Univerzita Karlova v Praze
Přírodovědecká fakulta

Studijní program: Chemie
Studijní obor: Modelování chemických vlastností nanostruktur a biostruktur

Bc. Michal Trachta

Modelování bioanorganických rozhraní
Modeling of bio-inorganic interfaces

Diplomová práce

Vedoucí závěrečné práce: RNDr. Ota Bludský, CSc.

Praha, 2016

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 3. 5. 2016

Podpis

ACKNOWLEDGEMENT

I would like to thank my thesis supervisor Ota Bludský for his patient leadership and valuable advices. I also want to thank my parents, grandparents and my sister for supporting me in times when I need it.

ABSTRACT

Dynamic atomistic description of bio-inorganic interfaces represents a challenging problem for contemporary computational chemistry. A detailed analysis of processes occurring on the interface between biomolecule and inorganic material can help our understanding of various processes, ranging from chromatography and protein separation to protein immobilization techniques and their effect on enzyme activity or protein conformational stability. High complexity of bio-inorganic interfaces prevents detailed investigation using accurate, but computationally demanding ab initio methods. Since reliable empirical potentials are not available for these systems, the aim of this work is to develop force fields based on ab initio data as well as a general methodology for parameterization of such force fields. Our potential fitting procedure was carried out in an automated fashion based on molecular dynamics simulation. The resulting potentials were applied for investigation of inorganic material's influence on polypeptide conformations.

Detailní atomistický popis bioanorganických rozhraní zahrnující dynamický pohled je jednou z výzev současné výpočetní chemie. Analýza procesů na rozhraní mezi biomolekulami a anorganickým materiálem může pomoci našemu porozumění nejrůznějších procesů, chromatografií a separací proteinů počínaje a imobilizací proteinů a efektem na konformační stabilitu proteinů konče. Složitost bioanorganických rozhraní bohužel brání rozsáhlejšímu nasazení nejpřesnějších metod výpočetní chemie. Vzhledem k tomu, že kvalitní empirické potenciály pro tyto systémy nejsou k dispozici, je naším cílem tyto potenciály vyvinout a s tím i obecnou metodologii pro jejich vývoj. Parametrizace těchto potenciálů byla provedena automatizovaným postupem, přičemž vzorkování struktur bylo založeno na simulacích molekulové dynamiky. Výsledné potenciály byly použity na zkoumání konformačních závislostí polypeptidů interagujících s anorganickými materiály.

CONTENTS

ACRONYMS

ADOR	Assembly-Disassembly-Organization-Reassembly
AIFF	ab initio force field
BLYP	functional of Becke, Lee, Yang and Parr
CBS	complete basis set
CC	coupled-cluster
CI	configuration interaction
D2	D2 dispersion correction of Grimme
D3	D3 dispersion correction of Grimme
DC-DFT	dispersion-corrected DFT
DFT	density functional theory
DFT/CC	dispersion corrected DFT method with the CC-based correction
FF	force field
GGA	generalized gradient approximation (in DFT)
genFF	generic FF (combination of force fields without reparameterization)
HF	Hartree-Fock method
LJ	Lennard-Jones potential
LDA	local density approximation (in DFT)
LZF	layered zeolite framework
M06-L	local version of the functional of Truhlar and Zhao
MC	Monte Carlo
MD	molecular dynamics
MM	molecular mechanics
MOF	molecular organic framework
MP	Møller-Plesset perturbation theory
MP2	second-order Møller-Plesset perturbation theory
newFF	new force field (our reparameterized force field)
OMC	ordered mesoporous carbon
PAW	projector augmented wave
PBE	density functional of Perdew, Burke, Ernzerhof
PME	particle-mesh Ewald method for calculation of electrostatics
post-HF	post-Hartee-Fock methods
PW	plane wave basis set
RMSD	root-mean-square deviation
RP-RKHS	reciprocal power reproducing kernel Hilbert space interpolation
SAPT	symmetry-adapted perturbation theory
tempFF	temporary force field used in particular round of fitting procedure
TS	dispersion correction of Tkatchenko and Scheffler
vdW	van der Waals (may refer to dispersion)
vdW-DF2	second version of the van der Waals functional
vdW-DF/CC	version of the DFT/CC approach
XC	exchange-correlation

1 INTRODUCTION

Intermolecular interactions represent one of the main research fields of computational chemistry. Every molecule influences its neighborhood and the neighborhood also affects the molecule, so molecular properties (*e.g.* chemical reactivity, conformational stability, various molecular spectra) depend on both the molecule and its chemical environment in a complex manner. The role of the chemical environment may be easily illustrated on proteins, large biopolymers formed from amino acid units in living organisms. After synthesis of the amino acid chain the protein folds into its native state (native structure).[1] The folding process is strongly affected by the surroundings (solvent may reduce the energy barriers of folding)[2] and also the stability of the native protein structure depends on surroundings (solvent removal can lead to denaturation of globular proteins), as can be illustrated on diverse protein stabilities on the water-vacuum interface.[3,4] The water-vacuum interface is an interesting example how the protein properties change if we take the protein from its cellular environment and place it in some quite different environment. In this work we focus on the interaction of biomolecules (amino acids, peptides, proteins) with such systems, namely with the surfaces of inorganic materials.

The interaction with inorganic surfaces is present even in living cells as can be illustrated on function of antifreeze proteins, *e.g.* the flat surface of antifreeze protein type III interacts strongly with ice crystal, preventing thus further growth of the crystal.[5,6] The interaction of inorganic materials with proteins plays also its role in development of biocompatible materials.[7,8] Another interesting research area is drug delivery. Many organic and inorganic materials are investigated for possible applications as drug carriers.[9] Protein separation techniques making use of properties of inorganic materials, especially pore size and accessible surface area, can be enhanced using functionalization or modifications of the surfaces.[10] For industrial applications the proteins can be immobilized on/inside inorganic materials. The immobilization of protein reduces the leaching rate and affects its rigidity, both the conformational stability and enzyme activity may be modified in positive or negative way.[11] The protein immobilization has also applications in the growing area of biosensors.[12]

All the above mentioned applications may benefit from a more detailed knowledge of the processes occurring on the bio-inorganic interface. Unfortunately, experimental techniques are not able to describe the interfaces with atomistic accuracy or at least they have serious problems in doing so. As a result, we can obtain either some averaged data from experiment, or we can get atomistic description (*e.g.* atomic force microscopy), but without any information about time evolution of the system.[13] In contrast, computational chemistry can easily describe these systems at the atomic scale level, but it suffers from other difficulties. First, it is generally hard to model a material which structure is not well-defined. The materials often contain various defects and its chemical composition may vary. The interaction of defects with water or oxygen molecules, for instance, profoundly affects the surface properties which are quite different from the bulk material. We restricted our work to some well-defined or idealized materials to avoid this problem. Second, the bio-inorganic interface is a complex heterogeneous system which can only be modeled using periodic models with truly large unit cells. Unfortunately, the use of the large cells in periodic calculations makes almost impossible to perform dynamical studies using the state-of-the-art methods of computational chemistry. To reduce the computational cost, less reliable methods such as force fields have to be used. Since the reliability of empirical potentials is often questionable, we have parameterized the interaction between the material and the biomolecule on very accurate ab initio data (*i.e.* ab initio force field – AIFF) to get the performance of our approach as close to high-level ab initio molecular dynamics as possible.

2 METHODS AND MODELS

2.1 Theoretical background

2.1.1 Brief overview of computational chemistry methods

The goal of this overview is to briefly present main areas of computational chemistry and thus prepare theoretical ground for the following chapters. Some important aspects relevant to this work will be highlighted. This overview is not intended to be exhaustive, more detailed information can be found in Cramer,[14] Szabo and Ostlund,[15] Parr and Yang[16] or Piela.[17]

The approximately 100 years long history of quantum chemistry resulted in a better understanding of molecular structure and properties, and consequently in a creation and expansion of many different methods of computational chemistry. In a slightly simplified view, these methods may be divided into four main classes:

1. Molecular mechanics
2. Semi-empirical methods
3. Wavefunction-based methods
4. Density functional theory methods

These approaches differ in their accuracy, applicability and physical background. The first group is molecular mechanics (MM). It represents an extremely simplified method of molecule/system description. Using a reference data from higher levels of theory (2-4) or experimental data, it tries to mathematically express the potential energy, resulting in empirical potentials (force fields; FFs). Molecular mechanics usually adopts functional forms from some basic physical or quantum chemistry models (*e.g.* quantum linear harmonic oscillator), but it can also be understood as a purely mathematical concept. The biggest advantage of molecular mechanics is its low computational cost. Thanks to it the theory of molecular mechanics was extended to (classical) molecular dynamics (MD) – a method capable of description of dynamical evolution of the studied system. Among the other advantages of MM is the simple separability of energy contributions thanks to the pairwise form of the potential energy function (energy contributions due to particular pairs of atoms), easy introduction of superpositions (atom may be half silicon and half germanium) and even the possibility to unite groups of

atoms and treat them together (united-atom approach, coarse-grained force fields[18]). Although some of these features may be applied even for the other levels of theory, for a combination of MM and MD they are much more intuitive and easier to implement.

Semi-empirical methods were developed as an attempt to reduce the computational cost of wavefunction-based methods (will be mentioned later), namely the Hartree-Fock method. They use either some approximations instead of explicit calculation of complicated integrals or these integrals are fully neglected. Some of these methods target on even higher accuracy than the Hartree-Fock theory (that means they try to cover the electron correlation), for instance the performance of PM6 is quite decent for various applications.[19]

Wavefunction-based methods are based on Schrödinger's equation and its solution – the wavefunction. The fundamental theory of wavefunction-based methods is the Hartree-Fock method (HF), which searches for the wavefunction in a form of Slater determinant. The equation of energy for electron (ε_i) in the HF orbital is

$$\left(-\frac{1}{2}\Delta + v + j - k\right)\psi_i = \varepsilon_i\psi_i \tag{1}$$

where the first contribution corresponds to the kinetic energy, second to the potential energy in the field of nuclei, third to the Coulomb electron-electron repulsion (uncorrelated electron motion) and fourth to the exchange energy, which accounts for the correlation of motion of electrons with the same spin (Fermi correlation arising from the Pauli antisymmetry principle). Note that the correlation of motion between electrons with opposite spins is not covered in the HF theory. The error in energy due to this deficiency is called a correlation energy and is defined as follows:

$$E_C = E_{exact} - E_{HF\ limit} \tag{2}$$

The E_C stands for the correlation energy (in the complete basis set limit), which is a difference between the exact nonrelativistic energy and the Hartree-Fock limit (the HF energy with infinite basis set). The correlation energy is always negative (or zero), thus stabilizing the system.

Although the performance of HF theory is not very good, the HF theory is the first step to more accurate methods (post-HF methods), such as the Møller-Plesset perturbation theory

(MP), symmetry-adapted perturbation theory (SAPT), coupled-cluster (CC) and configuration interaction (CI). The possibility to systematically improve the description of the system and thus approach the exact energy is the biggest advantage of wavefunction-based methods. On the other hand, the computational cost necessary for such accuracy increases extremely steeply with system size and thus the post-HF methods remain limited for (more or less) small systems.

The density functional theory (DFT) represents a computationally cheaper alternative to wavefunction-based methods. Together the DFT and the wavefunction-based methods are commonly denoted as ab initio methods (in some cases this term is used only for the latter), because these two approaches are based on strictly physical fundamentals (calculation from first principles) and they may (in principle) lead to the exact energy without any approximations and any empirical parameters. Instead of the wavefunction, the DFT uses the number of electrons per unit volume – the electron density. The simplest implementation of the DFT is the local density approximation (LDA), which was derived from the uniform electron gas and where the ground state energy is expressed as a local functional of the electron density. As an extension, the generalized gradient approximation (GGA) was developed, which uses not only the electron density, but also its gradient and thus it better describes regions with inhomogeneous electron density (semi-local density functional). The prevailing way of the DFT implementation is through the Kohn-Sham equations, which introduce the orbitals and the wavefunction present already in the HF theory. The energy of electron in the Kohn-Sham orbital is described by equation

$$\left(-\frac{1}{2}\Delta + v + v_{XC} + j\right)\psi_i = \varepsilon_i\psi_i. \tag{3}$$

From the comparison with Equation 1, we see that most of the terms are the same, but the exchange term is not present here, instead the exchange-correlation potential, v_{XC}, was introduced. This added XC potential, together with the coulombic repulsion j, mimics the electron-electron interaction and corrects the error in the kinetic energy introduced by this approach. Thanks to the use of the exchange-correlation potential, the DFT is less computationally demanding and also much more accurate than the HF theory, because the correlation of electronic motion is well-described (in principle exact) in the DFT. In reality, however, there is always an error introduced in the exchange energy (self-interaction is not

properly subtracted) and the correlation energy is accounted for in the local sense only – the short-range correlation effect is covered well, but the error increases with the distance. This missing contribution from the non-local electron correlation (dispersion) was the reason for development of dispersion-corrected density functional theory methods (DC-DFT) – these methods will be described in Section 2.1.3.

2.1.2 Cluster calculations

The development of force fields requires a reliable reference theory, which should be reproduced by the force field as well as possible. Such reference theory has to be chosen very carefully – it must be computationally affordable even for large periodic systems and allow sufficient number (typically hundreds) of potential energy evaluations. At the same time, the accuracy must be good enough for our purpose, since the force field error is given as a sum of reference theory error and error of the parameter fitting.

Since the most accurate post-HF methods are not computationally affordable for extended systems, they cannot be used for the force field parameterization. However, they may be used as a tool to assess the quality of various methods and thus help us in choosing the reference level of theory. For this assessment the extremely accurate CCSD(T)/CBS data on small cluster models were compared to various dispersion-corrected DFT methods (DC-DFT) – among them the reference theory will be chosen, because DC-DFT methods represent probably the best compromise between accuracy and computational cost.

The CCSD(T)/CBS data were calculated using the formula*

$$CCSD(T)/CBS = HF/AVQZ + \Delta MP2/CBS + \Delta CCSD(T)/AVDZ \,. \quad (4)$$

The CCSD(T)/CBS data obtained by this procedure are often denoted as CCSD(T)/CBS//MP2 to point out the way how they are obtained, but in this work the shorter CCSD(T)/CBS label will be used. The HF/aug-cc-pVQZ energy (HF/AVQZ) is a quite accurate estimate of the HF limit energy ($E_{HF\ limit}$). The ΔMP2/CBS is the correlation energy on the MP2 level of theory

* Please note that all energy values mentioned in this work are interaction/binding energy values (that means energy differences) without deformation contribution.

extrapolated to the complete basis set from the AVTZ and AVQZ data using the following equation:[20]

$$E_C^{AVXZ} = E_C^{CBS} + AX^{-3}\,, \tag{5}$$

where X represents the cardinal number of the basis set. The remaining term is the correction from MP2 to CCSD(T) which is evaluated using the AVDZ basis set; this difference in many cases converges quite quickly:[21]

$$\Delta CCSD(T)/AVDZ = CCSD(T)/AVDZ - MP2/AVDZ \tag{6}$$

2.1.3 Dispersion-corrected DFT

As was already mentioned, the DFT methods usually cover only the local part of electron correlation. It is quite a challenge to include the non-local electron correlation effects to the DFT and the methods trying to achieve this goal are called the dispersion-corrected DFT methods (DC-DFT). The DC-DFT approaches differ by the level of empiricism, ranging from empirical corrections to physically-sound non-local density functionals. To introduce at least some of them:

- vdW-DF2 calculates dispersion using a non-local term accounting for the response of electron densities in two points of space.[22]
- Functionals of Zhao and Truhlar (M05, M06) are parameterized to reproduce the energy including the dispersion, thus overestimating the effect of electron densities overlap on the interaction energy. Such approach is unphysical and often leads to serious problems in equilibrium distance prediction.[23,19,24]
- Grimme's D2 correction[25] represents a pairwise correction scheme, where the dispersion correction is added to the energy of a particular functional, for which the scaling constant must be determined.
- Grimme's D3 correction represents a more complex and less empirical approach than D2, some problems of D2 method are solved – *e.g.* D3 has the right asymptotic behavior and it introduces different atomic types (thanks to estimate of coordination number) to increase the accuracy.[26]
- The correction of Tkatchenko and Scheffler is in some aspects similar to Grimme's D2 correction, but it uses also the electron density.[27]

- DFT/CC[28] is quite different from the previously mentioned approaches. This correction scheme does not have ambitions to have general applicability (in the sense of transferability). Instead, it defines system-specific pairwise correction functions, which, when added to the GGA interaction energy, should recover the coupled-cluster accuracy. These corrections are parameterized on particular reference systems and can be transferred only over a small range of similar systems. The main advantages of this approach are high accuracy and the possibility to correct all errors in the DFT functional (DFT/CC does not correct only the dispersion). The further development of this scheme resulted in the vdW-DF/CC method, which has a much simpler parameterization procedure.[29]

For this work, the Grimme's D2 correction scheme[25] and the DFT/CC[28] method are particularly important. The D2 correction is defined as follows:

$$E_{D2} = -s_6 \sum_{i<j} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}) \tag{7}$$

In this equation s_6 represents a scaling constant, which depends on a given DFT functional (0.75 for the PBE functional),[25] but the use of values different from one is responsible for wrong asymptotical behavior of this method, as it was already pointed out. The sum runs over all pairs of atoms within a certain cut-off distance. The C_6 parameter defines the strength of the dispersion interaction and may be different for every pair of atomic types. The R_{ij} term corresponds to the interatomic distance. Finally, the damping function f_{dmp} switches off the D2 dispersion correction for short distances, where the local electron correlation dominates and the underlying GGA functional works reasonably well. For longer distances the damping function gradually changes its value from 0 to 1, which corresponds to switching on the correction to compensate the DFT deficiencies in the description of the non-local electron correlation.

The DFT/CC method[28] is based on the pairwise representability of the DFT error, ΔE, defined as the difference between the CCSD(T) and DFT interaction energies. The assumption of pairwise representability leads to the following equation

$$\Delta E = \sum_{i<j} \varepsilon_{ij}(R_{ij}) \quad (8)$$

where ε_{ij} are the DFT/CC correction functions and R_{ij} are the interatomic distances. The pairwise correction functions are obtained by means of the reciprocal power reproducing kernel Hilbert space interpolation (RP-RKHS)[30]; no a priori functional form of correction functions is assumed except for the asymptotic behavior given by the RKHS kernel (R^{-6} + R^{-8}).

2.1.4 Force fields

The force fields allow us to model large systems typically not accessible by ab initio methods and to perform much longer dynamic simulations compared to ab initio molecular dynamics. The computational efficiency of force fields has been attracting the attention of computational chemists for a long time. For illustration the Amber ff99SB-ILDN force field[31] employed in this work is in fact a continuation of the Amber ff94 force field[32]. The more than 10-year long history resulted in many changes and reparameterizations of the original force field (Amber ff99, Amber ff99SB[33]). The Amber-family force fields are among the most popular for modeling of proteins.

For the description of zeolitic materials we used the force field of Bushuev and Sastre,[34,35] which is based on the popular ClayFF.[36] The force field of Bushuev and Sastre (as well as many other force fields used for zeolites, clays, MOFs, *etc.*) is very different from the Amber family of force fields. The bonds are modeled in a non-bonded fashion (except the silanol O-H bond, which use the harmonic potential). The electrostatic attraction between silicon and oxygen atoms is balanced by the Lennard-Jones repulsion, thus leading to a bond-type potential. This approach may lead to serious issues, since the bond dissociation is not prevented by the infinite energy barrier as in the case of the harmonic potential. Badly parameterized potential or higher temperature can thus lead to "tearing" atoms out of the material.[36,37]

The combination of force fields is usually not recommended, because it may lead to various artifacts. In our case the choice of the force field of Bushuev and Sastre was not completely random – besides a good performance for zeolites (and particularly silanol

groups[38]), this force field has more reasonable charges on silicon and oxygen (2.10 and -1.05, respectively) than for instance the force field of Sanders, Leslie and Catlow[39,†] (4.0 for silicon, 0.86902 for oxygen core and -2.86902 for oxygen shell). Thus, it is more likely that it will work reasonably with the Amber force field. Moreover, the Amber force field use the TIP3P water, which has a quite similar parameterization as the flexible SPC water model recommended for the zeolite force field.[35] The interaction between organic molecule and the material can be reparameterized on accurate reference data, and thus it should not be the source of any problems.

Since we are dealing with interactions between biomolecules and inorganic material (usually physisorption), we are most concerned about non-bonded interactions. Throughout this work we do not attempt to reparameterize bonded energy contributions (bonds, angles, dihedrals). Therefore, the most relevant energy contributions are electrostatic and dispersion terms, usually expressed as the Coulomb potential and the 12-6 Lennard-Jones potential. The Coulomb potential is defined as

$$E_{Coulomb} = \frac{1}{4\pi\varepsilon_0}\sum_{i<j}\frac{q_i q_j}{R_{ij}}, \tag{9}$$

where ε_0 is vacuum permittivity, q_i and q_j are the charges of the interacting particles and R_{ij} is the distance separating them. Electrostatics represents a more problematic term than dispersion, since its decay with the distance is very slow and its amplitude is large, leading to effects such as polarization. Cut-offs commonly used for dispersion interactions result in large errors in the case of electrostatic interactions and the advanced summation techniques (*e.g.* PME[40]) are thus necessary. To best of our knowledge, all commonly used software packages have the charge as an universal parameter for each particular atom. This represents a serious problem in combining force fields with different approaches of the charge determination – the force fields do not have modular design (or their implementation does not support modular design). As a consequence, we do not attempt any fitting of the electrostatic charges

† This polarizable force field was used in some of our previous works and lead to good accuracy, but its unreasonable charges and its core-shell nature prevent usage in combination with other force fields (such as the force fields of Amber family), because it would lead to wrong electrostatics, instabilities and complicated control over the accuracy of the resulting combination of potentials. Moreover, this potential doesn't contain parameters for the silanol group.

(*e.g.* restrained electrostatic potential – RESP[41]), since it is not possible to implement them. When we want to combine force fields, we must rely on some level of "compatibility" in their implementation of electrostatics. Fortunately, the problem may be partially corrected by appropriate parameterization of dispersion, which can compensate for the error in the short-range part of electrostatics. Moreover, any pairwise potential may be defined using the tabulated potentials in GROMACS,[42] so the short-range electrostatics correction of the physical 1/R shape may be used.

As was already mentioned, the dispersion is usually expressed using the 12-6 Lennard-Jones potential:

$$E_{LJ} = \sum_{i<j} 4\,\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}} \right)^{6} \right] \tag{10}$$

The ε_{ij} determines the depth of the potential for a particular pair of atoms i and j, whereas the σ value gives the potential its "shape". The R_{ij} is the interatomic distance. The square brackets contain two terms – the repulsive (positive) exchange and the attractive (negative) dispersion term.

2.2 Materials

In this work two kinds of materials are considered, namely the carbon and zeolite materials. The carbon materials include diamond (sp^3), graphite (sp^2), and porous materials with less defined structures (amorphous carbon). The diamond is not in the focus of this work, despite its extraordinary physical properties. For adsorption of biomolecules more interesting materials are graphitic nanostructures such as fullerenes and carbon nanotubes, and porous carbon materials.[43] The graphite is a black conducting material, which is composed of weakly bound graphene layers. These layers consisting of hexagonally arranged sp^2 carbon atoms (D_{6h} symmetry) are stacked in ABAB manner in hexagonal graphite (*e.g.* highly oriented pyrolytic graphite), whereas the stacking in rhombohedral graphite is ABCA. The fullerene structure resembles the graphene, the carbon nanotubes are in fact rolled up graphene layers and share some properties with graphite (*e.g.* electrical conductivity utilized in their applications in biosensors[12]), but they have much larger surface area. The size (diameter) of fullerenes and nanotubes may vary depending on synthesis conditions. Even higher variability is observed

among amorphous carbon materials. Amorphous carbon is usually prepared by carbonization of organic (carbon rich) matter, such as saccharides or polymers[43], or by chemical vapor deposition on/inside materials[44]. The porosity may be tuned from micropores over mesopores to macropores, but for larger pores (meso-, macro-) the presence of a template material is necessary. Commonly used templates are siliceous materials, where the amorphous carbon fills the voids in the siliceous material which is then removed (dissolved), leading to amorphous carbon material.[43] The channel structure in amorphous carbon may or may not be regular. The possible applications as molecular sieves, adsorbents and electrodes have been investigated.[44] Since metal atoms may be easily introduced into carbon support, catalytic properties of these materials may be of interest.[44] Appealing properties of carbon materials are their thermal and chemical stabilities.

In contrast to carbon materials, zeolites and related materials are chemically heterogeneous systems. A general building block in these materials is the TO_4 tetrahedron, where T stands for Si, Al, P, Ti, B, Ge, *etc.*[45,46] This variability in composition results in a different chemical behavior (*e.g.* presence of Al increases acidity of the material, exchange of these "acidic" hydrogens by ions alters the properties as well). More than 230 different zeolite frameworks were prepared to this day[47] which represents quite a large number of various channel architectures. Zeolites are frequently used as heterogeneous catalysts, molecular sieves and adsorbents.[48,49] Compared to other molecular sieves (*e.g.* MOFs) zeolites are thermally and chemically stable. Unfortunately, there are some aspects limiting the use of these materials as environmentally-friendly catalysts. Since the zeolite synthesis is far from being fully understood, our ability to control the properties (composition, structure) of synthesized material is very limited.[46] Another problem is that the diffusion limits hinder the conversion rate.[49] To deal with this problem, huge research effort was spent on synthesis of zeolites and zeolite-like materials with larger pores, leading to a significant progress:

- Group of extra-large pore zeolites was synthesized (including the UTL zeolite).[50,51]
- Synthesis of amorphous zeolite-like structures with mesoporous and hierarchical channel structures was performed. The hierarchical structures contain at least two different sizes of interconnected channels, the smaller (micropore) channel is responsible for the reaction and the larger (mesopore) channel allows faster diffusion.

For synthesis of hierarchical siliceous material, a template is needed. Porous carbons, polymers or surfactants are commonly used.[48,49]

- The diffusion limits may also be overcome by the use of layered zeolite frameworks (LZFs),[52,53,38] two-dimensional zeolitic layers with very good accessibility. A possible manipulation with these layers opens up even larger possibilities in material tailoring.[52]

The LZFs may be prepared by three different routes: (i) some of the three-dimensional zeolites are prepared via layered precursors which are formed during the synthesis and subsequently interconnected during the calcination step (*e.g.* FER and CDO, which are formed from different arrangements of the same layer[54]), (ii) LZFs may also be prepared by a restriction of the zeolite growth using surfactants,[55] and (iii) based on decomposition of the 3D zeolite into 2D layers. This approach is one of fundamental steps in the Assembly-Disassembly-Organization-Reassembly (ADOR) process. In principle, many zeolites may be decomposed this way (ITG, ITH, ITR, IWR, IWV, IWW, UOV).[56]

The particular materials considered in this work are graphene/graphite, carbon nanotubes, α-quartz (fully reconstructed α-quartz (001) surface model of Goumans and Catlow[57]), UTL lamella (IPC-1P; LZF derived from 3D UTL zeolite)[52] and 3D zeolites of the UTL family (UTL, OKO, PCR frameworks). All zeolitic materials are modeled as having pure siliceous composition. Such assumption is quite reasonable as germanosilicate UTL zeolite is susceptible to hydrolysis and thus unsuitable for industrial applications, but its composition can be adjusted by post-synthesis treatment to a stable pure silica zeolite.[58] The UTL lamella was prepared by removal of Ge from germanosilicate UTL zeolite (see Figure 1), leading to pure siliceous IPC-1P.[52]

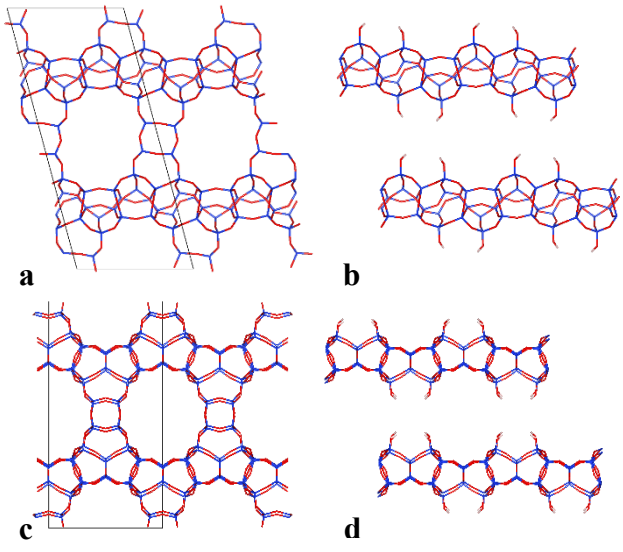


Figure 1. Disassembly of 3D UTL into IPC-1P lamellae. (a) UTL from (010) direction and (b) disassembled. (c) UTL from (001) direction and (d) disassembled.

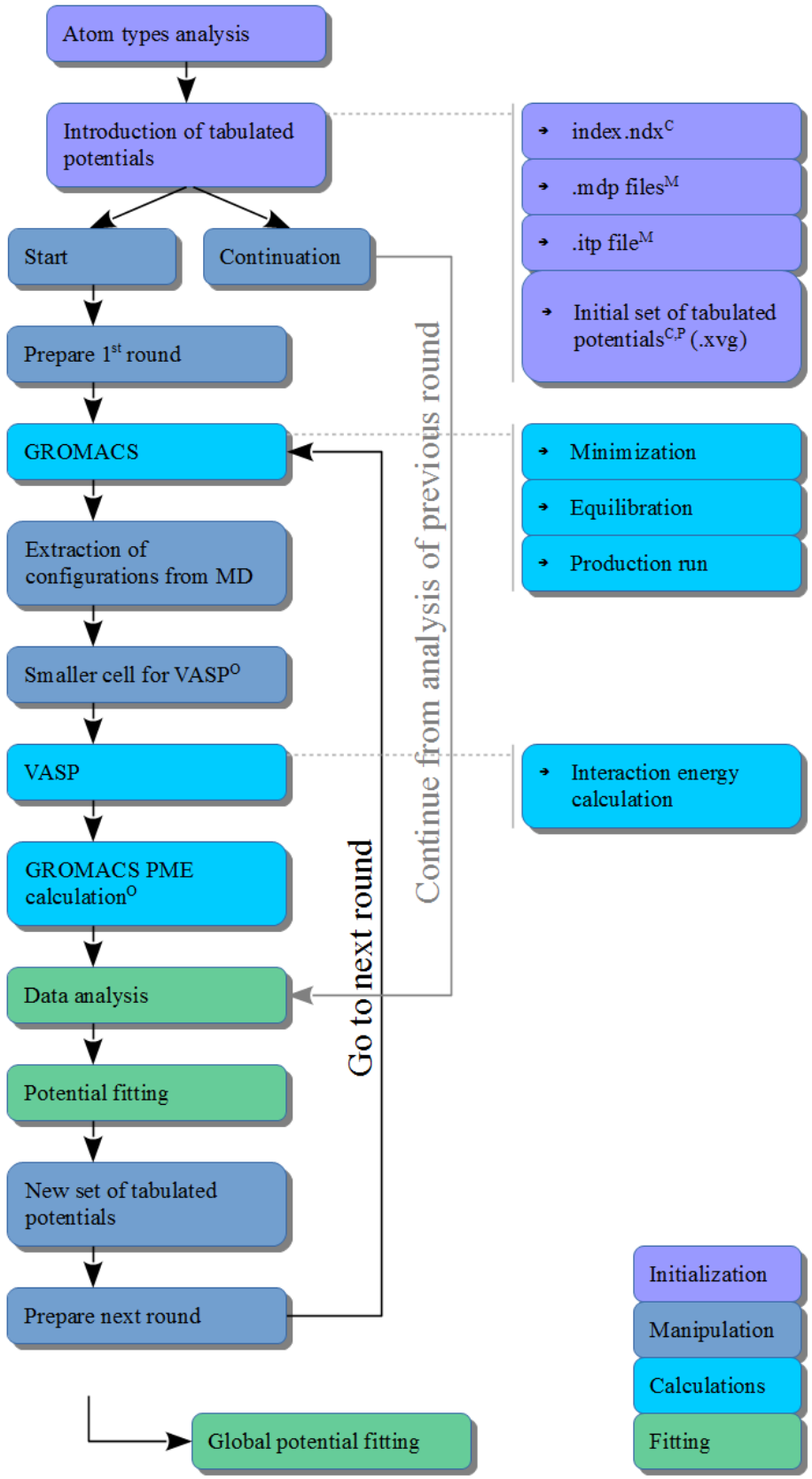
The selected materials have some other interesting aspects. Graphene can serve as a model of graphitic surfaces and the carbon nanotube represents a very

simplified model of pores in the ordered mesoporous carbon (OMC). The UTL zeolite and the UTL lamella are important materials for investigating the ADOR process,[56] which is a new systematic way to synthesize zeolites. A better understanding of interactions between LZFs and organic molecules may thus help to prepare novel materials with tailored properties.

2.3 AIFF development

The force field development is currently one of the fastest growing research areas of computational chemistry. Recently, a number of various approaches capable of force field parameterization from ab initio data appeared in literature.[59,60,61,37] These codes for the AIFF development may differ by their applicability, function forms used and by the choice of a reference structures. Some codes use the minimum energy structure and the corresponding Hessian to determine the shape of the potential energy surface in the vicinity of that minimum.[59,61] A different approach was chosen by Prampolini, Livotto and Cacelli, which is based on reproduction of a much larger part of the potential energy surface. They used MD or MC simulations to generate a set of structures from which they choose geometries for the AIFF fitting using the PICKY procedure.[60] Although this approach is probably not so accurate in the minimum region, it may provide a better description further away from the minimum and it certainly outperforms the "Hessian" approach in the case of multiple minima. Concerning bio-inorganic interfaces, even in the case of AceNMe interacting with anthracene (the smallest cluster model of a peptide on graphitic surfaces) the potential energy surface has quite a complex shape. Therefore, our force field development procedure is also based on sampling from MD calculations, where simulated annealing (increase in temperature) and pulling potential techniques are employed to expand the sampled regions.

Our force field fitting procedure is illustrated in Scheme 1. First, the initialization step connected with the atom type analysis and modifications of GROMACS input files (index.ndx, protocol .mpd and interaction potentials .itp files) is performed. Every atom with a unique symbol and charge is considered as a distinct atom type, atoms in different residues are also distinguished. The atom types used by the fitting procedure and the susequent applications have thus labels like SIZ (silicon in zeolite) and ACE-CT-a (1st CT-type carbon in acetyl group). Additionally, another atom type was introduced for the silicon in the silanol groups. The introduction of atom types requires changes in index file and .mdp files (definition of



Scheme 1. Scheme of our potential fitting procedure
[C] creation; [M] modification; [O] optional
[P] Some potentials from previous parameterizations may be used, these potentials are invariable during this parameterization

interacting energy groups). The non-bonded interaction parameters ε and σ in .itp files were changed to 0.25 kJ/mol and 1 nm, respectively, for all the reparameterized interactions. This simplifies the work with tabulated potentials, since we get rid of prefactors of the vdW potential in tabulated potentials, for all other (non-parameterized) interactions, the combination rules remain valid. The initial set of tabulated potentials is identical to pure combination of force fields – generic FF (genFF – see Section 2.4.2; some tabulated potentials may be used from previous fittings).

This way, in every iteration a new temporary FF (tempFF) is created and this FF is used for the next iteration of the fitting procedure. This approach should sample all relevant parts of the potential energy surface. Finally, a global fit on all structures sampled during all rounds is done, resulting in the newFF (see Figure 2 and Figure A1). The whole force field development is handled by approximately 200 kB of GNU Octave[62] code, with a small part of Linux Shell.

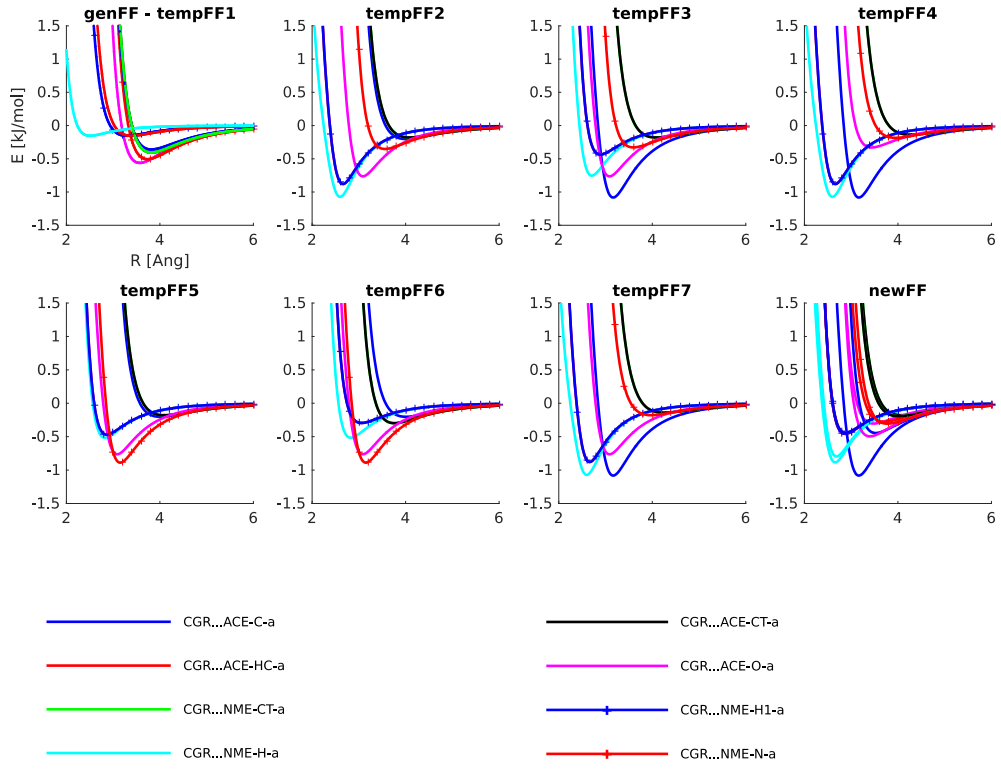


Figure 2. Changes in the potential during the parameterization procedure for AceNMe on graphene. The genFF is used as initial guess for the first round, but in the following rounds different temporary force fields are used for the sampling (only the first tempFFs are depicted). The final global fit results in the newFF. Please note that the parameters for acetyl CT carbon and NMe CT carbon on graphene (CGR) were held the same during the fitting, differing thus only for genFF. The same holds for the HC and H1 hydrogens.

Our newFF is usually a combination of genFF electrostatics (not present for carbon materials), LJ repulsion and scaled D2 correction of Grimme.[25] The D2 correction includes even the damping function, since change of LJ repulsion would otherwise have pathological consequences on the depth of the potential. Using the LJ repulsion and the scaled D2 correction with universal scaling parameter has the advantage of reducing the number of parameters (approximately one half) compared to standard LJ parameterizations. We have found that scaling s_6 values of 0.9525 and 0.9375 are optimal for carbon and siliceous materials, respectively, corresponding to multiplication of the D2 correction (PBE) by the factor of 1.27 and 1.25, respectively. Using the D2 or D3 correction of Grimme as a part of a force field was already suggested in literature.[59,63]

Finally, we briefly overview some other important aspects of our AIFF fitting procedure:

- Fragmentation scheme is used for parameterization of larger molecules, such as peptides or proteins. For instance, we first parameterize the interaction parameters for AceNMe on graphene and subsequently we parameterize the interaction of AceGlyNMe with graphene with the AceNMe parameters fixed, parameterizing thus only the Gly fragment. Although the parameterization of Gly is contaminated by the AceNMe parameterization error, it is probably the best way to parameterize large molecules (see Section 3.3 for the transferability test).
- The global fitting procedure is a standalone code. The global fitting procedure supports various function forms (R^{-1}, R^{-6}, R^{-8}, R^{-10}, R^{-12}, D2 correction of Grimme[25], user-supplied pairwise potential, …) and it gives the best control over the fitting parameters. It may also test various settings with a few additional parameters allowed to change (*e.g.* fitting all of the R^{-12} coefficients and two R^{-1} parameters allowed to change – the code will try all combinations of two R^{-1} parameters and choose the setting minimizing the RMSD).
- In principle, our AIFF fitting procedure may be used to reparameterize the interaction between water and the material, which may improve the accuracy of solvated interface simulations.

- Two cutoff values are implemented: one for the fitted force field and the second for evaluation of dispersion correction for the reference energy. This way, we may parameterize the force field that effectively accounts for the dispersion to 30 Å when using only 12 Å cutoff (this setting was used in major part of this work). This can lead for instance to more accurate adsorption enthalpies, but it is also an inconsistency that may lead to minor errors. Since the energy contribution coming from the asymptote is small (extending the cutoff from 12 to 20 Å in the case of methane in UTL at the PBE-D2 level results in lowering the energy by approximately 0.3 kJ/mol which is roughly 1.5 % of the interaction energy) and we work with periodic materials, we believe that the error of this approximation is negligible.

After the initialization, the fitting is performed in an iterative manner. In every iteration MD simulation is performed. From calculated trajectories a preselected number of frames is extracted and used for the DC-DFT calculation in VASP. Since the GROMACS use a minimum image convention, for a distance cutoff of 12 Å for the vdW potential we need a cell with a minimum size of 24 × 24 × 24 Å. Combination of this condition with the periodicity of the material often leads to very large cells. DC-DFT calculations of such systems would be too computationally demanding, so a conversion of large cell used in GROMACS to a smaller cell for VASP is performed. Such conversion leads to disruption in the periodic boundary conditions, but it has been minimized by application of position restraints on silicon atoms in zeolite (graphene was always frozen during the fitting, so it is not affected).

With very limited possibilities of making changes to the electrostatics, the interaction energy coming from the genFF electrostatics has been calculated (not for graphene where it is zero) and subtracted from the fitted energy, leaving the electrostatics unchanged (except for the short-range correction).

Using the calculated DC-DFT interaction energy values (without deformation), the fitting of AIFF has been performed. We restricted the function forms to those, for which the problem may be linearized. The fitting itself was done by GNU Octave[62] code, which is based on a linear regression and confidence interval (we adopted the regress function from Octave SourceForge[64]) on the significance level $\alpha = 0.05$. The code takes advantage of restriction of

the fitted values according to the following rules:

- Values exceeding the chosen limits are set to the limiting value
- If the confidence interval contains the default value (from genFF), the default value is used
- If the confidence interval contains the limiting value, the limiting value is used

The restrictions were used primarily on the repulsive part of LJ potential, where it keeps the repulsion between 10 and 1000 % of the original term in genFF,[‡] which is extremely important for numerical stability of the AIFF fitting, since the LJ repulsion often balance the electrostatics and switching off or changing the sign of this term may result to unphysical behavior and crashes. Unfortunately, the restriction used for LJ repulsion was not sufficient for some simulations of peptides on α-quartz surface. Especially in the case of carbonyl oxygen the change of LJ repulsion parameters sometimes led to problems, since the strong electrostatic interaction was not balanced by the LJ repulsion anymore. As a consequence, the carbonyl oxygen attracted one of the quartz silicon atoms too strongly, slightly disrupting the Si – O bond in the material. Such structures were partially covalently bound (using the DC-DFT description) and were removed from the reference set. It is quite interesting that this problem occurred almost exclusively in the second iteration of force field parameterization, when the procedure tried to improve the bad sampling of genFF from the first iteration.

2.4 Computational details

2.4.1 Ab initio calculations

The CCSD(T) data on cluster models were calculated with the Molpro program[65]. The related DFT data (except vdW-DF2 and PBE-TS) were calculated with the Gaussian09 program[66] using Dunning's augmented correlation-consistent pVQZ (aug-cc-pVQZ, AVQZ) basis set.[67] These interaction energy calculations were done using the frozen-monomer approach on the MP2/cc-pVTZ optimized geometry and were corrected for the basis set superposition error using the counterpoise procedure.[68] To reduce the computational cost of the DFT calculations and both the HF and MP2 at AVTZ/AVQZ levels the density fitting was applied.[69]

[‡] Since genFF does not have any LJ repulsion on the hydrogen of silanol group, we applied a small LJ repulsion parameter to increase the numerical stability during the fitting procedure.

The vdW-DF2 and PBE-TS calculations were done as a periodic calculation in VASP[70] with sufficiently large cell to minimize the interaction between images, the plane wave basis set is denoted as PW in this work.

The periodic DFT calculations were performed using VASP program[70] using the projector augmented plane waves (PAW).[71] The ENMAX values for oxygen, nitrogen, carbon and hydrogen were 400, 400, 400 and 250 eV, respectively. The kinetic energy cutoff ENCUT was set to 400 eV. The precision (PREC) was set to High. The Brillouin-zone sampling was restricted to the Γ point throughout this work.

Please note that all binding energies are reported without deformation energy contributions. Since only the intermolecular potential was parameterized, such definition is more appropriate.

2.4.2 Molecular dynamics simulations

All molecular dynamics simulations were performed using GROMACS 5.0.5.[72] Initial peptide structures were generated using PyMOL software package and then converted to GROMACS input files (conf.gro, topol.top) using the GROMACS 5.1 pdb2gmx module (since version 5.0.5 is affected by an Amber ff99SB-ILDN-related bug). Initial structures of carbon materials were generated using our Matlab[73] codes, structures of zeolitic materials were used/derived from the International Zeolite Association database.[47] The material topology file was created using our Matlab code.

For the description of a peptide the Amber ff99SB-ILDN force field[31] (in GROMACS 5.0.5 denoted as AMBER99SB-ILDN) was adopted. The water was modeled using the rigid TIP3P model.[74] Carbon materials were treated as frozen (unless stated otherwise) and the Amber parameters for aromatic carbon were used for the non-covalent interactions ($\sigma = 3.39967 \cdot 10^{-1}$ nm, $\varepsilon = 3.59824 \cdot 10^{-1}$ kJ/mol) and for bond-stretching, angle-bending and torsional deformations. The zeolitic materials were modeled using the force field of Bushuev and Sastre.[34,35,§] The pure combination of these force fields, where the cross non-bonded

§ This force field is not implemented in GROMACS, so our own implementation was used. For the exclusion of electrostatics within the O-H of silanol groups the tabulated potentials were used, where the electrostatic interaction is switched off for the short distances of O-H bond. However, all the results in this work were made

interaction terms are based on the Lorentz-Berthelot combination rules, were used as a "generic" force field (genFF). The force field, where the interaction between material and peptide was reparameterized, is denoted as "newFF".

The initial systems were energy minimized and subsequent 0.5 ns equilibration was performed. All bonds with hydrogen atoms were kept fixed using the LINCS algorithm[75,76] throughout this work enabling 2 fs time step. The Coulomb and vdW interactions were treated using PME-User and User methods, respectively. For generic force fields PME[40] and cut-off methods were used, because these simulations did not require the use of the tabulated potentials.[42] Both vdW and real-space PME cut-offs were set to 1.2 nm. The neighbor list for non-bonded interactions was updated every 10 steps leading to negligible error in interaction energy.

The carbon material was usually kept rigid using the "freeze groups" method, the C-C bond length was set to 1.4 Å and the C-C-C angle to 120 deg. As a consequence of "freeze groups" (and carbon material periodicity) some box dimensions cannot change, so anisotropic Berendsen barostat[77] was necessary. These dimensions were held frozen using zero compressibility value, the tau_p values of 0.2 ps for equilibration and 0.5 ps for production run were used, and the reference pressure was set to 1 bar. The material was separated from the rest of the system in the temperature coupling v-rescale algorithm,[78] the material and the rest of the system had the reference temperature set to 300 K unless stated otherwise, tau_t 0.05 ps for equilibration and 0.5 ps for production run. The frequency for pressure coupling (nstpcouple) was set to 3 since higher values resulted in simulation instabilities in the initial phase of equilibration and even more strict energy minimization did not help.

Due to different character of the modeled systems some variations in the settings were necessary:

- The fitting of the force field was carried out in iterations. Every iteration (round in Scheme 1) consisted of minimization and equilibration followed by 4 ns production run, the simulation was performed in vacuum - NVT ensemble. To enhance the

with a fixed O-H bond length, so the introduction of tabulated potential for the material was not necessary, since the O-H bond contributes to the potential by a constant term.

sampling, the pulling and simulated annealing techniques were used. Harmonic potential acting in increasing distance from the graphene sheet was added, maximal displacement 0.4 nm with respect to starting position of the peptide, force constant 1000 kJ/mol·nm^2. The simulated annealing repeatedly increased and decreased the temperature to 400 and 300 K, respectively, to sample all regions at various temperatures.

- The transferability test consisted of NpT equilibration at 350 K followed by 25 ns NVT simulation during which the temperature was linearly decreased to 0 K. The decrease in pressure caused a formation of empty cavities in the solvent, but these cavities were located far from water-graphene interface, where the peptide is located, so the peptide was solvated throughout the simulation. Water molecules were removed from the system obtained from simulated annealing procedure and graphene-peptide interaction energy on various levels of theory were calculated. The size of simulation box in GROMACS was 4 × 4 × 4 nm, for VASP calculations the cell size of approximately 2 × 2 × 2 nm was used.
- The adsorption of methane in UTL zeolite was modeled using CH_4 model derived from Amber parameters[31] as implemented in GROMACS[72]. We used a fixed bond length of 1.09 Å and flexible angle with equilibrium value 109.5 degrees and force constant 292.880 kJ/mol·rad^2. The simulations were performed with both rigid and flexible UTL zeolites. The tau_t value was set to 0.05 ps for both equilibration and production run for better convergence to the 298 K reference temperature. Neighbor list was updated in every step to obtain the best possible accuracy. The length of production run was 20 and 100 ns for rigid and flexible zeolite, respectively. Despite the very long simulation time and monitoring the energy every 20 ps the energy was still not fully converged and an error in the order of few tenths of kJ/mol may be expected. The interaction energy was corrected for the center of mass motion removal effects.
- The study of N-acetyl-N-methylamine on graphene sheet was performed both in vacuum (NVT) and in water (NpT). The simulation time of 10 ns was modeled; the statistics were done over approximately 9000 points. The size of the cell was approximately 2.5 × 2.5 × 3 nm.

- In the conformational studies strong position restraints were applied on the peptides during equilibration to prevent significant structural changes in the starting structure. Simulation time was 200 ns. The simulations of peptides in nanotubes (water present) were carried out in the NVT ensemble. Since the simulation was started from extended peptide, the initial size of the cell dimension was set to the length of extended peptide plus 1 nm. Since the peptide usually sample other (more compact) regions, it should provide sufficient space to minimize the interaction of peptide images. In the nanotubes the peptides were aligned to the x axis and the box dimensions in y and z axis were set to 4 nm. To determine the effect of nanotube flexibility, the simulation was repeated with relaxed nanotubes (NVT). In that case the nanotube was optimized first, its diameter increased by approximately 0.02 nm, then the peptide and water was inserted and the whole procedure of minimization, equilibration and production run was carried out. The analyses were done over time from 50 to 200 ns, the capping groups and two terminal amino acids on each side of the peptide were removed from the analysis to minimize the effect of terminal groups. The secondary structures were determined using DSSP program (version 2.0.4).[79]

3 RESULTS

The nature of the interaction between organic molecules/biomolecules and selected graphitic and siliceous materials depends on properties of the investigated materials:

- The interaction with carbon materials is dominated by dispersion forces. Among some more specific interactions occurring on the carbon surfaces we should mention the interaction of positively charged hydrogen with the carbon π-electron system and the π-stacking (*e.g.* π-π interaction between the graphene and the aromatic ring of phenylalanine).
- The UTL zeolite and fully reconstructed α-quartz surface also interact mainly through dispersion, but the character of the interaction reflects also the partial negative charge on oxygen atoms and twice as large positive charge on silicon atoms, so the electrostatics plays an important role here.
- The UTL lamella has some specific properties compared to the previous case, as it contains the silanol groups that may act as donors and acceptors of hydrogen bonds. Moreover, the silanol groups are sensitive to pH, in alkaline solution they deprotonate, so the lamella gains a negative charge. The protonation of silanol groups occurs rather unwillingly, even in highly acidic solutions.[80,81]

The interaction of materials with N-acetyl-N-methylamine (AceNMe) is illustrated by Figure 3. It is obvious, that the dispersion (D2 correction) term represents the major contribution to the interaction with graphene, since the PBE functional gives only weak binding or repulsive interaction (due to exchange repulsion). The situation is more complicated on quartz, where the interaction is composed mainly from dispersion and electrostatics. Since the most important interaction terms covered by PBE are electrostatics and exchange repulsion (but not dispersion), the correlation between PBE and generic FF electrostatics (genFF PME) may indicate the quality of electrostatics description of genFF. Indeed, such correlation was observed for the most attractive regions for both quartz and IPC-1P. Unfortunately, the correlation is not very good and especially in the repulsive regions the relative error of genFF electrostatics seems to be very high. The D2 correction plus the genFF electrostatics can be

used as a rough estimate of interaction energy, but this approach fails dramatically in the exchange repulsion region.

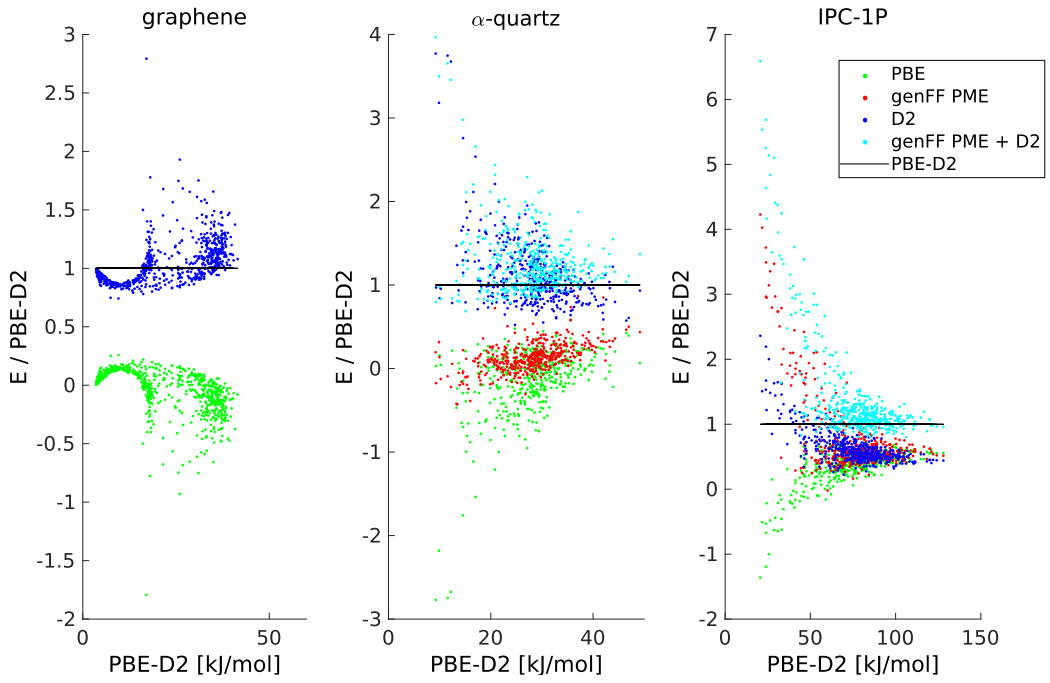


Figure 3. Comparison of the studied systems. Our reference data for AceNMe on materials (graphene, relaxed α-quartz and relaxed IPC-1P) were used for the analysis. Various energy terms as a fraction of the reference PBE-D2 binding energy are shown as a functions of the absolute binding energy. The electrostatics of the genFF (genFF PME) was added to assess the quality of the electrostatics description by genFF.

3.1 Performance of ab initio methods

To assess the quality of various ab initio methods, the interaction of N-acetyl-N-methylamine (AceNMe; used as a model of peptide bond) with cluster models of investigated materials was evaluated at the CCSD(T), MP2 and DC-DFT levels. Three configurations of AceNMe on anthracene (graphene model) and two configurations of AceNMe on 2T model (silica model representing zeolitic material) were considered. The three distinctive structures of AceNMe on anthracene may be characterized as “O down”, “H down” and “parallel” (see Figure 4). The first is the least stable, since the partially negatively charged oxygen atom interacts with

the negative π-electron cloud of anthracene, thus leading to repulsive electrostatic contribution acting against the attractive dispersion. In contrast, the second structure maximizes the electrostatic attraction, since the positive hydrogen atom of AceNMe is attracted by the π-electron system. The third structure is the most stable due to the maximized dispersion interaction in the π-π stacking arrangement. Interestingly, our calculations suggest that the energetic order of these structures is different for naphthalene (these results are not presented here), where the "H down" structure has lower energy than the "parallel" structure. This difference is caused by a weaker dispersion in the case of naphthalene, which influences the energy of "parallel" structure more than the "H down" structure. Most likely, the stacking arrangement will be the most stable structure for all acenes larger than naphthalene.

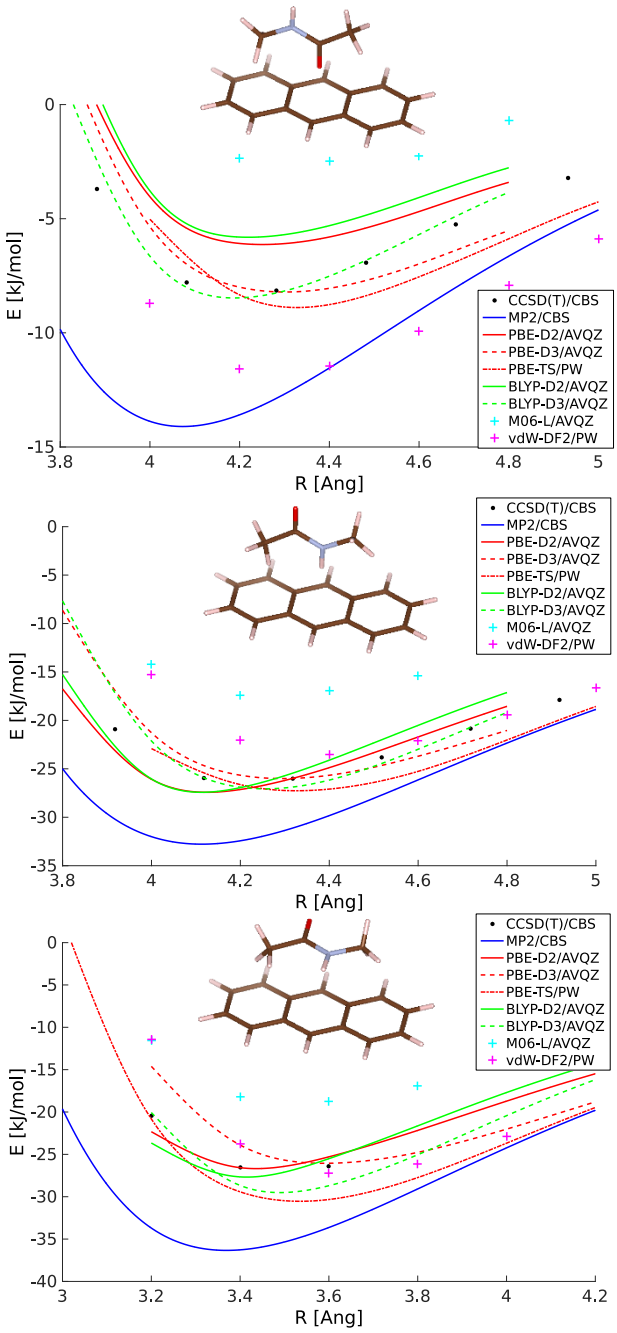


Figure 4. Structures of AceNMe on anthracene and the potential energy in vicinity of the minimum.

For the 2T model, the hydrogen bond-like structure is strongly preferred. From Figures 4 and 5, we can draw some conclusions about the accuracy of various methods. The MP2/CBS method severely overestimates the dispersion contribution on anthracene, but its performance on the 2T model is highly satisfactory. The M06-L method gives lower binding energies for both systems. The vdW-DF2 and PBE-TS methods systematically overbind on the 2T model, their performance on anthracene is also worse than the performance of the empirical PBE-D2 and PBE-D3 methods. The performance of BLYP and PBE methods corrected by D2 or D3 dispersion correction is more or less the same. Thus, the simplest PBE-D2 method was chosen as our reference theory for the force field parameterization.

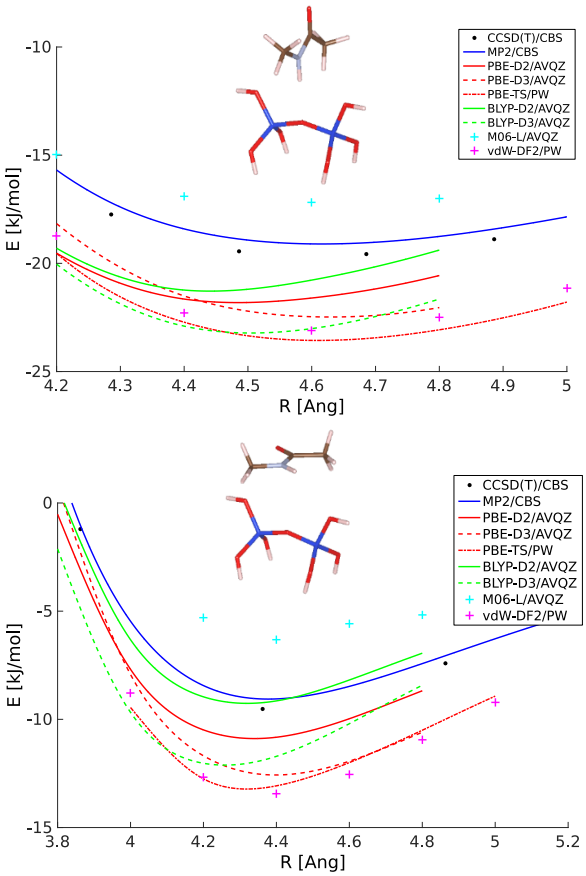


Figure 5. Structures of AceNMe on the 2T silica model and the potential energy in vicinity of the minimum.

3.2 Methane in the UTL zeolite: a case study

To explore the limits of our approach, we tried to model the physisorption of methane in the UTL zeolite. Since the physisorption of methane in pure silica zeolite framework is very weak, extremely high accuracy is necessary as even small absolute error of 1 kJ/mol corresponds to

Table 1. Parameterization of methane in UTL.

System	Reference theory	# of structures	RMSD	Maximal positive error	Maximal negative error	Reference[a] Max.	Reference[a] Avg.
CH_4 – UTL	PBE-D2	330	0.77	2.42	-3.63	23.76	15.56
CH_4 – UTL	PBE/CC	330	0.81	2.91	-4.61	19.64	12.26

[a] Maximal and average binding energy values for the corresponding reference theory were added for comparison. Please note that these values are affected by the fitting protocol (*i.e.* pulling potential) and also by the quality of temporary FFs.

large relative error (5-10%). Because of the high desired accuracy, we applied not only the PBE-D2 reference, but we also used our DFT/CC scheme to parameterize extremely accurate PBE/CC method.** Please note that even the state-of-the-art DC-DFT methods may be inadequate for some applications. Our tests suggest that PBE-D2 (and also PBE-D3) significantly overestimates the binding energy (at least in the case of methane in pure silica materials) and the deficiencies of the PBE-D2 method cannot be ignored for methane physisorption (see Figure 6). For peptides on silica materials, however, the relative error is much lower as it is obvious from the results reported in Section 3.1.

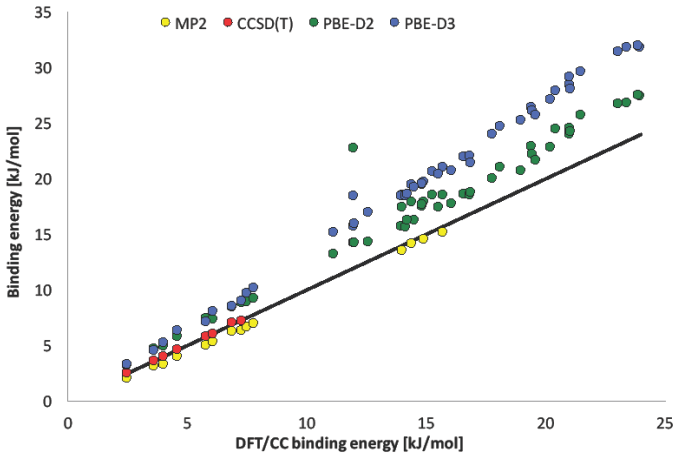


Figure 6. Methods comparison for methane on various sites in zeolites of the UTL family and their cluster models.

The same set of structures was used for both parameterizations (PBE-D2 and PBE/CC), only the reference energy differs (see Table 1). The CH_4 – UTL interaction was parameterized directly in UTL, but some tests on CH_4 – quartz and CH_4 – UTL have been performed to clearly show that force field parameterized on quartz is transferable to UTL and vice versa. In this case we used the same cutoff for the force field and for the reference theory (12 Å, see Section 2.3), since we add the exact energy contribution coming from the asymptotic region as a correction. The averaged CH_4 – UTL interaction energies (including the deformation contribution) for frozen and relaxed zeolite differ only marginally. Much larger error is introduced by the choice of reference theory – averaged interaction energy between -13.3 and -13.6 kJ/mol was calculated for AIFF with the PBE/CC reference and -17.0 kJ/mol for AIFF with the PBE-D2.

** More details about the DFT/CC parameterization will be presented in a forthcoming paper from our group (in preparation).

3.3 Physisorption of peptides on graphene

Peptides on graphene were among the first bio-inorganic systems parameterized in this work. Since most of the technical details were already mentioned in previous sections, only the results of the parameterization will be summarized. Our best newFF potentials obtained to this date were not available when we tested the transferability and accuracy of our AIFFs, and therefore the results presented in Table 3 and Figures 7, 8 and A2 were calculated with potentials of slightly worse quality (see Table A1).

From Table 2 it is obvious that the accuracy of our newFF is very good, much better that the accuracy of the generic FF (genFF). The biggest difference between newFF and genFF description was found in the case of AceLeuNMe, where the original genFF has the root-mean-square deviation (RMSD) over 10 kJ/mol and the reparameterized newFF has RMSD less than 2 kJ/mol with respect to the reference level of theory (PBE-D2). The genFF also fails dramatically for aromatic Phe and Tyr amino acids. Our results clearly show that genFF systematically overbinds the interaction with graphene (mean signed errors in binding energy are 0.8, 3.1, 2.6, 3.2, 6.3 and 3.9 kJ/mol for AceNMe, AceGlyNMe, AceAlaNMe, AceValNMe, AcePheNMe and AceSerNMe, respectively). Please note that the largest errors of genFF come from the exchange repulsion region of PBE-D2, which indicates that the genFF tends to different equilibrium distances than the PBE-D2 method.

Table 2. Parameterization of binding energies of biomolecules on graphene. All energy values in kJ/mol.

System	***newFF***			***genFF***			***reference set***		
								PBE-D2[a]	
Peptide	RMSD	Maximal positive error	Maximal negative error	RMSD	Maximal positive error	Maximal negative error	# of struct.	Max.	Avg.
AceNMe	0.7	2.9	-3.5	3.1	6.9	-19.7	684	41.5	22.6
AceGlyNMe	1.6	10.0	-5.2	5.0	13.2	-17.4	630	67.6	40.5
AceAlaNMe	1.6	5.6	-8.4	5.9	12.7	-25.2	681	68.4	40.4
AceValNMe	1.9	6.4	-6.8	7.0	14.9	-44.3	730	78.1	45.0
AcePheNMe	1.1	4.0	-4.8	8.5	19.9	-26.7	750	91.6	57.3
AceLeuNMe	1.8	6.0	-7.6	10.2	16.6	-58.9	280	85.3	50.9
AceIleNMe	1.5	7.9	-5.5	6.6	15.4	-36.0	480	85.2	47.7
AceSerNMe	2.1	11.8	-8.7	7.5	15.1	-65.8	495	74.1	41.4
AceTyrNMe	1.2	5.9	-5.6	9.8	19.7	-33.3	600	98.6	62.7
AceCysNMe	1.4	5.8	-5.4	6.8	19.2	-23.9	520	72.7	43.5
AceProNMe	1.4	4.7	-6.6	8.0	14.8	-43.0	350	70.8	42.4

[a] Maximal and average PBE-D2 binding energy values were added for comparison. Please note that these values are affected by the fitting protocol (*i.e.* pulling potential) and also by the quality of temporary FFs.

As was already mentioned, the parameters for particular amino acids were parameterized using capped amino acids (Ace-amino_acid-NMe), where the parameters for Ace and NMe were taken from previous parameterization and these parameters remained unchanged. Since the accuracy and transferability of such approach may be questioned, a transferability test was performed. Capped amino acid dimers ($AceGly_2NMe$, $AceAla_2NMe$, $AceVal_2NMe$ and $AcePhe_2NMe$) on graphene surface solvated by water were modeled. The simulated annealing for both the newFF and genFF (see Computational details) was performed and the resulting structures (with solvent removed) were used for a calculation of interaction energies at the PBE-D2 level. The results of the transferability test in Table 3 show an excellent performance of newFF (reproducing the PBE-D2 data within 2 kJ/mol) on structures from both newFF and genFF simulations. The overbindning of genFF is obvious from Table 3. Quite interestingly, the error of genFF is much lower for the structures obtained from newFF than for its own structures – genFF performs rather badly in the region of the minimum structure. Seemingly, genFF leads to structures with larger PBE-D2 binding energy, but it is probably just an artifact of overbinding between peptide and graphene, so the balance between peptide interaction with graphene and water is disturbed, strongly favoring the first.

In summary, we have presented benchmarking of newFF and genFF parameterizations. The analysis has been mainly targeted on differences in performance from structural and energetic points of view, but how these differences affect the actual sampling in molecular dynamics simulations remains to be seen. For this purpose, we modeled AceNMe on graphene in vacuum and solvated in water, the sampled regions were compared for newFF and genFF parameterizations. The results are illustrated in Figure 7. In vacuum the genFF entirely ignores the N-H…π interaction. In contrast, there is no significant difference between genFF and newFF in water. This may be explained by the effect of the water, which strongly interacts

Table 3. Transferability test of one-residue-longer peptides on graphene. All binding energies in kJ/mol.

System	***newFF geometry***			***genFF geometry***		
Peptide	PBE-D2	newFF	genFF	PBE-D2	newFF	genFF
$AceGly_2NMe$	95.96	93.79	104.04	90.63	88.72	107.14
$AceAla_2NMe$	75.81	76.36	76.66	88.47	87.91	110.02
$AceVal_2NMe$	84.59	86.16	90.24	95.52	93.13	105.13
$AcePhe_2NMe$	123.68	122.43	141.37	133.93	131.84	160.64

with the carbonyl oxygen. Consequently, the dynamics in water is driven by the water-AceNMe interaction.

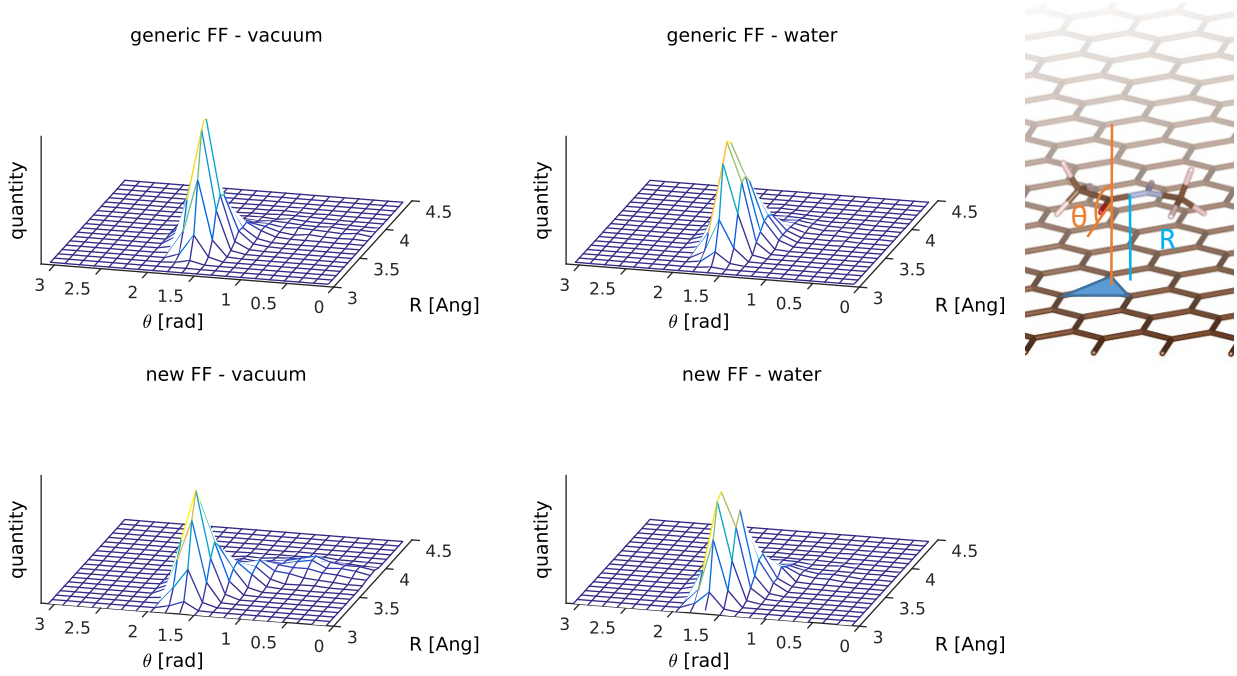


Figure 7. Simulation of AceNMe on graphene – the difference in sampling caused by a potential.

3.4 Conformational study of peptides in carbon nanotubes

We have investigated the dynamics of polypeptides (AceGly$_{15}$NMe, AceAla$_{15}$NMe, AceVal$_{15}$NMe and AcePhe$_{15}$NMe) in confined space of carbon nanotubes. A good transferability of graphene parameters to curved graphitic surfaces (carbon nanotubes) has been assumed throughout this study. The initial tests indicated that the interaction of polyphenylalanine side chains with nanotube walls is too strong – preventing any interesting conformational changes and resulting in the polyphenylalanine adsorbed randomly on the nanotube walls at the time scale of our MD simulations (~500 ns). Consequently, the phenylalanine was not included in the following conformational analysis. For the remaining three amino acids, the formation of helical secondary structures as a function of nanotube diameter was thoroughly investigated. For polyglycine, 3_{10}-helix formation propensity has been observed for small diameter nanotubes, with increasing nanotube diameter α-helix starts to form and even larger nanotubes support the formation of π-helices. From certain diameter,

however, the tendency to form helices steeply drops. The situation is quite different for polyalanine, which only prefers to form α-helix for a wide range of nanotube diameters. Polyvaline tends to form 3_{10}-helix for smaller diameters. With increasing diameter, a small fraction of α-helix also appears, but the propensity for α-helix seem to be much weaker than for polyalanine. Comparison between relaxed and frozen nanotubes was also made (see Figure 8 and Figure A2), but it did not show any significant differences. Only small changes in propensity and in the particular ranges were observed, but these changes may be the result of statistical error.

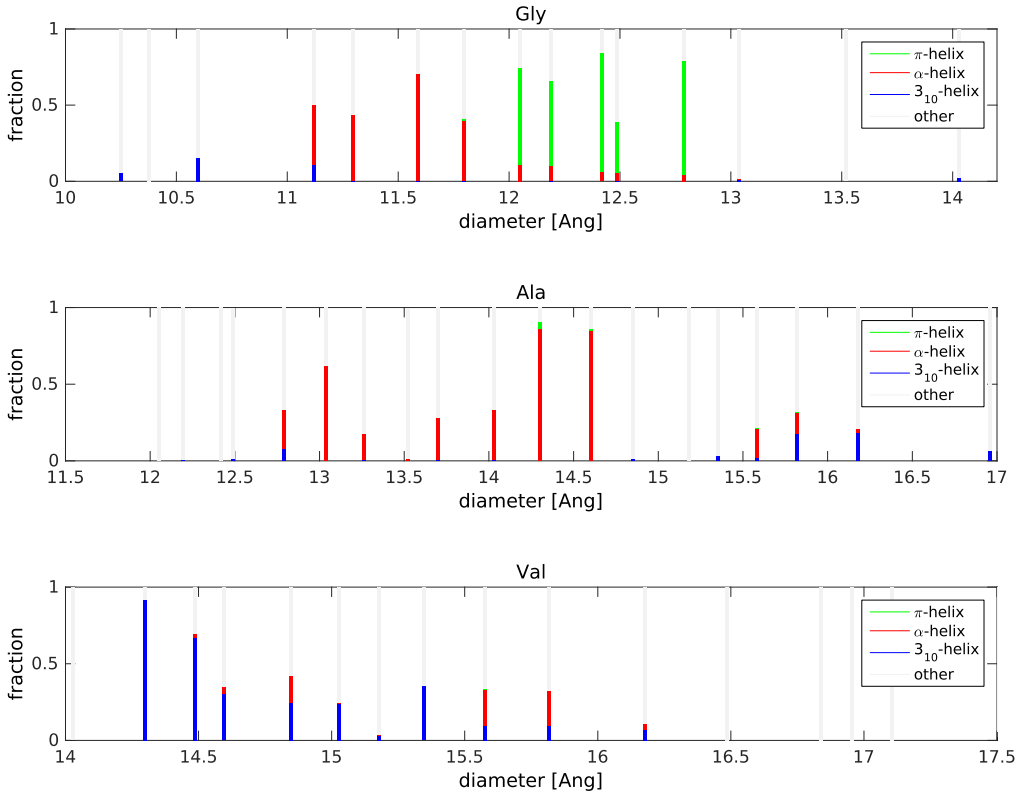


Figure 8. Formation of polypeptide helical structures in relaxed carbon nanotubes as a function of nanotube diameter.

The stabilization of helical structures in confined space of cylindrical symmetry was already studied by Ziv, Haran and Thirumalai, where the effect of ribosome exit tunnel on helical propensity of synthesized polypeptide chains was investigated.[82] The authors claims that the helices are entropically stabilized, because the cylindrical symmetry dramatically restricts the conformation space of various coil states, but the conformational space of various helices remains almost unchanged. Despite the very simple model used in that study (coarse-

grained model without explicit solvent) and very different system, the results may be of interest in context of our work, since the authors admitted that the role of symmetry may be even more important than specific interactions between ribosome and the peptide.[82] The stabilization of helical structure should be largest for a tunnel diameter of approximately 1.3 times the critical helix diameter (smallest diameter for which the helix may be formed). However, the presence of solvent may affect these trends, as suggested in literature. Sorin and Pande[83] investigated $AceAla_{21}NMe$ helix formation in hydrated carbon nanotubes, their work thus explores almost the same system as ours, the main difference is in the employed nanotube diameters. The diameters used by Sorin and Pande are 14.9, 17.6, 20.3, 25.8 and 35.3 Å, so their nanotubes are much larger than in our case. According to their results, the observed helicity is low (due to solvent entropy), increasing with nanotube diameter and favoring 3_{10}-helix over α-helix. Our polyalanine results with nanotube diameter >15 Å also shows small fractions of α-helix and 3_{10}-helix, but we cannot make conclusions about these large nanotube diameters, since we were mostly concerned about narrower nanotubes and we do not have enough data for those diameters. The region of strong α-helix preference predicted by us was not observed by Sorin and Pande, since even their smallest nanotube with diameter of 14.9 Å is too large (please note the significant drop in helicity near 15 Å in Figure 8). At first sight, the nanotube diameter of 15 Å may seem to be ideal for polyalanine α-helix formation, since it corresponds to the optimal diameter based on vdW radii. To stabilize the helical structure even more, however, the nanotube should be much narrower (or the helix wider, but it would significantly weaken the intramolecular hydrogen bonds), which optimizes the total interaction of side chain atoms (and backbone) with the nanotube walls.

3.5 Adsorption of peptides on 2D silica surfaces

The parameterization of AIFF for biomolecules on the fully-reconstructed α-quartz surface and especially on the IPC-1P lamella represents a much more complicated task than the parameterization for graphitic surfaces described in the previous sections. The crucial role of electrostatic interactions is not the only problem. In contrast to graphene, we attempted to parameterize the interaction on flexible material to capture the effect

Table 4. Parameterization of binding energies of AceNMe on 2D silica surfaces. Comparison between relaxed and frozen materials. All energy values in kJ/mol.

System	***newFF***			***genFF***			***reference set***		
								PBE-D2[a]	
Material	RMSD	Maximal positive error	Maximal negative error	RMSD	Maximal positive error	Maximal negative error	# of struct.	Max.	Avg.
quartz	2.2	8.6	-8.9	10.5	1.9	-51.3	450	49.2	28.0
IPC-1P	8.8	28.6	-45.8	40.1	6.1	-201.1	547	128.0	61.2
quartz[b]	1.3	4.8	-7.6	7.6	1.7	-32.9	525	39.1	23.9
IPC-1P[b]	4.1	18.9	-10.6	12.2	14.1	-26.5	390	84.8	58.2

[a] Maximal and average PBE-D2 binding energy values were added for comparison. Please note that these values are affected by the fitting protocol (*i.e.* pulling potential) and also by the quality of temporary FFs.
[b] Frozen material. Interaction capabilities of rigid IPC-1P are significantly restricted by the rigid silanol groups.

of the hindered rotation of silanol groups. Moreover, the number of interactions to parameterize is much larger than in the case of graphene – instead of one carbon atom type in graphene, there are two atom types in quartz (Si and O) and five in IPC-1P lamella (Si, O and silanol Si, O, H). To simplify the problem, we assumed transferability of the quartz Si and O parameters to IPC-1P.

From Table 4 we see that the parameterizations on two-dimensional silica materials lead to much higher RMSD values than were observed for graphene. The genFF fails dramatically for silica materials, with the largest errors of genFF again in the exchange repulsion region. It seems that the genFF parameterization systematically underbinds on silica materials (as is illustrated in Figure A3), which is in sharp contrast to graphene, where genFF overbinds. The newFF parameterization is significantly better, on rigid materials it seems to perform better than on the flexible material. These conclusions should be taken with care, since the frozen silanol groups drastically limit the range of possible interactions. The quality of the AceNMe – quartz potential is reasonably good for both the relaxed and frozen materials, but the accuracy for AceNMe on IPC-1P is not so satisfactory. There are various possible reasons for the difficulties observed in the case of IPC-1P:

- Too many interaction parameters leading to unreliable fit
- Wrong electrostatics of the silanol groups
- Polarization effects
- The geometry of the subsystems (peptide, graphene) predicted by the FF is too far from PBE-D2 minimum – such incompatibility of structures may lead to interaction polluted by partially covalent binding, because the structures are too “unreasonable” for DFT

- The robustness of the sampling may be bad, leading to too many structures in the repulsive range, such structures may be negatively affecting the fit
- The potential forms allowed does not have sufficient flexibility to describe the interaction accurately

It is still a subject of ongoing analysis which of these aspects have the largest negative effect on the force field quality. One of the tests performed is the transferability of AIFF between relaxed and frozen materials. The results are presented in Table 5. It seems that the potential fitted for the relaxed materials works quite reasonably even on the rigid materials, but the AIFF fitted on rigid IPC-1P gives much worse performance on the relaxed lamella. Omission of the most repulsive structures from the reference also leads to a significant improvement.

Since the quality of AceNMe description on quartz was reasonable, we have continued to parameterize amino acids on quartz (see Table 6). The quality of these parameterizations is worse than on graphene, but still quite satisfactory – with RMSD values between 2.6 and 4.4 kJ/mol. The worse performance of the Ala and Gly parameterizations is probably caused by presence of more repulsive structures in the reference set.

In summary, the AIFF parameterization of peptides on 2D silica surfaces is much more challenging than on graphene, especially when silanol groups are present. Although the RMSD values obtained may seem to be too large (especially for parameterizations on IPC-1P), the improvement of newFF with respect to genFF is still significant. Further work is planned to increase the accuracy of the newFF potential (*e.g.* by incorporating polarization effects).

Table 5. Transferability test between parameterizations on frozen and relaxed materials, results in kJ/mol.

Reference structures		RMSD frozen	RMSD relaxed	***Reference structures***		RMSD frozen	RMSD relaxed
System	Material	FF	FF	System	Material	FF	FF
AceNMe – quartz	relaxed	3.0	*2.2*[a]	AceNMe – quartz	frozen	*1.3*[a]	2.3
AceNMe – IPC-1P	relaxed	14.1	*8.8*[a]	AceNMe – IPC-1P	frozen	*4.1*[a]	5.6

[a] not transferred parameterizations for comparison

Table 6. Parameterization of binding energies of small peptides on relaxed α-quartz. All energy values in kJ/mol.

System	***newFF***			***genFF***			***reference set***		
								PBE-D2[a]	
Peptide	RMSD	Maximal positive error	Maximal negative error	RMSD	Maximal positive error	Maximal negative error	# of struct.	Max.	Avg.
AceGlyNMe	4.0	15.0	-30.9	38.3	2.69	-250.7	295	67.4	45.2
AceAlaNMe	4.4	16.6	-32.0	31.6	---	-231.6	293	74.5	49.0
AcePheNMe	2.6	7.0	-8.1	15.4	---	-54.2	540	94.5	67.9
AceLeuNMe	3.7	11.2	-26.4	19.4	---	-153.5	540	82.4	58.6
AceGlnNMe	3.4	15.7	-17.8	17.2	---	-112.5	550	98.8	61.8

[a] Maximal and average PBE-D2 binding energy values were added for comparison. Please note that these values are affected by the fitting protocol (*i.e.* pulling potential) and also by the quality of temporary FFs.

4 CONCLUSIONS

Ab initio quantum chemistry methods are commonly used computational tools to describe interaction of guest molecules with inner and outer surfaces of porous materials. Although these methods are capable of calculating equilibrium structures and binding energies very accurately, their use for predicting macroscopic properties such as isosteric heats of adsorption is impractical due to very high computational costs. Therefore, the classical simulation methods such as Grand Canonical Monte Carlo and Molecular Dynamics mostly rely on empirical force fields parameterized on a set of experimental data augmented with calculated (typically ab initio) benchmarks.

In this work, we have focused mainly on the development of parameterization strategies for bio-inorganic interfaces. These strategies are based solely on reliable ab initio data yielding system-specific but highly accurate first-principle force fields. Dispersion-corrected density functional theory methods have been employed in our parameterization procedure as they represent the only practical alternative for extended systems. Two types of inorganic materials have been investigated, graphitic nanostructured materials (graphene and carbon nanotubes) and silica based nanomaterials (2D zeolites). The performance of ab initio methods for proteins acting as guest molecules has been assessed on the interaction of N-acetyl-N-methylamine and non-polar amino acid side chains in the C_β representation with cluster models of investigated materials. Several configurations of AceNMe on graphene model (anthracene) and on the 2T model representing siliceous materials were evaluated at the CCSD(T), MP2 and various DC-DFT levels of theory. Based on these benchmarks, the PBE-D2 method has been selected as a reasonably accurate and computationally efficient compromise. The developed AIFFs (dubbed as newFF) have been successfully applied to a few interesting processes involving bio-inorganic systems, namely physisorption of peptides on graphene, conformational dynamics of peptides in confined space of carbon nanotubes, and adsorption of peptides on siliceous 2D materials. From the methodology point of view, the main emphasis has been given to comparison between generic FF (genFF) constructed as a combination of common force fields without reparameterization and newly developed AIFFs (newFF). The results presented in Section 3 clearly show that the newFF parameterization is vastly superior to genFF proving thus the usefulness of our approach.

Several issues related to development AIFF have been addressed, namely: (i) when the accuracy of the underlying DC-DFT is insufficient and (ii) in situations (*e.g.* hindered rotational dynamics of surface silanol groups) where the functional forms of currently implemented force fields are inadequate to reproduce ab initio data quantitatively. While the first problem was completely solved by employing the DFT/CC method at modest computational costs involving cluster calculations at the CCSD(T) level, the search for sufficiently flexible functional forms has proved to be a tough problem. A list of possible solutions of this problem include: (i) improved electrostatics of the silanol groups, (ii) incorporating polarization effects, (iii) better FF description of the subsystem geometries (peptide, graphene, 2D zeolite), (iv) enhanced robustness of the sampling eliminating difficulties with too many structures in the repulsive region, (v) improved flexibility of functional form (other than electrostatics and polarization). Work along these lines is currently in progress.

Very high accuracy of our potentials might be beneficial in various research fields, especially in adsorption, separation, and catalysis where our future work will be targeted. Besides further improvement and automation of our AIFF procedure, extension to charged species (amino acids and organic structure directing agents) and other materials (MOFs, aluminosilicates, *etc.*) would be highly desirable.

5 REFERENCES

1. Duan, Y.; Kollman, P. A., Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **1998,** *282* (5389), 740-744.
2. Levy, Y.; Onuchic, J. N., Water and proteins: A love-hate relationship. *P. Natl. Acad. Sci. USA* **2004,** *101* (10), 3325-3326.
3. Lu, J. R.; Su, T. J.; Thomas, R. K.; Penfold, J.; Webster, J., Structural conformation of lysozyme layers at the air/water interface studied by neutron reflection. *J. Chem. Soc. Faraday T.* **1998,** *94* (21), 3279-3287.
4. Wolynes, P. G., Biomolecular folding in vacuo!!!(?). *P. Natl. Acad. Sci. USA* **1995,** *92* (7), 2426-2427.
5. Meister, K.; Strazdaite, S.; DeVries, A. L.; Lotze, S.; Olijve, L. L. C.; Voets, I. K.; Bakker, H. J., Observation of ice-like water layers at an aqueous protein surface. *P. Natl. Acad. Sci. USA* **2014,** *111* (50), 17732-17736.
6. Prabhu, N.; Sharp, K., Protein-solvent interactions. *Chem. Rev.* **2006,** *106* (5), 1616-1623.
7. Castner, D. G.; Ratner, B. D., Biomedical surface science: Foundations to frontiers. *Surf. Sci.* **2002,** *500* (1-3), 28-60.
8. Fenoglio, I.; Fubini, B.; Ghibaudi, E. M.; Turci, F., Multiple aspects of the interaction of biomacromolecules with inorganic surfaces. *Adv. Drug Deliver. Rev.* **2011,** *63* (13), 1186-1209.
9. Vallet-Regi, M.; Ramila, A.; del Real, R. P.; Perez-Pariente, J., A new property of MCM-41: Drug delivery system. *Chem. Mater.* **2001,** *13* (2), 308-311.
10. Nawrocki, J., The silanol group and its role in liquid chromatography. *J. Chromatogr. A* **1997,** *779* (1-2), 29-71.
11. Mateo, C.; Palomo, J. M.; Fernandez-Lorente, G.; Guisan, J. M.; Fernandez-Lafuente, R., Improvement of enzyme activity, stability and selectivity via immobilization techniques. *Enzyme Microb. Technol.* **2007,** *40* (6), 1451-1463.
12. Wang, J., Carbon-nanotube based electrochemical biosensors: A review. *Electroanal.* **2005,** *17* (1), 7-14.
13. Orasanu-Gourlay, A.; Bradley, R. H., Protein adsorption by basal plane graphite surfaces: Molecular images and nano-structured films. *Adsorpt. Sci. Technol.* **2006,** *24* (2), 117-130.
14. Cramer, C. J., *Essentials of computational chemistry: theories and models*. John Wiley & Sons: 2013.
15. Szabo, A.; Ostlund, N. S., *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation: 1989.
16. Parr, R. G.; Yang, W., *Density-Functional Theory of Atoms and Molecules*. Oxford University Press: 1994.
17. Piela, L., *Ideas of quantum chemistry*. Elsevier: 2013.
18. Sokkar, P.; Choi, S. M.; Rhee, Y. M., Simple Method for Simulating the Mixture of Atomistic and Coarse-Grained Molecular Systems. *J. Chem. Theory Comput.* **2013,** *9* (8), 3728-3739.
19. Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P., Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories. *Chem. Rev.* **2010,** *110* (9), 5023-5063.
20. Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K., Basis-set convergence in correlated calculations on Ne, N2, and H2O. *Chem. Phys. Lett.* **1998,** *286* (3-4), 243-252.
21. Jurecka, P.; Hobza, P., On the convergence of the (Delta E-CCSD(T)-Delta E-MP2) term for complexes with multiple H-bonds. *Chem. Phys. Lett.* **2002,** *365* (1-2), 89-94.
22. Lee, K.; Murray, E. D.; Kong, L. Z.; Lundqvist, B. I.; Langreth, D. C., Higher-accuracy van der Waals density functional. *Phys. Rev. B* **2010,** *82* (8), 4.
23. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008,** *120* (1-3), 215-241.
24. Goerigk, L., How Do DFT-DCP, DFT-NL, and DFT-D3 Compare for the Description of London-Dispersion Effects in Conformers and General Thermochemistry? *J. Chem. Theory Comput.* **2014,** *10* (3), 968-980.
25. Grimme, S., Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006,** *27* (15), 1787-1799.

26. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010,** *132* (15), 19.
27. Tkatchenko, A.; Scheffler, M., Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009,** *102* (7), 4.
28. Bludsky, O.; Rubes, M.; Soldan, P.; Nachtigall, P., Investigation of the benzene-dimer potential energy surface: DFT/CCSD(T) correction scheme. *J. Chem. Phys.* **2008,** *128* (11), 8.
29. Hermann, J.; Bludsky, O., A novel correction scheme for DFT: A combined vdW-DF/CCSD(T) approach. *J. Chem. Phys.* **2013,** *139* (3), 6.
30. Soldan, P.; Hutson, J. M., On the long-range and short-range behavior of potentials from reproducing kernel Hilbert space interpolation. *J. Chem. Phys.* **2000,** *112* (9), 4415-4416.
31. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78* (8), 1950-1958.
32. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J. Am. Chem. Soc.* **1996,** *118* (9), 2309-2309.
33. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65* (3), 712-725.
34. Bushuev, Y. G.; Sastre, G., Atomistic Simulation of Water Intrusion-Extrusion in ITQ-4 (IFR) and ZSM-22 (TON): The Role of Silanol Defects. *J. Phys. Chem. C* **2011,** *115* (44), 21942-21953.
35. Bushuev, Y. G.; Sastre, G., Atomistic Simulations of Structural Defects and Water Occluded in SSZ-74 Zeolite. *J. Phys. Chem. C* **2009,** *113* (25), 10877-10886.
36. Cygan, R. T.; Liang, J. J.; Kalinichev, A. G., Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field. *J. Phys. Chem. B* **2004,** *108* (4), 1255-1266.
37. Burger, S. K.; Lacasse, M.; Verstraelen, T.; Drewry, J.; Gunning, P.; Ayers, P. W., Automated Parametrization of AMBER Force Field Terms from Vibrational Analysis with a Focus on Functionalizing Dinuclear Zinc(II) Scaffolds. *J. Chem. Theory Comput.* **2012,** *8* (2), 554-562.
38. Grajciar, L.; Bludsky, O.; Roth, W. J.; Nachtigall, P., Theoretical investigation of layered zeolite frameworks: Interaction between IPC-1P layers derived from zeolite UTL. *Catal. Today* **2013,** *204*, 15-21.
39. Sanders, M.; Leslie, M.; Catlow, C., Interatomic potentials for SiO 2. *J. Chem. Soc. Chem. Comm.* **1984,** (19), 1271-1273.
40. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald - An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.
41. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - The RESP Model. *J. Phys. Chem.* **1993,** *97* (40), 10269-10280.
42. Lemkul, J. Tabulated Potentials. www.gromacs.org/Documentation/How-tos/Tabulated_Potentials.
43. Ryoo, R.; Joo, S. H.; Kruk, M.; Jaroniec, M., Ordered mesoporous carbons. *Adv. Mater.* **2001,** *13* (9), 677-681.
44. Lee, J.; Kim, J.; Hyeon, T., Recent progress in the synthesis of porous carbon materials. *Adv. Mater.* **2006,** *18* (16), 2073-2094.
45. Valtchev, V.; Majano, G.; Mintova, S.; Perez-Ramirez, J., Tailored crystalline microporous materials by post-synthesis modification. *Chem. Soc. Rev.* **2013,** *42* (1), 263-290.
46. Schaack, B. B.; Schrader, W.; Schuth, F., How are Heteroelements (Ga and Ge) Incorporated in Silicate Oligomers? *Chem.-Eur. J.* **2009,** *15* (24), 5920-5925.
47. Meier, W. M.; Olson, D. H.; Baerlocher, C., Atlas of zeolite structure types. *Zeolites* **1996,** *17* (1-2), 1-229.
48. Egeblad, K.; Christensen, C. H.; Kustova, M.; Christensen, C. H., Templating mesoporous zeolites. *Chem. Mater.* **2008,** *20* (3), 946-960.
49. Hartmann, M., Hierarchical zeolites: A proven strategy to combine shape selectivity with efficient mass transport. *Angew. Chem. Int. Edit.* **2004,** *43* (44), 5880-5882.
50. Corma, A.; Diaz-Cabanas, M. J.; Jiang, J.; Afeworki, M.; Dorset, D. L.; Soled, S. L.; Strohmaier, K. G., Extra-large pore zeolite (ITQ-40) with the lowest framework density containing double four- and double three-rings. *P. Natl. Acad. Sci. USA* **2010,** *107* (32), 13997-14002.

51. Shamzhy, M. V.; Shvets, O. V.; Opanasenko, M. V.; Yaremov, P. S.; Sarkisyan, L. G.; Chlubna, P.; Zukal, A.; Marthala, V. R.; Hartmann, M.; Cejka, J., Synthesis of isomorphously substituted extra-large pore UTL zeolites. *J. Mater. Chem.* 2012, *22* (31), 15793-15803.
52. Chlubna, P.; Roth, W. J.; Greer, H. F.; Zhou, W. Z.; Shvets, O.; Zukal, A.; Cejka, J.; Morris, R. E., 3D to 2D Routes to Ultrathin and Expanded Zeolitic Materials. *Chem. Mater.* 2013, *25* (4), 542-547.
53. Shvets, O. V.; Nachtigall, P.; Roth, W. J.; Cejka, J., UTL zeolite and the way beyond. *Micropor. Mesopor. Mat.* 2013, *182*, 229-238.
54. Roth, W. J.; Dorset, D. L., The role of symmetry in building up zeolite frameworks from layered zeolite precursors having ferrierite and CAS layers. *Struct. Chem.* 2010, *21* (2), 385-390.
55. Roth, W. J.; Nachtigall, P.; Morris, R. E.; Cejka, J., Two-Dimensional Zeolites: Current Status and Perspectives. *Chem. Rev.* 2014, *114* (9), 4807-4837.
56. Roth, W. J.; Nachtigall, P.; Morris, R. E.; Wheatley, P. S.; Seymour, V. R.; Ashbrook, S. E.; Chlubna, P.; Grajciar, L.; Polozij, M.; Zukal, A.; Shvets, O.; Cejka, J., A family of zeolites with controlled pore size prepared using a top-down method. *Nat. Chem.* 2013, *5* (7), 628-633.
57. Goumans, T. P. M.; Wander, A.; Brown, W. A.; Catlow, C. R. A., Structure and stability of the (001) alpha-quartz surface. *Phys. Chem. Chem. Phys.* 2007, *9* (17), 2146-2152.
58. Xu, H.; Jiang, J. G.; Yang, B. T.; Zhang, L.; He, M. Y.; Wu, P., Post-Synthesis Treatment gives Highly Stable Siliceous Zeolites through the Isomorphous Substitution of Silicon for Germanium in Germanosilicates. *Angew. Chem. Int. Edit.* 2014, *53* (5), 1355-1359.
59. Grimme, S., A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* 2014, *10* (10), 4497-4514.
60. Prampolini, G.; Livotto, P. R.; Cacelli, I., Accuracy of Quantum Mechanically Derived Force-Fields Parameterized from Dispersion-Corrected DFT Data: The Benzene Dimer as a Prototype for Aromatic Interactions. *J. Chem. Theory Comput.* 2015, *11* (11), 5182-5196.
61. Vanduyfhuys, L.; Vandenbrande, S.; Verstraelen, T.; Schmid, R.; Waroquier, M.; Van Speybroeck, V., QuickFF: A Program for a Quick and Easy Derivation of Force Fields for Metal-Organic Frameworks from Ab Initio Input. *J. Comput. Chem.* 2015, *36* (13), 1015-1027.
62. Eaton, J. W.; Bateman, D.; Hauberg, S. *GNU Octave version 3.8.2 manual: a high-level interactive language for numerical computations*, 3.8.2; 2015.
63. Grimme, S., Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* 2004, *25* (12), 1463-1473.
64. Hadisoeseno, W. P. Y. *Statistics package for GNU Octave - regress.m function*, 1.2.4; 2006.
65. Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. *MOLPRO, version 2010.1, a package of ab initio programs*, 2010.1.
66. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.
67. Dunning, T. H., Gaussian-basis sets for use in correlated molecular calculations .1. The atoms boron through neon and hydrogen. *J. Chem. Phys.* 1989, *90* (2), 1007-1023.
68. Boys, S. F.; Bernardi, F., The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* 1970, *19* (4), 553-566.

69. Werner, H. J.; Manby, F. R.; Knowles, P. J., Fast linear scaling second-order Moller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.* 2003, *118* (18), 8149-8160.
70. Kresse, G.; Hafner, J., Ab-initio molecular-dynamics simulation of the liquid-metal amorphous-semiconductor transition in germanium. *Phys. Rev. B* 1994, *49* (20), 14251-14269.
71. Blochl, P. E., Projector augmented-wave method. *Phys. Rev. B* 1994, *50* (24), 17953-17979.
72. Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R., GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 1995, *91* (1-3), 43-56.
73. *Matlab*, 8.3.0.532 (R2014a); The MathWorks Inc.
74. Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L., What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday Discuss.* 2009, *141*, 251-276.
75. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J., LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 1997, *18* (12), 1463-1472.
76. Hess, B., P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 2008, *4* (1), 116-122.
77. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* 1984, *81* (8), 3684-3690.
78. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J. Chem. Phys.* 2007, *126* (1), 7.
79. Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.; Joosten, R. P.; Vriend, G., A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* 2015, *43* (D1), D364-D368.
80. Gao, Q.; Xu, W. J.; Xu, Y.; Wu, D.; Sun, Y. H.; Deng, F.; Shen, W. L., Amino acid adsorption on mesoporous materials: Influence of types of amino acids, modification of mesoporous materials, and solution conditions. *J. Phys. Chem. B* 2008, *112* (7), 2261-2267.
81. Rimola, A.; Costa, D.; Sodupe, M.; Lambert, J. F.; Ugliengo, P., Silica Surface Features and Their Role in the Adsorption of Biomolecules: Computational Modeling and Experiments. *Chem. Rev.* 2013, *113* (6), 4216-4313.
82. Haran, G.; Thirumalai, D., Ribosome exit tunnel can entropically stabilize¿-helices. *P. Natl. Acad. Sci. USA* 2005, (52), 18956-18961.
83. Sorin, E. J.; Pande, V. S., Nanotube confinement denatures protein helices. *J. Am. Chem. Soc.* 2006, *128* (19), 6316-6317.

6 APPENDIX

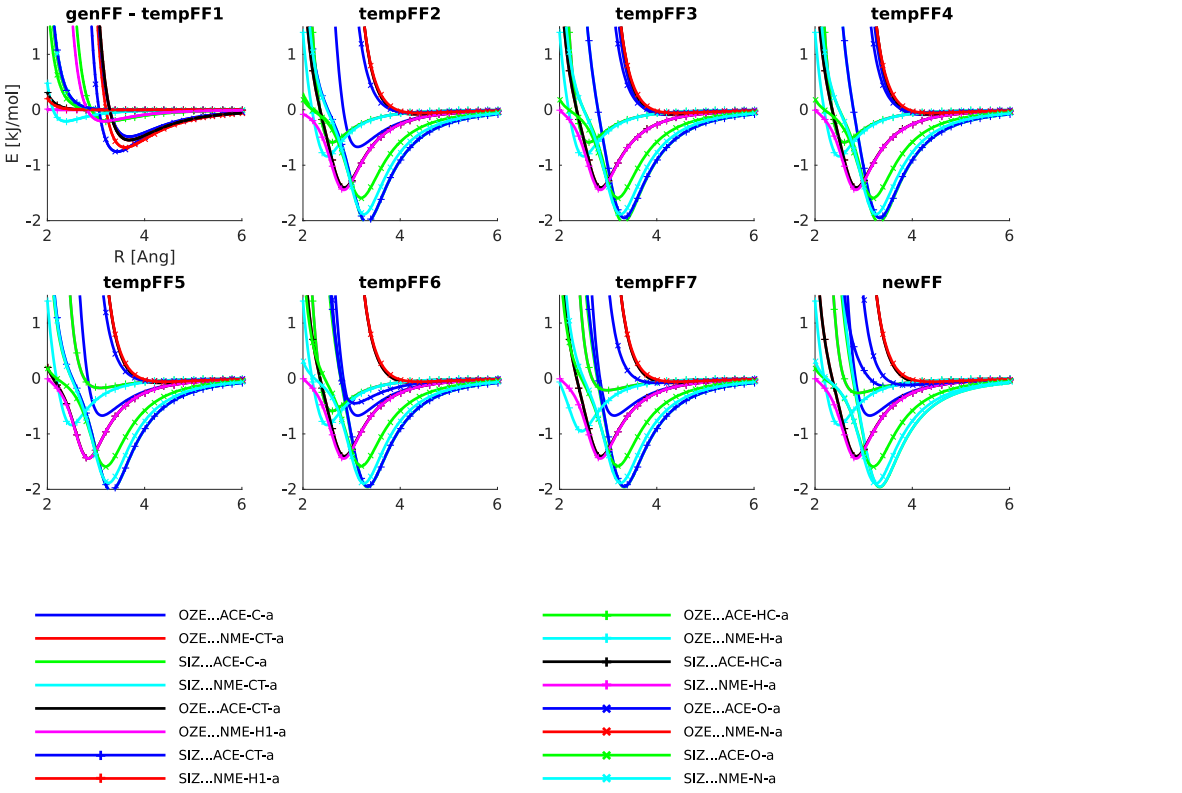


Figure A1. Changes in the potential during the parameterization procedure for AceNMe on relaxed α-quartz. The genFF is used as initial guess for the first round, but in the following rounds temporary force fields are used for the sampling. The final global fit results in the newFF. The potential is plotted without the genFF electrostatics.

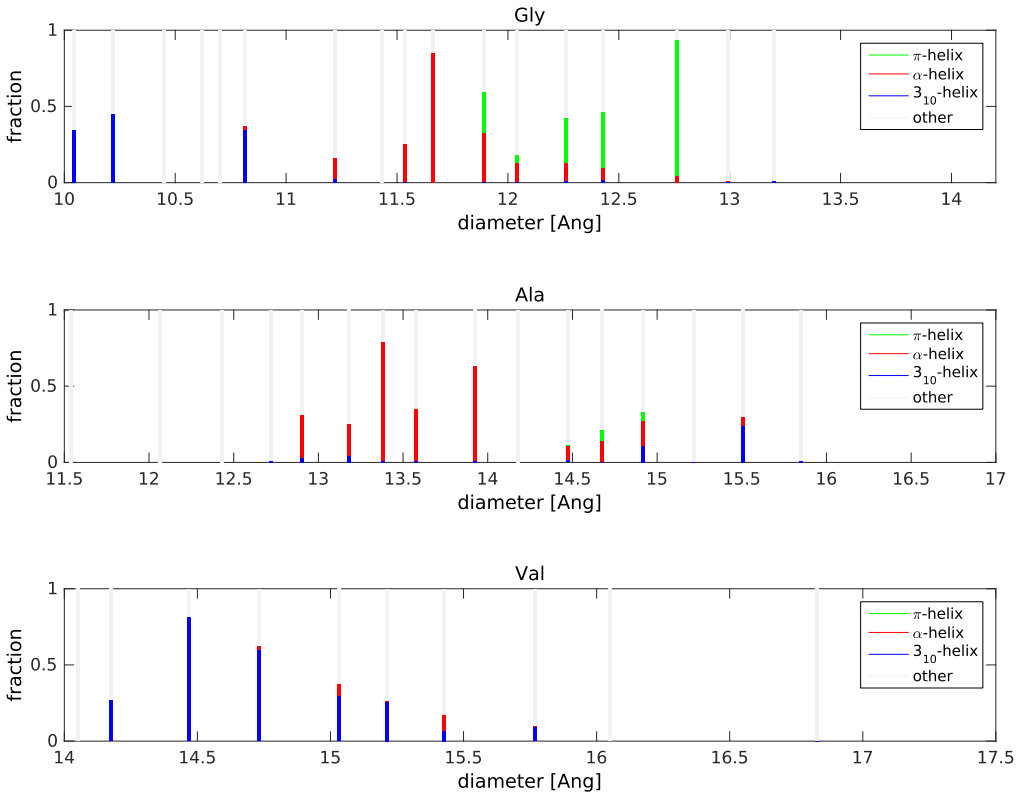


Figure A2. Formation of polypeptide helical structures in frozen carbon nanotubes as a function of nanotube diameter.

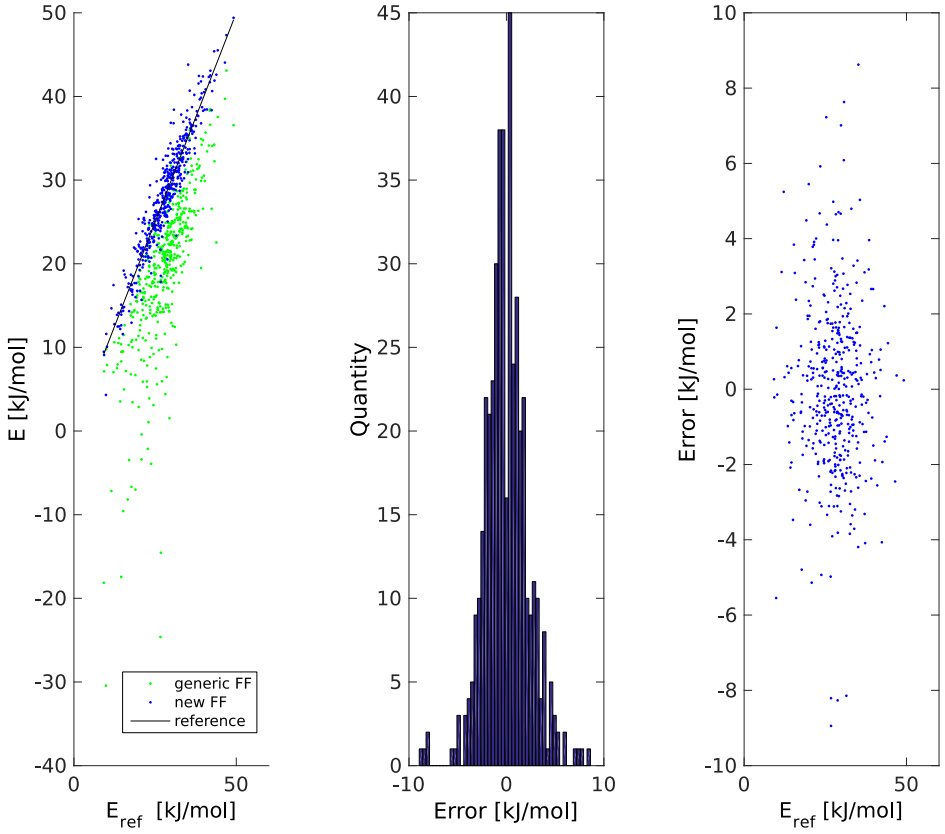


Figure A3. Fitting of AceNMe on relaxed α-quartz surface. Left: Comparison between binding energy of reference PBE-D2 method with newFF and genFF. Middle: Histogram of errors of newFF. Right: NewFF error as a function of reference binding energy.

Table A1. Performance of older newFF version, RMSD of binding energy in kJ/mol.

System	RMSD
AceNMe – graphene	1.0
AceGlyNMe – graphene	2.1
AceAlaNMe – graphene	2.1
AceValNMe – graphene	2.3
AcePheNMe – graphene	1.9

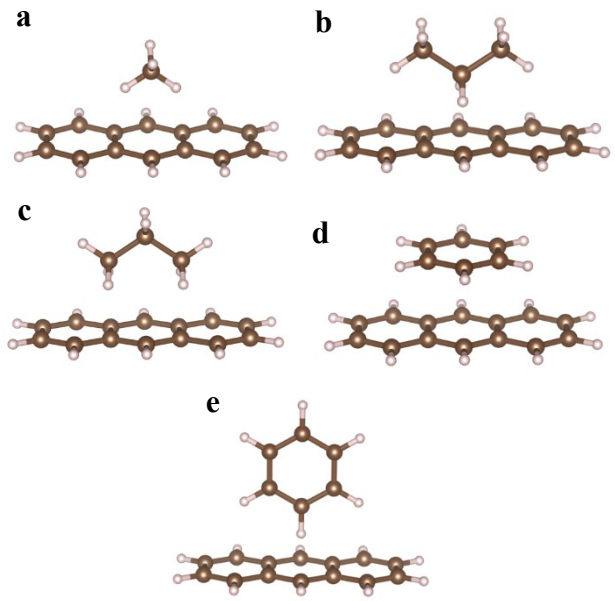


Figure A4 One dimensional potential energy curves calculated in the direction of the normal of the anthracene plane for the (a) methane, (b-c) propane and (d-e) benzene.

Table A2 The interaction energies in (kJ/mol) and equilibrium distances in (Å) of anthracene…molecule complexes.[a] The equilibrium distances are reported in parentheses.

Molecule	PBE-D2	PBE-D3bj	vdW-DF2	DFT-SAPT	MP2	CCSD(T)	Fig A4.
CH_4	-9.2 (3.45)	-9.1 (3.60)	-9.3 (3.60)	-8.8 (3.55)	-11.9 (3.43)	-7.5 (3.54)	a
C_3H_8	-14.0 (3.94)	-13.8 (4.07)	-14.4 (4.08)	-13.5 (4.01)	-18.6 (3.90)	-13.4 (4.02)	b
	-14.7 (3.84)	-15.0 (3.97)	-16.5 (3.97)	-14.8 (3.93)	-20.2 (3.81)	-14.7 (3.91)	c
C_6H_6	-16.1 (3.59)	-19.1 (3.65)	-21.7 (3.67)	-18.4 (3.65)	-36.5 (3.39)	-19.6 (3.59)	d
	-16.3 (4.79)	-16.0 (4.90)	-15.0 (4.94)	-15.2 (4.89)	-23.9 (4.69)	-16.6 (4.84)	e

[a] Equilibrium distances are defined as a center of mass separation between interacting molecules. The MP2 and CCSD(T) calculations are reported at the CBS limit. DFT and DFT-SAPT calculations were performed with AVQZ and AVTZ basis sets, respectively.