A Fragment-Based Construction of a Neural Network Potential for Metal-Organic Frameworks

Dissertation

for the award of the degree

“Doctor rerum naturalium"

of the Georg-August-Universität Göttingen

within the doctoral program Chemistry

of the Georg-August University School of Science (GAUSS)

submitted by

Marius Herbold

from **Hann. Münden**

Göttingen, **08.08.2022**

Thesis Committee

Supervisor:
Prof. Dr. Jörg Behler
Theoretische Chemie
Institut für Physikalische Chemie

Second Supervisor:
Prof. Dr. Ricardo A. Mata
Computerchemie und Biochemie
Institut für Physikalische Chemie

Examination Board

Reviewer:
Prof. Dr. Jörg Behler
Theoretische Chemie
Institut für Physikalische Chemie

Second Reviewer:
Prof. Dr. Ricardo A. Mata
Computerchemie und Biochemie
Institut für Physikalische Chemie

Further Members of the Examination Board
Prof. Dr. Burkhard Geil
Biophysikalische Chemie
Institut für Physikalische Chemie

Prof. Dr. Philipp Vana
Makromolekulare Chemie
Institut für Physikalische Chemie

Jun.-Prof. Dr. Daniel Obenchain
Physikalische Chemie
Institut für Physikalische Chemie

Jun.-Prof. Dr. Anna Krawczuk
Anorganische Chemie
Institut für Anorganische Chemie

Date of the oral examination: 19.09.2022

Oath

I hereby declare that I have prepared this thesis all by myself, not used any sources or tools except for those explicitly stated, and marked all quotes, be it in literal or analogous form, accordingly.

I declare that this thesis has not, nor in excerpts, been submitted to this or any other university in the context of a failed examination.

Göttingen, **08.08.2022**

(**Marius Herbold**)

Abstract

In recent years, many types of machine learning potentials (MLPs) have been developed, which are used to represent the high-dimensional potential-energy surface (PES) of a chemical system with similar accuracy as electronic structure methods. Commonly used MLPs rely on atomic energy contributions dependent on the local chemical environments. Frequently, in addition to the total energies, also atomic forces are used to construct the potentials, as these provide detailed local information about the PES. Since many systems are too large for electronic structure calculations, the MLP training is based on smaller subsystems like molecular fragments or clusters, providing reliable reference forces. Additionally this procedure can substantially simplify the construction of the training sets.

In this work, a well-defined method is proposed to determine structurally converged molecular fragments providing reliable training forces for high-dimensional neural network potentials (HDNNPs) based on the analysis of the Hessian. The Hessian permits the investigation of the atomic force dependency on the local environment and thus, the method serves as a locality test and allows to estimate the importance of long-range interactions. The procedure is illustrated for a series of simple, quasi-one-dimensional molecular model systems and the metal-organic frameworks IRMOF-1 (commonly known as MOF-5), -10 and-16 as examples for complex organic-inorganic hybrid materials. A fragment radius is dervied to construct size-converged molecular fragments as the foundation of a HDNNP data set.

In the formalism of the HDNNP, the atomic force components depend on twice the cutoff radius compared to the atomic energy contributions. Because of this relation another set of size-reduced molecular fragment is derived to construct another HDNNP data set. Both data sets can be represented with similar accuracy. The validation of the resulting HDNNPs illustrates the equivalence of the predictions. Consequently, very efficient small molecular fragments are proposed for the construction of HDNNP data set.

Contents

Glossary

1D one-dimensional IRMOF-1 system

ACE Atomic cluster expansion

ACSF Atom-centered symmetry function

BDC^{2-} Benzene-1,4-dicarboxylate

BM Birch-Murnaghan equation-of-state

BOA Born-Oppenheimer approximation

BPDC^{2-} Biphenyl-4,4′-dicarboxylate

BZ Brillouin zone

DFT Density functional theory

EOS Equation-of-state

FHI-aims Fritz Haber Institute *ab initio* molecular simulations

GAP Gaussian approximation potential

GGA Generalized gradient approximation

GTO Gaussian type orbital

HD Hexadecane

HDNNP High-dimensional neural network potential

HDOE (3E,5E,7E,9E,11E,13E)-Hexadeca-1,3,5,7,9,11,13,15-octaene

HF Hartree-Fock method

HK Hohenberg-Kohn theorems

HPC High-performance computing

I10 Reference fragment for Hessian analysis of IRMOF-10

I16 Reference fragment for Hessian analysis of IRMOF-16

IR Isoreticular

IRMOF Isoreticular metal-organic framework

KS Kohn-Sham approach

LCAO Linear combination of atomic orbitals

LDA Local density approximation

LSDA Local spin density approximation

MD Molecular dynamics

MLP Machine learning potential

MM Molecular mechanic

MOF Metal-organic framework

MTP Moment tensor potential

NAO Numerical atomic orbital

NN Neural network

NNP Neural network potential

PBC Periodic boundary conditions

PBE Perdew-Becke-Ernzerhof E_{xc} functional

PES Potential energy surface

QM Quantum mechanic

QPO pairwise-orthogonal $1,1':4',1'':4'',1''':4''',1'''':4'''',1'''''$-Quinquephenyl conformer

QPP all in-plane $1,1':4',1'':4'',1''':4''',1'''':4'''',1'''''$-Quinquephenyl conformer

RMSE Root-mean squared error

RPBE Revised Perdew-Becke-Ernzerhof E_{xc} functional

SBU Secondary building unit

SCF Self-consistent field

SNAP Spectral neighbor analysis potential

SOAP Smooth overlap of atomic positions

STO Slater type orbital

TF Thomas-Fermi model

TPDC^{2-} Terphenyl-4,4″-dicarboxylate

TS Tkatchenko-Scheffler method

vdW van der Waals

ZORA Zeroth-order regular approximation

Chapter 1

Introduction

Our modern everyday life is based on computers and related to devices including microchips in general. Although, computers became unavoidable in our modern society, most people would repel the idea of computers or machines pervading our daily life. We use computers to prepare projects for school, university, work and our private life; we write formal and informal letters, e-mails and list; we use computers for learning from written and video tutorials; and we use computers for entertainment, to mention only a few points among many others [1].

Also the ever-increasing computational power described by Moore's law [2], boosted the importance of computers. Furthermore, computers are of huge importance in industry [3] and also scientific research. Exemplary, well-known applications of computer simulations in science are related to the physics of the weather [4], modelling infectious diseases [5] and the virtual human body [6].

In addition, computers are used in chemistry to find stable atomic arrangements of molecules, calculate energy differences to derive thermodynamic properties of chemical reactions and to analyze the interactions of molecules. Thus, the computer is often used as a tool to solve the many-body problem as formulated by the electronic *Schrödinger equation* (cha. 2) providing the energy and atomic forces of a chemical system [7].

Among the static, equilibrium properties, also the evolution of a chemical system over time in simulations is of interest. Despite the ever increasing computational power over the last decades, simulation procedures like *molecular dynamic* (MD) simulations remain a challenging task, because of the growing complexity of investigated problems being inherently connected to larger and more realistic model systems with an increasing need of accuracy. The bottleneck of these simulations is the efficient and accurate description of the atomic interactions. The functional mathematical expression describing these atomic interactions is the *potential energy surface* (PES), which is a multi-dimensional, real valued function defining the potential energy, its gradient – the atomic forces – of a system dependent on the atomic positions. Furthermore, the PES is fully determined by solving the electronic Schrödinger equation for all atomic arrangements. First principle electronic structure methods like *Density Functional Theory* (DFT) allow accurate calculations of arbitrary points on the PES, which is mostly related to the electronic ground state but not restricted to it. Nevertheless, for applications like MD simulations, the energy and atomic forces for a huge number of atomic configurations are required *on-the-fly*, which is only feasible for low-dimensional small systems on short timescales by electronic structure methods, due to the demanding calculations. Another option is provided by simple, analytic potentials. These types of inter atomic potentials define a direct relation between the structure and the corresponding energy of the system. However, they introduce physical approximations and formulate a compromise between efficiency and accuracy. A third option is formed by *machine learned potentials* (MLPs) [8–10], which combine the advantages of electronic structure methods – the accuracy – and the simple, analytic potentials – the efficiency. Machine learning methods as a subtopic of artificial intelligence are used in autonomous vehicles, robotics and gaming [11] to mention only a few examples.

Machine learning methods can also be used to represent any complex PES function based on some provided energies and forces – the training data – derived from demanding electronic structure calculations for the specific system of interest. In general, MLPs consist of three parts: training data describing the relation of energy and structure, the descriptors transforming the structure data into a readable format of the machine learning method and the machine learning method itself. For neural networks (NN) – one of the first machine learning methods – it could be proven to represent any mathematical function [12,13].

Atomic interactions are invariant with respect to the translation and rotation of the underlying system, to the permutation of identically charged nuclei and to any type of point group symmetry, since the effective coordinates of the atoms remain unchanged by these operations. Consequently, these invariances are also valid for the PES and in conclusion, the descriptors have to fulfill these conditions to ensure the effective coordinates resulting from symmetry or permutational equivalent structures are related to the same input for the MLP and thus to the same energy and atomic forces. Additionally, the descriptors itself have to avoid further artificial symmetries to provide a biunique mapping of the structures. In contradiction to that, this is only assured for a one-to-one correspondence between the structural degrees of freedom and the descriptors, introducing an undesired dependency on the system size. Along with these fundamental conditions, some functional conditions are essential like the need of being differentiable with respect to the atomic coordinates for the analytic determination of atomic forces.

Extending applicability of machine learning methods increased the attention on this field and thus, different MLPs were developed, which can be categorized in four generations [10]. The first generation of MLPs came up with the advent of NNPs [14], which were limited to low-dimensional systems, due to the NN architecture dependence on the system size applied. Suitable descriptors for these systems were already quite challenging, since fundamental conditions like the permutational invariance were not fulfilled. To overcome the permutational invariance problem and the limitation to only low-dimensional systems, the second generation of MLPs – the *high-dimensional neural network potentials* (HDNNPs) [10, 15–18] – were invented and applied to systems of tens of thousands of atoms. Furthermore, the concept of the *nearsightedness* in quantum chemistry [19] was introduced, hence the total potential energy E_{tot} of the system can be formulated as a sum of M strictly local atomic energy contributions

$$E_{\text{tot}} = \sum_{A=1}^{M} E_A \quad , \tag{1.1}$$

which are calculated by element-specific atomic NNs. Moreover, the local atomic environment – the input information of the element-specific atomic NNs – is described by a new class of descriptors the *atom-centered symmetry functions* (ACSFs) [20]. The ACSFs fulfill the translational, rotational and permutational invariance and describe the environment of an atom up to a predefined cutoff radius r_{cut}. As a consequence, the HDNNP architecture is fixed, due to the fixed architecture of the element-specific atomic NNs and the HDNNP scales linearly with increasing system size.

Also other types of MLPs were invented within this second generation like *Gaussian approximation potentials* (GAPs) [21, 22], *moment tensor potentials* (MTPs) [23], *spectral neighbor analysis potentials* (SNAPs) [24], *atomic cluster expansion* (ACE) [25] and many others [26–30] based on different types of structural descriptors like the already mentioned ACSFs, the *bispectrum* [21], the *Coulomb matrix* [27], *smooth overlap of atomic positions* (SOAP) [31] and others [29, 30, 32–36] with similar predictive and descriptive power [37]. Second generation MLPs consider the major part of the atomic interactions by the strict locality, however the predictions for systems with sufficiently large long-range interactions like electrostatics might fail, since all interactions beyond the cutoff radius are truncated and equation 1.1 is not valid anymore. To include these truncated interactions, the third generation of MLPs needs to be considered, including long-range electrostatic contributions calculated by environment-dependent atomic charges [38–40]. Although this type of MLPs include long-range electrostatic interactions described by another set of element-specific atomic NNs, non-local structural or electronic effects on the charge of a certain atom are not included. Accordingly, the equation 1.1 also cannot be hold and other methods are needed for systems with present long-range charge transfer, guiding to the recent forth generation of MLPs [40–44]. In summary, there is a highly active research in this field since the advent of second generation HDNNPs.

As already mentioned an essential improvement in second generation MLPs is to split the complex full-dimensional PES into lower-dimensional atomic energy contributions (eq. 1.1). Although, these atomic energy contributions are no physical observables but mathematical auxiliary quantities, the MLP training is performed by the total potential energy of the training structures with an implicit partitioning onto the atomic energy contributions by the MLP. Especially, this separation into atomic energy contributions is not unique for large systems and a source of error compensation, since the atomic energy contributions just need to represent the total potential energy of the specific structure without further restrictions. Error compensation effects reduce the transferability of the potential. Among the descriptors and the machine learning methods, thus also the training data set is a very crucial and effort demanding element of the MLP development. Therefore, it is desired to preserve as much information as possible from the demanding electronic structure calculations, reducing the overall number of training structures, which increases simultaneously the efficiency of the MLP development. Nowadays, it is common practice to use atomic forces also for the MLP training procedure [45–48]. Additionally, the atomic forces are physically meaningful oberservables, which provide local information of the PES. Hence, each structure provides $3M$ atomic force components as additional information for the PES representation, if atomic forces are considered in the MLP training procedure.

For the application of MLPs for example in MD simulation, the atomic forces provided by the MLP are analytically calculated directly as the negative gradient of the PES with respect to the atomic positions, which ensures energy conversation trough the simulation. Consequently, high consistent energies and atomic forces are mandatory in the training data set and thus, highly converged electronic structure calculations, including the settings for the calculations but also the size of the training structures, are required for the training data set. Besides, the strict locality of atomic energy contributions enables the common procedure to use small systems for the MLP training process, although the resulting MLP will be applied to larger systems as demonstrated by HDNNPs for bulk systems [46, 49] to molecules [50]. However, this raises the question for the sufficient minimum training system size to ensure accurate atomic forces within a predefined convergence range. Definitely, this minimum size will be highly system dependent, due to very different atomic interactions in different molecules and solid-state materials. A method to analyze the locality of atomic interactions probed by environment dependency of the atomic force components will result in the minimum system size required for the MLP training systems and additionally will asses the applicability of the MLP generation type needed, thus the applicability of equation 1.1 and second generation MLPs or more recent generation types.

In this work the development of a method to analyze the dependency of the atomic forces on the local environment is of special interest. A well-defined procedure is aspired as the foundation of the method. Ideally, the individual dependence of the atomic force components on each specific atom and its coordinates can be derived and the method circumvents any statistical sampling of atomic configurations to avoid randomized factors. Finally, the method provides a clear and simple technique how to construct minimum sized MLP training systems including size-converged atomic forces. Additionally, the method assesses the crucial atomic interactions to derive the MLP generation, which in principle accurately represents the training data. Finally, a HDNNP is constructed and validated.

Outline of this Work

To address these problems, three *metal-organic framework* (MOF) structures – MOF-5 also known as IRMOF-1 [51, 52] and the related homologous IRMOF-10 and -16 [53] with increased linker molecules – are chosen as challenging benchmark systems, due to the huge bulk unit cells and the complex atomic interactions within the systems. MOFs form a class of organic-inorganic hybrid materials with fascinating properties and a wide range of applications, such as gas storage and separation, catalysis and optical devices [54–58]. In general, MOFs consist of organic linker molecules and inorganic metal-oxo clusters – the secondary building units (SBUs) – to form highly-ordered, nanoporous, crystalline and covalently bond three-dimensional structures. A huge diversity of SBUs and linker molecules exists, which can be connected to a huge amount of different MOFs [53, 54, 59–61]. Even MOFs combining different SBUs and/or linker molecules have been reported [59, 62]. Furthermore, post synthetic modifications [54, 63, 64], different functionalizations [54, 57] and MOF composites [65] increase the diversity and structural options for this class of materials. Also theoretical investigations are of high interest in order to develop new MOFs and to analyze and to predict their properties [66]. For such theoretical studies reliable and accurate interatomic potentials are needed [67, 68].

The desired size-converged HDNNP IRMOF training fragments are derived from the DFT optimized bulk unit cells (sec. 4.1) to construct a HDNNP based on DFT reference calculations, which is applied to the three IRMOF bulk structures. Thus, for each in-equivalent atomic site of the bulk structures, a size converged fragment, describing the bulk-like environment of this specific atomic site, needs to be defined. In section 4.2, drawbacks of the convergence test for the atomic force vector of the central bulk-like atom, as a function of the environment radius, are discussed. Owing to this drawbacks, the Hessian-based locality test of the atomic forces is illustrated in section 4.3. Firstly, the locality test is applied to a series of simple one-dimensional model systems including different types of bonding, which affects the interaction range (sec. 4.4). Coupled with the analysis of the IRMOF structures and the in-equivalent atomic sites, size-converged HDNNP training fragments are determined to construct a HDNNP training data set.

Because of the dependency of the atomic forces on twice the cutoff radius, another set of size-reduced molecular fragments are used to construct a HDNNP training data set (sec. 4.7). The validation of these resulting HDNNPs are performed to reveal conceptional significant differences of the HDNNPs.

Chapter 2

Theoretical Background

The main focus in computational chemistry is to understand and predict material properties without experimental data. The tools to gain this understanding are first-principle calculations, which are used to solve the time-independent, non-relativistic *Schrödinger equation* [69, 70]

$$\mathcal{H}\,\Psi = E\,\Psi. \tag{2.1}$$

Here, the wave function Ψ defines all properties and contains all information about the underlying system, which contains N electrons and M nuclei. Thus, the wave function depends on $N + M$ particle coordinates $\mathbf{x}$ [71–73]. The Hamilton operator (Hamiltonian) $\mathcal{H}$ decomposes into different parts of energetic contributions, given in atomic units

$$\begin{aligned} \mathcal{H} &= \widehat{T}_\mathrm{e} + \widehat{T}_\mathrm{n} + \widehat{V}_\mathrm{en} + \widehat{V}_\mathrm{ee} + \widehat{V}_\mathrm{nn}, \\ &= -\frac{1}{2}\sum_{i}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A}^{M}\frac{1}{M_A}\nabla_A^2 - \sum_{i}^{N}\sum_{A}^{M}\frac{Z_A}{r_{iA}} + \sum_{i}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A}^{M}\sum_{B>A}^{M}\frac{Z_A\,Z_B}{r_{AB}}, \\ M_A &= \frac{m_A}{m_e}. \end{aligned} \tag{2.2}$$

The kinetic energy is determined by the $\widehat{T}$ operators for the i electrons and A nuclei including the squared *nabla* operator ∇^2 and the mass ratio M_A of atom A. The $\widehat{V}$ operators characterize the potential energy contributions for the electron-nucleus ($\widehat{V}_\mathrm{en}$) for the nucleus A and the electron i dependent on their distance $\mathbf{r}_{iA}$ and the nuclear charge Z_A, respectively for inter electronic ($\widehat{V}_\mathrm{ee}$) and the inter nuclear interactions ($\widehat{V}_\mathrm{nn}$). The solution of the Schrödinger equation provides the total energy of the specific system. However, there are only a few special cases with analytical solutions to equation 2.1, due to the huge number of variables determining the wave function. Since the N electrons and M nuclei depend on three spatial $\mathbf{r} = (x, y, z)$ and one spin $s = \alpha, \beta$ coordinate, summing up to $4N + 4M$ coordinates $\mathbf{x} = (\mathbf{r}, s)$ of the wave function. The *Born-Oppenheimer Approximation* (BOA) exploits the huge mass difference of nuclei and electrons to separate slow nuclear and fast electronic variables from each other [74]. Even the lightest nucleus – a proton ^{1}H – is 1800 times heavier ($M_A = 1800$) than an electron. Hence, in BOA the electrons experience a field of fixed nuclei and respond instantaneously to any changes of the nuclear positions. Furthermore, it decouples the electronic and nuclear degrees of freedom, leading to the electronic *Schrödinger equation* and the electronic Hamiltonian

$$\begin{aligned} \mathcal{H}_\mathrm{elec}\,\Psi_\mathrm{elec} &= E_\mathrm{elec}\,\Psi_\mathrm{elec}, \\ \mathcal{H}_\mathrm{elec} &= \widehat{T}_\mathrm{e} + \widehat{V}_\mathrm{en} + \widehat{V}_\mathrm{ee}. \end{aligned} \tag{2.3}$$

Accordingly, the kinetic energy of the nuclei vanishes $\widehat{T}_\mathrm{n} = 0$ and the nuclear interactions remain unchanged $\widehat{V}_\mathrm{nn} = \mathit{const.}$ but depend parametrically on the nuclear positions, as well as the electronic wave function Ψ_elec. As a result, the concept of the *potential energy surface* (PES) is introduced, defining the energy landscape of the nuclear arrangement. In the following, the subscript *elec* will be omitted, thus $\mathcal{H}$ and Ψ refer to their electronic versions as given in equation 2.3, since the nuclear wave function is commonly not of interest.

The huge number of $4N$ wave function variables and thus the complexity of the wave function methods is a big disadvantage, hence, analytic solutions are still lacking. The first attempt to solve equation 2.3 numerically was suggested with the *Hartree-Fock* method (HF) [75–77], which states the expectation value for any trial wave function $\tilde{\Psi}$, provides a larger trial energy $\tilde{E}$ and thus an upper boundary for the exact energy E_0 resulting from the exact wave function Ψ_0. This is known as the variational principle [78]

$$E_0 \leq \tilde{E} = \left\langle \tilde{\Psi} \middle| \mathcal{H} \middle| \tilde{\Psi} \right\rangle . \tag{2.4}$$

The wave function itself is not an observable and only a mathematical concept describing all properties of a specific system. Its squared modulus can be interpreted as the probability density specifying to find the electrons $1, 2, ..., N$ within the small volume elements $d\mathbf{r}_1, d\mathbf{r}_2, ..., d\mathbf{r}_N$ at the same time

$$\left|\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)\right|^2 d\mathbf{x}_1\, d\mathbf{x}_2 \ldots d\mathbf{x}_N . \tag{2.5}$$

2.1 Density Functional Theory (DFT)

The quantum mechanical framework, which uses the probability density ρ instead of the wave function Ψ is called *Density Functional Theory* (DFT). The probability density is obtained by integrating the squared modulus of the wave function over all electronic coordinates but one spatial coordinate $\mathbf{r}_i$ and scaled by the number of electrons N

$$\rho(\mathbf{r}) = N \int \left|\Psi(\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N})\right|^2 ds_1\, d\mathbf{x_2} \ldots d\mathbf{x}_N . \tag{2.6}$$

Since all electrons are indistinguishable, each electron i illustrates the same probability density and all electronic coordinates are equivalent. To be exact, in this case the probability density ρ specifies the probability density of the presence of any electron i inside an incremental volume element $d\mathbf{r}$ – commonly known as the electron density – while the system is in state Ψ. The electron density integrates to the total number of electrons N in the system; for distances $\mathbf{r}_{1A}$ of electron 1 to any nucleus A near infinity the density vanishes and furthermore, over the whole space the density shows non-negative values

$$\begin{aligned} \int \rho(\mathbf{r})\, d\mathbf{r} &= N, \\ \lim_{\mathbf{r_{1A}} \to \infty} \rho(\mathbf{r}) &= 0, \\ \rho(\mathbf{r}) &\geq 0. \end{aligned} \tag{2.7}$$

At the nucleus positions $\mathbf{r}_A$ the density shows cusps providing information about the nuclear charge Z_A

$$\lim_{\mathbf{r_{1A}} \to 0} \left[\frac{\partial}{\partial r} + 2Z_A \right] \bar{\rho}(\mathbf{r}) = 0. \tag{2.8}$$

Consequently, the electron density ρ contains all information about the system (total number of nuclei M, type of nuclei via their charges Z_A, nuclei positions $\mathbf{r}_A$, the total number of electrons N and related quantities like the overall charge) and defines it completely. Additionally the electron density is an observable, by e. g. X-ray diffraction, contrary to the wave function. The reduction to 3 spatial coordinates is independent on the system size and one of the major advantages of the electron density. Moreover, only one- and two-electron terms occur in the Hamiltonian (eq. 2.2), so the explicit consideration for each single electron is not necessary in terms of indistinguishability of the electrons i. Vice versa, the wave function may provides redundant information and the electron density appears to be a more central and simplified variable of quantum mechanical systems. This offers the opportunity, to investigate much larger systems, including more atoms and electrons, like complex bulk structures in materials science or enzymes and macromolecules in biology.

The first model trying to connect the density to the total energy of the system via a functional H

$$E = H\left[\rho\right], \tag{2.9}$$

is given by the *Thomas-Fermi Model* (TF). However, this model is only of historical interest, since the TF is a crude approximation. Nevertheless, the energy is based only on the electron density and classical terms for electron-electron and electron-nuclear interactions. The TF defines the ground state of the system is given by the wave function Ψ_0, which is related to the ground state electron density ρ_0 and thus to the ground state energy E_0 via a functional (eq. 2.9). Additionally, to determine the required ground state density ρ_0, the TF deploys the variational principle, although the proof has still been lacking.

2.1.1 Hohenberg-Kohn Theorems

The foundation of modern DFT methods is given by the *Hohenberg-Kohn Theorems* (HK) [79], which do not provide any instructions to identify the ground state electron density ρ_0, nor any hint for the functional H (eq. 2.9). Anyway, HK are the proof of existence for this functional H and declare the ground state energy E_0 is deducible from the ground state electron density ρ_0 in general.

The First Hohenberg-Kohn Theorem

The physical legitimization of replacing the wave function by the electron density is given by the first theorem. It defines two different nuclear arrangements, resulting in two different external potentials V_{ext} and $V^{'}_{\text{ext}}$ – different by more than just a constant – lead to the same electron density. From this follows, two different wave functions, two different Hamiltonians and two different ground state energies

$$\mathcal{H} = \widehat{T} + \widehat{V}_{\text{ee}} + \widehat{V}_{\text{ext}} \quad \text{and} \quad \mathcal{H}^{'} = \widehat{T} + \widehat{V}_{\text{ee}} + \widehat{V}^{'}_{\text{ext}}, \tag{2.10}$$

$$E_0 = \langle\Psi|\mathcal{H}|\Psi\rangle \quad \neq \quad E^{'}_0 = \left\langle\Psi^{'}\middle|\mathcal{H}^{'}\middle|\Psi^{'}\right\rangle. \tag{2.11}$$

Although, these two wave functions are different, they result in the same electron density, as assumed

$$V_{\text{ext}} \Rightarrow \mathcal{H} \Rightarrow \Psi \Rightarrow \rho_0(\mathbf{r}) \Leftarrow \Psi^{'} \Leftarrow \mathcal{H}^{'} \Leftarrow V^{'}_{\text{ext}}. \tag{2.12}$$

Since the squared modulus of the wave function, used for the construction of the electron density, is not biunique, the assumption (eq. 2.12) can be valid. Nonetheless, these two wave functions $\Psi^{'}$ and Ψ can be inserted in equation 2.3, together with the different Hamiltonians

$\mathcal{H}$ and $\mathcal{H}^{'}$

$$
\begin{aligned}
E_0 &< \left\langle \Psi^{'} \middle| \mathcal{H} \middle| \Psi^{'} \right\rangle = \left\langle \Psi^{'} \middle| \mathcal{H}^{'} \middle| \Psi^{'} \right\rangle + \left\langle \Psi^{'} \middle| \mathcal{H} - \mathcal{H}^{'} \middle| \Psi^{'} \right\rangle , \\
E_0 &< E_0^{'} + \int \rho(\mathbf{r})(V_{\text{ext}} - V_{\text{ext}}^{'}) d\mathbf{r}, \qquad (2.13) \\
\mathcal{H}^{'} : \ E_0^{'} &< E_0 - \int \rho(\mathbf{r})(V_{\text{ext}} - V_{\text{ext}}^{'}) d\mathbf{r}, \qquad (2.14) \\
E_0 + E_0^{'} &< E_0^{'} + E_0. \qquad (2.15)
\end{aligned}
$$

This results in two equations for $\mathcal{H}$ and $\mathcal{H}^{'}$ (eq. 2.13 and eq. 2.14), which sum up to a physical and mathematical contradiction (eq. 2.15) and states the assumption (eq. 2.12) wrong. As a consequence, the ground state wave function determines uniquely the ground state electron density and therefore, the ground state energy, including all its components being functionals of the ground state density

$$
\begin{aligned}
H\left[\rho_0\right] &= V_{\text{ext}}\left[\rho_0\right] + T\left[\rho_0\right] + V_{\text{ee}}\left[\rho_0\right], \\
&= \int \rho_0(\mathbf{r}) V_{\text{ext}} d\mathbf{r} + T\left[\rho_0\right] + V_{\text{ee}}\left[\rho_0\right], \\
&= \int \rho_0(\mathbf{r}) V_{\text{ext}} d\mathbf{r} + F_{\text{HK}}\left[\rho_0\right], \qquad (2.16) \\
F_{\text{HK}}\left[\rho_0\right] &= T\left[\rho_0\right] + V_{\text{ee}}\left[\rho_0\right]. \qquad (2.17)
\end{aligned}
$$

The universal and system independent *Hohenberg-Kohn functional* F_{HK} forms the holy grail of DFT methods. In reality, nothing is known about the mathematical of F_{HK}. Only the classical Coulomb interaction $J\left[\rho\right]$

$$
V_{\text{ee}}\left[\rho\right] = J\left[\rho\right] + E_{\text{ncl}}\left[\rho\right] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{\mathbf{r}_{12}} d\mathbf{r}_1 \, d\mathbf{r}_2 + E_{\text{ncl}}\left[\rho\right], \qquad (2.18)
$$

is known, while $E_{\text{ncl}}\left[\rho\right]$ summarizes the non-classical part of the electron-electron interactions like self-interaction correction (since $J\left[\rho\right] \neq 0$ in an one-electron system), exchange and correlation effects.

The Second Hohenberg-Kohn Theorem

All properties of the system can be derived from the ground state density, formally also for the excited states, but by different functionals, since F_{HK} connects to the ground state energy. However, commonly DFT is called a ground state method, because F_{HK} provides only the ground state energy for the ground state electron density. For all other trial electron densities $\tilde{\rho}$, also fulfilling the restrictions in equation 2.7, the functional H connects to energies higher than the ground state energy E_0, which is exactly the variational principle, shortly introduced above, known from wave function methods

$$
E_0 \leq H\left[\tilde{\rho}\right]. \qquad (2.19)
$$

This concept *the ground state energy is the minimum energy of the system* is restricted only to the ground state and cannot be applied to excited states.

2.1.2 Kohn-Sham Approach

In addition to the Hohenberg-Kohn theorems, the *Kohn-Sham Approach* (KS) is very fundamental in DFT methods, since KS offers the lacking scheme for obtaining the ground state electron density and energy [80]. A non-interacting reference system (nic) based on one-electron orbitals is introduced to provide a simple and efficient way for calculating the kinetic energy contributions similar to wave function methods. A system of non-interacting electrons moving in an effective potential is exactly described by a single *Slater determinant* as in HF [81]. Consequently, the Hamiltonian for this system is formulated as

$$\mathcal{H}_{\mathrm{nic}} = -\frac{1}{2}\sum_{i}^{N} \nabla^2 + \sum_{i}^{N} V_{\mathrm{nic}}(\mathbf{r}_i), \tag{2.20}$$

and the related wave function Θ_{nic} is constructed by a set of one-electron spin orbitals $\{\psi\}$

$$\Theta_{\mathrm{nic}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \cdots & \psi_N(\mathbf{x}_N) \end{vmatrix}. \tag{2.21}$$

The following mathematics is equivalent to HF, thus the optimization of Θ_{nic} is performed by varying $\{\psi\}$ to minimize the resulting energy, restricted by the orthogonality $\langle \psi_i | \psi_j \rangle = \delta_{ij}$

$$\frac{\partial E_{\mathrm{nic}}}{\partial \psi} = \frac{\partial}{\partial \psi} \langle \Theta_{\mathrm{nic}} | \mathcal{H}_{\mathrm{nic}} | \Theta_{\mathrm{nic}} \rangle \overset{!}{=} 0. \tag{2.22}$$

This procedure leads to the *Kohn-Sham operator* $\widehat{f}_{\mathrm{KS}}$ and the *Kohn-Sham orbitals* ψ_i

$$\widehat{f}_{\mathrm{KS}}\, \psi_i = \epsilon_i\, \psi_i, \tag{2.23}$$

$$\widehat{f}_{\mathrm{KS}} = -\frac{1}{2}\nabla^2 + V_{\mathrm{nic}}(\mathbf{r}). \tag{2.24}$$

The connection of the real and non-interacting system is given by the potential V_{nic} (eq. 2.20) satisfying the condition

$$\rho_{\mathrm{nic}}(\mathbf{r}) = \sum_{i}^{N} |\psi_i(\mathbf{r})|^2 = \rho_0(\mathbf{r}). \tag{2.25}$$

Equivalent to the classical Coulomb part in V_{ee} (eq. 2.18), KS describes parts of T (eq. 2.15) with the introduced non-interacting system

$$T_{\mathrm{nic}} = -\frac{1}{2}\sum_{i}^{N} \langle \psi_i | \nabla^2 | \psi_i | \psi_i | \nabla^2 | \psi_i \rangle \neq T_{\mathrm{exact}}. \tag{2.26}$$

Definitely, there is a difference to the exact kinetic energy T_{exact}. Nevertheless, KS separates all known contributions from F_{HK} (eq. 2.15) and merges the unknown parts to the exchange-correlation functional E_{xc}

$$F\left[\rho\right] = T_{\mathrm{nic}}\left[\rho\right] + J\left[\rho\right] + E_{\mathrm{xc}}\left[\rho\right], \tag{2.27}$$

$$E_{\mathrm{xc}}\left[\rho\right] = T_{\mathrm{exact}}\left[\rho\right] - T_{\mathrm{nic}}\left[\rho\right] + V_{\mathrm{ee}}\left[\rho\right] - J\left[\rho\right]. \tag{2.28}$$

The above required potential V_{nic} of the non-interacting system, is determined by minimizing

$$\begin{aligned} H\left[\rho\right] &= T_{\text{nic}}\left[\rho\right] + J\left[\rho\right] + E_{\text{xc}}\left[\rho\right] + E_{\text{ne}}\left[\rho\right], \\ &= -\frac{1}{2}\sum_{i}^{N}\langle\psi_i|\nabla^2|\psi_i|\psi_i|\nabla^2|\psi_i\rangle + \frac{1}{2}\sum_{i}^{N}\sum_{A}^{M}\int\int|\psi_i(\mathbf{r}_1)|^2\frac{1}{\mathbf{r}_{12}}|\psi_j(\mathbf{r}_2)|^2\,d\mathbf{r}_1\,d\mathbf{r}_2, \\ &+ E_{\text{xc}}\left[\rho\right] - \sum_{i}^{N}\sum_{A}^{M}\int\frac{Z_A}{\mathbf{r}_{1A}}|\psi_i(\mathbf{r}_1)|^2\,d\mathbf{r}_1, \end{aligned} \tag{2.29}$$

which is performed by varying ψ_i. Resulting in

$$\left(-\frac{1}{2}\nabla^2 + \left[\int\frac{\rho(\mathbf{r}_2)}{\mathbf{r}_{12}}\,d\mathbf{r}_2 + V_{\text{xc}}(\mathbf{r}_1) - \sum_{A}^{M}\frac{Z_A}{\mathbf{r}_{1A}}\right]\right)\psi_i = \left(-\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{r}_1)\right)\psi_i = \epsilon_i\,\psi_i, \tag{2.30}$$

and stating by comparison of the equations 2.24 and 2.30

$$V_{\text{nic}}(\mathbf{r}_1) \equiv V_{\text{eff}}(\mathbf{r}_1). \tag{2.31}$$

Since the solution of equation 2.29 depends on the ground state electron density, which is determined by the one-electron orbitals, defined in equation 2.24 being again dependent on the effective potential and thus on the electron density, an iterative solution process is required similar to HF. Up to this point KS and in general DFT methods are accurate, since no approximations are included. However, the lacking knowledge about the mathematical expression for E_{xc}, causes the need of approximations leading to inaccuracies and to non-exact DFT methods, in practice.

2.1.3 Approximations to the Exchange-Correlation Functional

The major contributions are known exactly in the KS (eq. 2.29), whereas the only exception is conglomerated in the exchange-correlation functional E_{xc}. Thus, the accuracy of DFT methods exclusively depends on the approximation for E_{xc}. In contrast to wave function methods, the systematic instruction for continuously improving the accuracy of the DFT *ab initio* results is not existent, being disadvantageous. Nevertheless, diverse E_{xc} functionals are used being different in the complexity with a trend of enhanced accuracy [82].

A first common approach is the *local (spin) density approximation* (LDA or LSDA) based on a homogeneous electron gas. A concept of a system filling up an infinite volume with an infinite number of electrons, charge neutralized by a positive background charge. Over the whole space, the electron density illustrates a finite, but constant value. The underlying functional $E_{\text{xc}}^{\text{LDA}}/E_{\text{xc}}^{\text{LSDA}}$ for this model concept is known high in accuracy

$$E_{\text{xc}}^{\text{LDA}}\left[\rho\right] = \int\rho(\mathbf{r})\,\epsilon_{\text{xc}}\left(\rho(\mathbf{r})\right)d\mathbf{r}, \tag{2.32}$$

$$E_{\text{xc}}^{\text{LSDA}}\left[\rho_\alpha,\rho_\beta\right] = \int\rho(\mathbf{r})\,\epsilon_{\text{xc}}(\rho_\alpha(\mathbf{r}),\rho_\beta(\mathbf{r}))\,d\mathbf{r}, \tag{2.33}$$

including the exchange-correlation energy per electron ϵ_{xc}. LDA only considers the local value of the electron density, in contrast to the *generalized gradient approximation* (GGA), which includes also the gradient of the electron density $\nabla\rho$ in the exchange-correlation contributions

$$E_{\mathrm{xc}}^{\mathrm{GGA}}\left[\rho_\alpha \rho_\beta\right] = \int f(\rho_\alpha(\mathbf{r}), \rho_\beta(\mathbf{r}), \nabla\rho_\alpha(\mathbf{r}), \nabla\rho_\beta(\mathbf{r}))\, d\mathbf{r}. \tag{2.34}$$

The integrand f may also include parameters adjusting theoretical results to experimental data contrary to functionals like PBE solely based on physical quantities. Also higher order derivatives like the Laplacian ∇^2 are used in E_{xc} functionals, which are called meta-GGA functionals considering the kinetic energy density $\nabla^2\rho$.

2.2 Basis Set

Among the exchange-correlation functional E_{xc}, also a set of basis functions $\{\psi\}$ is required to be defined in order to construct the Slater determinant in equation 2.21. Several atom-centered basis function can be combined by the *linear combination of atomic orbitals* (LCAO) to describe molecular orbitals, describing bonding, non-bonding, anti-bonding or free electron pairs, while thinking in terms of a molecule. Different types of atom-centered basis functions exist. *Slater type orbitals* (STOs), are the analytical solution of hydrogen-like systems. However, STO are computationally demanding, when calculating overlap integrals. This disadvantage is overcome by *Gaussian type orbitals* (GTOs) being more efficient in calculating overlap integrals, since LCAO of GTOs does not change the underlying mathematics. Furthermore, GTOs can be contracted to adapt the basis functions. Combining the basis functions results in different basis sets like *split-valence* basis sets [83] increasing the basis functions to describe valence electrons, *correlation-consistent* basis sets [84] for a systematic converging to the complete basis set limit of post HF methods or *polarization-consistent* basis sets [85] as the DFT counterpart of the latest. Furthermore, there are non-atom-centered basis functions like *plane waves*, which are common for PBC systems. Among these analytic basis functions, there are also numerical basis functions, called numerical atomic orbitals (NAO) with a highly flexible form due to the pre-tabulated values. These basis functions can be quite efficient in calculations, although numerical integration is needed.

2.3 Dispersion Corrections

Among inter atomic interactions, between directly bonded atoms, there are also interactions between atoms not directly bonded to each other. These interactions can be separated into polar/electrostatic interactions E_{static}, which tackle the attraction of oppositely charged atoms or molecular parts and non-polar, also named *van der Waals* (vdW) interactions. The energy E_{vdW} resulting from these interactions becomes zero for large distances d and very repulsive for short distances, because of overlapping, negatively charged electron clouds Nevertheless, such vdW interactions are interesting, since at intermediate distances, a slight attractive region arises based on induced dipole-dipole contributions. Also higher multipoles are involved, but the major part arises from dipole contributions asymptotically vanishing with d^{-6}. The resulting forces from this interaction are named *London dispersion forces* [86]. For rare gas atoms, these are the only interactions and effect the atoms to form a liquid and a solid phase. Also for non-polar molecules like hydrocarbons, vdW contributions are the main interactions. Thus, it is expected these long-range contributions to become important for the MOF structures used in this work. The general form of the vdW interactions can be described as the difference of the repulsion and attraction term, described by the pairwise dispersion coefficient C_{AB}^6 of atom/molecular fragment/molecule

A and B

$$E_{\mathrm{vdW}} = E_{\mathrm{repulsion}}(d_{AB}) - \frac{C^6_{AB}}{(d_{AB})^6}. \tag{2.35}$$

Commonly used E_{xc} functionals do not correctly represent these long-range interactions and in DFT calculations it became a standard procedure to include dispersion correction schemes to gain the wanted accurate results. Different types of dispersion correction schemes are present in modern DFT methods [87]. In this work the *Tkatchenko-Scheffler method* (TS) [88] is used for including the dispersion corrections by an additional term just summed up with the energy E_{KS} of the *Kohn-Sham Approach*

$$\begin{aligned} E_{\mathrm{corrected}} &= E_{\mathrm{KS}} + E^{\mathrm{TS}}_{\mathrm{vdW}}, \\ E^{\mathrm{TS}}_{\mathrm{vdW}} &= -\frac{1}{2}\sum_{A,B} f_{\mathrm{damp}}(d_{AB})\frac{C^6_{AB}}{d^6_{AB}}, \end{aligned} \tag{2.36}$$

including the electron density dependent atom pair A–B dispersion coefficient C^6_{AB} for the internuclear distance d_{AB}. The damping function f_{damp} eliminates singularities at short distances.

2.4 Relativistic Corrections

The two main aspects of special relativity theory are the constant velocity of light and the invariance of physical laws in different inertial frames of reference and deals with inertial frames with constant velocity to each other. So the special relativity is a good approximation for the movement of electrons around a nucleus, when the electrons' velocity is a significant fraction of the speed of light. Since the speed of light c is constant, the relativistic mass m_{rel} of objects increases with its velocity v compared to the rest mass m_0

$$m_{\mathrm{rel}} = m_0 \sqrt{1 - \frac{v^2}{c^2}}^{-1}. \tag{2.37}$$

In a one electron-system with nuclear charge Z_A, like a hydrogen atom, the total energy of the 1s-electron is

$$E^{\mathrm{1s}}_{\mathrm{tot}} = -\frac{Z_A^2}{2}, \tag{2.38}$$

which is due to the *virial theorem* equivalent to its negative kinetic energy $T^{\mathrm{1s}}_{\mathrm{e}}$ resulting in a classical velocity

$$\begin{aligned} E^{\mathrm{1s}}_{\mathrm{tot}} &= -T^{\mathrm{1s}}_{\mathrm{e}}, \\ -\frac{Z_A^2}{2} &= -\frac{m_0 v^2}{2} \text{ with } m_0 = 1, \\ v &= Z_A. \end{aligned} \tag{2.39}$$

Thus, the velocity of the 1s-electron in heavy-core elements becomes significantly large in comparison to the speed of light $c = 137.036\, a.\, u.$, given in atomic units. In turn, relativistic effects increase with Z_A and get more significant. The relativistic wave function can be expressed as four-dimensional vector

$$\Psi = \begin{pmatrix} \Psi_{\mathrm{L}\alpha} \\ \Psi_{\mathrm{L}\beta} \\ \Psi_{\mathrm{s}\alpha} \\ \Psi_{\mathrm{s}\beta} \end{pmatrix}, \tag{2.40}$$

including large components $\Psi_{\mathrm{L}\alpha,\beta}$ and small components $\Psi_{\mathrm{s}\alpha,\beta}$. Because of the vectorial form of the wave function the solution for the large component is dependent on the small component

$$\begin{aligned} c(\boldsymbol{\sigma} \cdot \mathbf{p})\Psi_{\mathrm{s}} + \mathbf{V}\Psi_{\mathrm{L}} &= E\Psi_{\mathrm{L}}, \\ c(\boldsymbol{\sigma} \cdot \mathbf{p})\Psi_{\mathrm{L}} + (-2mc^2 + \mathbf{V})\Psi_{\mathrm{s}} &= E\Psi_{\mathrm{s}}. \end{aligned} \tag{2.41}$$

Here, $\boldsymbol{\sigma}$ defines a *Pauli spin matrix*, $\mathbf{p}$ the momentum operator and $\mathbf{V}$ an electric potential, e. g. based on the nuclei. Ψ_{s} can be solved and expressed by terms of Ψ_{L} with the factor $\mathbf{K}$

$$\begin{aligned} \Psi_{\mathrm{s}} &= \mathbf{K}\frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{2mc}\Psi_{\mathrm{L}}, \\ \mathbf{K} &= \left(1 + \frac{E - \mathbf{V}}{2mc^2}\right)^{-1}, \end{aligned} \tag{2.42}$$

leading to

$$\left[\frac{1}{2mc}(\boldsymbol{\sigma} \cdot \mathbf{p})\mathbf{K}(\boldsymbol{\sigma} \cdot \mathbf{p}) + (\mathbf{V} - E)\right]\Psi_{\mathrm{L}} = 0 \tag{2.43}$$

by inserting equation 2.42 in equation 2.41. For the non-relativistic case $c \to \infty$ and thus $\mathbf{K} = 1$, equation 2.43 simplifies to an equivalent of the already known *Schrödinger equation* (eq. 2.1). However, the current representation of $\mathbf{K}$ leads to divergent behaviour of the wave function near the nuclei and can be avoided by representing Ψ_{s} and $\mathbf{K}$ as

$$\begin{aligned} \Psi_{\mathrm{s}} &= \mathbf{K}'\frac{c(\boldsymbol{\sigma} \cdot \mathbf{p})}{2mc^2 - \mathbf{V}}\Psi_{\mathrm{L}}, \\ \mathbf{K}' &= \left(1 + \frac{E}{2mc^2} - \mathbf{V}\right)^{-1}. \end{aligned} \tag{2.44}$$

$\mathbf{K}'$ can be expressed by an power series expansion, for which the zeroth order approximation leads to $\mathbf{K}' = 1$, which is named the *Zeroth-Order Regular Approximation* (ZORA)

2.5 Extended Systems

Solid state materials, like MOFs, are from an atomistic view endless in each spatial direction. For sure, a system with an infinite number of atoms cannot be described by *ab initio* methods. Anyway, the infinite large structure follows a certain kind of translational symmetry, inherently connected to the structure. The symmetry, in general, defines the unit cell, which includes all information about the underlying system. Based on the unit cell the endless and infinite structure can be reconstructed in combination with the translational symmetry. In theory, only the unit cell is considered including *Periodic Boundary Conditions* (PBC).

2.5.1 Reciprocal Space

The symmetry of the structure defines a 3D lattice, which is filled up by the in-equivalent atomic sites related to the lattice point $\mathbf{r}$. All lattice points are, as mentioned above, equivalent, due to the translational symmetry

$$\begin{gathered} \mathbf{r} = \mathbf{r} + \mathbf{R} = \mathbf{r} + i\,\mathbf{a}_1 + j\,\mathbf{a}_2 + l\,\mathbf{a}_3, \\ i, j, l \in \mathbb{Z}. \end{gathered} \tag{2.45}$$

The basis vectors $\mathbf{a}_i$ $(i \in \{1,2,3\})$ of the real lattice (position space) are defined by the unit cell and simultaneously define how to reach the equivalent lattice points $\mathbf{R}$. In contrast, the electronic structure of the considered material is characterized in the reciprocal space (momentum space), which is determined by the basis vectors $\mathbf{b}_i$ $(i \in \{1,2,3\})$. These two different typs of basis vectors are related to each other via

$$\mathbf{b}_1 = \frac{2\pi}{V}(\mathbf{a}_2 \times \mathbf{a}_3), \quad \mathbf{b}_2 = \frac{2\pi}{V}(\mathbf{a}_3 \times \mathbf{a}_1), \quad \mathbf{b}_3 = \frac{2\pi}{V}(\mathbf{a}_1 \times \mathbf{a}_2). \tag{2.46}$$

In analogy to the real lattice points, the reciprocal lattice points follow a certain type of translational symmetry and thus the reciprocal lattice points are equivalent for

$$\begin{gathered}\mathbf{g} = \mathbf{g} + \mathbf{G} = \mathbf{g} + m\,\mathbf{b}_1 + n\,\mathbf{b}_2 + o\,\mathbf{b}_3, \\ m, n, o \in \mathbb{Z}.\end{gathered} \tag{2.47}$$

Among the common unit cell, there are primitive unit cells, including only one lattice point at a time. A specific construction is called the *Wigner-Seitz* primitive unit cell, whose counter part in the reciprocal space is called *Brillouin Zone* (BZ). The components k of the wave vector $\mathbf{k}$ are restricted to

$$-\frac{\pi}{|\mathbf{a}_i|} \leq k \leq \frac{\pi}{|\mathbf{a}_i|}, \tag{2.48}$$

in the first BZ. That means for a mathematical description of an infinite solid, it is necessary to consider all possible $\mathbf{k}$, thus to integrate over all $\mathbf{k}$. The integration for such a complex task is very demanding and a more convenient procedure is to map the complete BZ onto certain points of $\mathbf{k}$, which get summed up. This mapping is called $\mathbf{k}$ *point sampling* and has to be checked specifically for the system of interest.

2.5.2 Bloch Theorem

Equivalent to the structure, also the atomic potential for the electrons is repeated with the unit cells

$$V(\mathbf{r}) = V(\mathbf{r} + \mathbf{R}). \tag{2.49}$$

For this periodic potentials, there are specific functions solving the electronic *Schrödinger equation*, which are called *Bloch Functions*

$$\varphi(\mathbf{r}, \mathbf{k}) = \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}}\, e^{i\mathbf{G}\mathbf{r}} e^{i\mathbf{k}\mathbf{r}} = u(\mathbf{r}, \mathbf{k}) e^{i\mathbf{k}\mathbf{r}} = u(\mathbf{r} + \mathbf{R}, \mathbf{k}) e^{i\mathbf{k}\mathbf{r}}. \tag{2.50}$$

A translation by $\mathbf{R}$ or a multiplication by a phase factor $e^{i\mathbf{k}\mathbf{R}}$ is stated as equivalent

$$\varphi(\mathbf{r} + \mathbf{R}, \mathbf{k}) = u(\mathbf{r} + \mathbf{R}, \mathbf{k}) e^{i\mathbf{k}(\mathbf{r}+\mathbf{R})} = u(\mathbf{r} + \mathbf{R}, \mathbf{k}) e^{i\mathbf{k}\mathbf{r}} e^{i\mathbf{k}\mathbf{R}} = u(\mathbf{r}, \mathbf{k}) e^{i\mathbf{k}\mathbf{r}} e^{i\mathbf{k}\mathbf{R}} = \varphi(\mathbf{r}, \mathbf{k}) e^{i\mathbf{k}\mathbf{R}}, \tag{2.51}$$

which is known as the *Bloch Theorem.*

2.6 High-Dimensional Neural Networks

As already mentioned in the introduction, MLPs can be classified to one of four MLP generations. Starting with the first generation applied only to small systems. Although, the predicted results of such NNPs are very accurate, this method is quite limited due to missing invariances like permutational invariance, but also due to its limited system size, because of the size-dependent NNPs.

2.6.1 The Energy Expression

In 2007, Behler and Parrinello introduced the second generation of MLPs as *high-dimensional neural network potentials* (HDNNP) [15,17,18] to over come the limitations of the first generation. One important modification is the splitting of the total potential energy

$$E_{\text{tot}} = \sum_{A=1}^{M} E_A \quad , \tag{2.52}$$

into a sum of M artificial atomic energies E_A. In contrast to the total potential energy the atomic energy contributions are no observable physical quantities, only auxiliary quantities for the mathematical concept for the construction of the total potential energy. Each atomic energy contribution is predicted by an element-specific atomic NN (fig. 2.1) and depends on the specific atomic environment $\mathbf{S}_A$. Introducing environment dependent descriptors – the second modification – called *atom-centered symmetry functions* (ACSFs) ensures the local dependence of the atomic energies up to the cutoff radius r_{cut}. With this concept, the limited size of first generation NNPs can be overcome, since the atomic energy contributions are predicted strictly depending on the local environment, but independent on the total number of nuclei M. Thus, adding and removing an atom B does not effect the description of atom A, if the atomic distance is larger than the cutoff radius

$$d_{AB} > r_{\text{cut}} \quad . \tag{2.53}$$

Many different ACSFs are combined to form a vector $\mathbf{S}_A$, describing the specific environment of atom A, and translate the commonly used Cartesian coordinates into internal-like relative coordinates (distances and angles).

The atomic environment, thus the symmetry function vector $\mathbf{S}_A$, is used as the input for the element-specific atomic feed-forward NN (fig. 2.1), which processes the input via diverse hidden layers, two in this example, including n_1 and n_2 neurons, and predicts the atomic energy as the output of the last layer – the output layer. Each component $n_{\mathbf{S}}$ of $\mathbf{S}_A$ is connected to the n_1 neurons of the first hidden layer via the weights $a^{01}_{n_{\mathbf{S}}n_1}$. Similarly, any neuron μ of any layer ϵ is connected to the neurons ν in the following layer σ via the weights $a^{\epsilon\sigma}_{\mu\nu}$. Additionally, all the neurons, but the input neurons, are connected to the bias node via the bias weights b^{ϵ}_{μ}. The mathematical expression for each neuron value y^{σ}_{ν} is given by

$$y^{\sigma}_{\nu} = f^{\sigma}_{\nu}\left(b^{\sigma}_{\nu} + \sum_{\mu}^{n_\epsilon} a^{\epsilon\sigma}_{\mu\nu} y^{\epsilon}_{\mu}\right) \quad , \tag{2.54}$$

and thus, neuron ν of layer σ is evaluated by summing up all values of the previous layer neurons y^{ϵ}_{μ}, which are multiplied with the related connecting weight $a^{\epsilon\sigma}_{\mu\nu}$ and its bias weight b^{σ}_{ν}, followed by executing the activation function f^{σ}_{ν}. In case of HDNNP, the hyperbolic tangent

$$f(x) \equiv \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad , \tag{2.55}$$

is used as the activation function for the neurons in the hidden layers, introducing the non-linear character, and for the output neuron a linear activation function, which is numerically unrestricted in the range of the output energies. The functional form of the

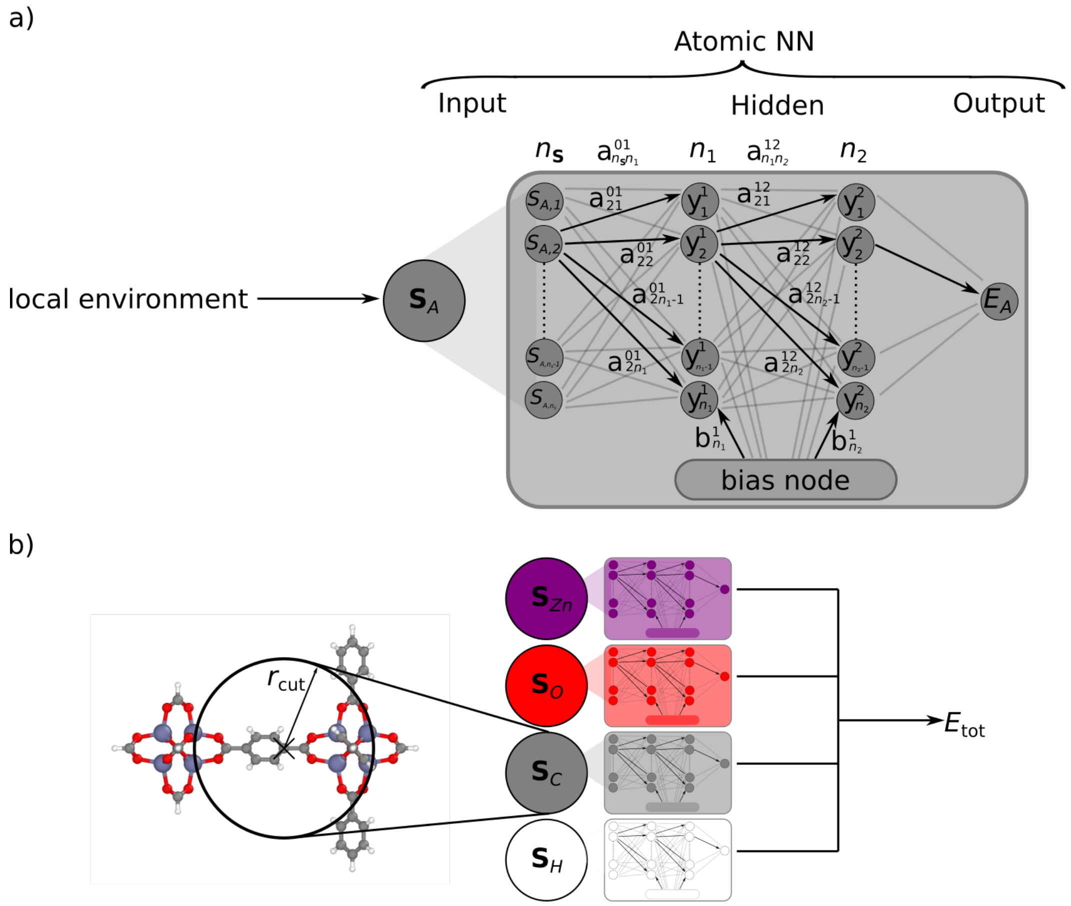

Figure 2.1: a) An atomic neural network as a part of a high-dimensional neural network (HDNNP). Atomic NNs determine the atomic energy E_A of atom A (in this case equivalent to element A) by processing the local atomic environment, which is described by the ACSF vector $\mathbf{S}_A$ containing the $n_\mathbf{S}$ ACSFs $S_{A,n_\mathbf{S}}$, via the hidden layers. The two hidden layers *1* and *2* consist of n_1 and n_2 neurons, respectively. The neurons of layer *1* are connected to the neurons in layer *2* by weights $a^{12}_{n_1 n_2}$ and are additionally connected to a bias note via the bias weights $b^1_{n_1}$. b) The environment of a specific carbon atom is described by an element specific ACSF vector $\mathbf{S}_\mathrm{C}$ and its energy is calculated by the element specific atomic NN. All atomic energies E_A are finally summed up to the total energy of the system E_tot. The figures in this work were created by Ovito [89], matplotlib [90] and inkscape [91].

atomic energy contributions of the element-specific NN is given by

$$E_A \quad = \quad f^3_1 \left\{ b^3_1 \sum_k^6 a^{23}_{k1} \cdot f^2_k \left[b^2_k + \sum_{j=1}^{2} a^{12}_{jk} \cdot f^1_j \left(b^1_j + \sum_{i=1}^{1} a^{01}_{ij} \cdot S_{A,i} \right) \right] \right\} \quad , \qquad (2.56)$$

which correlates the structural information, the local environment of atom A up to the cutoff radius r_cut given by the element specific symmetry function vector $\mathbf{S}_A$, to the atomic energy as a part of the total potential energy sum E_tot (eq. 2.52). As already mentioned, the system usually includes different types of elements, but for the equation 2.56 a sub or super script specifying the element is excluded for clarity. Nevertheless, the weights and biases, the symmetry functions and also in general the architecture of the atomic NN is element specific, although the number of hidden layers, the including nodes and the cutoff radius are the same in most practical cases.

2.6.2 Atom-Centered Symmetry Functions as Structural Descriptors

The already mentioned atom-centered symmetry functions (ACSF) translate the local atomic environment dependent on the nuclear Cartesian coordinates up to a pre-defined cutoff radius into a NN-readable format. The energy of a system is invariant with respect to the translation and rotation of the total system and also to the order of atoms – the permutational invariance. These invariances have to be fulfilled by the descriptors, as provided by ACSFs, of the structures as well. If not, different input coordinates, like a translated or rotated structure, would be related to the same energy, leading to contradictions. Furthermore, the symmetry function vector $\mathbf{S}_A$ dimensionality is pre-defined, fixed and independent on the number of atoms included in the system.

Among many types of symmetry functions, two ACSFs commonly used are referred to the radial and angular ACSF [20]. The radial ACSF (fig. 2.2) describe the distances d_{AB} between the specific atom A and its neighboring atoms B in each spatial direction around atom A,

$$S_A^{\mathrm{rad}} = \sum_B e^{-\eta(d_{AB}-r_{\mathrm{shift}})^2} \cdot f_{\mathrm{cut}}\left(d_{AB}\right) \quad , \tag{2.57}$$

with the real, non-negative parameter η, determining the width of the Gaussian part. The shifting parameter r_{shift} switches the maximum of the radial ACSF and increases the resolution in this certain region. The angular ACSFs (fig. 2.3) describe the orientation of the neighboring atoms B and C around atom A, depending on the angle θ_{ABC} between

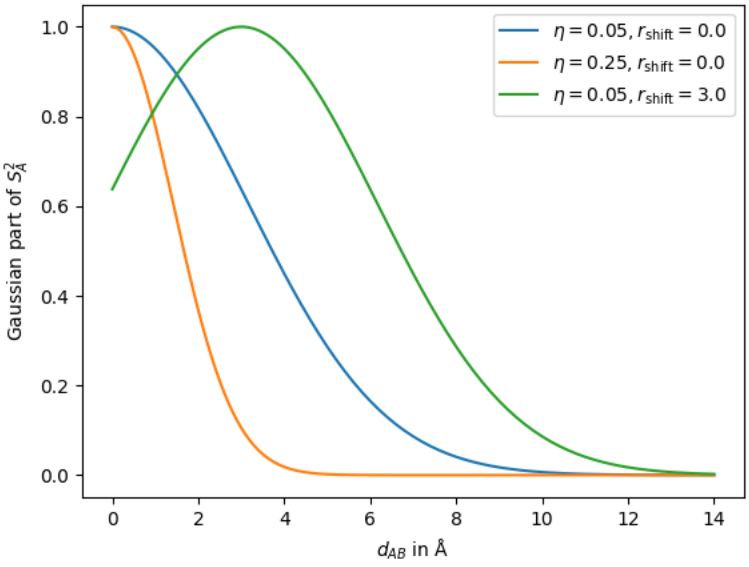

Figure 2.2: The Gaussian part of the radial ACSF (eq. 2.57) illustrating the behavior of the parameter η and r_{shift}.

the atoms, which is centered at atom A

$$S_A^{\text{ang}} = 2^{1-\zeta} \sum_B \sum_C \left[1 + \lambda \cdot \cos\left(\theta_{ABC}\right)\right]^{\zeta} \cdot e^{-\eta\left(d_{AB}^2 + d_{AC}^2 + d_{BC}^2\right)} \cdot f_{\text{cut}}\left(d_{AB}\right) \cdot f_{\text{cut}}\left(d_{AC}\right) \cdot f_{\text{cut}}\left(d_{BC}\right) \quad . \tag{2.58}$$

Similar to η, the parameter ζ (usually $\zeta \in \{1, 2, 4, 16\}$) determines the width of the cosine part. The normalization factor $2^{1-\zeta}$ guarantees a fixed value range of the cosine function, which changes with different ζ otherwise. The parameter $\lambda \in \{-1, 1\}$ inverts the cosine function and offers an additional option to increase the resolution for the angular environment. The cutoff function (fig. 2.4),

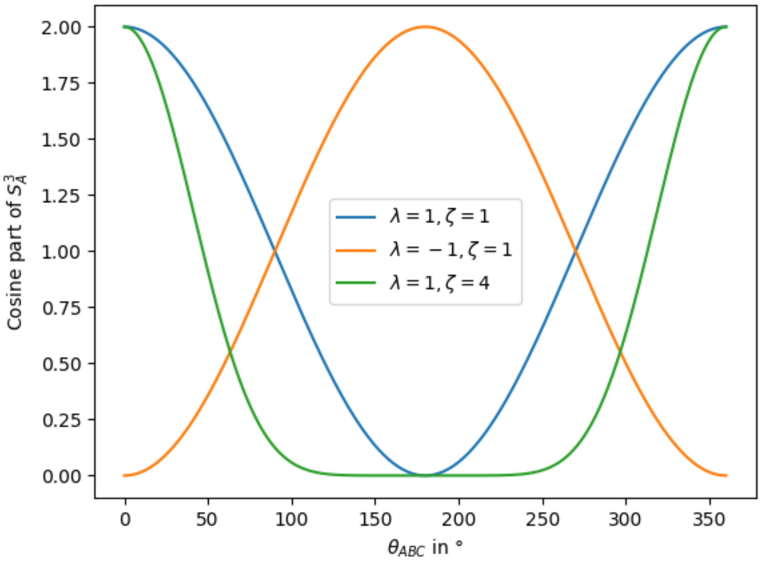

Figure 2.3: The cosine part of the angular ACSF (eq. 2.58) illustrating the behavior of the parameter λ and ζ.

$$f_{\text{cut}}\left(d_{AB}\right) = \begin{cases} 1 & \text{for} \quad d_{AB} \leq r_{\text{inner,cut}} \\ 0.5 \cdot \left[\cos\left(\pi x\right) + 1\right] & \text{for} \quad r_{\text{inner,cut}} \leq d_{AB} \leq r_{\text{cut}} \\ 0 & \text{for} \quad d_{AB} > r_{\text{cut}} \quad , \end{cases} \tag{2.59}$$

occuring in the mathematical expression of the ACSF (eq. 2.57 and 2.58) decays the ACSF smoothly to zero for $d_{AB} > r_{\text{cut}}$ in value and in slope. This ensures the continuity and differentiability of the ACSF, required for the optimization of the weights and biases, but also for the calculation of the forces, as presented below. The cutoff radius itself has to be chosen sufficiently large to include all relevant atomic interactions, usually around $6 - 10$ Å, but as small as possible to reduce the dimensionality of the configurational space, which needs to be covered by the training data set. Usually, the inner cutoff radius $r_{\text{inner,cut}}$ is set to zero.

All ACSF parameters defining the spatial shape are not changed during the training procedure, but pre-defined and fixed. Also the dimensionality of the element-specific ACSF vector is fixed and part of the NN architecture. Thus, the set of ACSFs is different for the element combinations occuring in the system. The aim of the ACSF vector is to provide a unique description for each important atomic environment used as input for the atomic

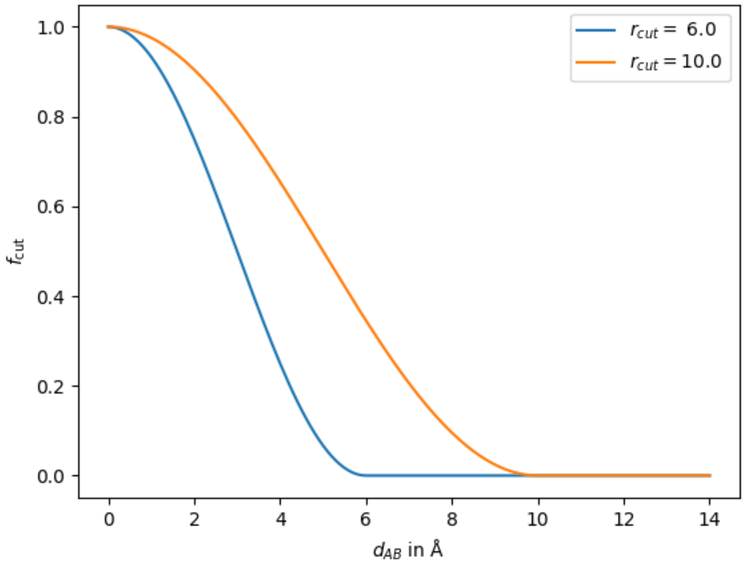

Figure 2.4: The cutoff function f_{cut} (eq. 2.59) illustrated for two different cutoff radii r_{cut}.

NN. Hence, the HDNNP is not only a universal approximator for the PES, but also a kind of classifier assigning a certain ACSF vector to the total potential energy via the auxiliary quantity E_A, the atomic energy contribution. Furthermore, HDNNPs can describe bond breaking and bond formation, since only the positions and the elements of all nuclei are used as input and no further definitions of bonding like in other classical potentials.

2.6.3 The Optimization of the Weights and Biases

To reproduce energies and forces of structures, a meaningful training data set including representative structures, sampling all significant degrees of freedom, needs to be constructed, which is used to optimize the weights and biases resulting in a good representation of the training data. This procedure is commonly known as *training the NN* or in this case as *training all atomic NNs simultaneously*, also *fitting the NN* is a common term. Mathematically, a *loss* or *error function* is defined as

$$f^{\mathrm{loss}} = \frac{1}{N_{\mathrm{struc}}^{\mathrm{train}}} \sum_{i=1}^{N_{\mathrm{struc}}^{\mathrm{train}}} \left(E_i^{\mathrm{NNP}} - E_i^{ref} \right)^2 \quad , \tag{2.60}$$

in case only the energy values are used for the training procedure, but will change when energies and forces are used. The error function is minimized with respect to the weights and biases. For this minimization problem, the adaptive, global, extended Kalman filter is consulted, performing the optimization iteratively [92, 93]. The performance of the resulting HDNNP needs to be carefully validated and if required, the training data set is extended referred to as *active learning* (sec. 3.4). The first validation of the HDNNP quality is assessed by the *root-mean squared error* (RMSE), which is defined as the sum of the squared energy differences between the HDNNP prediction E_i^{NNP} and the reference energy value E_i^{ref} of the $N_{\mathrm{struc}}^{\mathrm{train}}$ training data points. For the $N_{\mathrm{comp}}^{\mathrm{train}}$ force components f_i^{NNP}

the RMSE is defined equivalently

$$\mathrm{RMSE}(E^{\mathrm{train}}) = \sqrt{\frac{1}{N^{\mathrm{train}}_{\mathrm{struc}}} \sum_{i}^{N^{\mathrm{train}}_{\mathrm{struc}}} \left(E_i^{\mathrm{NNP}} - E_i^{ref}\right)^2} \quad , \tag{2.61}$$

$$\mathrm{RMSE}(f^{\mathrm{train}}) = \sqrt{\frac{1}{N^{\mathrm{train}}_{\mathrm{comp}}} \sum_{i}^{N^{\mathrm{train}}_{\mathrm{comp}}} \left(f_i^{\mathrm{NNP}} - f_i^{ref}\right)^2} \quad . \tag{2.62}$$

The RMSE values can be analyzed for each iteration of the optimization. As the fitting procedure progresses, the RMSE values are decreasing, which can lead to *overfitting* (fig. 2.5). Thus, the training data is represented perfectly, but the predictive power for data unknown to the HDNNP is reduced. To avoid these situations, $\sim 10\,\%$ of the training data set is extracted to form the testing data set. During the fitting procedure, the testing data is part of the HDNNP training procedure. The RMSE values are defined in the same manner as for the training data (eq. 2.61 and 2.62). The most accurate HDNNP representation can be derived, for small RMSE values of the training data set and small deviations to the RMSE testing data set are small. Only in this case, the HDNNP will provide an accurate representation of the training data with sufficient predictive power for related but unknown testing data.

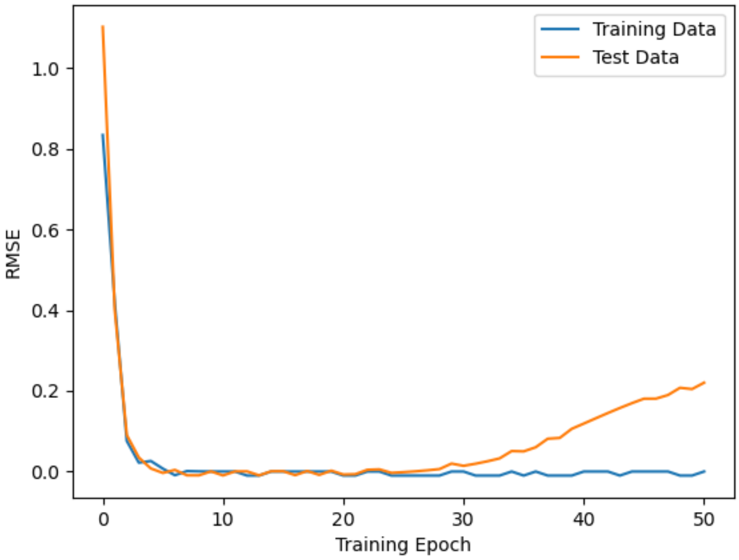

Figure 2.5: Convergence of an examplary RMSE. At training epoch ~ 25 overfitting can be detected, since the predictive power for the test data decreases stated by the increasing RMSE value.

2.6.4 Force calculation

Since the energy expression of the HDNNP and also the ACSF are known analytically, the force components f_{A_α} can be calculated analytically by the negative derivative of the total potential energy E_{tot} with respect to the coordinate A_α of atom A

$$f_{A_\alpha} = -\frac{\partial E_{tot}}{\partial A_\alpha} = -\sum_B \frac{\partial E_B}{\partial A_\alpha} \quad . \tag{2.63}$$

Not all atomic energies E_B are dependent on the coordinate A_α, but all atoms B within $d_{AB} \leq r_{\text{cut}}$. Consequently, the force component f_{A_α} formally depends on $2\,r_{\text{cut}}$, since the atomic energy E_B depends on all atoms within the cutoff radius of atom B.

2.7 Molecular Dynamics

The dynamical behaviour of an atomic system can be studied classically by applying Newton's equation of motion

$$f_{A_\alpha} = M_A \ddot{R}_{A_\alpha} \quad . \tag{2.64}$$

The force component f_{A_α} is the result of multiplying the mass M_A of atom A with the second time derivative – the accleration – $\ddot{R}_{A_\alpha}$ of its positional component α. Again, the force component f_{A_α} is related to the negative derivative of the energy, thus the PES provides the values for f_{A_α} as described in equation 2.63. Due to the many-body problem, the time evolution of a many-body system, must be solved numerically, since an analytical solution for this system does not exist. Instead the coupled motions are propagated using finite time steps dt by applying different time integrator algorithms [94–98]. This method is known as classical molecular dynamics (MD) simulations [99, 100].

The often implemented *Velocity Verlet algrithm* considers the current atomic position R_A and its velocity v_A for the time propagation

$$\mathbf{R}_A(t+dt) = \mathbf{R}_A(t) + \mathbf{v}_A(t)\,dt + \frac{\mathbf{f}_A(t)}{2M_A}dt^2 \quad , \tag{2.65}$$

$$\mathbf{v}_A(t+dt) = \mathbf{v}_A(t) + \frac{\mathbf{f}_A(t) + \mathbf{f}_A(t+dt)}{2M_A}dt \quad . \tag{2.66}$$

After the calculation of the propagated positions $R_A(t+dt)$, the forces for this new configuration is determined by the PES and finally the velocity can be updated for the current time step $t+dt$. The big advantage of the Velocity Verlet algorithm is the time-reversibility conserving the total energy among other conserved quantities, over a long period of time. This fact improves the quality of the underlying simulations.

Chapter 3

Computational Details

3.1 FHI-aims

The DFT calculations within this work are performed by the FHI-aims program package, version 171221 [101]. Approximations to the echange-correlation functional E_{xc} (eq. 2.29) are given by the revised version of the GGA functional PBE, called RPBE [102]. The DFT parameters are converged to a sub-meV accuracy ensured by the change of the total energy difference per atom $\Delta\Delta E_{\mathrm{tot}}^{\mathrm{atom}} \leq 0.001\,\mathrm{eV}$. This includes the NAO basis set, the related confinement radius, the number of radial integration shells, the angular integration grids, the *radial_ multiplier* keyword, the expansion of the atom-centered charge density and the **k**-point grid for IRMOF bulk calculations. Related convergence tests are summarized in section A.1, which also provides further keywords to construct the FHI-aims *control.in* file. During the SCF procedure, the electron density, the energy eigenvalue sum, the total energy and the atomic forces are converged to 10^{-4}, $10^{-2}\,\mathrm{eV}$, $10^{-6}\,\mathrm{eV}$ and $10^{-2}\,\mathrm{eV\,Å^{-1}}$, respectively. Dispersion corrections were included by TS as implemented in FHI-aims. FHI-aims recommends to include relativistic treatments for elements heavier than calcium, being indeed true for zinc. Furthermore, the relativistic corrections affect the energy convergence (tab A.1 andA.2) and thus, the *atomic_ zora* approach is included in the calculations. For structural optimizations, the FHI-aims implemented *bfgs* algorithm optimized the atomic positions until the atomic forces converged below $10^{-2}\,\mathrm{eV\,Å^{-1}}$, basically performed in combination with the *relaxed_ unit_ cell fixed_ angles* keyword to keep the cubic structure of the IRMOF bulks (fig. 4.1).

For the calculations of the numerical Hessian for molecular structures, stricter DFT parameters were used, being equivalent to the FHI-aims *tight* recommendations for the elements including additionally the first basis function of the second tier for zinc. Furthermore, the SCF cycle was forced to converged below 10^{-6}, $10^{-4}\,\mathrm{eV}$, $10^{-8}\,\mathrm{eV}$ and $10^{-4}\,\mathrm{eV\,Å^{-1}}$ for the electron density, the energy eigenvalue sum, the total energy and the atomic forces, respectively. This inconvenience in the DFT parameters is based on inconsistencies of the available compilers on the used *high-performance computing* (HPC) cluster, which was only found by coincidence.

3.2 Fragment Approach

The fragment approach in this work is based on the following rules. To construct a molecular fragment, one of the M in-equivalent atomic bulk sites is defined as the central atom A of the fragment F1. Every atom of the bulk structure within a sphere defined by the radius r_{frag} centered at atom A is included in the fragment structure (fig. 3.1a). The radius r_{frag} is called the fragment radius basically defining the size of the fragment. Furthermore, additional atoms are included in the fragment structure dependent on the already included atoms. The SBU, the phenylene ring and the carboxyl groups form entities, which are in-

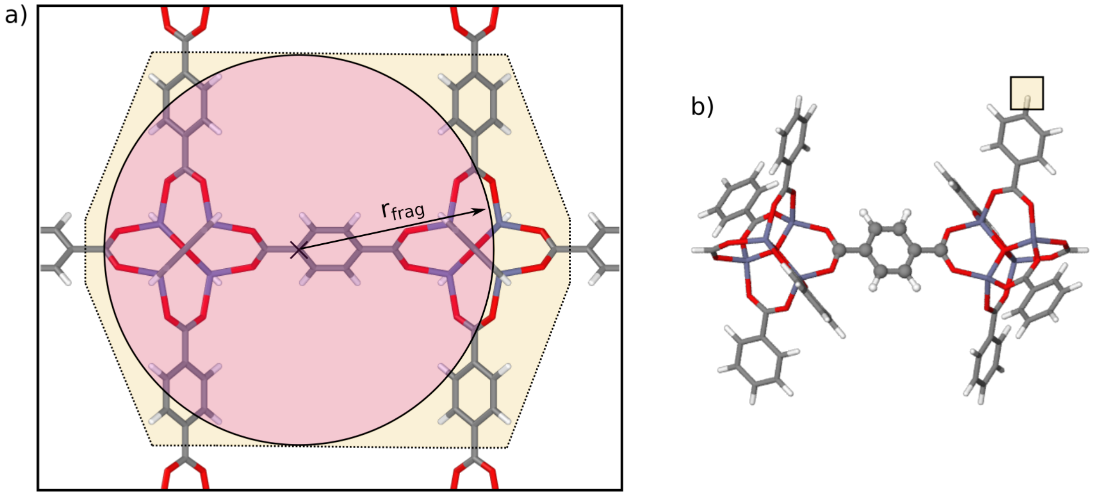

Figure 3.1: a) A 2D-projection of the IRMOF-1 periodic bulk structure (fig. 4.1a) with the marked central atom A = C2 as an exemplary in-equivalent atomic bulk position (black cross) with the sphere of radius r_{frag}, centered at the central atom (magenta-shaded, black circle around the cross) and the additionally included atoms (orange shaded area) to complete the partially added entities. b) The resulting molecular fragment structure highlighting the central atoms as balls and the remaining structure as sticks together with the saturation hydrogens (exemplary marked by the orange-shaded, black square). Zinc atoms are illustrated in violet, oxygen in red, carbon in grey, hydrogen in white in this work.

or excluded completely in the fragment structure. If any atom of an entity is located within the sphere around the central atom A, all remaining atoms of this entity will be included additionally, independent on the location within or outside the sphere. From a chemical point of view, these entities can be understood as functional groups based on a concept especially known in organic chemistry. This procedure ensures minimal deviations of the bulk and fragment electronic structure, which is crucial to keep the atomic interactions within the fragment similar to the bulk structure and thus accurately map atomic bulk interactions onto the atomic fragment interactions. The fragment radius r_{frag} at once describes the size of the molecular fragments and the bulk-like environment of the central atoms, and defines an upper boundary of the cutoff radius for the HDNNP descriptors, the ACSFs. Following these instructions for each in-equivalent atomic bulk site results in a set $\mathbf{F} = \{F_1, F_2, ..., F_M\}$ of molecular fragments, which ensures M in-equivalent atomic sites are embedded in a bulk-like environment at least in one fragment structure. However, larger fragments can include smaller, redundant fragments, leading to a set of non-redundant fragment structures. This set $\mathbf{F}' = \{F_1, F_2, ..., F_{M'}\}$ of non-redundant fragment structures fulfills also the criterion, that M in-equivalent atomic bulk sites are embedded in bulk-like environments at least within one of the M' ($M' < M$) non-redundant molecular fragment structures.

When constructing fragments from larger reference systems – like the bulk structure – bonds will be broken and may result in large changes of the electronic structure, which effect the interactions of the remaining atoms. Also these changes in the electronic structure need to be reduced to a minimum for an accurate mapping of the larger reference system interactions and properties onto the smaller fragment structure. In this work the broken bonds are saturated by hydrogen atoms (fig. 3.1b), which are placed along the broken bond in a 1.05 Å distance.

In quantum mechanics/molecular mechanics (QM/MM) methods and the related construction of the quantum mechanically treated region, similar problems occur. If a large system, e. g. an enzyme, should be investigated theoretically, the chemical interesting region of this large structure, the QM region, often includes only a few tens of atoms in contrast to the outer environment, including hundreds or even thousands of atoms, which is referred to the MM region. The different regions are treated by different levels of theory. The QM region is treated by a quantum mechanical method and the MM region only by classical molecular mechanics method. This spatial separation and the different levels of computational treatment reduce the effort to describe the whole structure theoretically. Advantageously, important long-range interactions – like electrostatics – determining certain properties of the QM region, can be considered in this ansatz. Hence, the goal of describing the large system, i. e. the enzyme, accurately by theoretical methods is in principle satisfied, because all atomic interactions are considered. Nevertheless, the problem arising here is to find solutions to merge accurately energetic contributions described on different levels of theory [103].

This can be transferred to the description of a larger reference structure, e. g. periodic IR-MOF bulk structure. Since the definition of a molecular fragment structure, being similar to QM region, based on the larger reference structure, which is the MM region equivalent, is comparably similar. However, all interactions of the molecular fragment (QM region) with the remaining environment of the large reference structure (MM region) introduces information from beyond the cutoff radius r_{cut}, which is indeed not applicable to MLPs, since this violates the locality ansatz (sec. 2.6). Furthermore, these beyond-cutoff radius assumptions would lead to contradictory information in the MLP data set, because the local environment can remain unchanged, while structural rearrangements beyond the cutoff radius may change atomic interactions and thus the force on the central atom. These contradictions cannot be resolved by MLPs due to the locality.

In summary, the molecular fragment structure must not include any information of the large reference structure beyond the cutoff radius. Additionally, any assumptions of the atomic structure beyond the cutoff radius must be avoided to ensure the locality ansatz. As a consequence, embedding schemes successfully applied in QM/MM methods are not applicable to MLPs.

3.3 RuNNer

The HDNNPs within this work (sec. 4.7) were constructed with the in-house program package RuNNer [15, 17, 18][1]. The atom-centered symmetry functions (ACSFs), used as structural descriptors, are summarized in the tables A.12 and A.13 for the HDNNPs $r'_{\mathrm{frag}}-2{-}\mathrm{SF1}$ and $r_{\mathrm{frag}}-2{-}\mathrm{SF1}$, as well as in tables A.14 to A.16. Each atomic NN consists of two hidden layers with 15 nodes each. For the hidden layers the activation function is defined by the hyperbolic tangent, in contrast to the output node, which is processed by a linear activation function. Furthermore, 10 % of the whole data set was separated for the test data set to identify and further avoid overfitting epochs during the HDNNP training process. The weights were optimized by the extended Kalman filter [92, 93] based on the reference total binding energies and atomic forces, as the binding energies defined

[1] In private communication a modified version of RuNNer was used in this work, based on changes by Alea Miako Tokita related to the Kalman filter.

as the difference of the total energies and the sum of the free atom energies. To improve the numerical stability of the training procedure, the ACSFs were rescaled to the interval $[0;1]$ and shifted by its average values to the origin, thus to the non-linear region of the hyperbolic tangent activation function. Furthermore, the random, initial weights were adapted to fit the reference energy average and the related standard deviation to reduce the RMSE of the energies in the starting epoch. During the training procedure the weights are optimized on a random order of the training points and an additional weight update based on the reference binding energies after the update based on the atomic forces, for each epoch.

3.4 Active Learning Procedure

For all MLPs, the reference data set forms the holy grail, since all information about atomic interactions are exclusively stored in the reference data set. A reliable and accurate representation of the PES by MLPs is crucially dependent on the training data set. MLPs permit accurate representations of the data set itself, due to the high flexibility of MLPs in representing any mathematical expression [12, 13], however this does not include an accurate representation of the PES simultaneously. The reasons for this fact are related to the overfitting problem mentioned in section 2.6 and the missing physical foundation of MLPs. Thus, an accurate and reliable MLP – more specific in this work a HDNNP – balances the accurate representation of the data set and the predictive power for the PES in general. Besides, the predictive power is affected by the resolution of the PES, which is related to the distribution of the data points over the PES. The more data points, the higher the resolution of the PES sampling and thus, the more information of the relation between the structure and the energy are included in the data set.

Commonly, a finite region of the PES, equivalent to a finite region of the full configurational space, is of interest depending on the specific application of the resulting HDNNP. In the following the phrases *PES* and *configurational space* relates to this finite region. Consequently, the data set has to sample the PES with a sufficient number of data points, but for efficiency reasons as less data points as possible, due to the computational effort of electronic structure calculations.

A systematic procedure to sample the PES is only possible for very small chemical systems including only a few degrees of freedom. In contrast, the high-dimensionality of the PES for larger systems prevents its systematic sampling, because of the high amount of possible atomic arrangements. Nevertheless, the PES can be sampled for high-dimensional systems with an iterative procedure converging to a complete data set, which is referred to as *active learning* (fig. 3.2). Based on initial structures, which can be generated from *ab initio* MD simulations as in this work, experimental data, as well as by many other procedures, and the reference electronic structure method, a reference data set can be constructed to train a HDNNP. Every MLP needs to be validated, which might be done by the analysis of the MLP training procedure. A simple quality feature of a MLP is given by the RMSE values of the training data set and the accurate representation of the individual training and test energies along with the atomic forces. Furthermore, a more detailed analysis of the data set and its representation can be performed by investigating the atomic energy range, the maximum atomic force components and the distribution of the ACSF values, to mention only a few. This may identify high energy, non-physical or missing intermediate structures. Further validation steps like applying the MLP to calculate certain kinds of properties as lattice constants, rotational barriers and many more can be performed.

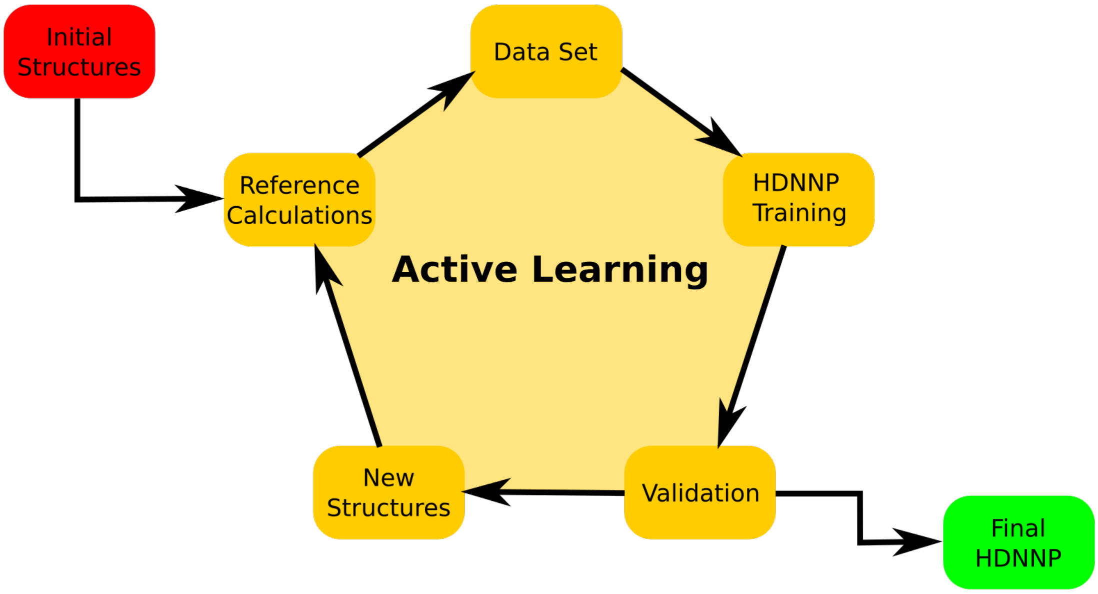

Figure 3.2: Iterative construction scheme for the reference data set, starting with initial structures, whose reference energies and atomic forces are calculated by the chosen reference electronic structure method to construct a data set for the HDNNP training procedure. The completeness of the data set and the PES representation accuracy need to be validated to ensure an accurate description of the essential degrees of freedom for the underlying system. Identified crucial but non-considered structures or structures of unreliably resolved PES regions can be added to the data set to increase the applicability of the HDNNP. Thus, another cycle of the active learning procedure is necessary, starting with new reference calculations based on the identified structures in the validation.

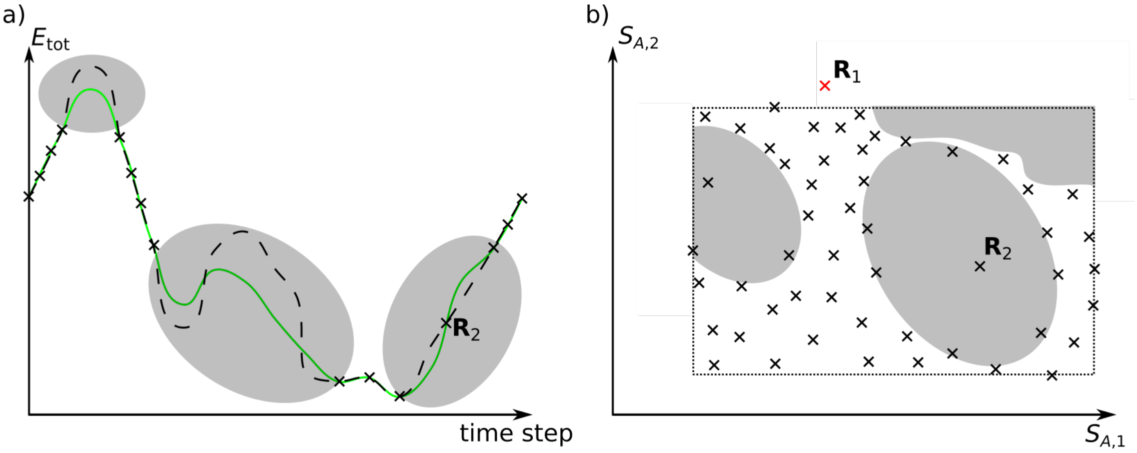

Figure 3.3: a) The total energy of a system E_{tot} over several trajectory time steps of a simulation for two different HDNNPs (green solid and dashed black line) with deviating predictions for the marked time steps (grey shaded areas) and some exemplary training data points (black crosses). The atomic arrangement $\mathbf{R}_2$ demonstrates the consequences of unreliably resolved PES regions leading to interpolation errors. b) Configurational space spanned by a two component ($S_{A,1}$ and $S_{A,2}$) atom-centered symmetry function (ACSF) vector (black dotted line) of a training data set including several training data points (black crosses) with unreliably resolved PES regions (grey shaded areas), due to few training data points. The atomic arrangement $\mathbf{R}_1$ causes extrapolation errors, because of ACSF values outside the trained range (dotted black line) and $\mathbf{R}_2$ causes interpolation errors, related to the insufficient resolution of the PES region, due to few training data points within the specific the region (grey shaded area).

The overall goal of these validation steps is to identify inaccurately represented data points, non-converged reference calculations or high-energy configurations, due to non-physical short bonds, which can be removed to improve the overall representation of the data set and the predictions of the PES. Additionally, inaccurately resolved regions of the PES can be identified, whose resolution need to be improved by adding structures to the training data set to increase the applicability of the MLP. Technically, these regions are determined by *extrapolation* and *interpolation* errors of the MLP (fig. 3.3). For HDNNPs, extrapolation errors are simple to identify and occur for an atomic arrangement $\mathbf{R}_1$ described by ACSF values outside the ACSF range of the training data set. Therefore, the HDNNP is obviously not trained to this kind of atomic arrangements and thus, accurate representations for the energy and atomic forces cannot be expected. Hence, this specific atomic arrangement $\mathbf{R}_1$ can be added to the training data set to extend the applicability of the HDNNP. Contrary, interpolation errors are more challenging to identify and occur for an atomic arrangement $\mathbf{R}_2$ described by ACSF values within the training data ranges. However, the predictions of two independent HDNNPs can deviate by more than a pre-defined tolerance for the total energy or atomic forces, demonstrating an inaccurate representation of these structural arrangements and a insufficient resolution for the related domain of the PES. To increase the resolution of the related domain, the structure $\mathbf{R}_2$ can be added to the training data set.

In this work, the *RuNNerActiveLearn* tool [49, 104] was used with a slight modification explained below, which was required for the active learning procedure in combination with the fragment approach (sec.3.2). The tool created different MD simulations based on the initial DFT optimized bulk IRMOF structures (fig. 4.1) in the isothermal-isobaric (NPT) ensemble processed by the Nose-Hoover barostat at a pressure of 1 bar, a time step of 1 fs and a total simulation duration of 200 ps. Furthermore, the range of simulation temperatures was iteratively increased up to 100 K. The program package n2p2 [105] processed the different MD simulations in this work based on the trained HDNNPs. Structures exceeding the ACSF space of the training data set (*extrapolation errors*) were analyzed by the tool to suggest additional structures for the extension of the training data set. Additionally, the predictions for two different and independent HDNNPs – different NN architecture, different seed for splitting data set and initializing the NN weights – were compared to identify atomic energy and atomic force component deviations by more than $\Delta E_{\text{atom}} \leq 0.00015\,\text{Ha}\,\text{atom}^{-1} \approx 0.004\,\text{eV}\,\text{atom}^{-1} \approx 5\text{RMSE}(E_{\text{atom}})$ and $\Delta f \leq 0.02\,\text{Ha}\,\text{a}_0{}^{-1} \approx 1\,\text{eV}\,\text{Å}^{-1} \approx 5\text{RMSE}(f)$ and the tool suggested further structures to extend the data set. Here, the *RuNNerActiveLearn* tool was modified to explicitly label atoms exceeding the ACSF space and deviating atomic force components to construct the related molecular fragments, which were added to the data set. Identified bulk structures by deviating atomic energies were rejected, because of the cohesive energy offsets mentioned by Eckhoff *et al.* [49].

3.5 Equation-of-State

For most HDNNP applications low energy structures around the ground state are most significant. Thus, the atomic positions and the lattice constants a of the bulk IRMOF structures are optimized by DFT. In addition, the equilibrium lattice constant can be determined by the volume-energy relation described by an *equation-of-state* (EOS). In this work, the Birch-Murnaghan [106] (BM) EOS is used

$$E^{\text{BM}}(V) = E_0 + \frac{9\,V_0\,B_0}{16}\left\{\left[\left(\frac{V_0}{V}\right)^{\frac{2}{3}} - 1\right]^3 B_0^{'} + \left[\left(\frac{V_0}{V}\right)^{\frac{2}{3}} - 1\right]^2 \left[6 - 4\left(\frac{V_0}{V}\right)^{\frac{2}{3}}\right]\right\}. \tag{3.1}$$

In cubic cases as for the selected IRMOF structures, the lattice constant is determined by the simple relation $V = a^3$. Among the lattice constant, also the bulk modulus and its pressure derivative of the material are determined. Additionally, the required structures of the EOS fit, can be used to validate the HDNNPs as shown in section 4.7.

Chapter 4

Results and Discussion

Within the HDNNP formalism (sec. 2.6) the atomic energies, summing up to the total energy, are strictly dependent on the local environment/structure of the atom. MOFs are built up by two types of building block, which can easily be separated into molecular MOF fragments by chemical intuition, which include inherently information about the local atomic environment and the energy. Since also atomic forces, as atomic observables providing local information about the PES, can be used for the HDNNP training, to increase the information per reference calculation, which reduces simultaneously the total amount of needed reference calculations and thus, the computational effort for the preparation of the HDNNP training set. Nevertheless, the quality of the atomic forces provided by the molecular fragments need to be accurate in comparison to the bulk structure for which the resulting HDNNP should finally be applied to.

A well-defined starting point for the determination of size-converged molecular fragments is of crucial importance. Since, for most applications of interatomic potentials, the equilibrium region is of special interest. In a first step, the IRMOF bulk unit cells are optimized by DFT and finally used for the determination of size-converged molecular fragments.

4.1 Bulk structures

The metal-organic frameworks (MOFs) of interest in this work – MOF-5 also known as *iso-reticular* MOF-1 (IRMOF-1), IRMOF-10 and IRMOF-16 – are illustrated in figure 4.1. These structures are built up by two building blocks. The secondary building unit (SBU) – an oxide-centered Zn_4O^{6+} tetrahedron, which is connected to six carboxylate groups forming an octahedron-shaped SBU. Each of the six carboxylate groups at the tetrahedron edges is connected to a dicarboxylate linker – the other building block, which is benzene-1,4-dicarboxylate (BDC^{2-}) in case of IRMOF-1, biphenyl-4,4′-dicarboxylate ($BPDC^{2-}$) for IRMOF-10 and terphenyl-4,4″-dicarboxylate ($TPDC^{2-}$) for IRMOF-16. Each unit cell contains eight units of $SBU(linker)_3$. The phrase *reticular* describes the netlike structure of the IRMOFs and *iso* the similarity, thus the term *isoreticular* describes the relation of the netlike structures with the same building scheme. The bulk unit cells and the related space group no. 225 $Fm\overline{3}m$ underline this similarity. Although, IRMOF-16 crystallizes in space group no. 221 $Pm\overline{3}m$, since the phenylene rings are repelling each other and prevent the terminating carboxyl groups in a co-planar orientation as the BDC^{2-}-linker of IRMOF-1. This reduces the unit cell symmetry of IRMOF-16 as demonstrated by Eddaoudi *et al.* [53]. Nevertheless, for comparison reasons the structure is represented in the same space group as IRMOF-1 and -10, which plays only a minor role for the construction of the molecular fragments in this work.

The unit cells (fig. 4.1) contain 424, 664 and 904 atoms with 7, 10 and 13 in-equivalent atomic-sites for IRMOF-1, -10 and -16, respectively. To determine the equilibrium lattice parameter a the bulk unit cell is optimized by DFT with respect to the lattice parameter

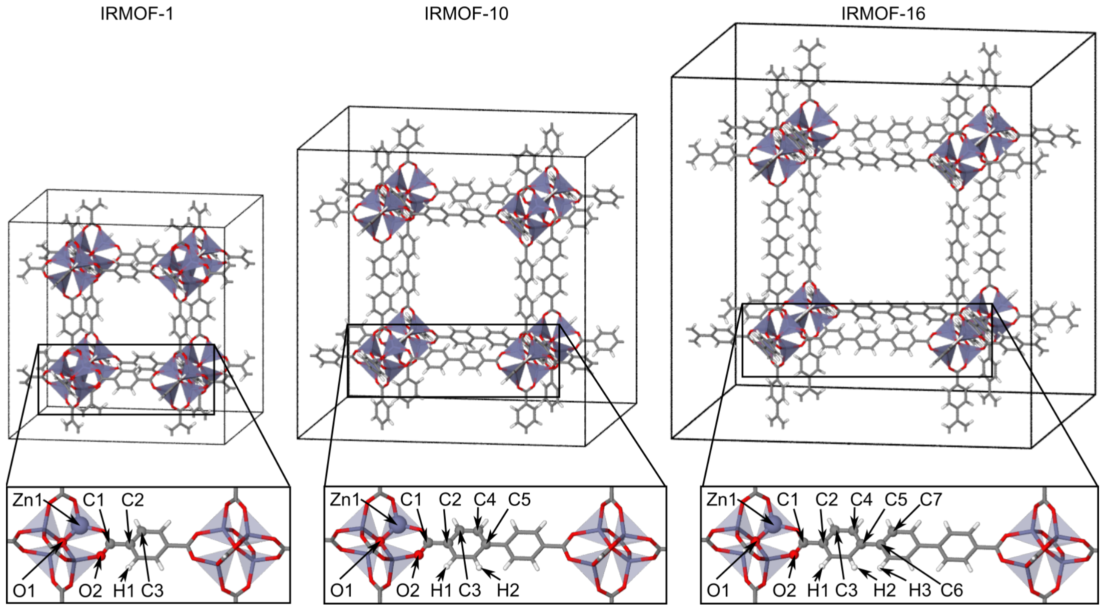

Figure 4.1: Bulk structures of the *isoreticular* (IR) metal-organic frameworks (MOFs) IRMOF-1, IRMOF-10 and IRMOF-16 in the upper panel and the different in-equivalent atomic sites in the lower panel. Zinc atoms are illustrated in violet, oxygen in red, carbon in grey, hydrogen in white in this work.

and the atomic positions. Furthermore, an equation of state (sec. 3.5) can be used to determine the equilibrium lattice parameter by fitting the resulting total energy of the bulk unit cell against the volume V, which depends on the lattice parameter in cubic systems by $V = a^3$. For the IRMOF structures, the scaling of the unit cell, performed by a scaling factor $\sigma \in \{0.95 - 1.10\}$ in steps of 0.01, changes also the atomic bonds within the unit cell. However, the atoms can relax within the scaled unit cell volume due to the large pores of the IRMOF structures. Thus, an additional relaxation of the atomic positions follows the expansion or compression of the unit cell and results in a data set of 16 scaled bulk structures with relaxed atomic positions for each of the three IRMOF bulk structures. The resulting lattice parameters a (tab. 4.1) are slightly overestimated by RPBE around

	lit. [53]	DFT	BM(DFT)
$a_{\mathrm{IRMOF-1}}$	25.8302	26.296	26.289
$a_{\mathrm{IRMOF-10}}$	34.2807	35.063	35.061
$a_{\mathrm{IRMOF-16}}$	42.9806	43.832	43.832
$\Delta a_{\mathrm{IRMOF-1}}$	0.0000	-0.4658	-0.4586
$\Delta a_{\mathrm{IRMOF-10}}$	0.0000	-0.7823	-0.7807
$\Delta a_{\mathrm{IRMOF-16}}$	0.0000	-0.8510	-0.8514
$\Delta' a_{\mathrm{IRMOF-1}}$	0.0000	-0.0180	-0.0178
$\Delta' a_{\mathrm{IRMOF-10}}$	0.0000	-0.0228	-0.0228
$\Delta' a_{\mathrm{IRMOF-16}}$	0.0000	-0.0198	-0.0198

Table 4.1: Compilation of the literature [53], DFT and Birch-Murnaghan (BM) EOS equilibrium lattice parameter a, the absolute deviation $\Delta a = a^{\mathrm{lit.}} - a$ compared to the literature lattice parameters in Å and the relative deviation $\Delta' a = 1 - \frac{a}{a^{\mathrm{lit.}}}$ calculated by DFT and the BM EOS fit.

2 %.

4.2 Fragment Construction by Convergence of Force Differences

The dependence of the local environment is strongly dependent on the underlying electronic structure. Thus, for each in-equivalent atomic site A, a molecular fragment is constructed based on the optimized bulk structures. The central atom A is embedded in a bulk-like environment up to the fragment radius r_{frag}, determining the size of the molecular fragment (sec. 3.2). Furthermore, the atomic force $\mathbf{f}_A^{\text{frag}}$ of atom A depends on the fragment radius. If the fragment radius converges to infinity, the fragment will converge to the periodic bulk structure and the molecular atomic force will be equivalent to the bulk force $\mathbf{f}_A$

$$\lim_{r_{\text{frag}}\to\infty} \mathbf{f}_A^{\text{frag}} = \mathbf{f}_A \quad . \tag{4.1}$$

Hence, the fragment radius, at which the atomic force will be converged to the atomic bulk force within a pre-defined tolerance $\Delta\mathbf{f}_A^{\max}$, is of crucial importance. Thus, the atomic force and the force difference $\Delta\mathbf{f}_A$ are functions of the fragment radius, which need to be analyzed to find the fragment radius vanishing the atomic force difference

$$||\Delta\mathbf{f}_A|| = ||\mathbf{f}_A - \mathbf{f}_A^{\text{frag}}|| \overset{!}{\leq} ||\Delta\mathbf{f}_A^{\max}|| \quad , \tag{4.2}$$

$$\lim_{r_{\text{frag}}\to\infty} ||\Delta\mathbf{f}_A|| = 0 \quad . \tag{4.3}$$

An obvious disadvantage of this ansatz is highlighted by atoms in a symmetric environment as O1 (fig. 4.1). The atomic force $\mathbf{f}_{\text{O1}}$ is independent on the fragment radius, although the force difference is below the pre-defined tolerance, a meaningful fragment radius is undefined (fig. 4.2). Furthermore, no detailed information about the central atom force dependence on the neighboring atoms is gained. In cases of central atoms effected by a few atoms in large distance via electrostatic interactions, large fragment radii, enforcing large molecular fragments, will be derived to include these interacting atoms. Simultaneously,

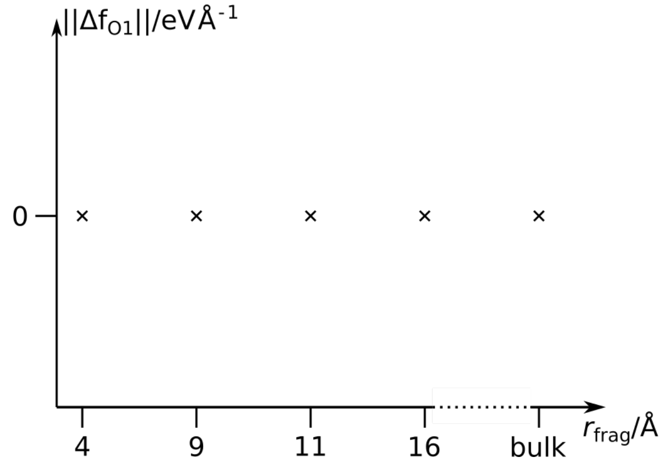

Figure 4.2: Compilation of the force error norm $||\Delta\mathbf{f}_{\text{O1}}|| = ||\mathbf{f}_{\text{O1}} - \mathbf{f}_{\text{O1}}^{\text{frag}}||$ as the norm of the difference between the atomic bulk force $\mathbf{f}_{\text{O1}}$ and the related atomic fragment force $\mathbf{f}_{\text{O1}}^{\text{frag}}$ of the in-equivalent atomic site O1 (fig. 4.1) for different fragments of IRMOF-1.

the large fragments lead to computationally demanding reference calculations needed for the HDNNP training.

An alternative locality test is presented for amorphous carbon [107]. In this approach fluctuations of atomic forces are analyzed for the central atom A embedded in a frozen environment up to a certain radius. The atomic positions outside this radius are varied to investigate the effect on the central atom. The frozen environment is increased until the changes of the atomic force of the central atom decrease below a predefined threshold value. Although, this method is quite universal and it permits the analysis of symmetrically embedded atoms, it relies on a large number of electronic structure calculations, which need to sample a set of representative atomic arrangements. Furthemore, this approach prevents a detailed analysis of the specific neighboring atoms.

For this reasons, a well-defined method is of interest to analyze atoms in symmetric environments and specific neighboring atoms. A systematic way for analyzing the dependence of the atomic force f_A of atom A on the atomic coordinate B_β of the neighboring atom B is illustrated by the derivative of the atomic forces with respect to the atomic positions. This is equivalent to the second derivative of the energy with respect to the atomic positions, which is commonly denoted as the Hessian.

4.3 The Hessian-Based Assessment of Atomic Forces

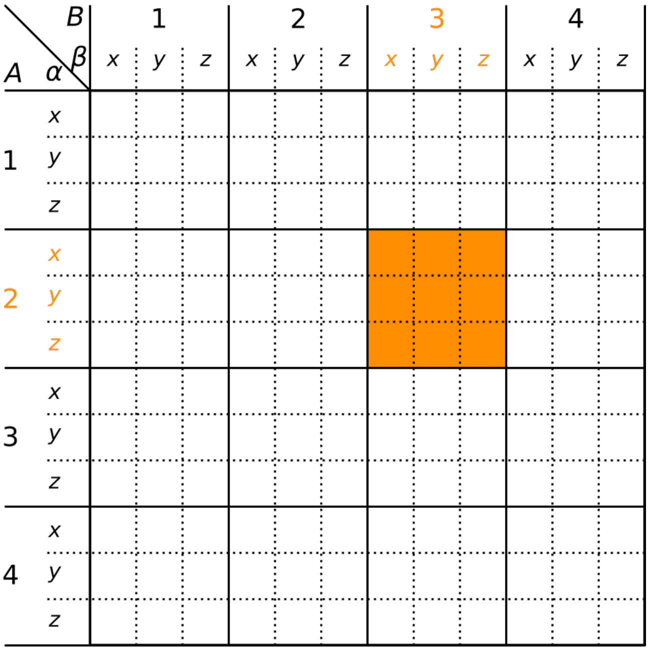

Figure 4.3: Structure of the Hessian matrix $\mathbf{H}$ for a system containing $M = 4$ atoms. The interaction between atom $A = 2$ and atom $B = 3$, defined by equation 4.4 is represented by the atomic Hessian submatrix h_{23} highlighted in orange. Adapted from [108] with permission from ©2022 AIP Publishing.

To estimate the dependence of the force vector $\mathbf{f}_A$ of the reference atom A on the atomic coordinates B_β of the neighboring atom B, the $3M \times 3M$-dimensional Hessian matrix (fig. 4.3) is calculated [1] and contains the elements

$$H_{A_\alpha B_\beta} = \frac{\partial^2 E}{\partial A_\alpha \partial B_\beta} = -\frac{\partial f_{B_\beta}}{\partial A_\alpha} = -\frac{\partial f_{A_\alpha}}{\partial B_\beta} \quad , \tag{4.4}$$

which describe the dependence of each force component f_{A_α} on the Cartesian coordinate B_β of the neighboring atom B, with $\alpha, \beta \in \{x, y, z\}$. Each atomic interaction is depicted by a 3×3 submatrix $\mathbf{h}_{AB}$ and quantified by the norm of the atomic Hessian submatrix

$$||\mathbf{h}_{AB}|| = \sqrt{\sum_{\alpha=x,y,z} \sum_{\beta=x,y,z} h^2_{A_\alpha B_\beta}} \quad . \tag{4.5}$$

The scalar atomic Hessian submatrix norm is assumed to decrease with increasing atomic distance d_{AB}, between the interacting atoms A and B, and vice versa.

4.3.1 Hessian Group Matrix

Starting from the bulk or alternatively from an arbitrary large molecular fragment as the reference structure, the crucial point is, how much of the outermost, nearly spherical environment can be removed, leading to the smaller fragment, without introducing significant

[1] The results discussed and presented in this section 4.3, the following section 4.4 and to a significant part the results of section 4.5 were obtained in my recent publication [108] and are shown for completeness with permission from ©2022 AIP Publishing.

errors for force components on the central atom. Thus, atoms in all spatial directions need to be removed for a decreasing spherical environment and therefore the cumulative effect of the removed atoms on the central atom is of interest. This defines the Hessian group matrix

$$\mathbf{G}_A^g = \sum_{B \in g} \mathbf{h}_{AB} \quad , \tag{4.6}$$

as the sum of atoms present in group g, which re removed from the reference fragment. Hence, the Hessian group matrix $\mathbf{G}_A^g$ is the sum of all atomic interactions $\mathbf{h}_{AB}$ of the central atom A and the neighboring atoms B of the reference system beyond a specific fragment radius r_{frag} and thus, defines all missing interactions in the smaller fragment, determined by r_{frag}, compared to the reference fragment. In analogy to equation 4.5, the Hessian group matrix is quantified by its norm. While summing up different atomic Hessian submatrices $\mathbf{h}_{AB}$, specific contributions of neighboring atoms B can cancel each other, in general. The more atoms are removed, the more atoms are include in the groups and the Hessian group matrix norm increases. Additionally, the fragment size-decreases and the truncated atomic interactions increase. This results in a system-related dependence of the Hessian group matrix on the atomic environment. Diverse bonding situations, depend differently on the specific environment.

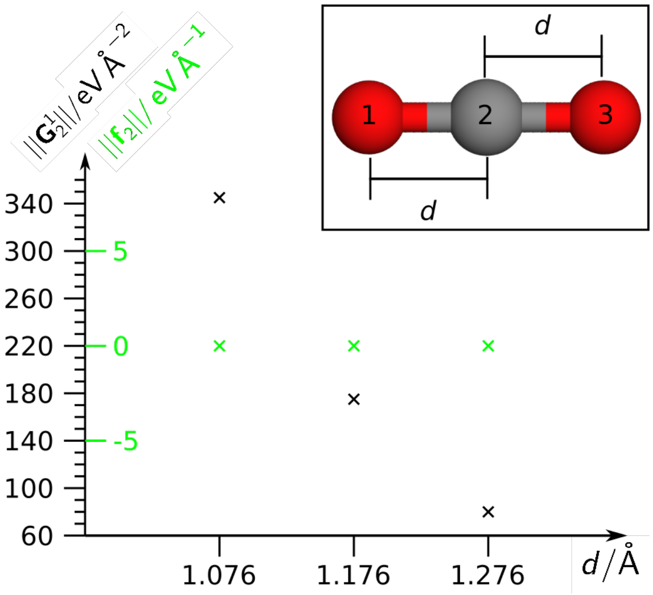

Figure 4.4: Simple carbon dioxide showcase for the Hessian group matrix. The force (green) and the Hessian group matrix norm (black) are shown for the central carbon. The DFT equilibrium bond length of $d = 1.176$ Å is expanded and shrinked symmetrically by $\Delta d \pm 0.1$ Å. Related to the symmetric environment of the central carbon atom, the force value is vanishing for all the structures. The Hessian group matrix norm instead describes the different atomic interactions. Adapted from [108] with permission from ©2022 AIP Publishing.

The concept of the *Hessian group matrix* is explained by a simple showcase (fig. 4.4). The atomic Hessian submatrices cover the individual atom-atom interactions between the three atoms, while the summarized interactions of the central carbon $A = 2$ is covered by the Hessian group matrix

$$\mathbf{G}_2^1 = \sum_{B \in g} \mathbf{h}_{2B} = \mathbf{h}_{21} + \mathbf{h}_{23} \quad , \tag{4.7}$$

the group $g = 1$ includes the two oxygen atoms $B = 1, 3$. As already stated in section 4.2, atoms in symmetric environments like the central carbon, experiences symmetric force contributions, which cancel each other independent on the considered atomic environment, in this case independent from the distance d. Thus, the force difference to a symmetric reference structure will not change and the range of interaction is hard to determine using these force values and has to be checked carefully. The Hessian group matrix norm instead, changes for different symmetric atomic arrangements of the two oxygen atoms and decreases for $d = 1.276$ Å and increases for $d = 1.076$ Å, providing a useful quantity to determine interaction ranges also for symmetric environments. However, the carbon dioxide molecule is a very simple showcase, since all atoms but the central carbon are included in the group $g = 1$ and no bonds are broken. For reasons described in section 3.2, broken bonds or dangling bonds, occurring in fragment construction, are saturated by hydrogen atoms in this work. Thus, in addition to the atoms included in group g, also these added and artificial interactions $\mathbf{h}_{Ab}$ of the central atom A and the saturation hydrogen atoms b needs to be considered in the cumulative atomic interactions by subtracting the contributions from the Hessian group matrix, leading to the effective Hessian group matrix

$$\mathbf{G}'^{g}_{A} = \sum_{B \in \mathrm{g}} \mathbf{h}_{AB} - \sum_{b} \mathbf{h}_{Ab} \quad . \tag{4.8}$$

Another option is to neglect these interactions $\mathbf{h}_{Ab}$ and examine the interaction range by applying $\mathbf{G}^{g}_{A}$ directly. Nevertheless, $\mathbf{h}_{Ab}$ is usually very small for meaningful-sized fragments, leading to a very small difference of $\mathbf{G}^{g}_{A}$ and $\mathbf{G}'^{g}_{A}$ in practice.

4.4 Model Systems

The electronic structure determines the range of interaction, and this will be different for diverse bonding situations. For a first investigation different model systems, differing in their bonding situation and in their electronic structure, are chosen as idealized structures. Real and more complex systems, like IRMOF structures, can be viewed as combinations of these idealized model systems. However, the effect of the electronic structure needs to be figured out, achieved by the analysis of different covalent bonding situations in the model system.

4.4.1 Structures

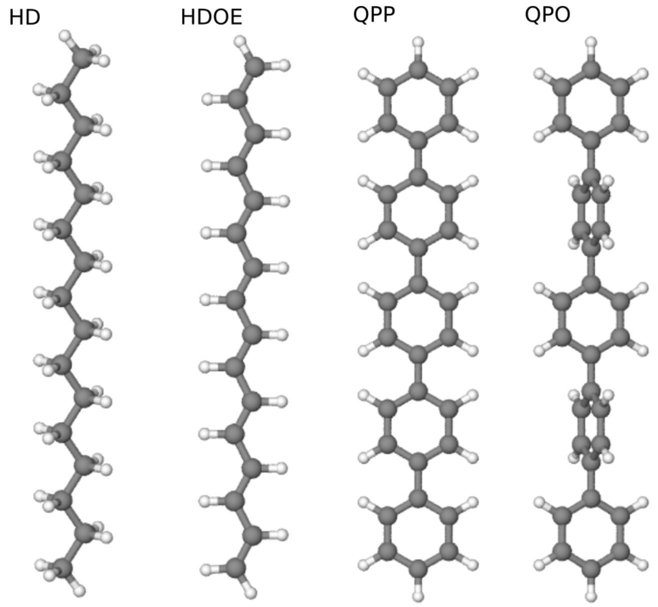

Figure 4.5: Chosen model systems for the Hessian-based assessment being exemplary for bonding situations in the IRMOF structures. Hexadecane (HD) represents typical covalent single bonds, (3E,5E,7E,9E,11E,13E)-hexadeca-1,3,5,7,9,11,13,15-octaene (HDOE) a conjugated π-electron system extending over the whole molecule, while the 1,1′:4′,1″:4″,1‴:4‴,1⁗:4⁗,1′′′′′-quinquephenyl all in-plane conformer (QPP), which provides also an extended π-electron system over the whole structure and a maximum amount of resonance stabilization together with the out-of-plane conformer (QPO) decoupling the π-system of the specific phenylene rings, represent aromatic systems. Adapted from [108] with permission from ©2022 AIP Publishing.

Quasi-1D hydro carbon structures (fig. 4.5) are used to clarify the distance dependence of the central atomic force. The model systems include typical covalent single bonds covered by the hexadecane (HD), a conjugated π-electron system extending the whole model structure as given in (3E,5E,7E,9E,11E,13E)-hexadeca-1,3,5,7,9,11,13,15-octaene (HDOE) and two conformers of 1,1′:4′,1″:4″,1‴:4‴,1⁗:4⁗,1′′′′′-quinquephenyl. An all in-plane conformer (QPP) also providing an extended π-electron system over the whole structure and a maximum amount of resonance stabilization. Contrary, the out-of-plane conformer (QPO) is based on pairwise-orthogonal neighboring phenylene-ring, without inter phenylene ring π-electron resonance, due to the missing inter phenylene overlap of the p-orbitals.

For HD and HDOE, the molecular structures are fully minimized in energy, whereas the QPP and QPO are based on a fully minimized benzene molecule, which is replicated and inter connected in *para*-position with a carbon-carbon distance of 1.45 Å.

4.4.2 Atomic Hessian Matrix Norm

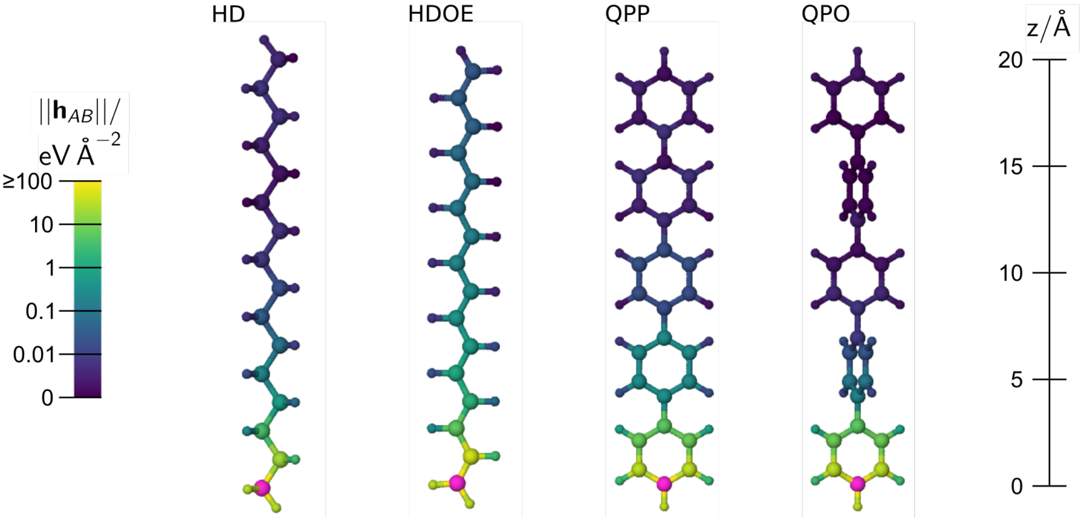

Figure 4.6: Atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$, describing the interaction between the carbon reference atom (magenta) and all neighboring atoms B for the different model systems HD, HDOE, QPP and QPO. Adapted from [108] with permission from ©2022 AIP Publishing.

The Hessian matrix is calculated by DFT for the different model systems and the atomic interactions are analyzed. Figure 4.6 states a decrease of the atomic Hessian submatrix norm $||\mathbf{h}_{AB}||$, describing the interaction between the terminal carbon reference atom A (highlighted in magenta) and the neighboring atoms B, with an increasing spatial distance d_{AB} between these atoms. These results are in line with the expectation of weak atomic interactions at large distances and vice versa, strong interactions at small distances. Thus, with increasing atomic distance of the neighboring atom B and the reference atom A, the less significant is atom B effecting the reference atom force $\mathbf{f}_A$ (fig. A.1 and 4.7). Of course, the position of the reference atom A illustrates the maximum influence on the force $\mathbf{f}_A$ as expected and stated by the submatrix $\mathbf{h}_{AA}$. Figure 4.7 emphasizes also the differences related to the range of interaction and the decay with increasing distance. Although, the model systems qualitatively behave similarly, differences in the quantitative behaviour of the model systems highlight the electronic structure differences. Especially, the carbon-carbon interactions decays much slower in HDOE and QPP, than in comparison to HD and QPO. Thus, the interactions in HDOE QPO demonstrate a long-ranged character in comparison HD and QPO, respectively. This behaviour is based on the conjugated π-system of HDOE and QPP, causing the increased interactions. The comparison of QPP and QPO confirms the effect of the π-system. The molecular structure differs in second phenylene ring, around $d_{AB} \approx 4$ Å. Additionally, this change is also obvious in the carbon interactions (fig. 4.7, green and black line with marking circles), being very similar for distances $d_{AB} < 4$ Å and start to deviate for larger distances $d_{AB} \geq 4$ Å. A special case occurs in the region of the second phenylene ring $d_{AB} \approx 4 - 7$ Å, interactions within the QPO start to vanish, in contrast to QPP demonstrating nearly constant interaction with the reference atom. Thus, the increased electron delocalization in QPP, due to the in-plane orientation of the phenylene rings, increases drastically the effect on the force $\mathbf{f}_A$ for the carbon atoms localized in the second phenylene ring. In QPO, these interactions are truncated, because of the orthogonal orientations of neighboring phenylene rings.

The hydrogen interactions of the reference atom at the given distances are smaller than the carbon interactions discussed above. Because of the additional bond transmitting information of structural changes, related to the specific neighboring hydrogen atom, towards the reference atom, compared to the neighboring carbon atom at a similar distance. But the major part is related to the π-electron system, to which hydrogen does not contribute, due to its missing π-electrons. For this reason, the interactions of hydrogen and the reference atom are reduced. As discussed above, the π-system dictates the degree of long-range character. Hence, hydrogen atom interactions are mostly dependent on the distance to the reference atom as shown by very similar reference carbon-hydrogen interactions in HD and HDOE and a nearly indistinguishable interactions in QPP and QPO, also in the pronounced region $d_{AB} \approx 4 - 7\,\text{Å}$ of the carbon interactions.

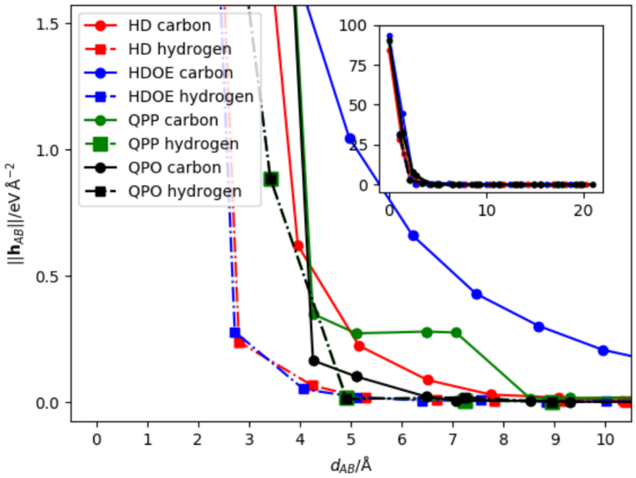

Figure 4.7: The atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ of the four model systems HD, HDOE, QPP and QPO as a function of the distance d_{AB} between the reference carbon atom A as defined in figure 4.6 and all neighboring atoms B. Separated curves are given for the interactions of atom A with neighboring carbon and hydrogen atoms. The inset shows the data for the interaction of A with all atoms in the entire molecules. A separated plot for each model system is shown in figure A.2. Adapted from [108] with permission from ©2022 AIP Publishing.

4.4.3 Hessian Group Matrix Norm

The results show decreasing atomic interactions and thus decreasing dependence of the force $\mathbf{f}_A$ on the neighboring atoms B with increasing atomic separation. Thus, in principle weakly interacting atoms can be determined by choosing a threshold value for the atomic Hessian submatrix norm $||\mathbf{h}_{AB}||$. Furthermore, all atoms showing $||\mathbf{h}_{AB}||$ below this threshold can be eliminated from the reference structure, assumed without significant changes for the atomic force $\mathbf{f}_A$ provided by the resulting smaller fragment. However, this procedure would lead to arbitrary bond cutting, may resulting in large changes of the elec-

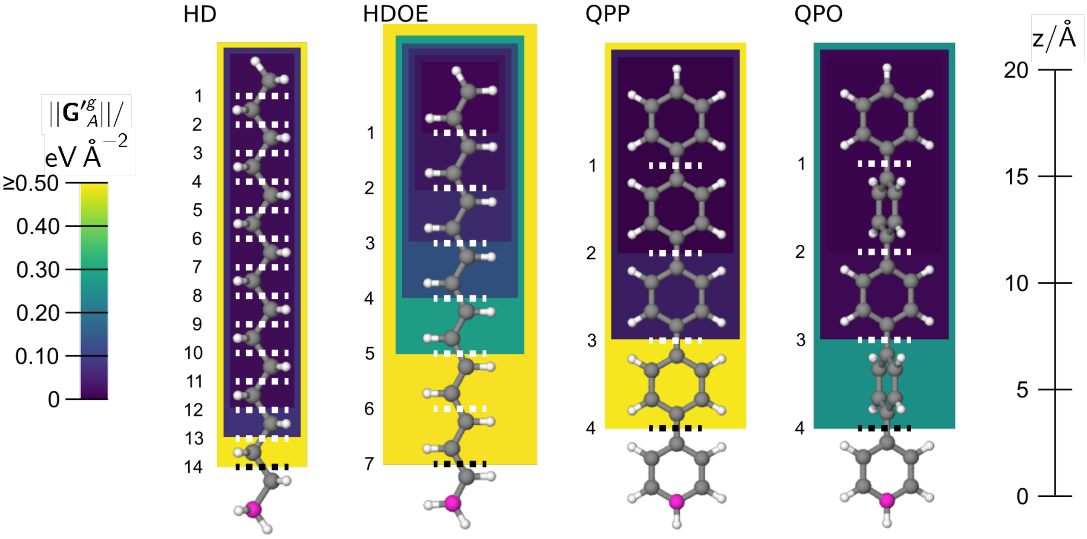

Figure 4.8: Effective Hessian group matrix norm $||\mathbf{G}'^{g}_{A}||$ for all atomic groups (represented by the colored rectangles) with respect to the reference carbon atoms A shown in magenta for the model systems HD, HDOE, QPP and QPO. The bonds, which are cut to form increasing groups of removed atoms, are shown as white dashed lines along with the numbering of the resulting groups from the top to the bottom. The black dashed lines indicate the bond to be cut for the smallest considered fragment corresponding to the largest atomic group of removed atoms. Note that each group is included in the next larger group when more atoms are removed from the system. Adapted from [108] with permission from ©2022 AIP Publishing.

tronic structure, if π-bonds get broken, for example. This will affect the atomic interactions and the energy , as well as the atomic forces. Therefore, a definition of a threshold based on the atomic Hessian submatrix cannot be applied as a criterion for indicating weakly and thus non-significant atomic interactions.

Using the concept of functional groups (sec. 3.2 and the effective Hessian group matrix norm, significant and non-significant interactions can be distinguished. Moreover, the changes of the electronic structure, comparing to the reference system, are minimized for the smaller fragment. Furthermore, this results in discrete smaller fragments by removing the functional groups step-wise from the reference structure. For the model systems, the chemical groups/entities, which are allow to be separated from the reference structure, are defined as follows:

- HD: terminal CH_3- and each CH_2-entity,
- HDOE: terminal C_2H_3- and each C_2H_2-entity to sustain the extended π-electron system,
- QPP and QPO: each phenylene for the aromatic character of the phenylene rings.

In addition, fragments, which are created by breaking bonds to the reference atom, are not considered. Consequently, there are 15, 8, 5 and 5 fragments, whose construction is based on the for model systems HD, HDOE, QPP and QPO. Following this procedure, the entities are cut one after the other from the reference system, while the removed atoms are then included in 14, 7, 4 and 4 groups g of removed atoms. Thus, the fragment size decreases, while the number of atoms in the groups g increases, since all groups form a sub group of the next following larger group (fig. 4.8). The colors of the rectangular boxes is defined by the effective Hessian group matrix norm (eq. 4.8 and 4.5). Similar to the results of the atomic interactions above, the effective Hessian group matrix norm decreases with increasing distance of the reference atom A and the group g. Again, there is a similar qualitative, but different quantitative behaviour among the model systems. For the constructed fragments, the force $\mathbf{f}_A^g$ can be compared to the force value $\mathbf{f}_A$ of the reference system, resulting in the force error

$$\Delta \mathbf{f}_A^g = \mathbf{f}_A - \mathbf{f}_A^g \quad , \tag{4.9}$$

of the fragments. The broken bond is saturated along the broken carbon-carbon bond with a carbon-hydrogen distance of 1.05 Å.Figure A.3 illustrates a similar decay of the force error norm and the effective Hessian group matrix norm for the four model systems, giving explicit evidence of a correlation for these quantities.

The effect of the saturation hydrogen on the Hessian group matrix norm is investigated exemplary for the model system HD (fig. A.4). For distances of the saturating hydrogen atoms to the reference atom larger than ~ 5 Å, its contribution is rather small and the effect of the hydrogen saturation on the machine learning data set, typically constructed with an environment radius of $5 - 6$ Å can be neglected.

4.4.4 Force-Convergence Threshold

As described in section 4.4.1 above, the investigated structures of the model systems are near-equilibrium structures. A more diverse data set is obtained by rescaling the model system structures. The scaling factor $\sigma \in \{0.90, 0.95, 1.00, 1.05, 1.10, 1.15, 1.20\}$ expands and contracts the whole molecular structures and results in non-equilibrium atomic arrangements. Furthermore, the Hessian is calculated for these non-equilibrium structures, followed by the analysis in the same manner as for the near-equilibrium structure ($\sigma = 1.00$). The correlation of the Hessian group matrix norm $||\mathbf{h}_{AB}||$ and the force error norm $||\Delta \mathbf{f}_A^g||$ is shown in figure 4.9 for all model system structures generated by employing σ.

Approximately, the two quantities $||\Delta \mathbf{f}_A^g||$ and $||\mathbf{G}_A'^g||$ correlate linearly, even for the highly compressed structures ($\sigma = 0.90$, blue curves in fig. 4.9). Large differences of this approximated linear behaviour is shown by the smallest investigated HD fragments, which can be understood, since the construction of these fragments include already structural changes near the reference atom – the second nearest carbon is removed. Nevertheless, the force errors in these specific atomic arrangements are unexpectedly small, indicating only weak interactions of the reference atom in the expanded ($\sigma \in \{1.10, 1.15, 1.20\}$) HD structures. All model systems show the largest force errors for their highly compressed structures ($\sigma = 0.90$). These strong interactions of the reference atom within the smallest fragments are perfectly in line with the strong atomic repulsion for short atomic distances as present in these compressed structures.

Based on the approximately linear correlation and the definition of the required force accuracy criterion $||\mathbf{f}^{\max}||$, which needs to be fulfilled by the fragments, a threshold value Γ of the Hessian group matrix norm can be determined. Following, all atomic interactions accumulating to a smaller effective Hessian group matrix norm as the derived threshold Γ, can be removed to result in the minimum, size-converged fragments, which provide the desired force accuracy in relation to the reference system. All atoms contributing to an

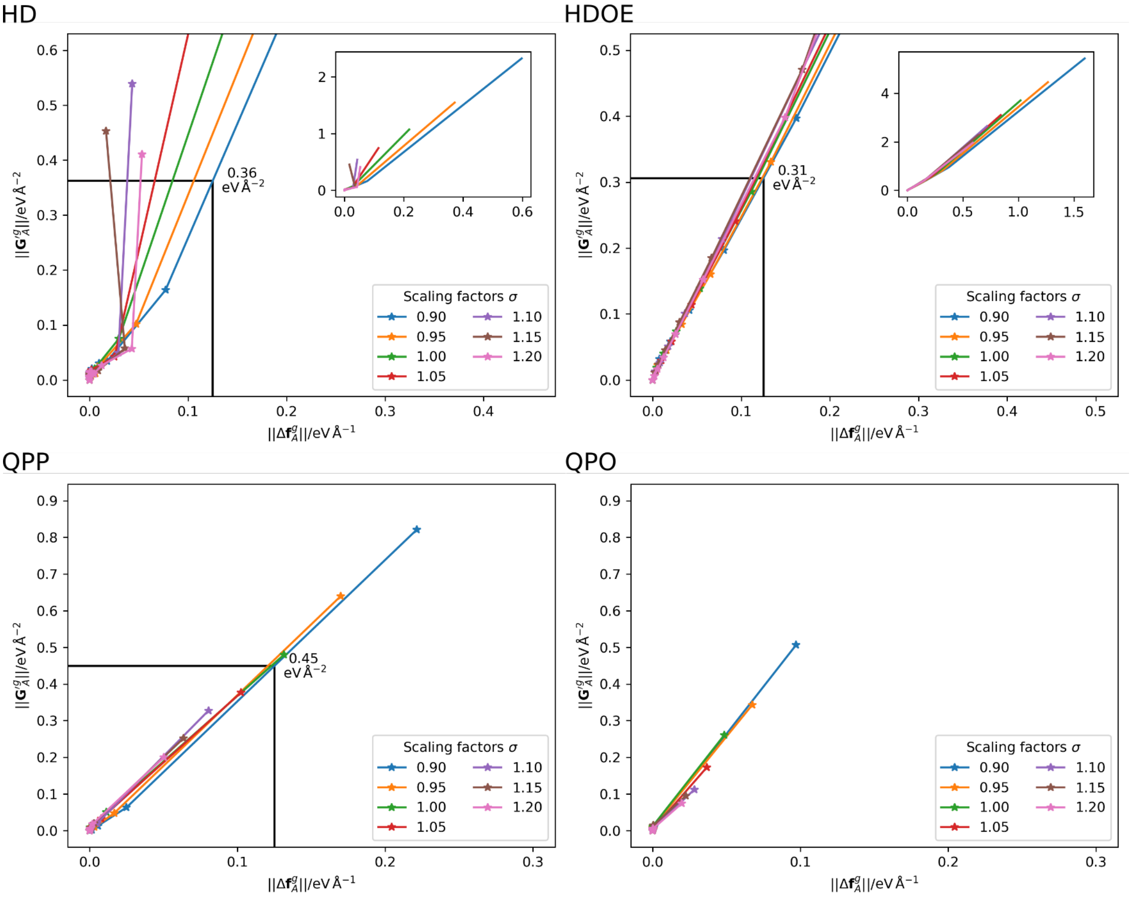

Figure 4.9: The effective Hessian group matrix norm $||\mathbf{G}'^g_A||$ in relation to the norm of the force error $||\Delta\mathbf{f}^g_A||$ of the reference carbon atom A for the model systems HD, HDOE, QPP and QPO. For HD and HDOE, the inset shows the complete data range and the main plot focuses on the region near the origin. These plots summarize the results for all employed scaling factors $\sigma = 0.90 - 1.20$. The norm of the force error $||\Delta\mathbf{f}^g_A||$ and the effective Hessian group matrix norm $||\mathbf{G}'^g_A||$ decrease towards the origin, describing the increase of the molecular fragment up to the reference system at the origin. The threshold values Γ of the effective Hessian group matrix norm $||\mathbf{G}'^g_A||$ are represented by the black lines in the panels of HD, HDOE and QPP, ensuring size-converged fragments with a force accuracy criterion of $||\mathbf{f}^{\max}|| = 0.125\,\text{eV}\,\text{Å}^{-1}$. In case of QPO, already the smallest fragment is stated as size-converged. Adapted from [108] with permission from ©2022 AIP Publishing.

effective Hessian group matrix norm larger in value than the threshold Γ, interact significantly with the reference atom and must not be removed to keep the desired force accuracy. Up to now, it is not yet clear, if this threshold Γ is strongly dependent on the underlying bonding situation. A distinct tendency for such a dependence prevents the application of the threshold Γ to different types of bonding. Consequently, for each type of bonding a specific threshold needs to be derived. For a threshold Γ being nearly independent on the bonding type, size-converged fragments can be determined for a wide range of many different bonding situations, because of the universal character. A desired force accuracy for HDNNP training fragments is chosen to be $||\Delta\mathbf{f}^{\max}|| \leq 0.125\,\text{eV}\,\text{Å}$, which is typically achieved during HDNNP training procedure as stated by the RMSE of the force components. The related threshold Γ of the effective Hessian group matrix norm, providing size-converged fragments with the desired force accuracy, is given by the correlation of $||\Delta\mathbf{f}_A^g||$ and $||\mathbf{G}'^g_A||$ (fig. 4.9). Here, the most compressed structures ($\sigma = 0.90$) are decisive for the determination of the threshold Γ, since the largest force errors occur for the most compressed structures. For QPO, even the smallest fragment (benzene) is already size-converged within the chosen force accuracy, due to the very weak interactions of the reference atom and the atoms of the second phenylene ring, as discussed above (sec. 4.4.2 and 4.4.3). The resulting threshold values Γ for the effective Hessian group matrix norm of the different model systems and thus for different bonding situations, are quite similar with only slight absolute differences (HD: 0.36 eV Å^{-2}, HDOE: 0.31 eV Å^{-2} and QPP: 0.45 eV Å^{-2}). Hence, to a good approximation, a general, universal threshold can be used to define size-converged fragments for the different model systems. Below the tightest convergence criterion of HDOE (0.31 eV Å^{-2}), the reference atom forces for all model systems are converged in the resulting fragment. While adding a safety margin to the threshold Γ of ±0.02 eV Å^{-2} (sec. A.2), the threshold value for the effective Hessian group matrix norm is chosen to be $\Gamma = 0.29$ eV Å^{-2}, which can be adjusted for stricter convergence of the forces, if needed. For the HDOE model system this leads to the size-converged fragment HDOE$_5$ ((1E,3E,5E)-hexatriene) (tab. 4.2) with an atomic force error of the reference atom A below the desired convergence criterion $||\Delta\mathbf{f}^{\max}|| = 0.1111$ eV Å^{-1} ≤ 0.125 eV Å^{-1} with an approximated fragment radius 6.2 Å. Furthermore, the resulting size-converged fragments of the model systems determined by the threshold $\Gamma = 0.29$ eV Å^{-2} are HD$_{13}$ (propane), QPP$_3$ (biphenyl) and QPO$_4$ (benzene), with the approximated fragment radii 2.6, 7.1 and 2.8 Å, respectively. Similar to the enlarged long-range character of HDOE and QPP, the increased environment dependence results in larger fragments compared to HD and QPO.

Table 4.2: Compilation of the force component errors $\Delta f^{\mathrm{HDOE}_g}_{A_{x,y,z}}$ and the total force errors $||\Delta\mathbf{f}^{\mathrm{HDOE}_g}_A||$ in eV Å^{-1} for the reference carbon atom in the model system HDOE (fig. 4.8 and 4.9, $\sigma = 1.00$). Further, the effective Hessian group matrix norm $||\mathbf{G}'^g_A||$ is given in eV Å^{-2}. Numbers outside the intended convergence criterion are given in bold.Adapted from [108] with permission from ©2022 AIP Publishing.

g	$\Delta f^{\mathrm{HDOE}}_{A_\mathrm{x}}$	$\Delta f^{\mathrm{HDOE}_g}_{A_\mathrm{y}}$	$\Delta f^{\mathrm{HDOE}_g}_{A_\mathrm{z}}$	$\lvert\lvert\Delta\mathbf{f}^{\mathrm{HDOE}_g}_A\rvert\rvert$	$\lvert\lvert\mathbf{G}'^g_A\rvert\rvert$
ref	0.0000	0.0000	0.0000	0.0000	0.00
1	0.0000	0.0014	-0.0043	0.0045	0.02
2	0.0000	0.0042	-0.0116	0.0124	0.04
3	0.0000	0.0094	-0.0247	0.0264	0.07
4	0.0000	0.0195	-0.0495	0.0532	0.14
5	0.0000	0.0414	-0.1031	0.1111	0.29
6	0.0000	0.0986	**−0.2474**	**0.2663**	**0.71**
7	0.0000	**0.1877**	**−1.0018**	**1.0192**	**3.71**

4.5 IRMOF-Structures

To transferring the knowledge from the idealized and simple model systems to a more complex system like the IRMOF structures, a simplified, one-dimensional IRMOF-1 system is constructed and analyzed. The overall goal is again the construction of size-converged fragments with an accurate description of the reference atom forces within the size-converged fragments based on the threshold γ.

4.5.1 1D-IRMOF-1: Proof of Principle

The one-dimensional IRMOF-1 structure (1D) is constructed from an IRMOF-1 supercell as shown in figure 4.10. It contains six SBU entities, which are connected by five linker molecules including hydrogen saturation for the broken bonds. The Hessian analysis is performed for different reference atoms of the 1D reference system. The defined smaller fragments $1D_g$ are labeled as $1D_1 - 1D_9$ for the reference atoms occurring in the first SBU (fig. 4.10b, fragment $1D_9$). Reference atoms localized in the first linking phenylene ring (fig. 4.10b, fragment $1D_{9'}$) depend on two adapted low-radius fragments, which are redefined as $1D'_8$ and $1D'_9$. With increasing group index g, the number of atoms included in the group increases, whereas the fragment size decreases. Occurring broken bonds within the fragments are also saturated with hydrogen atoms. The fragments $1D_g$ follow the same construction rules as given in section 3.2.

The C1-like atomic sites are analyzed in more detail and the remaining atomic positions are summarized in the appendix and will be referenced, respectively. Because of the reduced symmetry of 1D, three C1-like positions occur in the 1D model system – $C1'$, $C1''$ and $C1'''$. Similar to the results above, the Hessian atomic submatrix norm $||\mathbf{h}_{AB}||$ decreases with in creasing atomic distance d_{AB} between the reference atom A (magenta) and the neighboring atom B (fig. 4.11). Equivalent results are shown for the remaining atoms in appendix (fig. A.5 and A.6). Although, a more detailed analysis offers an increase of the Hessian submatrix norm with increasing distance. This feature occurs in figure A.5 j): the phenylene carbon atoms ($z \approx 10$ Å) show rather weak interactions (blue color) to the reference atom, in comparison to the more distant atoms of the following linking carboxyl group ($z \approx 14$ Å), which demonstrate increased interactions with the reference atom, despite the larger distance. This underlines diverse effects of different in-equivalent atomic sites on the reference atom, which did not emerge within the model systems above. In fact, the neighboring carbon atoms of the model systems illustrate very similar environments, which smears their in-equivalence. This results in similar effects of all neighboring carbon atoms on the reference carbon atom and thus, the interactions are only dependent on the distance to the reference carbon atom (fig. 4.6 and 4.7).

In general, atoms in the spatial proximity of the reference atom – in the same SBU or the same linker – show significant interactions (yellow) in all cases. Perfectly in line with the QPP model system, the π-electron system of the phenylene linker mediates the atomic interactions over large distances. Contrary to the mediating effect of the linker π-electron system, the more ionic character of the SBU locks the range of the atomic interactions, especially the zinc atoms. For the $C1'''$ position, the atomic interactions beyond the zinc atoms of the next SBU ($z \approx 16$ Å) vanishes, which is also observed in figure A.5 and A.6, panels j), k) and l). Investigating the Hessian group matrix norm, similar results occur: significant long-ranged interactions occur for atoms contributing to the π-electron system, presented by $C1'''$ and figure A.7 and A.8, panel n). Thus, $C1'''$ is a special case compared to the other atomic sites, because of the most long-ranged interactions, due to its position

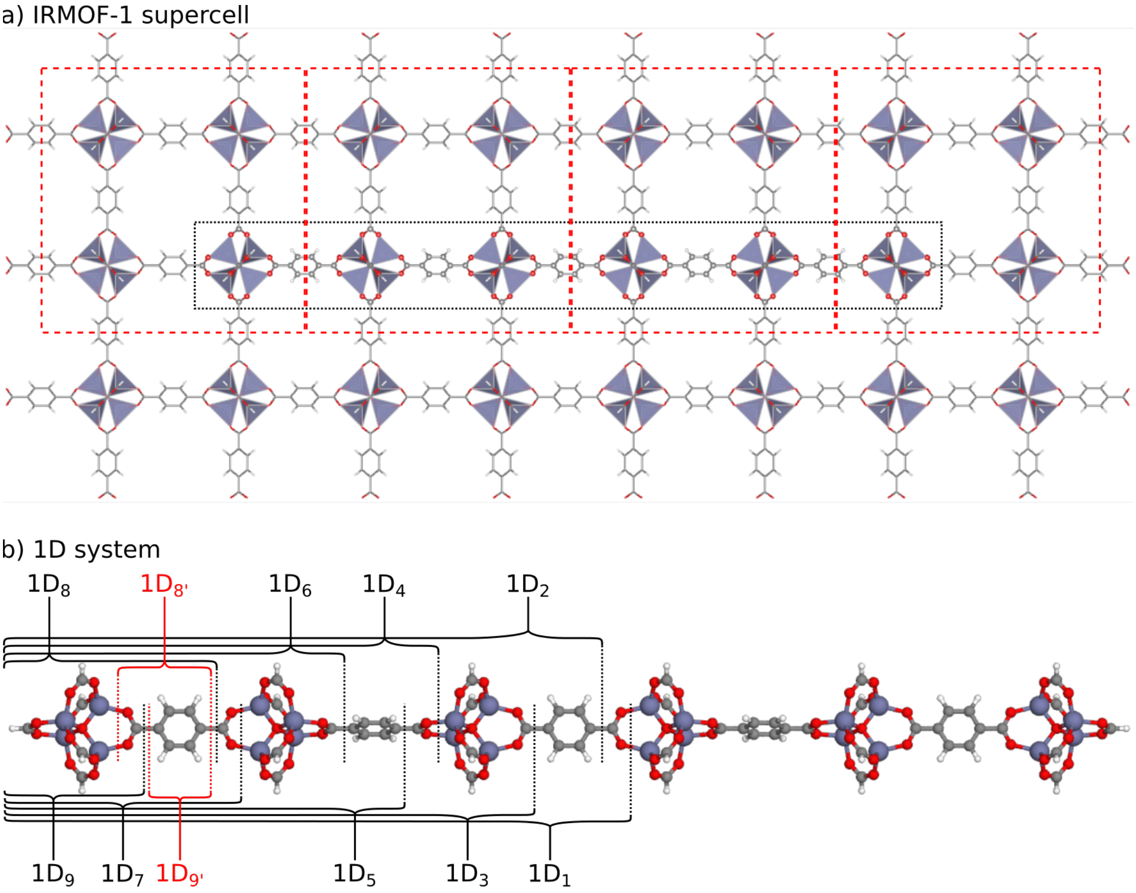

Figure 4.10: a) The IRMOF-1 supercell used to construct the 1D system with marked IRMOF-1 unit cells (red dashed line) and the 1D system embedded in the periodic environment (black dotted line). b) The 1D structure without the periodic environment, but instead with the hydrogen saturated broken bonds and the diverse fragments of different size (black dashed lines) used in the Hessian analysis labeled as $1D_{1-9}$. For the reference atoms occurring in the first phenylene ring, the smallest two molecular fragments change for those positions and are labeled as $1D_{8'}$ and $1D_{9'}$ (red dashed lines). Atomic colors: zinc violet, oxygen red, carbon gray and hydrogen white. Adapted from [108] with permission from ©2022 AIP Publishing.

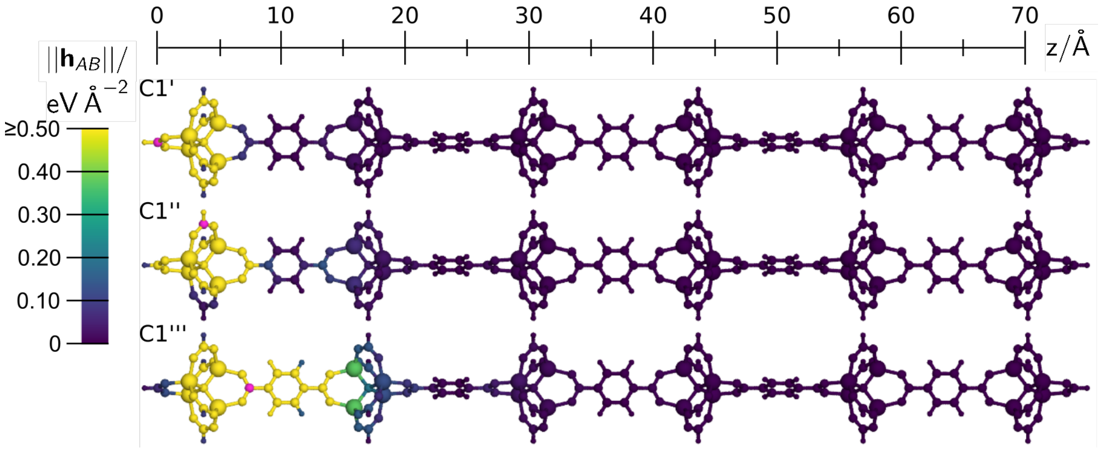

Figure 4.11: The atomic Hessian submatrix norm $||\mathbf{h}_{AB}||$ of the three C1-like atoms C1$'$, C1$''$ and C1$'''$ (magenta) of 1D related to the C1 atomic site in IRMOF-1. The remaining atomic sites are shown in appendix (fig. A.5 and A.6). Adapted from [108] with permission from ©2022 AIP Publishing.

in the carboxyl group bridging the gap between SBU and linker in IRMOF-1.

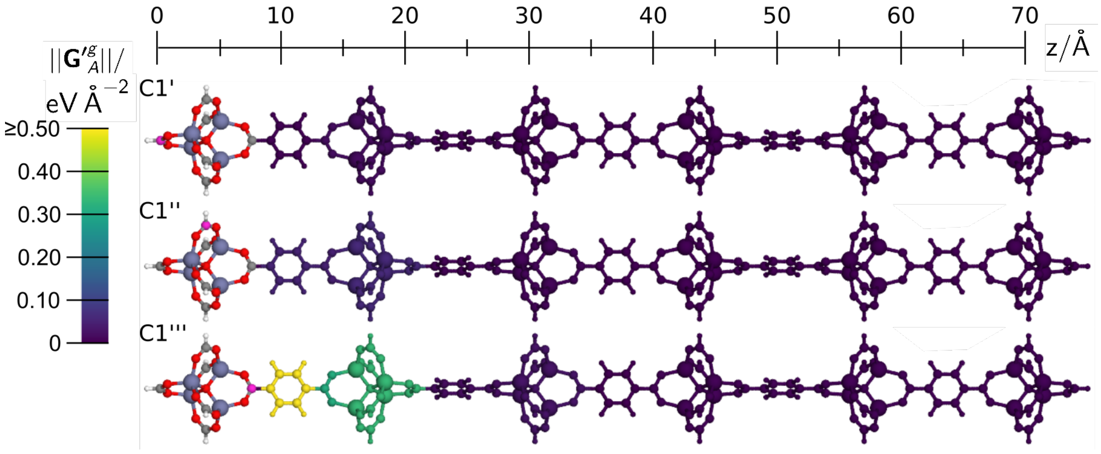

Figure 4.12: The effective Hessian group matrix norm $||\mathbf{G}'^{g}_{A}||$ of the three C1-like atoms C1$'$, C1$''$ and C1$'''$ (magenta) of 1D related to the C1 atomic site in IRMOF-1. The remaining atomic sites are shown in appendix (fig. A.7 and A.6). Adapted from [108] with permission from ©2022 AIP Publishing.

Furthermore, size-converged fragments are derived by employing $\Gamma = 0.29\,\mathrm{eV\,Å^{-2}}$ based on the analysis of the model systems above. For C1$'''$, the threshold Γ predicts at least the fragment $1D_6$ and larger fragments as size-converged, providing accurate forces for the C1$'''$ position as shown in table 4.3. The force error of C1$'''$ in $1D_8$ is indeed lower than the chosen convergence criterion $f = 0.125 < 0.0016$. Repeating the conclusion of the model systems: *the threshold* Γ *is a universal threshold, resulting in size-converged fragments for a variety of bonding types*, which is confirmed by the results of the 1D model system and thus Γ is even extendable to more complex systems like IRMOF structures. The resulting molecular fragments represent the forces in the desired accurate manner.

Table 4.3: Compilation of the force component errors $\Delta f^{1\mathrm{D}_g}_{\mathrm{C1}'''_{x,y,z}}$ and the total force errors $||\Delta \mathbf{f}^{1\mathrm{D}_g}_{\mathrm{C1}'''}||$ in eV Å^{-1} for the C1$'''$ reference atom in different fragments of the one-dimensional IRMOF-1 system (1D) shown in figure 4.11. Further, the effective Hessian group matrix norm $||\mathbf{G}'^{g}_{\mathrm{C1}'''}||$ is given in eV Å^{-2} as shown in figure 4.12. Numbers outside the intended convergence are given in bold. Adapted from [108] with permission from ©2022 AIP Publishing.

g	$\Delta f^{1\mathrm{D}_g}_{\mathrm{C1}'''_x}$	$\Delta f^{1\mathrm{D}_g}_{\mathrm{C1}'''_y}$	$\Delta f^{1\mathrm{D}_g}_{\mathrm{C1}'''_z}$	$\|\|\Delta \mathbf{f}^{1\mathrm{D}_g}_{\mathrm{C1}'''}\|\|$	$\|\|\mathbf{G}'^{g}_{\mathrm{C1}'''}\|\|$
ref	0.0000	0.0000	0.0000	0.0000	0.00
1	0.0001	-0.0001	0.0000	0.0002	0.01
2	0.0000	0.0000	-0.0013	0.0013	0.01
3	0.0000	0.0000	0.0001	0.0001	0.00
4	0.0001	0.0001	0.0005	0.0005	0.03
5	0.0000	0.0000	-0.0069	0.0069	0.02
6	0.0000	0.0000	0.0016	0.0016	0.02
7	0.0006	-0.0006	0.0003	0.0008	**0.34**
8	0.0000	0.0000	**−0.1492**	**0.1492**	**0.31**
9	-0.0003	0.0001	**3.0328**	**3.0328**	**21.80**

4.5.2 3D-IRMOF Fragments

The results from the simplified 1D system, proved the principle and the effective Hessian group matrix norm threshold Γ, which is based on the four idealized model systems above (sec. 4.4.4). And the threshold Γ can additionally be applied to more complex structures. Thus, the next step is the derivation of size-converged three-dimensional IRMOF fragments predicted by the threshold $\Gamma = 0.29$. However, for the reference system no periodic bulk structure is used to avoid interferences of periodic images, which may introduce artificial periodic effects affecting the results of the environment dependencies on the central reference atom force. Extending the periodic bulk structure to a periodic supercell, increases the computational effort enormously and shifts the ratio of computational effort to knowledge earnings far to the effort side. Instead, very large molecular fragments are constructed for each in-equivalent atomic site (fig. 4.1). These fragments need to be as large to show nearly zero interactions to the outer most shell of atoms, ensured by the atomic Hessian submatrix norm $||\mathbf{h}_{AB}|| < 0.1\,\mathrm{eV\,Å^{-2}}$, which is derived from the 1D system above.

IRMOF-1 The resulting very large molecular fragments ($\mathrm{C1_{ref}}$, $\mathrm{Zn1_{ref}}$, $\mathrm{O1_{ref}}$, $\mathrm{O2_{ref}}$, $\mathrm{C2_{ref}}$, $\mathrm{C3_{ref}}$, and $\mathrm{H1_{ref}}$) are used as reference structures for the different in-equivalent atomic sites (C1, Zn1, 01, O2, C2, C3 and H1 as defined in fig. 4.1). With a fragment radius around $r_{\mathrm{frag}} = 10 - 12$ Å. As a decisive example, the effect of the environment for the atomic site C1 is discussed in detail. The results of the remaining in-equivalent sites are also shown in the appendix and referenced, respectively.

Similar to the four model systems and the 1D system, the atomic Hessian submatrix norm $||\mathbf{h}_{AB}||$ and the atomic distance d_{AB} of the reference atom A and the neighboring atom B correlate negatively. While the distance increases, the atomic Hessian submatrix norm decreases (fig. 4.13, panel a)). Additionally, the atomic Hessian submatrix norm decreases similar to the results of the 1D case (fig. 4.11, panel c)). The groups g of the fragments (fig. 4.14) form a shell-like structure around the central reference atom. Shells near to

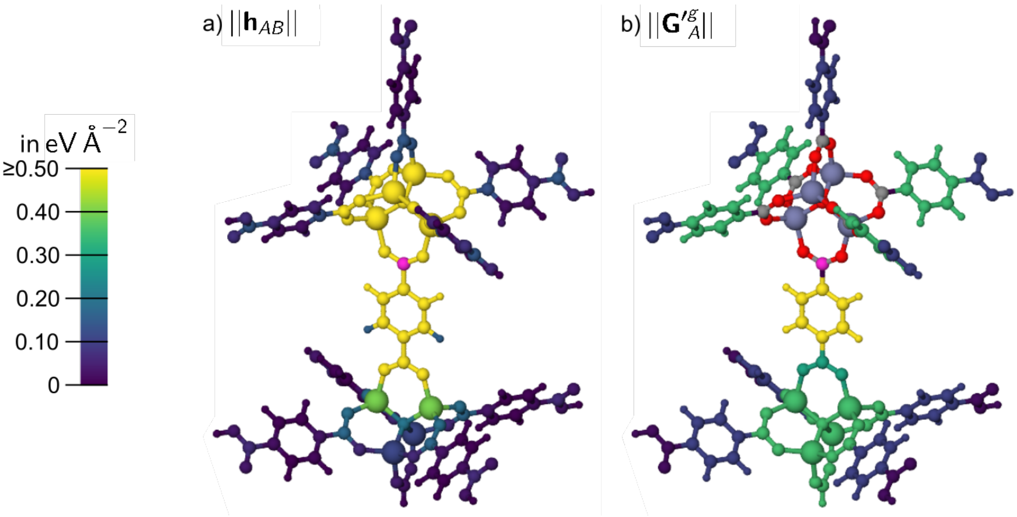

Figure 4.13: a) The atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G'}^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C1 (magenta) in reference structure C1$_{\text{ref}}$. $||\mathbf{G'}^{g}_{\text{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements. Adapted from [108] with permission from ©2022 AIP Publishing.

the reference atom include also the more distant atoms, equivalent to the groups in the model systems and the 1D system. Due to the three-dimensionality of the structures, more atoms are included in the groups than in the 1D system, increasing the truncated atomic interactions for the constructed fragments, which may be reflected by the larger values of the effective Hessian group matrix norm. Thus, the reference atom interactions with its environment may be increased in total, compared to the 1D system. This has to be considered for the definition of the size-converged fragment.

Table 4.4: Compilation of the force error components $\Delta f^{\text{C1}_g}_{\text{C1}_{x,y,z}}$ and of the force vector $||\Delta\mathbf{f}^{\text{C1}_g}_{\text{C1}}||$ of the reference atom C1 (fig. 4.13) for different fragments in eV Å^{-1}. Further, the effective Hessian group matrix norm $||\mathbf{G'}^{g}_{\text{C1}}||$ is given in eV Å^{-2}. Numbers outside the intended convergence level are given in bold. The fragments C1$_{1-7}$ are shown in figure 4.14. Adapted from [108] with permission from ©2022 AIP Publishing.

g	$\Delta f^{\text{C1}_g}_{\text{C1}_x}$	$\Delta f^{\text{C1}_g}_{\text{C1}_y}$	$\Delta f^{\text{C1}_g}_{\text{C1}_z}$	$\|\|\Delta\mathbf{f}^{\text{C1}_g}_{\text{C1}}\|\|$	$\|\|\mathbf{G'}^{g}_{\text{C1}}\|\|$
ref	0.0000	0.0000	0.0000	0.0000	0.00
1	-0.0018	0.0022	0.0165	0.0167	0.02
2	-0.0019	0.0022	-0.0733	0.0734	0.10
3	-0.0019	0.0022	-0.0630	0.0630	0.11
4	-0.0013	0.0016	-0.0599	0.0599	**0.36**
5	-0.0024	0.0027	0.0172	0.0176	**0.35**
6	-0.0018	0.0022	**0.1666**	**0.1666**	**0.30**
7	-0.0016	0.0016	**−3.0135**	**3.0135**	**21.83**

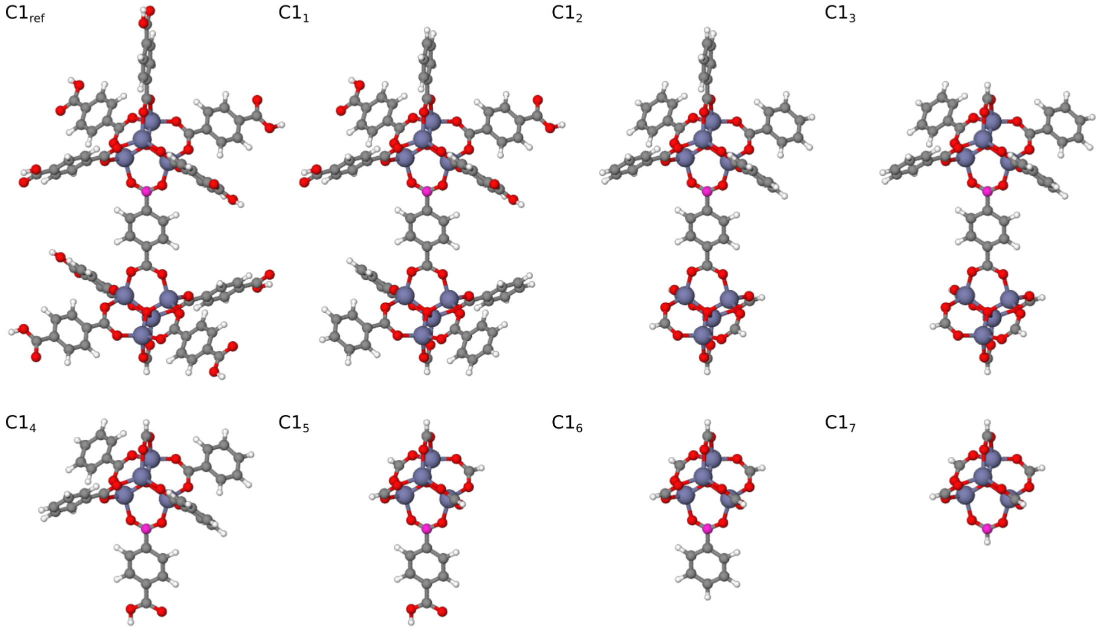

Figure 4.14: The reference structure $C1_{\text{ref}}$ and the smaller molecular fragments $C1_{1-7}$ constructed for the IRMOF-1 atomic site C1 (magenta) including the hydrogen atoms saturating the broken bonds. Adapted from [108] with permission from ©2022 AIP Publishing.

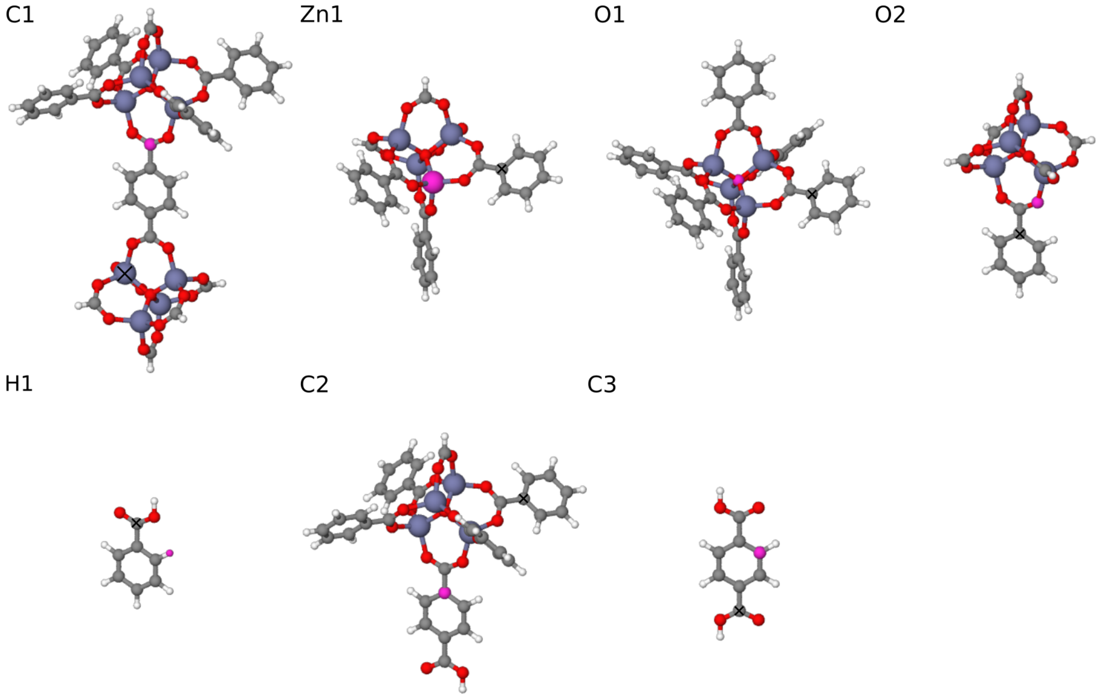

Figure 4.15: Fragment structures, reduced to the atoms with significant interaction for the chosen force accuracy criterion $||\Delta\mathbf{f}^{\mathrm{max}}|| = 0.125\,\mathrm{eV}\,\mathrm{Å}^{-1}$, resulting from the Hessian analysis of IRMOF-1 for the in-equivalent positions C1, Zn1, O1, O2, H1, C2 and C3 (magenta). The cross marks the atom with the largest $||\mathbf{h}_{AB}||$ of the outermost group g. Adapted from [108] with permission from ©2022 AIP Publishing.

Employing the effective Hessian group matrix norm threshold derived from the model systems, $\mathrm{C1}_3$ is stated as the size-converged fragment for the in-equivalent atomic site C1 with a force error $||\Delta\mathbf{f}_{\mathrm{C1}}^{\mathrm{C1}_3}|| = 0.0630\,\mathrm{eV}\,\mathrm{Å}^{-1}$. Also the results of the remaining positions (fig. A.9 to A.15, tab. A.5 and A.6) are similar to the results of the 1D system. Hence employing the effective Hessian group matrix norm threshold results in size-converged fragments (fig. 4.15) with well converged forces (tab. A.5 and A.6). Nevertheless, the environment dependence and thus, the fragment size is very diverse for the chosen reference atomic site, which highlights the difference in boding type and the connected local electronic structure. As a spill over effect, some of the smaller sized fragments are redundantly included in the more extended fragments. Thus, the size-converged fragments of C1 and O1 (fig. 4.15) are the non-redundant fragments and provide effectively size-converged DFT forces for all remaining atomic sites. These two fragments could be the foundation of a data set, which includes different structural configurations of these two fragments and is used for HDNNP training.

Nevertheless, this approach of creating a data set suffers from the different effective atom environment represented by the fragments for the atomic sites. But for an accurate description of the PES by HDNNP or more general MLP, the environment of each atom should be known up to the same radius. To ensure this for all atomic sites, the largest appearing fragment radius r_{frag} of the size-converged fragments is decisive to gain a general fragment radius for all of those atomic positions. Within this chosen fragment radius each, fragment provides accurate DFT forces for its specific central reference atom. An simple procedure to derive the fragment radius, is to take the distance from the central reference atom to its most distant neighboring atom. Conversely, many of these atoms in the derived fragments show only very weak interactions with the reference atom and are included just because this atom is part of a non-divisible chemical entity, like a phenylene-ring, a carboxyl-group or a SBU. A more sensitive definition of the fragment radius is defined by the distance d_{AB} of the reference atom A and the neighboring atom B, contributing with the largest atomic Hessian submatrix norm $||\mathbf{h}_{AB}||$ to the effective Hessian group matrix norm $||\mathbf{G}'^{g}_{A}||$ of the outer most group g. This procedure leads to the fragment radii between the reference atom A and the specific neighbor atom B as marked in figure 4.15 and summarized in table 4.5.

Table 4.5: Fragment radii r_{frag} in Å obtained for the fragments shown in figure 4.15 for the seven atomic sites in IRMOF-1. g is the number of the converged fragment and simultaneously of the outer most group. Adapted from [108] with permission from ©2022 AIP Publishing.

A	g	r_{frag}
Zn1	3	4.333
O1	3	5.165
O2	6	2.379
C1	3	8.502
C2	3	7.304
C3	4	3.817
H1	5	2.725

IRMOF-10 Similar to IRMOF-1, a very large reference molecular fragment (I10) for all atomic sites of IRMOF-10 (fig. 4.1) is defined, with nearly zero interactions of the reference atoms and the outer most fragment atoms, ensured by the atomic Hessian submatrix norm $||\mathbf{h}_{AB}|| < 0.1\,\text{eV}\,\text{Å}^{-2}$.

In figure 4.16 the results of the C1 position are shown, because of the prominent long-range character in contrast to the other atomic positions (fig. A.16 to A.25). Fragments used for the analysis of the C1 atomic site occurring in IRMOF-10 are shown in figure 4.14. The overall behaviour for each atomic site occurring in IRMOF-10, is similar to the results of IRMOF-1. Thus, no qualitative differences to the IRMOF-1 results arise in the IRMOF-10 results and even for the atomic sites Zn1, O1, O2 and H1, the results of IRMOF-1 (fig. A.9, A.10, A.11 and A.15) and -10 (fig. A.16, A.17, A.18 and A.24) are nearly equivalent. Employing the threshold of the effective Hessian group matrix norm results in the fragment I10$_2$ (fig. 4.17) with a force error of $||\Delta\mathbf{f}^{\text{I10}_g}_{\text{C1}}|| = 0.0236\,\text{eV}\,\text{Å}^{-1} < 0.125\,\text{eV}\,\text{Å}^{-1}$ (tab. 4.6). Also for the remaining atomic sites, fragments fulfilling the desired force error convergence are derived by the effective Hessian group matrix norm threshold, which are shown in figure 4.18. The related force errors and effective Hessian group matrix norms are summarized in table A.7 and A.8. The derived minimum fragments (fig. 4.15 and 4.18) for the atomic sites Zn1, O1 and H1 are equivalent for the IRMOF-1 and -10 structure, whereas the fragments for C1 and O2 are still similar in both cases. For C1, the slight differences are related to the structural linker changes – the additional phenylene ring – going from IRMOF-1 to IRMOF-10 and for the O2 site, low radius fragments where not considered as detailed in the IRMOF-10 case as for IRMOF-1. For the atomic sites C2 and C3, more differences are arising, also related to the structural changes within the linker and thus changing the low distance environment and the local electronic structure, mostly effecting these atoms. For the remaining atomic sites (C4, C5 and H2) no equivalents are existent within the IRMOF-1 structure. The environment dependence of the

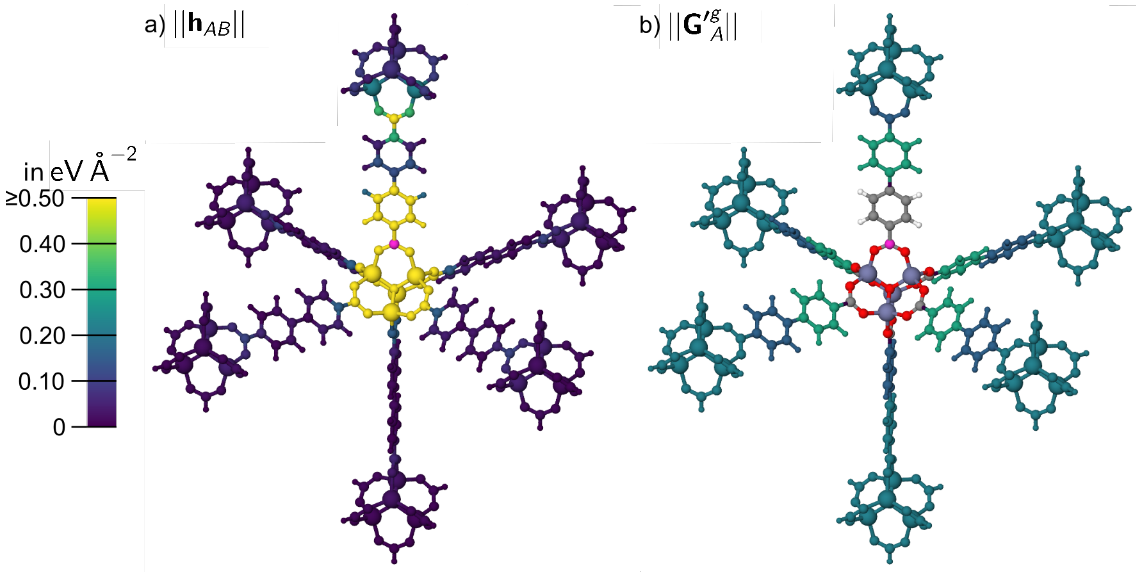

Figure 4.16: a) The atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A = \mathrm{C1}$ (magenta) in reference structure I10$_{\mathrm{ref}}$. $||\mathbf{G}'^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

Table 4.6: Compilation of the force error components $\Delta f^{\mathrm{I10}_g}_{\mathrm{C1}_{x,y,z}}$ and of the force vector $||\Delta\mathbf{f}^{\mathrm{C1}_g}_{\mathrm{C1}}||$ of the reference atom C1 (fig. 4.13) for different fragments in eV Å^{-1}. Further, the effective Hessian group matrix norm $||\mathbf{G}'^{g}_{\mathrm{C1}}||$ is given in eV Å^{-2}. Numbers outside the intended convergence level are given in bold. The fragments I10$_g$ for the reference site C1 are shown in figure 4.17.

g	$\Delta f^{\mathrm{I10}_g}_{\mathrm{C1}_x}$	$\Delta f^{\mathrm{I10}_g}_{\mathrm{C1}_y}$	$\Delta f^{\mathrm{I10}_g}_{\mathrm{C1}_z}$	$\|\|\Delta\mathbf{f}^{\mathrm{I10}_g}_{\mathrm{C1}}\|\|$	$\|\|\mathbf{G}'^{g}_{\mathrm{C1}}\|\|$
ref	0.0000	0.0000	0.0000	0.0000	0.00
1	-0.0013	0.0020	0.0430	0.0431	0.22
2	-0.0040	0.0049	0.0905	0.0907	0.16
3	-0.0013	0.0020	-0.0235	0.0236	**0.30**

atomic sites, is as in the IRMOF-1 case (tab. 4.5) very different from each other resulting in molecular fragments different in size, which satisfy the defined force error convergence criterion $||\Delta\mathbf{f}^{\mathrm{I10}_g}_{\mathrm{A}}|| < 0.125\,\mathrm{eV}\,\text{Å}^{-1}$. The fragment radii of the IRMOF-10 minimum structures (tab. 4.7) are determined by the same procedure as described and applied for the IRMOF-1 minimum fragment structures. In comparison to the IRMOF-1 fragment radii (tab. 4.5) the environment dependence of the C2 position is decreased in the IRMOF-10, based on the changed linker structure, which increases the distance between the C2 atomic site and the carboxyl group on the opposite site of the linker, but simultaneously decreases the interactions with these carboxyl group atoms in the IRMOF-10 case. Thus, the strong interactions between the C2 site and the carboxyl group atoms are decreased (fig. A.20) in comparison to the IRMOF-1 case (fig. A.13) and not important for the construction of a size-converged molecular fragment. Another important aspect, the C2 sites are only to a zero order approximation equivalent, since the IRMOF linker molecules are different,

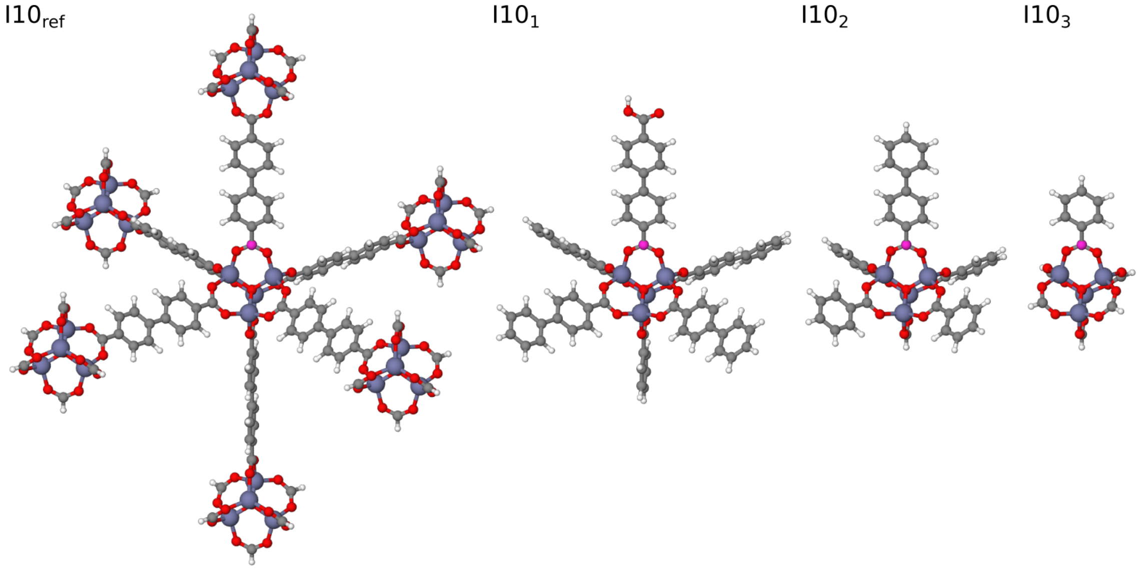

Figure 4.17: The reference structure I10$_{\text{ref}}$ and the smaller molecular fragments I10$_{1-3}$ constructed for the IRMOF-10 atomic site C1 (magenta) including the hydrogen atoms saturating the broken bonds.

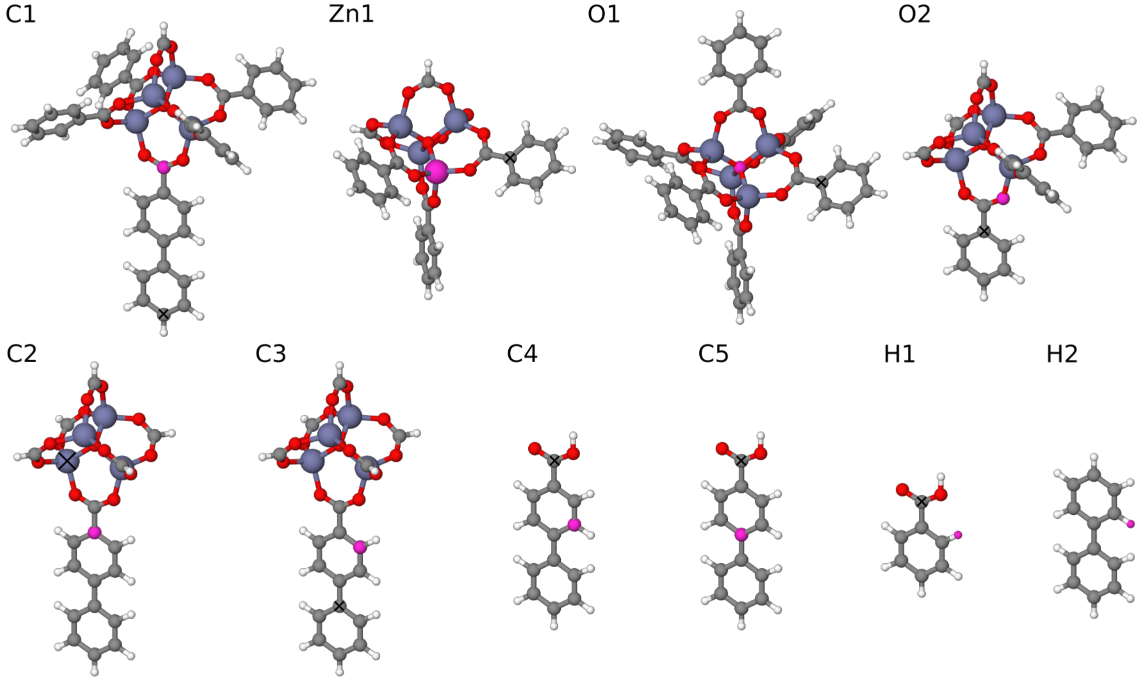

Figure 4.18: Fragment structures, reduced to the atoms with significant interaction for the chosen force accuracy criterion $||\Delta\mathbf{f}^{\max}|| = 0.125\,\text{eV}\,\text{Å}^{-1}$, resulting from the Hessian analysis of IRMOF-10 for the in-equivalent positions C1, Zn1, O1, O2, C2, C3, C4, C5, H1 and H2 (magenta). The cross marks the atom with the largest $||\mathbf{h}_{AB}||$ of the outermost group g. For the H2 position, the smallest considered molecular fragment (biphenylene) is size-converged and a fragment radius is not defined. Considering still a smaller fragment (benzene, compare to H1), would just increase the computational effort without further insight.

which for sure will affect the environment dependence of this site in the different IRMOF structures. For the atomic sites Zn1, O1, O2, C1, C3 and H1, the derived fragment radii are very similar for the IRMOF-1 and -10 case. As mentioned above, for the C4, C5 and H2 sites, there are no equivalents within the IRMOF-1 structure to compare with.

Table 4.7: Fragment radii r_{frag} in Å obtained for the fragments shown in figure 4.18 for the ten atomic sites in IRMOF-10. g is the number of the converged fragment and simultaneously of the outer most group. For the H2 position, no atom is marked by a cross, since the smallest considered molecular fragment (biphenylene) is already size-converged. Considering still a smaller fragment (benzene, compare to H1), would just increase the computational effort without further insight.

A	g	r_{frag}
Zn1	3	4.328
O1	3	5.160
O2	3	2.376
C1	2	8.718
C2	3	4.328
C3	2	3.830
C4	2	3.816
C5	2	4.364
H1	3	2.731
H2	–	–

IRMOF-16 Equivalent to the reference structure I10 as used above, a large reference molecular structure (I16) is constructed for the analysis of all in-equivalent atomic sites of IRMOF-16 (fig. 4.1). Again, the size of the fragment does not illustrate any strong atomic interactions of a reference atom with its outermost neighboring atoms, ensured by the atomic Hessian submatrix norm $||\mathbf{h}_{AB}|| < 0.1\,\mathrm{eV}\,\text{Å}^{-2}$. Similar to the IRMOF-1 and -10, the C1 position demonstrates the most prominent long-range character of all in-equivalent atomic sites in IRMOF-16. Therefore, the C1 results are portrayed exemplary in figure 4.19. Nevertheless, the qualitative relation of atomic Hessian submatrix norm and the atomic distance remains unchanged in IRMOF-16: increasing the atomic distance between the reference and the neighboring atom, but decreasing the interaction of these atoms as demonstrated by the decreasing atomic Hessian submatrix norm (fig. 4.19 a)).This can also be confirmed by the effective Hessian group matrix norm (fig. 4.19 b)) being similar for all remaining in-equivalent atomic site (fig. A.26 to A.38). Thus, in addition to the similarities of IRMOF-1 and -10 results, also IRMOF-16 provides nearly equivalent results for the atomic positions Zn1, O1, O2 and H1 (IRMOF-1: fig. A.9, A.10, A.11 and A.15; IRMOF-10: fig. A.16, A.17, A.18 and A.24; IRMOF-16: fig. A.26, A.27, A.28 and A.36). Furthermore, the atomic positions C2, C3, C4, C5 and H2 indicate nearly equivalence by comparing the IRMOF-10 (fig. A.20, A.21, A.22, A.23 and A.25) and IRMOF-16 results (fig. A.30, A.31, A.32, A.33 and A.37). For the in-equivalent positions of C6, C7 and H3 of IRMOF-16, no equivalents to compare with are included in the structures IRMOF-1 and -10. Employing the effective Hessian group matrix norm threshold $\Gamma = 0.29\mathrm{eV}\,\text{Å}^{-2}$ provides the size-converged fragments in figure 4.20 fulfilling the force error criterion of $||\Delta\mathbf{f}_{\mathrm{A}}^{\mathrm{I16}_g}|| < 0.125\,\mathrm{eV}\,\text{Å}^{-1}$ as indicated by tables A.9 and A.10. The size-converged fragments and the related fragment radii (tab.4.9) are equivalent to the IRMOF-10 results and similar to IRMOF-1 with the same differences as already mentioned above for IRMOF-10.

Surprisingly, the size-converged fragments for the C5 position differ in IRMOF-16 and -10 by a terminal carboxyl group being present for the C5 size-converged fragment in IRMOF-10 (fig. 4.18), but not in the C5 size-converged fragment of IRMOF-16 (fig. 4.20). A reason for this deviation, may be based on the less detailed analysis of the low radius fragments for the C5 position. However, the resulting fragment radius for C5 would be small in comparison to the other atomic positions (C1 in IRMOF-1 and -10) and thus, its accurate fragment radius is not of further interest.

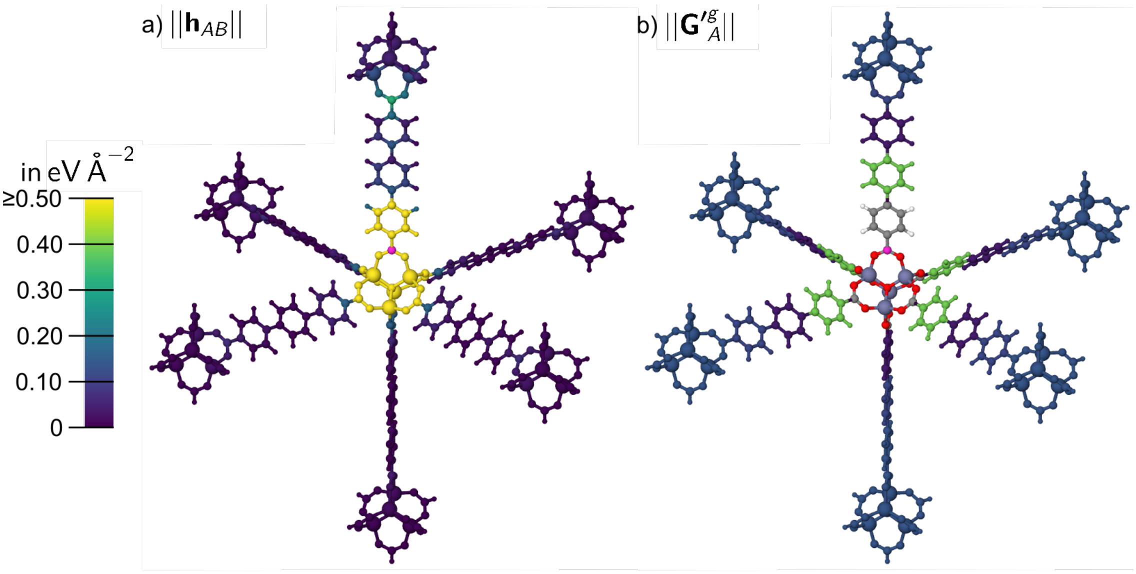

Figure 4.19: a) The atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A = \text{C1}$ (magenta) in reference structure I16$_{\text{ref}}$. $||\mathbf{G}'^{g}_{\text{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

Table 4.8: Compilation of the force error components $\Delta f^{\text{I16}_g}_{\text{C1}_{x,y,z}}$ and of the force vector $||\Delta\mathbf{f}^{\text{I16}_g}_{\text{C1}}||$ of the reference atom $A = \text{C1}$ (fig. 4.19) for different fragments in eV Å^{-1}. Further, the effective Hessian group matrix norm $||\mathbf{G}'^{g}_{\text{I16}}||$ is given in eV Å^{-2}. Numbers outside the intended convergence level are given in bold.

g	$\Delta f^{\text{I16}_g}_{\text{C1}_x}$	$\Delta f^{\text{I16}_g}_{\text{C1}_y}$	$\Delta f^{\text{I16}_g}_{\text{C1}_z}$	$\lvert\lvert\Delta\mathbf{f}^{\text{I16}_g}_{\text{C1}}\rvert\rvert$	$\lvert\lvert\mathbf{G}'^{g}_{\text{C1}}\rvert\rvert$
ref	0.0000	0.0000	0.0000	0.0000	0.00
1	-0.0012	0.0013	0.0317	0.0318	0.14
2	0.0000	-0.0001	-0.0244	0.0244	0.11
3	-0.0001	-0.0001	-0.0308	0.0308	0.04
4	0.0000	-0.0001	-0.0007	0.0008	**0.40**

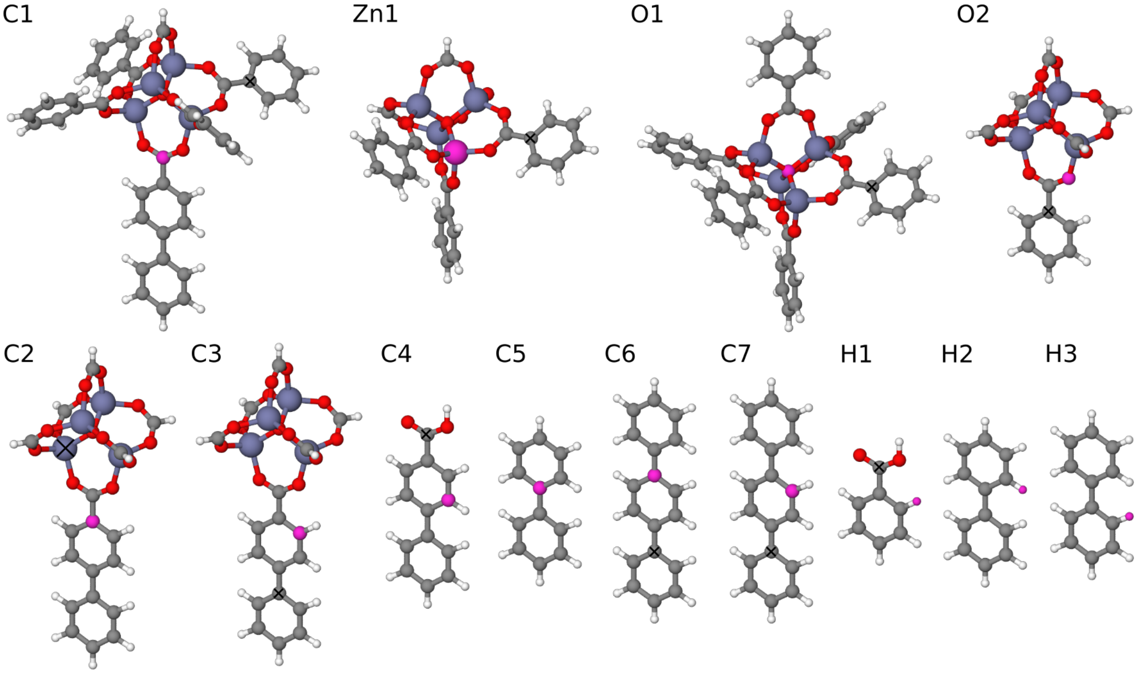

Figure 4.20: Fragment structures, reduced to the atoms with significant interaction for the chosen force accuracy criterion $||\Delta\mathbf{f}^{\max}|| = 0.125\,\text{eV}\,\text{Å}^{-1}$, resulting from the Hessian analysis of IRMOF-16 for the in-equivalent positions C1, Zn1, O1, O2, C2, C3, C4, C5, C6, C7, H1, H2 and H3 (magenta). The cross marks the atom with the largest $||\mathbf{h}_{AB}||$ of the outermost group g. For the positions C5, H2 and H3, the smallest considered molecular fragments (biphenylene) are size-converged and a fragment radius is not defined. Considering still a smaller fragment (benzene, compare to H1), would just increase the computational effort without further insight.

Table 4.9: Fragment radii r_{frag} in Å obtained for the fragments shown in 4.20 for the ten atomic sites in IRMOF-10. g is the number of the converged fragment and simultaneously of the outer most group. For the positions C5, H2 and H3, the smallest considered molecular fragments (biphenylene) are size-converged and a fragment radius is not defined. Considering still a smaller fragment (benzene, compare to H1), would just increase the computational effort without further insight.

A	g	r_{frag}
Zn1	4	4.333
O1	4	5.158
O2	4	2.375
C1	3	6.316
C2	3	4.327
C3	3	3.827
C4	2	4.811
C5	–	–
C6	3	4.389
C7	5	3.830
H1	4	2.730
H2	–	–
H3	–	–

4.6 HDNNP-Training Fragments

For each of the IRMOF bulk structures, size-converged fragments (fig. 4.15, 4.18 and 4.20) are defined, which embedding the in-equivalent atomic sites (fig. 4.1) in the minimum required spatial environment. The different in-equivalent atomic sites demonstrate different dependencies on the atomic environment (tab. 4.5, 4.7 and 4.5). The most long-ranged obtained environment dependence is pronounced by the C1 atomic site of IRMOF-10 with a fragment radius $r_{\text{frag}} = 8.718\,\text{Å}$. This fragment radius deals as a global fragment radius for all atomic sites to construct size-converged fragments, fulfilling the force error criterion for the central reference atoms of the fragments (fig.,4.21). For IRMOF-1, the fragment of Zn1 includes the O1 and O2 fragment, while the C2 fragment includes the fragments of C1, C3 and H1. For IRMOF-10, the Zn1 fragment includes the fragments of O1, O2, C1 and H1, as the fragment of C5 includes the C2, C3, C4, H2 and additionally also the C1 and H1 fragments. For IRMOF-16, the fragment of Zn1 is equivalent to the Zn1 fragment of IRMOF-10 and includes the O1, O2, C1, C2 and H1 fragments, while the C3 fragment includes the fragments of C4, C5, C6, H2, H3 and additionally also of C1, C2 and H1. Furthermore, the IRMOF-16 C7 atomic site is only accurately represented by the C7 fragment. Consequently, there remain only six non-redundant fragments for the accurate description of all atomic sites occurring in the three IRMOF bulk structures, which are summarized in figure 4.21. The advantage of these non-redundant fragments is to provide converged reference data for a HDNNP training set for different atomic sites simultaneously and to reduce the amount of computational effort compared to atomic site specific fragments. Related to the uncertainties of the fragment radius definition, the resulting fragments are definitely not sensitive to a specific definition, since the main structural IRMOF building blocks – the

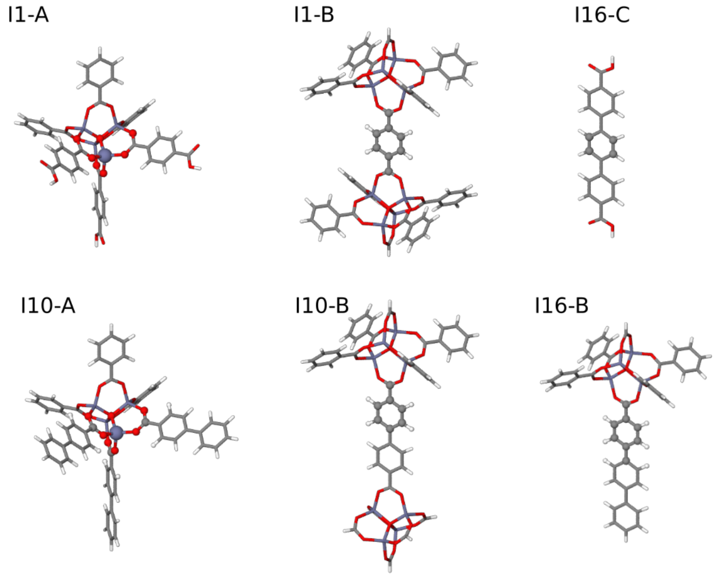

Figure 4.21: Non-redundant $r_{\text{frag}} = 8.718\,\text{Å}$ fragments of for the in-equivalent positions of IRMOF-1, -10 and -16 bulk structures, which can be used as foundation for a HDNNP training data set. I1-A and I1-B are based on Zn1 and C2 of IRMOF-1, I10-A and I10-B on Zn1 and C5 of IRMOF-10, I16-B and I16-C on C3 and C7 of IRMOF-16, respectively. The molecular fragment structures are shown by sticks and the element specific color, but the central atoms of the fragments, which are shown by balls and are embedded in the same environment as in the bulk up to a radius of $r_{\text{frag}} = 8.718\,\text{Å}$.

SBU, the phenylene ring and the carboxyl group – are in- or excluded completely within a molecular fragment structure to avoid significant changes of the electronic structure. For the definition of the HDNNP cutoff radius equal to the fragment radius $r_{\mathrm{cut}} = r_{\mathrm{frag}}$, not all atoms included in the fragment structure effect the atomic energy contribution E_A of a selected central reference atom A (magenta). Only atoms B (orange) within an atomic distance d_{AB} smaller than the cutoff radius $d_{AB} \leq r_{\mathrm{cut}}$ effects the atomic energy contribution E_A, while the remaining atoms (blue) are only included due to restrictions in the fragment construction procedure (fig. 4.23).

Besides the fragments in figure 4.21, more efficient fragments structures can be defined by increasing the number of atoms in a bulk-like environment by just slightly increasing the total number of atoms. For example, the fragment I1-A (fig. 4.21) shows only one zinc atom in a bulk-like environment, because for the remaining three zinc atoms the terminating carboxyl groups are missing to complete the bulk-like environment up to a radius of $r_{\mathrm{frag}} = 8.718$ Å. Including the missing carboxyl group atoms leads to the fragment I1-A′ with four bulk-like zinc atoms (fig. 4.22). Hence, by a moderate increase of the total number of atoms and thus a moderate increase of the computational effort, the information obtained by a DFT calculation for this fragment is strongly increased. Table 4.10 summarizes the number of total and bulk-like atoms, as well as their ratio, which is increased for all efficiency increased fragments. For the fragments I1-B and I16-C, a more efficient version of the molecular fragment structure is not existent as in the aspect of the fragment I1-A and these two fragments remain unaffected.

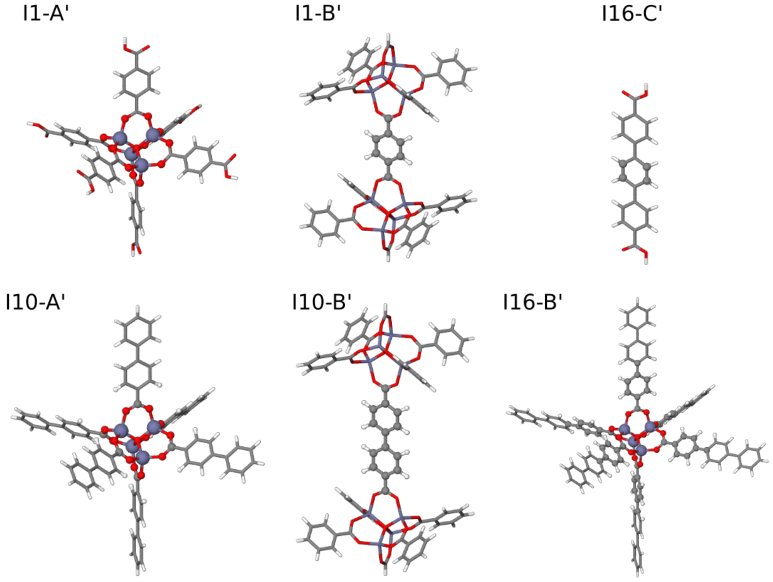

Figure 4.22: Non-redundant, increased efficiency $r_{\mathrm{frag}} = 8.718$ Å fragments of for the inequivalent positions of IRMOF-1, -10 and -16 bulk structures, which can be used as foundation for a HDNNP training data set. I1-A′ and I1-B′ (equivalent to I1-B) are based on Zn1 and C2 of IRMOF-1, I10-A′ and I10-B′ on Zn1 and C5 of IRMOF-10, I16-B′ and I16-C′ (equivalent to I16-C) on C3 and C7 of IRMOF-16, respectively. The molecular fragment structures are shown by sticks and the element specific color, but the central atoms of the fragments, which are shown by balls and are embedded in the same environment as in the bulk up to a radius of $r_{\mathrm{frag}} = 8.718$ Å.

Table 4.10: Compilation for M the total number of atoms, M_{bulk} the number bulk-like atoms within a cutoff radius $r_{\text{cut}} = 8.718\,\text{Å}$ and their ratio $\frac{M_{\text{bulk}}}{M}$ for the $r_{\text{frag}} = 8.718\,\text{Å}$ based (fig.4.21) and increased efficiency fragments (fig.4.22).

$r_{\text{frag}} = 8.718\,\text{Å}$ based	I1-A	I1-B	I10-A	I10-B	I16-B	I16-C
M	98	145	119	116	99	38
M_{bulk}	8	12	17	11	14	4
$\frac{M_{\text{bulk}}}{M}$	0.08	0.08	0.14	0.09	0.14	0.11
increased efficiency	I1-A′	I1-B′	I10-A′	I10-B′	I16-B′	I16-C′
M	107	145	149	156	209	38
M_{bulk}	17	12	35	22	101	4
$\frac{M_{\text{bulk}}}{M}$	0.16	0.08	0.23	0.14	0.48	0.11

4.7 Construction of a HDNNP-Based on Molecular Fragments Structures

From the DFT summary in section 4.6, a fragment radius r_{frag} is obtained being crucial for the construction of size-converged molecular fragments based on IRMOF-1, -10 and -16 structure. The central atoms of the fragments are characterized by a bulk-like environment up to a radius of $r_{\text{frag}} = 8.718\,\text{Å}$ and accurate atomic forces, which are particularly similar to the related bulk forces within the periodic bulk-structure. Thus, these non-redundant fragments form the foundation of a HDNNP training set including different configurational atomic arrangements probably created by MD simulations. This training set provides accurate IRMOF bulk forces for the central atoms of the fragments and can be used for the prediction of bulk properties [49], like bulk forces as demonstrated in figure 4.23 a) and b). As known from the theoretical background of HDNNP (sec. 2.6), the total potential energy E_{tot} of a system is separated into atomic energy contributions E_A of atoms A (eq. 2.52).

The non-observable auxiliary quantities – the atomic energy contributions – E_A depend on the local atomic environment described by a set of ACSFs up to the HDNNP cutoff radius r_{cut}. The atomic force component f_{A_α} (sec. 2.6.4) is the negative derivative of the total potential energy with respect to the atomic coordinate A_α of atom A (eq. 2.63)

$$f_{A_\alpha} = -\frac{\partial E_{tot}}{\partial A_\alpha} = -\sum_B \frac{\partial E_B}{\partial A_\alpha} \quad . \tag{4.10}$$

However, not all atoms B contribute to the force component f_{A_α} in the formalism of HDNNP, since only atomic energy contributions E_B depend on the atomic coordinate A_α, if the atomic distance d_{AB} is smaller than the cutoff radius,

$$\frac{\partial}{\partial A_\alpha} E_B \neq 0 \,\text{for}\, d_{AB} \leq r_{\text{cut}} \quad , \tag{4.11}$$

and vice versa for atoms C the distance d_{AC} is beyond the the cutoff radius,

$$\frac{\partial}{\partial A_\alpha} E_C = 0 \,\text{for}\, d_{AC} > r_{\text{cut}} \quad . \tag{4.12}$$

Thus, the ACSFs strictly ensures the locality approach for the atomic energy contributions. All interactions for a specific atom A with its neighboring atoms B beyond the cutoff radius ($d_{AB} > r_{\text{cut}}$), are truncated. Nevertheless, the atomic energy contributions E_B depend on all atoms within within their cutoff radius and thus, the force component f_{A_α} depends formally on all atoms within a $2r_{\text{cut}}$ environment. Consequently, the HDNNP formalism defines the atomic energy contribution a function of the single cutoff radius $E_A = f(r_{\text{cut}})$ and the atomic force as function of twice the cutoff radius $\mathbf{f}_A = f(2r_{\text{cut}})$. Since the derived fragment radius $r_{\text{frag}} = 8.718\,\text{Å}$, provides size-converged fragment with respect to accurate atomic forces compared to the bulk structures, the derived fragment radius is equivalent to the twice cutoff radius,

$$r_{\text{frag}} = 8.718\,\text{Å} = 2r_{\text{cut}} \quad , \tag{4.13}$$

$$r_{\text{cut}} = 4.359\,\text{Å} \quad . \tag{4.14}$$

For this reason, another set of molecular fragments can be constructed with the strong predictive power for the much larger bulk structures based on $r'_{\text{frag}} = 4.359\,\text{Å}$ (fig. 4.24), including less atoms than the fragments based on $r_{\text{frag}} = 8.718\,\text{Å}$ (fig. 4.22).

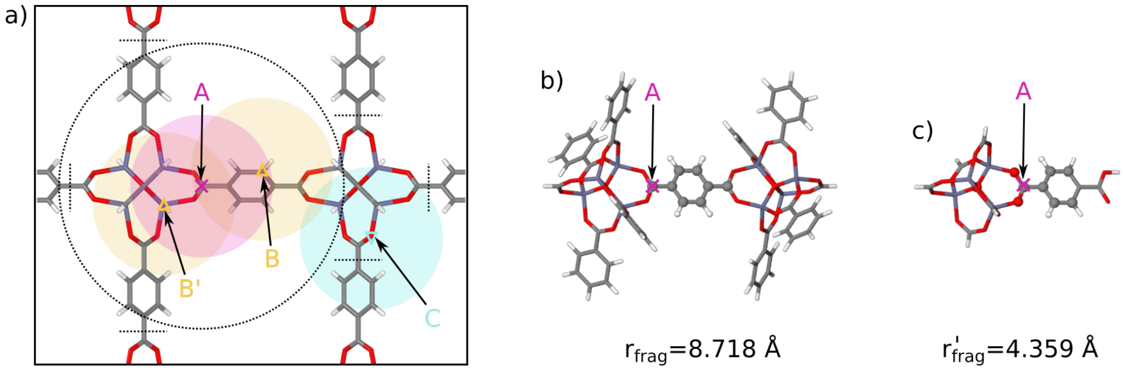

Figure 4.23: a) A 2D-projection of the IRMOF-1 structure with a marked C1 atom of interest A (magenta cross), two exemplary neighboring atoms B and B' (orange triangles) within the cutoff radius environment (magenta/orange/blue shaded circle) and a neighboring atom C (blue triangle) even outside the $2r_{\text{cut}}$ environment (black dashed circle) of atom A. The cutoff radii of atoms A, B and B' demonstrate the formal $2r_{\text{cut}}$ dependence of the force component f_{A_α}. The black dashed lines highlight the broken bonds for the size-converged fragment of the atomic site C1 and underline the integration of the complete $2r_{\text{cut}}$ environment within this fragment. b) The non-redundant fragment I1-B' for the accurate description of the in-equivalent atomic site C1 of IRMOF-1, being used together wit further fragment structures by the HDNNP to predict the atomic bulk force $\mathbf{f}_A$. c) The non-redundant fragment I1-Bs for the description of the in-equivalent atomic site C1 of IRMOF-1, based on $r'_{\text{frag}} = 4.359\,\text{Å}$, being used together wit further fragment structures by the HDNNP to predict the atomic bulk force $\mathbf{f}_A$.

The fragment radius defines the upper boundary for the cutoff radius, since up to this radius the central atoms of the fragments are embedded in a bulk-like environment. Comparing the bulk-like atoms within a cutoff radius $r_{\mathrm{cut}} = 4.359\,\text{Å}$, the $r_{\mathrm{frag}} = 8.718\,\text{Å}$ fragments include more bulk-like atoms than their $r_{\mathrm{frag}} = 4.359\,\text{Å}$ counter parts (tab. 4.11). To image the accurate atomic bulk forces by a molecular fragment structure, definitely the size-converged fragments based on $r_{\mathrm{frag}} = 8.718\,\text{Å}$ (fig. 4.22) demonstrate the significant minimum environment, which needs to be considered for the specific atomic sites of the IRMOF structures. This is related to the underlying physics of the IRMOF systems and comparable to map the periodic bulk properties onto molecular fragment structures as used for example in the cluster approach to model surfaces [109] and related to the already mentioned problems in QM/MM approaches in section 3.2: mapping properties of a computational demanding real system, like a complex periodic surface structure or an enzyme including a huge number of atoms onto a computationally non-demanding model system. However, this imaging of periodic bulk properties, like atomic forces onto molecular fragments, is not the main aspect here. The goal is to train a HDNNP based on molecular fragment data, resulting in the predictive power of the HDNNP for bulk properties, among others also for the atomic bulk forces. Nevertheless, the predictive power is not related to any further requirements for the molecular fragments itself, but the correct description of the bulk environment up to the cutoff radius ($r_{\mathrm{frag}} = 8.718\,\text{Å} = 2r_{\mathrm{cut}} \to r_{\mathrm{cut}} = 4.359\,\text{Å}$) and the correct relation of the structure and the total binding energy of a system (eq. 2.56). Thus the molecular fragments do not need to illustrate the accurate atomic bulk forces for their central atoms and the restriction

Table 4.11: Compilation for M the total number of atoms, M_{bulk} the number bulk-like atoms within a cutoff radius $r_{\mathrm{cut}} = 4.359\,\text{Å}$ and their ratio $\frac{M_{\mathrm{bulk}}}{M}$ for the $r_{\mathrm{frag}} = 4.359\,\text{Å}$-based (fig.4.24) and the fragments based on $r_{\mathrm{frag}} = 8.718\,\text{Å}$ (fig. 4.22).

	I1-As	I1-Bs	I10-Bs	I16-Cs	I1-A'	I1-B'	I10-A'	I10-B'	I16-B'	I16-C'
M	59	42	49	32	107	146	149	156	209	38
M_{bulk}	17	11	17	16	65	62	101	68	161	24
$\frac{M_{\mathrm{bulk}}}{M}$	0.29	0.26	0.35	0.50	0.61	0.42	0.68	0.44	0.77	0.63

$$\mathbf{f}_A^{\mathrm{frag}} \overset{!}{=} \mathbf{f}_A^{\mathrm{bulk}} \quad , \tag{4.15}$$

does not need to be fulfilled for the molecular fragments. It might be possible for the HDNNP to predict the accurate atomic bulk forces, while trained on molecular fragments as reference data, which do not provide the accurate atomic bulk forces of a chosen central atom (fig. 4.23c). This is related to the HDNNP prediction of the bulk forces, which is based on the total potential bulk energy. As long as the prediction of the bulk energy is accurate, this should be valid for the atomic bulk forces.

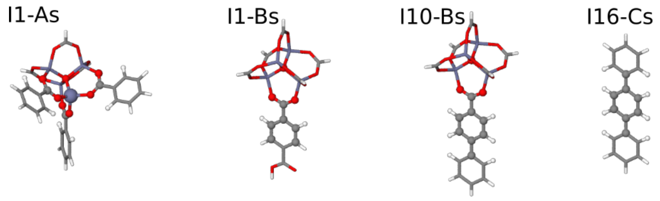

Figure 4.24: Non-redundant, $r'_{\mathrm{frag}} = 4.359\,\text{Å}$ fragments for the in-equivalent positions of IRMOF-1, -10 and -16 bulk structures. I1-As describes the Zn1, O1 and O2 positions for all three IRMOF structures; I1-Bs describes also the O1 and O2 position, as well as C1 and H1 for all IRMOF structures, additionally C2 and C3 for IRMOF-1; I10-Bs desribes also O1, O2, C1 and H1 for all IRMOF structures and C2, C3, C4, C5 and H2 for IRMOF-10 and -16; I16-Cs describes also C5 and H2 for IRMOF-10 and -16 and C6, C7 and H3 for IRMOF-16. The molecular fragment structures are shown by sticks and the element specific color, but the central atoms of the fragments, which are shown by balls and are embedded in the same environment as in the bulk up to a radius of $r'_{\mathrm{frag}} = 4.359\,\text{Å}$.

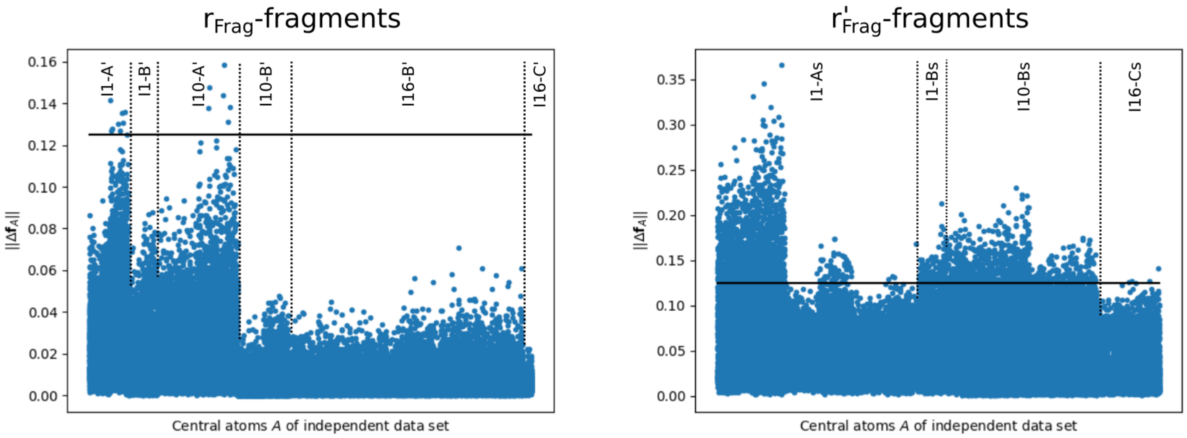

Figure 4.25: Norm of the force error $||\Delta\mathbf{f}_A||$ for all central atoms of the fragment between the true bulk and fragment force of the r_{frag}- (fig. 4.22) and r'_{frag}-fragments (fig. 4.24). This data sets are independent from HDNNP training as introduced in section 4.9.1. The predefined force criterion $||\Delta\mathbf{f}^{\mathrm{max}}|| \leq 0.125\,\text{eV Å}$ is indicated by the black line.

The atomic forces of the central atoms are verified by a bulk IRMOF data set being introduced in section 4.9.1. Based on these bulk structures, r_{frag}- and r'_{frag}-fragments are constructed and the atomic fragment forces are compared to the related bulk forces (fig. 4.25). The accuracy of the atomic fragment forces is different for the diverse fragments, which is also related to the construction of the underlying bulk data set. Nevertheless, r_{frag}-fragments describe their central atoms rather accurately with force error norms similar to the pre-defined force criterion $||\Delta\mathbf{f}_A|| \approx ||\Delta\mathbf{f}^{\max}|| \leq 0.125\,\mathrm{eV}\,\text{Å}$ in the analysis of the Hessian (sec. 4.9). Contrary, the r'_{frag}-fragments show large deviations of their related central atom forces with significantly large deviations to the bulk forces $||\Delta\mathbf{f}_A|| \approx 3 \times ||\Delta\mathbf{f}^{\max}|| \leq 0.125\,\mathrm{eV}\,\text{Å}$.

4.8 Training of the HDNNPs

For the r'_{frag} based fragments (fig .4.24) an initial training set HDNNP-1′ is created by *ab initio* MD simulations of the *FHI-aims* program package [101] at a moderate temperature of 600 K, resulting in 3498 structures (I1-As: 682, I1-Bs: 908, I10-Bs: 908, I16-Cs: 1000) or in terms of the HDNNP training points. The neural network architecture is selected as $15 \times 15 \times 1$ with two hidden layers containing 15 nodes each and an one-node output layer. For the activation function the hyperbolic tangent is chosen for the hidden layers and a linear activation function for the output layer, respectively. Exact details are summarized in sections sec. 3.3 and A.7.

4.8.1 r'_{frag}-Fragments Based HDNNP

For the ACSFs, the HDNNP cutoff radius is defined as $r_{\mathrm{cut}} = 4.359\,\text{Å}$ as already mentioned above. For each element combination five radial ACSFs are defined, while for the angular ACSFs eight different sets of parameters for chosen per element combination. To reduce the amount of ACSFs for certain element combinations, the DFT Hessian data is analyzed demonstrating no significant interactions between zinc and hydrogen atoms (fig. A.9, A.15, A.16, A.24, A.25, A.26, A.36, A.37 and A.38), only less significant hydrogen-hydrogen interactions (fig. A.15, A.24, A.25, A.36, A.37 and A.38) and hydrogen-oxygen interactions (fig. A.10, A.11, A.15, A.17, A.18, A.24, A.25, A.27, A.28, A.36, A.37 and A.38), resulting in the set of 20 radial ACSFs for carbon, 18 for oxygen, 15 for zinc and eleven for hydrogen atoms as mentioned in table A.12 with the cutoff radius $r_{\mathrm{cut}} = 4.359\,\text{Å}= 8.237\,\mathrm{a}_0$, the shifting parameter $r_{\mathrm{shift}} = 0.000\,\text{Å}= 0.000\,\mathrm{a}_0$ and the inner cutoff radius $r_{\mathrm{inner,cut}} = 0.000\,\text{Å}= 0.000\,\mathrm{a}_0$. For the angular ACSF, the element combinations as summarized in table A.13 are expanded by all different combinations of the parameter $\zeta \in \{1, 2, 4, 16\}$, defining the width of the cosine part and the parameter $\lambda \in \{-1, 1\}$ inverting the cosine part. For the element combinations *C-Zn-Zn*, *Zn-C-C*, *Zn-C-Zn*, *Zn-O-Zn* and *Zn-Zn-Zn* the angular ACSF with the parameter combination of $\eta/\lambda = -1/16$ are neglected, since these do not provide any input information for the HDNNP and the underlying data set. A first HDNNP represents the initial data with a high accuracy (tab. 4.12), low errors for the total potential energy of the training $\mathrm{RMSE}(E_{\mathrm{tot}}^{\mathrm{train}}) = 0.0009\,\mathrm{eVatom}^{-1}$ and test data set $\mathrm{RMSE}(E_{\mathrm{tot}}^{\mathrm{test}}) = 0.0008\,\mathrm{eVatom}^{-1}$, as well as acceptable errors for the force components of the training $\mathrm{RMSE}(f^{\mathrm{train}}) = 0.1450\,\mathrm{eV}\,\text{Å}^{-1}$ and test data set $\mathrm{RMSE}(f^{\mathrm{test}}) = 0.1450\,\mathrm{eV}\,\text{Å}^{-1}$, respectively. Also the individual errors of the fragments I1-As, I1-Bs and I10-Bs are acceptably small, which underlines the accurate representation of the data set, although the errors for the I16-Cs fragment are increased but still acceptable. Based on the initial HDNNP fit and data set of r'_{frag}-1, the data set is iteratively expanded to sample the important regions of the configurational space. More details are given in the sections 3.4. Finally the data set r'_{frag}-2 of 13220 data points is observed, containing 3092 structures of I1-As, 3049 of I1-Bs, 4000 of I10-Bs and 3079 of I16-Cs, which is represented by the same HDNNP architecture and ACSF as the initial data set, resulting in similar RMSEs of the total data set and for the individual fragment structures (tab. 4.12).

4.8.2 r_{frag}-Fragments Based HDNNP

Based on the initial HDNNP fit r'_{frag}-1-SF1 (tab. 4.12), a data set for the r_{frag}-fragments (fig.4.22) is created by MD simulations (sec. 3.4). For the ACSF also a cutoff radius of $r_{\mathrm{cut}} = 4.359\,\text{Å}$ can be defined with the same ACSFs as summarized in tables A.12 and A.13. However, the atomic environment of the central atoms within these fragments is equivalent to the bulk environment up to a radius $r_{\mathrm{frag}} = 8.718\,\text{Å}$, which defines the

Table 4.12: *Root-mean squared error* (RMSE) for the total potential energy of the training ($E_{\mathrm{tot}}^{\mathrm{train}}$) and test data set ($E_{\mathrm{tot}}^{\mathrm{test}}$) in eV atom^{-1} and for the force components of the training (f^{train}) and test data set (f^{test}) in eV Å^{-1} summarized for the initial r'_{frag}-1-SF1 and the final data set r'_{frag}-2-SF1, which are based on the r'_{frag}-fragments as shown in figure 4.24, on the atom-centered symmetry functions summarized in tables A.12 and A.13 and a 15 × 15 × 1 NNP architecture (details given in sec. 3.3 and A.7). For the individual fragments I1-As, I1-Bs, I10-Bs and I16-Cs, the individual RMSE values are shown, resulting from the parameters fitted for the complete data set presented by r'_{frag}-1-SF1 and r'_{frag}-2-SF1.

HDNNP/	RMSE		RMSE		data
data set	$E_{\mathrm{tot}}^{\mathrm{train}}$	$E_{\mathrm{tot}}^{\mathrm{test}}$	f^{train}	f^{test}	points
r'_{frag}-1-SF1	0.0012	0.0013	0.1436	0.1453	3498
I1-As	0.0009	0.0008	0.1450	0.1450	682
I1-Bs	0.0009	0.0008	0.1483	0.1472	908
I10-Bs	0.0008	0.0008	0.1224	0.1229	908
I16-Cs	0.0018	0.0016	0.1626	0.1638	1000
r'_{frag}-2-SF1	0.0014	0.0015	0.1267	0.1295	13220
I1-As	0.0013	0.0013	0.1291	0.1315	4370
I1-Bs	0.0011	0.0010	0.1167	0.1181	2860
I10-Bs	0.0013	0.0012	0.1212	0.1173	4314
I16-Cs	0.0022	0.0021	0.1571	0.1572	1676

upper boundary for the ACSF cutoff radius $r_{\mathrm{cut}}^{\mathrm{max}}$. Because of the atomic force dependency on twice the cutoff radius $\mathbf{f}_A = f(2r_{\mathrm{cut}})$, the HDNNP predicts atomic forces dependent on $2r_{\mathrm{cut}} = 17.436$ Å environment. Consequently, the increased cutoff radius and thus, the inherently increased information of the local atomic environment input to the HDNNP should in principle increase the accuracy of the predicted results, apart from technical issues like the increased configurational space to sample. For a set of ACSFs SF-2, providing a similar radial resolution as SF-1 (tab. A.12 and A.13), the increased number of radial ACSFs is inherently related to the increased the cutoff radius $r_{\mathrm{cut}} = 8.718$ Å. Thus, 17 radial ACSF are defined for each element combination, while again zinc-hydrogen interactions are neglected, hydrogen-hydrogen and hydrogen-oxygen ACSFs are reduced resulting in 68 radial ACSF for carbon, 54 for oxygen, 51 for zinc and 23 for hydrogen (tab. A.14 and A.15) with the cutoff radius $r_{\mathrm{cut}} = 8.718$ Å$= 16.475\,\mathrm{a}_0$, the shifting parameter $r_{\mathrm{shift}} = 0.000$ Å$= 0.000\,\mathrm{a}_0$ and the inner cutoff radius $r_{\mathrm{inner,cut}} = 0.000$ Å$= 0.000\,\mathrm{a}_0$. For the angular ACSF, the element combinations as summarized in table A.16 are expanded in analogy to SF-1. For the element combinations *C-Zn-Zn*, *O-O-O*, *Zn-C-C*, *Zn-C-Zn*, *Zn-O-Zn*, *Zn-Zn-Zn* and *H-O-O* the angular ACSF with the parameter combination of $\eta/\lambda = -1/16$ are neglected, since these do not provide any input information for the HDNNP for the underlying data set. Based on the initial HDNNP fit and data set, the data set is iteratively expanded independently from the data based on r'_{frag}, to sample the important regions of the configurational space. More details are given in the sections 3.4. Finally the data set r_{frag}-2-SF2 of 13503 data points is observed, containing 1279 structures of I1-A', 3271 of I1-B', 1240 of I10-A', 4314 of I10-B', 1938 of I16-B' and 1461 of I16-C', resulting in RMSE values as the HDNNP r'_{frag}-2-SF1 (tab. 4.12 and 4.13).

Table 4.13: *Root-mean squared error* (RMSE) for the total potential energy of the training ($E_{\text{tot}}^{\text{train}}$) and test data set ($E_{\text{tot}}^{\text{test}}$) in eVatom^{-1} and for the force components of the training (f^{train}) and test data set (f^{test}) in eVÅ^{-1} summarized for the initial r_{frag}-1-SF2 and the final data set r_{frag}-2-SF2, which are based on the r_{frag}-fragments as shown in figure 4.22, on the ACSF summarized in tables A.14, A.15 and A.16 and a 15 × 15 × 1 NNP architecture (details given in sec. 3.3 and A.7). Additionally, the errors are summarized for the final data set r_{frag}-2-SF1, which based on the ACSF summarized in tables A.12 and A.13. For the individual fragments I1-A', I1-B', I10-B' and I16-C', the individual RMSE values are shown, resulting from the parameters fitted for the complete data set presented by r_{frag}-1-SF2 and r'_{frag}-2-SF2.

HDNNP/	RMSE		RMSE		data
data set	($E_{\text{tot}}^{\text{train}}$)	($E_{\text{tot}}^{\text{test}}$)	(f^{train})	(f^{test})	points
r_{frag}-2-SF2	0.0011	0.0012	0.1654	0.1636	13 503
I1-A'	0.0007	0.0007	0.1323	0.1314	1 279
I1-B'	0.0007	0.0008	0.1351	0.1337	3 271
I10-A'	0.0010	0.0010	0.1666	0.1585	1 240
I10-B'	0.0009	0.0009	0.1606	0.1614	4 314
I16-B'	0.0010	0.0010	0.2014	0.1974	1 938
I16-C'	0.0023	0.0024	0.2268	0.2233	1 461
r_{frag}-2-SF1	0.0012	0.0011	0.1448	0.1441	13 503
I1-A'	0.0008	0.0007	0.1174	0.1150	1 279
I1-B'	0.0008	0.0008	0.1225	0.1204	3 271
I10-A'	0.0009	0.0011	0.1422	0.1425	1 240
I10-B'	0.0009	0.0009	0.1408	0.1436	4 314
I16-B'	0.0010	0.0010	0.1720	0.1733	1 938
I16-C'	0.0025	0.0021	0.2047	0.1946	1 461

4.9 Application of the HDNNPs

The resulting HDNNPs r'_{frag}-2-SF1, r_{frag}-2-SF1 and r_{frag}-2-SF2 presented in the section above are used to predict properties for the periodic bulk IRMOF structures (fig. 4.1) and for the different molecular fragment structures based on r_{frag} (fig. 4.22) and r'_{frag} (fig. 4.24), which are derived by the DFT Hessian results (sec. 4.6 and related). The underlying data sets are independent from the HDNNP training data sets.

4.9.1 Energies and Forces of the Fragments and the Bulk Structures

For each of the threeIRMOF structures, an independent data set of 502 structures/data points is constructed based on two independent MD simulations within the NPT ensemble at normal pressure and at temperatures of 200 K and 500 K for IRMOF-1, 200 K and 450 K for IRMOF-10, as well as 200 K and 350 K for IRMOF-16, which are generated by the HDNNP r'_{frag}-2-SF1. These temperatures are chosen to ensure the MD trajectory stays in a reasonable region of the configurational space to avoid extrapolation warning for the ACSFs marking less accurate sampled regions of the configurational space resulting in unreasonable energies and forces and this unreasonable structures of the IRMOFs. More details are given in section 3.4. From each MD trajectory 251 structures separated by 2 ps ($2000\,\mathrm{TS} \times 1\,\mathrm{fs}\,\mathrm{TS}^{-1}$) are combined for the IRMOFs to result in the HDNNP training independent bulk data set.

Bulk Structure Predictions

Predicted energies and forces of the two HDNNPs r'_{frag}-2-SF1 and r_{frag}-2-SF1 for IRMOF-1 are summarized in figure 4.26, for IRMOF-10 in figure A.40 and for IRMOF-16 in figure A.41. The structures of the two independent MD trajectories at different temperatures can clearly be separated by the two starting structures as stated by the data points 1 and 252 (panel a) of fig. 4.26, A.40 and A.41). The atomic energy differences ΔE of the HDNNP predictions and the DFT references are properly small (panel a) of fig. 4.26, A.40 and A.41; $\Delta E \lesssim 0.0025\,\mathrm{eV}$) and comparable to the RMSE of the energies of the HDNNPs (tab. 4.12 and 4.13). Also for the forces, the error RMSE(f) as stated by the RMSE of the force components for each data point are in an acceptable order of magnitude (panel c) of fig. 4.26, A.40 and A.41; RMSE(f) $\lesssim 0.16\,\mathrm{eV\,\AA^{-1}}$) as expected from the HDNNP RMSE of the force components. Furthermore, there no significant quantitative differences between the two HDNNP predictions. In general, with increase of the temperature of the MD trajectory and thus the kinetic energy of the atoms, also the errors increase, because this additional kinetic energy leads to regions of the configurational space, which are not sampled properly.

r'_{frag}-Fragments Predictions

Based on the generated bulk data mentioned above, r'_{frag}-fragment structures are constructed for each of the 502 data points to result in HDNNP training independent data sets of the fragment structures. For the I1-As fragment, representing – among other central atoms – the in-equivalent site Zn1 in the three IRMOF bulk structures, a data set of 1506 data points are combined from 251 data points of the two MD trajectories for each of the three IRMOFs. The data set (502 data points) for the I1-Bs fragment structures is combined from 251 data points of the two IRMOF-1 trajectories; the I10-Bs data set (1004 data points) is combined from 251 data points of the IRMOF-10 and -16 trajectories; the I16-Cs data set (502 data points) is combined from 251 data points of the two IRMOF-16 trajectories. The predicted energies and forces for the data set of I1-As are summarized in

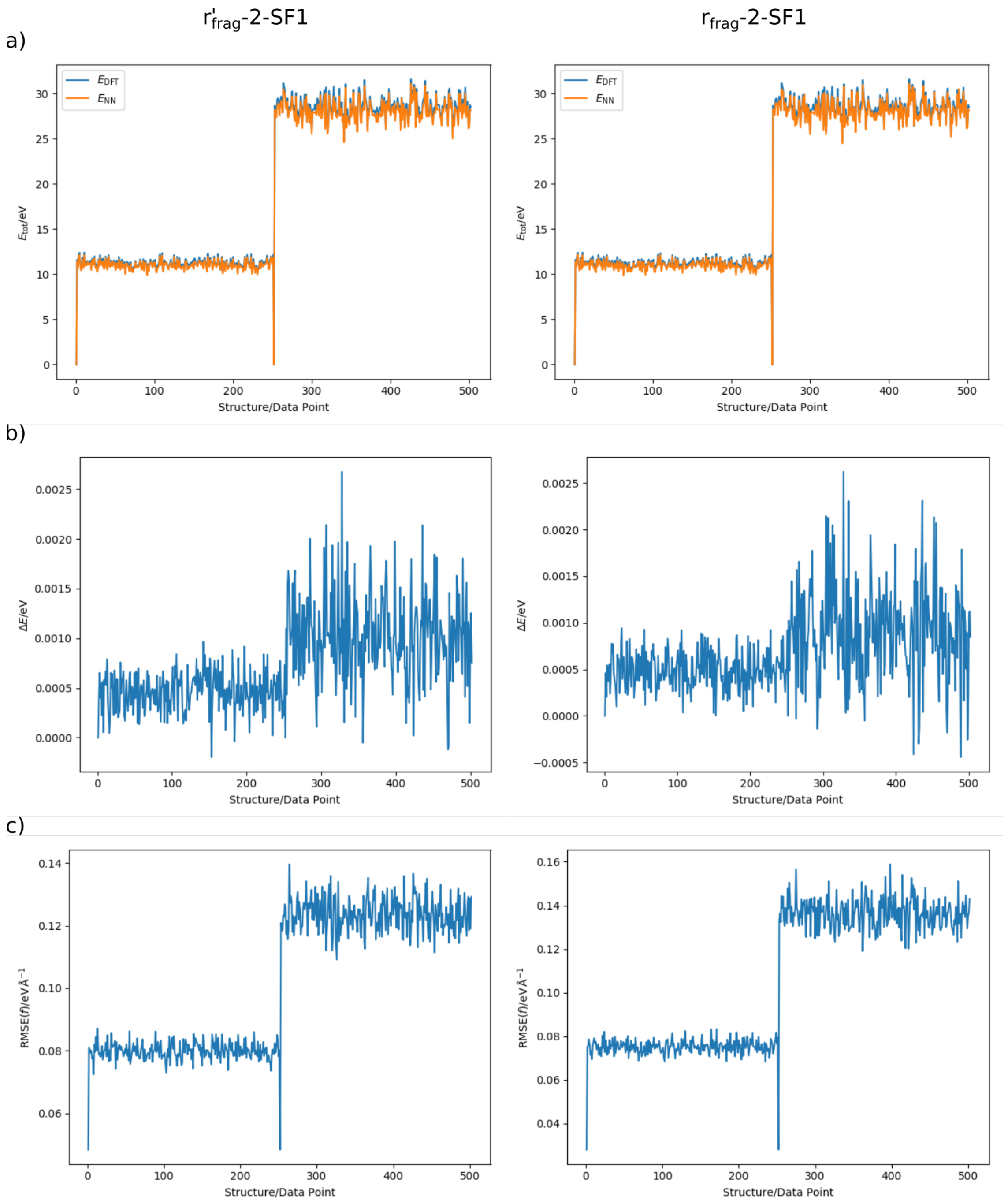

Figure 4.26: Compilation of the IRMOF-1 bulk predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) of a HDNNP training independent data set (502 structures/data points), combined of 251 data points for each of two MD simulations in the NPT ensemble (sec. 3.4) at normal pressure and temperatures $T \in \{200, 500\}$ in K. a) Demonstrates the total potential energy of the IRMOF-1 bulk structure over the data set. The two independent MD simulations can be separated by the two starting points of the simulations (data point 1 and 252). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

figure 4.27, for I1-Bs in figure A.42, for I10-Bs in figure A.40 and for I16-Cs in figure A.44. The atomic energy differences are low in value (panel b) of fig. 4.27, A.42, A.40 and A.44; $\Delta E \lesssim 0.004\,\mathrm{eV}$) similar to the RMSE of the force components per data point (panel c) of fig. 4.27, A.42, A.40 and A.44; $\mathrm{RMSE}(f) \lesssim 0.010\,\mathrm{eV\,Å^{-1}}$) and again comparable to the RMSE values of the HDNNP training (tab. 4.13 and 4.13). Also here, the errors increase with the temperature of the underlying MD trajectory, due to the increased kinetic energy. For the I16-Cs fragment data set, the errors are marginally increased ($\Delta E \lesssim 0.006\,\mathrm{eV}$ and $\mathrm{RMSE}(f) \lesssim 0.250\,\mathrm{eV\,Å^{-1}}$), which is in line with the increased RMSE values of the HDNNP training data set of I16-Cs. Furthermore, the two differentHDNNPs does not predict significantly different values.

r_{frag}-Fragments Predictions

For each of the r_{frag}-fragments also an independent data set is constructed based on the specific bulk MD trajectories, including 502 data points (251 from each trajectory). The results for I1-A′ are summarized in figure 4.28, for I1-B′ in figure A.45, for I10-A′ in figure A.46, for I10-B′ in figure A.47, for I16-B′ in figure A.48 and for I16-C′ in figure A.49. The errors are similar to the r_{frag}- fragments presented above and in accurate agreement with the training RMSEs. Some outliers arise e. g. for I1-A′ resulting in large force errors ($\mathrm{RMSE}(f) \approx 0.300\,\mathrm{eV\,Å^{-1}}$ at data point   400), demonstrating a fragment structure for which the configurational space is less accurately sampled resulting in this large force error. Nevertheless, equivalent to the results of the bulk and r'_{frag}-fragment structures, both HDNNPs illustrate similar results and no significant differences for the forces and energies of the r_{frag}-fragment structures.

4.9.2 The Lattice Parameter of the IRMOF Structures

To determine the equilibrium lattice parameter a the data set presented in section 4.1 is used. The predicted energies and forces of the HDNNPs are summarized for IRMOF-1 in figure 4.29, for IRMOF-10 in figure A.50 and for IRMOF-16 in figure A.51, respectively. The predictions around the equilibrium structure (data point 6) are accurate, as this region of the configurational space is sampled with high resolution by the active learning scheme (sec. 3.4). Larger deviations of the energy predictions ($\Delta E \gtrsim -0.005\,\mathrm{eV}$) occur for the compressed structures and for the expanded structures ($\Delta E \lesssim 0.015\,\mathrm{eV}$) predicted by r'_{frag}-2-SF1, whereas r_{frag}-2-SF1 predicts more accurate energies for the expanded structures ($\Delta E \lesssim 0.002\,\mathrm{eV}$). The same behaviour is observed for the error of the forces as in the equilibrium regions the force predictions are accurate ($\mathrm{RMSE}(f) \lesssim 0.060\,\mathrm{eV\,Å^{-1}}$) compared to the expanded and compressed structures ($\mathrm{RMSE}(f) \lesssim 0.200\,\mathrm{eV\,Å^{-1}}$). The resulting lattice parameters are in accurate agreement with the DFT results (tab. 4.14) with an absolute deviation of the lattice parameter $|\Delta a| \lesssim 0.050\,\mathrm{Å}$ and insignificantly relative deviations $|\Delta' a| \lesssim 0.0011$.

4.9.3 Phenylene Rotations within the IRMOF Structures

To analyze the rotational barrier so-called dumbbell fragments of the IRMOFs are constructed (fig. 4.30) including a linker of the specific IRMOF being embedded in the SBU bulk environment. For sure, periodic bulk effects are not considered in these dumbbell model fragments, but the computational effort is decreased drastically compared to the analysis of the phenylene rotations within the periodic bulk structures. Nevertheless, in this comparison the goal is to find different and common features of the predictions of two HDNNP with a conceptional different approach for the data set. An accurate description of the IRMOFs rotational barriers is not focused. The data sets for the rotational

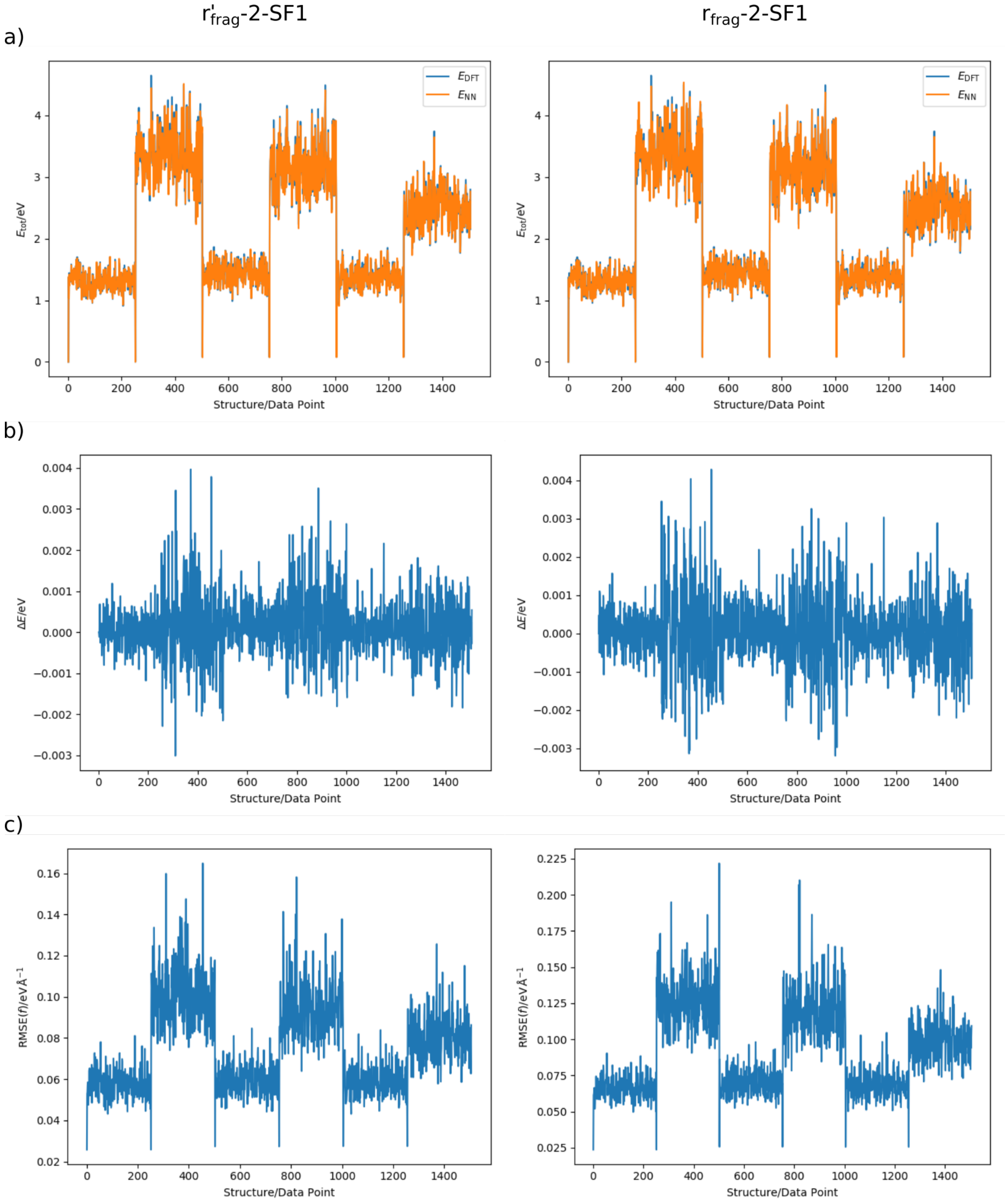

Figure 4.27: Compilation of the I1-As fragment (fig. 4.24) predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) based on the HDNNP training independent data sets for the three individual IRMOFs (in total 1506 I1-As fragment structures/data points, 502 for the individual IRMOFs). a) Demonstrates the total potential energy of the I1-As fragment structure over the data set. The individual MD simulations can be separated by the starting points of the simulations (data point 1, 252, 503, 754, 1005 and 1256). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

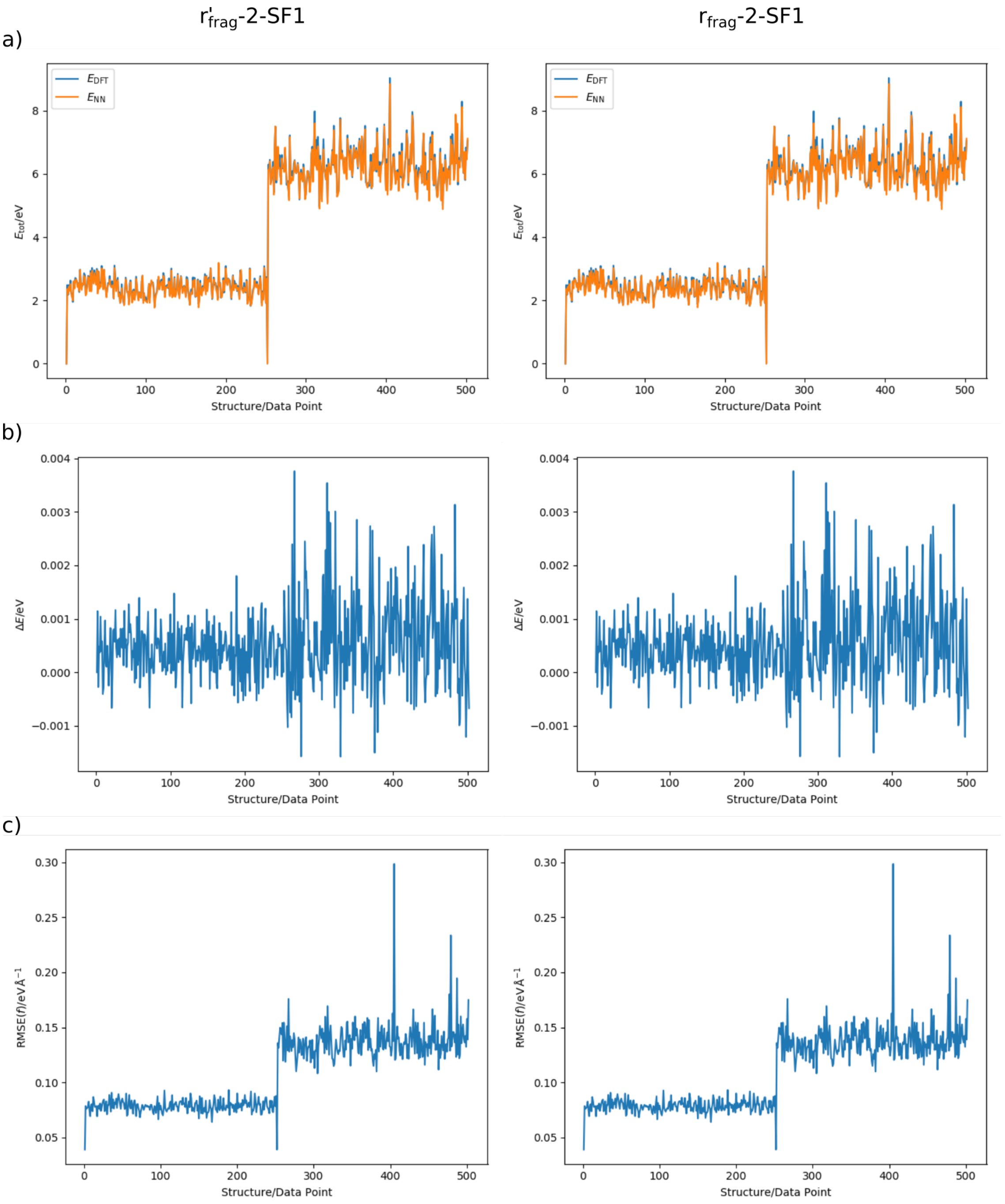

Figure 4.28: Compilation of the I1-A' fragment (fig. 4.22) predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) based on the HDNNP training independent data sets for IRMOF-1 (in total 502 I1-A' fragment structures/data points). a) Demonstrates the total potential energy of the I1-A' fragment structure over the data set. The individual MD simulations can be separated by the starting points of the simulations (data point 1 and 252). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

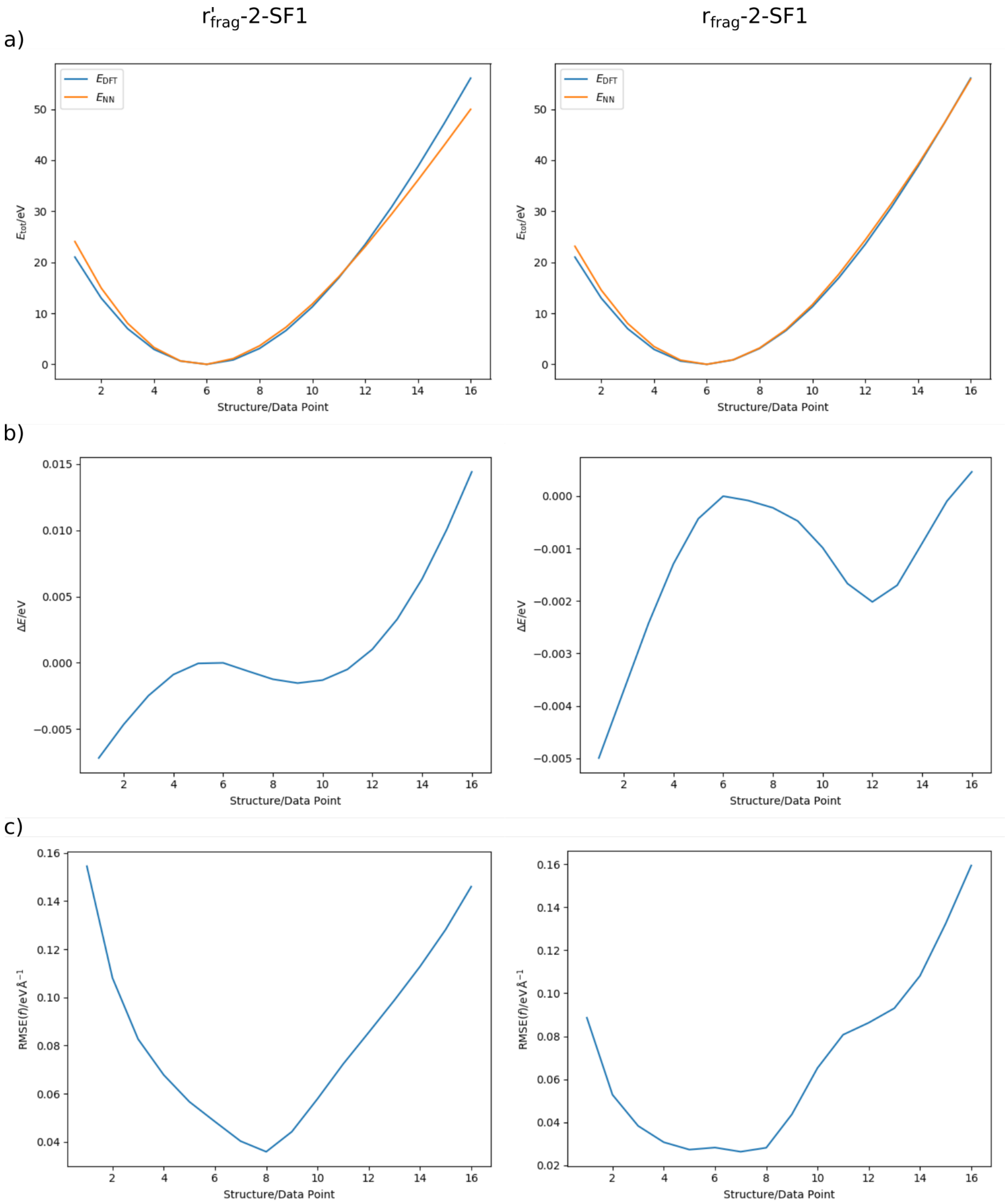

Figure 4.29: Compilation of the IRMOF-1 bulk predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) of a HDNNP training independent data set (16 structures/data points), based on expanded and compressed bulk structures by a scaling factor $\sigma \in \{0.95 - 1.10\}$ in steps of 0.01 with DFT optimized atomic positions. a) Demonstrates the total potential energy of the IRMOF-1 bulk structure over the data set, b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

Table 4.14: Compilation of the DFT and Birch-Murnaghan (BM) EOS equilibrium lattice parameter a, the absolute deviation $\Delta a = a^{\mathrm{DFT}} - a$ compared to the DFT lattice parameter in Å and the relative deviation $\Delta' a = 1 - \frac{a}{a^{\mathrm{DFT}}}$ predicted by DFT and the HDNNPs r'_{frag}-2-SF1, r_{frag}-2-SF1 and r_{frag}-2-SF2 BM EOS fit.

		BM	BM	BM	BM
	DFT	DFT	r'_{frag}-2-SF1	r_{frag}-2-SF1	r_{frag}-2-SF2
$a_{\mathrm{IRMOF-1}}$	26.296	26.289	26.293	26.298	26.382
$a_{\mathrm{IRMOF-10}}$	35.063	35.061	35.074	35.094	34.973
$a_{\mathrm{IRMOF-16}}$	43.832	43.832	43.835	43.882	43.742
$\Delta a_{\mathrm{IRMOF-1}}$	0.000	0.007	0.003	-0.002	-0.0857
$\Delta a_{\mathrm{IRMOF-10}}$	0.000	0.002	-0.011	-0.031	0.0896
$\Delta a_{\mathrm{IRMOF-16}}$	0.000	0.000	-0.004	-0.050	0.0894
$\Delta' a_{\mathrm{IRMOF-1}}$	0.0000	0.0003	0.0001	-0.0001	-0.0033
$\Delta' a_{\mathrm{IRMOF-10}}$	0.0000	0.0000	-0.0003	-0.0009	0.0026
$\Delta' a_{\mathrm{IRMOF-16}}$	0.0000	0.0000	-0.0001	-0.0011	0.0020

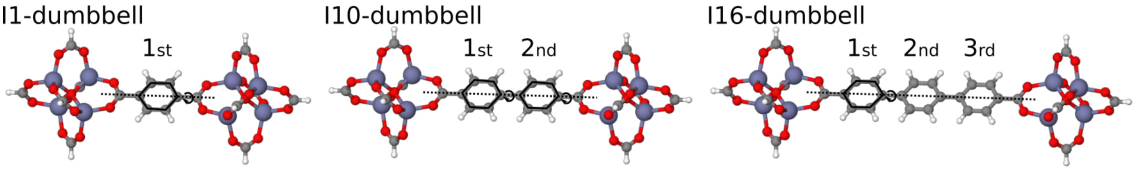

Figure 4.30: IRMOF-dumbbell fragments used for the analysis of the phenylene ring rotations with the two-fold rotational axis (dashed line), the phenylene rings being rotated (solid line and rotational arrow) and the labeling of the phenylene rings.

barriers are based on the DFT equilibrium structure (data point 1) combined with 36 structures of a full rotation for the phenylene rings in steps of 10 ° without any further relaxation steps. Contrary to the one-dimensional analysis of IRMOF-1 and -16 (37 data points in the data set), the phenylene-ring rotation of IRMOF-10 is analyzed in its two dimensional manner resulting in 1369 data points (37 orientations for the first and the second phenylene ring). The symmetry of the phenylene ring is not explicitly considered here, although the symmetry can reduce the computational effort by reducing the total number of structures necessary to map the PES related to the phenylene ring rotations, since the phenylene ring includes a two-fold rotational axis, thus rotations beyond 180 ° are rotationally equivalent. The DFT reference data illustrate the minimum of the phenylene ring rotation for IRMOF-1 in the co-planar orientation of the phenylene ring and the two carboxyl groups of the embedding SBUs as expected [53]. For IRMOF-10 the global minimum of the rotation in not the co-planar orientation as assumed in section 4.1 and in the literature [53], although in this orientation the overlap of the p-orbitals portray the most bonding character. Also the interactions of the hydrogen atoms of the neighboring phenylene rings demonstrate the most repulsive character in this orientation. Thus, an interplay of both effects shifts the minimum for the phenylene ring orientation out of the co-planar structure to a slightly twisted orientation of the phenylene rings and the carboxyl groups, i. e. the first phenylene ring in an orientation 160 ° and the second phenylene ring in an orientation 200 ° and symmetry related orientations, respectively. This orientation results in a 40 ° dihedral angle between the neighboring phenylene rings, being perfectly in line with the dihedral angle of 38 ° within the structure of a biphenyl molecule [110]. For IRMOF-16, the global minimum fo the phenylene ring orientation is also not located at the co-planar orientation of the phenylene rings and the carboxyl groups of the embedding SBUs for the same reasons as pronounced for IRMOF-10, being completely in line with the structure from [53]. The results of the one-dimensional IRMOF-1 phenylene-ring rotation are summarized in figure A.53; for IRMOF-16 only the first phenylene-ring rotation is analyzed (fig. A.54) exemplary for its three dimensional phenylene-ring rotation problem and the results of IRMOF-10 are summarized in figure 4.31. The equilibrium structures around data point 1 (and symmetry related 19) of IRMOF-1 and -16, as well as the 0 °-orientation of both phenylene rings for IRMOF-10, respectively, are again described accurately by both HDNNPs. In contrast to the less accurate description of the high energy structures, thus 90 ° rotation of the phenylene rings for IRMOF-1 and -16, as well as the 90 ° co-planar rotation of both phenylene rings, respectively. Also the turning points of the rotational barriers around 60 ° rotation – for IRMOF-10 the co-planar 60 ° rotation, respectively – demonstrate large deviations in energy from the reference data. In general, the HDNNP r'_{frag}-2-SF1 predicts more accurate energies for IRMOF-10 ($|\Delta E| \lesssim 0.04\,\mathrm{eV}$) and -16 ($|\Delta E| \lesssim 0.02\,\mathrm{eV}$), but not for IRMOF-1 ($|\Delta E| \lesssim 0.05\,\mathrm{eV}$), which is describe more precisely by r_{frag}-2-SF1 ($|\Delta E| \lesssim 0.035\,\mathrm{eV}$). However, the HDNNP r_{frag}-2-SF1 predicts larger energy deviations for IRMOF-10 and -16 ($|\Delta E| \lesssim 0.04\,\mathrm{eV}$). The forces are comparable to the HDNNP RMSE values (tab. 4.12 and4.13) with in general smaller force errors for r'_{frag}-2-SF1 compared to r_{frag}-2-SF1.

4.9.4 Application of HDNNP r_{frag}-2-SF2

The predicted results for the HDNNP r_{frag}-2-SF2 are qualitatively comparable to the results predicted by r'_{frag}-2-SF1 and r_{frag}-2-SF1. Force and energy predictions for the bulk structures (fig. A.55) and the related r_{frag}-based fragments (fig. A.56 and A.57) of the HDNNP training independent data sets are similar in all three HDNNPs, but the force errors are slightly increased for r_{frag}-2-SF2. For the r'_{frag}-based fragments, no predictions are performed, since for these structures no central atom is embedded in a bulk-like environment because of the enlarged cutoff radius $r_{\mathrm{cut}} = 8.718\,\text{Å}$, while the fragment radius is smaller $r_{\mathrm{frag}} = 4.359\,\text{Å}$. Also the predictions for the lattice parameter determination (fig. A.58) illustrate increased force errors resulting in larger deviations for the determined lattice parameter a (tab 4.14). Furthermore, for the phenylene rotations (fig. A.59) the errors are increased of the HDNNP r_{frag}-2-SF2 predictions.

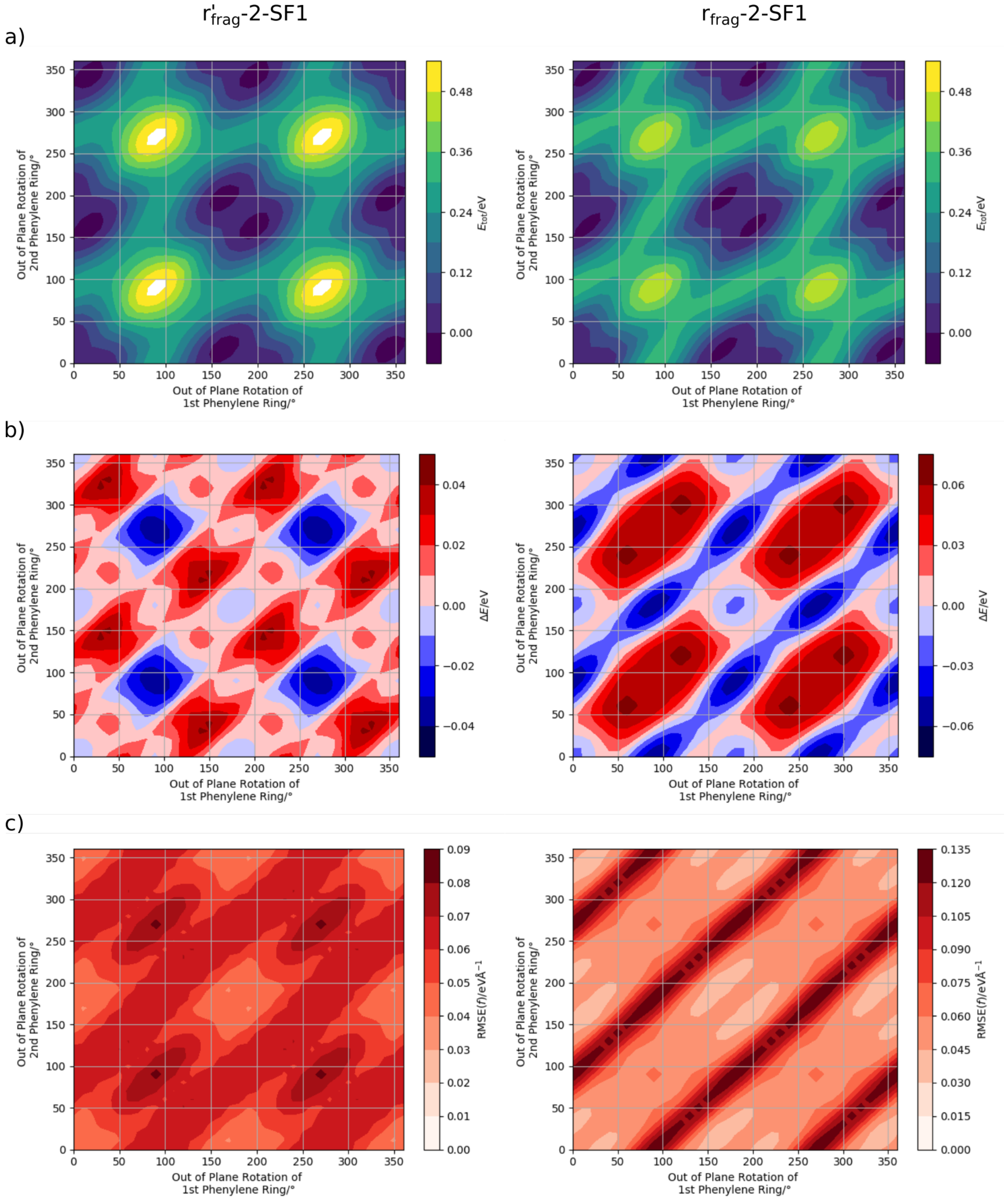

Figure 4.31: Compilation of the IRMOF-10 dumbbell model fragment predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) of a HDNNP training independent data set (1369 structures/data points), based on the rotation of the phenylene rings in steps of 10 °. a) Demonstrates the total potential energy predicted by the HDNNP of the IRMOF-10 dumbbell model fragment structure over the data set (DFT reference total potential energy given in figure A.52), b) The total energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

4.9.5 Suggestion for Efficient Construction of HDNNPs

In summary, the predictions of the HDNNP r'_{frag}-2-SF1 and r_{frag}-2-SF1 are similar to each other only with slight deviations of the quantitative description. However, the predictions for the HDNNP r_{frag}-2-SF2 illustrates qualitatively the same behavior as observed for r'_{frag}-2-SF1 and r_{frag}-2-SF1, but the RMSE values are increased. The increased cutoff radius $r_{\mathrm{cut}} = 8.718\,\text{Å}$ increases the configurational space, which needs to be sampled with an increased amount of data points. Nevertheless, this HDNNP and its results are not further interest.

Although, the data set r_{frag}-2 in combination with the cutoff radius $r_{\mathrm{cut}} = 4.359\,\text{Å}$ contains approximately five times the amount of bulk-like atomic environments of the data set r_{frag}-2 (tab. 4.15), the HDNNP r_{frag}-2-SF1 predicts similar errors as r'_{frag}-2-SF1. This behaviour is counter intuitive, but is not related to differences of the training data sets, since both data sets contain similar structural information as demonstrated by the ACSF averages and the ranges spanned by the ACSFs (fig. 4.32). Nevertheless, deviations and differences of the ACSF averages and ranges are related to the different types of HDNNP training fragments (fig. 4.22 and 4.24). However, the independently and iteratively extended data sets cover a similar region of the configurational space. Indeed, these two data sets and thus the similar sampling of the configurational space are perfectly in-line with the predictions of the two HDNNPs $r_{\mathrm{frag}}-2-\mathrm{SF1}$ and $r'_{\mathrm{frag}}-2-\mathrm{SF1}$ presented and discussed above. Intuatively, a larger amount of bulk-like atomic environments of the HDNNP training fragments should be more efficient in sampling the configurational space. In practice, this should only be valid if different symmetry equivalent bulk-like central atoms are embedded in significantly different atomic environments. Hence, the symmetry equivalent bulk-like atomic environments of the central atom A are analyzed within the fragment I16-B′, which is included in the data set $r_{\mathrm{frag}}-2-\mathrm{SF1}$, e. g. each I16-B′ structure contains four Zn1 atoms in a bulk-like environment as summarized in table 4.16.

Table 4.15: Compilation of the bulk-like atomic environments $M_{\mathrm{bulk,data}}$ of the two different data sets $r_{\mathrm{frag}}-2-\mathrm{SF1}$ and $r'_{\mathrm{frag}}-2-\mathrm{SF1}$ calculated by the sum of the multiplication for the bulk-like central atoms M_{bulk} and the amount of fragments N_{frag} included in the data sets.

	M_{bulk}	N_{frag}	$M_{\mathrm{bulk,data}}$
I1-A′	65	1 279	83 135
I1-B′	62	3 271	202 802
I10-A′	101	1 240	125 240
I10-B′	68	4 314	293 352
I16-B′	161	1 938	312 018
I16-C′	24	1 461	35 064
$\Sigma_{r_{\mathrm{frag}}-2-\mathrm{SF1}}$		13 503	1 051 611
I1-As	17	4 370	74 290
I1-Bs	11	2 860	31 460
I10-Bs	17	4 314	73 338
I16-Cs	16	1 676	26 816
$\Sigma_{r'_{\mathrm{frag}}-2-\mathrm{SF1}}$		13 220	205 904

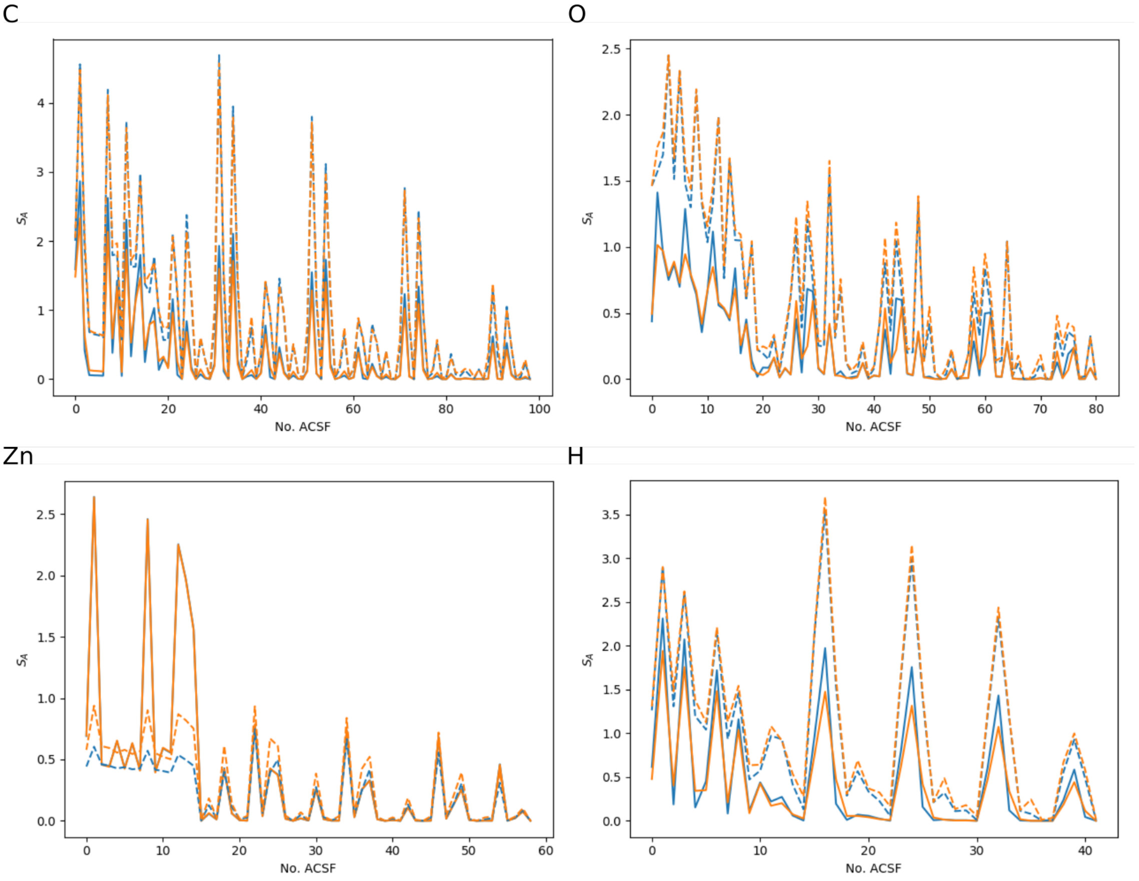

Figure 4.32: Compilation of the element specific (carbon C, oxygen O, zinc Zn and hydrogen H) ACSFs average (solid lines) and the related ACSF ranges (dashed lines) for the training data sets r_{frag}-2-SF1 (blue) and r'_{frag}-2-SF1 (orange).

Table 4.16: Compilation of the N_{frag} fragment structures I16-B′, I1-As and I10-Bs included in the related data set, exhibiting $N_{>}$ structures with a larger norm $||\mathbf{S}'^{\sigma}_{A,\mathrm{frag}}||$ (fig. 4.33 and 4.34) of the range-scaled ACSF standard deviation vector for the $M_{\mathrm{bulk,A}}$ bulk-like central atoms A within each fragment structure compared to the norm $||\mathbf{S}'^{\sigma}_{A,\mathrm{data}}||$ of the range-scaled ACSF standard deviation vector for the bulk-like central atoms A over all structures of this fragment type. Additionally, the counter part $N_{<}$, as well as the ratios $\frac{N_{>}}{N^{\mathrm{frag}}}$ and $\frac{N_{<}}{N^{\mathrm{frag}}}$ are given. Data in bold is also shown in figures 4.33 and 4.34.

	A	$M_{\mathrm{bulk},A}$	N_{frag}	$N_{>}$	$N_{<}$	$\frac{N_{>}}{N_{\mathrm{frag}}}$	$\frac{N_{<}}{N_{\mathrm{frag}}}$	$\lvert\lvert\mathbf{S}'^{\sigma}_{A,\mathrm{data}}\rvert\rvert$
I16-B′	**Zn1**	**4**	**1938**	**189**	**1749**	**0.10**	**0.90**	**0.791**
	O2	12	1938	617	1321	0.32	0.68	0.417
	C1	6	1938	521	1417	0.27	0.73	0.411
	C2	6	1938	471	1467	0.24	0.76	0.213
	C3	**12**	**1938**	**685**	**1253**	**0.35**	**0.65**	**0.181**
	C4	12	1938	475	1463	0.25	0.75	0.242
	C5	12	1938	397	1541	0.20	0.80	0.223
	C6	12	1938	397	1541	0.20	0.80	0.236
	C7	24	1938	479	1459	0.25	0.75	0.264
	H1	**12**	**1938**	**677**	**1261**	**0.35**	**0.65**	**0.260**
	H2	**24**	**1938**	**448**	**1490**	**0.23**	**0.77**	**0.326**
	H3	24	1938	469	1469	0.24	0.76	0.330
	Ø		1938	485.4	1452.6	0.25	0.75	
I1-As	**Zn1**	**1**	**4370**	**0**	**4370**	**0.00**	**1.00**	**1.046**
	O2	6	4370	979	3391	0.22	0.78	0.511
	C1	3	4370	696	3674	0.16	0.84	0.569
	H1	**6**	**4370**	**108**	**3283**	**0.25**	**0.75**	**0.316**
	Ø		4370	690.5	3679.5	0.16	0.84	
I10-Bs	O2	2	4314	358	3956	0.08	0.92	0.419
	C1	1	4314	0	4314	0.00	1.00	0.458
	C2	1	4314	0	4314	0.00	1.00	0.259
	C3	**2**	**4314**	**426**	**3888**	**0.10**	**0.90**	**0.195**
	C4	2	4314	342	3972	0.08	0.92	0.318
	C5	2	4314	14	4300	0.00	1.00	0.271
	H1	2	4314	584	3730	0.14	0.86	0.269
	H2	**4**	**4314**	**424**	**3890**	**0.10**	**0.90**	**0.453**
	Ø		4314	268.5	4045.5	0.06	0.94	

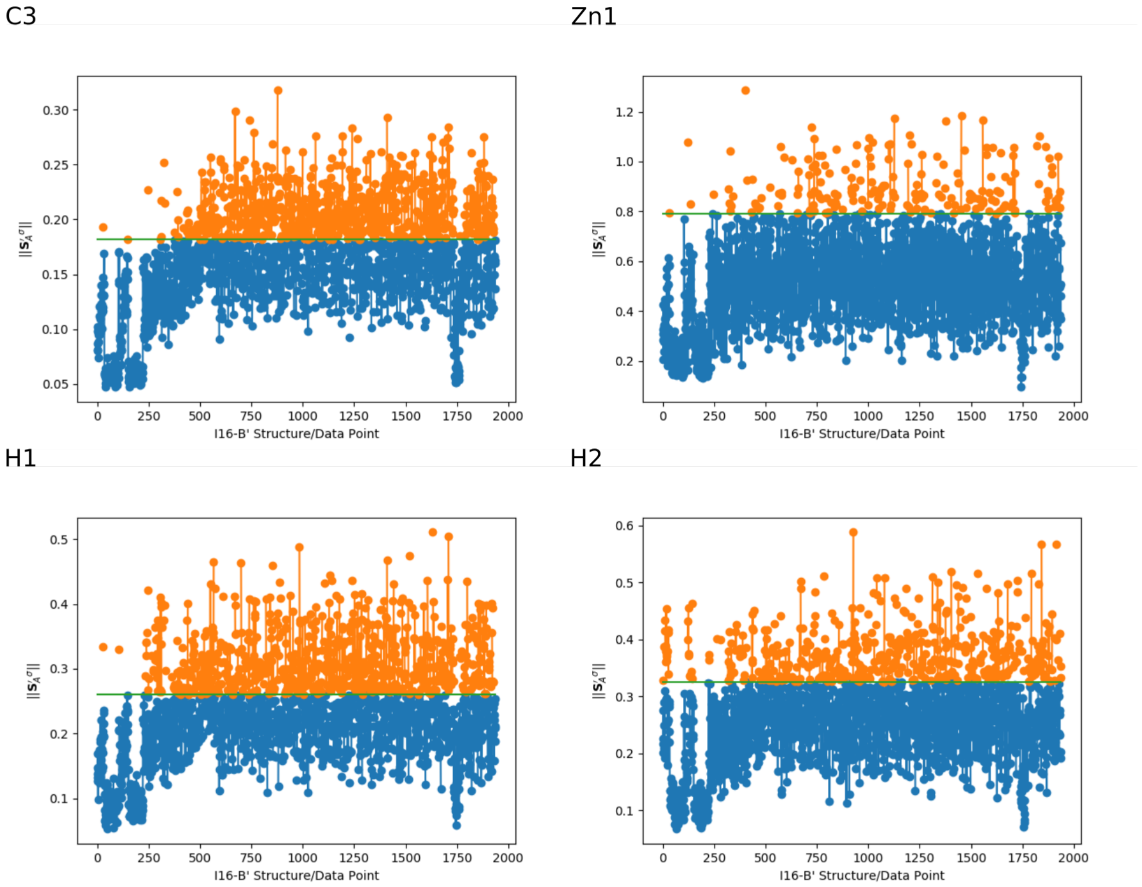

Figure 4.33: Norm $||\mathbf{S}'^{\sigma}_{A,\text{frag}}||$ of the ACSF standard deviation vector for all symmetry equivalent bulk-like central atoms $A \in \{\text{C3}, \text{Zn1}, \text{H1}, \text{H2}\}$ in each fragment structure I16-B′ included in the data set r_{frag}-2-SF1 (blue and orange dots) compared to the norm $||\mathbf{S}'^{\sigma}_{A,\text{data}}||$ of the ACSF standard deviation vector for all symmetry equivalent bulk-like central atoms $A \in \{\text{C3}, \text{Zn1}, \text{H1}, \text{H2}\}$ over all fragments I16-B′ included in the data set r_{frag}-2-SF1 (green line). If $||\mathbf{S}'^{\sigma}_{A,\text{frag}}|| > ||\mathbf{S}'^{\sigma}_{A,\text{data}}||$ the structure is colored in orange, else in blue.

For each I16-B′ structure, an averaged environment for these Zn1 atoms and the related standard deviation vector $\mathbf{S}'^{\sigma}_{A,\text{frag}}$ can be determined based on the ACSF. To balance the numerical effect of different ACSFs, the ACSF values are scaled by their ranges. The norm $||\mathbf{S}'^{\sigma}_{A,\text{frag}}||$ of this standard deviation vector defines a scalar value, which describes the difference of the four Zn1 atoms, since a vanishing norm ($||\mathbf{S}'^{\sigma}_{A,\text{frag}}|| \equiv \vec{O}$) relates to four Zn1 atoms perfectly embedded in the same atomic environment. Furthermore, an averaged environment for these Zn1 atoms and the related standard deviation vector $\mathbf{S}'^{\sigma}_{A,\text{data}}$ can be defined for all bulk-like Zn1 atoms of all 1938 I16-B′ structures (fig. 4.33). The norm of these two standard deviations vectors $||\mathbf{S}'^{\sigma}_{A,\text{frag}}||$ and $||\mathbf{S}'^{\sigma}_{A,\text{data}}||$ are now compared to each other. If

$$||\mathbf{S}'^{\sigma}_{A,\text{frag}}|| > ||\mathbf{S}'^{\sigma}_{A,\text{data}}||, \tag{4.16}$$

the Zn1 atoms of this specific I16-B′ structure differ significantly from each other, since the averaged difference of the atomic environments is larger within this specific I16-B′ structure than the averaged difference within all I16-B′ structures of the whole data set. Vice versa, if

$$||\mathbf{S}'^{\sigma}_{A,\text{frag}}|| < ||\mathbf{S}'^{\sigma}_{A,\text{data}}||, \tag{4.17}$$

the Zn1 atoms are non-significantly different to each other and are embedded in non-significantly different atomic environments, thus the information of the bulk-like Zn1 central atoms, provided as input to the HDNNP training data set by the ACSF vector, is approximately equivalent. Therefore, each I16-B$'$ fragment structure can be classified to one of these categories ($N_>$ and $N_<$). Obviously, only a fraction of 0.10 of the I16-B$'$ fragment structures contain significantly different Zn1 bulk-like central atoms (tab. 4.16). Moreover, only a minor part – an averaged fraction of 0.25 and at maximum 0.35 for the C3 and H1 bulk-like central atoms – of the structures for the remaining central atoms A of fragment I16-B$'$, contain significantly different bulk-like atomic environments. In general, it is not a surprising fact, that differences of symmetry equivalent bulk-like atoms within the same fragment structure are small, in comparison to the differences of the symmetry equivalent bulk-like atoms of two independent fragment structures. As a consequence, not all bulk-like central atoms of a specific fragment are independent from each other. This reduces the efficiency of the larger fragments in terms of sampling different atomic environments, which is indeed essential for the HDNNP training data set. Likewise, the central atoms A of the fragments I1-As and I10-Bs can be analyzed, which demonstrates a similar behaviour of the symmetry equivalent bulk-like central atoms. Hence, the amount of $N_>$-structures, representing a larger norm of the standard deviation vector per fragment, is drastically reduced – for I1-As averaged to a fraction of 0.16 and for I10-Bs to 0.06 (fig. 4.34 and tab. 4.16). In summary, the number of bulk-like central atoms A for a specific fragment structure is of minor importance for the HDNNP training set; the major part of different atomic environments is included via many different and independent structures of a specific fragment. For this reason, the r'_{frag}-fragments are preferred for the construction of a HDNNP data set, because these fragments are computationally less demanding: faster reference and ACSF value calculation, which decreases the effort for the construction of the data set and for the training procedure itself. The HDNNPs based on those fragments predict non-significant deviations compared to the HDNNP based on the r_{frag}-fragments. Furthermore, these fragments are easy to handle for the extension of the data set, since inaccurately described atomic environments can be added to the data set by a minimum amount of atoms (r'_{frag}-fragments) compared to the may redundant atomic environments additionally included in the larger r_{frag}-fragments. In addition, the HDNNP error compensation effect as described by Eckhoff *et al.* [49] is reduced for smaller fragments, in general.

Nevertheless, the increased computational effort for the r_{frag}-based fragments and the required additional ACSFs due to the increased cutoff radius, are not mandatory to yield a HDNNP providing accurate predictions of the bulk properties. A HDNNP based on the r'_{frag}-fragments (fig. 4.24) predicts accurate energies and forces for bulk and fragment structures. Indeed, large errors arise for high energy structures (EOS: fig. 4.29, A.50 and A.51 and the phenylene rotations: fig. 4.31, A.53 and A.54) for which the current data set does not properly sample the high energy regions of the configurational space. Nevertheless, this lack of information can easily, quickly and accurately be added to the data set by the r'_{frag}-fragments to improve the predictions.

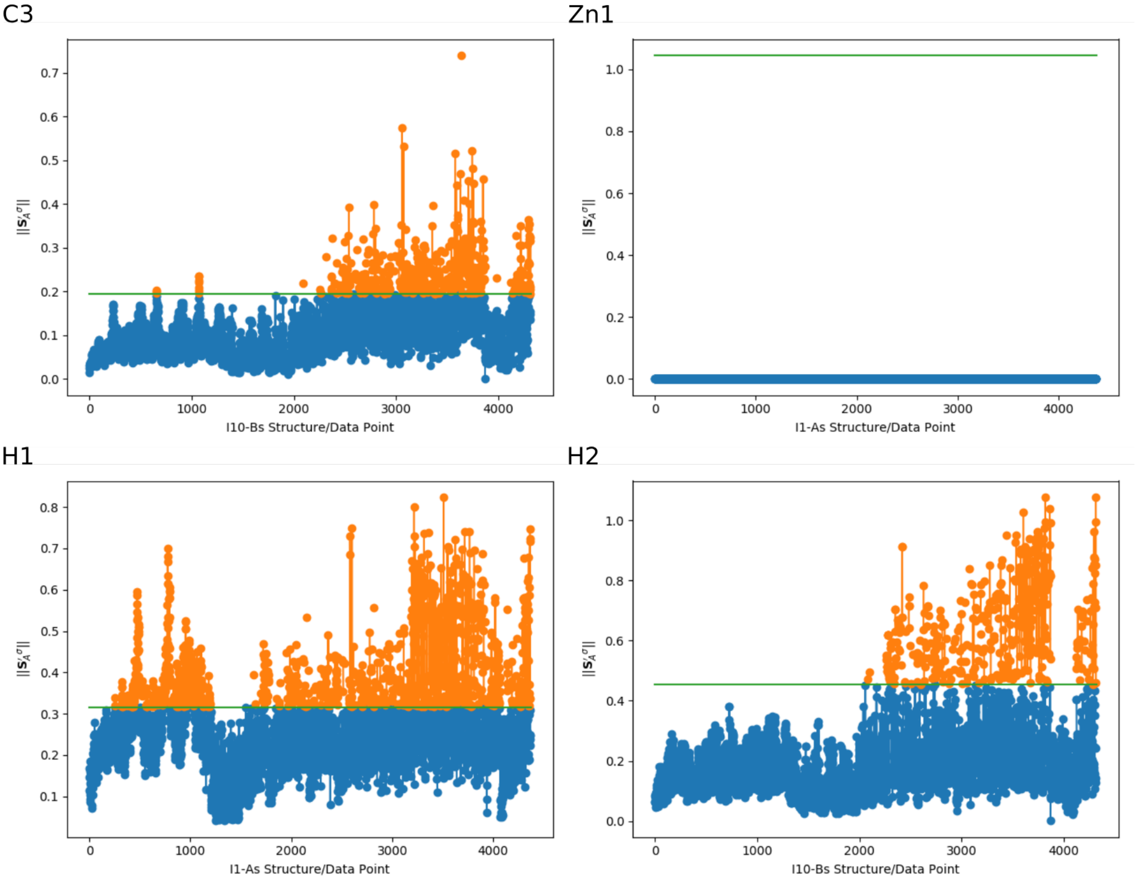

Figure 4.34: Norm $||\mathbf{S}'^{\sigma}_{A,\mathrm{frag}}||$ of the range-scaled ACSF standard deviation vector for all symmetry equivalent bulk-like central atoms $A \in \{\mathrm{Zn1}, \mathrm{H1}\}$ in each fragment structure I1-As and $A \in \{\mathrm{C3}, \mathrm{H2}\}$ in each fragment structure I10-Bs included in the data set r'_{frag}-2-SF1 (blue and orange dots) compared to the norm $||\mathbf{S}'^{\sigma}_{A,\mathrm{data}}||$ of the range-scaled ACSF standard deviation vector for all symmetry equivalent bulk-like central atoms A over all fragments I1-As ($A \in \{\mathrm{Zn1}, \mathrm{H1}\}$ and I10-Bs ($A \in \{\mathrm{C3}, \mathrm{H2}\}$) included in the data set r_{frag}-2-SF1 (green line). If $||\mathbf{S}'^{\sigma}_{A,\mathrm{frag}}|| > ||\mathbf{S}'^{\sigma}_{A,\mathrm{data}}||$ the structure is colored in orange, else in blue.

Chapter 5

Conclusion and Outlook

The main goal of this work was the development of a method to analyze the dependency of the atomic forces on the local environment to determine minimum-sized MLP training data providing size-converged atomic forces. As exemplary and challenging benchmark systems three IRMOF structures (IRMOF-1, -10 and -16) were selected. For these systems, a HDNNP, based on molecular training fragment data, should be developed, which is applicable to the bulk structures. Firstly, the drawbacks for the convergence of the atomic forces as a function of the environmental radius were discussed. The essential quantity defining the fragment size is the fragment radius r_{frag}. Increasing molecular fragments, satisfying some restrictions to reduce changes of the electronic structure, were constructed, which describe the local environment of the bulk-like central atom. In the limit of an infinite fragment radius, the molecular fragment becomes similar to the bulk. For this reason, the atomic forces provided by the fragments were compared to the reference atomic bulk forces to derive a minimum fragment size. For a certain fragment radius, the atomic force difference was assumed to vanish within a predefined convergence range. However, an obvious drawback of this method was illustrated by the O1 position of the IRMOF bulk structures, embedded in a symmetric environment demonstrating an independent force difference on the fragment radius. Furthermore, a detailed analysis of each neighboring atom is lacking.

Therefore, a Hessian-based locality test of the atomic forces was developed from scratch and illustrated by some one-dimensional model systems. As expected, the range of interactions strongly depended on the underlying electronic structure. Especially, an electronic π-system determines the range of atomic interactions. In analogy to the fragment approach, chemical entities were removed from the one-dimensional model systems to result in smaller molecular fragments and to remain the electronic structure unchanged. The force difference vector norm in relation to the effective Hessian group matrix norm conclude an effective Hessian group matrix norm threshold $\Gamma = 0.29\,\text{eV}\,\text{Å}^{-2}$, which was used to determine size-converged molecular fragments of the model systems.

As an intermediate step of transferring these results to the IRMOF systems, a one-dimensional IRMOF-1 model system (1D) was constructed and analyzed. The Hessian-based locality test provided a detailed insight on the atomic interactions of IRMOF-1. Different atomic positions were analyzed, showing the same overall behaviour of decreasing atomic interactions with increasing atomic distances. Again the π-system of the benzene-1,4-dicarboxylate (BDC^{2-}) linker mediated atomic interactions over long ranges, in contrast to the zinc atoms of the secondary building unit (SBU), which locked the atomic interactions to shorter ranges, due to the ionic character. Employing the derived threshold Γ, size-converged molecular fragments were derived providing accurate atomic forces.

The analysis of the three-dimensional IRMOF structures were perfectly in line with the results from the 1D model. Size-converged molecular fragments were derived for each inequivalent atomic site of the IRMOF bulk structures, leading to a series of fragments.

For obvious reasons, each of the atomic sites depended to a different extent on the local environment. In fact, different fragment radii were derived from these fragments. The strongest dependence on the local environment was found for the C1 position in IRMOF-10 ($r_{\mathrm{frag}} = 8.718\,\mathrm{Å}$). Nevertheless, the local environment of each atomic site have to be known to the same extent. Thus, six non-redundant molecular fragments were derived, based on $r_{\mathrm{frag}} = 8.718\,\mathrm{Å}$, as a foundation for a HDNNP training data set, which describe all atoms of the three IRMOF bulk structures. Furthermore, the type of the HDNNP generation could be assessed by the Hessian-based locality test.

The analytic forces in the second generation HDNNP formalism depend on twice the HDNNP cutoff radius, which was assumed to be equivalent to the fragment radius ($2r_{\mathrm{cut}} = 8.718\,\mathrm{Å}$). However, to train an accurate HDNNP, based on molecular fragments, with a predictive power for bulk structures, it was assumed to be also possible by another series of four molecular fragments based on half the derived fragment radius ($r'_{\mathrm{frag}} = 4.359\,\mathrm{Å}$), since for the application to bulk structures the molecular training fragments are not restricted to provide accurate bulk forces. Two independent data sets were constructed iteratively, based on half and the full fragment radius, respectively. Three HDNNPs (r'_{frag}-2-SF1, r_{frag}-2-SF1 and r_{frag}-2-SF2) were developed, validated and compared to each other.

The training of the data sets was performed by the same HDNNP architecture with differing sets of atom-centered symmetry functions (ACSFs) indicated as *-SF1* ($r_{\mathrm{cut}} = 4.359\,\mathrm{Å}$) and *-SF2* ($r_{\mathrm{cut}} = 8.718\,\mathrm{Å}$). The root-mean squared error (RMSE) values are in a comparable range and in the same order of magnitude with slightly increased values for the forces of r_{frag}-2-SF1 and r_{frag}-2-SF2. Moreover, the HDNNPs were validated for several independent data sets, resulting in the same qualitative predictions with quantitatively increased deviations for r_{frag}-2-SF2 compared to r'_{frag}-2-SF1 and r_{frag}-2-SF1.

Although, the data set r_{frag}-2-SF1 includes more bulk like atomic environments than r'_{frag}-2-SF1, the predictive power is equivalent as demonstrated by the validation. It could be illustrated, the sampling of the potential energy surface (PES) is similar for the both data sets and the analysis of the symmetry equivalent bulk-like central atoms in r_{frag}-2-SF1 revealed dependencies of these symmetry equivalent atoms underlined by a similar local environment of these atoms. Thus, the fragments based on the fragment radius $r'_{\mathrm{frag}} = 4.359\,\mathrm{Å}$ are favored compared to the enlarged fragments based on $r_{\mathrm{frag}} = 8.718\,\mathrm{Å}$, due to less demanding reference calculations and an efficient handling during the HDNNP construction.

In summary, a method to analyze the dependence of the atomic forces on the local environment was developed without the need of statistical sampling of atomic configurations but detailed information about the dependence on all neighboring atoms. The Hessian-based locality test provides minimum-sized molecular fragment structures as a foundation for a HDNNP training data set. However, this method is not restricted to molecular fragment structures and not to HDNNPs, but in principle it can also be applied to drive minimum-sized periodic training systems, which will be represented by other MLP methods. Nevertheless, this method is quite recent and other types of bonding situations like metals, electrostatic or charge-transfer dominated materials need to be analyzed to extend the knowledge and thus the applicability of this general and well-defined procedure. Furthermore, this method derived fragment radii providing molecular fragments with accurate representation of the reference forces. Moreover, it could be illustrated HDNNP training fragments do not need to include accurate bulk forces for the accurate prediction of these.

In principle, smaller reference training systems can be used to construct the HDNNP increasing the efficiency of the HDNNP construction. Additionally, this method assess the atomic interactions to derive the accurate generation of MLPs to represent the system accurately. To extend the applicability of the developed HDNNPs, further molecular fragments introducing a huge diversity of bonding situations of different MOFs can be included to the data set employing the effective Hessian group matrix norm threshold.

In the same manner as the atomic forces are calculated by MLPs, also the Hessian should be available. The analysis of the MLP Hessian will show the dependency of the atomic forces on the local environment in terms of the MLP formalism. The comparison of these two local dependencies, resulting from the reference electronic structure method and the MLP, may reveal differences in the description of the atomic forces, which further can be analyzed to improve the MLP predictions. However, for the MLP construction not only the energy and its derivative with respect to the atomic positions – the atomic forces – can be used in the training procedure, but also the second derivative of the energy with respect to the atomic positions – the Hessian. This increases the information per reference electronic structure calculation for the MLP training, reduces the amount of required reference data and consequently, increases the efficiency of MLP constructions, in general.

Bibliography

[1] Importance of Computers in Our Daily Life. URL: https://importantindia.com/24006/importance-of-computers-in-our-daily-life/, Accessed: September 9, 2022.

[2] Mooresches Gesetz: Defintion und Ende von Moore's Law. URL: https://www.giga.de/ratgeber/specials/mooresches-gesetz-defintion-und-ende-von-moore-s-law-einfach-erklaert/, Accessed: September 9, 2022.

[3] Was ist Industrie 4.0? URL: https://www.plattform-i40.de/IP/Navigation/DE/Industrie40/WasIndustrie40/was-ist-industrie-40.html, Accessed: September 9, 2022.

[4] Numerische Modellierung. URL: https://www.dwd.de/DE/forschung/wettervorhersage/-num_modellierung/numerischemodelierung_node.html, Accessed: September 9, 2022.

[5] Hethcote, H. W. The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653 (2000).

[6] The Virtual Body That Could Make Clinical Trials Unnecessary. URL: https://www.theatlantic.com/sponsored/vmware-2017/virtual-body/1625/, Accessed: September 9, 2022.

[7] Jensen, F. *Introduction to Computational Chemistry* (Wiley, West Sussex, England, 2007), 2 edn.

[8] Behler, J. Perspective: Machine learning potentials for atomistic s imulations. *The Journal of Chemical Physics* **145**, 170901 (2016).

[9] Mueller, T., Hernandez, A. & Wang, C. Machine learning for interatomic potential models. *The Journal of Chemical Physics* **152**, 050902 (2020).

[10] Behler, J. Four generations of high-dimensional neural network potentials. *Chem. Rev.* **121**, 10037–10072 (2021).

[11] AI Applications: Top 14 Artificial Intelligence Applications in 2022. URL: https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/artificial-intelligence-applications, Accessed: September 9, 2022.

[12] Hornik, K., Stinchcombe, M. & White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**, 359–366 (1989).

[13] Hornik, K., Stinchcombe, M. & White, H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks* **3**, 551–560 (1990).

[14] Blank, T. B., Brown, S. D., Calhoun, A. W. & Doren, D. J. Neural network models of potential energy surfaces. *J. Chem. Phys.* **103**, 4129–4137 (1995).

[15] Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).

[16] Behler, J. Representing potential energy surfaces by high-dimensional neural network potentials. *J. Phys. Condens. Matter* **26**, 183001 (2014).

[17] Behler, J. Constructing high-dimensional neural network potentials: A tutorial review. *Int. J. Quantum Chem.* **115**, 1032–1050 (2015).

[18] Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem. Int. Ed.* **56**, 12828 (2017).

[19] Kohn. Density functional and density matrix method scaling linearly with the number of atoms. *Phys. Rev. Lett.* **76**, 3168–3171 (1996).

[20] Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).

[21] Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).

[22] Bartók, A. P. & Csányi, G. Gaussian approximation potentials: A brief tutorial introduction. *Int. J. Quantum Chem.* **115**, 1051–1057 (2015).

[23] Shapeev, A. V. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**, 1153–1173 (2016).

[24] Thompson, A. P., Swiler, L. P., Trott, C. R., Foiles, S. M. & Tucker, G. J. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **285**, 316–330 (2015).

[25] Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **99**, 014104 (2019).

[26] Balabin, R. M. & Lomakina, E. I. Support vector machine regression (ls-svm)–an alternative to artificial neural networks (anns) for the analysis of quantum chemistry data? *Phys. Chem. Chem. Phys.* **13**, 11710–11718 (2011).

[27] Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).

[28] Balabin, R. M. & Lomakina, E. I. Support vector machine regression (ls-svm)–an alternative to artificial neural networks (anns) for the analysis of quantum chemistry data? *Phys. Chem. Chem. Phys.* **13**, 11710–11718 (2011).

[29] Pronobis, W., Tkatchenko, A. & Müller, K.-R. Many-body descriptors for predicting molecular properties with machine learning: Analysis of pairwise and three-body interactions in molecules. *J. Chem. Theory Comput.* **14**, 2991–3003 (2018).

[30] Faber, F. A., Christensen, A. S., Huang, B. & von Lilienfeld, O. A. Alchemical and structural distribution based representation for universal quantum machine learning. *J. Chem. Phys.* **148**, 241717 (2018).

[31] Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).

[32] Langer, M. F., Goeßmann, A. & Rupp, M. Representations of molecules and materials for interpolation of quantum-mechanical simulations via machine learning. *npj Comput. Mater.* **8** (2022).

[33] Kocer, E., Mason, J. K. & Erturk, H. A novel approach to describe chemical environments in high-dimensional neural network potentials. *J. Chem. Phys.* **150**, 154102 (2019).

[34] Jindal, S., Chiriki, S. & Bulusu, S. S. Spherical harmonics based descriptor for neural network potentials: Structure and dynamics of au147 nanocluster. *The Journal of Chemical Physics* **146**, 204301 (2017).

[35] Jenke, J., Subramanyam, A. P. A., Densow, M., Hammerschmidt, T., Pettifor, D. G. & Drautz, R. Electronic structure based descriptor for characterizing local atomic environments. *Phys. Rev. B* **98** (2018).

[36] Gastegger, M., Schwiedrzik, L., Bittermann, M., Berzsenyi, F. & Marquetand, P. wacsf-weighted atom-centered symmetry functions as descriptors in machine learning potentials. *J. Chem. Phys.* **148**, 241709 (2018).

[37] Parsaeifard, B., Sankar De, D., Christensen, A. S., Faber, F. A., Kocer, E., De, S., Behler, J., Anatole von Lilienfeld, O. & Goedecker, S. An assessment of the structural resolution of various fingerprints commonly used in machine learning. *Mach. Learn.: Sci. Technol.* **2**, 015018 (2021).

[38] Artrith, N., Morawietz, T. & Behler, J. High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B* **83** (2011).

[39] Yao, K., Herr, J. E., Toth, D. W., Mckintyre, R. & Parkhill, J. The tensormol-0.1 model chemistry: A neural network augmented with long-range physics. *Chem. Sci.* **9**, 2261–2269 (2018).

[40] Ko, T. W., Finkler, J. A., Goedecker, S. & Behler, J. General-purpose machine learning potentials capturing nonlocal charge transfer. *Acc. Chem. Res.* **54**, 808–817 (2021).

[41] Ko, T. W., Finkler, J. A., Goedecker, S. & Behler, J. General-purpose machine learning potentials capturing nonlocal charge transfer. *Acc. Chem. Res.* **54**, 808–817 (2021).

[42] Ko, T. W., Finkler, J. A., Goedecker, S. & Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nature communications* **12**, 398 (2021).

[43] Ghasemi, S. A., Hofstetter, A., Saha, S. & Goedecker, S. Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network. *Phys. Rev. B* **92** (2015).

[44] Xie, X., Persson, K. A. & Small, D. W. Incorporating electronic information into machine learning potential energy surfaces via approaching the ground-state electronic energy as a function of atom-based electronic populations. *J. Chem. Theory Comput.* **16**, 4256–4270 (2020).

[45] Witkoskie, J. B. & Doren, D. J. Neural network models of potential energy surfaces: Prototypical examples. *J. Chem. Theory Comput.* **1**, 14–23 (2005).

[46] Artrith, N. & Behler, J. High-dimensional neural network potentials for metal surfaces: A prototype study for copper. *Phys. Rev. B* **85**, 045439 (2012).

[47] Le, H. M., Huynh, S. & Raff, L. M. Molecular dissociation of hydrogen peroxide (hooh) on a neural network ab initio potential surface with a new configuration sampling method involving gradient fitting. *J. Chem. Phys.* **131**, 014107 (2009).

[48] Chmiela, S., Sauceda, H. E., Poltavsky, I., Müller, K.-R. & Tkatchenko, A. sgdml: Constructing accurate and data efficient molecular force fields using machine learning. *Comp. Phys. Comm.* **240**, 38–45 (2019).

[49] Eckhoff, M. & Behler, J. From molecular fragments to the bulk: Development of a neural network potential for mof-5. *J. Chem. Theory Comput.* **15**, 3793–3809 (2019).

[50] Gastegger, M., Kauffmann, C., Behler, J. & Marquetand, P. Comparing the accuracy of high-dimensional neural network potentials and the systematic molecular fragmentation method: A benchmark study for all-trans alkanes. *J. Chem. Phys.* **144**, 194110 (2016).

[51] Yang, L.-M., Vajeeston, P., Ravindran, P., Fjellvåg, H. & Tilset, M. Theoretical investigations on the chemical bonding, electronic structure, and optical properties of the metal-organic framework mof-5. *Inorg. Chem.* **49**, 10283–10290 (2010).

[52] Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **402**, 276–279 (1999).

[53] Eddaoudi, M., Kim, J., Rosi, N., Vodak, D., Wachter, J., O'Keeffe, M. & Yaghi, O. M. Systematic design of pore size and functionality in isoreticular mofs and their application in methane storage. *Science* **295**, 469–472 (2002).

[54] Furukawa, H., Cordova, K. E., O'Keeffe, M. & Yaghi, O. M. The chemistry and applications of metal-organic frameworks. *Science* **341**, 1230444 (2013).

[55] Horcajada, P., Gref, R., Baati, T., Allan, P. K., Maurin, G., Couvreur, P., Férey, G., Morris, R. E. & Serre, C. Metal-organic frameworks in biomedicine. *Chem. Rev.* **112**, 1232–1268 (2012).

[56] Kuppler, R. J., Timmons, D. J., Fang, Q.-R., Li, J.-R., Makal, T. A., Young, M. D., Yuan, D., Zhao, D., Zhuang, W. & Zhou, H.-C. Potential applications of metal-organic frameworks. *Coord. Chem. Rev.* **253**, 3042–3066 (2009).

[57] Li, B., Chrzanowski, M., Zhang, Y. & Ma, S. Applications of metal-organic frameworks featuring multi-functional sites. *Coord. Chem. Rev.* **307**, 106–129 (2016).

[58] Wang, L., Han, Y., Feng, X., Zhou, J., Qi, P. & Wang, B. Metal–organic frameworks for energy storage: Batteries and supercapacitors. *Coord. Chem. Rev.* **307**, 361–381 (2016).

[59] Li, M., Li, D., O'Keeffe, M. & Yaghi, O. M. Topological analysis of metal-organic frameworks with polytopic linkers and/or multiple building units and the minimal transitivity principle. *Chem. Rev.* **114**, 1343–1370 (2014).

[60] Eddaoudi, M., Moler, D. B., Li, H., Chen, B., Reineke, T. M., O'Keeffe, M. & Yaghi, O. M. Modular chemistry: secondary building units as a basis for the design of highly porous and robust metal-organic carboxylate frameworks. *Acc. Chem. Res.* **34**, 319–330 (2001).

[61] Tranchemontagne, D. J., Mendoza-Cortés, J. L., O'Keeffe, M. & Yaghi, O. M. Secondary building units, nets and bonding in the chemistry of metal-organic frameworks. *Chem. Soc. Rev.* **38**, 1257–1283 (2009).

[62] Haldar, R. & Maji, T. K. Metal–organic frameworks (mofs) based on mixed linker systems: structural diversities towards functional materials. *CrystEngComm* **15**, 9276–9295 (2013).

[63] Wang, Z. & Cohen, S. M. Postsynthetic modification of metal-organic frameworks. *Chem. Soc. Rev.* **38**, 1315–1329 (2009).

[64] Kalaj, M. & Cohen, S. M. Postsynthetic modification: An enabling technology for the advancement of metal-organic frameworks. *ACS Cent. Sci.* **6**, 1046–1057 (2020).

[65] Zhu, Q.-L. & Xu, Q. Metal-organic framework composites. *Chem. Soc. Rev.* **43**, 5468–5512 (2014).

[66] Coudert, F.-X. & Fuchs, A. H. Computational characterization and prediction of metal–organic framework properties. *Coord. Chem. Rev.* **307**, 211–236 (2016).

[67] Chong, S., Lee, S., Kim, B. & Kim, J. Applications of machine learning in metal-organic frameworks. *Coord. Chem. Rev.* **423**, 213487 (2020).

[68] Jablonka, K. M., Ongari, D., Moosavi, S. M. & Smit, B. Big-data science in porous materials: Materials genomics and machine learning. *Chem. Rev.* **120**, 8066–8129 (2020).

[69] Schrödinger, E. Quantisierung als eigenwertproblem. *Ann. Phys.* **384**, 361–376 (1926).

[70] Schrödinger, E. Quantisierung als eigenwertproblem. *Ann. Phys.* **384**, 489–527 (1926).

[71] Heisenberg, W. über quantentheoretische umdeutung kinematischer und mechanischer beziehungen. *Z. Phys.* **33**, 879–893 (1925).

[72] Born, M. & Jordan, P. Zur quantenmechanik. *Z. Phys.* **34**, 858–888 (1925).

[73] Born, M., Heisenberg, W. & Jordan, P. Zur quantenmechanik. ii. *Z. Phys.* **35**, 557–615 (1926).

[74] Born, M. & Oppenheimer, R. Zur quantentheorie der molekeln. *Ann. Phys.* **389**, 457–484 (1927).

[75] Gaunt, J. A. A theory of hartree's atomic fields. *Proc. Camb. Phil. Soc.* **24**, 328–342 (1928).

[76] Fock, V. Nherungsmethode zur lsung des quantenmechanischen mehrkrperproblems. *Z. Phys.* **61**, 126–148 (1930).

[77] Szabo, A. & Ostlund, N. S. *Modern Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover Publications Inc., Mineola, New York, 1996), 1 edn.

[78] Ritz, W. Über eine neue methode zur lösung gewisser variationsprobleme der mathematischen physik. *J. für Reine Angew. Math.* **1909**, 1–61 (1909).

[79] Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964).

[80] Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).

[81] Slater, J. C. The theory of complex spectra. *Phys. Rev.* **34**, 1293–1322 (1929).

[82] Perdew, J. P. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.* **577**, 1–20 (2001).

[83] Ditchfield, R., Hehre, W. J. & Pople, J. A. Self–consistent molecular–orbital methods. ix. an extended gaussian–type basis for molecular–orbital studies of organic molecules. *J. Chem. Phys.* **54**, 724–728 (1971).

[84] Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007–1023 (1989).

[85] Jensen, F. Polarization consistent basis sets: Principles. *J. Chem. Phys.* **115**, 9113–9125 (2001).

[86] Eisenschitz, R. & London, F. ber das verhltnis der van der waalsschen krfte zu den homopolaren bindungskrften. *Z. Physik* **60**, 491–527 (1930).

[87] Grimme, S. Density functional theory with london dispersion corrections. *WIREs Comput. Mol. Sci.* **1**, 211–228 (2011).

[88] 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).

[89] Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool. *Model. Simul. Mater. Sci. Eng.* **18**, 015012 (2010).

[90] Hunter, J. Matplotlib: A 2d graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

[91] Inkscape. https://inkscape.org. Version 0.92.5.

[92] Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45 (1960).

[93] Blank, T. B. & Brown, S. D. Adaptive, global, extended kalman filters for training feedforward neural networks. *J. Chemom.* **8**, 391–407 (1994).

[94] Verlet, L. Computer "experiments´´ on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.* **159**, 98–103 (1967).

[95] Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **76**, 637–649 (1982).

[96] Hockney, R. & Eastwood, J. *Computer Simulation Using Particles* (CRC Press, 1988).

[97] Beeman, D. Some multistep methods for use in molecular dynamics calculations. *J. Chem. Phys.* **20**, 130–139 (1976).

[98] Gear, C. W. *Numerical initial value problems in ordinary differential equations*. Prentice-Hall series in automatic computation (Prentice-Hall, Englewood Cliffs, NJ, 1971), 2. print edn.

[99] Alder, B. J. & Wainwright, T. E. Phase transition for a hard sphere system. *J. Chem. Phys.* **27**, 1208–1209 (1957).

[100] Alder, B. J. & Wainwright, T. E. Studies in molecular dynamics. i. general method. *J. Chem. Phys.* **31**, 459–466 (1959).

[101] Blum, V., Gehrke, R., Hanke, F., Havu, P., Havu, V., Ren, X., Reuter, K. & Scheffler, M. Ab initio molecular simulations with numeric atom-centered orbitals. *Comp. Phys. Comm.* **180**, 2175–2196 (2009).

[102] Hammer, B., Hansen, L. B. & Nørskov, J. K. Improved adsorptoin energetics within density-functional theory using revised perdew-burke-ernzerhof functionals. *Phys. Rev. B* **59**, 7413–7421 (1999).

[103] Senn, H. M. & Thiel, W. Qm/mm methods for biomolecular systems. *Angew. Chem. - Int. Ed.* **48**, 1198–1229 (2009).

[104] Eckhoff, M. & Behler, J. High-dimensional neural network potentials for magnetic systems using spin-dependent atom-centered symmetry functions. *npj Comput. Mater.* **7** (2021).

[105] Singraber, A., Behler, J. & Dellago, C. Library-based lammps implementation of high-dimensional neural network potentials. *Journal of chemical theory and computation* **15**, 1827–1840 (2019).

[106] Birch, F. Finite elastic strain of cubic crystals. *Phys. Rev.* **71**, 809–824 (1947).

[107] Deringer, V. L. & Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* **95**, 094203 (2017).

[108] Herbold, M. & Behler, J. A hessian-based assessment of atomic forces for training machine learning interatomic potentials. *J. Chem. Phys.* **156**, 114106 (2022).

[109] Rantala, T. S., Lantto, V. & Rantala, T. T. A cluster approach for modelling of surface characteristics of stannic oxide. *Phys. Scr.* **T54**, 252–255 (1994).

[110] Solak, A. O., Ranganathan, S., Itoh, T. & McCreery, R. L. A mechanism for conductance switching in carbon-based molecular electronic junctions. *Electrochem. Solid-State Lett.* **5**, E43 (2002).

[111] Lock, N., Wu, Y., Christensen, M., Cameron, L. J., Peterson, V. K., Bridgeman, A. J., Kepert, C. J. & Iversen, B. B. Elucidating negative thermal expansion in mof-5. *The Journal of Physical Chemistry C* **114**, 16181–16186 (2010).

Appendix A

Appendix

A.1 FHI-aims Settings

For the convergence tests the energy difference between the IRMOF-1 unit cell ($a = 25.8247$ Å) [111] and an expanded structure ($1.05\,a = 25.8247$ Å) is investigated for the different DFT parameters. Although, FHI-aims provides default settings for the elements, which are called *light* for fast relaxation of preliminary test calculations, *tight* for meV-level accuracy and *really_tight* for high accuracy results, the convergence behavior in this work is analyzed explicitly to ensure accurate but also efficient FHI-aims settings. The *light* settings form a lower boundary for the derived and converged DFT settings used in this work (tab. A.1 and A.2). The basis function set includes for all elements the FHI-aims recommended functions of the minimal basis, of tier 1 and the first basis function of tier 2. The confinement potential is selected with the onset radius of 4 Å, a width of 2 Å and the scaling parameter 1. The number of radial shells for the numerical integration and the angular integration grids are equivalent to the *light* FHI-aims settings with a *radial_multiplier* of 1. The atom-centered charge density expansion is, as in the *light* settings, truncated at *l_hartree* = 4. These settings are used in addition to the following keywords to form the FHI-aims *control.in* file used in this work:

```
xc                          rpbe
spin                        none
relativistic                atomic_zora scalar
charge                                          0
occupation_type             gaussian            0.01
mixer                       pulay
sc_accuracy_rho                                 1E-04   # 1E-06 mol. Hessian calc.
sc_accuracy_eev                                 1E-02   # 1E-04 mol. Hessian calc.
sc_accuracy_etot                                1E-06   # 1E-08 mol. Hessian calc.
sc_accuracy_forces                              1E-04   # 1E-06 mol. Hessian calc.
sc_iter_limit                                   100
vdw_correction_hirshfeld
compute_forces .true.
#k_grid                                         1 1 1   # for bulk calculations
#relax_geometry             bfgs                1E-02   # for structure minimization
#relax_unit_cell            fixed_angles                # for structure minimization
```

Table A.1: Compilation of the DFT parameter convergence analyzed by energy difference ΔE_{tot} of an IRMOF-1 (a = 25.8247 Å) [111]) and an expanded structure (1.05 a = 25.8247 Å) in eV. Total energy difference per atom $\Delta E_{\text{tot}}^{\text{atom}}$ is converged below ≤ 0.001 eV as stated by energy difference $\Delta\Delta E_{\text{tot}}^{\text{atom}}$ to the following, more strict DFT settings. The basis functions in FHI-aims are grouped tiers: tier = 1 all basis functions of tier 1 used; tier = $2-1$ all basis functions up to the first in tier 2 used; until tier = 2 all basis function of tier 1 and 2 used. The parameter *cut_pot* defines the onset radius of the confinement potential, *radial_ base* the number of shells for numerical integration with the *light* settings, divisions define the specific angular grids for the numerical integration, *radial_ multiplier* factors additional integration shells and *l_ hartree* defines the angular momentum expansion of the atom-centered charge density. All parameters given in *italic* font are FHI-aims keywords.

parameter	ΔE_{tot}	$\Delta E_{\text{tot}}^{\text{atom}}$	$\Delta\Delta E_{\text{tot}}^{\text{atom}}$
tier			
1	-18.5815	-0.0438	-0.0039
2-1	-16.9138	-0.0399	0.0004
2-2	-17.0792	-0.0403	0.0000
2-3	-17.0965	-0.0403	-0.0006
2-4	-16.8430	-0.0397	0.0000
2	-16.8327	-0.0397	–
cut_pot			
3	19.1896	0.0453	-0.0003
4	19.3030	0.0455	0.0000
5	19.3102	0.0455	0.0000
6	19.3106	0.0455	0.0000
7	19.3100	0.0455	–
radial_ base			
-10	-19.3750	-0.0457	-0.0002
light	-19.3045	-0.0455	0.0000
+10	-19.3102	-0.0455	–
divisions			
light	-16.7556	-0.0395	0.0003
tight	-16.8658	-0.0398	-0.0001
really_ tight	-16.8327	-0.0397	–
radial_ multiplier			
1	-19.3495	-0.0456	-0.0001
2	-19.3102	-0.0455	0.0000
3	-19.3103	-0.0455	–

Table A.2: Continuation of table A.1. Compilation of the DFT parameter convergence analyzed by energy difference ΔE_{tot} of an IRMOF-1 ($a = 25.8247$ Å) [111]) and an expanded structure ($1.05\,a = 25.8247$ Å) in eV. Total energy difference per atom $\Delta E_{\mathrm{tot}}^{\mathrm{atom}}$ is converged below ≤ 0.001 eV as stated by energy difference $\Delta\Delta E_{\mathrm{tot}}^{\mathrm{atom}}$ to the following, more strict DFT settings. The basis functions in FHI-aims are grouped tiers: tier = 1 all basis functions of tier 1 used; tier = $2-1$ all basis functions up to the first in tier 2 used; until tier = 2 all basis function of tier 1 and 2 used. The parameter *cut_pot* defines the onset radius of the confinement potential, *radial_base* the number of shells for numerical integration with the *light* settings, divisions define the specific angular grids for the numerical integration, *radial_multiplier* factors additional integration shells and *l_hartree* defines the angular momentum expansion of the atom-centered charge density. All parameters given in *italic* font are FHI-aims keywords.

parameter	ΔE_{tot}	$\Delta E_{\mathrm{tot}}^{\mathrm{atom}}$	$\Delta\Delta E_{\mathrm{tot}}^{\mathrm{atom}}$
l_hartree			
4	-19.1258	-0.0451	0.0004
6	-19.2942	-0.0455	0.0000
8	-19.3102	-0.0455	0.0000
10	-19.3166	-0.0456	–
relativistic			
none	-19.3102	-0.0455	-0.0030
atomic_zora	-18.0530	-0.0426	
k_grid			
111	-19.3102	-0.0455	0.0000
222	-19.3102	-0.0455	–

A.2 Accuracy of the Hessian Group Matrix Norm

The simple carbon dioxide molecule already introduced in section 4.3.1 is used to verify the accuracy of the Hessian group matrix norm $\mathbf{G}_A^g$ with respect to the spatial orientation, being used as confidence intervall for the effective Hessian group matrix norm threshold Γ (sec. 4.4.4). The carbon dioxide is orientated along the Cartesian z-axis referred as orientation 1. A second orientation is gained by the rotation of 35 ° around the y-axis. The Hessian group matrix norm for both orientation is equivalent within the derived Hessian group matrix norm accuracy of $\Delta||G_2^1|| \pm 0.02\,\mathrm{eV\,Å^{-2}}$ as shown in table A.3.

Table A.3: Hessian group matrix norm $||G_2^1||$ in $\mathrm{eV\,Å^{-2}}$ for two different orientations of the carbon dioxide molecule, introduced in section 4.3.1, within the Cartesian space.

orientation	$\|\|G_2^1\|\|$
1	173.563
2	173.543

A.3 Hessian-Based Assessment: Model Systems

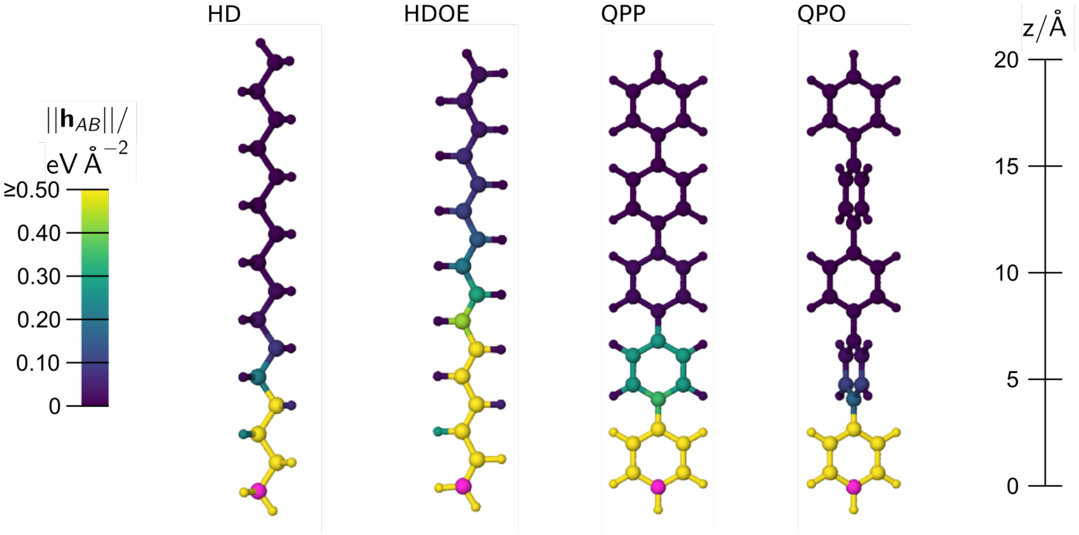

Figure A.1: The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ describing the interaction between the magenta reference carbon atoms A and all neighboring atoms B in the model systems HD, HDOE, QPP and QPO (linear scale). Adapted from [108] with permission from ©2022 AIP Publishing.

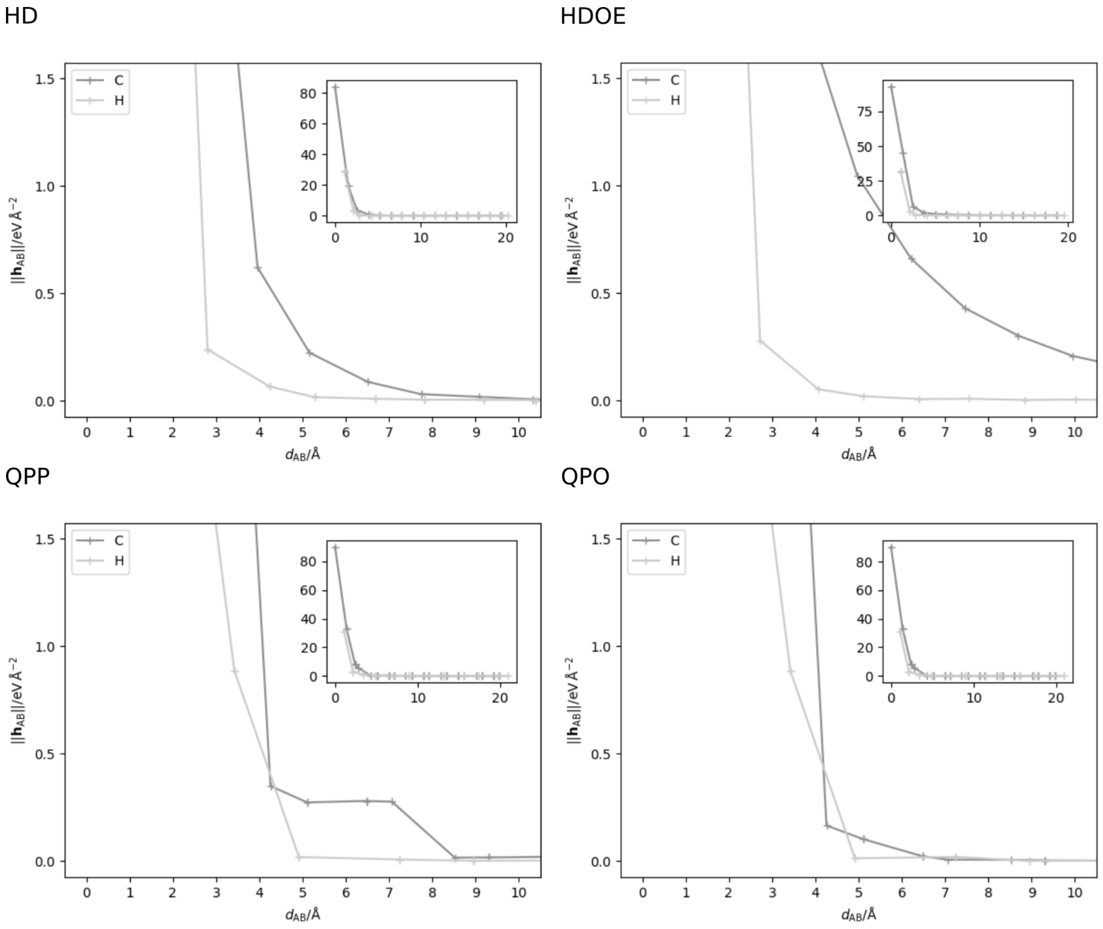

Figure A.2: The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ of the four model systems HD, HDOE, QPP and QPO as a function of the distance d_{AB} between the reference atom A as defined in figure 4.6 and neighboring atoms B. Separate curves are given for the interactions of atom A with neighboring carbon and hydrogen atoms. The inset shows the data for the interaction of A with all atoms in the entire molecules. Adapted from [108] with permission from ©2022 AIP Publishing.

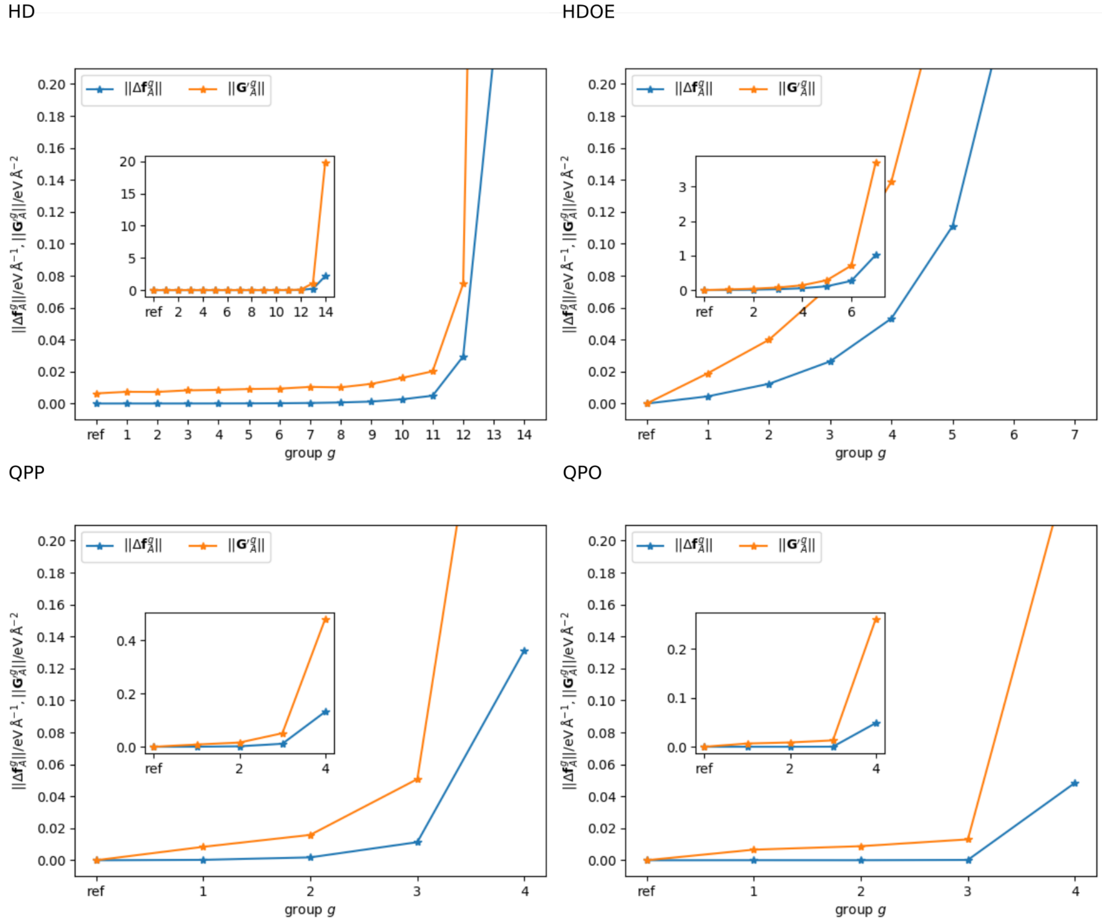

Figure A.3: The ffective Hessian group matrix norm values $||\mathbf{G'}^{g}_{A}||$ of group g and the norm of the force error $||\Delta\mathbf{f}_{A}^{Y_g}||$ of the reference carbon atom A in the fragment structure Y_g (reference structure Y without the removed atoms of group g, including hydrogen saturation) for the model systems HD, HDOE, QPP and QPO. The force difference vector is defined as $\Delta\mathbf{f}_{A}^{Y_g} = \mathbf{f}_{A}^{Y} - \mathbf{f}_{A}^{Y_g}$. Adapted from [108] with permission from ©2022 AIP Publishing.

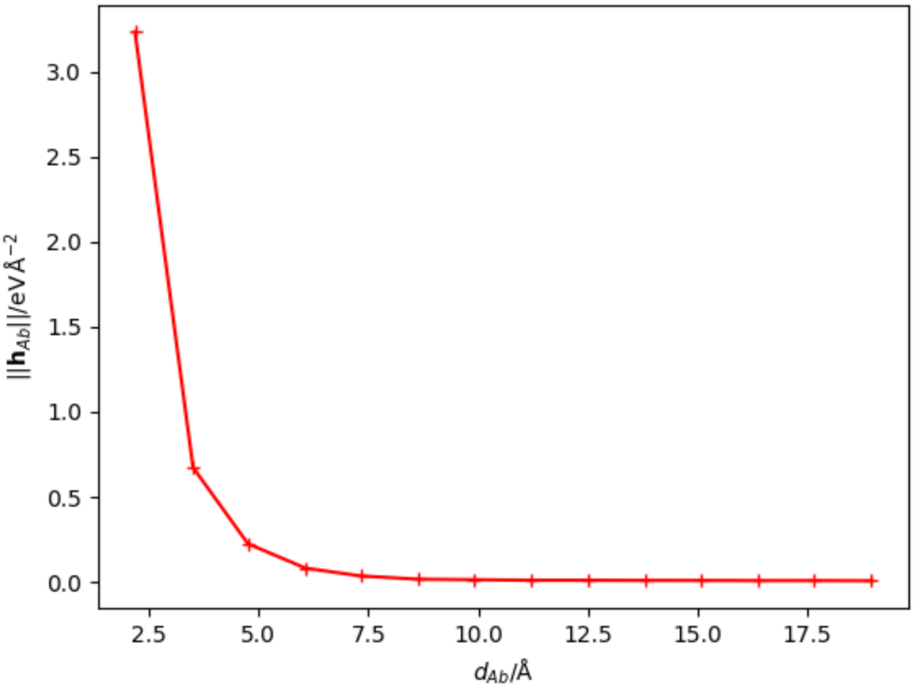

Figure A.4: The atomic Hessian sub matrix norm values $||\mathbf{h}_{Ab}||$ for the saturating hydrogen atom b as a function of the distance d_{Ab} to the reference carbon atom A in the different fragments of the HD model system. Adapted from [108] with permission from ©2022 AIP Publishing.

Table A.4: Compilation of the force component errors $\Delta f^{Y_g}_{A_{x,y,z}}$ and the total force errors $||\Delta \mathbf{f}^{Y_g}_{A}||$ in eV Å^{-1} for the reference carbon atoms in the model systems $Y =$ HD, HDOE, QPP, QPO (Fig. 6 and 7, $\sigma = 1.00$). Further, the effective Hessian group matrix norm $||\mathbf{G}'^{g}_{A}||$ is given in eV Å^{-2}. Numbers outside the intended convergence are given in bold. Adapted from [108] with permission from ©2022 AIP Publishing.

Y	g	$\Delta f^{Y_g}_{A_\mathrm{x}}$	$\Delta f^{Y_g}_{A_\mathrm{y}}$	$\Delta f^{Y_g}_{A_\mathrm{z}}$	$\|\|\Delta \mathbf{f}^{Y_g}_{A}\|\|$	$\|\|\mathbf{G}'^{g}_{A}\|\|$
HD	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0000	0.0000	0.0000	0.0000	0.01
	2	0.0000	−0.0001	0.0000	0.0001	0.01
	3	0.0000	−0.0001	0.0000	0.0001	0.01
	4	0.0000	−0.0001	0.0000	0.0001	0.01
	5	0.0000	−0.0001	0.0000	0.0001	0.01
	6	0.0000	−0.0001	0.0000	0.0001	0.01
	7	0.0000	−0.0002	0.0000	0.0002	0.01
	8	0.0000	−0.0004	−0.0001	0.0004	0.01
	9	0.0000	−0.0007	−0.0003	0.0007	0.01
	10	0.0000	−0.0011	−0.0006	0.0013	0.01
	11	0.0000	−0.0025	−0.0009	0.0027	0.02
	12	0.0000	−0.0039	−0.0030	0.0049	0.02
	13	0.0000	0.0054	−0.0291	0.0296	0.08
	14	0.0000	**−0.1981**	−0.0921	**0.2184**	**1.07**
HDOE	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0000	0.0014	−0.0043	0.0045	0.02
	2	0.0000	0.0042	−0.0116	0.0124	0.04
	3	0.0000	0.0094	−0.0247	0.0264	0.07
	4	0.0000	0.0195	−0.0495	0.0532	0.14
	5	0.0000	0.0414	−0.1031	0.1111	0.29
	6	0.0000	0.0986	**−0.2474**	**0.2663**	**0.71**
	7	0.0000	**0.1877**	**−1.0018**	**1.0192**	**3.71**
QPP	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0000	0.0000	−0.0003	0.0003	0.01
	2	0.0000	0.0000	−0.0018	0.0018	0.02
	3	0.0000	0.0000	−0.0113	0.0113	0.05
	4	0.0000	0.0000	**−0.1315**	**0.1315**	**0.48**
QPO	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0000	0.0000	0.0000	0.0000	0.01
	2	0.0000	0.0000	0.0000	0.0000	0.01
	3	0.0000	0.0000	−0.0002	0.0002	0.01
	4	0.0000	0.0000	−0.0484	0.0484	0.26

A.4 Hessian-Based Assessment: 1D-System

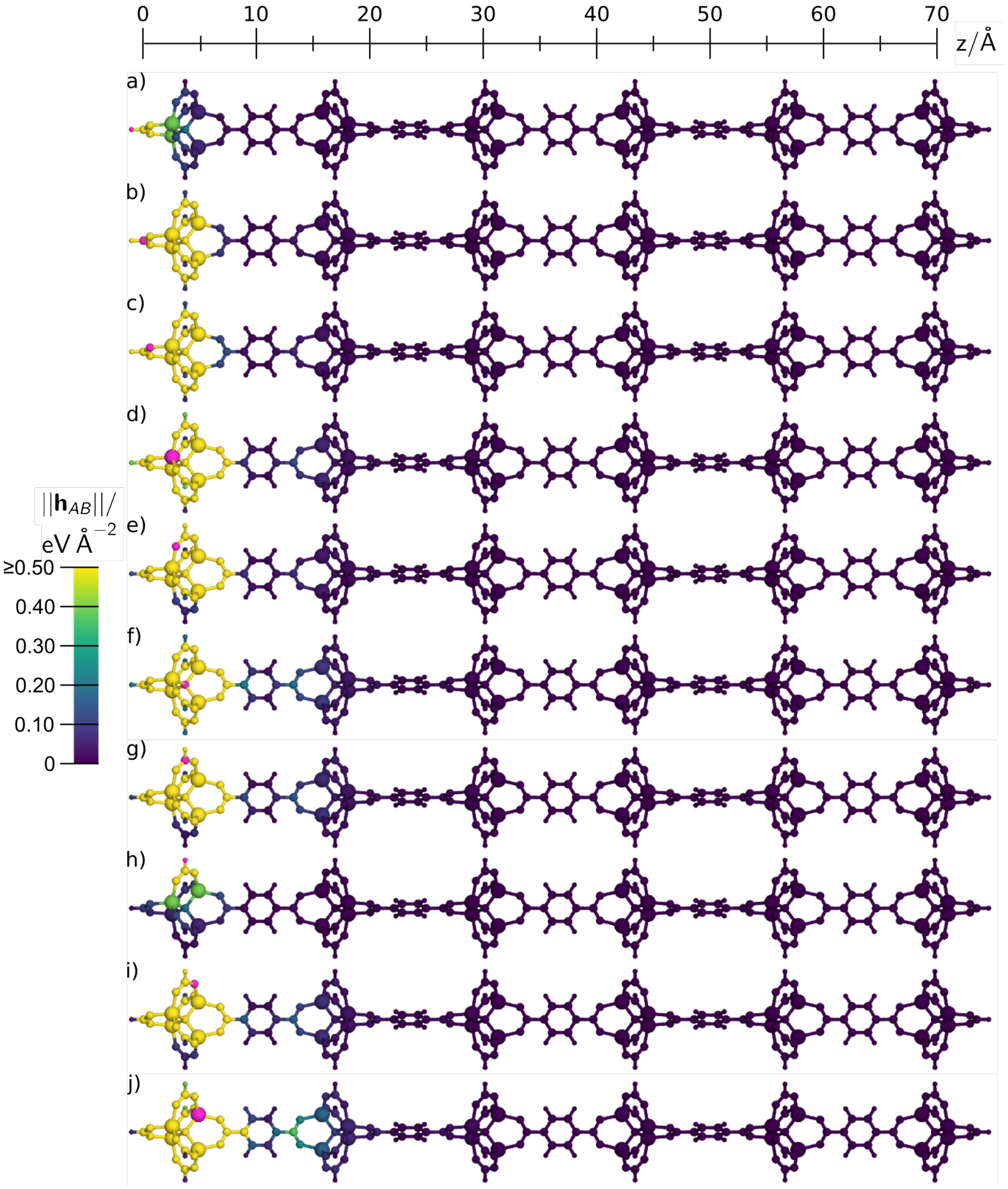

Figure A.5: a)–j) Atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ for different reference atoms A (magenta) of the one-dimensional reference structure 1D of IRMOF-1. Adapted from [108] with permission from ©2022 AIP Publishing.

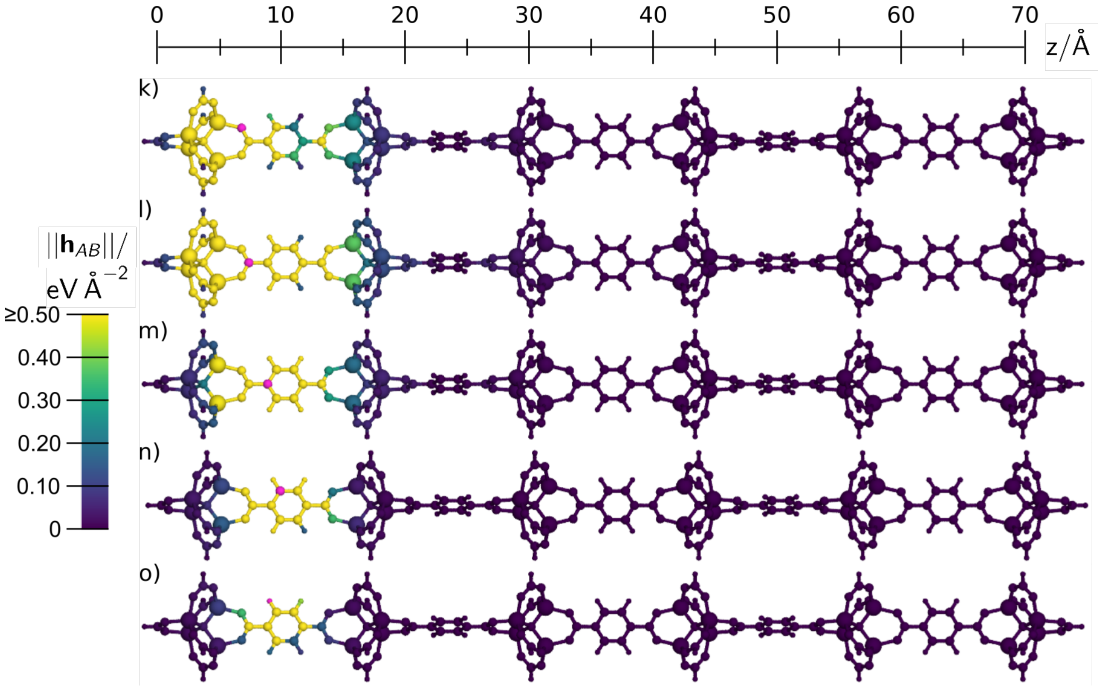

Figure A.6: Continuation of Fig. A.5. k)–o) Atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ for different reference atoms A (magenta) of the one-dimensional reference structure 1D of IRMOF-1. Adapted from [108] with permission from ©2022 AIP Publishing.

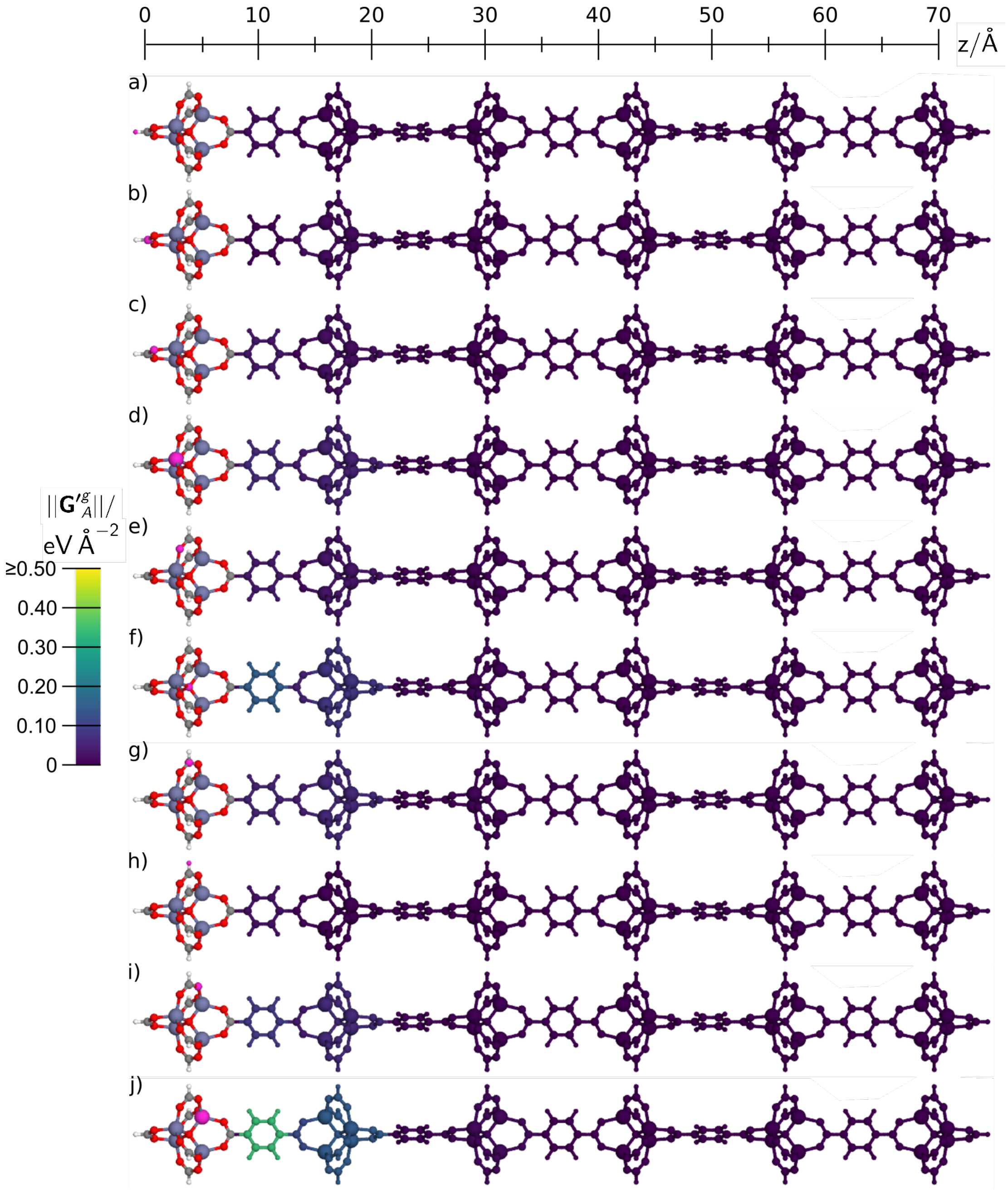

Figure A.7: a)–j) Effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ for different reference atoms A (magenta) of the one-dimensional reference structure 1D of IRMOF-1. The atomic colors of the smallest possible fragment are specified by the atom's element. Adapted from [108] with permission from ©2022 AIP Publishing.

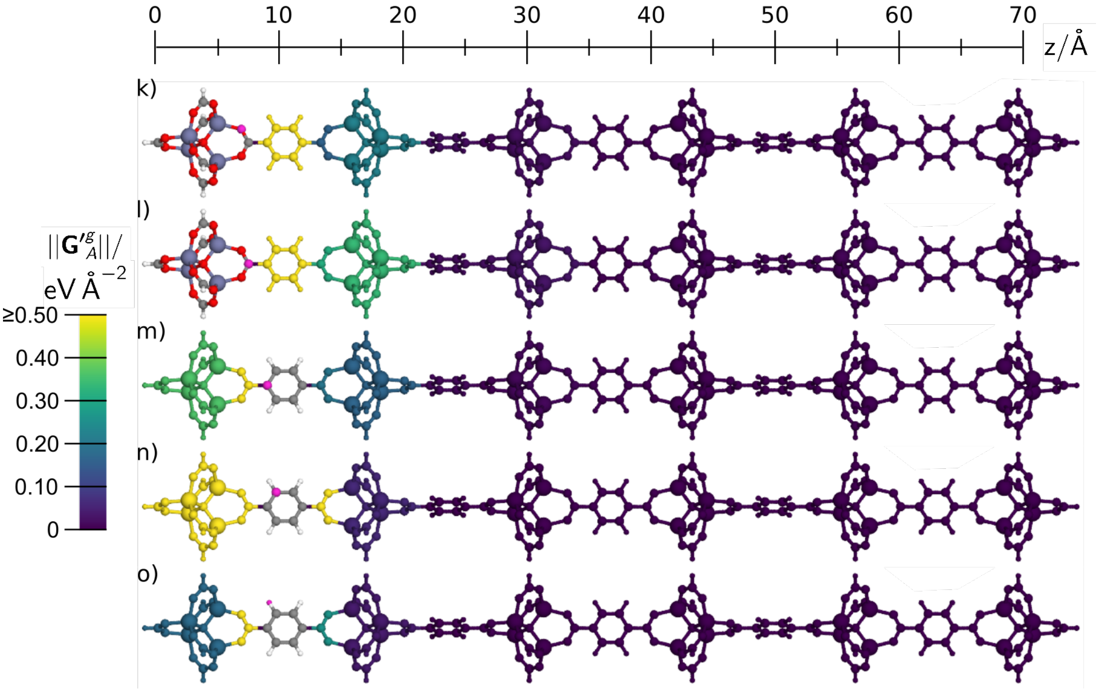

Figure A.8: ontinuation of Fig. A.7. k)–o) Effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ for different reference atoms A (magenta) of the one-dimensional reference structure 1D of IRMOF-1. The atomic colors of the smallest possible fragment are specified by the atom's element. Adapted from [108] with permission from ©2022 AIP Publishing.

A.5 Hessian-Based Assessment: 3D Fragments

A.5.1 IRMOF-1

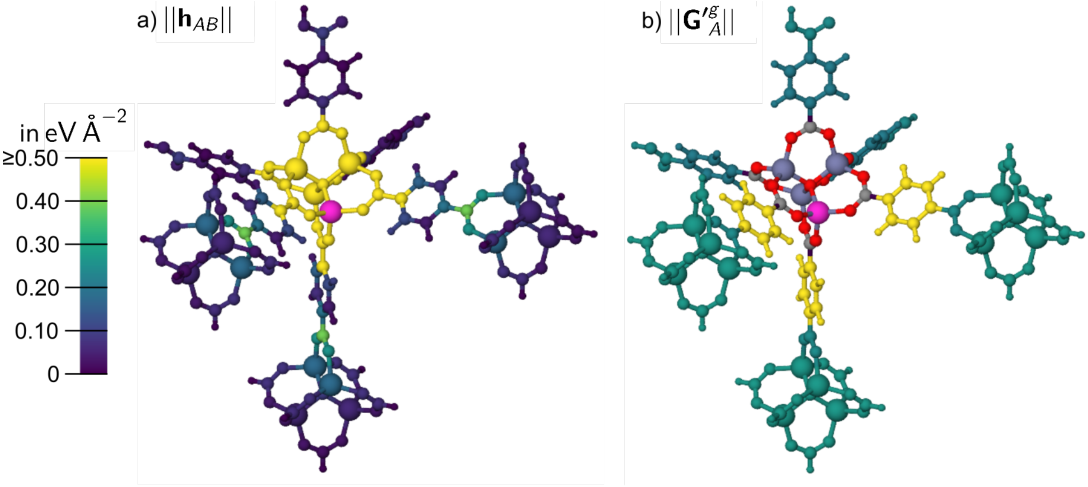

Figure A.9: a) The atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ and b) effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A = $ Zn1 (magenta) in reference structure Zn1$_{\text{ref}}$. $||\mathbf{G}'^{g}_{A}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements. Adapted from [108] with permission from ©2022 AIP Publishing.

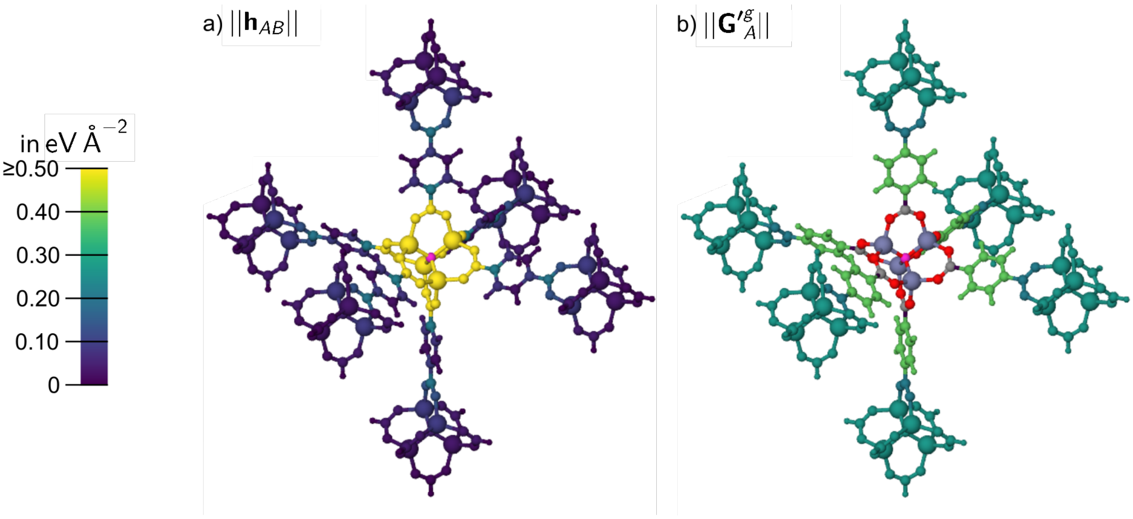

Figure A.10: a) The atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ and b) effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A = $ O1 (magenta) in reference structure O1$_{\text{ref}}$. $||\mathbf{G}'^{g}_{A}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements. Adapted from [108] with permission from ©2022 AIP Publishing.

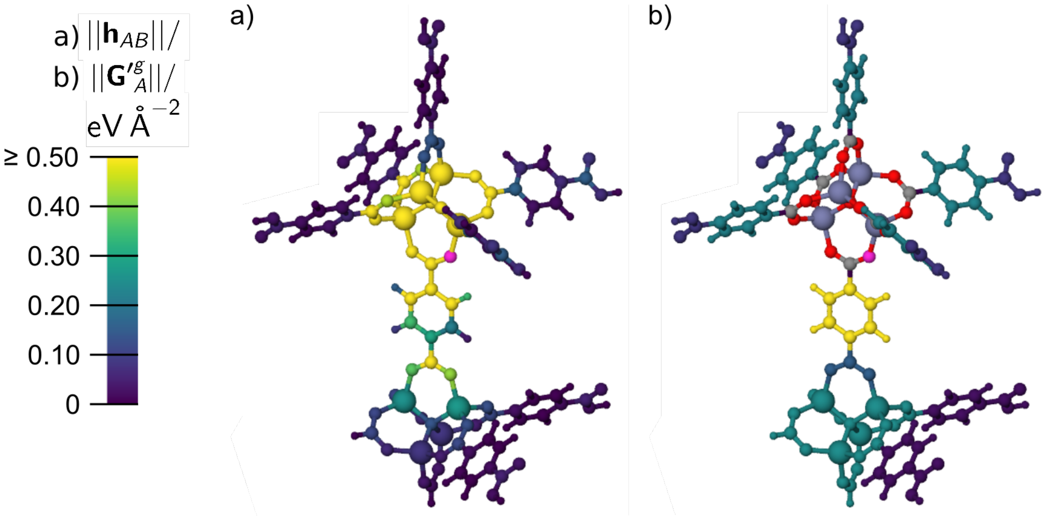

Figure A.11: a) The atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ and b) effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = O2 (magenta) in reference structure O2$_{\text{ref}}$. $||\mathbf{G}'^{g}_{A}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements. Adapted from [108] with permission from ©2022 AIP Publishing.

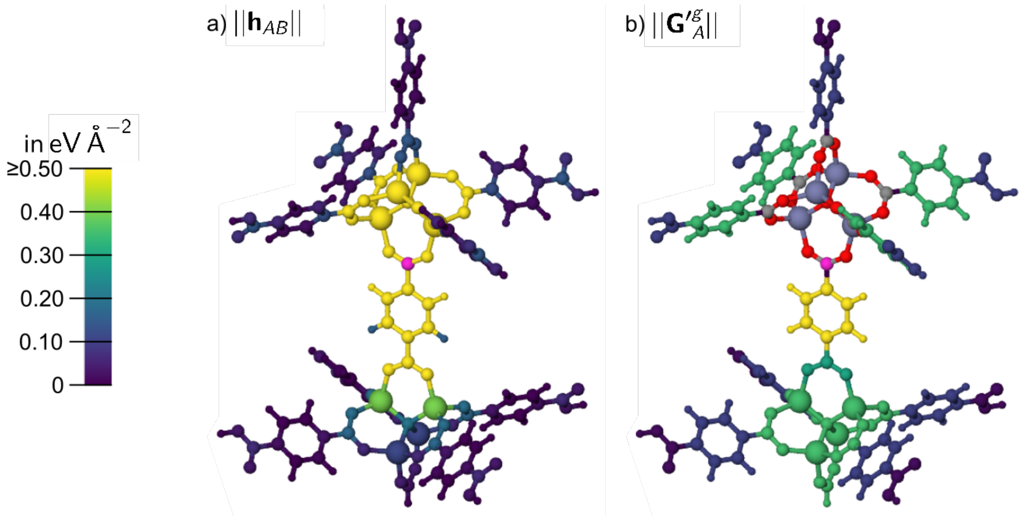

Figure A.12: a) The atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ and b) effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C1 (magenta) in reference structure C1$_{\text{ref}}$. $||\mathbf{G}'^{g}_{A}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements. Adapted from [108] with permission from ©2022 AIP Publishing.

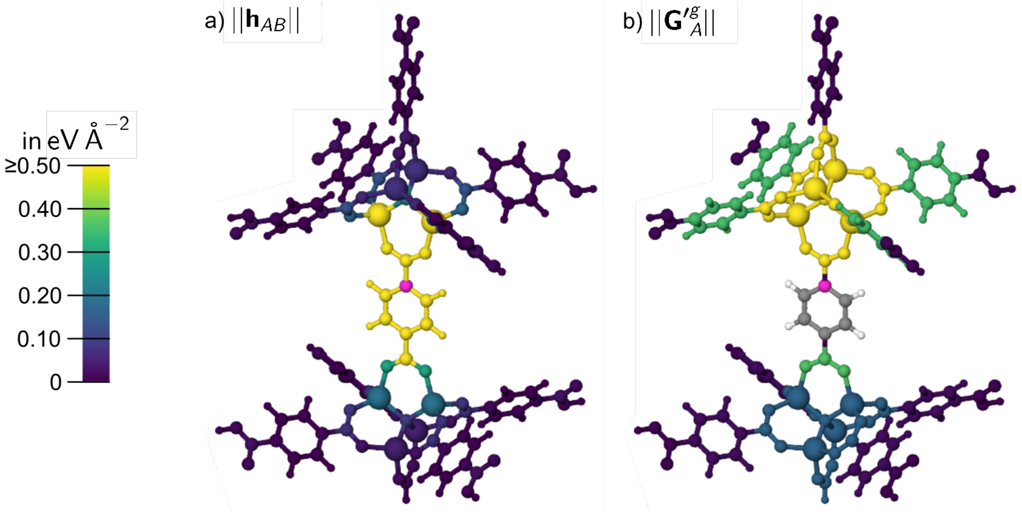

Figure A.13: a) The atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ and b) effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C2 (magenta) in reference structure C2$_{\text{ref}}$. $||\mathbf{G}'^{g}_{A}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements. Adapted from [108] with permission from ©2022 AIP Publishing.

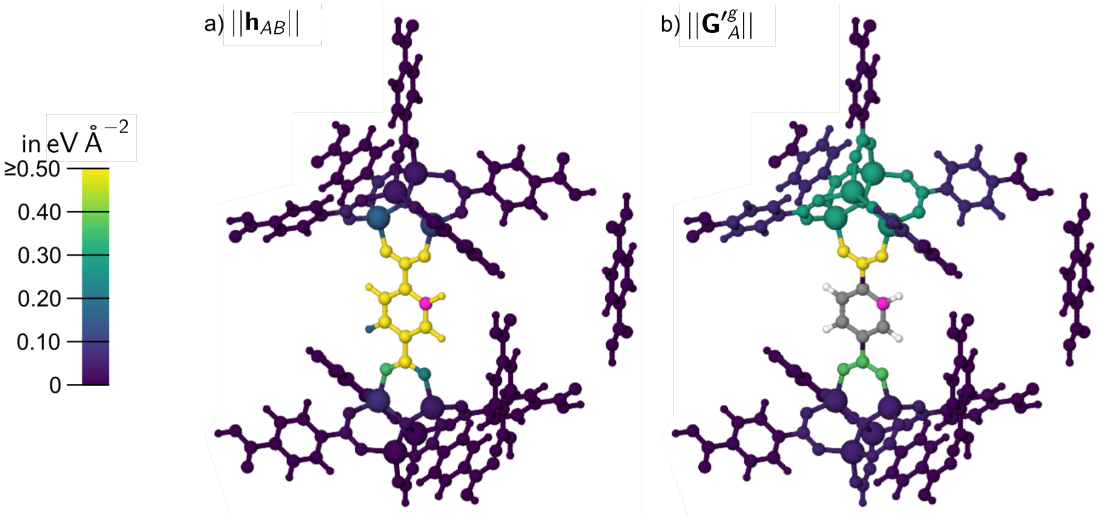

Figure A.14: a) The atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ and b) effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C3 (magenta) in reference structure C3$_{\text{ref}}$. $||\mathbf{G}'^{g}_{A}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements. Adapted from [108] with permission from ©2022 AIP Publishing.

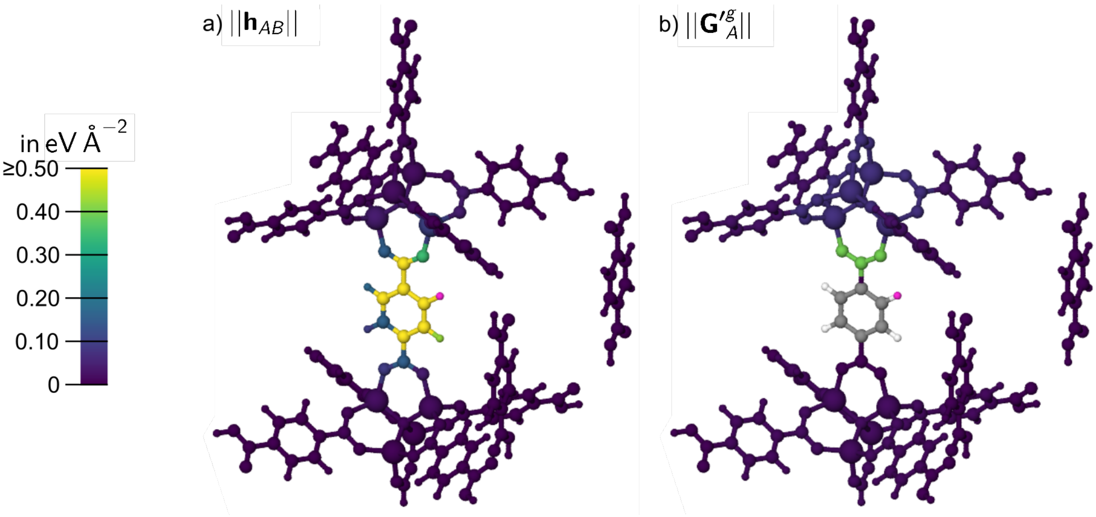

Figure A.15: a) The atomic Hessian submatrix norm values $||\mathbf{h}_{AB}||$ and b) effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = H1 (magenta) in reference structure H1$_{\text{ref}}$. $||\mathbf{G}'^{g}_{A}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements. Adapted from [108] with permission from ©2022 AIP Publishing.

Table A.5: Compilation of the force component errors $\Delta f^{Y_g}_{A_{x,y,z}}$ and the total force errors $||\Delta \mathbf{f}^{Y_g}_A||$ in eV Å^{-1} for the reference atoms Zn1, O1, O2, C1, C2, C3, H1 of the reference structures Zn1$_{\mathrm{ref}}$ (fig. A.9), O1$_{\mathrm{ref}}$ (fig. A.10), O2$_{\mathrm{ref}}$ (fig. A.11), C1$_{\mathrm{ref}}$ (fig. 4.13), C2$_{\mathrm{ref}}$ (fig. A.13), C3$_{\mathrm{ref}}$ (fig. A.14) and H1$_{\mathrm{ref}}$ (fig. A.15). Further, the effective Hessian group matrix norm $||\mathbf{G}'^{g}_{\mathrm{C1}'''}||$ is given in eV Å^{-2}. Numbers outside the intended convergence are given in bold. Adapted from [108] with permission from ©2022 AIP Publishing.

A/Y	g	$\Delta f^{Y_g}_{A_{\mathrm{x}}}$	$\Delta f^{Y_g}_{A_{\mathrm{y}}}$	$\Delta f^{Y_g}_{A_{\mathrm{z}}}$	$\|\|\Delta \mathbf{f}^{Y_g}_A\|\|$	$\|\|\mathbf{G}'^{g}_A\|\|$
Zn1	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0016	-0.0018	-0.0100	0.0103	0.27
	2	0.0065	0.0011	-0.0143	0.0158	0.26
	3	-0.0282	0.0358	0.0204	0.0499	0.23
	4	0.0005	0.0073	-0.0081	0.0109	**0.59**
O1	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0001	0.0066	-0.0001	0.0066	0.27
	2	0.0001	0.0000	-0.0001	0.0001	0.22
	3	0.0000	0.0002	-0.0001	0.0002	0.33
O2	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0028	0.0028	-0.0026	0.0048	0.02
	2	0.0220	-0.0233	0.0576	0.0659	0.09
	3	0.0268	-0.0281	0.0655	0.0762	0.25
	4	0.0143	-0.0155	0.0429	0.0478	0.22
	5	0.0058	-0.0070	-0.0115	0.0146	0.23
	6	-0.0290	0.0277	-0.0952	0.1033	0.15
	7	**0.1157**	**−0.1168**	**0.5454**	**0.5696**	**3.37**
C1	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0018	0.0022	0.0165	0.0167	0.02
	2	-0.0019	0.0022	-0.0733	0.0734	0.10
	3	-0.0019	0.0022	-0.0630	0.0630	0.11
	4	-0.0013	0.0016	-0.0599	0.0599	**0.36**
	5	-0.0024	0.0027	0.0172	0.0176	0.35
	6	-0.0018	0.0022	**0.1666**	**0.1666**	0.30
	7	-0.0016	0.0016	**−3.0135**	**3.0135**	**21.83**
C2	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0002	-0.0001	-0.0115	0.0115	0.01
	2	0.0002	-0.0001	0.0162	0.0162	0.02
	3	-0.0363	0.0364	0.0070	0.0519	0.17
	4	0.0366	-0.0365	-0.0068	0.0521	**0.36**
	5	**0.5746**	**−0.5738**	-0.0916	**0.8172**	**0.57**
	6	**−0.5443**	**0.5442**	-0.1055	**0.7769**	**0.37**
	7	-0.0002	0.0014	**1.8754**	**1.8754**	**20.2**

Table A.6: Continuation of table A.5. Compilation of the force component errors $\Delta f^{Y_g}_{A_{x,y,z}}$ and the total force errors $||\Delta \mathbf{f}^{Y_g}_A||$ in eV Å^{-1} for the reference atoms Zn1, O1, O2, C1, C2, C3, H1 of the reference structures Zn1$_{\mathrm{ref}}$ (fig. A.9), O1$_{\mathrm{ref}}$ (fig. A.10), O2$_{\mathrm{ref}}$ (fig. A.11), C1$_{\mathrm{ref}}$ (fig. 4.13), C2$_{\mathrm{ref}}$ (fig. A.13), C3$_{\mathrm{ref}}$ (fig. A.14) and H1$_{\mathrm{ref}}$ (fig. A.15). Further, the effective Hessian group matrix norm $||\mathbf{G}'^{g}_{\mathrm{C1}'''}||$ is given in eV Å^{-2}. Numbers outside the intended convergence are given in bold. Adapted from [108] with permission from ©2022 AIP Publishing.

A/Y	g	$\Delta f^{Y_g}_{A_\mathrm{x}}$	$\Delta f^{Y_g}_{A_\mathrm{y}}$	$\Delta f^{Y_g}_{A_\mathrm{z}}$	$\|\|\Delta \mathbf{f}^{Y_g}_A\|\|$	$\|\|\mathbf{G}'^{g}_A\|\|$
C3	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0022	-0.0023	0.0017	0.0036	0.00
	2	0.0029	-0.0030	0.0015	0.0045	0.02
	3	-0.0252	0.0250	-0.0655	0.0745	0.05
	4	0.0348	-0.0349	0.0668	0.0830	0.06
	5	-0.0236	0.0235	-0.0102	0.0348	0.29
	6	0.0026	-0.0030	0.0006	0.0040	**0.37**
	7	**0.1660**	**−0.1663**	**−0.3542**	**0.4251**	**1.99**
H1	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0006	0.0008	0.0002	0.0010	0.01
	2	0.0003	-0.0001	-0.0005	0.0006	0.02
	3	0.0010	-0.0007	-0.0127	0.0128	0.02
	4	-0.0035	0.0038	0.0188	0.0195	0.03
	5	-0.0025	0.0027	-0.0166	0.0170	0.01
	6	-0.0361	0.0362	0.0387	0.0641	0.07
	7	-0.0795	0.0797	-0.0035	0.1126	**0.40**

A.5.2 IRMOF-10

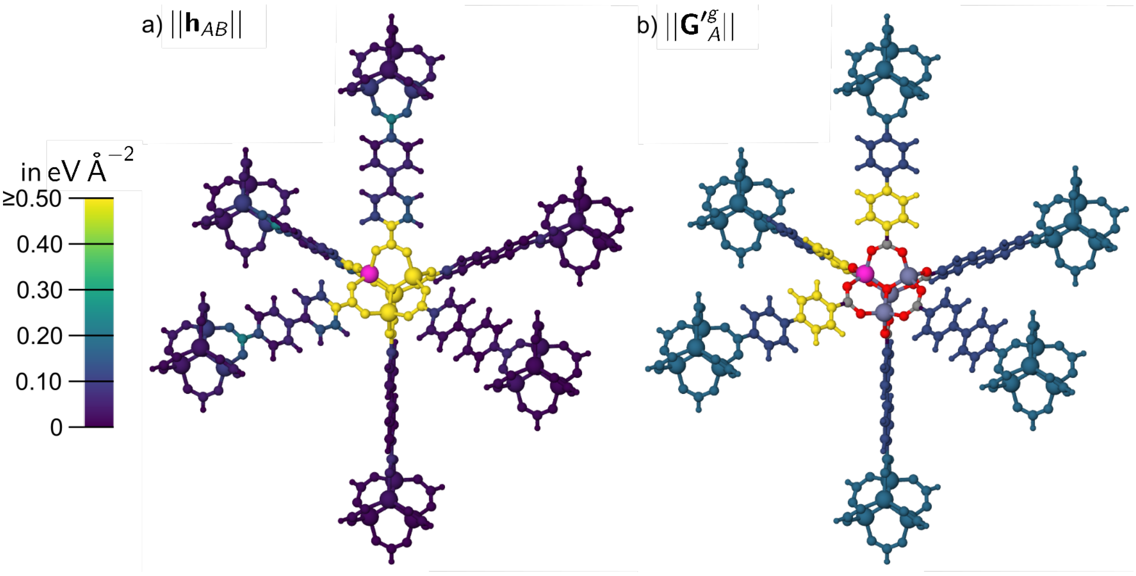

Figure A.16: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = Zn1 (magenta) in reference structure I10$_{\text{ref}}$. $||\mathbf{G}'^{g}_{\text{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

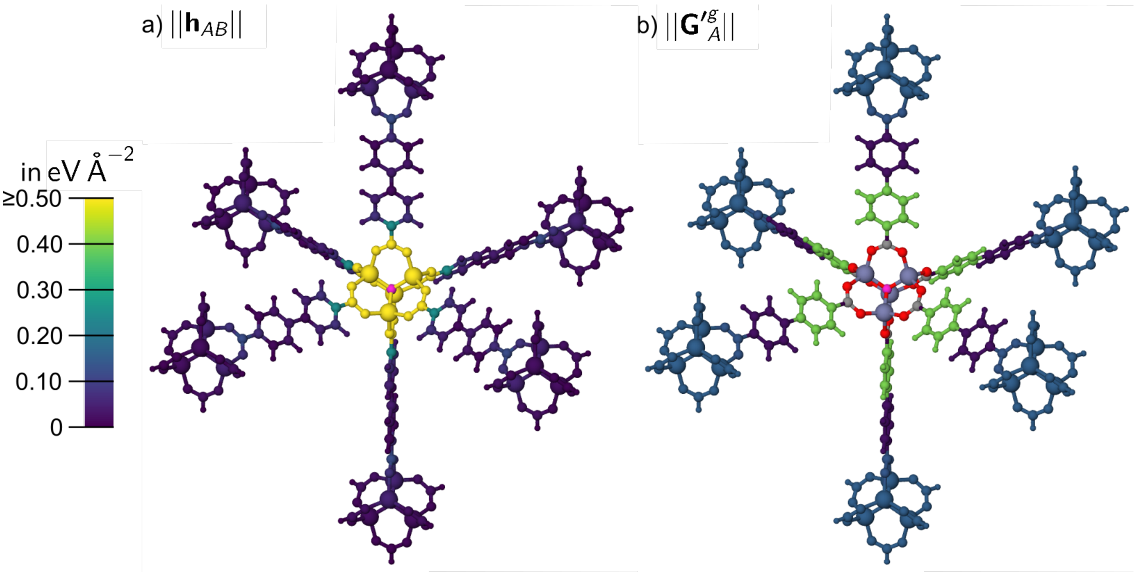

Figure A.17: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = O1 (magenta) in reference structure I10$_{\text{ref}}$. $||\mathbf{G}'^{g}_{\text{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

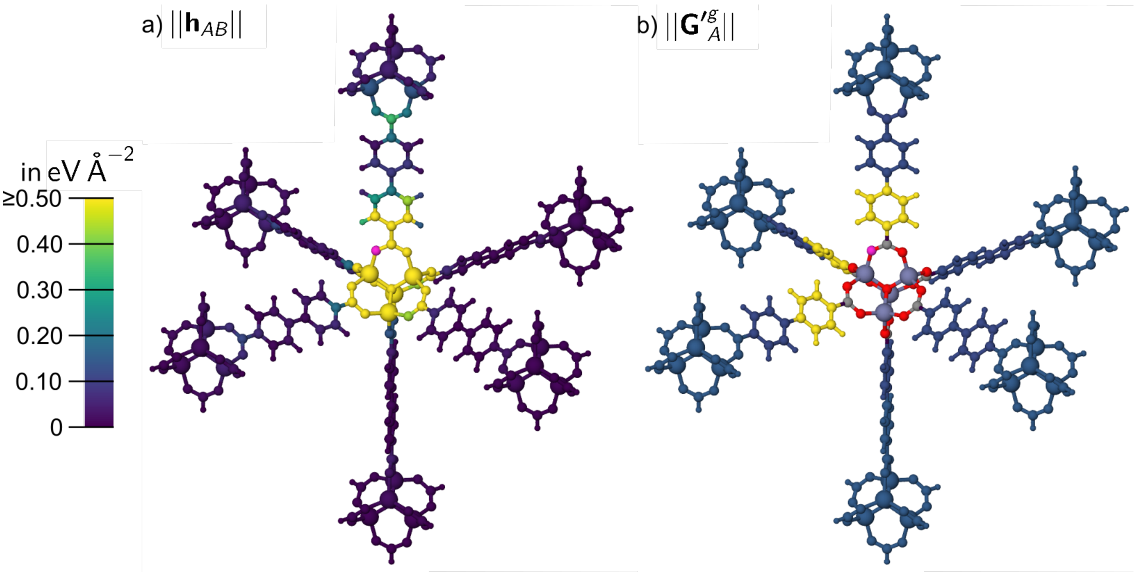

Figure A.18: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A = \mathrm{O2}$ (magenta) in reference structure $\mathrm{I10_{ref}}$. $||\mathbf{G}'^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

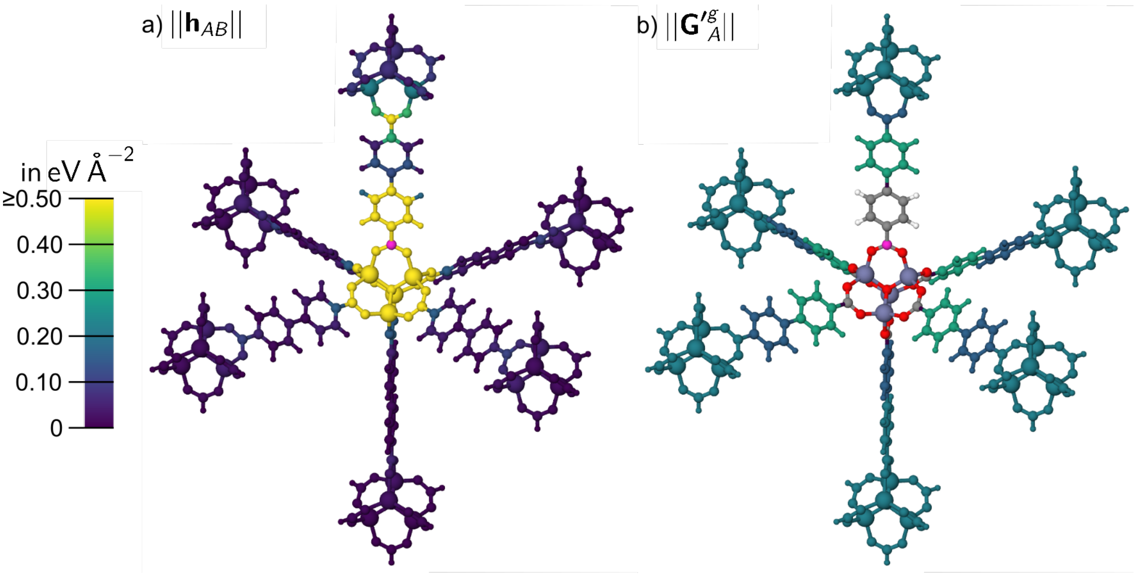

Figure A.19: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A = \mathrm{C1}$ (magenta) in reference structure $\mathrm{I10_{ref}}$. $||\mathbf{G}'^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

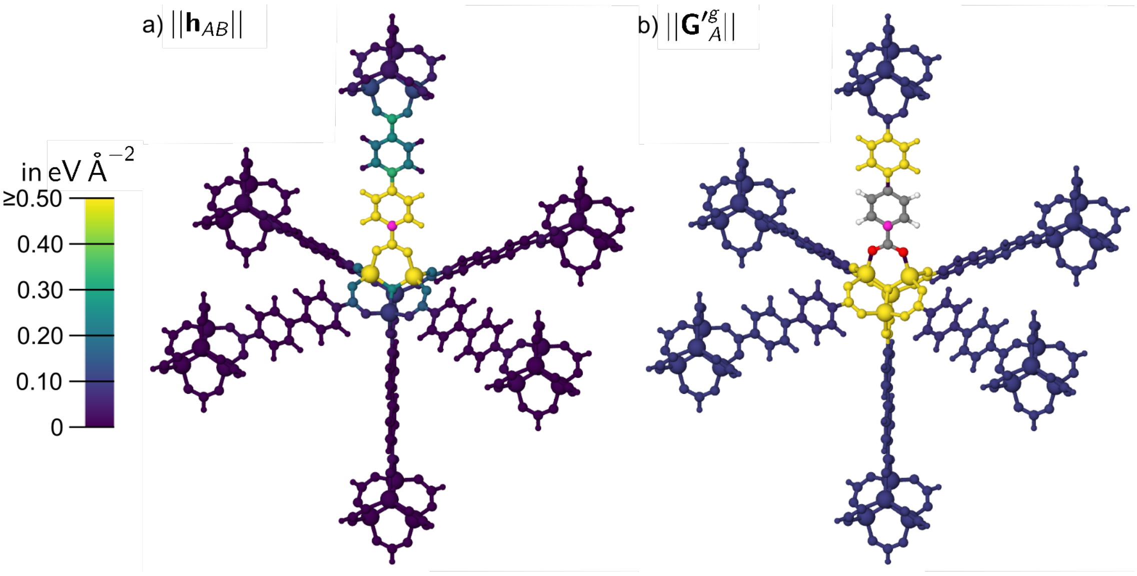

Figure A.20: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G'}^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C2 (magenta) in reference structure I10$_{\mathrm{ref}}$. $||\mathbf{G'}^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

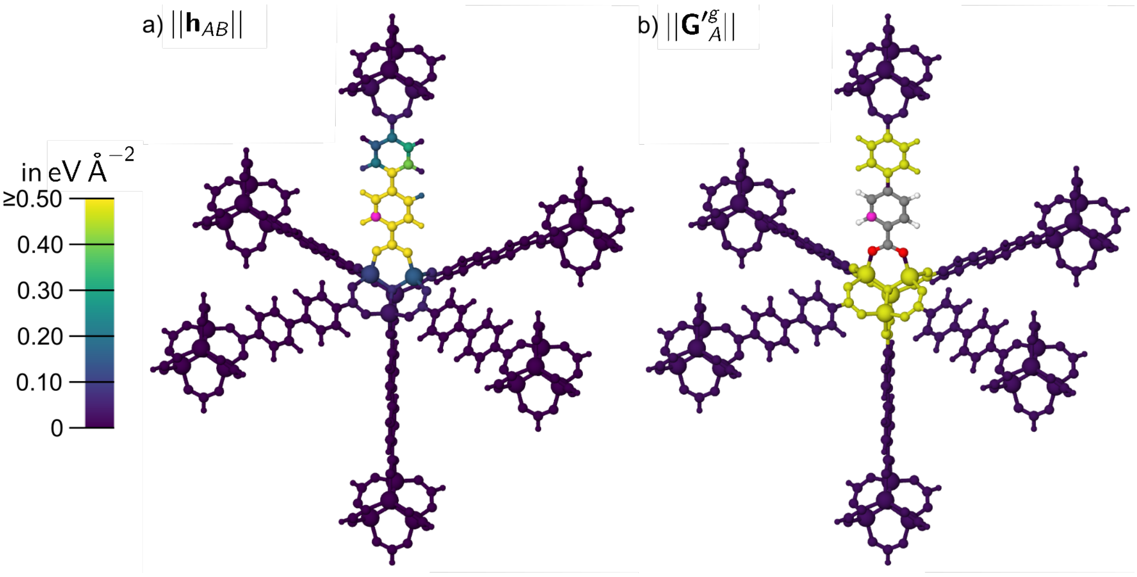

Figure A.21: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G'}^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C3 (magenta) in reference structure I10$_{\mathrm{ref}}$. $||\mathbf{G'}^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

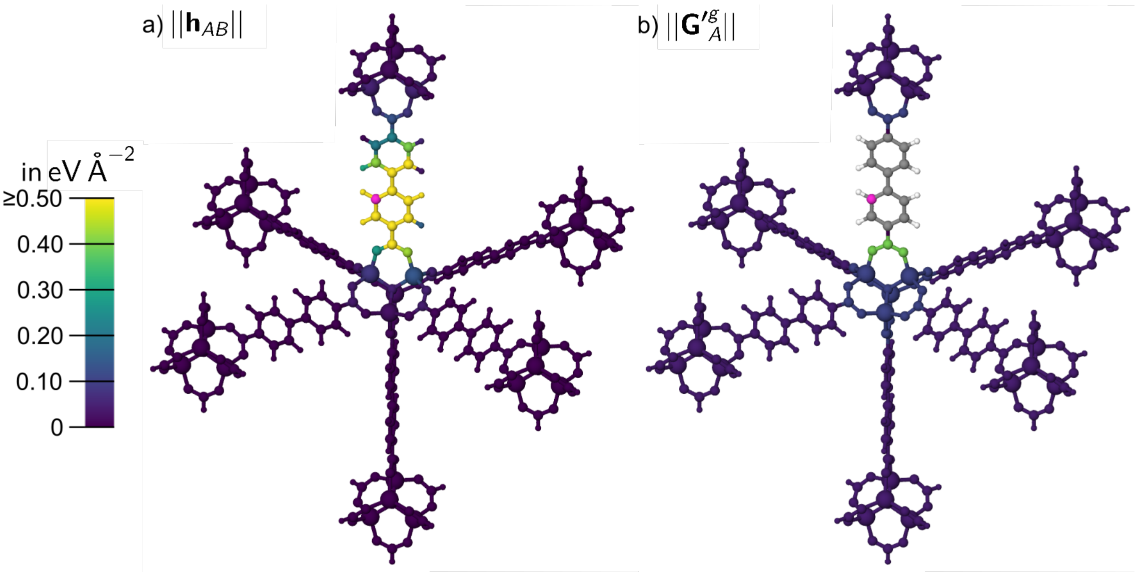

Figure A.22: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G'}^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A =$ C4 (magenta) in reference structure I10$_{\mathrm{ref}}$. $||\mathbf{G'}^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

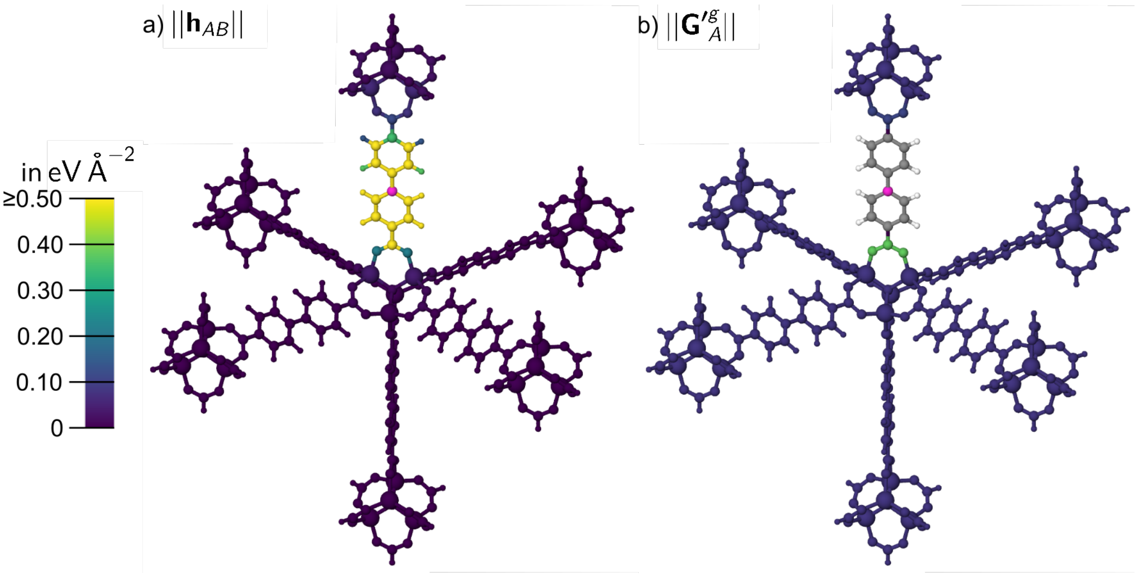

Figure A.23: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G'}^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A =$ C5 (magenta) in reference structure I10$_{\mathrm{ref}}$. $||\mathbf{G'}^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

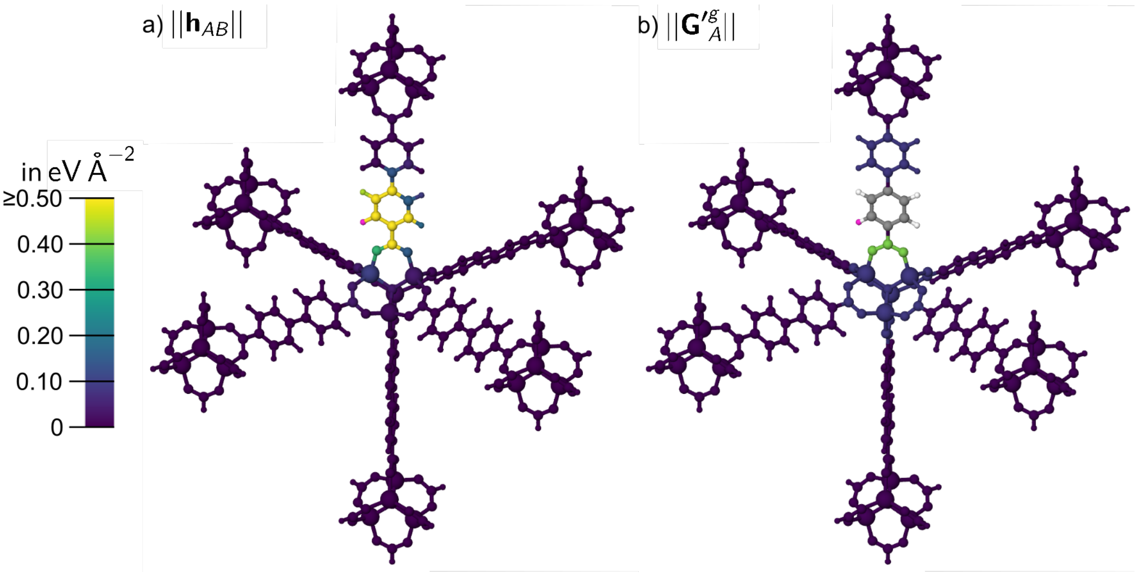

Figure A.24: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A = \mathrm{H1}$ (magenta) in reference structure I10$_{\mathrm{ref}}$. $||\mathbf{G}'^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

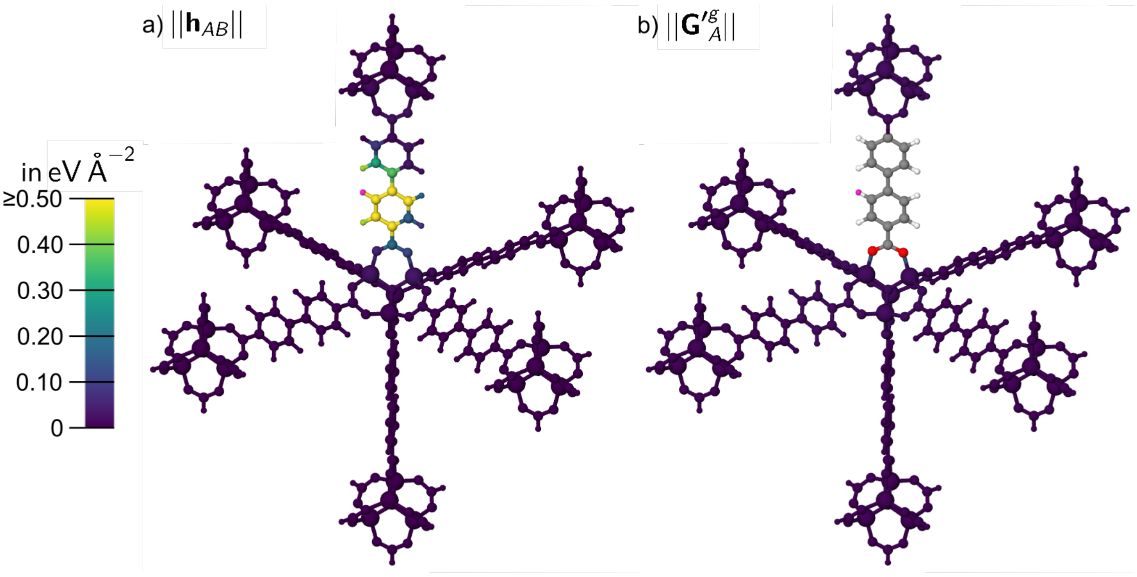

Figure A.25: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A = \mathrm{H2}$ (magenta) in reference structure I10$_{\mathrm{ref}}$. $||\mathbf{G}'^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

Table A.7: Compilation of the force component errors $\Delta f^{Y_g}_{A_{x,y,z}}$ and the total force errors $||\Delta \mathbf{f}^{Y_g}_A||$ in eV Å^{-1} for the IRMOF-10 reference atoms Zn1, O1, O2, C1, C2, C3, C4, C5, H1 and H2 of the reference structure I10 (fig. 4.16 and A.16 to A.25). Further, the effective Hessian group matrix norm $||\mathbf{G}'^{g}_{\mathrm{C1}'''}||$ is given in eV Å^{-2}. Numbers outside the intended convergence are given in bold.

A/Y	g	$\Delta f^{Y_g}_{A_\mathrm{x}}$	$\Delta f^{Y_g}_{A_\mathrm{y}}$	$\Delta f^{Y_g}_{A_\mathrm{z}}$	$\lvert\lvert\Delta \mathbf{f}^{Y_g}_A\rvert\rvert$	$\lvert\lvert\mathbf{G}'^{g}_A\rvert\rvert$
Zn1	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0164	0.0162	0.0128	0.0264	0.18
	2	-0.0020	0.0018	0.0018	0.0032	0.18
	3	0.0340	-0.0342	-0.0342	0.0591	0.12
	4	0.0054	-0.0050	-0.0050	0.0089	**0.70**
O1	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0001	0.0031	0.0001	0.0031	0.15
	2	0.0001	0.0001	0.0001	0.0002	0.14
	3	0.0001	0.0001	0.0001	0.0001	0.02
	4	-0.0001	0.0000	0.0000	0.0001	**0.40**
O2	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0095	0.0095	-0.0321	0.0348	0.15
	2	0.0118	-0.0117	0.0255	0.0304	0.12
	3	-0.0163	0.0164	0.0172	0.0288	0.12
	4	**−0.1481**	**0.1477**	**−0.5776**	**0.6143**	**3.52**
C1	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0013	0.0020	0.0430	0.0431	0.22
	2	-0.0040	0.0049	0.0905	0.0907	0.16
	3	-0.0013	0.0020	-0.0235	0.0236	**0.30**
C2	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0017	0.0017	-0.0076	0.0079	0.10
	2	0.0000	0.0000	0.0438	0.0438	0.09
	3	0.0000	0.0000	0.0667	0.0667	0.08
	4	**0.5441**	**−0.5444**	-0.0497	**0.7712**	**0.71**
C3	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0010	-0.0010	-0.0009	0.0017	0.02
	2	0.0059	-0.0058	-0.0045	0.0094	0.03
	3	-0.0407	0.0414	0.0858	0.1036	**0.47**
C4	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0004	-0.0004	0.0082	0.0082	0.04
	2	0.0179	-0.0178	-0.0546	0.0601	0.10
	3	-0.0060	0.0068	-0.1236	0.1239	**0.40**
C5	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0433	-0.0434	-0.0065	0.0616	0.08
	2	0.0433	-0.0434	-0.0065	0.0616	0.11
	3	0.0384	-0.0384	-0.0017	0.0543	**0.38**
	4	-0.0001	-0.0001	0.1109	0.1109	0.29

Table A.8: Continuation of table A.7. Compilation of the force component errors $\Delta f^{Y_g}_{A_{x,y,z}}$ and the total force errors $||\Delta \mathbf{f}^{Y_g}_{A}||$ in eV Å^{-1} for the IRMOF-10 reference atoms Zn1, O1, O2, C1, C2, C3, C4, C5, H1 and H2 of the reference structure I10 (fig. 4.16 and A.16 to A.25). Further, the effective Hessian group matrix norm $||\mathbf{G}'^{g}_{\mathrm{C1}'''}||$ is given in eV Å^{-2}. Numbers outside the intended convergence are given in bold.

A/Y	g	$\Delta f^{Y_g}_{A_\mathrm{x}}$	$\Delta f^{Y_g}_{A_\mathrm{y}}$	$\Delta f^{Y_g}_{A_\mathrm{z}}$	$\|\|\Delta \mathbf{f}^{Y_g}_{A}\|\|$	$\|\|\mathbf{G}'^{g}_{A}\|\|$
H1	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0003	0.0003	0.0014	0.0015	0.01
	2	0.0036	-0.0036	-0.0022	0.0055	0.01
	3	0.0356	-0.0356	-0.0526	0.0728	0.09
	4	0.0787	-0.0788	-0.0145	0.1123	**0.40**
H2	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0003	-0.0003	0.0008	0.0009	0.01
	2	0.0003	-0.0003	0.0028	0.0028	0.02
	3	0.0000	0.0000	-0.0130	0.0130	0.16
	4	0.0000	0.0000	-0.0130	0.0130	0.03
	5	0.0028	-0.0028	-0.0236	0.0239	0.15

A.5.3 IRMOF-16

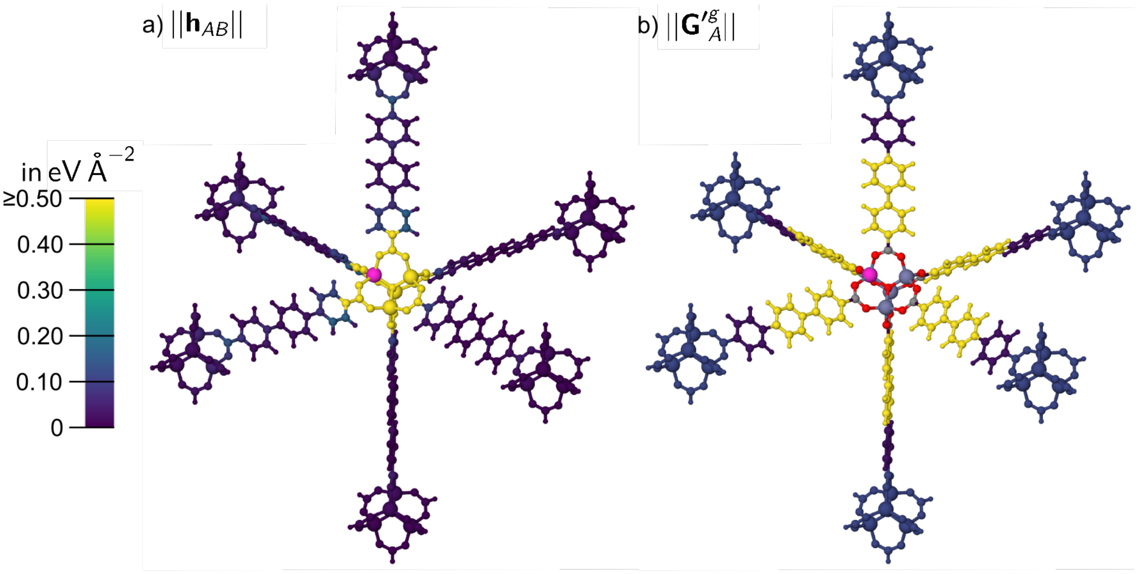

Figure A.26: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G'}^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A =$ Zn1 (magenta) in reference structure I16$_{\text{ref}}$. $||\mathbf{G'}^{g}_{\text{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

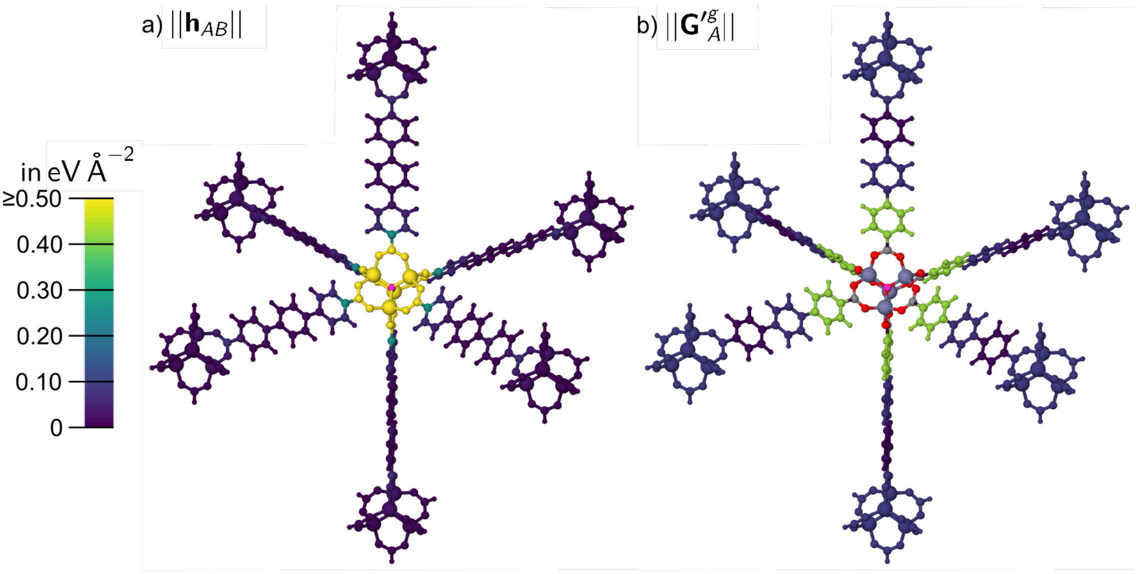

Figure A.27: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A = \mathrm{O1}$ (magenta) in reference structure I16$_{\mathrm{ref}}$. $||\mathbf{G}'^{g}_{A}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

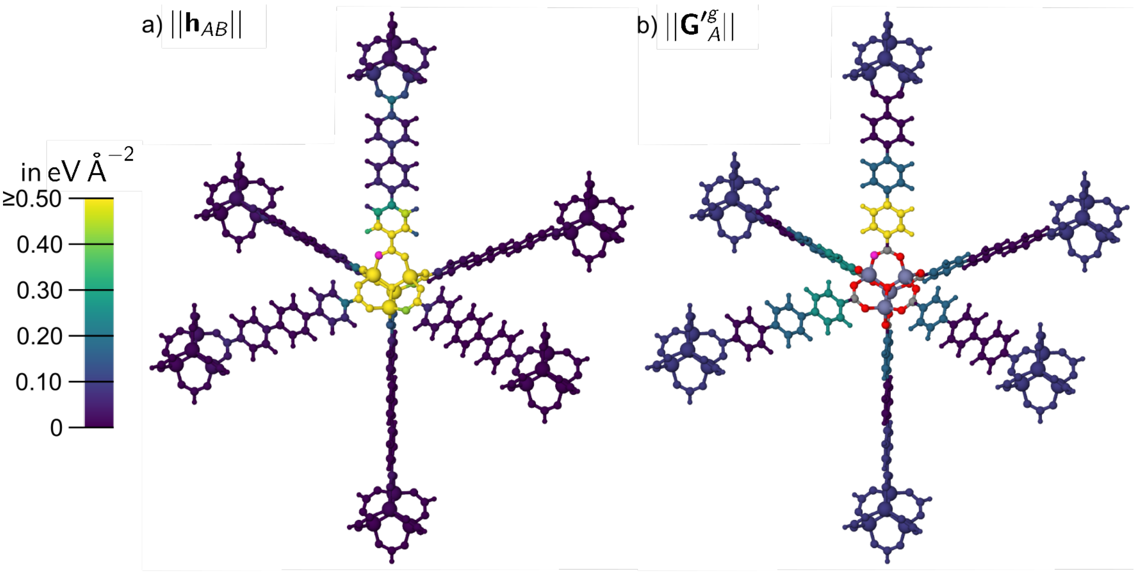

Figure A.28: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom $A = \mathrm{O2}$ (magenta) in reference structure I16$_{\mathrm{ref}}$. $||\mathbf{G}'^{g}_{A}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

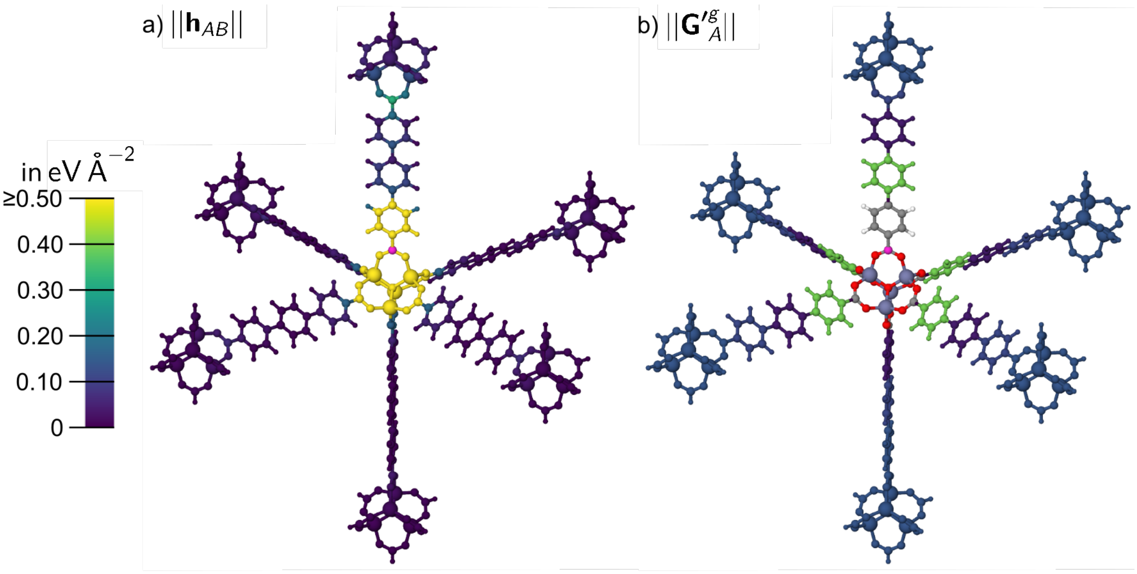

Figure A.29: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C1 (magenta) in reference structure I16$_{\text{ref}}$. $||\mathbf{G}'^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

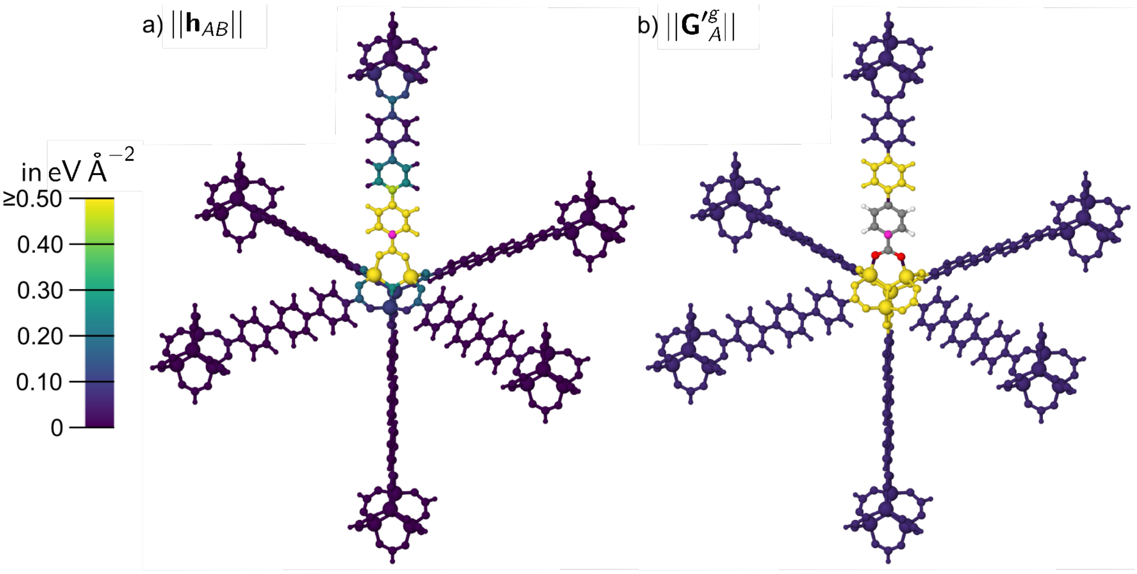

Figure A.30: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C2 (magenta) in reference structure I16$_{\text{ref}}$. $||\mathbf{G}'^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

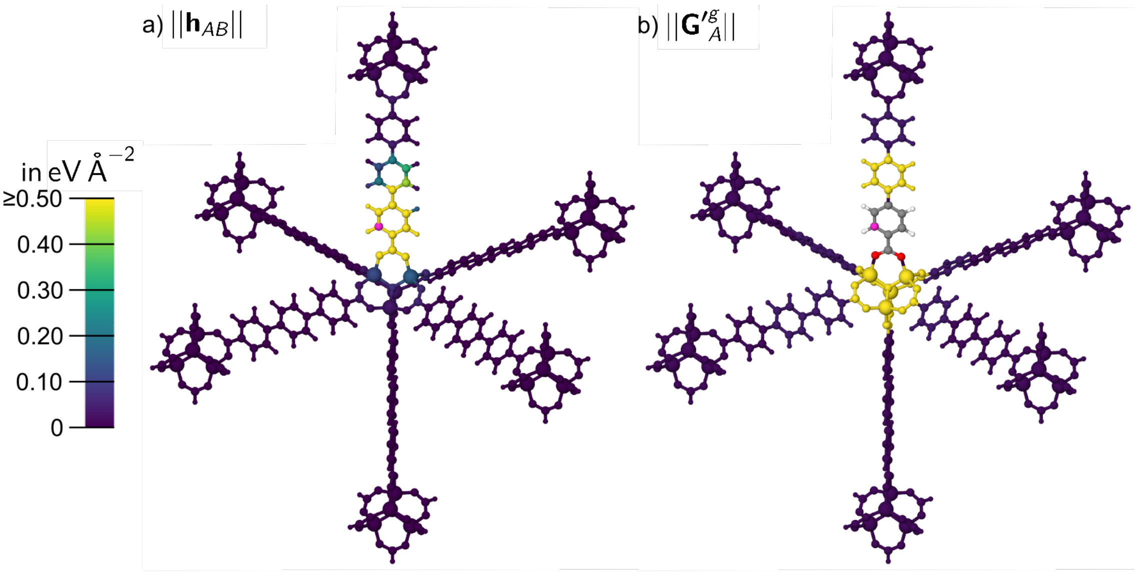

Figure A.31: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C3 (magenta) in reference structure I16$_{\text{ref}}$. $||\mathbf{G}'^{g}_{\text{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

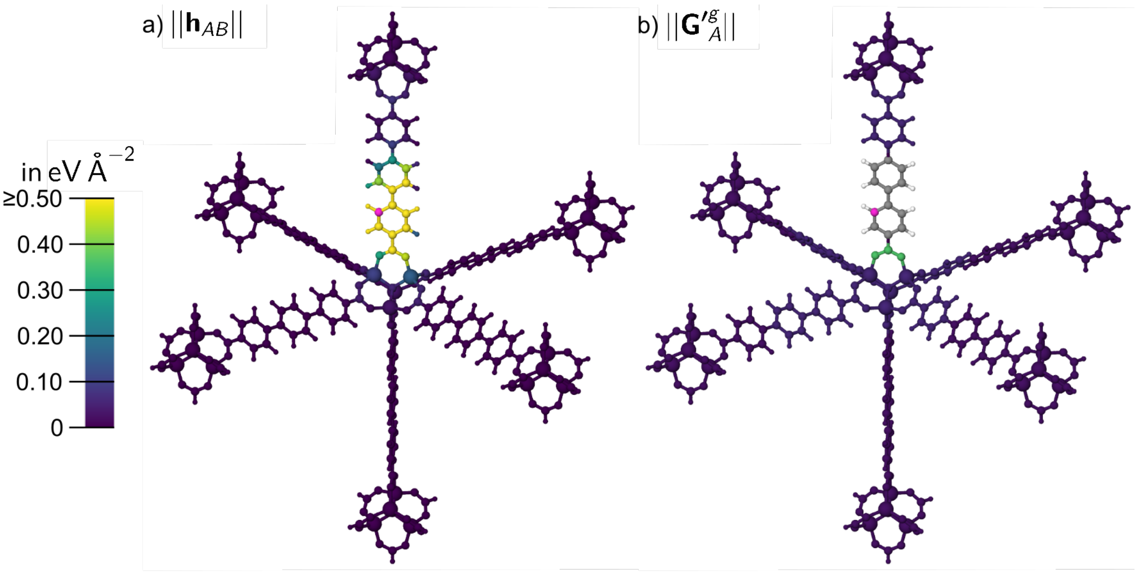

Figure A.32: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C4 (magenta) in reference structure I16$_{\text{ref}}$. $||\mathbf{G}'^{g}_{\text{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

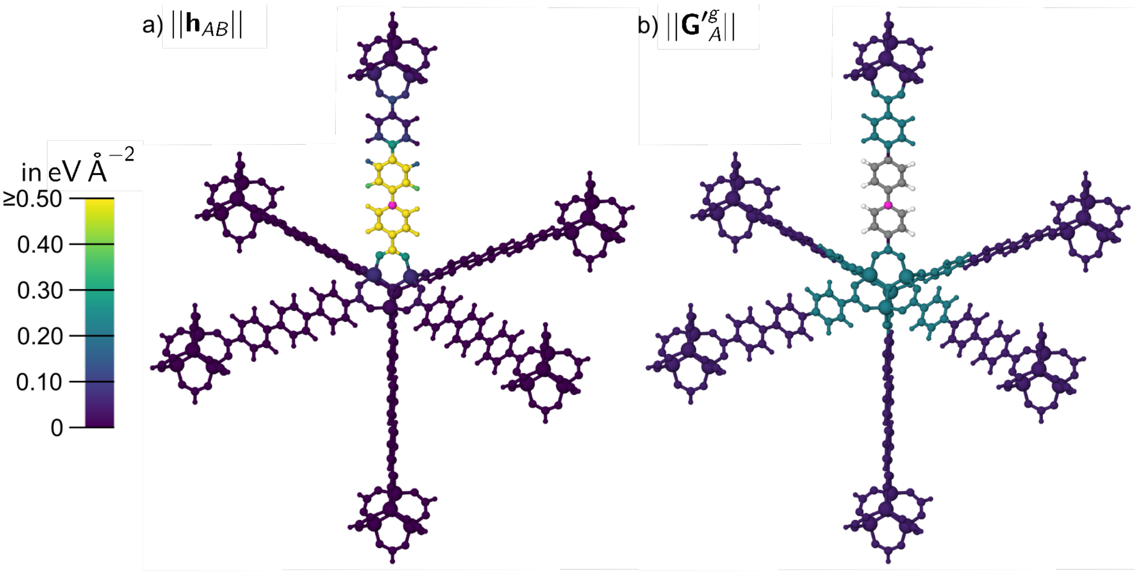

Figure A.33: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C5 (magenta) in reference structure I16$_{\mathrm{ref}}$. $||\mathbf{G}'^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

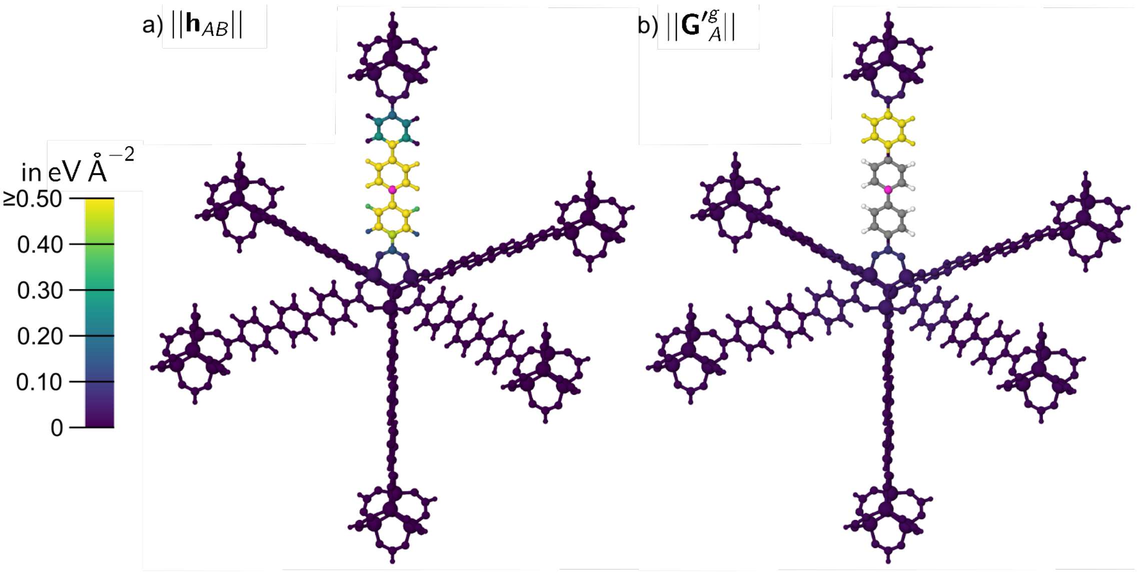

Figure A.34: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C6 (magenta) in reference structure I16$_{\mathrm{ref}}$. $||\mathbf{G}'^{g}_{\mathrm{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

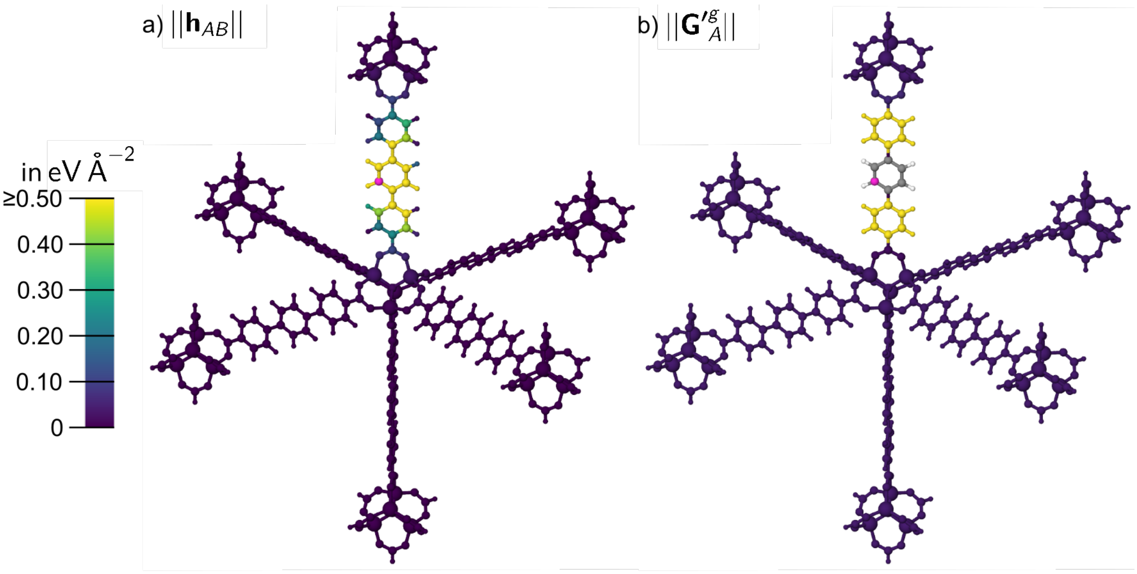

Figure A.35: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = C7 (magenta) in reference structure I16$_{\text{ref}}$. $||\mathbf{G}'^{g}_{\text{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

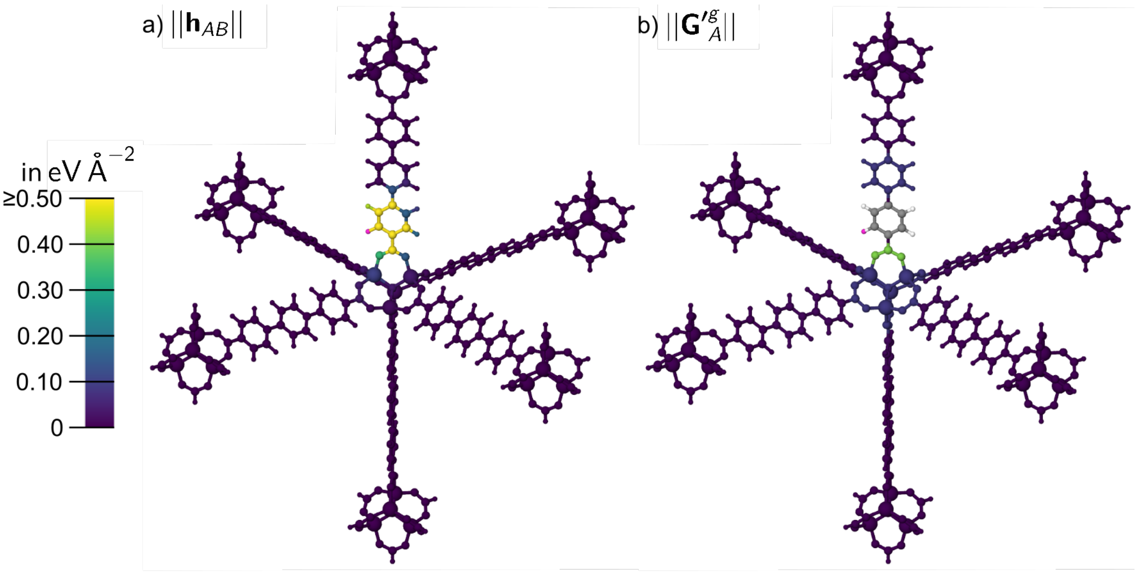

Figure A.36: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G}'^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = H1 (magenta) in reference structure I16$_{\text{ref}}$. $||\mathbf{G}'^{g}_{\text{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

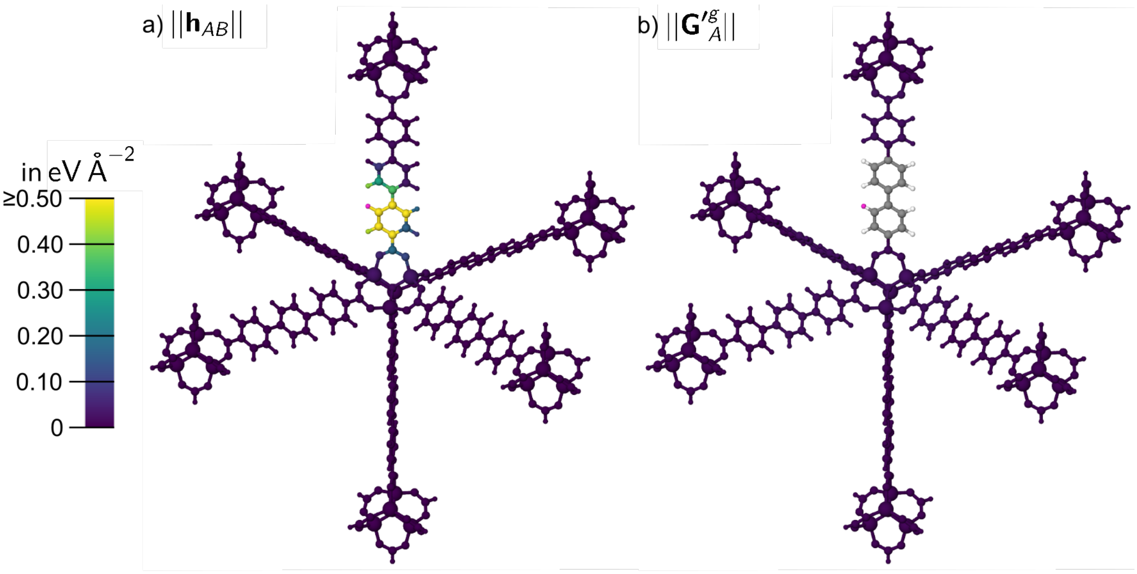

Figure A.37: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G'}^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = H2 (magenta) in reference structure I16$_{\text{ref}}$. $||\mathbf{G'}^{g}_{\text{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

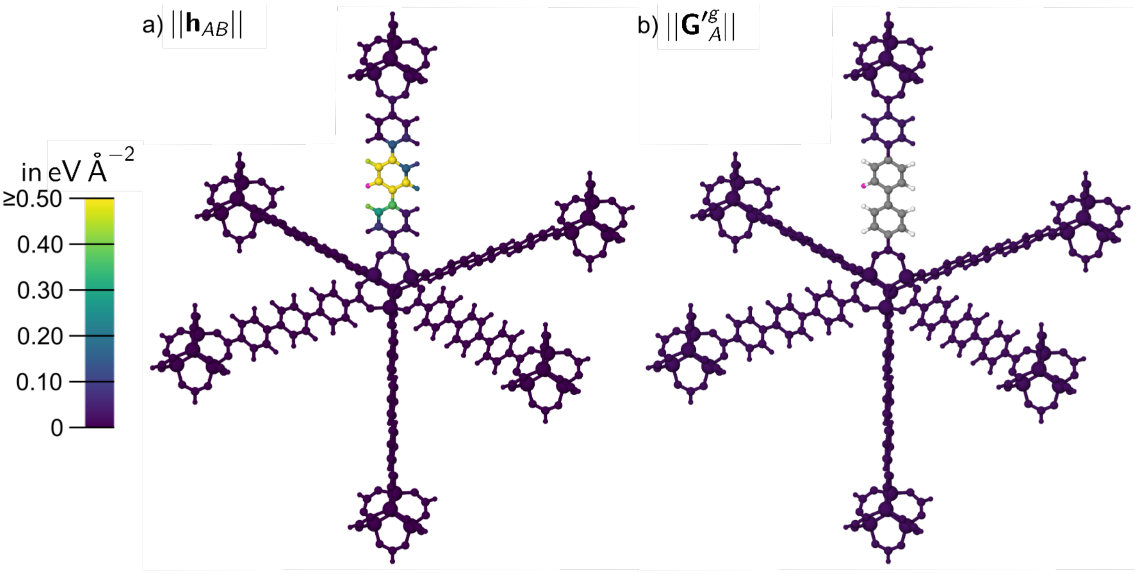

Figure A.38: a) The atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ and b) the effective Hessian group matrix norm values $||\mathbf{G'}^{g}_{A}||$ in eV Å^{-2} with respect to the central atom A = H3 (magenta) in reference structure I16$_{\text{ref}}$. $||\mathbf{G'}^{g}_{\text{A}}||$ defines the color for the closest atoms of a given group, which in addition also contains all atoms at larger distance. The colors of the smallest possible fragment in b) refer to the chemical elements.

Table A.9: Compilation of the force component errors $\Delta f^{\mathrm{I16}}_{A_{x,y,z}}$ and the total force errors $||\Delta\mathbf{f}^{\mathrm{I16}}_{A}||$ in eV Å^{-1} for the IRMOF-16 reference atoms Zn1, O1, O2, C1, C2, C3, C4, C5, C6, C7, H1, H2 and H3 of the reference structure I16 (fig. 4.19 and A.26 to A.37). Further, the effective Hessian group matrix norm $||\mathbf{G}'^{g}_{A}||$ is given in eV Å^{-2}. Numbers outside the intended convergence are given in bold.

A/Y	g	$\Delta f^{\mathrm{I16}}_{A_\mathrm{x}}$	$\Delta f^{\mathrm{I16}}_{A_\mathrm{y}}$	$\Delta f^{\mathrm{I16}}_{A_\mathrm{z}}$	$\|\|\Delta\mathbf{f}^{\mathrm{I16}}_{A}\|\|$	$\|\|\mathbf{G}'^{g}_{A}\|\|$
Zn1	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0100	0.0101	0.0082	0.0164	0.11
	2	-0.0008	0.0008	0.0008	0.0014	0.11
	3	-0.0006	0.0005	0.0005	0.0009	0.03
	4	0.0065	-0.0063	-0.0062	0.0110	**0.73**
O1	ref	0.0328	-0.0305	-0.0299	0.0539	0.00
	1	0.0327	-0.0290	-0.0300	0.0530	0.09
	2	0.0328	-0.0305	-0.0299	0.0539	0.09
	3	0.0328	-0.0305	-0.0299	0.0539	0.02
	4	0.0328	-0.0305	-0.0300	0.0539	0.07
	5	0.0328	-0.0305	-0.0299	0.0538	**0.42**
O2	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0066	0.0066	-0.0205	0.0226	0.09
	2	-0.0003	0.0003	0.0066	0.0066	0.01
	3	-0.0230	0.0230	0.0089	0.0337	0.17
	4	0.0002	-0.0001	0.0598	0.0598	0.26
	5	**−0.1579**	**0.1551**	**−0.5901**	**0.6302**	**3.57**
C1	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0012	0.0013	0.0317	0.0318	0.14
	2	0.0000	-0.0001	-0.0244	0.0244	0.11
	3	-0.0001	-0.0001	-0.0308	0.0308	0.04
	4	0.0000	-0.0001	-0.0007	0.0008	**0.40**
C2	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0000	-0.0001	-0.0099	0.0099	0.06
	2	0.0001	0.0001	0.0190	0.0190	0.06
	3	0.0001	0.0001	0.0293	0.0293	0.07
	4	**0.5438**	**−0.5444**	-0.0828	**0.7739**	**0.77**
C3	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0007	0.0008	0.0019	0.0021	0.02
	2	0.0005	-0.0005	-0.0048	0.0049	0.02
	3	0.0068	-0.0067	0.0101	0.0139	0.04
	4	-0.0416	0.0421	0.1013	0.1174	**0.50**
C4	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0008	0.0008	-0.0012	0.0016	0.02
	2	0.0175	-0.0178	-0.0772	0.0811	0.06
	3	-0.0068	0.0060	**−0.1458**	**0.1461**	**0.36**
C5	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0048	-0.0012	-0.0005	0.0050	0.05
	2	-0.0048	-0.0012	0.0018	0.0052	0.22

Table A.10: Continuation of table A.9. Compilation of the force component errors $\Delta f^{\mathrm{I16}}_{A_{x,y,z}}$ and the total force errors $||\Delta \mathbf{f}^{\mathrm{I16}}_{A}||$ in eV Å^{-1} for the IRMOF-16 reference atoms Zn1, O1, O2, C1, C2, C3, C4, C5, C6, C7, H1, H2 and H3 of the reference structure I16 (fig. 4.19 and A.26 to A.37). Further, the effective Hessian group matrix norm $||\mathbf{G}'^{g}_{A}||$ is given in eV Å^{-2}. Numbers outside the intended convergence are given in bold.

A/Y	g	$\Delta f^{Y_g}_{A_\mathrm{x}}$	$\Delta f^{Y_g}_{A_\mathrm{y}}$	$\Delta f^{Y_g}_{A_\mathrm{z}}$	$\|\|\Delta \mathbf{f}^{Y_g}_{A}\|\|$	$\|\|\mathbf{G}'^{g}_{A}\|\|$
C7	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	0.0006	-0.0006	-0.0045	0.0046	0.04
	2	-0.0104	0.0107	0.0069	0.0164	0.06
	3	-0.0268	0.0268	0.0164	0.0412	0.02
	4	-0.0268	0.0268	0.0164	0.0412	0.02
	5	-0.0268	0.0268	0.0164	0.0412	0.02
	6	-0.0369	0.0341	**0.1769**	**0.1839**	**0.52**
	7	-0.1091	0.1101	**0.6727**	**0.6903**	**2.28**
H1	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0391	0.0309	-0.0298	0.0581	0.01
	2	-0.0396	0.0314	-0.0286	0.0580	0.01
	3	-0.0387	0.0305	-0.0288	0.0570	0.02
	4	-0.0357	0.0275	-0.0342	0.0566	0.09
	5	-0.0034	-0.0046	-0.0847	0.0849	**0.41**
H2	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0360	0.0362	0.0214	0.0554	0.01
	2	-0.0356	0.0358	0.0217	0.0549	0.02
	3	-0.0378	0.0379	0.0085	0.0542	0.02
H3	ref	0.0000	0.0000	0.0000	0.0000	0.00
	1	-0.0399	0.0386	-0.0201	0.0590	0.02
	2	-0.0367	0.0354	-0.0185	0.0542	0.02
	3	-0.0370	0.0358	-0.0154	0.0538	0.04

A.6 Comparison Hessian Reference Systems

The approach of using sufficiently large fragment structures[1] for the Hessian analysis reference structure instead of the periodic bulk structure is validated by the comparison of the atomic Hessian sub matrix norms calculated from the large fragment $\mathrm{C1_{ref}}$ (fig. 4.13a) and the bulk structure (fig. 4.1a, short note: results are calculated with derived DFT settings, described in section A.1 and 3.1, are used and not the *tight*-like settings used for molecular numerical Hessian) which is illustrated in figure A.39. The atomic Hessian sub matrix norms, calculated from the $1\times1\times1$ unit cell, are shown in a $2\times2\times2$ supercell (fig. A.39b). All atoms included in $\mathrm{C1_{ref}}$ are marked by the grey background (fig. A.39b) embedded in the IRMOF-1 bulk environment. A direct comparison of the $\mathrm{C1_{ref}}$ fragment (fig. A.39c and fig. 4.13a) and the periodic results (fig. A.39d) show only very small differences (tab. A.11, $\Delta||\mathbf{h}_{AB}|| \lesssim 0.080\,\mathrm{eV\,Å^{-2}}$). Only the reference atom itself $B = 4$ and the directly neighboring oxygen $B = 11$ show larger deviations for the atomic Hessian sub matrix norms. However, these deviations do not effect the determination of the size-converged fragments.

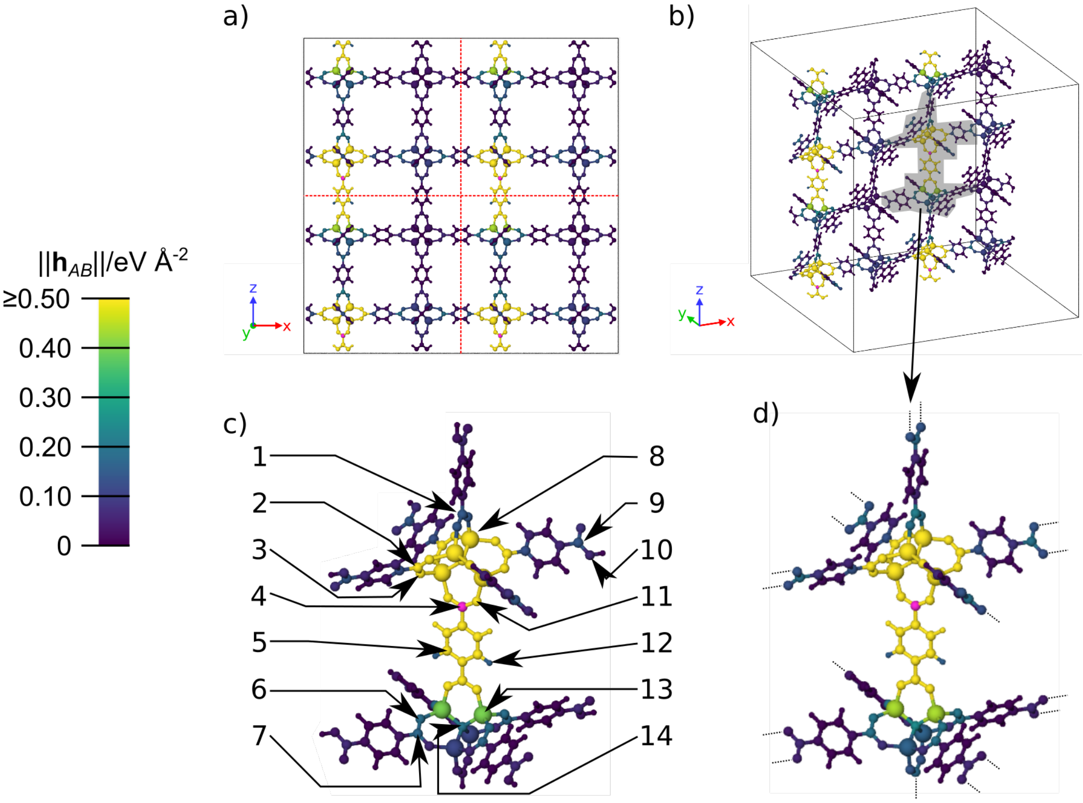

Figure A.39: a) An in-plane view of the periodic C1 (magenta) atomic Hessian sub matrix norm values of the IRMOF-1 bulk unit cell, shown by a $2\times2\times2$ replication of the simple unit cell (red dashed lines separate the simple unit cells), b) bent view of a 2D-slab cutout with a grey-marked $\mathrm{C1_{ref}}$ structure embedded in the periodic IRMOF-1 environment, c) C1 results of Fig. 11a with some marked atoms ($B = 1 - 14$) and d) the grey-marked $\mathrm{C1_{ref}}$ structure from b) without the surrounding environment and marked broken bonds by the dashed lines. Adapted from [108] with permission from ©2022 AIP Publishing.

[1] The results discussed and presented in this section were obtained in my recent publication [108] and are shown for completeness. Adapted from [108] with permission from ©2022 AIP Publishing.

Table A.11: Compilation of the atomic Hessian sub matrix norm values $||\mathbf{h}_{AB}||$ of the atoms $B = 1 - 14$ marked in fig. A.39c calculated by the reference structures $\mathrm{C1_{ref}}$ and the periodic bulk unit cell (fig. 4.1a), as well as the differences of these values $\Delta||\mathbf{h}_{AB}||$ given in $\mathrm{eV\,\AA^{-2}}$. Adapted from [108] with permission from ©2022 AIP Publishing.

B	element	$\|\|\mathbf{h}_{AB}^{\mathrm{C1_{ref}}}\|\|$	$\|\|\mathbf{h}_{AB}^{\mathrm{bulk}}\|\|$	$\Delta\|\|\mathbf{h}_{AB}\|\|$
1	C	0.143	0.218	−0.075
2	C	1.170	1.205	−0.036
3	O	1.218	1.265	−0.047
4	C	94.885	95.788	−0.903
5	C	1.098	1.106	−0.008
6	O	0.185	0.202	−0.017
7	C	0.172	0.193	−0.021
8	Zn	0.558	0.563	−0.005
9	C	0.142	0.193	−0.052
10	O	0.063	0.129	−0.066
11	O	42.074	42.630	−0.556
12	H	0.155	0.153	0.002
13	Zn	0.402	0.439	−0.037
14	O	0.200	0.244	−0.045

A.7 RuNNer Atom-Centered Symmetry Functions

Table A.12: Compilation of the radial atom-centered symmetry functions (ACSF) as given in equation 2.57 for the different element combinations of the ACSF set SF-1. A defines the element of the atom at which the ACSF is centered and B the element of the neighboring atom. The η parameter defining the width of the Gaussian is given in a_0^{-2} with the cutoff radius $r_{\text{cut}} = 4.359\,\text{Å} = 8.237\,\mathrm{a}_0$, the shifting parameter $r_{\text{shift}} = 0.000\,\text{Å} = 0.000\,\mathrm{a}_0$ and the inner cutoff radius $r_{\text{inner,cut}} = 0.000\,\text{Å} = 0.000\,\mathrm{a}_0$.

no.	A	B	η	A	B	η	A	B	η	A	B	η
1	C	H	0.000	O	H	0.000	Zn	C	0.000	H	H	0.000
2	C	C	0.000	O	C	0.000	Zn	O	0.000	H	C	0.000
3	C	O	0.000	O	O	0.000	Zn	Zn	0.000	H	O	0.000
4	C	Zn	0.000	O	Zn	0.000	Zn	Zn	0.001	H	C	0.007
5	C	Zn	0.002	O	O	0.003	Zn	C	0.002	H	O	0.010
6	C	Zn	0.003	O	Zn	0.004	Zn	Zn	0.002	H	H	0.014
7	C	Zn	0.005	O	C	0.006	Zn	C	0.003	H	C	0.020
8	C	C	0.006	O	O	0.007	Zn	Zn	0.003	H	O	0.041
9	C	O	0.006	O	Zn	0.009	Zn	O	0.004	H	C	0.053
10	C	H	0.007	O	H	0.010	Zn	Zn	0.004	H	H	0.079
11	C	Zn	0.007	O	O	0.013	Zn	C	0.005	H	C	0.182
12	C	C	0.015	O	C	0.016	Zn	C	0.007			
13	C	O	0.016	O	Zn	0.017	Zn	O	0.009			
14	C	H	0.020	O	O	0.021	Zn	O	0.017			
15	C	C	0.034	O	Zn	0.031	Zn	O	0.031			
16	C	O	0.039	O	C	0.039						
17	C	H	0.053	O	H	0.041						
18	C	C	0.084	O	C	0.106						
19	C	O	0.106									
20	C	H	0.182									

Table A.13: Compilation of the angular atom-centered symmetry functions (ACSFs) as given in equation 2.58 for the different element combinations of the ACSF set SF-1. A defines the element of the atom at which the ACSF is centered, B and C the element of the neighboring atoms. The η parameter defining the width of the Gaussian is defined as $\eta = 0.000\,\mathrm{a}_0^{-2}$ with the cutoff radius $r_{\text{cut}} = 4.359\,\text{Å} = 8.237\,\mathrm{a}_0$ and the inner cutoff radius $r_{\text{inner,cut}} = 0.000\,\text{Å} = 0.000\,\mathrm{a}_0$. The set of angular ACSF of each element combination is expanded by all different combinations of the parameter $\zeta \in \{1, 2, 4, 16\}$, defining the width of the cosine part and the parameter $\lambda \in \{-1, 1\}$ inverting the cosine part. For the element combinations *C-Zn-Zn*, *Zn-C-C*, *Zn-C-Zn*, *Zn-O-Zn* and *Zn-Zn-Zn* the angular ACSF with the parameter combination of $\eta/\lambda = -1/16$ are neglected, since these do not provide any input information for the HDNNP for the underlying data set.

no.	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
1	C	H	H	O	H	C	Zn	C	C	H	H	C
2	C	H	C	O	H	O	Zn	C	O	H	C	C
3	C	H	O	O	C	C	Zn	C	Zn	H	C	O
4	C	H	Zn	O	C	O	Zn	O	O	H	O	O
5	C	C	C	O	C	Zn	Zn	O	Zn			
6	C	C	O	O	O	O	Zn	Zn	Zn			
7	C	C	Zn	O	O	Zn						
8	C	O	O	O	Zn	Zn						
9	C	O	Zn									
10	C	Zn	Zn									

Table A.14: Compilation of the radial atom-centered symmetry functions (ACSFs) as given in equation 2.57 for the different element combinations of the ACSF set SF-2. A defines the element of the atom at which the ACSF is centered and B the element of the neighboring atom. The η parameter defining the width of the Gaussian is given in a_0^{-2} with the cutoff radius $r_{\mathrm{cut}} = 8.718\,\text{Å}= 16.475\,\mathrm{a}_0$, the shifting parameter $r_{\mathrm{shift}} = 0.000\,\text{Å}= 0.000\,\mathrm{a}_0$ and the inner cutoff radius $r_{\mathrm{inner,cut}} = 0.000\,\text{Å}= 0.000\,\mathrm{a}_0$.

no.	A	B	η	A	B	η	A	B	η	A	B	η
1	C	H	0.000	O	H	0.000	Zn	C	0.000	H	H	0.000
2	C	C	0.000	O	C	0.000	Zn	O	0.000	H	C	0.000
3	C	O	0.000	O	O	0.000	Zn	Zn	0.000	H	O	0.000
4	C	Zn	0.000	O	Zn	0.000	Zn	C	0.001	H	C	0.001
5	C	H	0.001	O	C	0.001	Zn	O	0.001	H	C	0.002
6	C	C	0.001	O	O	0.001	Zn	Zn	0.001	H	C	0.003
7	C	O	0.001	O	Zn	0.001	Zn	C	0.002	H	C	0.004
8	C	Zn	0.001	O	C	0.002	Zn	O	0.002	H	C	0.006
9	C	H	0.002	O	O	0.002	Zn	Zn	0.002	H	C	0.007
10	C	C	0.002	O	Zn	0.002	Zn	C	0.003	H	H	0.010
11	C	O	0.002	O	C	0.003	Zn	O	0.003	H	C	0.010
12	C	Zn	0.002	O	O	0.003	Zn	Zn	0.003	H	C	0.012
13	C	H	0.003	O	Zn	0.003	Zn	C	0.004	H	O	0.013
14	C	C	0.003	O	C	0.004	Zn	O	0.004	H	C	0.016
15	C	O	0.003	O	O	0.004	Zn	Zn	0.004	H	C	0.021
16	C	Zn	0.003	O	Zn	0.004	Zn	C	0.005	H	C	0.027
17	C	H	0.004	O	C	0.005	Zn	O	0.005	H	C	0.036
18	C	C	0.004	O	O	0.005	Zn	Zn	0.005	H	C	0.049
19	C	O	0.004	O	Zn	0.005	Zn	C	0.006	H	C	0.070
20	C	Zn	0.004	O	O	0.006	Zn	O	0.006	H	H	0.082
21	C	C	0.005	O	Zn	0.006	Zn	Zn	0.006	H	C	0.102
22	C	O	0.005	O	C	0.007	Zn	C	0.007	H	C	0.161
23	C	Zn	0.005	O	O	0.007	Zn	O	0.007	H	O	0.194
24	C	H	0.006	O	Zn	0.007	Zn	Zn	0.007			
25	C	C	0.006	O	O	0.008	Zn	C	0.008			
26	C	Zn	0.006	O	Zn	0.008	Zn	O	0.008			
27	C	H	0.007	O	C	0.009	Zn	Zn	0.008			
28	C	O	0.007	O	O	0.009	Zn	C	0.009			
29	C	Zn	0.007	O	O	0.010	Zn	Zn	0.009			
30	C	C	0.008	O	Zn	0.010	Zn	C	0.010			
31	C	Zn	0.008	O	C	0.011	Zn	O	0.010			
32	C	O	0.009	O	O	0.011	Zn	Zn	0.010			
33	C	Zn	0.009	O	O	0.012	Zn	C	0.011			
34	C	H	0.010	O	Zn	0.012	Zn	Zn	0.011			
35	C	C	0.010	O	H	0.013	Zn	C	0.012			
36	C	Zn	0.010	O	C	0.014	Zn	O	0.012			
37	C	O	0.011	O	O	0.015	Zn	Zn	0.012			
38	C	Zn	0.011	O	Zn	0.015	Zn	C	0.013			
39	C	H	0.012	O	C	0.018	Zn	Zn	0.013			
40	C	Zn	0.012	O	O	0.018	Zn	C	0.014			

Table A.15: Continuation of table A.14. Compilation of the radial atom-centered symmetry functions (ACSFs) as given in equation 2.57 for the different element combinations of the ACSF set SF2. A defines the element of the atom at which the ACSF is centered and B the element of the neighboring atom. The η parameter defining the width of the Gaussian is given in a_0^{-2} with the cutoff radius $r_{\mathrm{cut}} = 8.718\,\text{Å} = 16.475\,\mathrm{a}_0$, the shifting parameter $r_{\mathrm{shift}} = 0.000\,\text{Å} = 0.000\,\mathrm{a}_0$ and the inner cutoff radius $r_{\mathrm{inner,cut}} = 0.000\,\text{Å} = 0.000\,\mathrm{a}_0$.

no.	A	B	η	A	B	η	A	B	η	A	B	η
41	C	C	0.013	O	Zn	0.018	Zn	Zn	0.014			
42	C	Zn	0.013	O	O	0.021	Zn	O	0.015			
43	C	O	0.014	O	Zn	0.022	Zn	Zn	0.015			
44	C	Zn	0.014	O	C	0.023	Zn	C	0.016			
45	C	H	0.016	O	O	0.025	Zn	Zn	0.016			
46	C	C	0.016	O	Zn	0.028	Zn	C	0.018			
47	C	Zn	0.016	O	C	0.030	Zn	O	0.018			
48	C	O	0.018	O	Zn	0.034	Zn	O	0.022			
49	C	Zn	0.018	O	C	0.039	Zn	O	0.028			
50	C	H	0.021	O	Zn	0.044	Zn	O	0.034			
51	C	C	0.021	O	C	0.052	Zn	O	0.044			
52	C	O	0.023	O	C	0.073						
53	C	H	0.027	O	C	0.105						
54	C	C	0.027	O	H	0.194						
55	C	O	0.030									
56	C	C	0.035									
57	C	H	0.036									
58	C	O	0.039									
59	C	C	0.046									
60	C	H	0.049									
61	C	O	0.052									
62	C	C	0.062									
63	C	H	0.070									
64	C	O	0.073									
65	C	C	0.086									
66	C	H	0.102									
67	C	O	0.105									
68	C	H	0.161									

Table A.16: Compilation of the angular atom-centered symmetry functions (ACSFs) as given in equation 2.58 for the different element combinations of the ACSF set SF2. *A* defines the element of the atom at which the ACSF is centered, *B* and *C* the element of the neighboring atoms. The η parameter defining the width of the Gaussian is defined as $\eta = 0.000\,\mathrm{a}_0^{-2}$ with the inner cutoff radius $r_{\text{inner,cut}} = 0.000\,\text{Å}= 0.000\,\mathrm{a}_0$. There are two shells of angular ACSF: one shell is similar to SF1 (tab. A.13) with the cutoff radius $r_{\text{cut}} = 4.359\,\text{Å}= 8.237\,\mathrm{a}_0$ and a second shell with the cutoff radius $r_{\text{cut}} = 8.718\,\text{Å}= 16.475\,\mathrm{a}_0$. The set of angular ACSF of each element combination is expanded by all different combinations of the parameter $\zeta \in \{1, 2, 4, 16\}$, defining the width of the cosine part and the parameter $\lambda \in \{-1, 1\}$ inverting the cosine part. For the element combinations *C-Zn-Zn*, *O-O-O*, *Zn-C-C*, *Zn-C-Zn*, *Zn-O-Zn*, *Zn-Zn-Zn* and *H-O-O* the angular ACSF with the parameter combination of $\eta/\lambda = -1/16$ are neglected, since these do not provide any input information for the HDNNP for the underlying data set.

no.	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
1	C	H	H	O	H	C	Zn	C	C	H	H	C
2	C	H	C	O	H	O	Zn	C	O	H	C	C
3	C	H	O	O	C	C	Zn	C	Zn	H	C	O
4	C	H	Zn	O	C	O	Zn	O	O	H	O	O
5	C	C	C	O	C	Zn	Zn	O	Zn			
6	C	C	O	O	O	O	Zn	Zn	Zn			
7	C	C	Zn	O	O	Zn						
8	C	O	O	O	Zn	Zn						
9	C	O	Zn									
10	C	Zn	Zn									

A.8 Predictions of the HDNNPs

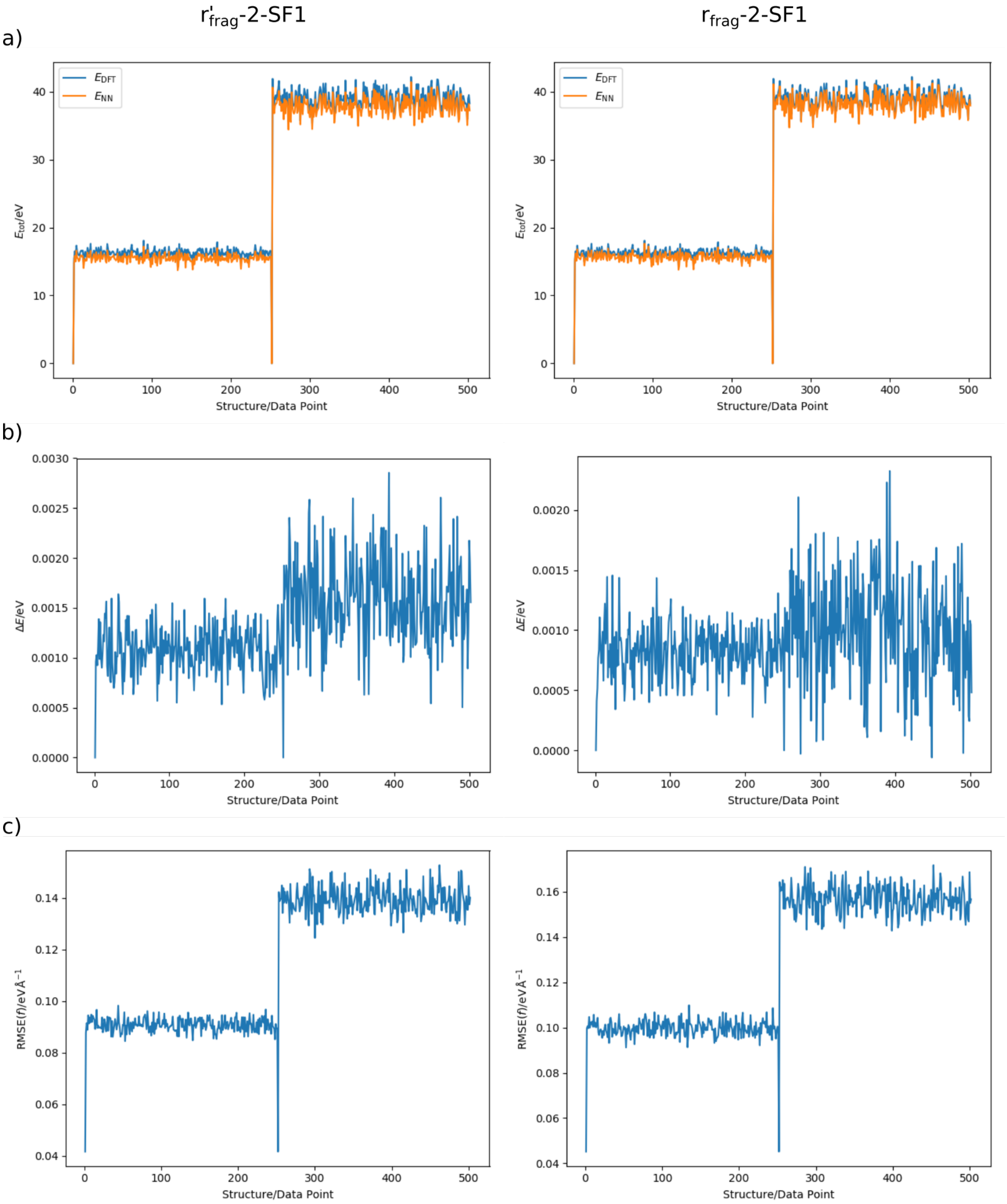

Figure A.40: Compilation of the IRMOF-10 bulk predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) of a HDNNP training independent data set (502 structures/data points), combined of 251 data points for each of two MD simulations in the NPT ensemble (sec. 3.4) at normal pressure and temperatures $T \in \{200, 450\}$ in K. a) Demonstrates the total potential energy of the glsirmof-1 bulk structure over the data set. The two independent MD simulations can be separated by the two starting points of the simulations (data point 1 and 252). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

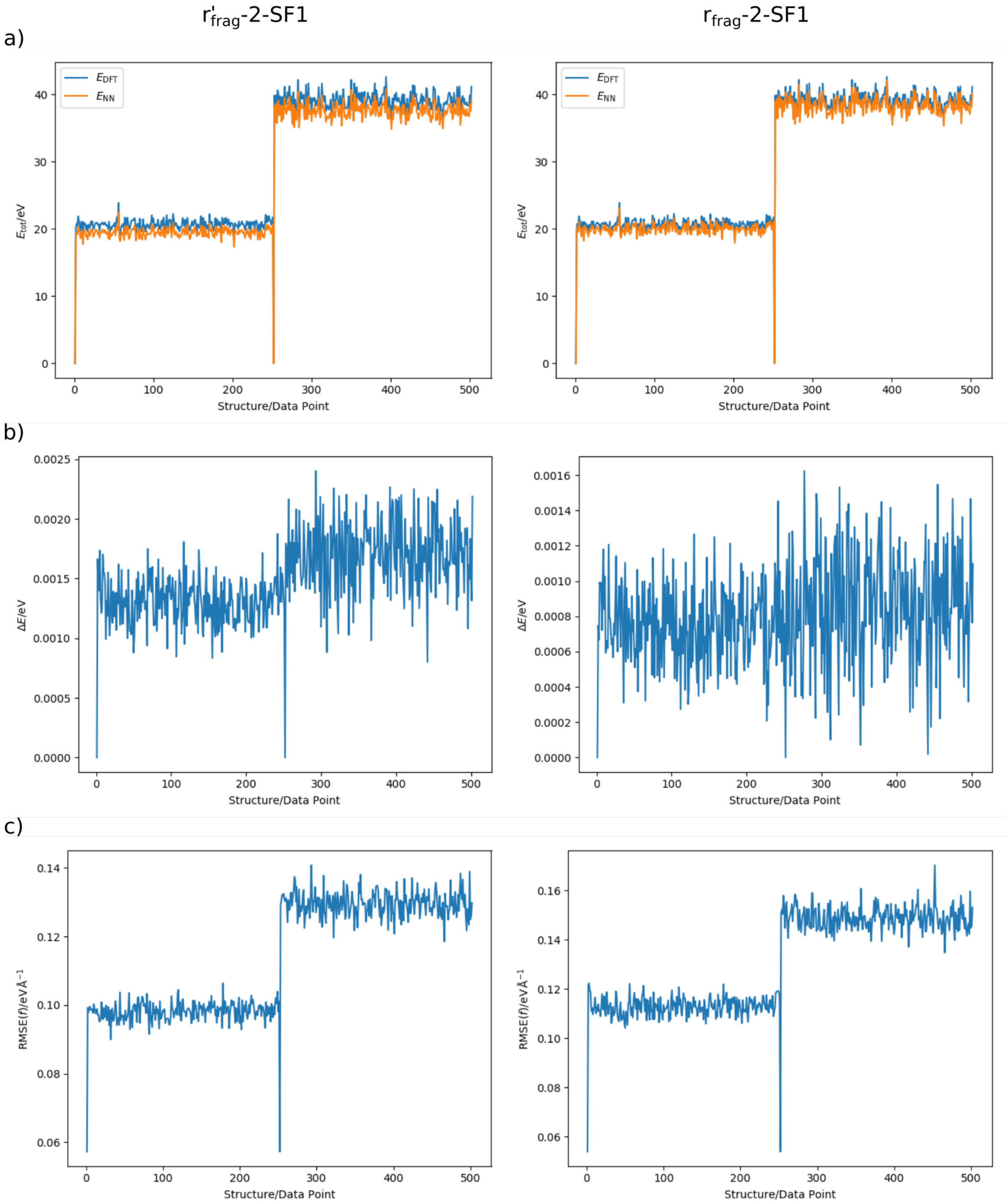

Figure A.41: Compilation of the IRMOF-16 bulk predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) of a HDNNP training independent data set (502 structures/data points), combined of 251 data points for each of two MD simulations in the NPT ensemble (sec. 3.4) at normal pressure and temperatures $T \in \{200, 350\}$ in K. a) Demonstrates the total potential energy of the glsirmof-1 bulk structure over the data set. The two independent MD simulations can be separated by the two starting points of the simulations (data point 1 and 252). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

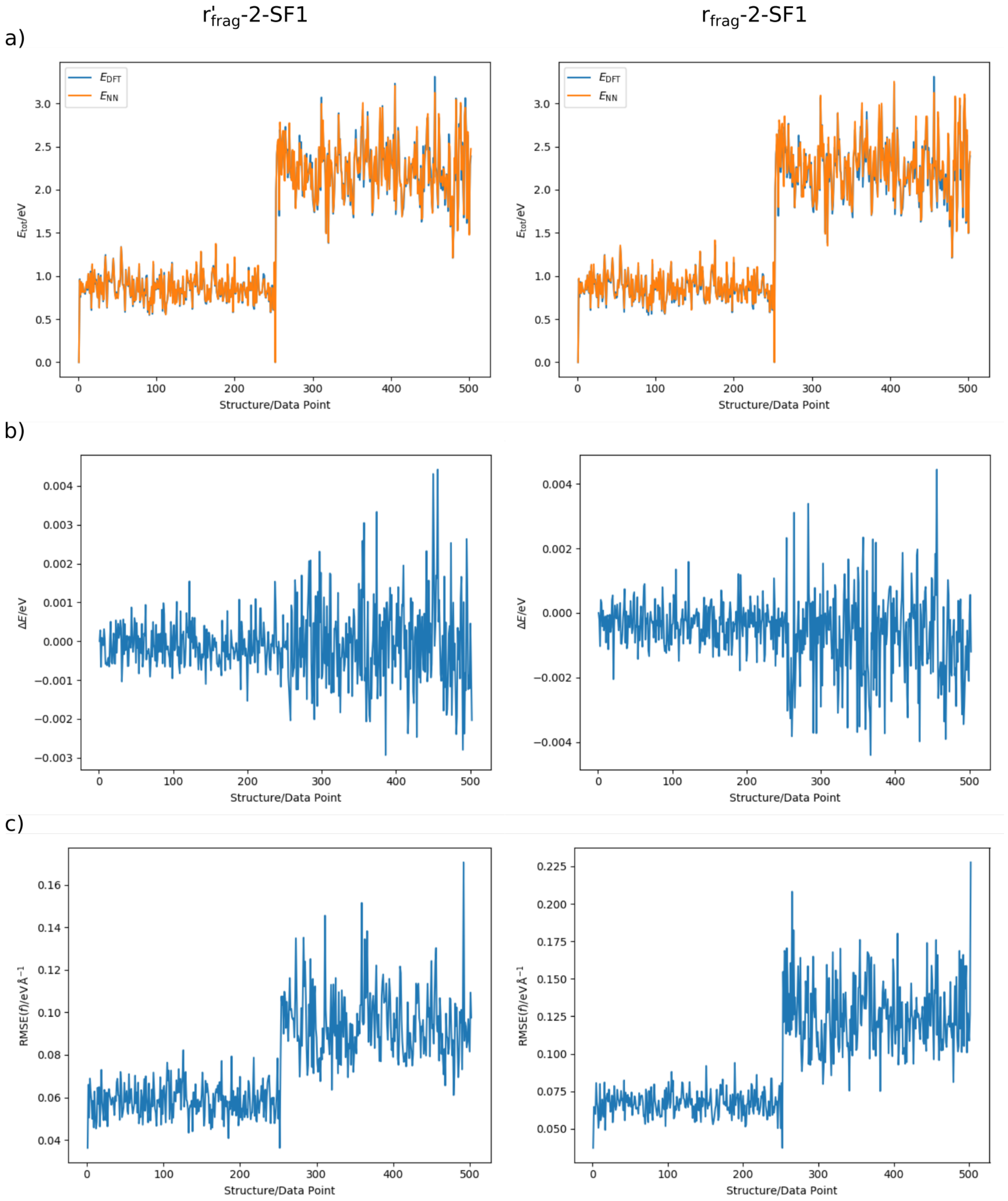

Figure A.42: Compilation of the I1-Bs fragment (fig. 4.24) predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) based on the HDNNP training independent data sets for IRMOF-1 (in total 502 I1-Bs fragment structures/data points). a) Demonstrates the total potential energy of the I1-Bs fragment structure over the data set. The individual MD simulations can be separated by the starting points of the simulations (data point 1 and 252). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

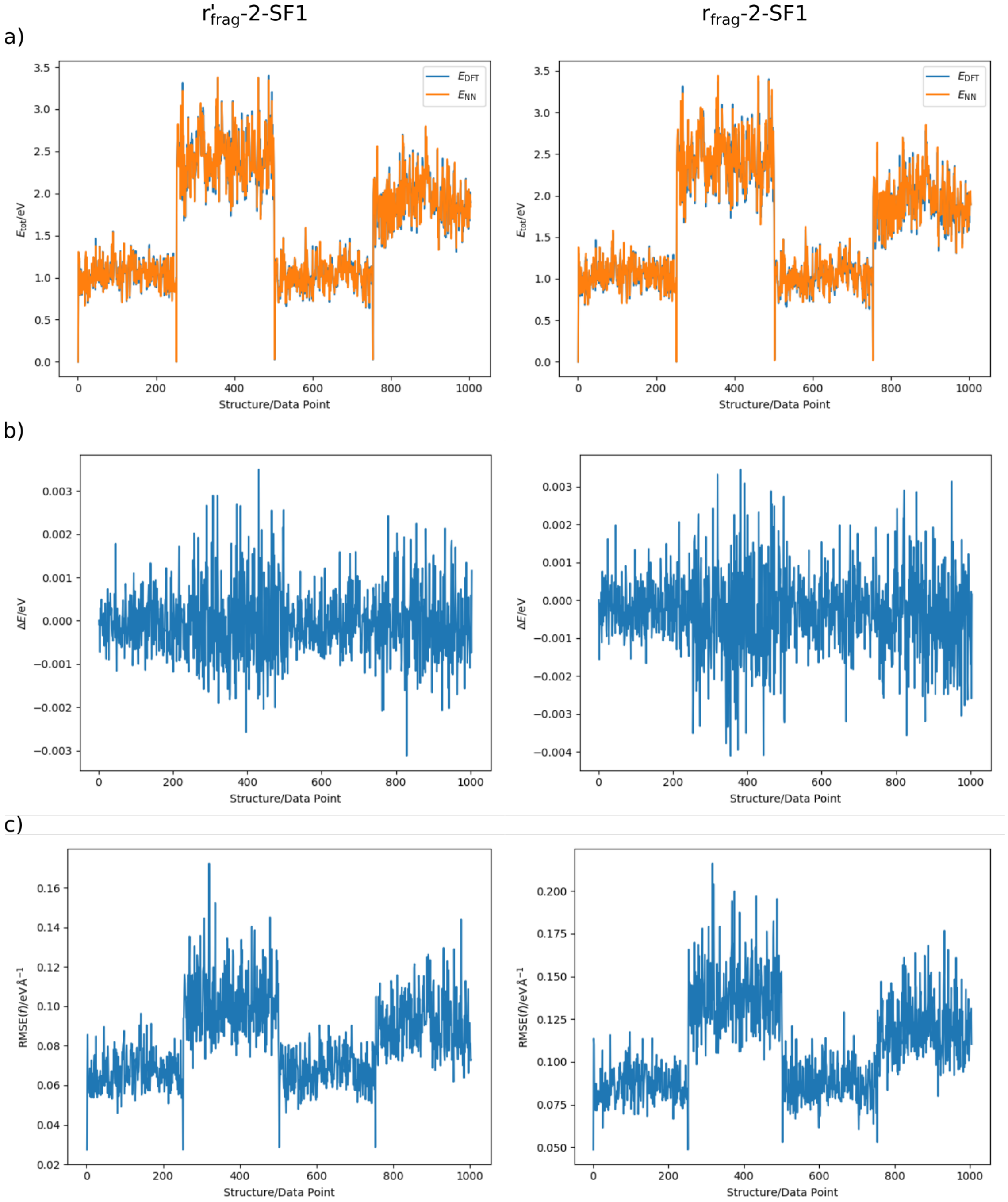

Figure A.43: Compilation of the I10-Bs fragment (fig. 4.24) predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) based on the HDNNP training independent data sets for IRMOF-1 and -16 (in total 1004 I10-Bs fragment structures/data points, 502 for the individual IRMOFs). a) Demonstrates the total potential energy of the I10-Bs fragment structure over the data set. The individual MD simulations can be separated by the starting points of the simulations (data point 1, 252, 503 and 754). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

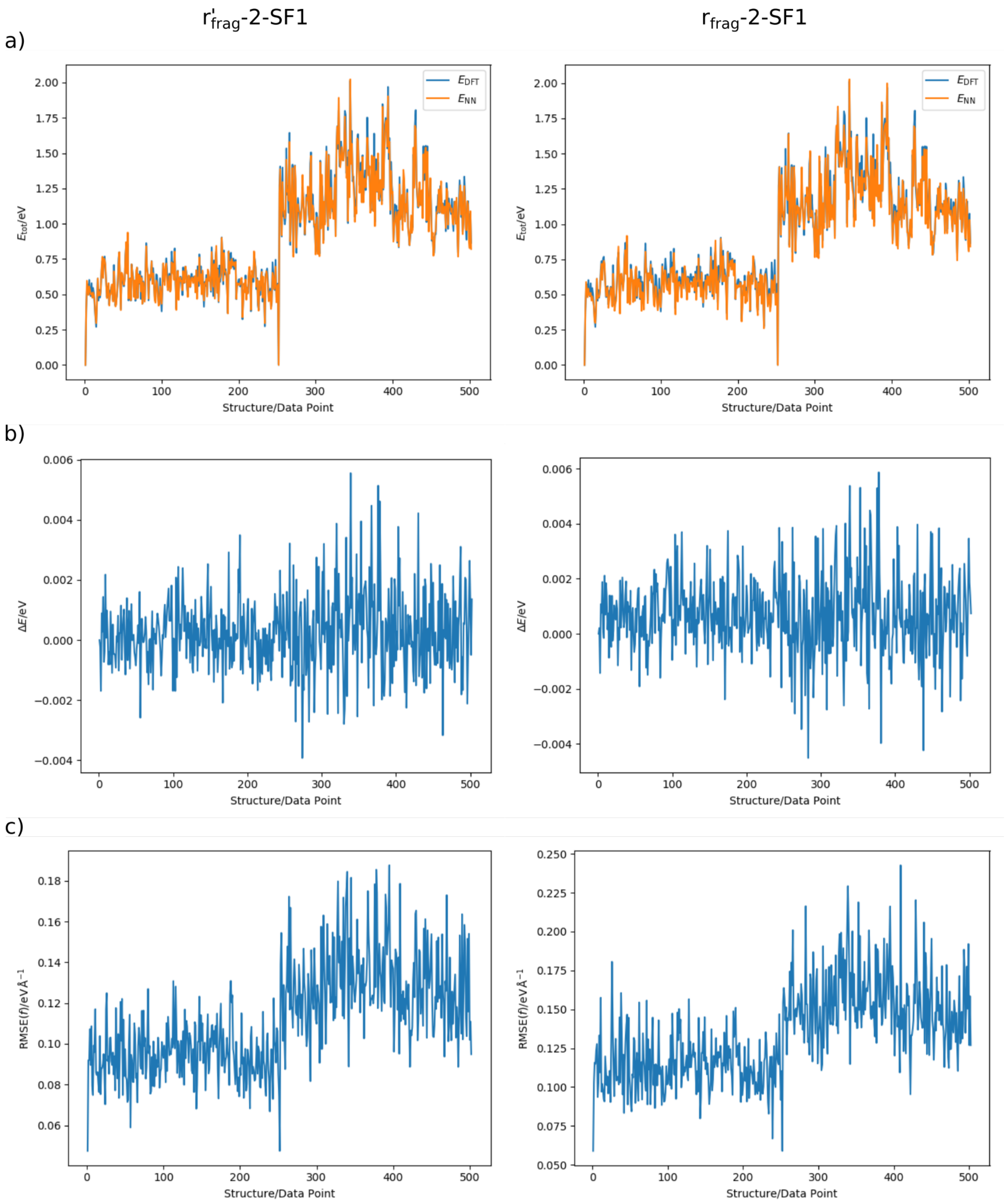

Figure A.44: Compilation of the I16-Cs fragment (fig. 4.24) predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) based on the HDNNP training independent data sets for IRMOF-16 (in total 502 I16-Cs fragment structures/data points). a) Demonstrates the total potential energy of the I16-Cs fragment structure over the data set. The individual MD simulations can be separated by the starting points of the simulations (data point 1 and 252). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

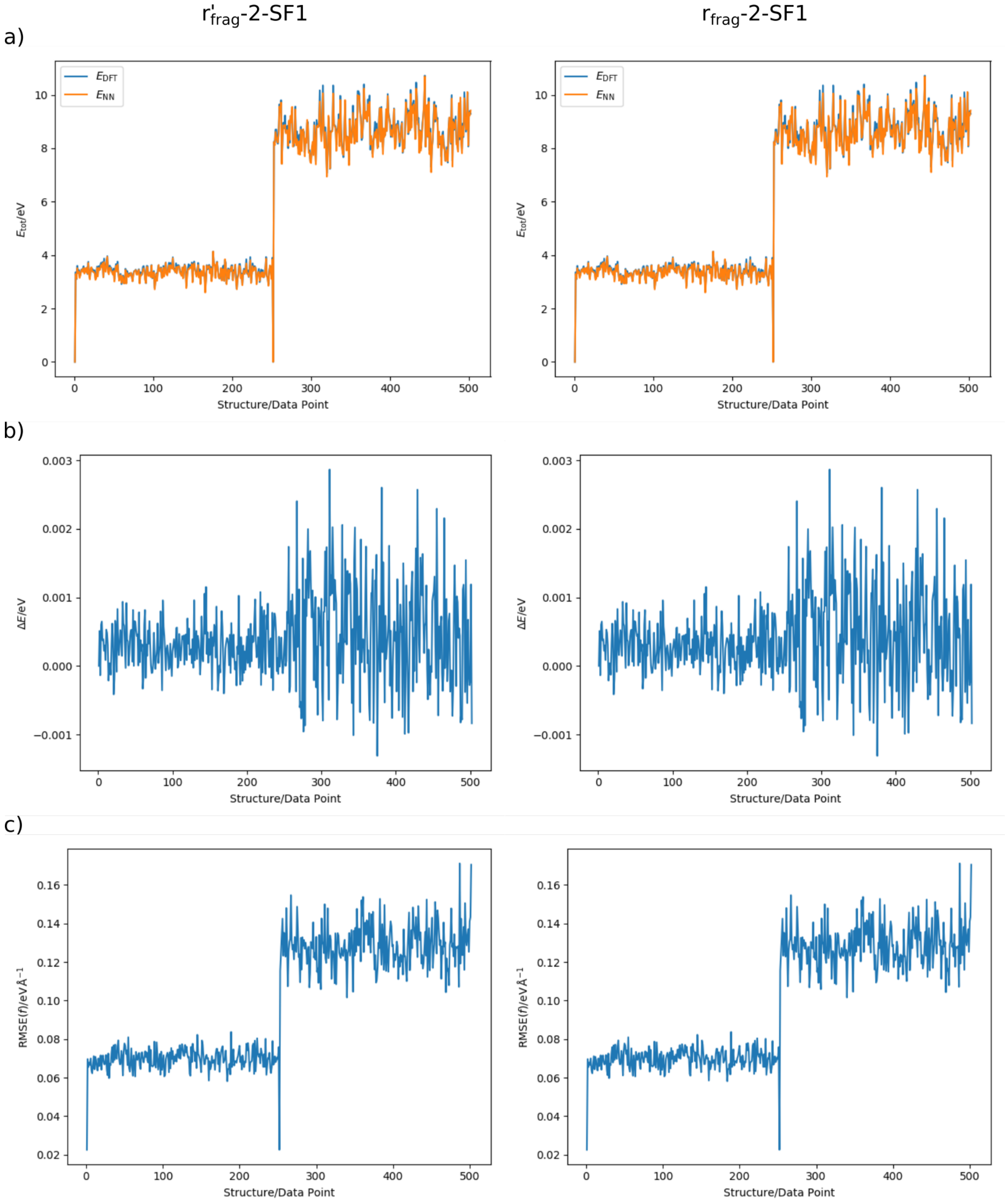

Figure A.45: Compilation of the I1-B’ fragment (fig. 4.22) predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) based on the HDNNP training independent data sets for IRMOF-1 (in total 502 I1-B’ fragment structures/data points). a) Demonstrates the total potential energy of the I1-B’ fragment structure over the data set. The individual MD simulations can be separated by the starting points of the simulations (data point 1 and 252). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

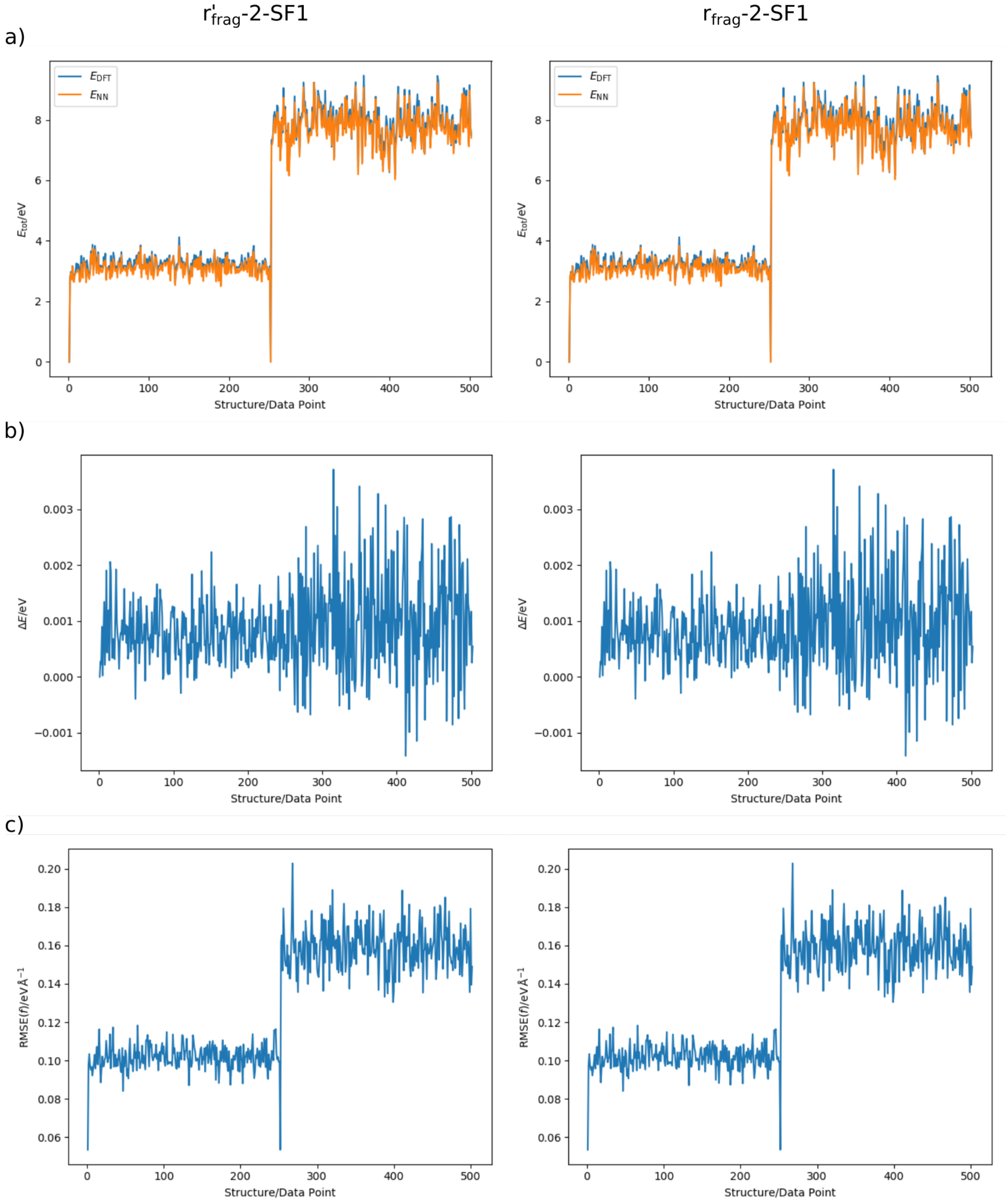

Figure A.46: Compilation of the I10-A' fragment (fig. 4.22) predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) based on the HDNNP training independent data sets for IRMOF-10 (in total 502 I10-A' fragment structures/data points). a) Demonstrates the total potential energy of the I10-A' fragment structure over the data set. The individual MD simulations can be separated by the starting points of the simulations (data point 1 and 252). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

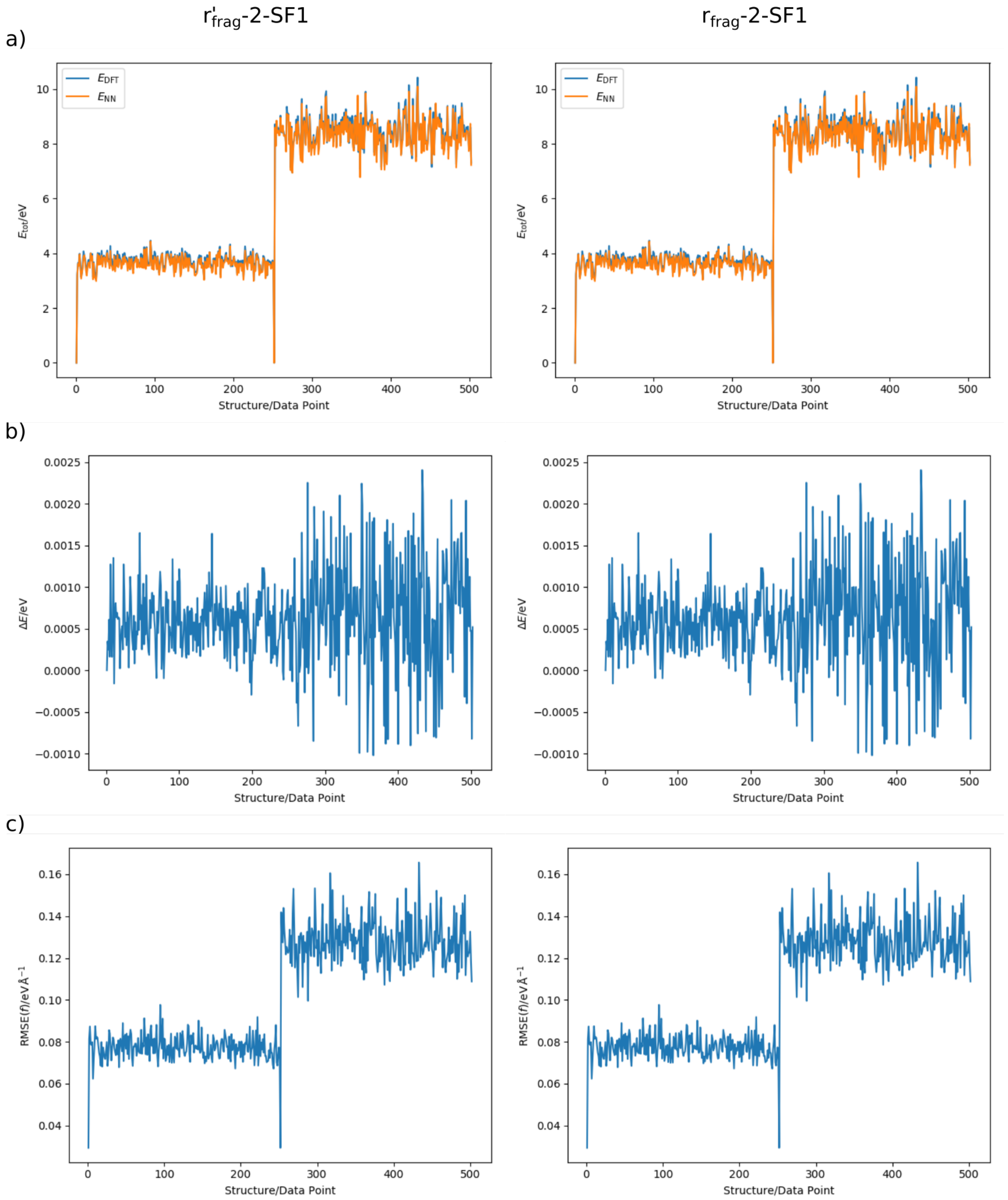

Figure A.47: Compilation of the I10-B' fragment (fig. 4.22) predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) based on the HDNNP training independent data sets for IRMOF-10 (in total 502 I10-B' fragment structures/data points). a) Demonstrates the total potential energy of the I10-B' fragment structure over the data set. The individual MD simulations can be separated by the starting points of the simulations (data point 1 and 252). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

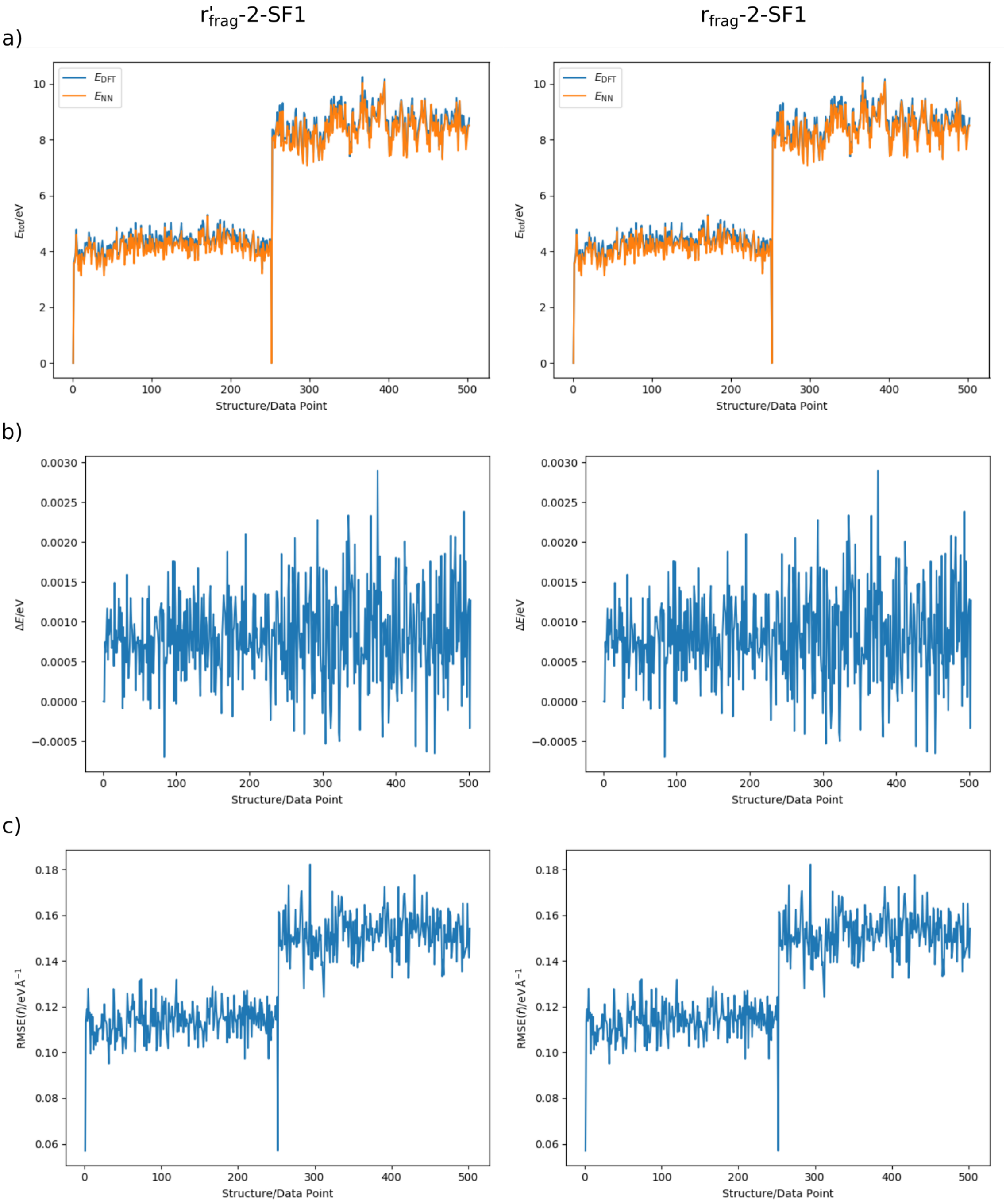

Figure A.48: Compilation of the I16-B’ fragment (fig. 4.22) predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) based on the HDNNP training independent data sets for IRMOF-16 (in total 502 I16-B’ fragment structures/data points). a) Demonstrates the total potential energy of the I16-B’ fragment structure over the data set. The individual MD simulations can be separated by the starting points of the simulations (data point 1 and 252). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

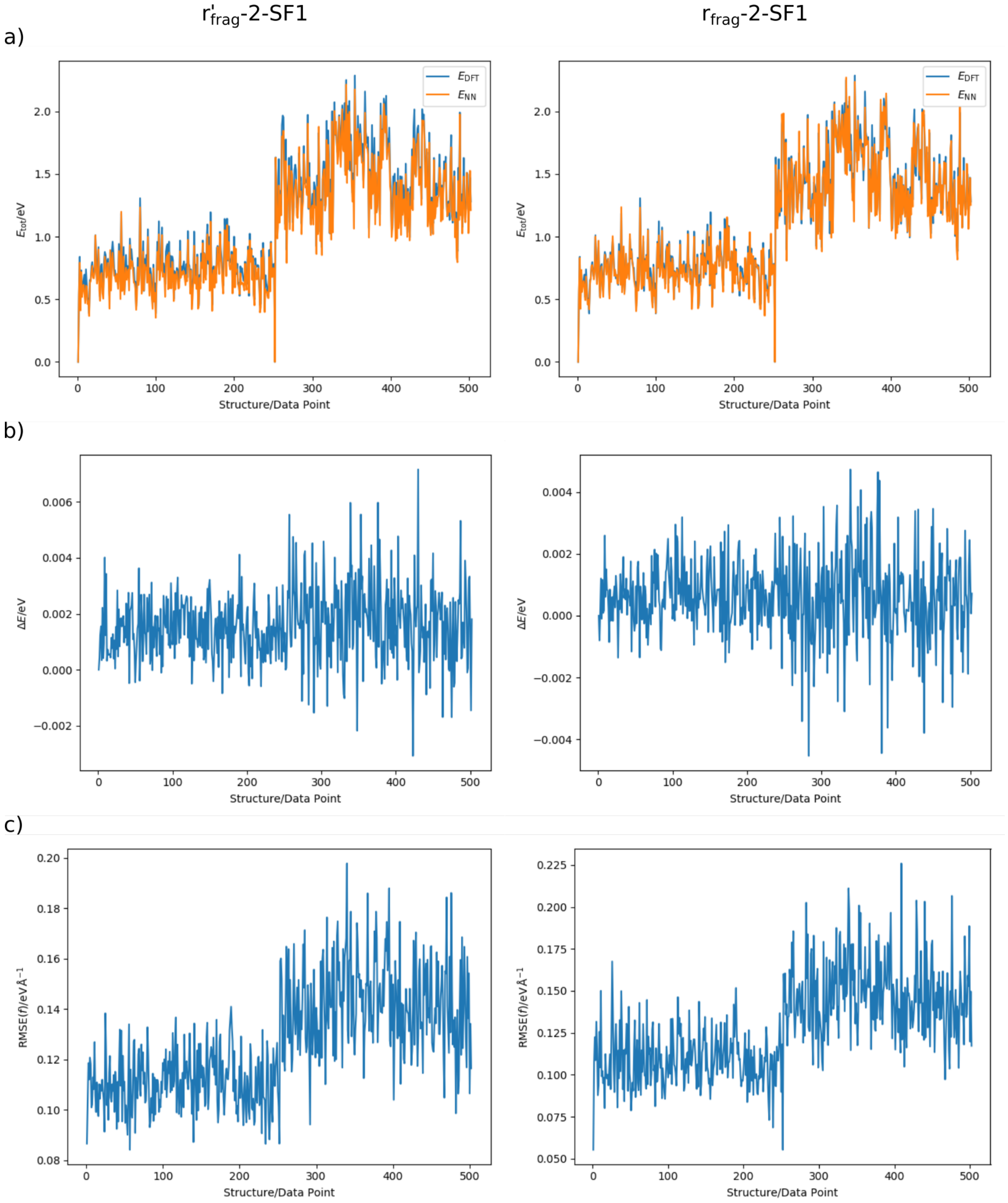

Figure A.49: Compilation of the I16-C' fragment (fig. 4.22) predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) based on the HDNNP training independent data sets for IRMOF-16 (in total 502 I16-C' fragment structures/data points). a) Demonstrates the total potential energy of the I16-C' fragment structure over the data set. The individual MD simulations can be separated by the starting points of the simulations (data point 1 and 252). b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

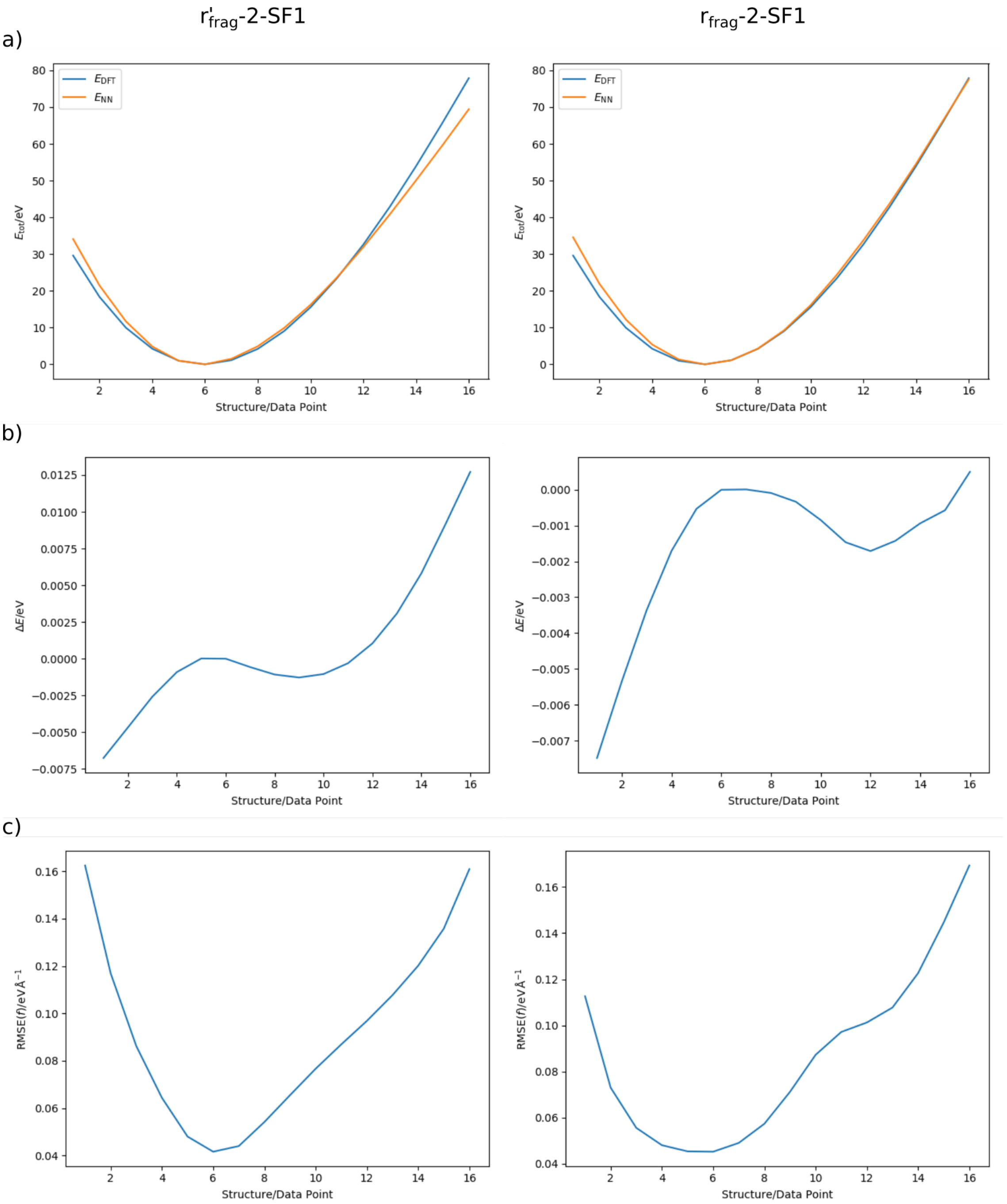

Figure A.50: Compilation of the IRMOF-10 bulk predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) of a HDNNP training independent data set (16 structures/data points), based on expanded and compressed bulk structures by a scaling factor $\sigma \in \{0.95 - 1.10\}$ in steps of 0.01 with DFT optimized atomic positions. a) Demonstrates the total potential energy of the IRMOF-10 bulk structure over the data set, b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

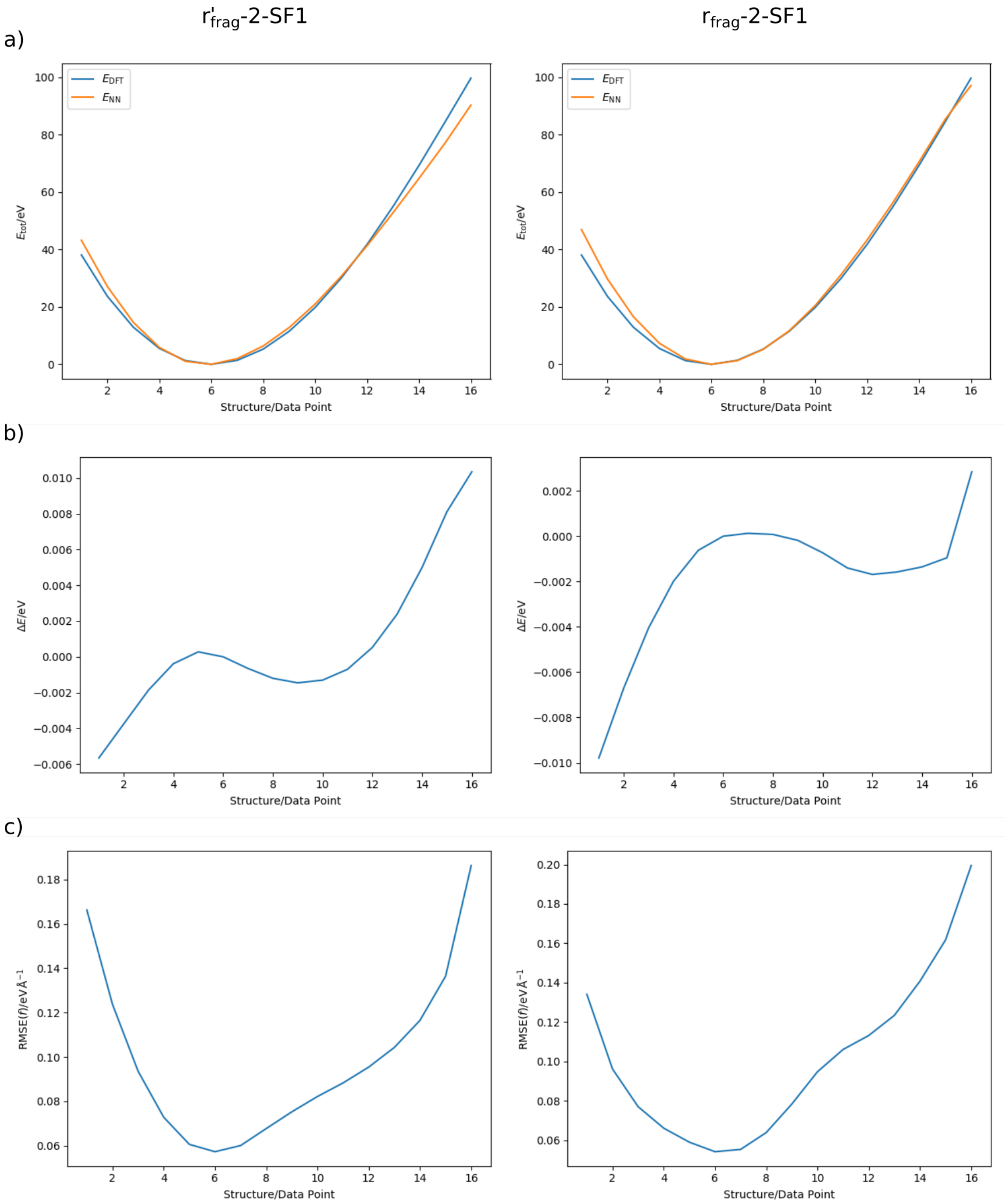

Figure A.51: Compilation of the IRMOF-16 bulk predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) of a HDNNP training independent data set (16 structures/data points), based on expanded and compressed bulk structures by a scaling factor $\sigma \in \{0.95 - 1.10\}$ in steps of 0.01 with DFT optimized atomic positions. a) Demonstrates the total potential energy of the IRMOF-16 bulk structure over the data set, b) The atomic energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

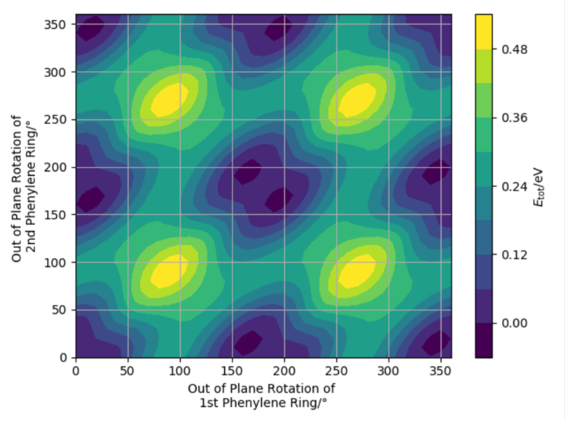

Figure A.52: The DFT reference total potential energy for the I10-dumbbell structure (fig. 4.30) of a HDNNP training independent data set (1369 structures/data points), based on the rotation of the phenylene rings in steps of 10 °.

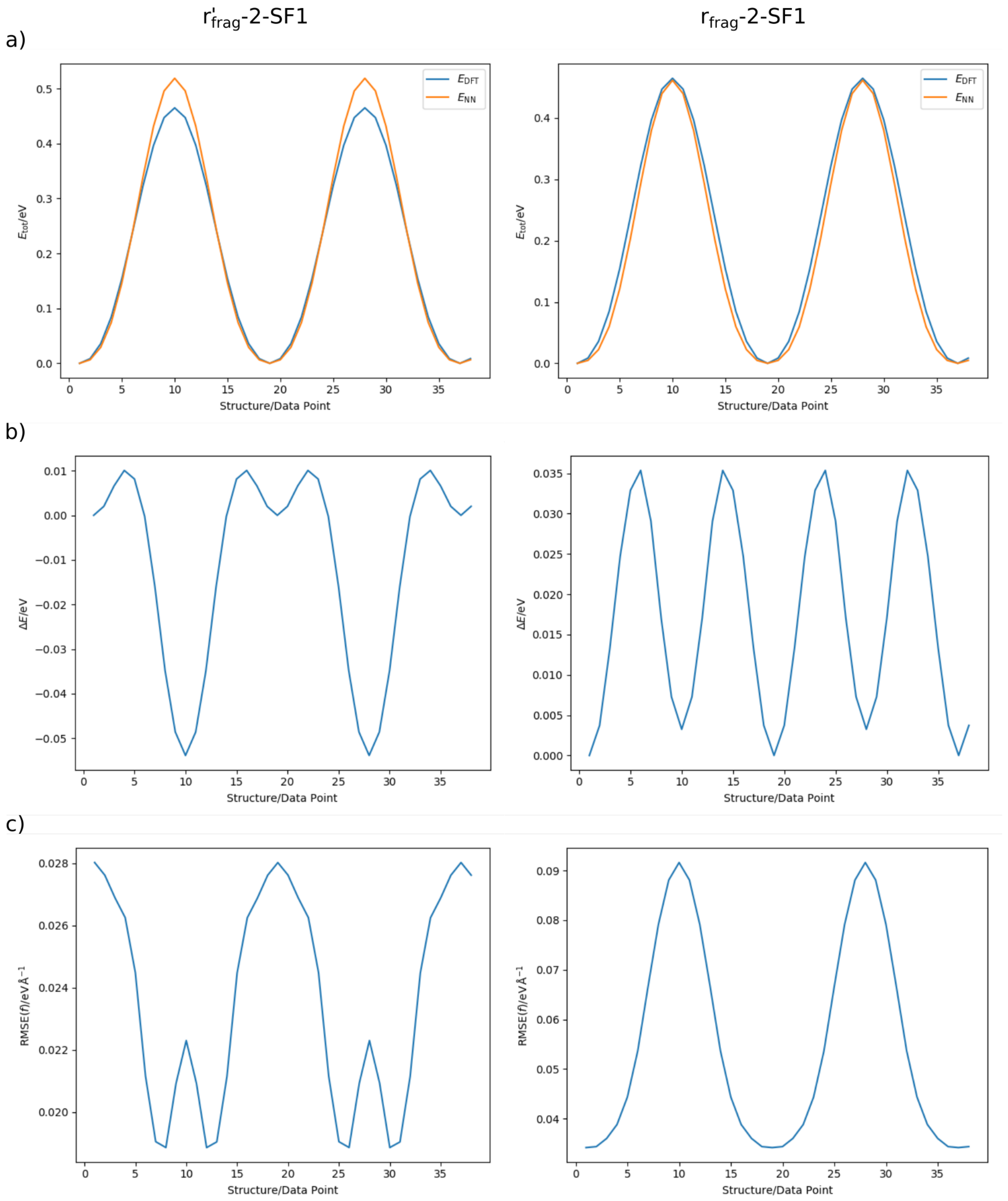

Figure A.53: Compilation of the I1-dumbbell predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) of a HDNNP training independent data set (37 structures/data points), based on the rotation of the phenylene ring in steps of 10 °. a) Demonstrates the total potential energy of the IRMOF-1 bulk structure over the data set, b) The total energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

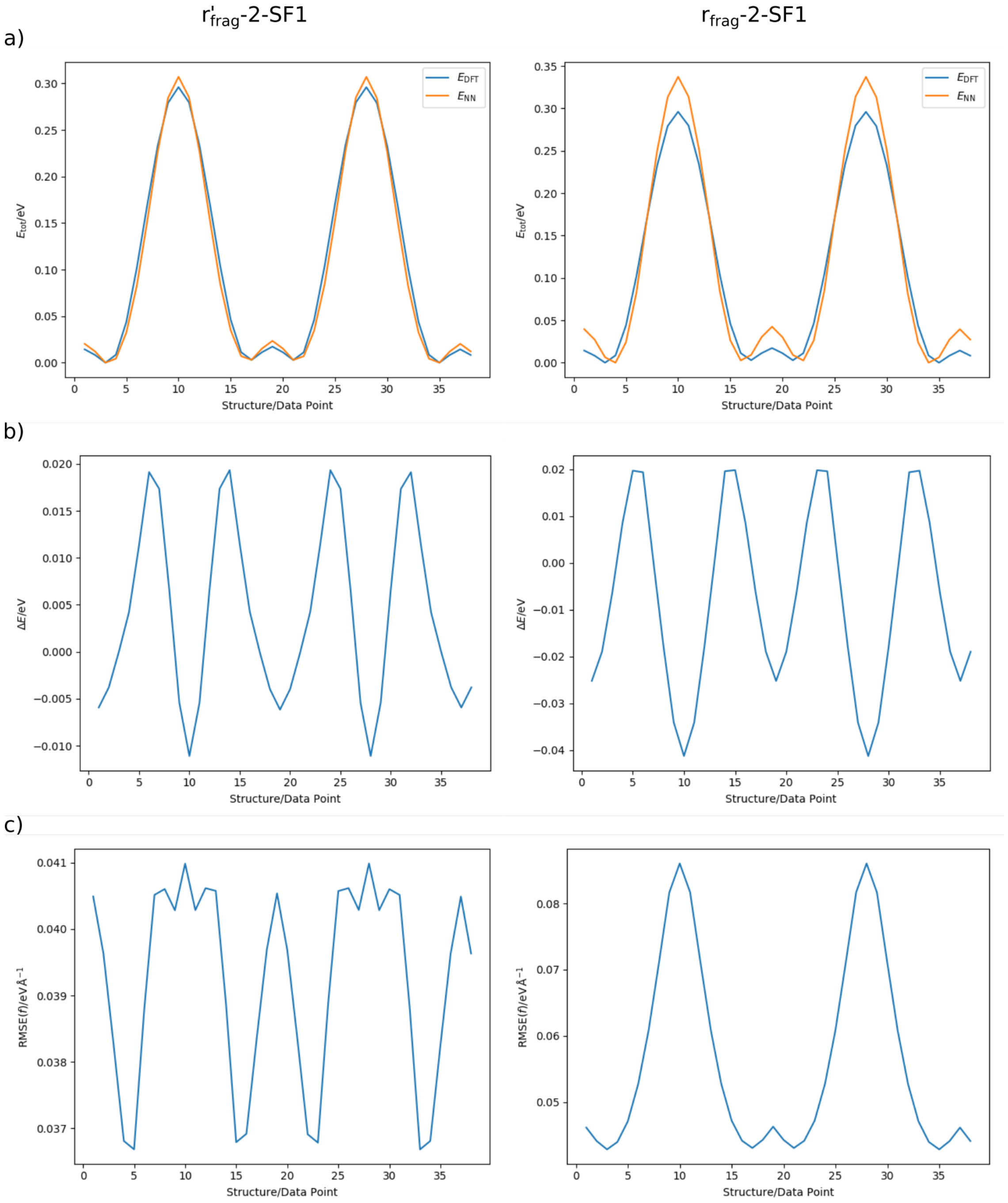

Figure A.54: Compilation of the I16-dumbbell predictions for the HDNNPs r'_{frag}-2-SF1 (tab. 4.12) and r_{frag}-2-SF1 (tab 4.13) of a HDNNP training independent data set (37 structures/data points), based on the rotation of the phenylene ring in steps of 10 °. a) Demonstrates the total potential energy of the IRMOF-16 bulk structure over the data set, b) The total energy error ΔE in eV for the data set and c) the root-mean squared error of the force components RMSE(f) for each data point in eV Å^{-1}.

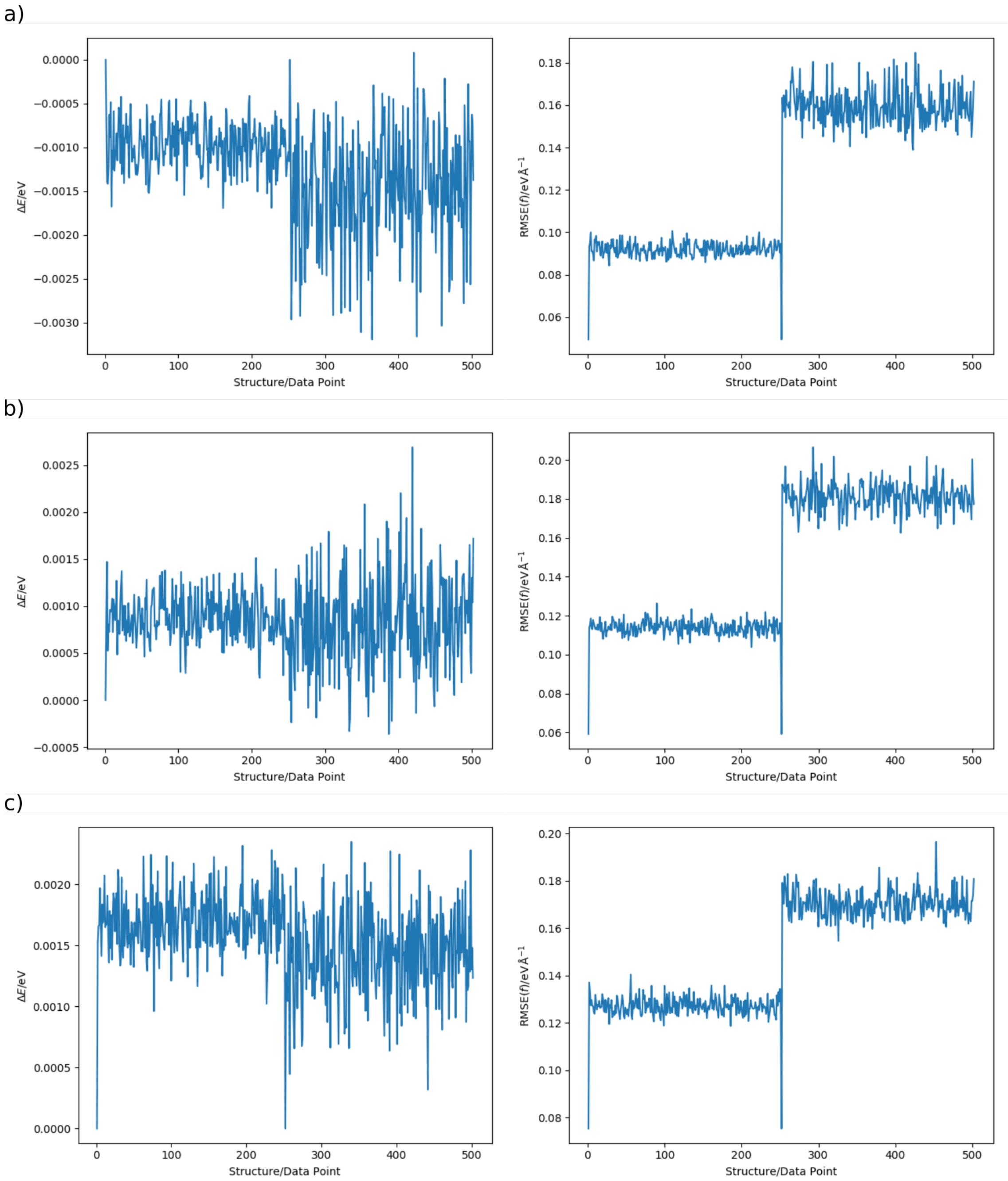

Figure A.55: Compilation of the atomic energy ΔE in eV and the force error RMSE(f) in eV Å^{-1} per data point of the HDNNP r_{frag}-2-SF2 (tab. 4.13) predictions for a) a HDNNP training independent data set (502 structures/data points) for IRMOF-1, combined of 251 data points for each of two MD simulations in the NPT ensemble (sec. 3.4) at normal pressure and temperatures $T \in \{200, 450\}$ in K, b) for IRMOF-10 at temperatures $T \in \{200, 450\}$ in K and c) for IRMOF-16 at temperatures $T \in \{200, 350\}$ in K.

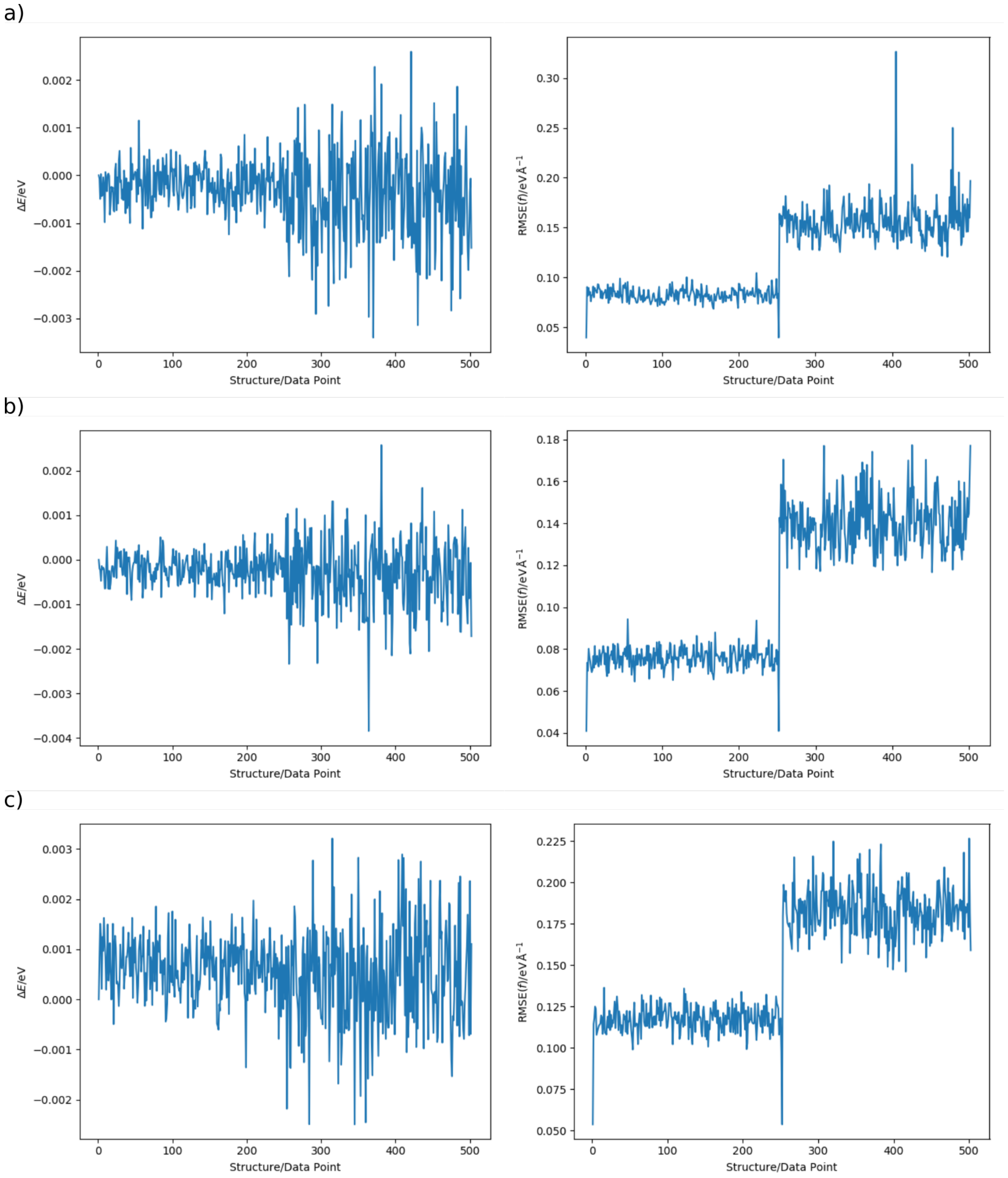

Figure A.56: Compilation of the atomic energy ΔE in eV and the force error RMSE(f) in eV Å^{-1} per data point of the HDNNP r_{frag}-2-SF2 (tab. 4.13) predictions for a) I1-A$'$ based on the HDNNP training independent data sets for IRMOF-1 (in total 502 I1-A' fragment structures/data points), b) I1-B$'$ based on the HDNNP training independent data sets for IRMOF-1 (in total 502 I1-B' fragment structures/data points), c) I10-A$'$ based on the HDNNP training independent data sets for IRMOF-10 (in total 502 I10-A' fragment structures/data points).

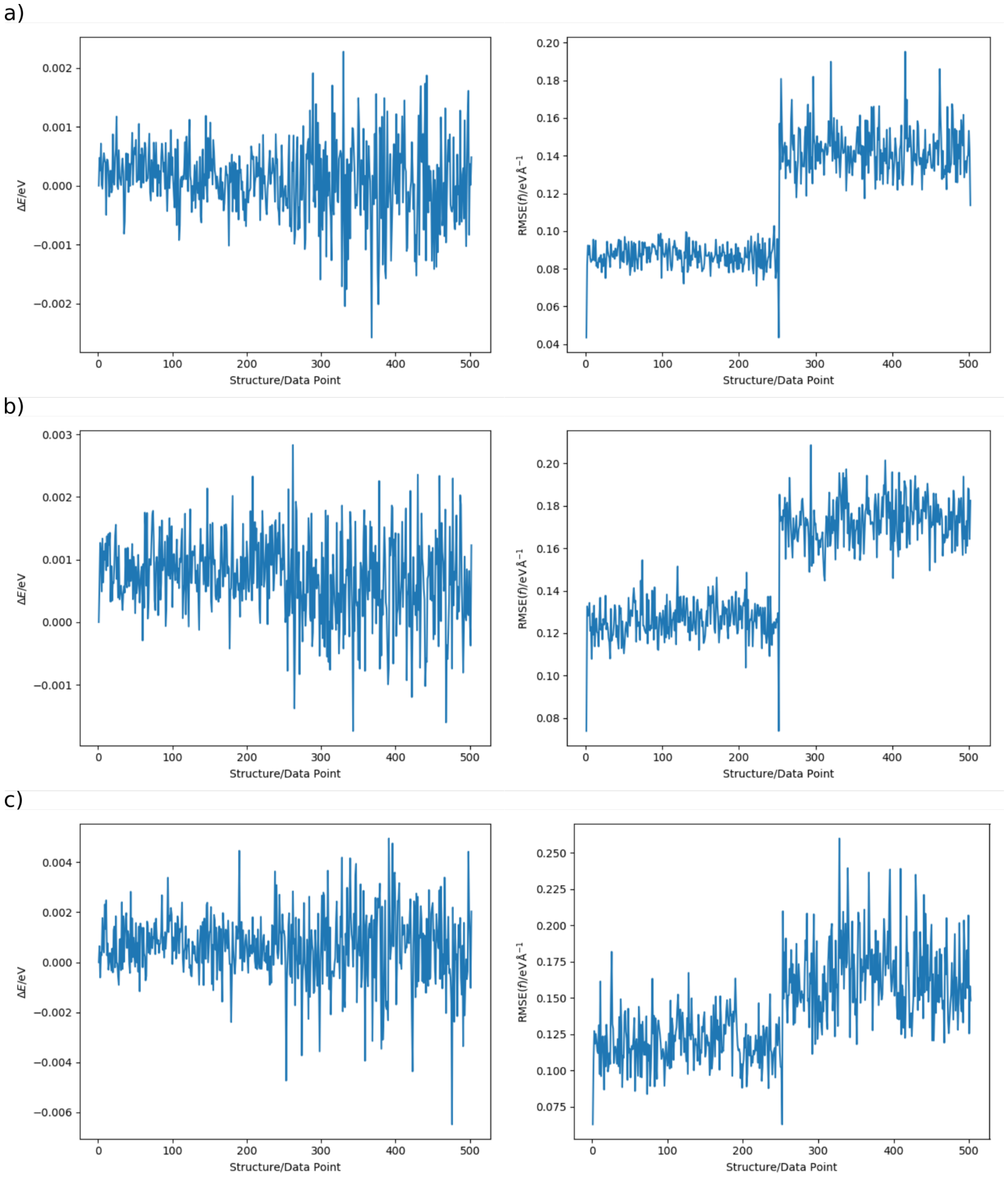

Figure A.57: Compilation of the atomic energy ΔE in eV and the force error RMSE(f) in eV Å^{-1} per data point of the HDNNP r_{frag}-2-SF2 (tab. 4.13 predictions for a) I10-B′ based on the HDNNP training independent data sets for IRMOF-10 (in total 502 I10-B' fragment structures/data points), b) I16-B′ based on the HDNNP training independent data sets for IRMOF-16 (in total 502 I16-B' fragment structures/data points), c) I16-C′ based on the HDNNP training independent data sets for IRMOF-16 (in total 502 I16-C' fragment structures/data points).

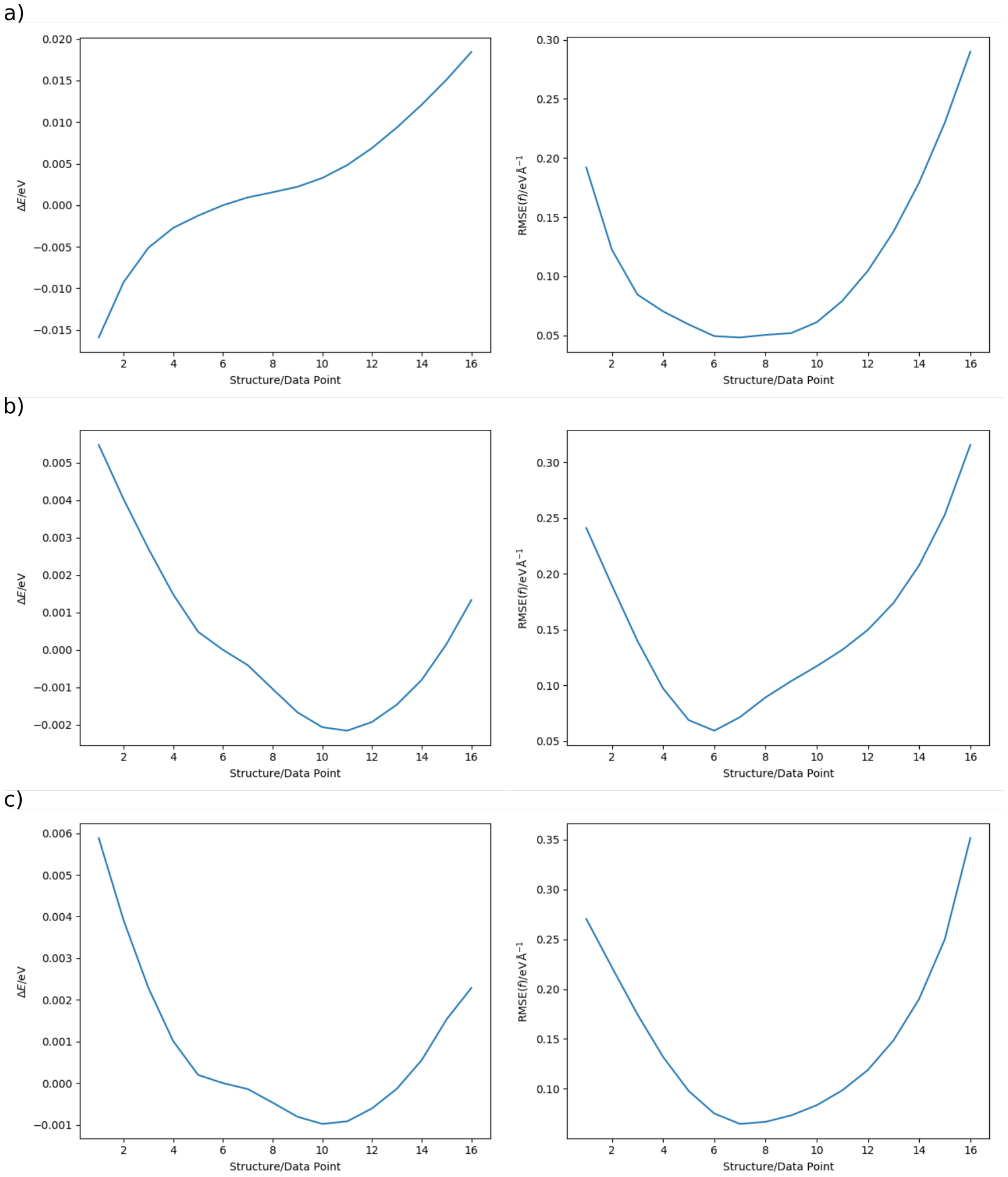

Figure A.58: Compilation of the atomic energy ΔE in eV and the force error RMSE(f) in eV Å^{-1} per data point of the HDNNP r_{frag}-2-SF2 (tab. 4.13) predictions for HDNNP training independent data set (16 structures/data points), based on expanded and compressed bulk structures by a scaling factor $\sigma \in \{0.95-1.10\}$ in steps of 0.01 with DFT optimized atomic positions for a) IRMOF-1, b) IRMOF-10 and c) IRMOF-16.

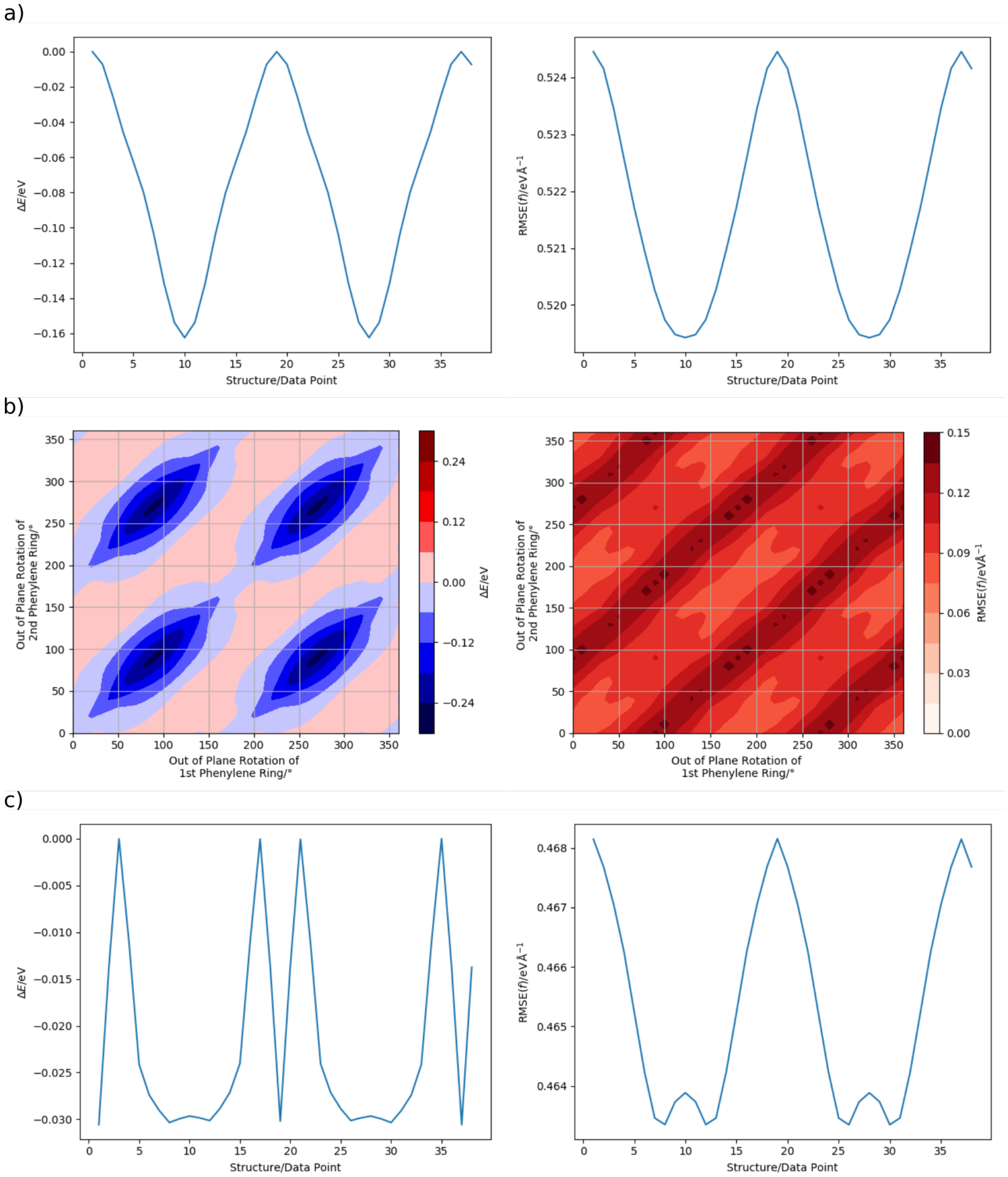

Figure A.59: Compilation of the atomic energy ΔE in eV and the force error RMSE(f) in eV Å^{-1} per data point of the HDNNP r_{frag}-2-SF2 (tab. 4.13) predictions for a HDNNP training independent data set (37 structures/data points and 1369 structures/data points for IRMOF-10, respectively), based on the rotation of the phenylene ring in steps of 10 ° for a) IRMOF-1, b) IRMOF-10 and c) IRMOF-16.