^{overlap} = \begin{cases} 1 - \left(1 + \zeta R + \frac{1}{2}(\zeta R)^2 + \frac{1}{6}(\zeta R)^3 + \frac{1}{24}(\zeta R)^4 + \frac{1}{120}(\zeta R)^5 + \frac{1}{720}(\zeta R)^6 + \frac{1}{5040}(\zeta R)^7\right) e^{-\zeta R}, & \zeta_i = \zeta_j \\ 1 - A^2 \left(1 + \zeta_i R + \frac{1}{2}(\zeta_i R)^2 + \frac{1}{6}(\zeta_i R)^3 + \frac{4}{105}(\zeta_i R)^4 + \frac{1}{210}(\zeta_i R)^5\right) e^{-\zeta_i R} - \\ B^2 \left(1 + \zeta_j R + \frac{1}{2}(\zeta_j R)^2 + \frac{1}{6}(\zeta_j R)^3 + \frac{4}{105}(\zeta_j R)^4 + \frac{1}{210}(\zeta_j R)^5\right) e^{-\zeta_j R} - \\ 2A^2 B \left(1 + \zeta_i R + \frac{3}{7}(\zeta_i R)^2 + \frac{2}{21}(\zeta_i R)^3 + \frac{1}{105}(\zeta_i R)^4\right) e^{-\zeta_i R} - \\ 2B^2 A \left(1 + \zeta_j R + \frac{3}{7}(\zeta_j R)^2 + \frac{2}{21}(\zeta_j R)^3 + \frac{1}{105}(\zeta_j R)^4\right) e^{-\zeta_j R}, & \zeta_i \neq \zeta_j \end{cases}$$

$$B = \frac{\zeta_i^2}{\zeta_i^2 - \zeta_j^2}, \quad A = \frac{\zeta_j^2}{\zeta_j^2 - \zeta_i^2}$$

Appendix B

Permanent Electrostatic Field:

(field at induced dipole i, due to permanent moments of atom j)

$$\mathbf{F}_i^{perm}(R) = Z_j \nabla \left(\frac{1}{R} \right) + Q_j \nabla \left(\frac{1}{R} f^{damp}(R) \right) + \boldsymbol{\mu}_j \cdot \nabla^2 \left(\frac{1}{R} f^{damp}(R) \right) + \boldsymbol{\Theta}_j : \nabla^3 \left(\frac{1}{R} f^{damp}(R) \right)$$

$$\nabla \left(\frac{1}{R} f^{damp}(R) \right) = -f_3^{damp} \frac{R_\alpha}{R^3}$$

$$\nabla^2 \left(\frac{1}{R} f^{damp}(R) \right) = f_5^{damp} \frac{3R_\alpha R_\beta}{R^5} - f_3^{damp} \frac{\delta_{\alpha\beta}}{R^3}$$

$$\nabla^3 \left(\frac{1}{R} f^{damp}(R) \right) = -f_7^{damp} \frac{15R_\alpha R_\beta R_\gamma}{R^7} + f_5^{damp} \frac{3(R_\alpha \delta_{\beta\gamma} + R_\beta \delta_{\alpha\gamma} + R_\gamma \delta_{\alpha\beta})}{R^5}$$

$$f_3^{damp} = 1 - \left(1 + \zeta_j R + \frac{1}{2} (\zeta_j R)^2 \right) e^{-\zeta_j R}$$

$$f_5^{damp} = 1 - \left(1 + \zeta_j R + \frac{1}{2} (\zeta_j R)^2 + \frac{1}{6} (\zeta_j R)^3 \right) e^{-\zeta_j R}$$

$$f_7^{damp} = 1 - \left(1 + \zeta_j R + \frac{1}{2} (\zeta_j R)^2 + \frac{1}{6} (\zeta_j R)^3 + \frac{1}{30} (\zeta_j R)^4 \right) e^{-\zeta_j R}$$

Induced Dipole Electrostatic Field:

(field at induced dipole i, due to induced dipole j)

$$\mathbf{F}_i^{ind}(R) = \boldsymbol{\mu}_j^{ind} \cdot \nabla^2 \left(\frac{1}{R} f^{overlap}(R) \right)$$

$$\nabla^2 \left(\frac{1}{R} f^{overlap}(R) \right) = f_5^{overlap} \frac{3R_\alpha R_\beta}{R^5} - f_3^{overlap} \frac{\delta_{\alpha\beta}}{R^3}$$

$$f_3^{overlap} = \begin{cases} 1 - \left(1 + \zeta R + \frac{1}{2} (\zeta R)^2 + \frac{7}{48} (\zeta R)^3 + \frac{1}{48} (\zeta R)^4 \right) e^{-\alpha R}, & \zeta_i = \zeta_j \\ 1 - A^2 \left(1 + \zeta_i R + \frac{1}{2} (\zeta_i R)^2 \right) e^{-\zeta_i R} - B^2 \left(1 + \zeta_j R + \frac{1}{2} (\zeta_j R)^2 \right) e^{-\zeta_j R} - \\ \quad 2A^2 B (1 + \zeta_i R) e^{-\zeta_i R} - 2B^2 A (1 + \zeta_j R) e^{-\zeta_j R}, & \zeta_i \neq \zeta_j \end{cases}$$

$$f_5^{overlap} = \begin{cases} 1 - \left(1 + \zeta R + \frac{1}{2} (\zeta R)^2 + \frac{1}{6} (\zeta R)^3 + \frac{1}{24} (\zeta R)^4 + \frac{1}{144} (\zeta R)^5 \right) e^{-\zeta R}, & \zeta_i = \zeta_j \\ 1 - A^2 \left(1 + \zeta_i R + \frac{1}{2} (\zeta_i R)^2 + \frac{1}{6} (\zeta_i R)^3 \right) e^{-\zeta_i R} - \\ \quad B^2 \left(1 + \zeta_j R + \frac{1}{2} (\zeta_j R)^2 + \frac{1}{6} (\zeta_j R)^3 \right) e^{-\zeta_j R} - \\ \quad 2A^2 B \left(1 + \zeta_i R + \frac{1}{3} (\zeta_i R)^2 \right) e^{-\zeta_i R} - \\ \quad 2B^2 A \left(1 + \zeta_j R + \frac{1}{3} (\zeta_j R)^2 \right) e^{-\zeta_j R}, & \zeta_i \neq \zeta_j \end{cases}$$

$$B = \frac{\zeta_i^2}{\zeta_i^2 - \zeta_j^2}, \quad A = \frac{\zeta_j^2}{\zeta_j^2 - \zeta_i^2}$$

Appendix C

Pauli Repulsion:

$$U_{ij} = \frac{K_i K_j}{R} S_{\text{total}}^2$$

$$\frac{S_{\text{total}}^2}{R} = \vec{M}_i \mathbf{T}_{ij}^{\text{repulsion}} \vec{M}_j$$

$$\mathbf{T}_{ij}^{\text{repulsion}} = \begin{bmatrix} 1 & \nabla & \nabla^2 \\ \nabla & \nabla^2 & \nabla^3 \\ \nabla^2 & \nabla^3 & \nabla^4 \end{bmatrix} (\text{T}^{\text{pauli}})$$

$$\text{T}^{\text{pauli}} = \frac{\zeta_i^3 \zeta_j^3}{R} f_1^{\text{rep}}$$

$$\nabla \text{T}^{\text{pauli}} = -f_3^{\text{rep}} R_\alpha$$

$$\nabla^2 \text{T}^{\text{pauli}} = f_5^{\text{rep}} 3 R_\alpha R_\beta - f_3^{\text{rep}} \delta_{\alpha\beta}$$

$$\nabla^3 \text{T}^{\text{pauli}} = -f_7^{\text{rep}} 15 R_\alpha R_\beta R_\gamma + f_5^{\text{rep}} 3 (R_\alpha \delta_{\beta\gamma} + R_\beta \delta_{\alpha\gamma} + R_\gamma \delta_{\alpha\beta})$$

$$\begin{aligned} \nabla^4 \text{T}^{\text{pauli}} &= f_9^{\text{rep}} 105 R_\alpha R_\beta R_\gamma R_\eta \\ &\quad - f_7^{\text{rep}} 15 (R_\alpha R_\beta \delta_{\gamma\eta} + R_\alpha R_\gamma \delta_{\beta\eta} + R_\alpha R_\eta \delta_{\beta\gamma} + R_\beta R_\gamma \delta_{\alpha\eta} + R_\beta R_\eta \delta_{\alpha\gamma} + R_\gamma R_\eta \delta_{\alpha\beta}) \\ &\quad + f_5^{\text{rep}} 3 (\delta_{\alpha\beta} \delta_{\gamma\eta} + \delta_{\alpha\gamma} \delta_{\beta\eta} + \delta_{\alpha\eta} \delta_{\beta\gamma}) \end{aligned}$$

$$f_1^{\text{rep}} = (f_{\text{exp}})^2$$

$$f_{\text{exp}} = \begin{cases} \frac{1}{\zeta^3} \left(1 + \frac{\zeta R}{2} + \frac{1}{3} \left(\frac{\zeta R}{2} \right)^2 \right) e^{-\frac{\zeta R}{2}}, & \zeta_i = \zeta_j \\ \frac{1}{2X^3 R} \left[\zeta_i (RX - 2\zeta_j) e^{-\frac{\zeta_j R}{2}} + \zeta_j (RX + 2\zeta_i) e^{-\frac{\zeta_i R}{2}} \right], & \zeta_i \neq \zeta_j \end{cases},$$

$$X = \left(\frac{\zeta_i}{2} \right)^2 - \left(\frac{\zeta_j}{2} \right)^2.$$

$$f_3^{\text{rep}} = 2 f_{\text{exp}} f'_{\text{exp}}$$

$$f'_{exp} = \begin{cases} \frac{1}{\zeta^3} \frac{1}{3} \left(\frac{\zeta}{2}\right)^2 \left(1 + \frac{\zeta R}{2}\right) e^{-\frac{\zeta R}{2}}, & \zeta_i = \zeta_j \\ \frac{1}{2X^3 R} \left[\left(\frac{1}{2} \zeta_i \zeta_j X - \frac{\zeta_i \zeta_j^2}{R} - \frac{2\zeta_i \zeta_j}{R^2}\right) e^{-\frac{\zeta_j R}{2}} + \left(\frac{1}{2} \zeta_i \zeta_j X + \frac{\zeta_j \zeta_i^2}{R} + \frac{2\zeta_i \zeta_j}{R^2}\right) e^{-\frac{\zeta_i R}{2}} \right], & \zeta_i \neq \zeta_j \end{cases}$$

$$f_5^{rep} = 2(f_{exp} f''_{exp} + f'_{exp} f'_{exp})$$

$$f''_{exp} = \begin{cases} \frac{1}{\zeta^3} \frac{1}{9} \left(\frac{\zeta}{2}\right)^4 e^{-\frac{\zeta R}{2}}, & \zeta_i = \zeta_j \\ \frac{1}{2X^3 R^2} \left[\left(\frac{1}{4} \zeta_i \zeta_j^2 X - \frac{\zeta_i \zeta_j^3}{2R} + \frac{\zeta_i \zeta_j X}{2R} - \frac{3\zeta_i \zeta_j^2}{R^2} - \frac{6\zeta_i \zeta_j}{R^5}\right) e^{-\frac{\zeta_j R}{2}} + \left(\frac{1}{4} \zeta_j \zeta_i^2 X + \frac{\zeta_j \zeta_i^3}{2R} + \frac{\zeta_j \zeta_i X}{2R} + \frac{3\zeta_j \zeta_i^2}{R^2} + \frac{6\zeta_j \zeta_i}{R^5}\right) e^{-\frac{\zeta_i R}{2}} \right], & \zeta_i \neq \zeta_j \end{cases}$$

$$f_7^{rep} = 2(f_{exp} f'''_{exp} + 3f''_{exp} f'_{exp})$$

$$f'''_{exp}$$

$$= \begin{cases} \frac{1}{\zeta^3} \frac{1}{45} \left(\frac{\zeta}{2}\right)^5 \frac{1}{R} e^{-\frac{\zeta R}{2}}, & \zeta_i = \zeta_j \\ \frac{1}{2X^3 R^3} \left[\left(\frac{1}{8} \zeta_i \zeta_j^3 X + \frac{3\zeta_i \zeta_j^2 X}{4} + \frac{3\zeta_i \zeta_j X}{2} - \frac{1\zeta_i \zeta_j^4}{4} - \frac{3\zeta_i \zeta_j^3}{R^2} - \frac{15\zeta_i \zeta_j^2}{R^3} - \frac{30\zeta_i \zeta_j}{R^4}\right) e^{-\frac{\zeta_j R}{2}} + \left(\frac{1}{8} \zeta_j \zeta_i^3 X + \frac{3\zeta_j \zeta_i^2 X}{4} + \frac{3\zeta_j \zeta_i X}{2} + \frac{1\zeta_j \zeta_i^4}{4} + \frac{3\zeta_j \zeta_i^3}{R^2} + \frac{15\zeta_j \zeta_i^2}{R^3} + \frac{30\zeta_j \zeta_i}{R^4}\right) e^{-\frac{\zeta_i R}{2}} \right], & \zeta_i \neq \zeta_j \end{cases}$$

$$f_9^{rep} = 2(f_{exp} f''''_{exp} + 4f'''_{exp} f'_{exp} + 3f''_{exp} f''_{exp})$$

$$f''''_{exp} = \begin{cases} \frac{1}{\zeta^3} \frac{1}{315} \left(\frac{\zeta}{2}\right)^5 \frac{1}{R^3} \left(1 + \frac{\zeta R}{2}\right) e^{-\frac{\zeta R}{2}}, & \zeta_i = \zeta_j \\ \frac{1}{2X^3 R^4} \left[\left(\frac{1}{16} \zeta_i \zeta_j^4 X + \frac{3\zeta_i \zeta_j^3 X}{4} + \frac{15\zeta_i \zeta_j^2 X}{4} + \frac{15\zeta_i \zeta_j X}{2} - \frac{1\zeta_i \zeta_j^5}{8} - \frac{5\zeta_i \zeta_j^4}{2} - \frac{45\zeta_i \zeta_j^3}{2} - \frac{105\zeta_i \zeta_j^2}{R^4} - \frac{210\zeta_i \zeta_j}{R^5}\right) e^{-\frac{\zeta_j R}{2}} + \left(\frac{1}{16} \zeta_j \zeta_i^4 X + \frac{3\zeta_j \zeta_i^3 X}{4} + \frac{15\zeta_j \zeta_i^2 X}{4} + \frac{15\zeta_j \zeta_i X}{2} + \frac{1\zeta_j \zeta_i^5}{8} + \frac{5\zeta_j \zeta_i^4}{2} + \frac{45\zeta_j \zeta_i^3}{2} + \frac{105\zeta_j \zeta_i^2}{R^4} + \frac{210\zeta_j \zeta_i}{R^5}\right) e^{-\frac{\zeta_i R}{2}} \right], & \zeta_i \neq \zeta_j \end{cases}$$

Appendix D

Table D.1: Data set 1. 165 molecules with liquid experimental data and Quantum calculation references. CID is the *Pubchem* (<https://pubchem.ncbi.nlm.nih.gov>) assigned number.

#ID	CID	Name	Formula	Functional Groups
1	6212	chloroform	CHCl3	chlorine
2	6370	dichloro(fluoro)methane	CHCl2F	chlorine; fluorine
3	3024	dibromomethane	CH2Br2	bromine
4	6344	dichloromethane	CH2Cl2	chlorine
5	712	formaldehyde	CH2O	aldehydes
6	284	formic acid	CH2O2	carboxylic_acids
7	6323	bromomethane	CH3Br	bromine
8	713	formamide	CH3NO	amides
9	6375	nitromethane	CH3NO2	nitros
10	887	methanol	CH4O	alcohols
11	6419	1,1,1,2,2-pentachloroethane	C2HCl5	chlorine
12	6591	1,1,2,2-tetrachloroethane	C2H2Cl4	chlorine
13	6366	1,1-dichloroethene	C2H2Cl2	chlorine
14	6574	1,1,2-trichloroethane	C2H3Cl3	chlorine
15	6342	methyl cyanide	C2H3N	nitriles
16	7839	1,2-dibromoethane	C2H4Br2	bromine
17	6365	1,1-dichloroethane	C2H4Cl2	chlorine
18	11	1,2-dichloroethane	C2H4Cl2	chlorine
19	7865	methyl formate	C2H4O2	esters
20	6332	bromoethane	C2H5Br	bromine
21	6337	chloroethane	C2H5Cl	chlorine
22	34	2-chloroethanol	C2H5ClO	chlorine; alcohols
23	178	acetamide	C2H5NO	amides
24	31254	N-methylformamide	C2H5NO	amides
25	6587	nitroethane	C2H5NO2	nitros
26	8254	dimethyl ether	C2H6O	ethers
27	702	ethanol	C2H6O	alcohols
28	10902	1,2-ethanedithiol	C2H6S2	thiols
29	12232	dimethyl disulfide	C2H6S2	disulfides
30	679	dimethyl sulfoxide	C2H6OS	sulfoxides
31	1068	methylsulfanylmethane	C2H6S	sulfides
32	700	2-aminoethanol	C2H7NO	alcohols; amines
33	3301	ethane-1,2-diamine	C2H8N2	amines
34	7855	prop-2-enenitrile	C3H3N	nitriles
35	7303	1,3-dioxolan-2-one	C3H4O3	esters; cyclics
36	7854	ethyl cyanide	C3H5N	nitriles
37	6553	1,2-dibromopropane	C3H6Br2	bromine
38	8881	1,3-dichloropropane	C3H6Cl2	chlorine
39	146261	(2R)-2-methyloxirane	C3H6O	epoxide
40	180	acetone	C3H6O	ketones

41	6584	methyl acetate	C3H6O2	esters
42	12586	1,3-dioxolane	C3H6O2	acetal
43	6362	2-iodopropane	C3H7I	iodine
44	7840	1-bromopropane	C3H7Br	bromine
45	6228	N,N-dimethylformamide	C3H7NO	amides
46	6582	N-methylacetamide	C3H7NO	amides
47	7903	1-nitropropane	C3H7NO2	nitros
48	398	2-nitropropane	C3H7NO2	nitros
49	8020	dimethoxymethane	C3H8O2	ethers
50	753	propane-1,2,3-triol	C3H8O3	alcohols
51	7852	propan-1-amine	C3H9N	amines
52	6363	propan-2-amine	C3H9N	amines
53	6360	isobutane	C4H10	hydrocarbons
54	9609	ethylsulfanyethane	C4H10S	thiols
55	8012	butane-1-thiol	C4H10S	thiols
56	263	butan-1-ol	C4H10O	alcohols
57	6386	2-methylpropan-2-ol	C4H10O	alcohols
58	8064	butane-1,4-diol	C4H10O2	alcohols
59	8117	2,2'-Oxydiethanol	C4H10O3	alcohols; ethers
60	8021	N-ethylethanamine	C4H11N	amines
61	8007	butan-1-amine	C4H11N	amines
62	6385	2-methylpropan-2-amine	C4H11N	amines
63	8113	2-(2-hydroxyethylamino)ethanol	C4H11NO2	alcohols; amines
64	9260	pyrimidine	C4H4N2	pyrimidine; aromatics
65	8029	furan	C4H4O	ethers; cyclics; aromatics
66	8030	thiophene	C4H4S	thiols; cyclics; aromatics; sulfides
67	8027	1H-pyrrole	C4H5N	amines; pyrroles; cyclics; aromatics
68	7904	ethenyl acetate	C4H6O2	esters
69	7302	oxolan-2-one	C4H6O2	esters
70	7918	acetyl acetate	C4H6O3	anhydride
71	8059	1,4-dichlorobutane	C4H8Cl2	chlorine
72	8028	oxolane	C4H8O	ethers; cyclics
73	8023	ethoxyethene	C4H8O	ethers
74	8857	ethyl acetate	C4H8O2	esters
75	31347	tetrahydrothiophene 1,1-dioxide	C4H8O2S	thiols; cyclics
76	1127	tetrahydrothiophene	C4H8S	thiols; cyclics
77	8002	1-bromobutane	C4H9Br	bromine
78	8005	1-chlorobutane	C4H9Cl	chlorine
79	31268	pyrrolidine	C4H9N	pyrroles
80	31374	N,N-dimethylacetamide	C4H9NO	amides
81	8083	morpholine	C4H9NO	ethers; amines; cyclics
82	1049	pyridine	C5H5N	pyridines; aromatics
83	8452	cyclopentanone	C5H8O	ketones; cyclics
84	13004	1-cyclopropylethanone	C5H8O	ketones; cyclics
85	31261	pentane-2,4-dione	C5H8O2	ketones
86	6658	methyl 2-methylprop-2-enoate	C5H8O2	esters

87	8061	pentanenitrile	C5H9N	nitriles
88	7749	ethyl propanoate	C5H10O2	esters
89	7766	diethyl carbonate	C5H10O3	esters
90	6276	pentan-1-ol	C5H12O	alcohols
91	11428	pentan-3-ol	C5H12O	alcohols
92	6405	2-methylbutan-2-ol	C5H12O	alcohols
93	8105	pentane-1,5-diol	C5H12O2	alcohols
94	12019	pentan-3-amine	C5H13N	amines
95	11084	1,2,3,4-tetrafluorobenzene	C6H2F4	fluorine; benzenes
96	16910	1,2,3,5-tetrafluorobenzene	C6H2F4	fluorine; benzenes
97	9741	1,3-difluorobenzene	C6H4F2	fluorine; benzenes; aromatics
98	9706	1,2-difluorobenzene	C6H4F2	fluorine; benzenes
99	10008	fluorobenzene	C6H5F	fluorine; benzenes; aromatics
100	7416	nitrobenzene	C6H5NO2	nitros; benzenes; aromatics
101	7240	2-chloroaniline	C6H6ClN	chlorine; benzenes
102	996	phenol	C6H6O	alcohols; benzenes; phenols; aromatics
103	7969	benzenethiol	C6H6S	thiols; benzenes
104	7975	2-methylpyridine	C6H7N	pyridines
105	7970	3-methylpyridine	C6H7N	pyridines; aromatics
106	7963	4-methylpyridine	C6H7N	pyridines
107	7967	cyclohexanone	C6H10O	ketones; cyclics
108	639661	(E)-hex-2-ene	C6H12	hydrocarbons
109	11583	hexan-2-one	C6H12O	ketones
110	31264	2,4,6-trimethyl-1,3,5-trioxane	C6H12O3	ethers; cyclics
111	7965	cyclohexanamine	C6H13N	amines; cyclics
112	7914	2-propan-2-yloxypropane	C6H14O	ethers
113	8150	1-methoxy-2-(2-methoxyethoxy)ethane	C6H14O3	ethers
114	6535	triethyl phosphate	C6H15O4P	phosphates
115	8471	N,N-diethylethanamine	C6H15N	amines
116	7912	N-propan-2-ylpropan-2-amine	C6H15N	amines
117	7368	trifluoromethylbenzene	C7H5F3	fluorine; benzenes
118	7505	benzonitrile	C7H5N	nitriles; benzenes
119	240	benzaldehyde	C7H6O	aldehydes; benzenes
120	1140	toluene	C7H8	hydrocarbons; benzenes; aromatics
121	7519	methoxybenzene	C7H8O	ethers; benzenes
122	244	phenylmethanol	C7H8O	alcohols; benzenes; phenols
123	335	2-methylphenol	C7H8O	benzenes; phenols
124	342	3-methylphenol	C7H8O	benzenes; phenols
125	2879	4-methylphenol	C7H8O	benzenes; phenols; aromatics
126	7761	diethyl propanedioate	C7H12O4	esters
127	11271	2,4-dimethylpentan-3-one	C7H14O	ketones
128	8051	heptan-2-one	C7H14O	ketones
129	7501	ethenylbenzene	C8H8	hydrocarbons; benzenes
130	7410	1-phenylethanone	C8H8O	ketones; benzenes

131	7150	methyl benzoate	C8H8O2	esters; benzenes
132	4133	methyl 2-hydroxybenzoate	C8H8O3	esters; benzenes; phenols
133	7500	ethylbenzene	C8H10	hydrocarbons; benzenes
134	7237	1,2-dimethylbenzene	C8H10	hydrocarbons; benzenes
135	7043	1,2-dimethoxybenzene	C8H10O2	ethers; benzenes
136	7953	2,4,6-trimethylpyridine	C8H11N	aromatics; pyridines
137	957	octan-1-ol	C8H18O	alcohols
138	8909	1-butoxybutane	C8H18O	ethers
139	8148	N-butylbutan-1-amine	C8H19N	amines
140	8405	isoquinoline	C9H7N	benzenes; quinolines
141	7047	quinoline	C9H7N	benzenes; quinolines
142	7406	(1-methylethyl)benzene	C9H12	hydrocarbons; benzenes
143	7247	1,2,4-trimethylbenzene	C9H12	hydrocarbons; benzenes
144	7958	2,6-dimethylheptan-4-one	C9H18O	ketones
145	7003	1-chloronaphthalene	C10H7Cl	chlorine; benzenes
146	7583	phenoxybenzene	C12H10O	ethers; benzenes
147	176	acetic acid	C2H4O2	carboxylic_acids
148	241	benzene	C6H6	benzenes
149	7961	bromobenzene	C6H5Br	bromine; benzenes
150	7964	chlorobenzene	C6H5Cl	chlorine; benzenes
151	9253	cyclopentane	C5H10	hydrocarbons
152	1068	dimethyl sulfide	C2H6S	thiols
153	6325	ethene	C2H4	hydrocarbons
154	795	imidazole	C3H4N2	nucleic_acids; aromatics
155	798	indole	C8H7N	nucleic_acids; aromatics
156	6329	methyl amine	CH5N	amines
157	6327	methyl chloride	CH3Cl	chlorine
158	11638	methyl fluoride	CH3F	fluorine
159	6345	methylene fluoride	CH2F2	fluorine
160	10041	neopentane	C5H12	hydrocarbons
161	8003	pentane	C5H12	hydrocarbons
162	5943	carbon tetrachloride	CCl4	chlorine
163	6115	aniline	C6H7N	aromatics; benzenes
164	297	methane	CH4	hydrocarbons
165	222	ammonia	H3N	amines

Table D.2: Data set 2. 339 molecules without liquid experimental data included. These molecules were selected from the NCIA and DES370k databases. The selection process is discussed in chapter 3.

	#ID	CID	Name	Formula
1	166	174	ethane-1,2-diol	C2H6O2
2	167	177	acetaldehyde	C2H4O
3	168	239	3-aminopropanoic acid	C3H7NO2
4	169	260	bromane	BrH
5	170	313	chlorane	ClH
6	171	402	hydrogen sulfide	H2S
7	172	428	propane-1,3-diamine	C3H10N2
8	173	527	propanal	C3H6O
9	174	563	3-methylsulfanylpropanoic acid	C4H8O2S
10	175	674	N-methylmethanamine	C2H7N
11	176	750	2-aminoacetic acid	C2H5NO2
12	177	757	2-hydroxyacetic acid	C2H4O3
13	178	768	hydrogen cyanide	CHN
14	179	807	molecular iodine	I2
15	180	867	propanedioic acid	C3H4O4
16	181	878	methanethiol	CH4S
17	183	1004	phosphoric acid	H3O4P
18	184	1031	propan-1-ol	C3H8O
19	185	1032	propanoic acid	C3H6O2
20	186	1088	2-(methylamino)acetic acid	C3H7NO2
21	187	1119	sulfur dioxide	O2S
22	188	1133	2-sulfanylacetic acid	C2H4O2S
23	189	1146	N,N-dimethylmethanamine	C3H9N
24	190	1567	2-mercaptoethanol	C2H6OS
25	191	1672	3-methoxypropan-1-amine	C4H11NO
26	192	3283	ethoxyethane	C4H10O
27	193	3776	propan-2-ol	C3H8O
28	194	4685	1,4-dichlorobenzene	C6H4Cl2
29	195	5558	bromoform	CHBr3
30	196	6058	2-aminoethanethiol	C2H7NS
31	197	6213	methylsulfonylmethane	C2H6O2S
32	198	6214	1,1,1,2,2,2-hexachloroethane	C2Cl6
33	199	6278	1,1,1-trichloroethane	C2H3Cl3
34	200	6324	ethane	C2H6
35	201	6326	acetylene	C2H2
36	202	6328	iodomethane	CH3I
37	203	6334	propane	C3H8
38	204	6335	prop-1-yne	C3H4

39	205	6340	iodoethane	C2H5I
40	206	6341	ethanamine	C2H7N
41	207	6343	ethanethiol	C2H6S
42	208	6346	diiodomethane	CH2I2
43	209	6368	1,1-difluoroethane	C2H4F2
44	210	6373	fluoroform	CHF3
45	211	6384	bromo(trifluoro)methane	CBrF3
46	212	6392	chloro(trifluoro)methane	CClF3
47	213	6393	tetrafluoromethane	CF4
48	214	6403	2,2-dimethylbutane	C6H14
49	215	6431	1,1,1,2,2,2-hexafluoroethane	C2F6
50	216	6514	3-sulfanylpropanoic acid	C3H6O2S
51	217	6556	2-methylbutane	C5H12
52	218	6568	butan-2-ol	C4H10O
53	219	6569	butan-2-one	C4H8O
54	220	6573	1,1-dichloropropane	C3H6Cl2
55	221	6578	propanamide	C3H7NO
56	222	6736	3-methyl-1H-indole	C9H9N
57	223	7239	1,2-dichlorobenzene	C6H4Cl2
58	224	7288	pentan-3-one	C5H10O
59	225	7296	methylcyclopentane	C6H12
60	226	7298	cyclopentanol	C5H10O
61	227	7804	1,4-dibromobenzene	C6H4Br2
62	228	7843	butane	C4H10
63	229	7844	but-1-ene	C4H8
64	230	7846	but-1-yne	C4H6
65	231	7848	propane-1-thiol	C3H8S
66	232	7911	propanediamide	C3H6N2O2
67	233	7950	1,3,5-trichlorobenzene	C6H3Cl3
68	234	7962	methylcyclohexane	C7H14
69	235	7966	cyclohexanol	C6H12O
70	236	8008	butanenitrile	C4H7N
71	237	8013	propane-1,3-dithiol	C3H8S2
72	238	8014	N'-methylethane-1,2-diamine	C3H10N2
73	239	8016	2-(methylamino)ethanol	C3H9NO
74	240	8018	2-methoxyethanamine	C3H9NO
75	241	8019	2-methoxyethanol	C3H8O2
76	242	8025	ethyl formate	C3H6O2
77	243	8058	hexane	C6H14
78	244	8070	N,N'-dimethylethane-1,2-diamine	C4H12N2
79	245	8071	1,2-dimethoxyethane	C4H10O2
80	246	8073	propyl formate	C4H8O2
81	247	8077	(ethyldisulfanyl)ethane	C4H10S2
82	248	8078	cyclohexane	C6H12
83	249	8081	1,3,5-trioxane	C3H6O3
84	250	8082	piperidine	C5H11N

85	251	8252	prop-1-ene	C3H6
86	252	8255	2-methylprop-1-ene	C4H8
87	253	8370	1,2,3,4,5,6-hexachlorobenzene	C6Cl6
88	254	8472	trimethyl phosphite	C3H9O3P
89	255	8894	tetrahydropyran	C5H10O
90	256	9086	3-aminopropan-1-ol	C3H9NO
91	257	9261	pyrazine	C4H4N2
92	258	9264	1,3,5-trithiane	C3H6S3
93	259	9620	fluoroethane	C2H5F
94	260	9633	1,1,1,2,2-pentafluoroethane	C2HF5
95	261	9696	1,2,3,4,5-pentafluorobenzene	C6HF5
96	262	9745	1,3,5-trifluorobenzene	C6H3F3
97	263	9805	1,2,3,4,5,6-hexafluorobenzene	C6F6
98	264	9868	1,1,1-trifluoroethane	C2H3F3
99	265	9890	1,1,2-trifluoroethane	C2H3F3
100	266	9998	1-fluoropropane	C3H7F
101	267	10039	carbonyl sulfide	COS
102	268	10419	but-2-yne	C4H6
103	269	10421	dimethyldiazene	C2H6N2
104	270	10442	propane-1,3-diol	C3H8O2
105	271	10448	3-methylsulfanylpropan-1-ol	C4H10OS
106	272	10450	1,3-dioxane	C4H8O2
107	273	10451	1,3-dithiane	C4H8S2
108	274	10452	1,4-dithiane	C4H8S2
109	275	10476	carbononitridic bromide	CBrN
110	276	10477	carbononitridic chloride	CClN
111	277	10478	carbononitridic iodide	CIN
112	278	10541	trimethyl phosphate	C3H9O4P
113	279	10553	2-methylbut-2-ene	C5H10
114	280	10892	1,4-difluorobenzene	C6H4F2
115	281	10899	1-chloropropane	C3H7Cl
116	282	10903	methoxyethane	C3H8O
117	283	10926	1,1-dichlorobutane	C4H8Cl2
118	284	10943	1,3-dichlorobenzene	C6H4Cl2
119	285	11124	methyl propanoate	C4H8O2
120	286	11168	methyl thiocyanate	C2H3NS
121	287	11182	1-methoxypropane	C4H10O
122	288	11201	1,1-dibromoethane	C2H4Br2
123	289	11240	2-methylbut-1-ene	C5H10
124	290	11250	2,3-dimethylbut-2-ene	C6H12
125	291	11575	iodobenzene	C6H5I
126	292	11646	isocyanomethane	C2H3N
127	293	11723	N,N-dimethylethanamine	C4H11N
128	294	11855	1,2,3,4,5-pentachlorobenzene	C6HCl5
129	295	12014	3-methylpent-2-ene	C6H12
130	296	12021	dimethyl carbonate	C3H6O3

131	297	12022	N-ethyl-N-methylethanamine	C5H13N
132	298	12051	N,N-diethylformamide	C5H11NO
133	299	12219	N-methylethanamine	C3H9N
134	300	12220	but-2-ene	C4H8
135	301	12223	1,2-difluoroethane	C2H4F2
136	302	12224	1,2-diiodoethane	C2H4I2
137	303	12230	methylsulfanylethane	C3H8S
138	304	12243	2-methylpent-2-ene	C6H12
139	305	12251	2-methoxyacetic acid	C3H6O3
140	306	12253	N-ethylacetamide	C4H9NO
141	307	12309	pent-1-yne	C5H8
142	308	12310	pent-2-yne	C5H8
143	309	12315	N-methylpropan-1-amine	C4H11N
144	310	12319	N-ethylformamide	C3H7NO
145	311	12356	2-methoxyethyl formate	C4H8O3
146	312	12376	2-formyloxyethyl formate	C4H6O4
147	313	12418	1,1,1,2-tetrachloroethane	C2H2Cl4
148	314	12585	pent-2-ene	C5H10
149	315	12703	N,N-diethylacetamide	C6H13NO
150	316	12753	1-oxidopyridin-1-ium	C5H5NO
151	317	12961	2-sulfanylacetamide	C2H5NOS
152	318	12965	N,N-dimethylpropanamide	C5H11NO
153	319	13129	1,1,1,2-tetrafluoroethane	C2H2F4
154	320	13130	methyl dihydrogen phosphate	CH5O4P
155	321	13134	dimethyl hydrogen phosphate	C2H7O4P
156	322	13195	5-methyl-1H-imidazole	C4H6N2
157	323	14470	N-methylpropanamide	C4H9NO
158	324	14818	3,5,7-trithia-1,2,4,6-tetraphosphatricyclo[2.2.1.0 _{2,6}]heptane	P4S3
159	325	14917	fluorane	FH
160	326	15367	thiane	C5H10S
161	327	15380	bis(methylsulfanyl)methane	C3H8S2
162	328	16592	1-(methyldisulfanyl)propane	C4H10S2
163	329	16843	trifluoro(iodo)methane	CF3I
164	330	16908	1-fluorobutane	C4H9F
165	331	19754	1-methylsulfanylpropane	C4H10S
166	332	20970	1,3-dithiolane	C3H6S2
167	333	22686	N-propylformamide	C4H9NO
168	334	23110	1,2-bis(methylsulfanyl)ethane	C4H10S2
169	336	24387	trichlorophosphane	Cl3P
170	337	24404	phosphane	H3P
171	338	24408	molecular bromine	Br2
172	339	24524	molecular fluorine	F2
173	340	24526	molecular chlorine	Cl2
174	341	24622	1,1,1-trichloropropane	C3H5Cl3
175	342	24807	chlorosulfanyl thiohypochlorite	Cl2S2

176	343	24813	phosphoryl trichloride	Cl3OP
177	344	24841	iodane	HI
178	345	25403	2,3-dimethylpent-2-ene	C7H14
179	346	31242	4-ethylphenol	C8H10O
180	347	31275	1,4-dioxane	C4H8O2
181	348	33629	hex-2-yne	C6H10
182	349	61236	N,N-dimethylpropan-1-amine	C5H13N
183	350	62540	methoxymethanol	C2H6O2
184	351	66978	N,N'-dimethylpropane-1,3-diamine	C5H14N2
185	352	67515	1,2-thiazole	C3H3NS
186	353	67899	1,1,1-trifluoropropane	C3H5F3
187	354	68152	3-hydroxypropanoic acid	C3H6O3
188	355	68980	1,1-diiodoethane	C2H4I2
189	356	68983	trimethylphosphane	C3H9P
190	357	69020	2-aminoacetamide	C2H6N2O
191	358	69021	2-hydroxyacetamide	C2H5NO2
192	359	69404	2-hydroxyethyl formate	C3H6O3
193	360	69657	N-(2-hydroxyethyl)formamide	C3H7NO2
194	361	69811	3-sulfanylpropanamide	C3H7NOS
195	362	70075	nitrosomethane	CH3NO
196	363	74116	3-methoxypropan-1-ol	C4H10O2
197	364	75367	3-amino-3-oxopropanoic acid	C3H5NO3
198	365	75551	2-methylsulfanylacetic acid	C3H6O2S
199	366	75606	2-formamidoacetic acid	C3H5NO3
200	367	75891	3-(methylamino)propanoic acid	C4H9NO2
201	368	77743	3-methylsulfanylpropan-1-amine	C4H11NS
202	369	78925	2-methylsulfanylethanol	C3H8OS
203	370	79045	dithiolane	C3H6S2
204	371	79079	azidomethane	CH3N3
205	372	80511	N'-methylpropane-1,3-diamine	C4H12N2
206	373	82641	ethyl dimethyl phosphate	C4H11O4P
207	374	83297	1,1,1-trichlorobutane	C4H7Cl3
208	375	87697	2-methylsulfanylethanamine	C3H9NS
209	376	88211	3-sulfanylpropan-1-ol	C3H8OS
210	377	89675	2-amino-N-methylacetamide	C3H8N2O
211	378	94671	disulfanylethane	C2H6S2
212	379	97436	3-aminopropane-1-thiol	C3H9NS
213	380	108196	hydrogen disulfide	H2S2
214	381	122370	methylsulfanylmethanethiol	C2H6S2
215	382	123046	phosphinine	C5H5P
216	384	123323	fluorosulfanyl thiohypofluorite	F2S2
217	385	123388	(methyldisulfanyl)ethane	C3H8S2
218	386	134442	3-methoxypropanoic acid	C4H8O3
219	387	136335	dithiane	C4H8S2
220	388	136492	methylperoxymethane	C2H6O2
221	389	136869	dicyanophosphanylformonitrile	C3N3P

222	390	137036	carbononitridic fluoride	CFN
223	392	137201	methyl cyanate	C2H3NO
224	393	138210	methoxymethyl formate	C3H6O3
225	394	138743	N-methylpropan-2-imine	C4H9N
226	395	138769	methyl(methylidene)phosphane	C2H5P
227	396	139605	fluoro thiohypofluorite	F2S
228	397	139636	difluorophosphane	F2HP
229	399	140060	2-methoxyacetamide	C3H7NO2
230	400	140180	1,3-dimethoxypropane	C5H12O2
231	401	141161	1,3-bis(methylsulfanyl)propane	C5H12S2
232	402	141892	2-acetamido-N-methylpropanamide	C6H12N2O2
233	403	151411	2-(methylamino)acetamide	C3H8N2O
234	404	170607	3-(methylamino)-3-oxopropanoic acid	C4H7NO3
235	405	192755	N,N'-dimethylpropanediamide	C5H10N2O2
236	406	192802	3-aminopropanamide	C3H8N2O
237	407	202285	2-(methylamino)ethanethiol	C3H9NS
238	408	223579	2-hydroxy-N-methylacetamide	C3H7NO2
239	409	226108	N-(2-formamidoethyl)formamide	C4H8N2O2
240	410	232267	2-formamidoacetamide	C3H6N2O2
241	411	263087	3-methylsulfanylpropanamide	C4H9NOS
242	412	300977	2-methoxy-N-methylethanamine	C4H11NO
243	413	324305	3-ethyl-1H-indole	C10H11N
244	414	350667	N-ethyl-N-methylformamide	C4H9NO
245	415	439506	2-acetamido-N-methylacetamide	C5H10N2O2
246	416	521081	difluoromethanethione	CF2S
247	417	521324	N-ethylpropanamide	C5H11NO
248	418	522059	disulfanylmethane	CH4S2
249	419	524894	methoxymethoxyethane	C4H10O2
250	420	525376	1-(methylsulfanylmethylsulfanyl)propane	C5H12S2
251	421	525377	methylsulfanylmethylsulfanylethane	C4H10S2
252	422	525458	2-methylsulfanylethanethiol	C3H8S2
253	423	525488	3-methylsulfanylpropane-1-thiol	C4H10S2
254	424	1174	uracil	C4H4N2O2
255	425	533889	3-formamidopropanoic acid	C4H7NO3
256	426	547873	2-methylsulfanylacetamide	C3H7NOS
257	427	641811	thioacetone	C3H6S
258	428	642906	N-methyl-3-sulfanylpropanamide	C4H9NOS
259	429	2463138	N-methyl-2-(methylamino)acetamide	C4H10N2O
260	430	3014644	2-methoxyethanethiol	C3H8OS
261	431	4047279	N-methyl-2-sulfanylacetamide	C3H7NOS
262	432	4145140	3-hydroxypropanamide	C3H7NO2
263	433	4431608	1-methoxy-2-methylsulfanylethane	C4H10OS
264	434	5252481	3-methoxypropanamide	C4H9NO2
265	435	6428842	1-(disulfanyl)propane	C3H8S2
266	436	9940735	methoxymethanethiol	C2H6OS
267	437	10011858	2-sulfanylethyl formate	C3H6O2S

268	438	10148986	3-(methylamino)propan-1-ol	C4H11NO
269	439	10153736	5-ethyl-1H-imidazole	C5H8N2
270	440	10877157	3-formyloxypropyl formate	C5H8O4
271	441	12056230	fluorophosphane	FH2P
272	442	12387217	3-hydroxypropyl formate	C4H8O3
273	443	12545136	1,1-difluoropropane	C3H6F2
274	444	12545984	3-methoxy-N-methylpropan-1-amine	C5H13NO
275	445	12548977	N-methyl-3-(methylamino)propanamide	C5H12N2O
276	446	12599323	1-(methoxymethoxy)propane	C5H12O2
277	447	12634510	3-(methylamino)propane-1-thiol	C4H11NS
278	448	12923964	methylsulfanylmethanamine	C2H7NS
279	449	13048467	3-formyloxypropanoic acid	C4H6O4
280	450	13050065	N-(methoxymethyl)formamide	C3H7NO2
281	451	13050070	N-(methylsulfanylmethyl)formamide	C3H7NOS
282	452	13229773	N-(2-sulfanylethyl)formamide	C3H7NOS
283	453	13319001	ethyl methyl hydrogen phosphate	C3H9O4P
284	454	13390905	N-ethyl-N-methylacetamide	C5H11NO
285	455	13553138	N-methyl-2-methylsulfanylacetamide	C4H9NOS
286	456	13561311	N-(2-aminoethyl)formamide	C3H8N2O
287	457	13586013	N-(3-hydroxypropyl)formamide	C4H9NO2
288	458	13637593	methylsulfanylmethanol	C2H6OS
289	459	13682918	thiophen-3-one	C4H4OS
290	460	14481877	2-formamido-N-methylacetamide	C4H8N2O2
291	461	14510872	N'-methylpropanediamide	C4H8N2O2
292	462	14512801	1,1,1-trifluorobutane	C4H7F3
293	463	14872190	2-methoxy-N-methylacetamide	C4H9NO2
294	464	14889074	3-amino-N-methylpropanamide	C4H10N2O
295	465	15089697	hydroxymethyl formate	C2H4O3
296	466	15561472	3-hydroxy-N-methylpropanamide	C4H9NO2
297	467	15678214	methoxy(methylsulfanyl)methane	C3H8OS
298	468	17764882	3-formamidopropanamide	C4H8N2O2
299	469	17778177	methoxymethanamine	C2H7NO
300	470	17932045	methylaminomethyl formate	C3H7NO2
301	471	18178029	N-methyl-3-methylsulfanylpropan-1-amine	C5H13NS
302	472	18387020	N-ethyl-N-methylpropanamide	C6H13NO
303	473	18445434	N-[2-(methylamino)ethyl]formamide	C4H10N2O
304	474	18670831	N-methyl-2-methylsulfanylethanamine	C4H11NS
305	475	18967886	1,1-difluorobutane	C4H8F2
306	476	19017369	formyloxymethyl formate	C3H4O4
307	477	19348483	N-(2-methoxyethyl)formamide	C4H9NO2
308	478	19762762	2-formamidoethyl formate	C4H7NO3
309	479	20025584	3-methoxypropyl formate	C5H10O3
310	480	20396353	N-(3-methylsulfanylpropyl)formamide	C5H11NOS
311	481	20481374	methylsulfanylmethyl formate	C3H6O2S
312	482	20652631	1-methoxy-3-methylsulfanylpropane	C5H12OS
313	483	20979435	1-methoxy-N-methylmethanamine	C3H9NO

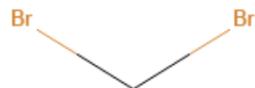
314	484	21258259	[2-(methylamino)-2-oxoethyl] formate	C4H7NO3
315	485	21258280	2-aminoethyl formate	C3H7NO2
316	486	21430458	3-methoxy-N-methylpropanamide	C5H11NO2
317	487	21444299	N-methyl-1-methylsulfanylmethanamine	C3H9NS
318	488	21711028	3-methoxypropane-1-thiol	C4H10OS
319	489	21878697	1,2-thiazolidin-3-one	C3H5NOS
320	490	22182162	3-(methylamino)propanamide	C4H10N2O
321	491	22323463	aminomethyl formate	C2H5NO2
322	492	22928613	2-formyloxyacetic acid	C3H4O4
323	493	22928648	(3-amino-3-oxopropyl) formate	C4H7NO3
324	494	22980384	methyl(propan-2-ylidene)phosphane	C4H9P
325	495	23187733	2-methylsulfanylethyl formate	C4H8O2S
326	496	23515101	3-formamido-N-methylpropanamide	C5H10N2O2
327	497	53671850	(2-amino-2-oxoethyl) formate	C3H5NO3
328	498	54014224	3-aminopropyl formate	C4H9NO2
329	499	54311466	2-(methylamino)ethyl formate	C4H9NO2
330	500	54336311	3-methylsulfanylpropyl formate	C5H10O2S
331	501	55286195	3-(methylamino)propyl formate	C5H11NO2
332	502	57222571	sulfanylmethyl formate	C2H4O2S
333	503	57305422	3-sulfanylpropyl formate	C4H8O2S
334	504	90984882	formamidomethyl formate	C3H5NO3
335	506	-15	O=COCCC(=O)NC	O=COCCC(=O)NC
336	507	-14	Cc1c[nH]cn1	Cc1c[nH]cn1
337	508	-13	CCc1c[nH]cn1	CCc1c[nH]cn1
338	509	-10	fluorobromocyanophosphine	N#CP(F)Br
339	512	-5	bis(difluorophosphanyl)thioether	FP(F)SP(F)F

Appendix E

Results for molecules selected from Data set 1 (Appendix D, Table D.1)

#3 Dibromomethane CH₂Br₂ CID: 3024

ref molpol	-10.86	-7.43	-6.78, avg	-8.36
molpol	10.68	7.54	7.28, avg	8.50
rms molpol	0.18	0.11	0.50, avg	0.14



Monomer potential fitting RMS: 0.24

##Dimer results - Fitting to QM datasets##

DESRES_3-water, energy values in kcal/mol

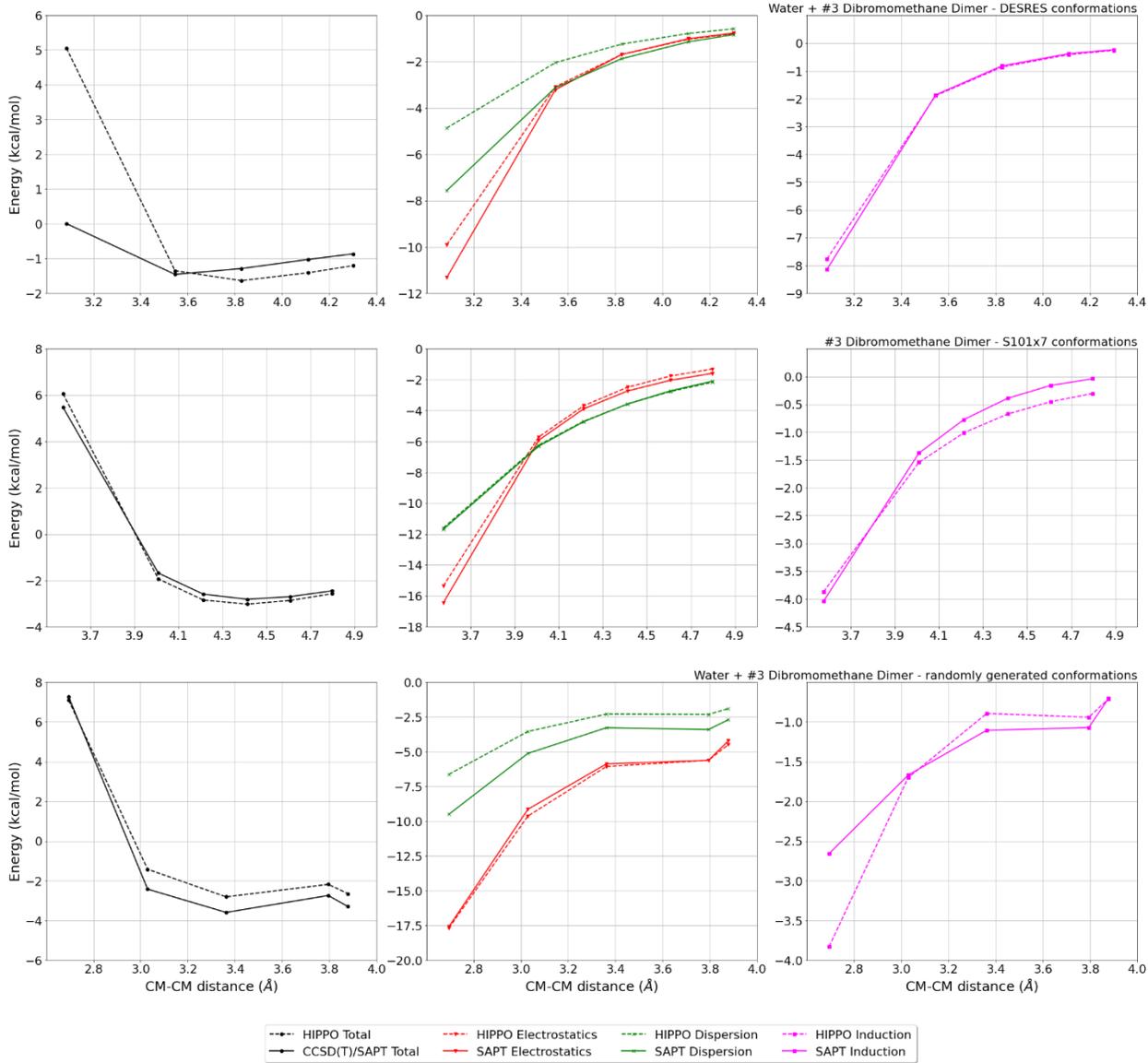
MAE	Std error	max error	#points	#count[err > 1]
0.185	0.356	2.9694	288	10

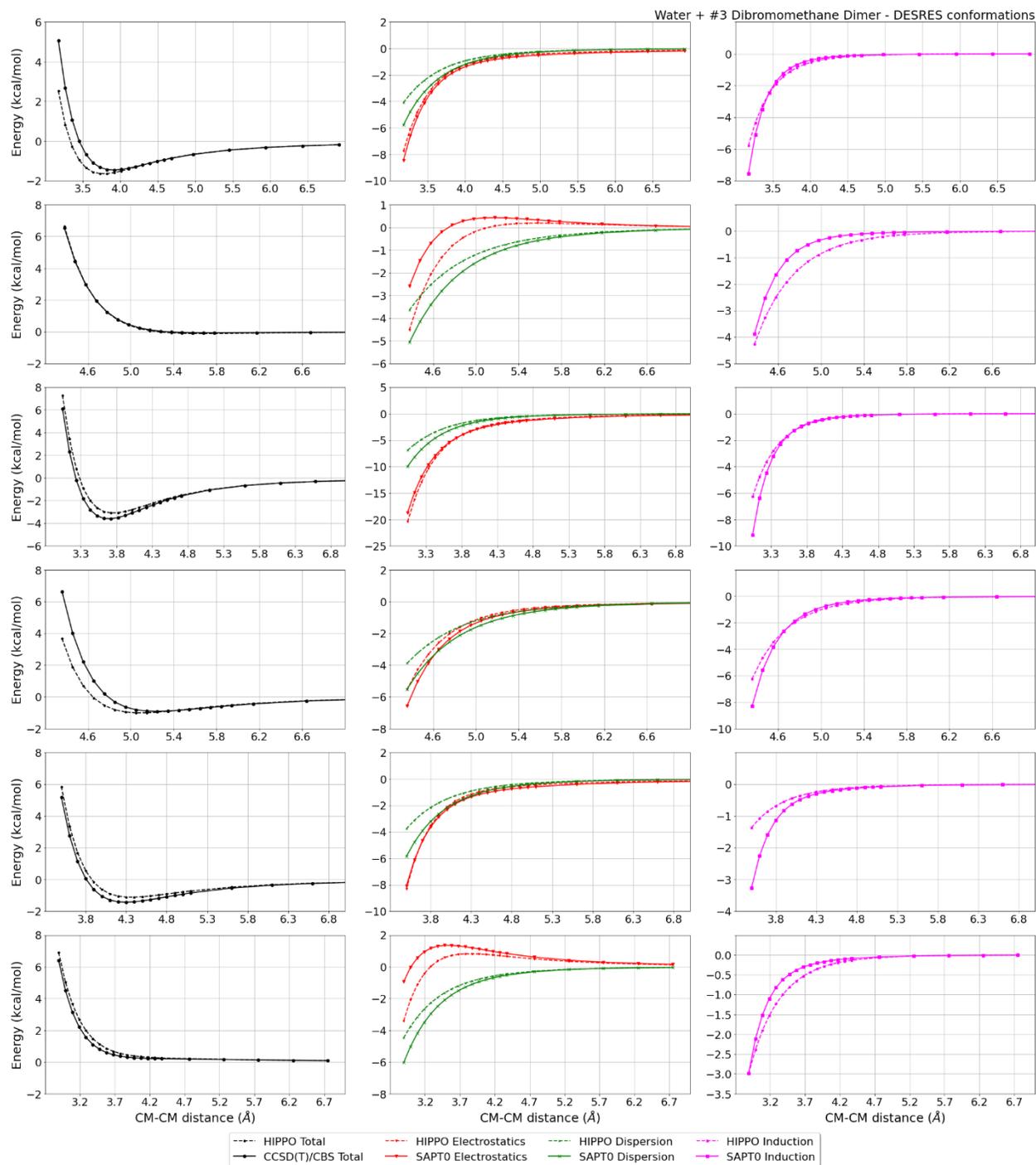
DESRES_3-3, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.447	0.334	1.3912	26	2

Liquid Dibromomethane @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
2510.81	2490.70	0.8	38.07	37.45	1.7	6.85	7.23	-5.3	74





#7 Bromomethane CH3Br CID: 6323



```
ref molpol  -6.54  -4.72  -4.72, avg  -5.33
  molpol     5.99   4.86   4.86, avg   5.23
rms molpol   0.55   0.14   0.14, avg   0.09
```

Monomer potential fitting RMS: 0.17

##Dimer results - Fitting to QM datasets##

DESRES_7-7, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.282	0.312	1.4662	24	1

R739x5_7-7, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
1.699	1.790	3.249	1.4592
1.781	0.373	1.257	0.8835
1.863	-0.372	0.159	0.5313
1.945	-0.724	-0.408	0.3157
2.110	-0.865	-0.760	0.1047

MAE	Std error	max error	#points	#count[err > 1]
0.659	0.476	1.4592	5	1

HB300SPXx10_7-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
2.371	1.008	2.312	1.3042
2.481	-1.614	-0.462	1.1518
2.592	-2.988	-2.014	0.9743
2.703	-3.588	-2.789	0.7992
2.814	-3.726	-3.085	0.6415
2.925	-3.602	-3.096	0.5063
3.036	-3.343	-2.948	0.3951
3.369	-2.377	-2.198	0.1786
3.924	-1.225	-1.178	0.0470
5.034	-0.398	-0.392	0.0057

MAE	Std error	max error	#points	#count[err > 1]
0.600	0.432	1.3042	10	2

DESRES_7-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.167	0.482	5.2953	294	11

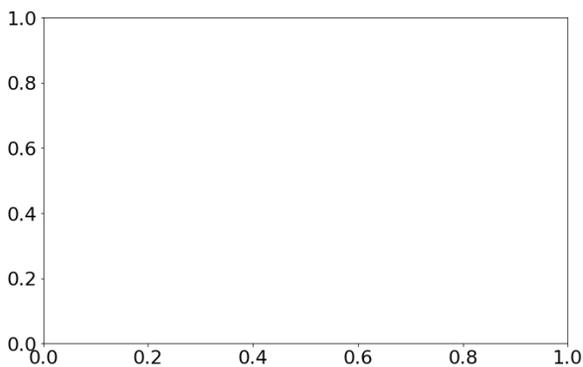
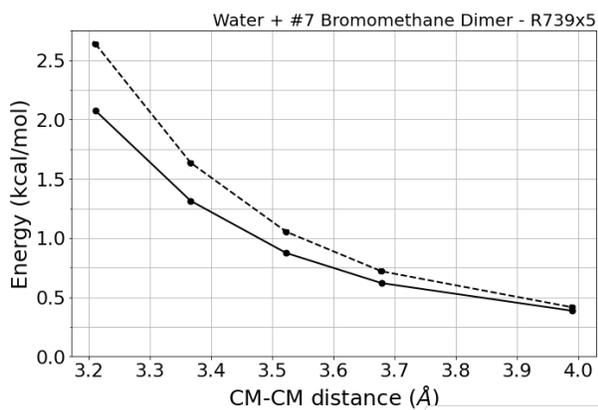
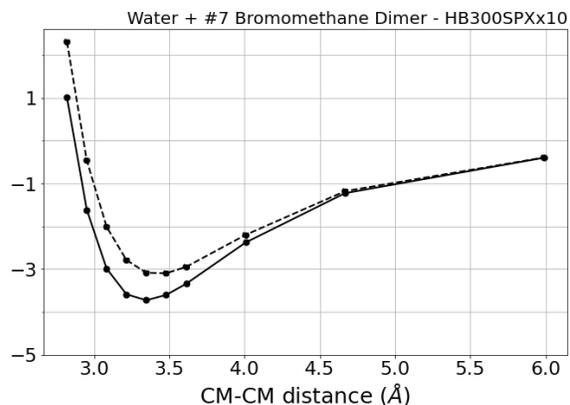
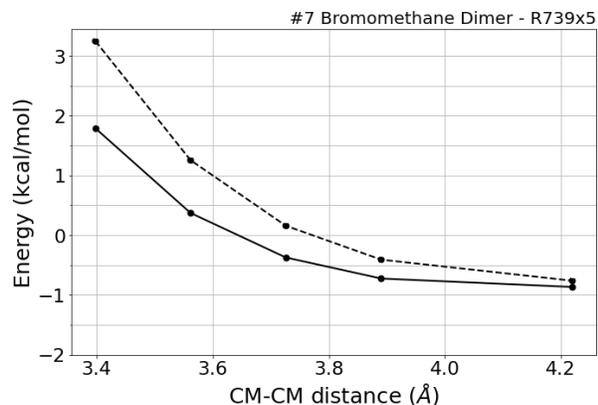
R739x5_7-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
2.699	2.075	2.638	0.5632
2.830	1.314	1.635	0.3214
2.961	0.874	1.054	0.1796
3.092	0.620	0.720	0.0997
3.354	0.386	0.415	0.0293

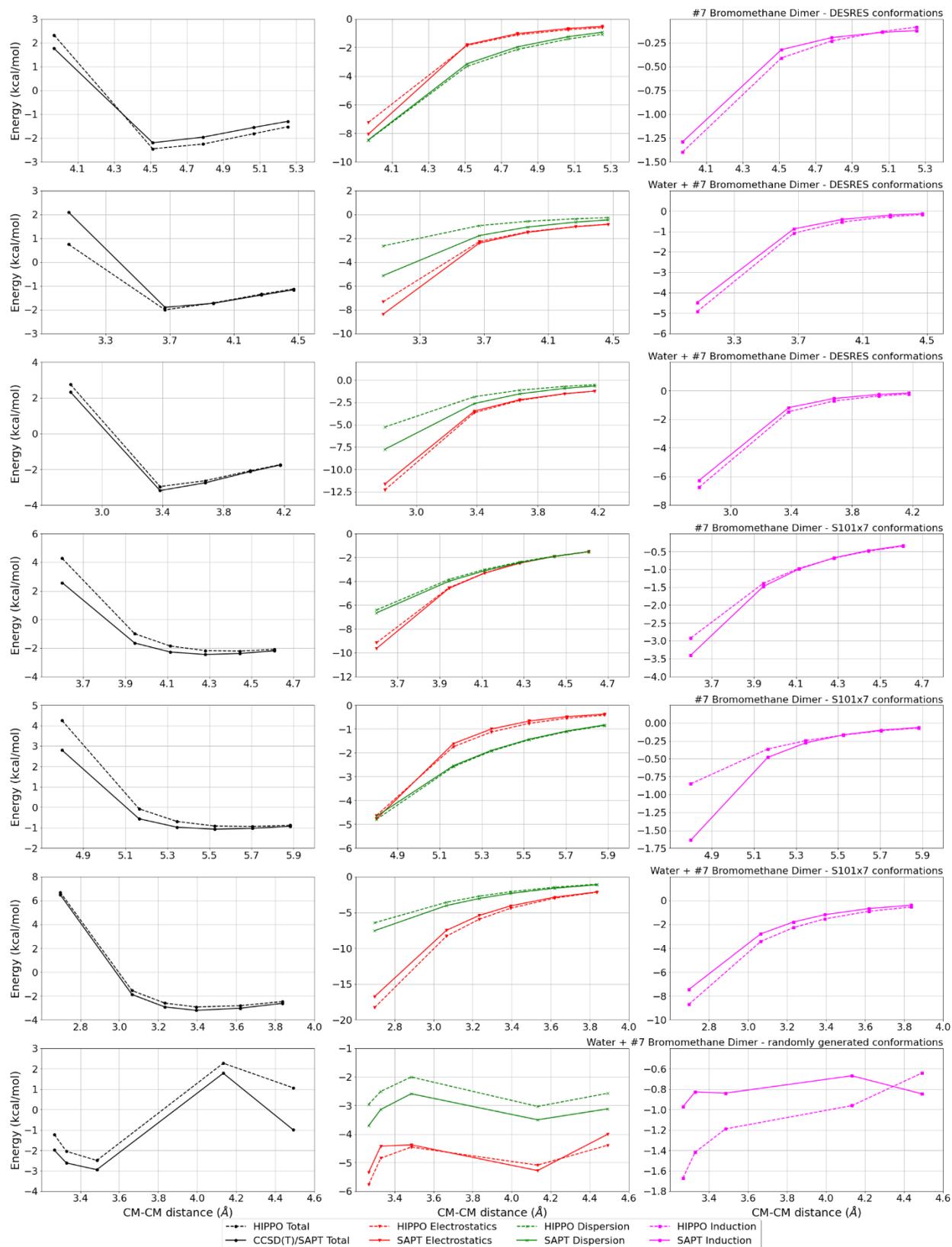
MAE	Std error	max error	#points	#count[err > 1]
0.239	0.189	0.5632	5	0

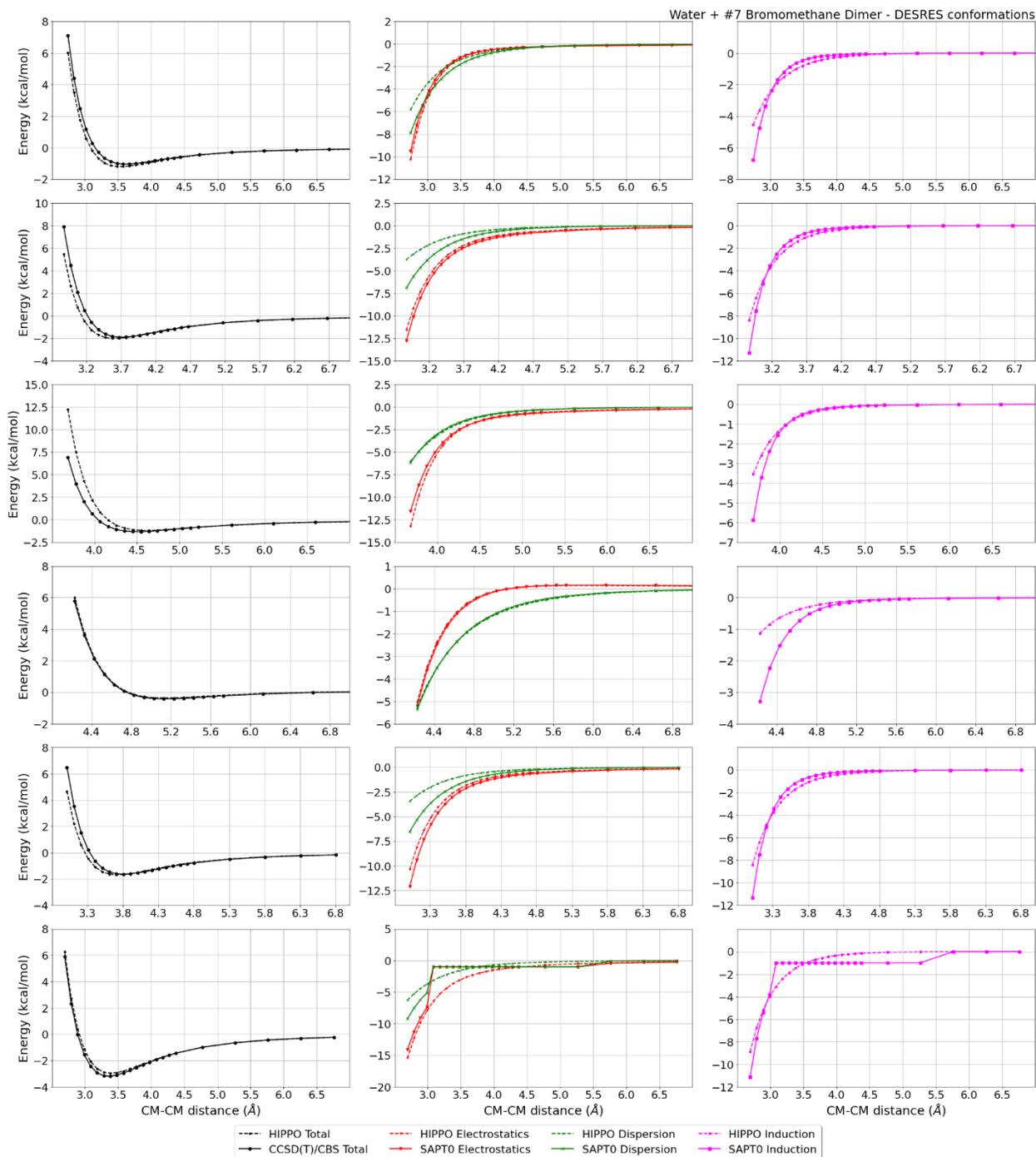
Liquid Bromomethane @ 276.65 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1734.92	1721.95	0.8	24.44	-1.00	0.0	7.49	-1.00	0.0	1000



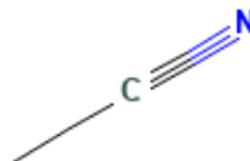
—●— CCSD(T) Total - - - ■ - - - HIPPO Total





#15 Acetonitrile C2H3N CID: 6342

ref molpol	-5.72	-3.55	-3.55, avg	-4.27
molpol	5.95	4.00	4.00, avg	4.65
rms molpol	0.23	0.45	0.45, avg	0.38



Monomer potential fitting RMS: 0.22

##Dimer results - Fitting to QM datasets##

DESRES_15-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.189	0.370	2.7454	287	14

HB375x10_15-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
1.788	-1.189	-1.635	-0.4465
1.868	-3.248	-3.262	-0.0139
1.947	-4.335	-4.108	0.2269
2.028	-4.800	-4.461	0.3393
2.108	-4.879	-4.508	0.3708
2.188	-4.730	-4.374	0.3563
2.269	-4.454	-4.137	0.3170
2.512	-3.414	-3.240	0.1740
2.919	-2.036	-1.998	0.0384
3.738	-0.824	-0.832	-0.0082

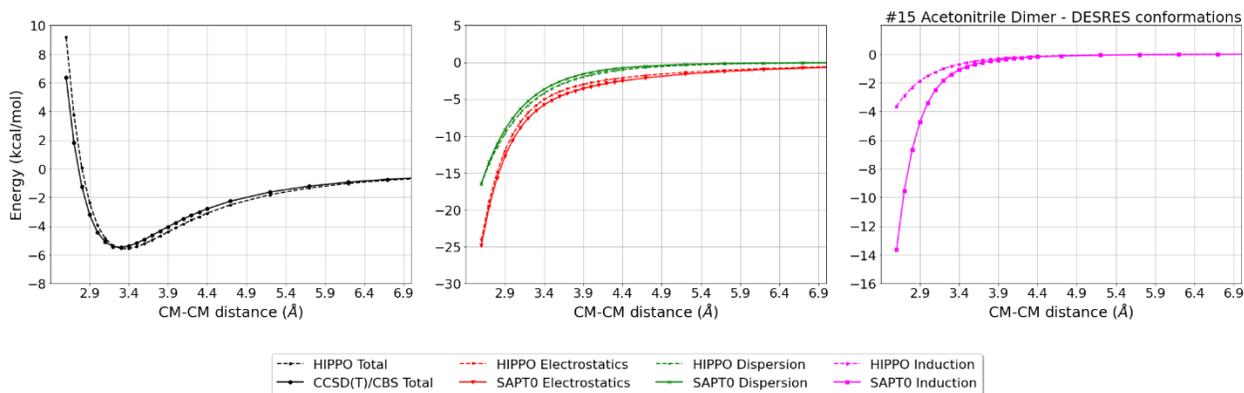
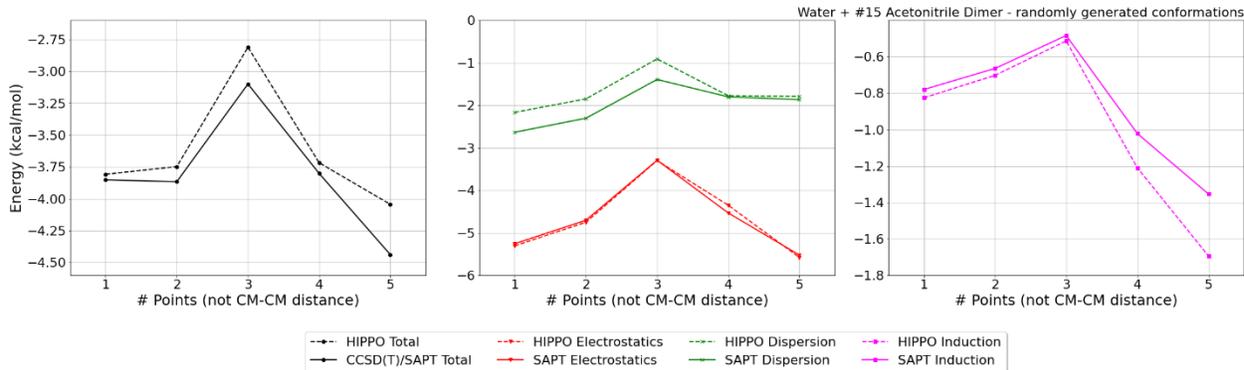
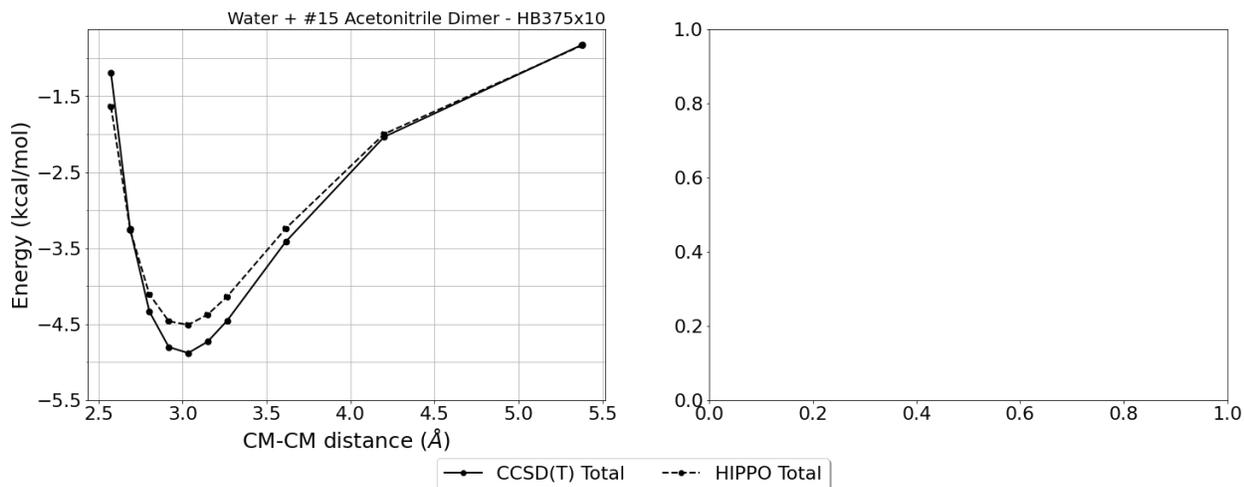
MAE	Std error	max error	#points	#count[err > 1]
0.229	0.154	0.4465	10	0

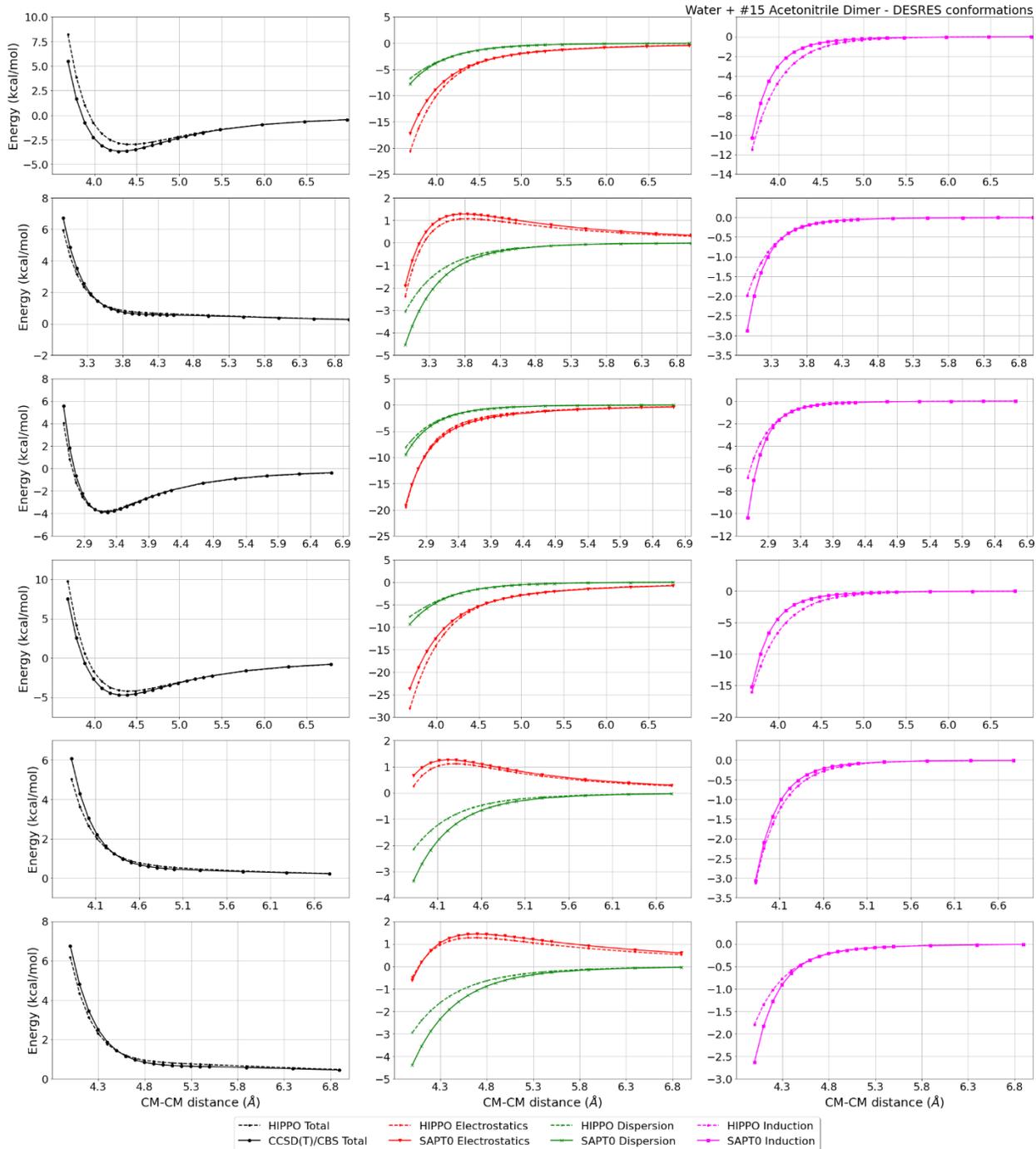
DESRES_15-15, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.479	0.627	2.8131	25	3

Liquid Acetonitrile @ 298.15 K

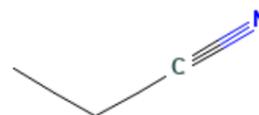
Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
779.82	776.00	0.5	34.84	33.23	4.9	32.77	35.69	-8.2	20000





#36 Propanenitrile C3H5N CID: 7854

ref molpol	-7.62	-5.46	-5.03, avg	-6.04
molpol	7.21	5.12	4.77, avg	5.70
rms molpol	0.42	0.34	0.26, avg	0.34



Monomer potential fitting RMS: 0.47

##Dimer results - Fitting to QM datasets##

DESRES_36-36, energy values in kcal/mol

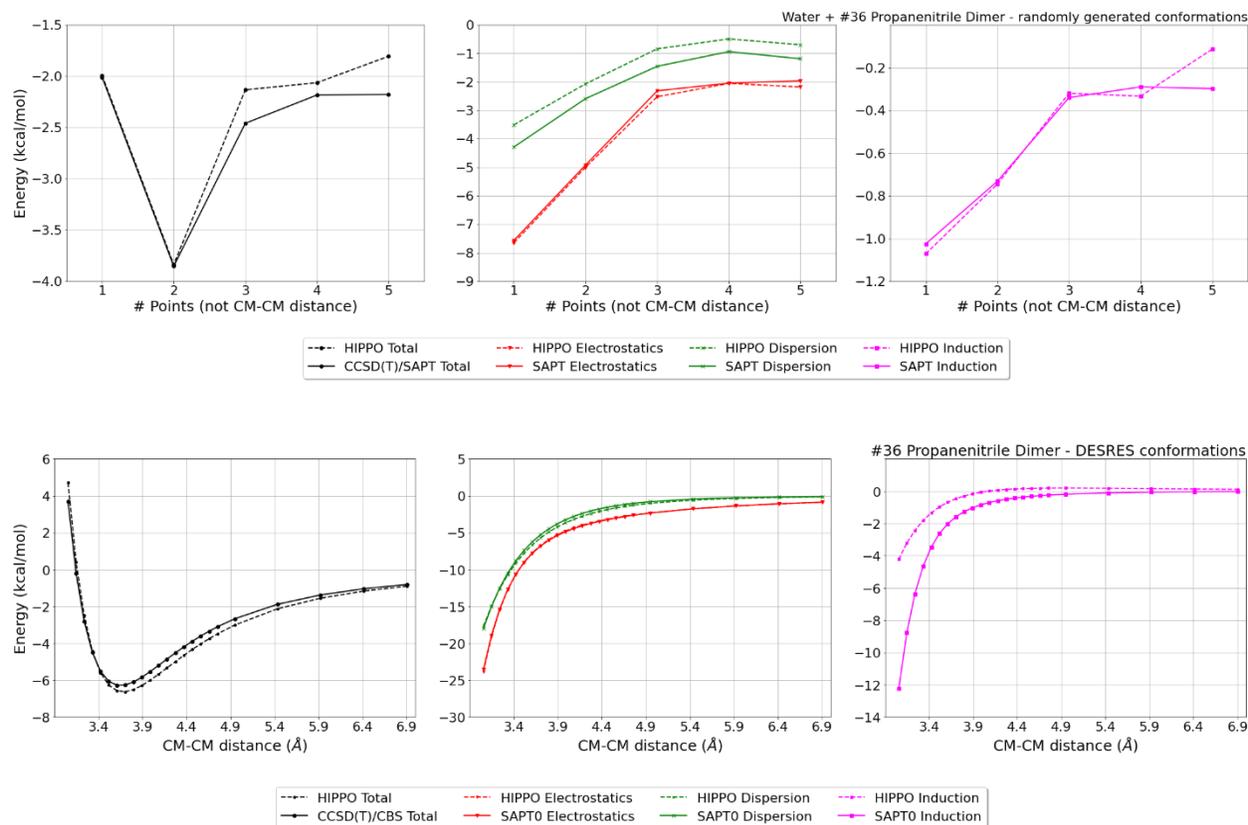
MAE	Std error	max error	#points	#count[err > 1]
0.551	0.239	0.8144	25	0

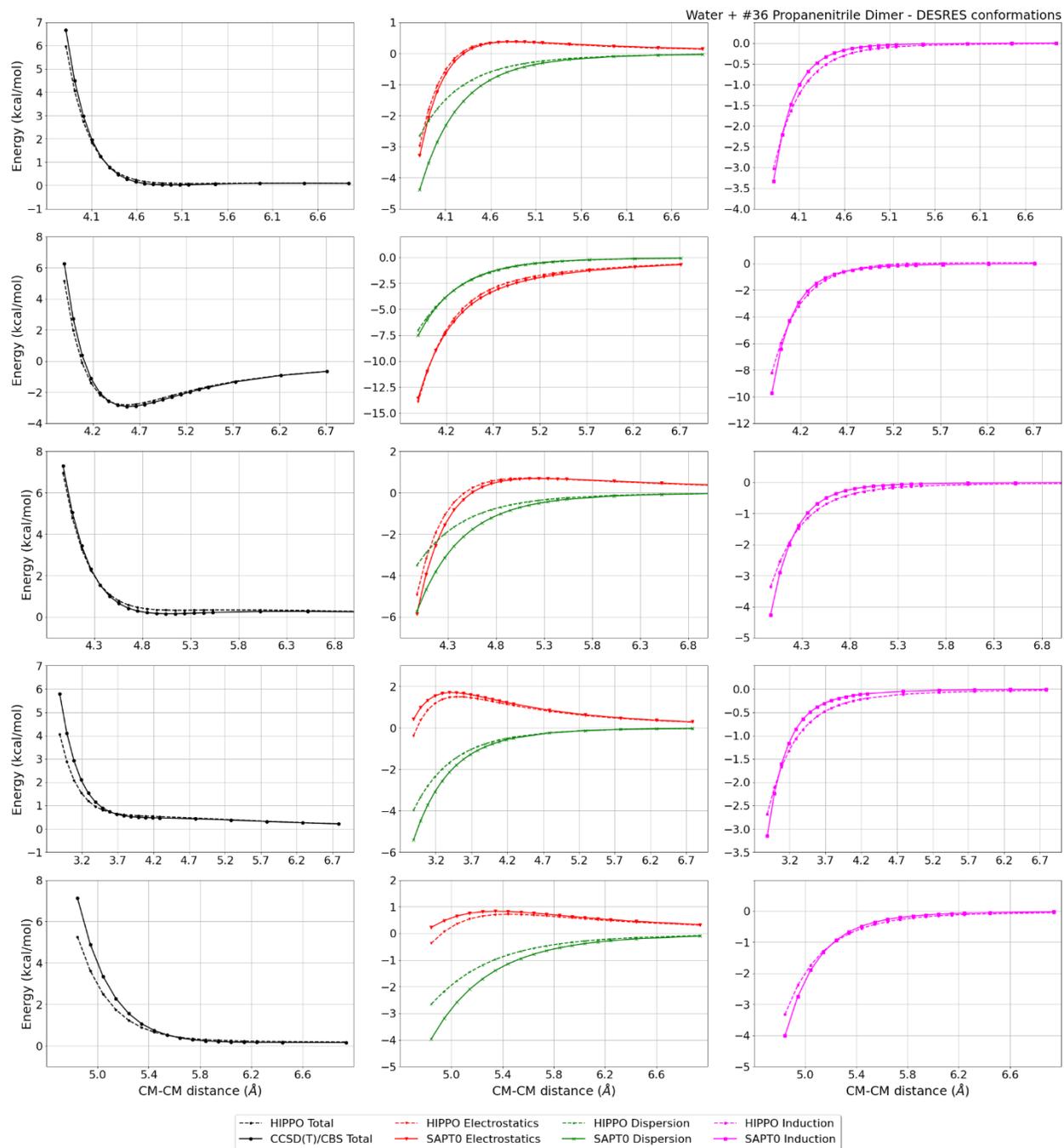
DESRES_36-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.143	0.238	1.8291	261	6

Liquid Propanenitrile @ 298.15 K

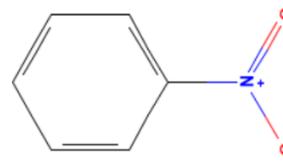
Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
781.50	776.40	0.7	36.33	36.03	0.8	24.10	29.32	-17.8	20000





#100 Nitrobenzene C6H5NO2 CID: 7416

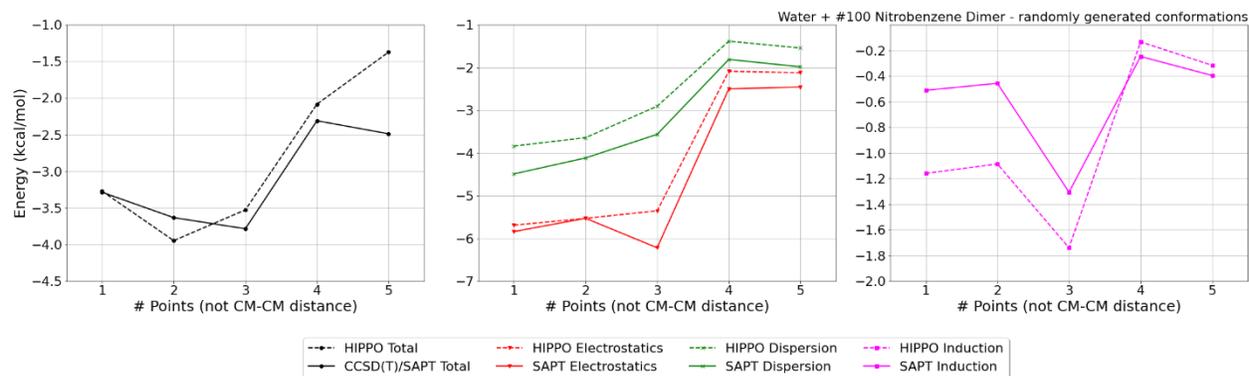
ref molpol	-16.70	-14.37	-7.39, avg	-12.82
molpol	18.03	15.32	7.74, avg	13.70
rms molpol	1.33	0.95	0.35, avg	0.87



Monomer potential fitting RMS: 0.39

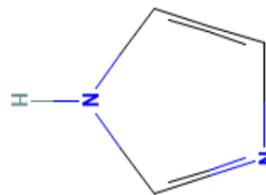
Liquid Nitrobenzene @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1134.82	1198.70	-5.3	57.92	55.01	5.3	16.39	34.81	-52.9	1000



#154 Imidazole C3H4N2 CID: 795

ref molpol	-8.49	-8.02	-5.04, avg	-7.18
molpol	8.51	8.00	5.02, avg	7.18
rms molpol	0.02	0.02	0.02, avg	0.00



Monomer potential fitting RMS: 1.00

##Dimer results - Fitting to QM datasets##

DESRES_154-154, energy values in kcal/mol

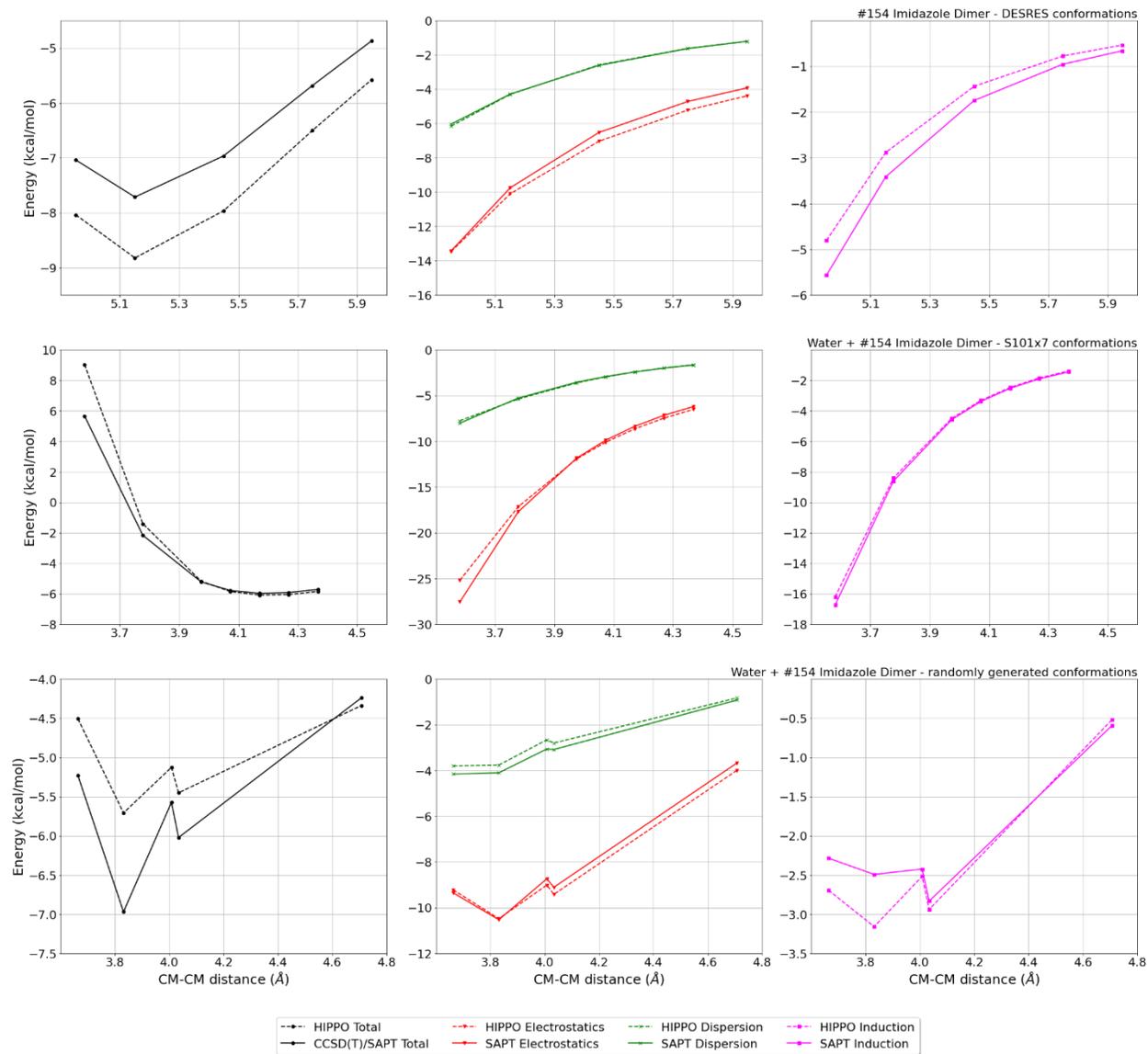
MAE	Std error	max error	#points	#count[err > 1]
0.404	0.511	3.7592	730	88

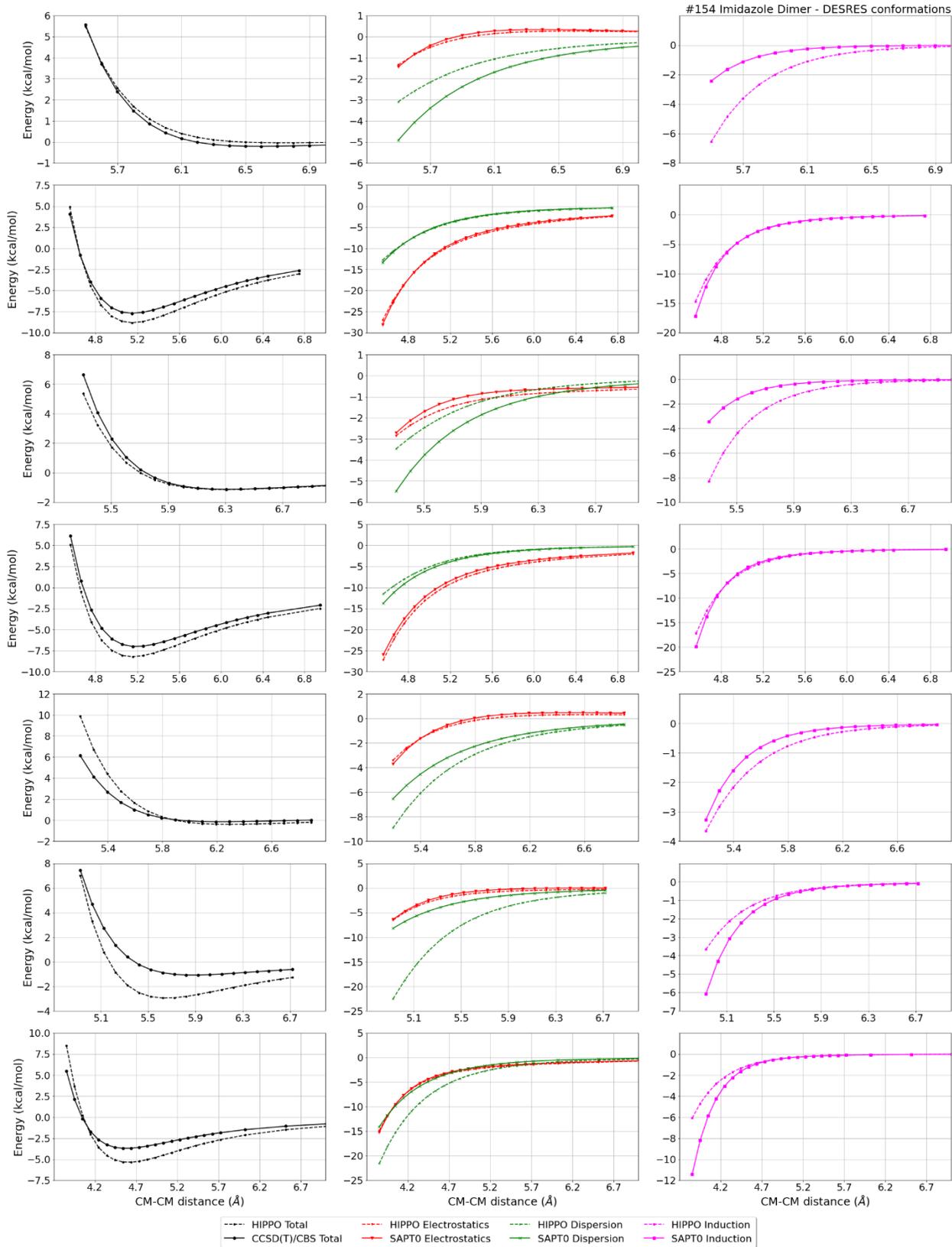
DESRES_154-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.219	0.410	3.8813	559	19

Liquid Imidazole @ 374.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1030.48	1030.30	0.0	72.91	-1.00	0.0	0.00	-1.00	0.0	17





#157 Methyl chloride CH3Cl CID: 6327

ref molpol	-5.26	-3.82	-3.82, avg	-4.30
molpol	4.88	3.95	3.95, avg	4.26
rms molpol	0.38	0.13	0.13, avg	0.04



Monomer potential fitting RMS: 0.24

##Dimer results - Fitting to QM datasets##

HB300SPXx10_157-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
1.997	0.595	0.463	-0.1325		
2.089	-1.805	-1.542	0.2633		
2.181	-3.072	-2.611	0.4614		
2.272	-3.631	-3.095	0.5356		
2.364	-3.761	-3.225	0.5357		
2.455	-3.645	-3.149	0.4963		
2.547	-3.399	-2.960	0.4391		
2.822	-2.465	-2.203	0.2615		
3.280	-1.313	-1.220	0.0930		
4.196	-0.448	-0.433	0.0149		
MAE	Std error	max error	#points	#count[err > 1]	
0.323	0.186	0.5357	10	0	

DESRES_157-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.238	0.522	4.1738	289	18

DESRES_157-157, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.073	0.068	0.2066	24	0

R739x5_157-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
2.305	2.068	2.880	0.8117		
2.414	1.299	1.877	0.5782		
2.523	0.854	1.264	0.4100		
2.633	0.597	0.886	0.2892		
2.852	0.362	0.504	0.1416		
MAE	Std error	max error	#points	#count[err > 1]	
0.446	0.232	0.8117	5	0	

R739x5_157-157, energy values in kcal/mol

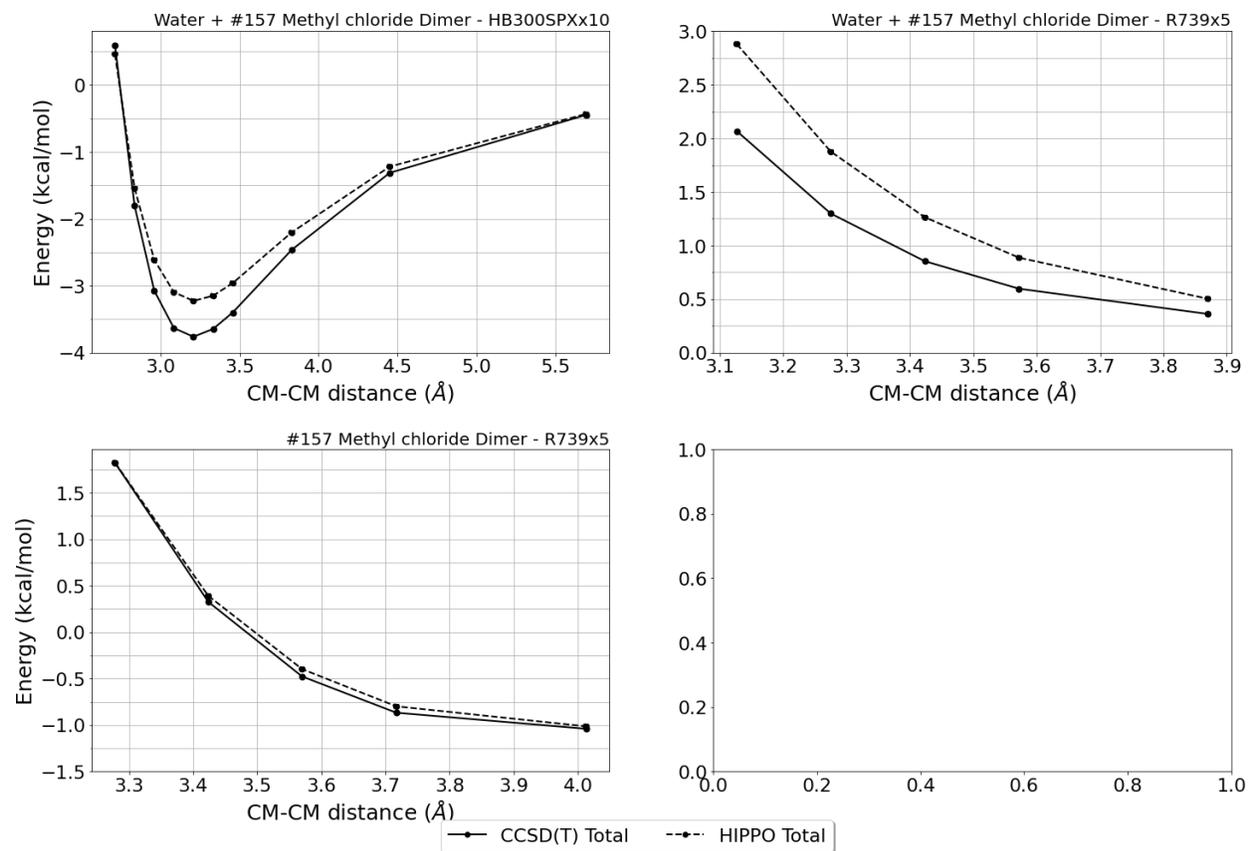
CM-CM (A)	Reference	HIPPO res	Abs diff
1.639	1.821	1.822	0.0006
1.712	0.325	0.390	0.0646

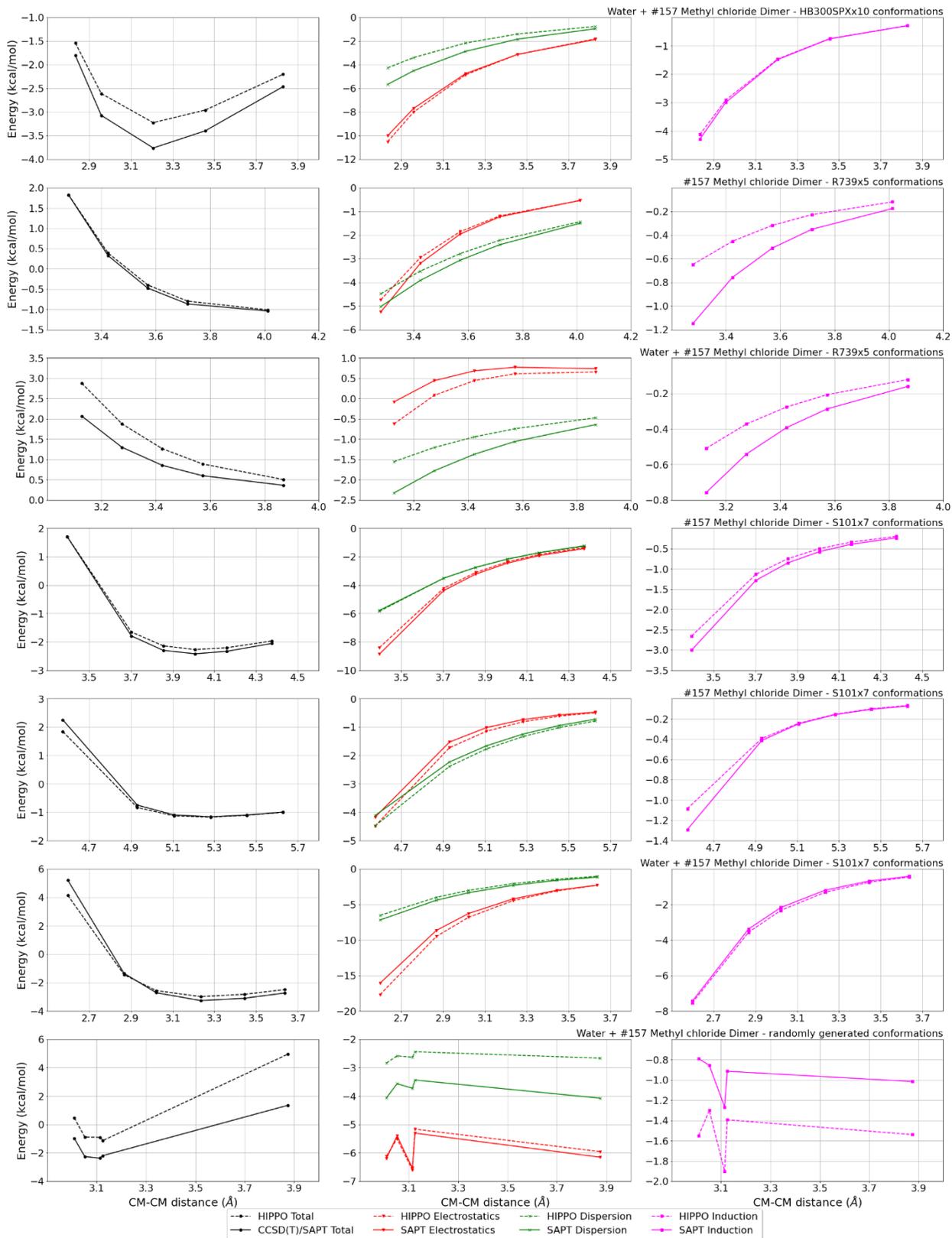
1.785	-0.477	-0.398	0.0791
1.858	-0.867	-0.798	0.0688
2.006	-1.040	-1.014	0.0263

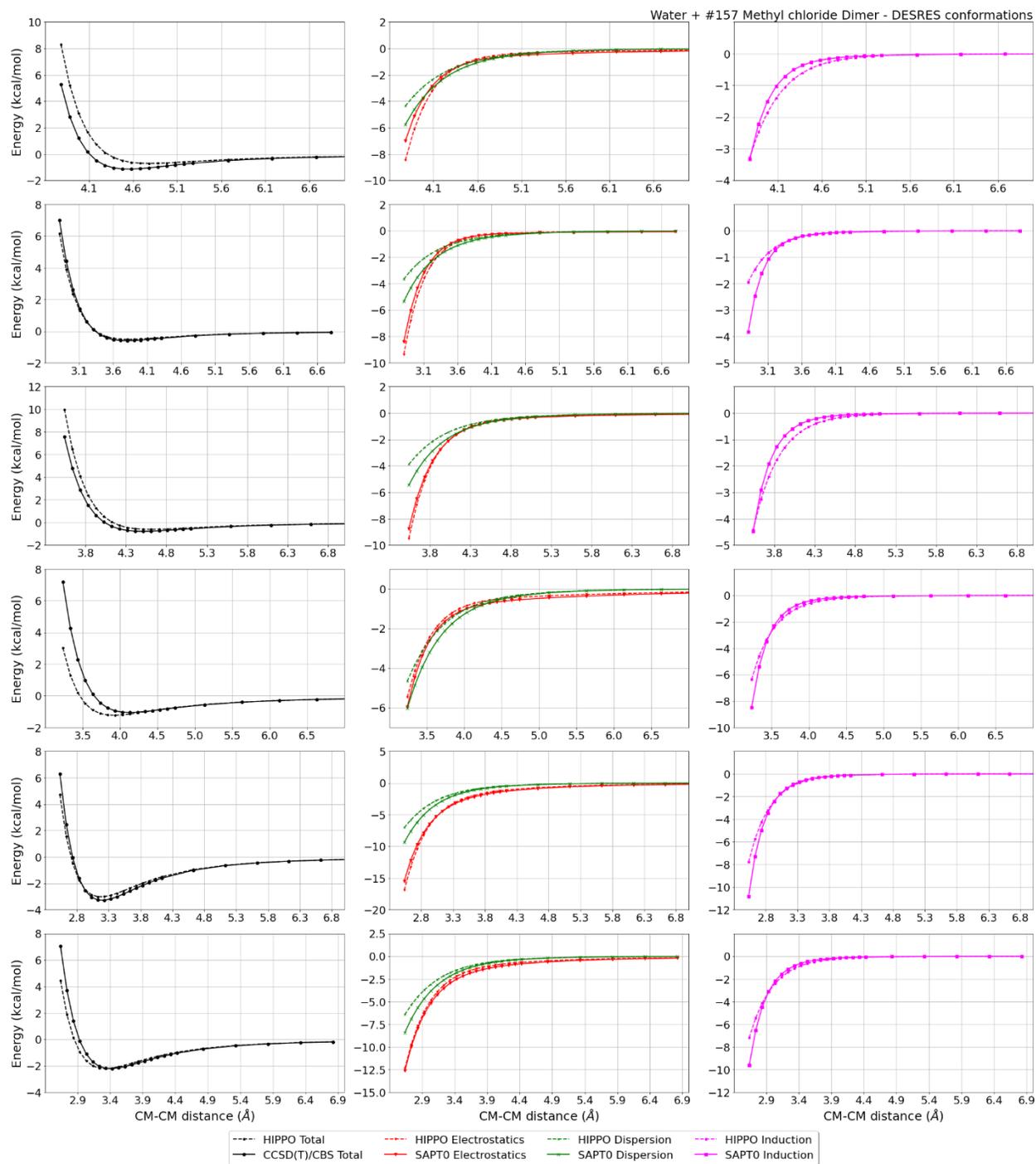
MAE	Std error	max error	#points	#count[err > 1]
0.048	0.030	0.0791	5	0

Liquid Methyl chloride @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
954.60	911.00	4.8	19.75	18.92	4.4	9.74	9.76	-0.2	3000

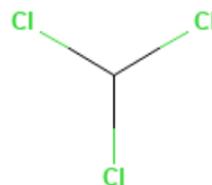






#1 Chloroform CHCl3 CID: 6212

ref molpol	-6.57	-9.09	-9.09, avg	-8.25
molpol	7.08	9.05	9.05, avg	8.39
rms molpol	0.52	0.04	0.04, avg	0.15



Monomer potential fitting RMS: 0.61

##Dimer results - Fitting to QM datasets##

DESRES_1-water, energy values in kcal/mol

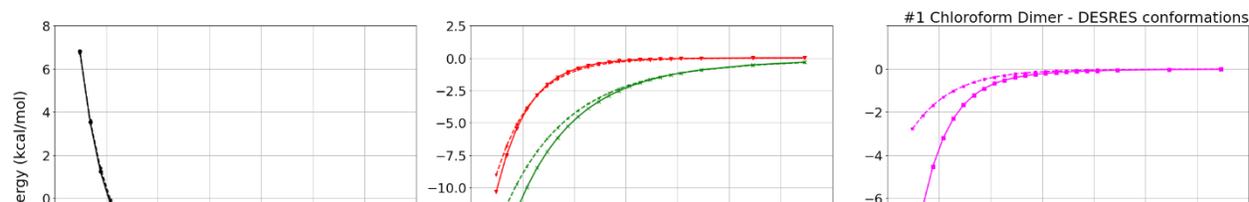
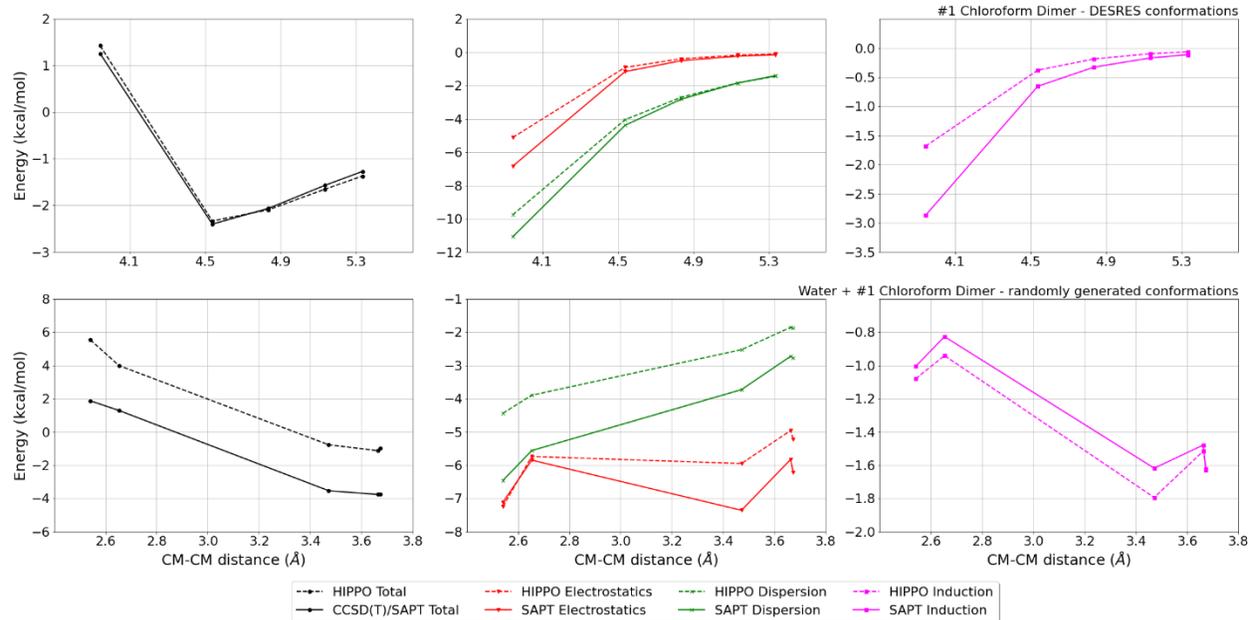
MAE	Std error	max error	#points	#count[err > 1]
0.215	0.389	2.6918	266	13

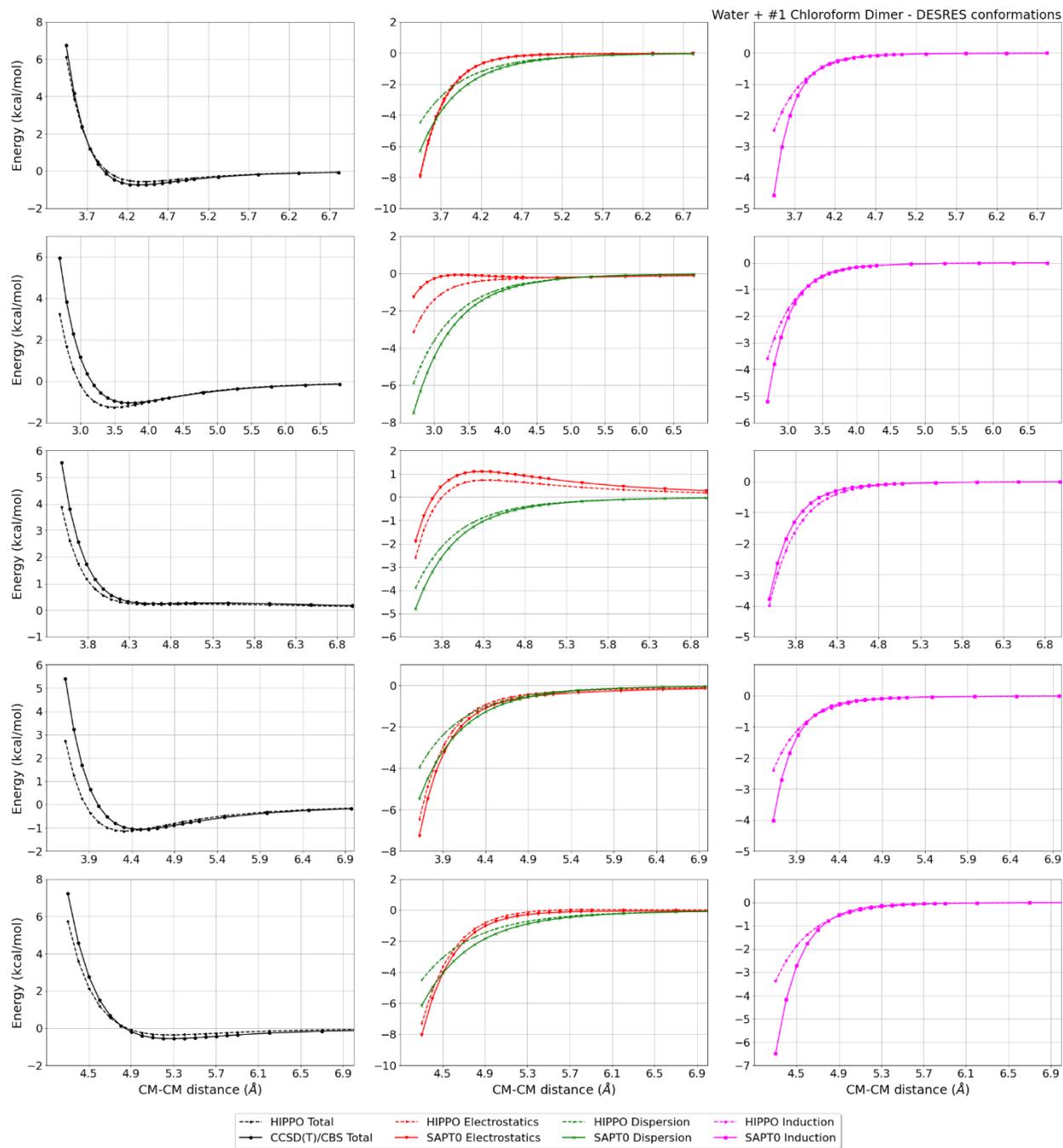
DESRES_1-1, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.091	0.052	0.2004	25	0

Liquid Chloroform @ 298.15 K

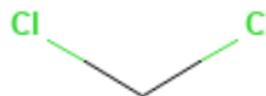
Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1477.95	1479.30	0.1	31.41	31.28	0.4	4.13	4.71	12.2	3000





#4 Dichloromethane CH2Cl2 CID: 6344

ref molpol	-7.93	-5.75	-5.15, avg	-6.28
molpol	7.81	5.83	5.51, avg	6.38
rms molpol	0.12	0.08	0.36, avg	0.11



Monomer potential fitting RMS: 0.14

##Dimer results - Fitting to QM datasets##

DESRES_4-4, energy values in kcal/mol

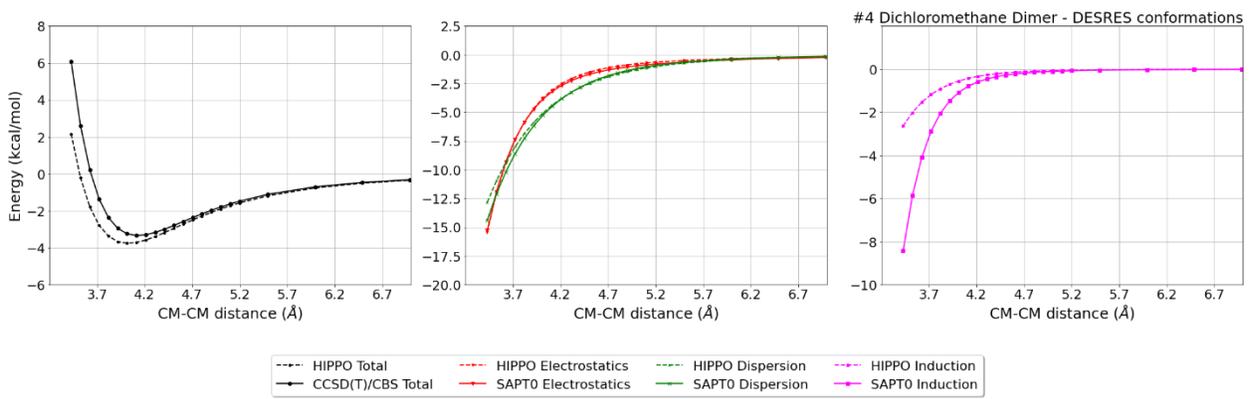
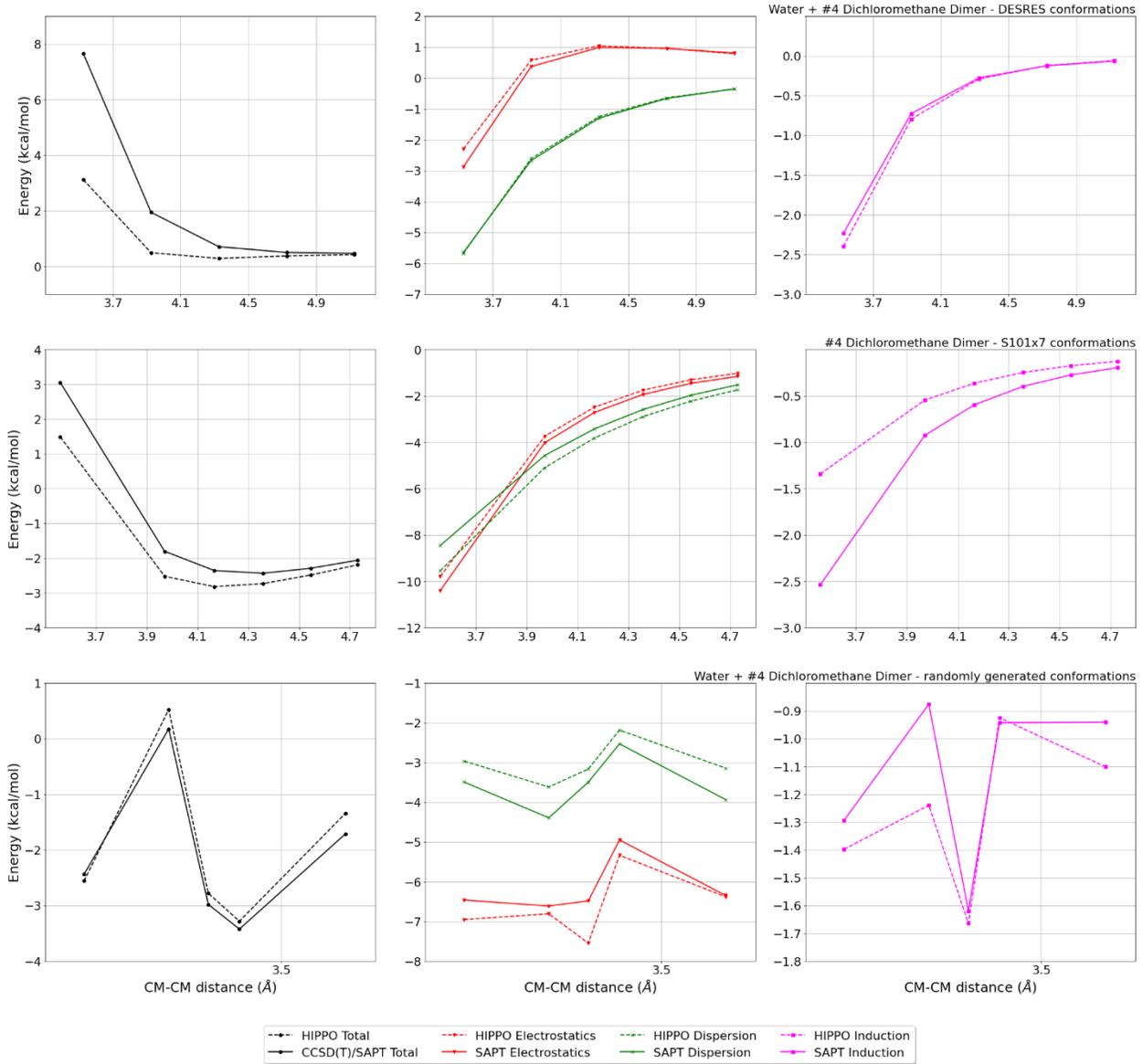
MAE	Std error	max error	#points	#count[err > 1]
0.587	0.961	3.9321	25	5

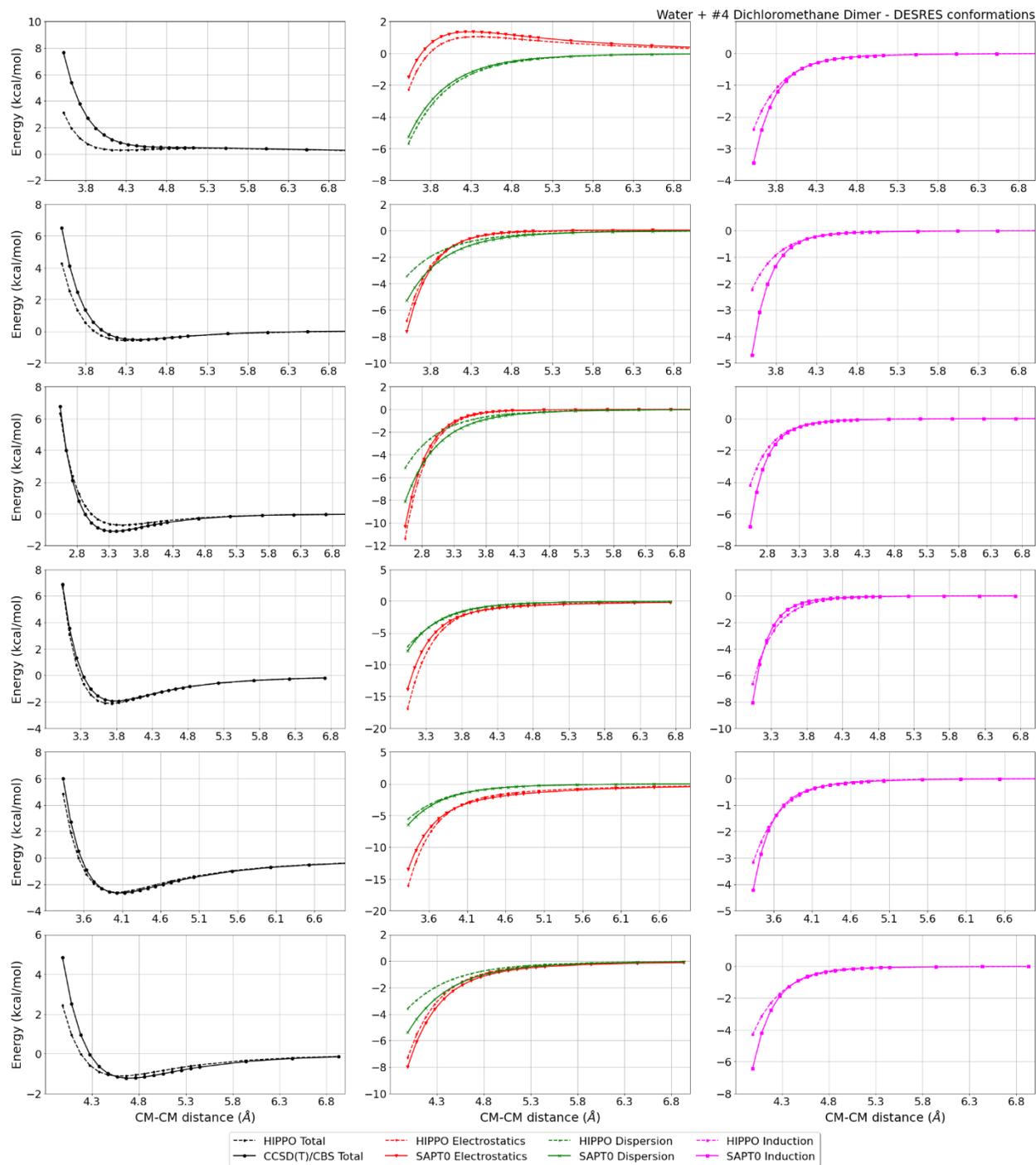
DESRES_4-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.217	0.471	4.5385	291	12

Liquid Dichloromethane @ 298.15 K

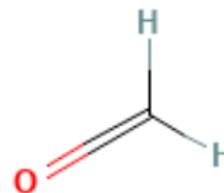
Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1322.72	1394.30	5.1	28.85	28.82	0.1	8.02	8.82	9.0	5000





#5 Methanal CH2O CID: 712

ref molpol	-2.60	-3.27	-1.91, avg	-2.59
molpol	2.68	2.67	2.07, avg	2.47
rms molpol	0.08	0.60	0.16, avg	0.12



Monomer potential fitting RMS: 0.28

##Dimer results - Fitting to QM datasets##

DESRES_5-water, energy values in kcal/mol

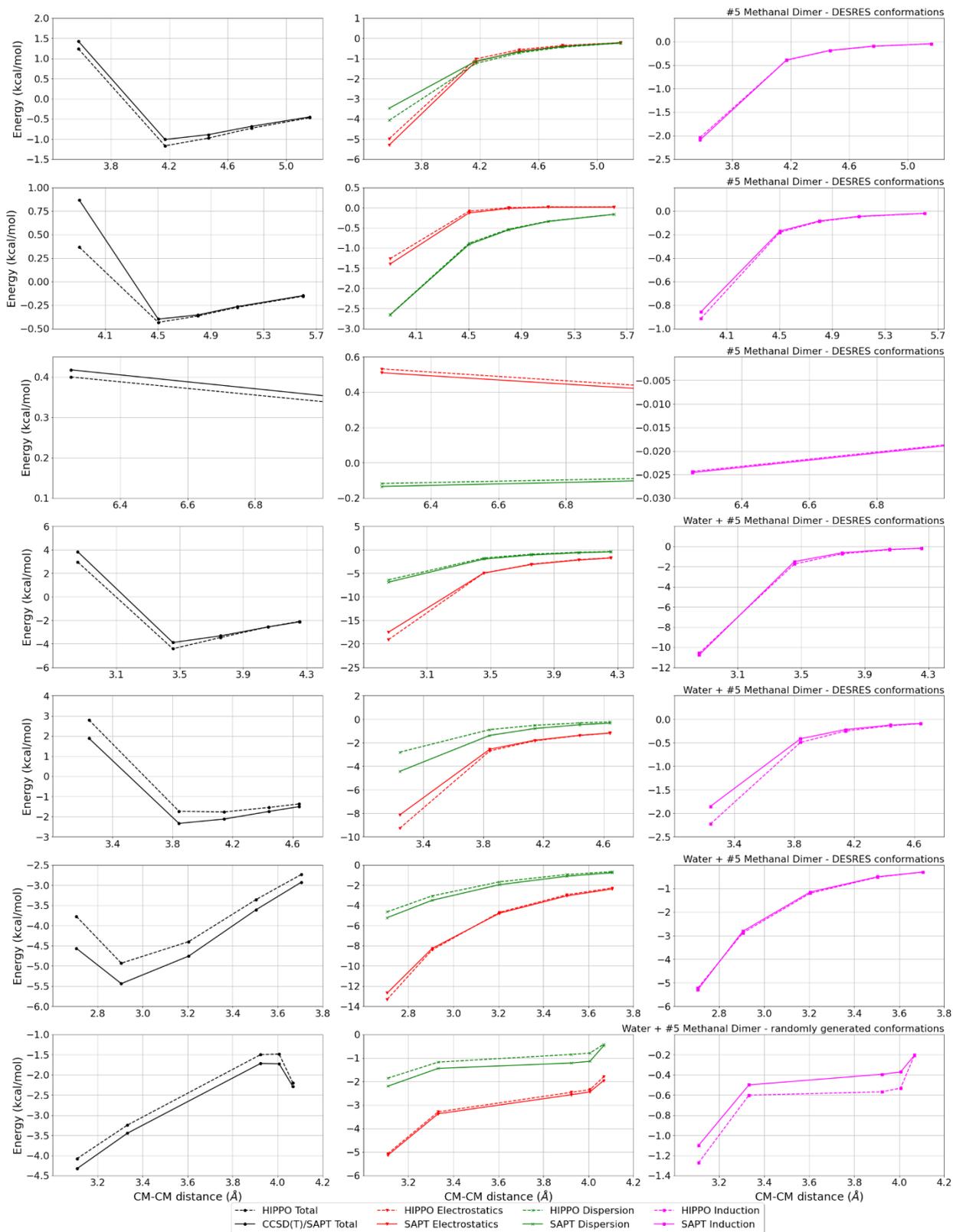
MAE	Std error	max error	#points	#count[err > 1]
0.289	0.498	4.6872	2661	235

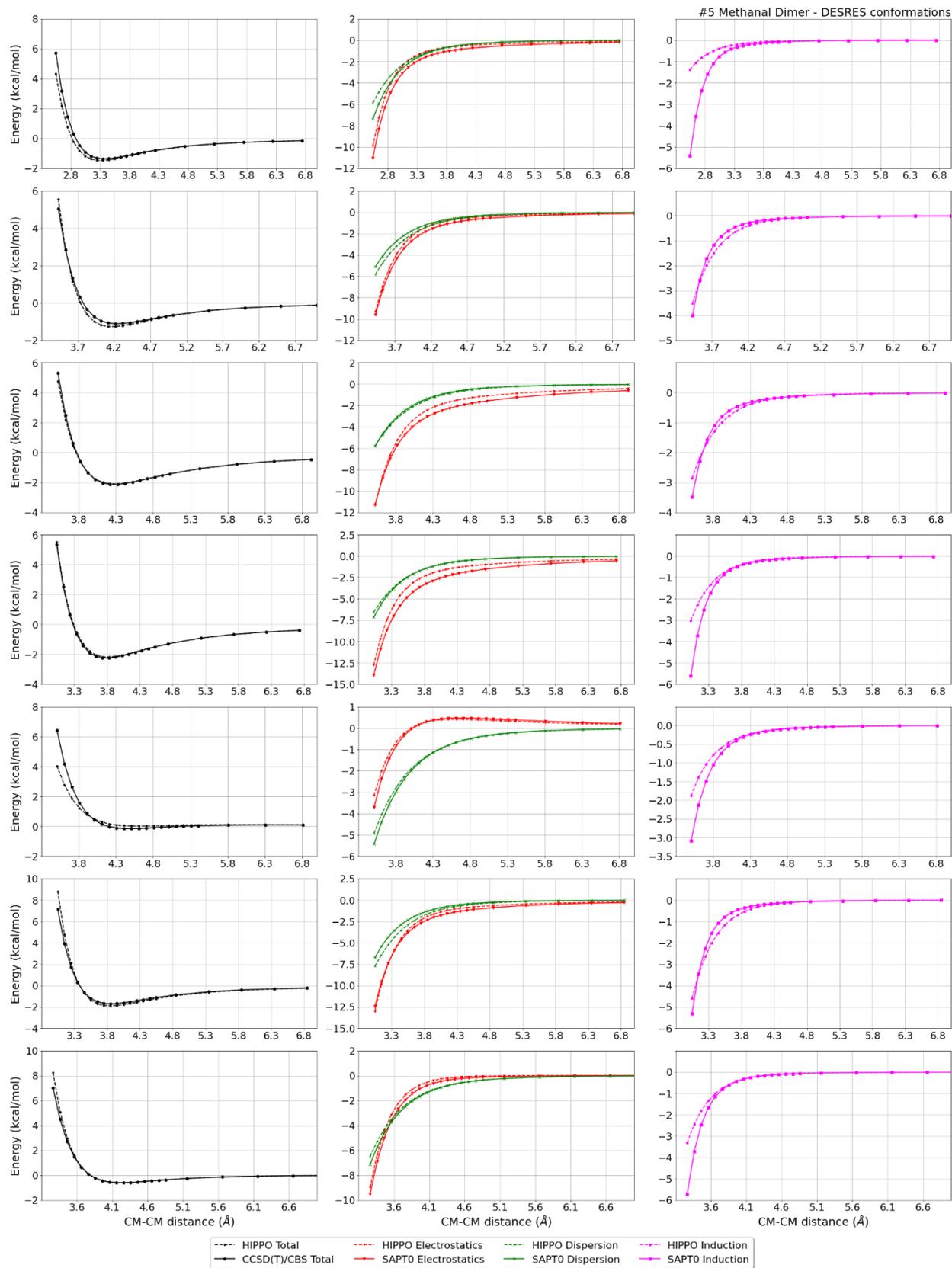
DESRES_5-5, energy values in kcal/mol

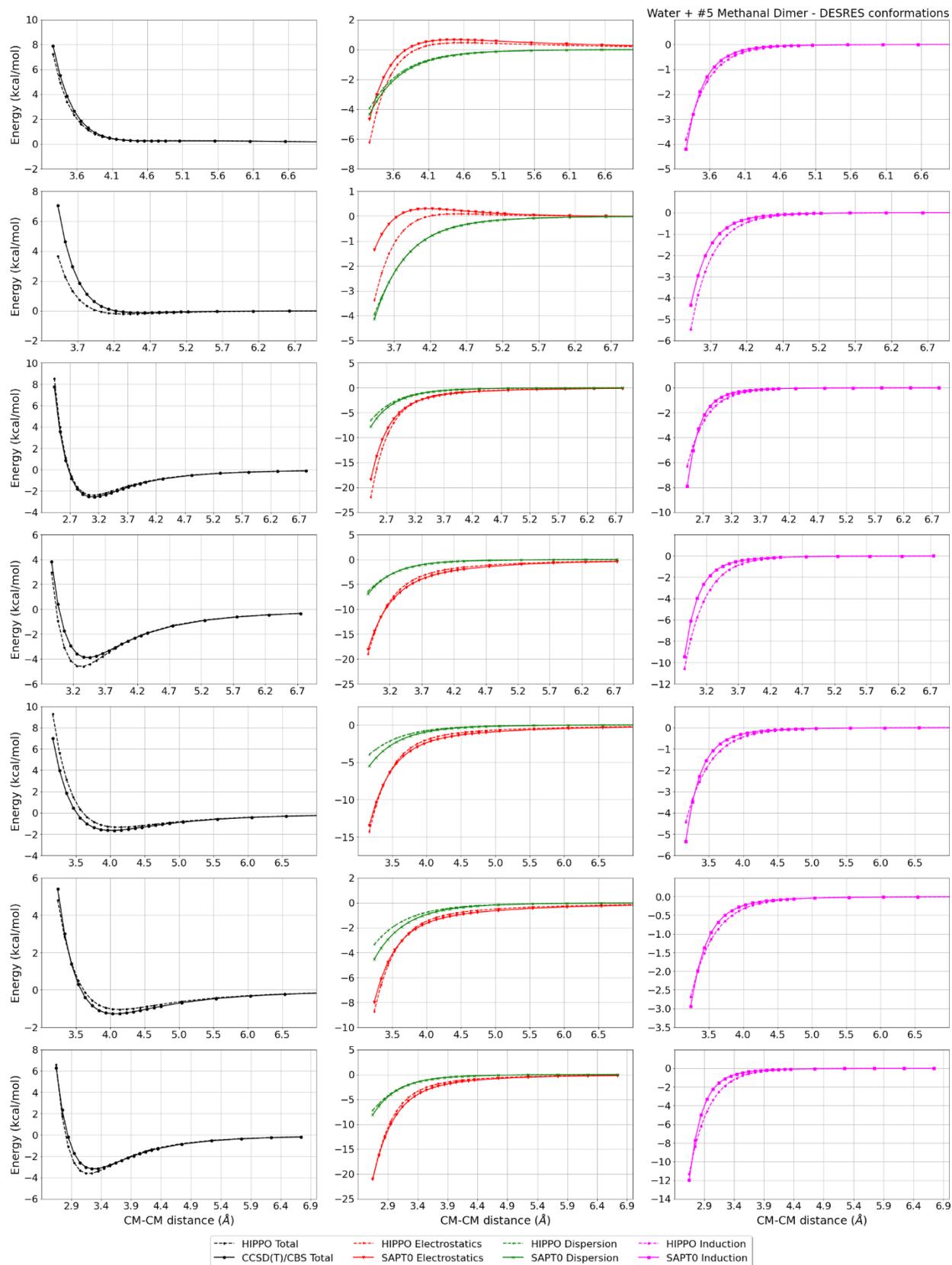
MAE	Std error	max error	#points	#count[err > 1]
0.143	0.413	5.2756	4258	159

Liquid Methanal @ 253.65 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	#nFrm
864.13	810.53	6.6	22.44	23.08	2.8	43.26	-1.00	3000

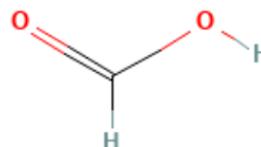






#6 Methanoic acid CH2O2 CID: 284

ref molpol	-4.09	-3.46	-2.40, avg	-3.32
molpol	4.03	3.47	2.77, avg	3.42
rms molpol	0.06	0.01	0.38, avg	0.11



Monomer potential fitting RMS: 2.65

##Dimer results - Fitting to QM datasets##

DESRES_6-water, energy values in kcal/mol

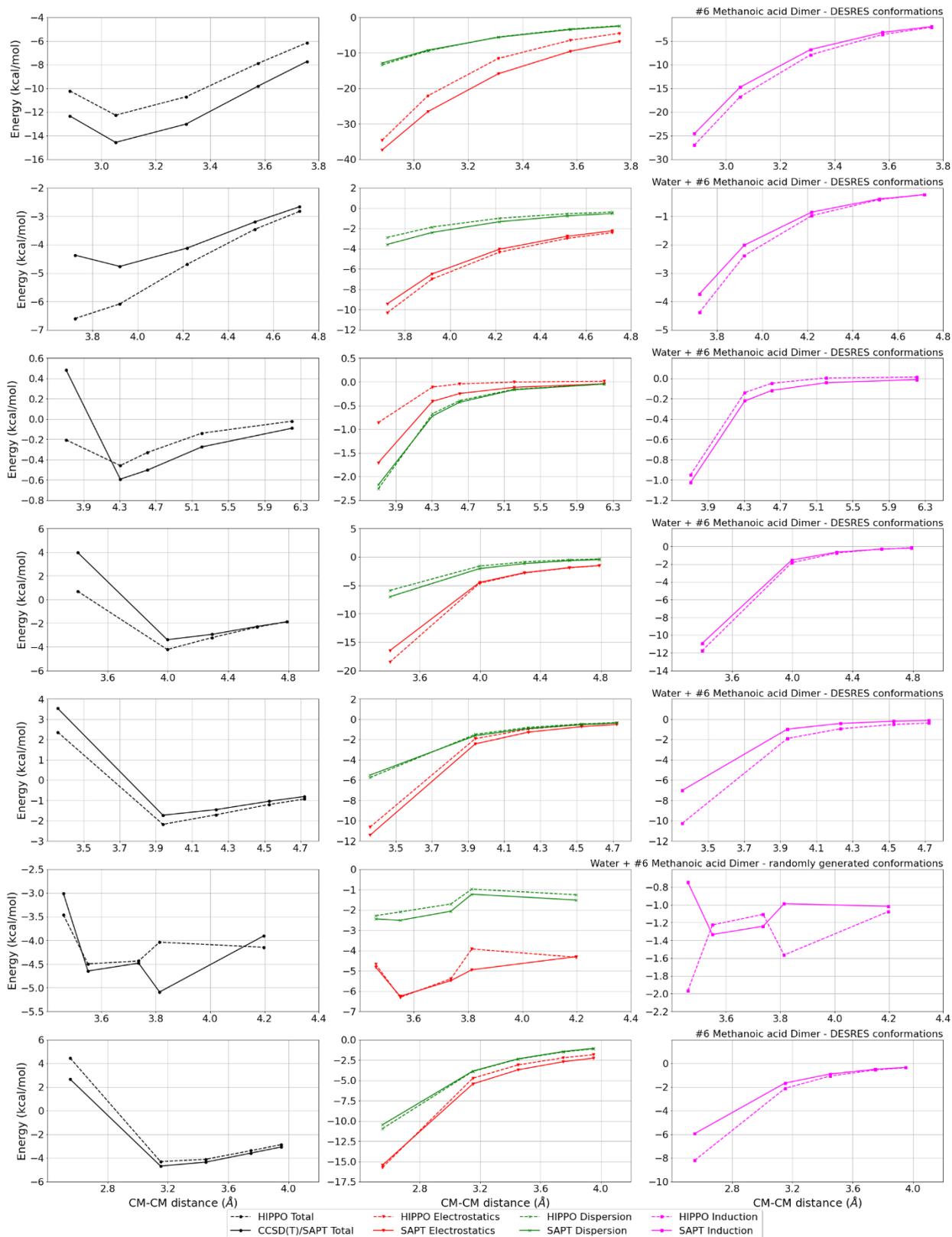
MAE	Std error	max error	#points	#count[err > 1]
0.657	0.965	7.3276	2711	624

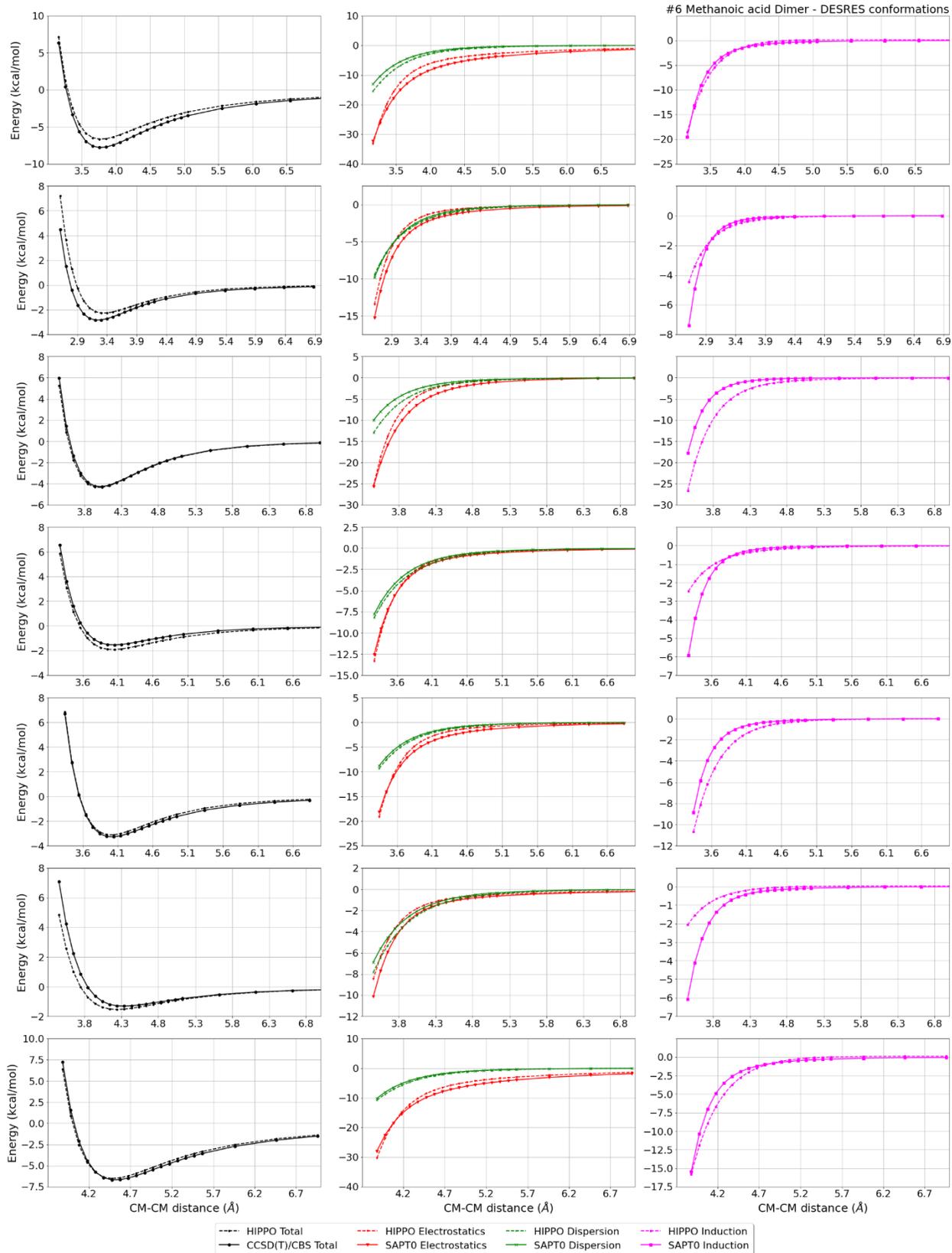
DESRES_6-6, energy values in kcal/mol

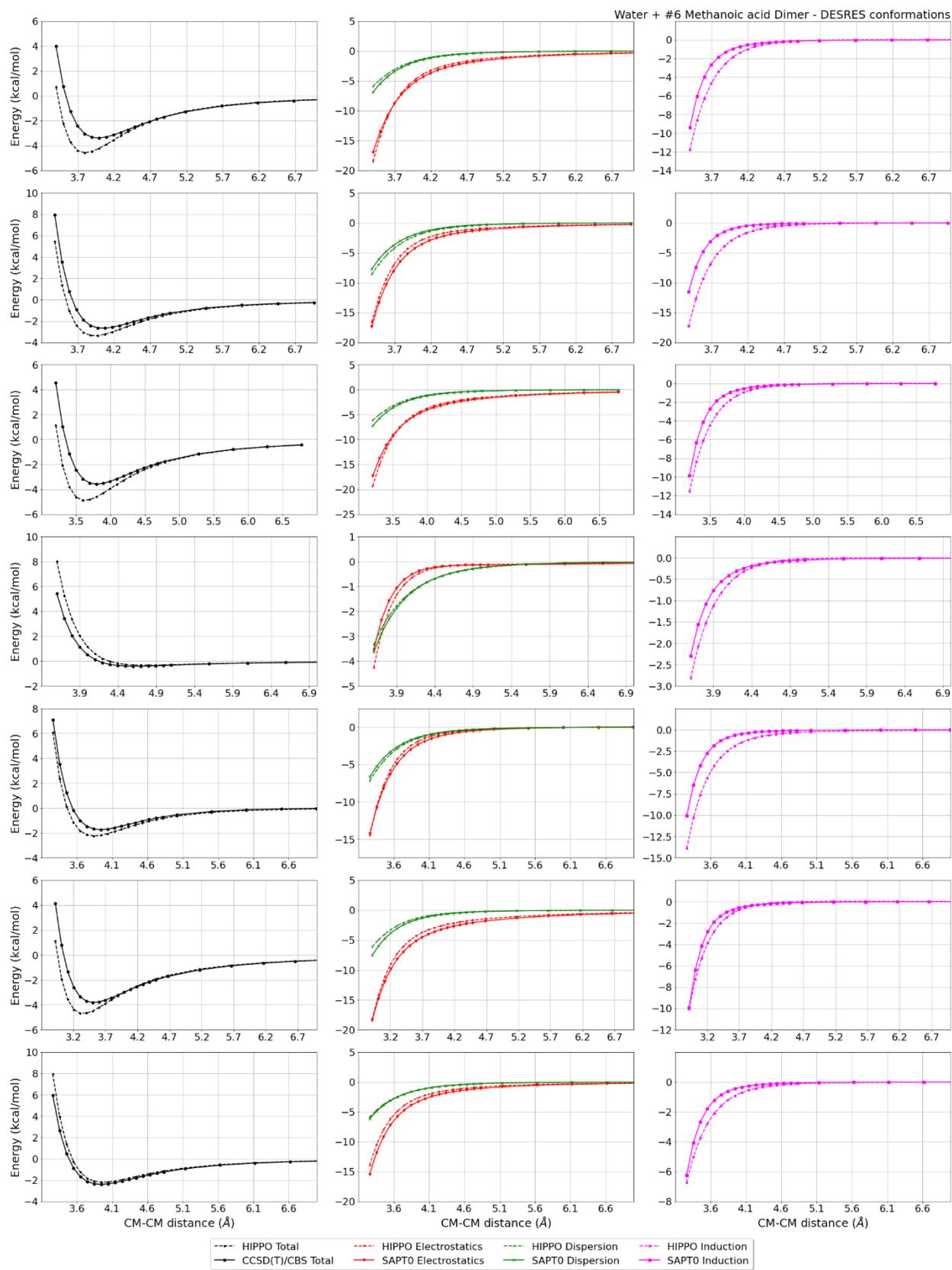
MAE	Std error	max error	#points	#count[err > 1]
0.215	0.427	4.7044	1626	64

Liquid Methanoic acid @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1221.93	1214.50	0.6	45.88	19.90	130.6	-1.00	51.10	102.0	-50

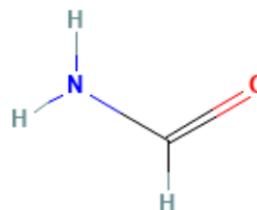






#8 Methanamide CH3NO CID: 713

ref molpol	-5.35	-4.03	-3.00, avg	-4.13
molpol	5.19	3.97	3.33, avg	4.16
rms molpol	0.16	0.07	0.32, avg	0.03



Monomer potential fitting RMS: 1.76

##Dimer results - Fitting to QM datasets##

DESRES_8-water, energy values in kcal/mol

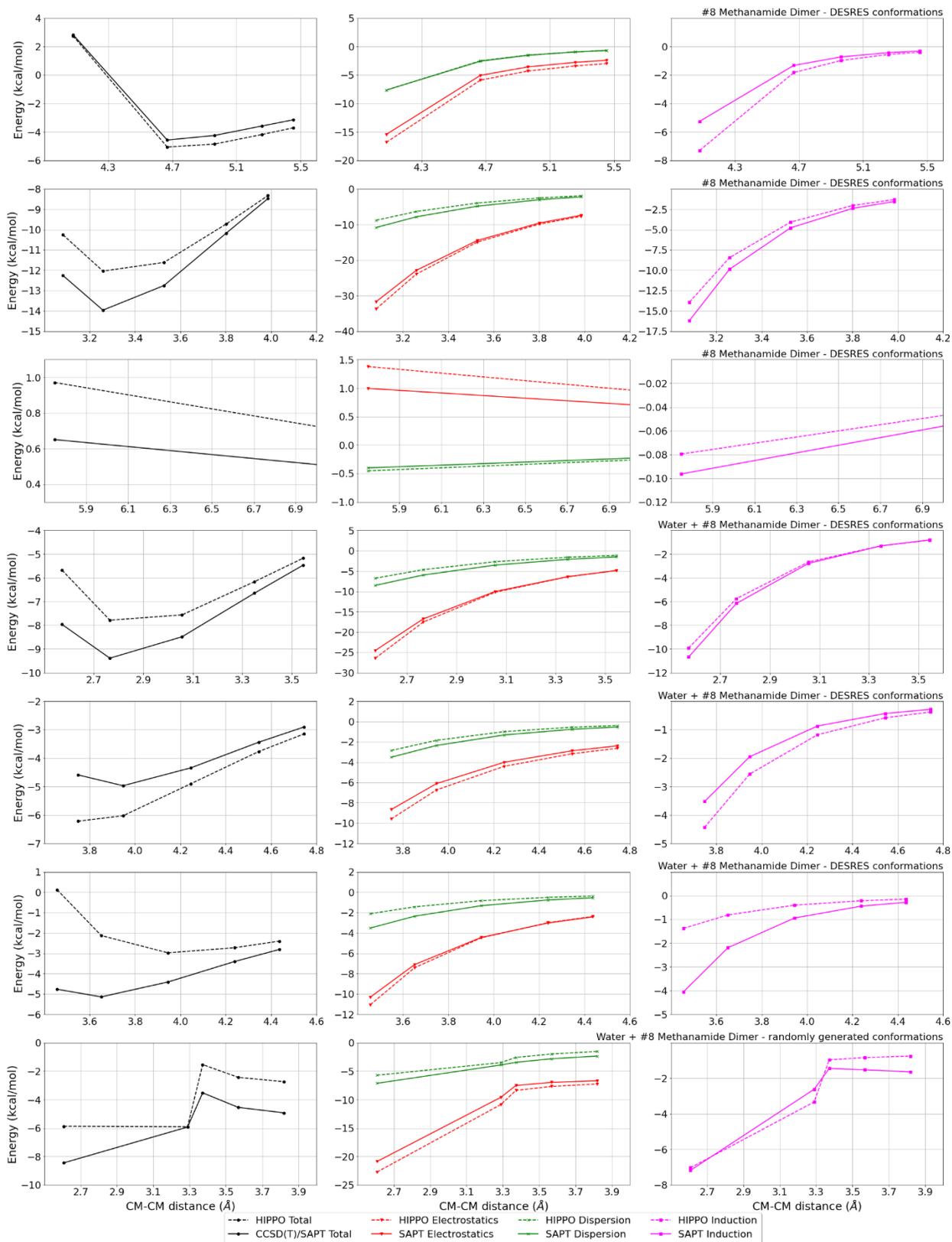
MAE	Std error	max error	#points	#count[err > 1]
0.496	1.124	12.5480	554	84

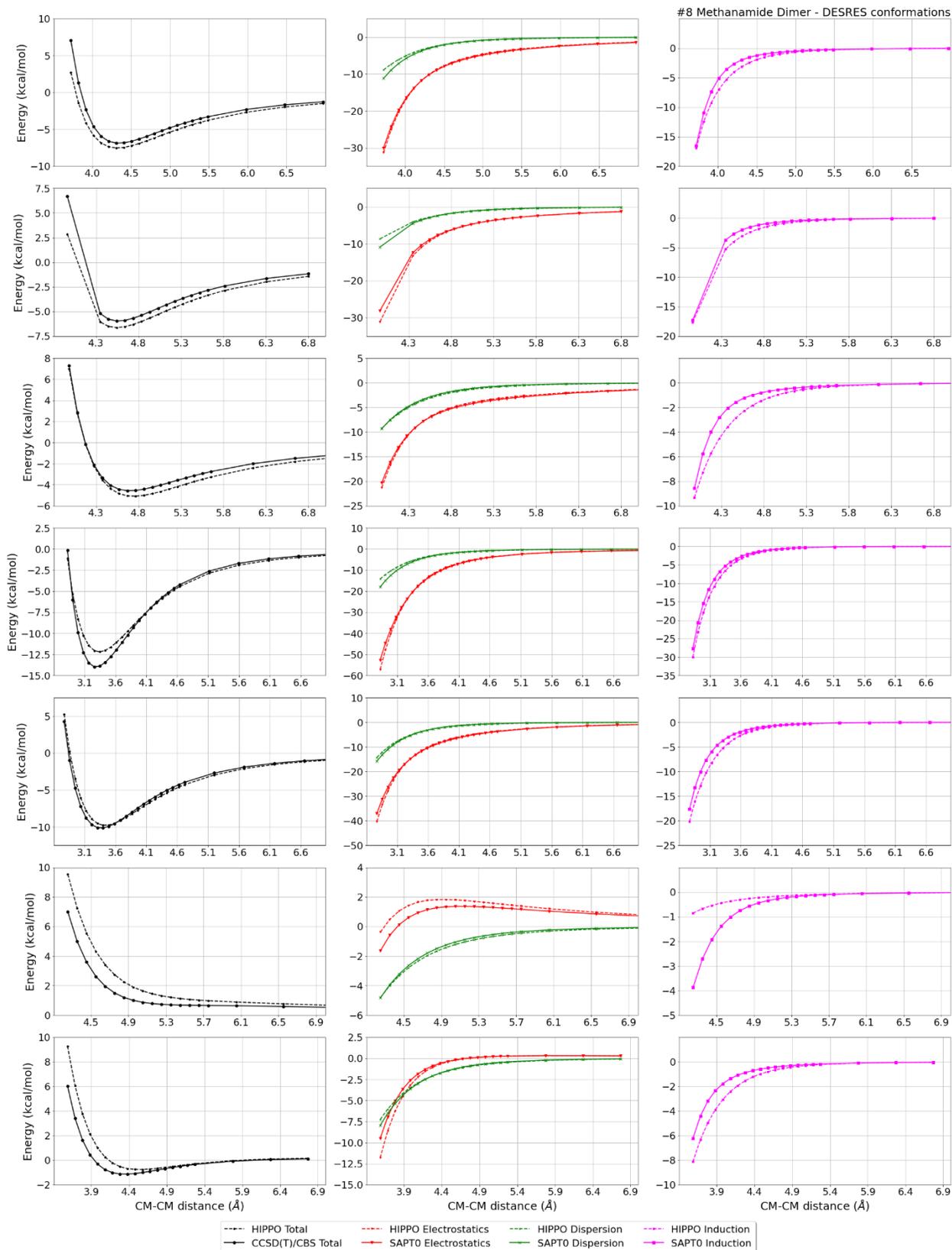
DESRES_8-8, energy values in kcal/mol

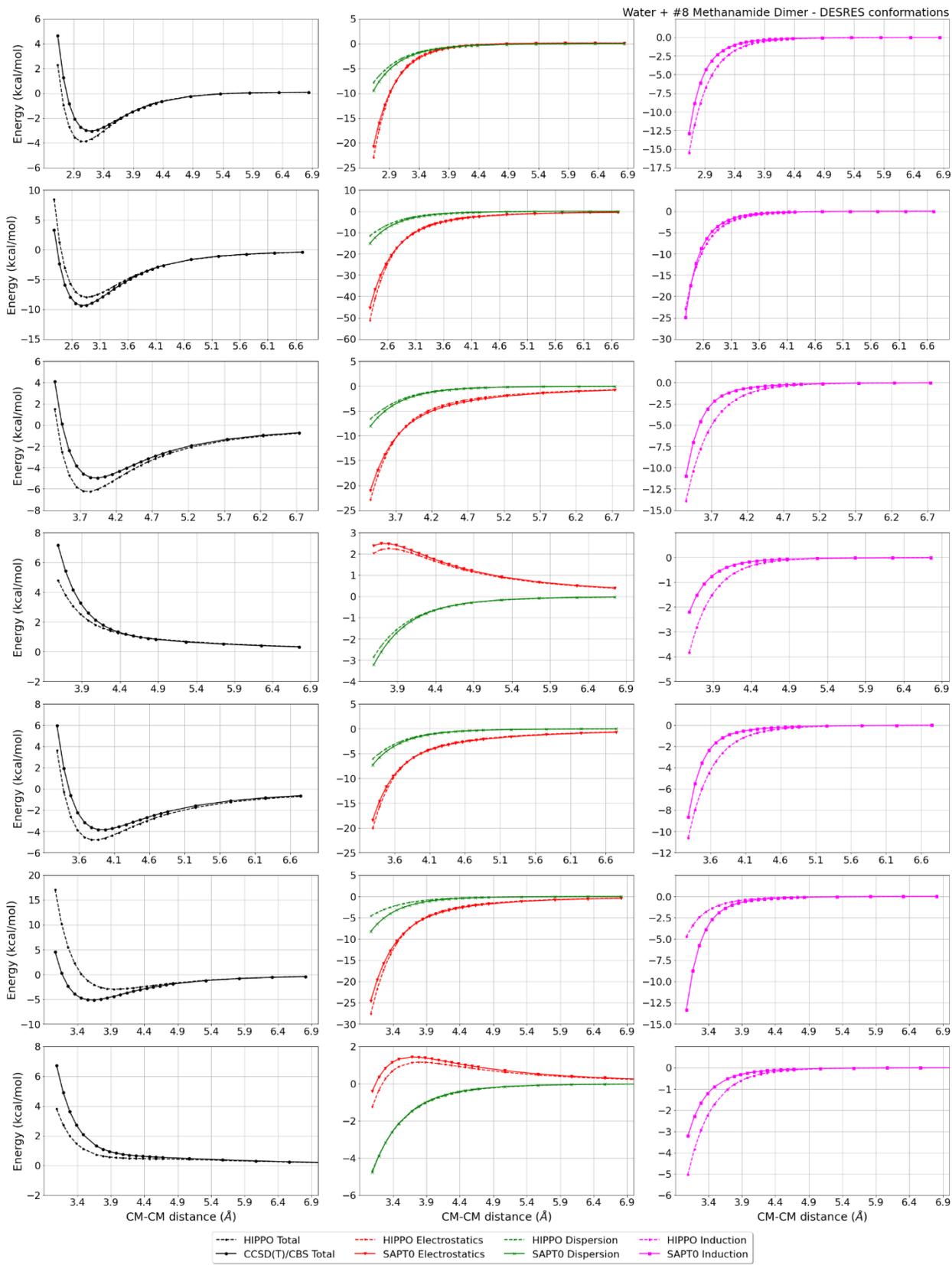
MAE	Std error	max error	#points	#count[err > 1]
0.269	0.414	4.3172	1314	60

Liquid Methanamide @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1.00	1128.80	99.9	1.00	60.57	98.3	-1.00	108.94	100.9	-50

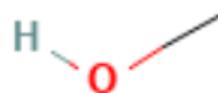






#10 Methanol CH4O CID: 887

ref molpol	-3.42	-2.97	-2.91, avg	-3.10
molpol	3.44	2.97	2.90, avg	3.10
rms molpol	0.02	0.01	0.01, avg	0.00



Monomer potential fitting RMS: 0.94

##Dimer results - Fitting to QM datasets##

R739x5_10-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
2.356	2.225	2.847	0.6217		
2.455	1.894	2.330	0.4362		
2.555	1.652	1.968	0.3156		
2.655	1.465	1.700	0.2351		
2.855	1.186	1.327	0.1414		
MAE	Std error	max error	#points	#count[err > 1]	
0.350	0.167	0.6217	5	0	

HB375x10_10-10, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
1.559	-1.784	-4.483	-2.6993		
1.604	-3.925	-6.124	-2.1991		
1.649	-5.128	-6.864	-1.7357		
1.694	-5.704	-7.059	-1.3551		
1.739	-5.871	-6.932	-1.0608		
1.784	-5.778	-6.620	-0.8420		
1.829	-5.528	-6.210	-0.6823		
1.966	-4.418	-4.841	-0.4230		
2.195	-2.730	-3.002	-0.2720		
2.656	-1.092	-1.237	-0.1449		
MAE	Std error	max error	#points	#count[err > 1]	
1.141	0.807	2.6993	10	5	

DESRES_10-10, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.141	0.363	4.4289	1619	51

R739x5_10-10, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
2.038	2.318	3.149	0.8310		
2.113	1.945	2.540	0.5952		
2.187	1.680	2.123	0.4426		
2.262	1.481	1.822	0.3407		
2.411	1.193	1.413	0.2204		
MAE	Std error	max error	#points	#count[err > 1]	
0.486	0.212	0.8310	5	0	

HB375x10_10-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
1.755	-1.669	-1.982	-0.3129
1.815	-3.809	-4.278	-0.4689
1.832	-1.232	-0.948	0.2844
1.874	-5.012	-5.448	-0.4356
1.893	-3.240	-3.224	0.0160
1.934	-5.590	-5.927	-0.3368
1.953	-4.368	-4.448	-0.0804
1.993	-5.761	-5.989	-0.2285
2.014	-4.914	-5.010	-0.0962
2.053	-5.673	-5.808	-0.1351
2.074	-5.081	-5.164	-0.0830
2.112	-5.429	-5.492	-0.0627
2.135	-5.009	-5.074	-0.0648
2.196	-4.794	-4.844	-0.0501
2.292	-4.339	-4.297	0.0423
2.379	-3.827	-3.867	-0.0396
2.592	-2.682	-2.639	0.0426
2.684	-2.364	-2.425	-0.0611
3.193	-1.078	-1.085	-0.0068
3.297	-0.958	-1.015	-0.0569

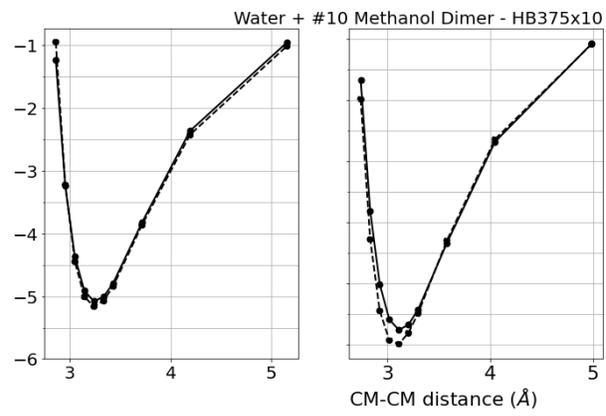
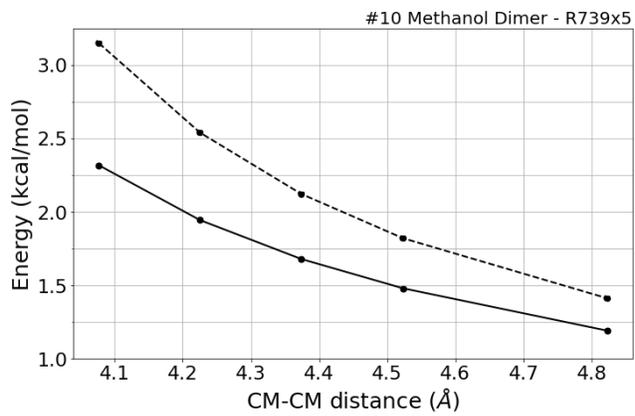
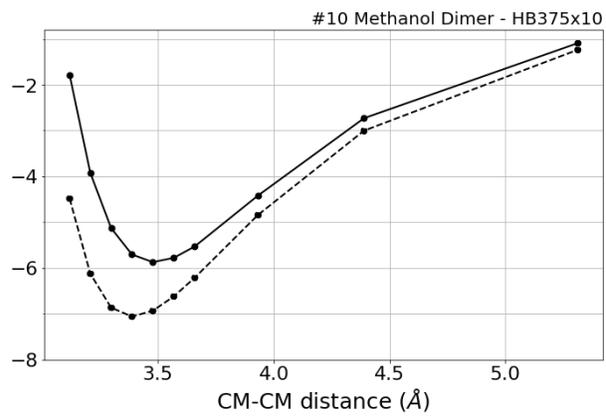
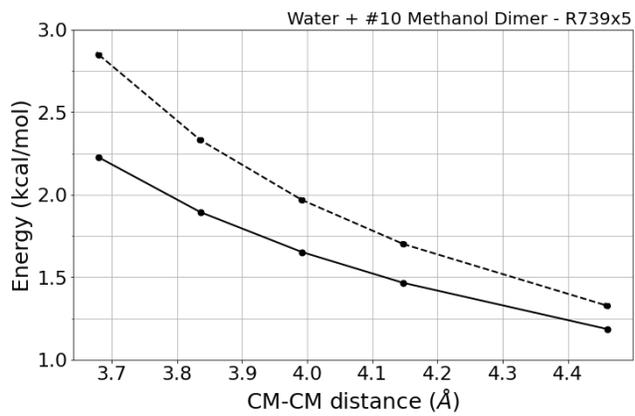
MAE	Std error	max error	#points	#count[err > 1]
0.145	0.141	0.4689	20	0

DESRES_10-water, energy values in kcal/mol

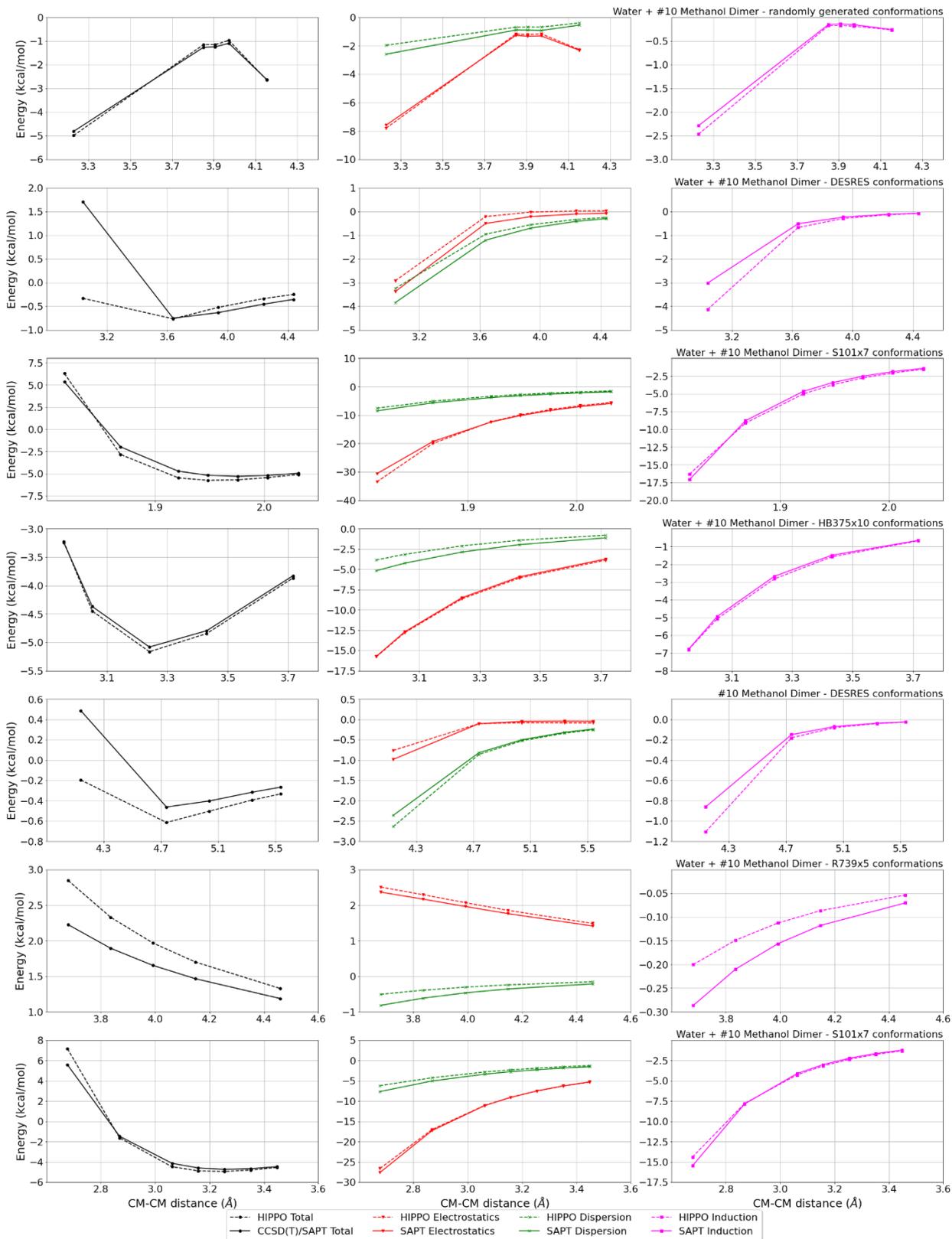
MAE	Std error	max error	#points	#count[err > 1]
0.259	0.527	6.0942	2652	183

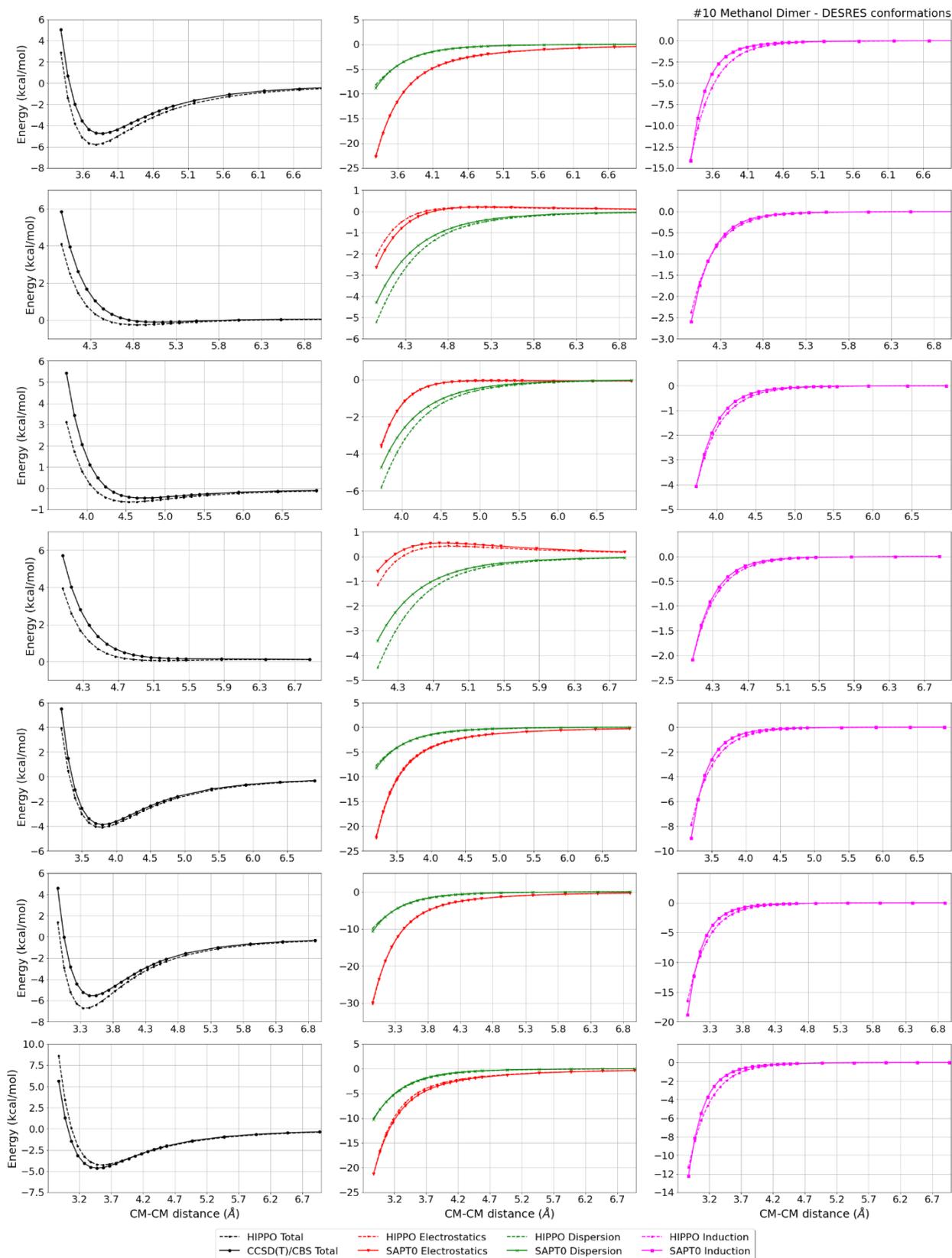
Liquid Methanol @ 298.15 K

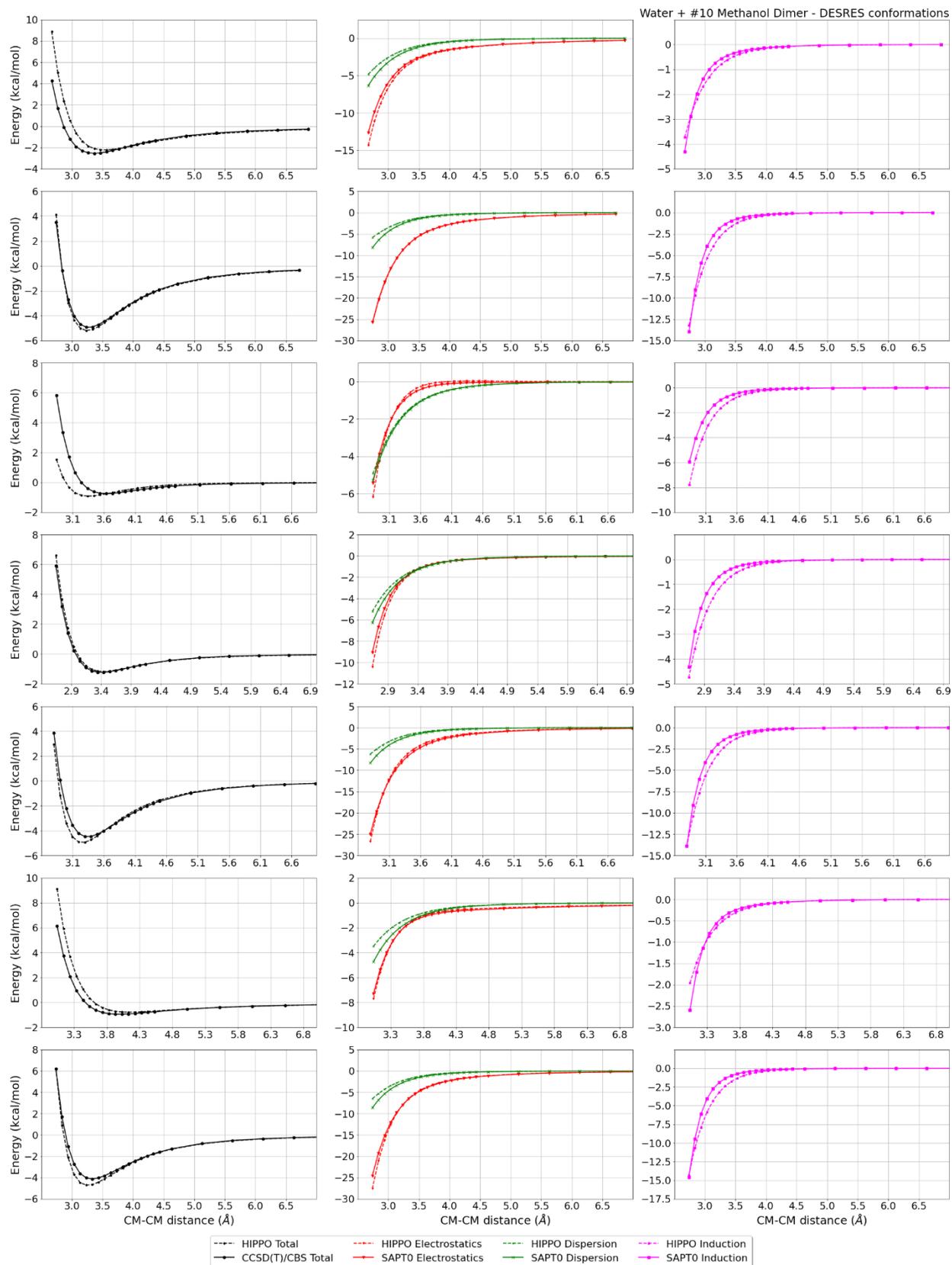
Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	#nFrm
824.78	787.20	4.8	54.17	37.43	44.7	8.14	-1.00	1746



—●— CCSD(T) Total - - -●- - HIPPO Total

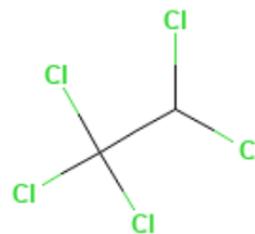






#11 1,1,1,2,2-pentachloroethane C2HCl5 CID: 6419

ref molpol -12.22 -14.87 -13.94, avg -13.68
 molpol 12.57 14.80 14.84, avg 14.07
 rms molpol 0.35 0.07 0.90, avg 0.39



Monomer potential fitting RMS: 0.56

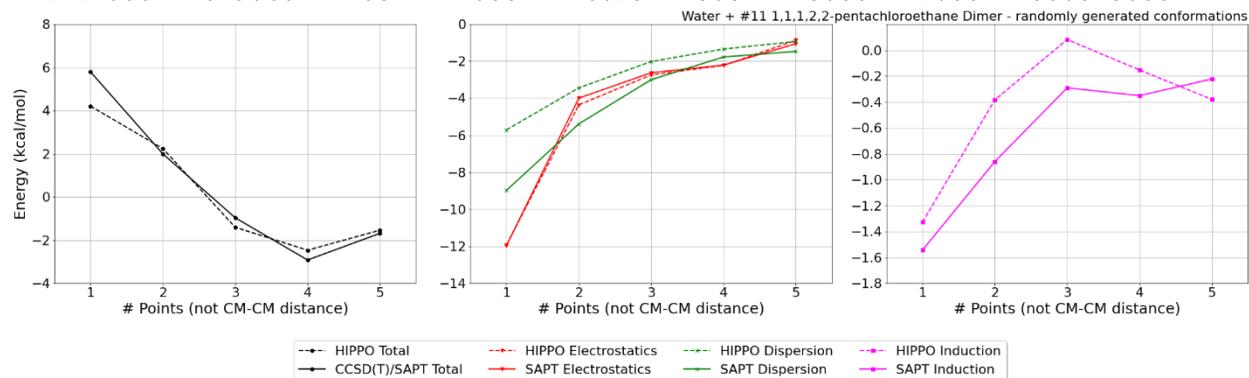
##Dimer results - Fitting to QM datasets##

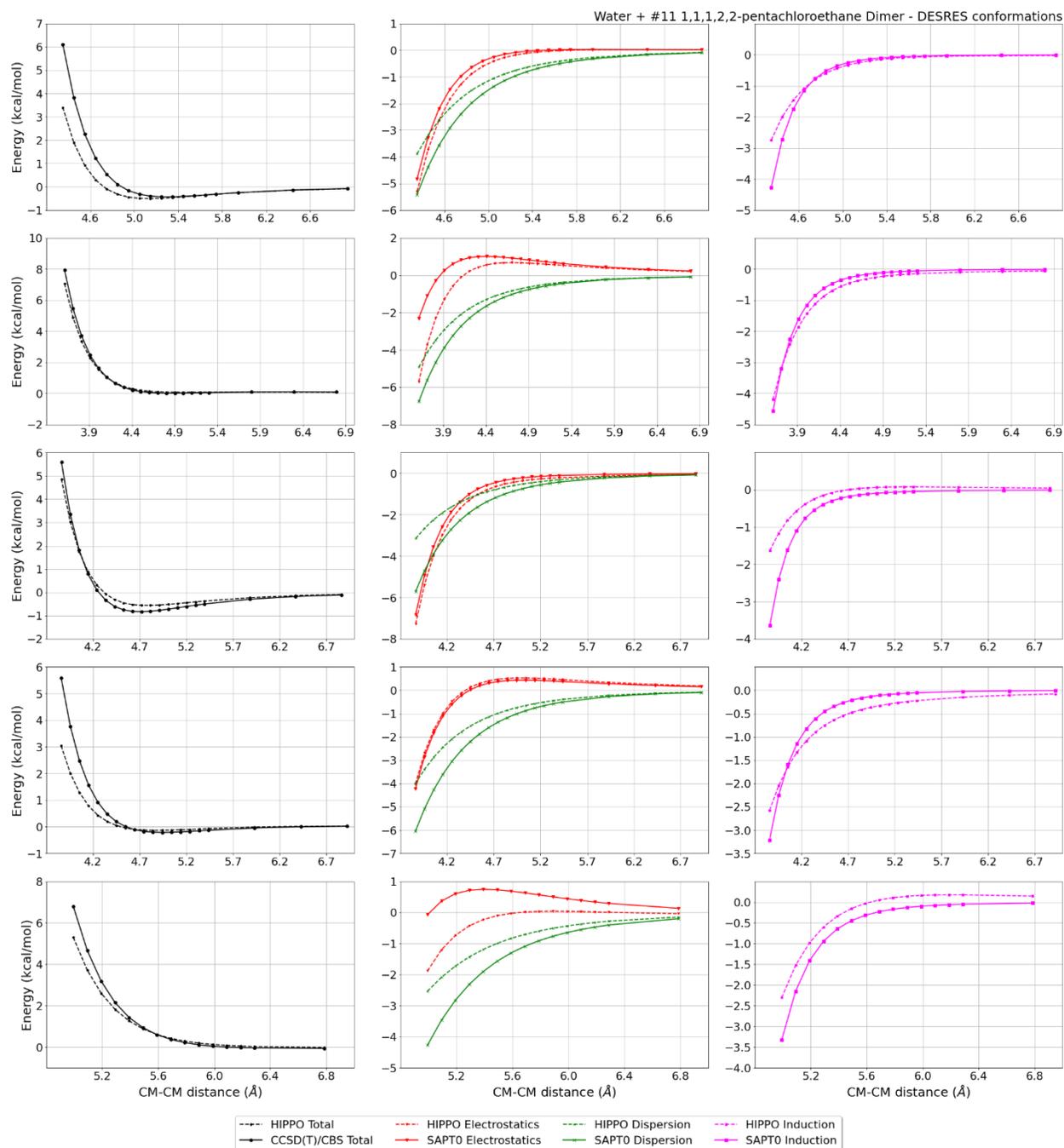
DESRES_11-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.272	0.490	2.7534	112	8

Liquid 1,1,1,2,2-pentachloroethane @ 293.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1722.00	1679.60	2.5	44.53	46.29	3.8	3.80	-1.00	-480.5	3000





#18 1,2-dichloroethane C2H4Cl2 CID: 11

ref molpol	-10.83	-7.03	-6.61, avg	-8.16
molpol	10.87	7.41	6.98, avg	8.42
rms molpol	0.04	0.38	0.37, avg	0.26



Monomer potential fitting RMS: 0.22

##Dimer results - Fitting to QM datasets##

DESRES_18-water, energy values in kcal/mol

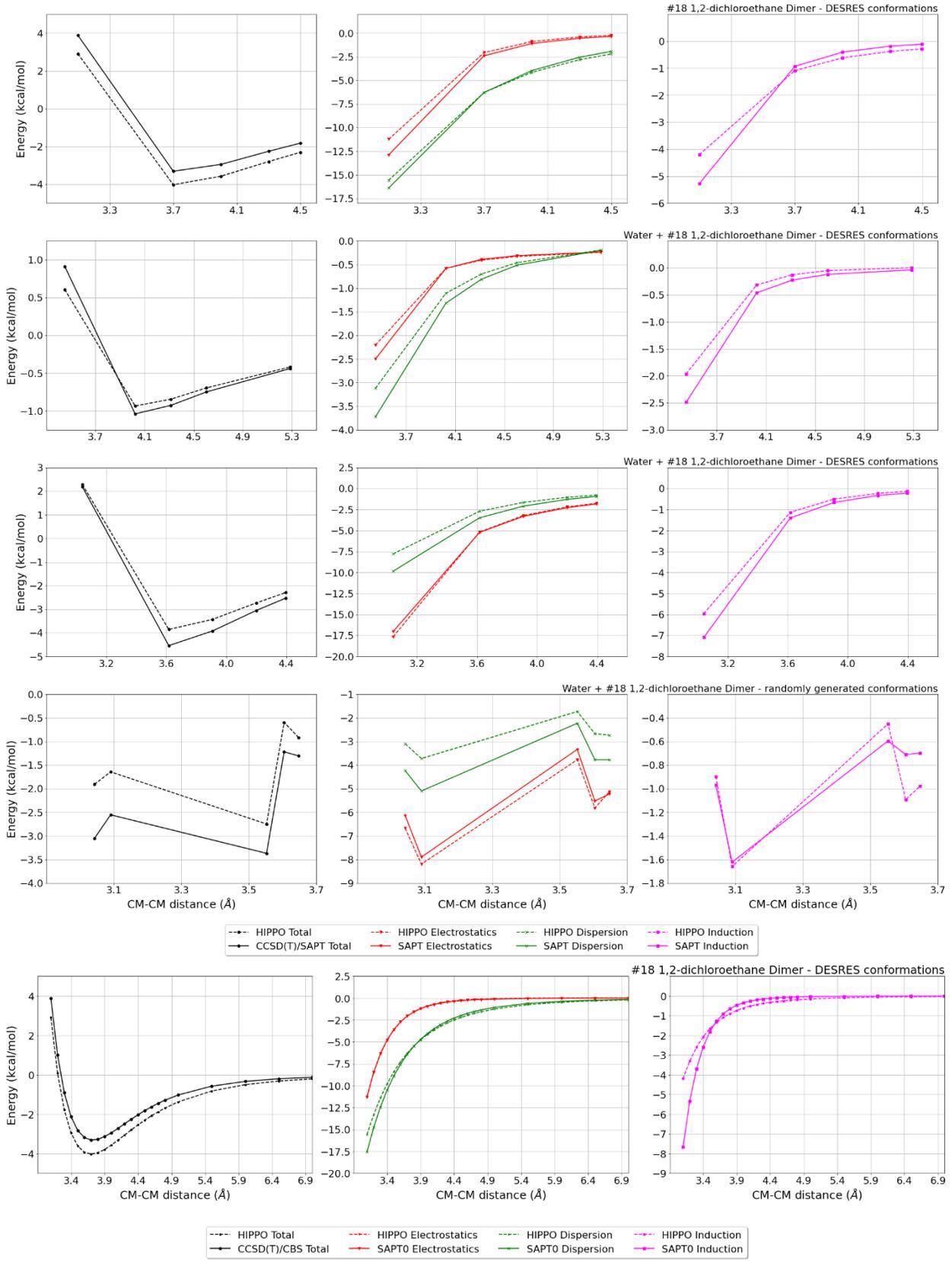
MAE	Std error	max error	#points	#count[err > 1]
0.195	0.320	1.9072	288	11

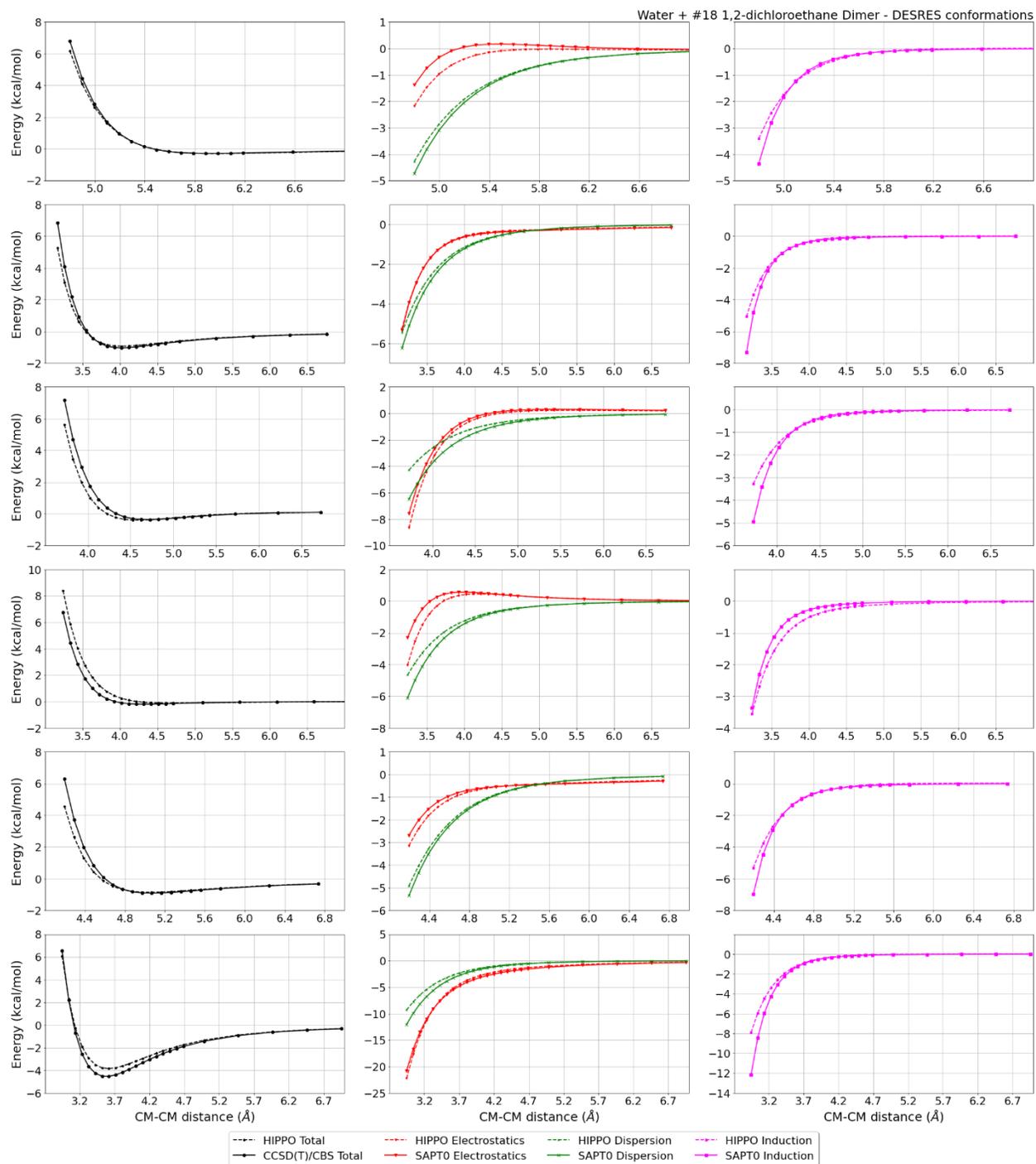
DESRES_18-18, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.398	0.167	0.6085	24	0

Liquid 1,2-dichloroethane @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1215.05	1246.30	2.5	35.43	35.16	0.8	10.60	10.13	4.6	3000





#20 Bromoethane C2H5Br CID: 6332

ref molpol	-8.73	-6.56	-6.17, avg	-7.15
molpol	8.73	6.56	6.19, avg	7.16
rms molpol	0.00	0.01	0.02, avg	0.01



Monomer potential fitting RMS: 0.14

##Dimer results - Fitting to QM datasets##

DESRES_20-20, energy values in kcal/mol

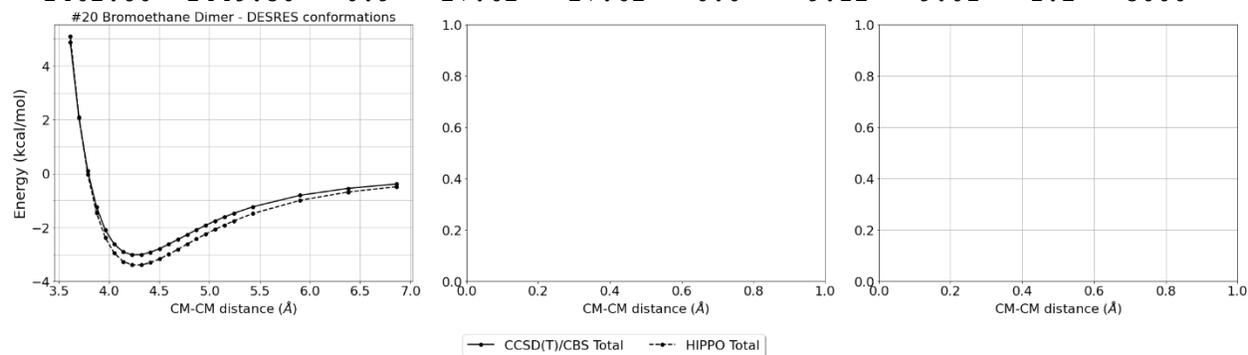
MAE	Std error	max error	#points	#count[err > 1]
0.261	0.113	0.3853	25	0

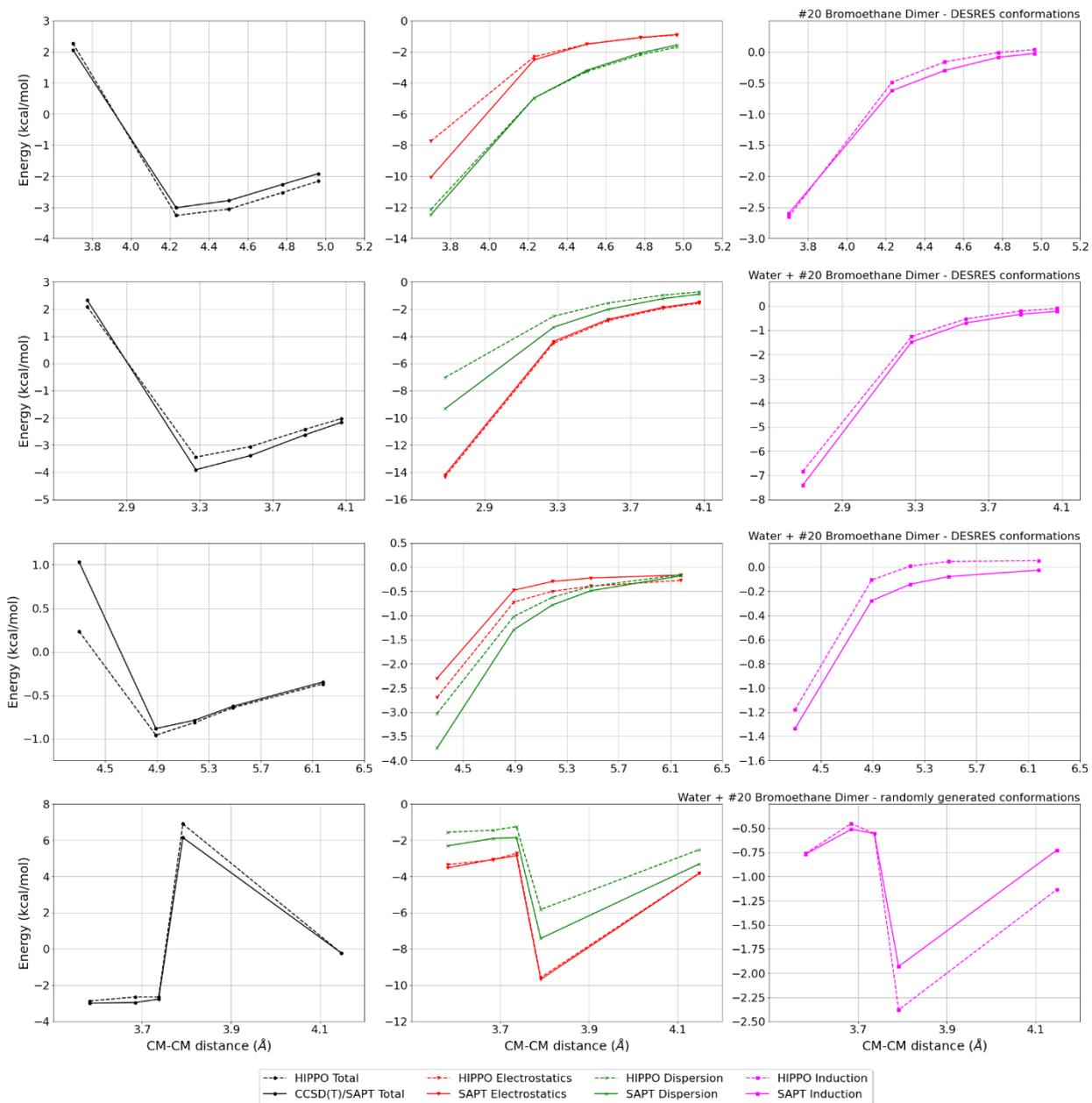
DESRES_20-water, energy values in kcal/mol

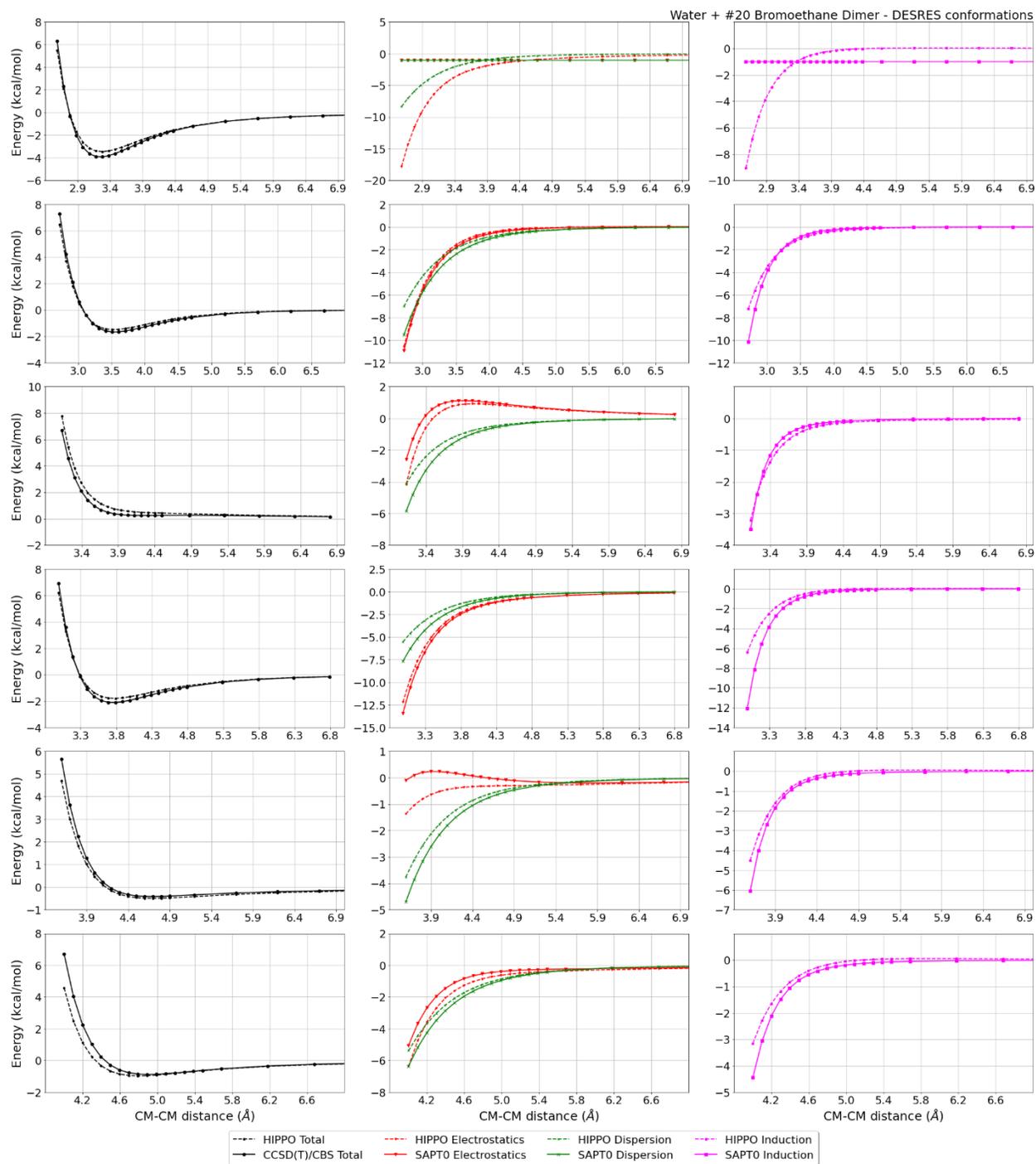
MAE	Std error	max error	#points	#count[err > 1]
0.188	0.276	2.3955	291	6

Liquid Bromoethane @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1462.80	1449.30	0.9	27.62	27.62	0.0	9.12	9.01	1.2	3000

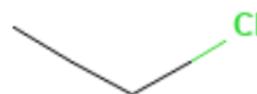






#21 Chloroethane C2H5Cl CID: 6337

ref molpol	-7.30	-5.70	-5.31, avg	-6.11
molpol	6.79	5.38	5.03, avg	5.73
rms molpol	0.52	0.32	0.29, avg	0.38



Monomer potential fitting RMS: 0.20

##Dimer results - Fitting to QM datasets##

DESRES_21-21, energy values in kcal/mol

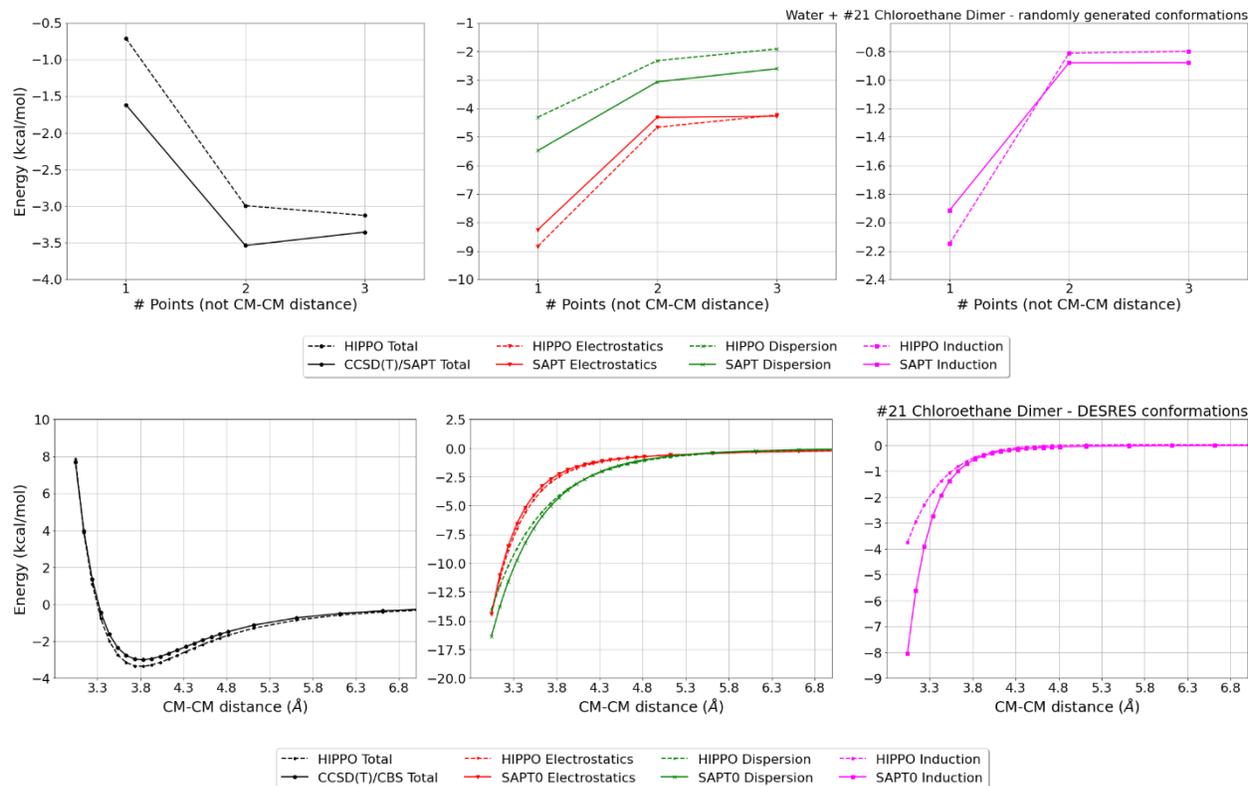
MAE	Std error	max error	#points	#count[err > 1]
0.280	0.138	0.4685	25	0

DESRES_21-water, energy values in kcal/mol

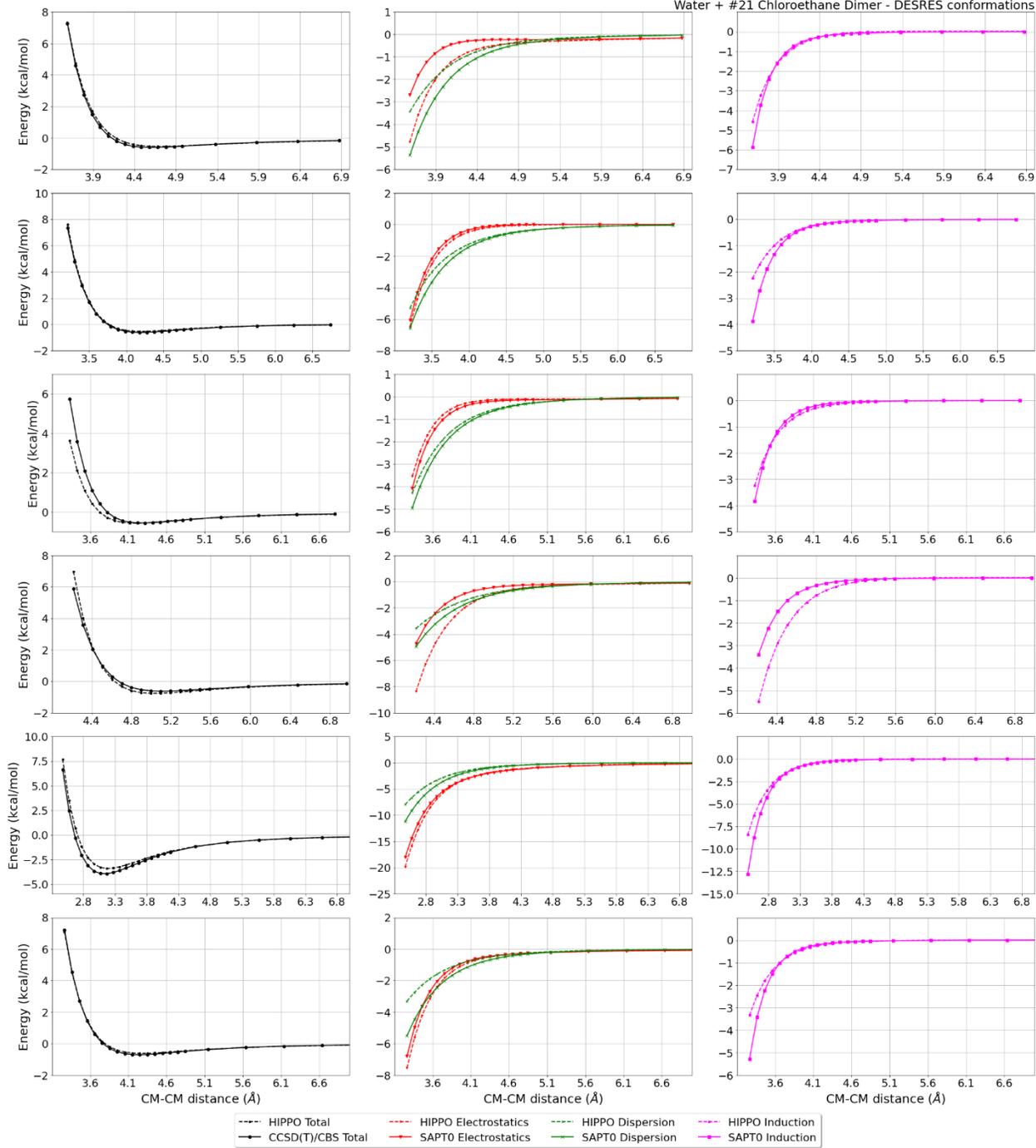
MAE	Std error	max error	#points	#count[err > 1]
0.131	0.245	2.1110	288	3

Liquid Chloroethane @ 273.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
919.78	923.90	0.4	25.43	25.39	0.2	10.13	10.41	2.7	4000

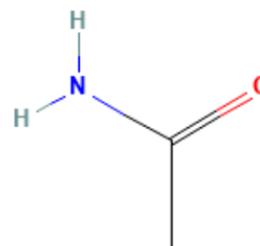


Water + #21 Chloroethane Dimer - DESRES conformations



#23 Ethanamide C₂H₅NO CID: 178

ref molpol	-6.69	-6.20	-4.49, avg	-5.80
molpol	6.64	6.15	4.47, avg	5.75
rms molpol	0.05	0.05	0.02, avg	0.04



Monomer potential fitting RMS: 1.74

##Dimer results - Fitting to QM datasets##

DESRES_23-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.618	1.351	13.1694	566	89

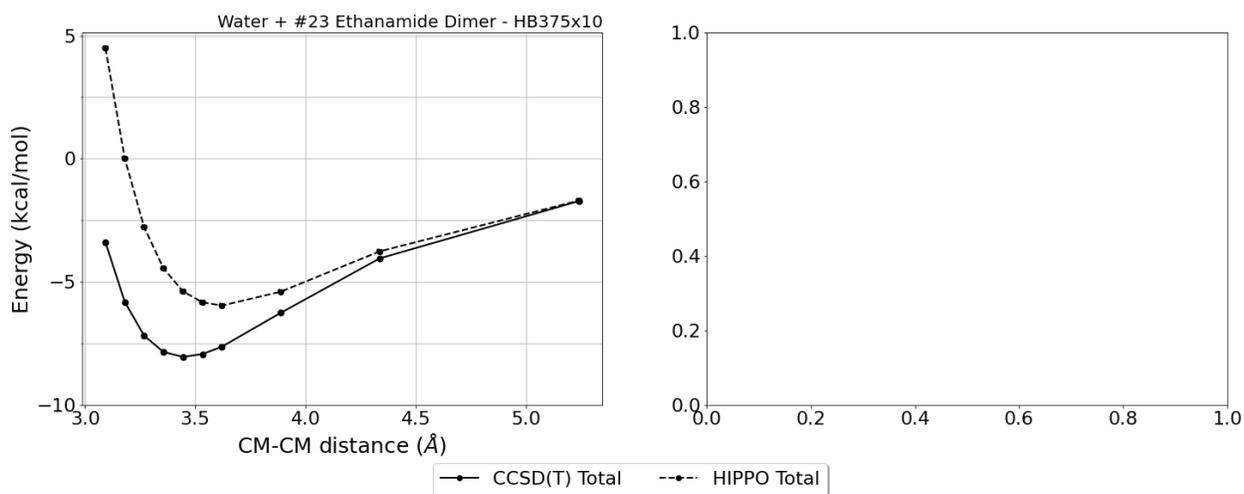
DESRES_23-23, energy values in kcal/mol

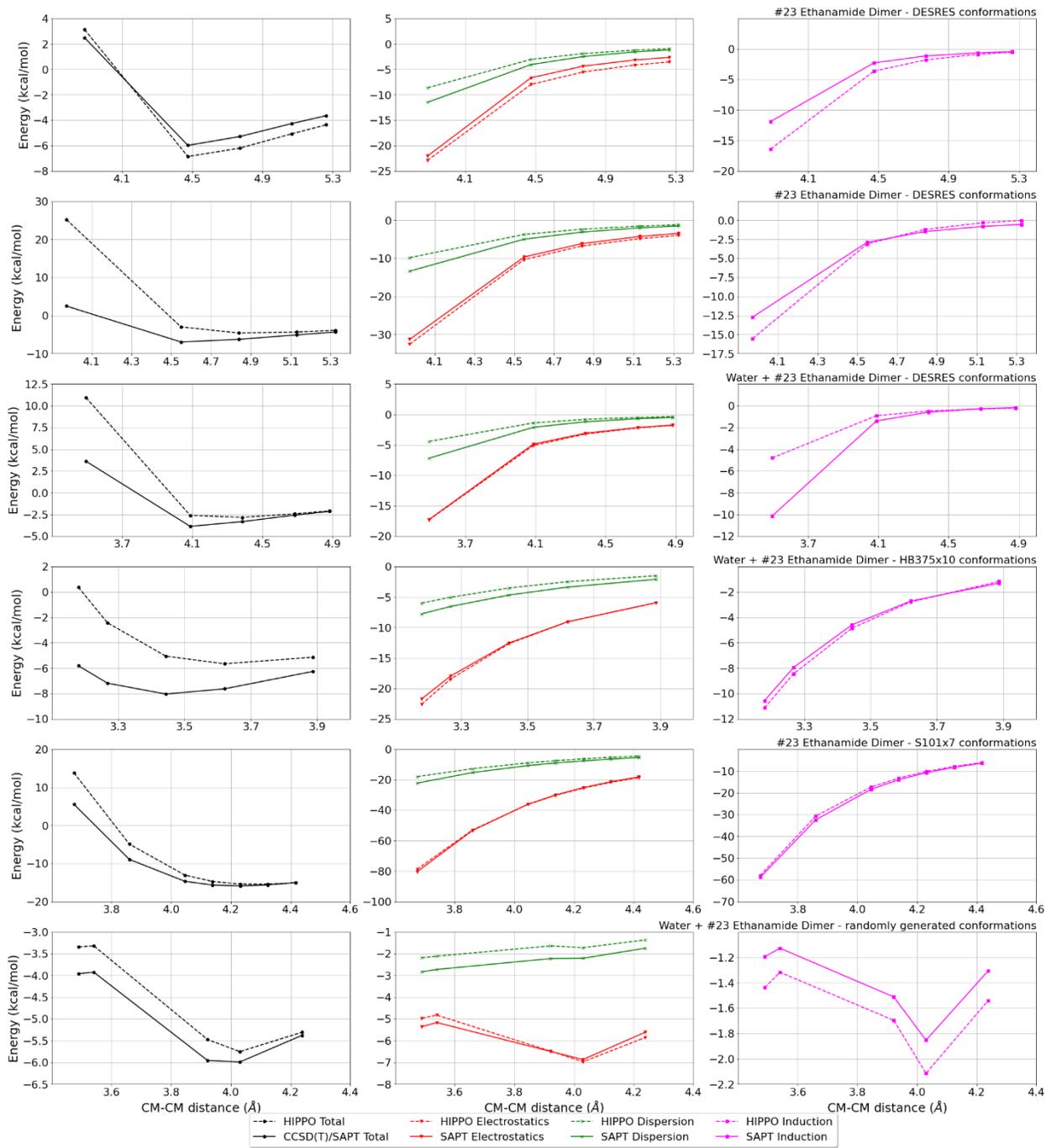
MAE	Std error	max error	#points	#count[err > 1]
0.665	1.425	21.6443	722	128

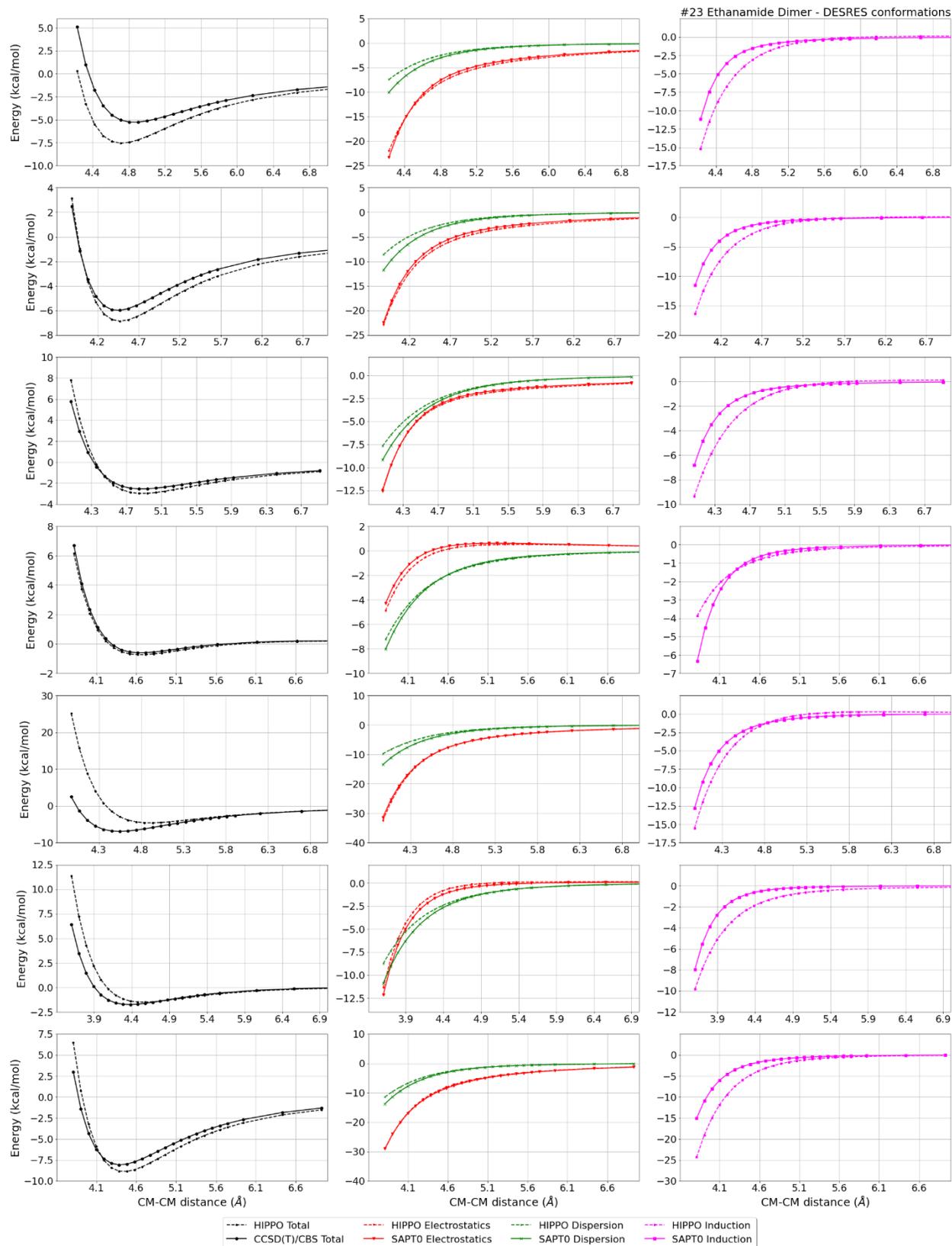
HB375x10_23-water, energy values in kcal/mol

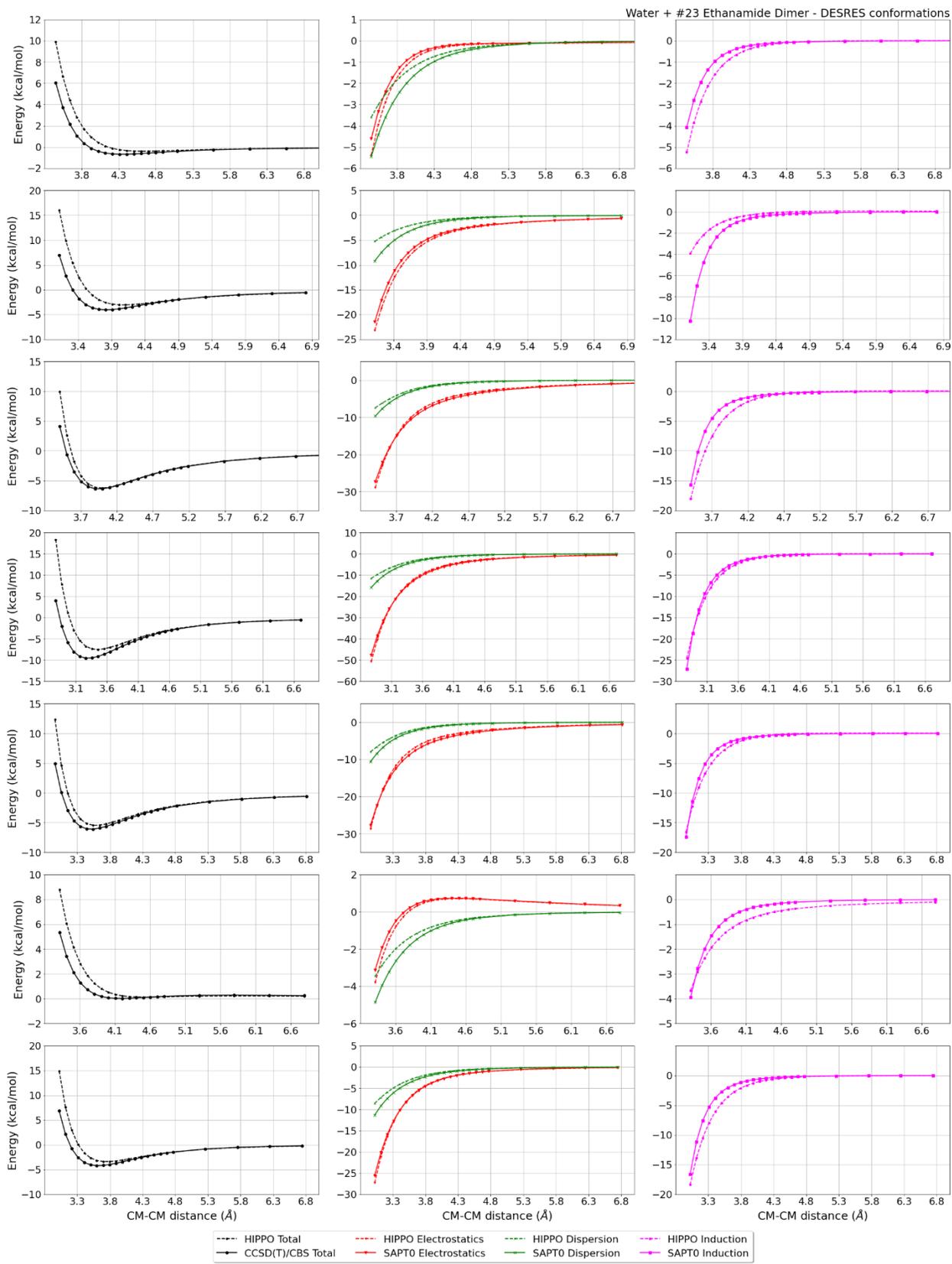
CM-CM (A)	Reference	HIPPO res	Abs diff
2.370	-3.394	4.492	7.8862
2.437	-5.811	0.013	5.8238
2.504	-7.184	-2.778	4.4063
2.571	-7.849	-4.451	3.3975
2.639	-8.044	-5.386	2.6577
2.707	-7.933	-5.834	2.0995
2.774	-7.631	-5.962	1.6693
2.979	-6.256	-5.399	0.8572
3.322	-4.049	-3.763	0.2857
4.014	-1.722	-1.700	0.0221

MAE	Std error	max error	#points	#count[err > 1]
2.911	2.396	7.8862	10	7



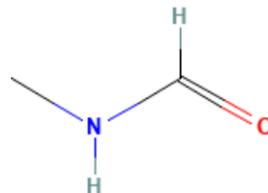






#24 N-methylformamide C2H5NO CID: 31254

ref molpol	-7.82	-5.47	-4.37, avg	-5.89
molpol	7.80	5.41	4.41, avg	5.87
rms molpol	0.03	0.06	0.04, avg	0.02



Monomer potential fitting RMS: 1.29

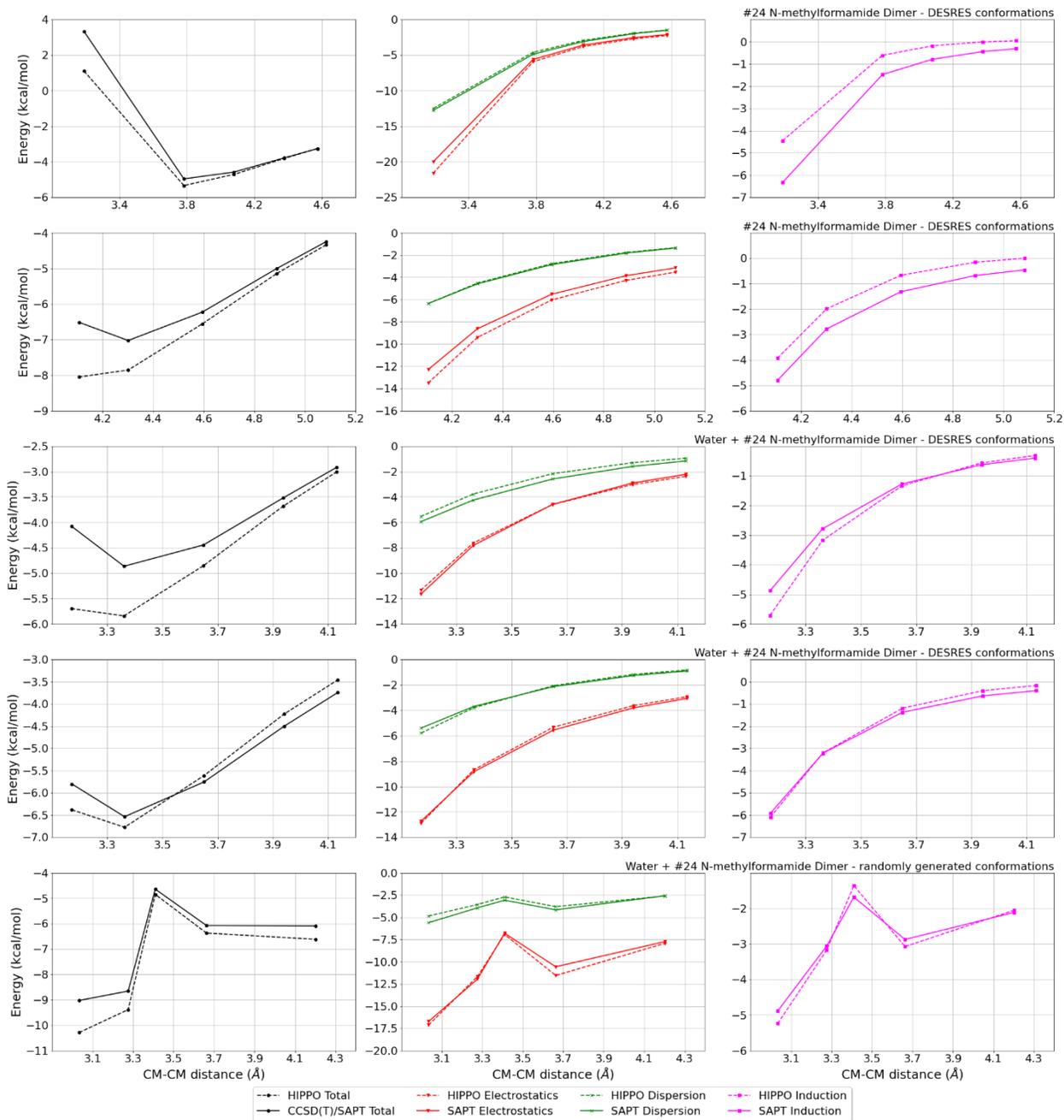
##Dimer results - Fitting to QM datasets##

DESRES_24-water, energy values in kcal/mol

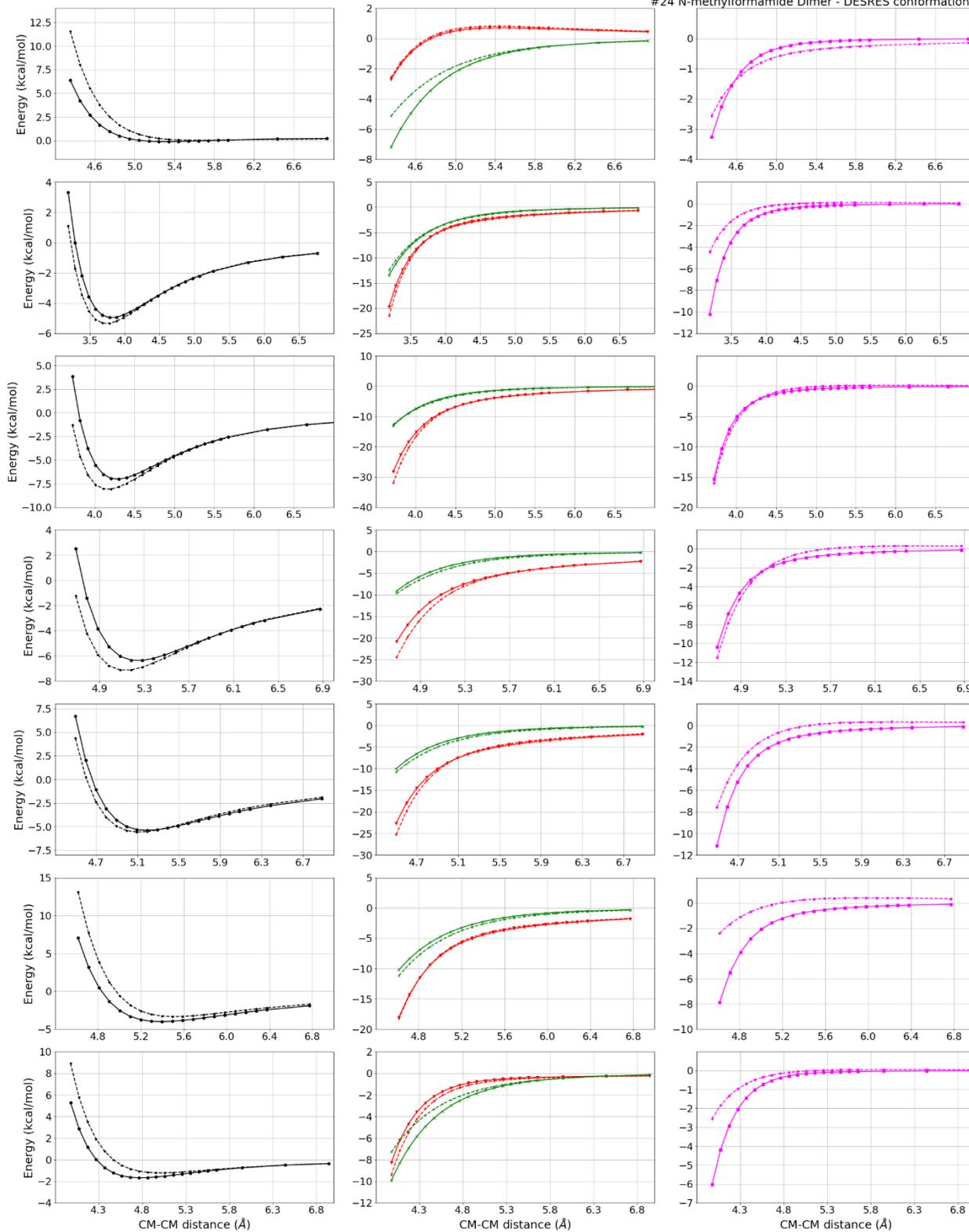
MAE	Std error	max error	#points	#count[err > 1]
0.338	0.608	4.1591	558	54

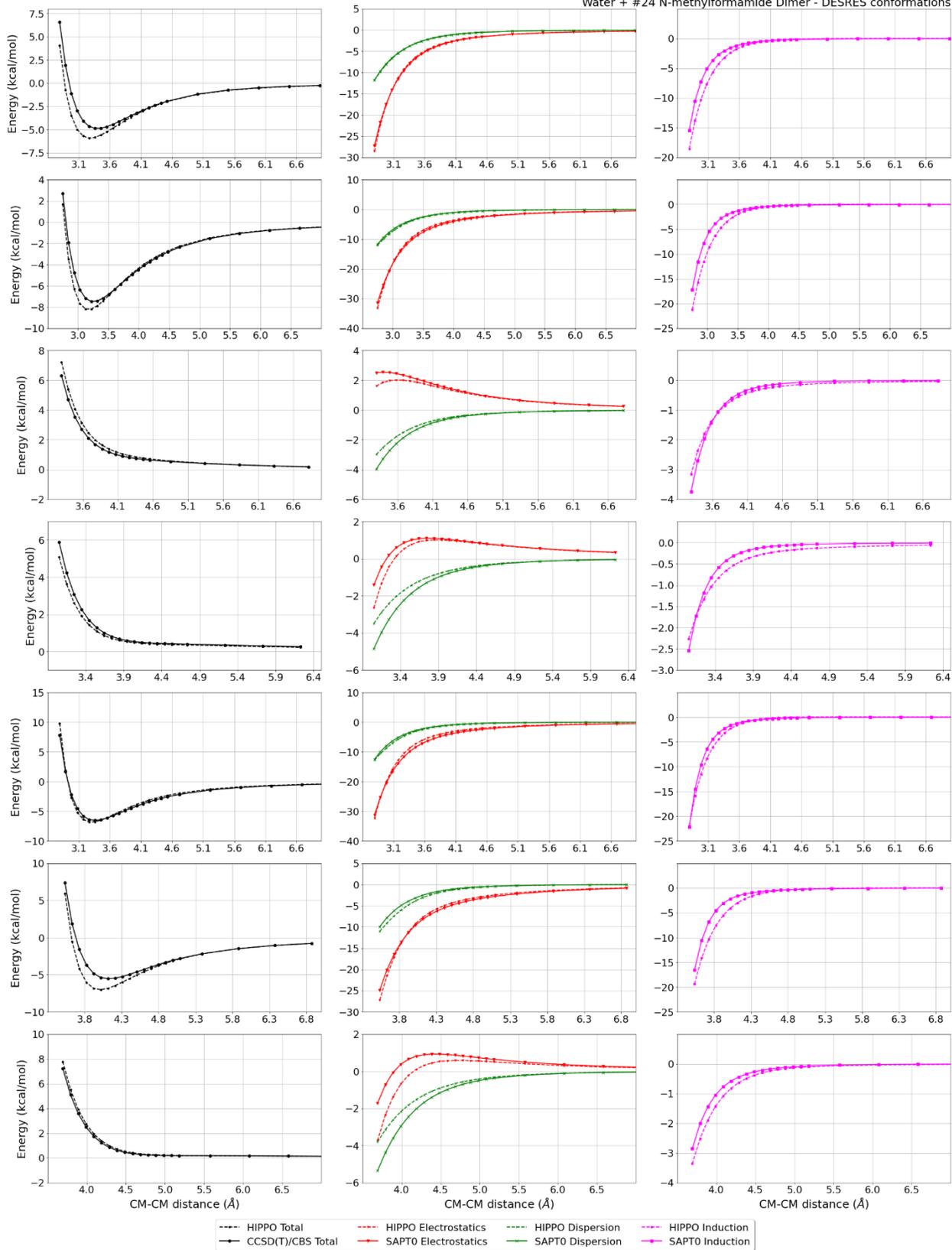
DESRES_24-24, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.432	0.795	5.7960	692	89



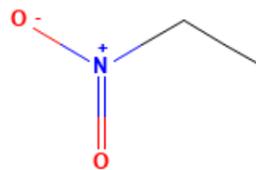
#24 N-methylformamide Dimer - DESRES conformations



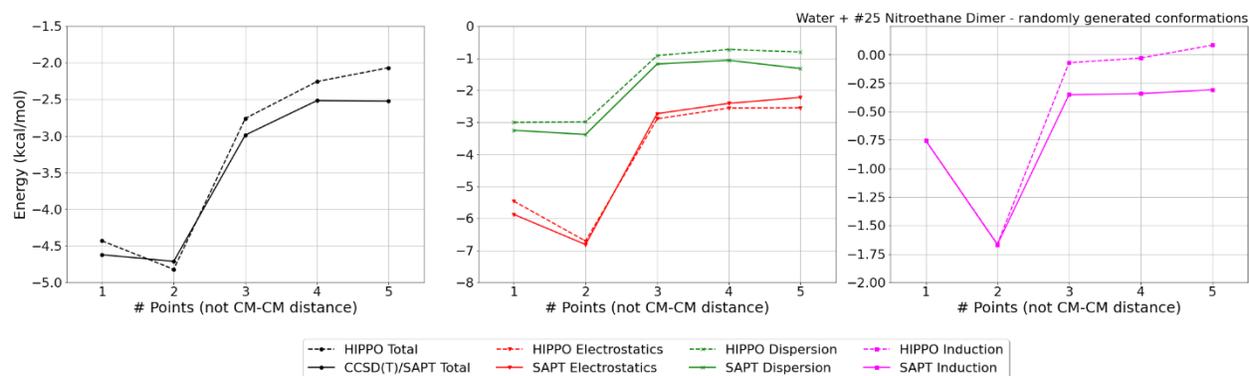


#25 Nitroethane C₂H₅NO₂ CID: 6587

ref molpol	-6.70	-7.81	-4.94, avg	-6.48
molpol	7.04	6.12	4.94, avg	6.03
rms molpol	0.34	1.69	0.01, avg	0.45

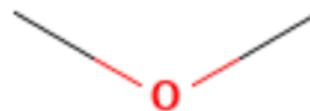


Monomer potential fitting RMS: 0.36



#26 Methoxymethane C2H6O CID: 8254

ref molpol	-5.58	-4.62	-4.49, avg	-4.90
molpol	5.62	4.57	4.53, avg	4.91
rms molpol	0.04	0.06	0.04, avg	0.01



Monomer potential fitting RMS: 0.40

##Dimer results - Fitting to QM datasets##

DESRES_26-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.143	0.331	3.6906	562	17

DESRES_26-26, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.130	0.334	3.4603	811	23

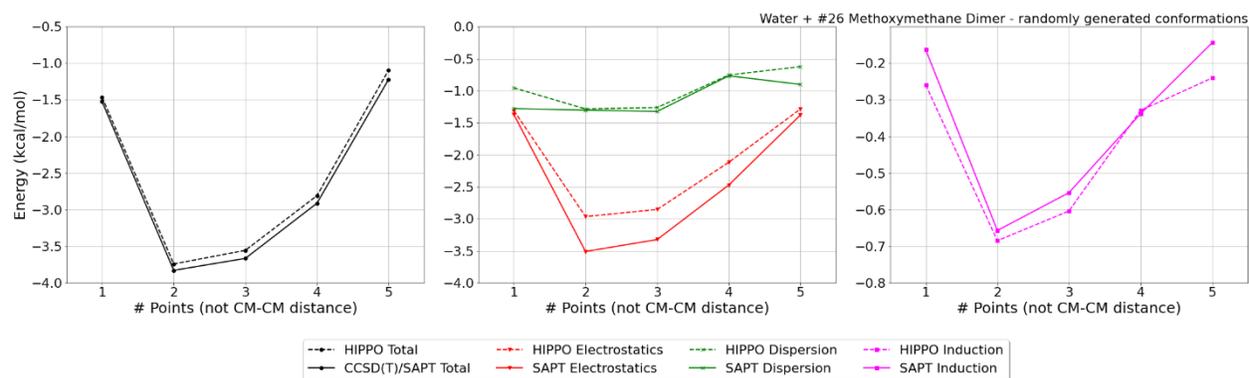
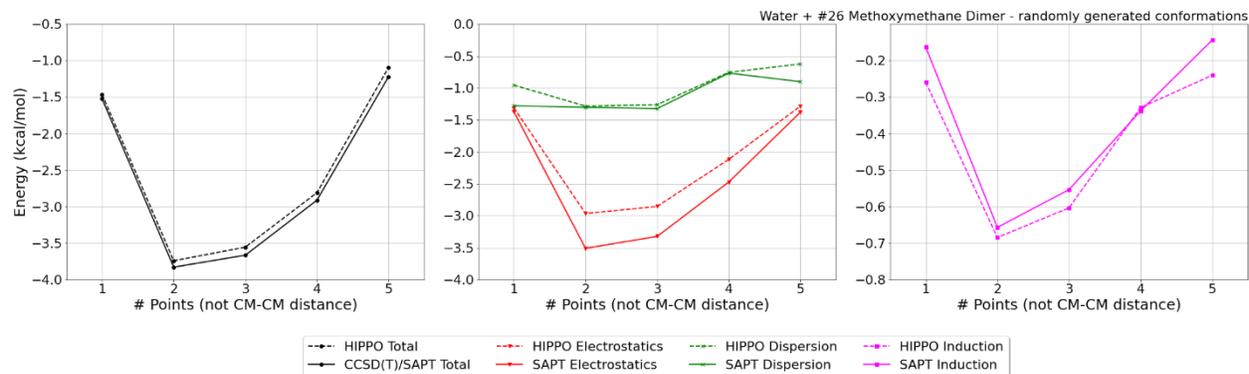
HB375x10_26-water, energy values in kcal/mol

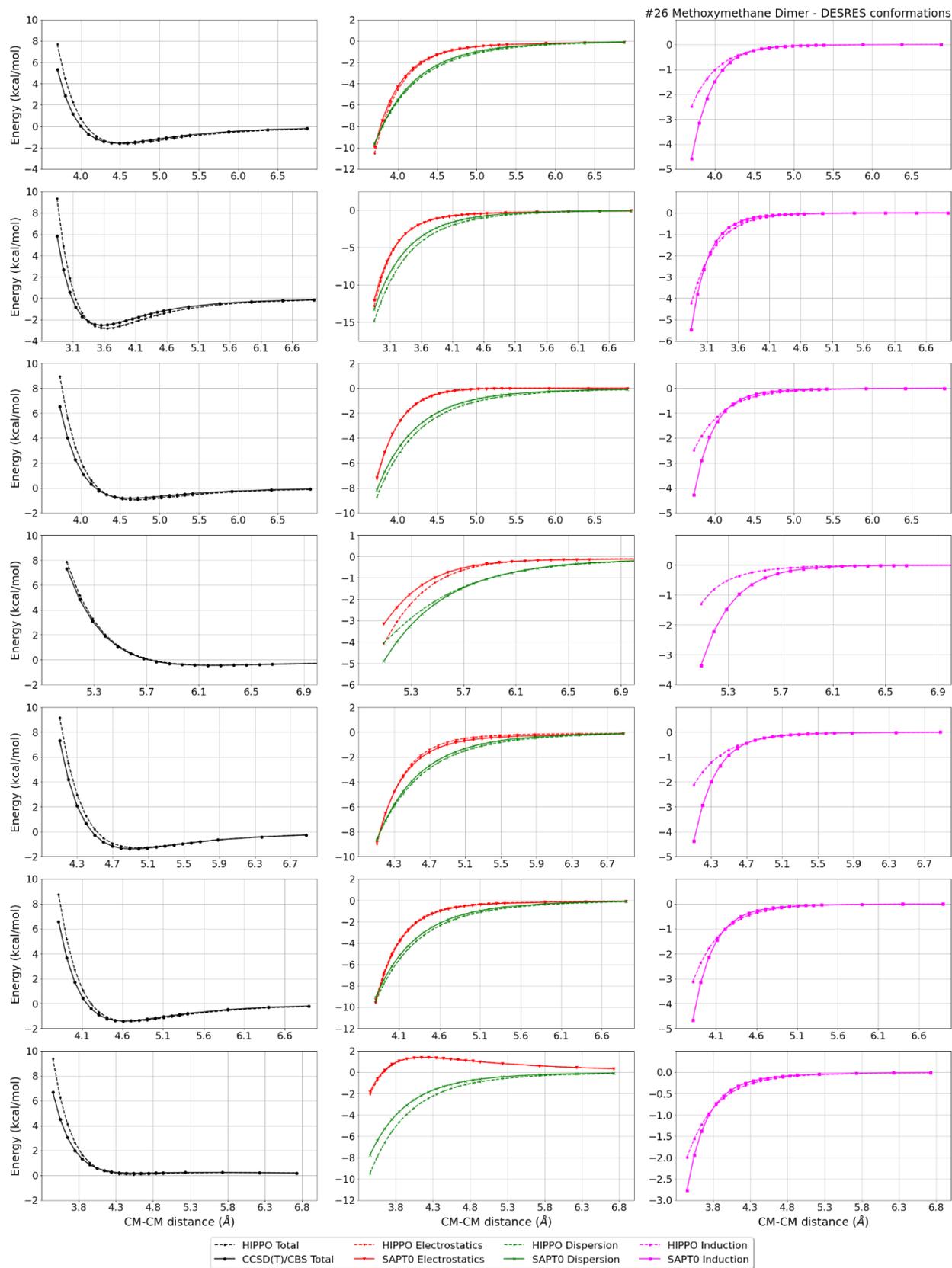
CM-CM (A)	Reference	HIPPO res	Abs diff
2.027	-1.892	-1.680	0.2120
2.094	-4.073	-4.235	-0.1622
2.161	-5.297	-5.539	-0.2420
2.228	-5.882	-6.076	-0.1944
2.295	-6.049	-6.153	-0.1040
2.362	-5.950	-5.961	-0.0114
2.428	-5.690	-5.623	0.0671
2.629	-4.538	-4.350	0.1884
2.964	-2.789	-2.613	0.1759
3.634	-1.096	-1.033	0.0628

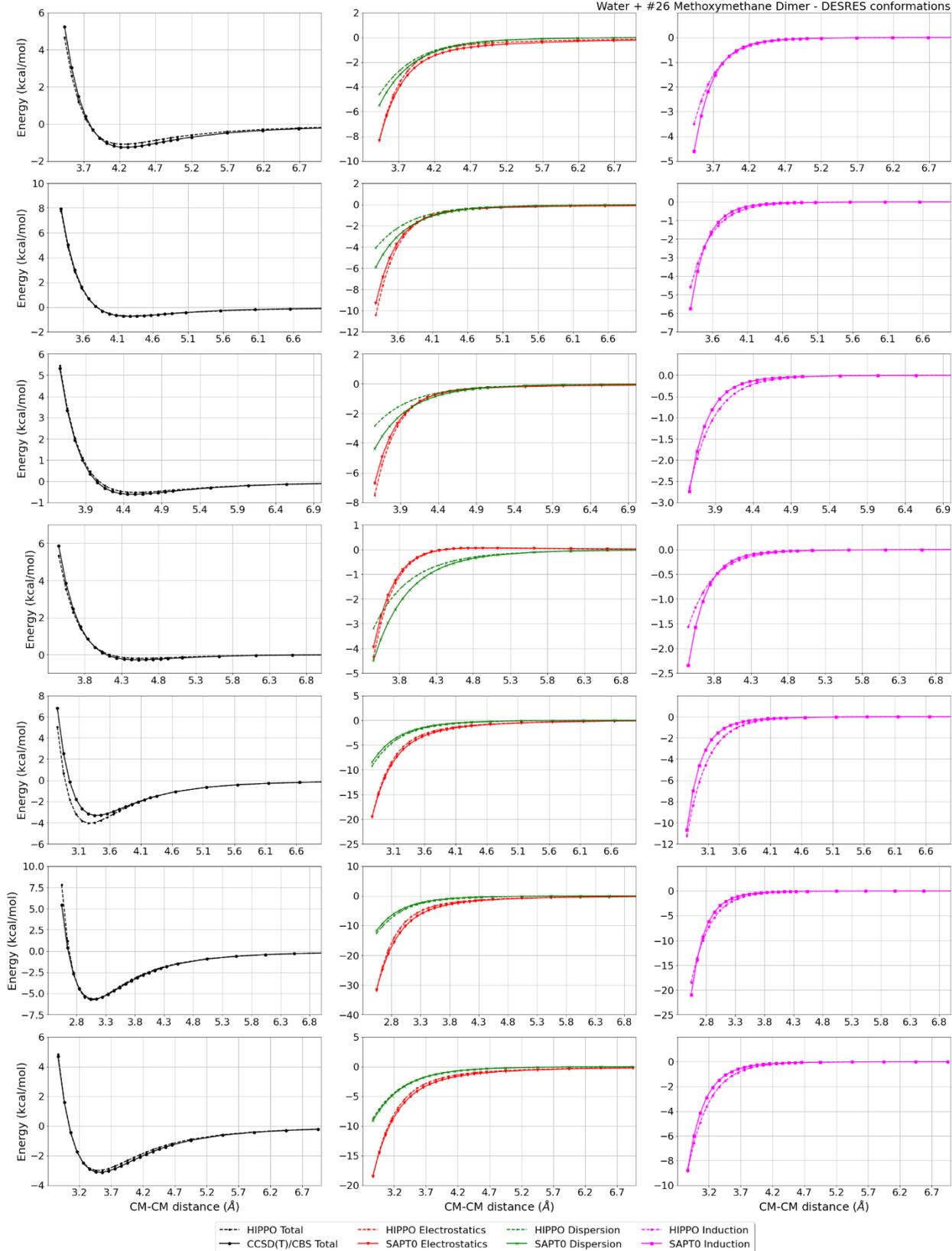
MAE	Std error	max error	#points	#count[err > 1]
0.142	0.072	0.2420	10	0

Liquid Methoxymethane @ 240.00 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
739.99	742.08	0.3	21.73	21.72	0.0	7.97	6.88	15.9	4000







#27 Ethanol C2H6O CID: 702

ref molpol	-5.38	-4.78	-4.46, avg	-4.87
molpol	5.36	4.70	4.45, avg	4.84
rms molpol	0.02	0.09	0.01, avg	0.04



Monomer potential fitting RMS: 0.61

##Dimer results - Fitting to QM datasets##

DESRES_27-water, energy values in kcal/mol

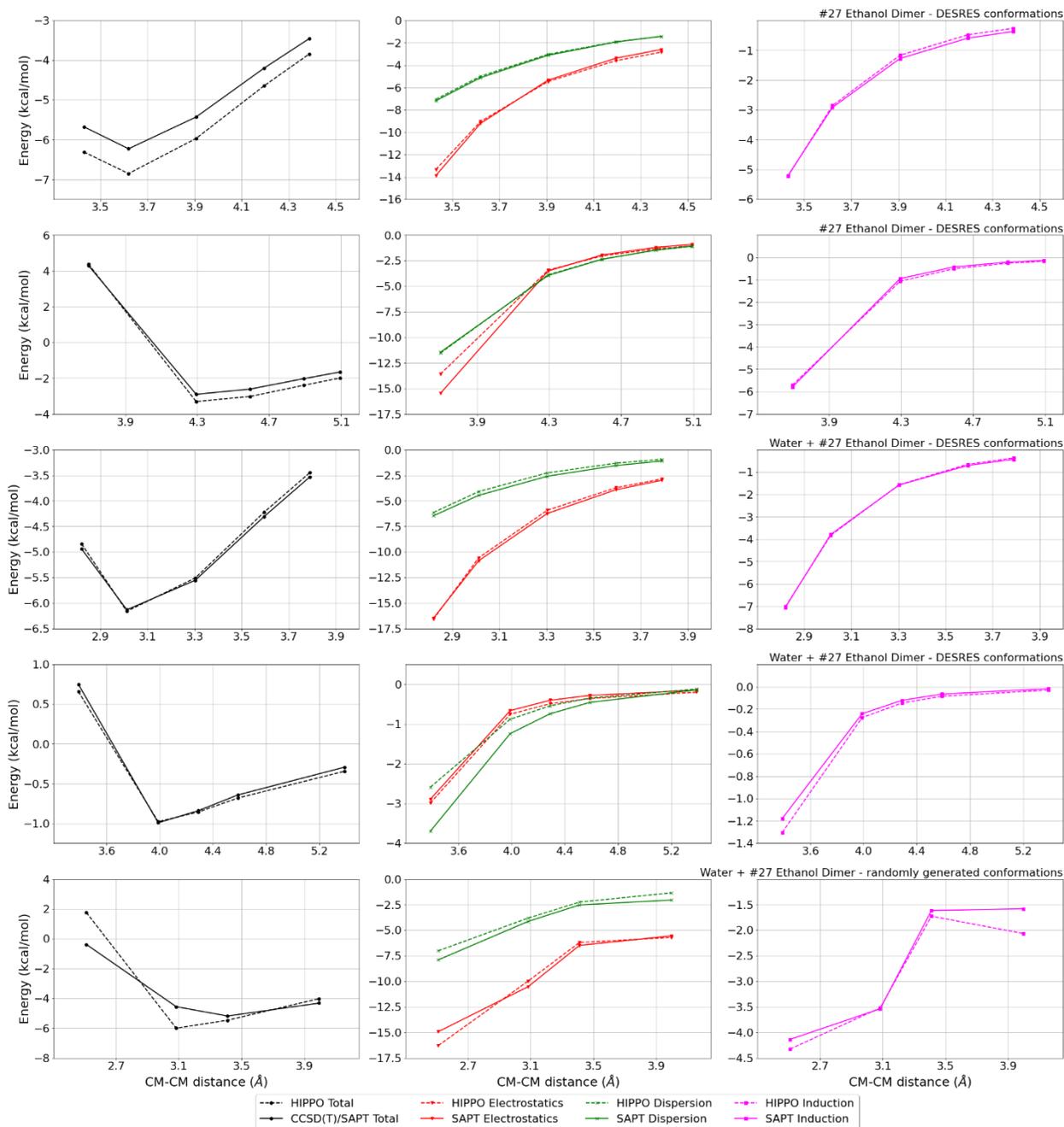
MAE	Std error	max error	#points	#count[err > 1]
0.229	0.517	4.7960	555	34

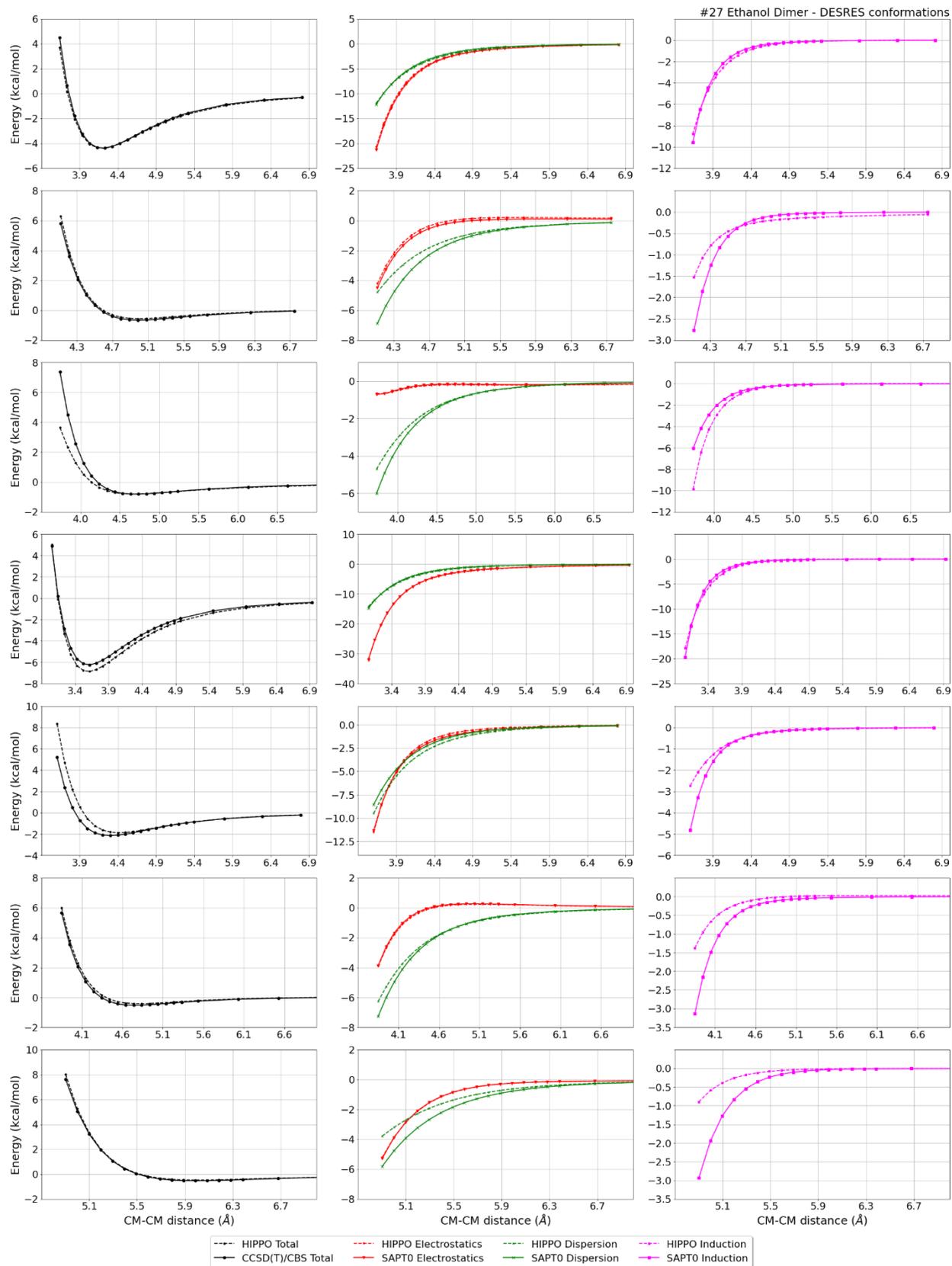
DESRES_27-27, energy values in kcal/mol

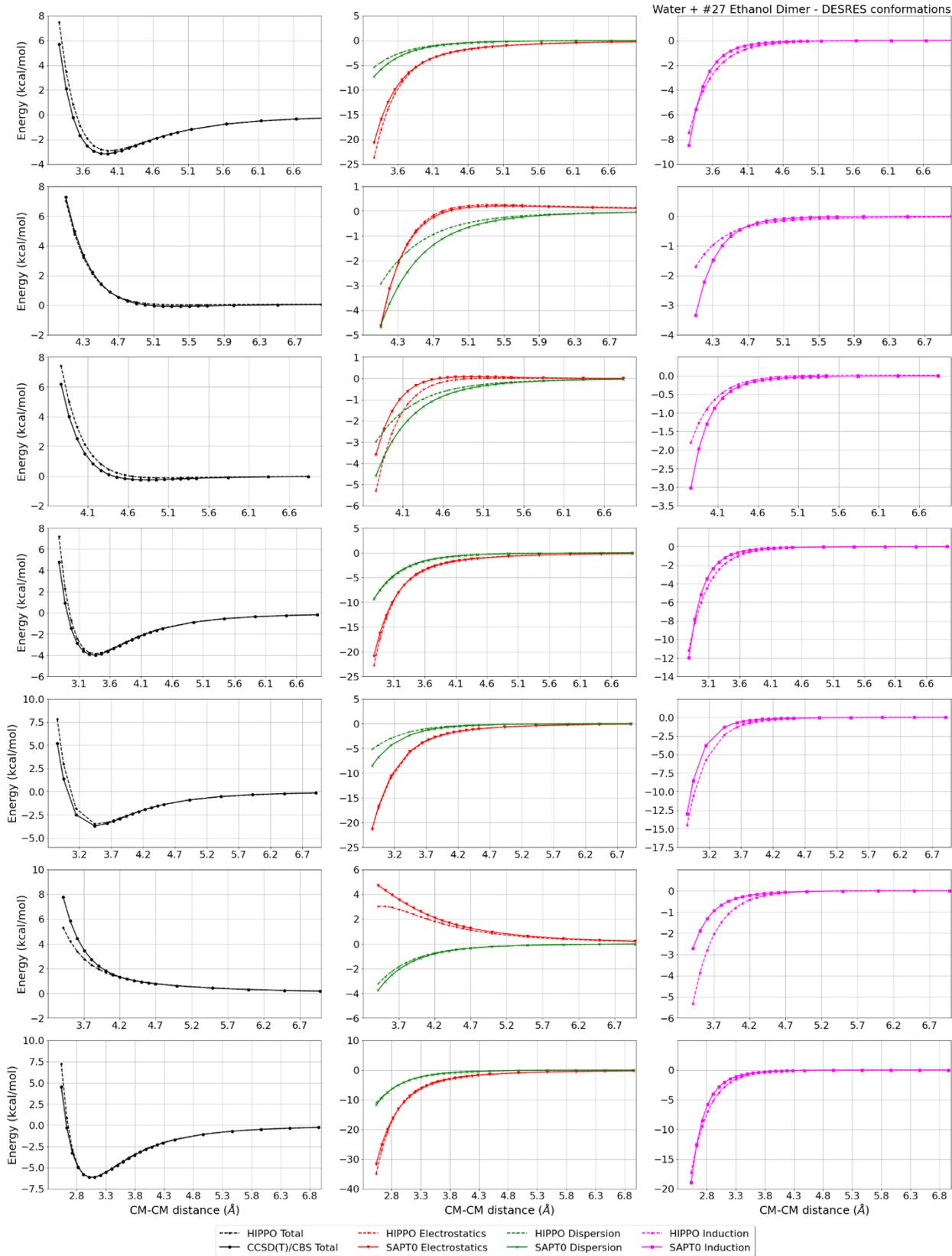
MAE	Std error	max error	#points	#count[err > 1]
0.150	0.337	4.1448	813	15

Liquid Ethanol @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
799.69	784.80	1.9	48.37	42.32	14.3	12.00	24.85	51.7	2000







#29 Methylsulfanylmethane C2H6S2 CID: 12232

ref molpol	-12.64	-8.95	-9.46, avg	-10.35
molpol	12.60	10.13	9.37, avg	10.70
rms molpol	0.04	1.18	0.09, avg	0.35



Monomer potential fitting RMS: 0.72

##Dimer results - Fitting to QM datasets##

DESRES_29-water, energy values in kcal/mol

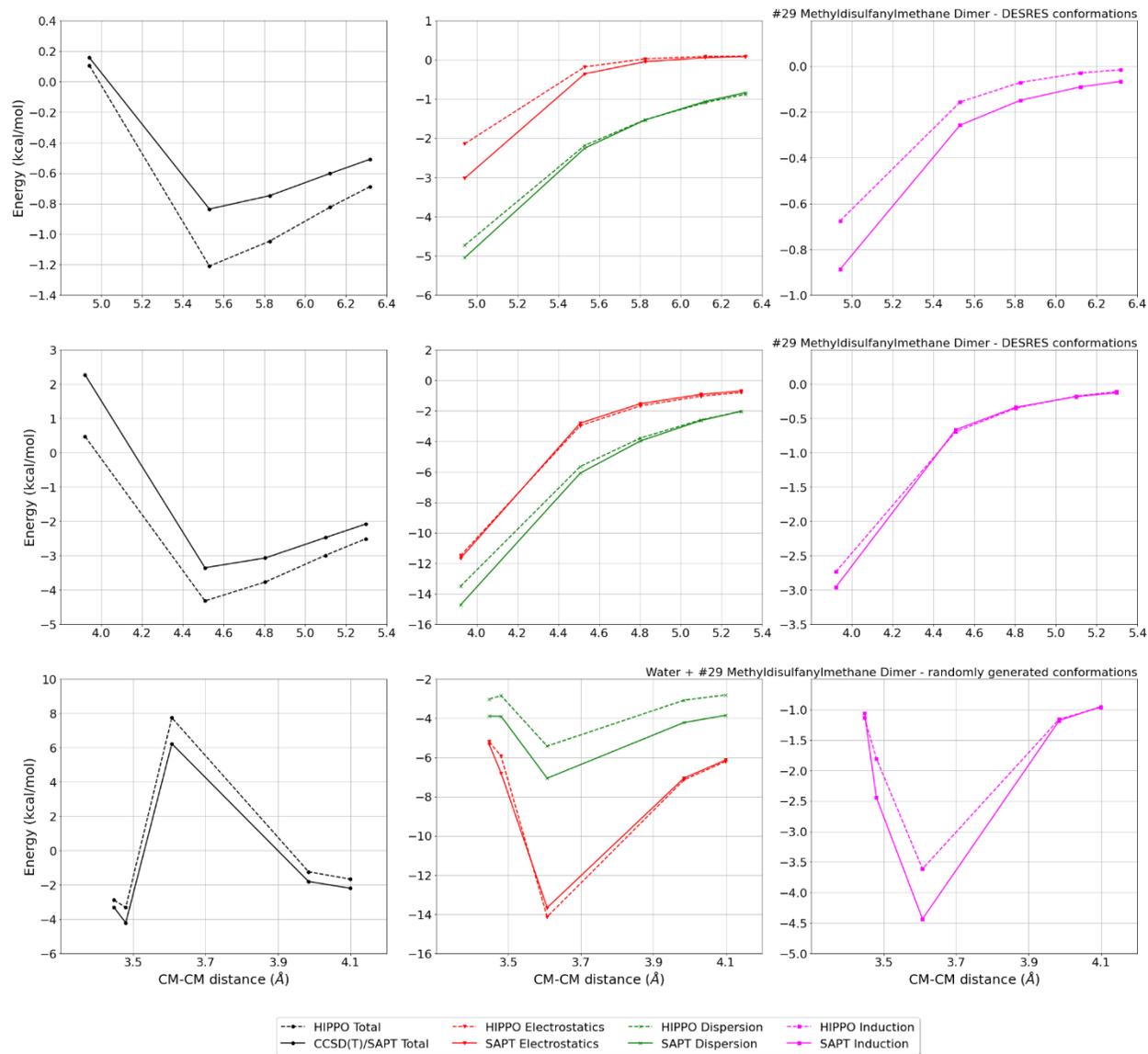
MAE	Std error	max error	#points	#count[err > 1]
0.269	0.545	5.8014	564	31

DESRES_29-29, energy values in kcal/mol

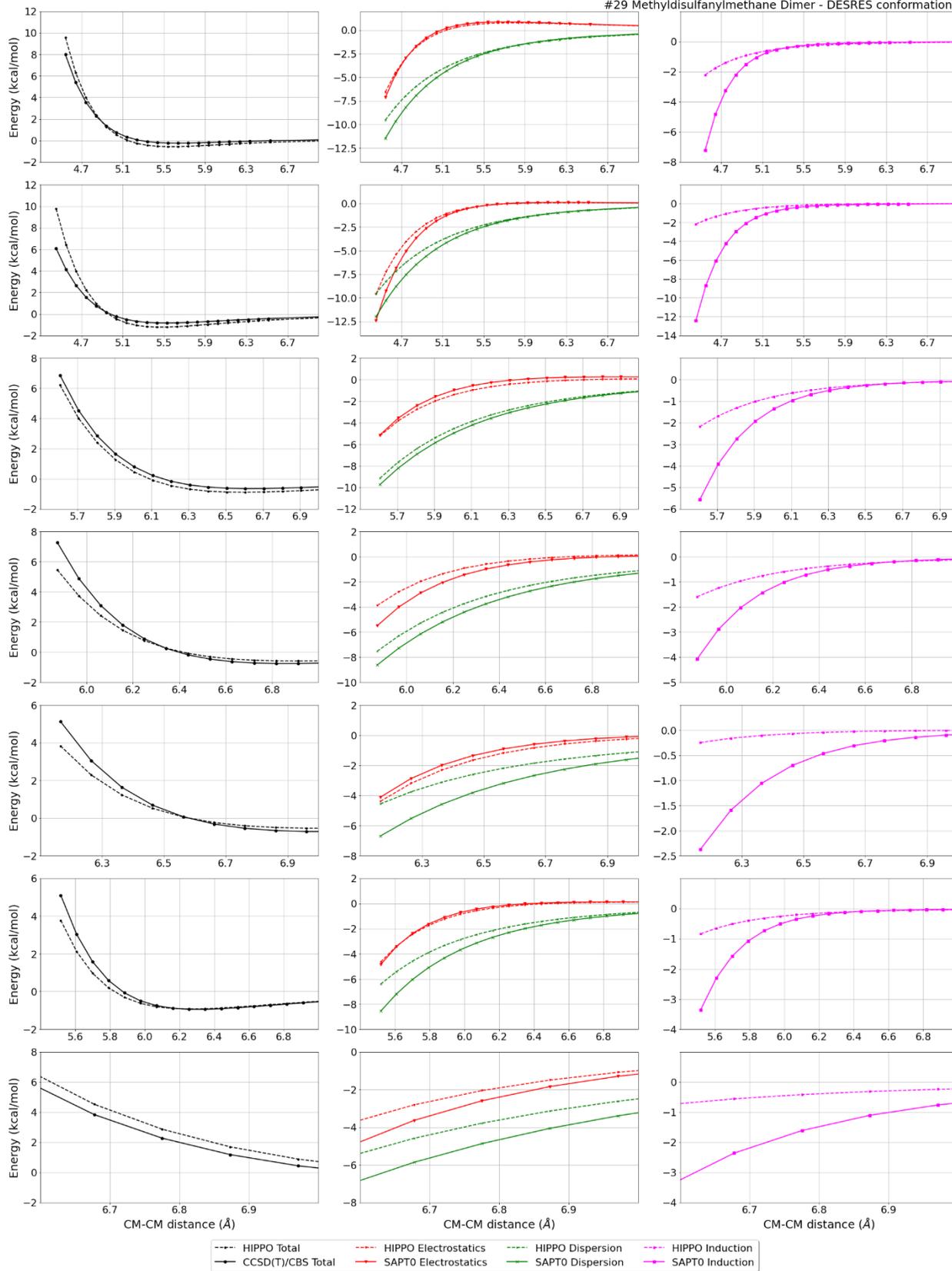
MAE	Std error	max error	#points	#count[err > 1]
0.247	0.365	3.5495	528	24

Liquid Methylsulfanylmethane @ 298.15 K

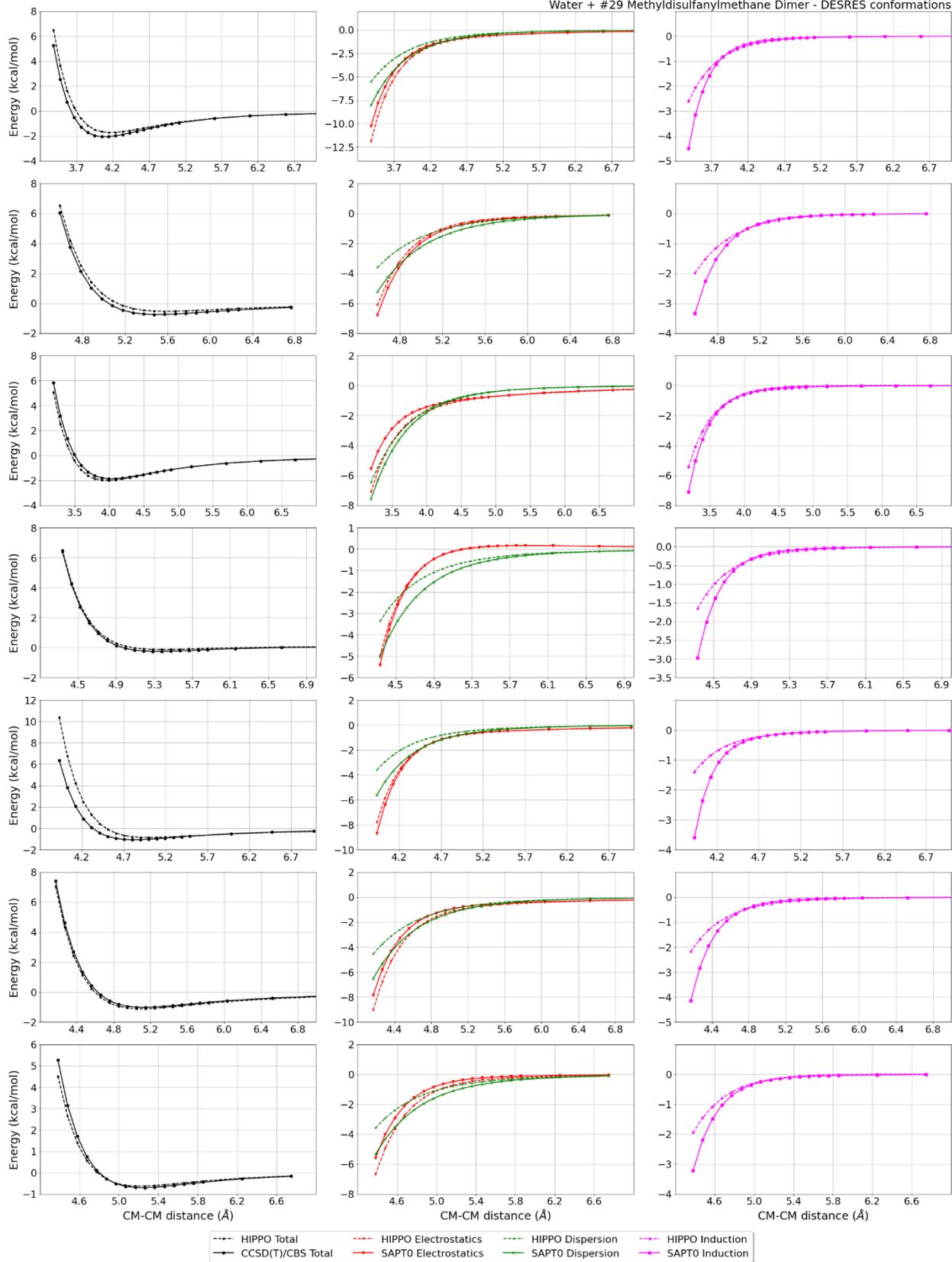
Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1066.43	1057.32	0.9	38.44	38.32	0.3	8.00	9.60	16.6	2000



#29 Methylsulfonylmethane Dimer - DESRES conformations

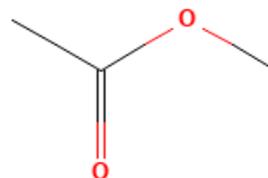


Water + #29 Methylsulfonylmethane Dimer - DESRES conformations



#41 Methyl acetate C3H6O2 CID: 6584

ref molpol	-7.98	-6.88	-5.36, avg	-6.74
molpol	7.97	6.89	5.37, avg	6.74
rms molpol	0.01	0.01	0.01, avg	0.00



Monomer potential fitting RMS: 0.68

##Dimer results - Fitting to QM datasets##

DESRES_41-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.222	0.368	4.5641	563	24

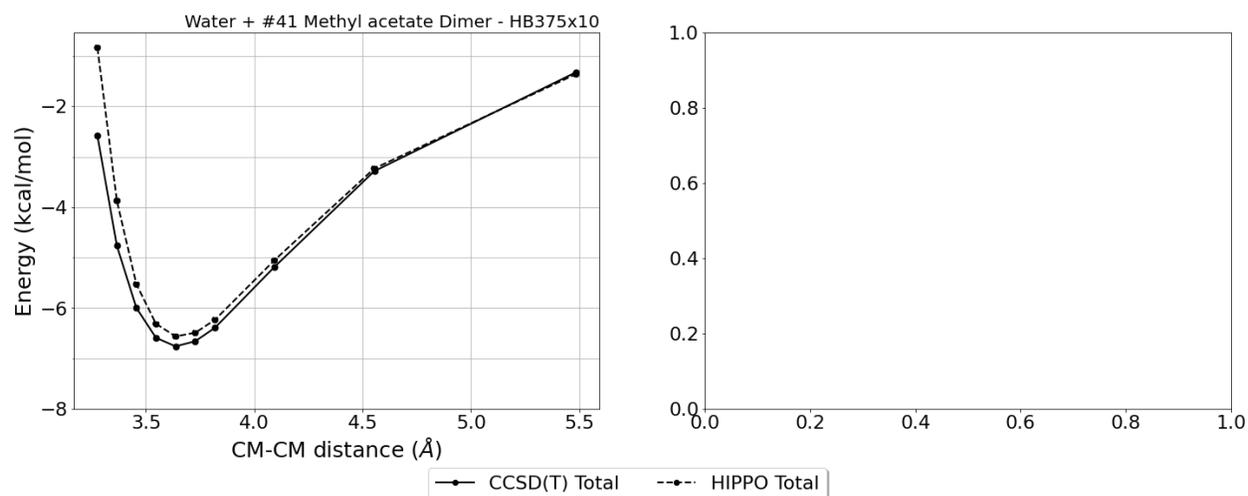
DESRES_41-41, energy values in kcal/mol

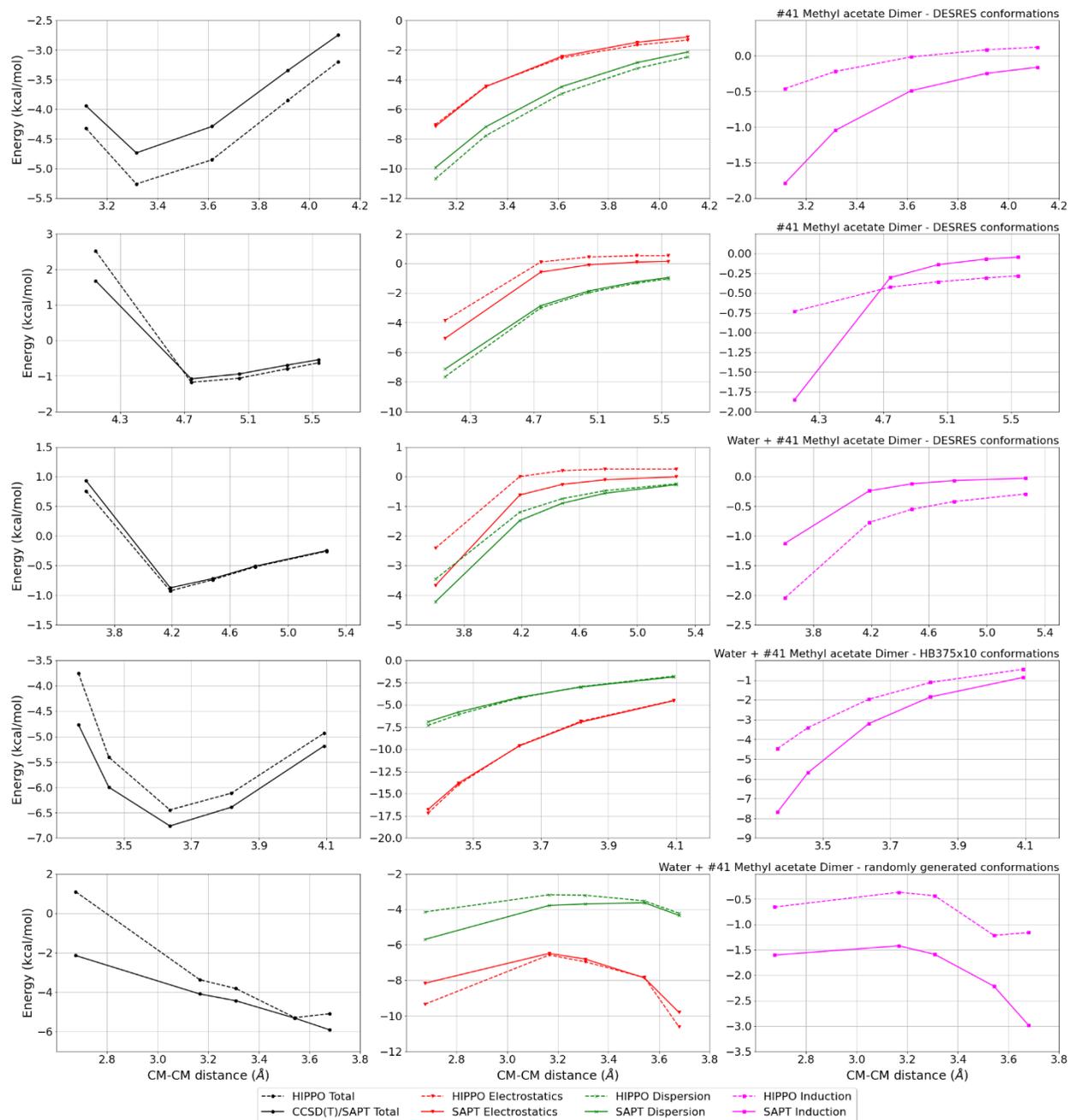
MAE	Std error	max error	#points	#count[err > 1]
0.434	0.645	4.7079	547	67

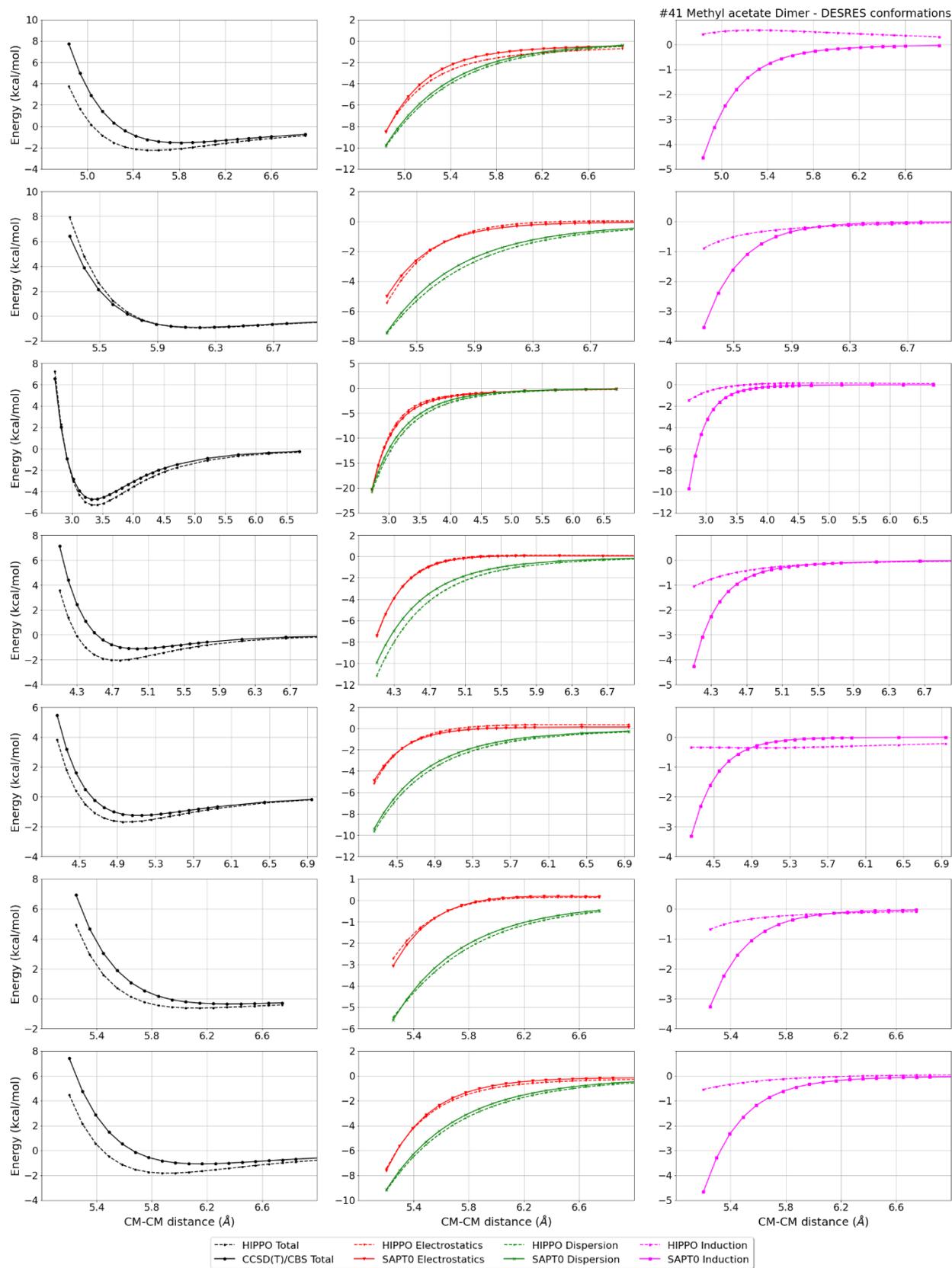
HB375x10_41-water, energy values in kcal/mol

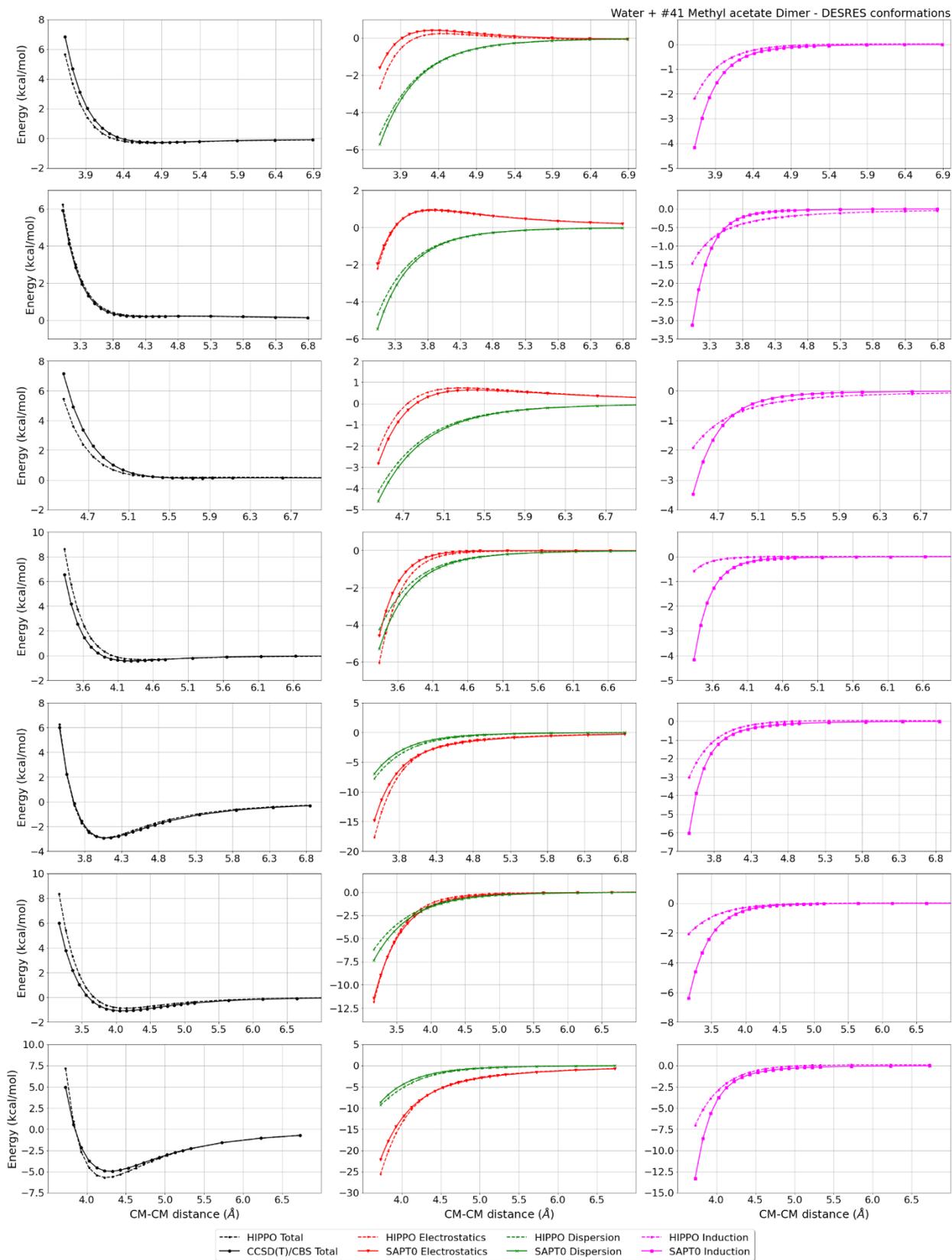
CM-CM (Å)	Reference	HIPPO res	Abs diff
2.635	-2.580	-0.836	1.7444
2.707	-4.765	-3.874	0.8915
2.780	-5.997	-5.527	0.4701
2.852	-6.590	-6.312	0.2779
2.925	-6.760	-6.562	0.1977
2.998	-6.659	-6.491	0.1683
3.071	-6.391	-6.233	0.1583
3.292	-5.185	-5.052	0.1333
3.663	-3.285	-3.232	0.0531
4.411	-1.326	-1.362	-0.0357

MAE	Std error	max error	#points	#count[err > 1]
0.413	0.504	1.7444	10	1









#42 1,3-dioxolane C3H6O2 CID: 12586

ref molpol	-7.23	-5.82	-6.30, avg	-6.45
molpol	7.18	6.13	6.02, avg	6.44
rms molpol	0.05	0.32	0.28, avg	0.00



Monomer potential fitting RMS: 0.62

##Dimer results - Fitting to QM datasets##

DESRES_42-water, energy values in kcal/mol

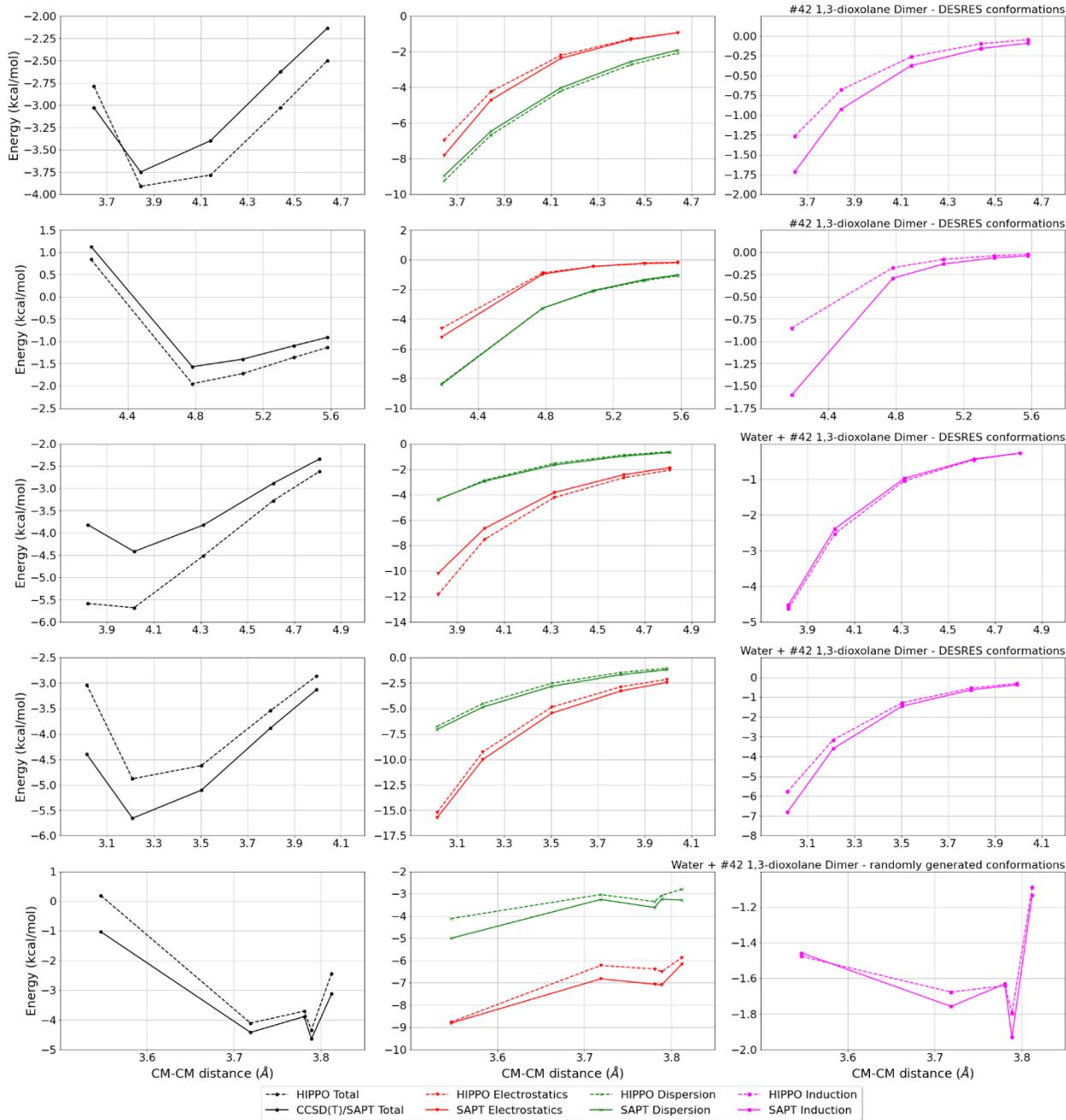
MAE	Std error	max error	#points	#count[err > 1]
0.273	0.498	5.2339	557	32

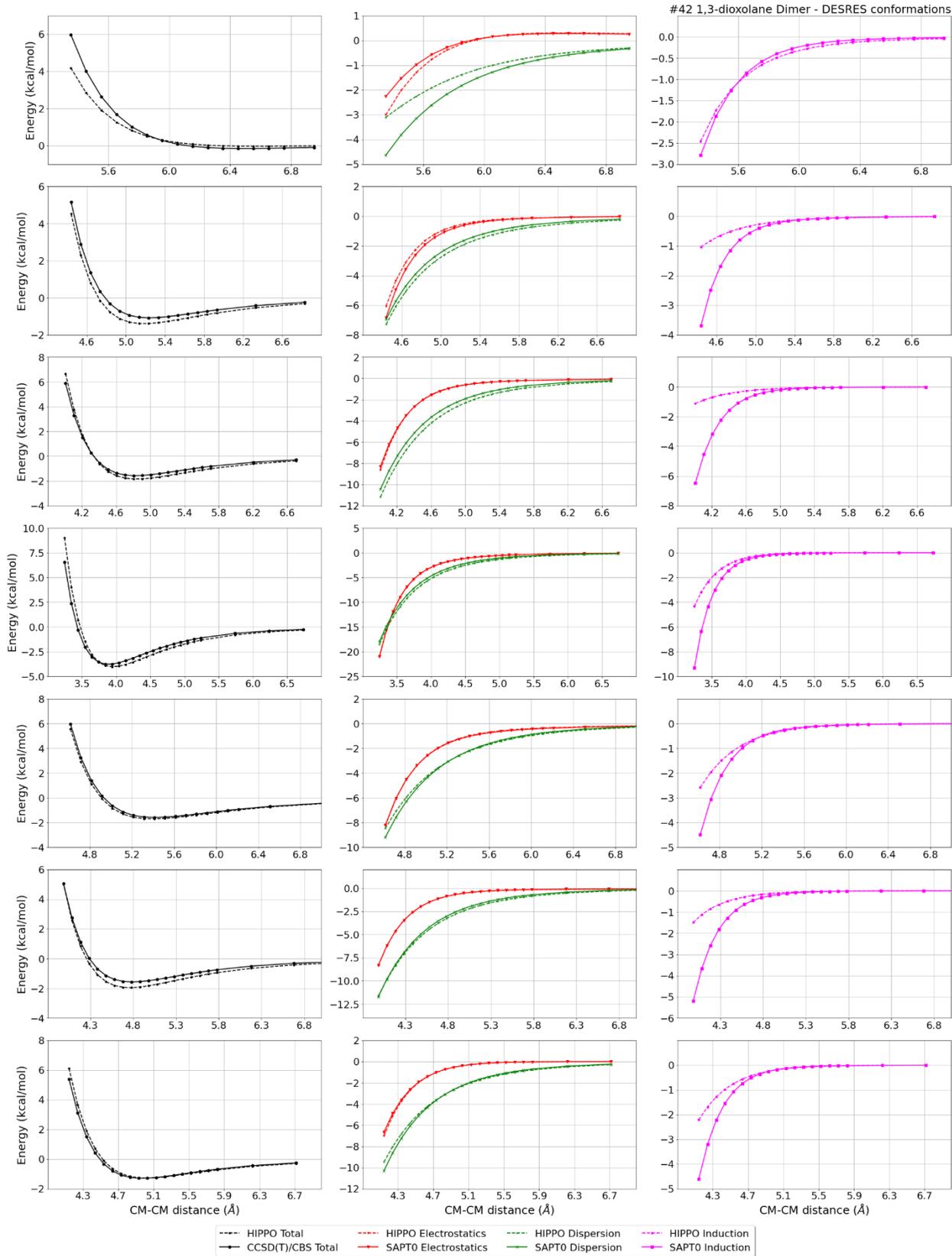
DESRES_42-42, energy values in kcal/mol

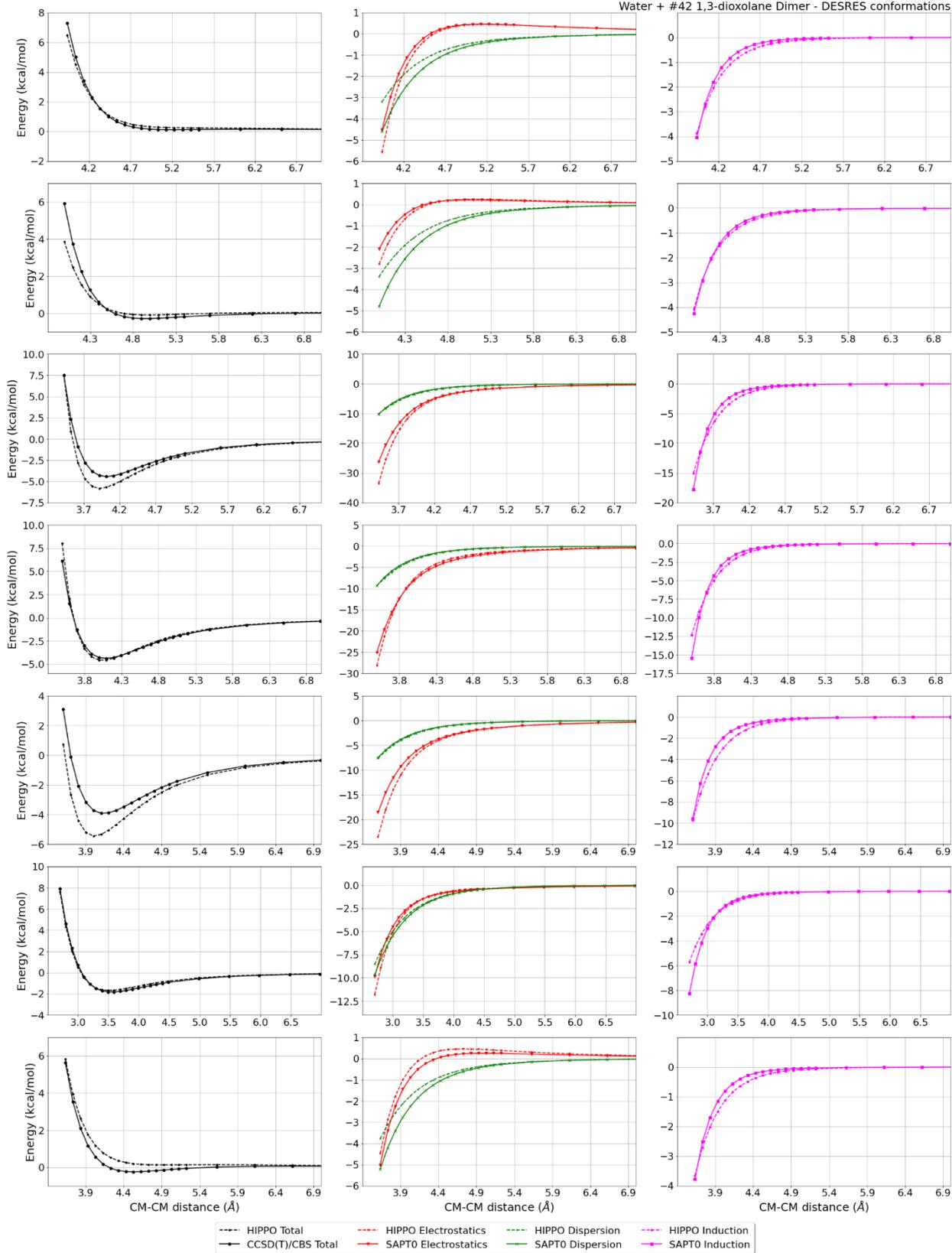
MAE	Std error	max error	#points	#count[err > 1]
0.204	0.290	2.3910	545	15

Liquid 1,3-dioxolane @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1062.13	1064.40	0.2	35.44	35.60	0.4	4.21	-1.00	-521.0	1000







#43 2-iodopropane C3H7I CID: 6362

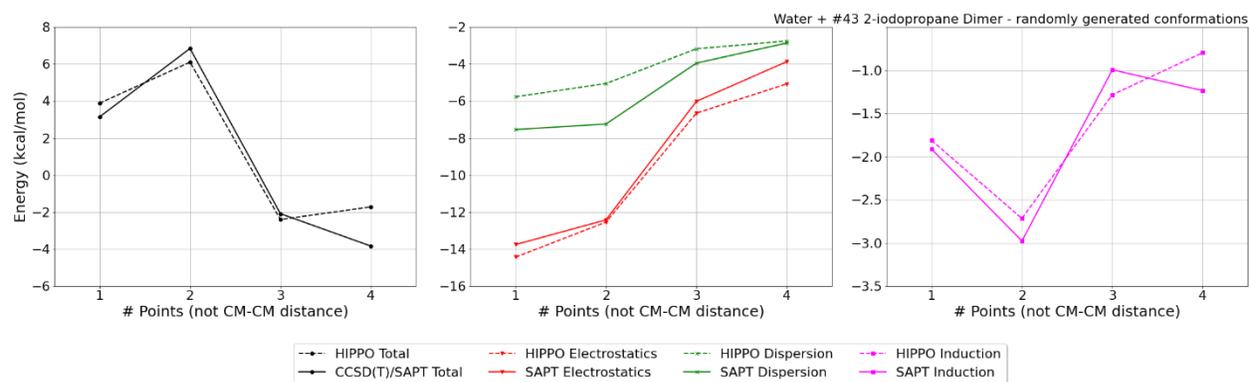
ref molpol	-13.18	-10.31	-9.27, avg	-10.92
molpol	13.05	10.24	9.31, avg	10.87
rms molpol	0.13	0.07	0.04, avg	0.05



Monomer potential fitting RMS: 0.75

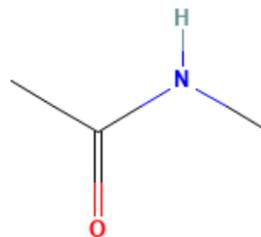
Liquid 2-iodopropane @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1333.11	1737.25	23.3	33.21	34.20	2.9	7.23	8.19	11.7	1000



#46 N-methylacetamide C3H7NO CID: 6582

ref molpol	-8.83	-7.86	-5.86, avg	-7.52
molpol	8.66	7.91	5.96, avg	7.51
rms molpol	0.17	0.05	0.10, avg	0.01



Monomer potential fitting RMS: 0.61

##Dimer results - Fitting to QM datasets##

DESRES_46-46, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.683	1.189	10.4610	509	81

DESRES_46-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.696	1.638	15.0406	552	86

HB375x10_46-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
2.581	-3.722	-2.801	0.9205
2.650	-6.107	-5.472	0.6346
2.719	-7.465	-6.987	0.4778
2.729	-1.783	0.318	2.1011
2.789	-8.126	-7.757	0.3688
2.809	-3.630	-1.654	1.9762
2.858	-8.321	-8.045	0.2756
2.889	-4.659	-2.876	1.7832
2.928	-8.214	-8.022	0.1916
2.969	-5.146	-3.589	1.5574
2.998	-7.916	-7.801	0.1146
3.050	-5.282	-3.957	1.3248
3.130	-5.199	-4.096	1.1032
3.209	-6.543	-6.597	-0.0536
3.210	-4.983	-4.083	0.8998
3.452	-4.048	-3.607	0.4410
3.564	-4.302	-4.435	-0.1330
3.856	-2.641	-2.562	0.0786
4.283	-1.860	-1.936	-0.0757
4.665	-1.225	-1.287	-0.0623

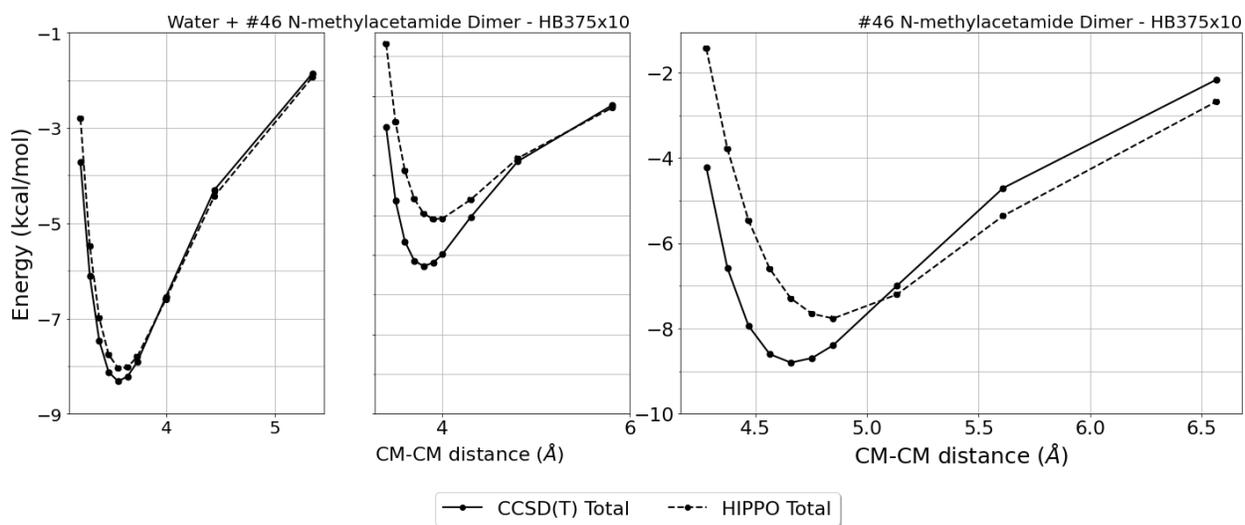
MAE	Std error	max error	#points	#count[err > 1]
0.729	0.673	2.1011	20	6

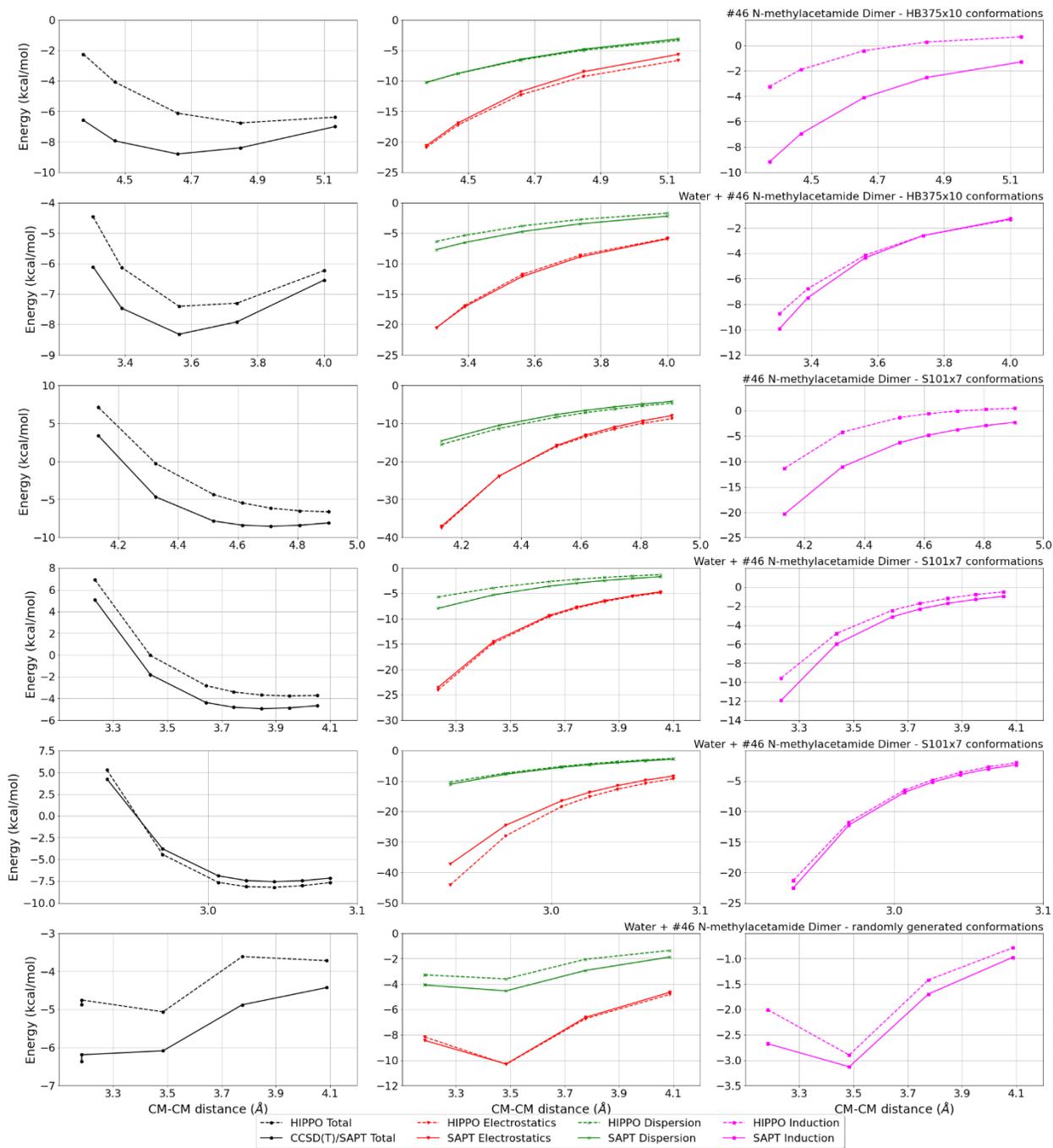
HB375x10_46-46, energy values in kcal/mol

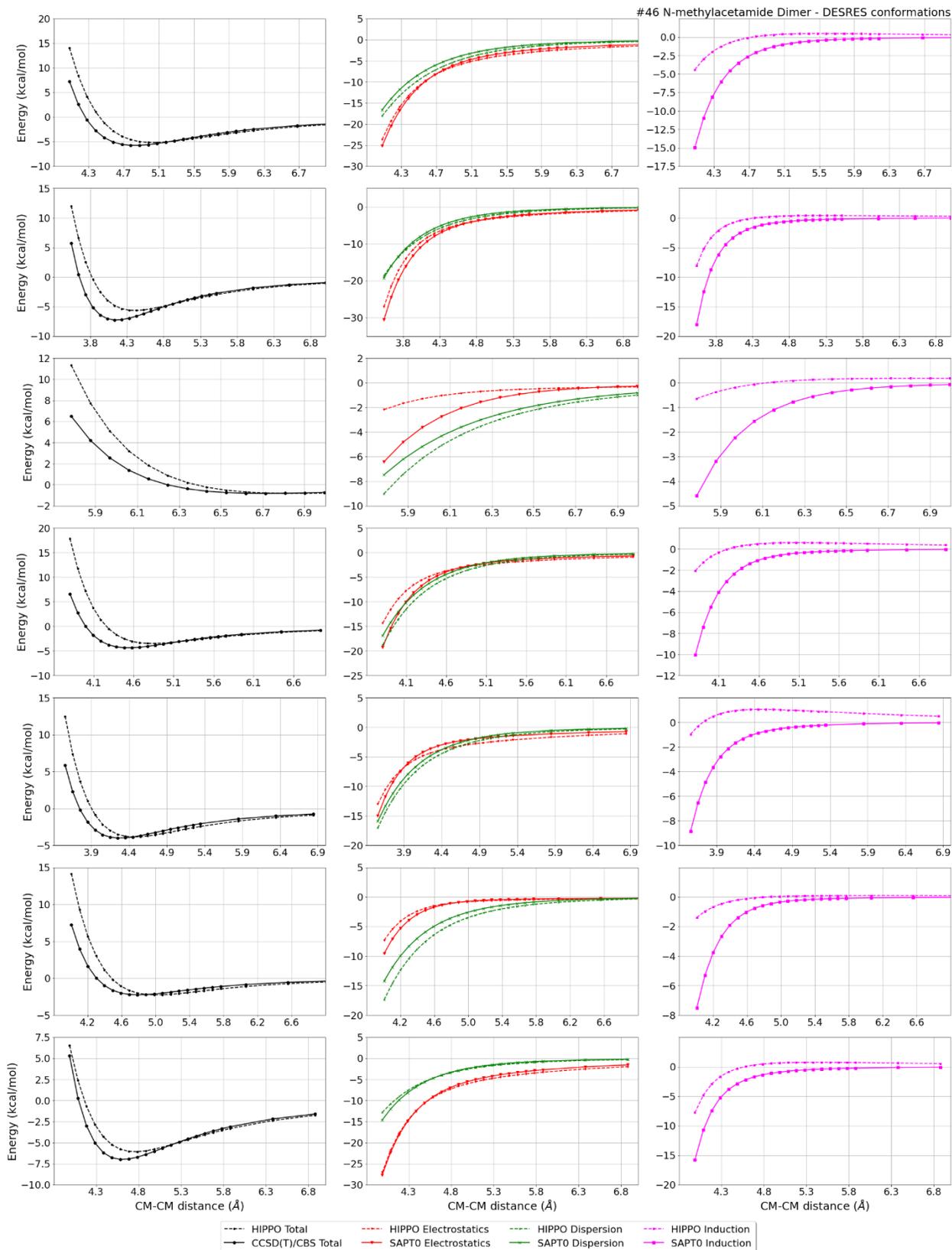
CM-CM (A)	Reference	HIPPO res	Abs diff
2.140	-4.212	-1.429	2.7826
2.187	-6.582	-3.783	2.7990
2.234	-7.938	-5.466	2.4723
2.282	-8.602	-6.598	2.0040
2.329	-8.801	-7.294	1.5065
2.377	-8.694	-7.655	1.0385
2.424	-8.393	-7.765	0.6278

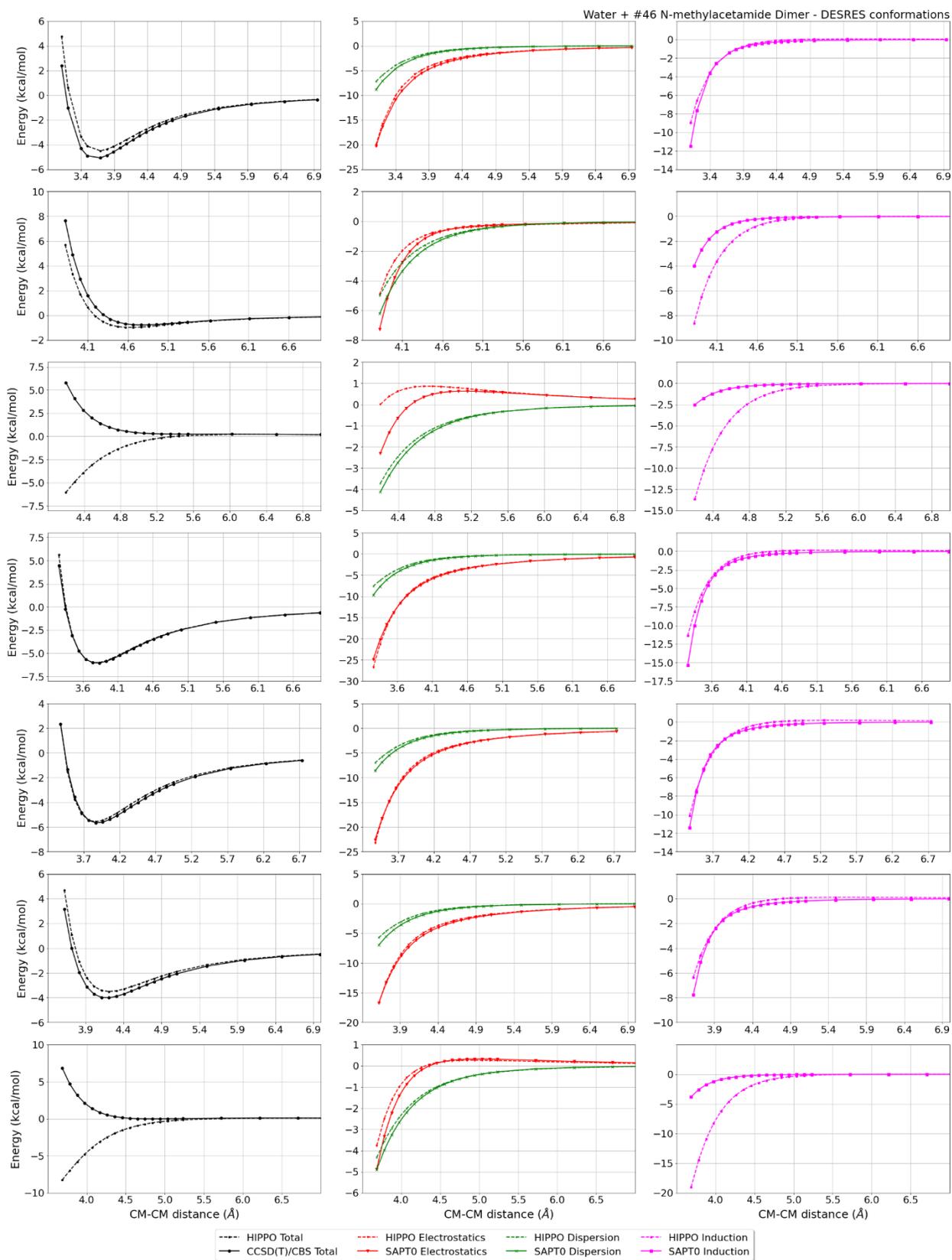
2.567	-7.000	-7.204	-0.2041
2.805	-4.706	-5.358	-0.6524
3.283	-2.164	-2.678	-0.5136

MAE	Std error	max error	#points	#count[err > 1]
1.460	0.941	2.7990	10	6



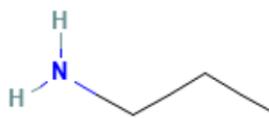






#51 Propan-1-amine C3H9N CID: 7852

ref molpol	-8.37	-6.98	-6.58, avg	-7.31
molpol	8.27	6.84	6.50, avg	7.20
rms molpol	0.11	0.13	0.07, avg	0.10



Monomer potential fitting RMS: 0.41

##Dimer results - Fitting to QM datasets##

DESRES_51-51, energy values in kcal/mol

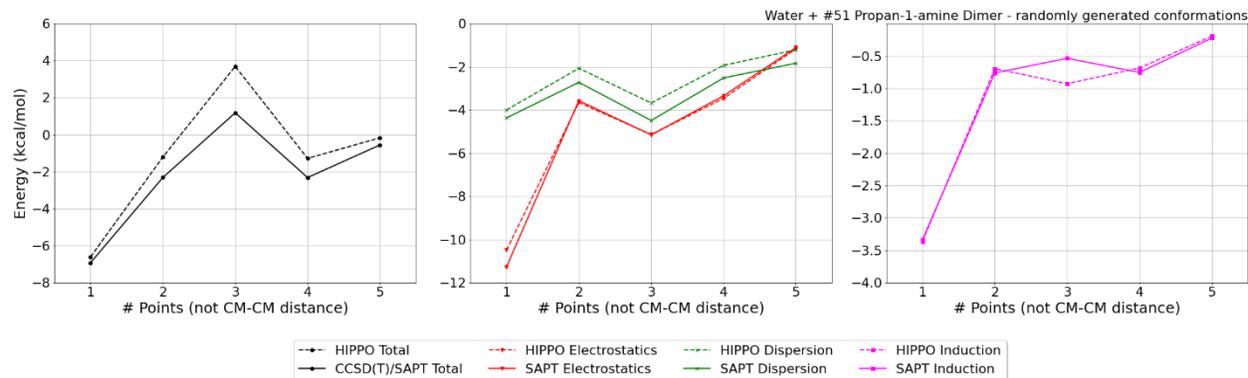
MAE	Std error	max error	#points	#count[err > 1]
0.164	0.262	3.5836	521	6

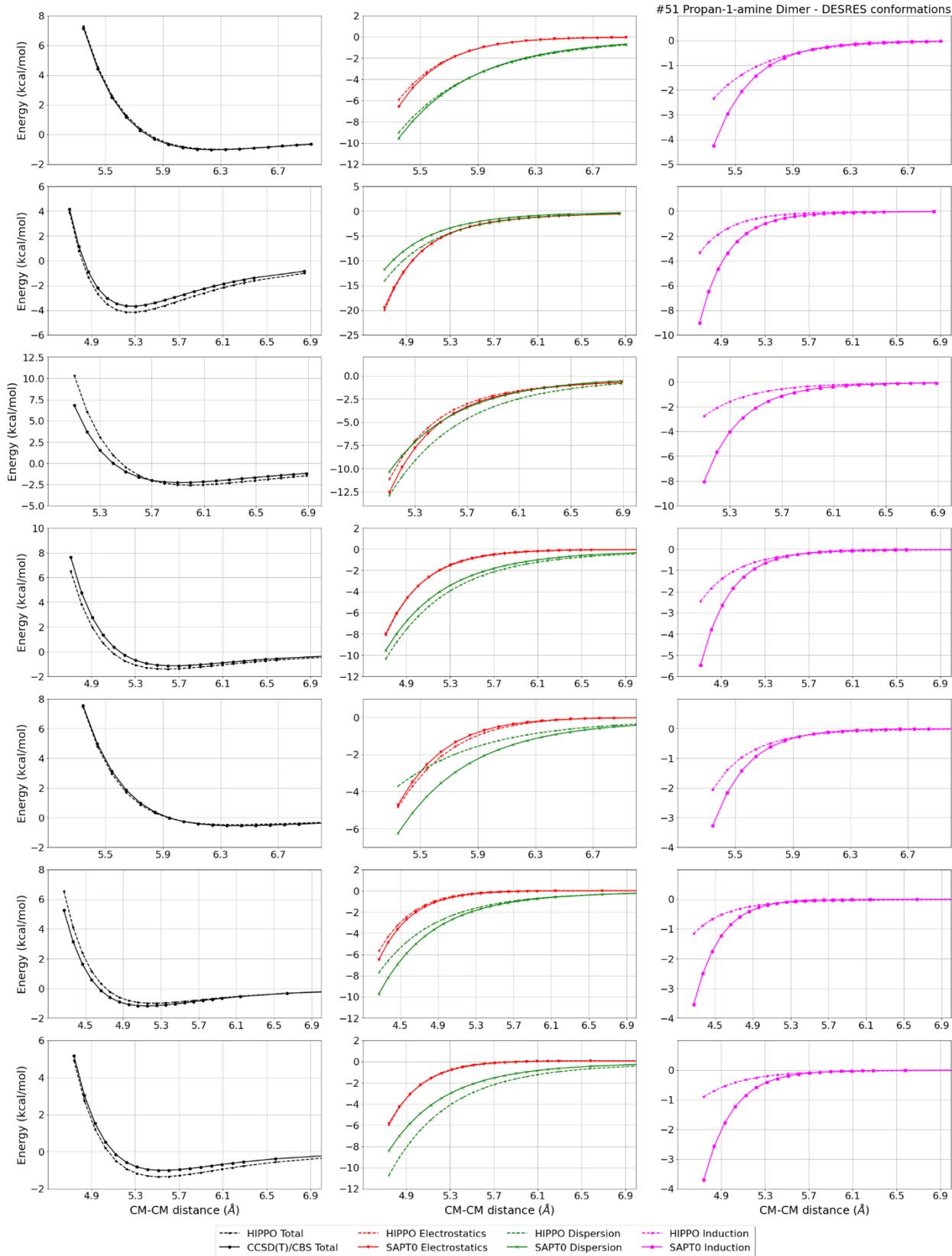
DESRES_51-water, energy values in kcal/mol

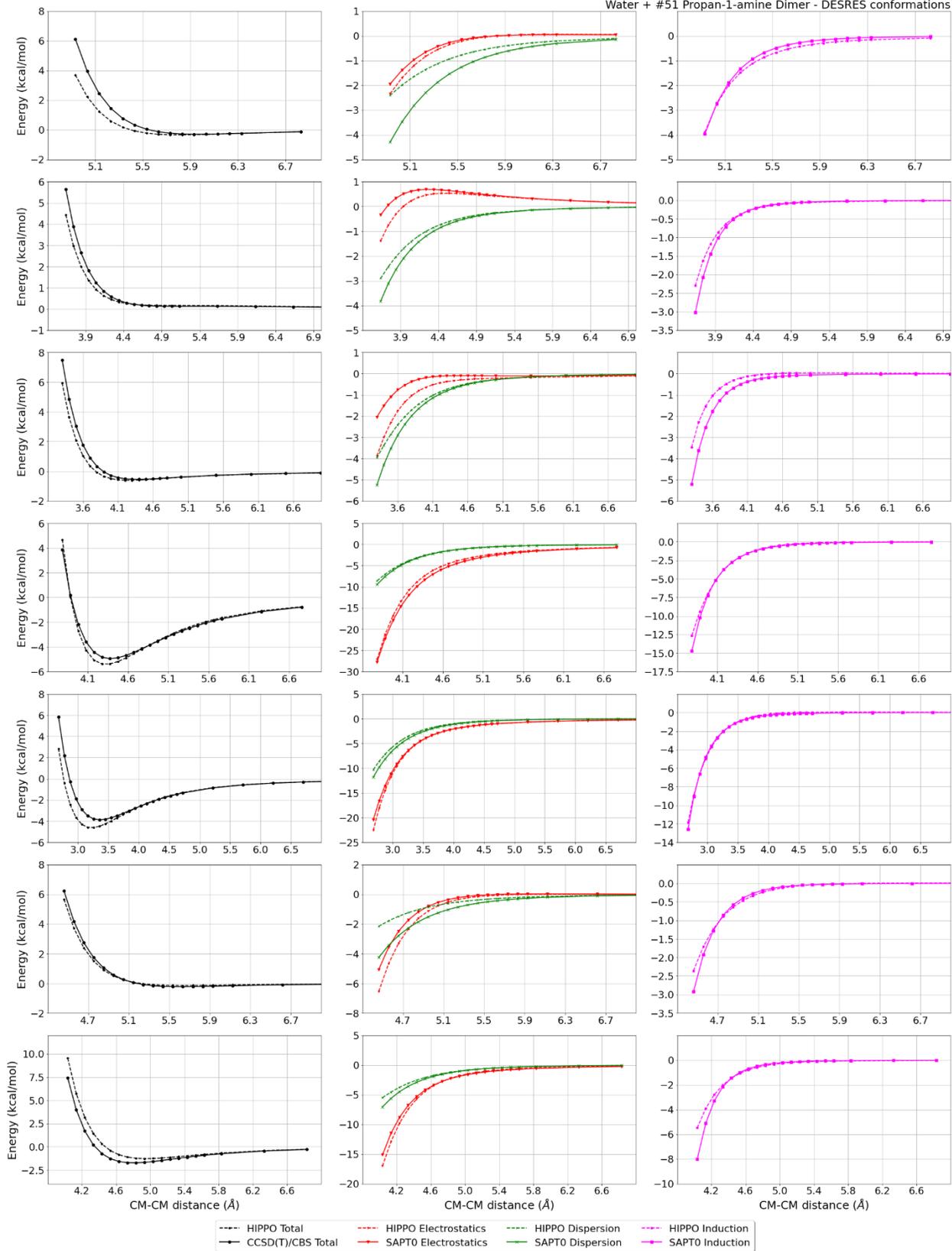
MAE	Std error	max error	#points	#count[err > 1]
0.240	0.429	3.1865	536	32

Liquid Propan-1-amine @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
695.41	711.47	2.3	29.69	30.98	4.2	4.67	5.11	8.6	5000

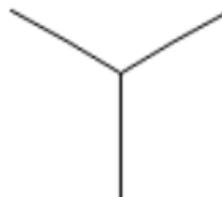






#53 2-methylpropane C4H10 CID: 6360

ref molpol	-7.05	-8.02	-8.02, avg	-7.70
molpol	7.06	8.01	8.01, avg	7.69
rms molpol	0.01	0.02	0.02, avg	0.01



Monomer potential fitting RMS: 0.21

##Dimer results - Fitting to QM datasets##

DESRES_53-water, energy values in kcal/mol

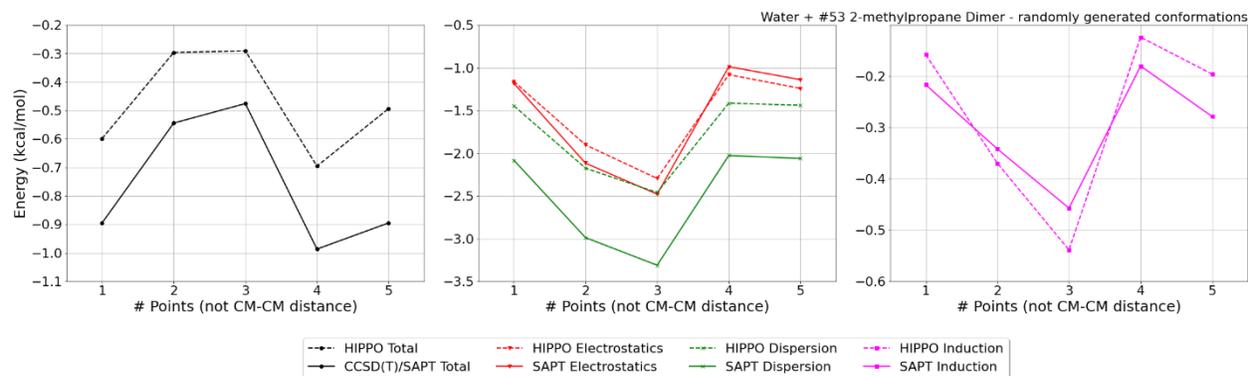
MAE	Std error	max error	#points	#count[err > 1]
0.165	0.388	4.5014	554	16

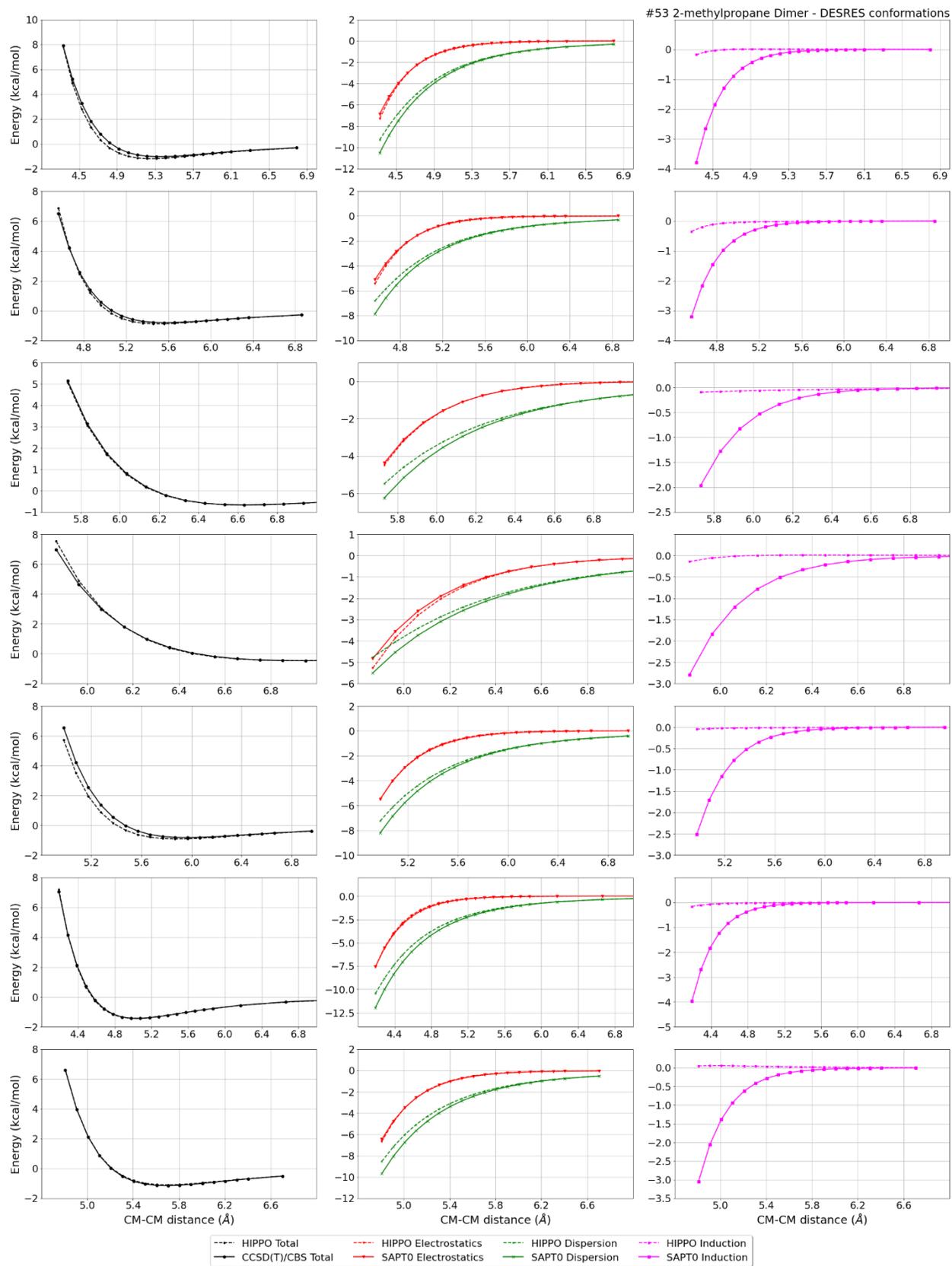
DESRES_53-53, energy values in kcal/mol

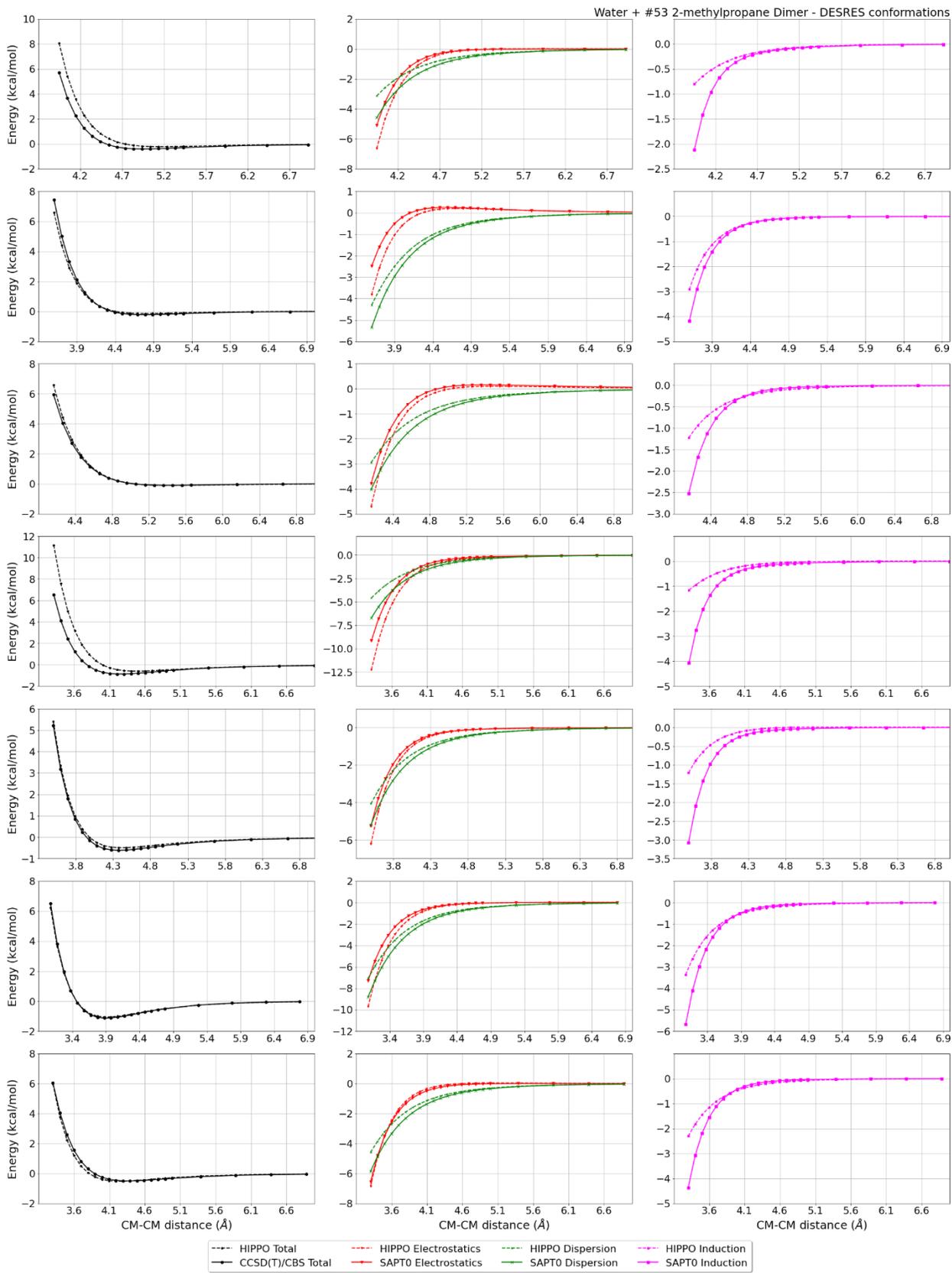
MAE	Std error	max error	#points	#count[err > 1]
0.064	0.121	0.9857	479	0

Liquid 2-methylpropane @ 243.65 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
623.72	613.53	1.7	21.05	22.35	5.8	1.82	1.85	1.6	6000

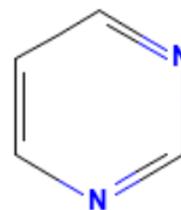






#64 Pyrimidine C4H4N2 CID: 9260

ref molpol	-10.33	-9.89	-5.48, avg	-8.57
molpol	10.36	9.92	5.74, avg	8.67
rms molpol	0.03	0.03	0.26, avg	0.11



Monomer potential fitting RMS: 0.27

##Dimer results - Fitting to QM datasets##

DESRES_64-64, energy values in kcal/mol

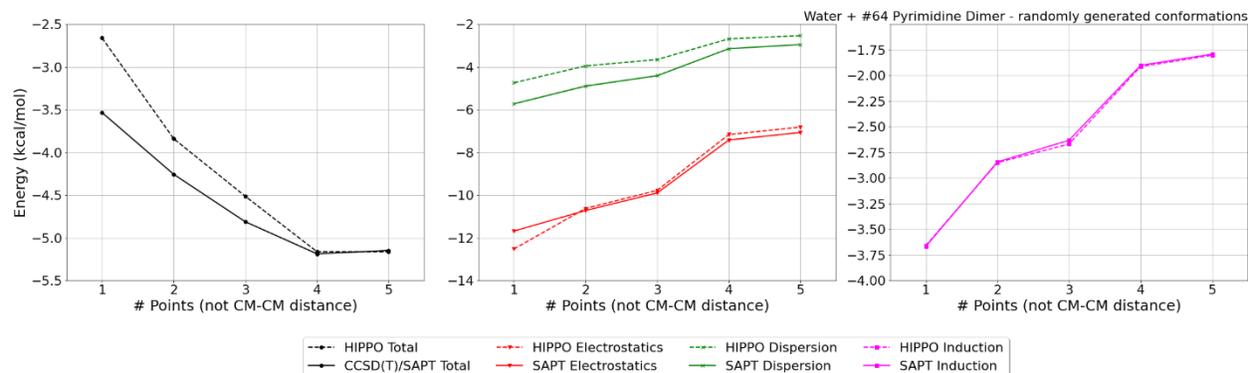
MAE	Std error	max error	#points	#count[err > 1]
0.314	0.561	6.5949	546	30

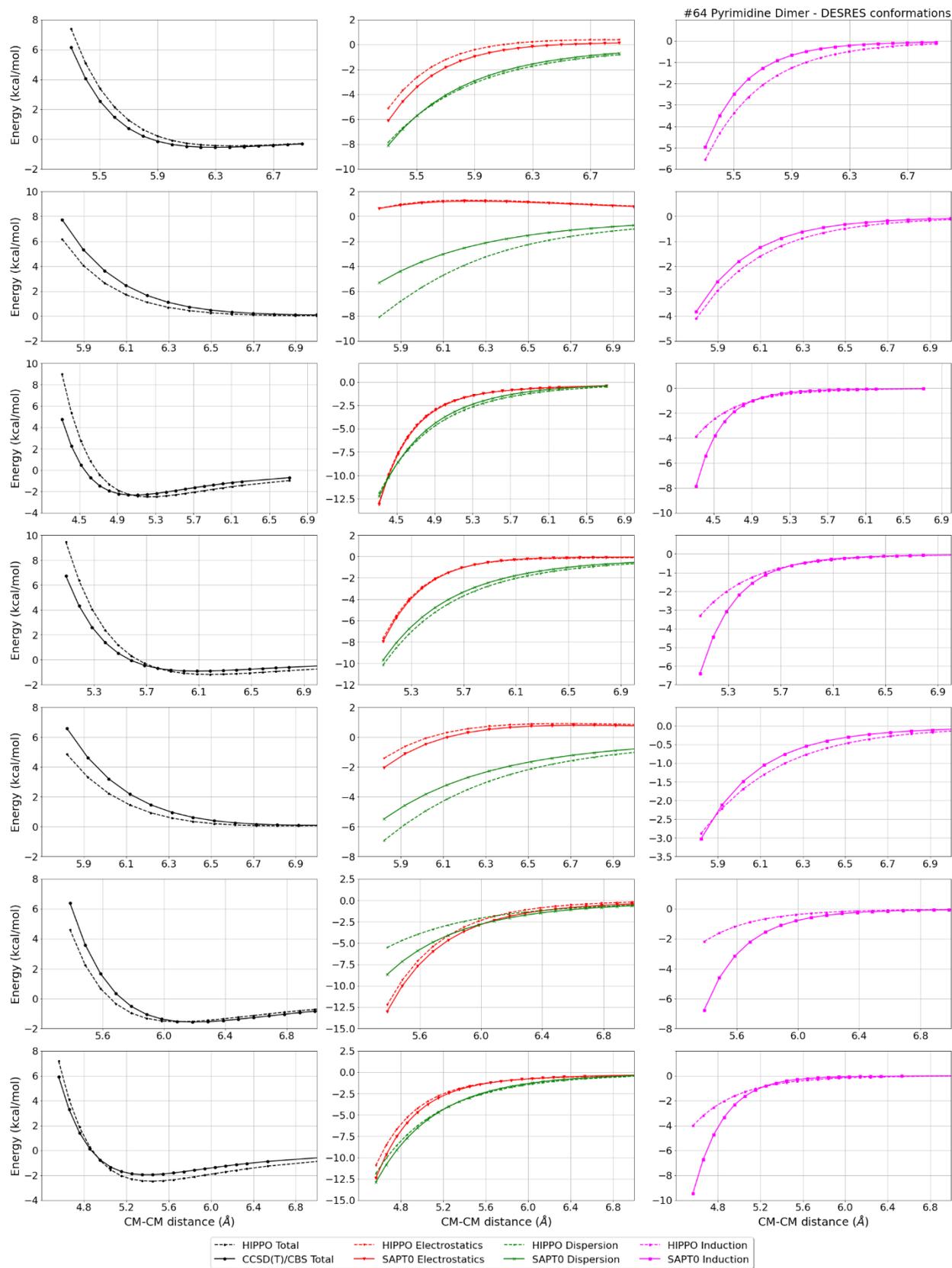
DESRES_64-water, energy values in kcal/mol

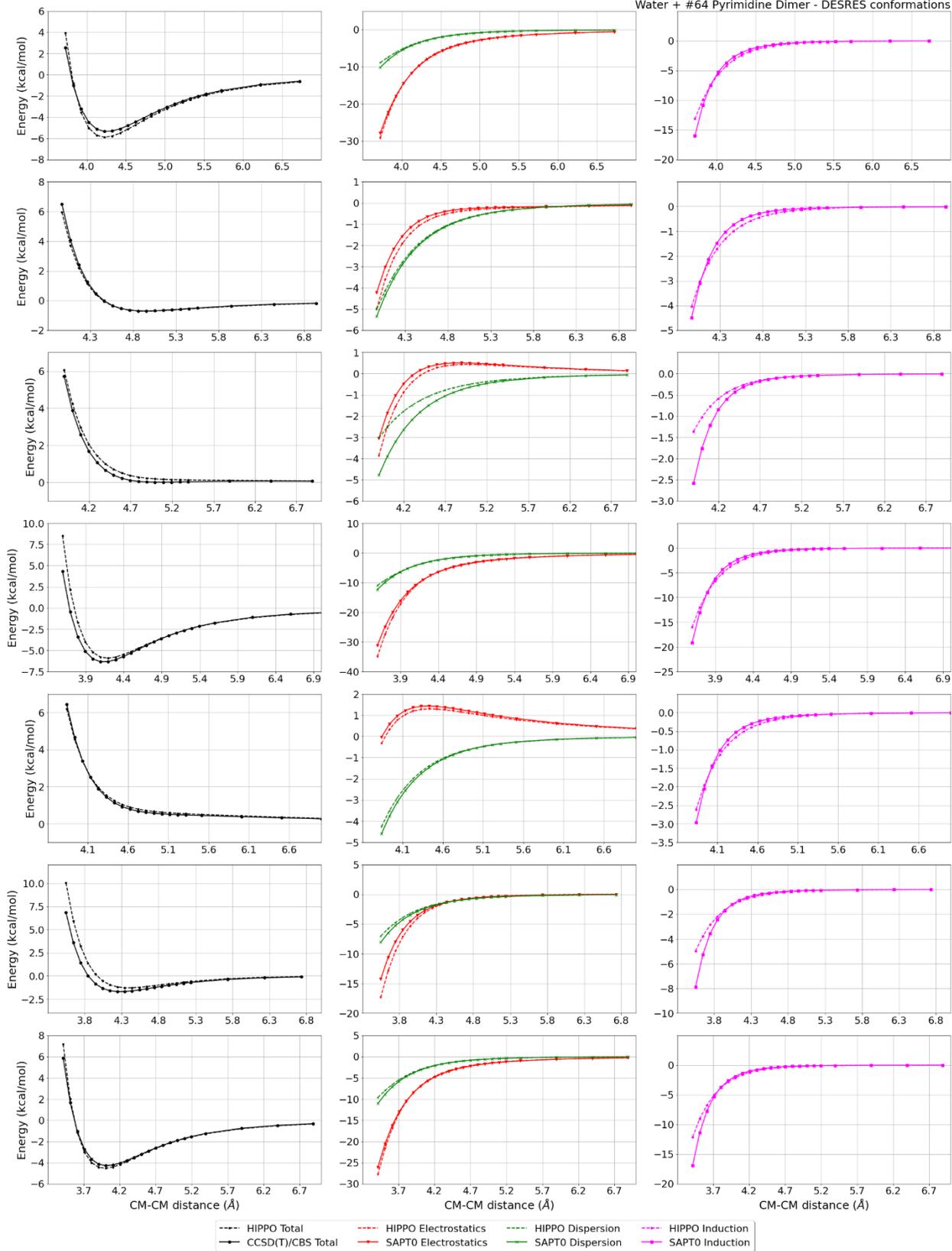
MAE	Std error	max error	#points	#count[err > 1]
0.227	0.421	4.1906	566	28

Liquid Pyrimidine @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1062.92	1016.40	4.6	49.23	49.81	1.2	20.24	-1.00	-2124.0	3000

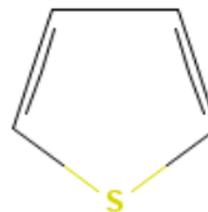






#66 Thiophene C4H4S CID: 8030

ref molpol	-10.40	-11.22	-6.59, avg	-9.40
molpol	11.07	11.07	6.47, avg	9.54
rms molpol	0.67	0.15	0.13, avg	0.13



Monomer potential fitting RMS: 0.16

##Dimer results - Fitting to QM datasets##

R739x5_66-water, energy values in kcal/mol

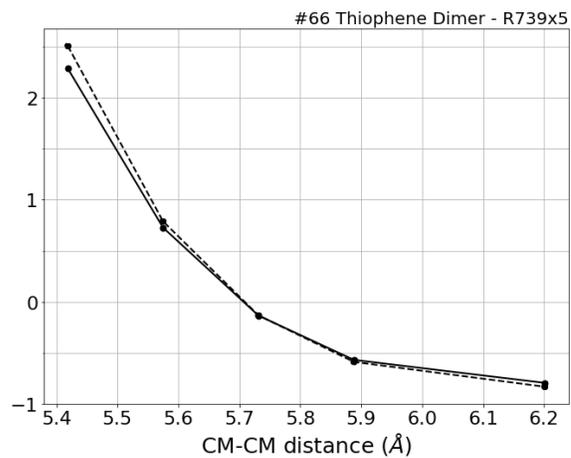
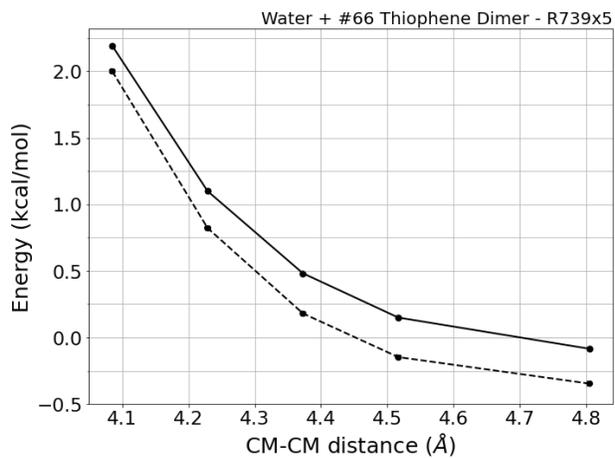
CM-CM (A)	Reference	HIPPO res	Abs diff		
3.365	2.190	1.999	-0.1911		
3.483	1.099	0.822	-0.2772		
3.602	0.482	0.181	-0.3013		
3.720	0.150	-0.147	-0.2969		
3.957	-0.084	-0.346	-0.2618		
MAE	Std error	max error	#points	#count[err > 1]	
0.266	0.040	0.3013	5	0	

R739x5_66-66, energy values in kcal/mol

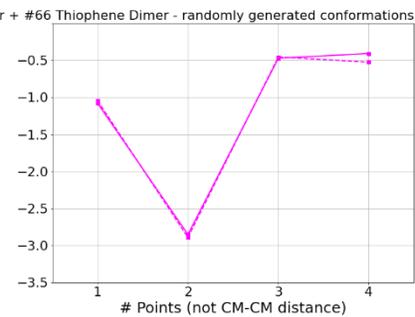
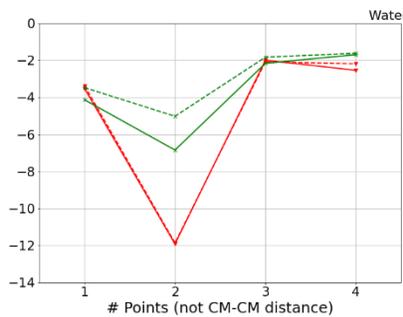
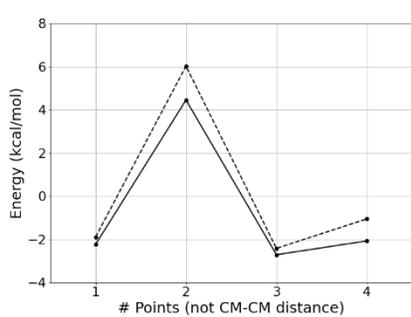
CM-CM (A)	Reference	HIPPO res	Abs diff		
2.709	2.282	2.503	0.2211		
2.787	0.727	0.788	0.0613		
2.866	-0.134	-0.131	0.0026		
2.944	-0.567	-0.587	-0.0199		
3.101	-0.792	-0.830	-0.0377		
MAE	Std error	max error	#points	#count[err > 1]	
0.069	0.079	0.2211	5	0	

Liquid Thiophene @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1043.82	1059.01	1.4	34.71	34.65	0.2	2.40	-1.00	-340.3	1000



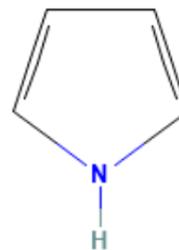
—●— CCSD(T) Total - - - ● - - - HIPPO Total



- - - ● - - - HIPPO Total - - - ● - - - HIPPO Electrostatics - - - ● - - - HIPPO Dispersion - - - ● - - - HIPPO Induction
 —●— CCSD(T)/SAPT Total —●— SAPT Electrostatics —●— SAPT Dispersion —●— SAPT Induction

#67 1h-pyrrole C4H5N CID: 8027

ref molpol	-9.33	-8.93	-5.79, avg	-8.02
molpol	9.35	8.94	5.48, avg	7.93
rms molpol	0.02	0.01	0.30, avg	0.09



Monomer potential fitting RMS: 0.32

##Dimer results - Fitting to QM datasets##

DESRES_67-67, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.309	0.553	4.7842	634	36

HB375x10_67-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
2.977	-1.626	-0.587	1.0394
3.056	-3.575	-2.823	0.7519
3.134	-4.661	-4.074	0.5869
3.212	-5.175	-4.691	0.4837
3.291	-5.319	-4.908	0.4112
3.369	-5.230	-4.878	0.3525
3.447	-5.002	-4.701	0.3007
3.682	-4.017	-3.843	0.1737
4.073	-2.551	-2.506	0.0453
4.856	-1.116	-1.135	-0.0192

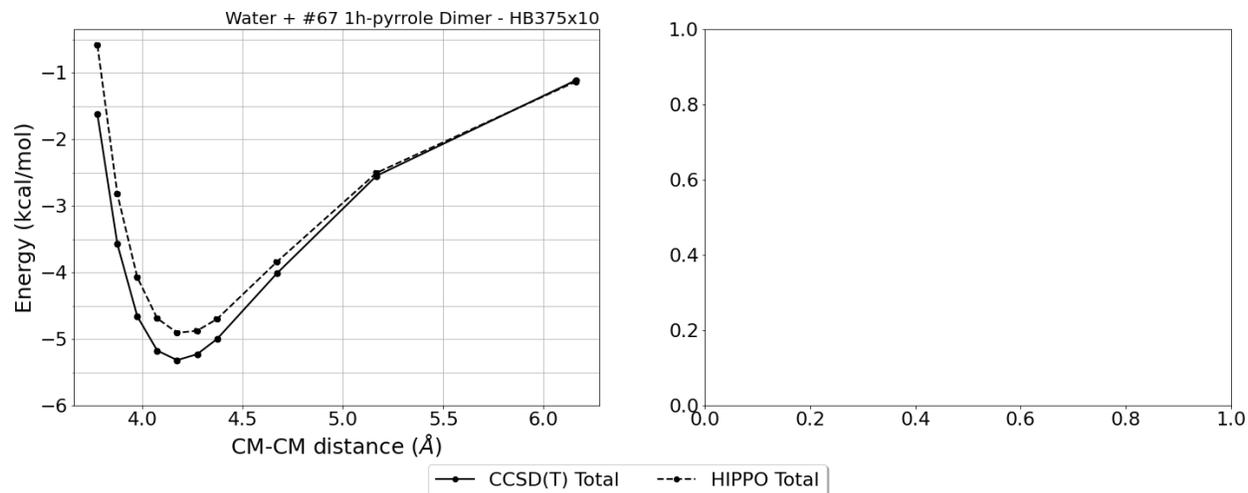
MAE	Std error	max error	#points	#count[err > 1]
0.416	0.301	1.0394	10	1

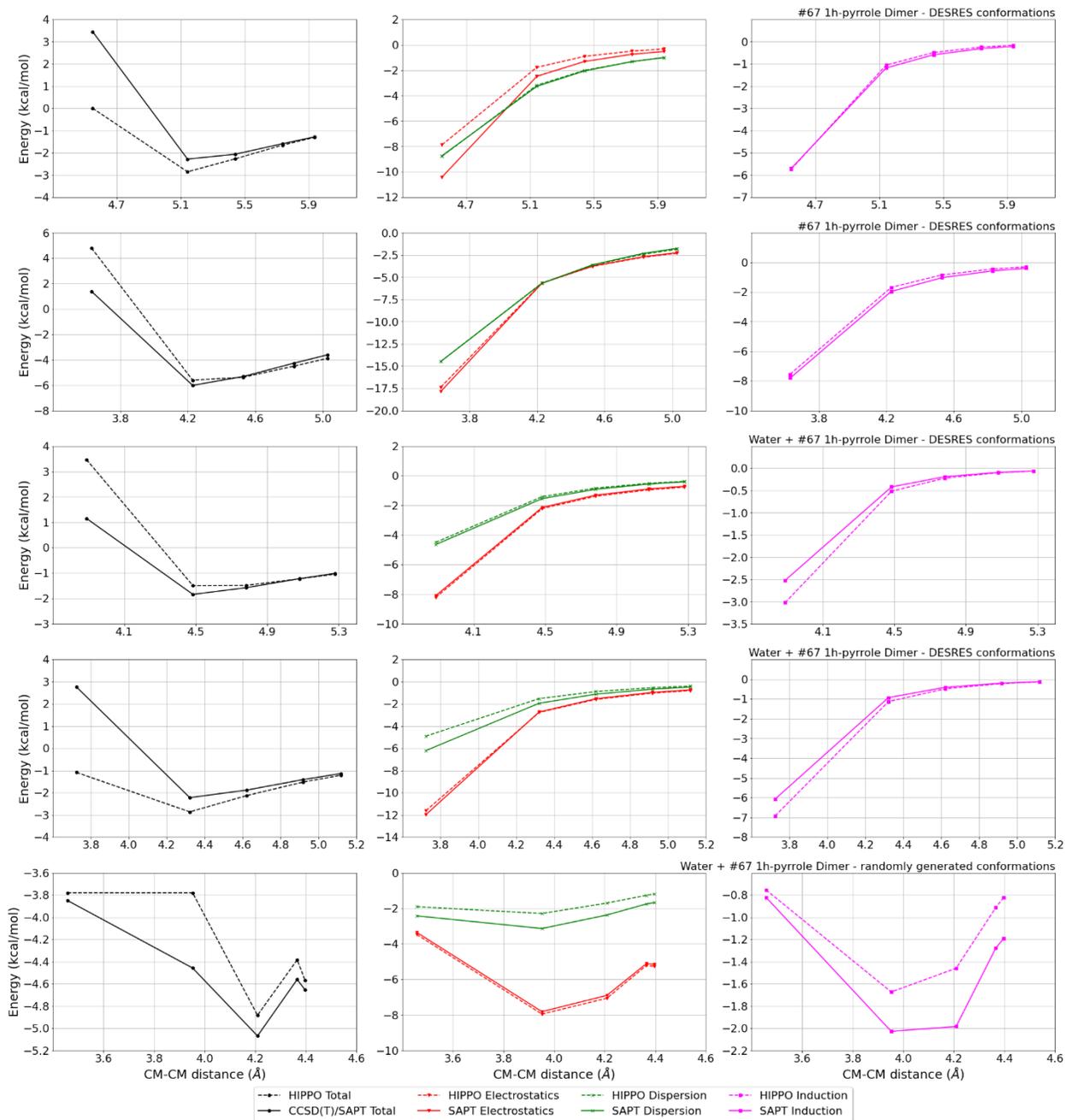
DESRES_67-water, energy values in kcal/mol

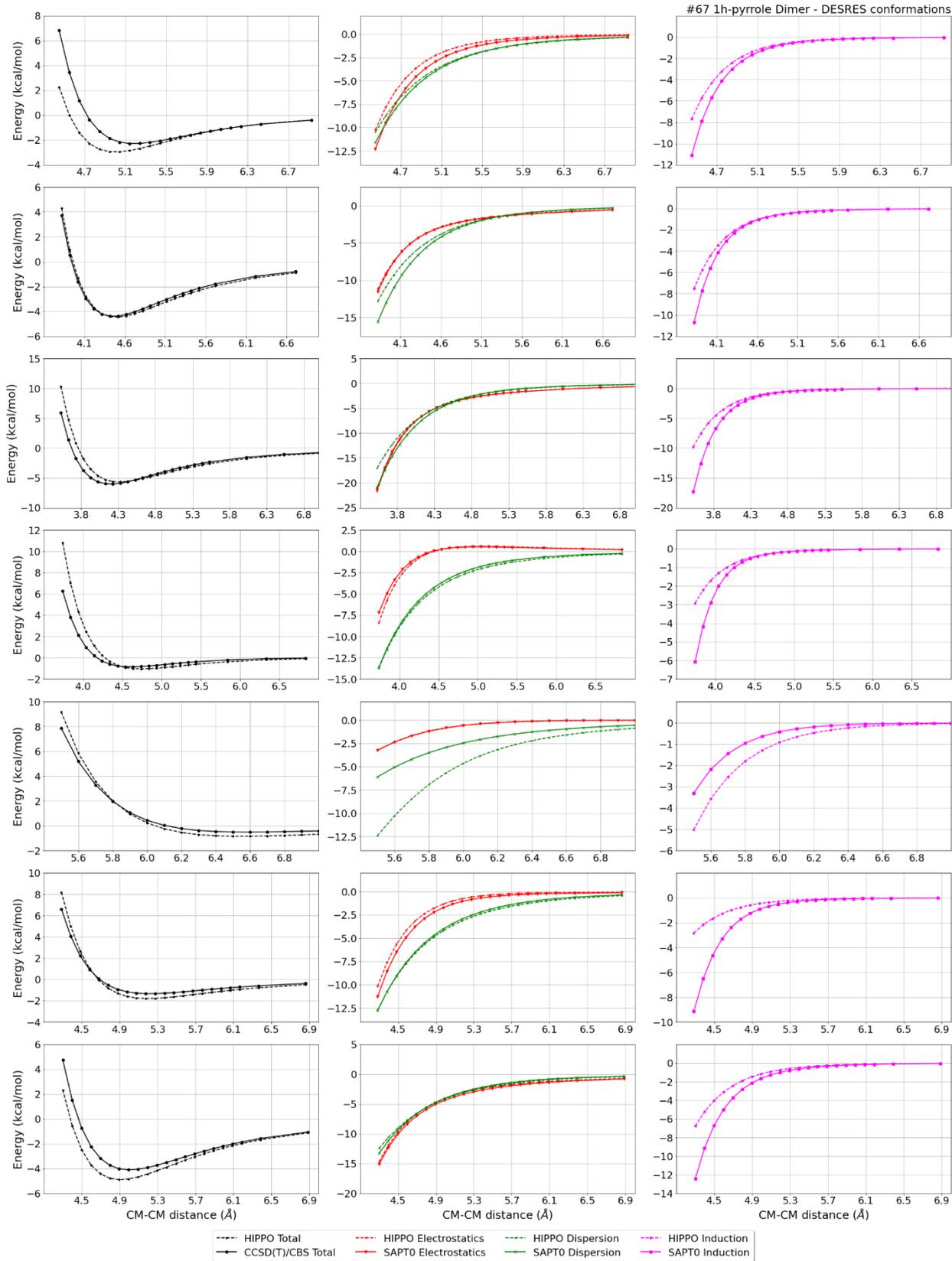
MAE	Std error	max error	#points	#count[err > 1]
0.276	0.549	4.9026	559	34

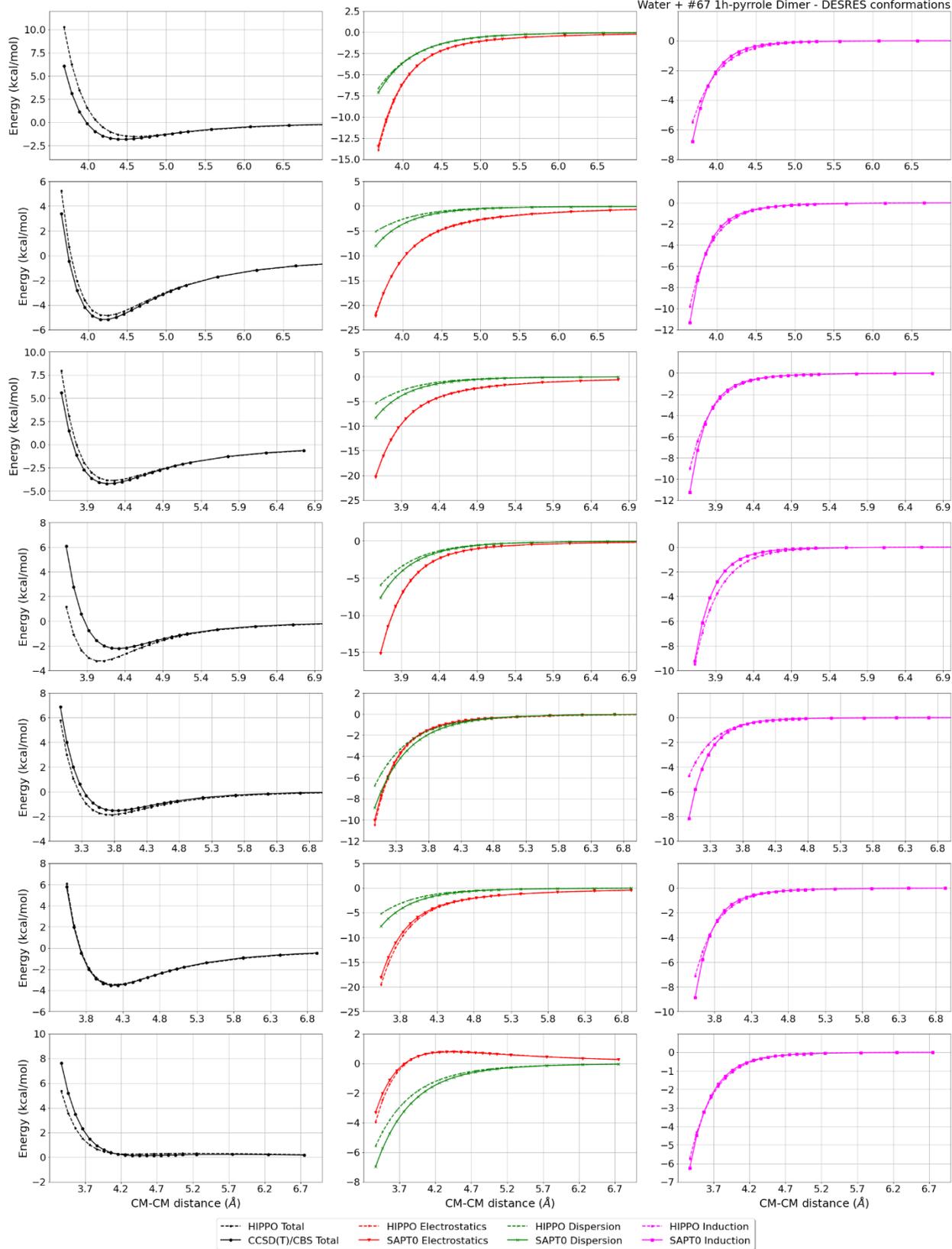
Liquid 1h-pyrrole @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
979.48	965.30	1.5	53.61	45.15	18.7	9.51	7.92	20.1	1000



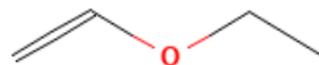






#73 Ethoxyethene C4H8O CID: 8023

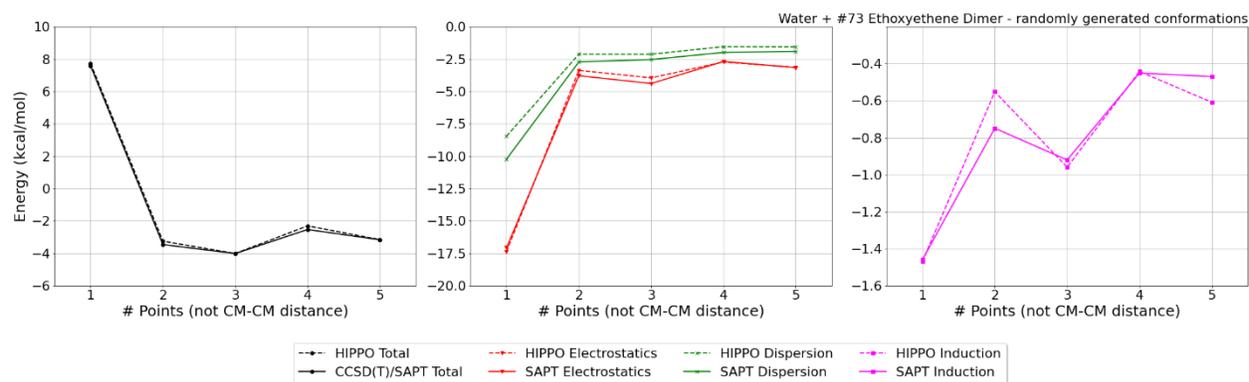
ref molpol	-11.46	-7.66	-6.84, avg	-8.65
molpol	11.43	7.67	6.85, avg	8.65
rms molpol	0.03	0.01	0.01, avg	0.01



Monomer potential fitting RMS: 0.17

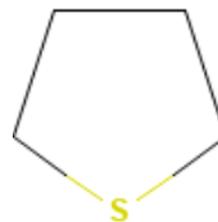
Liquid Ethoxyethene @ 293.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
713.31	758.90	6.0	27.83	27.84	0.0	6.25	-1.00	-724.7	1000



#76 Thiolane C4H8S CID: 1127

ref molpol	-8.50	-10.72	-10.43, avg	-9.88
molpol	8.31	10.58	10.60, avg	9.83
rms molpol	0.19	0.15	0.17, avg	0.05



Monomer potential fitting RMS: 0.15

##Dimer results - Fitting to QM datasets##

DESRES_76-water, energy values in kcal/mol

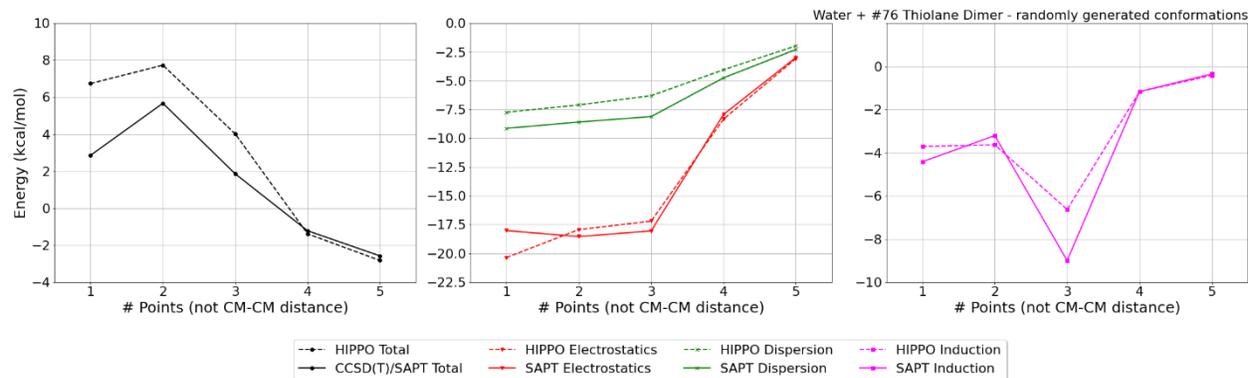
MAE	Std error	max error	#points	#count[err > 1]
0.208	0.424	5.4180	562	22

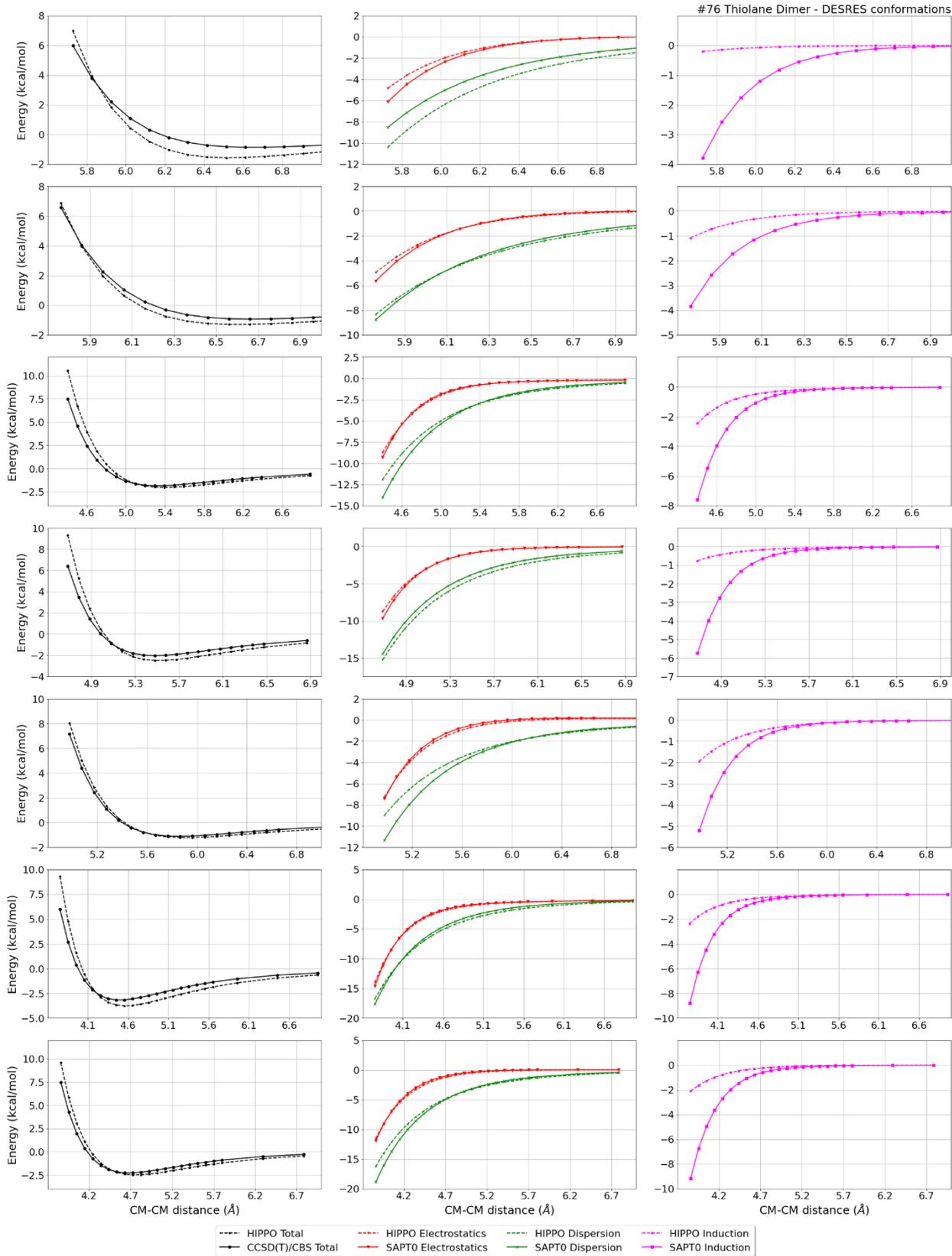
DESRES_76-76, energy values in kcal/mol

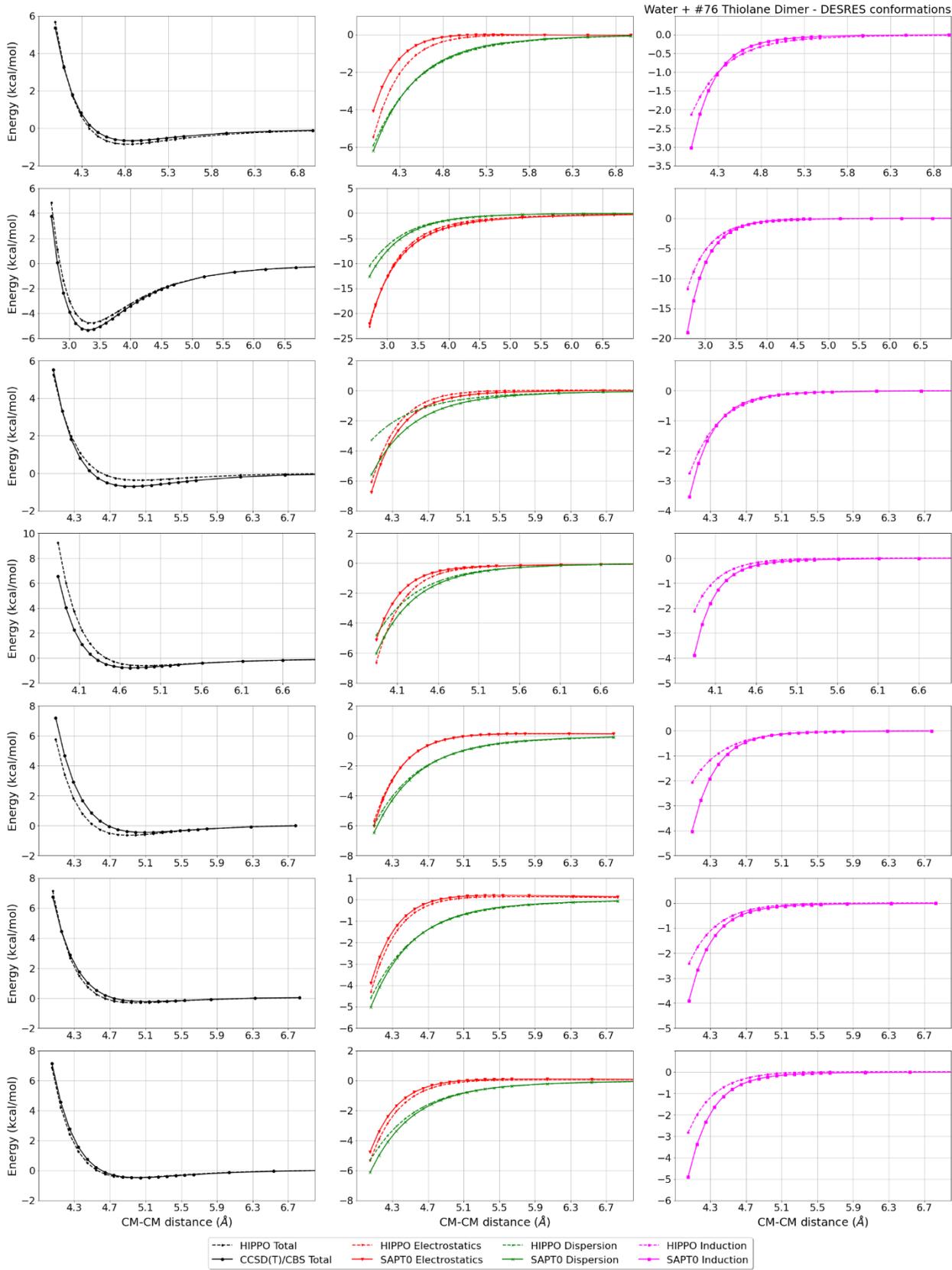
MAE	Std error	max error	#points	#count[err > 1]
0.390	0.524	4.5663	487	33

Liquid Thiolane @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1040.70	994.00	4.7	52.19	38.62	35.1	10.83	-1.00	-1183.1	1000

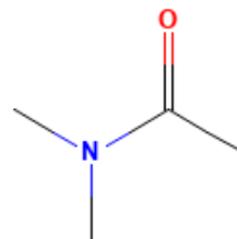






#80 N,n-dimethylacetamide C4H9NO CID: 31374

ref molpol -10.59 -10.02 -7.13, avg -9.24
molpol 10.60 10.00 7.09, avg 9.23
rms molpol 0.01 0.02 0.03, avg 0.01



Monomer potential fitting RMS: 1.04

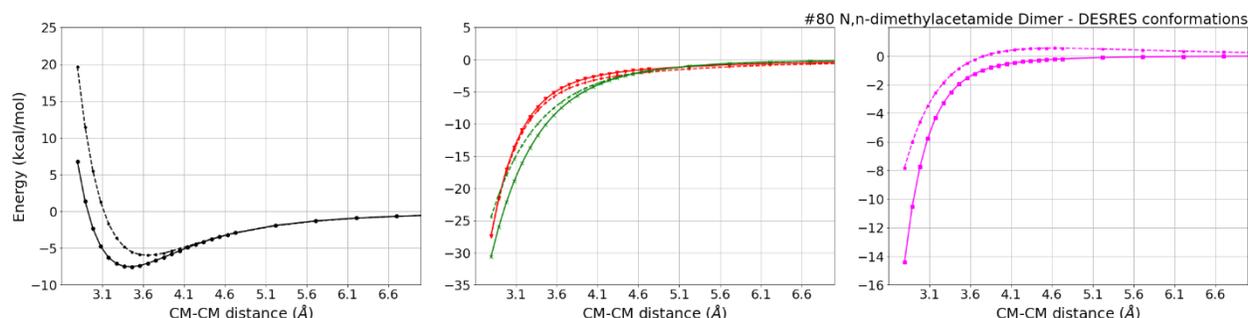
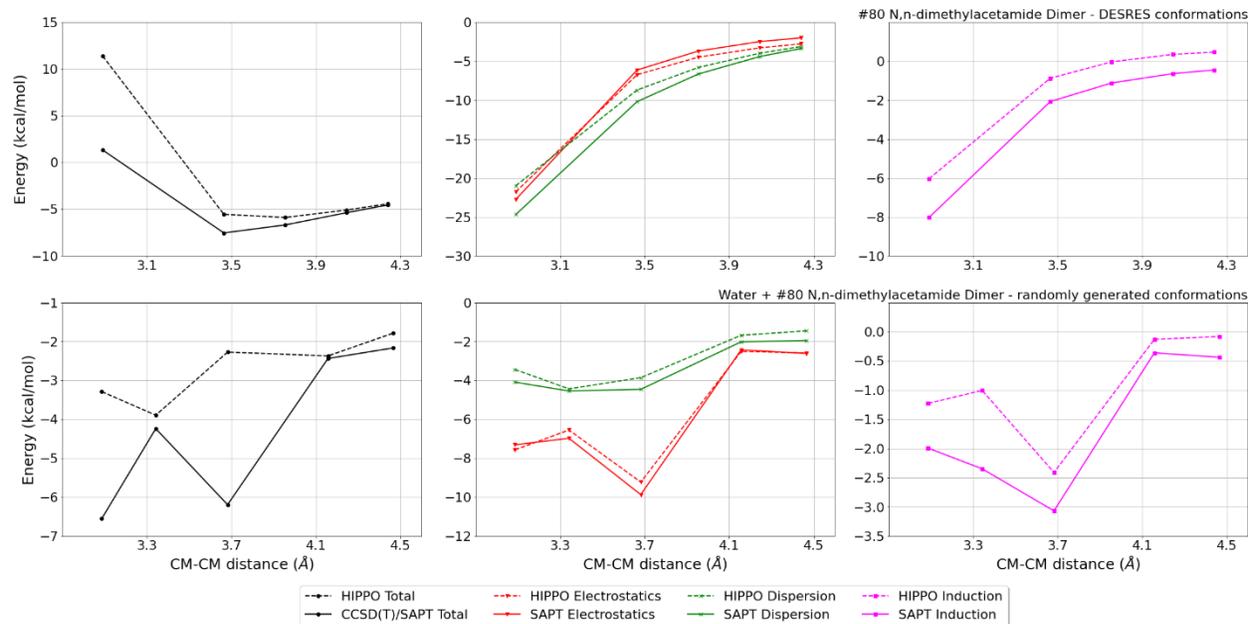
##Dimer results - Fitting to QM datasets##

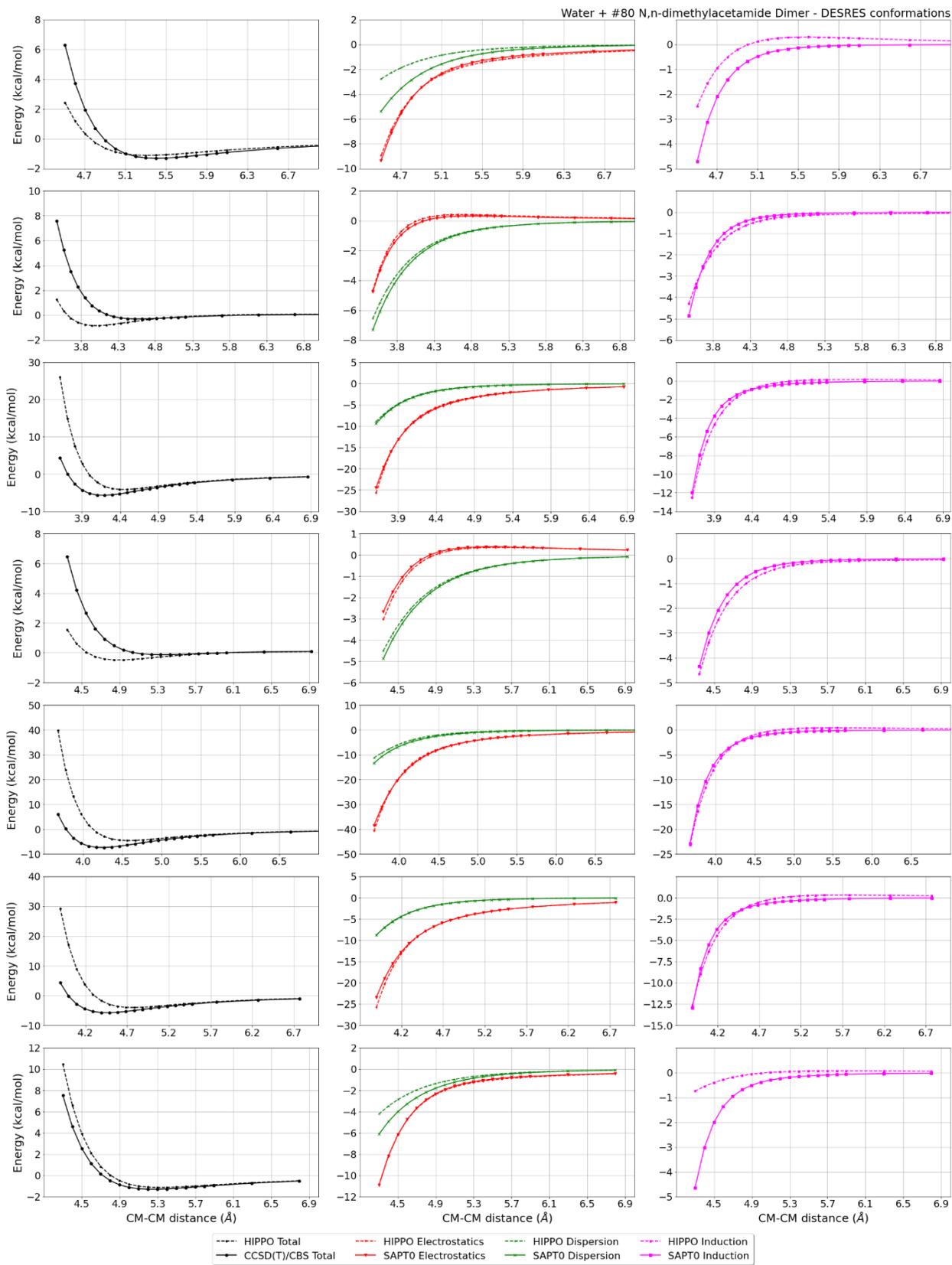
DESRES_80-80, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
1.681	2.751	10.9311	26	8

DESRES_80-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.769	2.533	31.0263	561	78





#82 Pyridine C5H5N CID: 1049

ref molpol -10.70 -11.34 -6.02, avg -9.35
molpol 10.86 10.85 6.17, avg 9.29
rms molpol 0.16 0.49 0.16, avg 0.06

Monomer potential fitting RMS: 0.30

##Dimer results - Fitting to QM datasets##

DESRES_82-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.322	0.657	5.4068	565	53

DESRES_82-82, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.203	0.410	3.9218	546	19

HB375x10_82-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
3.128	-2.162	-0.965	1.1972
3.206	-4.659	-4.251	0.4078
3.284	-6.103	-6.058	0.0448
3.361	-6.833	-6.911	-0.0782
3.439	-7.085	-7.166	-0.0810
3.517	-7.024	-7.058	-0.0337
3.595	-6.768	-6.740	0.0284
3.828	-5.504	-5.344	0.1595
4.217	-3.454	-3.285	0.1689
4.995	-1.372	-1.318	0.0537

MAE	Std error	max error	#points	#count[err > 1]
0.225	0.341	1.1972	10	1

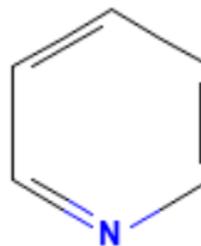
R739x5_82-82, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
3.111	2.403	2.025	-0.3778
3.198	1.987	1.630	-0.3570
3.284	1.694	1.374	-0.3196
3.370	1.473	1.199	-0.2741
3.543	1.161	0.969	-0.1924

MAE	Std error	max error	#points	#count[err > 1]
0.304	0.066	0.3778	5	0

R739x5_82-water, energy values in kcal/mol

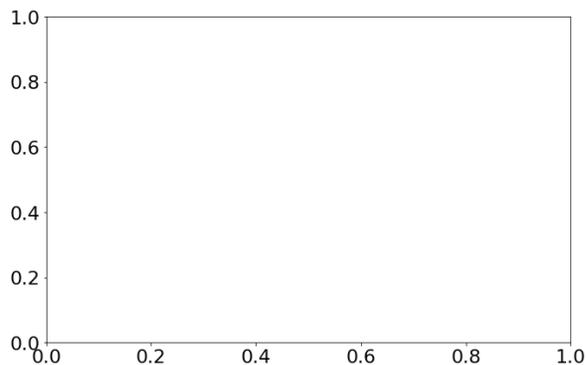
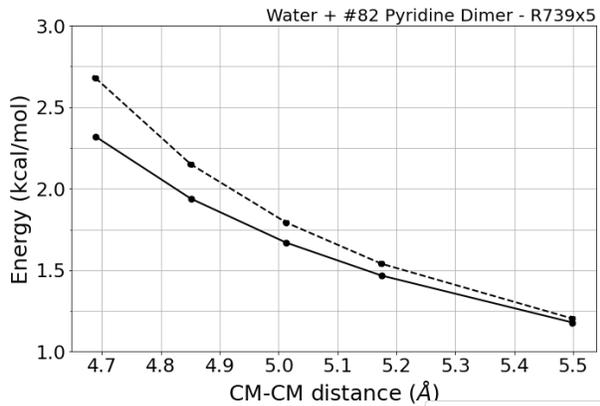
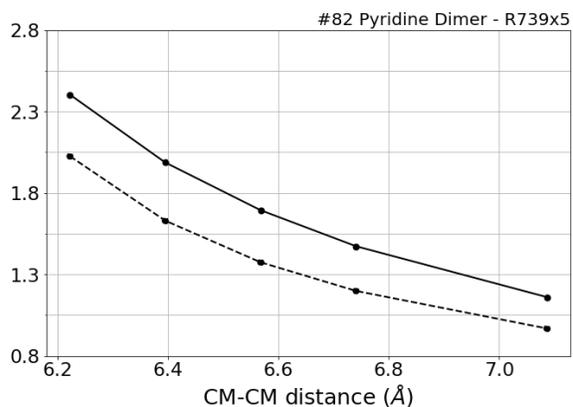
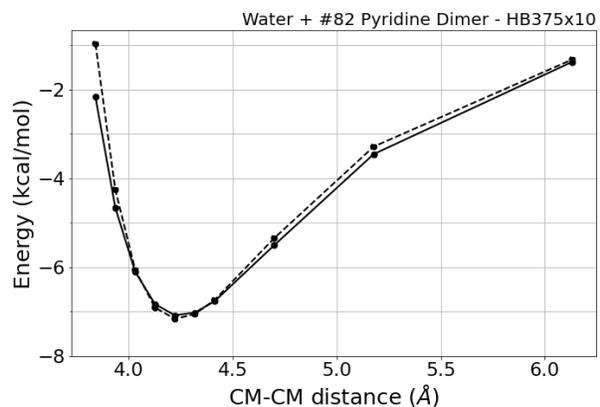
CM-CM (A)	Reference	HIPPO res	Abs diff
3.819	2.320	2.679	0.3594
3.951	1.939	2.150	0.2108
4.083	1.669	1.792	0.1235
4.215	1.467	1.540	0.0726
4.479	1.179	1.203	0.0242



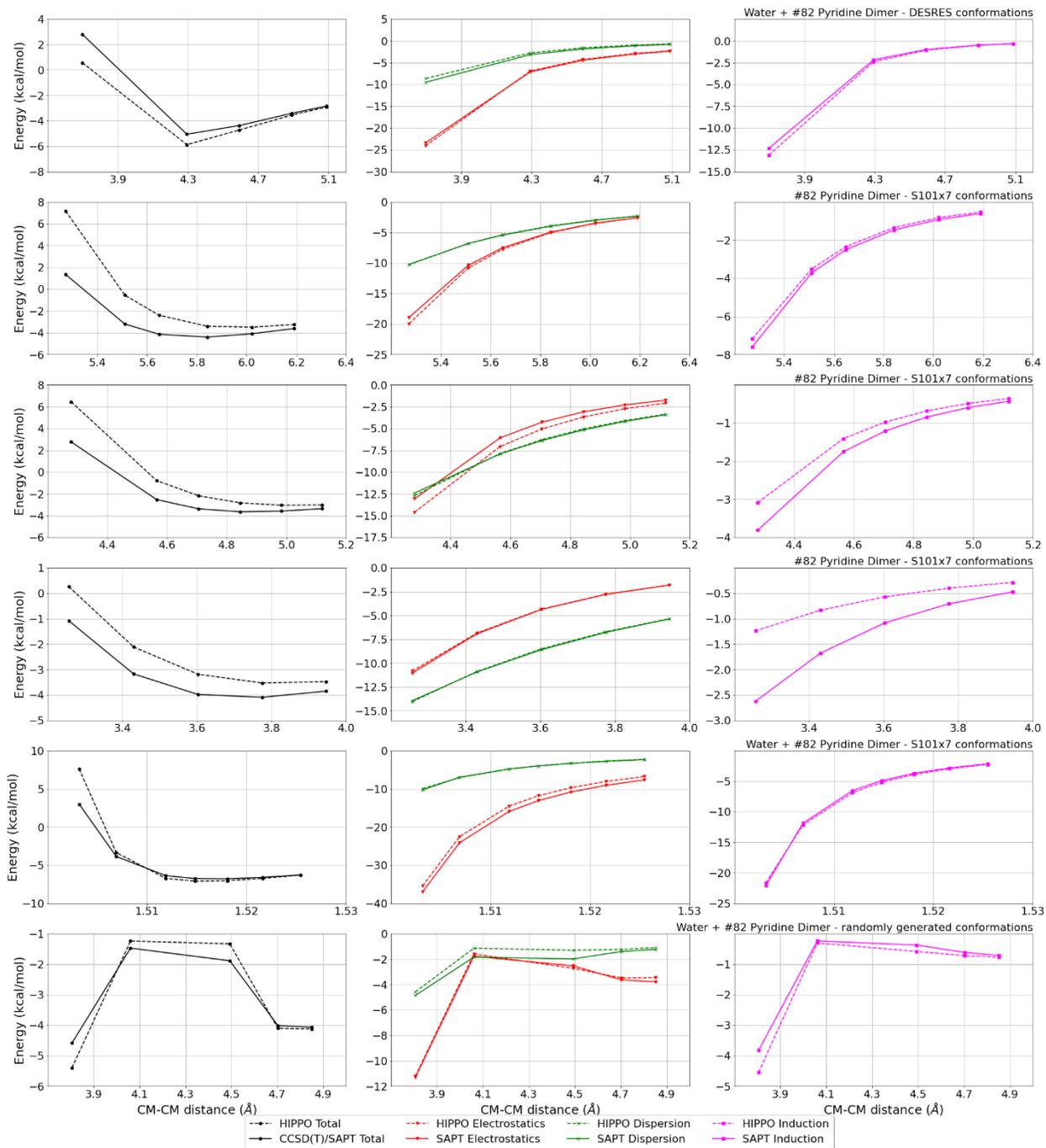
MAE	Std error	max error	#points	#count[err > 1]
0.158	0.118	0.3594	5	0

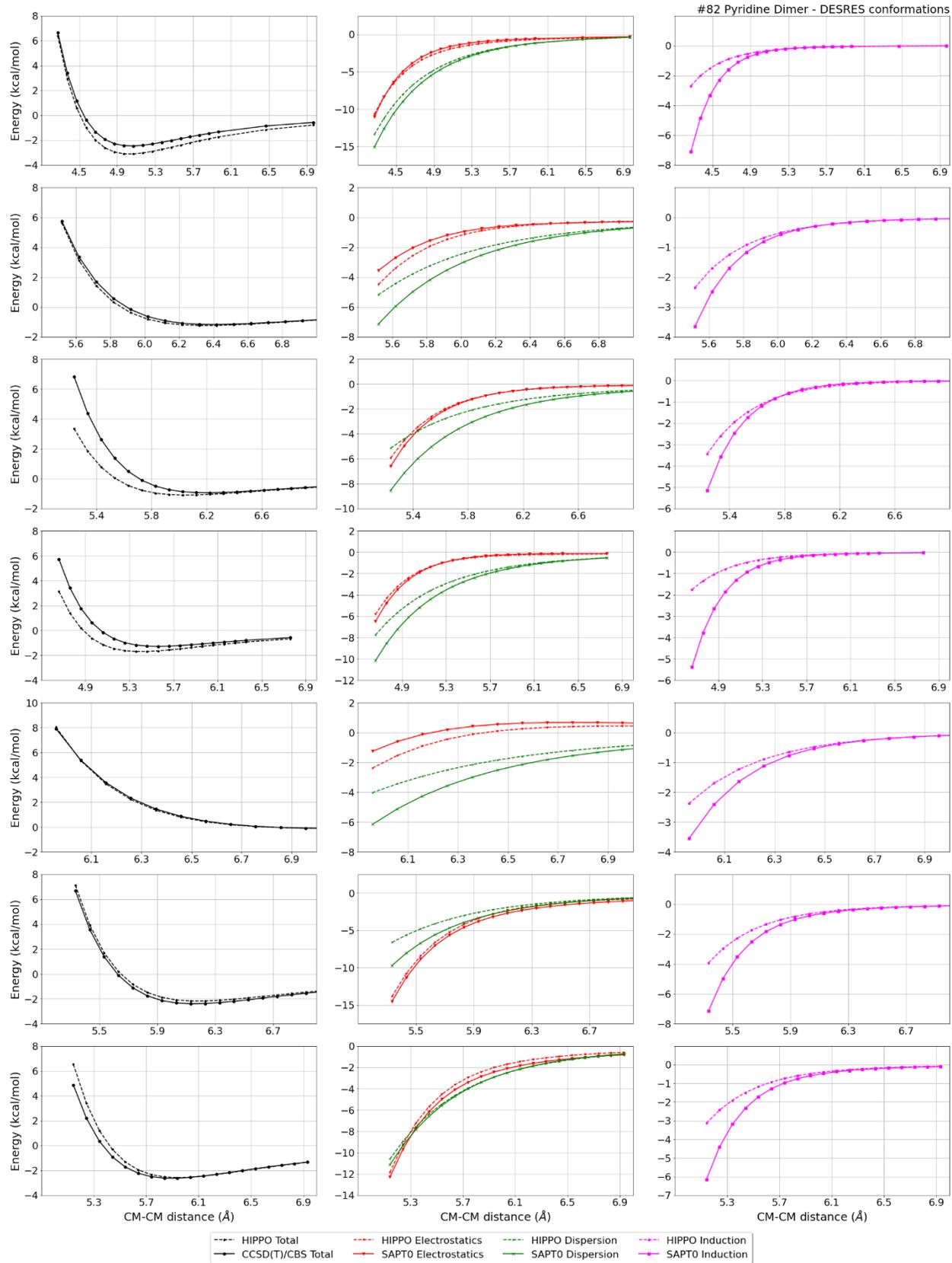
Liquid Pyridine @ 298.15 K

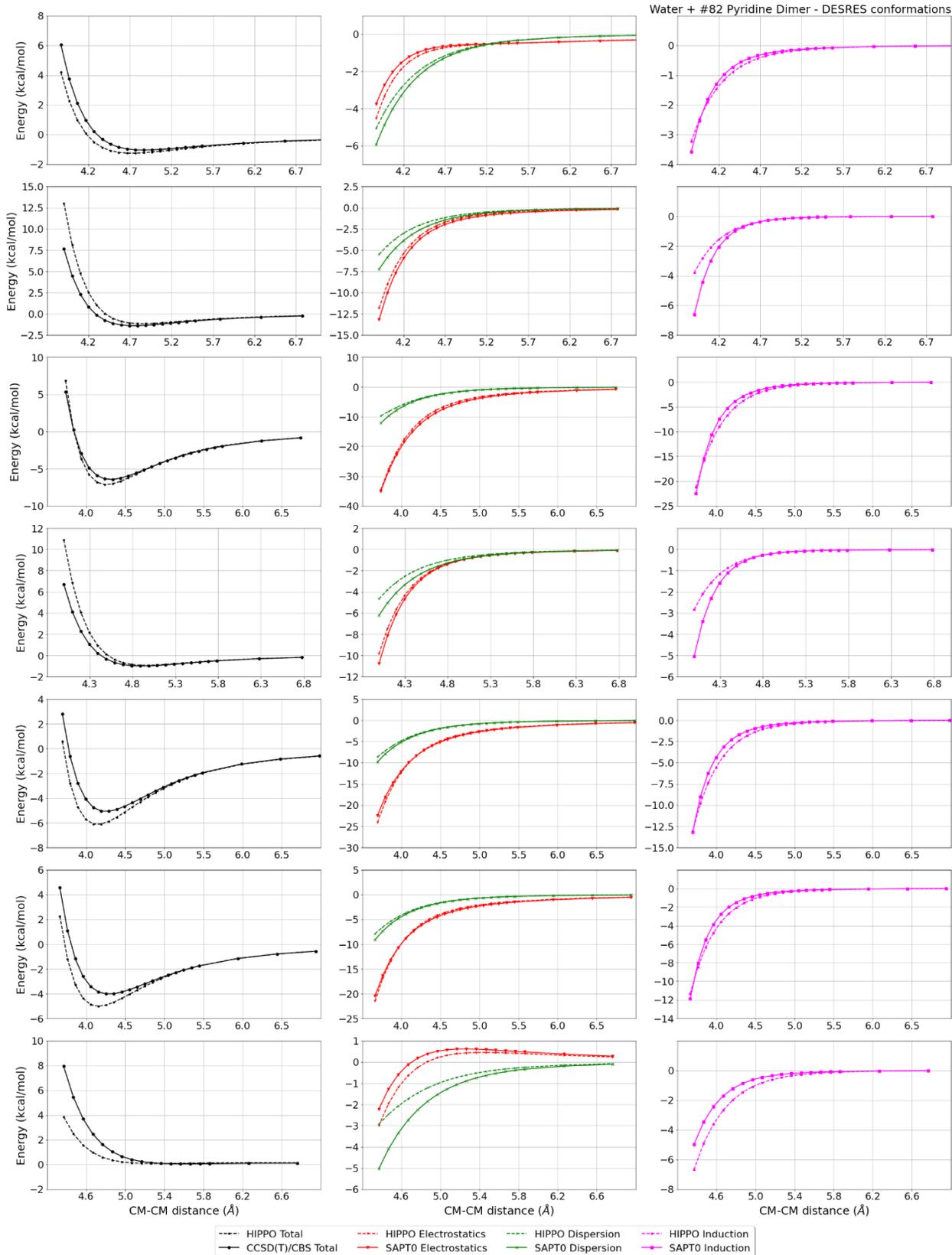
Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1008.40	977.80	3.1	45.32	40.15	12.9	16.73	12.98	28.9	1000



—●— CCSD(T) Total - - - ● - - - HIPPO Total

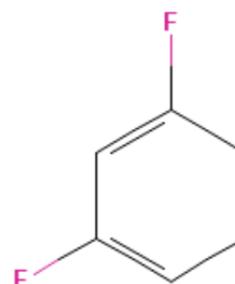






#97 1,3-difluorobenzene C6H4F2 CID: 9741

ref molpol -11.84 -12.20 -6.39, avg -10.14
molpol 11.94 11.93 6.61, avg 10.16
rms molpol 0.10 0.27 0.22, avg 0.02



Monomer potential fitting RMS: 0.10

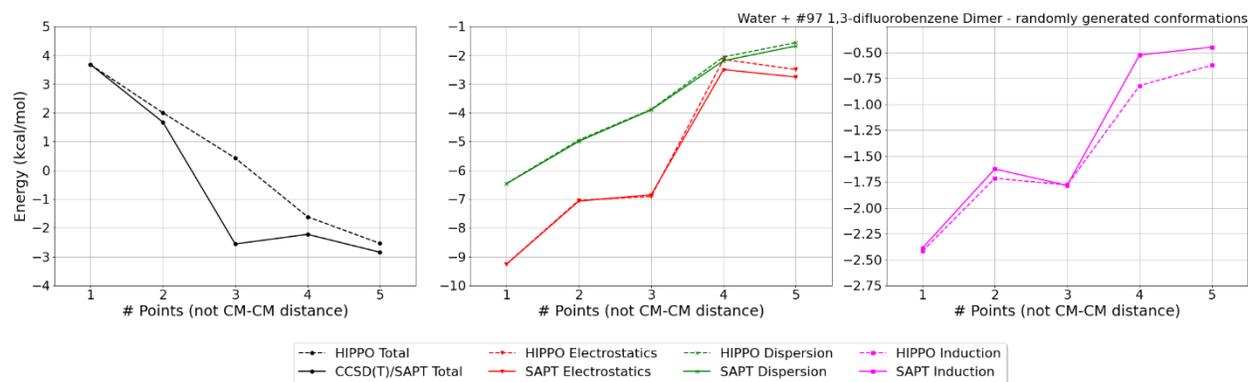
##Dimer results - Fitting to QM datasets##

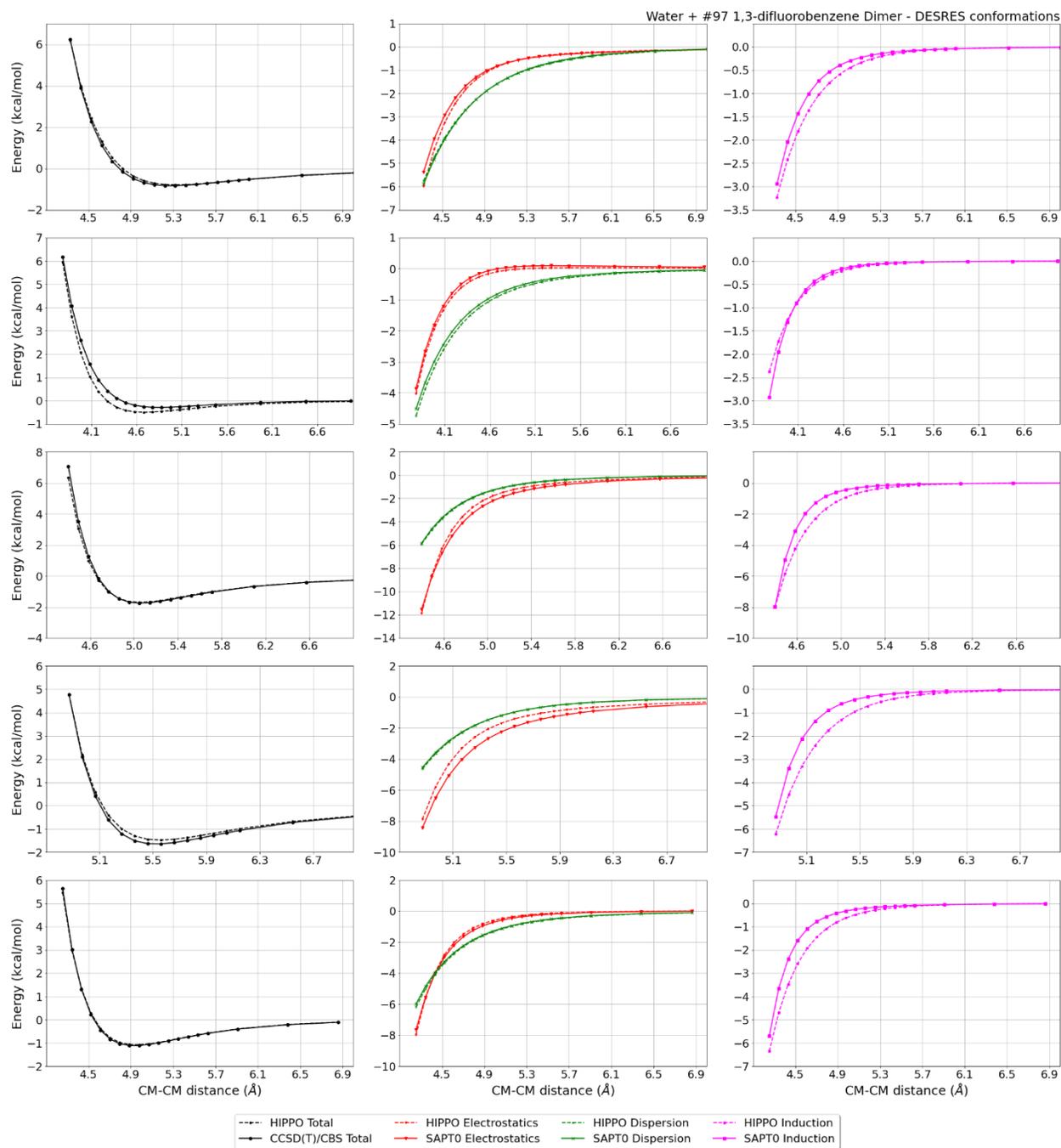
DESRES_97-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.103	0.135	0.7061	110	0

Liquid 1,3-difluorobenzene @ 298.15 K

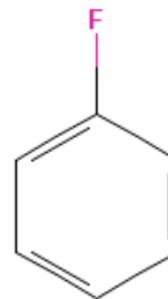
Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1168.53	1162.02	0.6	58.68	36.58	60.4	-1.00	5.06	119.8	-50





#99 Fluorobenzene C6H5F CID: 10008

ref molpol	-12.09	-11.79	-6.49, avg	-10.12
molpol	12.07	11.76	6.65, avg	10.16
rms molpol	0.02	0.03	0.16, avg	0.04



Monomer potential fitting RMS: 0.14

##Dimer results - Fitting to QM datasets##

DESRES_99-water, energy values in kcal/mol

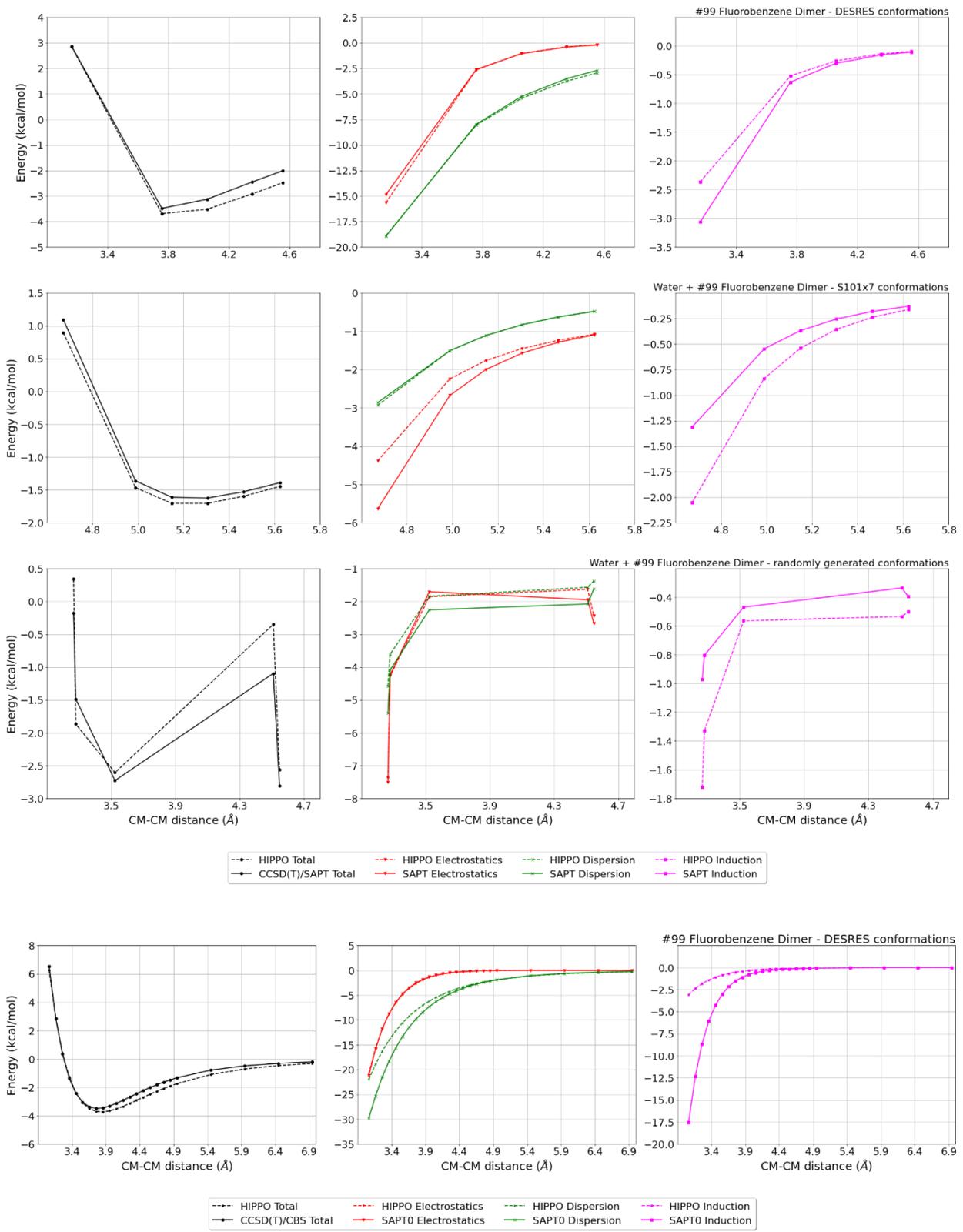
MAE	Std error	max error	#points	#count[err > 1]
0.237	0.496	4.6684	287	13

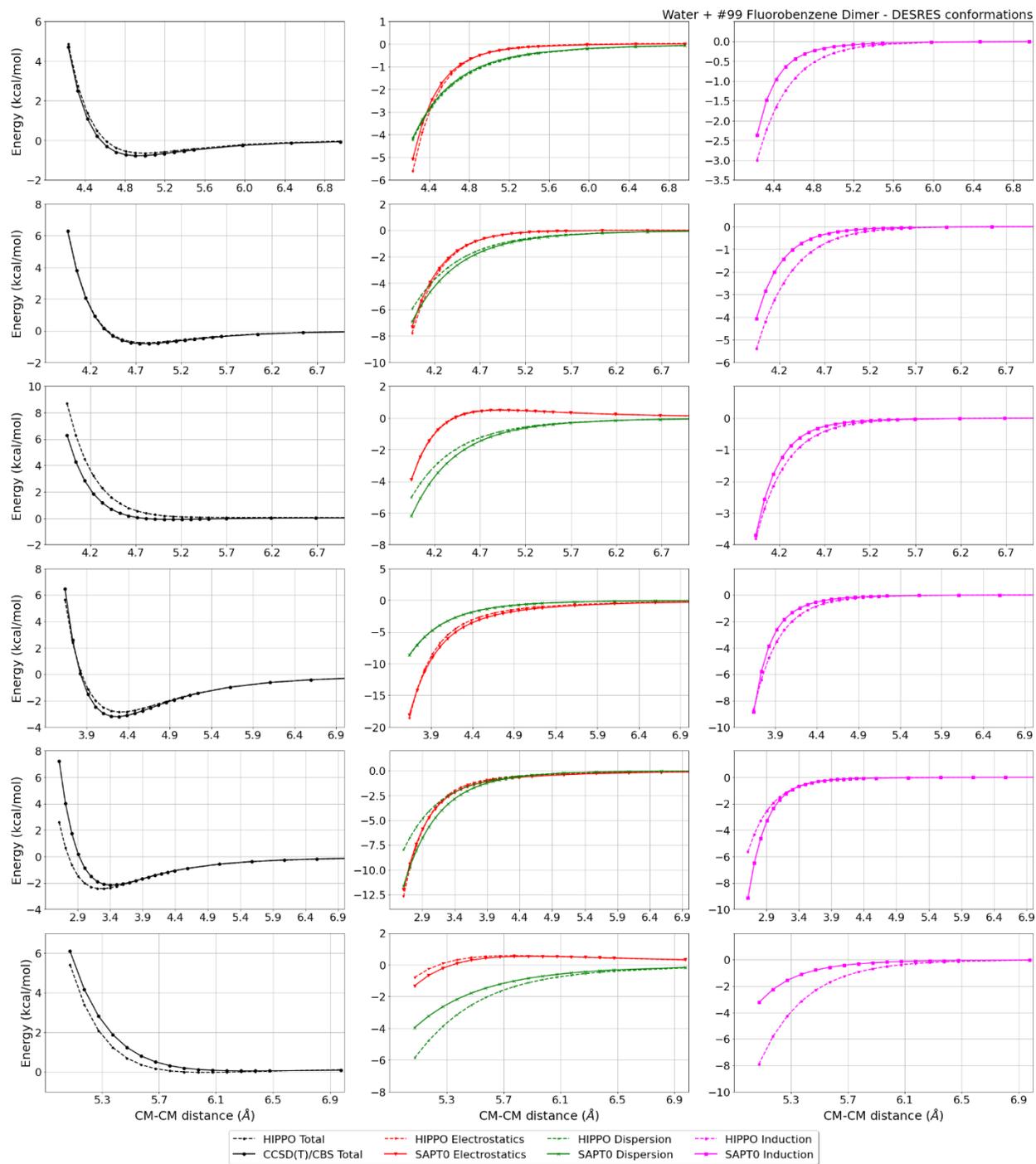
DESRES_99-99, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.270	0.162	0.4701	25	0

Liquid Fluorobenzene @ 298.15 K

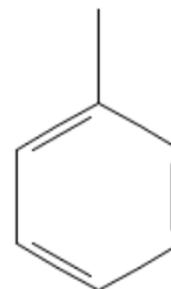
Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
1008.26	1019.10	1.1	37.94	34.58	9.7	4.74	5.34	11.2	1000





#120 Toluene C7H8 CID: 1140

ref molpol -15.08 -13.17 -7.94, avg -12.06
molpol 15.10 13.18 7.66, avg 11.98
rms molpol 0.02 0.01 0.28, avg 0.08



Monomer potential fitting RMS: 0.13

##Dimer results - Fitting to QM datasets##

DESRES_120-120, energy values in kcal/mol

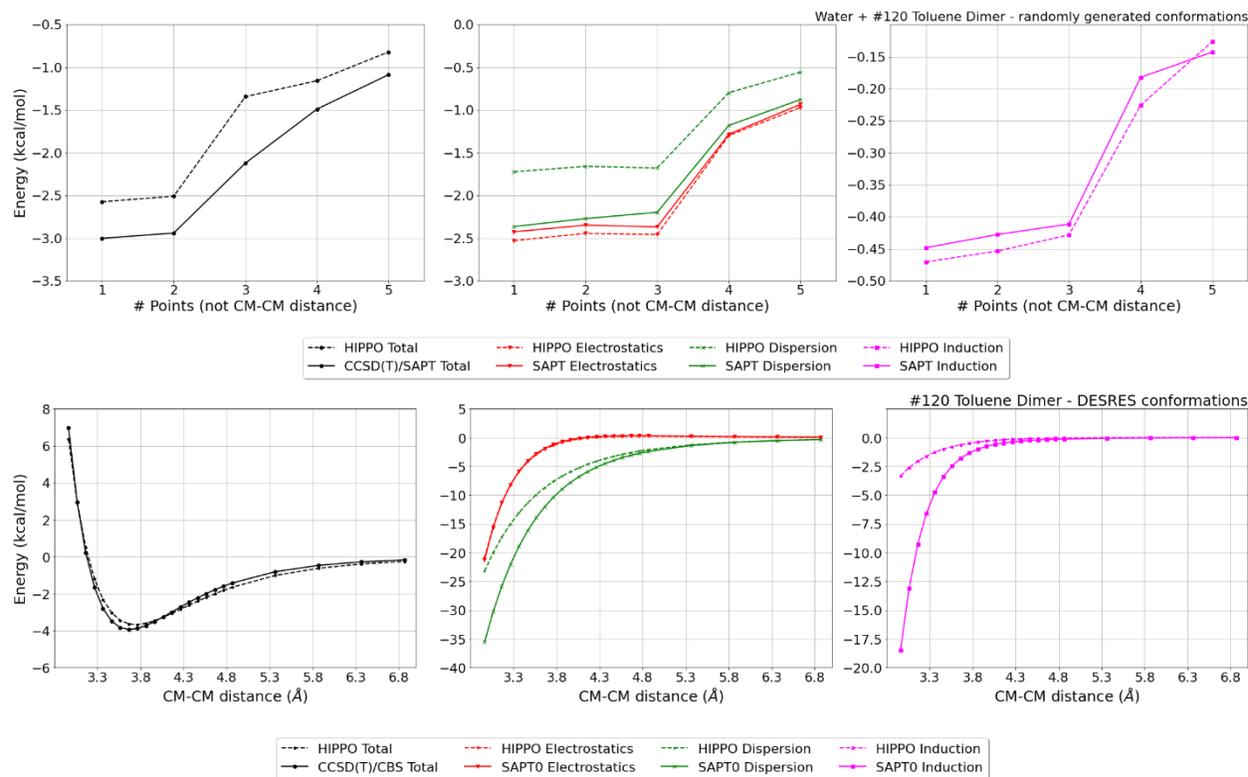
MAE	Std error	max error	#points	#count[err > 1]
0.222	0.151	0.8172	25	0

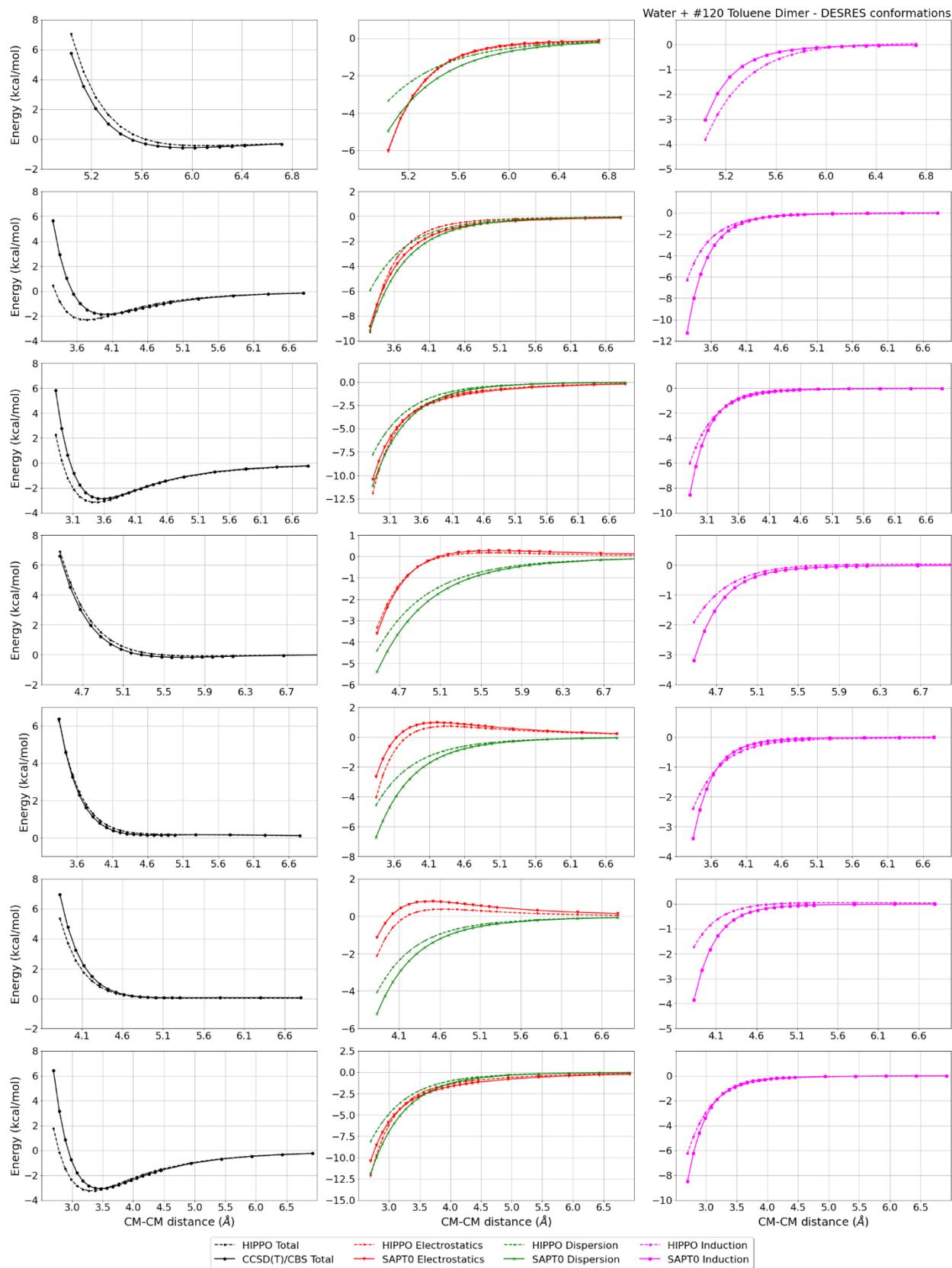
DESRES_120-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.304	0.595	5.1402	557	44

Liquid Toluene @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
850.10	861.90	1.4	34.89	37.99	8.2	2.36	2.37	0.3	3000





#151 Cyclopentane C5H10 CID: 9253

ref molpol	-7.80	-9.06	-9.07, avg	-8.64
molpol	7.79	9.07	9.08, avg	8.65
rms molpol	0.01	0.01	0.01, avg	0.00



Monomer potential fitting RMS: 0.27

##Dimer results - Fitting to QM datasets##

DESRES_151-151, energy values in kcal/mol

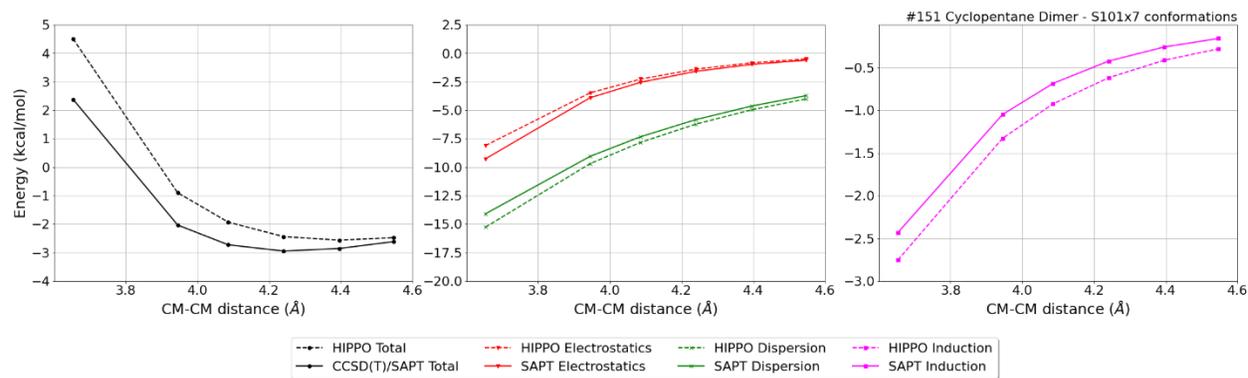
MAE	Std error	max error	#points	#count[err > 1]
0.105	0.223	2.3364	473	7

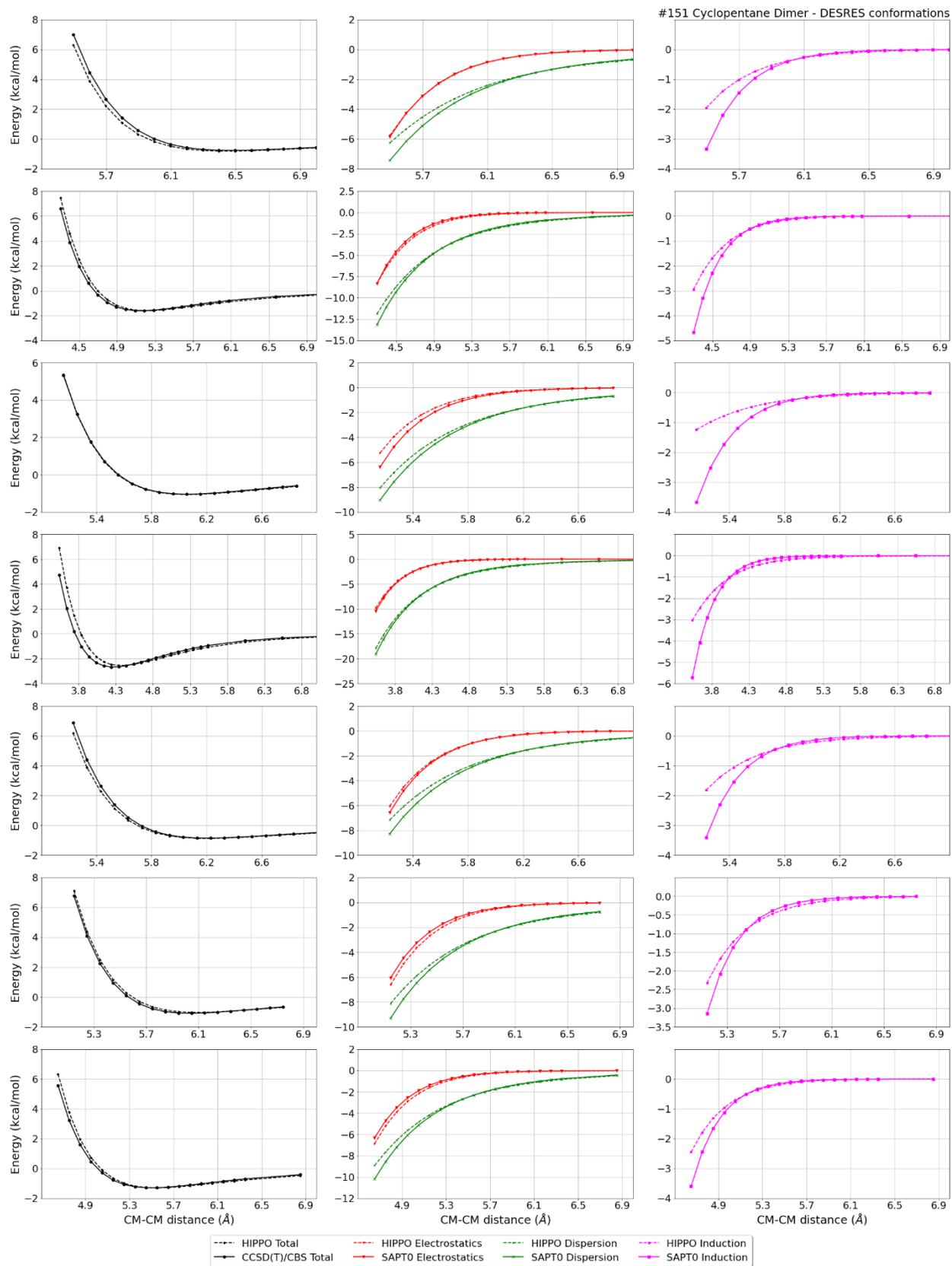
DESRES_151-water, energy values in kcal/mol

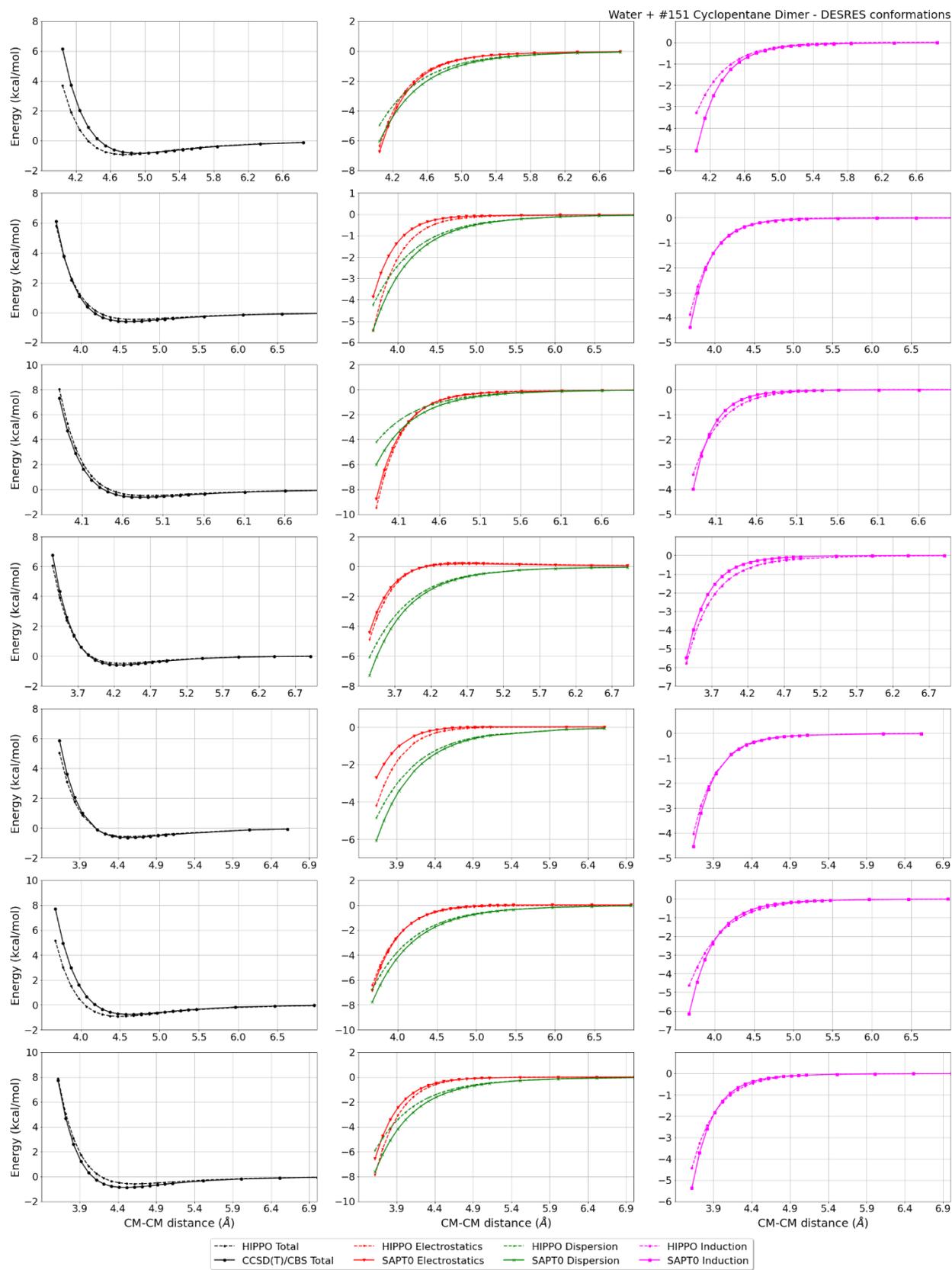
MAE	Std error	max error	#points	#count[err > 1]
0.151	0.261	2.5191	552	9

Liquid Cyclopentane @ 293.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
745.89	745.70	0.0	27.27	27.30	0.1	1.91	1.97	2.8	3000







#152 Dimethyl sulfide C2H6S CID: 1068

ref molpol	-8.27	-7.17	-6.21, avg	-7.22
molpol	8.23	7.16	6.26, avg	7.22
rms molpol	0.03	0.02	0.05, avg	0.00



Monomer potential fitting RMS: 0.62

##Dimer results - Fitting to QM datasets##

DESRES_152-152, energy values in kcal/mol

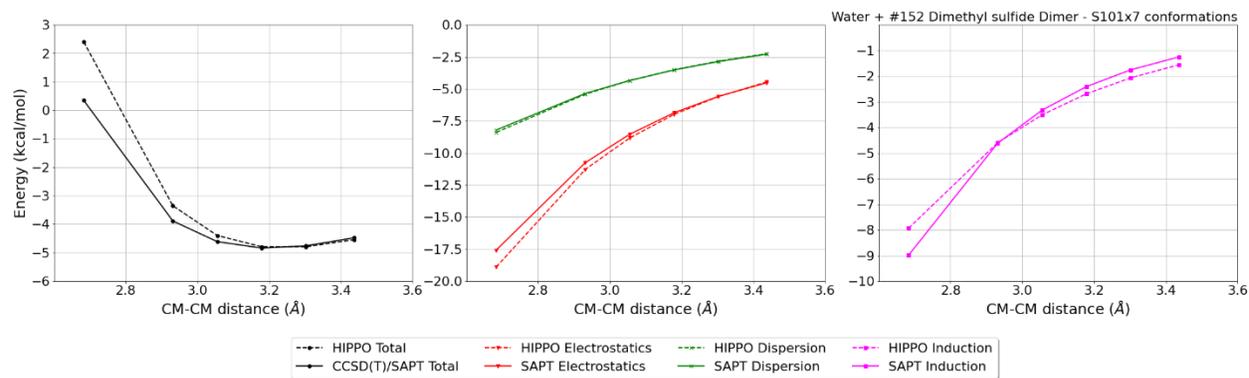
MAE	Std error	max error	#points	#count[err > 1]
0.369	0.824	8.1883	715	66

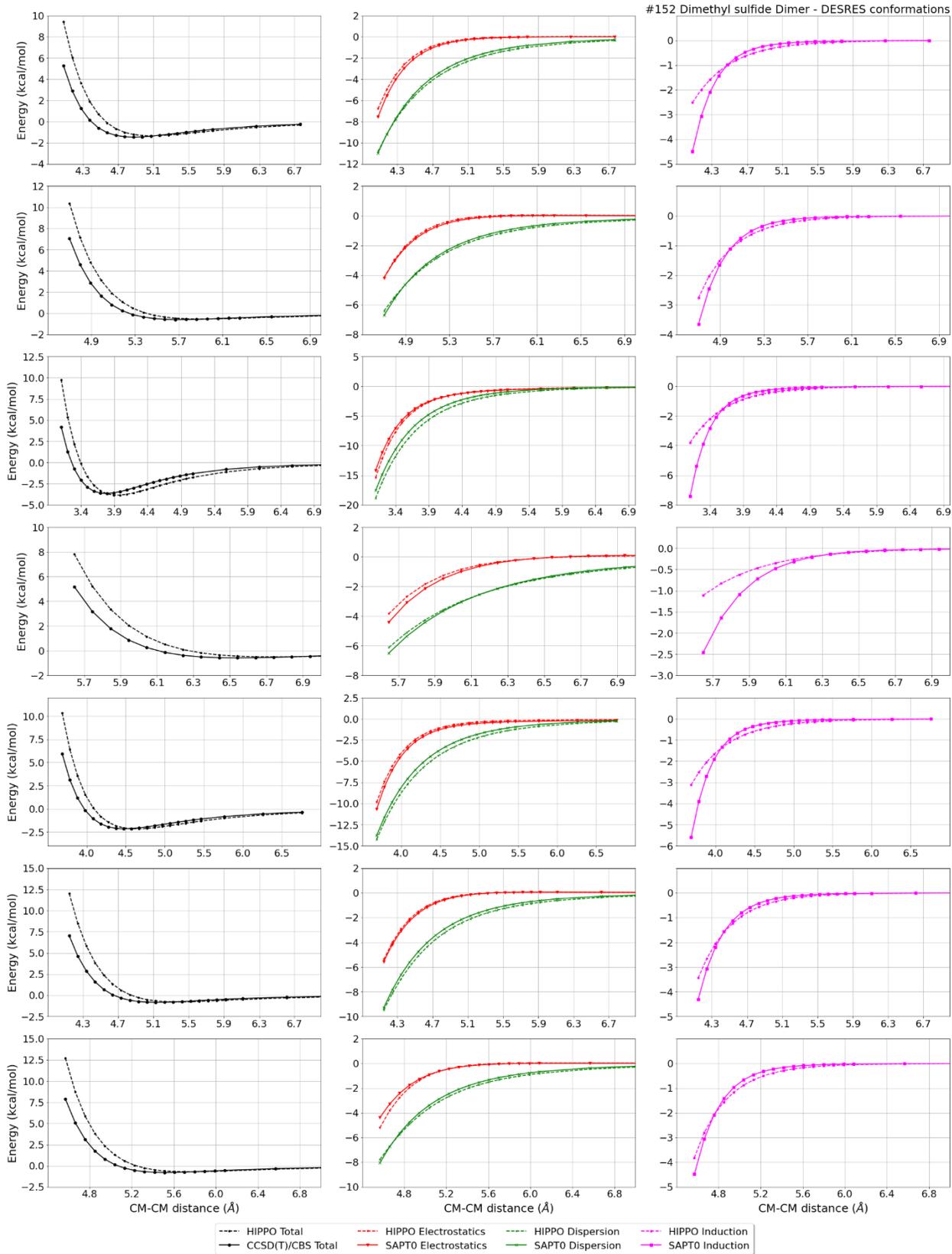
DESRES_152-water, energy values in kcal/mol

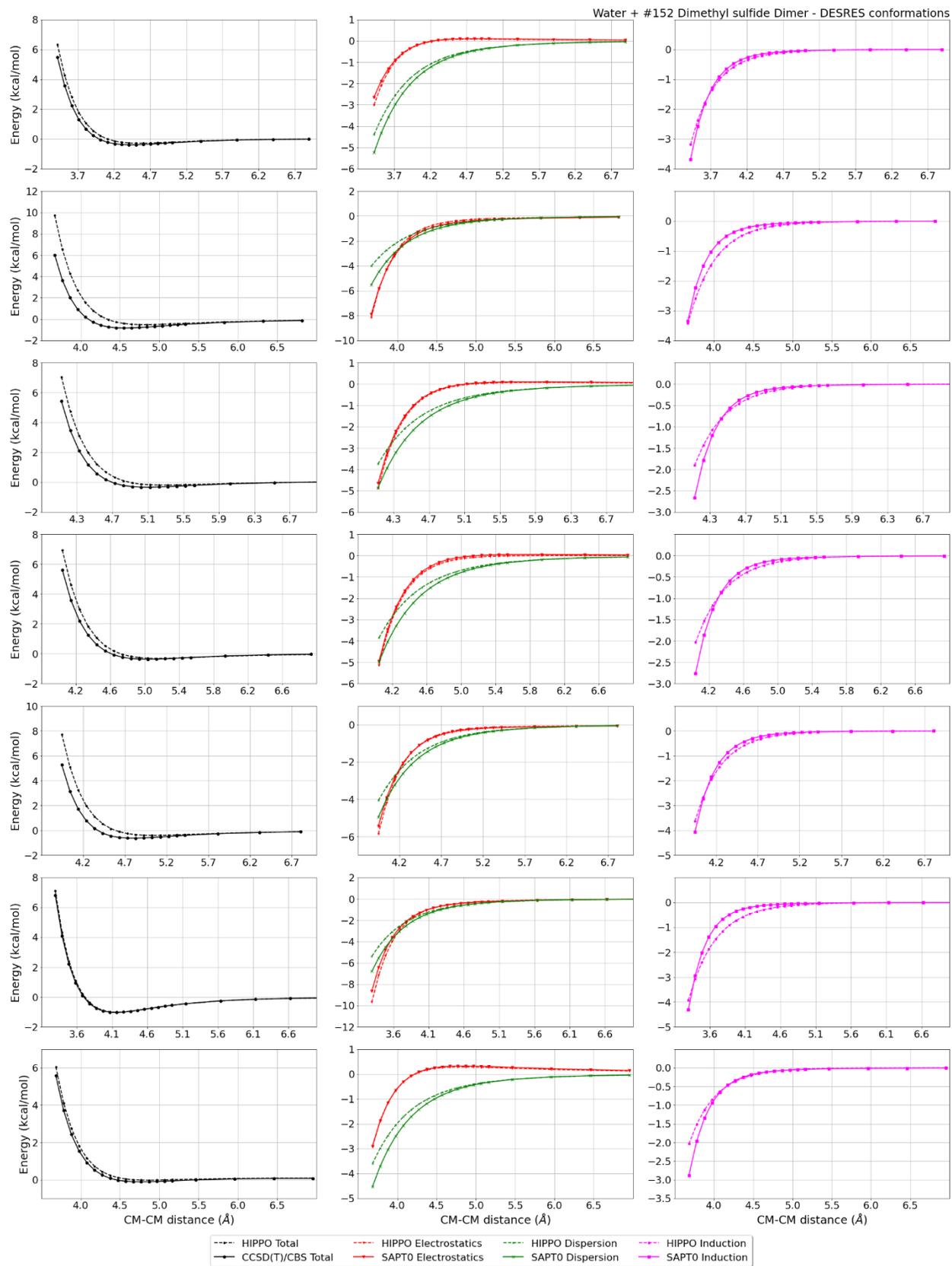
MAE	Std error	max error	#points	#count[err > 1]
0.261	0.450	3.7378	559	37

Liquid Dimethyl sulfide @ 293.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
847.29	848.30	0.1	32.88	27.65	18.9	6.42	6.70	4.1	3000







#153 Ethene C2H4 CID: 6325

ref molpol	-3.67	-3.30	-5.03, avg	-4.00
molpol	3.52	4.12	5.00, avg	4.22
rms molpol	0.15	0.82	0.03, avg	0.22

=====

Monomer potential fitting RMS: 0.18

##Dimer results - Fitting to QM datasets##

DESRES_153-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.491	0.781	4.7408	2679	456

R739x5_153-153, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
1.615	2.385	2.774	0.3885
1.682	1.242	1.198	-0.0443
1.750	0.538	0.273	-0.2648
1.819	0.122	-0.242	-0.3638
1.957	-0.227	-0.613	-0.3862

MAE	Std error	max error	#points	#count[err > 1]
0.290	0.131	0.3885	5	0

R739x5_153-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
1.840	2.229	3.347	1.1184
1.926	1.476	2.206	0.7301
2.012	1.011	1.492	0.4809
2.098	0.726	1.047	0.3206
2.271	0.445	0.594	0.1488

MAE	Std error	max error	#points	#count[err > 1]
0.560	0.339	1.1184	5	1

DESRES_153-153, energy values in kcal/mol

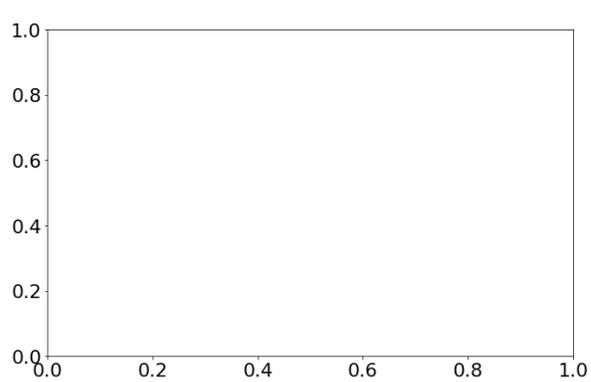
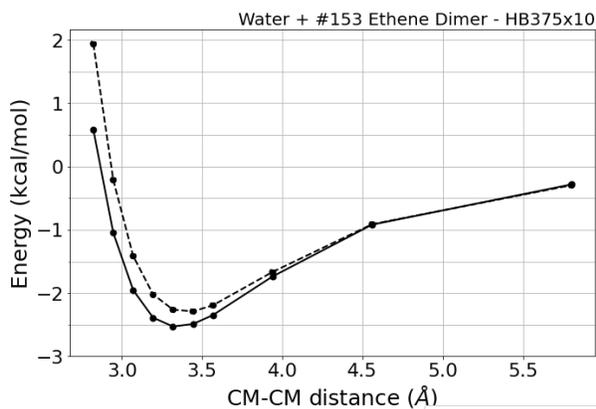
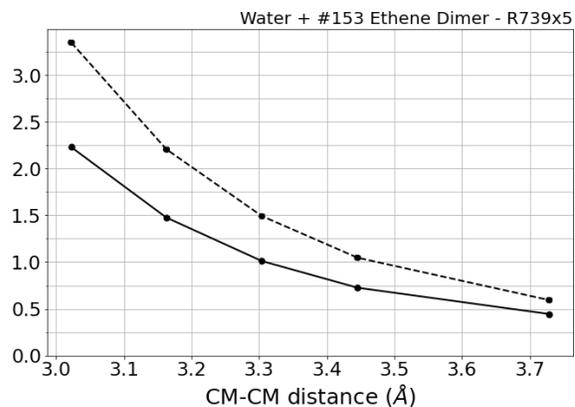
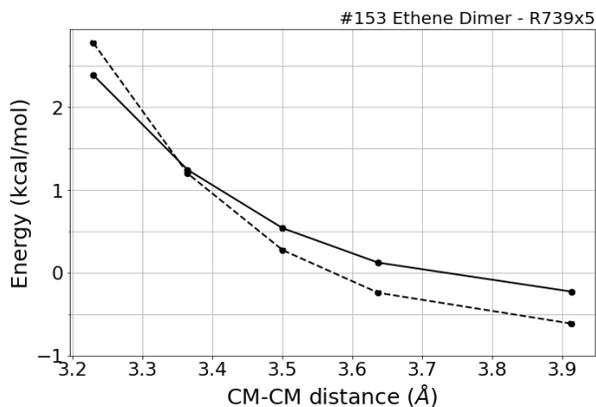
MAE	Std error	max error	#points	#count[err > 1]
0.328	0.523	3.3714	1627	164

HB375x10_153-water, energy values in kcal/mol

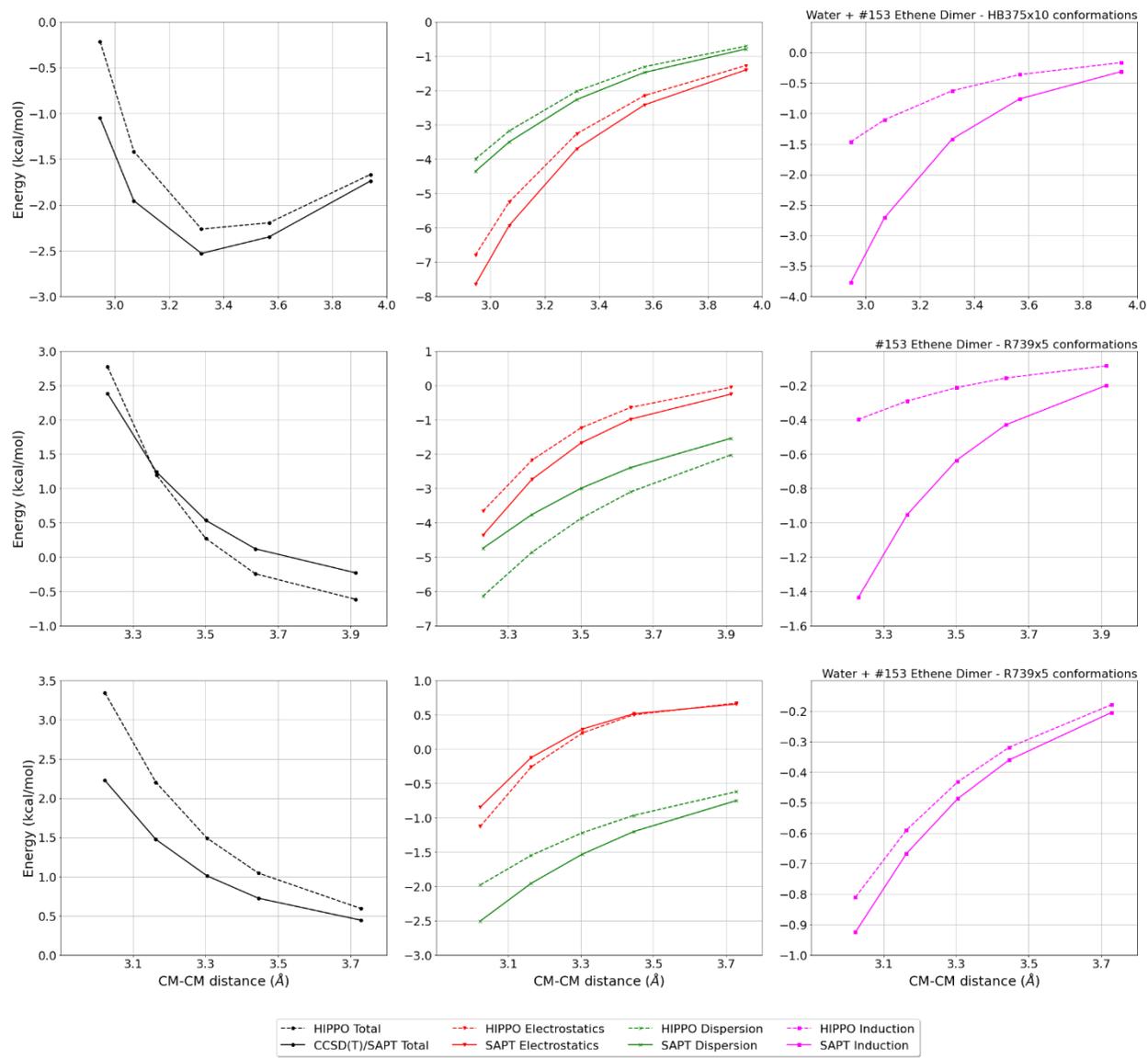
CM-CM (A)	Reference	HIPPO res	Abs diff
1.718	0.583	1.940	1.3567
1.794	-1.048	-0.213	0.8354
1.869	-1.954	-1.415	0.5391
1.945	-2.389	-2.021	0.3678
2.021	-2.529	-2.263	0.2660
2.096	-2.490	-2.289	0.2009
2.172	-2.349	-2.194	0.1548
2.399	-1.738	-1.667	0.0715
2.777	-0.919	-0.912	0.0069

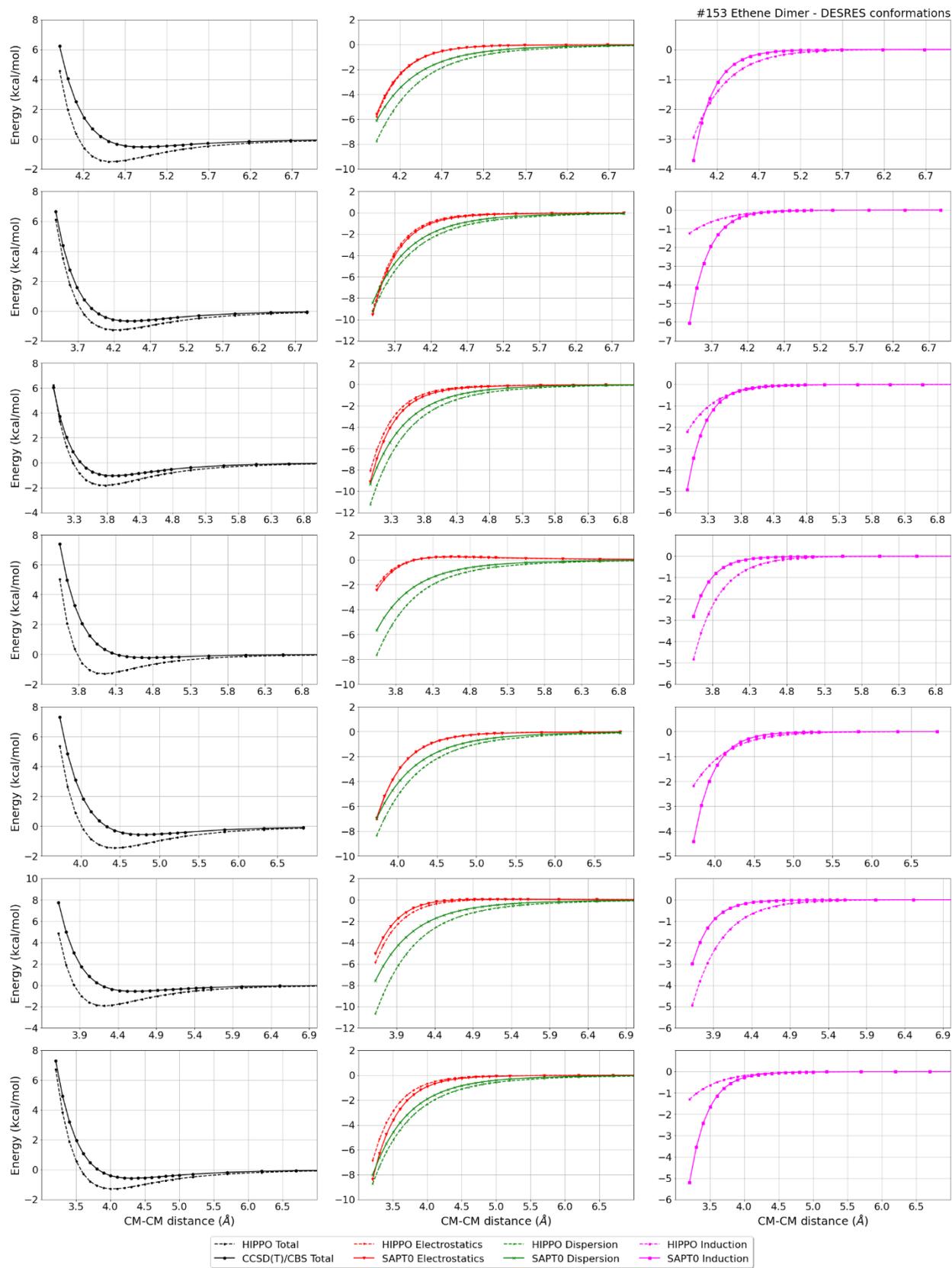
3.533 -0.284 -0.297 -0.0126

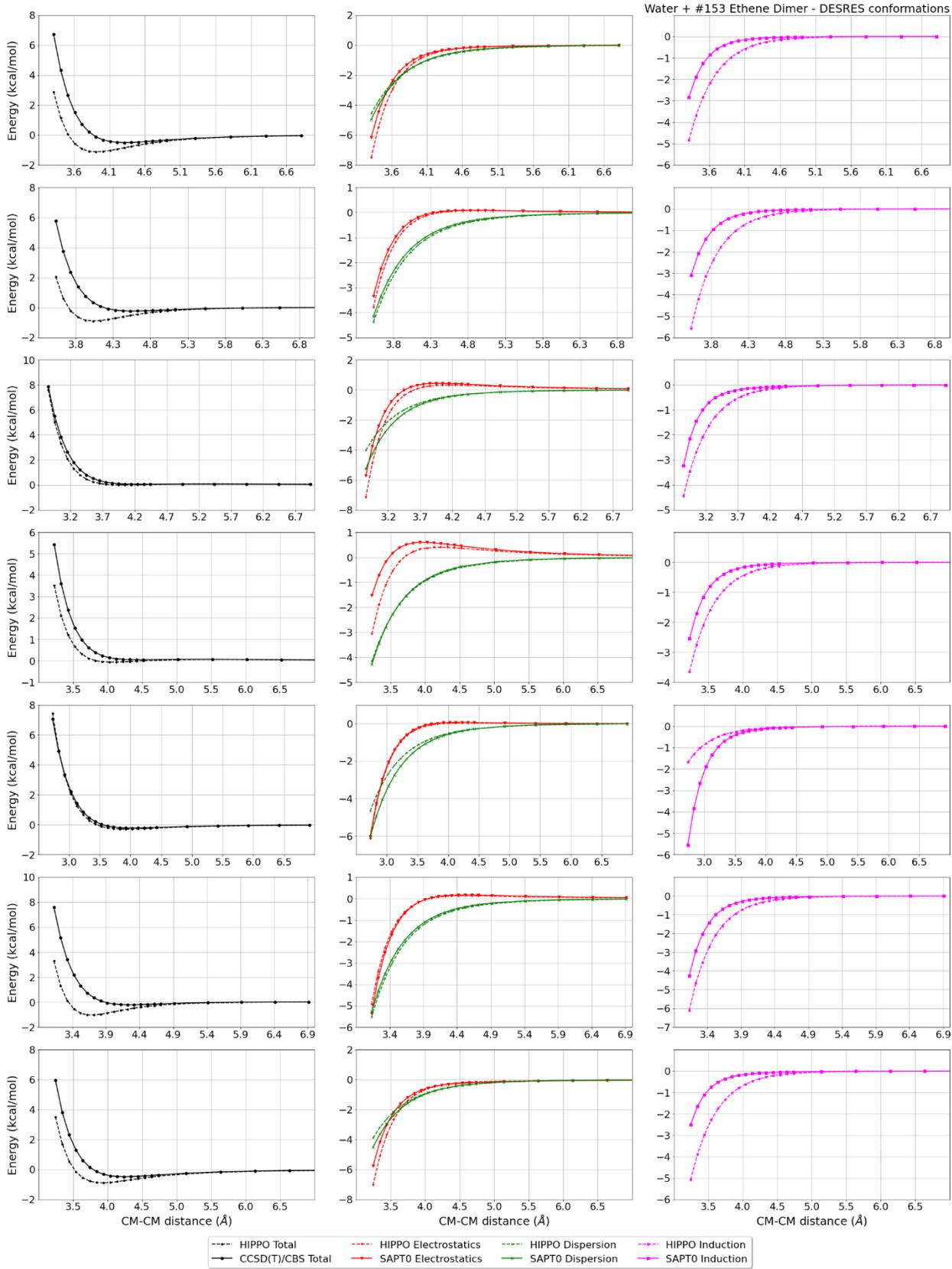
MAE	Std error	max error	#points	#count[err > 1]
0.381	0.406	1.3567	10	1



—●— CCSD(T) Total - - - ■ - - - HIPPO Total

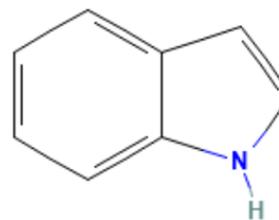






#155 Indole C8H7N CID: 798

ref molpol -20.48 -15.62 -8.87, avg -14.99
molpol 20.09 15.31 8.78, avg 14.73
rms molpol 0.39 0.32 0.09, avg 0.26

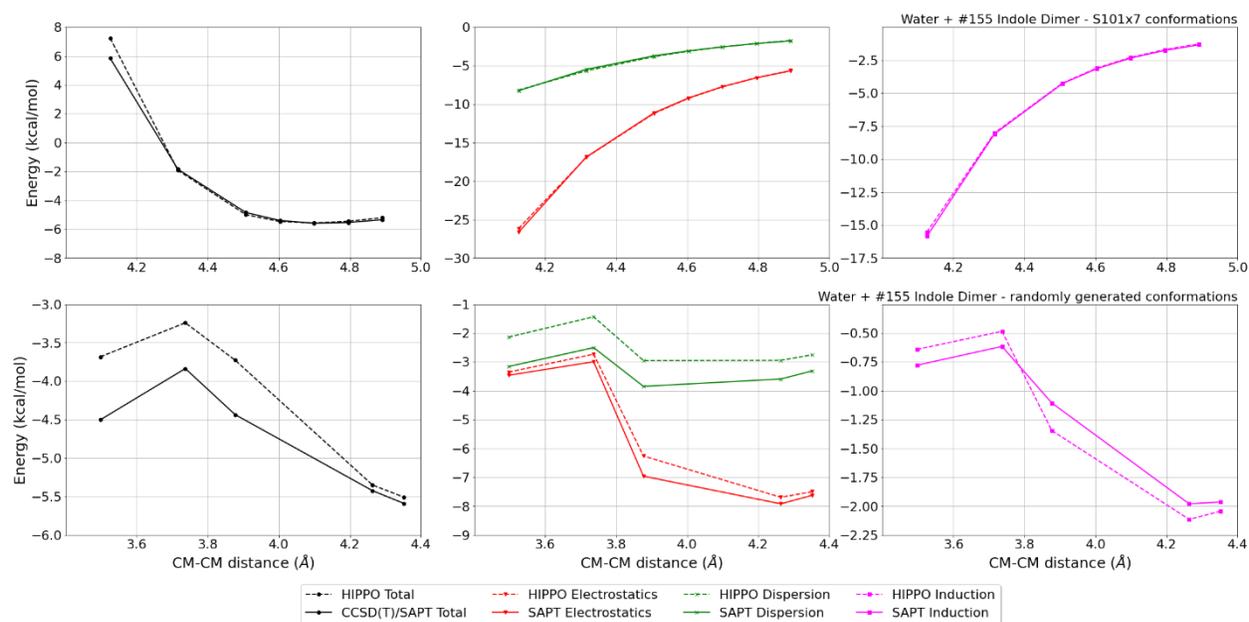


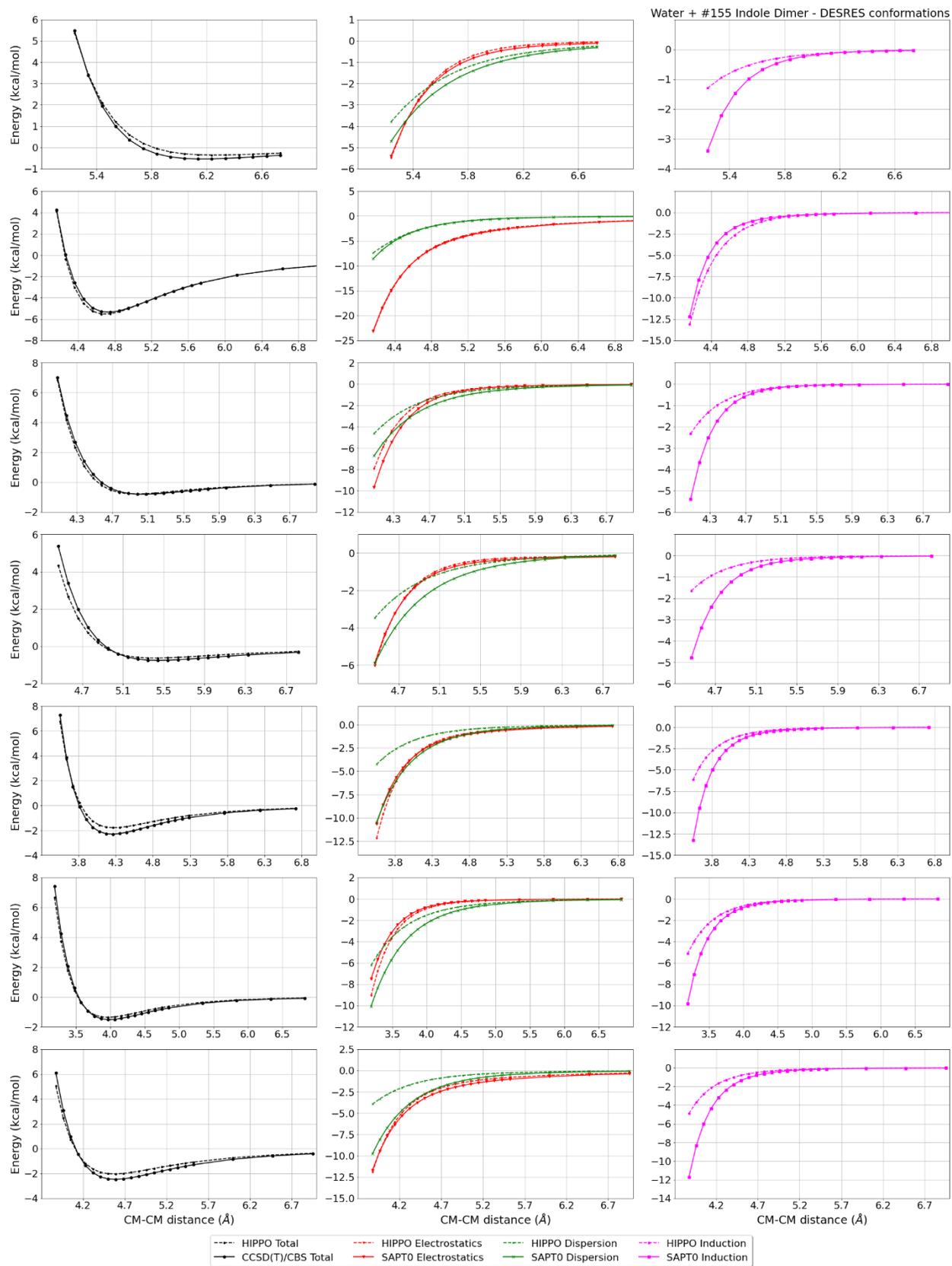
Monomer potential fitting RMS: 0.17

##Dimer results - Fitting to QM datasets##

DESRES_155-water, energy values in kcal/mol

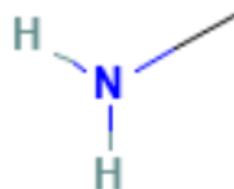
MAE	Std error	max error	#points	#count[err > 1]
0.255	0.376	5.0186	558	17





#156 Methyl amine CH5N CID: 6329

ref molpol	-3.53	-3.64	-4.16, avg	-3.77
molpol	3.55	3.56	4.12, avg	3.74
rms molpol	0.02	0.08	0.04, avg	0.04



Monomer potential fitting RMS: 0.12

##Dimer results - Fitting to QM datasets##

R739x5_156-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
2.362	2.333	2.691	0.3580		
2.467	1.960	2.192	0.2324		
2.572	1.685	1.842	0.1567		
2.677	1.473	1.582	0.1093		
2.888	1.161	1.219	0.0584		
MAE	Std error	max error	#points	#count[err > 1]	
0.183	0.105	0.3580	5	0	

R739x5_156-156, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
2.142	2.482	2.659	0.1770		
2.227	2.045	2.141	0.0961		
2.313	1.723	1.779	0.0561		
2.398	1.478	1.513	0.0348		
2.571	1.126	1.143	0.0174		
MAE	Std error	max error	#points	#count[err > 1]	
0.076	0.057	0.1770	5	0	

HB375x10_156-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
1.705	-2.777	0.664	3.4408		
1.765	-5.228	-3.429	1.7991		
1.824	-6.624	-5.744	0.8804		
1.883	-7.303	-6.907	0.3956		
1.943	-7.504	-7.343	0.1611		
2.002	-7.395	-7.330	0.0652		
2.062	-7.094	-7.052	0.0423		
2.241	-5.715	-5.620	0.0949		
2.541	-3.535	-3.404	0.1312		
3.141	-1.351	-1.300	0.0506		
MAE	Std error	max error	#points	#count[err > 1]	
0.706	1.053	3.4408	10	2	

DESRES_156-156, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.095	0.262	3.5873	1306	21

DESRES_156-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.212	0.508	7.3904	564	22

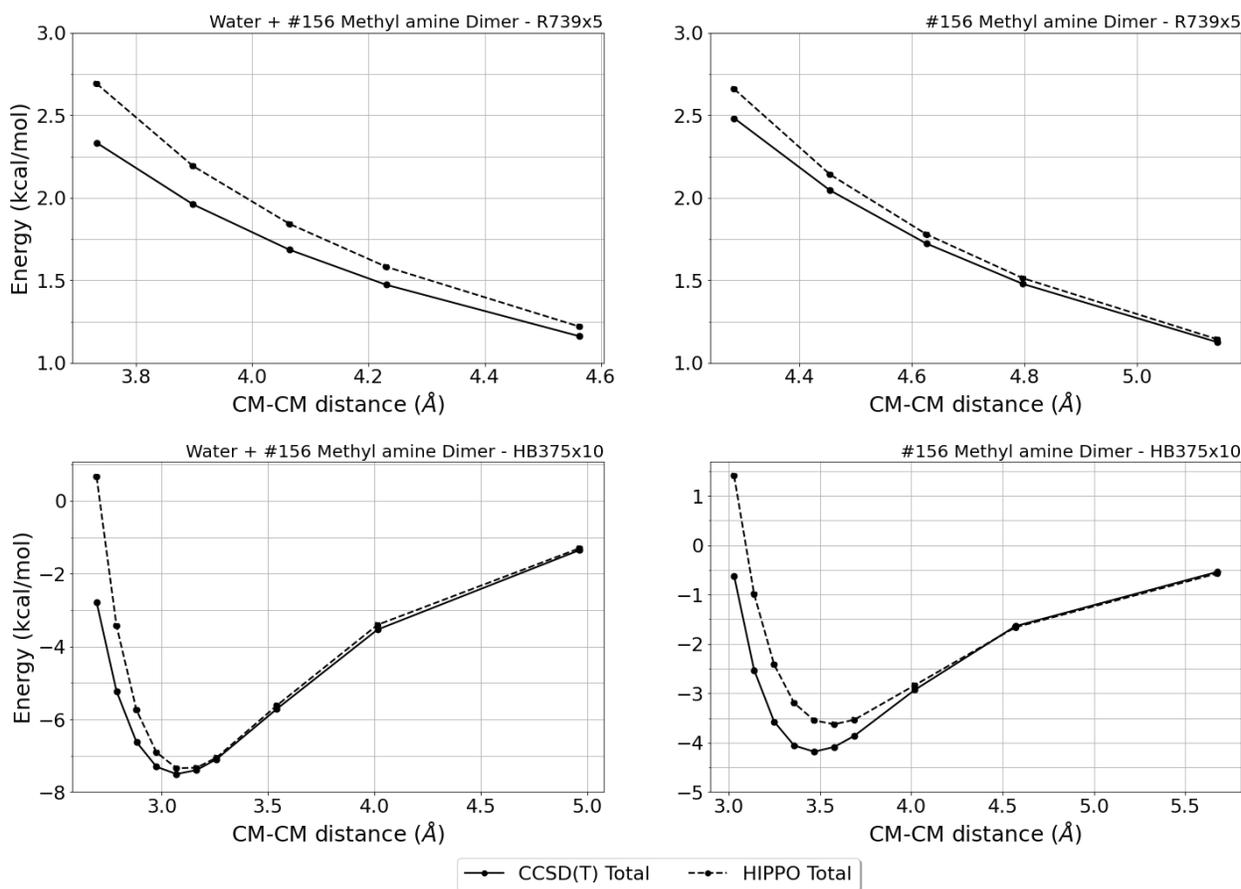
HB375x10_156-156, energy values in kcal/mol

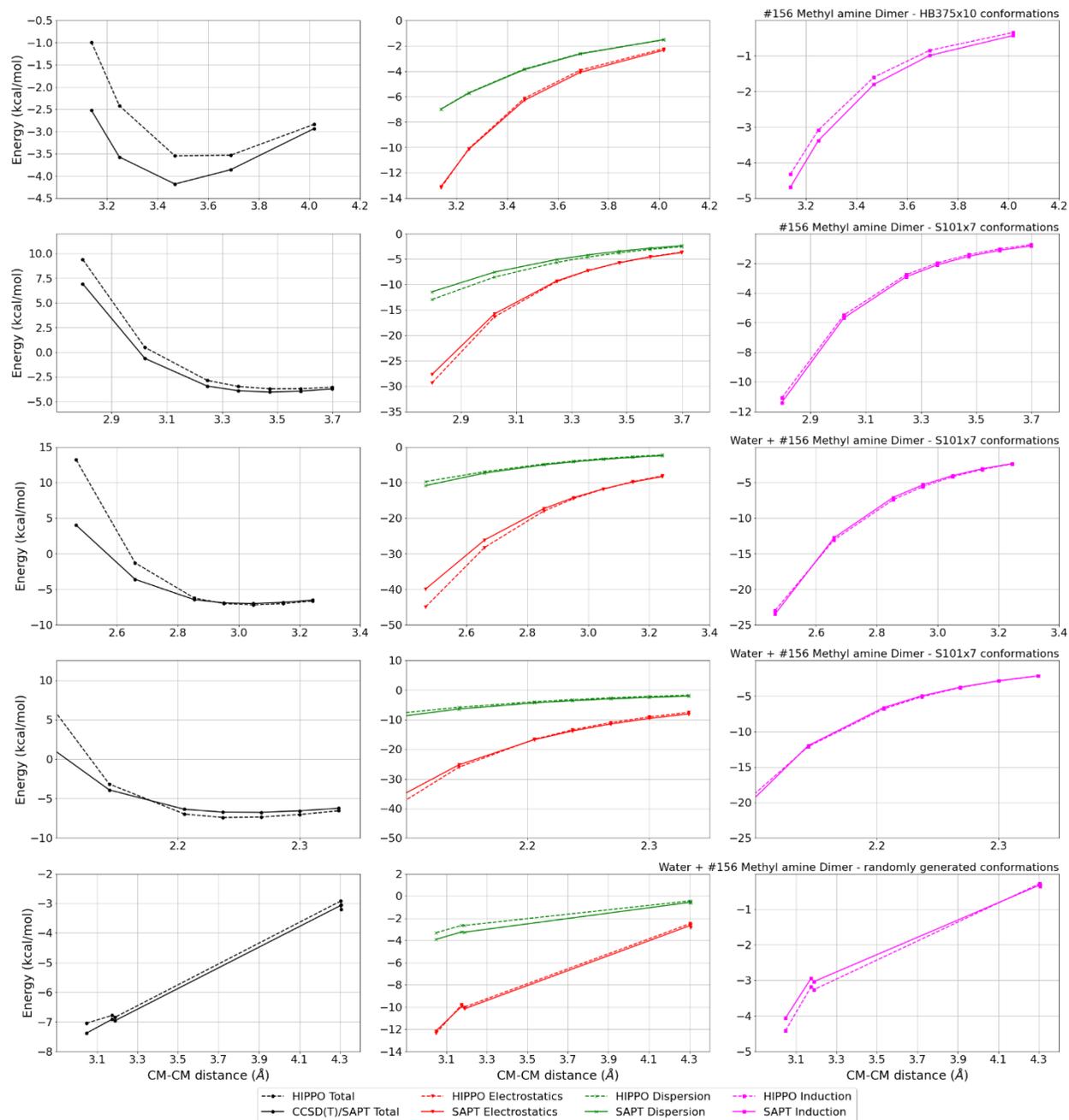
CM-CM (A)	Reference	HIPPO res	Abs diff
1.515	-0.621	1.405	2.0259
1.569	-2.527	-0.990	1.5366
1.624	-3.573	-2.418	1.1549
1.679	-4.057	-3.197	0.8598
1.734	-4.182	-3.549	0.6330
1.789	-4.085	-3.626	0.4586
1.844	-3.860	-3.534	0.3259
2.009	-2.932	-2.837	0.0952
2.285	-1.641	-1.664	-0.0227
2.837	-0.544	-0.575	-0.0313

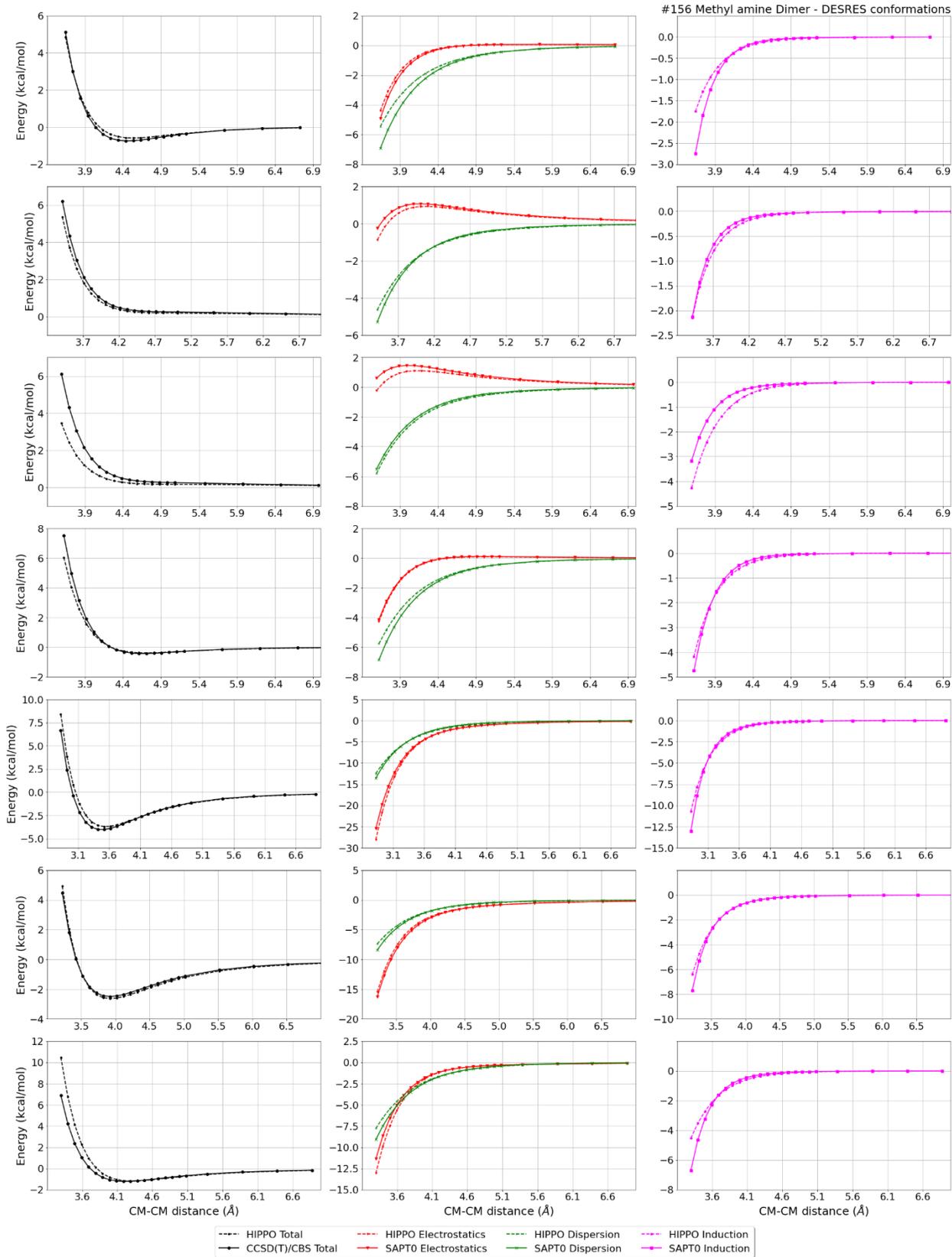
MAE	Std error	max error	#points	#count[err > 1]
0.714	0.645	2.0259	10	3

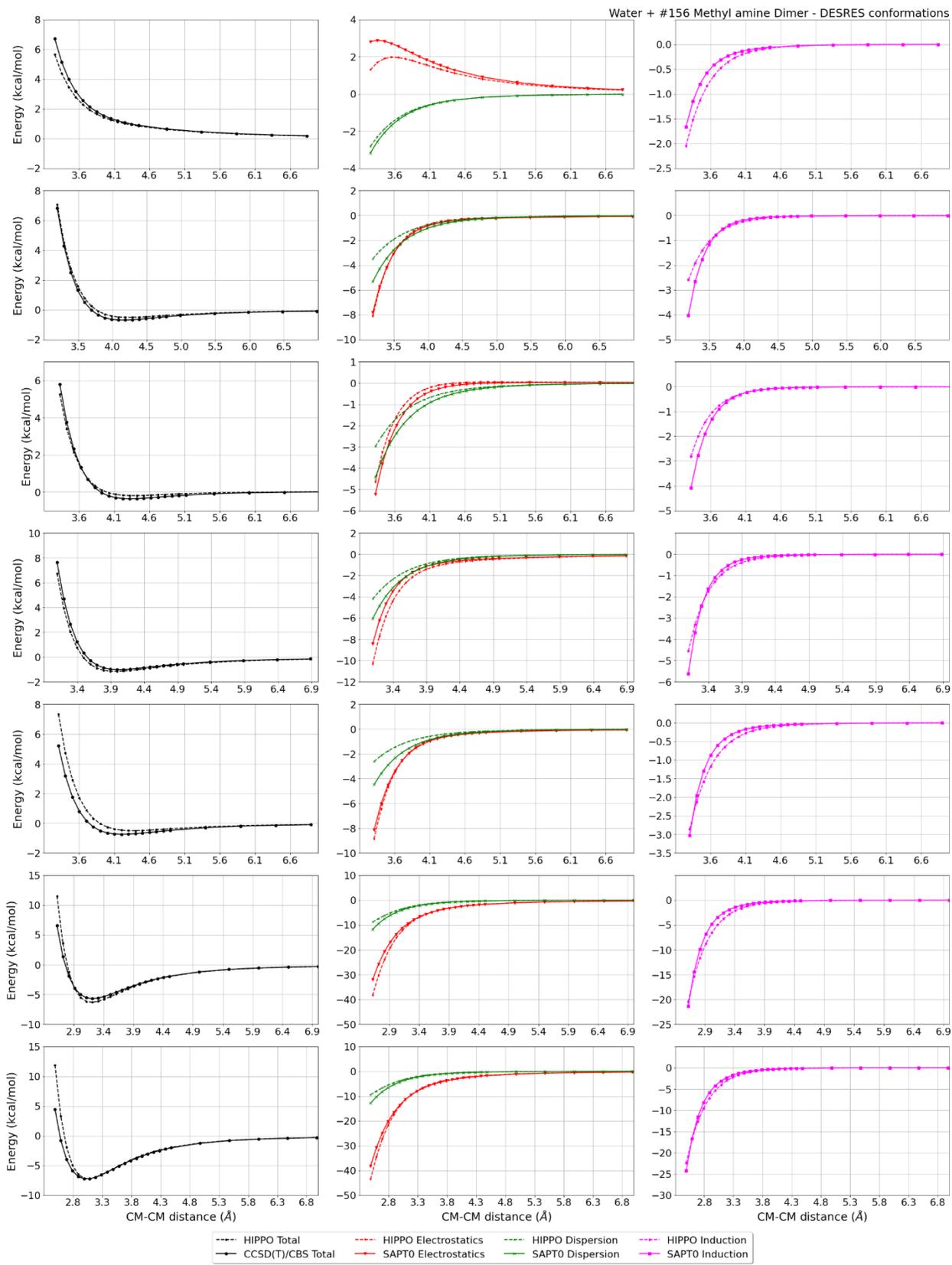
Liquid Methyl amine @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
701.34	656.00	6.9	27.21	23.37	16.4	8.26	16.55	50.1	3000









#157 Methyl chloride CH3Cl CID: 6327

ref molpol	-5.26	-3.82	-3.82, avg	-4.30
molpol	4.88	3.95	3.95, avg	4.26
rms molpol	0.38	0.13	0.13, avg	0.04

Monomer potential fitting RMS: 0.24

##Dimer results - Fitting to QM datasets##

HB300SPXx10_157-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
1.997	0.595	0.497	-0.0978		
2.089	-1.805	-1.515	0.2904		
2.181	-3.072	-2.589	0.4826		
2.272	-3.631	-3.079	0.5522		
2.364	-3.761	-3.212	0.5487		
2.455	-3.645	-3.138	0.5066		
2.547	-3.399	-2.952	0.4473		
2.822	-2.465	-2.199	0.2658		
3.280	-1.313	-1.218	0.0946		
4.196	-0.448	-0.433	0.0152		
MAE	Std error	max error	#points	#count[err > 1]	
0.330	0.195	0.5522	10	0	

DESRES_157-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.237	0.517	4.1283	289	16

DESRES_157-157, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.084	0.081	0.2467	24	0

R739x5_157-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
2.305	2.068	2.891	0.8232		
2.414	1.299	1.886	0.5870		
2.523	0.854	1.271	0.4167		
2.633	0.597	0.891	0.2944		
2.852	0.362	0.507	0.1448		
MAE	Std error	max error	#points	#count[err > 1]	
0.453	0.235	0.8232	5	0	

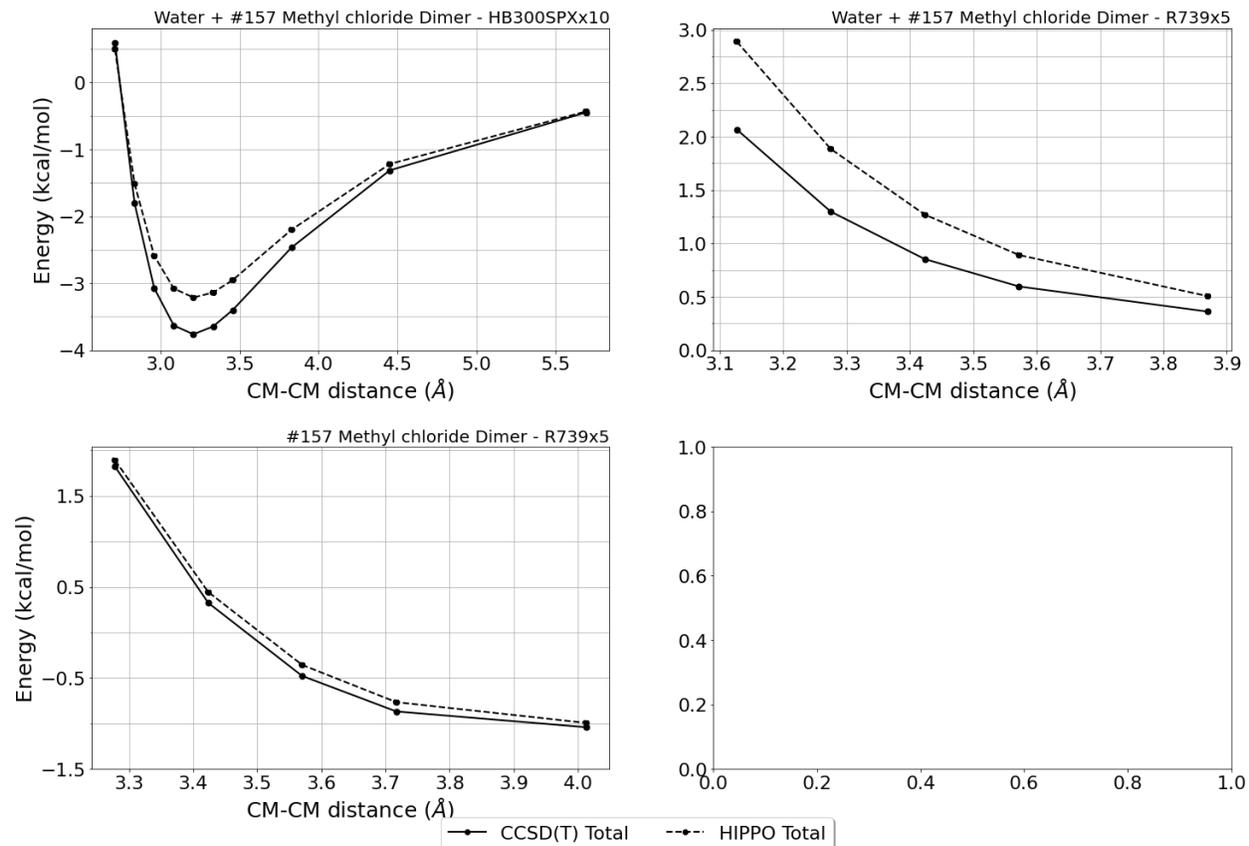
R739x5_157-157, energy values in kcal/mol

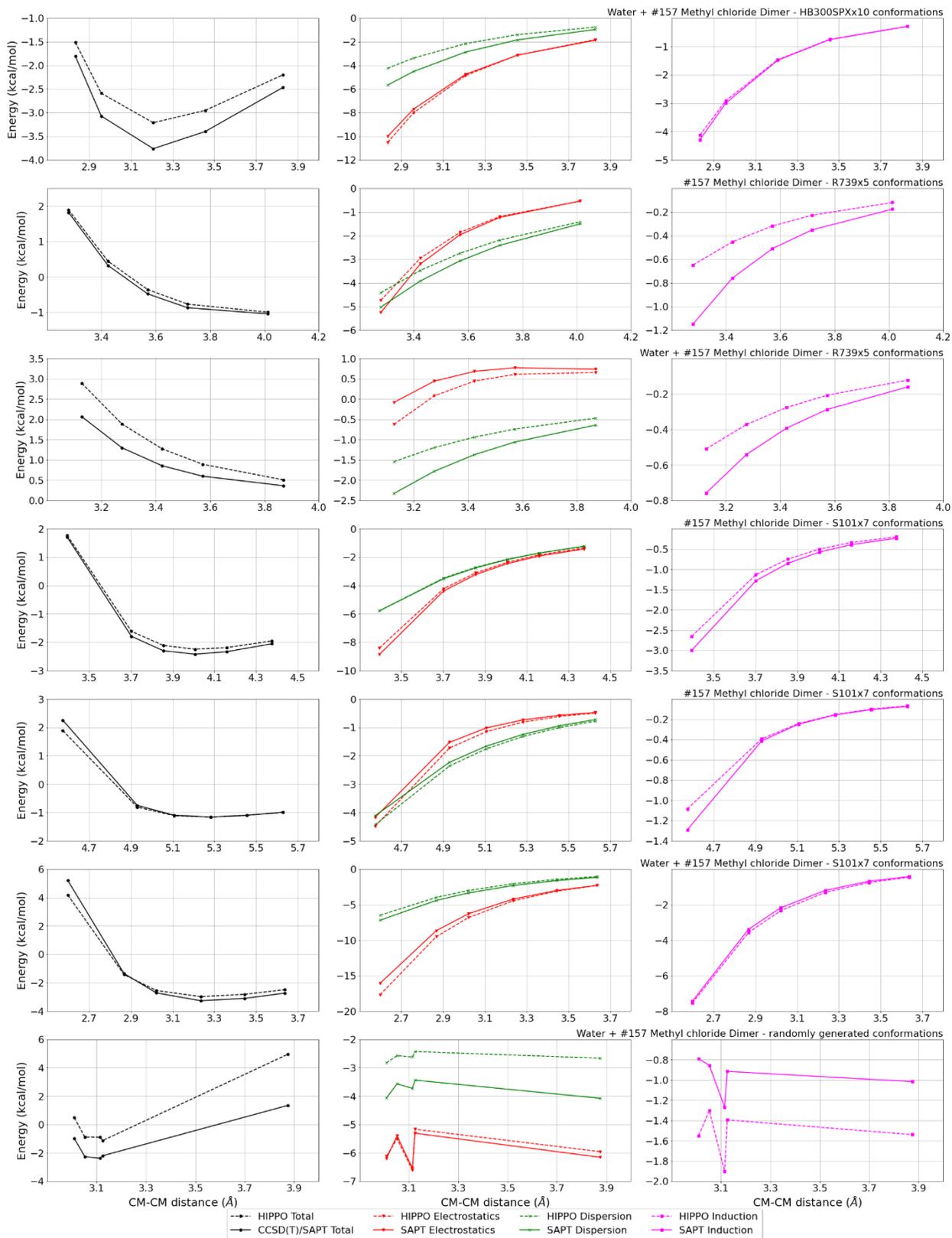
CM-CM (A)	Reference	HIPPO res	Abs diff
1.639	1.821	1.891	0.0702
1.712	0.325	0.444	0.1187
1.785	-0.477	-0.356	0.1213
1.858	-0.867	-0.765	0.1020
2.006	-1.040	-0.993	0.0473

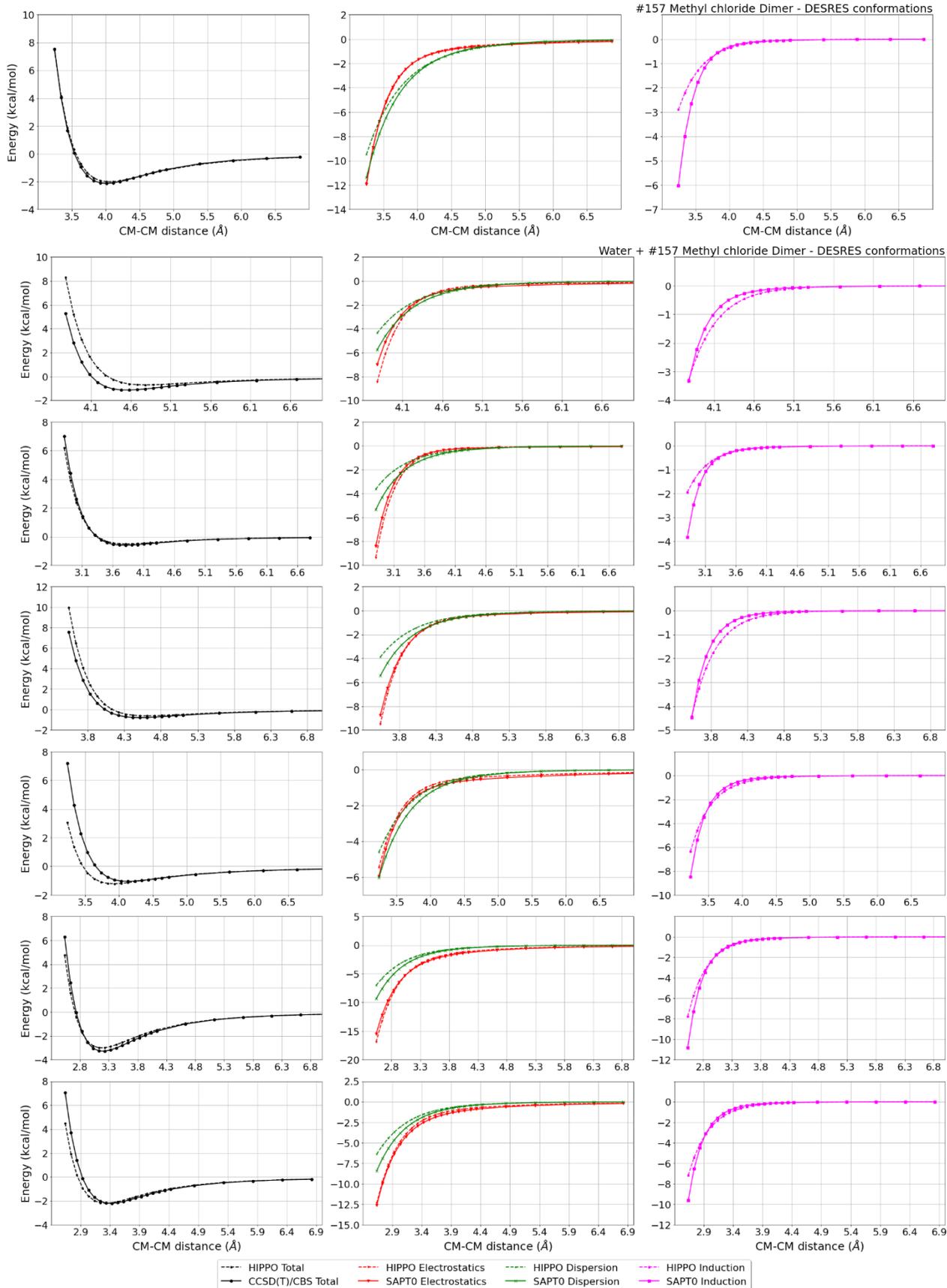
MAE	Std error	max error	#points	#count[err > 1]
0.092	0.029	0.1213	5	0

Liquid Methyl chloride @ 298.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
953.35	911.00	4.6	18.90	18.92	0.1	10.06	9.76	3.1	3000

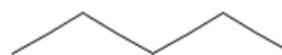






#161 Pentane C5H12 CID: 8003

ref molpol	-11.14	-8.98	-8.37, avg	-9.50
molpol	11.16	8.85	8.44, avg	9.48
rms molpol	0.02	0.14	0.08, avg	0.01



Monomer potential fitting RMS: 0.18

##Dimer results - Fitting to QM datasets##

DESRES_161-161, energy values in kcal/mol

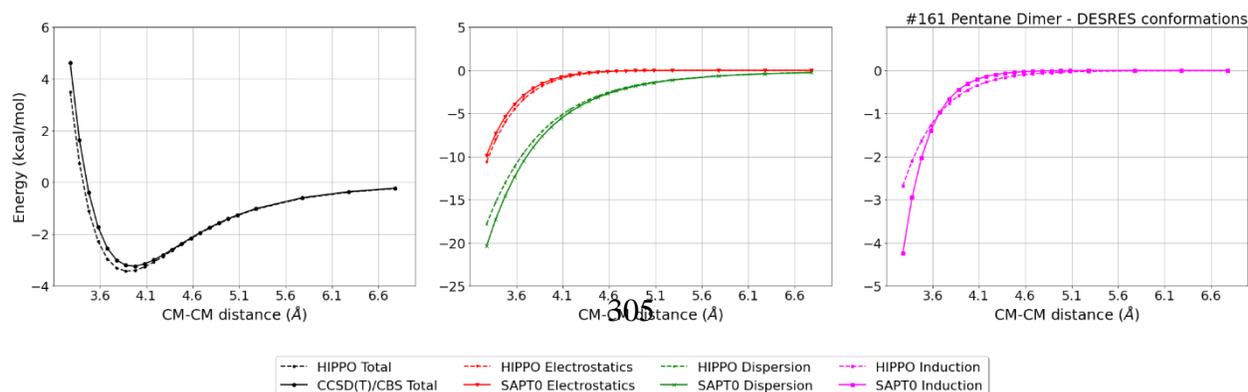
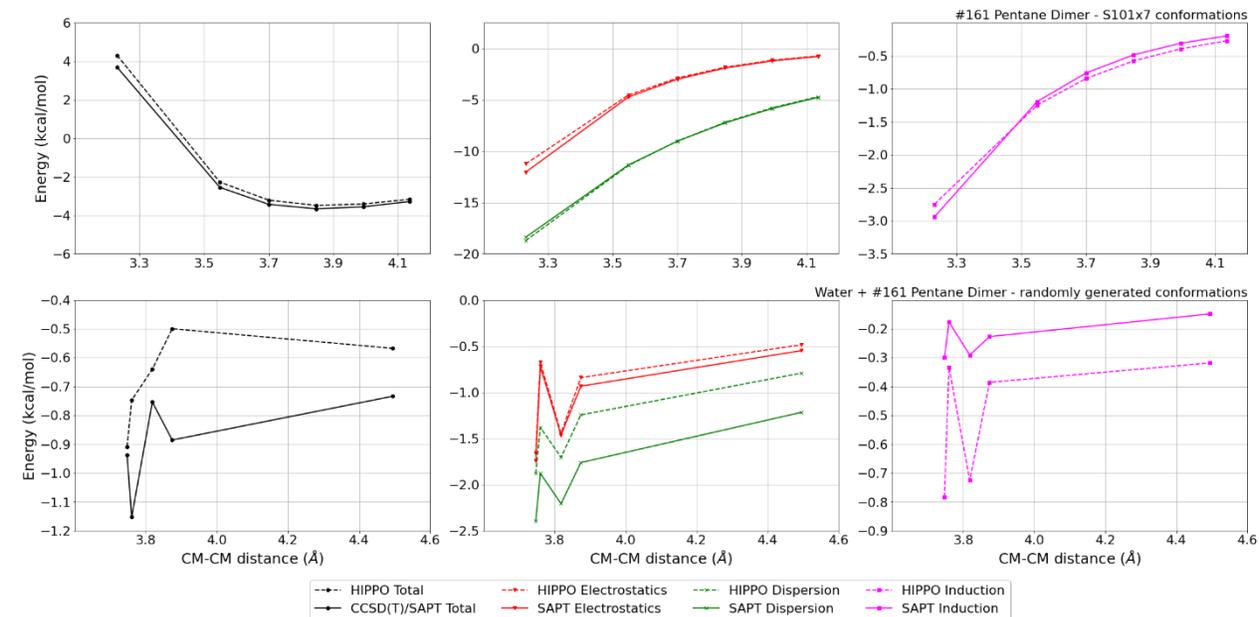
MAE	Std error	max error	#points	#count[err > 1]
0.106	0.179	1.2400	175	2

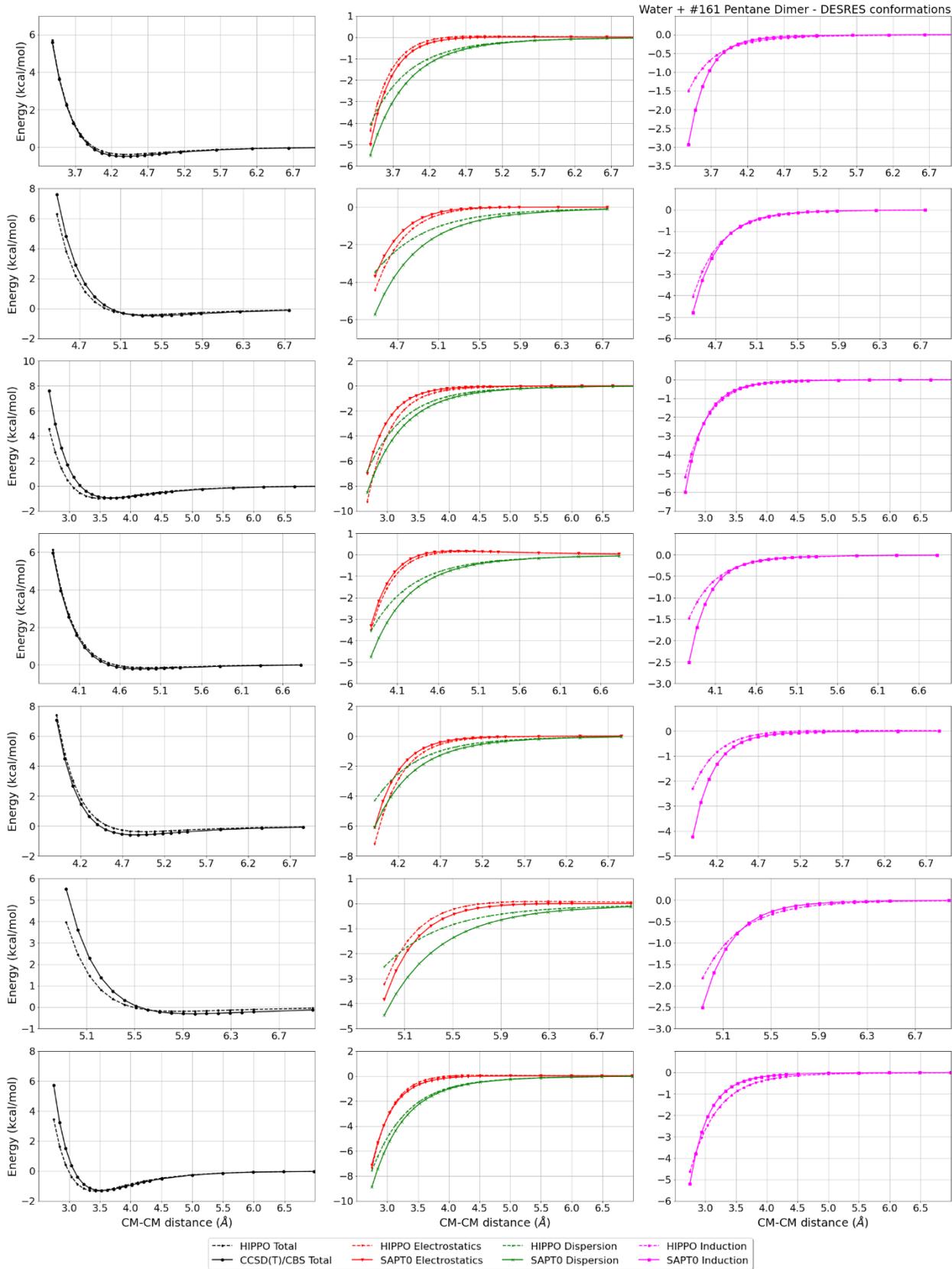
DESRES_161-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.180	0.361	3.3420	561	23

Liquid Pentane @ 293.15 K

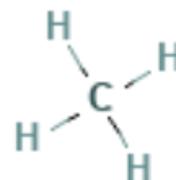
Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
645.70	626.20	3.1	23.63	26.43	10.6	1.86	1.84	1.2	5000





#164 Methane CH4 CID: 297

ref molpol	-2.43	-2.43	-2.43, avg	-2.43
molpol	2.65	2.65	2.65, avg	2.65
rms molpol	0.21	0.22	0.22, avg	0.22



Monomer potential fitting RMS: 0.06

##Dimer results - Fitting to QM datasets##

DESRES_164-164, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.194	0.341	4.4902	4276	202

R739x5_164-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
1.330	1.782	5.083	3.3011
1.388	0.870	3.208	2.3375
1.447	0.331	1.987	1.6560
1.450	2.217	4.275	2.0576
1.496	1.117	2.285	1.1681
1.505	0.028	1.203	1.1753
1.541	0.385	1.035	0.6502
1.587	-0.087	0.268	0.3553
1.623	-0.203	0.389	0.5915
1.677	-0.545	-0.440	0.1050

MAE	Std error	max error	#points	#count[err > 1]
1.340	0.951	3.3011	10	6

R739x5_164-164, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
1.498	1.673	3.554	1.8811
1.559	0.655	1.982	1.3272
1.621	0.062	0.991	0.9288
1.654	1.930	2.062	0.1321
1.682	-0.262	0.377	0.6389
1.703	1.069	0.974	-0.0952
1.752	0.485	0.283	-0.2019
1.800	0.098	-0.142	-0.2401
1.804	-0.481	-0.205	0.2757
1.877	2.856	1.329	-1.5271
1.898	-0.295	-0.523	-0.2276
1.917	2.019	0.430	-1.5890
1.956	1.392	-0.117	-1.5088
1.996	0.925	-0.432	-1.3571
2.075	0.333	-0.667	-0.9996

MAE	Std error	max error	#points	#count[err > 1]
0.862	0.614	1.8811	15	6

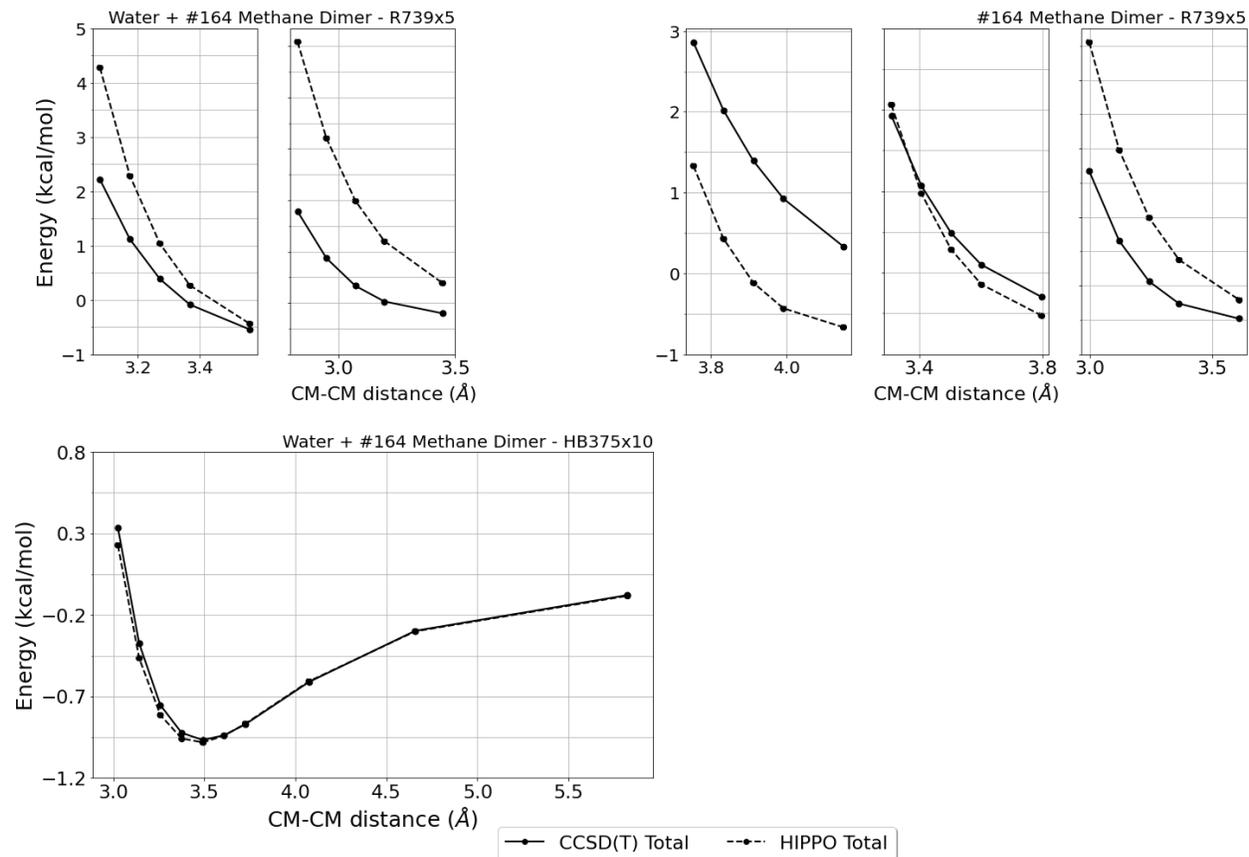
HB375x10_164-water, energy values in kcal/mol

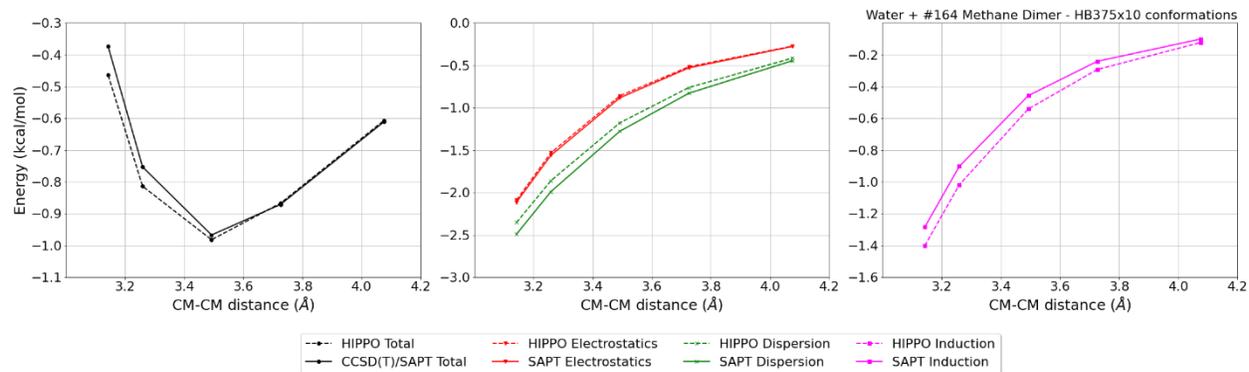
CM-CM (Å)	Reference	HIPPO res	Abs diff
1.426	0.336	0.228	-0.1076
1.480	-0.374	-0.464	-0.0903
1.535	-0.753	-0.814	-0.0608
1.590	-0.923	-0.958	-0.0345
1.645	-0.967	-0.982	-0.0150
1.700	-0.938	-0.941	-0.0029
1.755	-0.871	-0.867	0.0039
1.920	-0.611	-0.607	0.0043
2.194	-0.297	-0.301	-0.0036
2.743	-0.078	-0.082	-0.0045

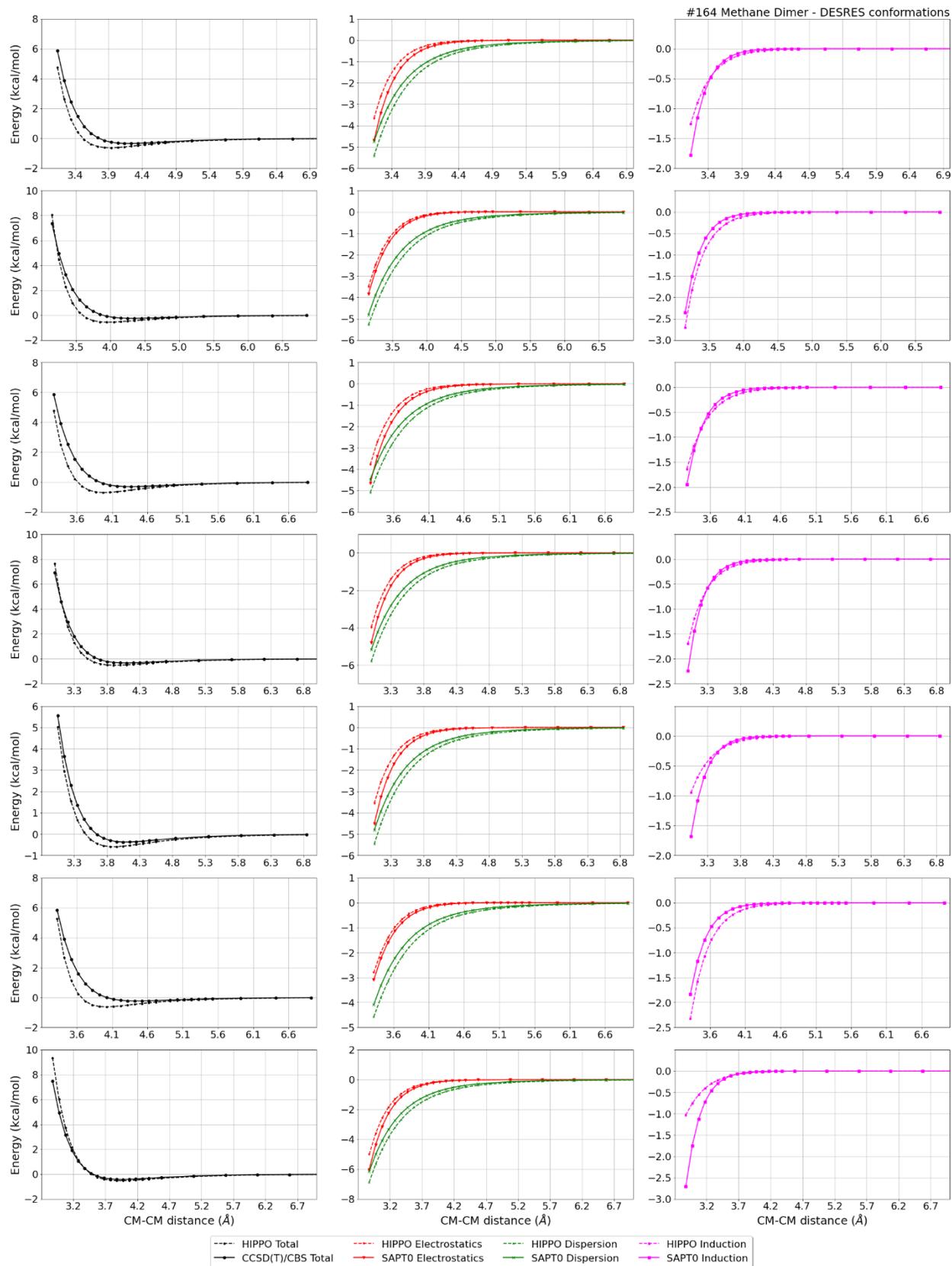
MAE	Std error	max error	#points	#count[err > 1]
0.033	0.038	0.1076	10	0

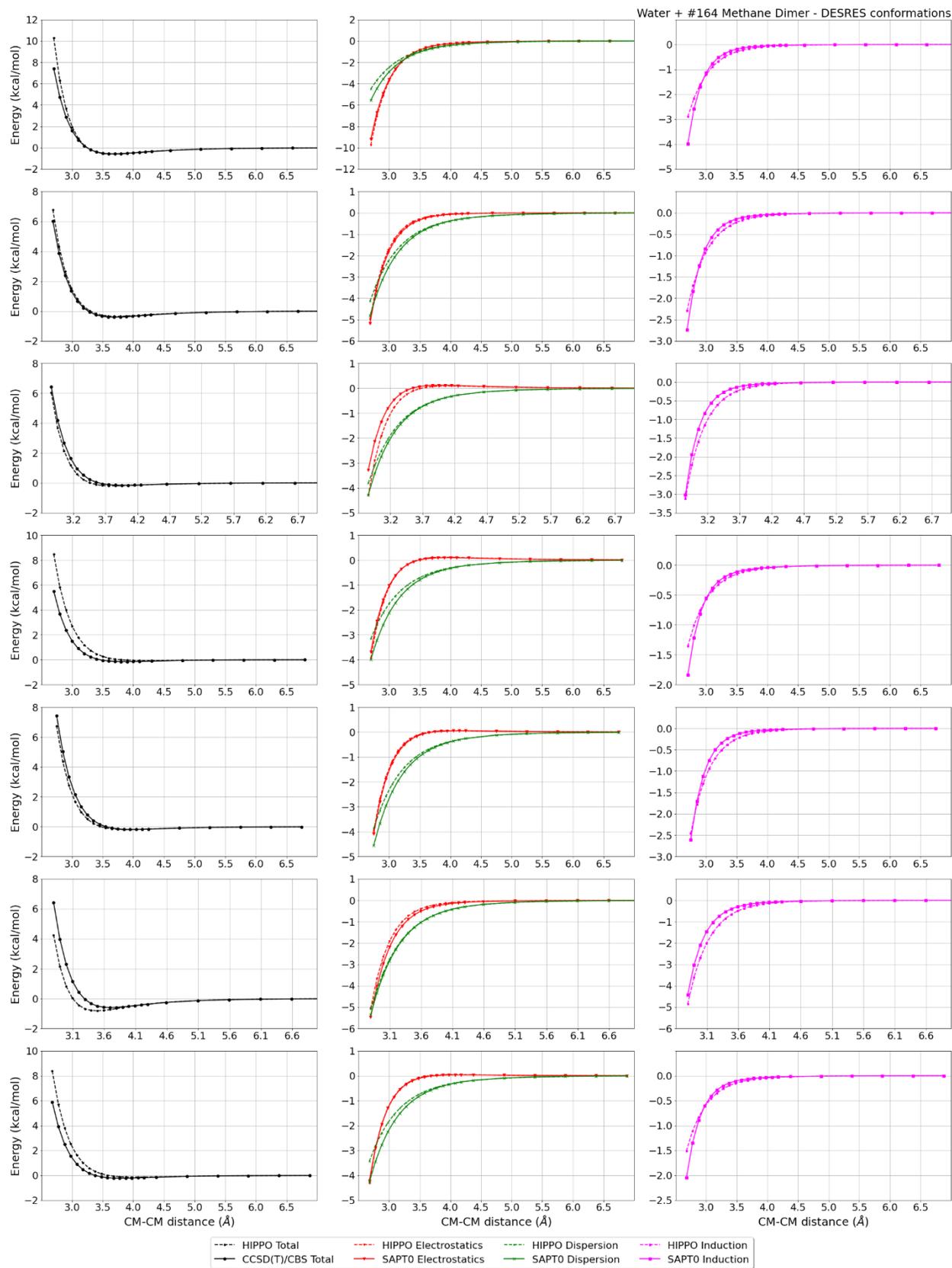
DESRES_164-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.318	0.650	6.5081	2655	282



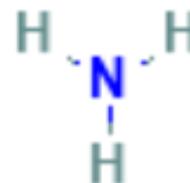






#165 Ammonia H3N CID: 222

ref molpol	-2.26	-2.01	-2.01, avg	-2.09
molpol	2.22	2.22	2.22, avg	2.22
rms molpol	0.04	0.21	0.21, avg	0.13



Monomer potential fitting RMS: 0.10

##Dimer results - Fitting to QM datasets##

DESRES_165-165, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.159	0.428	5.0302	4261	185

HB375x10_165-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
1.246	-2.138	-2.008	0.1299
1.294	-4.430	-4.907	-0.4767
1.342	-5.739	-6.417	-0.6783
1.389	-6.382	-7.056	-0.6741
1.437	-6.585	-7.161	-0.5756
1.484	-6.502	-6.949	-0.4474
1.532	-6.242	-6.562	-0.3205
1.674	-5.033	-5.082	-0.0493
1.912	-3.124	-3.043	0.0805
2.388	-1.222	-1.184	0.0380

MAE	Std error	max error	#points	#count[err > 1]
0.347	0.245	0.6783	10	0

DESRES_165-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.252	0.492	5.0265	2667	169

R739x5_165-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
1.764	2.207	2.344	0.1374
1.849	1.883	1.963	0.0801
1.934	1.635	1.682	0.0471
2.019	1.437	1.466	0.0286
2.188	1.138	1.147	0.0089

MAE	Std error	max error	#points	#count[err > 1]
0.060	0.045	0.1374	5	0

HB375x10_165-165, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
1.421	-0.097	-0.927	-0.8301
1.478	-1.694	-2.161	-0.4668
1.534	-2.585	-2.833	-0.2478
1.590	-3.016	-3.136	-0.1198

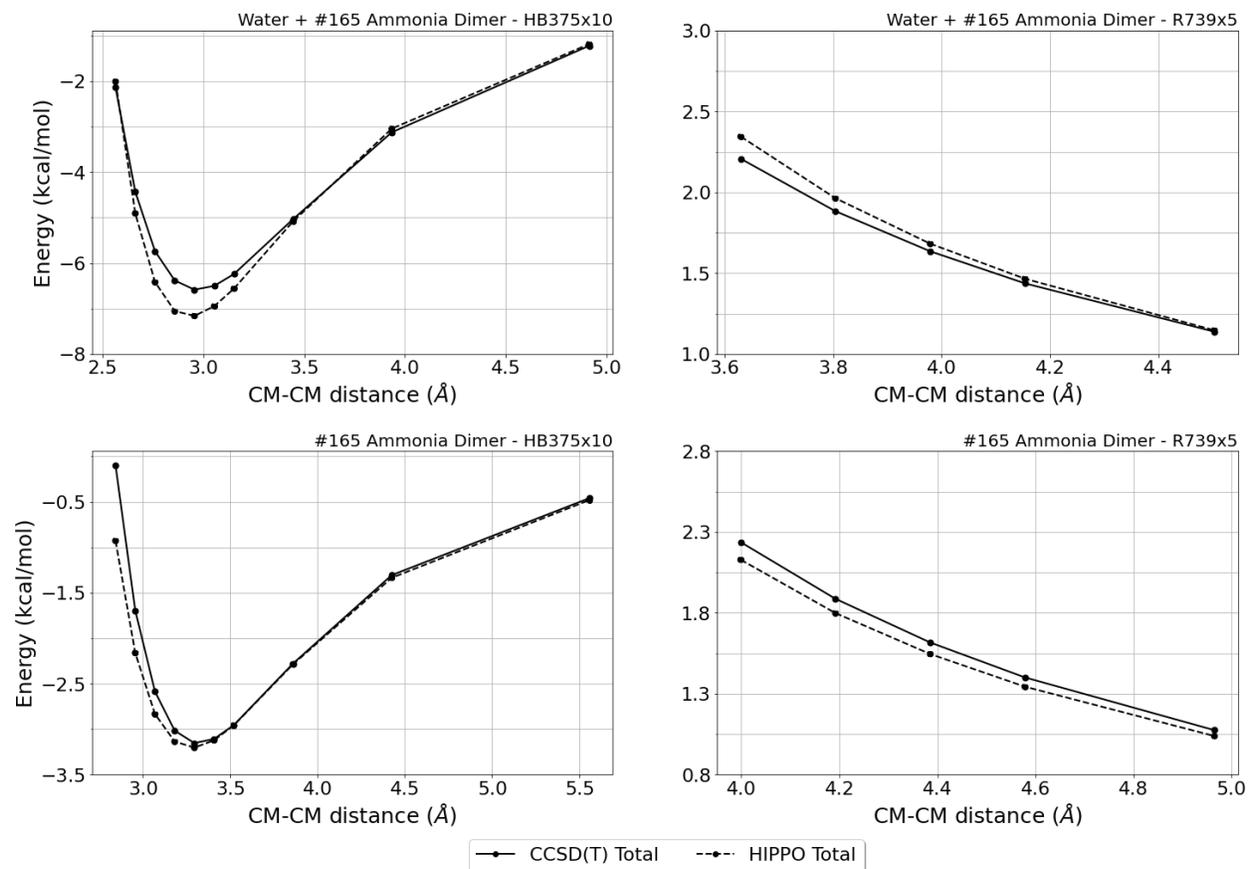
1.647	-3.153	-3.202	-0.0493
1.703	-3.107	-3.122	-0.0154
1.760	-2.955	-2.957	-0.0019
1.929	-2.278	-2.286	-0.0084
2.212	-1.306	-1.335	-0.0293
2.777	-0.460	-0.481	-0.0210

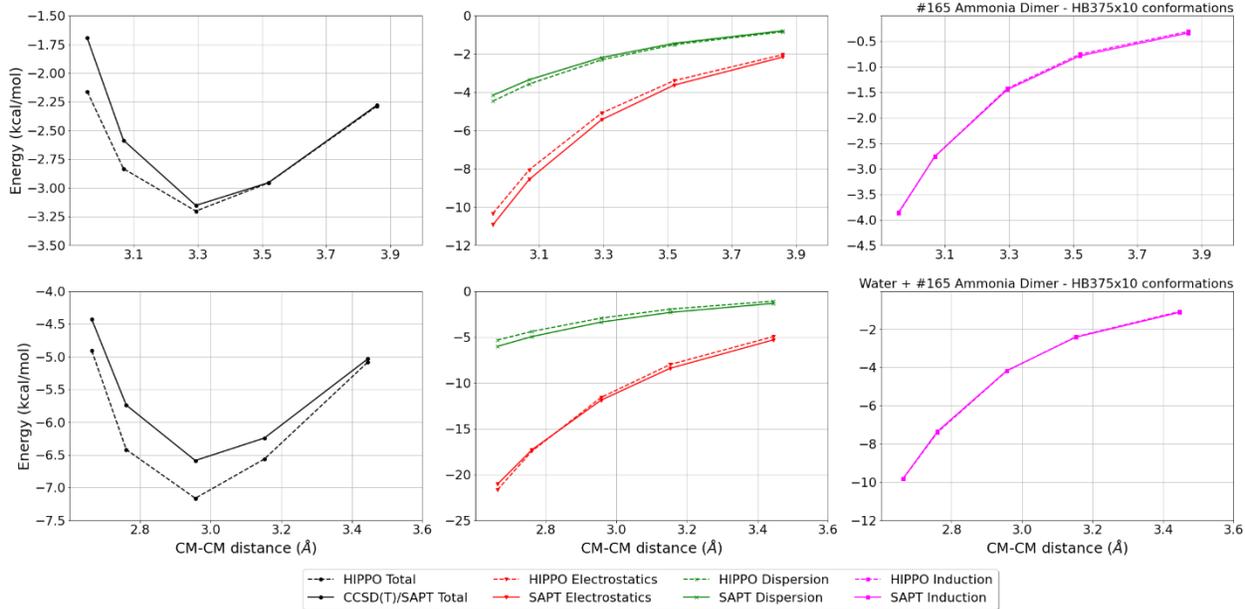
MAE	Std error	max error	#points	#count[err > 1]
0.179	0.258	0.8301	10	0

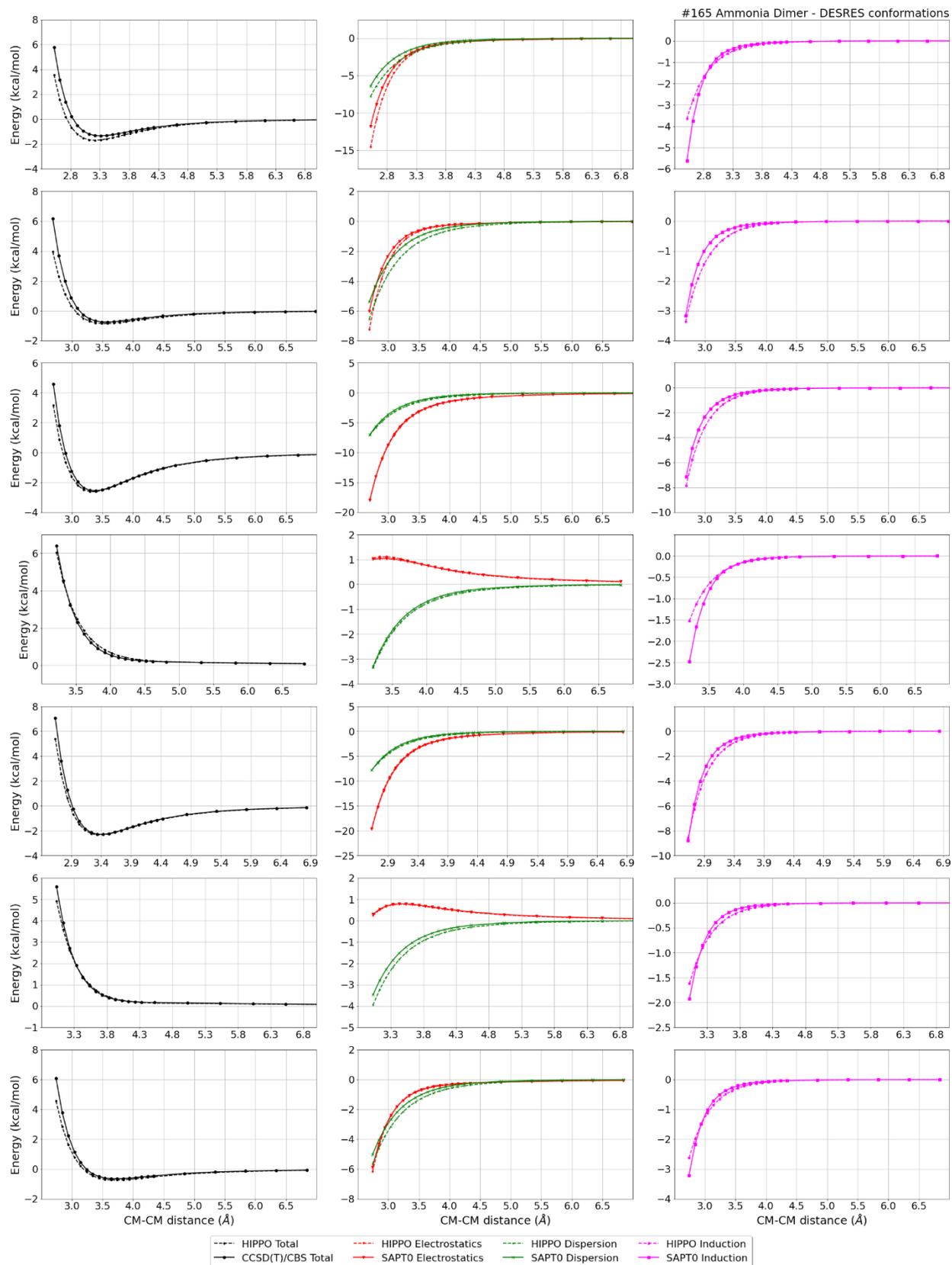
R739x5_165-165, energy values in kcal/mol

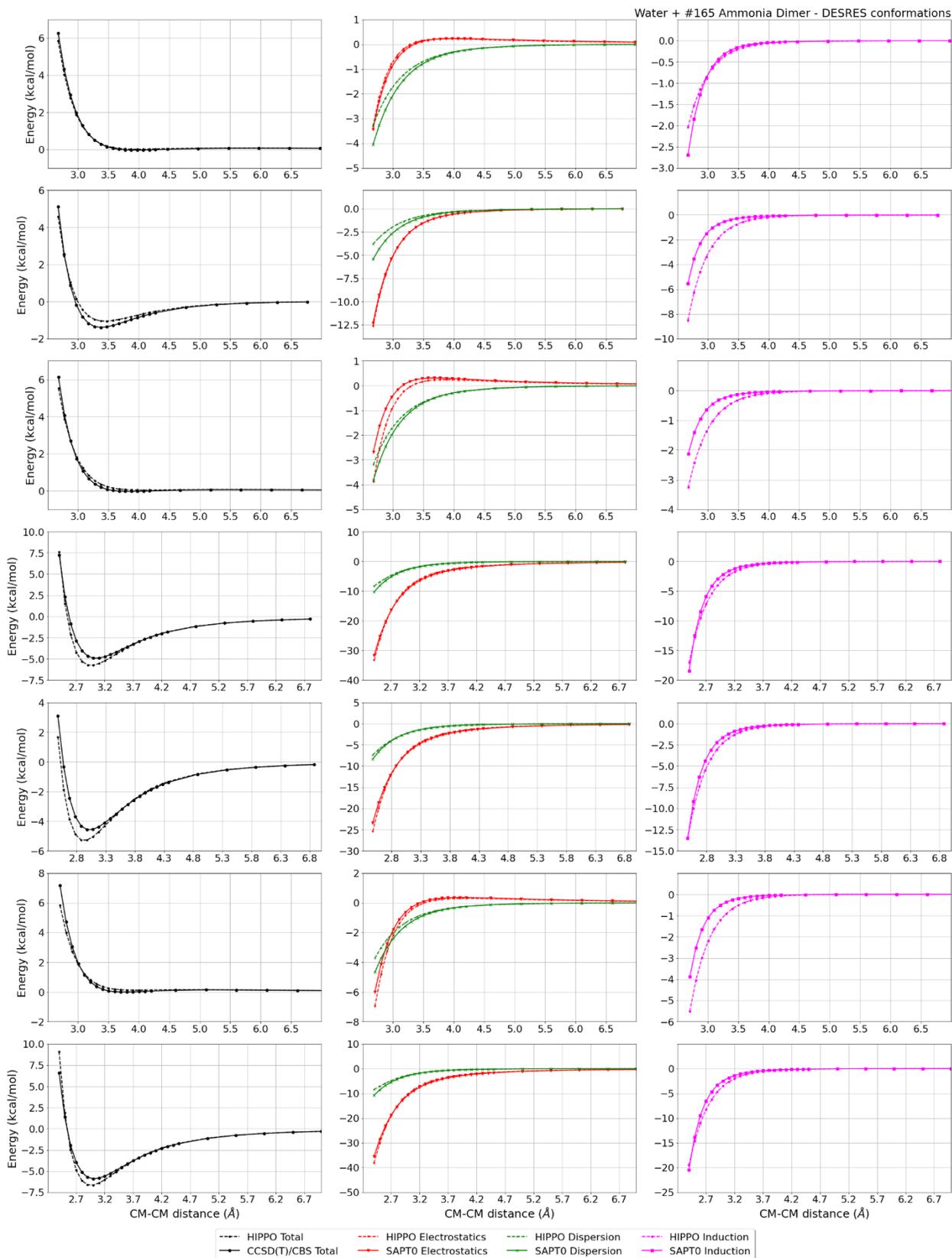
CM-CM (Å)	Reference	HIPPO res	Abs diff
1.999	2.237	2.128	-0.1086
2.096	1.887	1.799	-0.0884
2.193	1.617	1.546	-0.0713
2.289	1.401	1.344	-0.0572
2.483	1.075	1.039	-0.0363

MAE	Std error	max error	#points	#count[err > 1]
0.072	0.025	0.1086	5	0









#171 Sulfane H2S CID: 402



ref molpol	3.56	3.58	3.65, avg	3.60
molpol	3.30	3.46	3.53, avg	3.43
rms molpol	0.27	0.12	0.12, avg	0.17

Monomer potential fitting RMS: 0.23

##Dimer results - Fitting to QM datasets##

R739x5_171-171, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
1.666	2.256	1.919	-0.3368		
1.746	1.046	0.910	-0.1359		
1.826	0.383	0.373	-0.0097		
1.907	0.042	0.102	0.0597		
2.068	-0.169	-0.070	0.0986		
MAE	Std error	max error	#points	#count[err > 1]	
0.128	0.112	0.3368	5	0	

DESRES_171-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.191	0.528	10.1426	2773	107

DESRES_171-171, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.099	0.302	5.1787	4379	70

R739x5_171-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
2.012	2.109	2.298	0.1894		
2.108	1.226	1.337	0.1107		
2.204	0.727	0.792	0.0650		
2.301	0.455	0.489	0.0341		
2.493	0.245	0.240	-0.0048		
MAE	Std error	max error	#points	#count[err > 1]	
0.081	0.065	0.1894	5	0	

HB300SPXx10_171-171, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
1.748	1.467	3.993	2.5265
1.816	-0.070	1.356	1.4261
1.885	-0.948	-0.128	0.8200
1.953	-1.398	-0.908	0.4900
2.022	-1.580	-1.267	0.3125
2.090	-1.600	-1.383	0.2165
2.158	-1.526	-1.365	0.1612
2.364	-1.120	-1.033	0.0871
2.706	-0.563	-0.530	0.0333
3.390	-0.157	-0.155	0.0020

MAE	Std error	max error	#points	#count[err > 1]
0.608	0.763	2.5265	10	2

Liquid Sulfane @ 281.20 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
828.43	816.00	1.5	15.15	15.90	4.7	5.53	-1.00	-653.0	2000

171 failed to run analysis
Saving: True
241
Loading previous prmdict: newprms14.pickle

['dispersion', 'repulsion', 'chgtrn']
CCSD(T) DIMERS
SAPT DIMERS
#241 2-methoxyethanol C3H8O2 CID: 8019

ref molpol	-6.89	-8.80	0.00, avg	-5.23
molpol	6.88	5.41	5.22, avg	5.84
rms molpol	0.01	3.40	5.22, avg	0.61

Monomer potential fitting RMS: 0.58

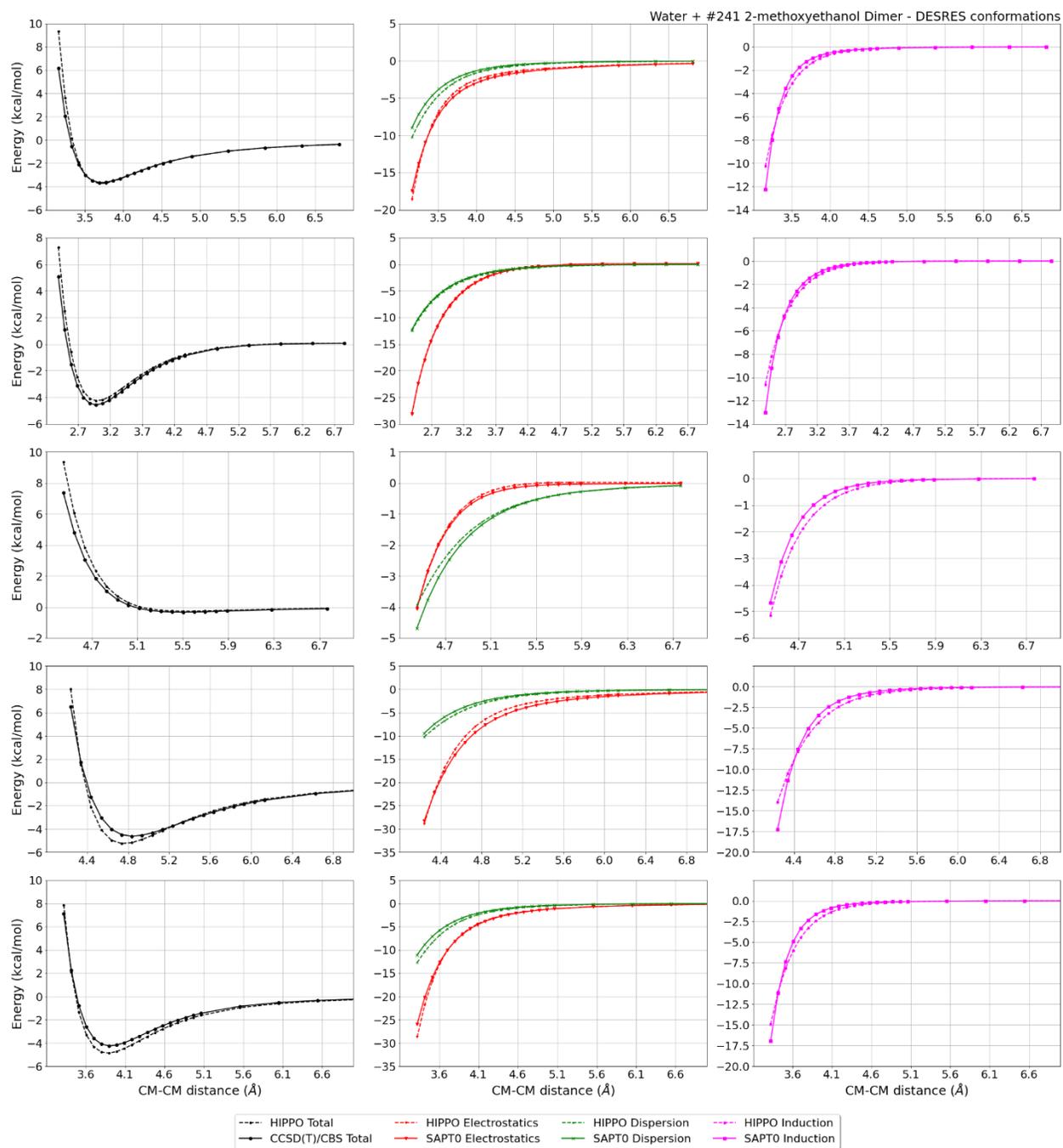
##Dimer results - Fitting to QM datasets##

DESRES_241-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.300	0.481	3.0579	121	8

Liquid 2-methoxyethanol @ 293.15 K

Density	Ref-Dens	%err	HV	Ref-HV	%err	Dielec	Ref-D	%err	#nFrm
840.99	964.70	12.8	45.14	40.85	10.5	7.74	-1.00	-874.4	1000



Results for molecules from Data set 2 (Appendix D, Table D.2)

#169 Hydrogen Bromide (HBr) CID: 260

ref molpol	3.69	3.42	3.42, avg	3.51
molpol	2.56	2.39	2.39, avg	2.45
rms molpol	1.10	1.03	1.03, avg	1.06



Monomer potential fitting RMS: 0.36

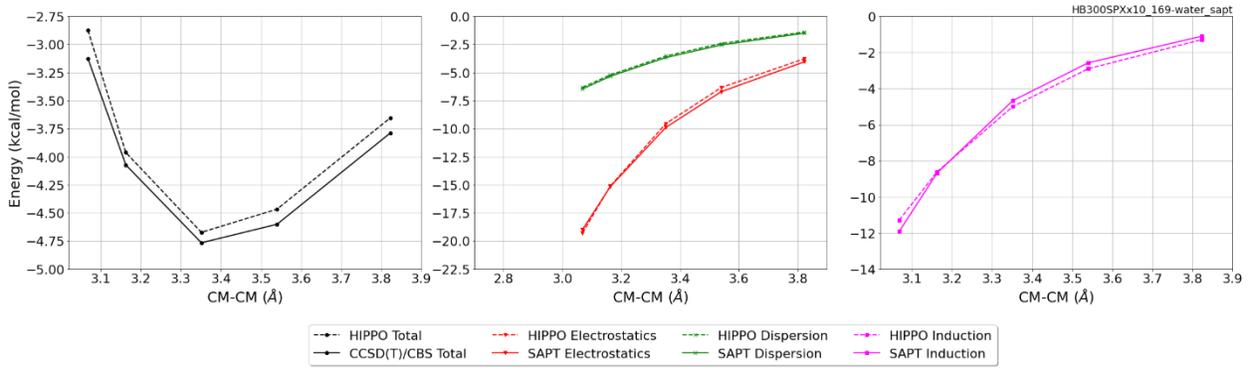
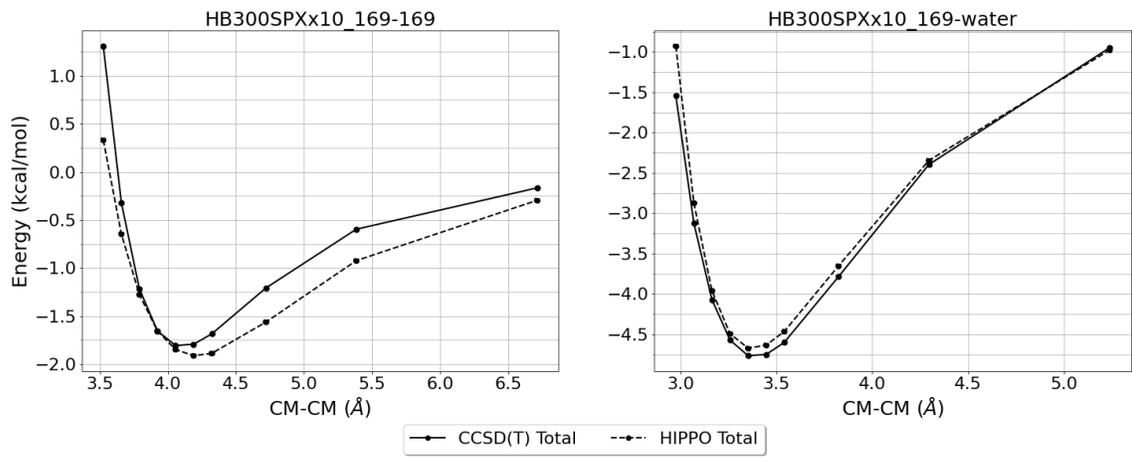
##Dimer results - Fitting to QM datasets##

HB300SPXx10_169-169, energy values in kcal/mol

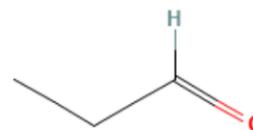
CM-CM (A)	Reference	HIPPO res	Abs diff		
1.763	1.309	0.332	-0.9771		
1.829	-0.320	-0.646	-0.3259		
1.896	-1.219	-1.278	-0.0595		
1.962	-1.656	-1.656	-0.0000		
2.028	-1.808	-1.849	-0.0413		
2.095	-1.794	-1.913	-0.1193		
2.161	-1.687	-1.890	-0.2025		
2.360	-1.209	-1.564	-0.3547		
2.692	-0.597	-0.924	-0.3271		
3.356	-0.167	-0.297	-0.1295		
MAE	Std error	max error	#points	#count[err > 1]	
0.254	0.270	0.9771	10	0	

HB300SPXx10_169-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff		
2.432	-1.541	-0.927	0.6137		
2.510	-3.125	-2.874	0.2506		
2.587	-4.072	-3.962	0.1101		
2.664	-4.573	-4.493	0.0795		
2.741	-4.766	-4.673	0.0929		
2.818	-4.751	-4.635	0.1161		
2.895	-4.601	-4.466	0.1347		
3.127	-3.787	-3.652	0.1346		
3.513	-2.396	-2.346	0.0504		
4.284	-0.948	-0.977	-0.0292		
MAE	Std error	max error	#points	#count[err > 1]	
0.161	0.161	0.6137	10	0	



#173 Propanal (C3H6O) CID: 527
ref molpol 5.95 7.36 4.98, avg 6.10
molpol 6.00 5.32 4.50, avg 5.27
rms molpol 0.01 2.03 0.48, avg 0.83



Monomer potential fitting RMS: 0.26

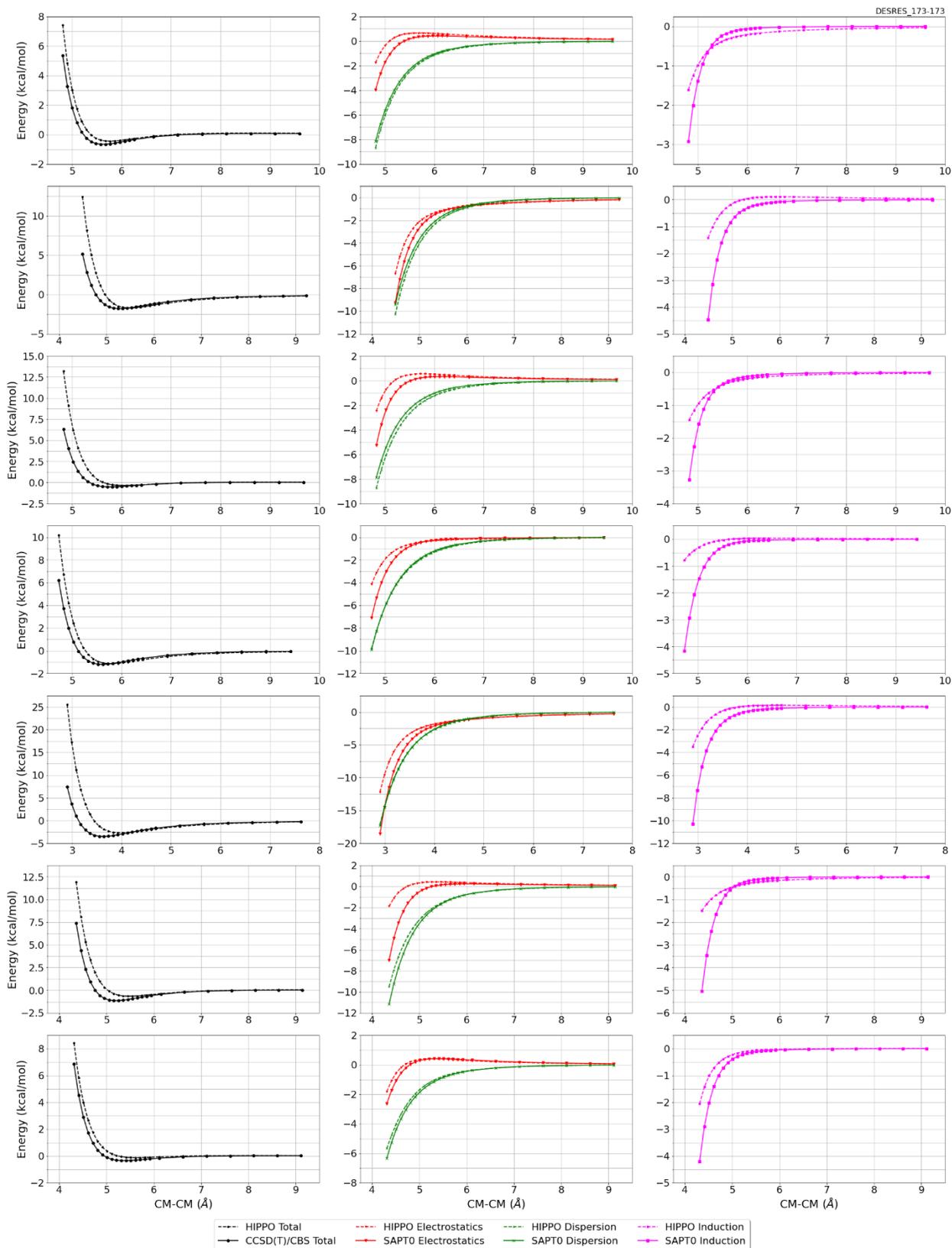
##Dimer results - Fitting to QM datasets##

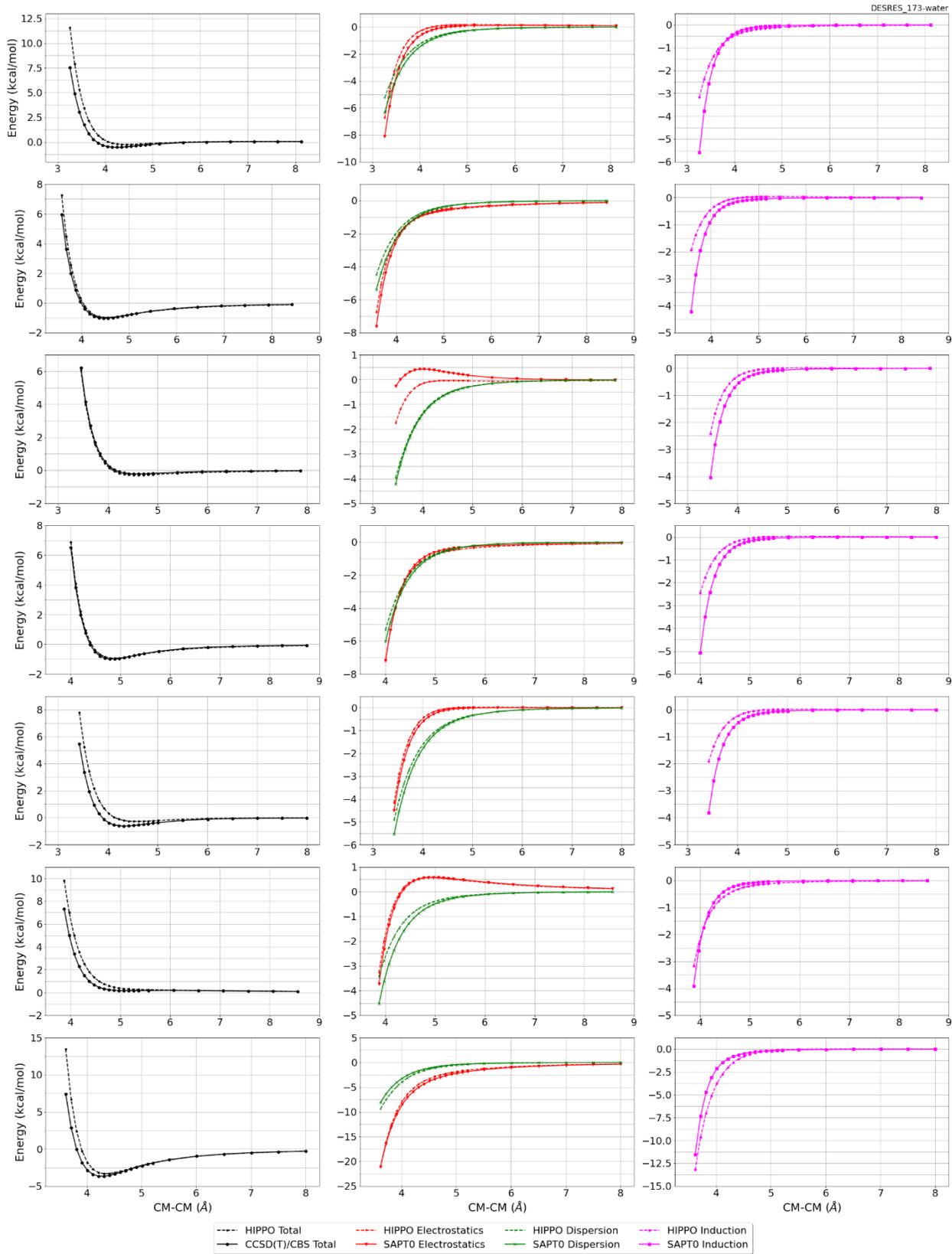
DESRES_173-173, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.555	1.516	17.6181	696	87

DESRES_173-water, energy values in kcal/mol

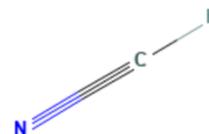
MAE	Std error	max error	#points	#count[err > 1]
0.319	0.565	5.7335	531	37





#178 Hydrogen Cyanide CID: 768

ref molpol	3.30	2.06	2.06, avg	2.48
molpol	2.08	1.73	1.73, avg	1.85
rms molpol	1.22	0.33	0.33, avg	0.63



Monomer potential fitting RMS: 0.09

##Dimer results - Fitting to QM datasets##

R739x5_178-178, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.118	0.063	0.1930	5	0

DESRES_178-178, energy values in kcal/mol

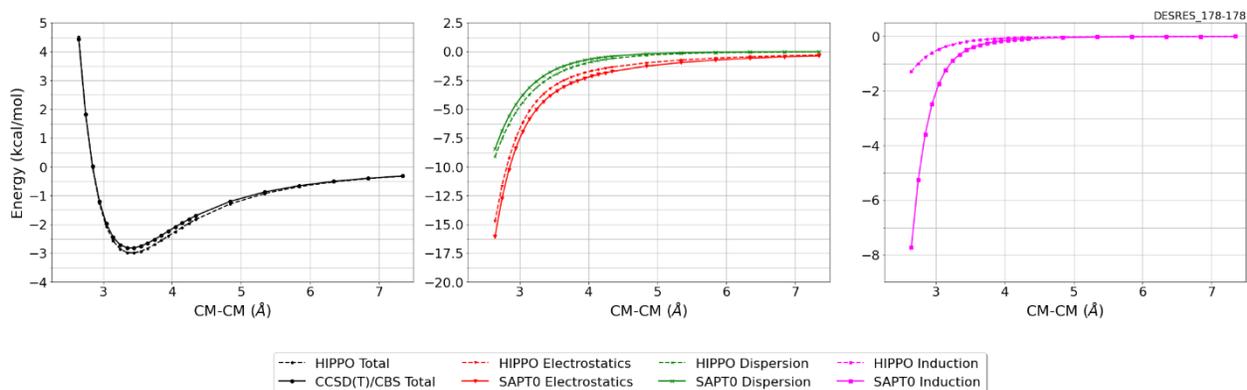
MAE	Std error	max error	#points	#count[err > 1]
0.108	0.063	0.1807	24	0

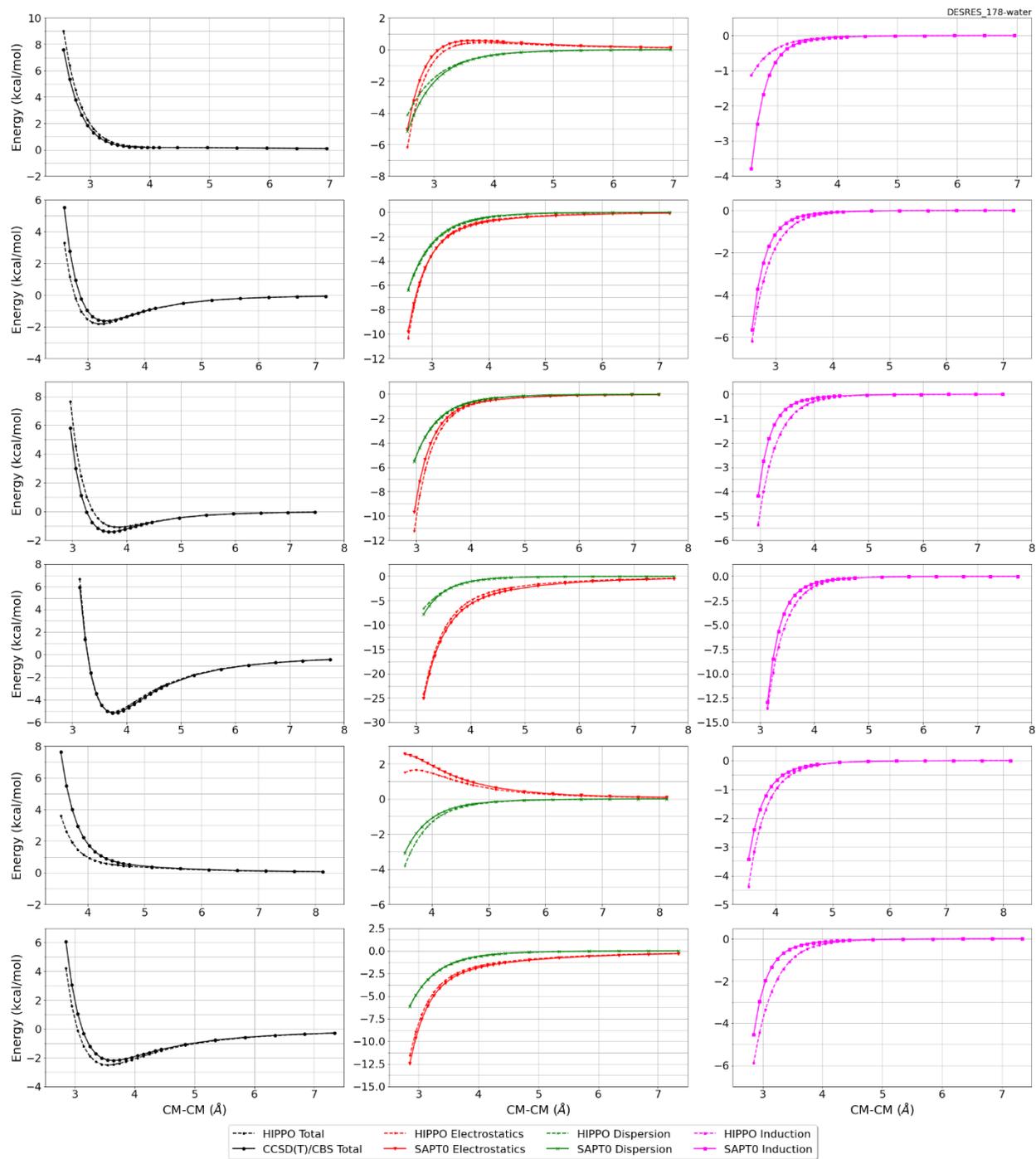
DESRES_178-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.210	0.460	4.0144	284	17

R739x5_178-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
1.790	0.369	2.3515	5	5





#181 Methanethiol (CH4S) CID: 878

ref molpol	4.93	6.25	4.93, avg	5.37
molpol	4.66	4.95	4.66, avg	4.76
rms molpol	0.27	1.30	0.27, avg	0.61



Monomer potential fitting RMS: 0.47

##Dimer results - Fitting to QM datasets##

R739x5_181-181, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
2.077	2.346	0.718	-1.6283
2.157	1.185	0.171	-1.0141
2.237	0.546	-0.085	-0.6307
2.316	0.215	-0.183	-0.3975
2.476	-0.001	-0.176	-0.1754

MAE	Std error	max error	#points	#count[err > 1]
0.769	0.511	1.6283	5	2

DESRES_181-181, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.147	0.299	3.6706	1265	29

R739x5_181-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
2.572	2.168	2.283	0.1149
2.681	1.450	1.573	0.1235
2.791	1.037	1.152	0.1153
2.901	0.800	0.900	0.0996
3.120	0.585	0.646	0.0607

MAE	Std error	max error	#points	#count[err > 1]
0.103	0.022	0.1235	5	0

DESRES_181-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.181	0.346	2.6450	625	27

HB300SPXx10_181-181, energy values in kcal/mol

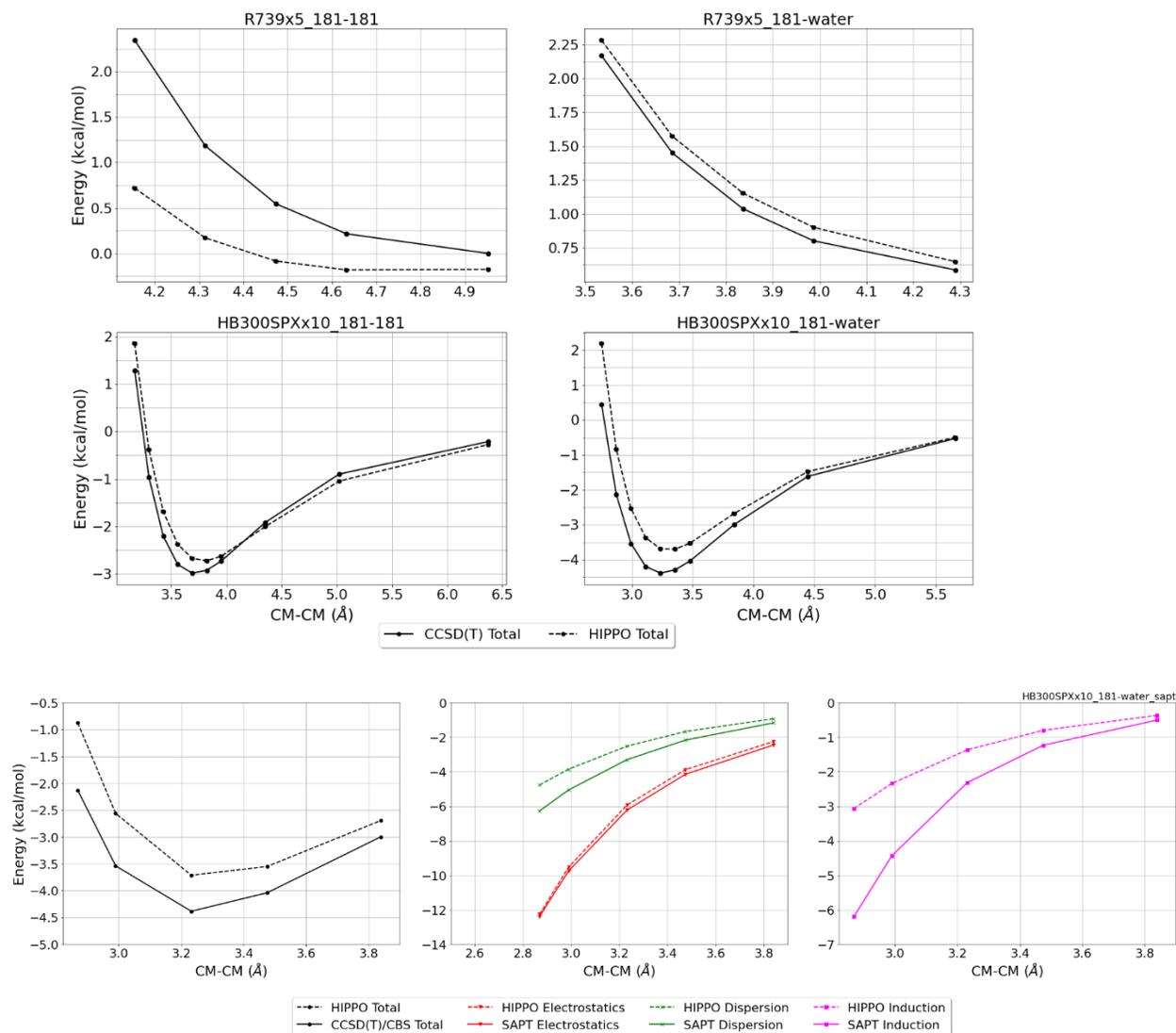
CM-CM (A)	Reference	HIPPO res	Abs diff
1.583	1.289	1.855	0.5656
1.648	-0.961	-0.383	0.5775
1.713	-2.206	-1.686	0.5205
1.778	-2.800	-2.376	0.4244
1.844	-2.984	-2.673	0.3114
1.910	-2.925	-2.724	0.2010
1.976	-2.731	-2.628	0.1026
2.175	-1.917	-2.006	-0.0886
2.510	-0.899	-1.046	-0.1470
3.187	-0.213	-0.275	-0.0623

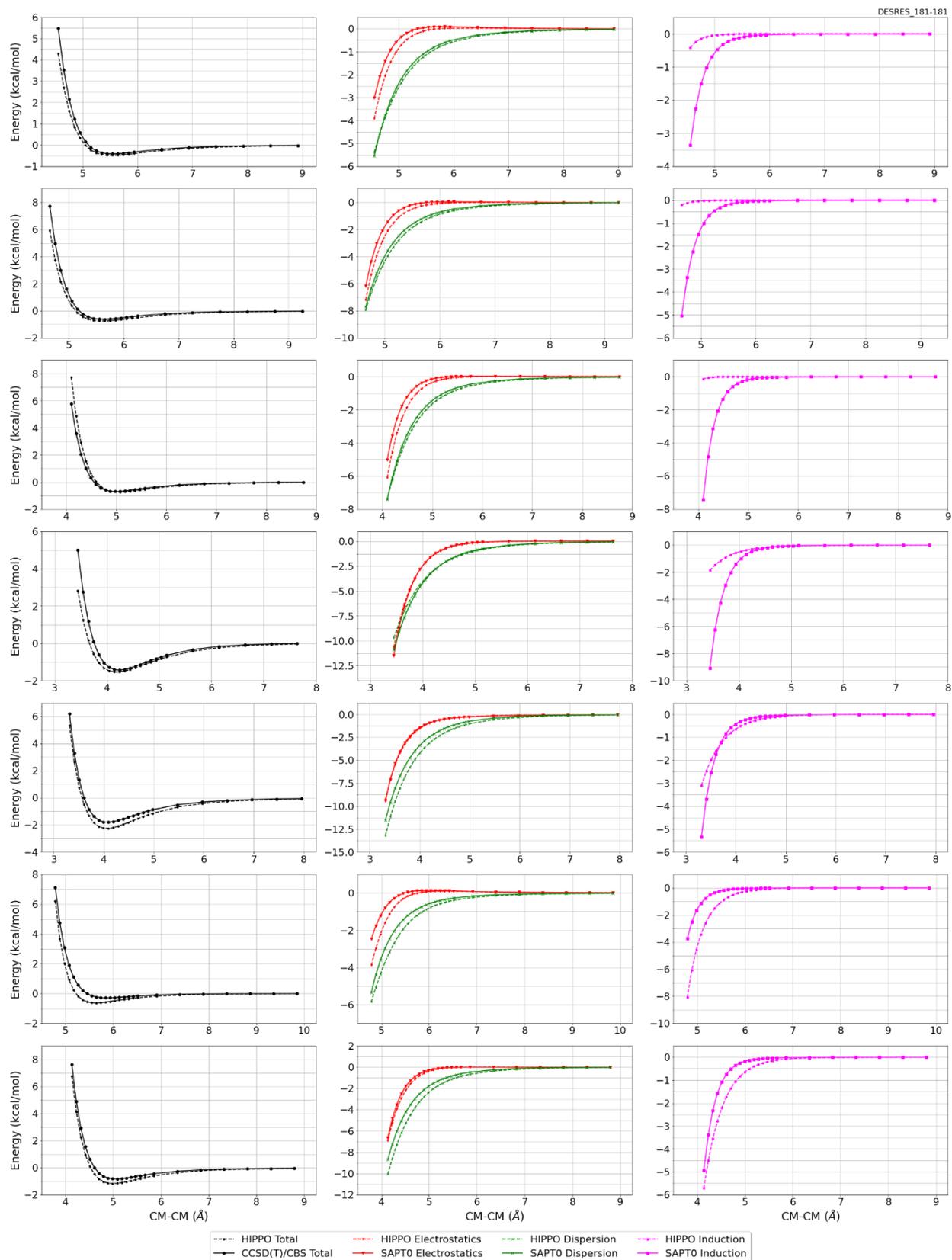
MAE	Std error	max error	#points	#count[err > 1]
0.300	0.196	0.5775	10	0

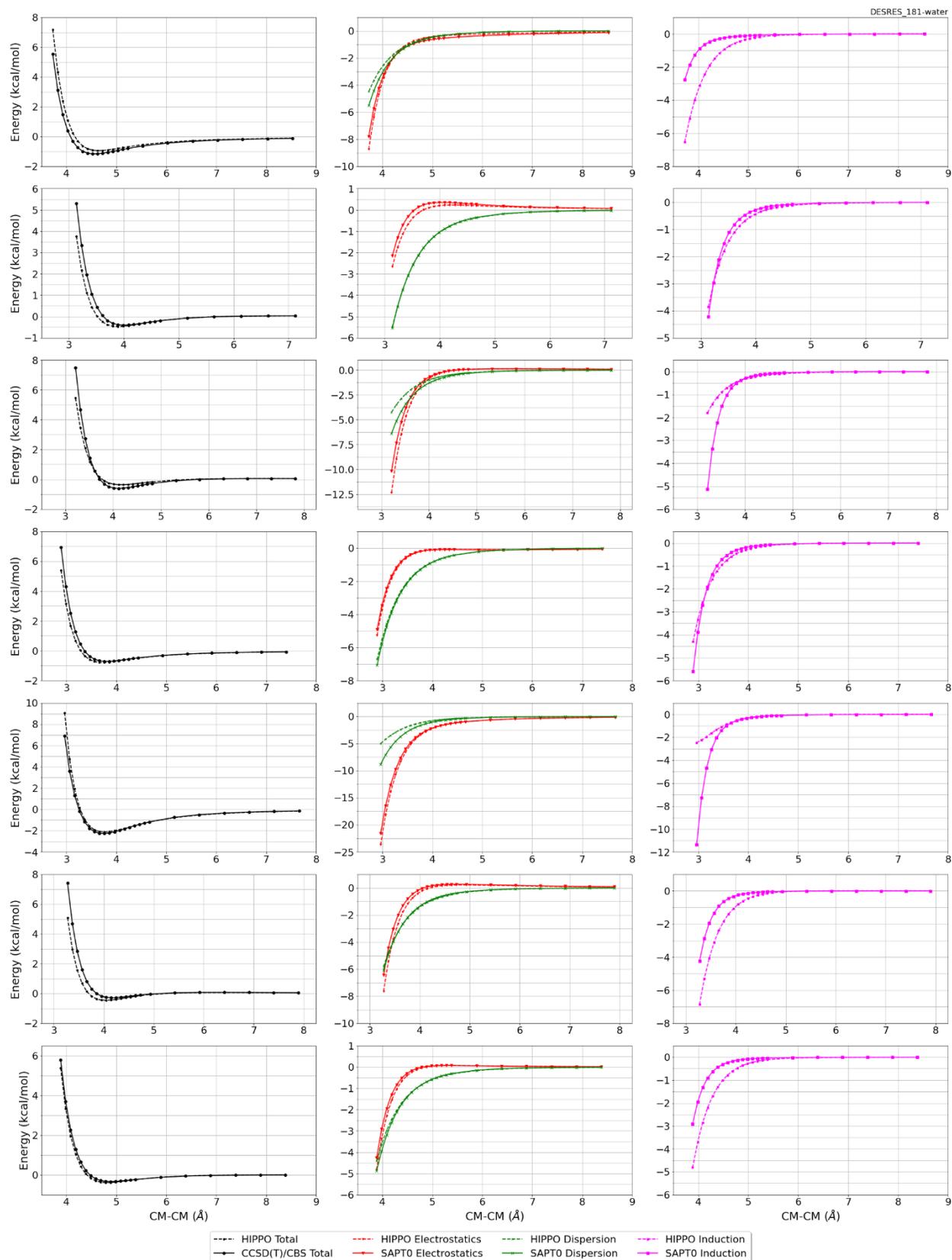
HB300SPXx10_181-water, energy values in kcal/mol

CM-CM (Å)	Reference	HIPPO res	Abs diff
1.999	0.446	2.189	1.7425
2.087	-2.131	-0.844	1.2874
2.175	-3.539	-2.531	1.0079
2.264	-4.197	-3.371	0.8256
2.352	-4.385	-3.690	0.6949
2.440	-4.292	-3.698	0.5936
2.528	-4.041	-3.532	0.5095
2.793	-2.999	-2.681	0.3182
3.234	-1.617	-1.480	0.1372
4.117	-0.531	-0.501	0.0302

MAE	Std error	max error	#points	#count[err > 1]
0.715	0.500	1.7425	10	3







#190 2-Sulfanylethanol (C2H6OS) CID: 1567

ref molpol	9.52	7.21	6.81, avg	7.85
molpol	9.55	7.09	6.48, avg	7.70
rms molpol	0.03	0.12	0.33, avg	0.15

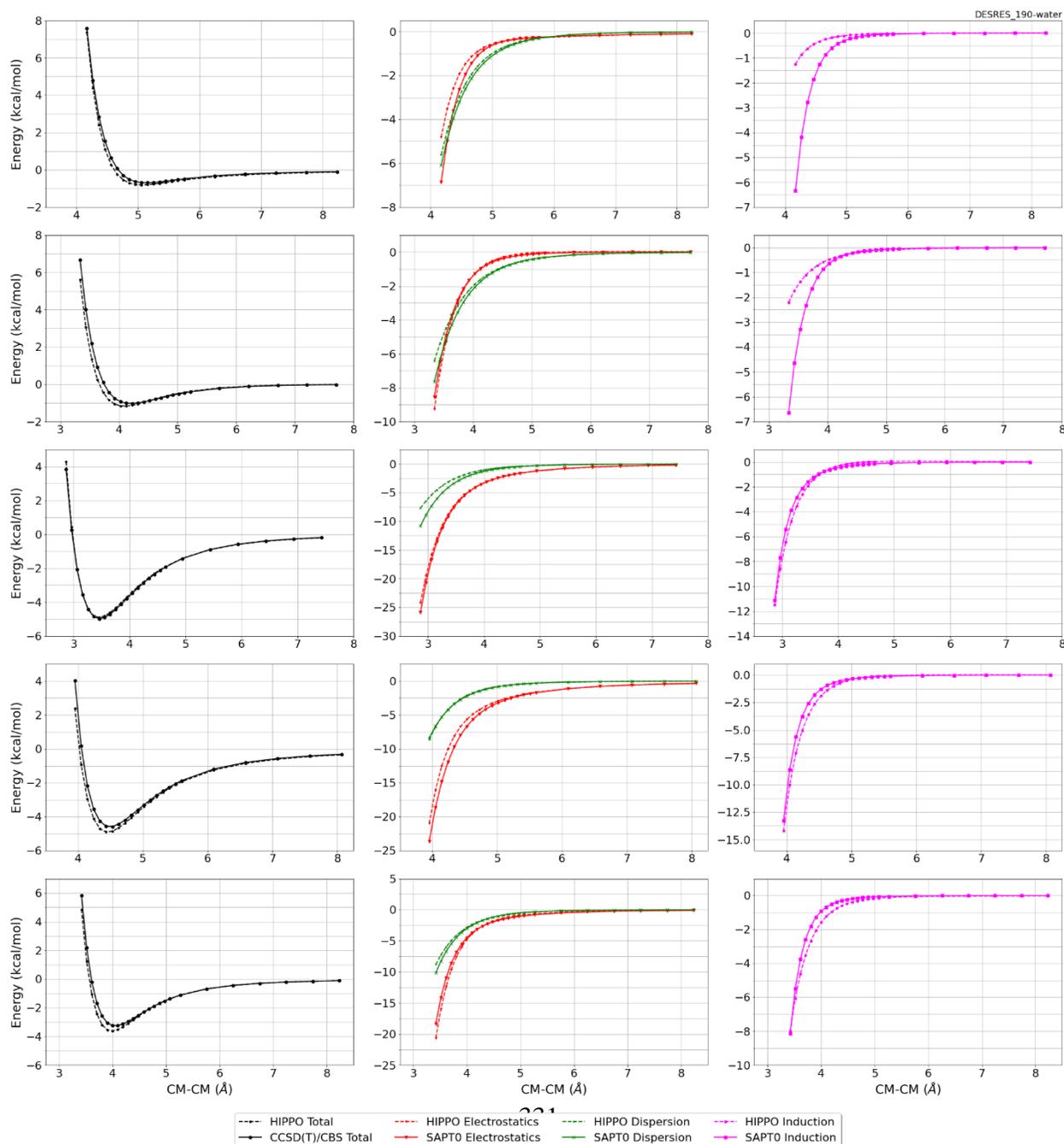


Monomer potential fitting RMS: 0.23

##Dimer results - Fitting to QM datasets##

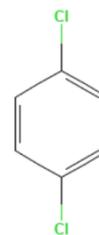
DESRES_190-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.174	0.236	1.8450	270	2



#194 1,4-Dichlorobenzene CID: 4685

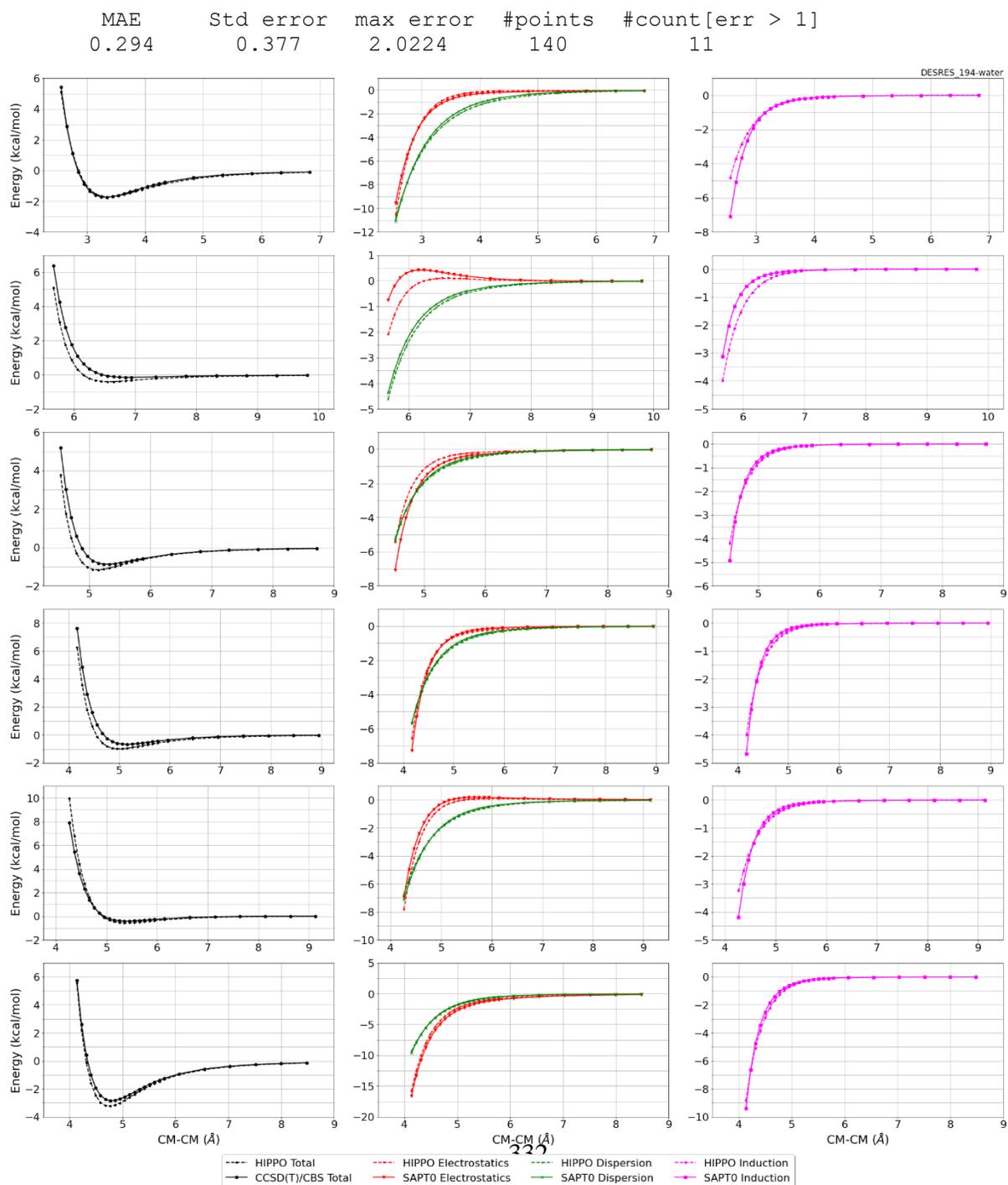
ref molpol	13.88	20.91	0.00, avg	11.60
molpol	14.05	12.43	7.66, avg	11.38
rms molpol	0.17	8.48	7.66, avg	0.22



Monomer potential fitting RMS: 0.15

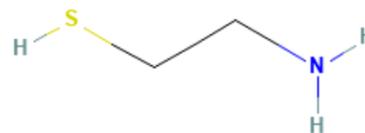
##Dimer results - Fitting to QM datasets##

DESRES_194-water, energy values in kcal/mol



#196 2-Aminoethanethiol CID: 6058

ref molpol	10.48	7.82	7.46, avg	8.59
molpol	10.53	7.84	7.17, avg	8.51
rms molpol	0.05	0.02	0.29, avg	0.08

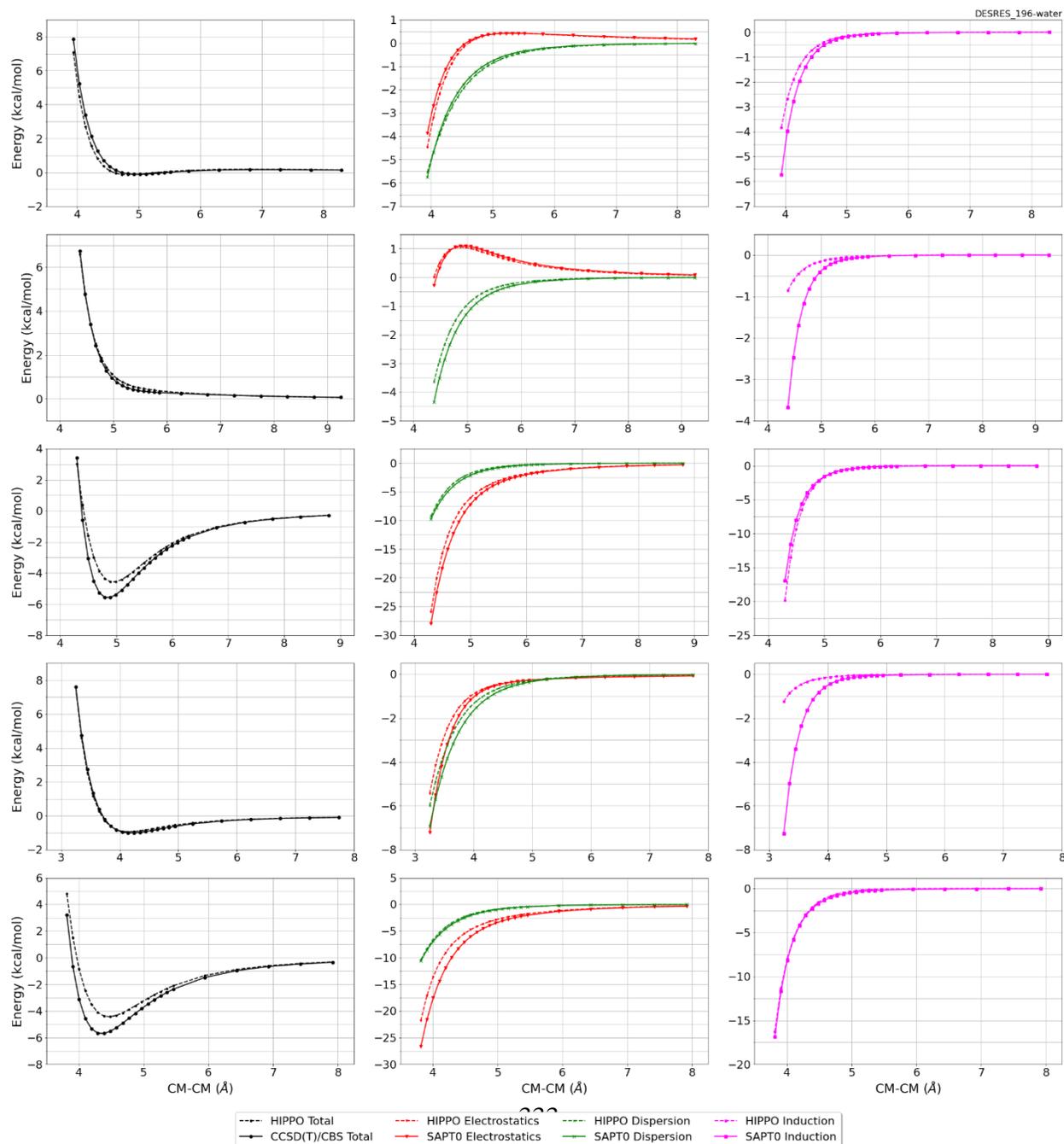


Monomer potential fitting RMS: 0.31

##Dimer results - Fitting to QM datasets##

DESRES_196-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.312	0.460	2.1357	119	11



#200 Ethane CID: 6324

ref molpol	3.96	3.93	4.50, avg	4.14
molpol	3.02	3.02	3.24, avg	3.09
rms molpol	0.94	0.94	1.26, avg	1.05

Monomer potential fitting RMS: 0.11

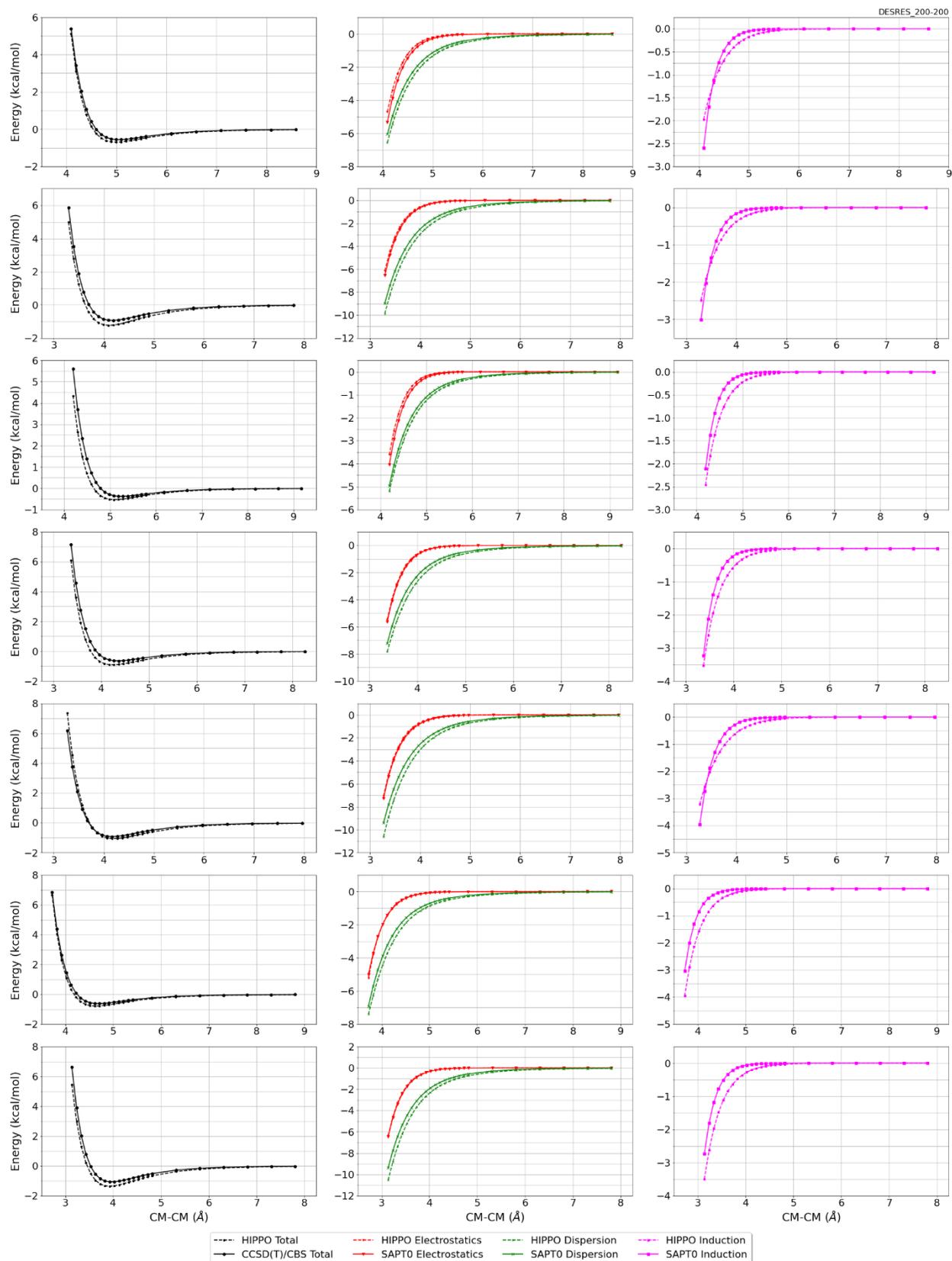
##Dimer results - Fitting to QM datasets##

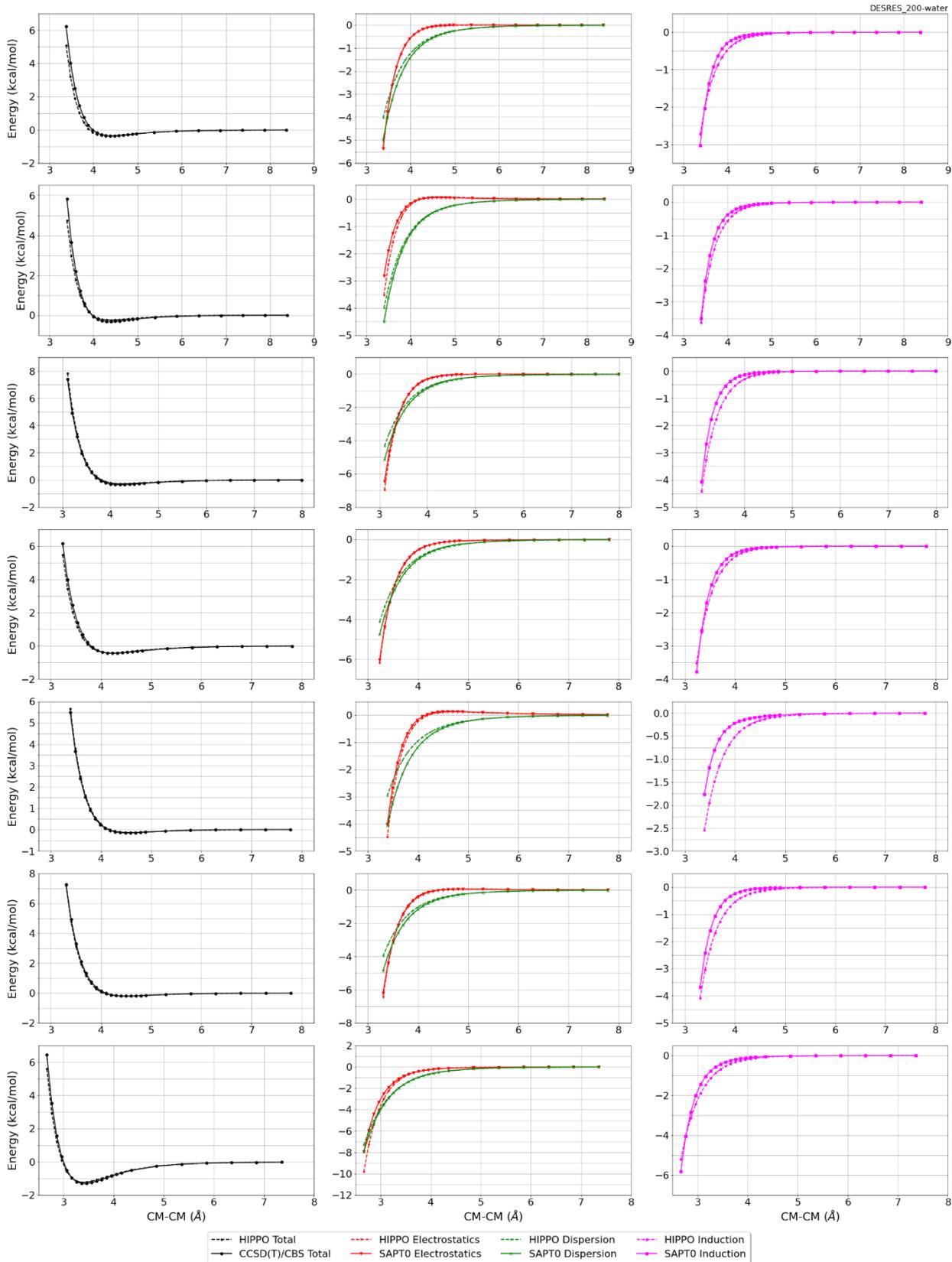
DESRES_200-200, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.107	0.174	1.9625	1171	9

DESRES_200-water, energy values in kcal/mol

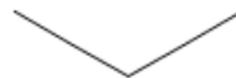
MAE	Std error	max error	#points	#count[err > 1]
0.090	0.214	2.5847	557	7





#203 Propane CID: 6334

ref molpol	5.77	6.45	5.45, avg	5.89
molpol	5.18	5.81	4.94, avg	5.31
rms molpol	0.59	0.64	0.51, avg	0.58



Monomer potential fitting RMS: 0.22

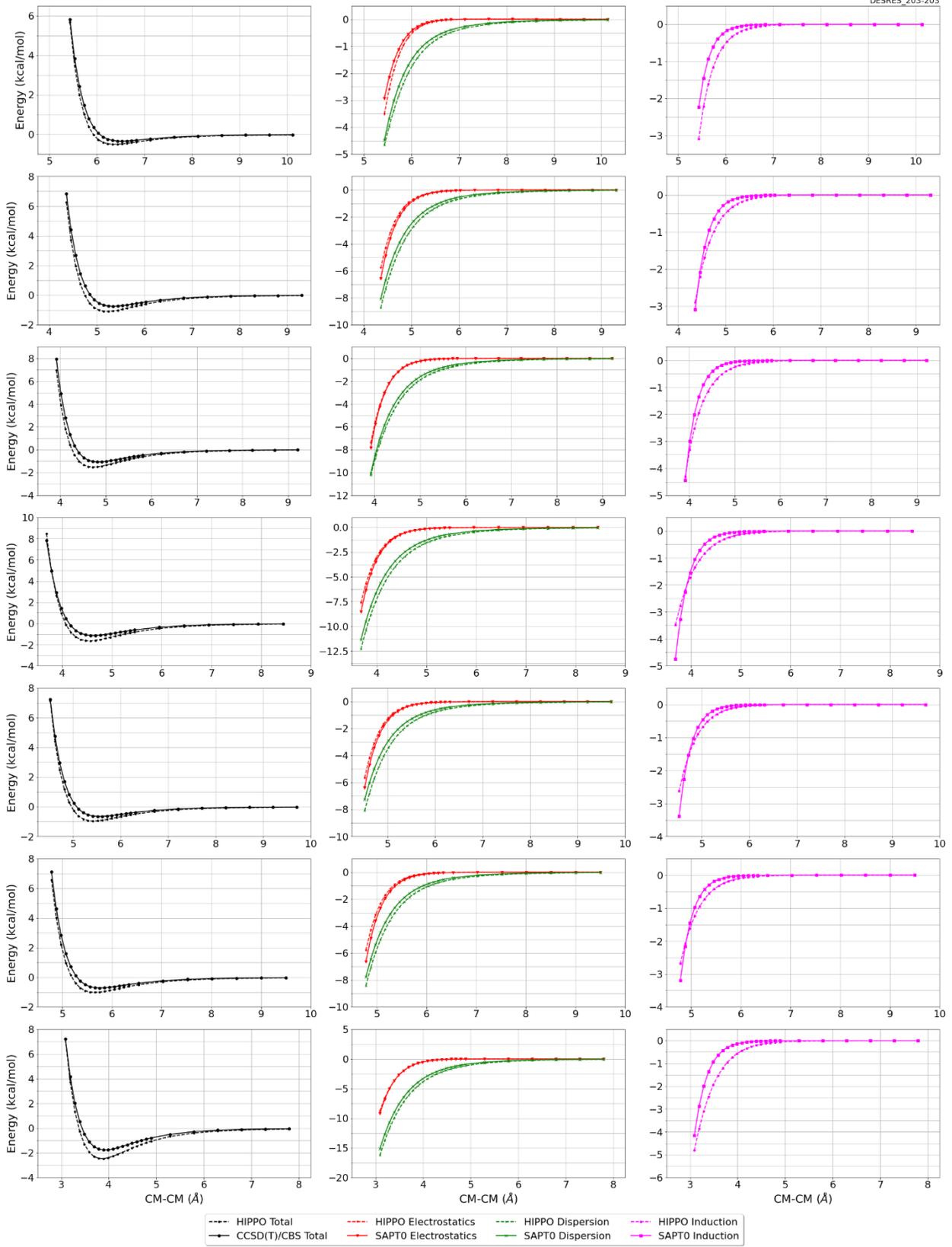
##Dimer results - Fitting to QM datasets##

DESRES_203-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.133	0.273	2.0186	558	16

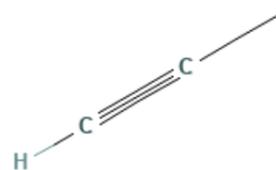
DESRES_203-203, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.229	0.267	1.5365	725	13



#204 Propyne CID: 6335

ref molpol	7.25	4.21	4.21, avg	5.22
molpol	4.25	3.23	3.23, avg	3.57
rms molpol	3.00	0.97	0.97, avg	1.65



Monomer potential fitting RMS: 0.13

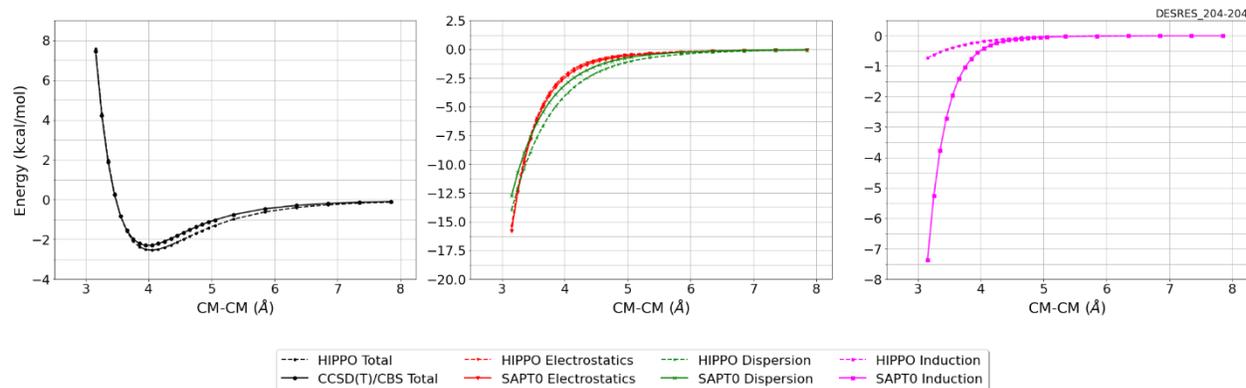
##Dimer results - Fitting to QM datasets##

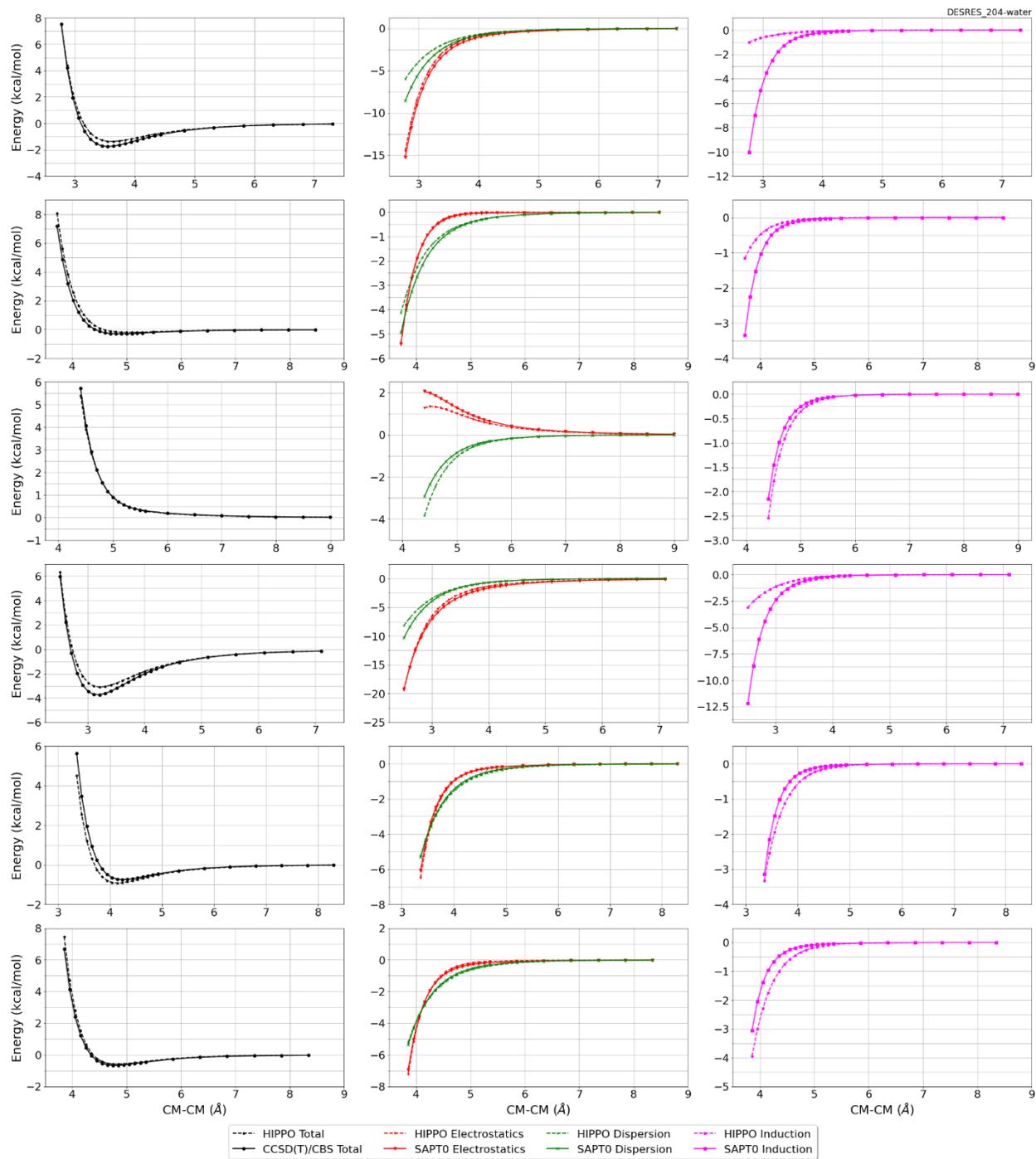
DESRES_204-204, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.190	0.113	0.3343	26	0

DESRES_204-water, energy values in kcal/mol

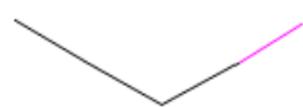
MAE	Std error	max error	#points	#count[err > 1]
0.142	0.200	1.1262	289	1





#205 Iodoethane CID: 6340

ref molpol	11.20	8.21	7.82, avg	9.08
molpol	10.97	8.35	7.97, avg	9.10
rms molpol	0.24	0.14	0.15, avg	0.02

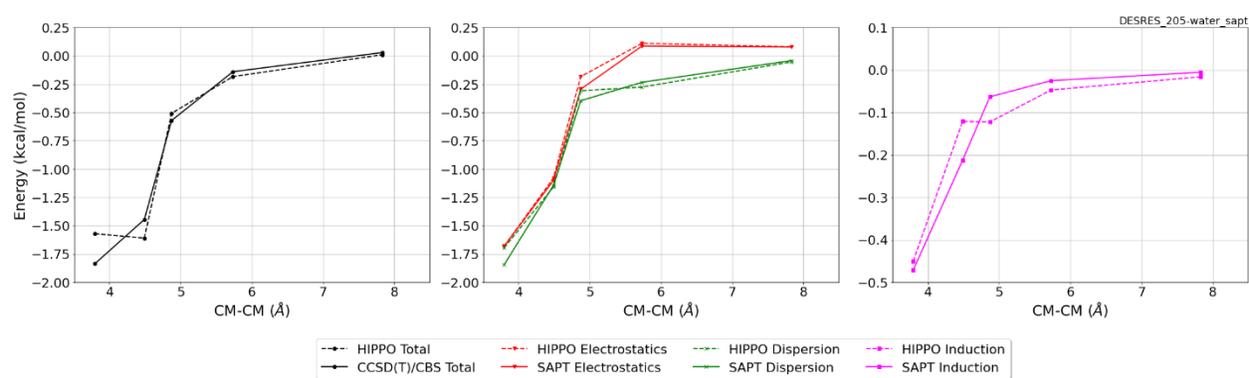


Monomer potential fitting RMS: 0.22

##Dimer results - Fitting to QM datasets##

DESRES_205-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.108	0.160	1.0900	150	1



#207 Ethanethiol CID: 6343

ref molpol	8.40	6.69	6.38, avg	7.16
molpol	8.42	6.61	6.06, avg	7.03
rms molpol	0.02	0.08	6.32, avg	0.13



Monomer potential fitting RMS: 0.11

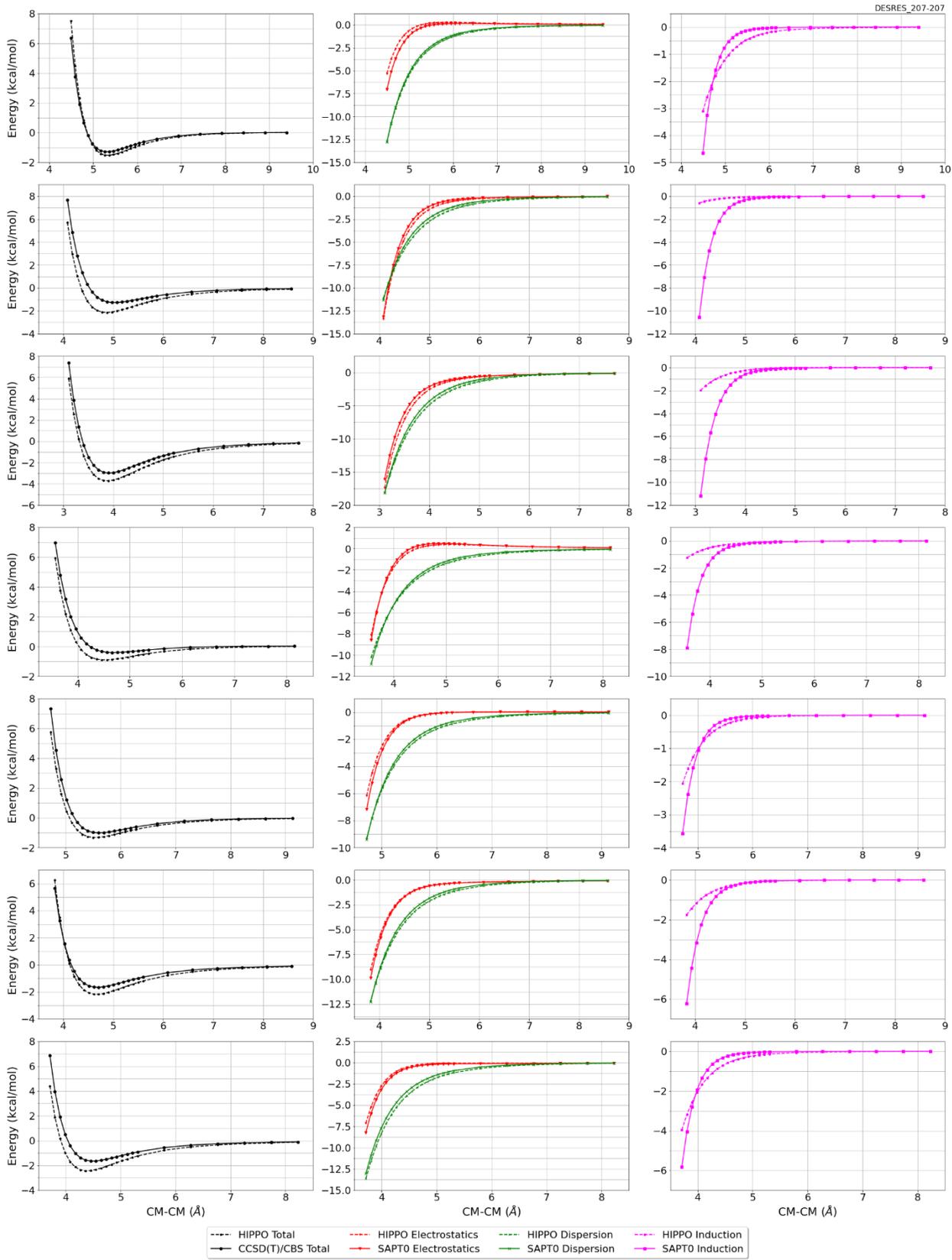
##Dimer results - Fitting to QM datasets##

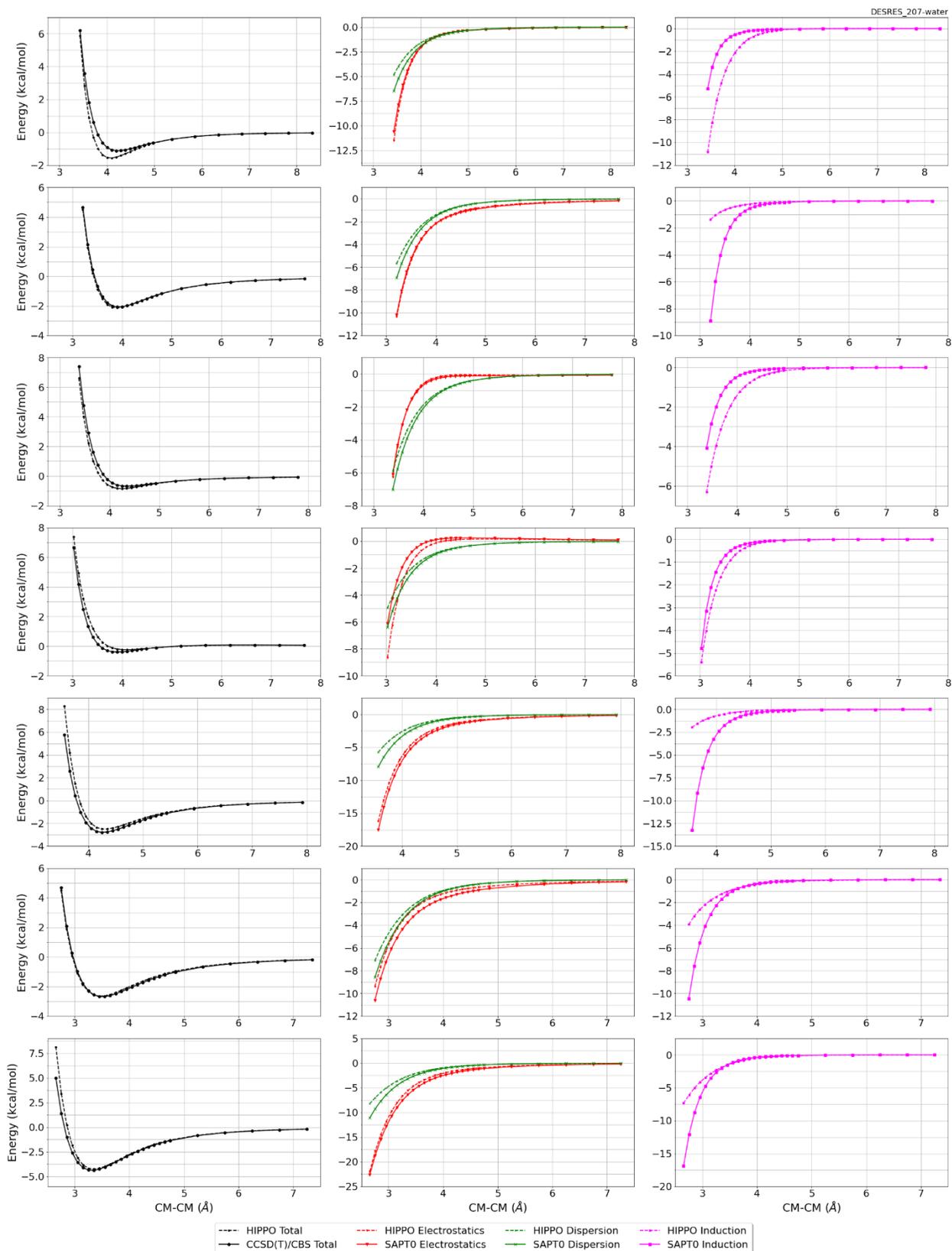
DESRES_207-207, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.286	0.360	2.3349	723	39

DESRES_207-water, energy values in kcal/mol

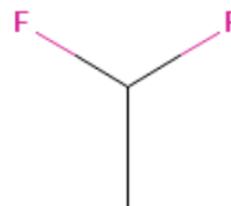
MAE	Std error	max error	#points	#count[err > 1]
0.204	0.361	3.3356	564	21





#209 1,1-Difluoroethane CID: 6368

ref molpol	4.09	4.30	4.47, avg	4.29
molpol	4.09	4.23	4.53, avg	4.28
rms molpol	0.00	0.07	0.06, avg	0.00



Monomer potential fitting RMS: 0.45

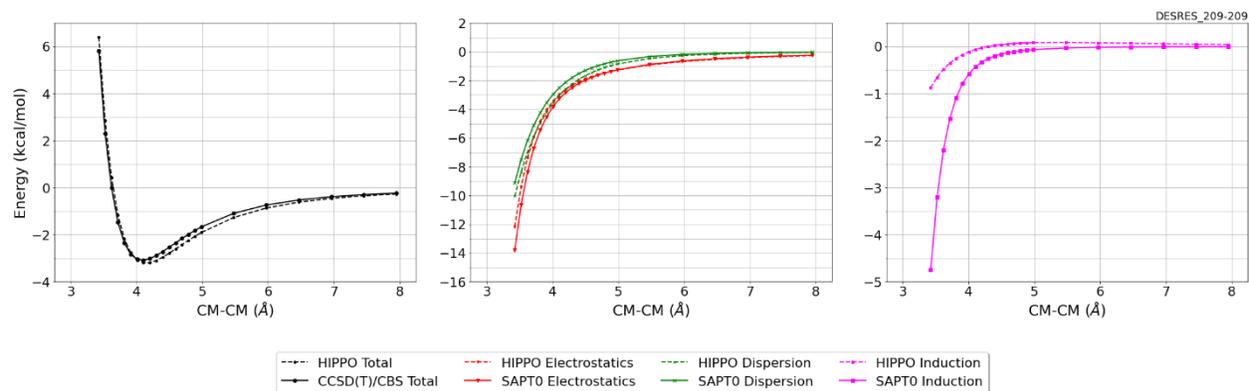
##Dimer results - Fitting to QM datasets##

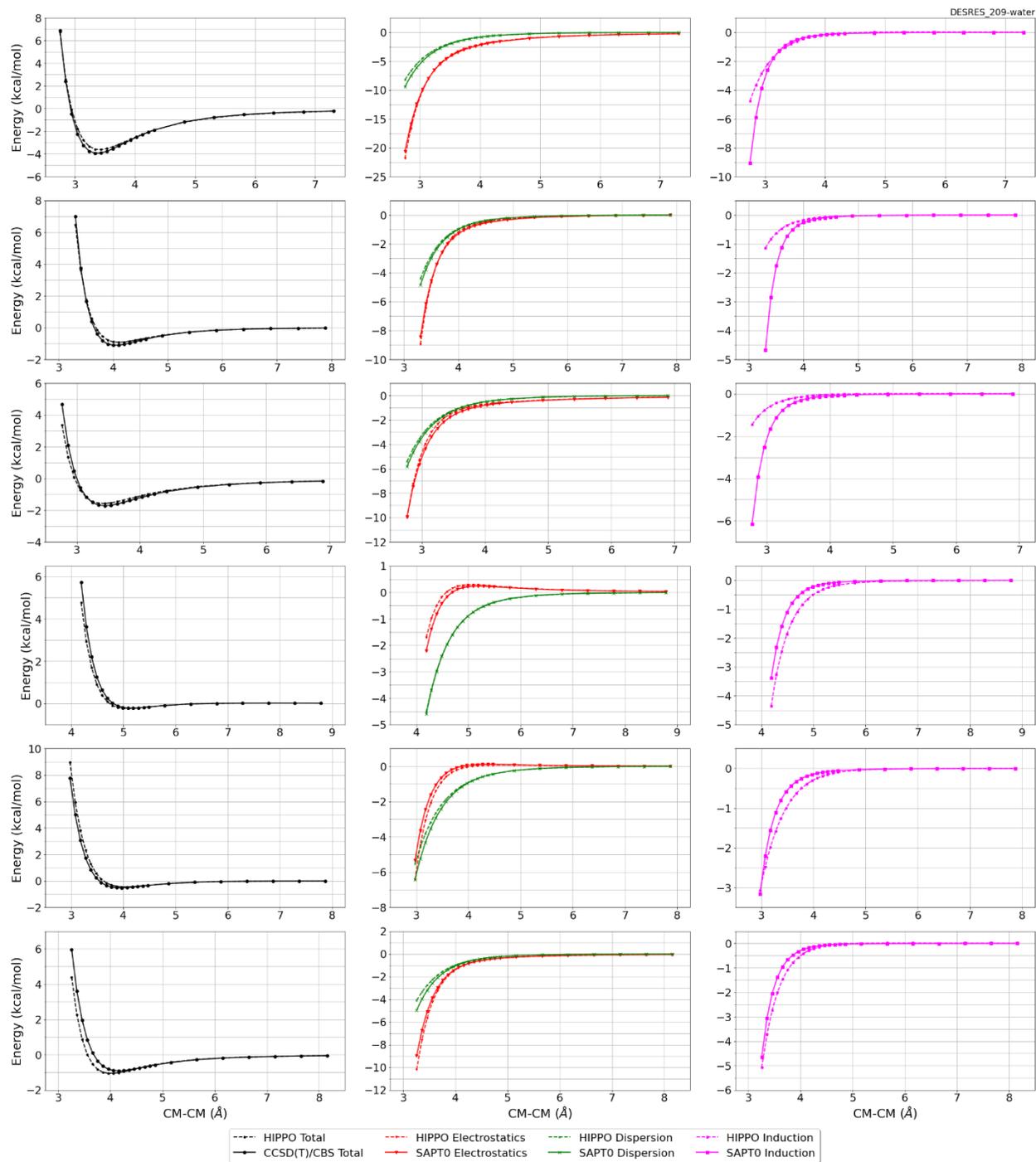
DESRES_209-209, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.239	0.115	0.4251	23	0

DESRES_209-water, energy values in kcal/mol

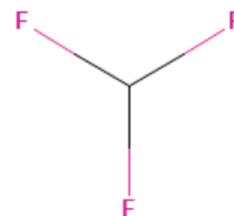
MAE	Std error	max error	#points	#count[err > 1]
0.145	0.244	1.6071	284	6





#210 Fluoroform CID: 6373

ref molpol	2.57	2.78	2.78, avg	2.71
molpol	2.60	2.77	2.77, avg	2.71
rms molpol	0.03	0.01	0.01, avg	0.01



Monomer potential fitting RMS: 0.20

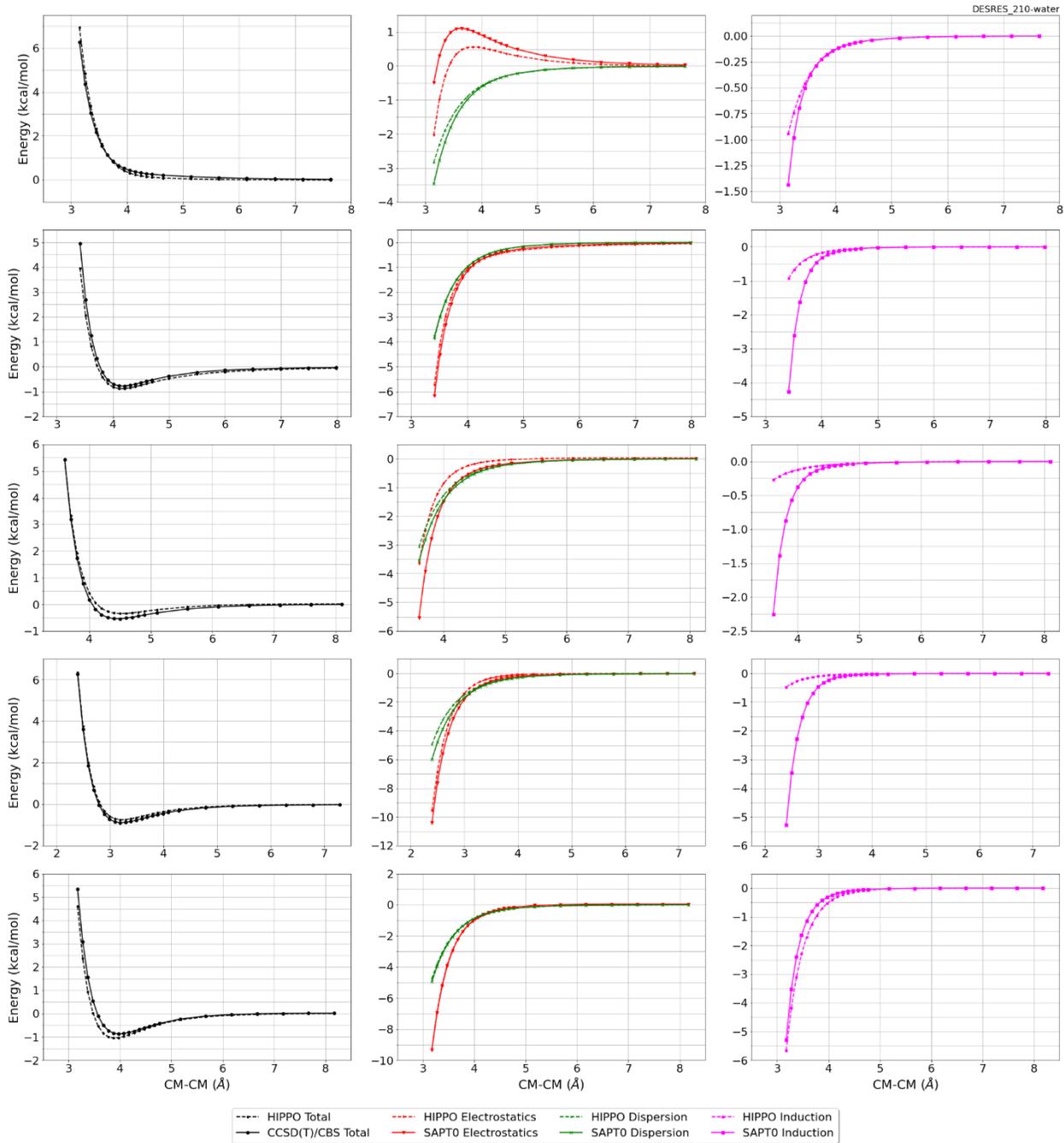
##Dimer results - Fitting to QM datasets##

DESRES_210-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.114	0.148	0.9967	262	0

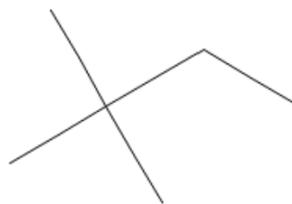
DESRES_210-210, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.037	0.022	0.0643	23	0



#214 2,2-Dimethylbutane CID: 6403

ref molpol	11.83	10.80	10.76, avg	11.13
molpol	11.83	10.78	10.77, avg	11.13
rms molpol	0.00	0.02	0.02, avg	0.00

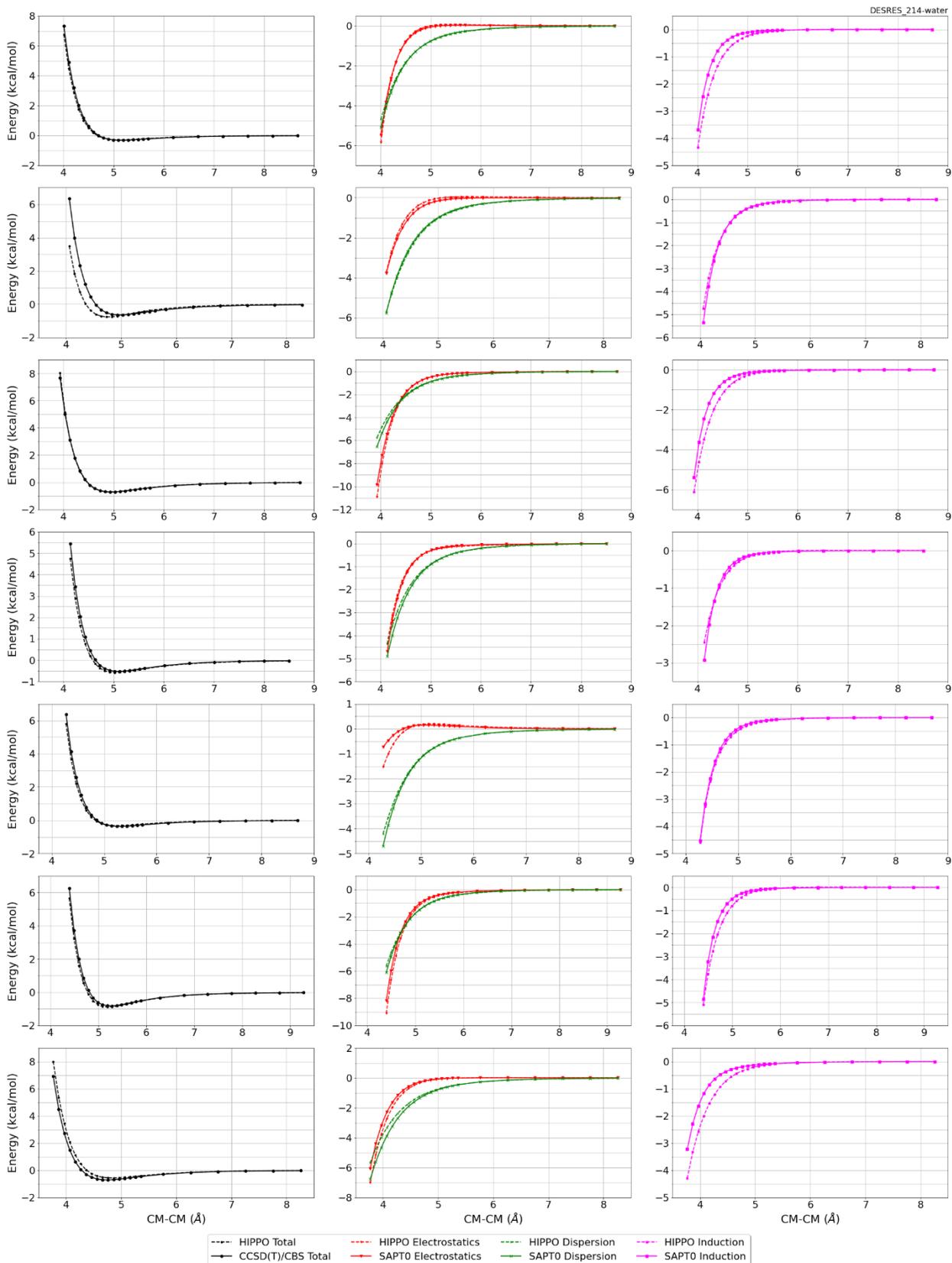


Monomer potential fitting RMS: 0.42

##Dimer results - Fitting to QM datasets##

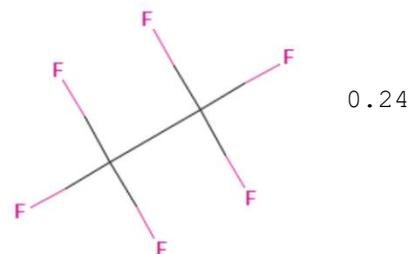
DESRES_214-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.133	0.249	2.9400	529	7



#215 Hexafluoroethane CID: 6431

ref molpol	4.71	4.77	4.77, avg	4.75
molpol	4.72	4.40	4.40, avg	4.51
rms molpol		0.01	0.37	



Monomer potential fitting RMS: 0.31

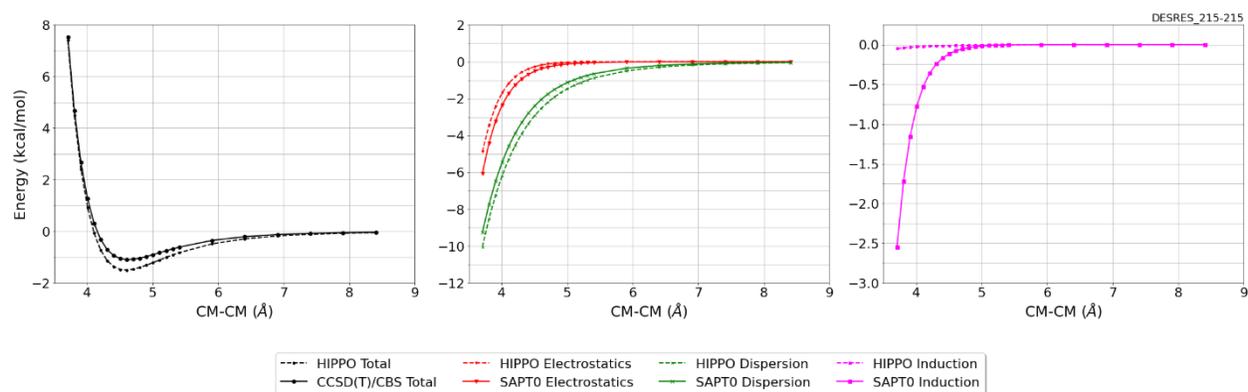
##Dimer results - Fitting to QM datasets##

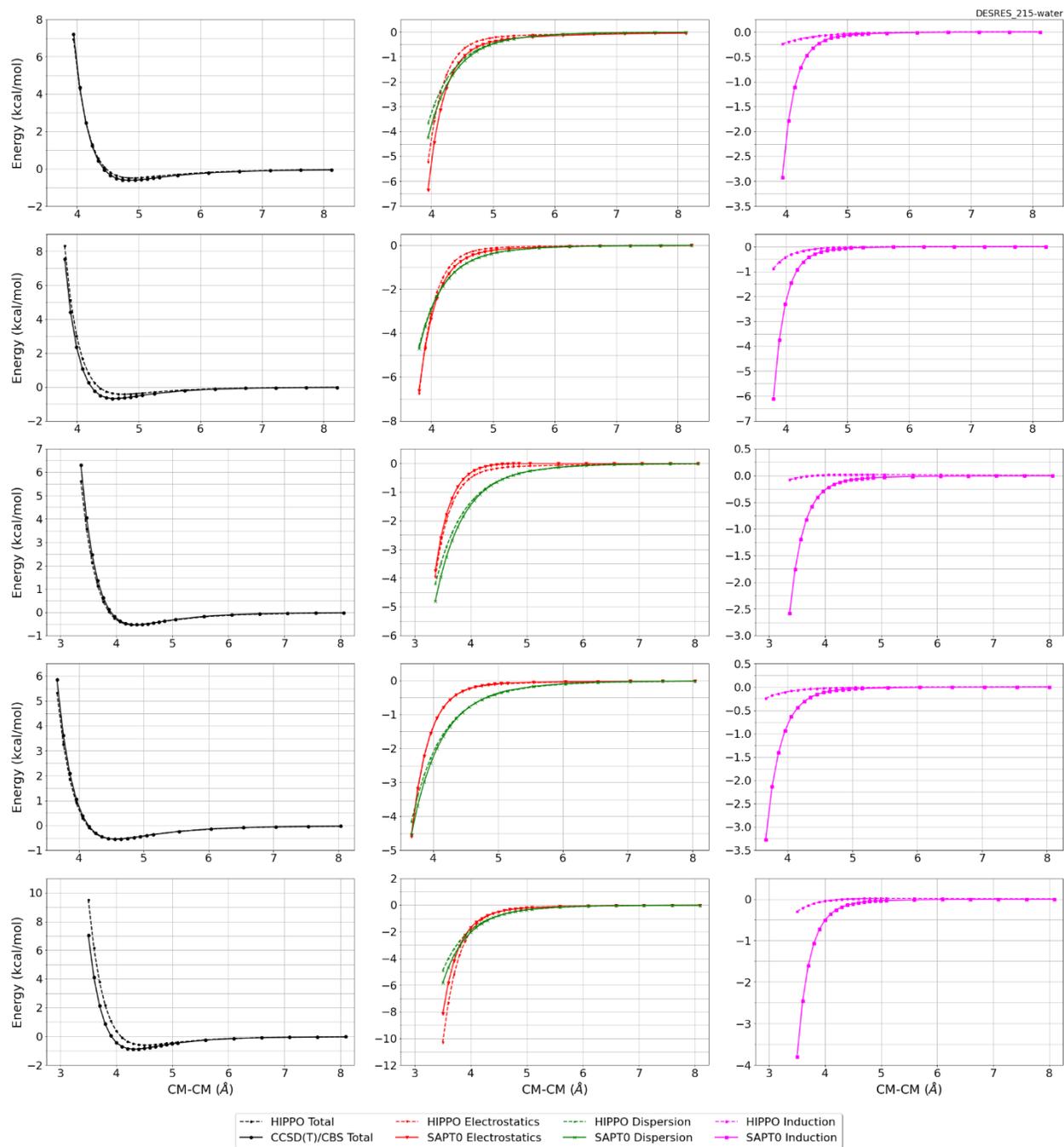
DESRES_215-215, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.251	0.137	0.4238	24	0

DESRES_215-water, energy values in kcal/mol

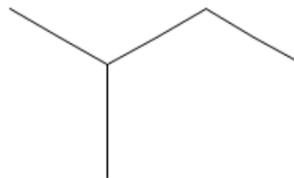
MAE	Std error	max error	#points	#count[err > 1]
0.153	0.276	2.4043	260	4





#217 2-Methylbutane CID: 6556

ref molpol	8.47	9.41	10.37, avg	9.42
molpol	8.46	9.42	10.38, avg	9.42
rms molpol	0.01	0.01	0.01, avg	0.00



Monomer potential fitting RMS: 0.62

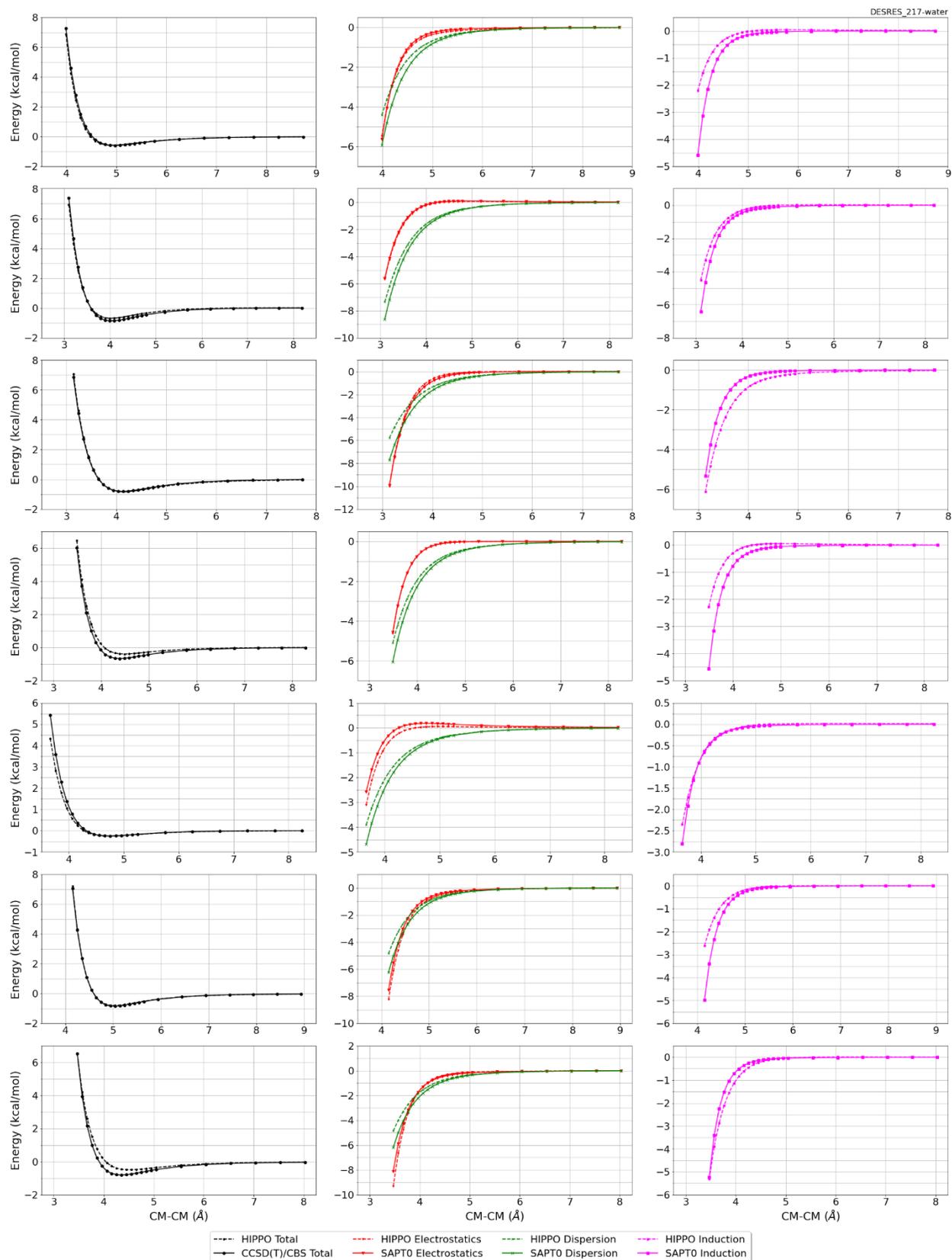
##Dimer results - Fitting to QM datasets##

DESRES_217-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.127	0.149	1.3292	537	1

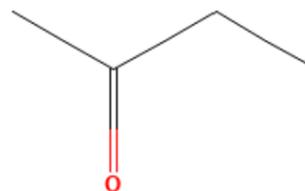
DESRES_217-217, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.054	0.072	0.3424	150	0



#219 Butan-2-one CID: 6569

ref molpol	8.12	8.95	0.00, avg	5.69
molpol	8.14	8.14	5.71, avg	7.33
rms molpol	0.02	0.81	5.71, avg	1.64



Monomer potential fitting RMS: 3.94

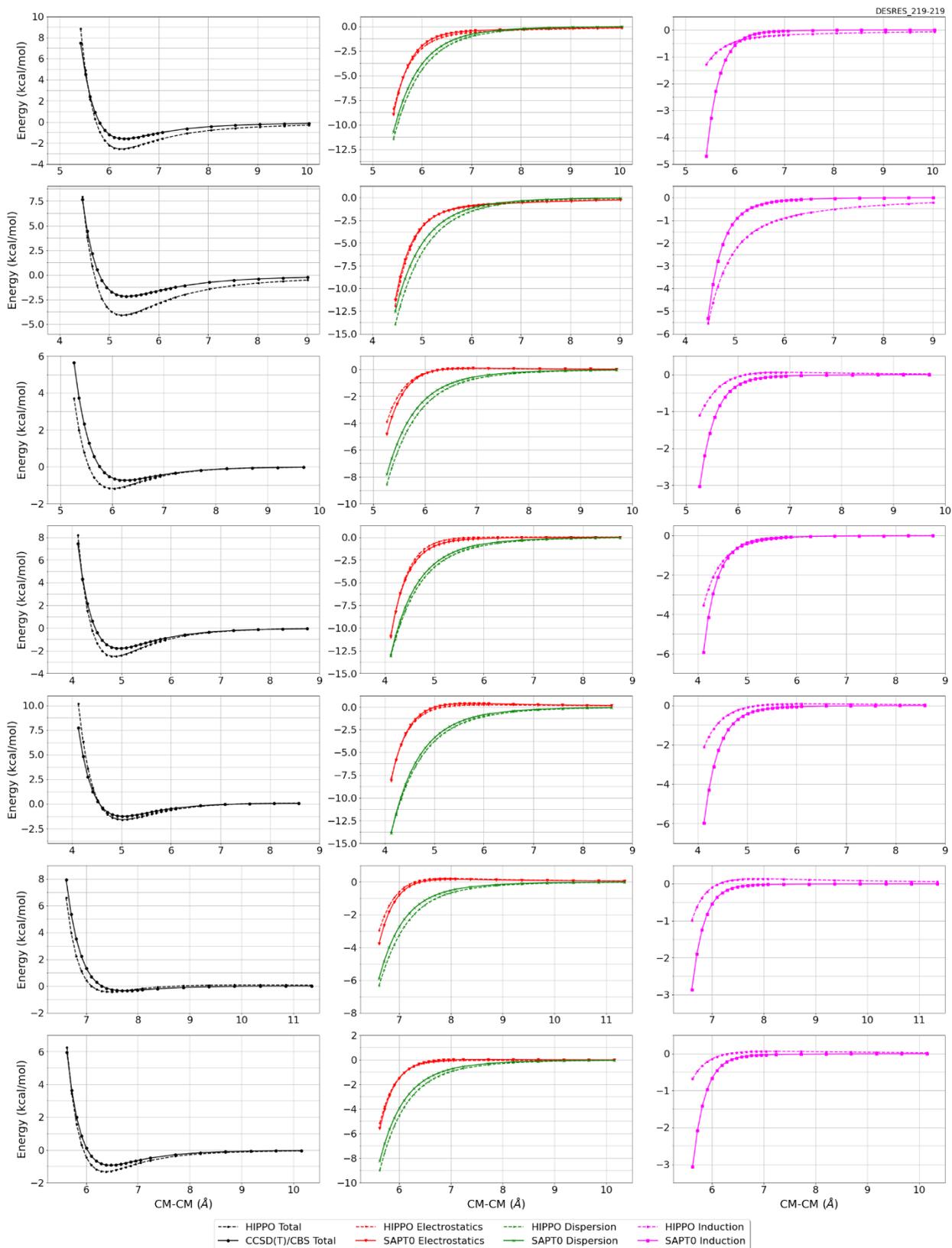
##Dimer results - Fitting to QM datasets##

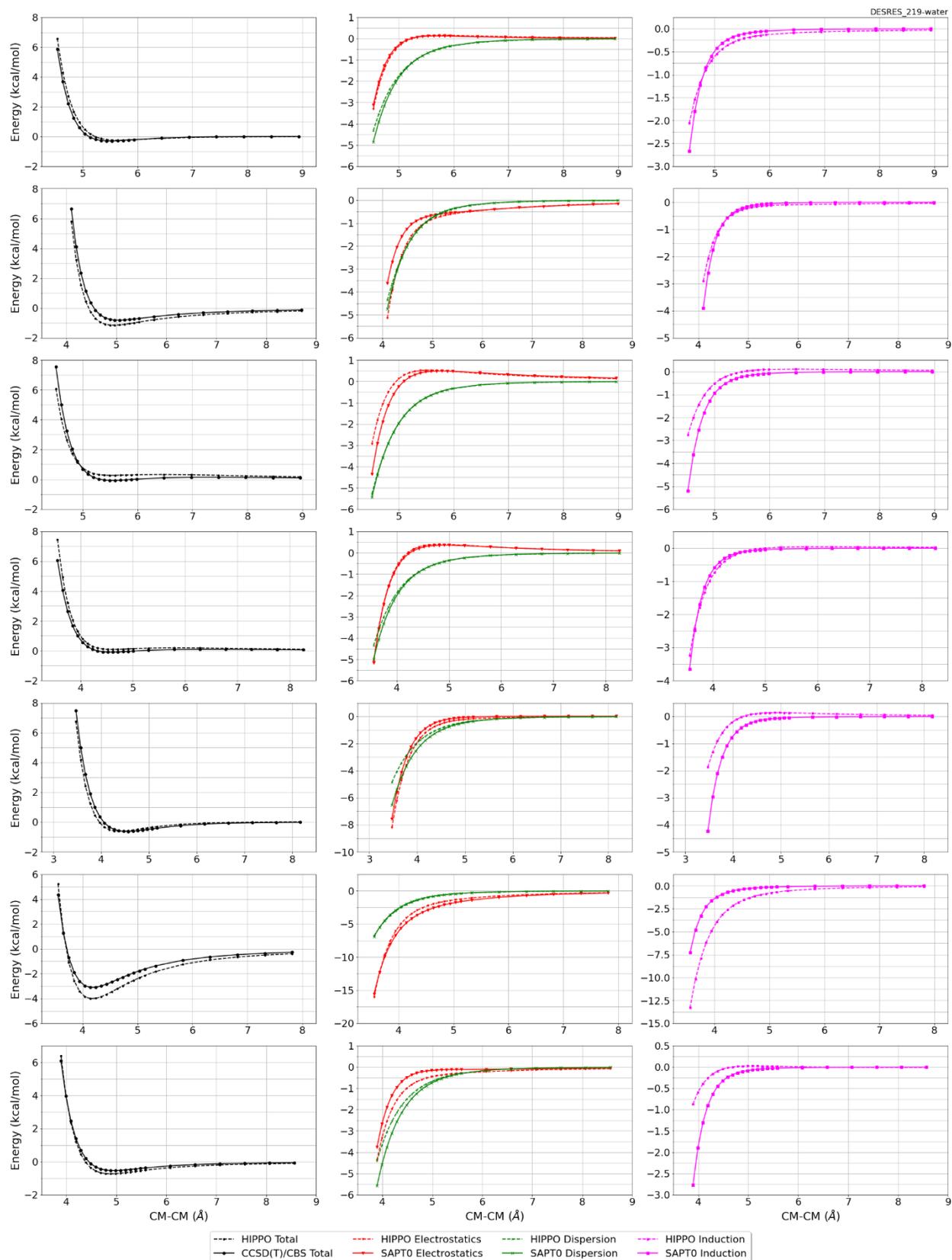
DESRES_219-219, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.435	0.447	2.4732	453	56

DESRES_219-water, energy values in kcal/mol

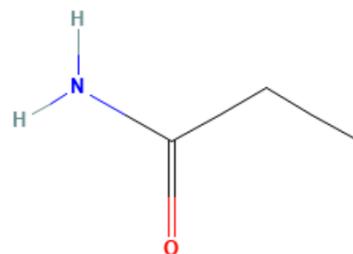
MAE	Std error	max error	#points	#count[err > 1]
0.271	0.330	2.7983	529	27





#221 Propanamide CID: 6578

ref molpol	7.73	8.82	0.00, avg	5.52
molpol	7.74	7.74	5.64, avg	7.04
rms molpol	0.01	1.08	5.64, avg	1.52



Monomer potential fitting RMS: 0.31

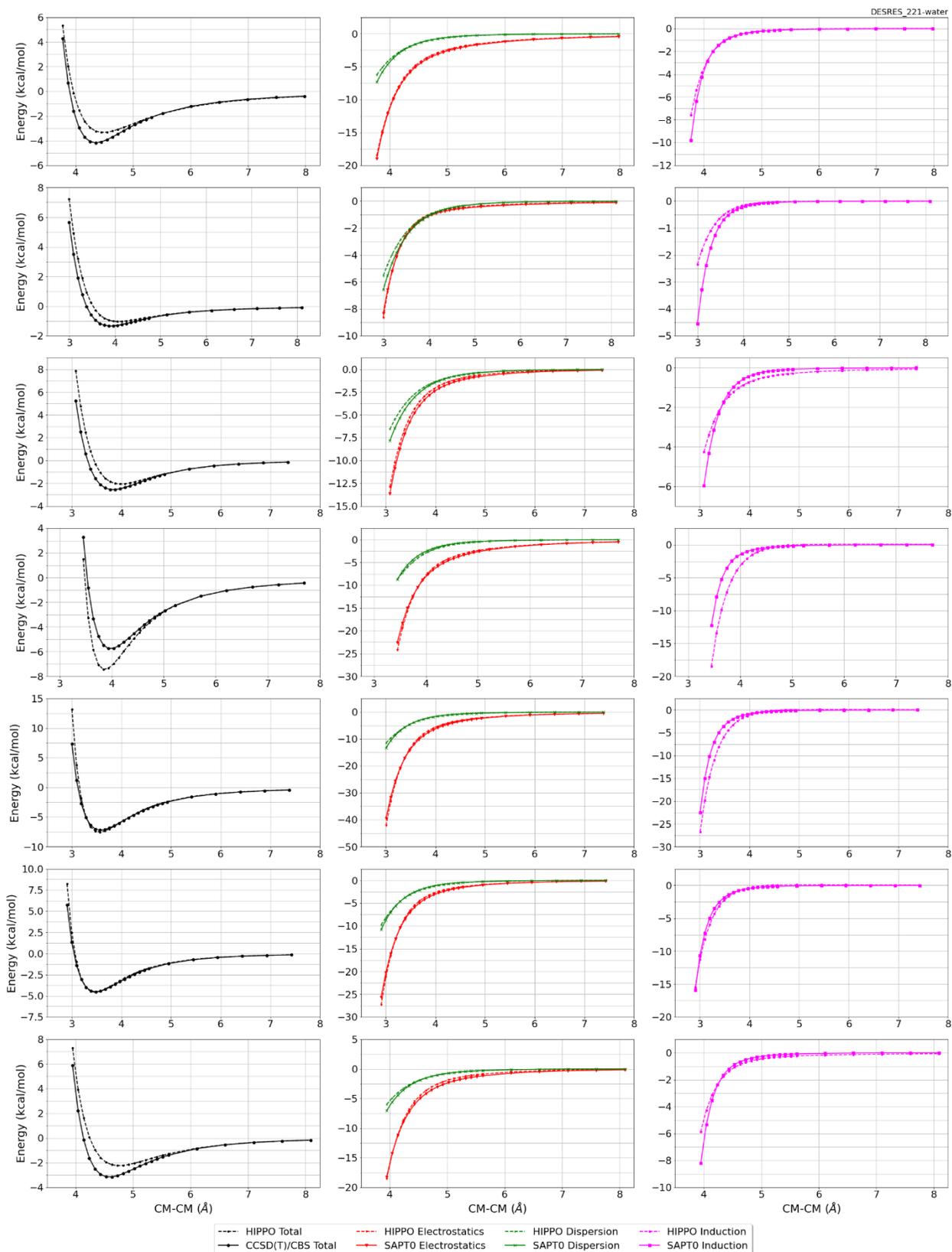
##Dimer results - Fitting to QM datasets##

DESRES_221-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.398	0.597	5.8420	540	71

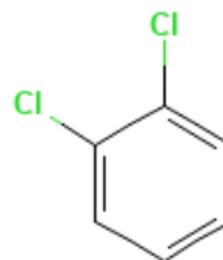
DESRES_221-221, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.475	0.632	4.7435	521	58



#223 1,2-Dichlorobenzene C₆H₄Cl₂ CID: 7239

ref molpol	18.28	15.90	0.00, avg	11.39
molpol	18.22	15.85	8.67, avg	14.25
rms molpol	0.06	0.04	8.67, avg	2.86

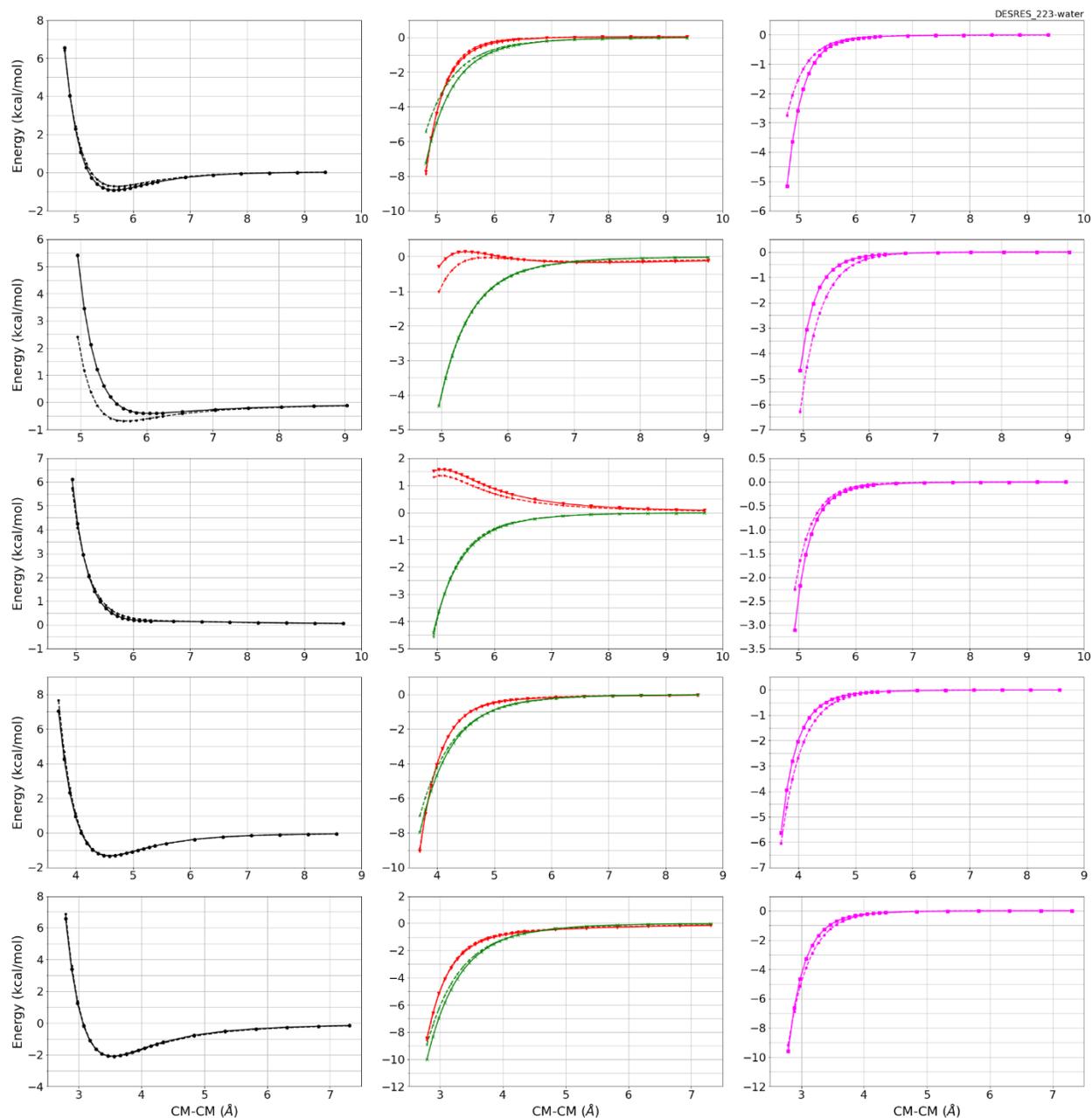


Monomer potential fitting RMS: 0.28

##Dimer results - Fitting to QM datasets##

DESRES_223-water, energy values in kcal/mol

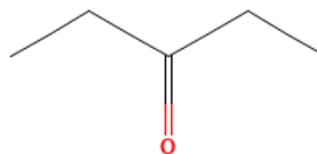
MAE	Std error	max error	#points	#count[err > 1]
0.181	0.413	2.9929	114	5



--- HIPPO Total	--- HIPPO Electrostatics	--- HIPPO Dispersion	--- HIPPO Induction
--- CCSD(T)/CBS Total	--- SAPTO Electrostatics	--- SAPTO Dispersion	--- SAPTO Induction

#224 Pentan-3-one C5H10O CID: 7288

ref molpol	11.15	9.61	0.00, avg	6.92
molpol	11.09	9.56	6.69, avg	9.11
rms molpol	0.06	0.05	6.69, avg	2.19

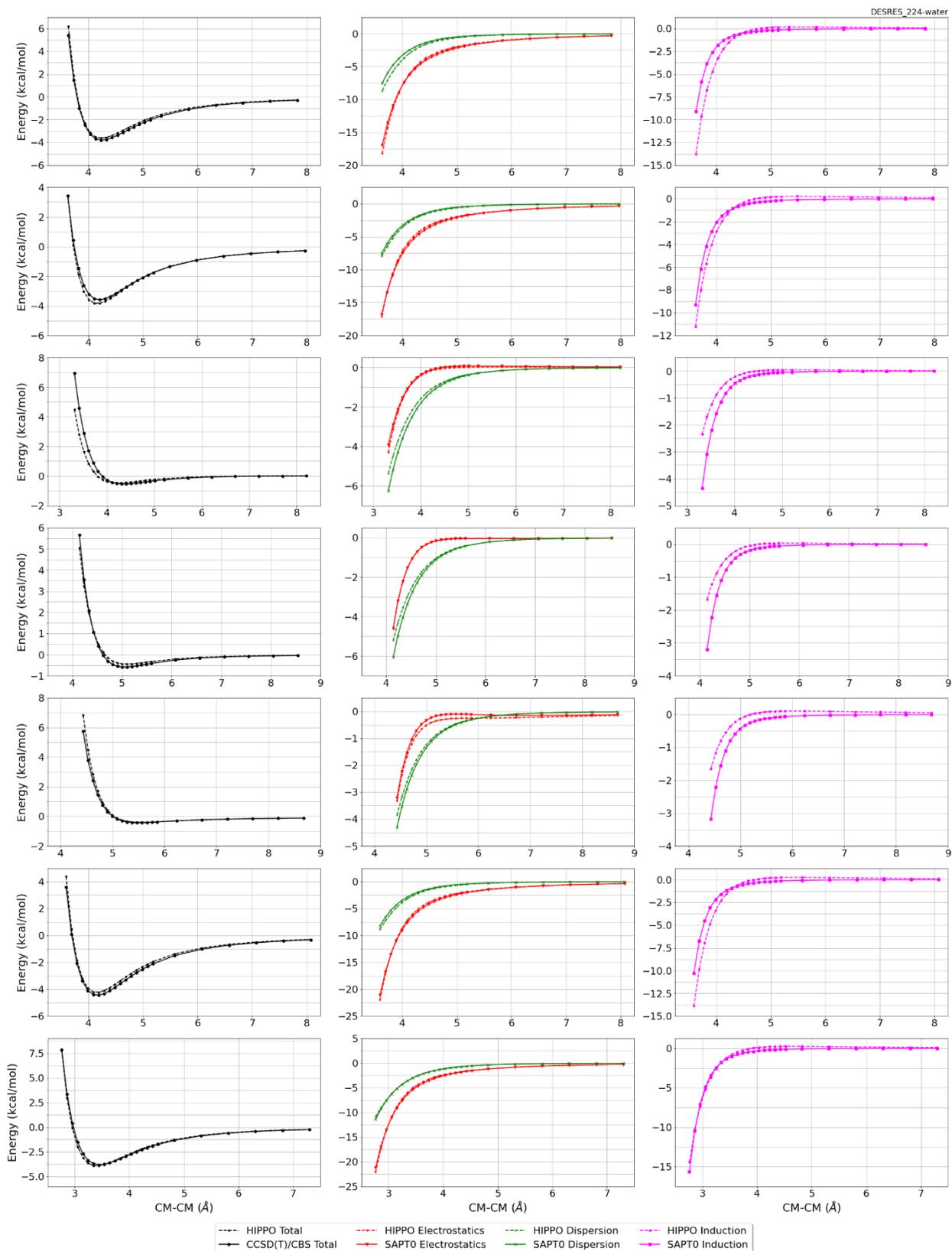


Monomer potential fitting RMS: 0.87

##Dimer results - Fitting to QM datasets##

DESRES_224-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.241	0.346	2.9407	531	21



#225 Methylcyclopentane C6H12 CID: 7296

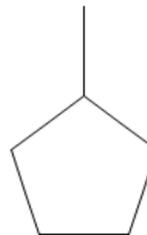
ref molpol	11.62	10.47	9.23, avg	10.44
molpol	11.51	10.40	9.17, avg	10.36
rms molpol	0.11	0.08	0.07, avg	0.08

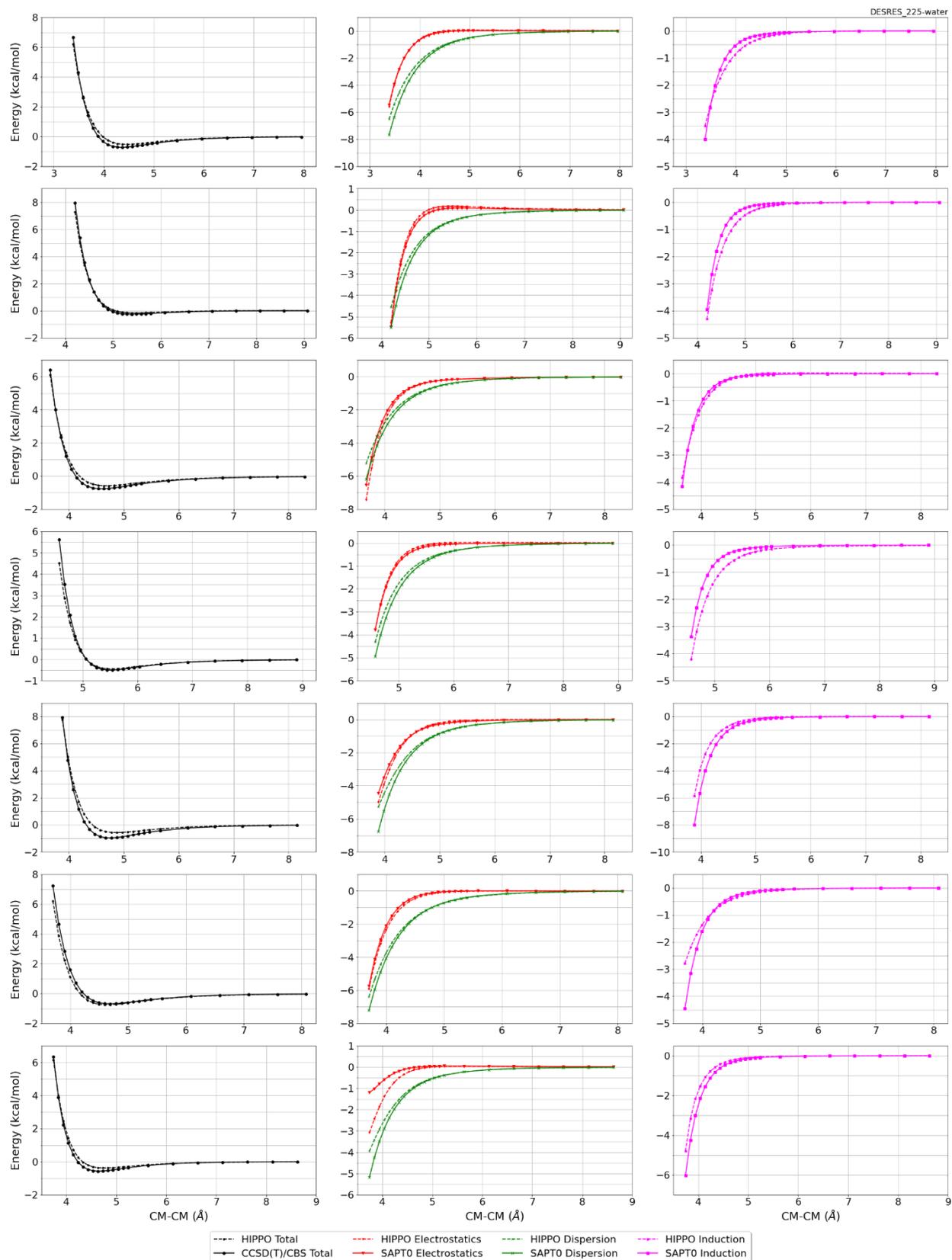
Monomer potential fitting RMS: 0.23

##Dimer results - Fitting to QM datasets##

DESRES_225-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.134	0.155	0.9964	535	0





#228 Butane C4H10 CID: 7843

ref molpol	7.33	8.86	0.00, avg	5.40
molpol	7.33	6.27	5.85, avg	6.48
rms molpol	0.01	2.59	5.85, avg	1.08



Monomer potential fitting RMS: 0.63

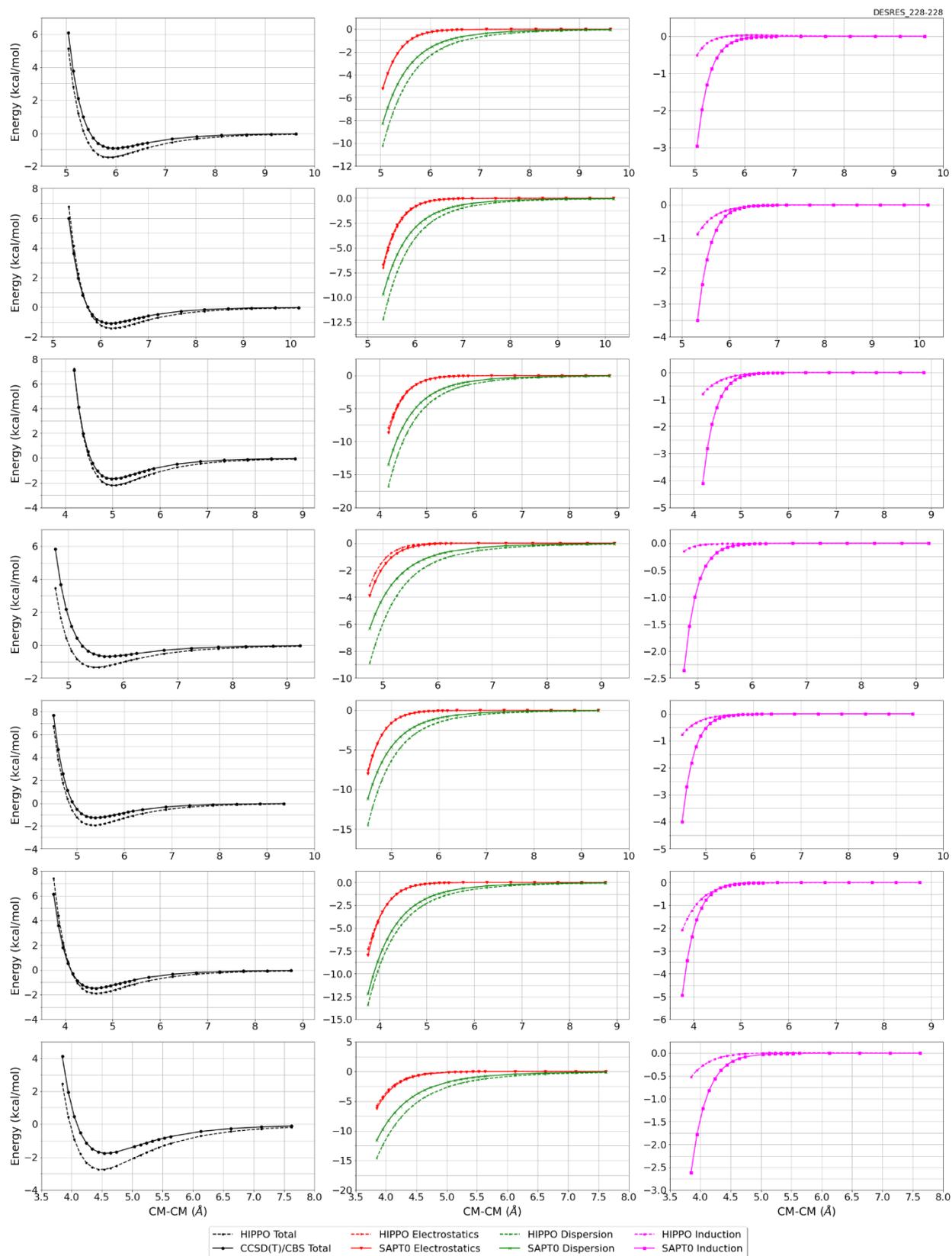
##Dimer results - Fitting to QM datasets##

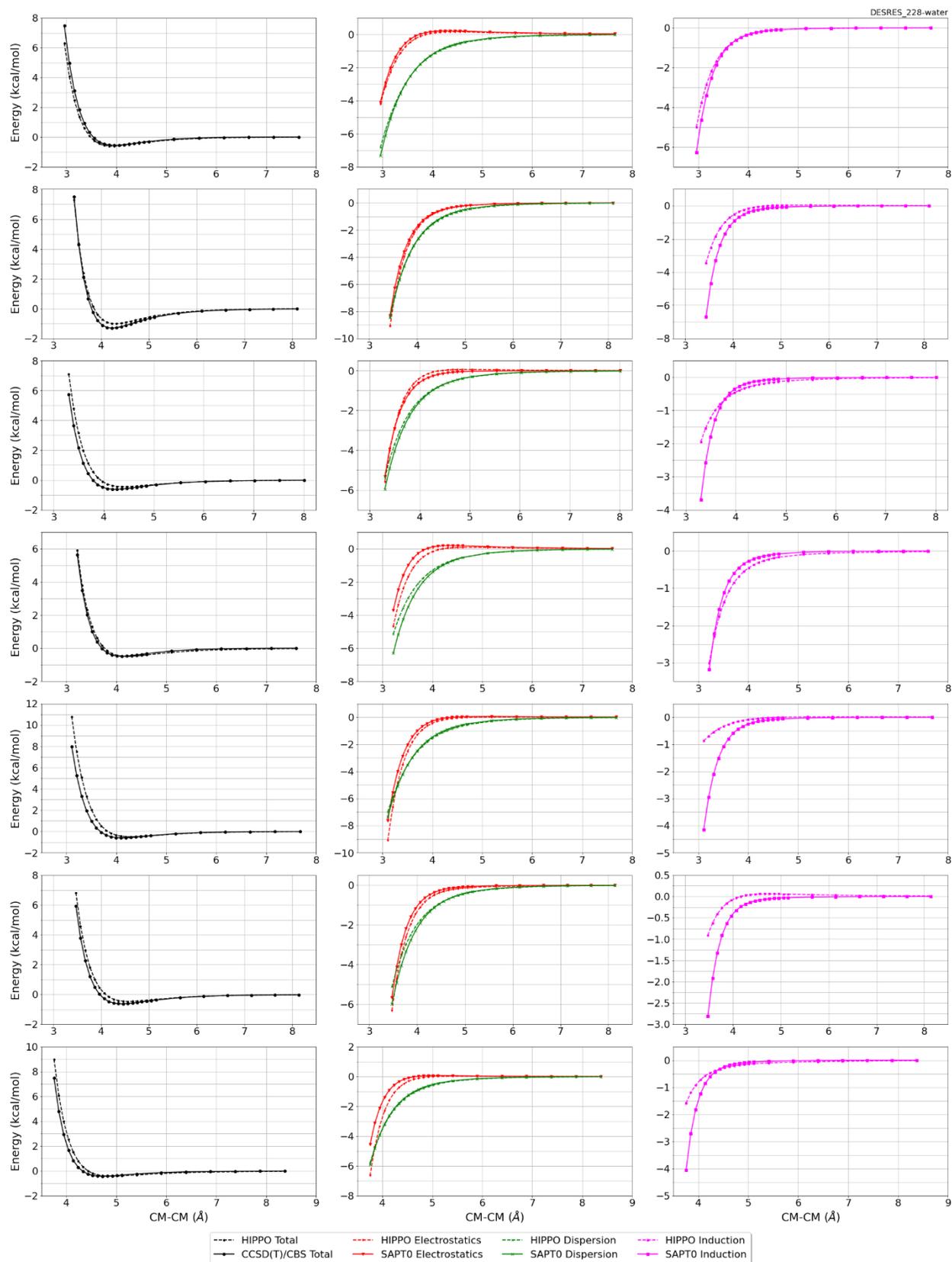
DESRES_228-228, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.400	0.352	2.3788	473	23

DESRES_228-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.166	0.298	2.8088	556	16



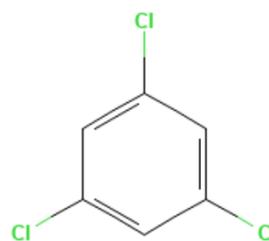


#233 1,3,5-Trichlorobenzene C6H3Cl3 CID: 7950

ref molpol 20.22 20.22 0.00, avg 13.48
molpol 20.19 20.19 10.65, avg 17.01
rms molpol 0.03 0.03 10.65, avg 3.53

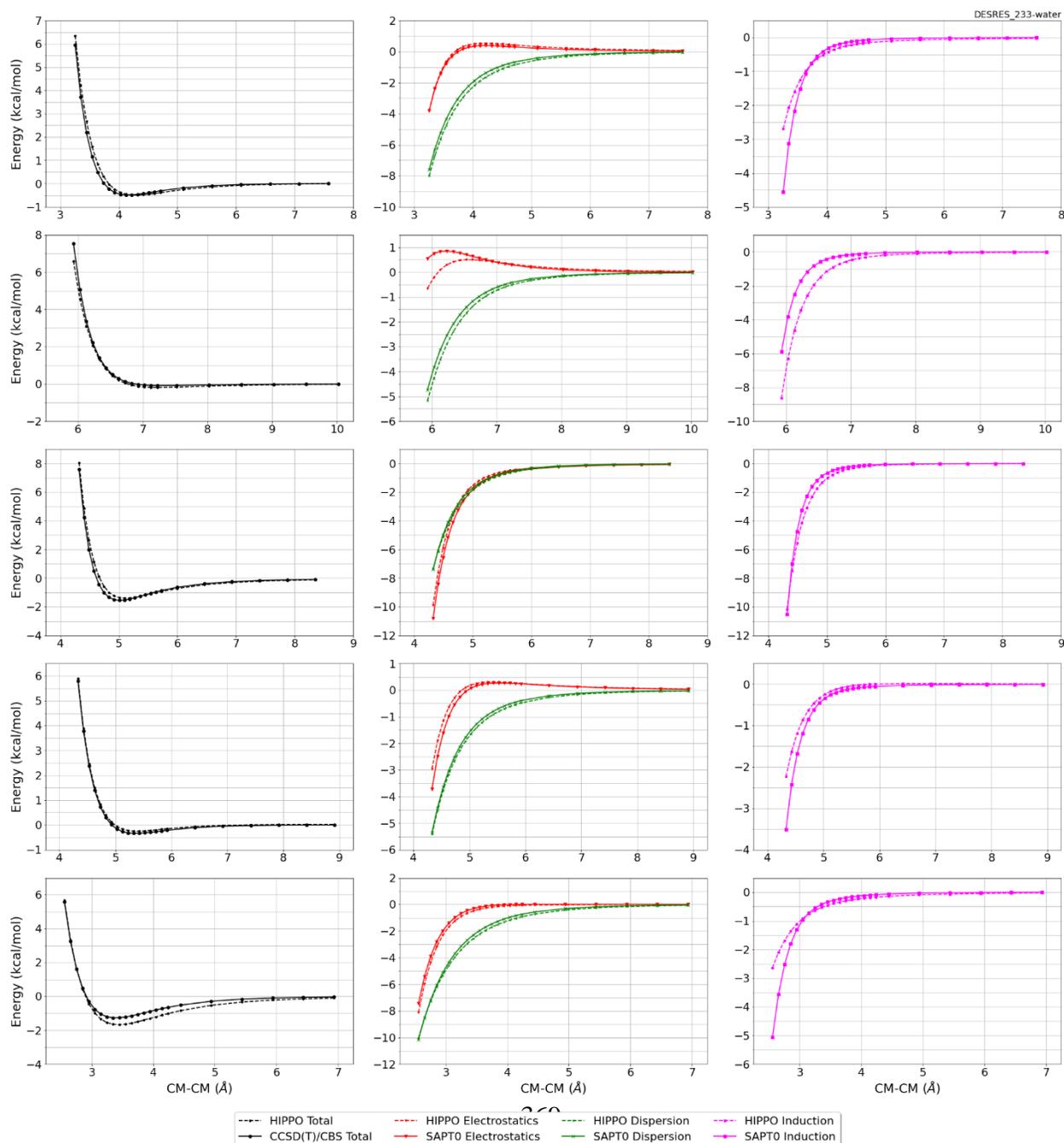
Monomer potential fitting RMS: 0.16

##Dimer results - Fitting to QM datasets##



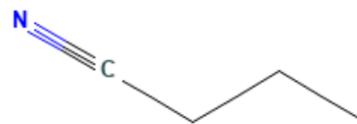
DESRES_233-water, energy values in kcal/mol

MAE 0.156 Std error 0.189 max error 0.7357 #points 112 #count[err > 1] 0



#236 Butanenitrile C4H7N CID: 8008

ref molpol	6.94	10.12	0.00, avg	5.69
molpol	6.93	6.07	5.82, avg	6.27
rms molpol	0.01	4.05	5.82, avg	0.58

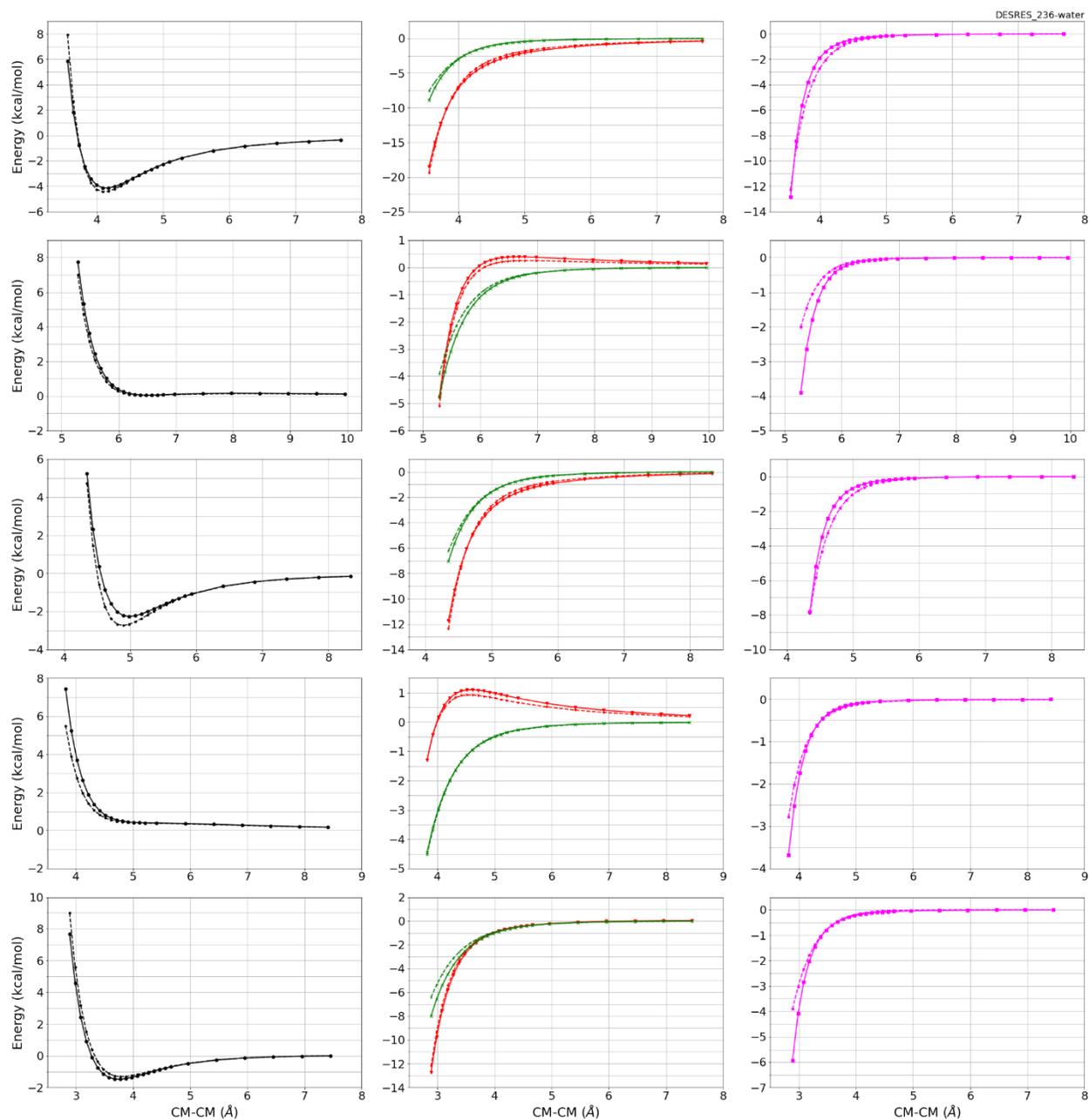


Monomer potential fitting RMS: 0.50

##Dimer results - Fitting to QM datasets##

DESRES_236-water, energy values in kcal/mol

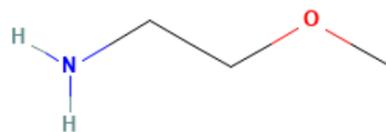
MAE	Std error	max error	#points	#count[err > 1]
0.166	0.321	2.8911	267	5



--- HIPPO Total - - - HIPPO Electrostatics - - - HIPPO Dispersion - - - HIPPO Induction
- - - CCSD(T)/CBS Total - - - SAPTO Electrostatics - - - SAPTO Dispersion - - - SAPTO Induction

#240 2-Methoxyethanamine C3H9NO CID: 8018

ref molpol	9.57	7.56	7.08, avg	8.07
molpol	9.59	7.52	7.10, avg	8.07
rms molpol	0.01	0.03	0.03, avg	0.00

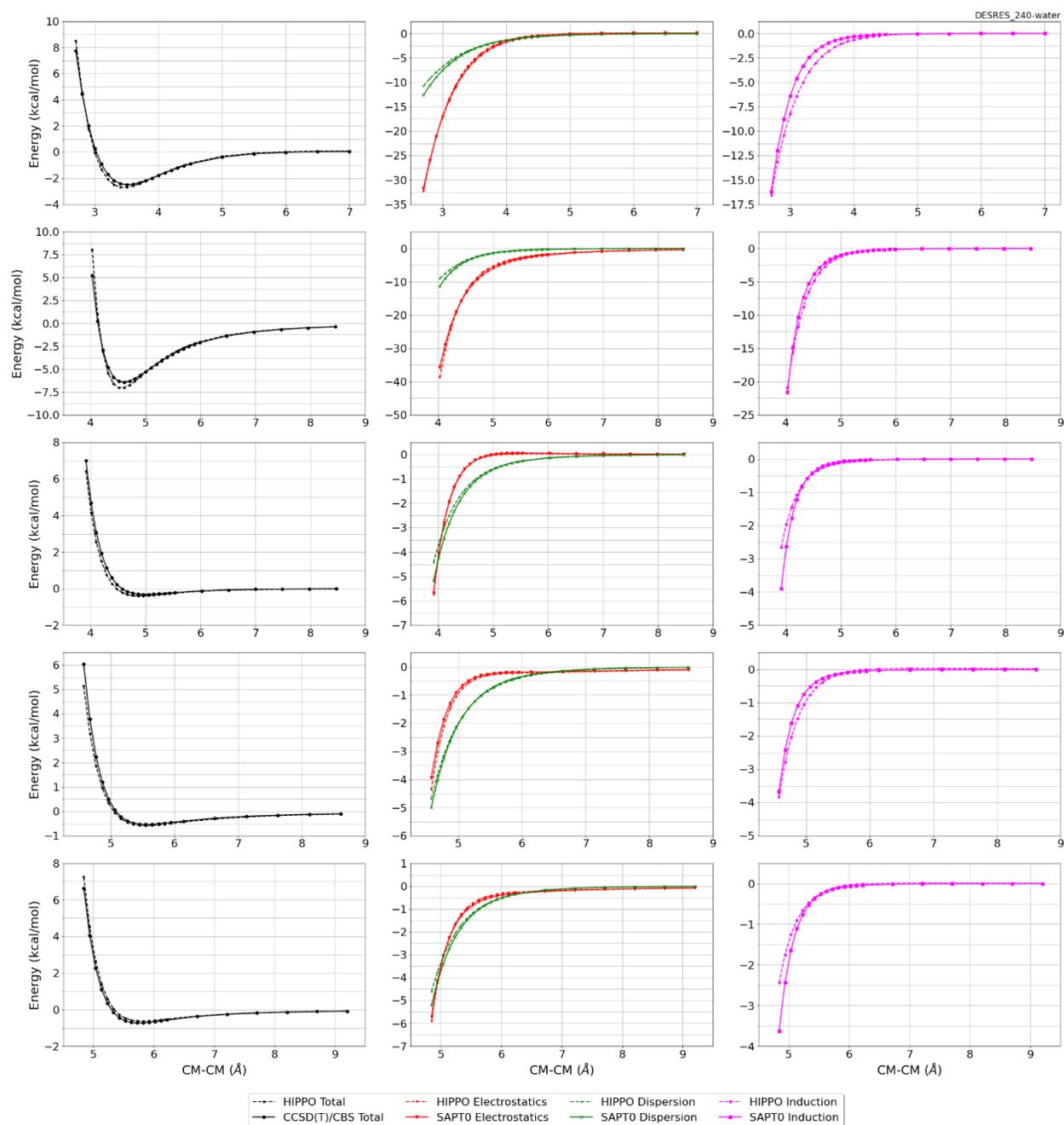


Monomer potential fitting RMS: 0.46

##Dimer results - Fitting to QM datasets##

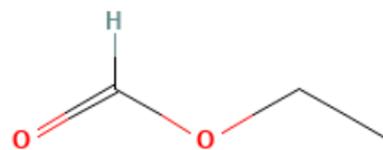
DESRES_240-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.185	0.358	3.0338	266	7



#242 Ethyl Formate C3H6O2 CID: 8025

ref molpol	9.04	6.44	0.00, avg	5.16
molpol	9.02	6.38	5.40, avg	6.93
rms molpol	0.02	0.06	5.40, avg	1.77



Monomer potential fitting RMS: 0.27

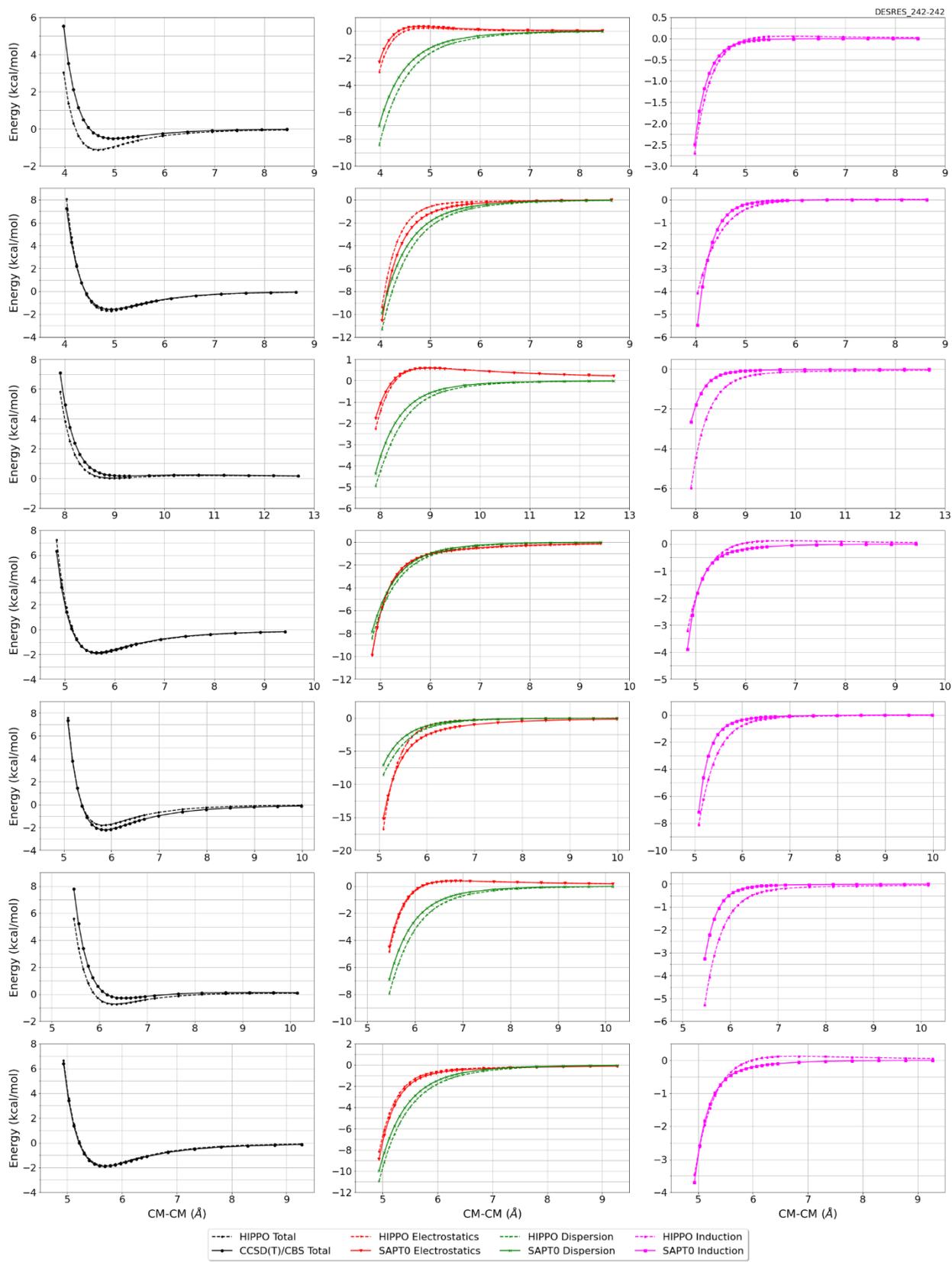
##Dimer results - Fitting to QM datasets##

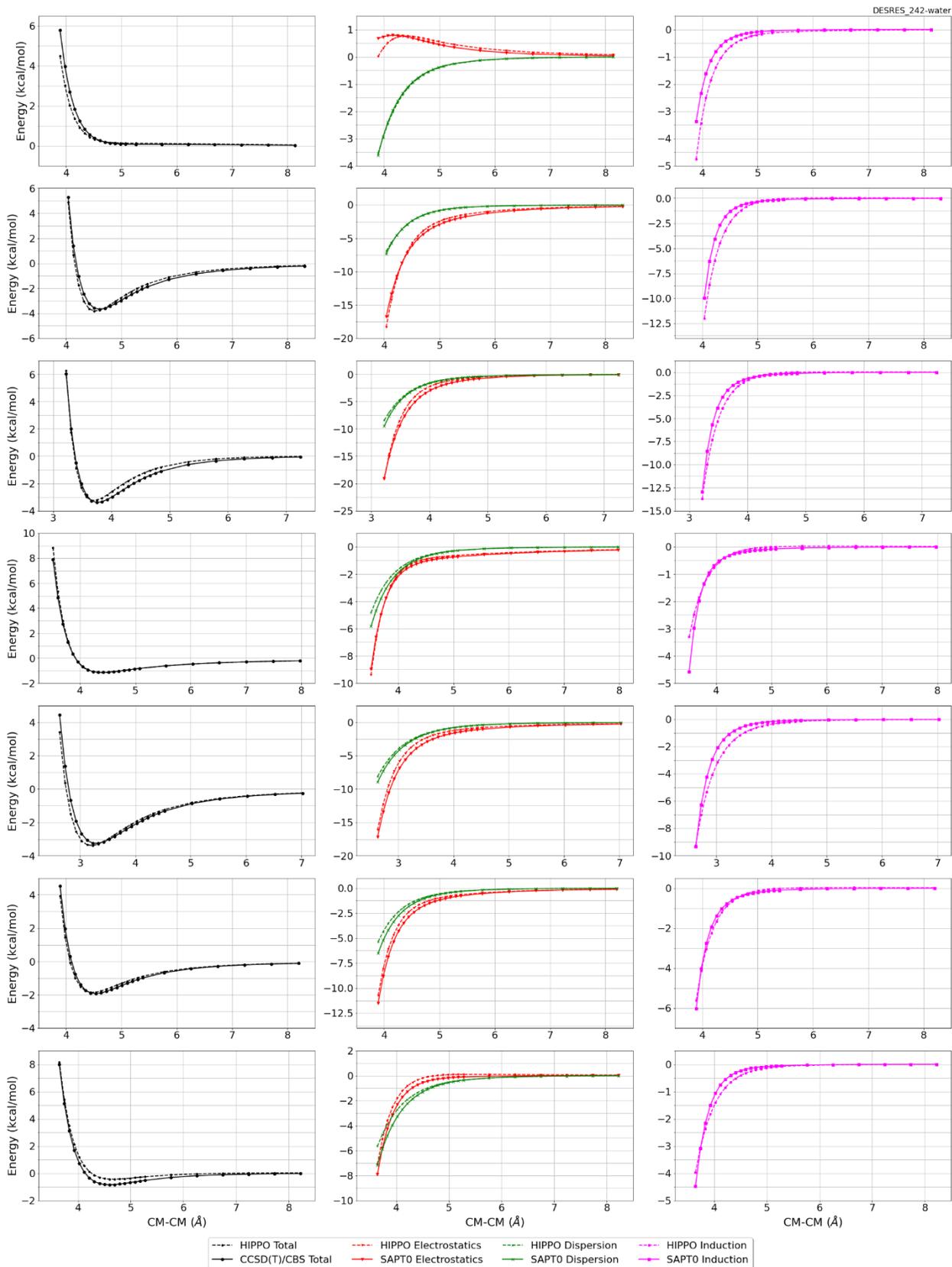
DESRES_242-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.197	0.276	2.7101	533	13

DESRES_242-242, energy values in kcal/mol

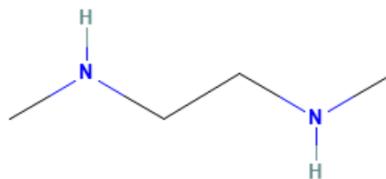
MAE	Std error	max error	#points	#count[err > 1]
0.259	0.384	3.0052	516	21





#244 N,N'-Dimethylethane-1,2-diamine C4H12N2 CID: 8070

ref molpol 12.81 9.90 9.01, avg 10.57
 molpol 13.07 9.17 8.95, avg 10.40
 rms molpol 0.26 0.74 0.06, avg 0.18

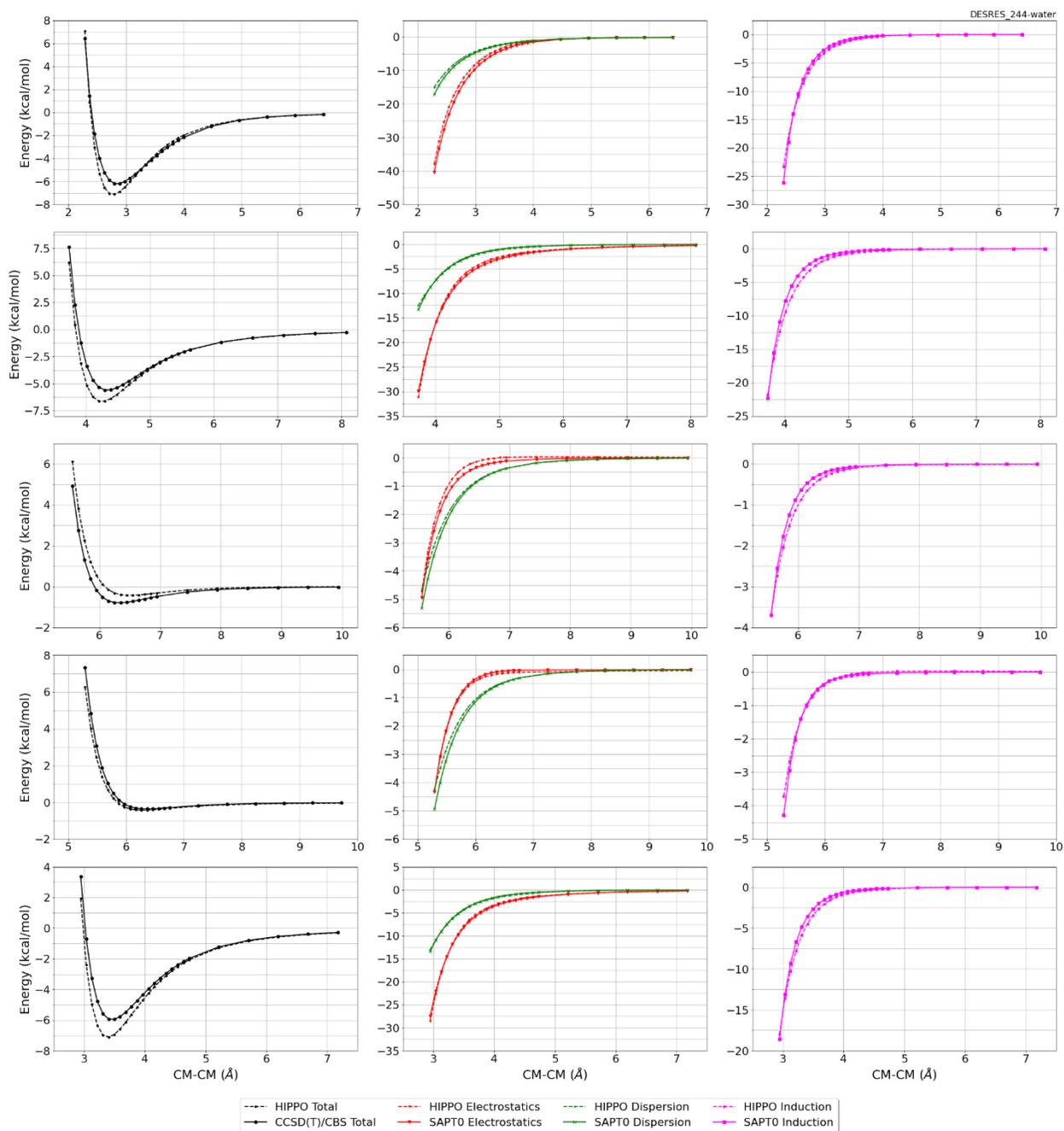


Monomer potential fitting RMS: 0.71

##Dimer results - Fitting to QM datasets##

DESRES_244-water, energy values in kcal/mol

MAE 0.377 Std error 0.419 max error 1.5307 #points 119 #count[err > 1] 16



#245 1,2-Dimethoxyethane C4H10O2 CID: 8071

ref molpol	8.37	11.42	0.00, avg	6.60
molpol	8.40	6.20	6.04, avg	6.88
rms molpol	0.03	5.22	6.04, avg	0.28

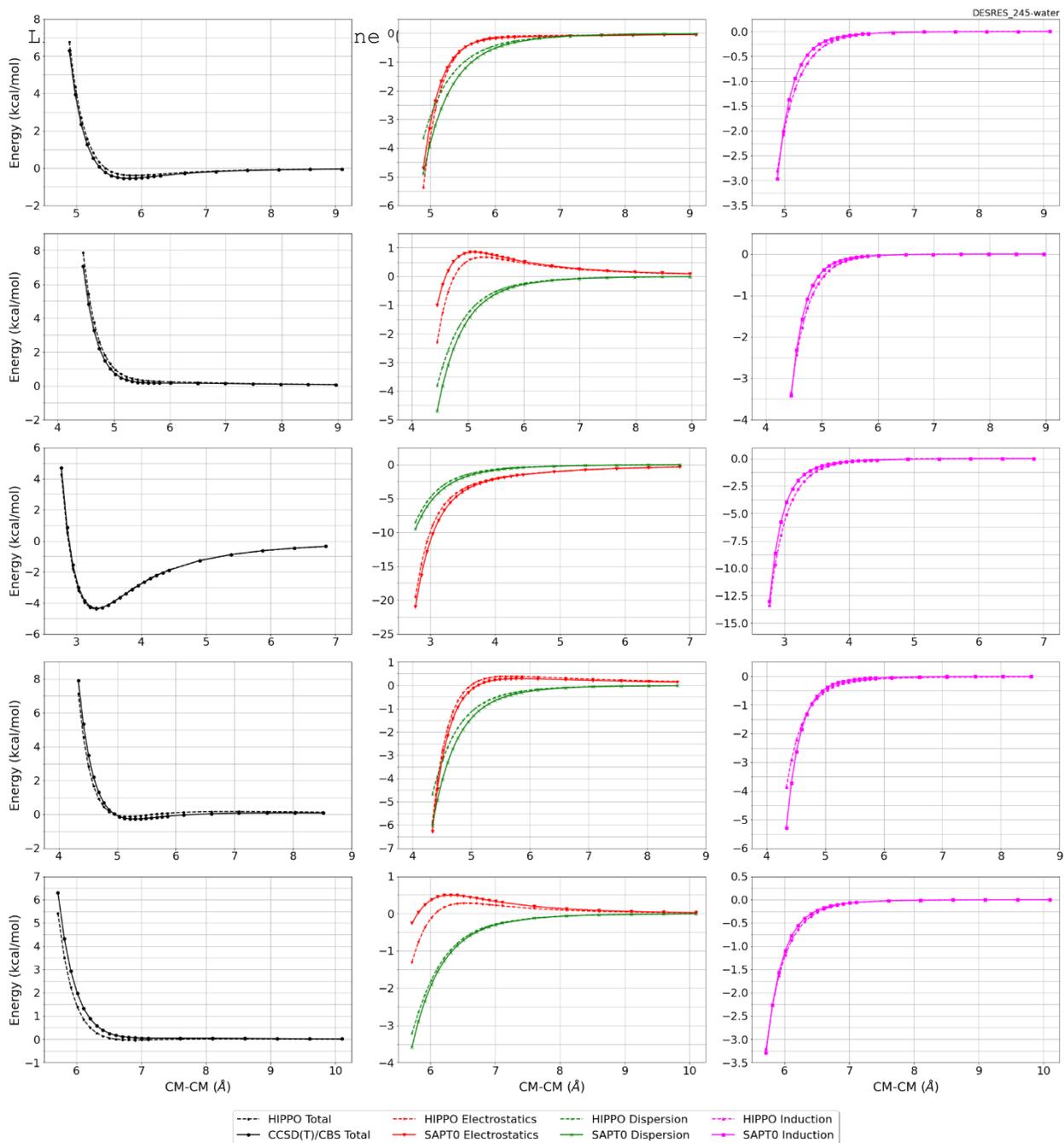


Monomer potential fitting RMS: 0.37

##Dimer results - Fitting to QM datasets##

DESRES_245-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.194	0.204	0.8977	113	0



#248 Cyclohexane C6H12 CID: 8078

ref molpol	9.38	10.80	10.80, avg	10.32
molpol	9.27	10.84	10.84, avg	10.32
rms molpol	0.10	0.04	0.04, avg	0.01

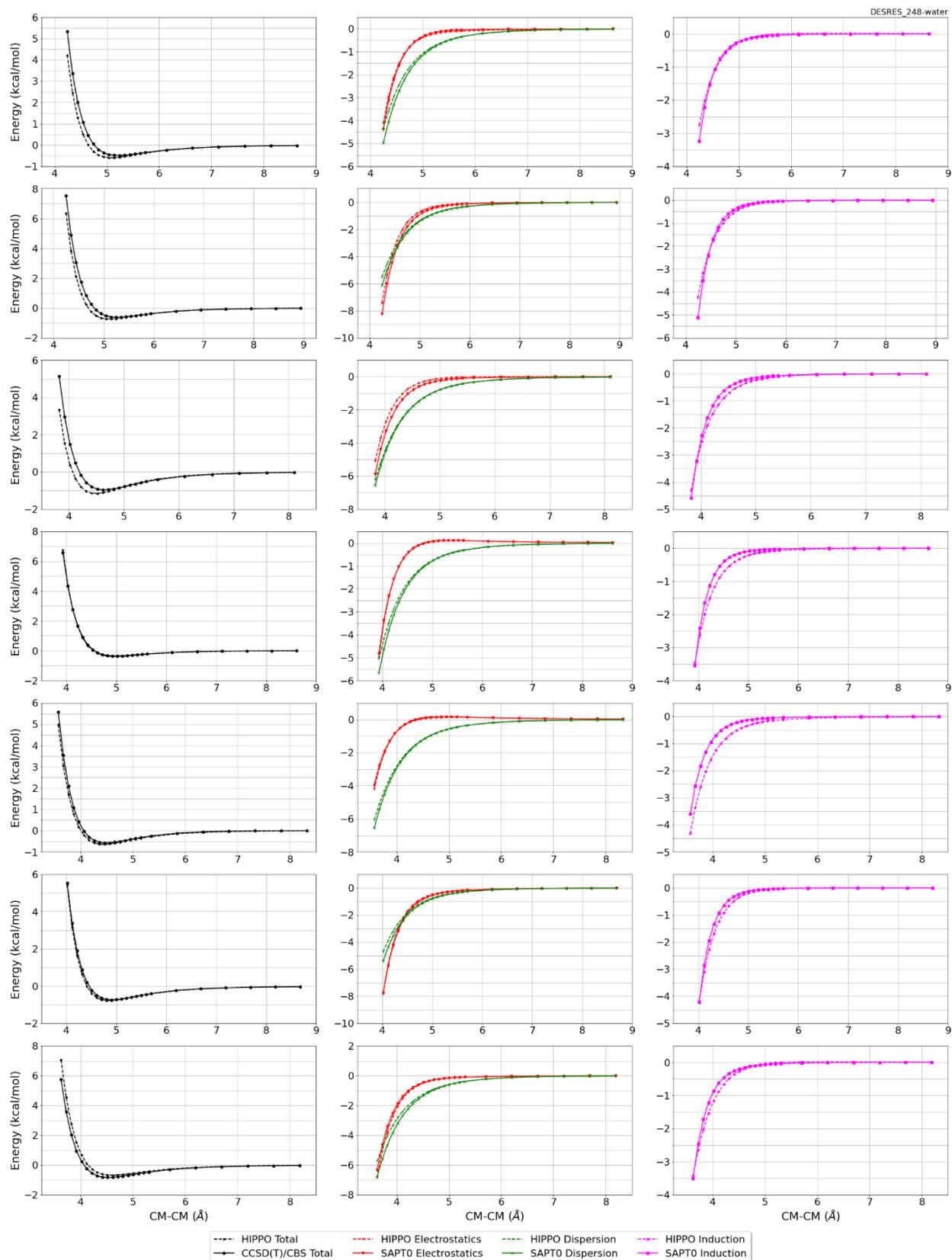


Monomer potential fitting RMS: 0.13

##Dimer results - Fitting to QM datasets##

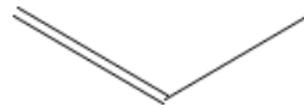
DESRES_248-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.120	0.210	1.6906	560	11



#251 Propene C3H6 CID: 8252

ref molpol	7.18	5.44	0.00, avg	4.21
molpol	6.95	5.65	4.56, avg	5.72
rms molpol	0.22	0.20	4.56, avg	1.51



Monomer potential fitting RMS: 0.28

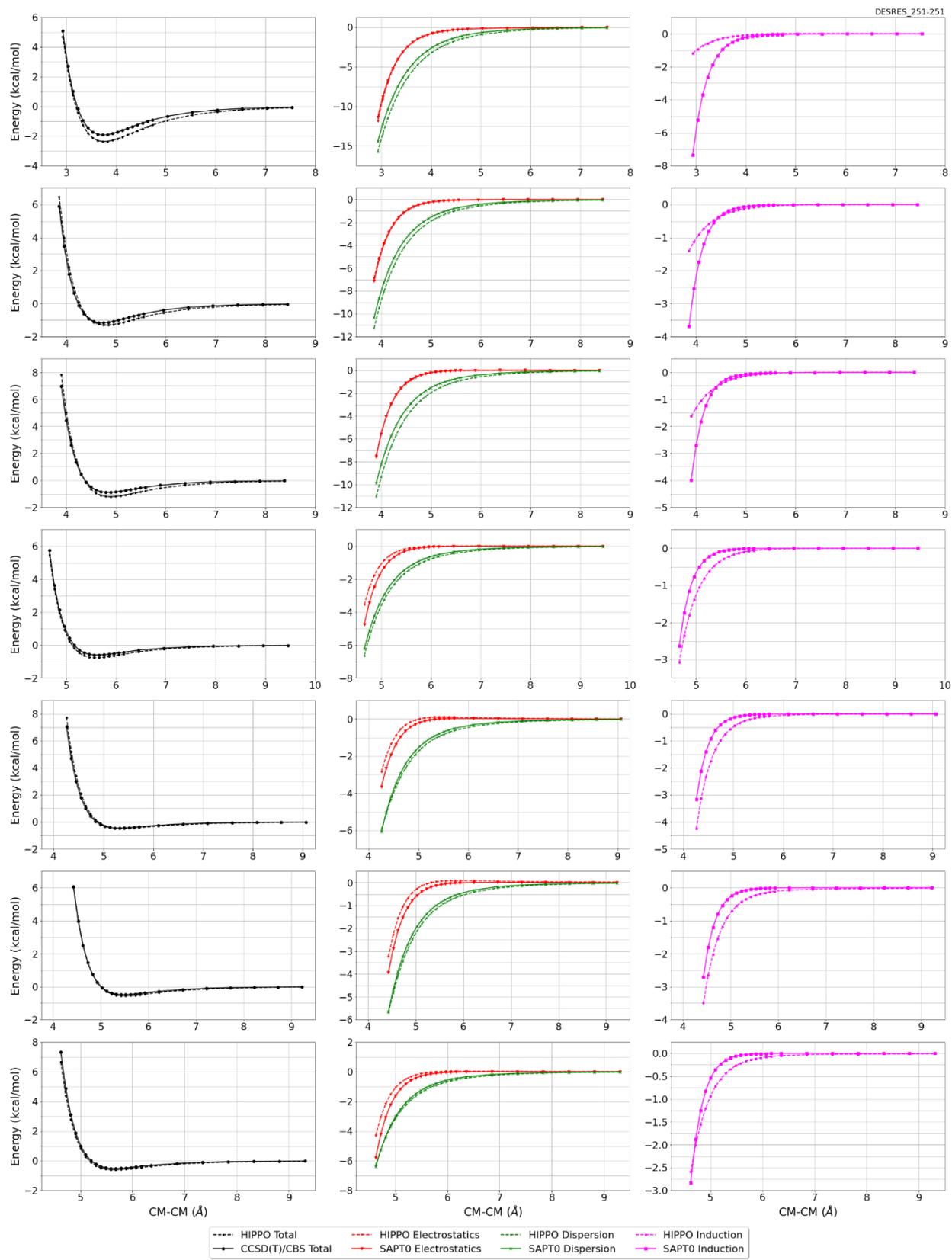
##Dimer results - Fitting to QM datasets##

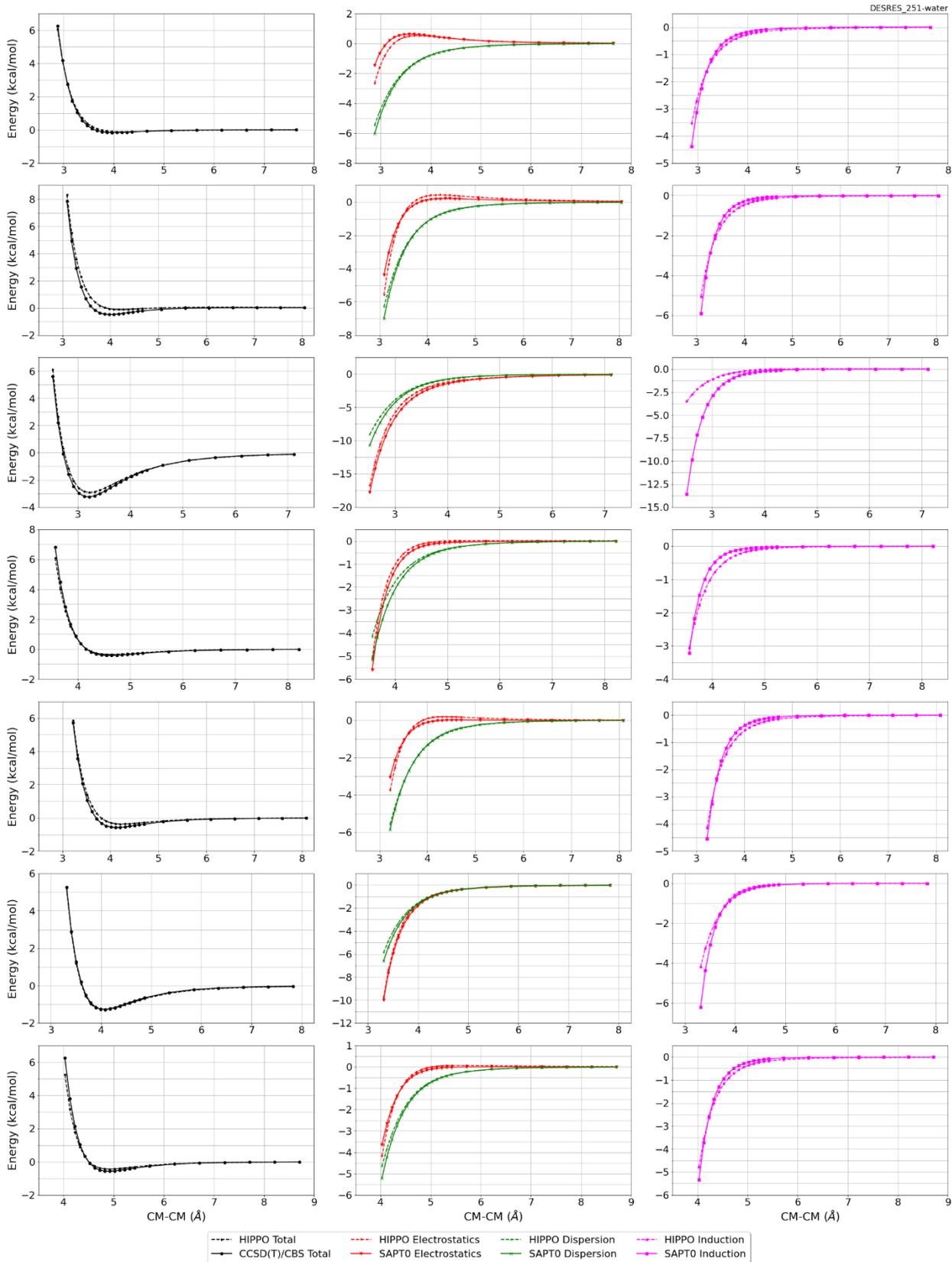
DESRES_251-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.124	0.169	1.5976	559	3

DESRES_251-251, energy values in kcal/mol

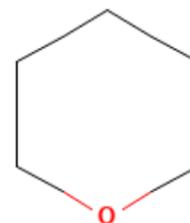
MAE	Std error	max error	#points	#count[err > 1]
0.170	0.217	1.3078	814	11





#255 Tetrahydropyran C5H10O CID: 8894

ref molpol	8.47	9.33	10.03, avg	9.27
molpol	8.43	9.34	10.06, avg	9.28
rms molpol	0.04	0.01	0.04, avg	0.00



Monomer potential fitting RMS: 0.41

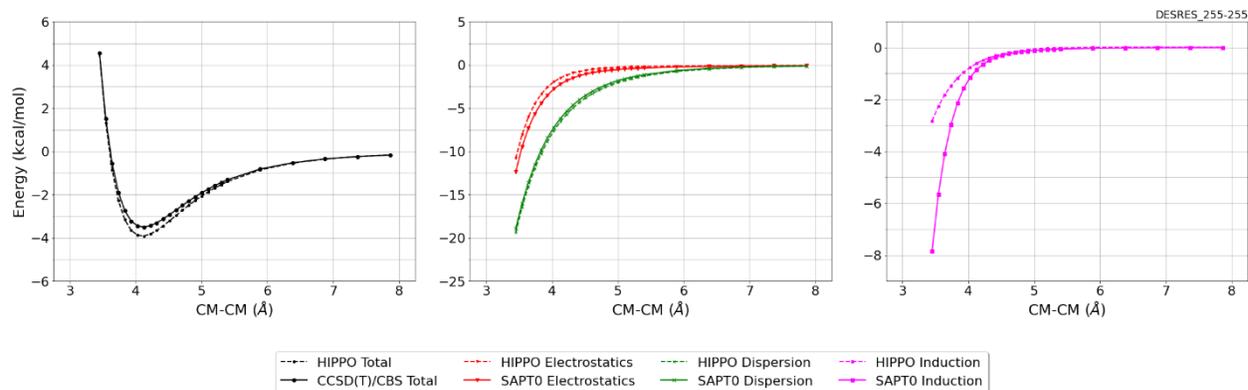
##Dimer results - Fitting to QM datasets##

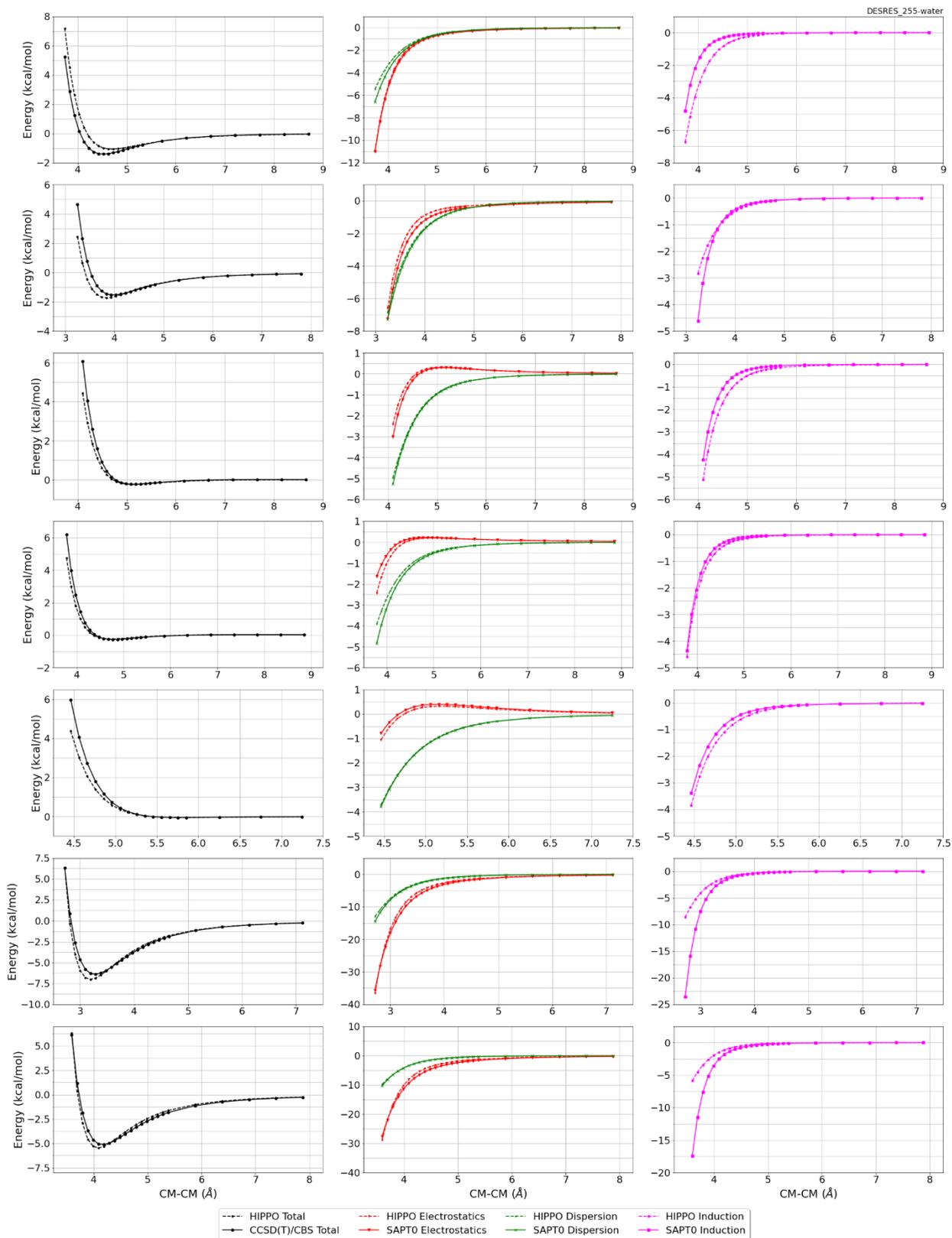
DESRES_255-255, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.120	0.138	0.5610	176	0

DESRES_255-water, energy values in kcal/mol

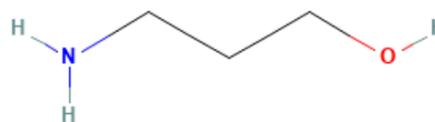
MAE	Std error	max error	#points	#count[err > 1]
0.204	0.363	2.5899	555	29





#256 3-Aminopropan-1-ol C3H9NO CID: 9086

ref molpol	9.36	7.61	7.02, avg	8.00
molpol	9.37	7.58	7.06, avg	8.00
rms molpol	0.01	0.04	0.04, avg	0.00

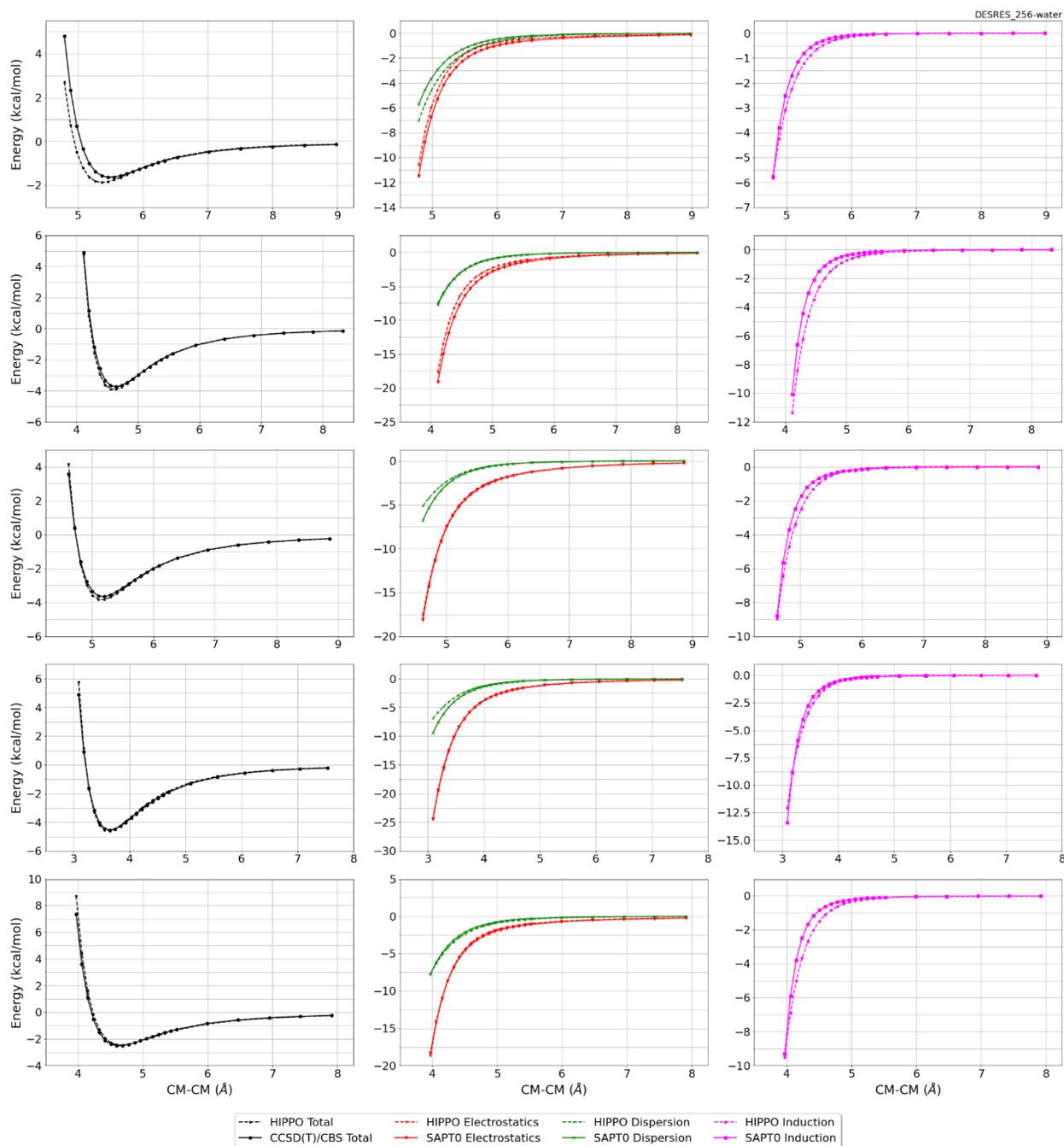


Monomer potential fitting RMS: 0.58

##Dimer results - Fitting to QM datasets##

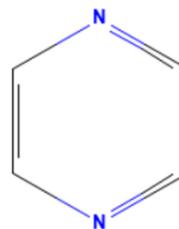
DESRES_256-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.187	0.289	2.1128	265	7



#257 Pyrazine C4H4N2 CID: 9261

ref molpol	10.88	9.56	0.00, avg	6.81
molpol	10.82	9.58	5.91, avg	8.77
rms molpol	0.06	0.02	5.91, avg	1.96



Monomer potential fitting RMS: 0.14

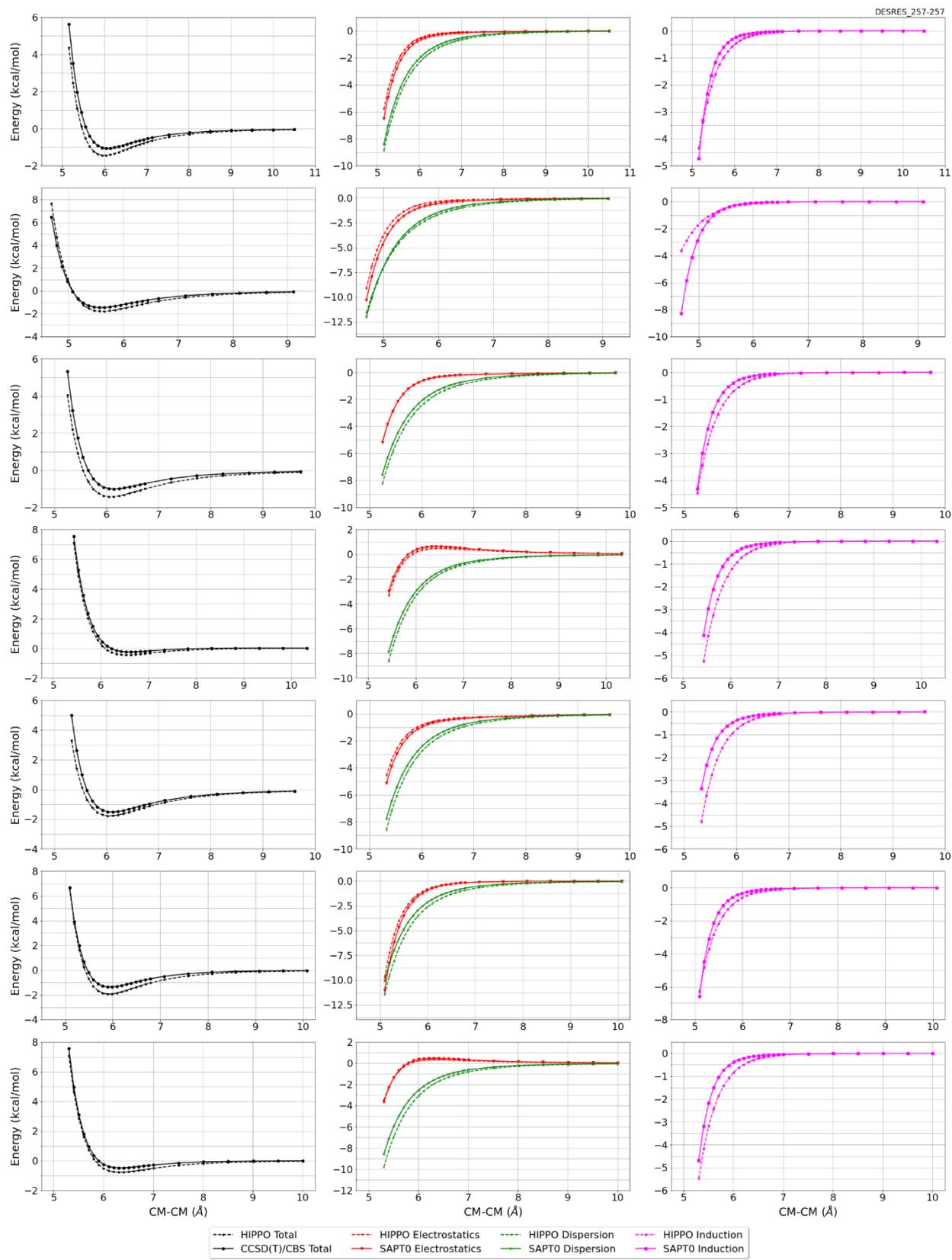
##Dimer results - Fitting to QM datasets##

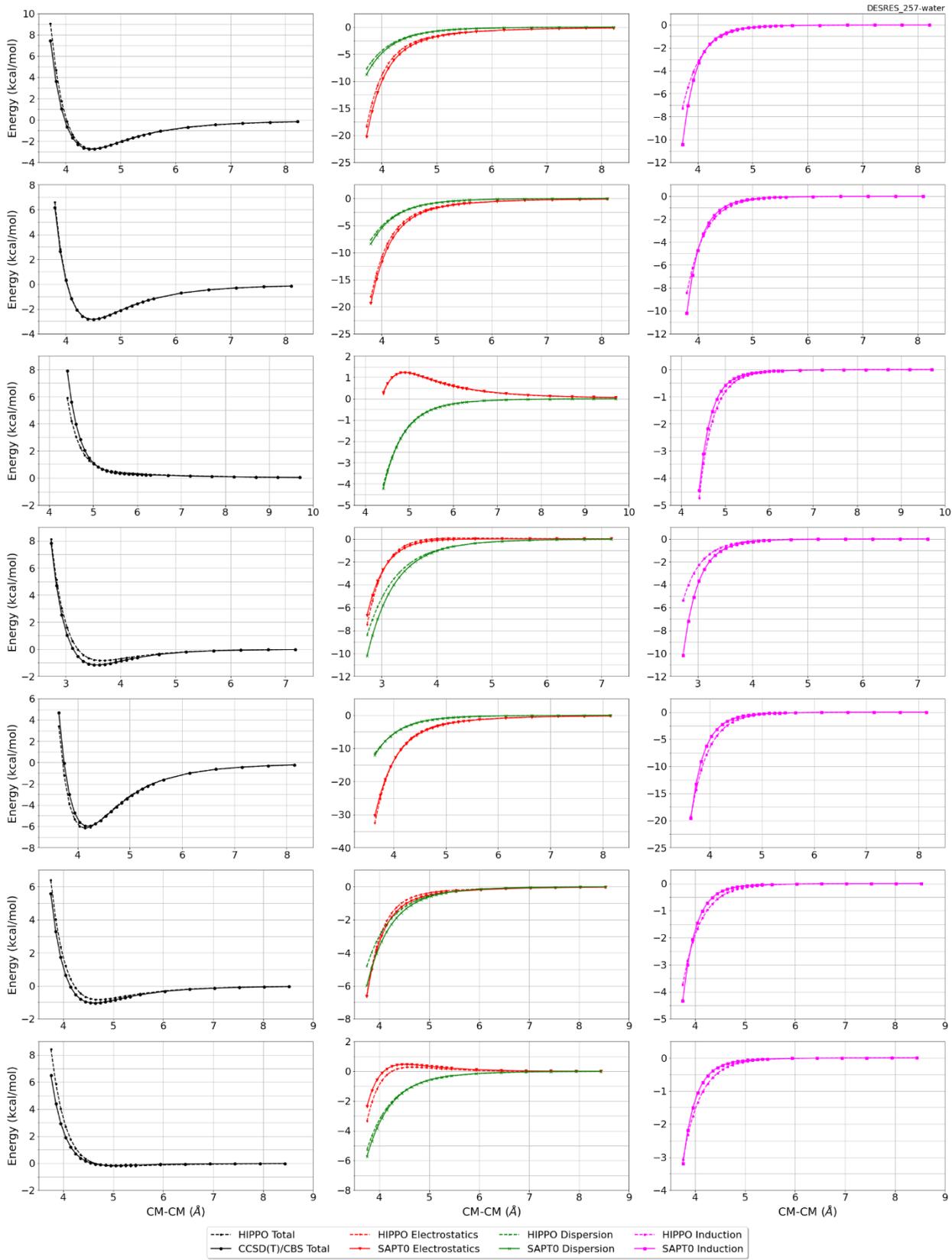
DESRES_257-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.226	0.404	3.5802	567	29

DESRES_257-257, energy values in kcal/mol

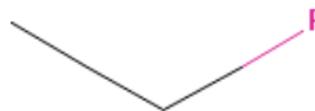
MAE	Std error	max error	#points	#count[err > 1]
0.315	0.326	1.7692	548	27





#259 Fluoroethane C2H5F CID: 9620

ref molpol	4.19	4.50	0.00, avg	2.90
molpol	4.20	3.70	3.52, avg	3.81
rms molpol	0.01	0.80	3.52, avg	0.91



Monomer potential fitting RMS: 0.09

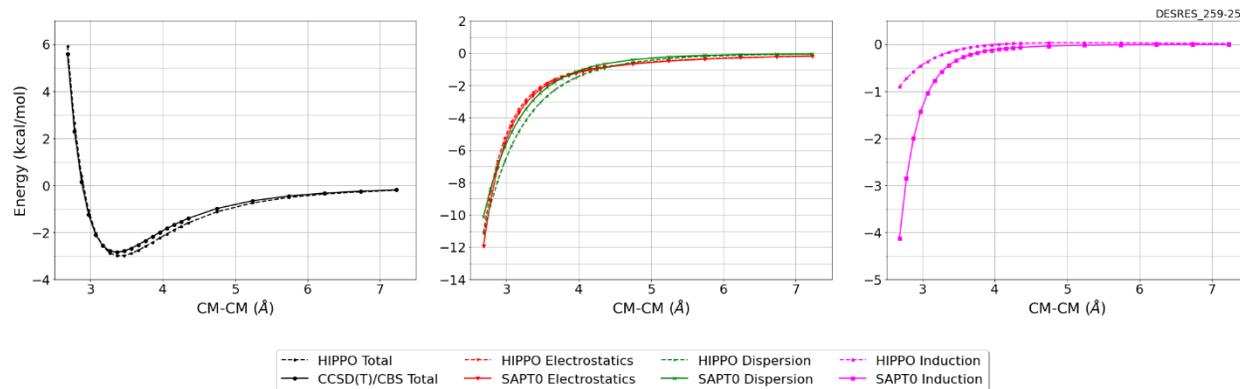
##Dimer results - Fitting to QM datasets##

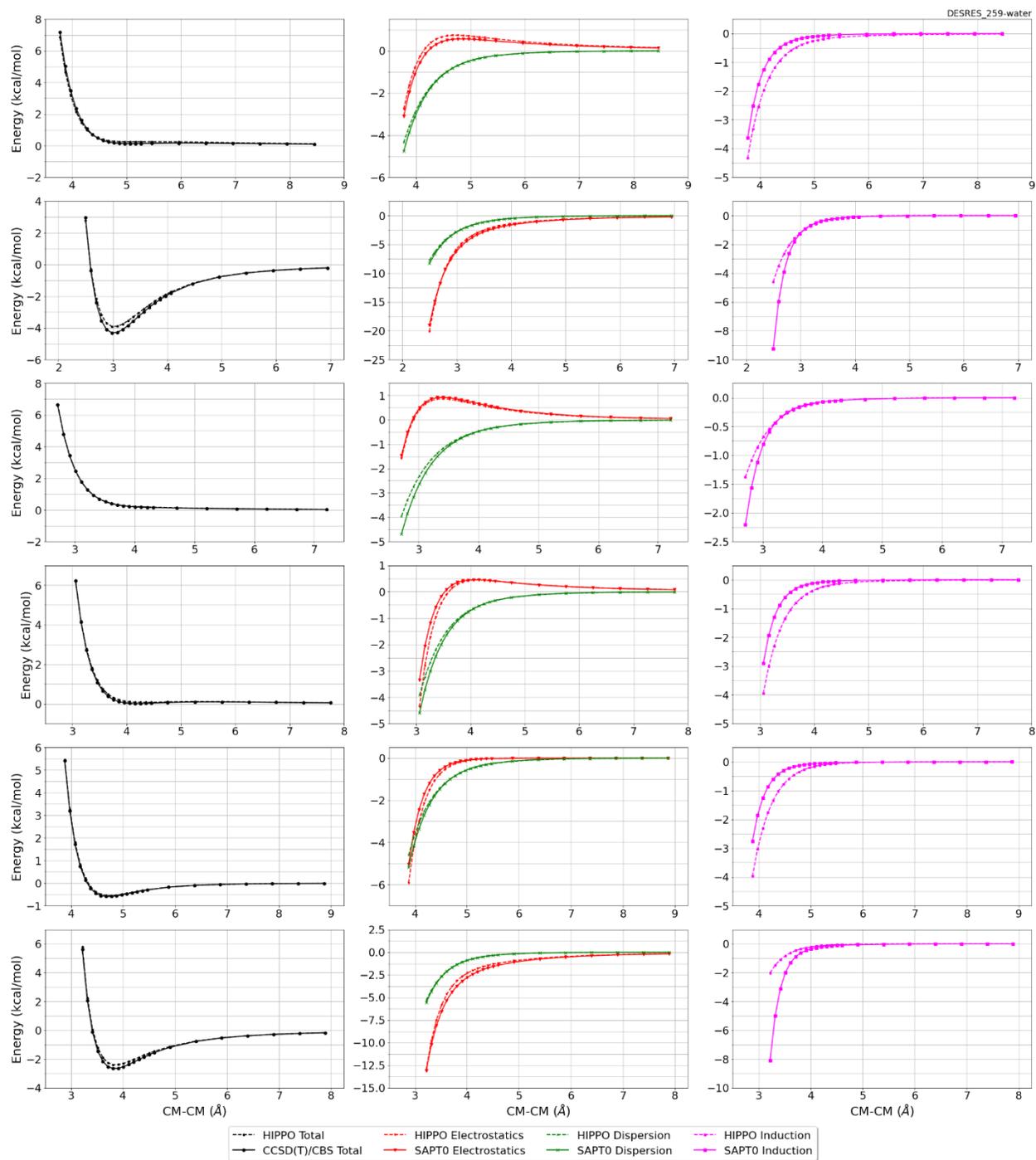
DESRES_259-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.076	0.088	0.7435	287	0

DESRES_259-259, energy values in kcal/mol

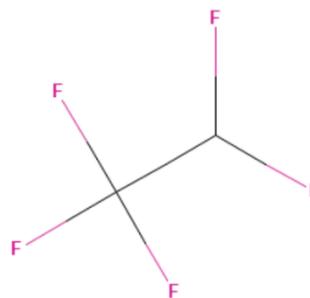
MAE	Std error	max error	#points	#count[err > 1]
0.186	0.097	0.3075	24	0





#260 Pentafluoroethane C2HF5 CID: 9633

ref molpol	4.45	4.70	4.64, avg	4.59
molpol	4.43	4.58	4.81, avg	4.61
rms molpol	0.02	0.11	0.17, avg	0.01

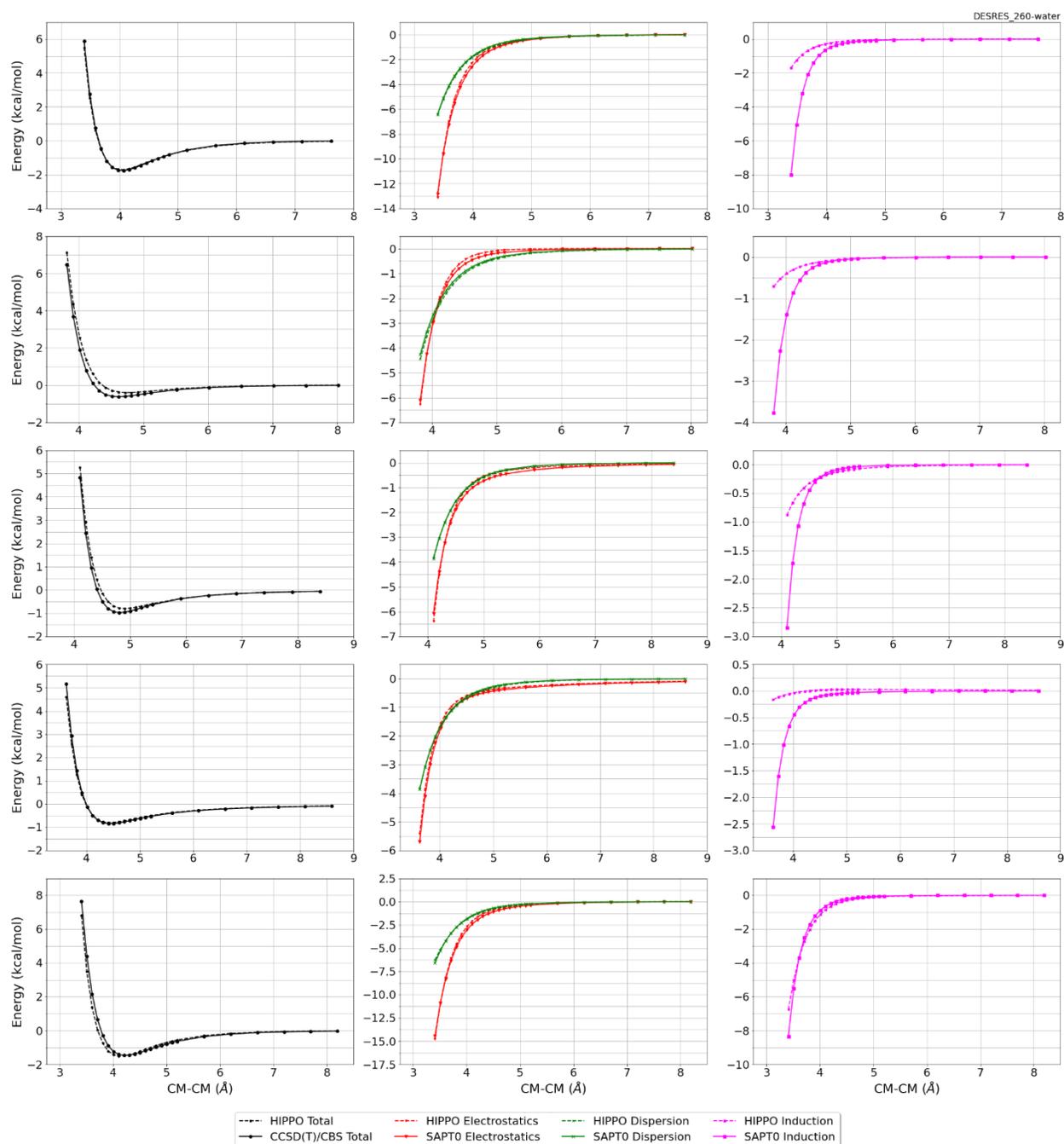


Monomer potential fitting RMS: 0.43

##Dimer results - Fitting to QM datasets##

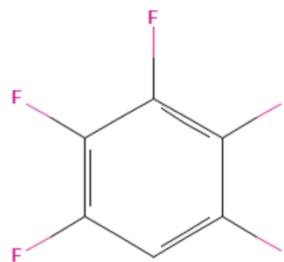
DESRES_260-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.117	0.178	0.8413	261	0



#261 Pentafluorobenzene C6HF5 CID: 9696

ref molpol	12.29	12.61	0.00, avg	8.30
molpol	12.38	12.38	6.92, avg	10.56
rms molpol	0.09	0.23	6.92, avg	2.26

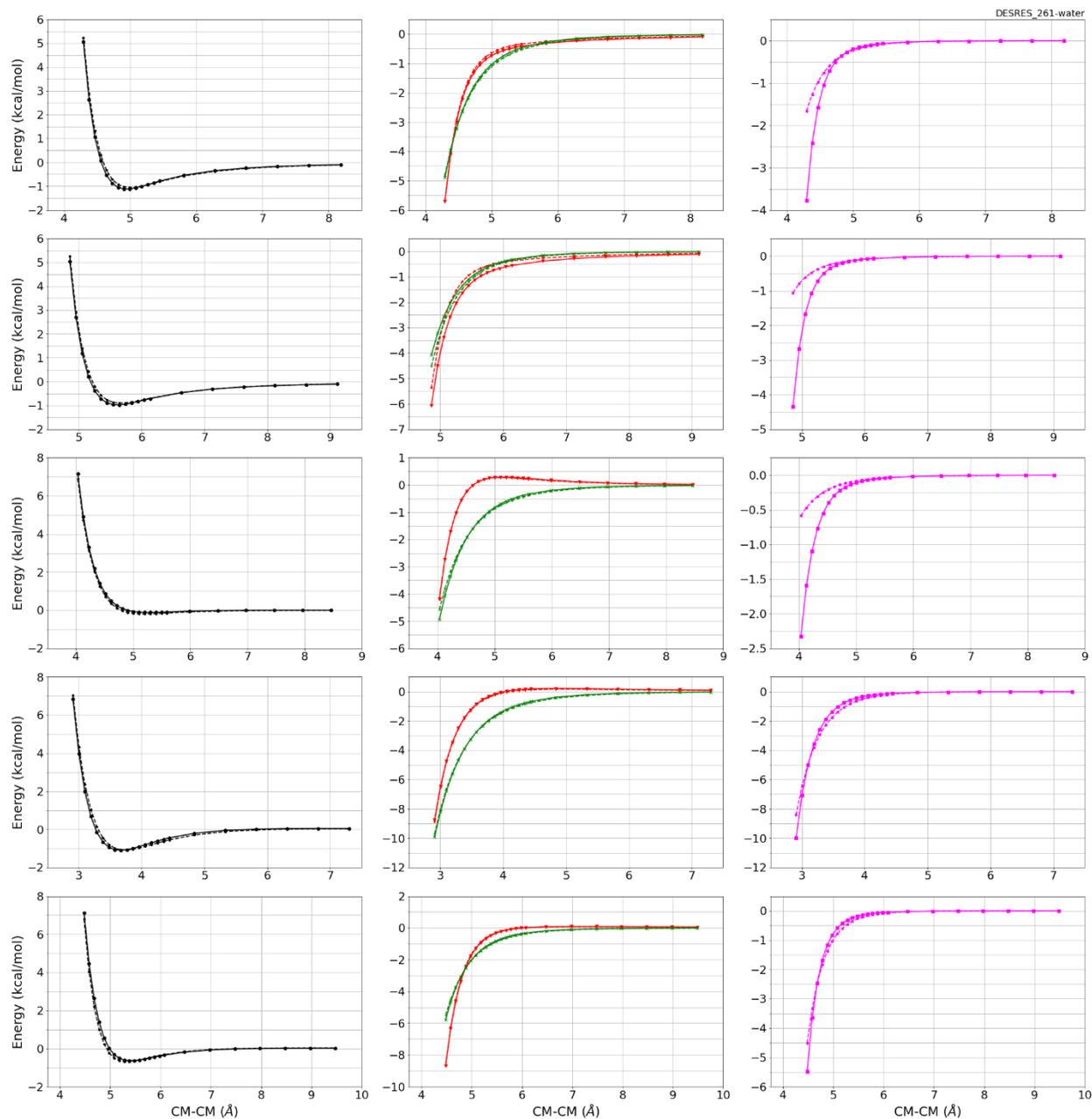


Monomer potential fitting RMS: 0.30

##Dimer results - Fitting to QM datasets##

DESRES_261-water, energy values in kcal/mol

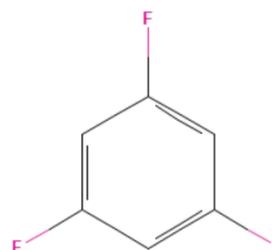
MAE	Std error	max error	#points	#count[err > 1]
0.106	0.109	0.4455	110	0



--- HIPPO Total - - - HIPPO Electrostatics - - - HIPPO Dispersion - - - HIPPO Induction
- - - CCSD(T)/CBS Total - - - SAPTO Electrostatics - - - SAPTO Dispersion - - - SAPTO Induction

#262 1,3,5-Trifluorobenzene C6H3F3 CID: 9745

ref molpol	12.13	12.13	0.00, avg	8.09
molpol	12.12	12.12	6.75, avg	10.33
rms molpol	0.02	0.02	6.75, avg	2.24



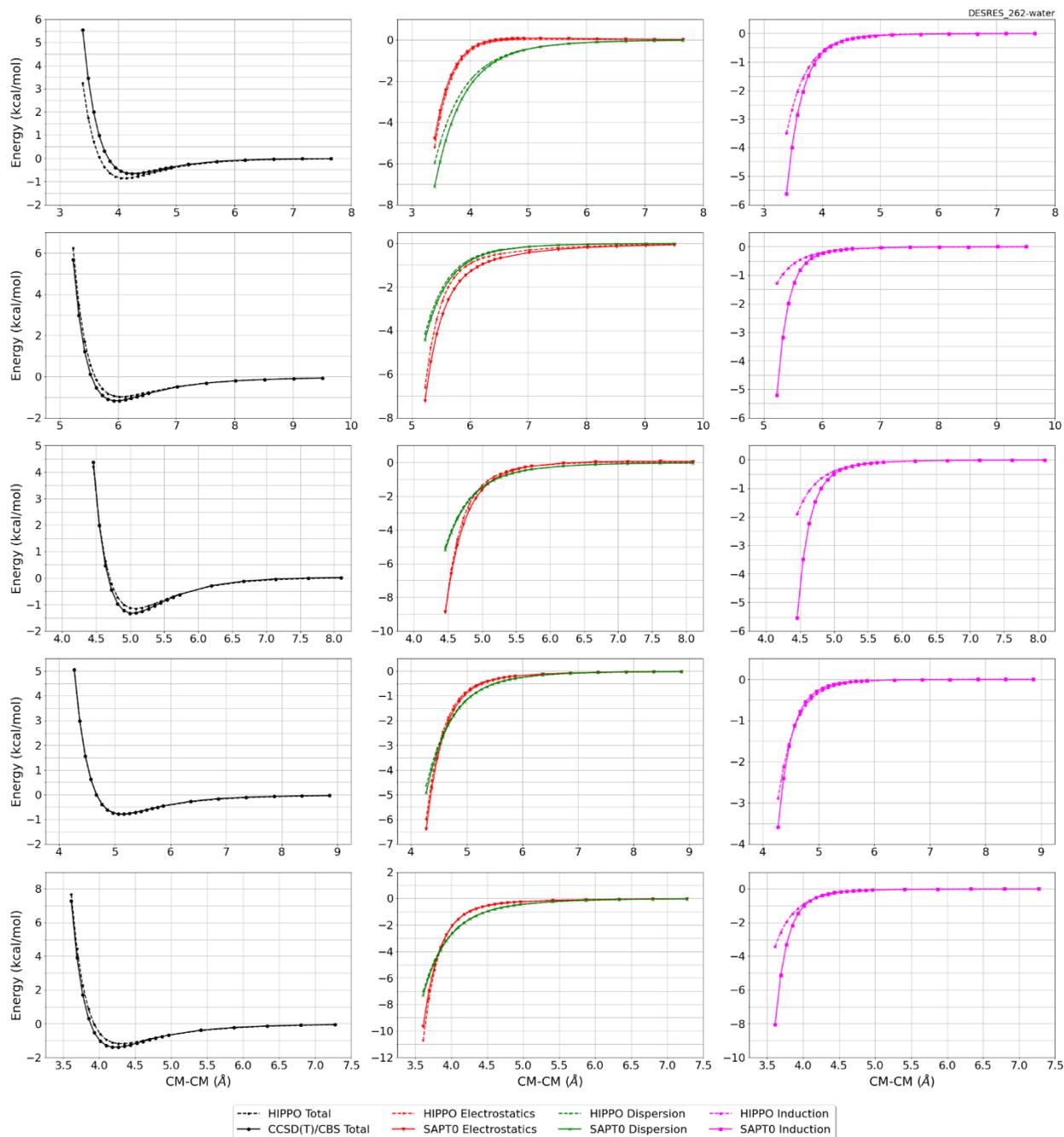
Monomer potential fitting RMS: 0.07

##Dimer results - Fitting to QM datasets##
DESRES_262-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.193	0.337	2.3107	108	3

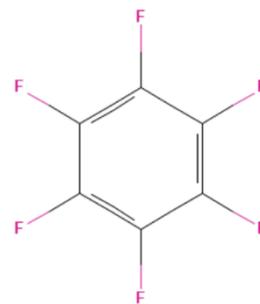
DESRES_262-262, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.808	0.437	1.3166	24	12



#263 Hexafluorobenzene C6F6 CID: 9805

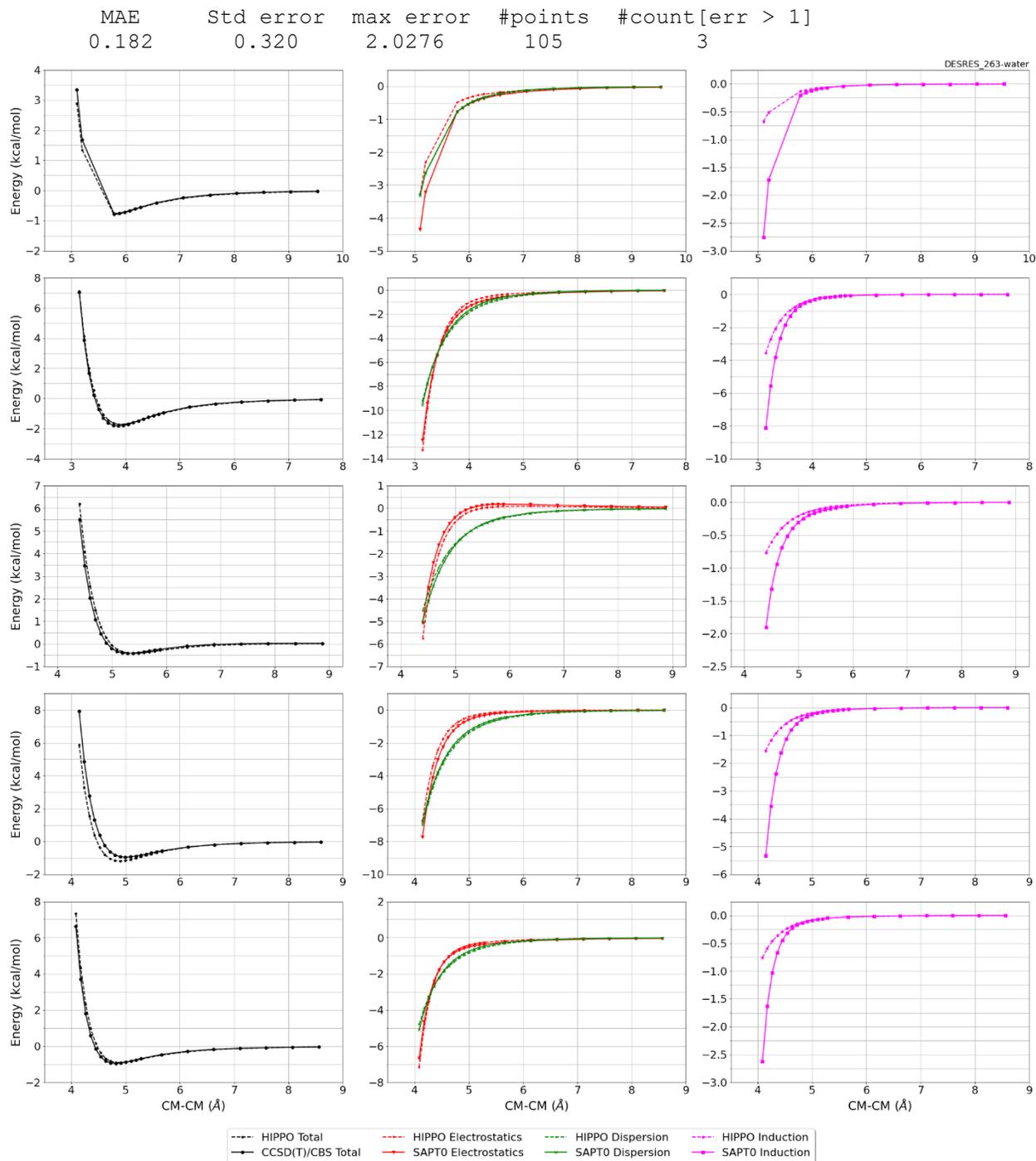
ref molpol	12.64	12.64	0.00, avg	8.43
molpol	12.60	12.60	6.96, avg	10.72
rms molpol	0.03	0.03	6.96, avg	2.30



Monomer potential fitting RMS: 0.32

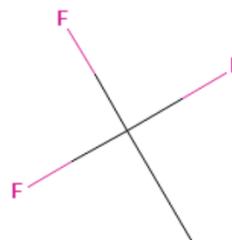
##Dimer results - Fitting to QM datasets##

DESRES_263-water, energy values in kcal/mol



#264 1,1,1-Trifluoroethane C2H3F3 CID: 9868

ref molpol	4.28	4.48	0.00, avg	2.92
molpol	4.27	4.04	4.04, avg	4.12
rms molpol	0.01	0.44	4.04, avg	1.20



Monomer potential fitting RMS: 0.15

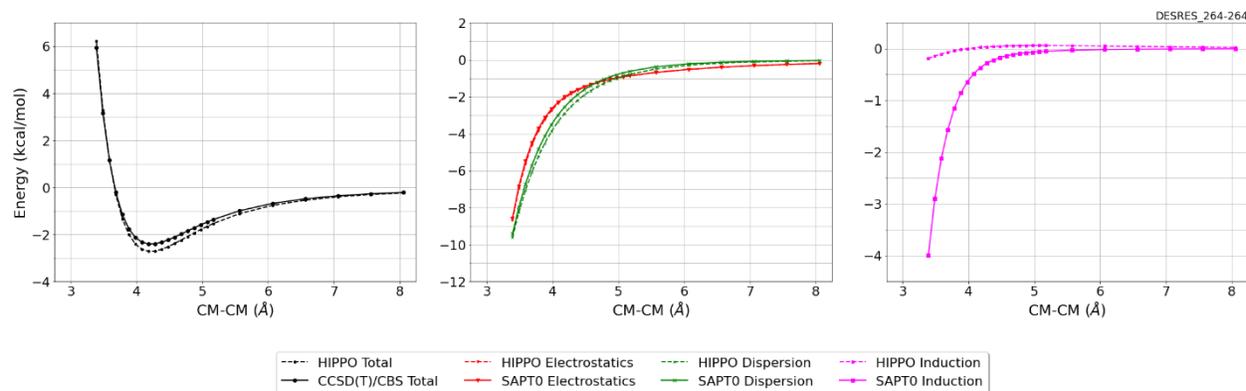
##Dimer results - Fitting to QM datasets##

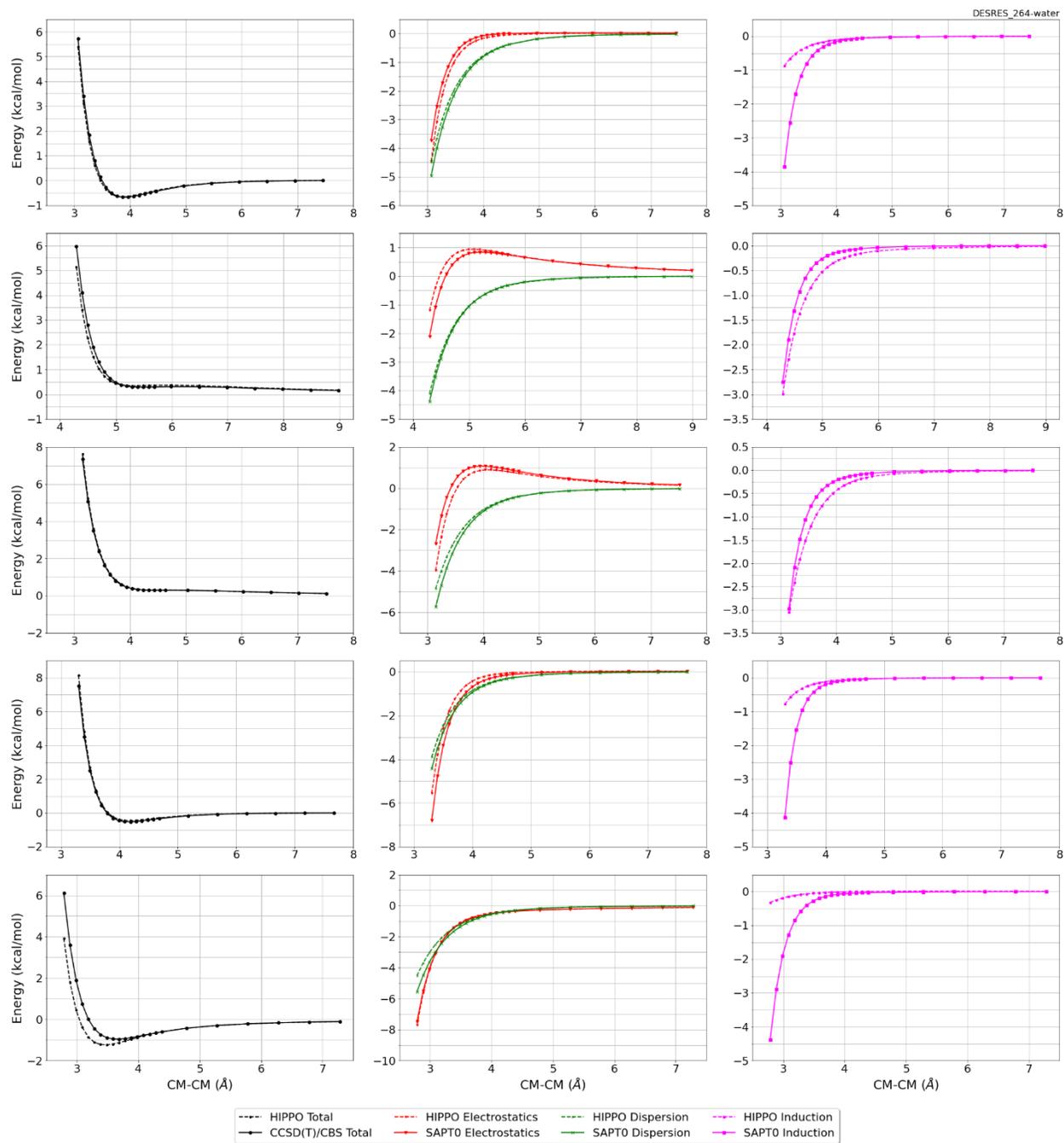
DESRES_264-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.121	0.244	2.2156	259	4

DESRES_264-264, energy values in kcal/mol

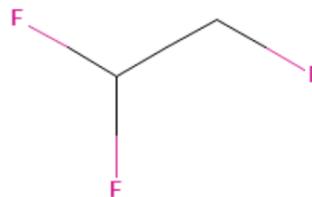
MAE	Std error	max error	#points	#count[err > 1]
0.241	0.130	0.4094	25	0





#265 1,1,2-Trifluoroethane C2H3F3 CID: 9890

ref molpol	4.12	4.59	4.39, avg	4.37
molpol	4.14	4.34	4.60, avg	4.36
rms molpol	0.02	0.25	0.22, avg	0.01

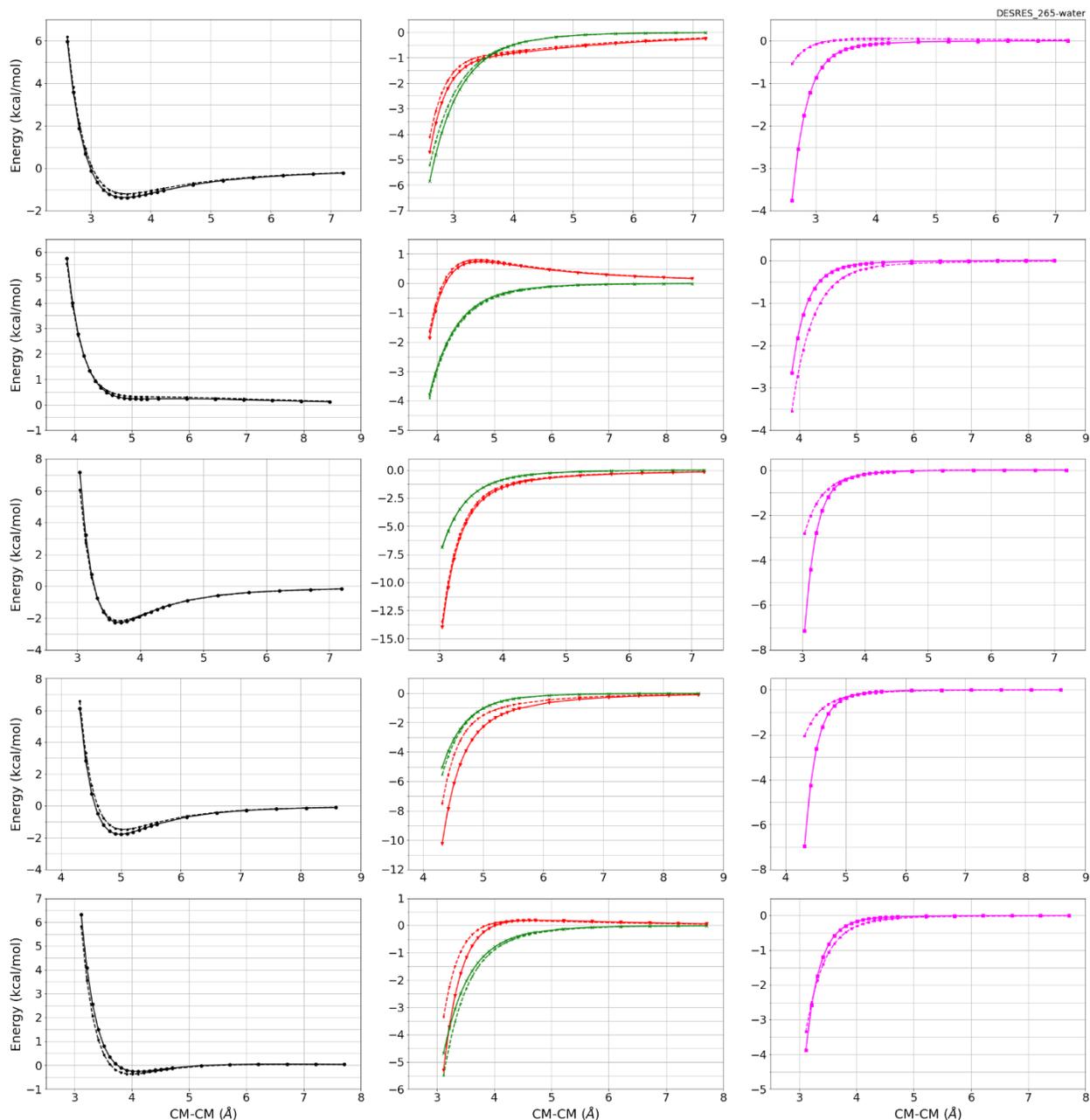


Monomer potential fitting RMS: 0.39

##Dimer results - Fitting to QM datasets##

DESRES_265-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.114	0.145	1.0760	260	1



--- HIPPO Total --- HIPPO Electrostatics --- HIPPO Dispersion --- HIPPO Induction
--- CCSD(T)/CBS Total --- SAPTO Electrostatics --- SAPTO Dispersion --- SAPTO Induction

#266 1-Fluoropropane C3H7F CID: 9998

ref molpol	5.78	6.61	0.00, avg	4.13
molpol	5.80	5.02	4.76, avg	5.19
rms molpol	0.02	1.59	4.76, avg	1.06



Monomer potential fitting RMS: 0.16

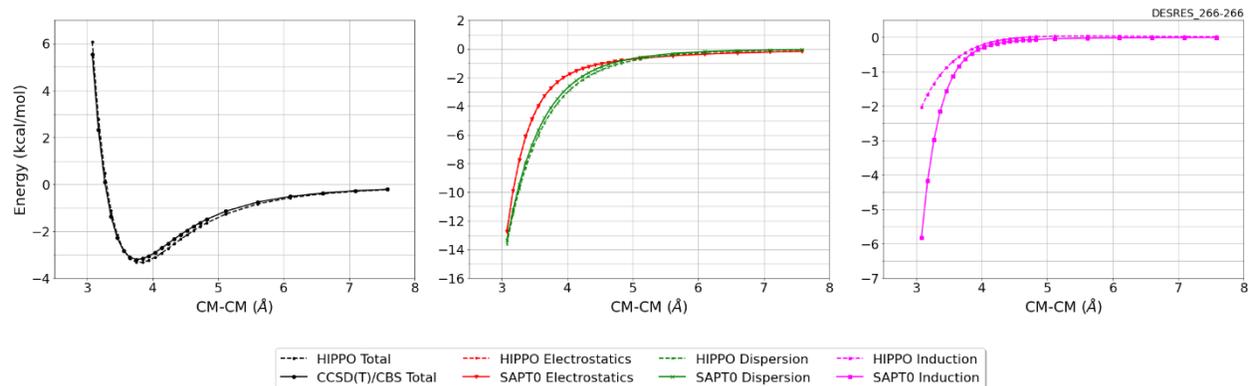
##Dimer results - Fitting to QM datasets##

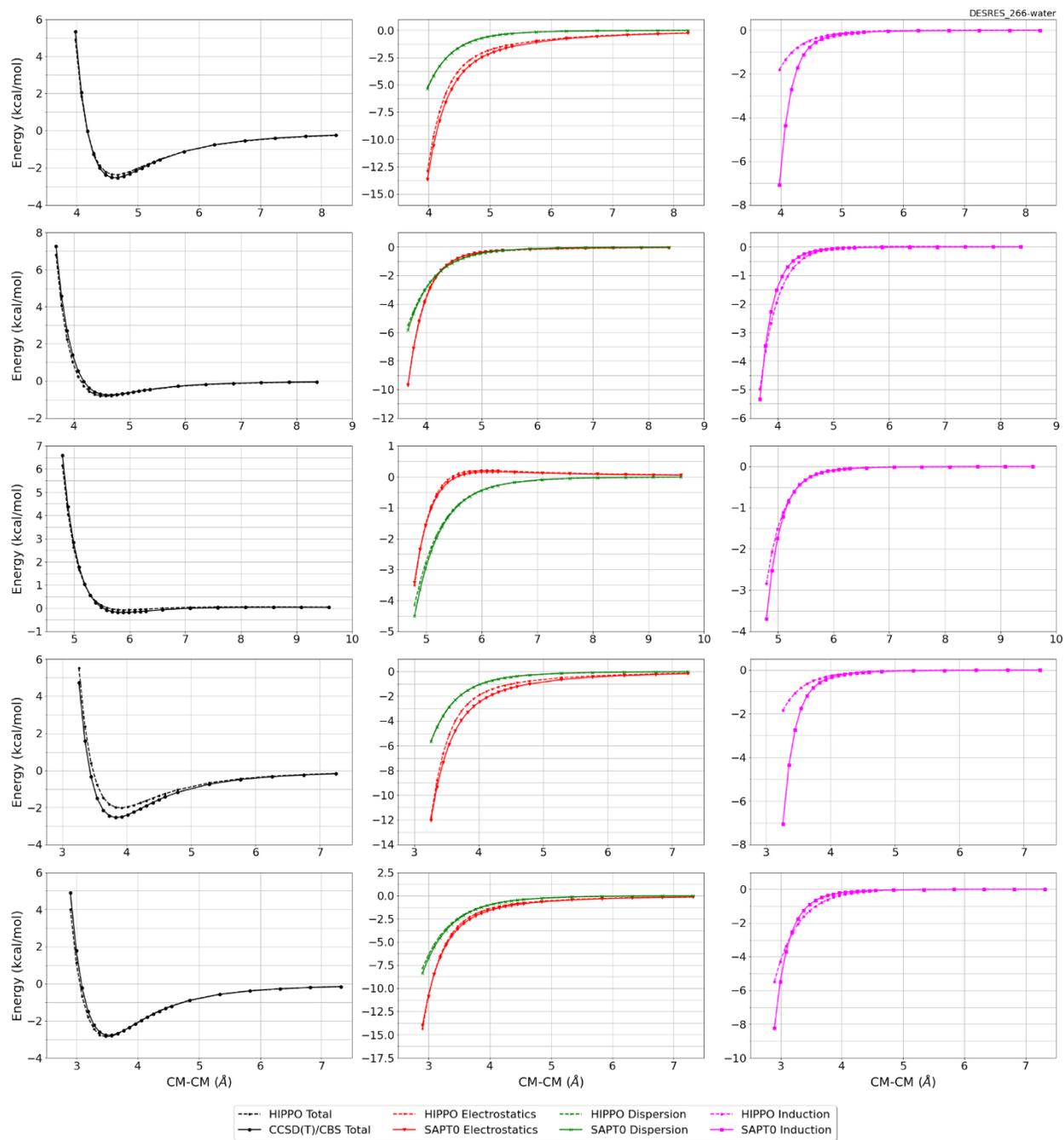
DESRES_266-266, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.188	0.096	0.3064	25	0

DESRES_266-water, energy values in kcal/mol

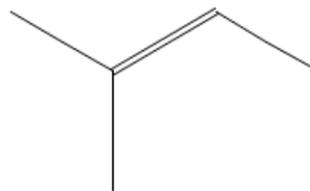
MAE	Std error	max error	#points	#count[err > 1]
0.140	0.203	1.4256	263	1





#279 2-Methylbut-2-ene C5H10 CID: 10553

ref molpol	11.13	9.36	0.00, avg	6.83
molpol	11.10	9.35	6.79, avg	9.08
rms molpol	0.03	0.01	6.79, avg	2.25



Monomer potential fitting RMS: 0.35

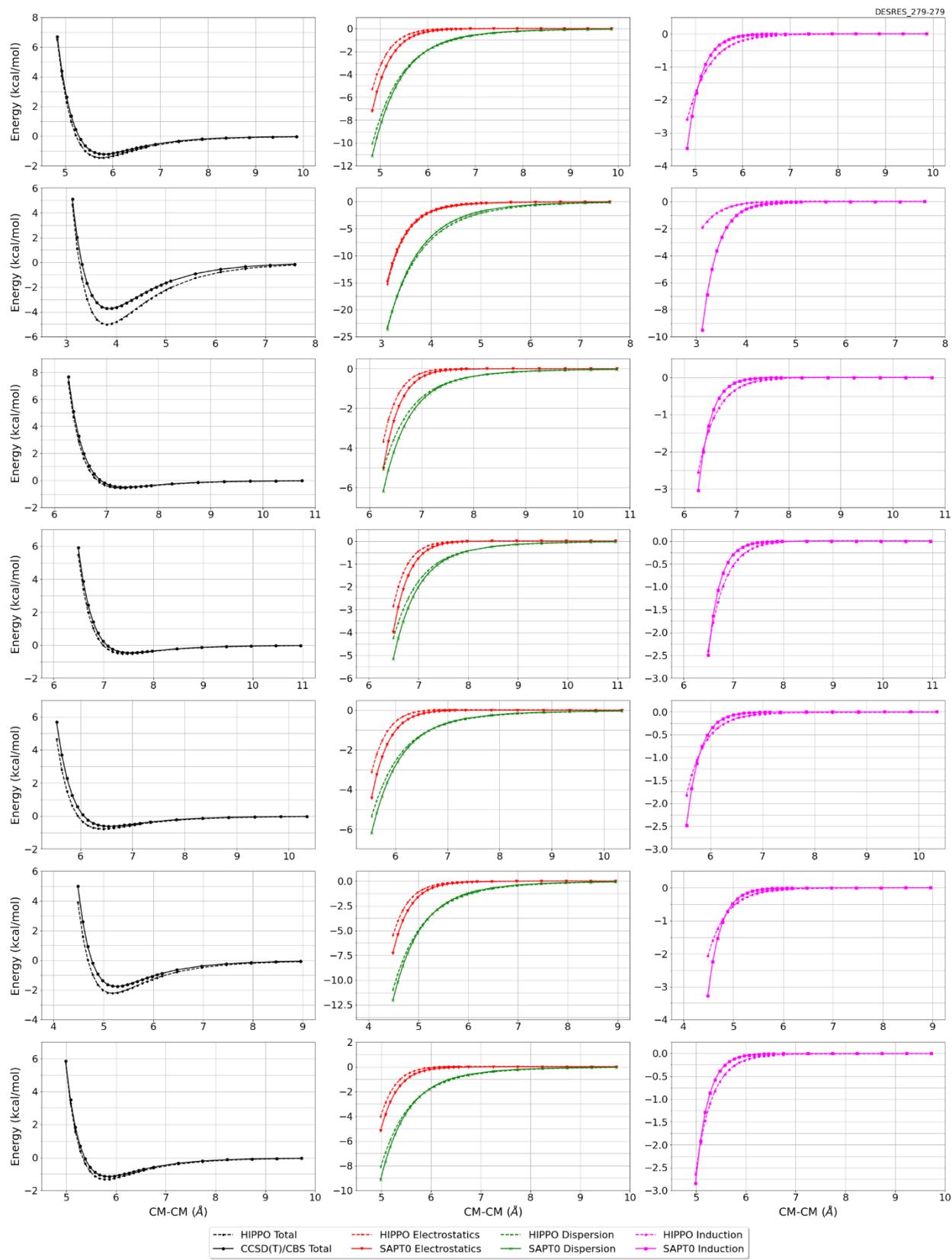
##Dimer results - Fitting to QM datasets##

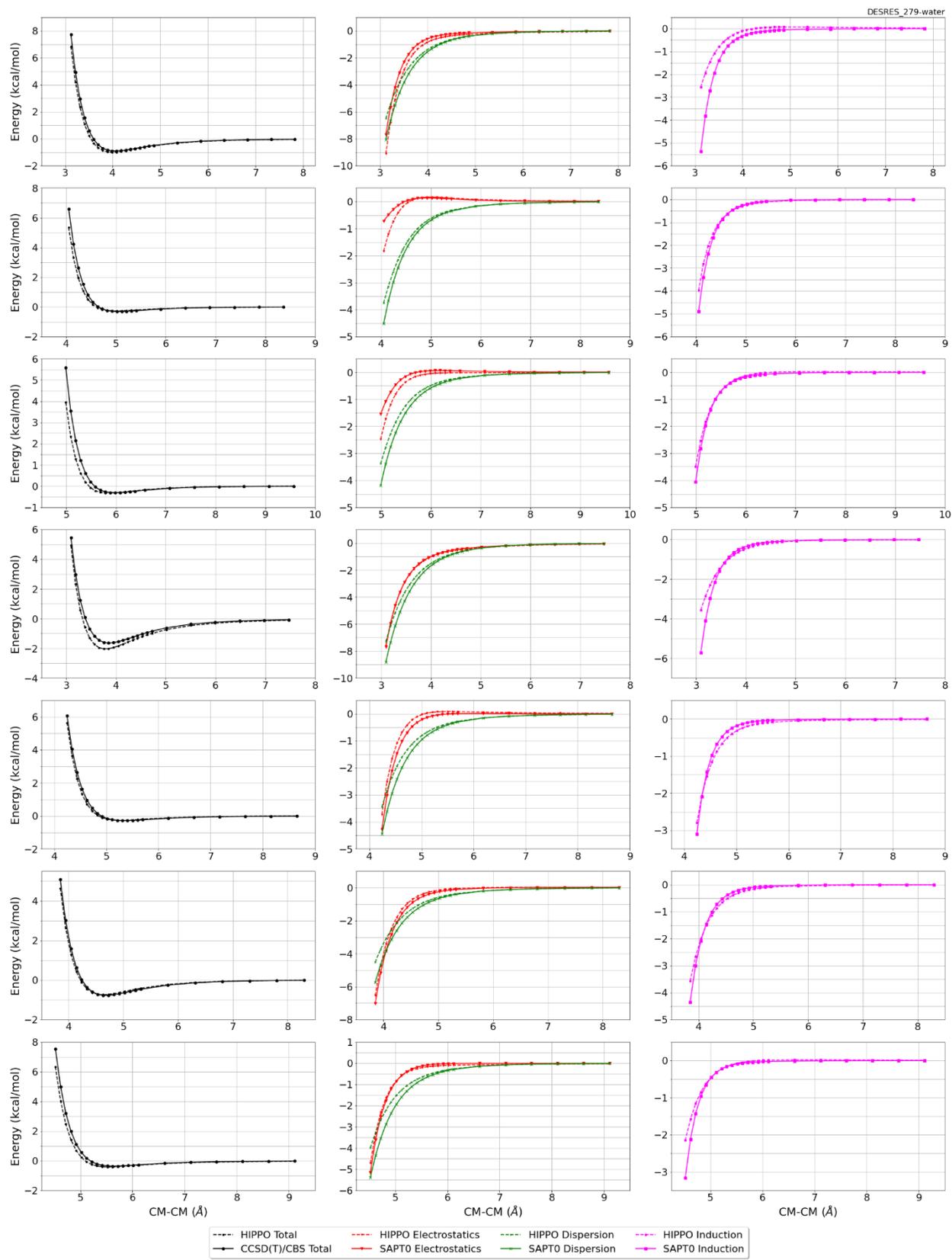
DESRES_279-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.158	0.240	1.7633	555	13

DESRES_279-279, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.248	0.298	1.5275	481	18





#281 1-Chloropropane C3H7Cl CID: 10899

ref molpol	7.17	9.80	0.00, avg	5.66
molpol	7.19	5.88	5.54, avg	6.20
rms molpol	0.02	3.92	5.54, avg	0.54



Monomer potential fitting RMS: 0.19

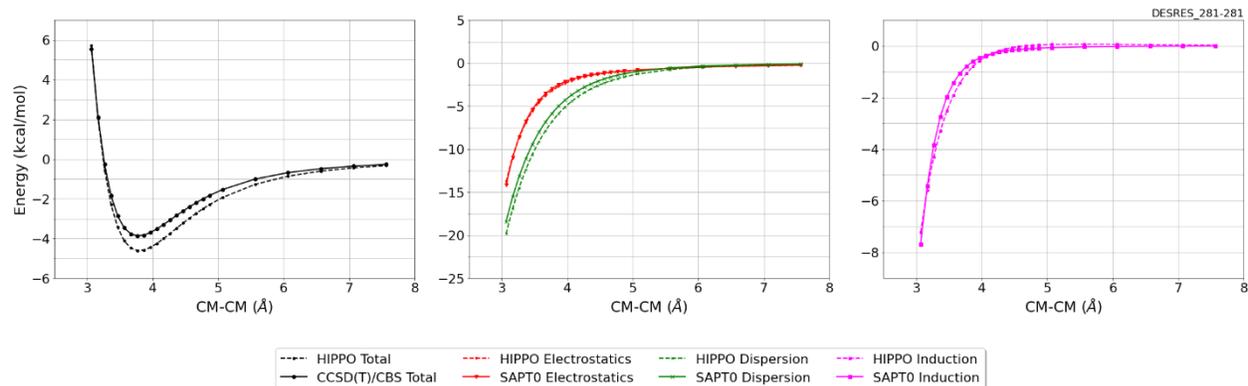
##Dimer results - Fitting to QM datasets##

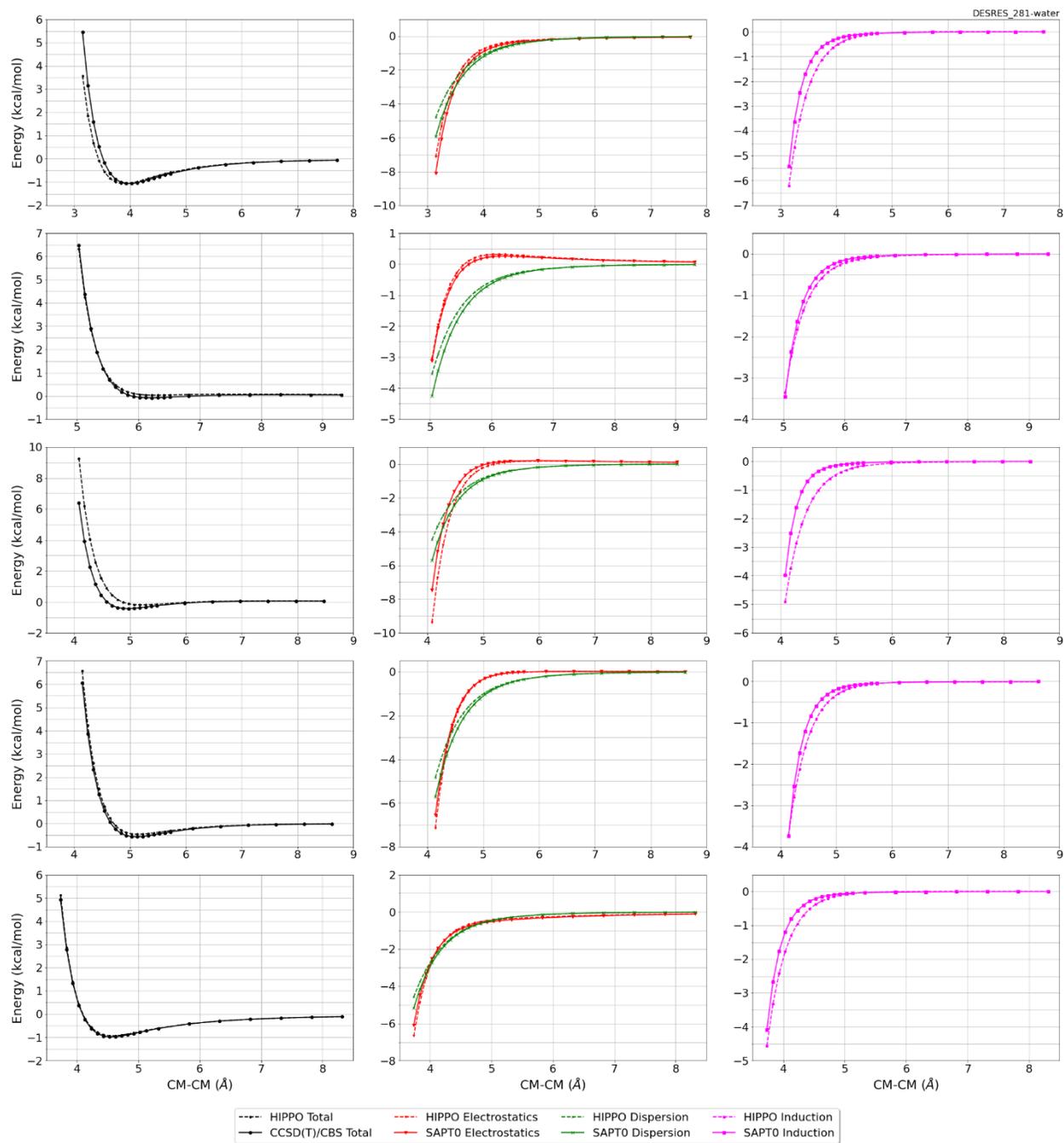
DESRES_281-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.169	0.344	2.9245	261	9

DESRES_281-281, energy values in kcal/mol

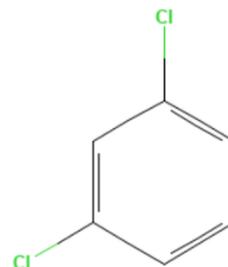
MAE	Std error	max error	#points	#count[err > 1]
0.620	0.282	0.9529	25	0





#284 1,3-Dichlorobenzene C6H4Cl2 CID: 10943

ref molpol	15.47	19.20	0.00, avg	11.56
molpol	15.49	15.49	8.07, avg	13.02
rms molpol	0.02	3.71	8.07, avg	1.46

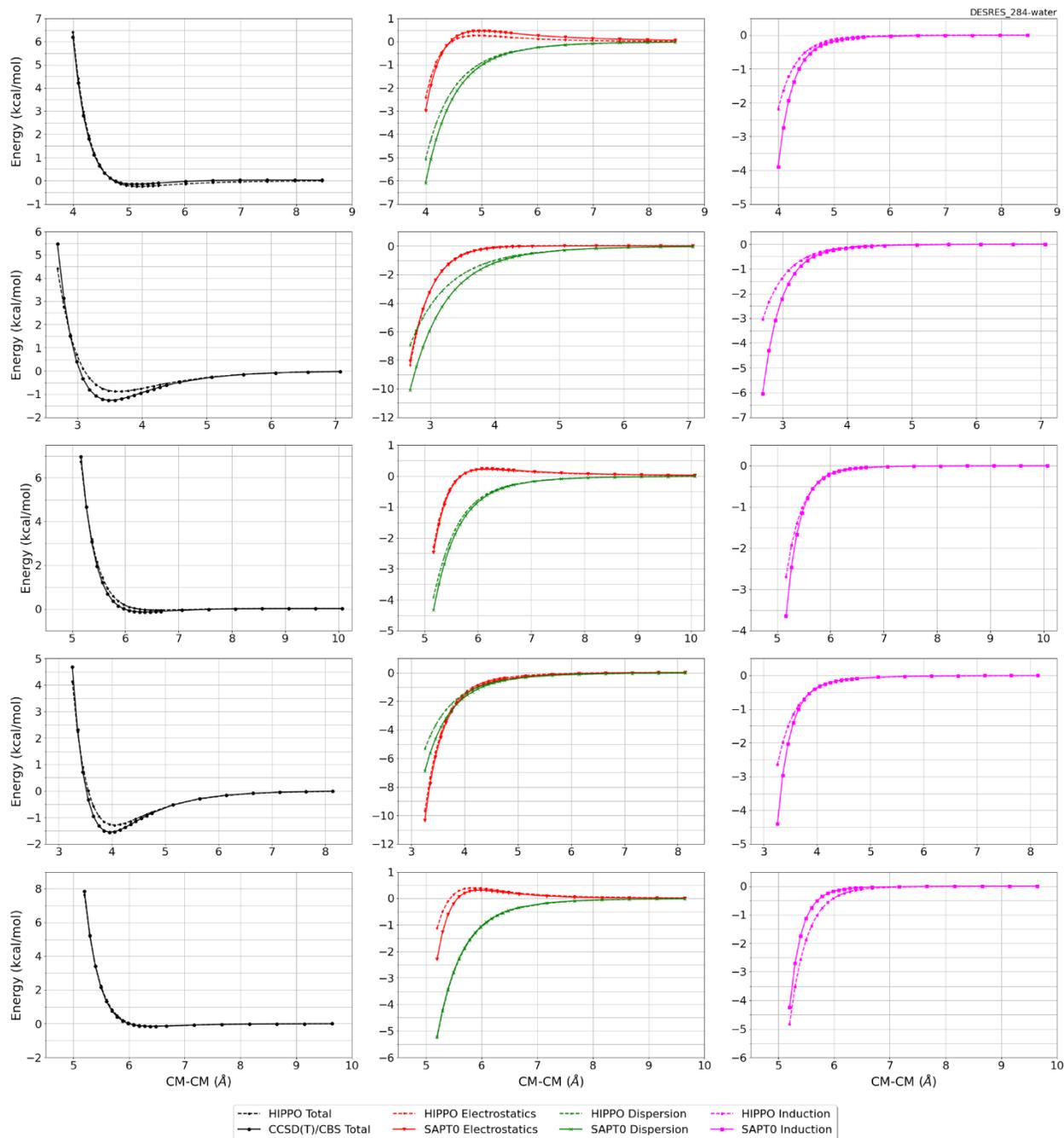


Monomer potential fitting RMS: 0.34

##Dimer results - Fitting to QM datasets##

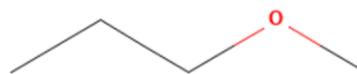
DESRES_284-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.136	0.155	1.0583	114	1



#287 1-Methoxypropane C4H10O CID: 11182

ref molpol	10.09	7.89	0.00, avg	5.99
molpol	10.10	7.21	6.78, avg	8.03
rms molpol	0.01	0.68	6.78, avg	2.04



Monomer potential fitting RMS: 0.37

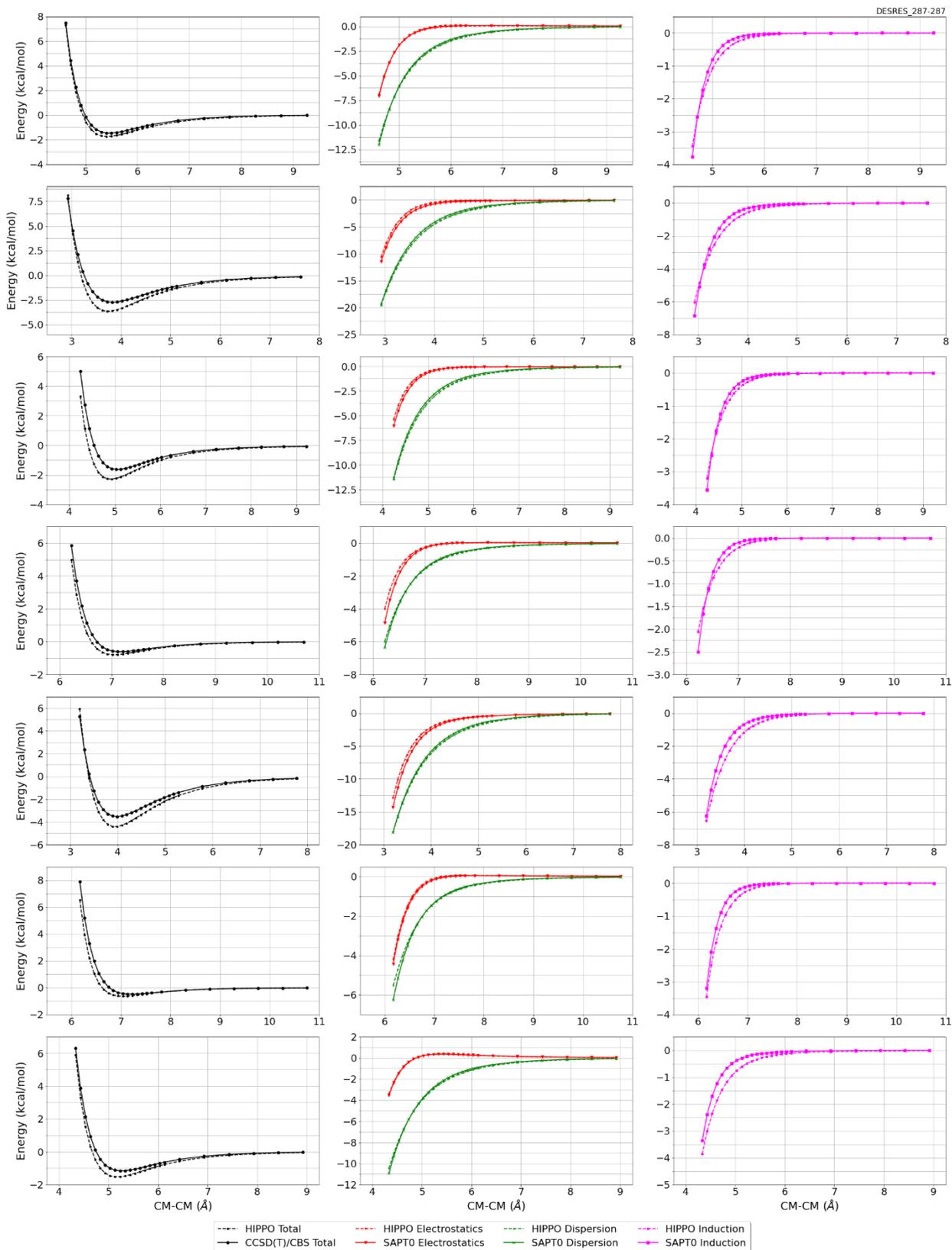
##Dimer results - Fitting to QM datasets##

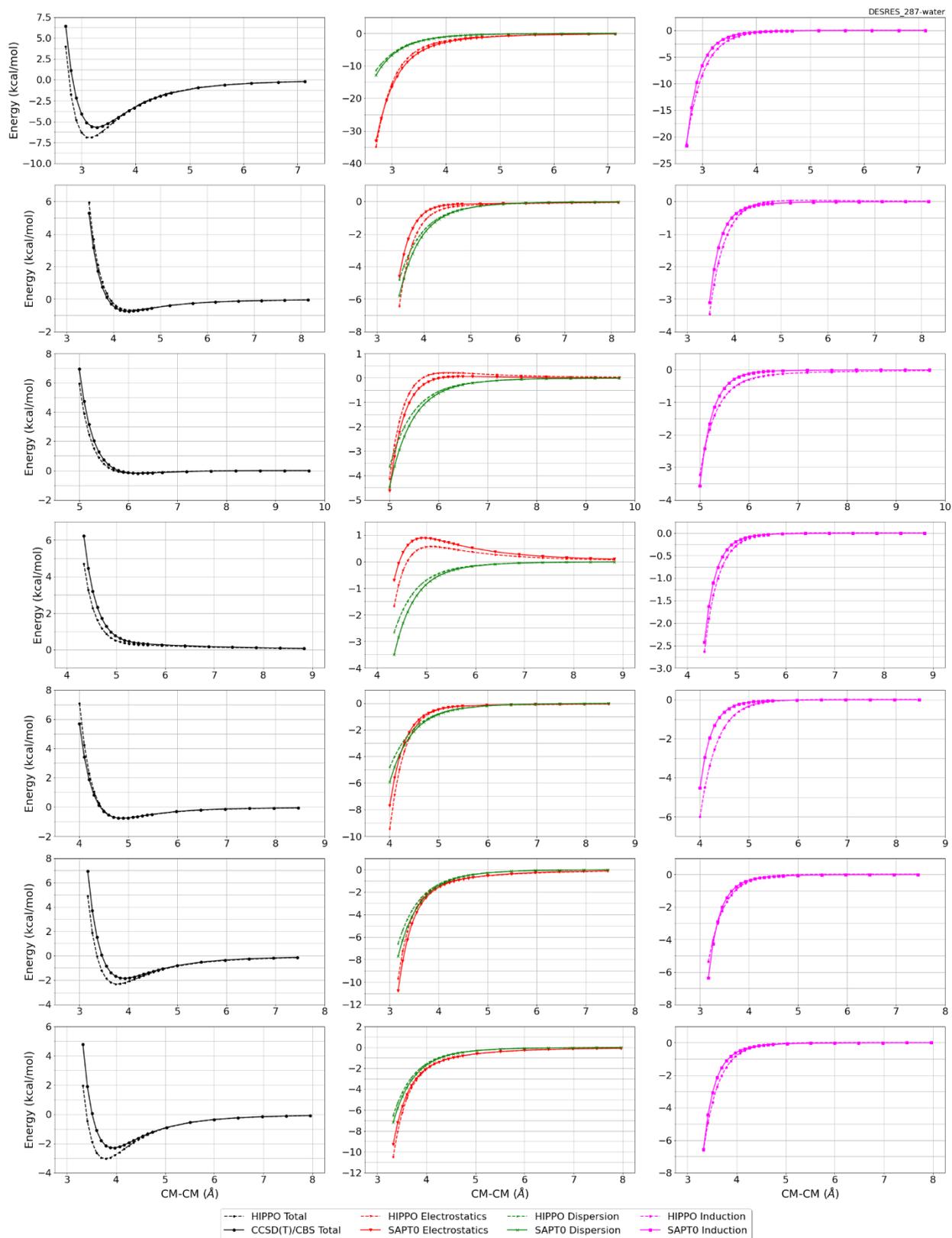
DESRES_287-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.267	0.554	3.9176	553	41

DESRES_287-287, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.316	0.338	1.8016	482	27





#290 2,3-Dimethylbut-2-ene C6H12 CID: 11250

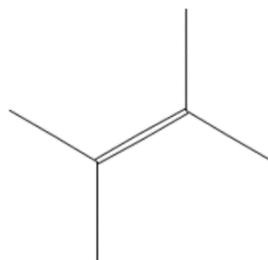
ref molpol	12.79	11.70	0.00, avg	8.16
molpol	12.63	11.83	7.84, avg	10.76
rms molpol	0.16	0.12	7.84, avg	2.60

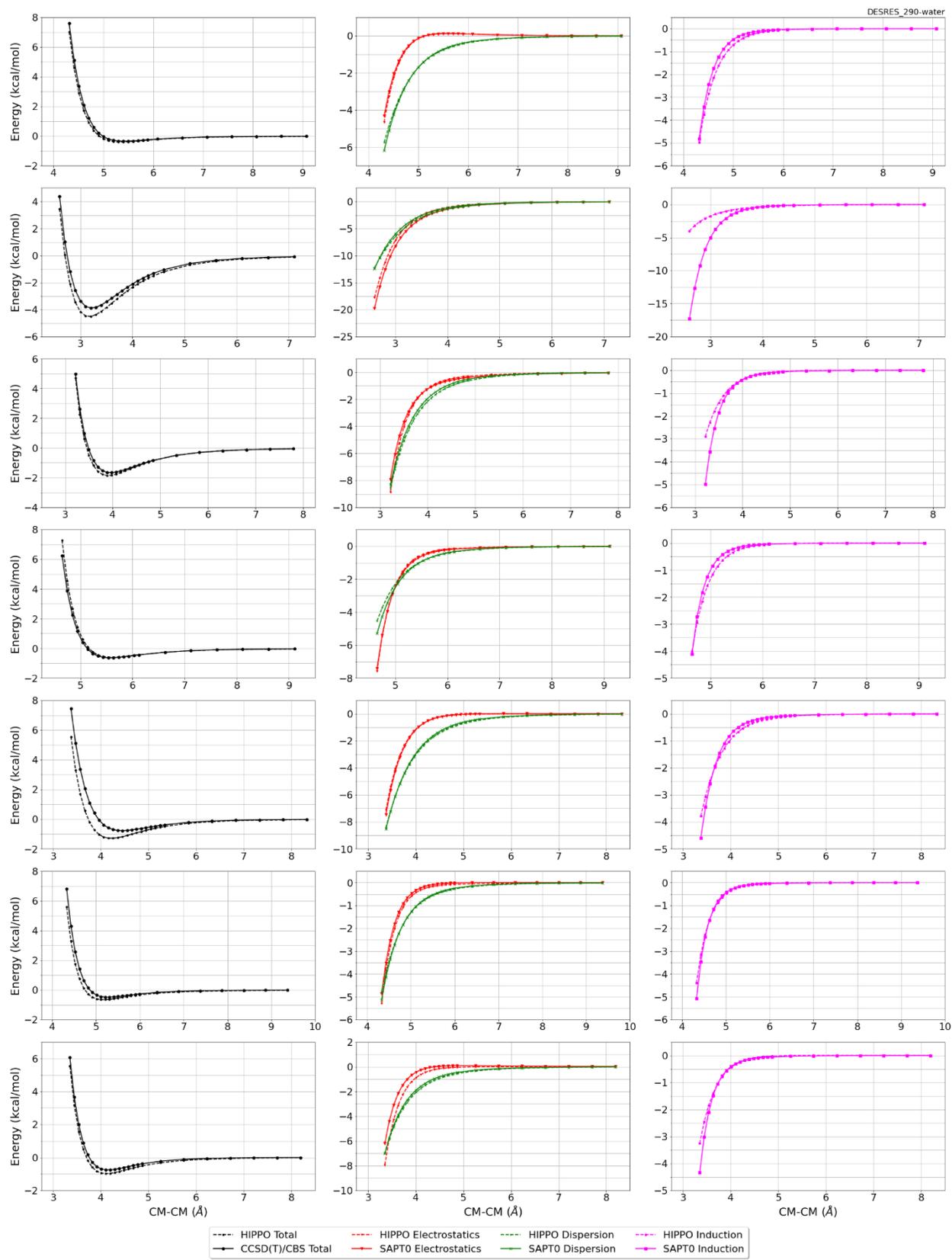
Monomer potential fitting RMS: 0.34

##Dimer results - Fitting to QM datasets##

DESRES_290-water, energy values in kcal/mol

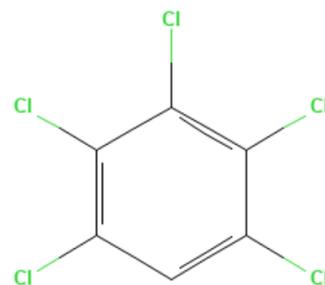
MAE	Std error	max error	#points	#count[err > 1]
0.184	0.269	1.7783	568	17





#294 Pentachlorobenzene C6HCl5 CID: 11855

ref molpol 23.82 26.93 0.00, avg 16.92
 molpol 23.94 23.94 11.83, avg 19.90
 rms molpol 0.12 2.99 11.83, avg 2.99

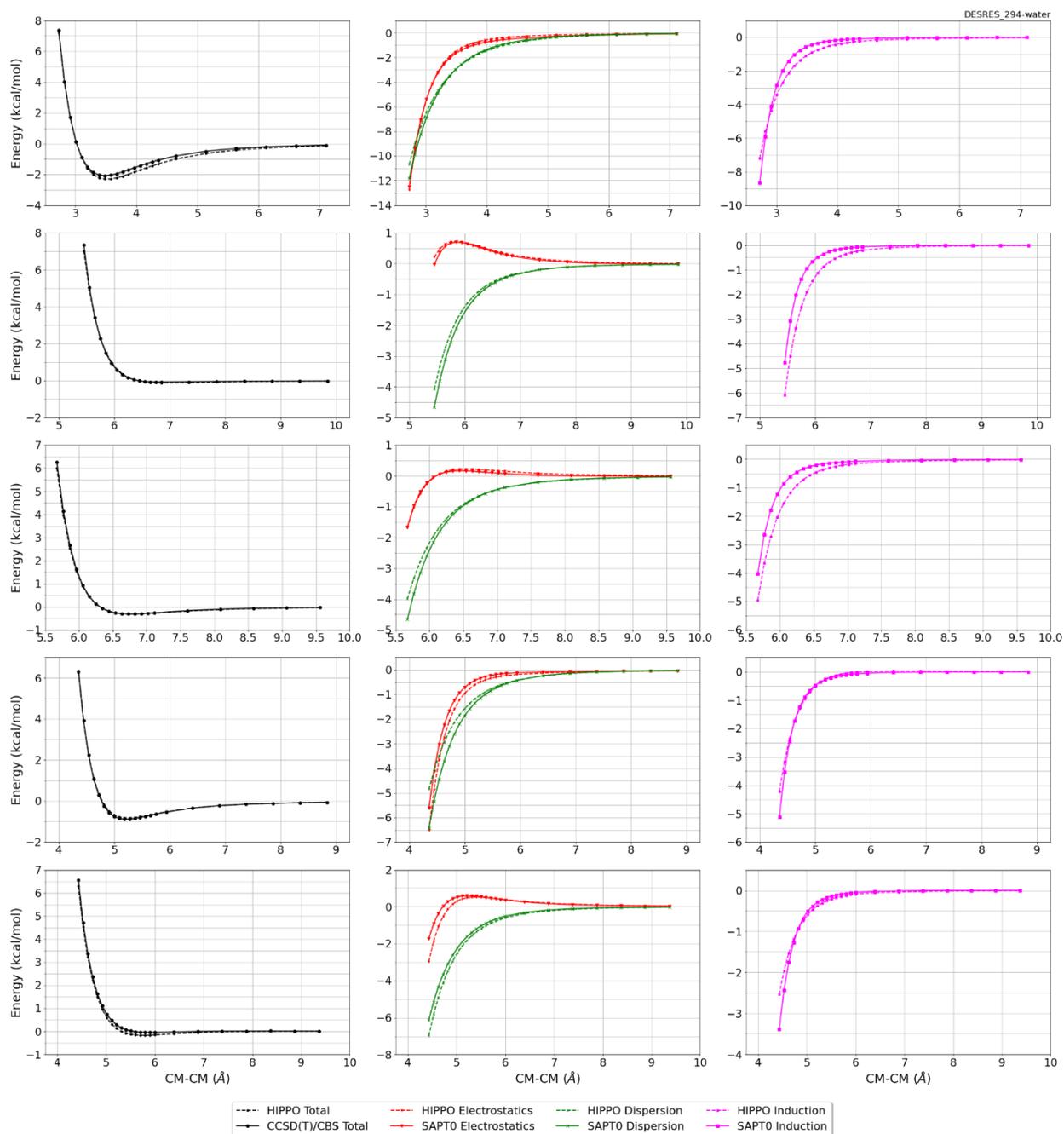


Monomer potential fitting RMS: 0.49

##Dimer results - Fitting to QM datasets##

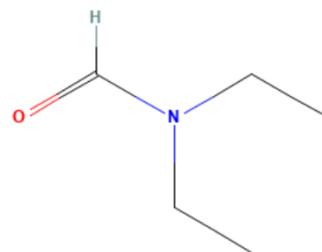
DESRES_294-water, energy values in kcal/mol

MAE 0.070 Std error 0.062 max error 0.2071 #points 113 #count[err > 1] 0



#298 N,N-Diethylformamide C5H11NO CID: 12051

ref molpol	12.42	11.67	8.92, avg	11.00
molpol	12.50	11.63	9.02, avg	11.05
rms molpol	0.09	0.04	0.09, avg	0.05

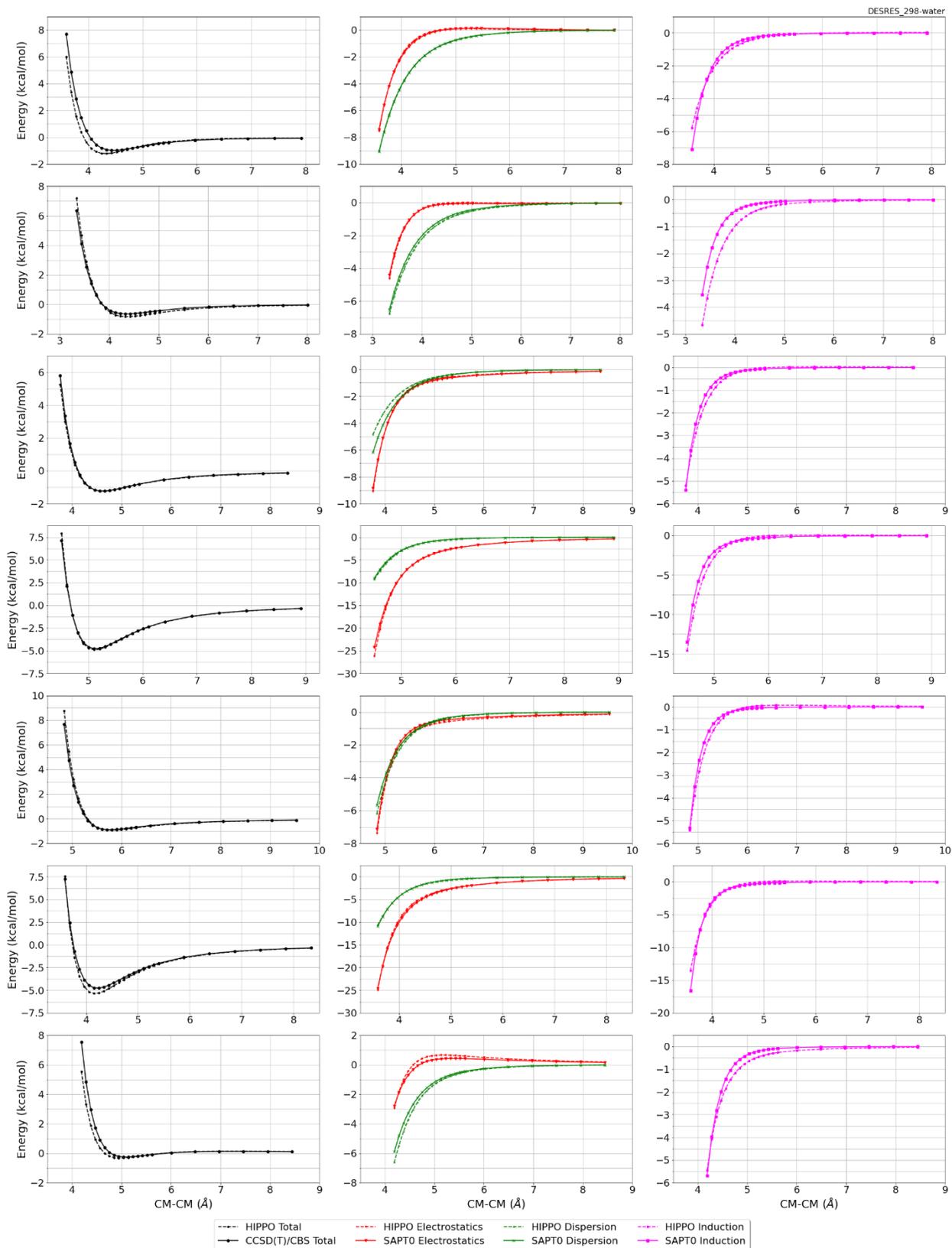


Monomer potential fitting RMS: 0.82

##Dimer results - Fitting to QM datasets##

DESRES_298-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.205	0.268	1.6993	536	15



#301 1,2-Difluoroethane C2H4F2 CID: 12223

ref molpol	4.57	4.28	0.00, avg	2.95
molpol	4.58	4.00	3.72, avg	4.10
rms molpol	0.02	0.28	3.72, avg	1.15



Monomer potential fitting RMS: 0.13

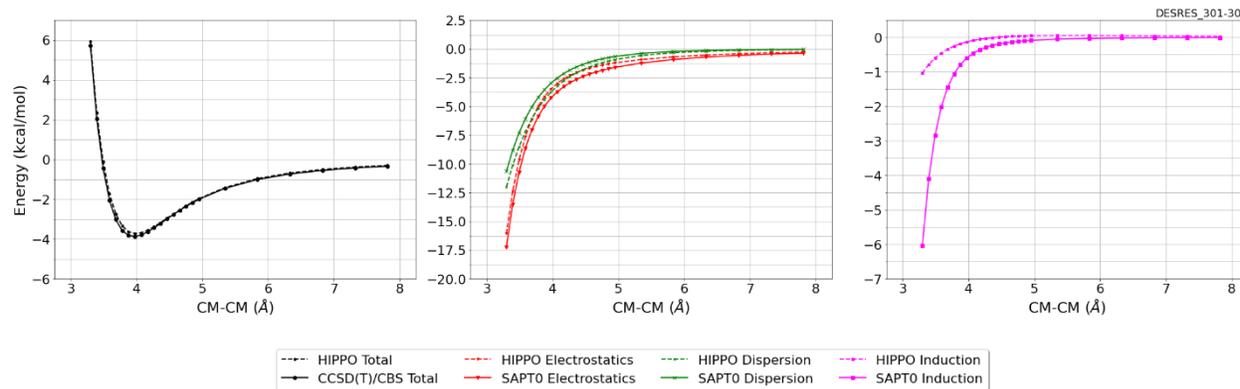
##Dimer results - Fitting to QM datasets##

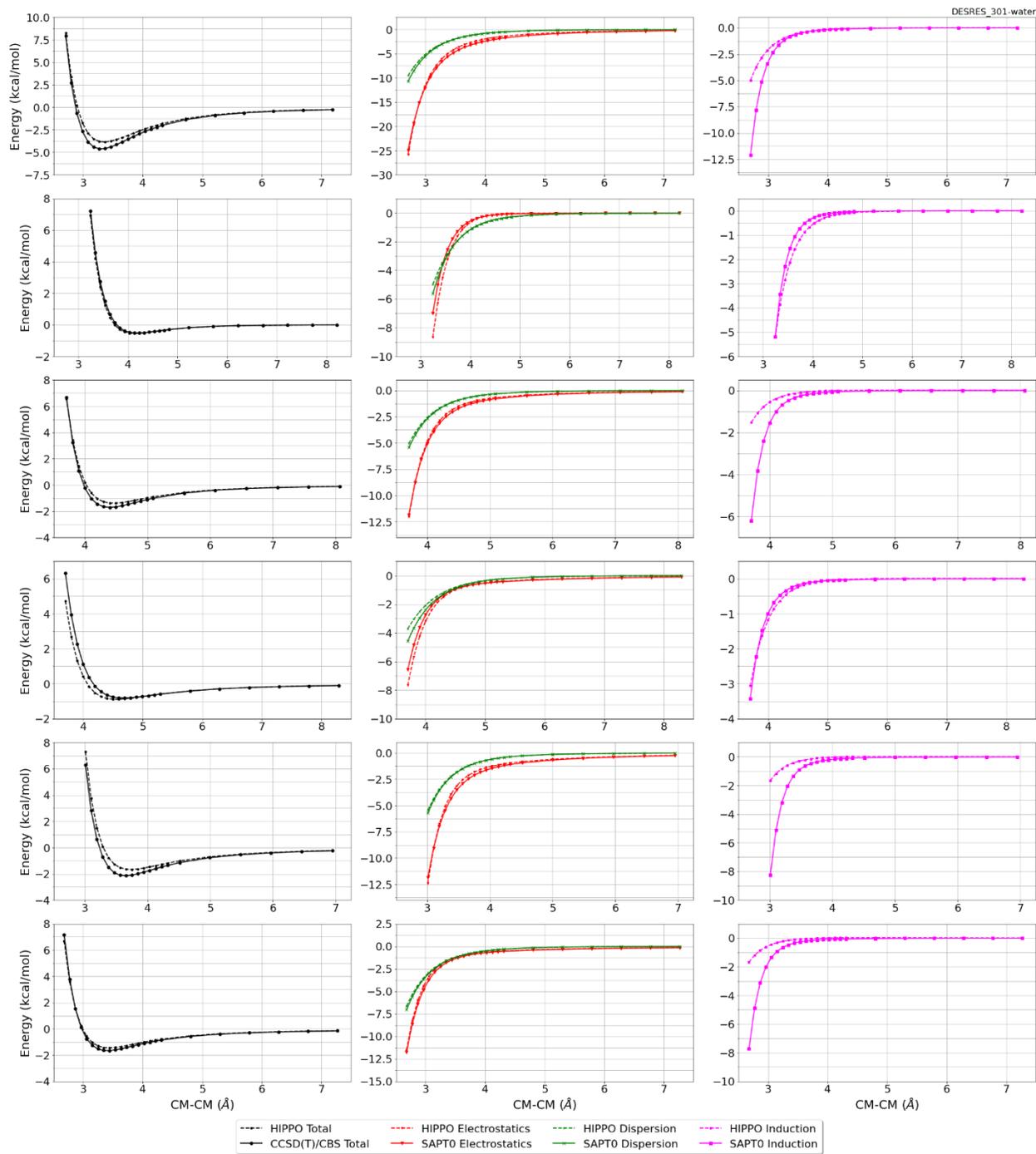
DESRES_301-301, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.078	0.086	0.2708	24	0

DESRES_301-water, energy values in kcal/mol

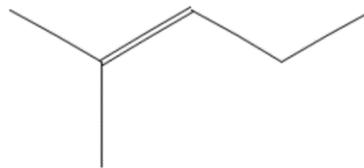
MAE	Std error	max error	#points	#count[err > 1]
0.151	0.232	1.6602	287	3





#304 2-Methylpent-2-ene C6H12 CID: 12243

ref molpol	10.89	13.73	0.00, avg	8.21
molpol	10.89	9.87	7.22, avg	9.33
rms molpol	0.01	3.86	7.22, avg	1.12

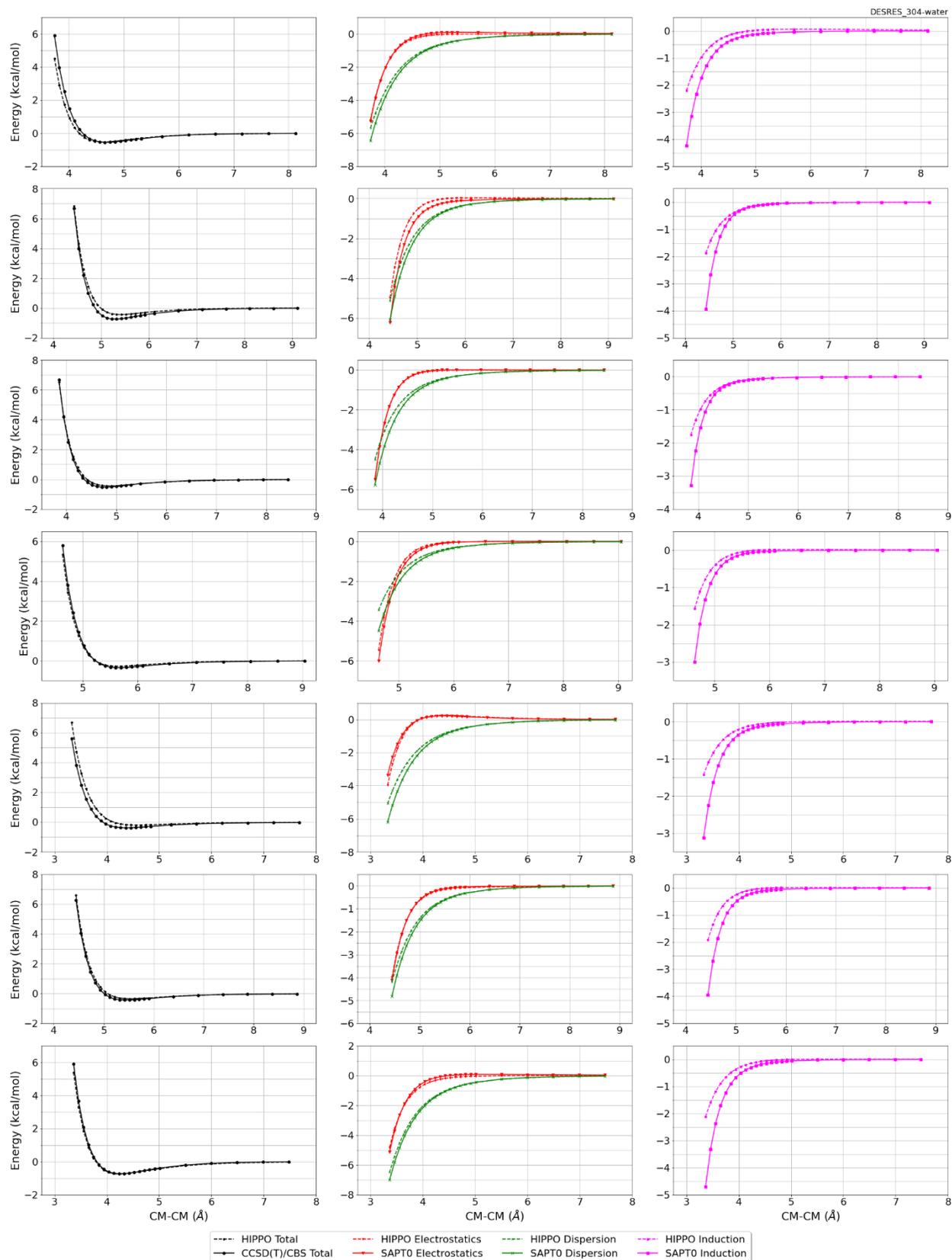


Monomer potential fitting RMS: 0.40

##Dimer results - Fitting to QM datasets##

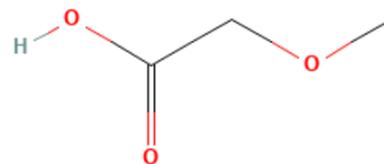
DESRES_304-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.153	0.188	1.3876	531	3



#305 2-Methoxyacetic Acid C3H6O3 CID: 12251

ref molpol	7.66	9.08	0.00, avg	5.58
molpol	7.74	6.83	5.42, avg	6.66
rms molpol	0.08	2.25	5.42, avg	1.09

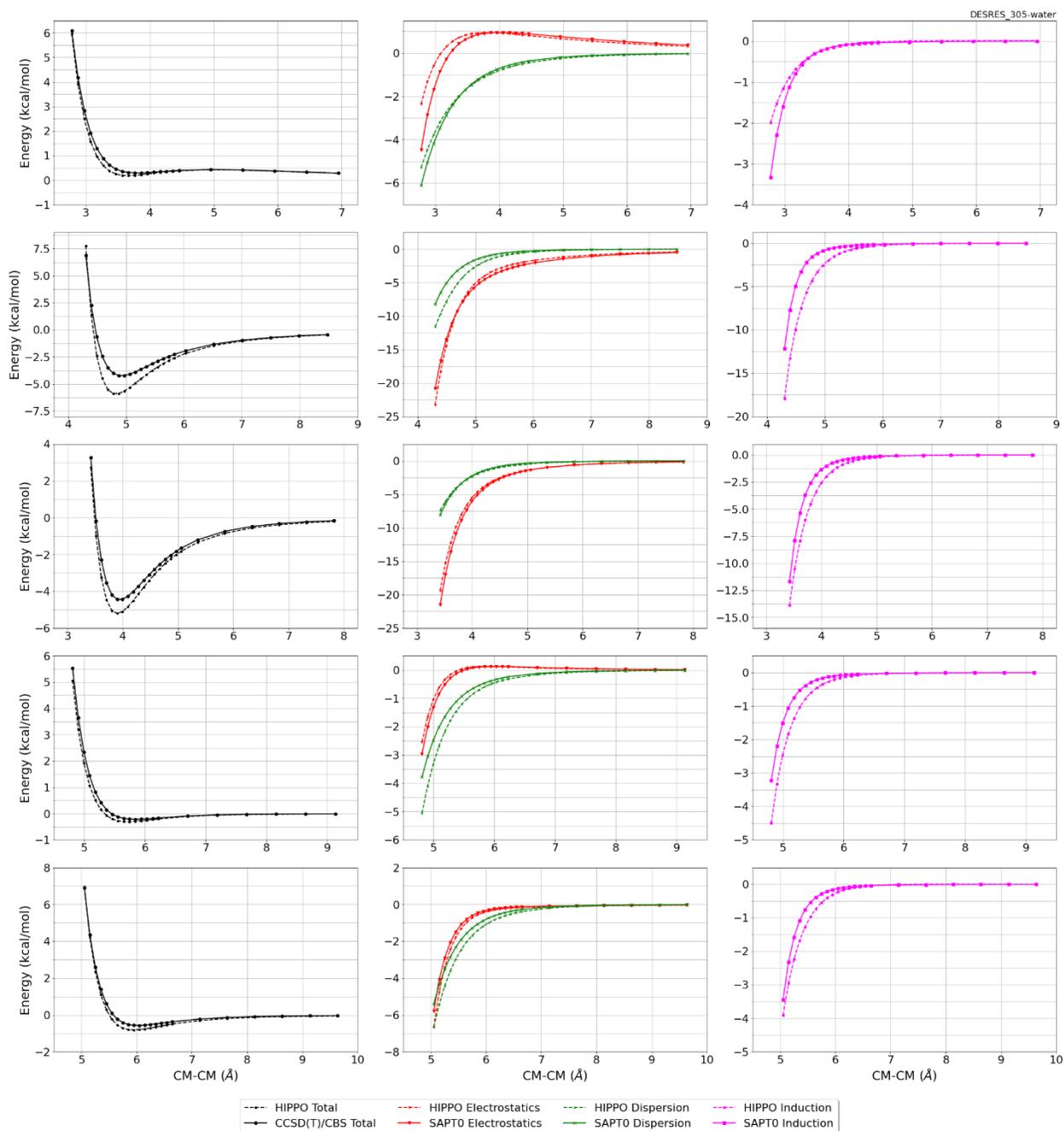


Monomer potential fitting RMS: 0.55

##Dimer results - Fitting to QM datasets##

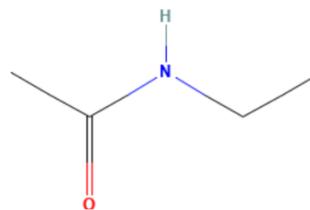
DESRES_305-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.284	0.388	1.8032	265	19



#306 N-Ethylacetamide C4H9NO CID: 12253

ref molpol	9.40	11.26	0.00, avg	6.89
molpol	9.43	7.83	6.28, avg	7.85
rms molpol	0.03	3.43	6.28, avg	0.96

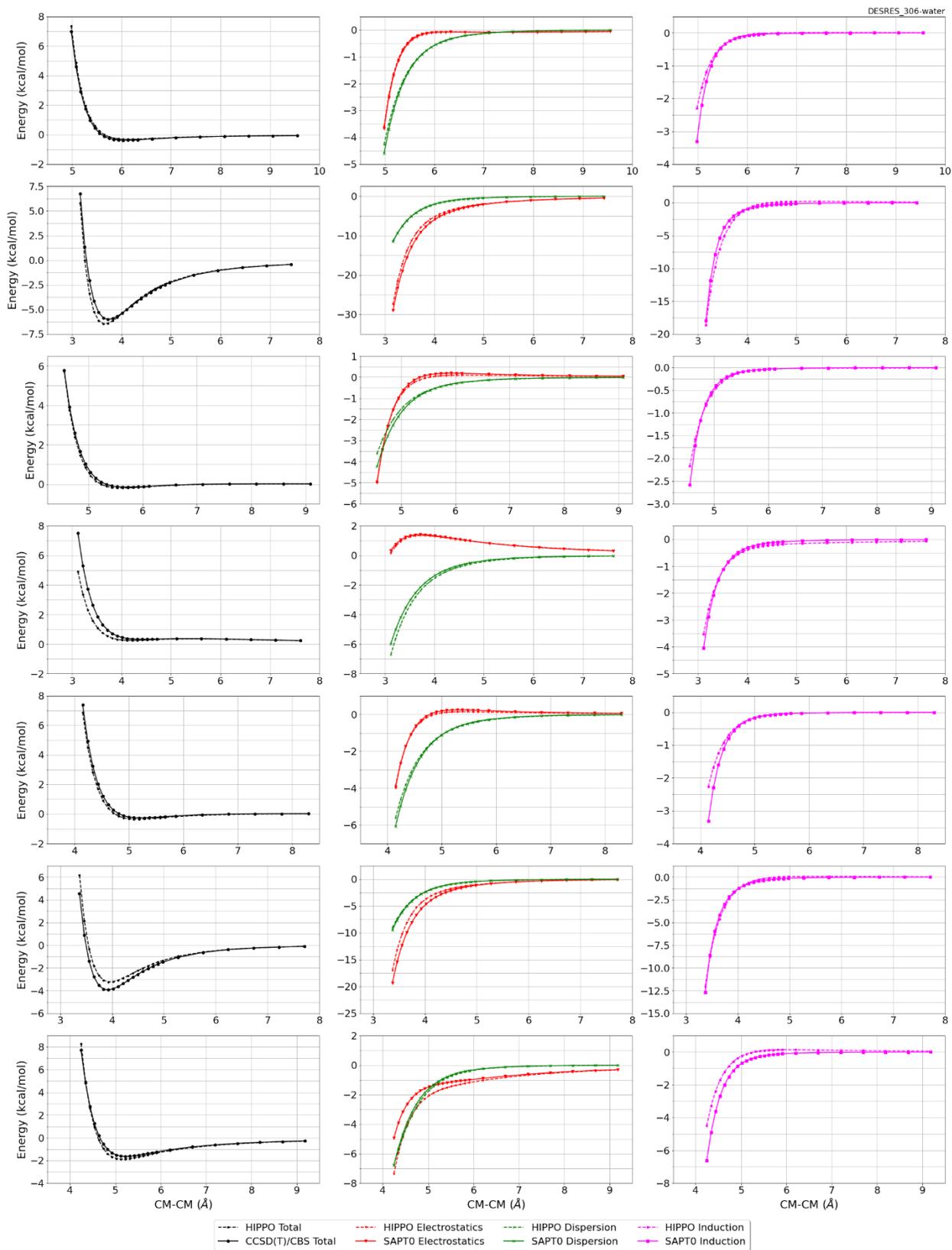


Monomer potential fitting RMS: 0.35

##Dimer results - Fitting to QM datasets##

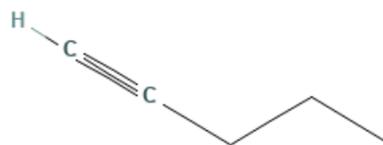
DESRES_306-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.269	0.404	2.6871	536	34



#307 Pent-1-yne C5H8 CID: 12309

ref molpol	7.69	11.87	0.00, avg	6.52
molpol	7.74	5.88	5.56, avg	6.39
rms molpol	0.05	5.99	5.56, avg	0.13

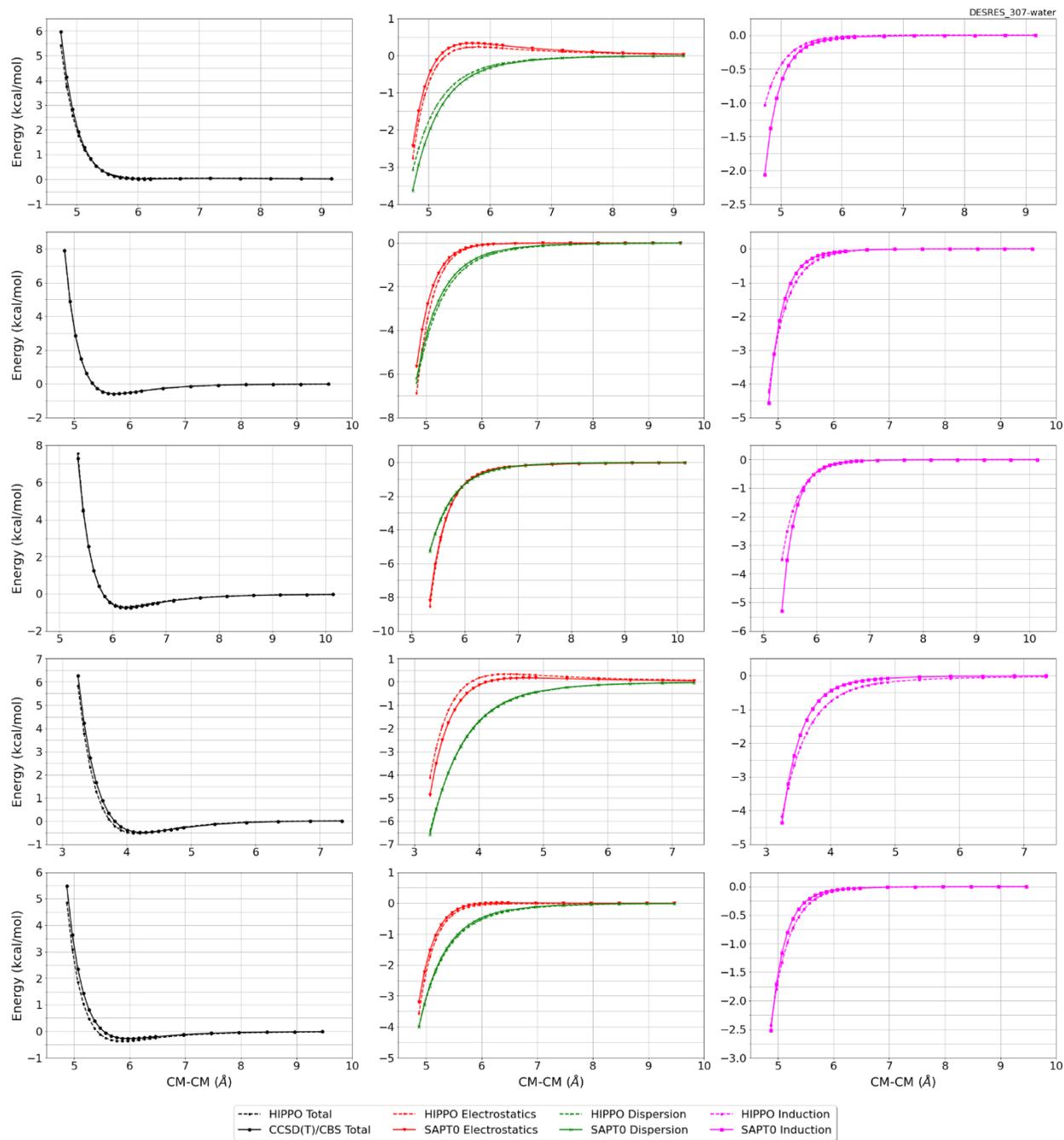


Monomer potential fitting RMS: 0.47

##Dimer results - Fitting to QM datasets##

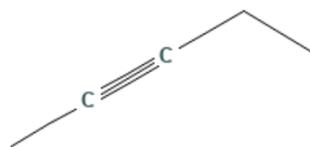
DESRES_307-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.083	0.130	0.6744	113	0



#308 Pent-2-yne C5H8 CID: 12310

ref molpol	12.45	7.62	0.00, avg	6.69
molpol	12.34	7.57	7.02, avg	8.98
rms molpol	0.11	0.05	7.02, avg	2.29

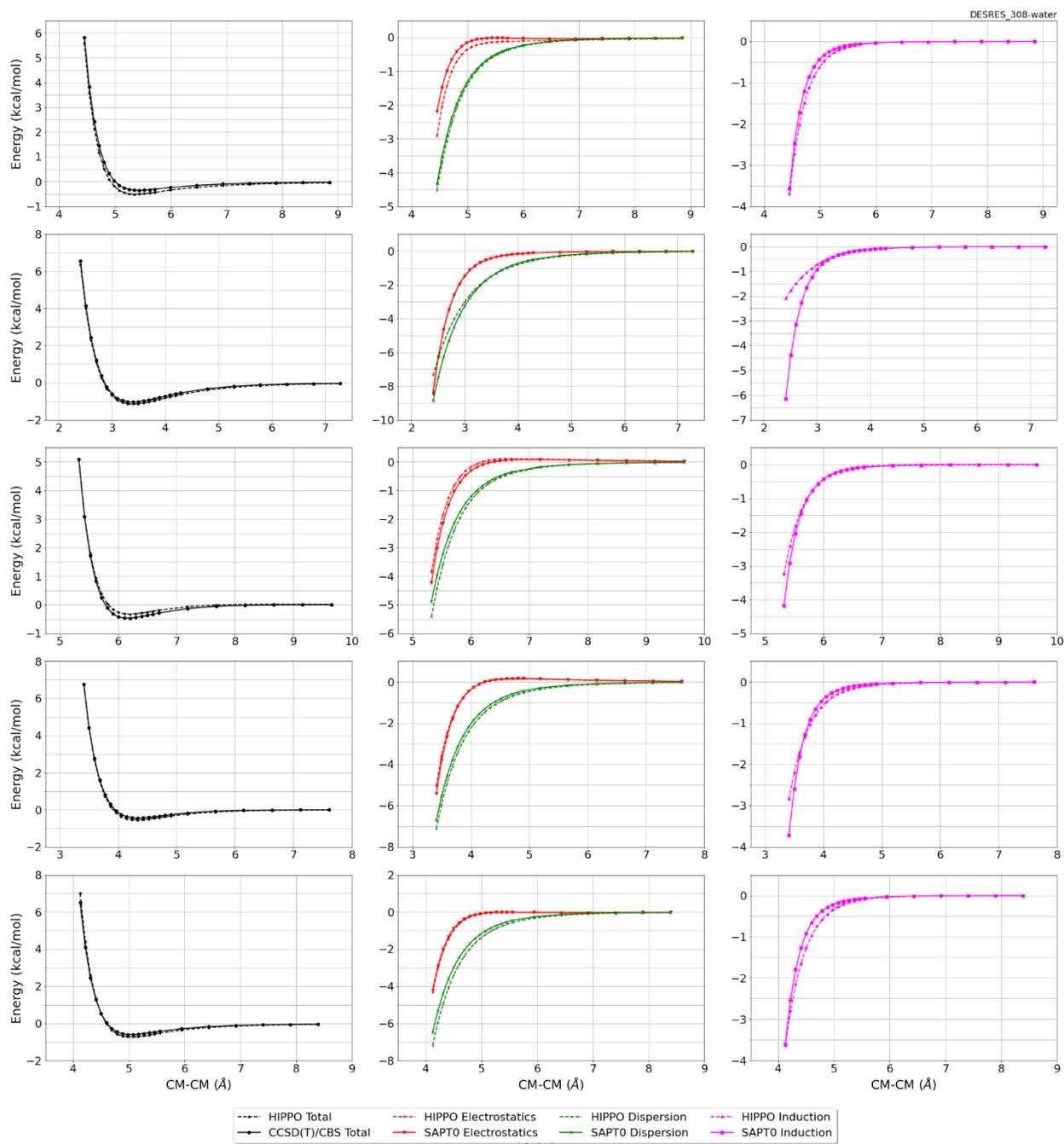


Monomer potential fitting RMS: 0.30

##Dimer results - Fitting to QM datasets##

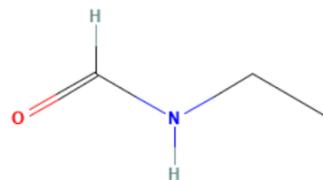
DESRES_308-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.104	0.105	0.6255	264	0



#310 N-Ethylformamide C3H7NO CID: 12319

ref molpol	10.05	7.19	0.00, avg	5.75
molpol	10.06	7.00	5.87, avg	7.64
rms molpol	0.01	0.19	5.87, avg	1.90



Monomer potential fitting RMS: 0.23

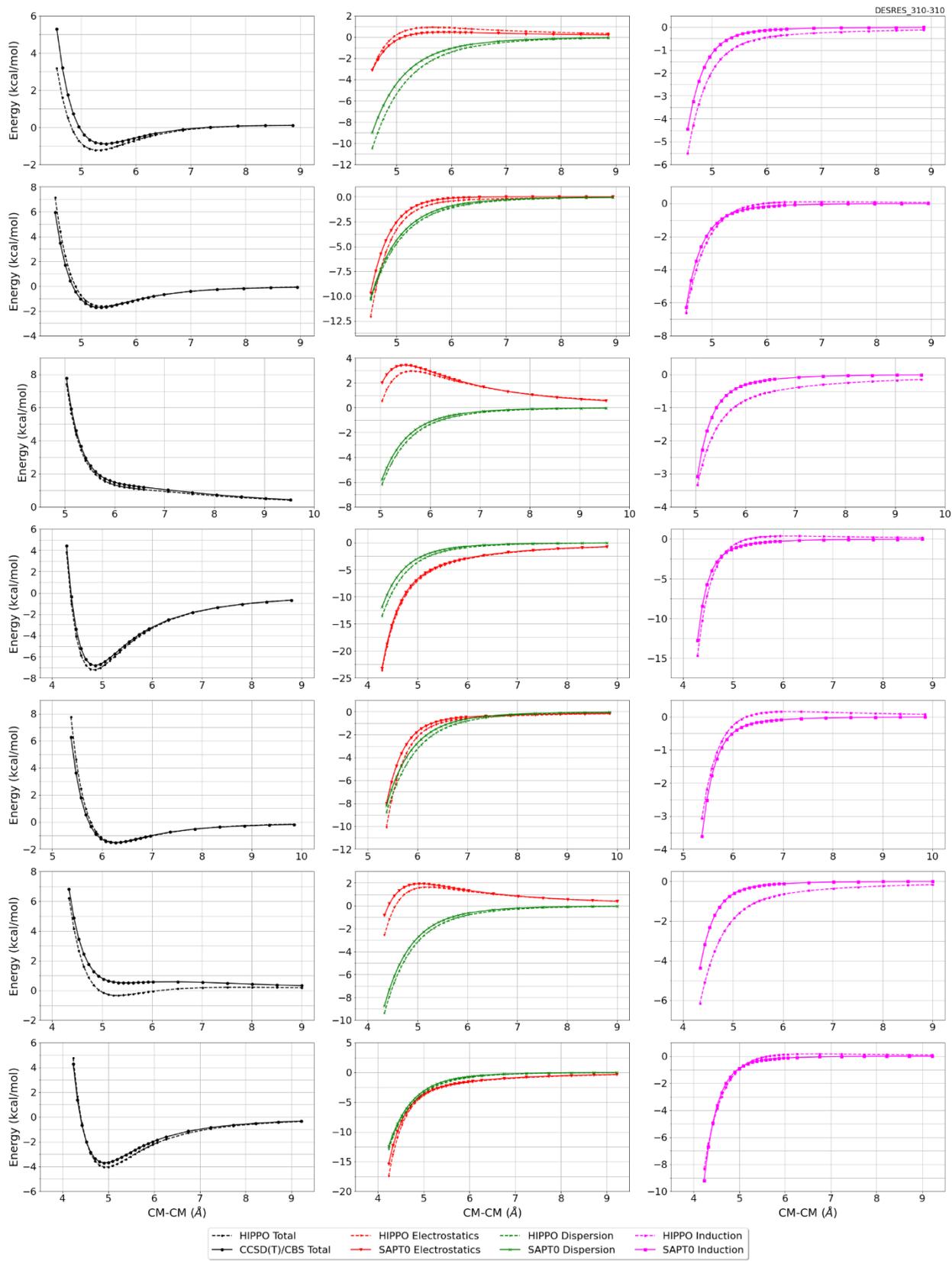
##Dimer results - Fitting to QM datasets##

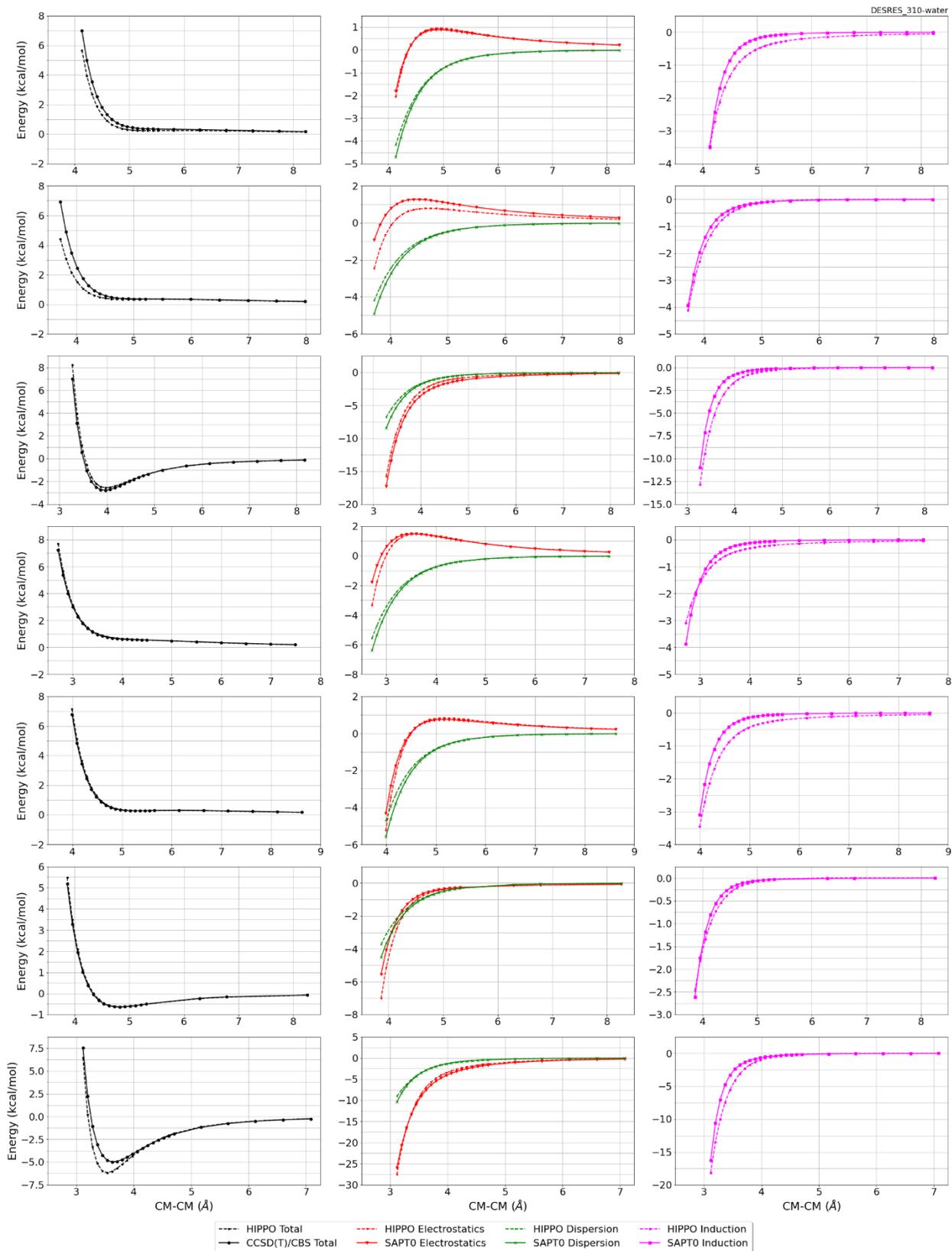
DESRES_310-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.223	0.414	2.5941	528	28

DESRES_310-310, energy values in kcal/mol

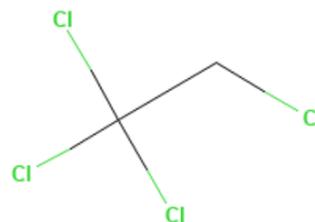
MAE	Std error	max error	#points	#count[err > 1]
0.278	0.391	3.0650	474	28





#313 1,1,1,2-Tetrachloroethane C2H2Cl4 CID: 12418

ref molpol	13.27	10.80	0.00, avg	8.02
molpol	13.25	10.49	10.14, avg	11.29
rms molpol	0.02	0.32	10.14, avg	3.27

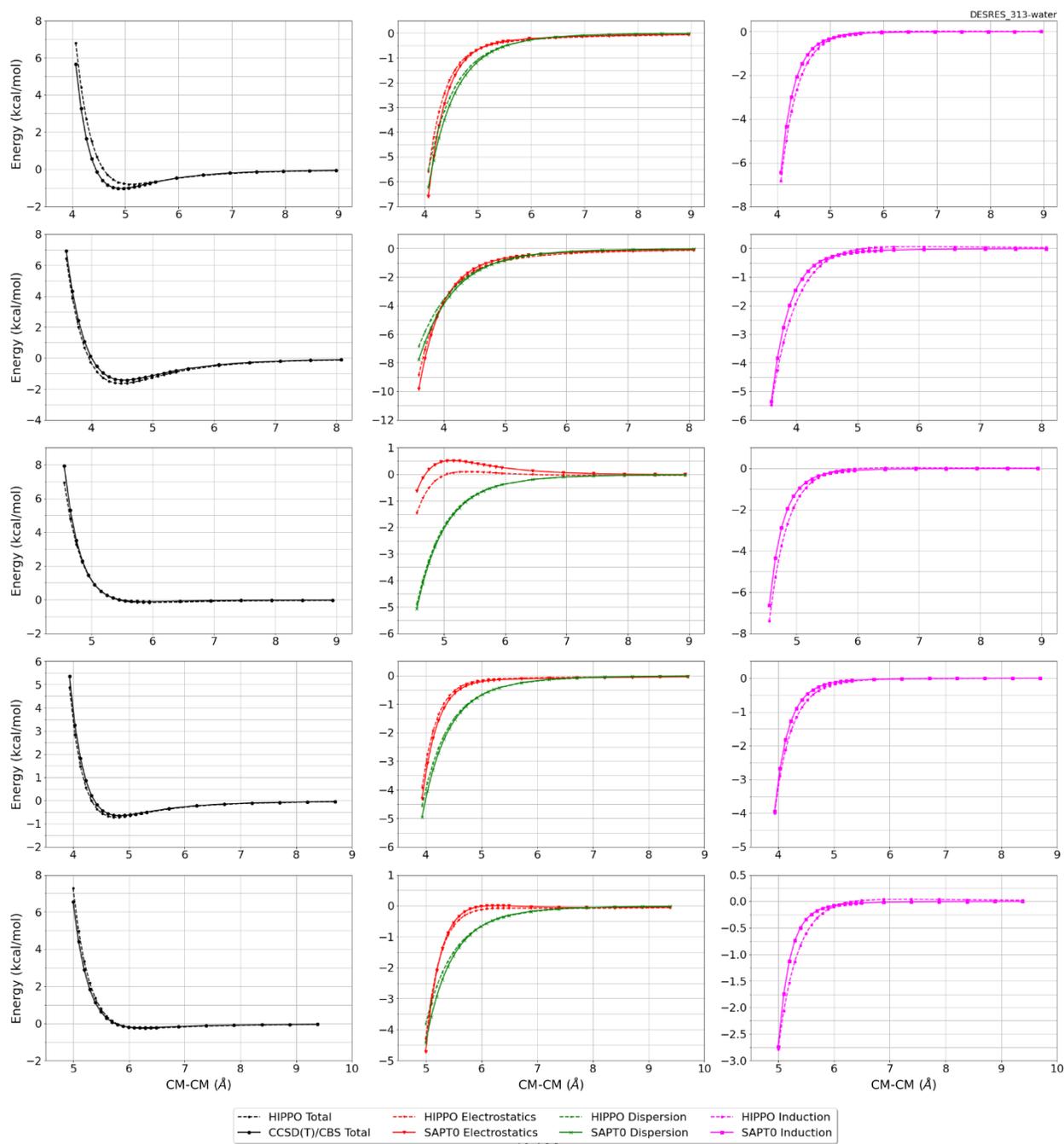


Monomer potential fitting RMS: 0.37

##Dimer results - Fitting to QM datasets##

DESRES_313-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.286	0.521	4.6908	263	15



#315 N,n-diethylacetamide C6H13NO CID: 12703

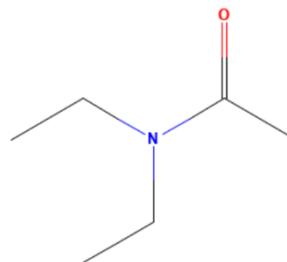
ref molpol -10.51 -13.07 -14.64, avg -12.74
molpol 10.48 13.12 14.62, avg 12.74
rms molpol 0.03 0.05 0.02, avg 0.00

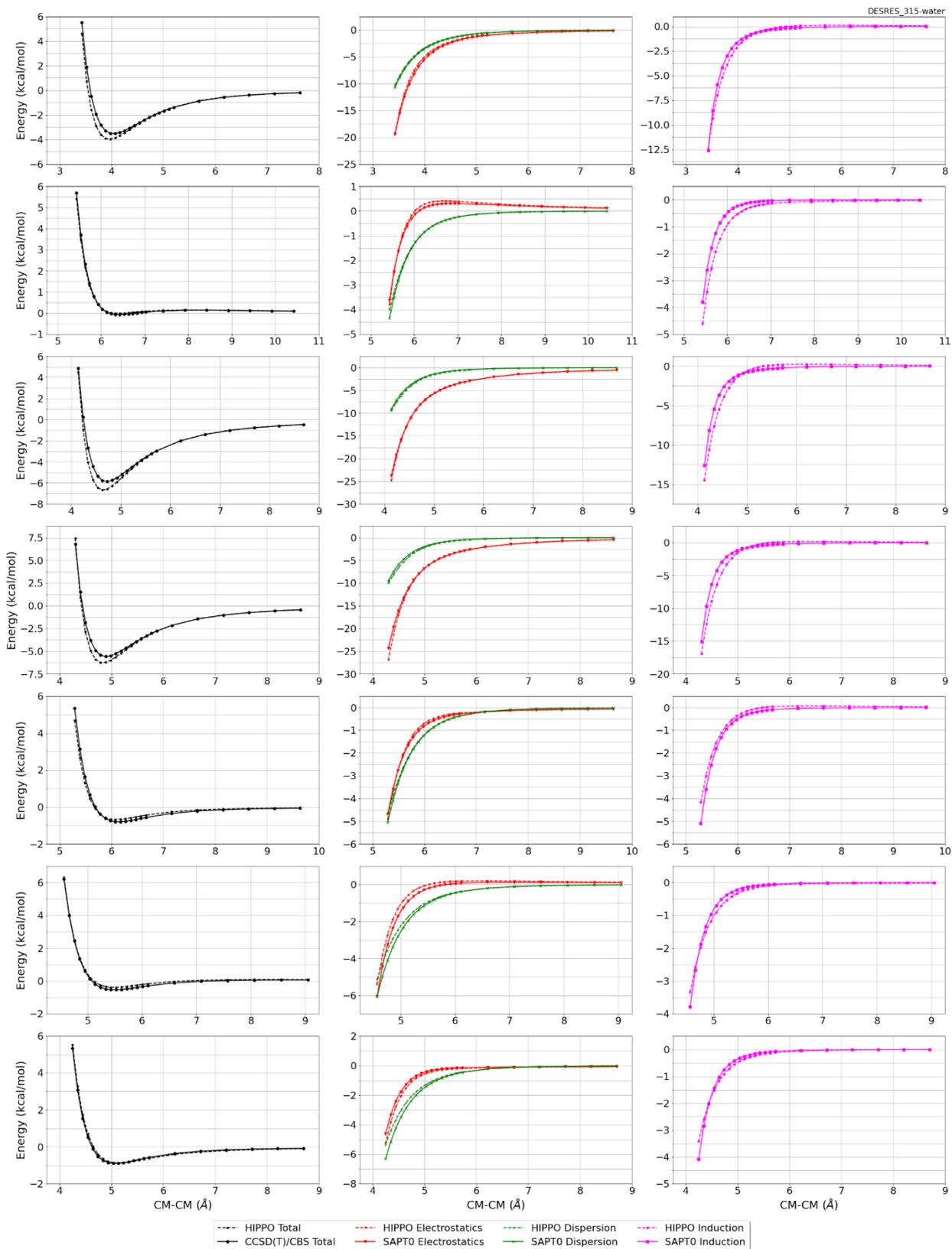
Monomer potential fitting RMS: 0.72

##Dimer results - Fitting to QM datasets##

DESRES_315-water, energy values in kcal/mol

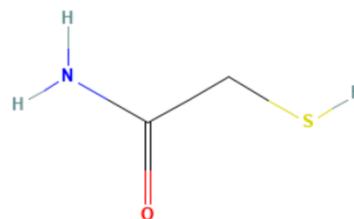
MAE	Std error	max error	#points	#count[err > 1]
0.271	0.426	2.4432	534	42





#317 2-Sulfanylacetamide C₂H₅NOS CID: 12961

ref molpol	10.83	8.58	0.00, avg	6.47
molpol	10.89	8.80	6.89, avg	8.86
rms molpol	0.06	0.22	6.89, avg	2.39

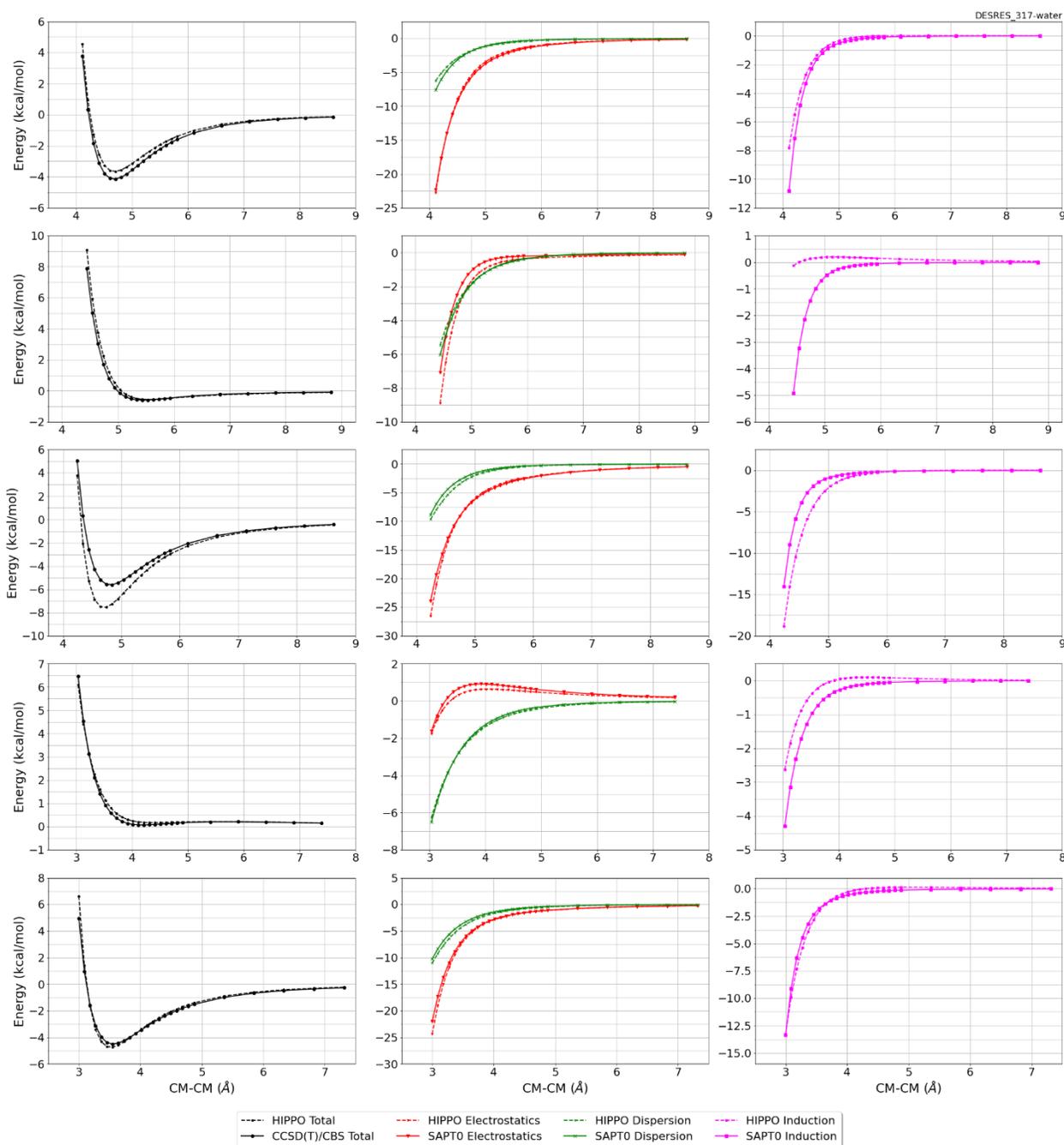


Monomer potential fitting RMS: 0.97

##Dimer results - Fitting to QM datasets##

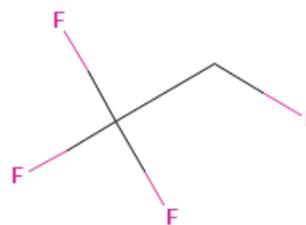
DESRES_317-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.309	0.452	2.3775	239	15



#319 1,1,1,2-Tetrafluoroethane C₂H₂F₄ CID: 13129

ref molpol	4.62	4.44	0.00, avg	3.02
molpol	4.62	4.11	3.90, avg	4.21
rms molpol	0.00	0.34	3.90, avg	1.19

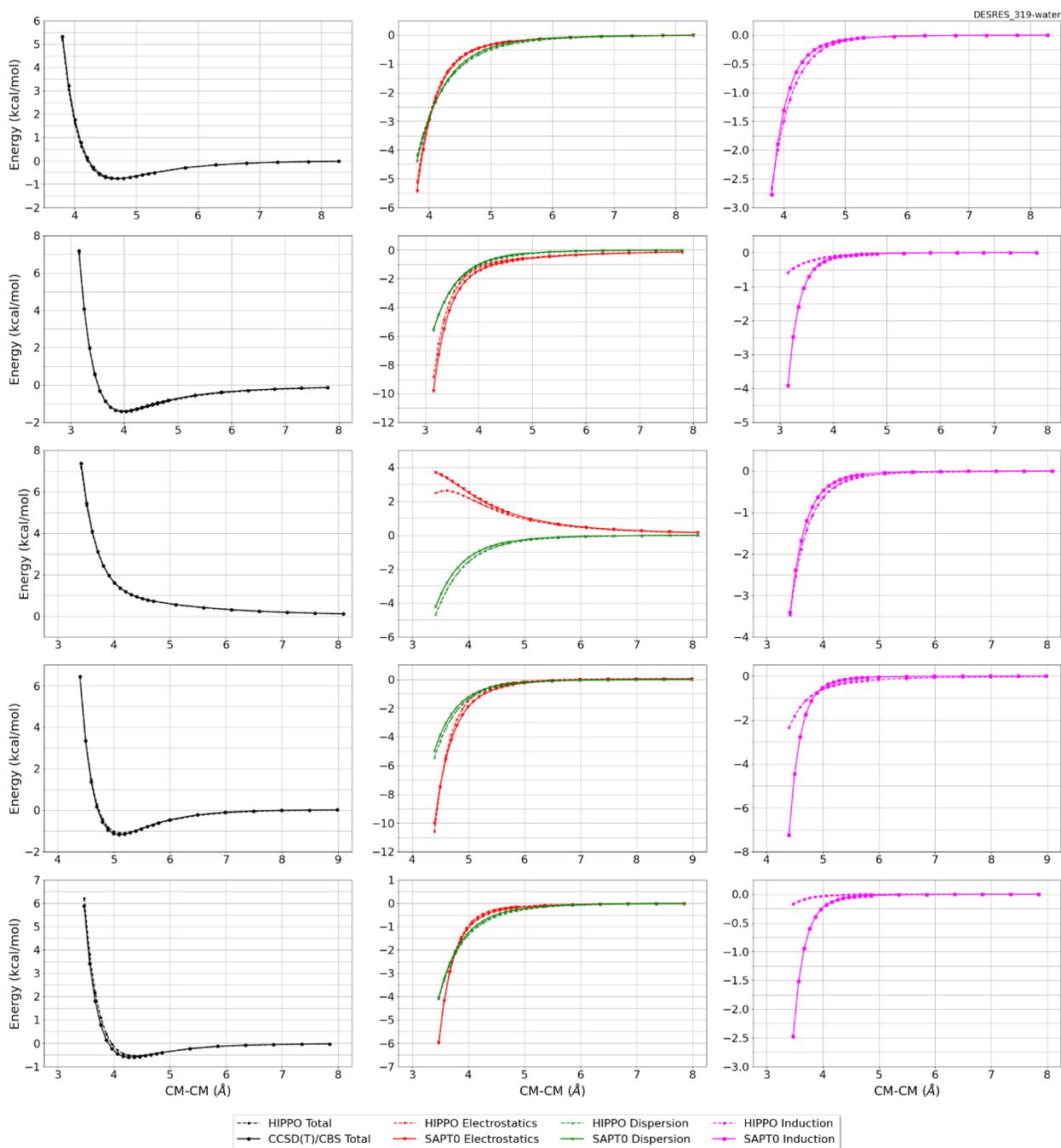


Monomer potential fitting RMS: 0.20

##Dimer results - Fitting to QM datasets##

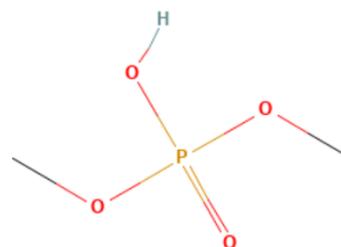
DESRES_319-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.070	0.097	0.5522	260	0



#321 Dimethyl Hydrogen Phosphate C2H7O4P CID: 13134

ref molpol	10.81	8.36	8.20, avg	9.12
molpol	10.79	8.39	8.19, avg	9.12
rms molpol	0.02	0.03	0.01, avg	0.00

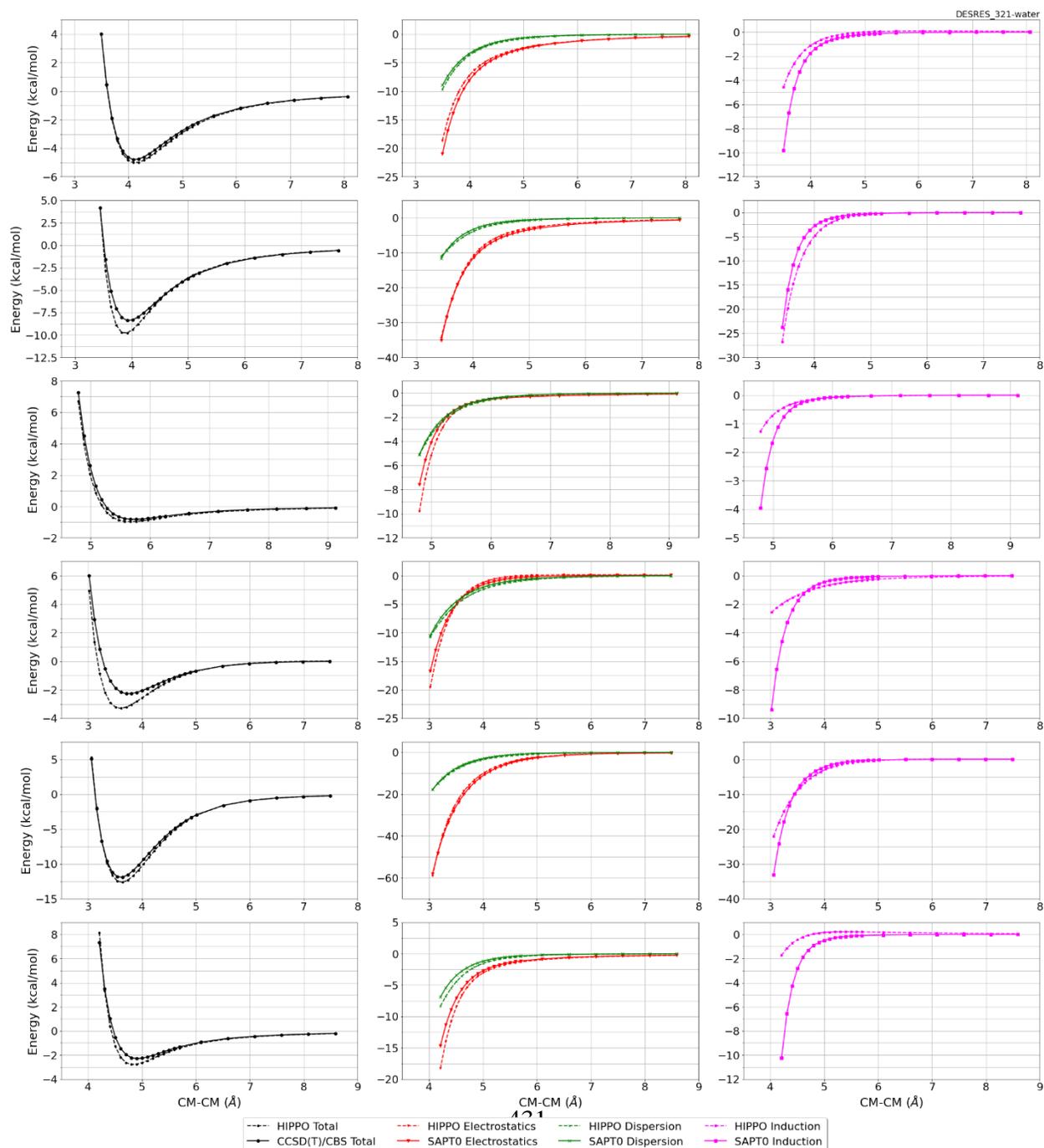


Monomer potential fitting RMS: 0.52

##Dimer results - Fitting to QM datasets##

DESRES_321-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.350	0.378	1.7454	145	13



#322 5-Methyl-1h-imidazole C4H6N2 CID: 13195

ref molpol	11.18	9.53	6.38, avg	9.03
molpol	11.03	9.32	5.88, avg	8.75
rms molpol	0.14	0.21	0.50, avg	0.29

Monomer potential fitting RMS: 0.69

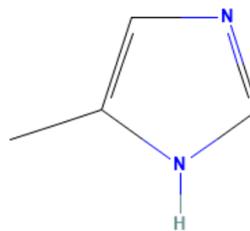
##Dimer results - Fitting to QM datasets##

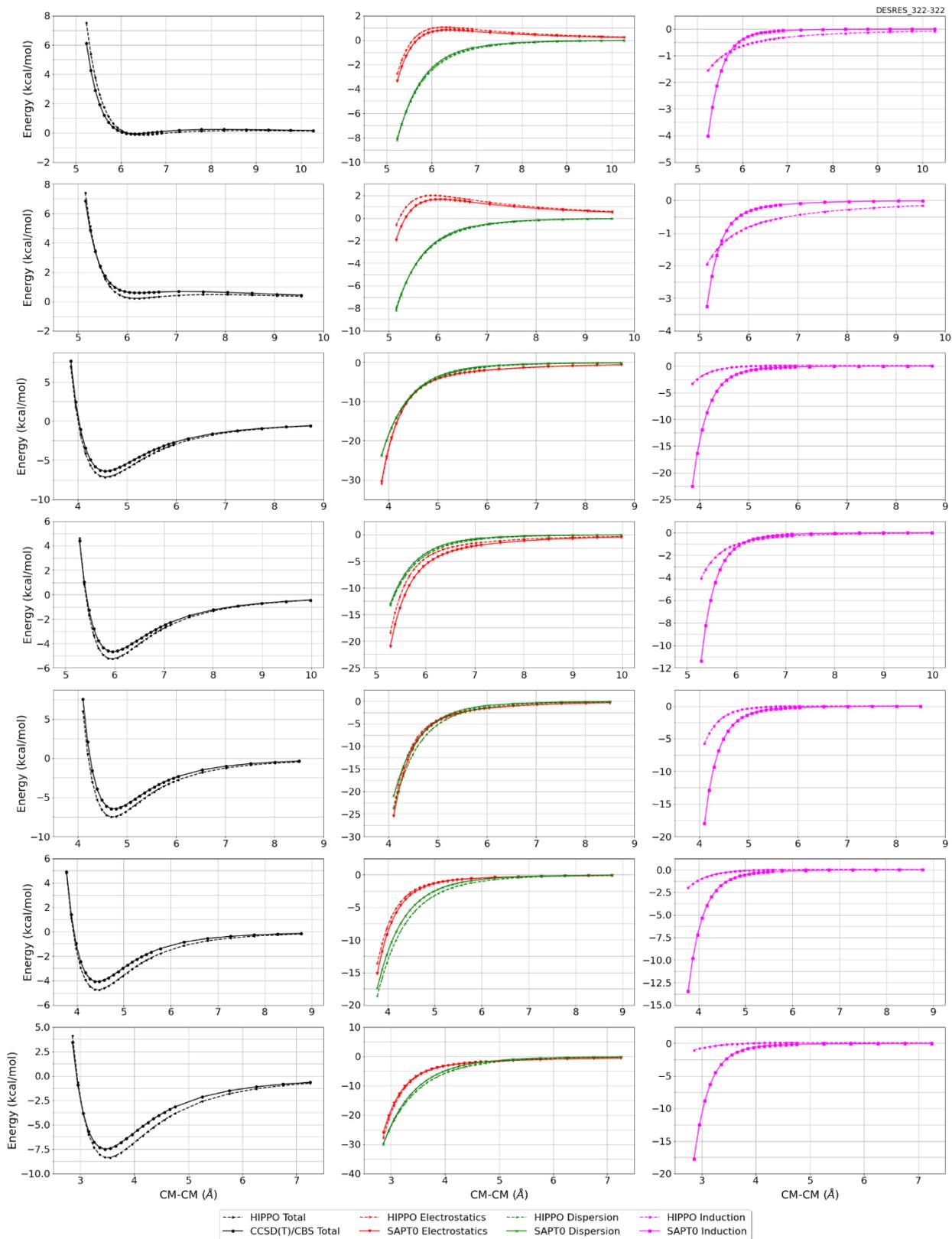
DESRES_322-322, energy values in kcal/mol

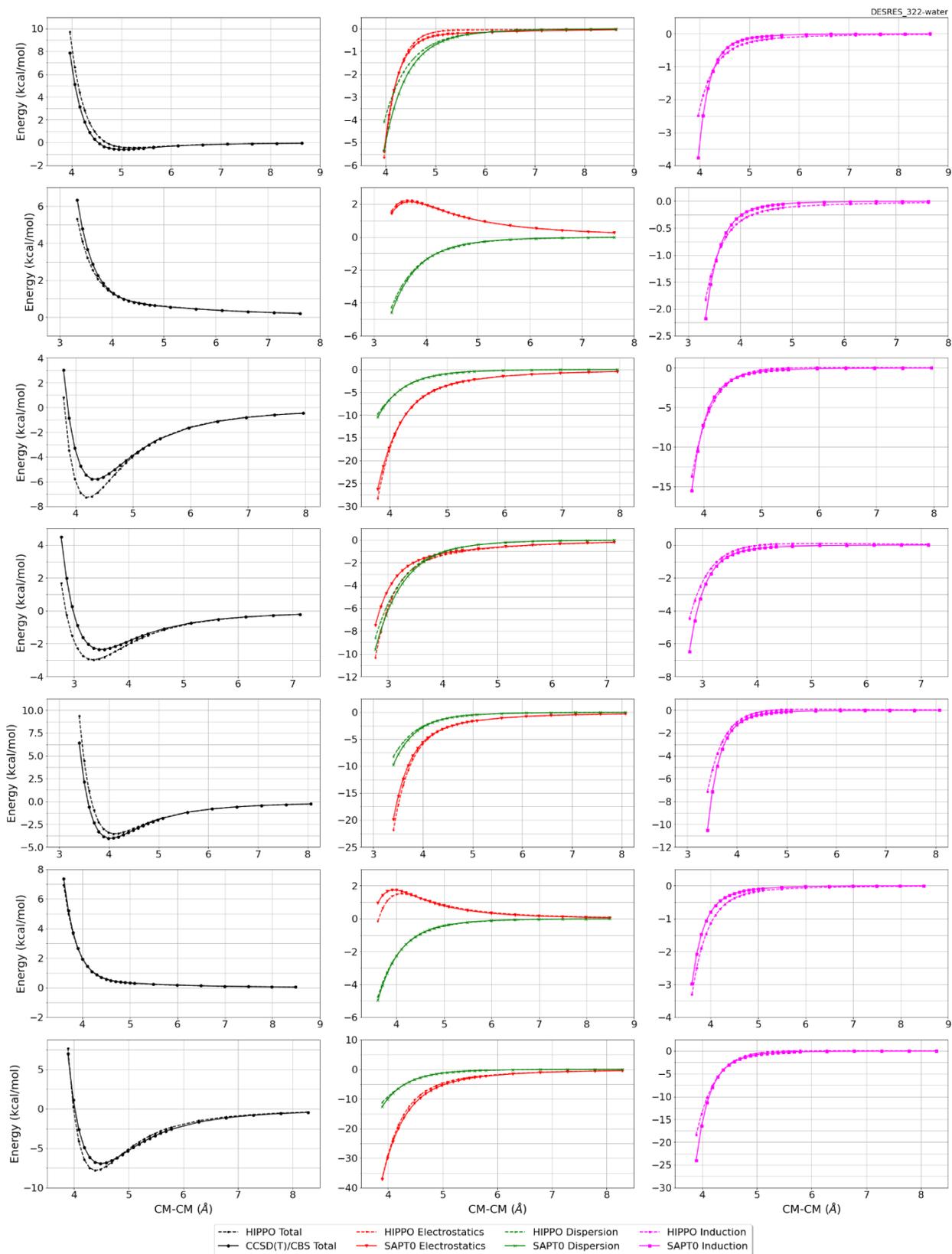
MAE	Std error	max error	#points	#count[err > 1]
0.542	0.610	3.1648	489	97

DESRES_322-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.373	0.641	3.4983	558	63







#328 1-(Methyldisulfanyl)propane C4H10S2 CID: 16592

ref molpol	17.95	11.82	12.61, avg	14.12
molpol	17.92	12.73	11.83, avg	14.16
rms molpol	0.03	0.91	0.78, avg	0.04

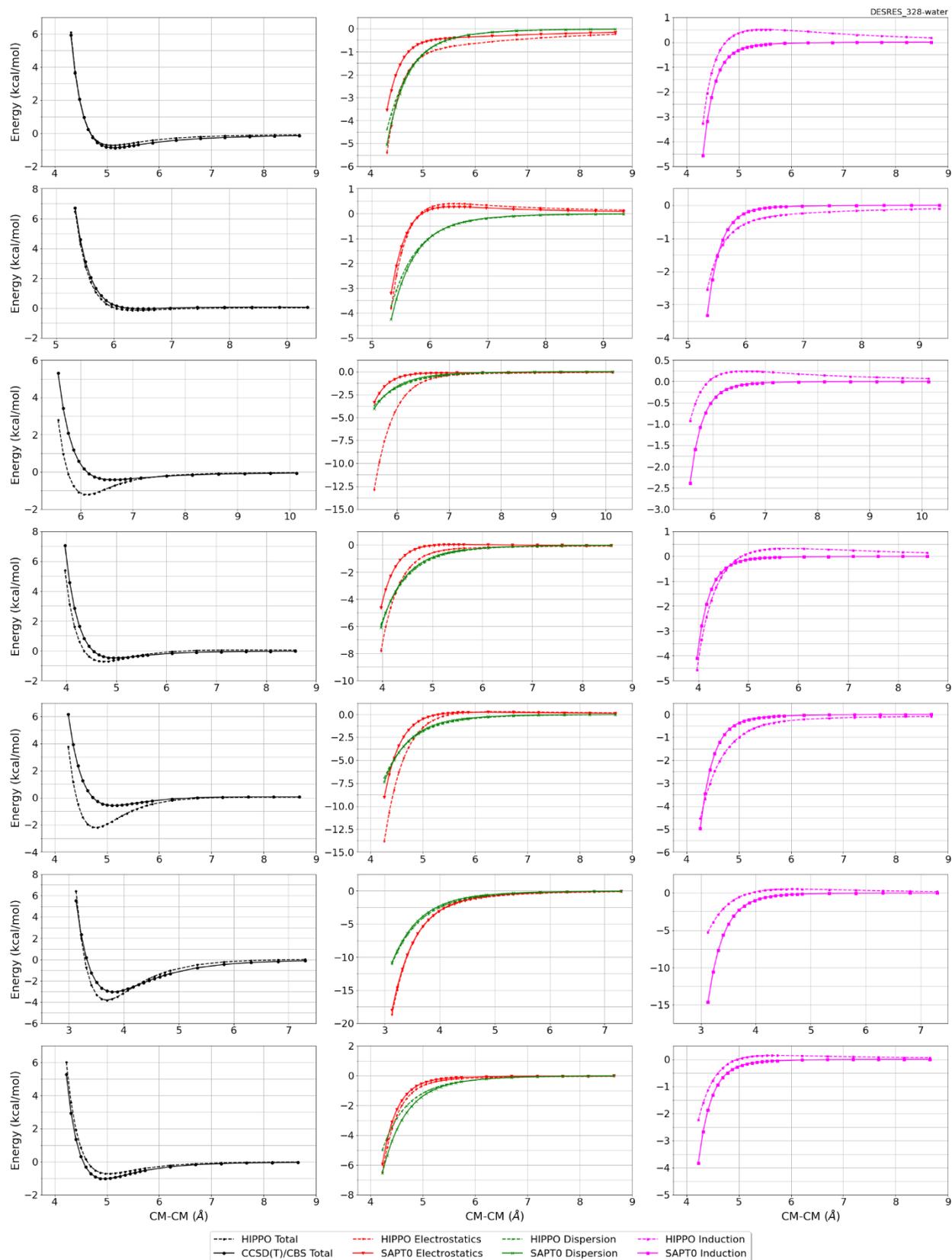


Monomer potential fitting RMS: 3.20

##Dimer results - Fitting to QM datasets##

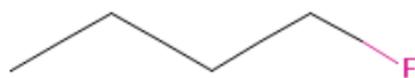
DESRES_328-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.562	0.755	3.5792	503	100



#330 1-Fluorobutane C4H9F CID: 16908

ref molpol	7.41	8.85	0.00, avg	5.42
molpol	7.42	6.10	5.78, avg	6.43
rms molpol	0.02	2.76	5.78, avg	1.01

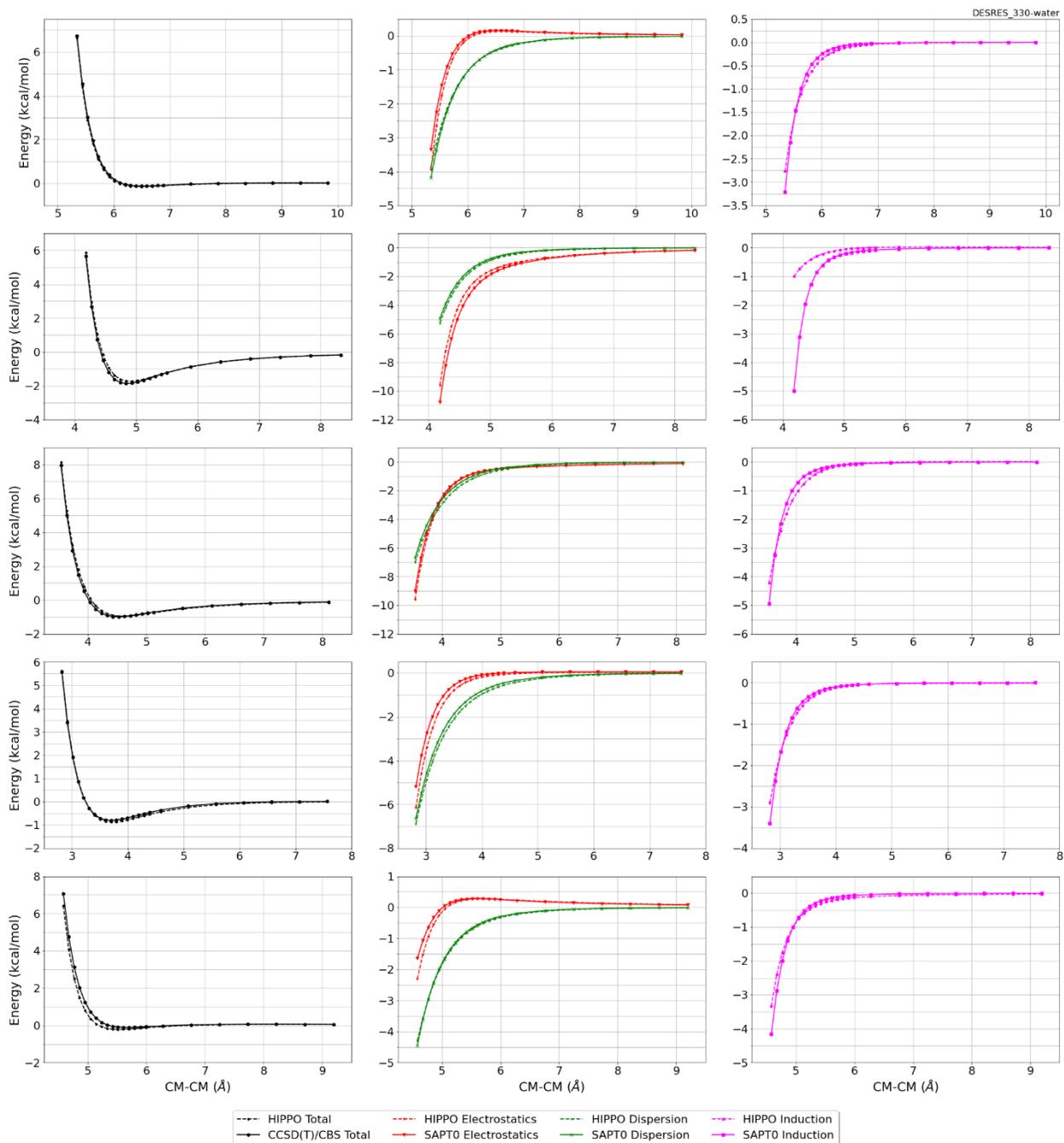


Monomer potential fitting RMS: 0.56

##Dimer results - Fitting to QM datasets##

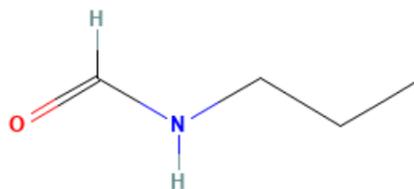
DESRES_330-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.107	0.134	0.8026	264	0



#333 N-Propylformamide C4H9NO CID: 22686

ref molpol	12.54	8.68	0.00, avg	7.07
molpol	12.57	8.41	7.24, avg	9.41
rms molpol	0.03	0.26	7.24, avg	2.33

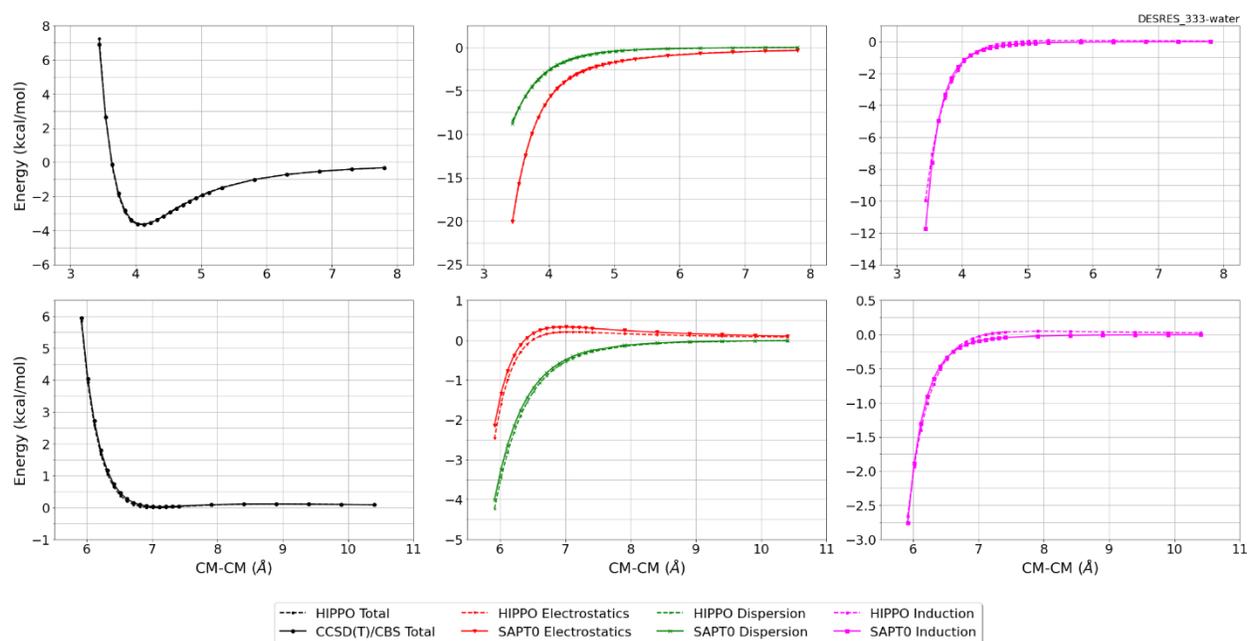


Monomer potential fitting RMS: 1.09

##Dimer results - Fitting to QM datasets##

DESRES_333-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.120	0.096	0.3582	46	0



#334 1,2-Bis(methylsulfanyl)ethane C4H10S2 CID: 23110

ref molpol 18.59 12.83 0.00, avg 10.47
molpol 18.61 12.49 10.98, avg 14.03
rms molpol 0.02 0.34 10.98, avg 3.55

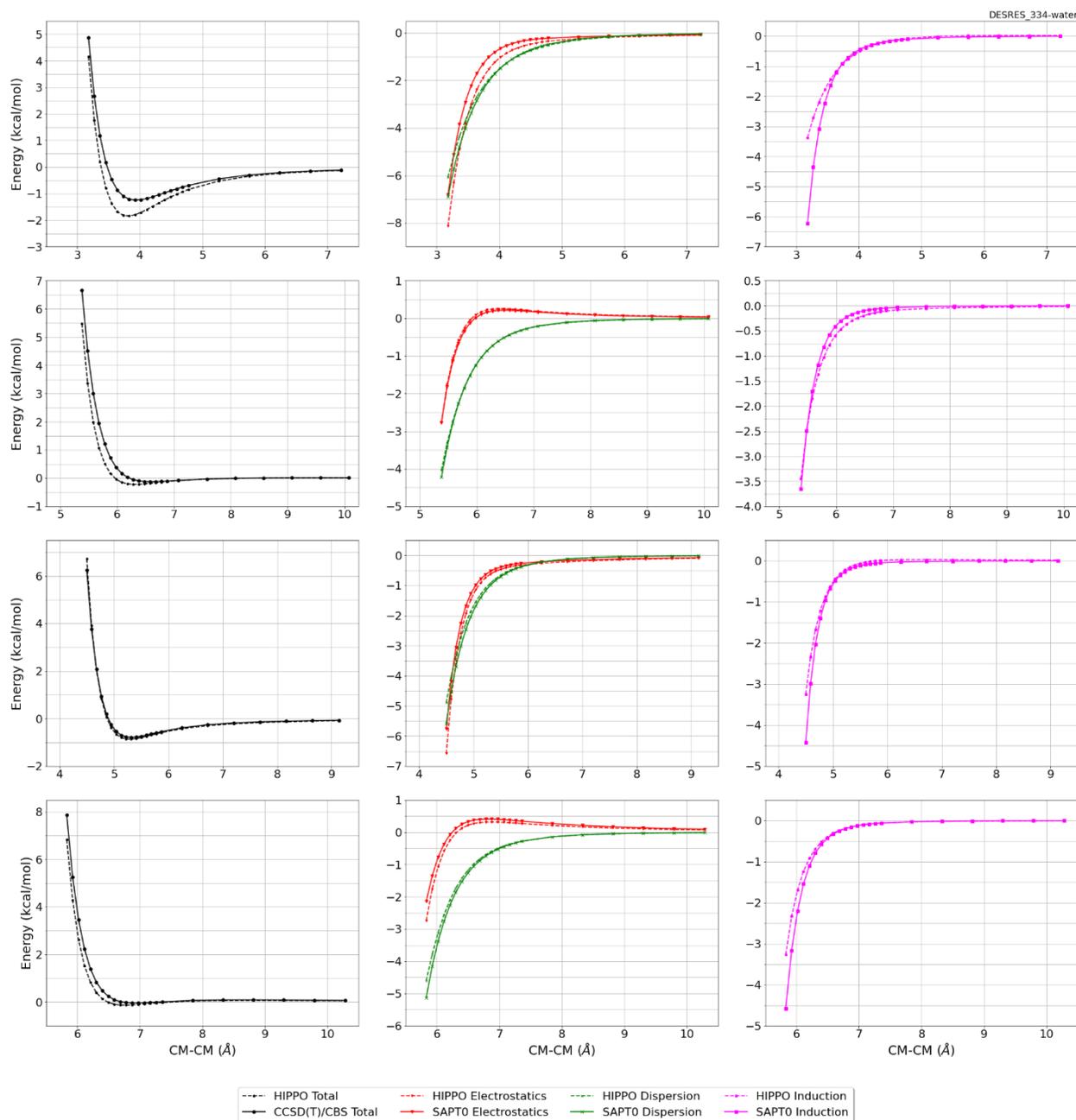


Monomer potential fitting RMS: 0.42

##Dimer results - Fitting to QM datasets##

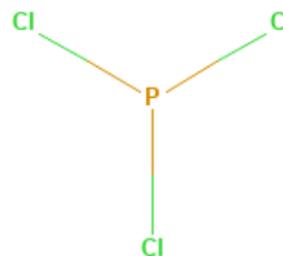
DESRES_334-water, energy values in kcal/mol

MAE 0.308 Std error 0.373 max error 1.2563 #points 92 #count[err > 1] 9



#336 Trichlorophosphane Cl3P CID: 24387

ref molpol	8.31	11.15	11.15, avg	10.20
molpol	8.66	11.12	11.12, avg	10.30
rms molpol	0.35	0.03	0.03, avg	0.10



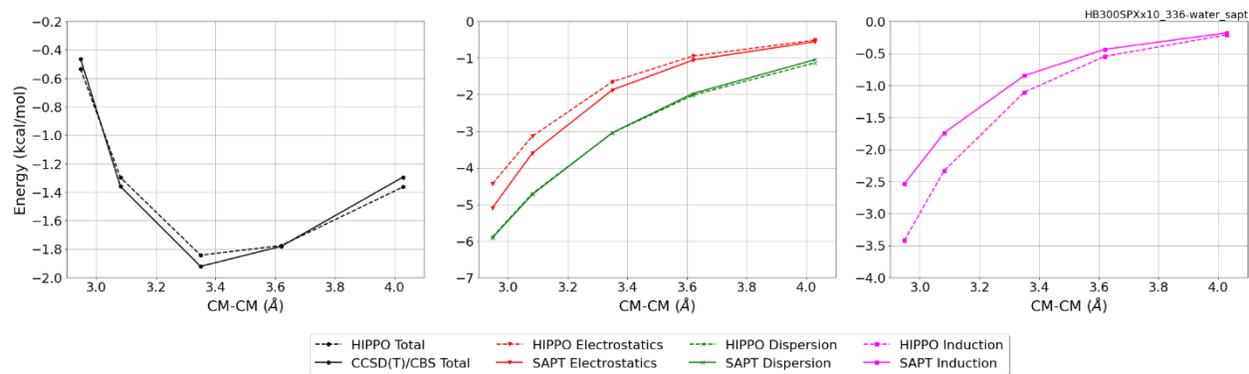
Monomer potential fitting RMS: 0.27

##Dimer results - Fitting to QM datasets##

HB300SPXx10_336-water, energy values in kcal/mol

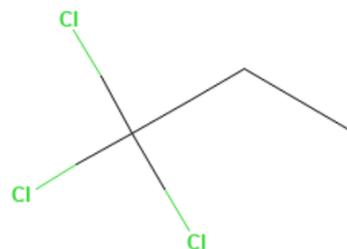
CM-CM (Å)	Reference	HIPPO res	Abs diff
2.489	1.179	0.794	-0.3850
2.606	-0.466	-0.535	-0.0693
2.724	-1.360	-1.297	0.0626
2.843	-1.784	-1.689	0.0954
2.961	-1.922	-1.843	0.0789
3.081	-1.896	-1.852	0.0441
3.200	-1.781	-1.776	0.0054
3.561	-1.295	-1.364	-0.0693
4.166	-0.684	-0.762	-0.0781
5.385	-0.231	-0.258	-0.0274

MAE	Std error	max error	#points	#count[err > 1]
0.092	0.101	0.3850	10	0



#341 1,1,1-Trichloropropane C3H5Cl3 CID: 24622

ref molpol	10.97	12.40	0.00, avg	7.79
molpol	11.01	10.44	9.95, avg	10.47
rms molpol	0.04	1.96	9.95, avg	2.68

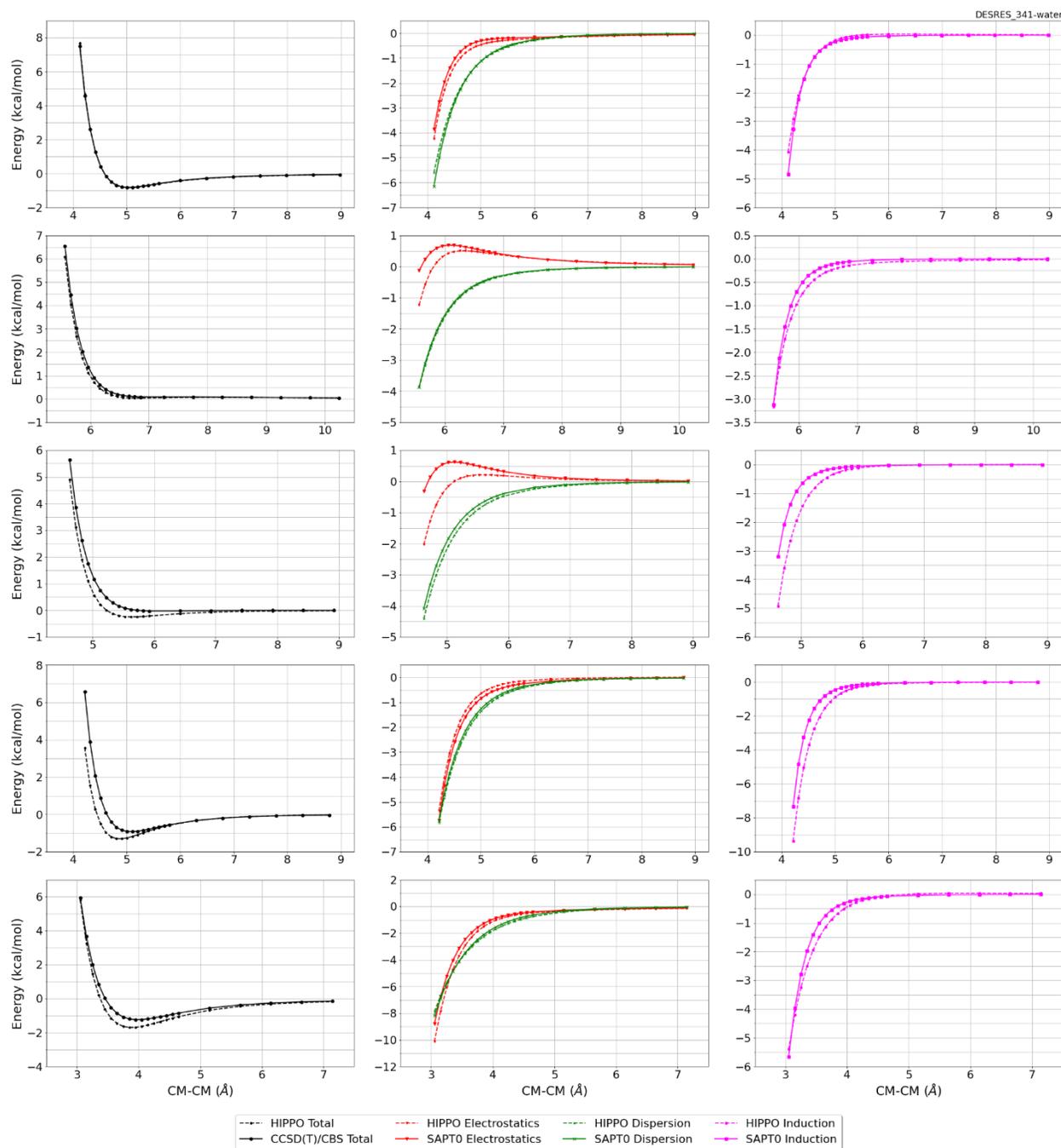


Monomer potential fitting RMS: 0.45

##Dimer results - Fitting to QM datasets##

DESRES_341-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.302	0.443	2.9980	109	5



#346 4-Ethylphenol C8H10O CID: 31242

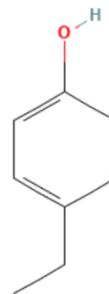
ref molpol	19.15	14.84	10.15, avg	14.71
molpol	19.18	14.86	9.38, avg	14.47
rms molpol	0.03	0.01	0.76, avg	0.24

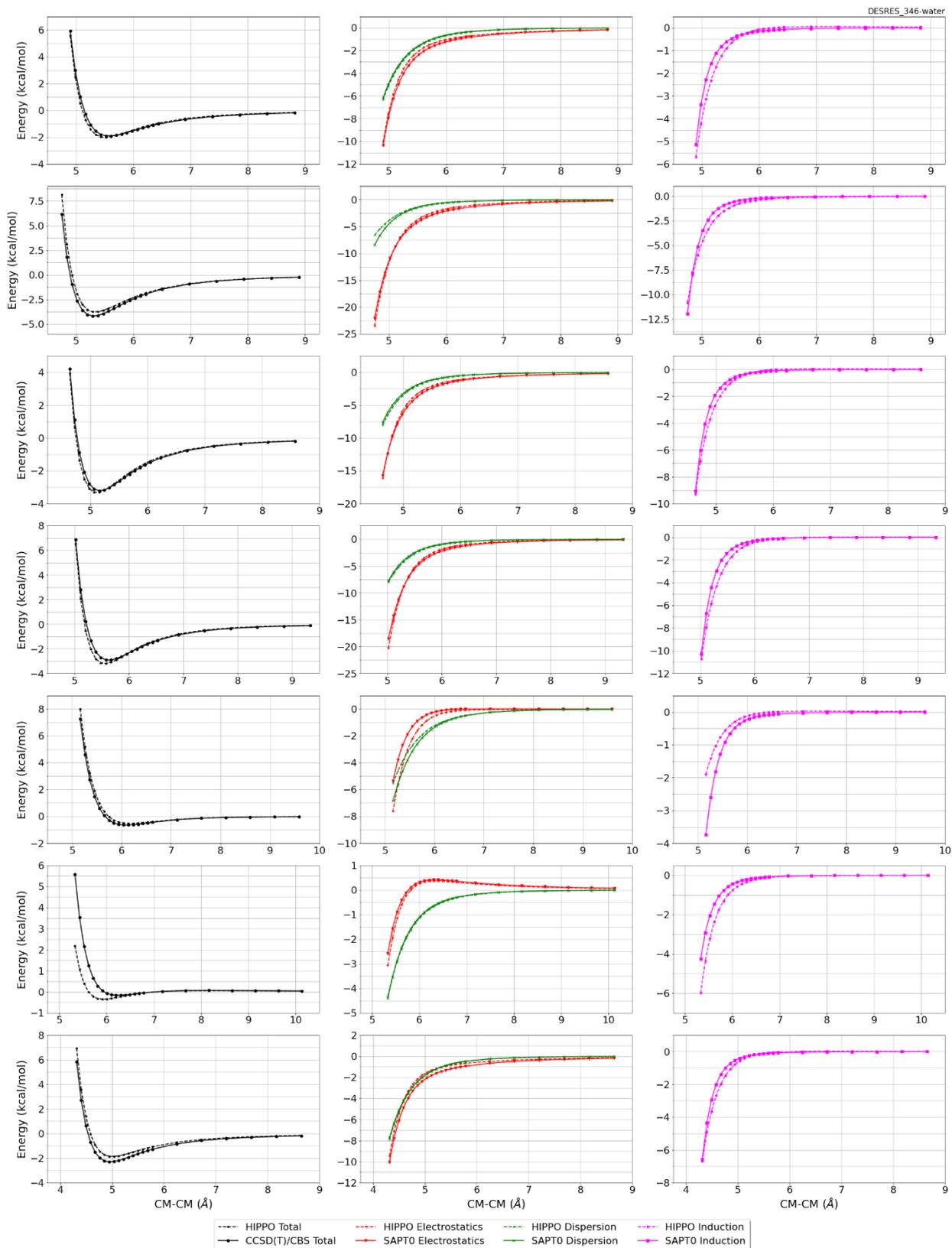
Monomer potential fitting RMS: 0.27

##Dimer results - Fitting to QM datasets##

DESRES_346-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.220	0.327	3.3462	530	15





#347 1,4-Dioxane C4H8O2 CID: 31275

ref molpol	9.25	7.53	7.93, avg	8.24
molpol	9.21	8.00	7.26, avg	8.16
rms molpol	0.05	0.48	0.67, avg	0.08

Monomer potential fitting RMS: 0.25

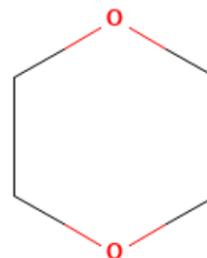
##Dimer results - Fitting to QM datasets##

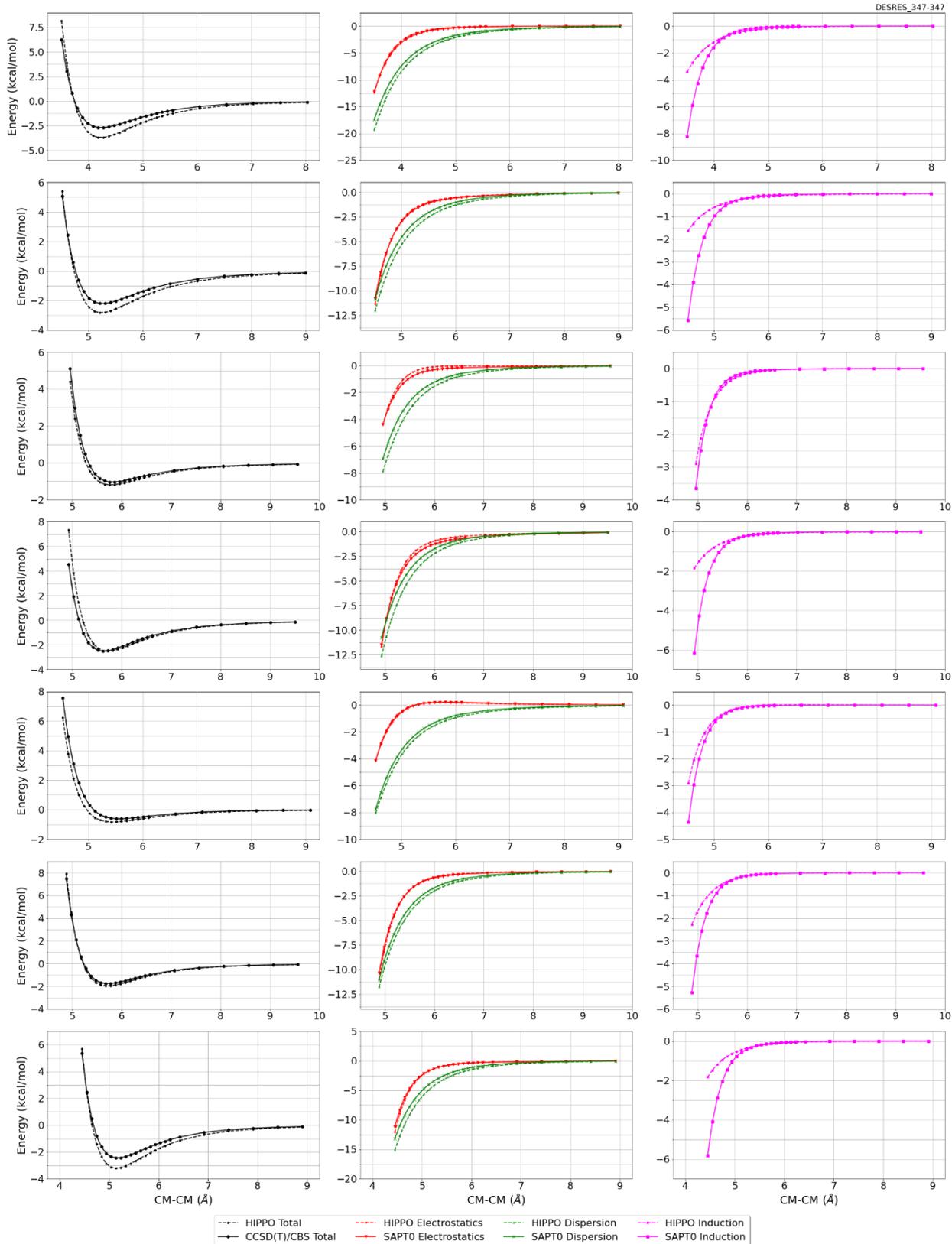
DESRES_347-water, energy values in kcal/mol

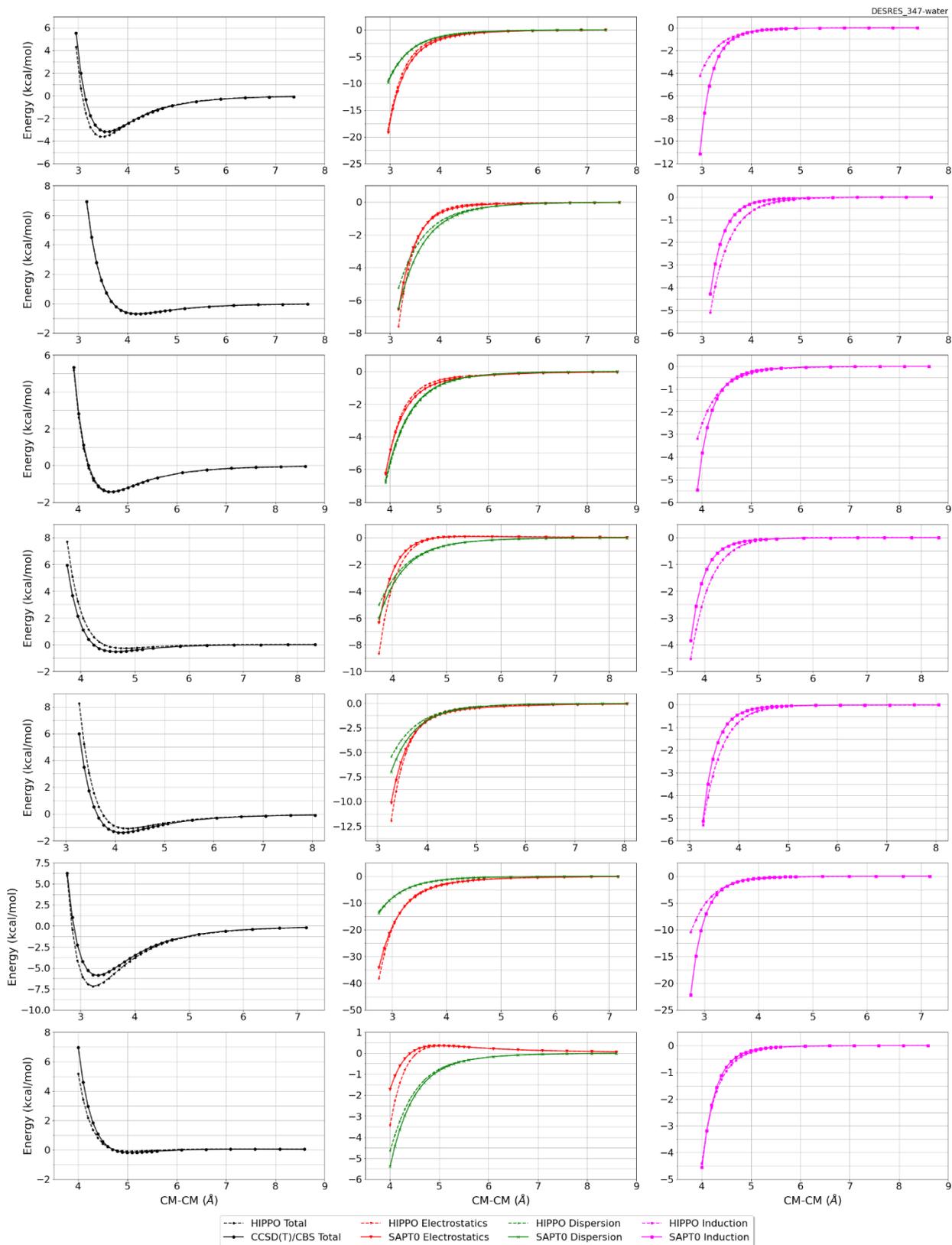
MAE	Std error	max error	#points	#count[err > 1]
0.166	0.311	2.2679	565	20

DESRES_347-347, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.292	0.353	2.8809	477	21







#350 Methoxymethanol C2H6O2 CID: 62540

ref molpol	5.24	6.69	0.00, avg	3.98
molpol	5.25	3.97	3.75, avg	4.32
rms molpol	0.00	2.73	3.75, avg	0.34

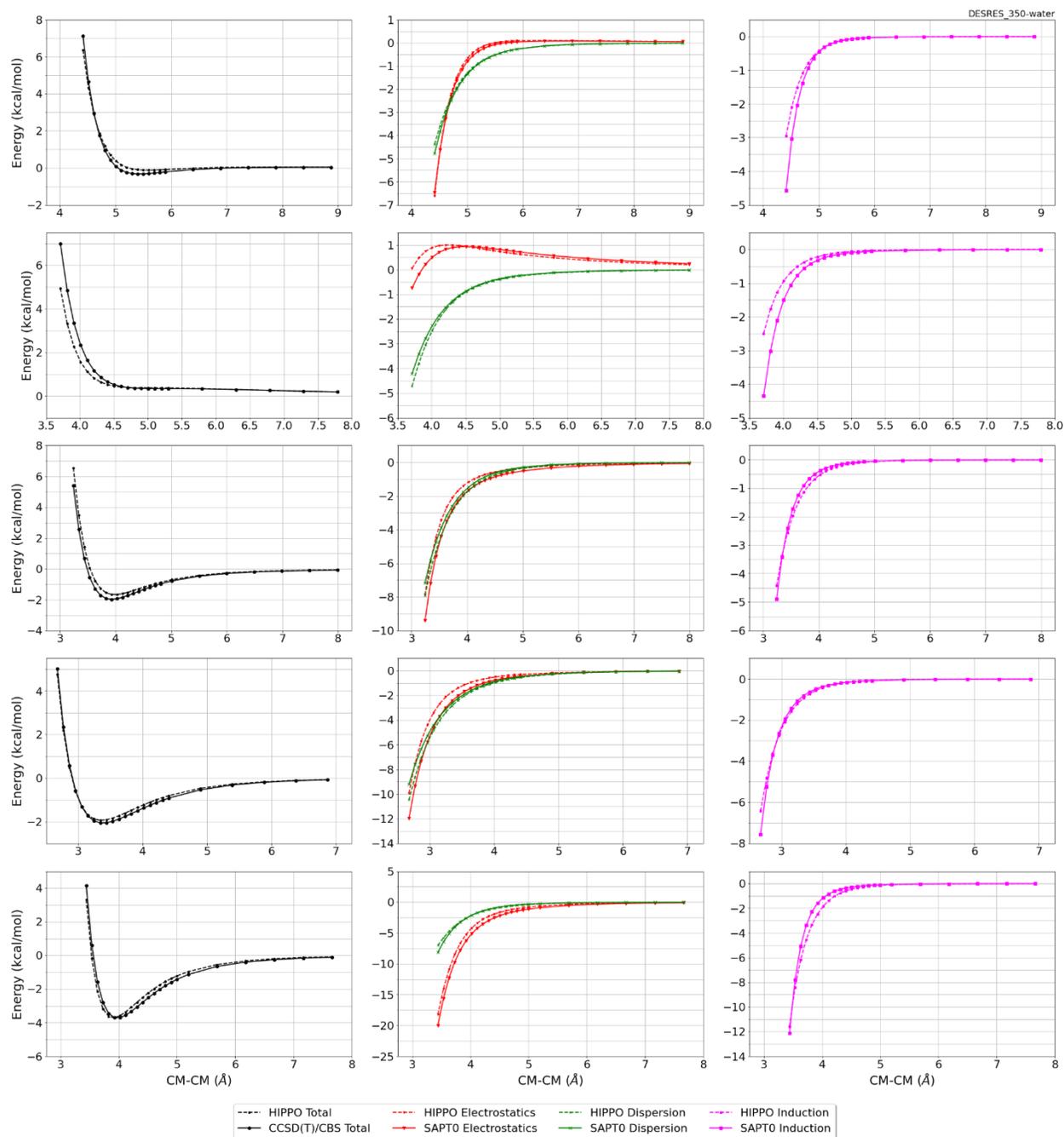


Monomer potential fitting RMS: 1.28

##Dimer results - Fitting to QM datasets##

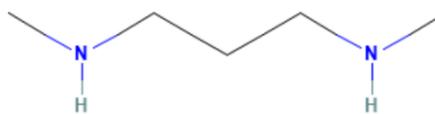
DESRES_350-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.204	0.287	2.0687	265	8



#351 N,N'-Dimethylpropane-1,3-diamine C5H14N2 CID: 66978

ref molpol 15.24 11.54 10.45, avg 12.41
molpol 15.26 10.74 10.55, avg 12.19
rms molpol 0.03 0.80 0.10, avg 0.22

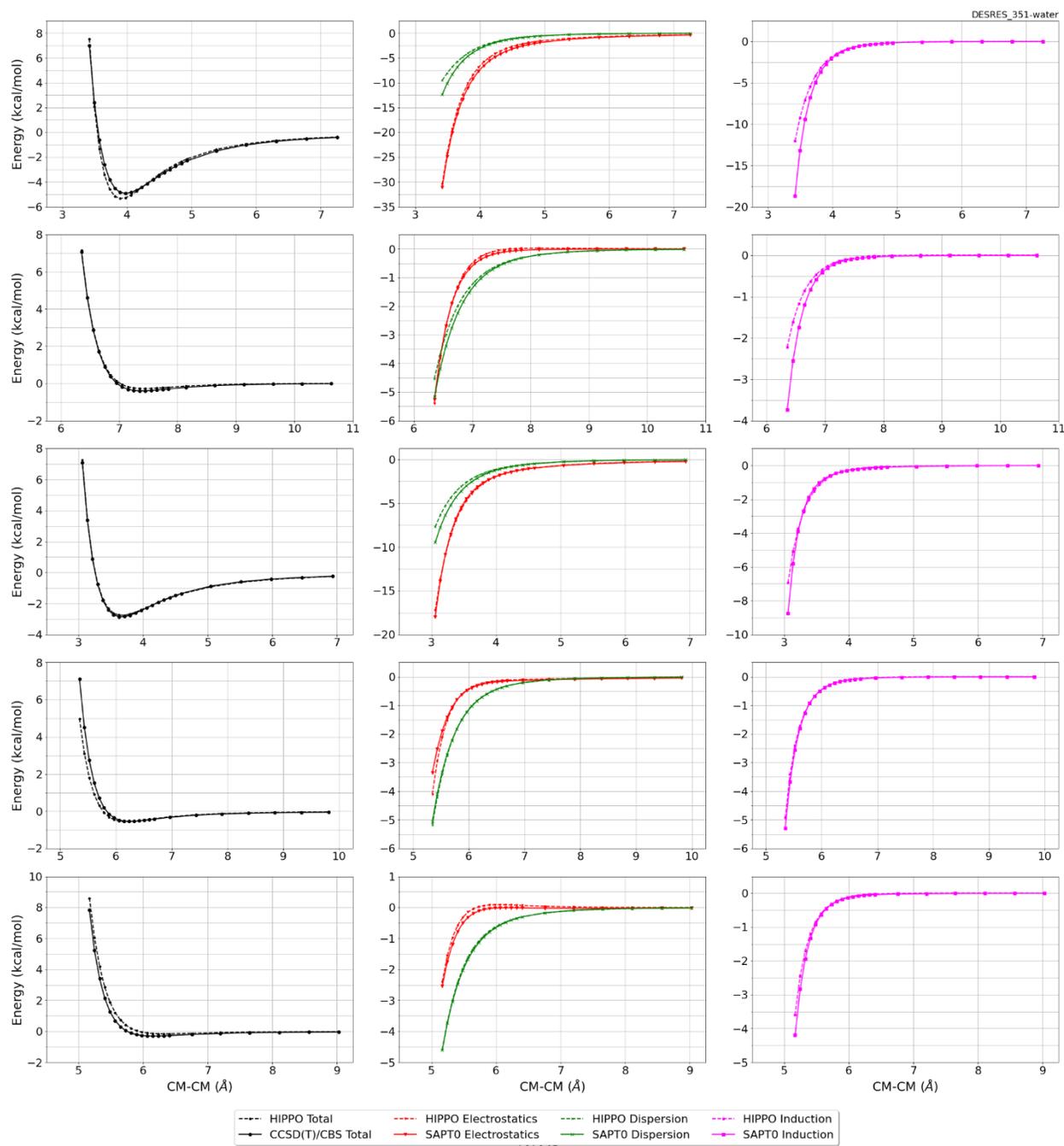


Monomer potential fitting RMS: 0.82

##Dimer results - Fitting to QM datasets##

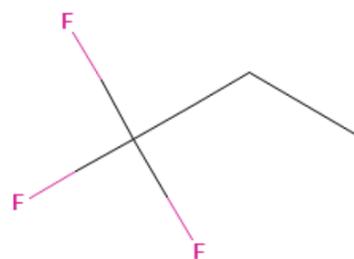
DESRES_351-water, energy values in kcal/mol

MAE 0.180 Std error 0.306 max error 2.1924 #points 115 #count[err > 1] 2



#353 1,1,1-Trifluoropropane C3H5F3 CID: 67899

ref molpol	6.55	5.84	5.69, avg	6.03
molpol	6.55	5.83	5.70, avg	6.03
rms molpol	0.01	0.01	0.01, avg	0.00



Monomer potential fitting RMS: 0.14

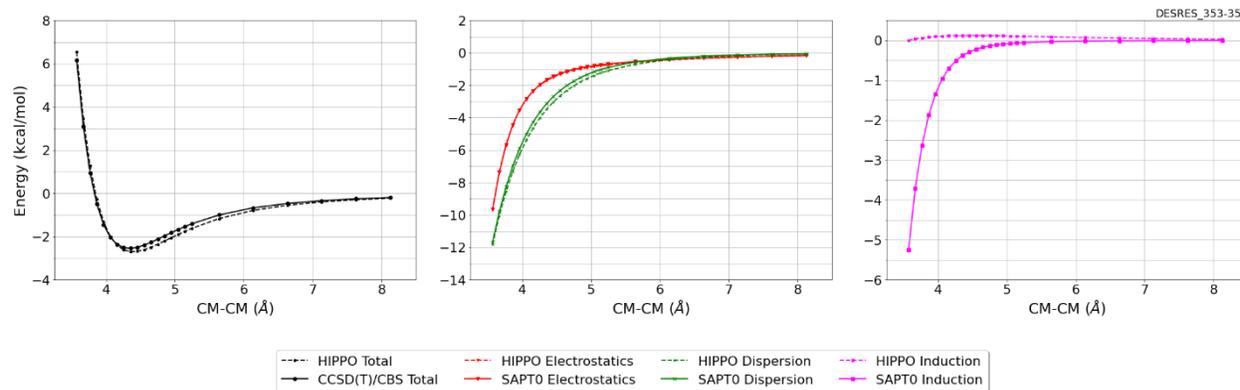
##Dimer results - Fitting to QM datasets##

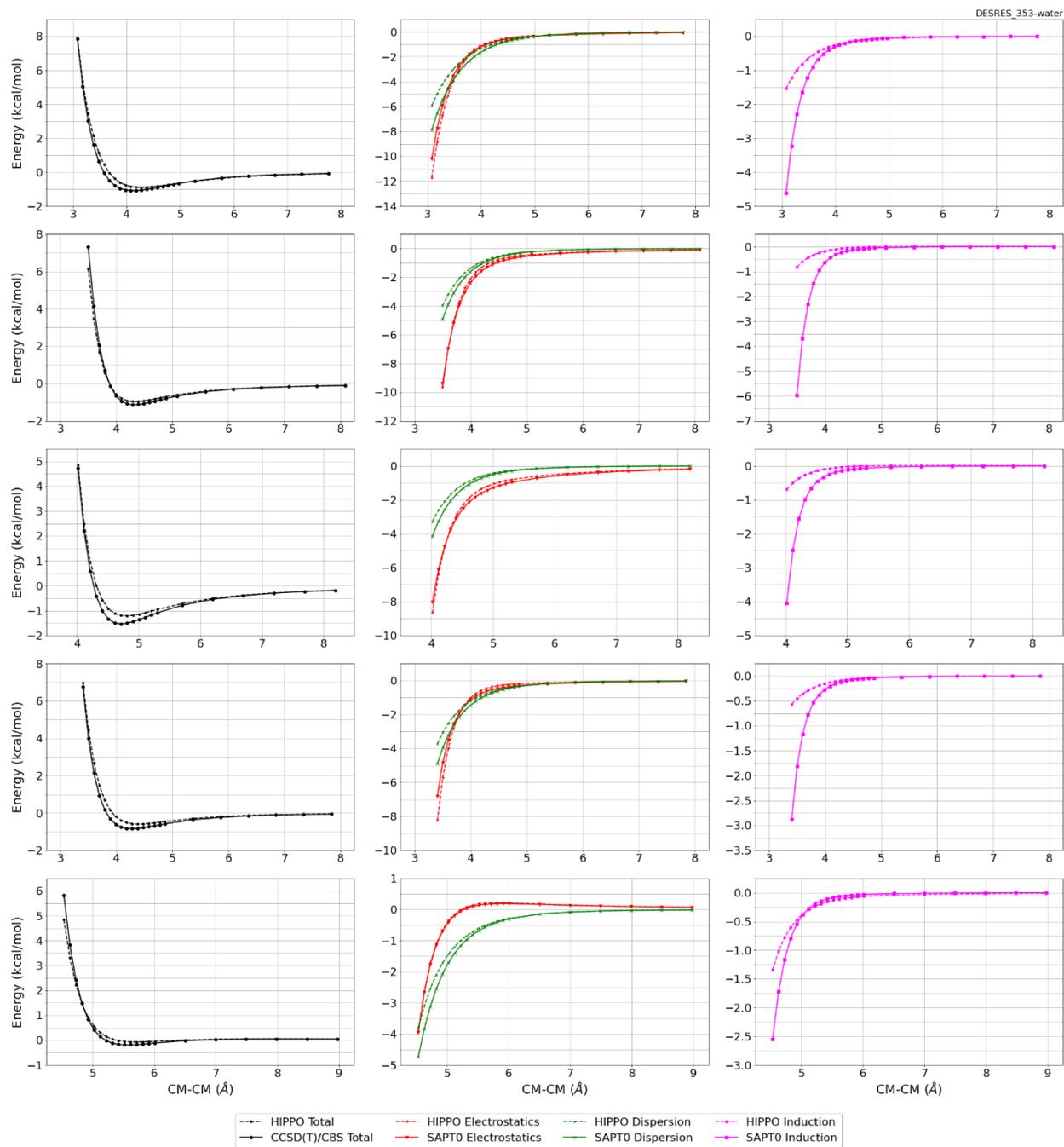
DESRES_353-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.144	0.177	1.2107	262	1

DESRES_353-353, energy values in kcal/mol

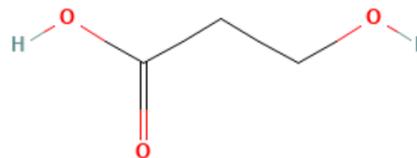
MAE	Std error	max error	#points	#count[err > 1]
0.204	0.116	0.3438	24	0





#354 3-Hydroxypropanoic Acid C3H6O3 CID: 68152

ref molpol	7.65	8.79	0.00, avg	5.48
molpol	7.73	7.63	5.37, avg	6.91
rms molpol	0.09	1.16	5.37, avg	1.43

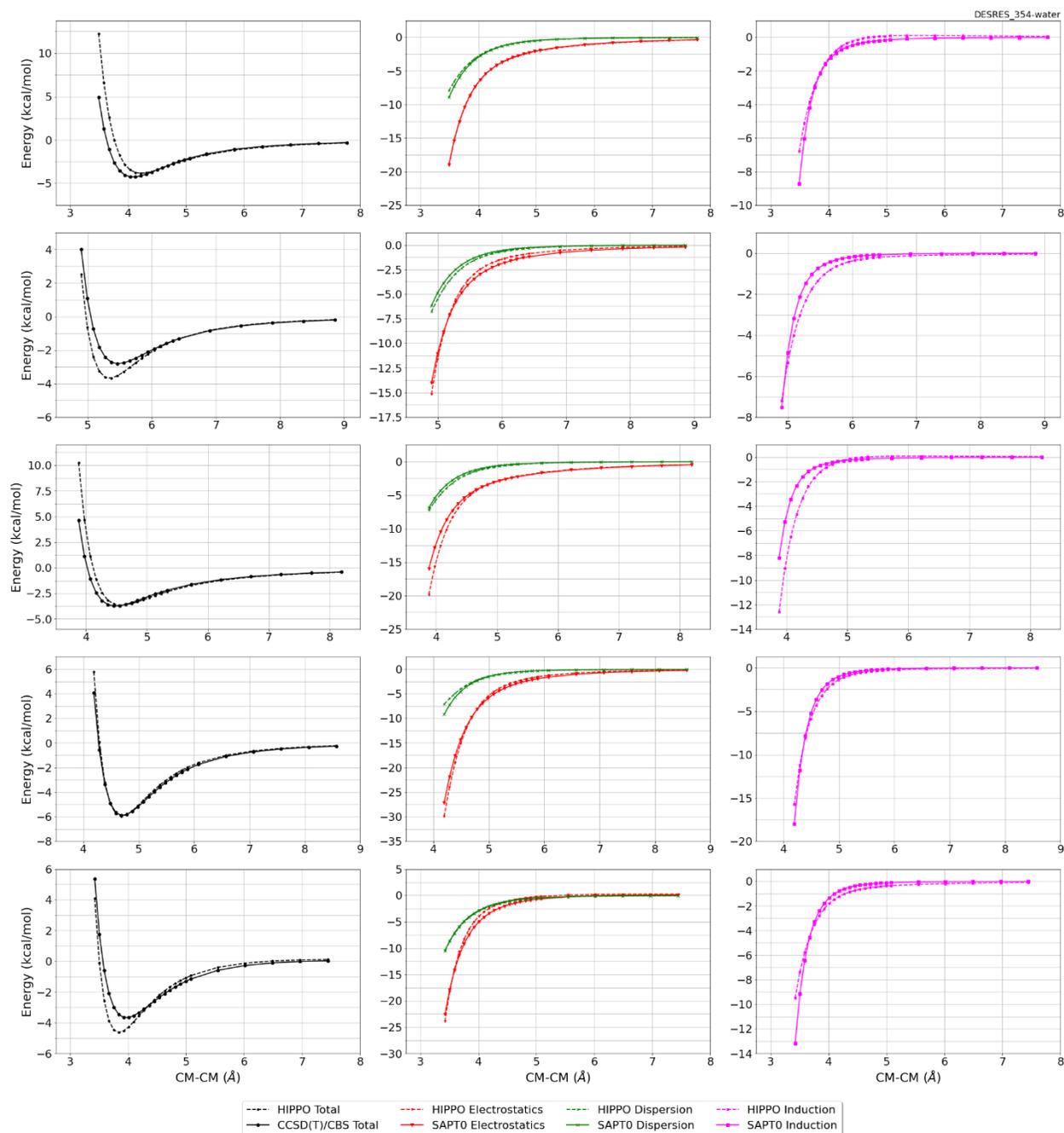


Monomer potential fitting RMS: 0.72

##Dimer results - Fitting to QM datasets##

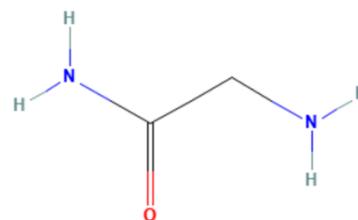
DESRES_354-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.363	0.728	6.9796	267	23



#357 2-Aminoacetamide C₂H₆N₂O CID: 69020

ref molpol	8.40	7.25	5.66, avg	7.10
molpol	8.40	7.25	5.66, avg	7.10
rms molpol	0.00	0.01	0.00, avg	0.00

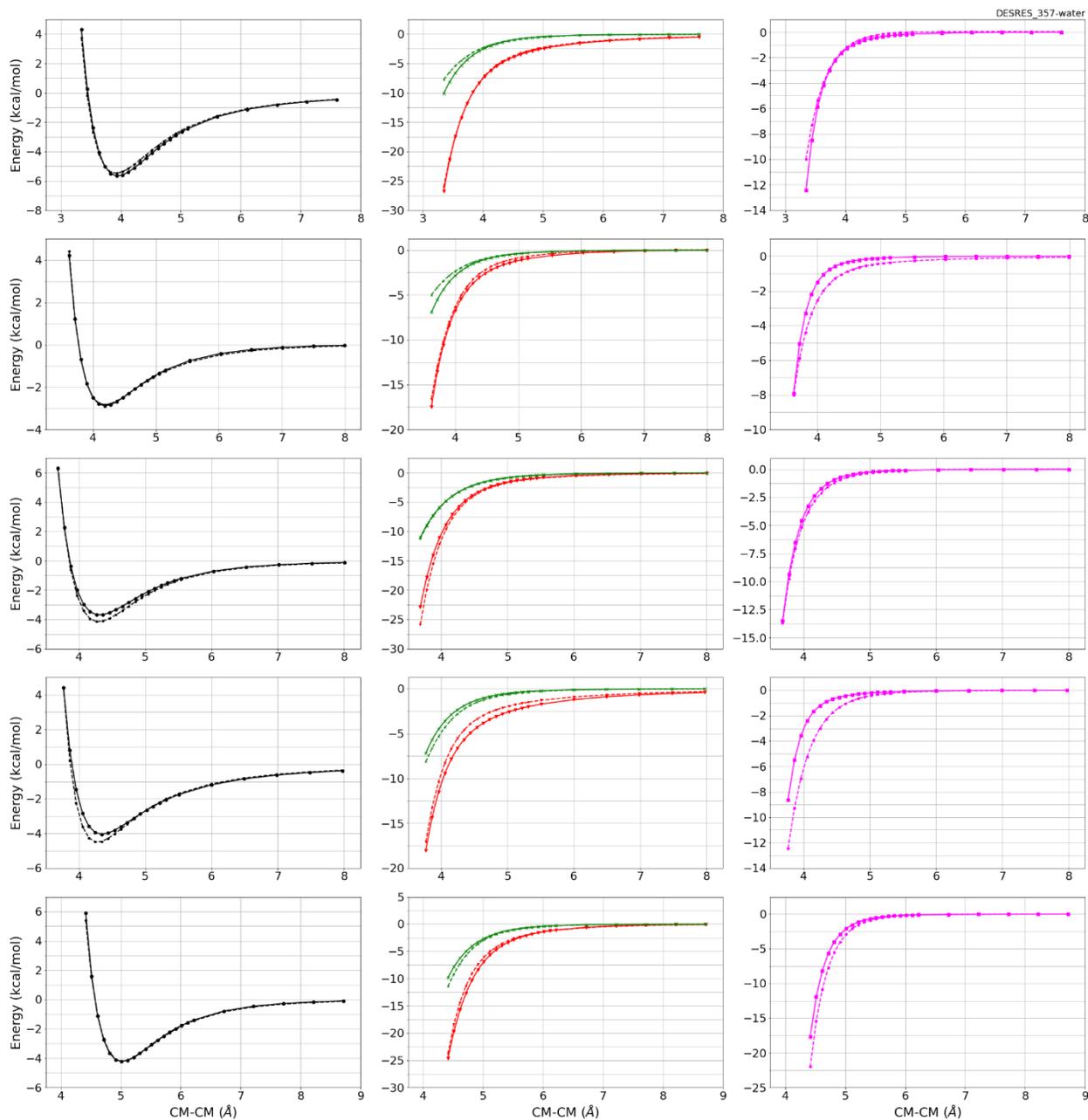


Monomer potential fitting RMS: 1.47

##Dimer results - Fitting to QM datasets##

DESRES_357-water, energy values in kcal/mol

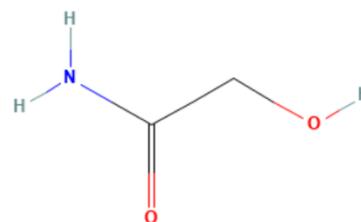
MAE	Std error	max error	#points	#count[err > 1]
0.196	0.263	1.8906	269	5



--- HIPPO Total	--- HIPPO Electrostatics	--- HIPPO Dispersion	--- HIPPO Induction
--- CCSD(T)/CBS Total	--- SAPTO Electrostatics	--- SAPTO Dispersion	--- SAPTO Induction

#358 2-Hydroxyacetamide C2H5NO2 CID: 69021

ref molpol	7.61	6.89	0.00, avg	4.83
molpol	7.59	6.85	4.87, avg	6.44
rms molpol	0.02	0.04	4.87, avg	1.60

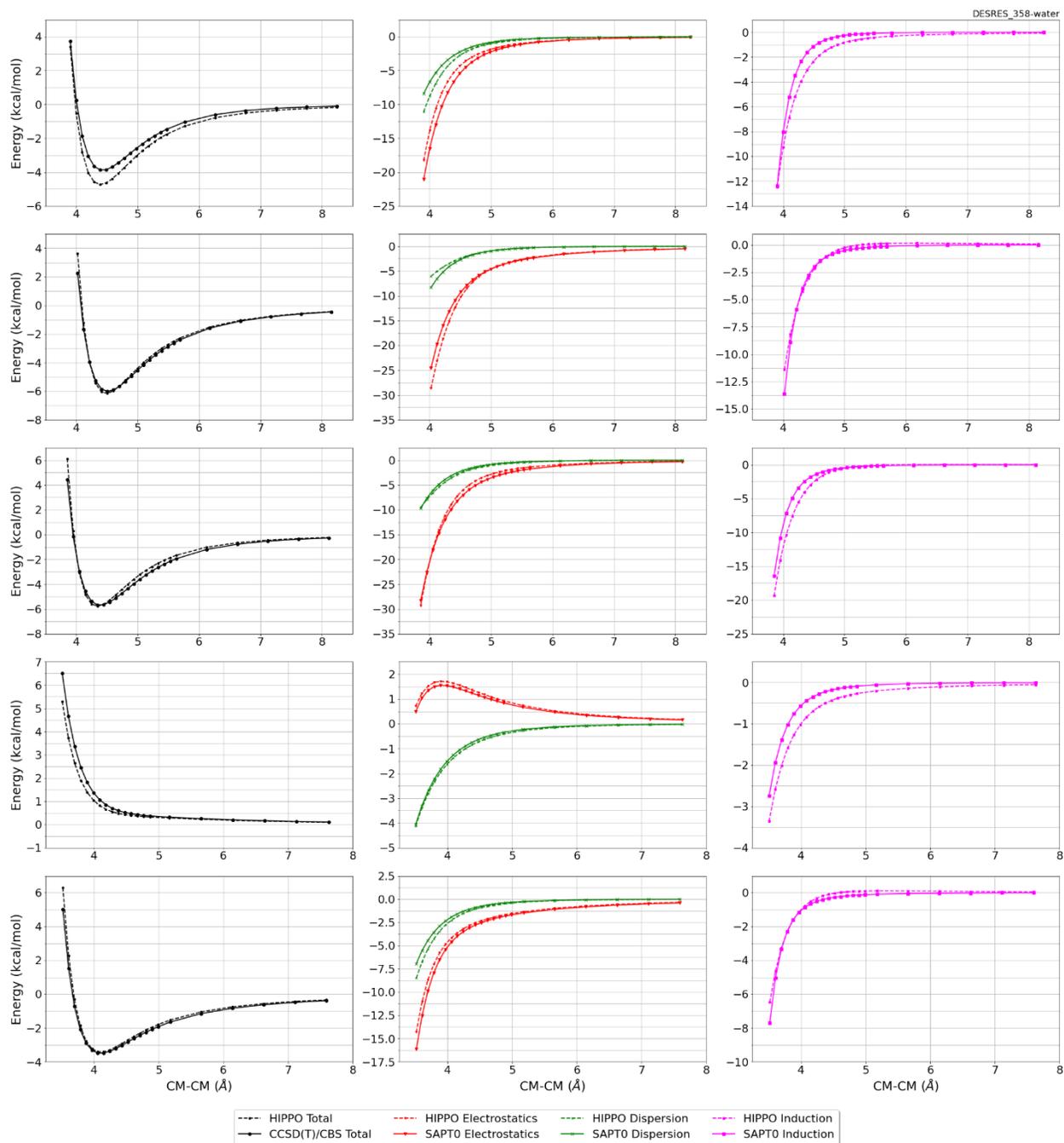


Monomer potential fitting RMS: 0.48

##Dimer results - Fitting to QM datasets##

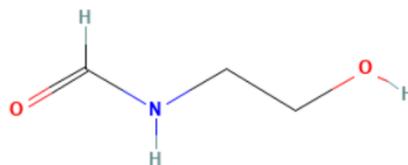
DESRES_358-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.194	0.271	1.6076	264	7



#360 N-(2-Hydroxyethyl)formamide C3H7NO2 CID: 69657

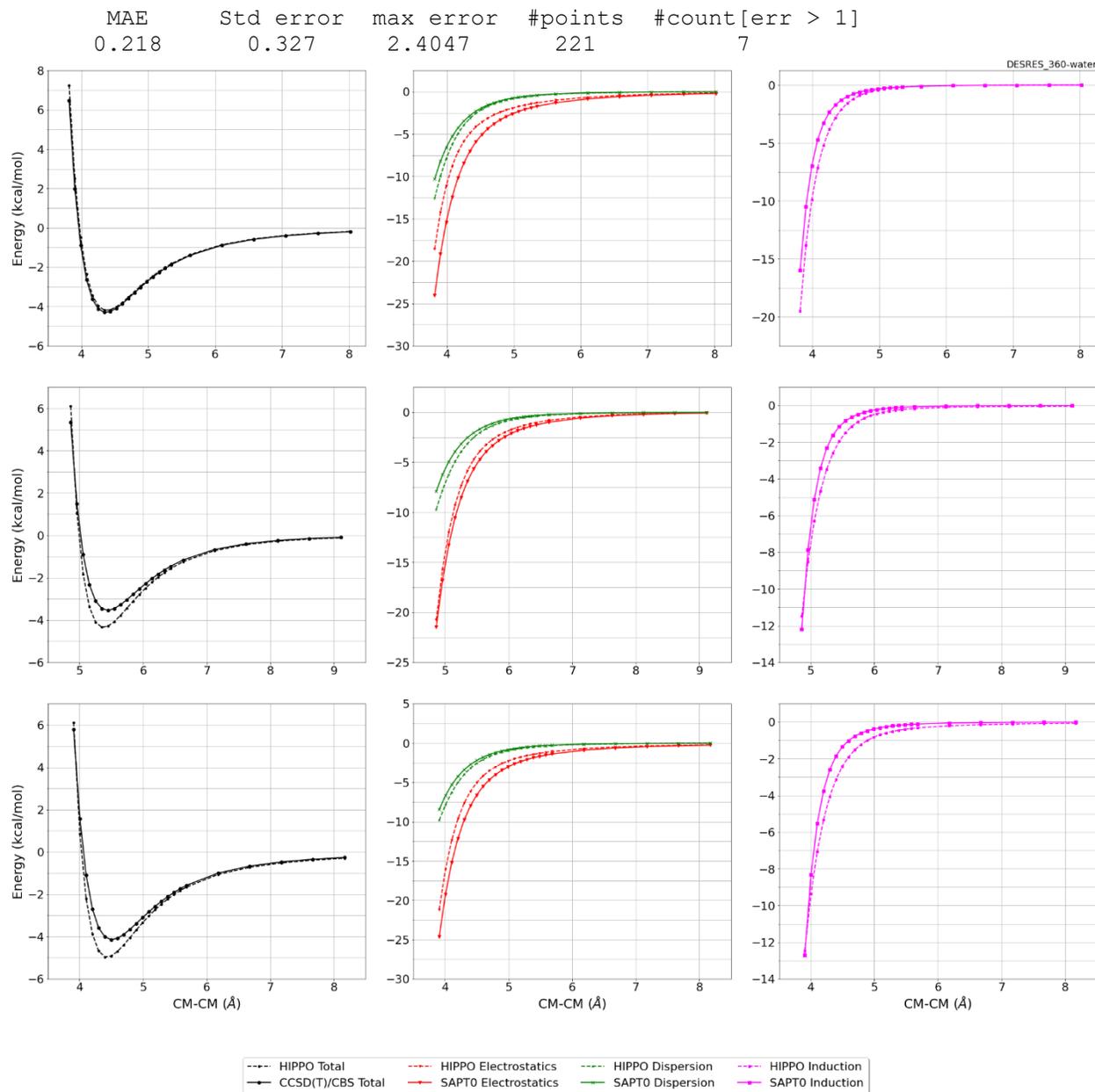
ref molpol	11.20	7.74	0.00, avg	6.31
molpol	11.22	7.53	5.96, avg	8.24
rms molpol	0.02	0.21	5.96, avg	1.92



Monomer potential fitting RMS: 0.29

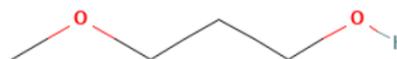
##Dimer results - Fitting to QM datasets##

DESRES_360-water, energy values in kcal/mol



#363 3-Methoxypropan-1-ol C4H10O2 CID: 74116

ref molpol 8.45 11.21 0.00, avg 6.55
molpol 8.46 5.99 5.58, avg 6.68
rms molpol 0.01 5.22 5.58, avg 0.12

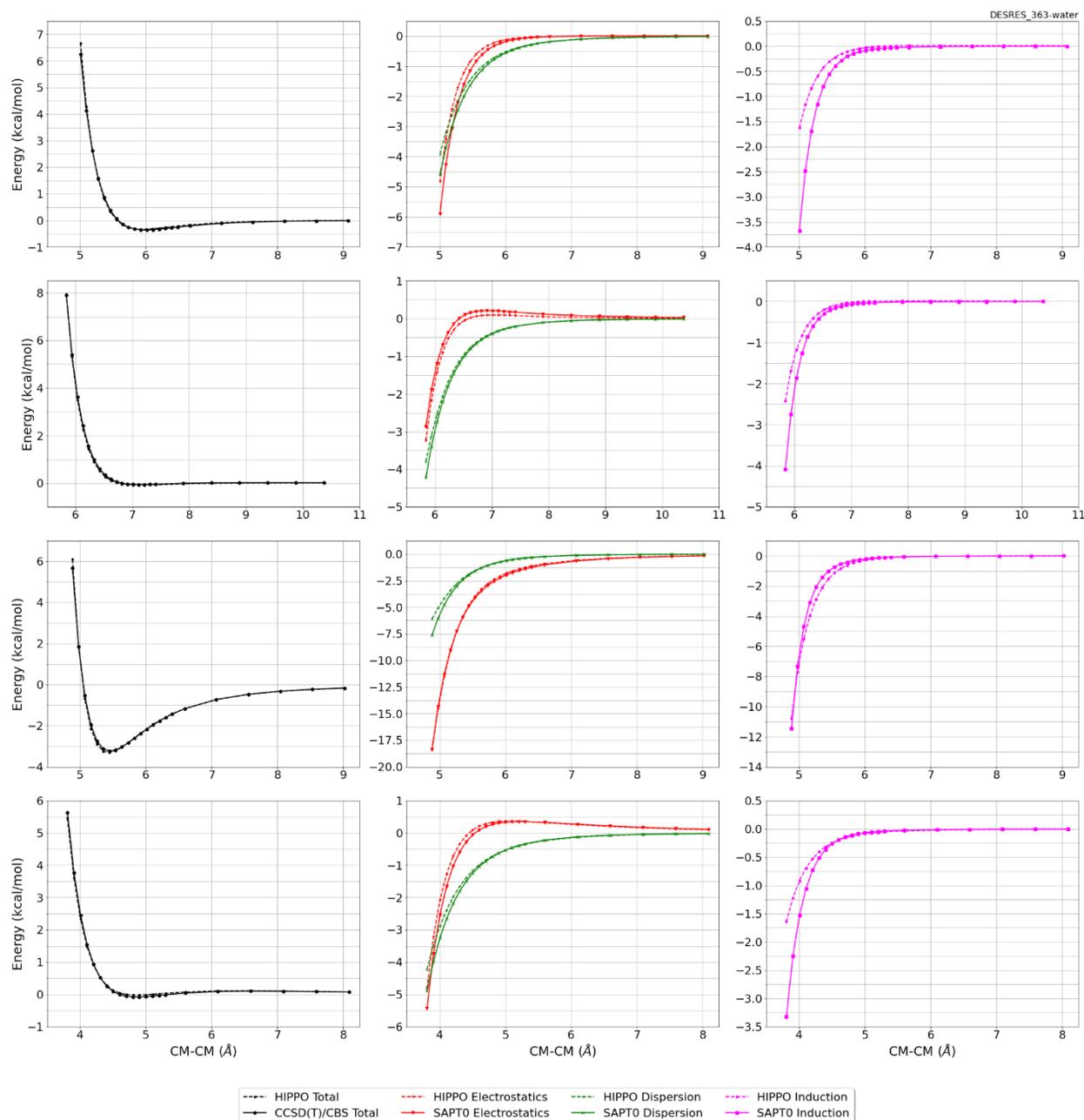


Monomer potential fitting RMS: 0.34

##Dimer results - Fitting to QM datasets##

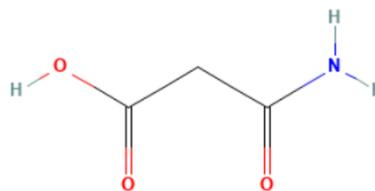
DESRES_363-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.055	0.068	0.4394	91	0



#364 3-Amino-3-oxopropanoic Acid C3H5NO3 CID: 75367

ref molpol	10.05	9.16	0.00, avg	6.40
molpol	10.03	9.13	6.01, avg	8.39
rms molpol	0.02	0.03	6.01, avg	1.99

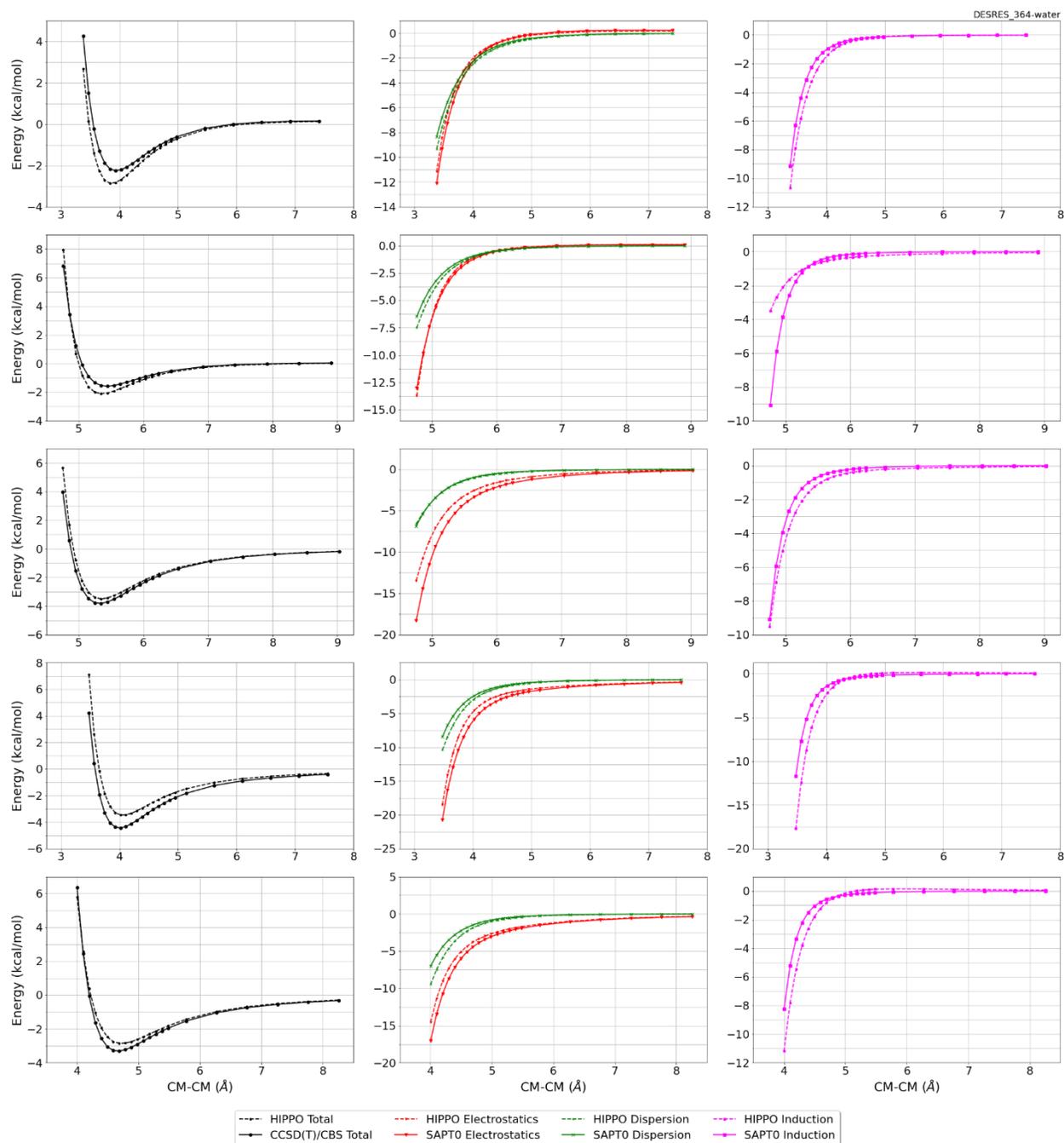


Monomer potential fitting RMS: 1.05

##Dimer results - Fitting to QM datasets##

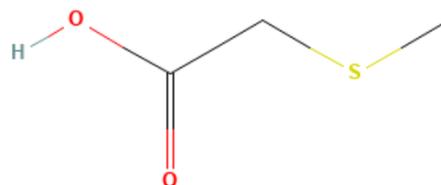
DESRES_364-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.364	0.721	8.2849	262	21



#365 2-Methylsulfanylacetic Acid C3H6O2S CID: 75551

ref molpol	9.77	12.12	0.00, avg	7.30
molpol	9.85	9.84	6.25, avg	8.65
rms molpol	0.08	2.27	6.25, avg	1.35

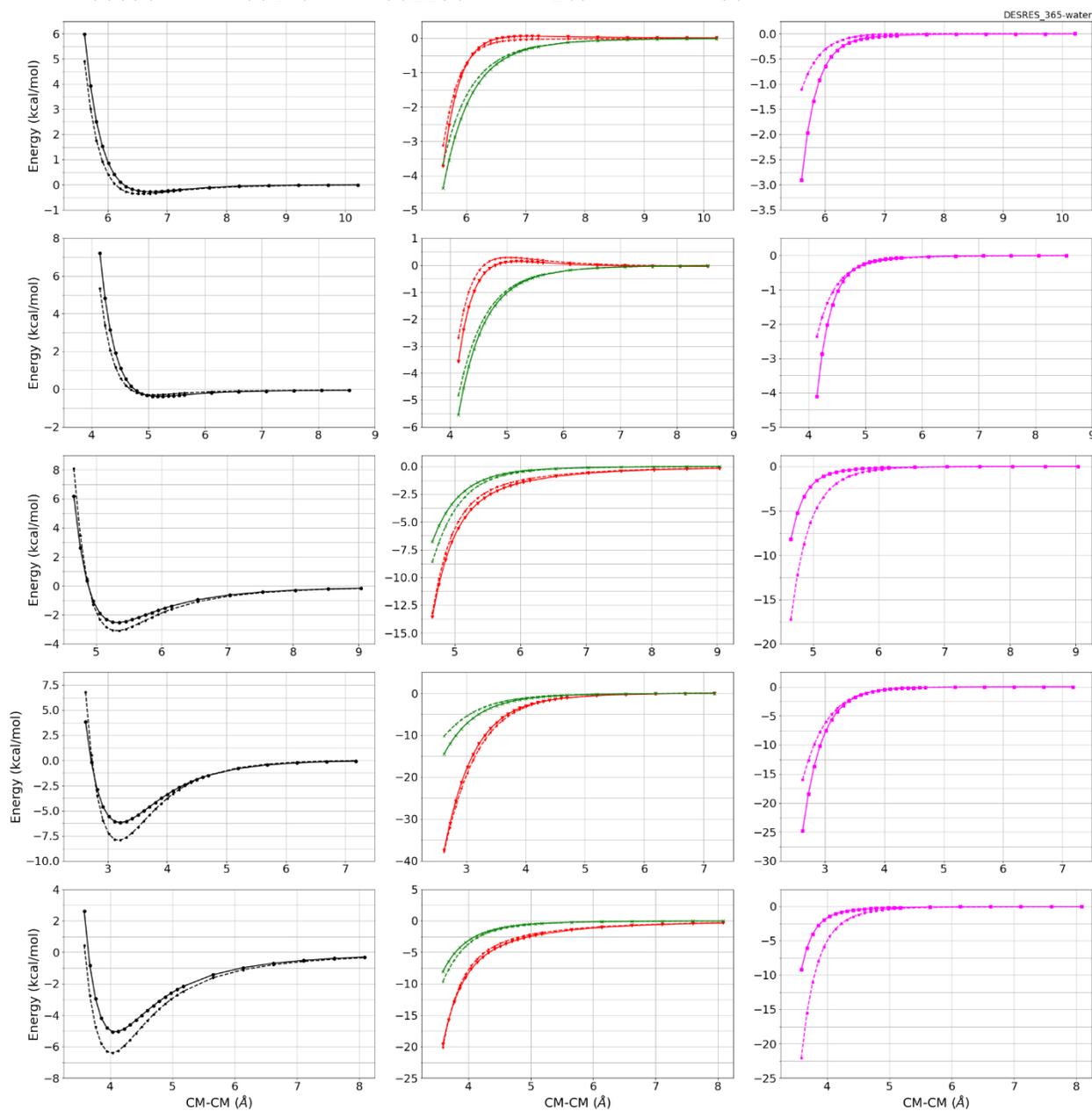


Monomer potential fitting RMS: 3.01

##Dimer results - Fitting to QM datasets##

DESRES_365-water, energy values in kcal/mol

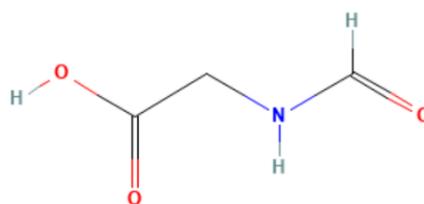
MAE	Std error	max error	#points	#count[err > 1]
0.395	0.470	3.2236	269	30



--- HIPPO Total - - - HIPPO Electrostatics - - - HIPPO Dispersion - - - HIPPO Induction
- - - CCSD(T)/CBS Total - - - SAPTO Electrostatics - - - SAPTO Dispersion - - - SAPTO Induction

#366 2-Formamidoacetic Acid C3H5NO3 CID: 75606

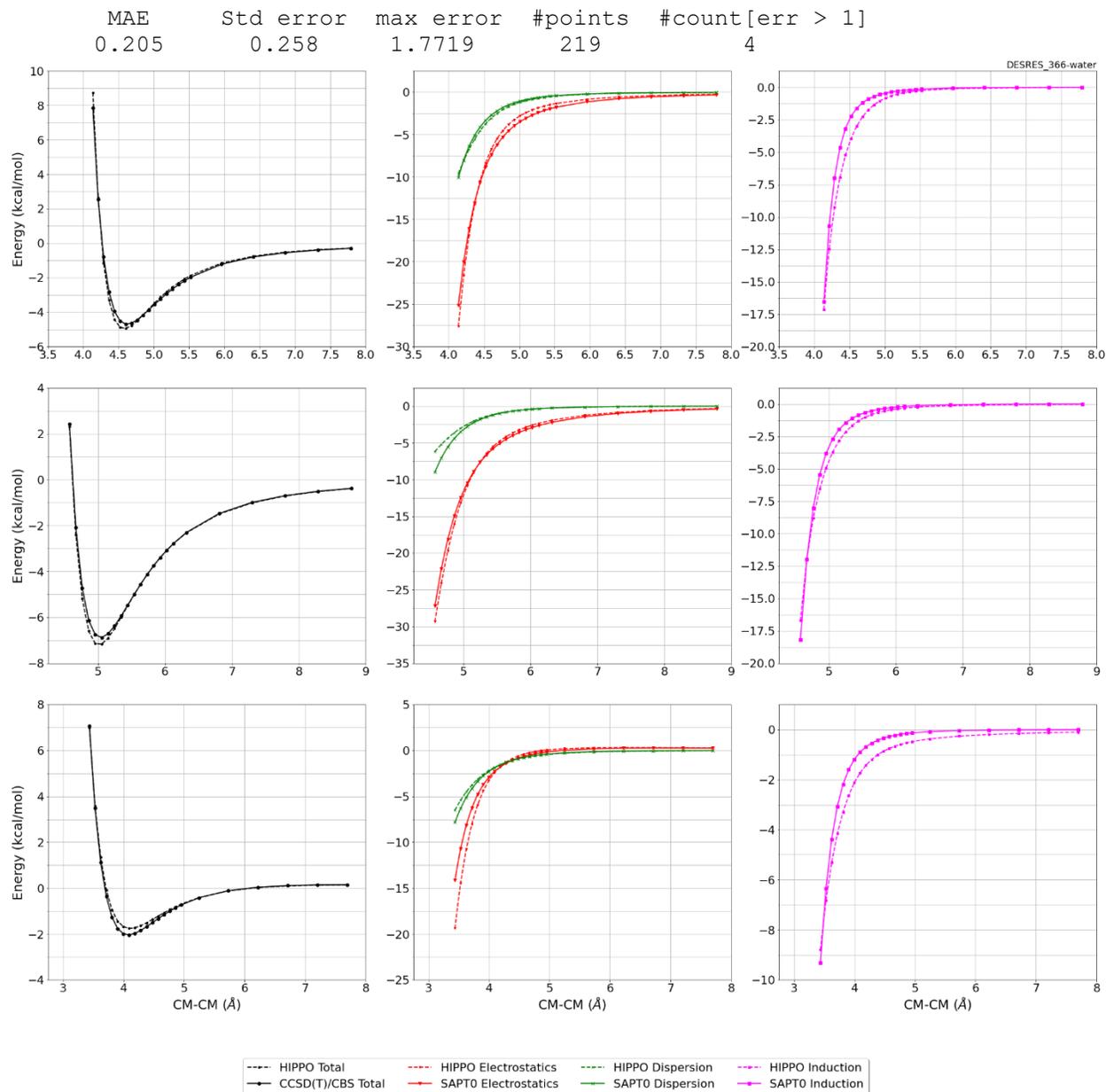
ref molpol	11.45	8.44	0.00, avg	6.63
molpol	11.43	8.41	6.17, avg	8.67
rms molpol	0.02	0.03	6.17, avg	2.04



Monomer potential fitting RMS: 0.47

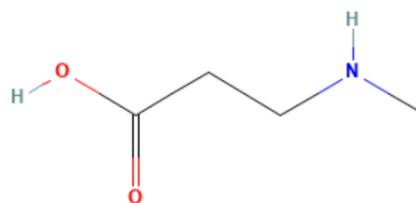
##Dimer results - Fitting to QM datasets##

DESRES_366-water, energy values in kcal/mol



#367 3-(Methylamino)propanoic Acid C4H9NO2 CID: 75891

ref molpol	12.07	9.86	7.91, avg	-9.94
molpol	12.06	9.86	7.92, avg	9.95
rms molpol	0.01	0.01	0.01, avg	0.00

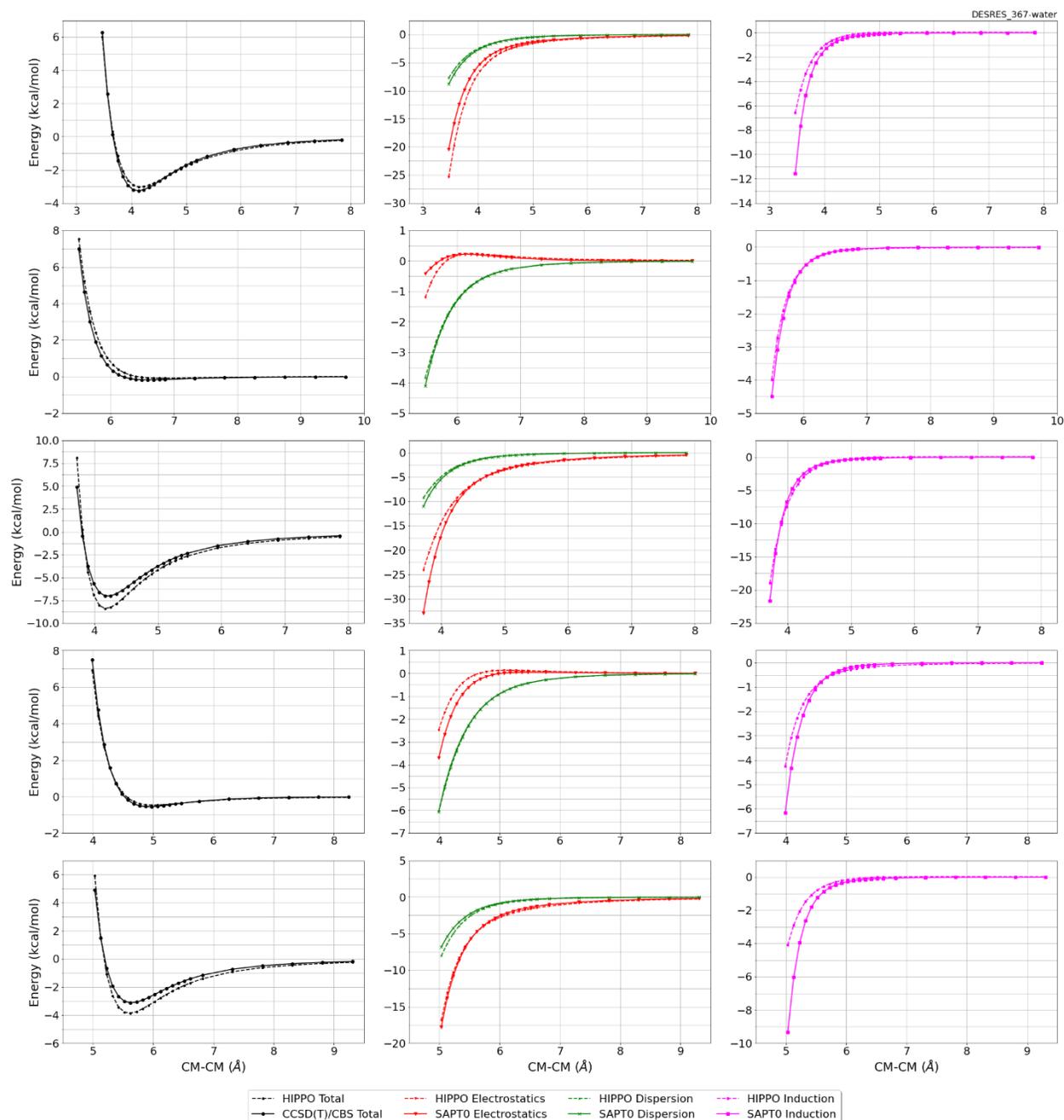


Monomer potential fitting RMS: 0.87

##Dimer results - Fitting to QM datasets##

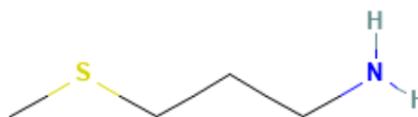
DESRES_367-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.357	0.428	3.0973	117	7



#368 3-Methylsulfanylpropan-1-amine C4H11NS CID: 77743

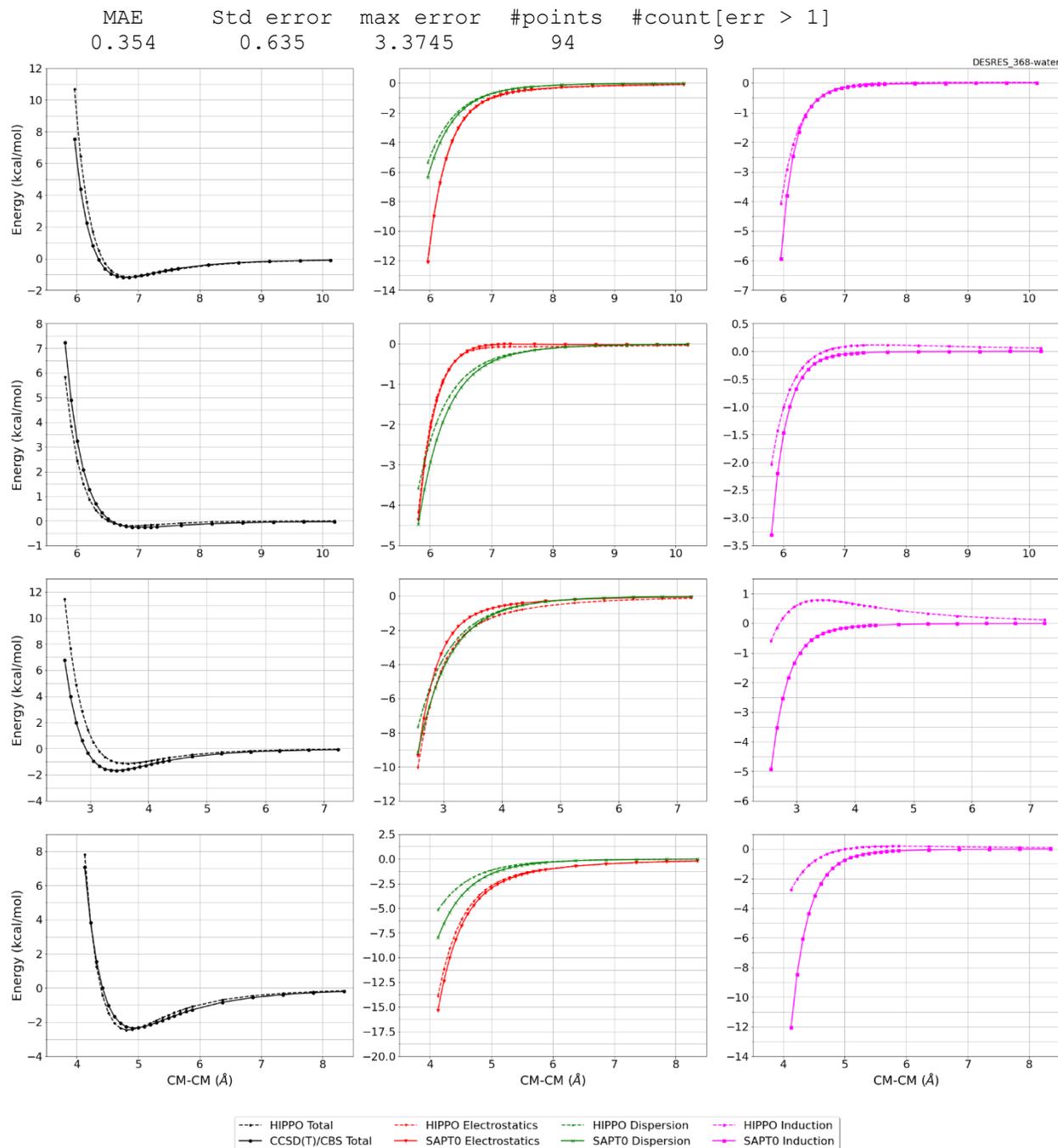
ref molpol 15.44 11.35 10.13, avg 12.31
molpol 15.46 11.35 10.10, avg 12.30
rms molpol 0.02 0.00 0.02, avg 0.00



Monomer potential fitting RMS: 2.23

##Dimer results - Fitting to QM datasets##

DESRES_368-water, energy values in kcal/mol



#369 2-Methylsulfanylethanol C3H8OS CID: 78925

ref molpol	11.97	9.19	0.00, avg	7.05
molpol	11.96	9.09	7.77, avg	9.61
rms molpol	0.01	0.10	7.77, avg	2.55

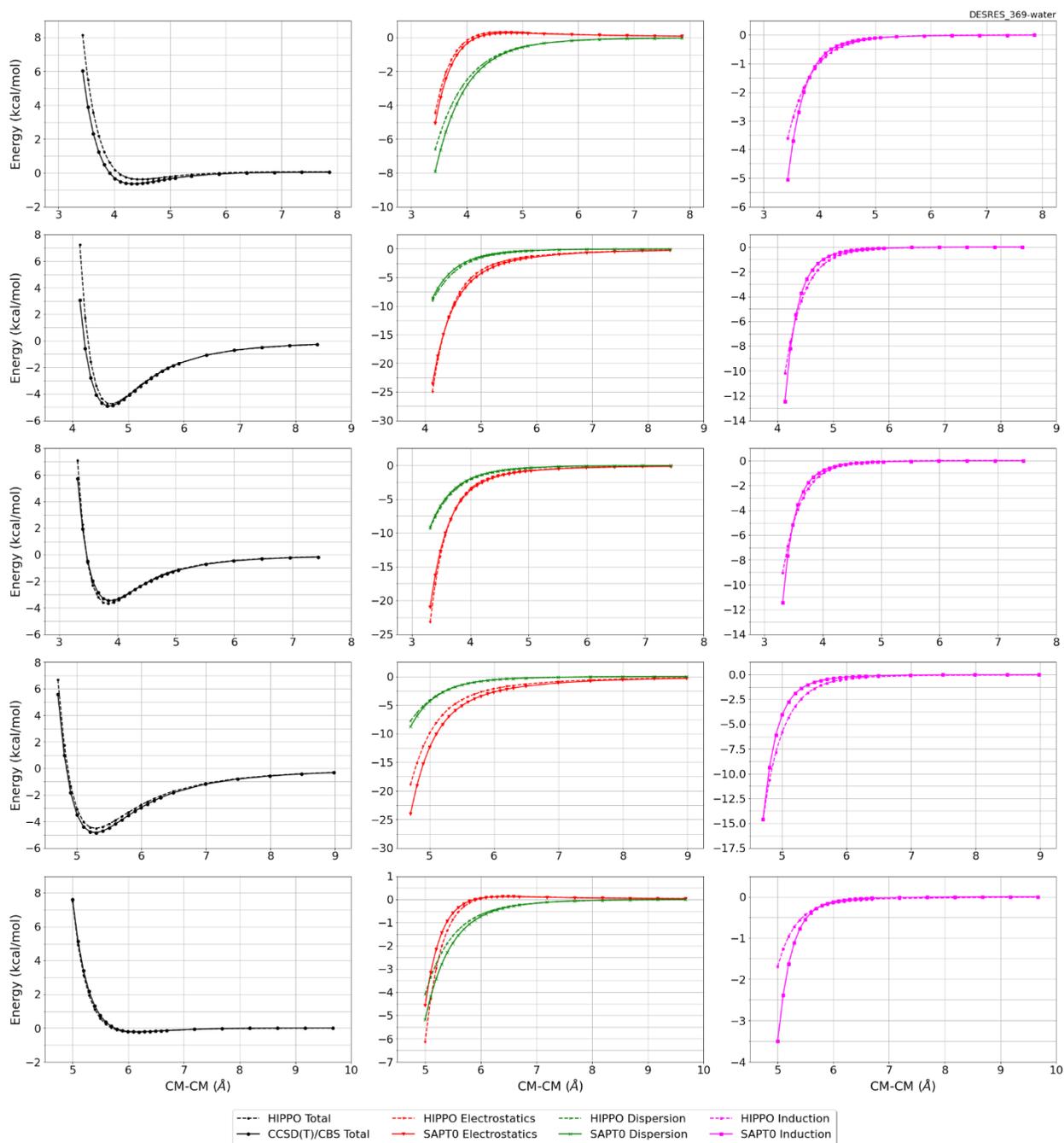


Monomer potential fitting RMS: 0.79

##Dimer results - Fitting to QM datasets##

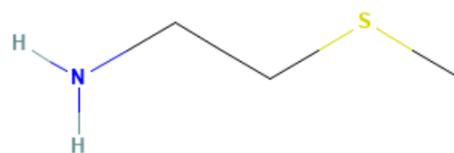
DESRES_369-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.240	0.429	4.3298	270	13



#375 2-Methylsulfanylethanamine C3H9NS CID: 87697

ref molpol	12.83	9.85	8.69, avg	10.46
molpol	12.83	9.86	8.68, avg	10.46
rms molpol	0.00	0.01	0.01, avg	0.00

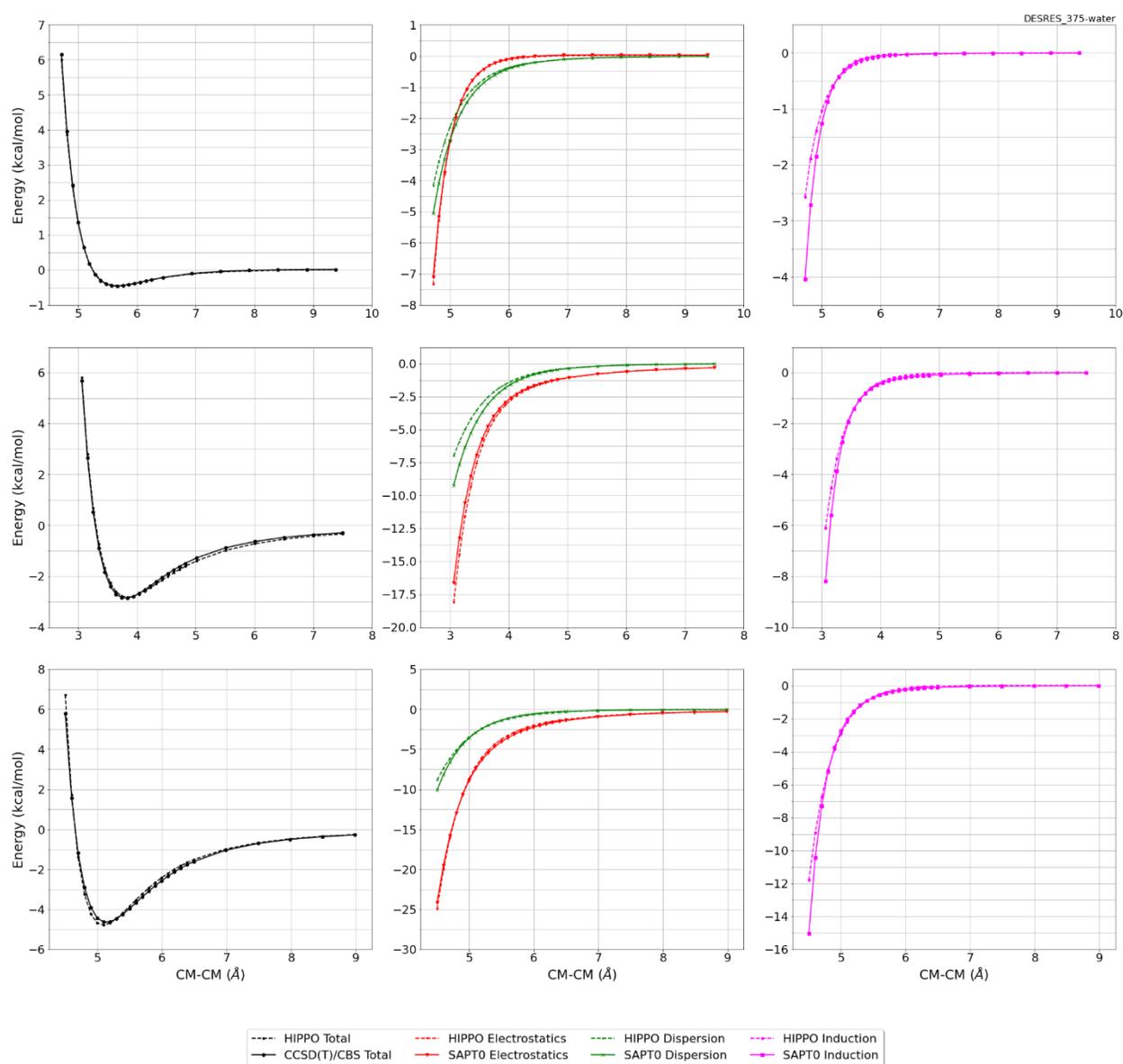


Monomer potential fitting RMS: 0.96

##Dimer results - Fitting to QM datasets##

DESRES_375-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.087	0.125	0.8515	75	0



#378 Disulfanylethane C2H6S2 CID: 94671

ref molpol	13.49	9.28	8.32, avg	10.36
molpol	13.47	9.25	8.37, avg	10.37
rms molpol	0.02	0.03	0.05, avg	0.00



Monomer potential fitting RMS: 3.18

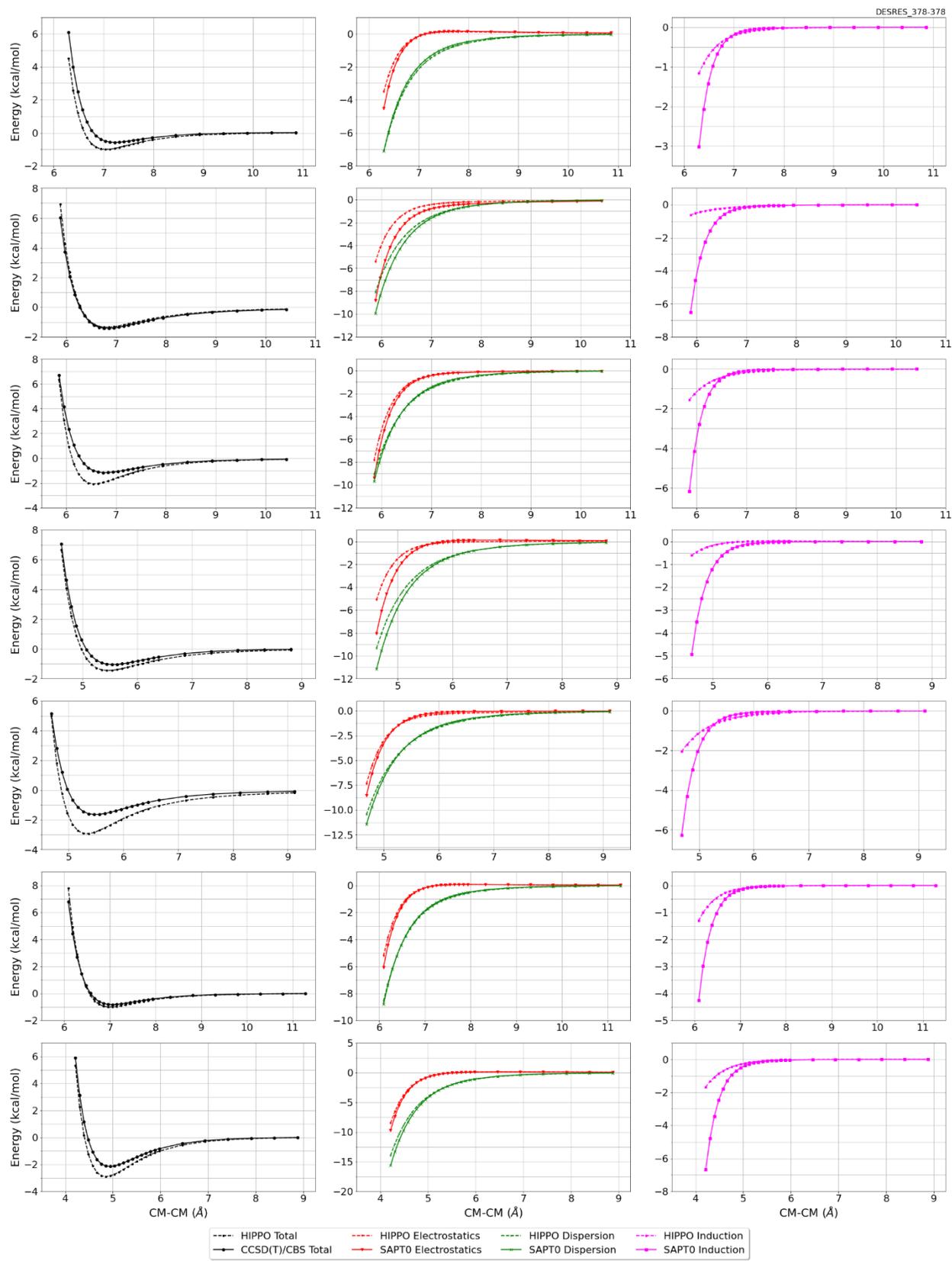
##Dimer results - Fitting to QM datasets##

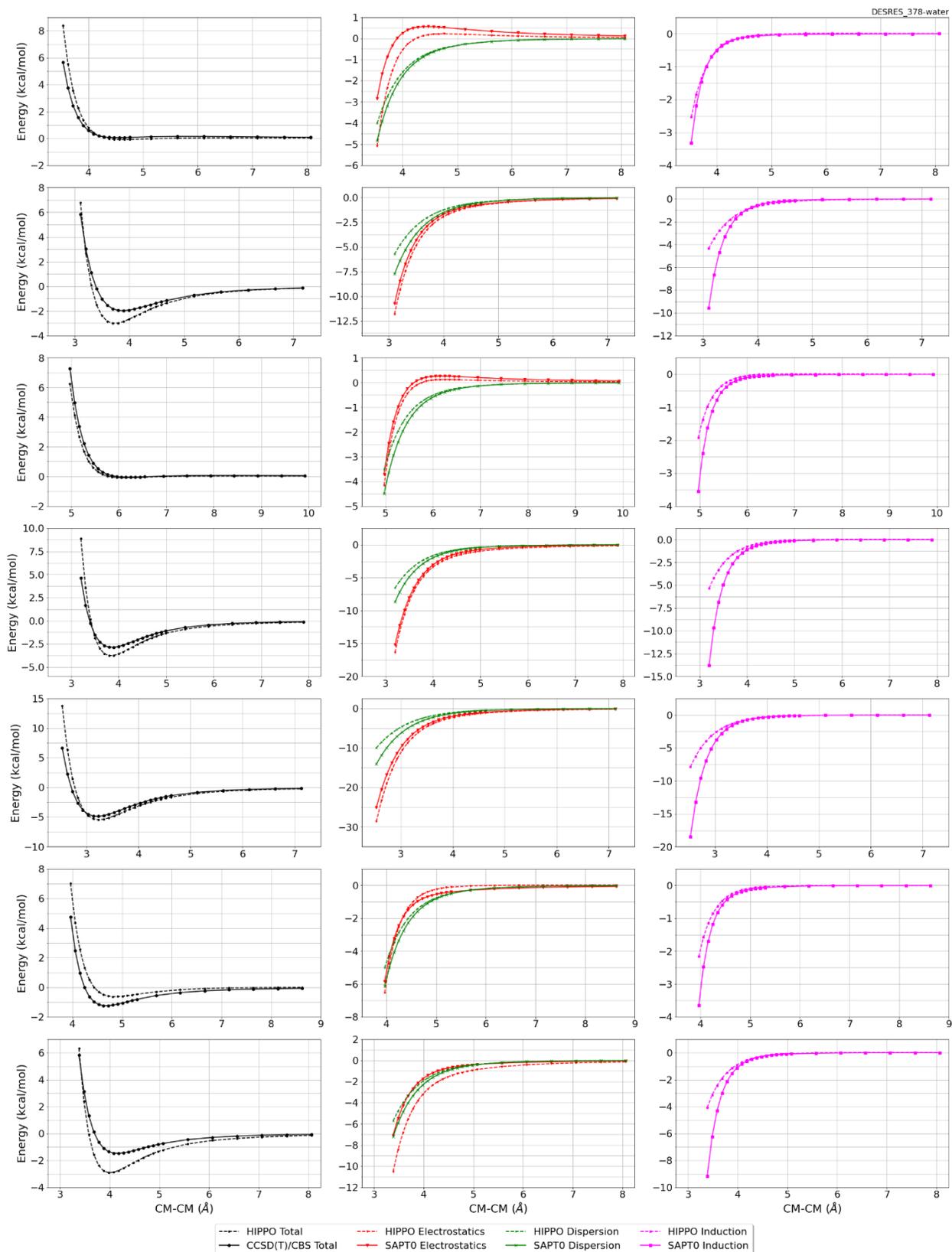
DESRES_378-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.433	0.607	7.3099	559	62

DESRES_378-378, energy values in kcal/mol

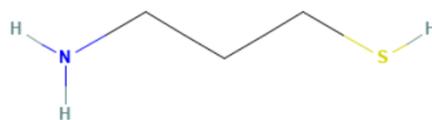
MAE	Std error	max error	#points	#count[err > 1]
0.448	0.521	2.6163	562	91





#379 3-Aminopropane-1-thiol C3H9NS CID: 97436

ref molpol 12.80 9.51 8.91, avg 10.41
molpol 12.80 9.53 8.90, avg 10.41
rms molpol 0.01 0.02 0.01, avg 0.00

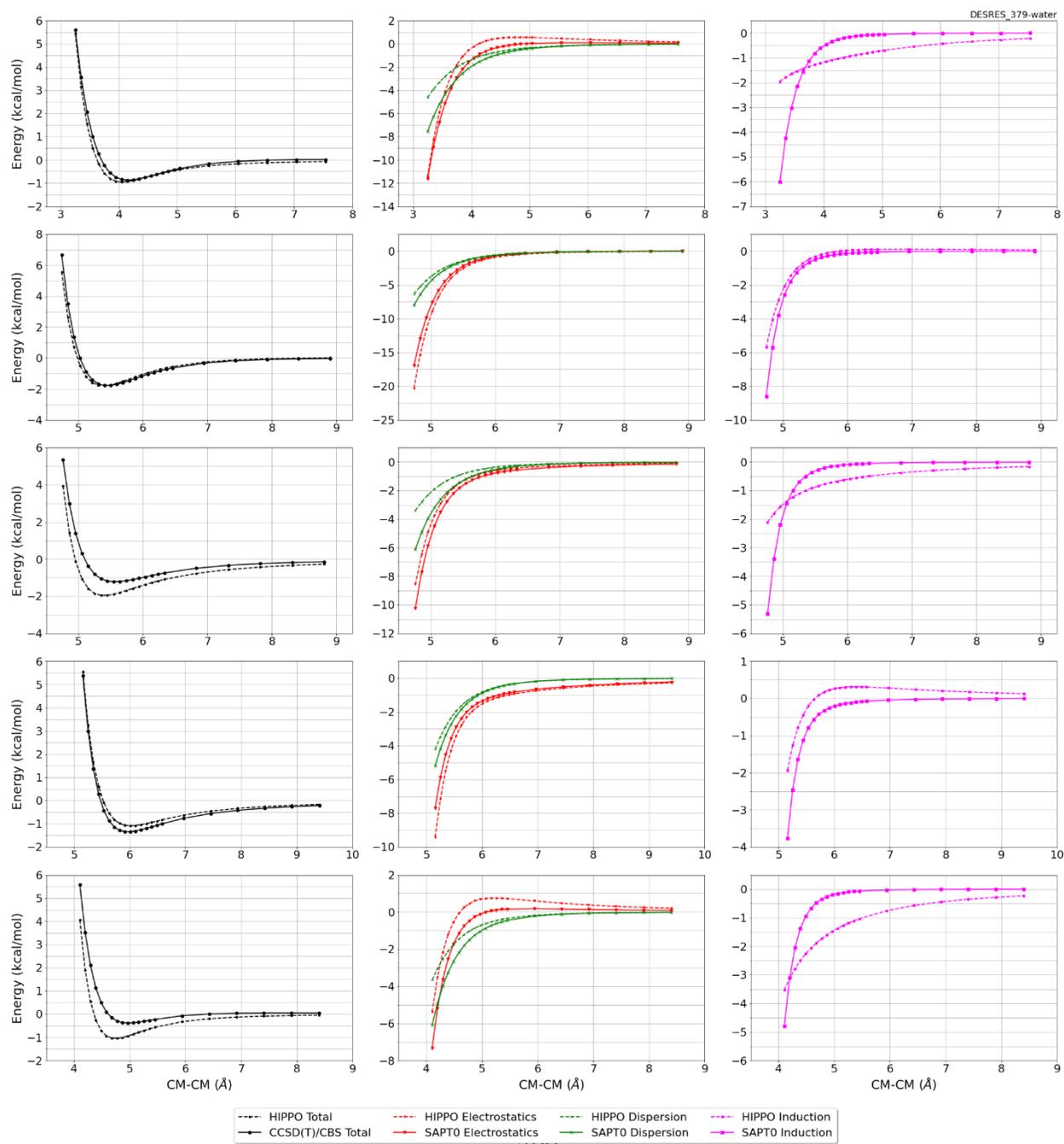


Monomer potential fitting RMS: 5.67

##Dimer results - Fitting to QM datasets##

DESRES_379-water, energy values in kcal/mol

MAE 0.182 Std error 0.231 max error 1.2203 #points 113 #count[err > 1] 1



#381 Methylsulfanylmethanethiol C2H6S2 CID: 122370

ref molpol	13.14	9.31	0.00, avg	7.48
molpol	13.11	9.23	8.14, avg	10.16
rms molpol	0.03	0.08	8.14, avg	2.68

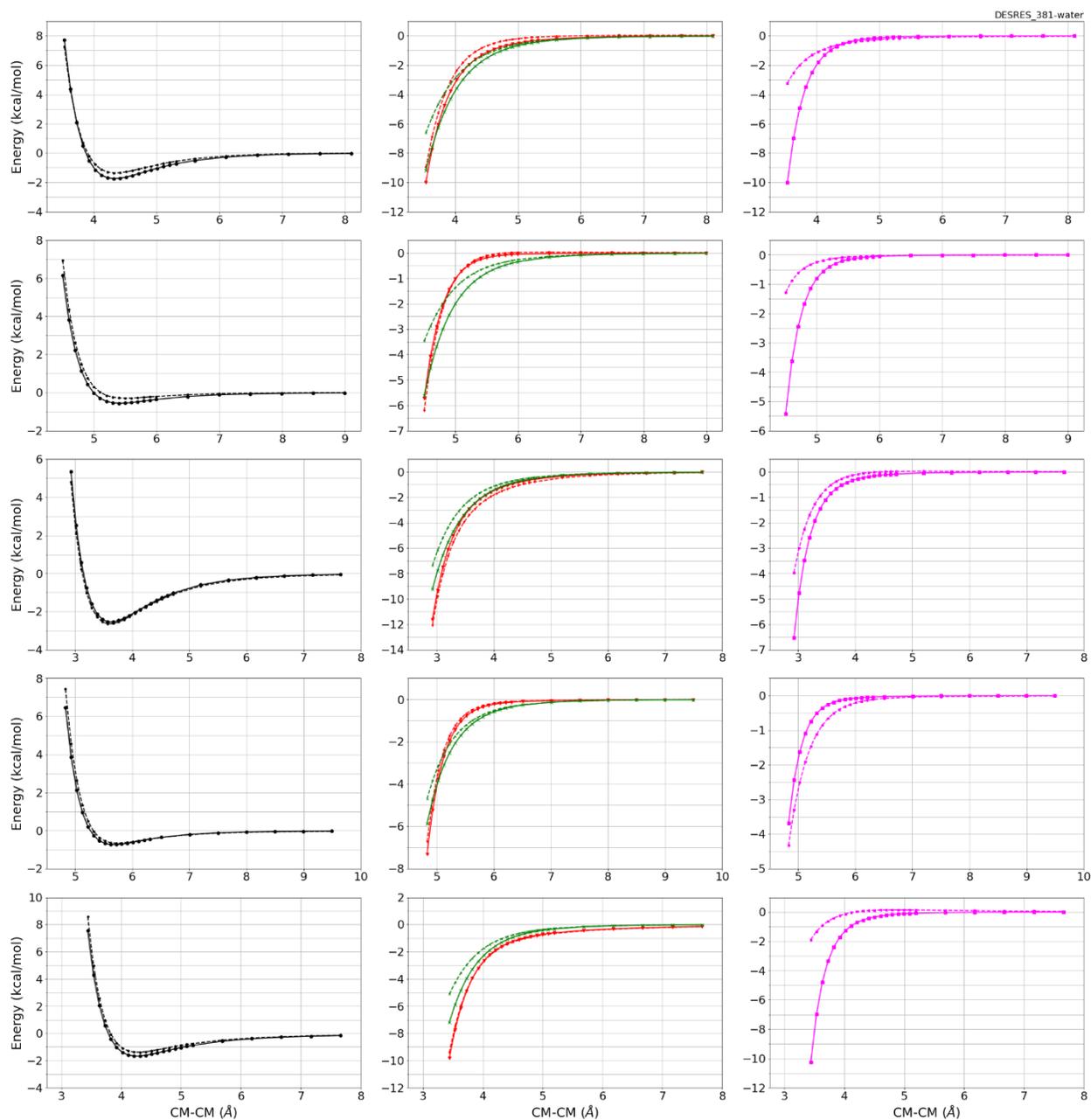


Monomer potential fitting RMS: 0.51

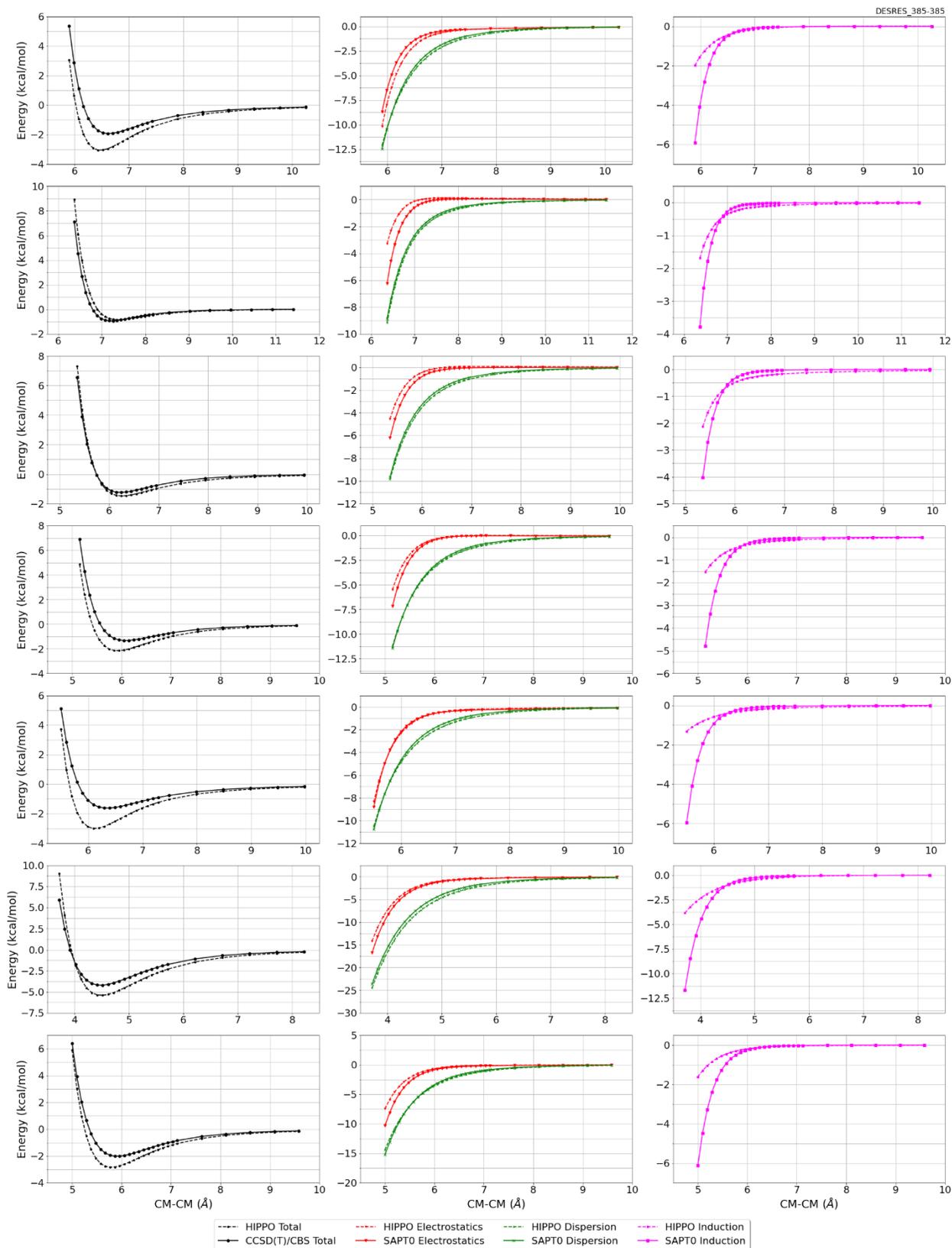
##Dimer results - Fitting to QM datasets##

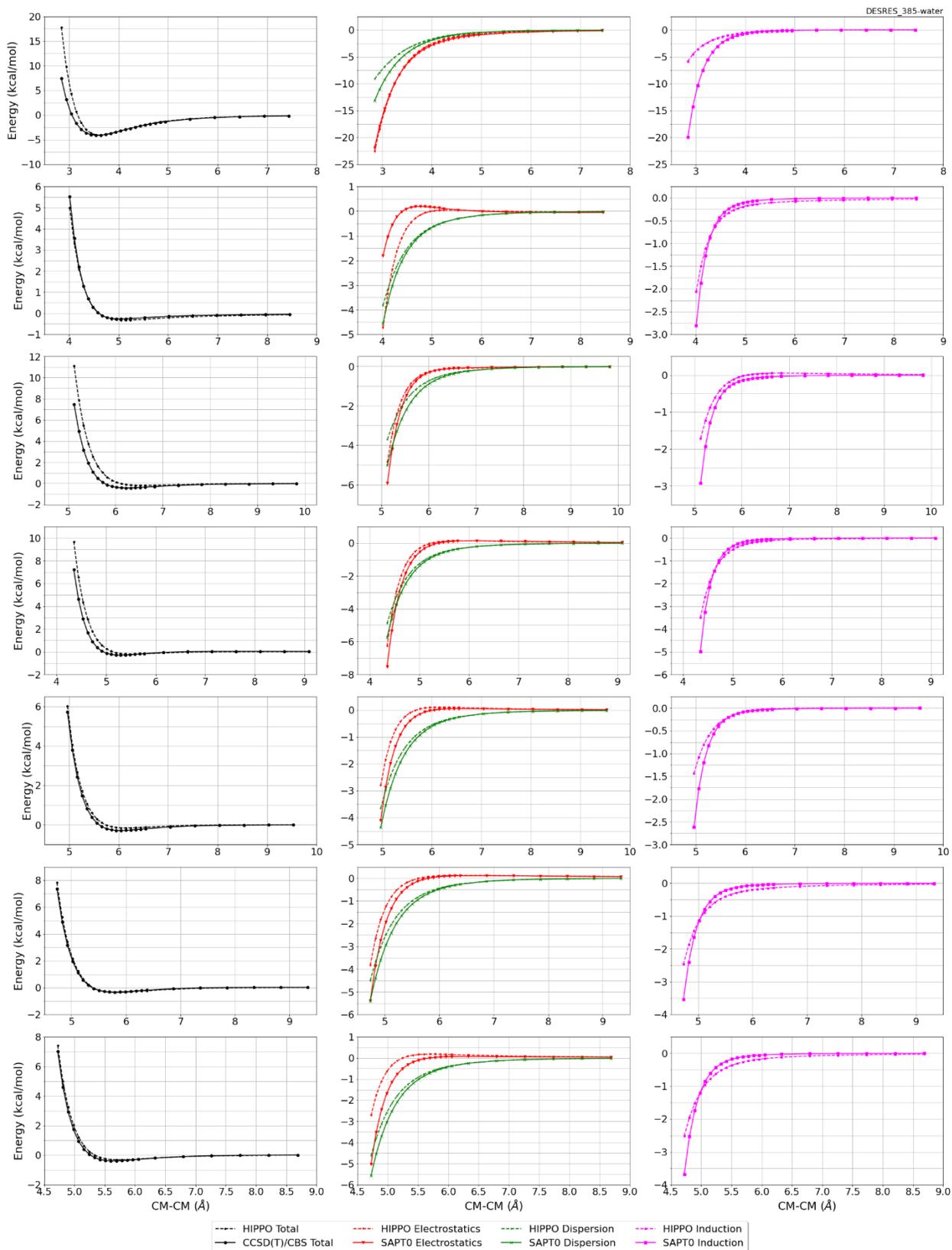
DESRES_381-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.170	0.183	1.0278	270	1



--- HIPPO Total --- HIPPO Electrostatics --- HIPPO Dispersion --- HIPPO Induction
--- CCSD(T)/CBS Total --- SAPTO Electrostatics --- SAPTO Dispersion --- SAPTO Induction





#387 Dithiane C4H8S2 CID: 136335

ref molpol	10.86	13.60	14.13, avg	12.86
molpol	10.75	13.64	14.17, avg	12.85
rms molpol	0.11	0.05	0.04, avg	0.01

Monomer potential fitting RMS: 1.69

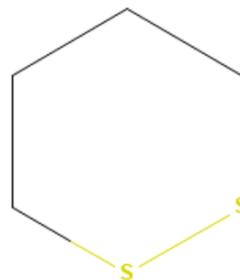
##Dimer results - Fitting to QM datasets##

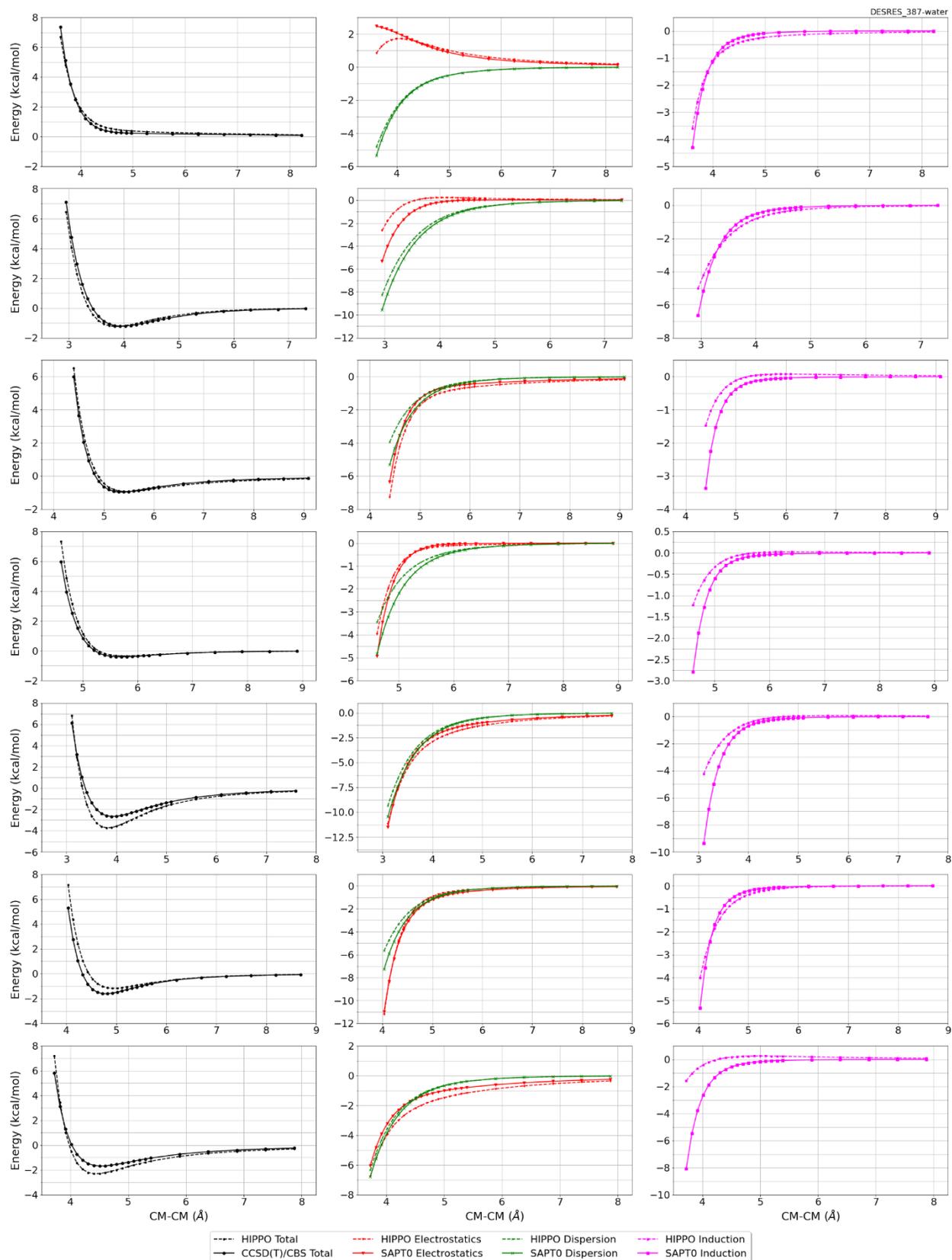
DESRES_387-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.358	0.594	8.5945	564	50

DESRES_387-387, energy values in kcal/mol

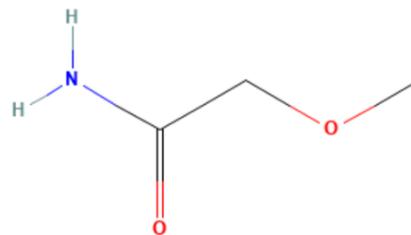
MAE	Std error	max error	#points	#count[err > 1]
1.446	0.777	3.5730	26	20





#399 2-Methoxyacetamide C3H7NO2 CID: 140060

ref molpol	8.59	9.92	0.00, avg	6.17
molpol	8.66	8.66	6.01, avg	7.78
rms molpol	0.08	1.26	6.01, avg	1.61

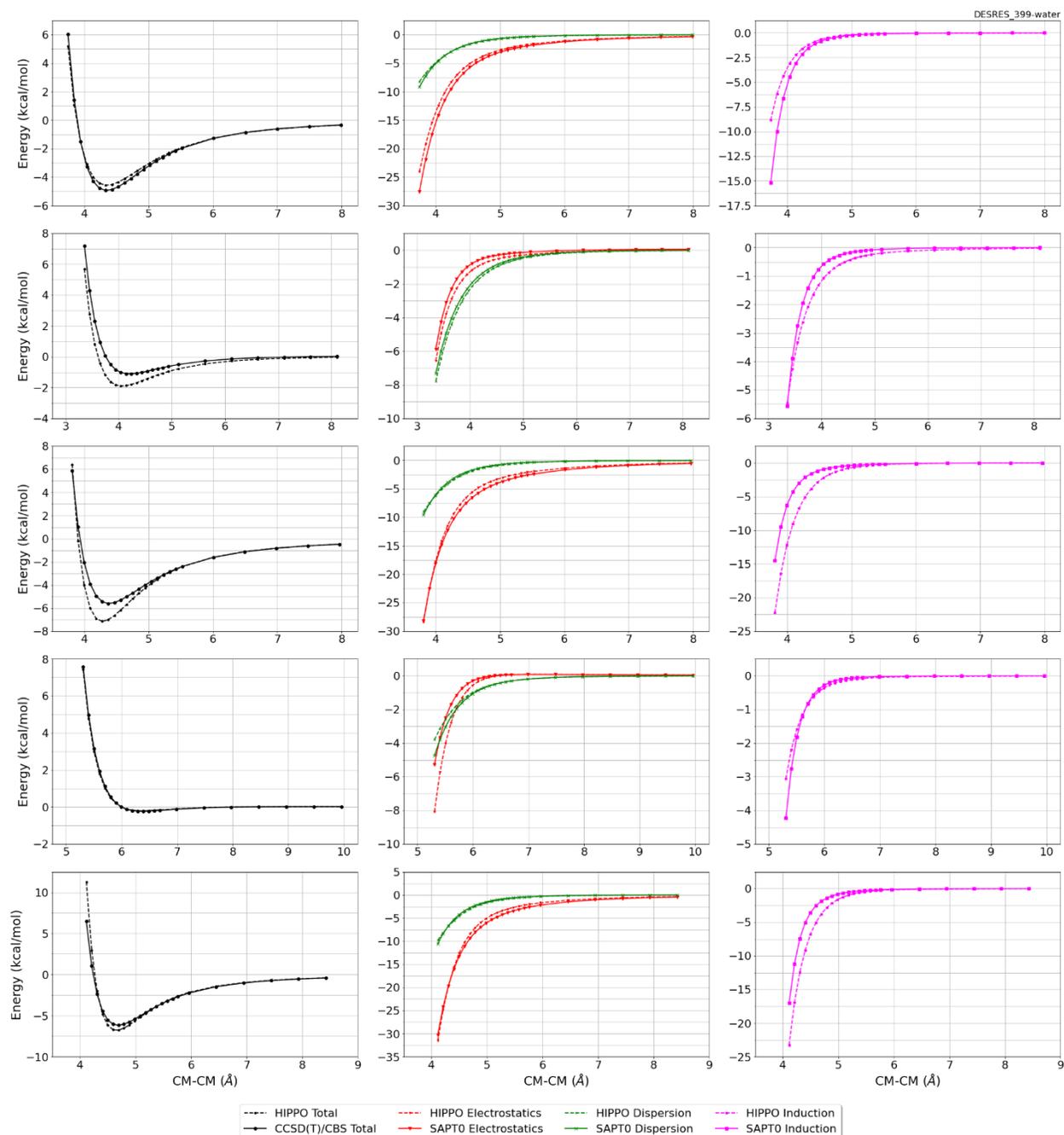


Monomer potential fitting RMS: 0.99

##Dimer results - Fitting to QM datasets##

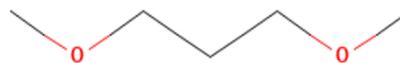
DESRES_399-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.293	0.475	4.8225	268	21



#400 1,3-Dimethoxypropane C5H12O2 CID: 140180

ref molpol 13.86 9.95 0.00, avg 7.94
molpol 13.87 8.66 8.09, avg 10.20
rms molpol 0.01 1.30 8.09, avg 2.27

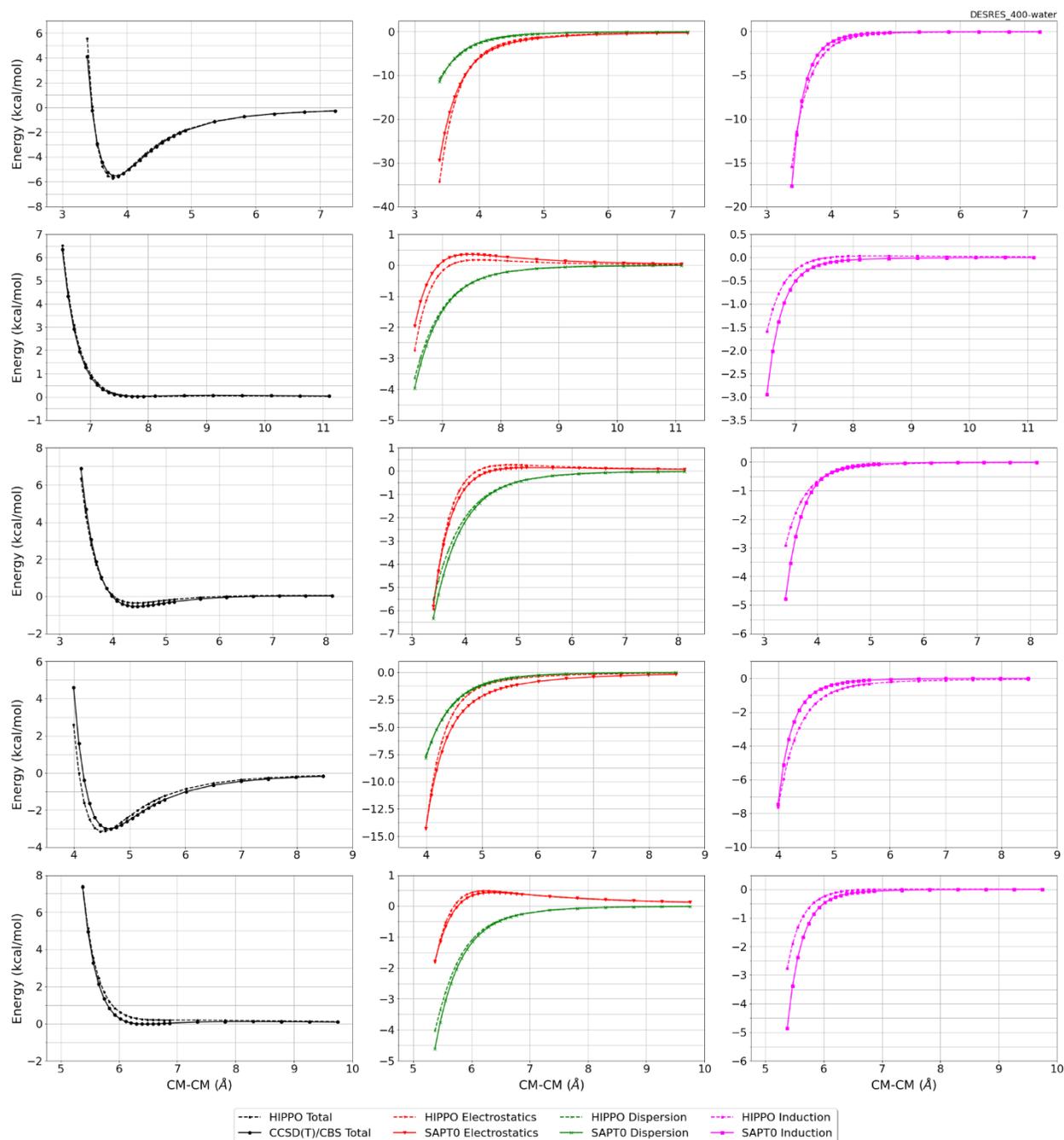


Monomer potential fitting RMS: 1.38

##Dimer results - Fitting to QM datasets##

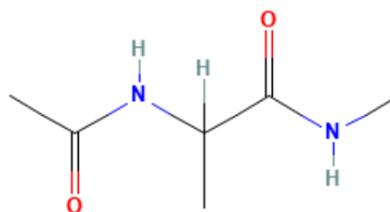
DESRES_400-water, energy values in kcal/mol

MAE Std error max error #points #count[err > 1]
0.183 0.227 1.5270 118



#402 2-Acetamido-N-methylpropanamide C6H12N2O2 CID: 141892

ref molpol	17.81	14.15	11.30, avg	14.42
molpol	17.81	14.15	11.30, avg	14.42
rms molpol	0.01	0.01	0.00, avg	0.00

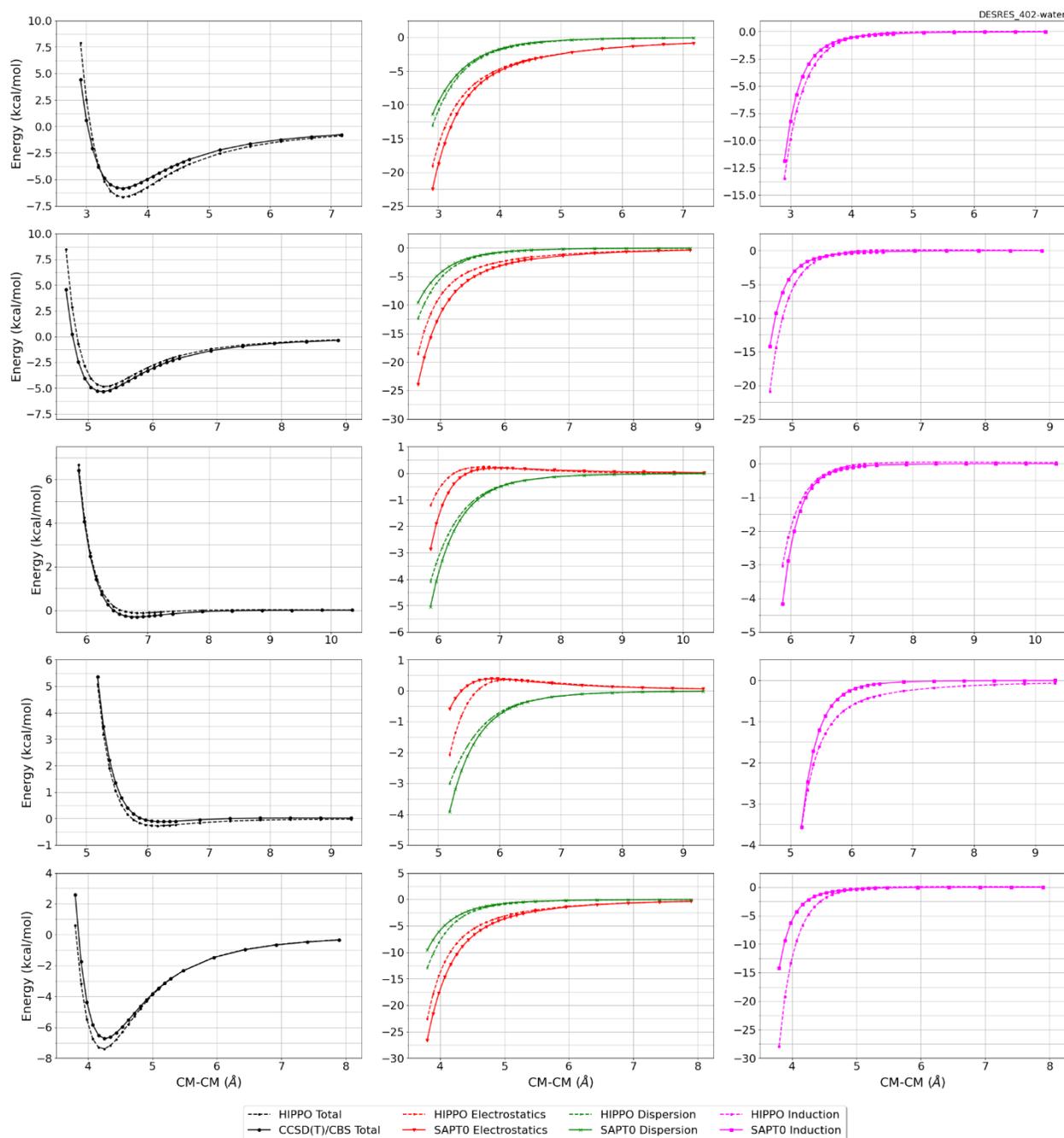


Monomer potential fitting RMS: 3.08

##Dimer results - Fitting to QM datasets##

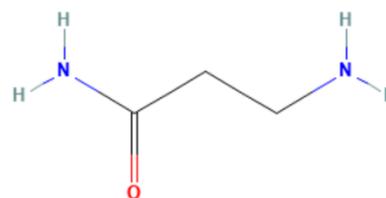
DESRES_402-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.366	0.508	2.7236	113	15



#406 3-aminopropanamide C3H8N2O CID: 192802

ref molpol	10.44	9.13	7.07, avg	8.88
molpol	10.45	9.12	7.07, avg	8.88
rms molpol	0.01	0.01	0.00, avg	0.00

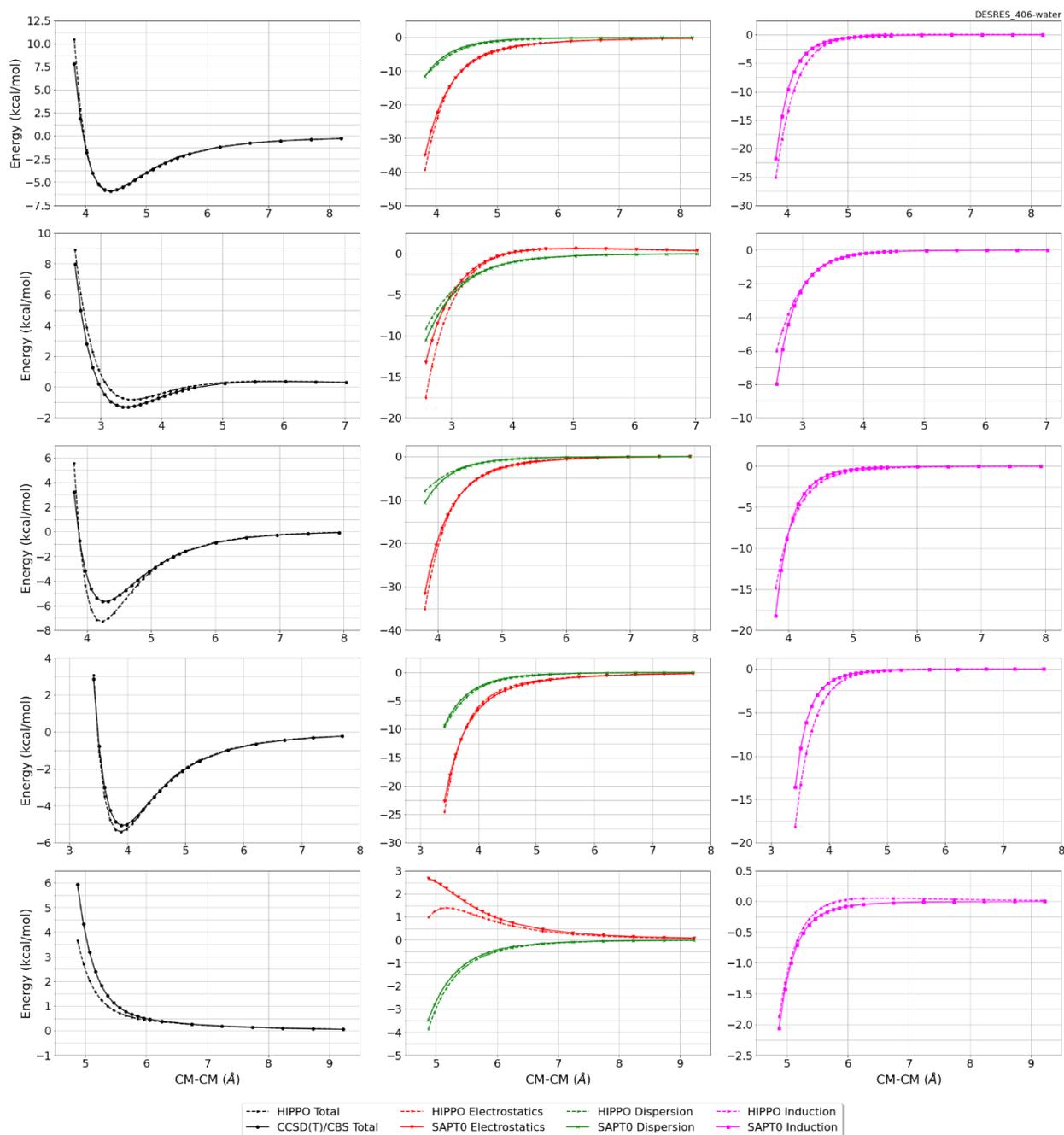


Monomer potential fitting RMS: 0.36

##Dimer results - Fitting to QM datasets##

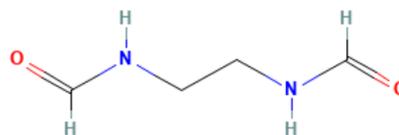
DESRES_406-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.370	0.479	2.7691	120	9



#409 N-(2-Formamidoethyl)formamide C4H8N2O2 CID: 226108

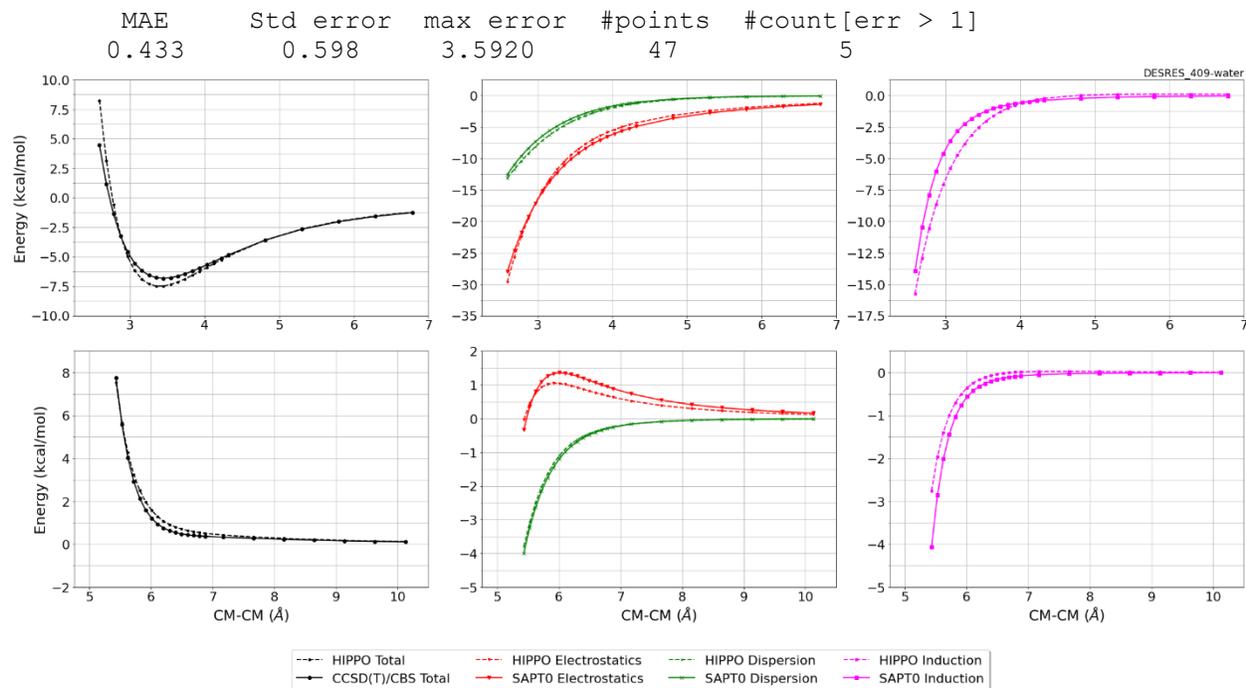
ref molpol 16.42 10.06 0.00, avg 8.83
molpol 16.41 9.96 7.82, avg 11.40
rms molpol 0.01 0.10 7.82, avg 2.57



Monomer potential fitting RMS: 0.29

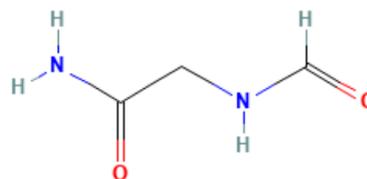
##Dimer results - Fitting to QM datasets##

DESRES_409-water, energy values in kcal/mol



#410 2-Formamidoacetamide C3H6N2O2 CID: 232267

ref molpol	12.61	9.12	0.00, avg	7.24
molpol	12.58	9.10	6.49, avg	9.39
rms molpol	0.02	0.02	6.49, avg	2.15

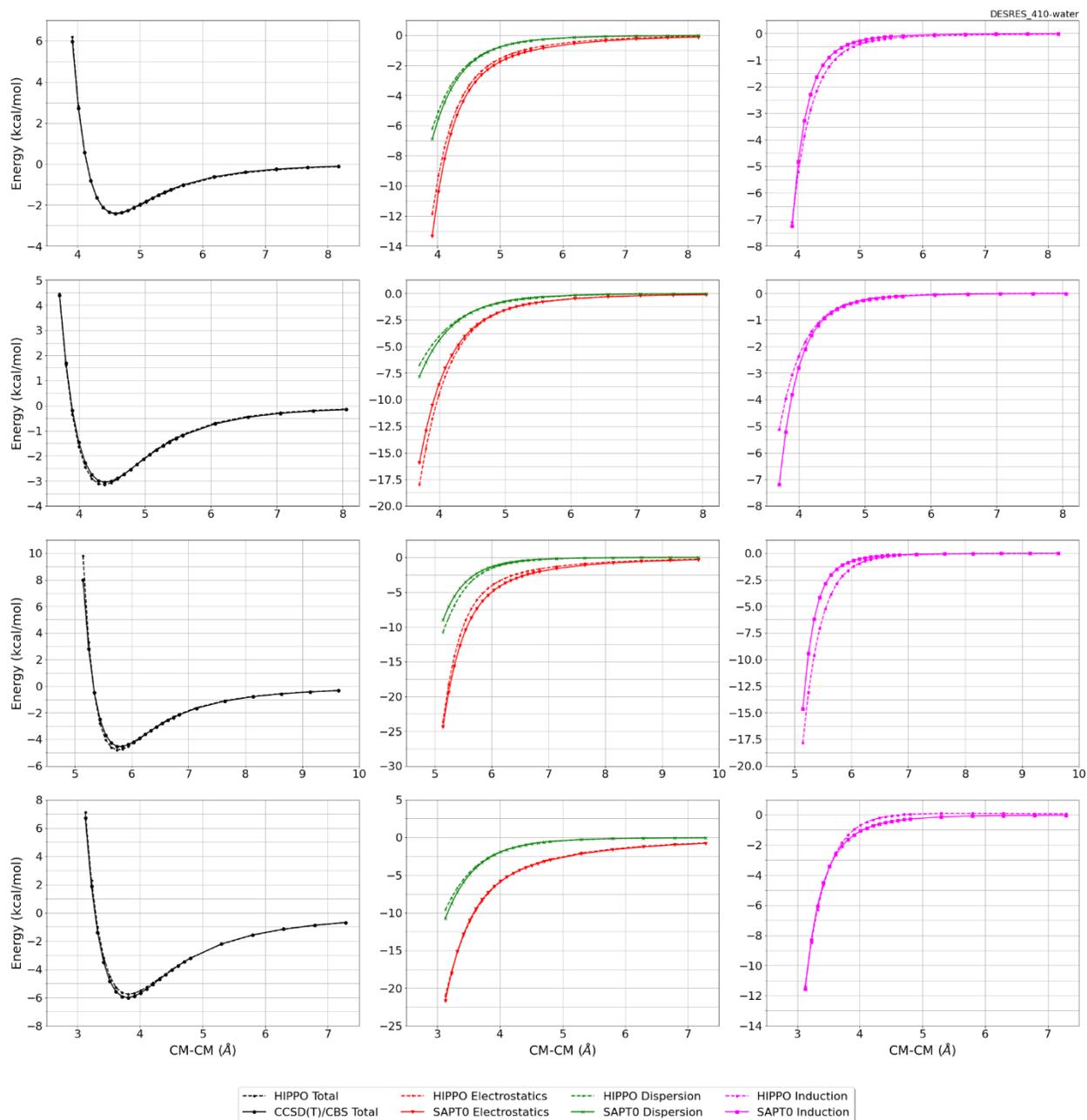


Monomer potential fitting RMS: 1.00

##Dimer results - Fitting to QM datasets##

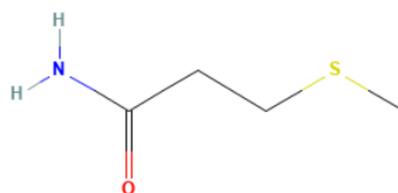
DESRES_410-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.090	0.180	1.6244	95	1



#411 3-Methylsulfanylpropanamide C4H9NOS CID: 263087

ref molpol	15.83	12.14	0.00, avg	9.32
molpol	15.80	12.12	8.36, avg	12.09
rms molpol	0.03	0.02	8.36, avg	2.77

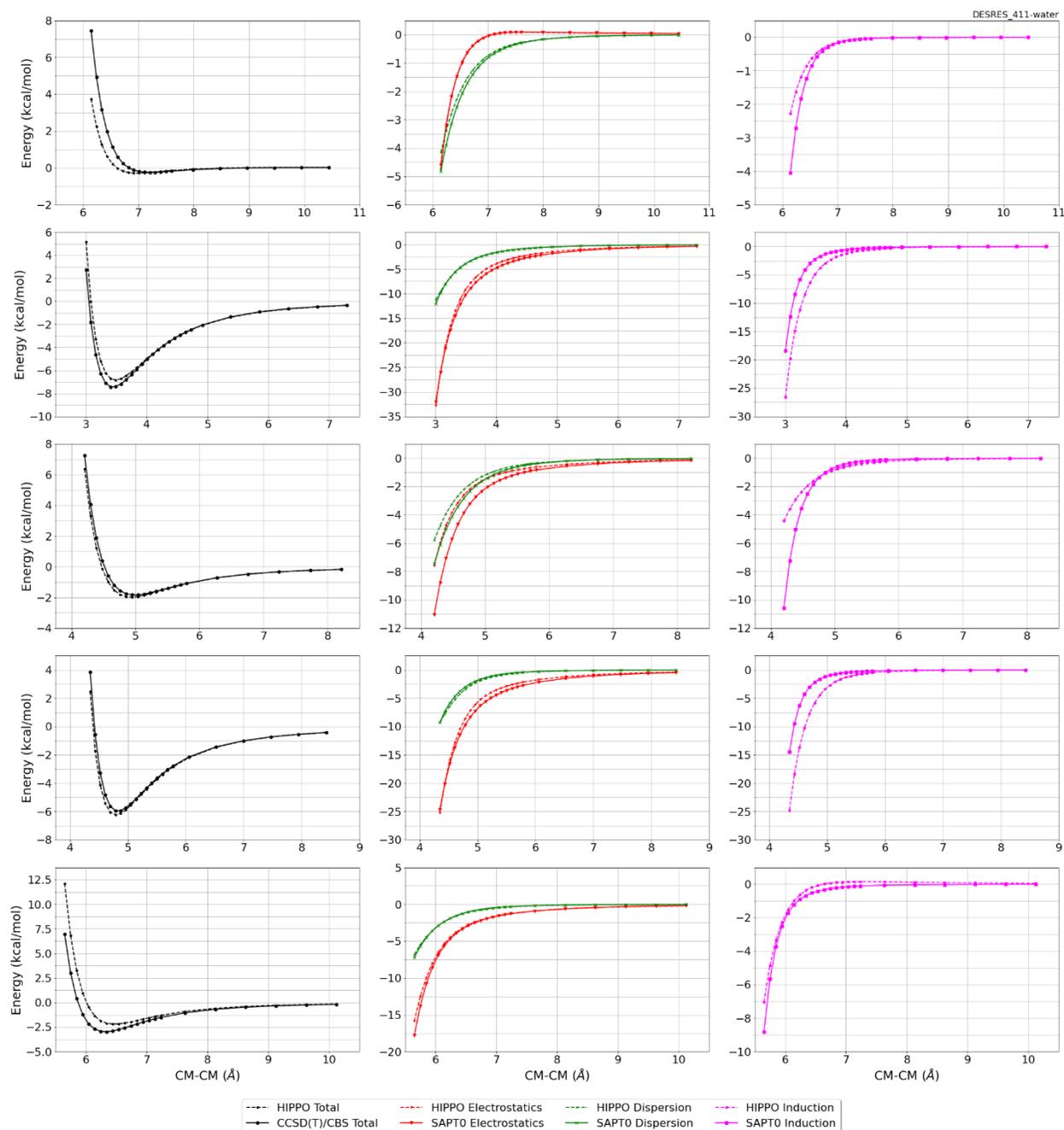


Monomer potential fitting RMS: 1.75

##Dimer results - Fitting to QM datasets##

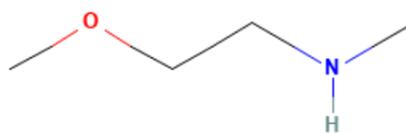
DESRES_411-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.448	0.836	4.4669	118	15



#412 2-Methoxy-N-methylethanamine C4H11NO CID: 300977

ref molpol	12.12	9.14	8.49, avg	9.92
molpol	12.16	8.71	8.58, avg	9.82
rms molpol	0.04	0.43	0.09, avg	0.10

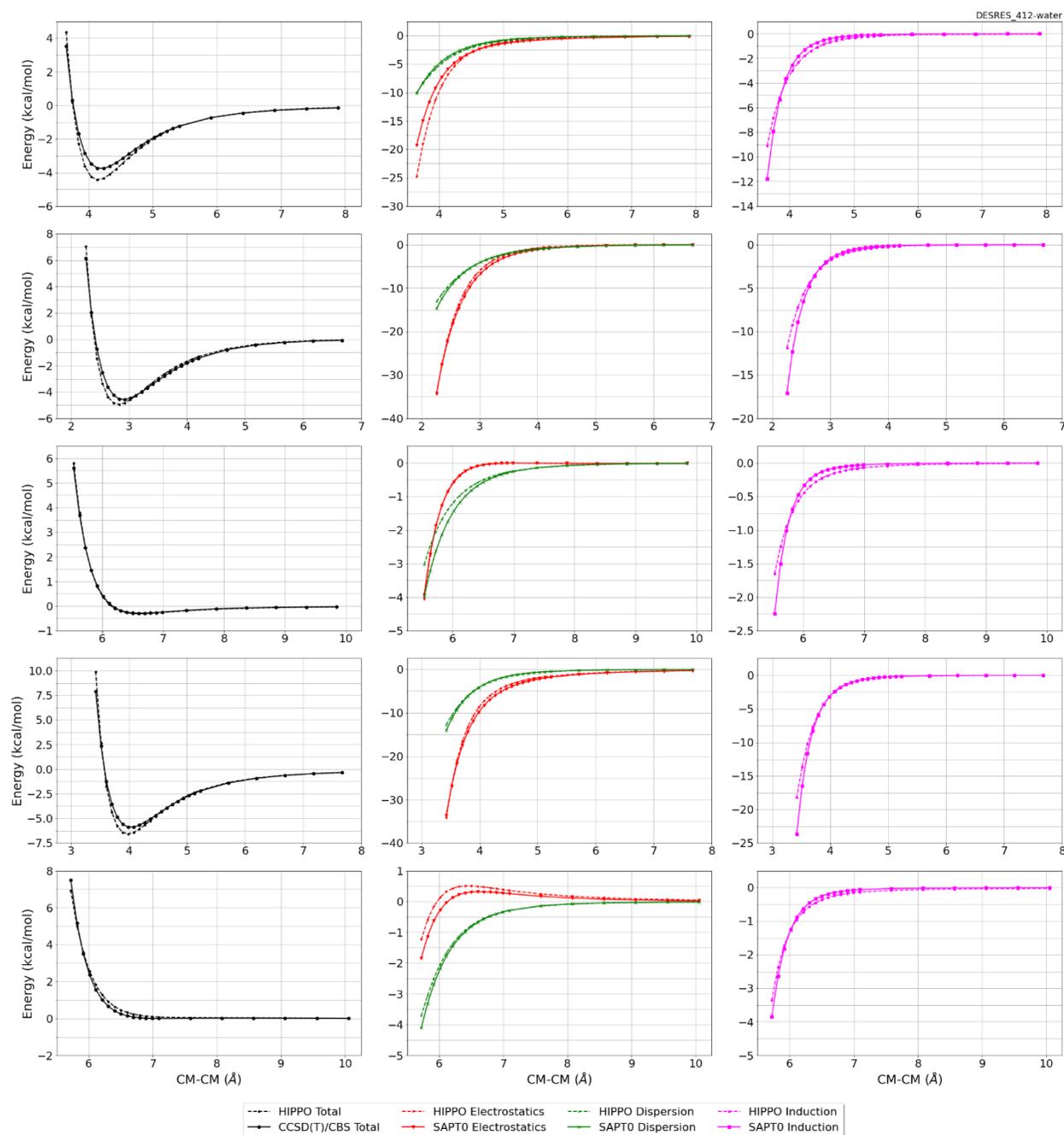


Monomer potential fitting RMS: 0.61

##Dimer results - Fitting to QM datasets##

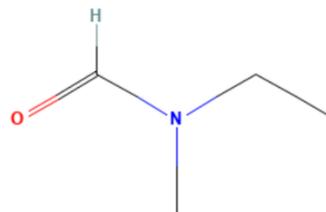
DESRES_412-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.199	0.290	2.1412	118	3



#414 N-Ethyl-N-methylformamide C4H9NO CID: 350667

ref molpol	11.54	9.25	0.00, avg	6.93
molpol	11.52	9.21	6.50, avg	9.08
rms molpol	0.02	0.05	6.50, avg	2.14

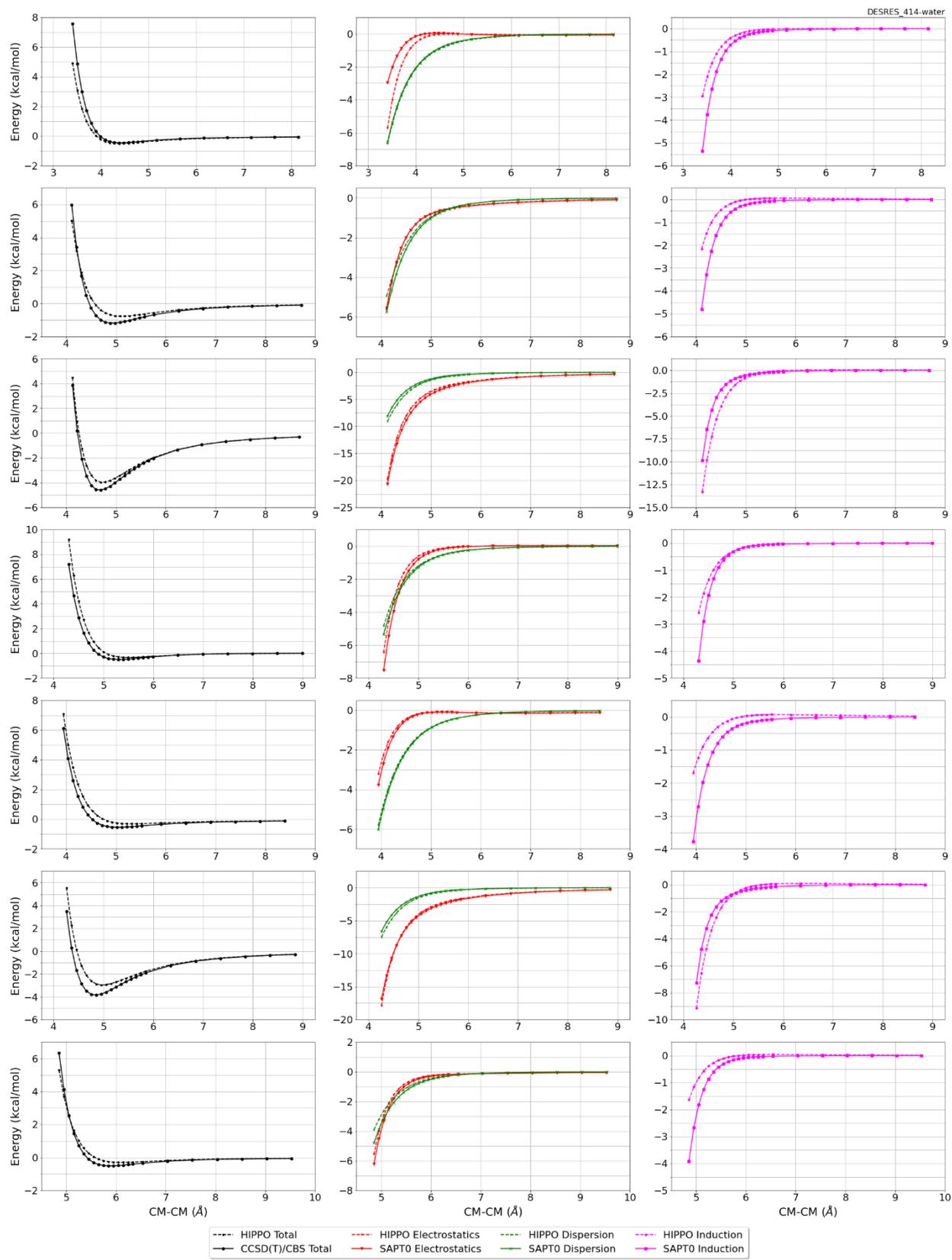


Monomer potential fitting RMS: 0.75

##Dimer results - Fitting to QM datasets##

DESRES_414-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.266	0.331	2.9084	536	26



#419 Methoxymethoxyethane C4H10O2 CID: 524894

ref molpol	11.48	8.42	0.00, avg	6.63
molpol	11.49	7.33	6.78, avg	8.53
rms molpol	0.01	1.09	6.78, avg	1.90



Monomer potential fitting RMS: 0.90

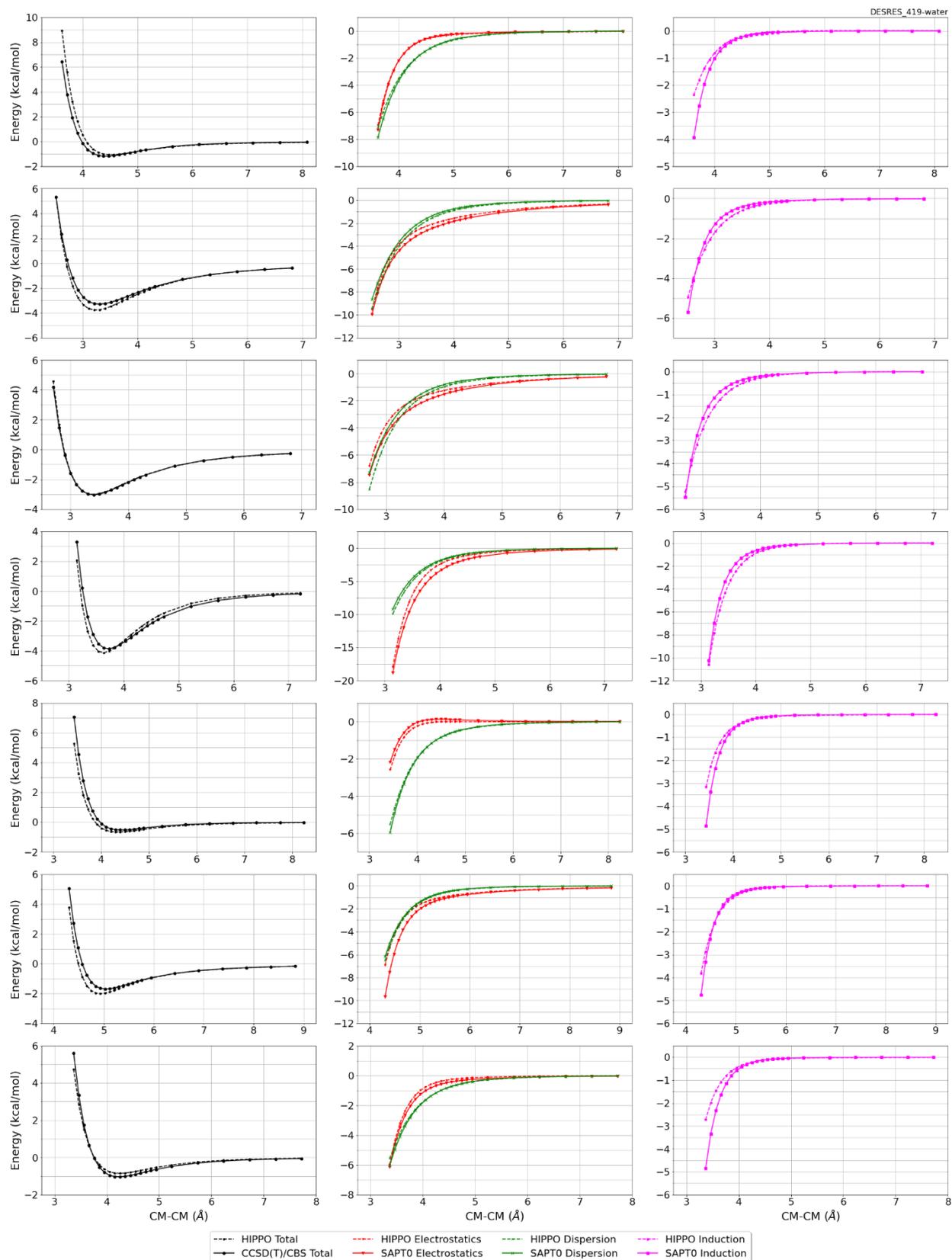
##Dimer results - Fitting to QM datasets##

DESRES_419-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.251	0.331	2.4398	535	23

DESRES_419-419, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.402	0.451	2.3430	147	18



#420 1-(Methylsulfanylmethylsulfanyl)propane C5H12S2 CID: 525376

ref molpol	21.36	14.36	0.00, avg	11.91
molpol	21.37	13.43	11.62, avg	15.47
rms molpol	0.02	0.93	11.62, avg	3.57

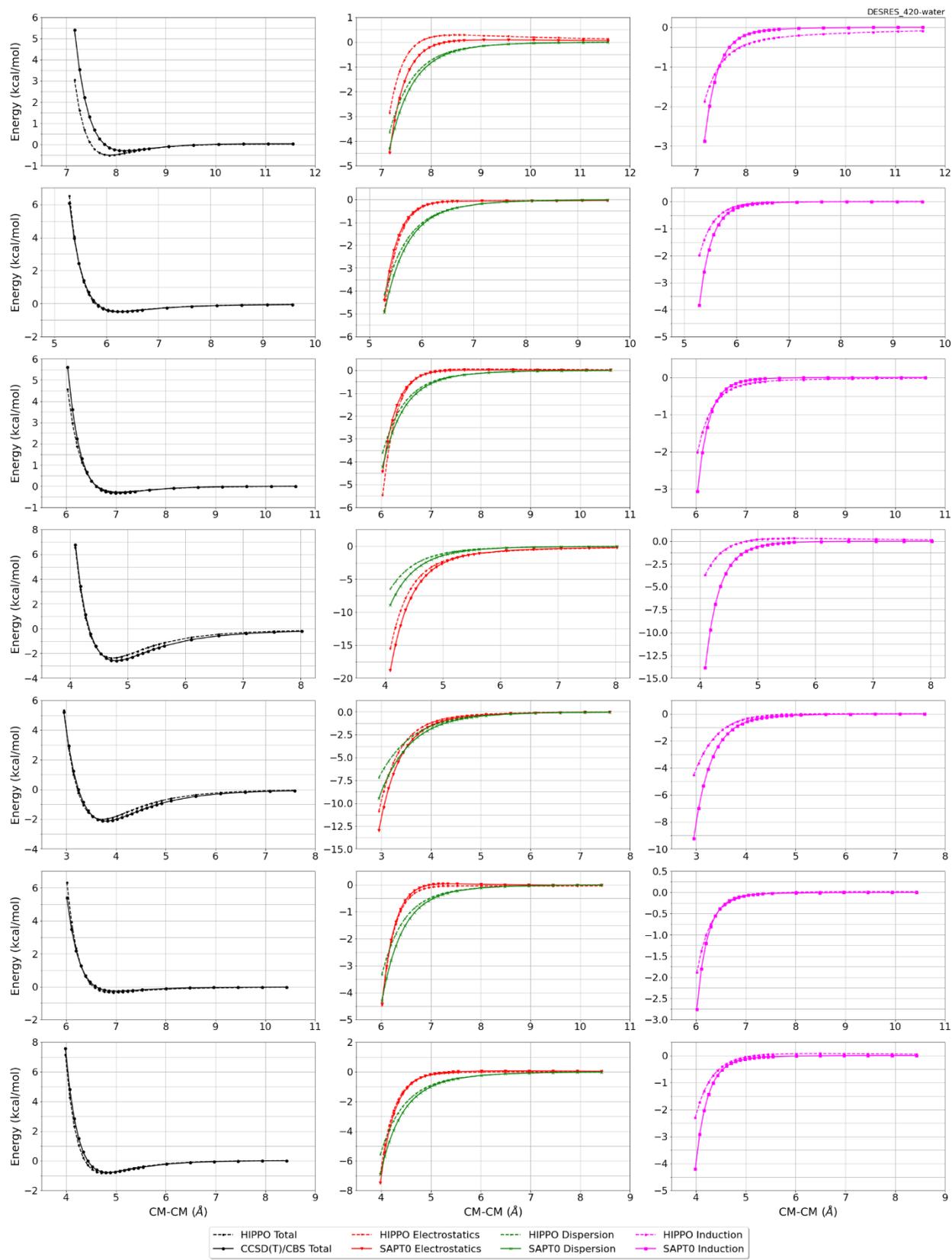


Monomer potential fitting RMS: 2.42

##Dimer results - Fitting to QM datasets##

DESRES_420-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.192	0.260	2.1482	531	11



#421 Methylsulfanylmethylsulfanylethane C4H10S2 CID: 525377

ref molpol	18.75	12.69	0.00, avg	10.48
molpol	18.78	12.12	10.42, avg	13.77
rms molpol	0.03	0.57	10.42, avg	3.29

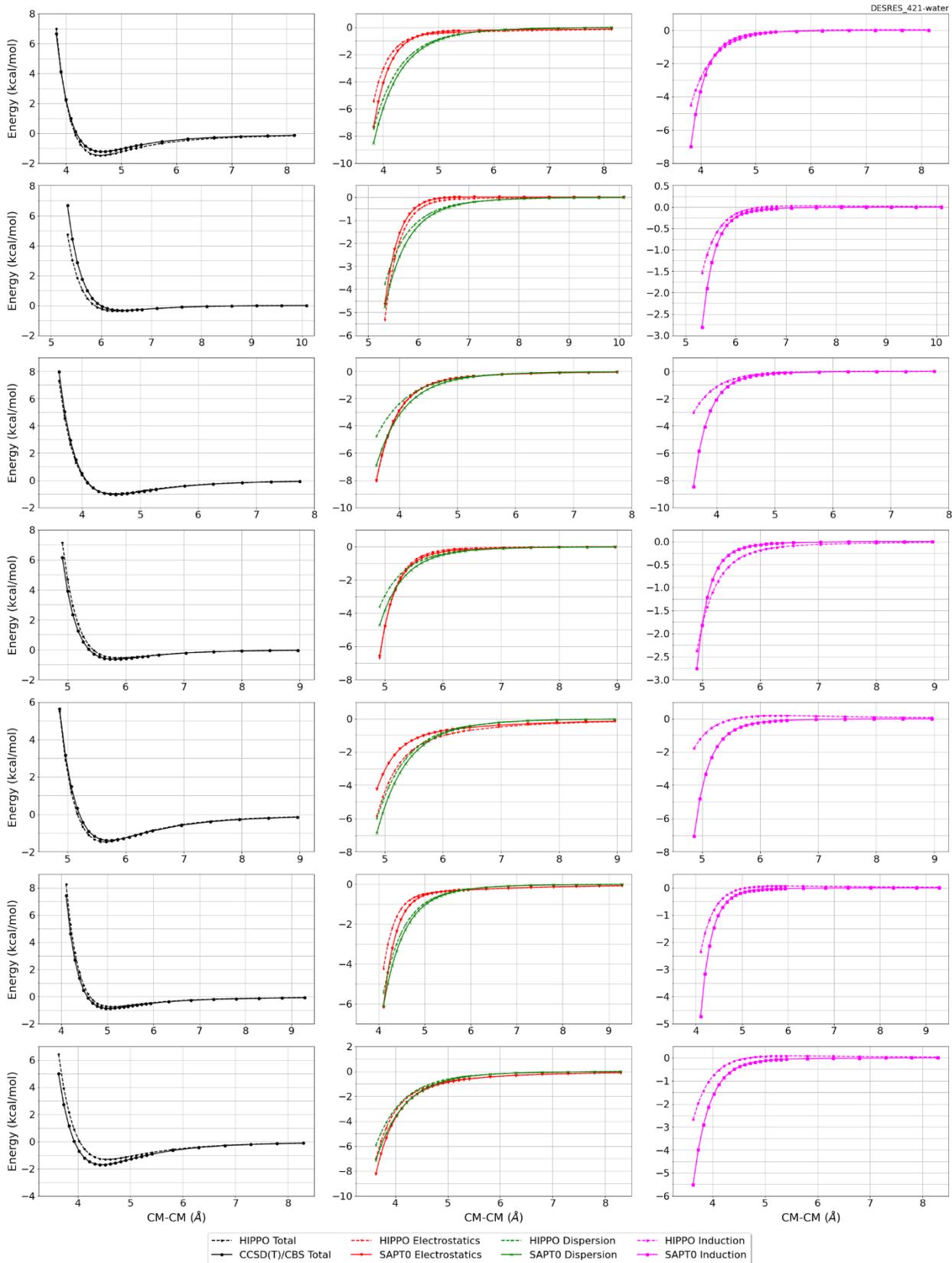


Monomer potential fitting RMS: 1.66

##Dimer results - Fitting to QM datasets##

DESRES_421-water, energy values in kcal/mol

MAE	Std error	max error	#points	#count[err > 1]
0.168	0.232	2.0398	539	5



#427 Thioacetone C3H6S CID: 641811

ref molpol	11.30	8.66	6.75, avg	8.91
molpol	10.37	8.86	7.09, avg	8.77
rms molpol	0.93	0.20	0.34, avg	0.14



Monomer potential fitting RMS: 0.54

##Dimer results - Fitting to QM datasets##

R739x5_427-water, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
3.337	1.985	3.049	1.0644
3.450	1.107	1.738	0.6314
3.563	0.632	1.004	0.3718
3.676	0.396	0.610	0.2139
3.902	0.263	0.319	0.0562

MAE	Std error	max error	#points	#count[err > 1]
0.468	0.354	1.0644	5	1

R739x5_427-427, energy values in kcal/mol

CM-CM (A)	Reference	HIPPO res	Abs diff
2.766	2.230	2.219	-0.0110
2.841	0.945	0.529	-0.4160
2.916	0.244	-0.328	-0.5715
2.990	-0.102	-0.707	-0.6045
3.140	-0.267	-0.802	-0.5348

MAE	Std error	max error	#points	#count[err > 1]
0.428	0.218	0.6045	5	0

