



Comparison of different approaches to derive classical bonded force-field parameters for a transition metal cofactor: a case study for non-heme iron site of ectoine synthase

Justyna Andrys¹ · Johann Heider² · Tomasz Borowski¹

Received: 1 March 2021 / Accepted: 9 June 2021 / Published online: 29 July 2021
© The Author(s) 2021

Abstract

Computational investigations into the structure and function of metalloenzymes with transition metal cofactors require proper preparation of the model, which requires obtaining reliable force field parameters for the cofactor. Here, we present a test case where several methods were used to derive amber force field parameters for a bonded model of the Fe(II) cofactor of ectoine synthase. Moreover, the spin of the ground state of the cofactor was probed by DFT and post-HF methods, which consistently indicated the quintet state is lowest in energy and well separated from triplet and singlet. The performance of the obtained force field parameter sets, derived for the quintet spin state, was scrutinized and compared taking into account metrics focused on geometric features of the models as well as their energetics. The main conclusion of this study is that Hessian-based methods yield parameters which represent the geometry around the metal ion, but poorly reproduce energy variance with geometrical changes. On the other hand, the energy-based method yields parameters accurately reproducing energy-structure relationships, but with bad performance in geometry optimization. Preliminary tests show that admixing geometrical criteria to energy-based methods may allow to derive parameters with acceptable performance for both energy and geometry.

Keywords Parametrization · Force field · Transition metals · Iron · Ectoine · Ectoine synthase

1 Introduction

Transition metal ions fall in between group 1 and group 2 metal cations, which form mainly ionic bonds, and the p-block elements of the periodic table, which form bonds of more covalent character. In the frame of classical force fields used to model biological macromolecules, the former group is described exclusively with non-bonded parameters (charge and Lennard–Jones vdW parameters), whereas elements from the p-block of the periodic table, e.g., C, N, O, P, S, Se, are treated with bonded parameters (explicit covalent bonds). Thus, it is, perhaps, not surprising that some

authors describe transition metal ions with the non-bonded model, whereas others with the bonded counterpart. Each of these approaches has its pros (as well as cons), e.g., the non-bonded model allows for dynamic change of coordination number, whereas the bonded model offers a good control over geometry of the first coordination sphere of the metals. Within the bonded model for transition metal ions one needs to determine force constants and reference values for metal–ligand bonds and all valence angles formed by the metal ion. Dihedral potential is usually not fitted and left to be zero. To derive these bond and angle parameters one resorts to quantum chemistry, as typically there is not enough available experimental data for the (often unique) composition of the ion first coordination shell. Here, we are presenting our efforts to find the best set of force field parameters to describe the geometry of the Fe(II) cofactor present in the active site of ectoine synthase (EctC). Experimental data strongly indicate that the Fe(II) ion plays a central role in catalytic activity of EctC [1, 2], thus we aim to obtain as precise description of the metal site as possible by testing three parameterization methods for bonded approach: the

✉ Tomasz Borowski
borowski@ikifp.edu.pl

¹ Jerzy Haber Institute of Catalysis and Surface Chemistry,
Polish Academy of Sciences, ul. Niezapominajek 8,
30-239 Kraków, Poland

² Philipps-Universität Marburg, Karl-von-Frisch Strasse 8,
35043 Marburg, Germany

projected Hessian method proposed by Seminario [3], Hessian-based methods developed and implemented within the ParmHess program by R. Wang et al. [4, 5] and the energy-based method implemented within the ParamFit program by Betz and Walker [6]. Over the years, thorough surveys of the metal site parameterization methods have been published. In 2011, Hu and Ryde analyzed five different approaches (both bonded and non-bonded) for MM parameters derivation while applying them to zinc metalloproteins for testing purposes [7]. They concluded, that before attempting the parameterization procedure, it is crucial to determine the future use of the parameters, as their performance will heavily depend on the nature of the metal-binding site, for example, they recommend using Norrby-Lijefors automated method for the catalytically important ions [7, 8]. In 2017, Li and Merz published a comprehensive review of the metal site parameterization methods [9]. Their goal was to cover the current state of the art in one paper in order to facilitate the decision-making over which approach to choose, without testing the performance of the methods in proteins. Our goal is somewhat similar to the work of Hu and Ryde, but in this case, we are focusing only on the bonded approaches, keeping in mind that the Fe(II) ion takes part in the enzymatic reaction and we need a good description of its coordination sphere for further work. Here, we try to find the method that will provide the best compromise between energetic and geometric accuracy. All of the methods tested here are discussed in the 2017 review [9], but here we apply them to perform parameterization and then we validate the performance of the resulting parameter sets.

The cupin-type protein EctC is one of four enzymes that are involved in the ectoine biosynthesis pathway, which is found mostly in bacterial but also archaeal species [2, 10]. Ectoine is a small molecule osmolyte and cell-protective agent preventing macromolecules from dehydration during osmotic stress [1, 11]. It is formed from N- γ -acetyl-L-2,4-diaminobutyric acid (N- γ -ADABA) by cyclo-condensation and water elimination catalyzed by EctC (Fig. 1). Recent work by Czech et al. reveals not only the presence of the metal-binding motifs in EctC amino acid sequence, but also points out that iron ion most likely plays an important role during enzymatic catalysis [1]. When compared to

the amino acid sequences identified as metal-coordinating motifs in other cupin-type proteins, the one found in the EctC sequence deviates from the consensus patterns in a unique way (Table 1). In the first motif, two usually conserved histidines are missing and only the metal-binding glutamate (Glu57) is conserved, while the second motif contains two metal-binding residues instead of one: tyrosine (Tyr84) replaces a canonical histidine and an additional histidine (His92) is present. To check the importance of these iron-binding residues for EctC enzymatic activity, they were individually replaced by alanine via site-directed mutagenesis, and the results showed a substantial reduction in enzymatic activity: A drop of 94.6%, 87.3% and 84.9% with respect to the activity of the WT form was observed for the E57A, H92A and Y84A mutants, respectively. Furthermore, the study indicated that the substrate-binding site is in close proximity to the iron and a direct N- γ -ADABA-iron interaction, via the carbonyl oxygen, was proposed. It is also interesting that no good quality crystals were obtained in the absence of iron, indicating a stabilizing effect of the metal on structural integrity [1].

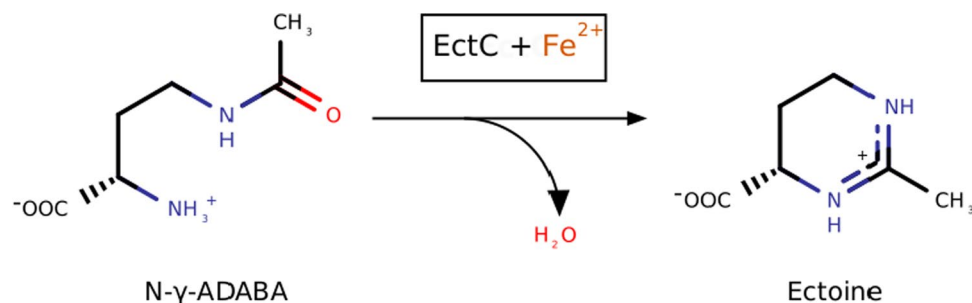
The crystal structure of *Paenibacillus lautus* ectoine synthase with substrate and Fe²⁺ ion bound (PDB code: 5ONN) was our starting point for the parametrization procedure. Although the overall resolution of the crystal structure is high (1.4 Å), the bond lengths found in the first coordination sphere of the iron are significantly longer than expected (2.8–2.9 vs. ~2.0 Å) [12]. In this structure occupation for Fe is only 0.38 and the B-factor for this atom is larger than for its nearest protein atoms (23.9 vs. 16.75–18.85 Å²), which might be the reason why this fragment of the structure is less reliable. Hence, the force field parameters, whose derivation is detailed beneath, will not only help us gain insights into

Table 1 Metal-coordinating consensus motifs established for cupin-type proteins vs EctC metal-binding motifs

Cupin consensus motifs	Variant EctC motifs
Motif 1 [G(X) ₅ HX H(X) _{3,4} E (X) ₆ G]	[G(X) ₅ WY(X) ₄ E (X) ₆ G]
Motif 2 [G(X) ₅ PXG(X) ₂ H (X) ₃ N]	[G(X) ₆ PG(X) ₂ Y (X) ₃ G(X) ₃ H]

Letters in bold indicate iron-binding amino acids [1]

Fig. 1 Simplified reaction scheme of cyclo-condensation of N- γ -ADABA to ectoine, catalyzed by EctC in the presence of Fe²⁺



the dynamic behavior of the EctC active site, but they will also aid in further refinement of the metal cofactor geometry, both of which are desirable, as the structure captured in this crystal is most likely not catalytically competent.

The rest of this article is organized as follows. First, models and quantum chemistry methods used to infer the spin of the ground state and to derive reference energy and geometry data are described, then methods used to derive force field bonded parameters are presented, followed by description of the MD protocol and metrics used to assess the performance of derived parameter sets. In the Results and discussion section, we first present quantum chemistry data for plausible spin states of the cofactor, and shortly present QM (ADMP) trajectories. Then, the major part of the manuscript is devoted to the derived parameter sets and their performance in reproducing stationary and dynamic geometries as well as energies. The report finishes with concise conclusions.

2 Computational models and methods

2.1 Models

Because of the unclear ionization state of Tyr84, which coordinates Fe^{2+} , two types of models were considered in this study; one with Tyr84 phenolic group ionized to a tyrosinate, hereafter called model 1 and model s1, and a second with a protonated Tyr84 phenolic group, hereafter called model 2 and model s2 (s stands for ‘small’) (see Fig. 2 and S1). Primary models, i.e., model 1 and model 2, were full optimized *in vacuo* with no geometrical constraints, whereas models

s1 and s2 were constructed from the optimized structures of model 1 and 2, respectively, by trimming the model, substituting terminal unsaturated atoms with H and manual adjustment of lengths for thus introduced H–X bonds.

2.2 Quantum chemistry methods

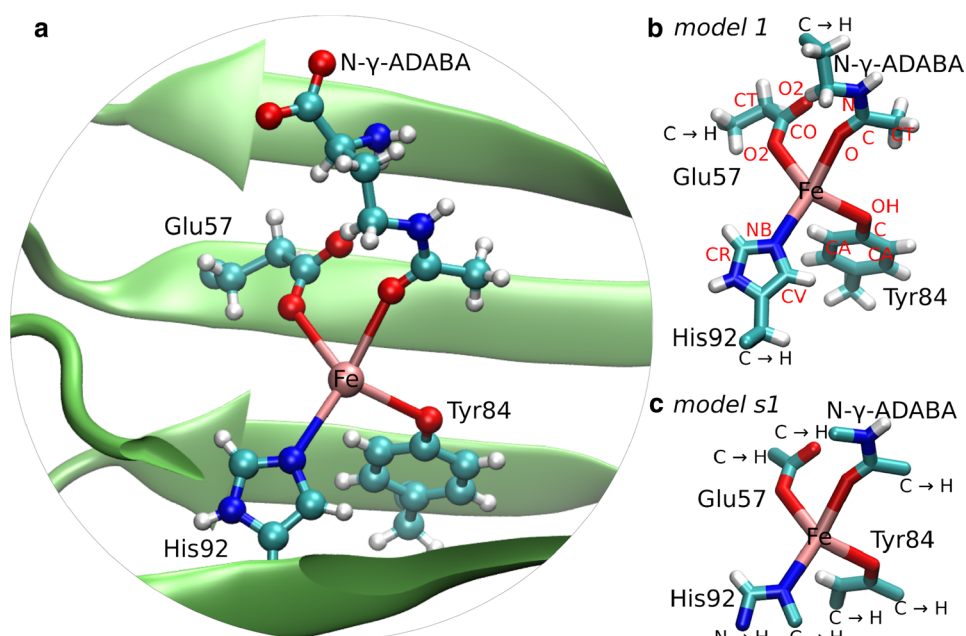
2.2.1 DFT

B3LYP-D3 [13, 14] combined with the Def2TZVP basis set [15] was used to obtain optimal geometries of the models and their electronic energy for three possible spin-states: singlet (low spin-state), triplet (intermediate spin-state) and quintet (high spin-state). As for model 1 and model 2, the singlet state was computed to lie very high in energy, 44.3 and 36.7 kcal/mol higher than the quintet ground state, respectively; therefore, the singlet state was not further considered. Single point energy computations were also done for the quintet and triplet states with the use of CAM-B3LYP-D3 [16] and MN15 [17] functionals combined with the Def2TZVP basis set. Computations were done with Gaussian 16 rev C.01 using default convergence criteria [18].

2.2.2 ADMP

Atom-centered Density Matrix Propagation (ADMP) method is a Car-Parinello method that allows for time-efficient calculations of ab initio molecular dynamics trajectories [19–22]. Here, we used the ADMP method at the UB3LYP-D3/Def2SVP level. The starting geometry was the optimal structure with iron in a quintet spin-state, obtained at the UB3LYP-D3/Def2SVP level, the initial nuclear kinetic energy was

Fig. 2 Fe^{2+} cofactor of ectoine synthase with bound N- γ -ADABA in the crystal structure (a) and its parts selected for the models used in this study: primary model 1 (b) and the minimal model s1 used for some post-HF calculations (c). Red letters show amber atom types of the key atoms; X \rightarrow H denotes that a given atom is replaced by hydrogen in the model. Model 2 and model s2 differ from model 1 and s1 only by the presence of a proton bound to the phenolic oxygen of Tyr84 (see Fig. S1)



set at the 0.43250 Hartree level, which is the oscillation energy obtained from the frequency calculations. 30 000 frames with default 0.1 fs time step were collected (3 ps), velocity scaling thermostat was used throughout the simulation (IOP(1/80) = 1000000), with temperature checked and scaled every 10 steps (IOP(1/81) = 10). The temperature of the thermostat was set at 300 K (IOP(1/82) = 300).

2.2.3 ONIOM

ONIOM(ROCCSD(T):UB3LYP-D3) computations were successfully performed only for model 1, as despite many attempts, the CCSD calculations did not converge for model 2 in the quintet state. The ONIOM model was partitioned into layers as shown in Fig. 2 i.e., the CCSD(T) calculations were done for model 1 s, whereas DFT calculations for model 1 and model 1 s. For CCSD(T), restricted open shell (RO) formalism was used and the basis set combined cc-pVTZ basis for Fe, N and O, and cc-pVDZ for C and H atoms. FreezeNobleGasCore option was used to correlate only the valence electrons. For the UB3LYP-D3 calculations, Def2TZVP basis set was used and the D3 correction with the Becke-Johnson damping.

2.2.4 DLPNO-UCCSD(T)

DLPNO-UCCSD(T) is an approximation to the canonical UCCSD(T) method that treats electronic correlations as local phenomena [23–26]. Thanks to its much more favorable scaling with the system size it could be applied directly to model 1 and model 2. The basis set used combined the cc-pwCVQZ basis for Fe and cc-pVTZ for all other atoms of the system. AutoAux option was used to generate an auxiliary basis set. As the basis used for Fe justifies it, 3 s and 3p electrons of Fe were included in the correlated calculations. The calculations were performed in two variants, first with ROHF orbitals, and second with UB3LYP orbitals. For all other settings, default values were used. The calculations were performed with ORCA 4.0.0 [27].

2.2.5 NEVPT2

NEVPT2 (*n*-electron valence state perturbation theory) is a multireference perturbation (second order) method build on top of a CASSCF wave function [28–31], which is very well suited for calculations of excitation energies [32]. Here, we have used its strongly contracted (SC) variant to calculate the quintet-triplet adiabatic energy difference using 6 electrons in 10 orbitals (3d + 4d) active space. The basis set used was the same as in the DLPNO-UCCSD(T) calculations, i.e., cc-pwCVQZ basis for Fe and cc-pVTZ for all other atoms. All electrons were correlated at the NEVPT2 stage. The calculations were performed with ORCA 4.0.0 [27].

2.3 Methods used to fit bonded force field parameters for the cofactor

2.3.1 Seminario

The Seminario method allows for direct evaluation of bond and angle force constants from the Hessian matrix and it can be viewed as a projection method from cartesian to internal coordinates [3]. The Hessian was computed at the B3LYP-D3/Def2TZVP level using Gaussian 16, whereas force constants of bonds and angles involving Fe were obtained from it with the use of the XYZViewer program, in which the Seminario method is implemented. Parameter sets derived with this method are labeled “Sem” throughout the manuscript.

2.3.2 ParmHess

The Hessian computed with a quantum chemistry method includes all components of interatomic interactions (bonded, vdW, electrostatic), whereas when deriving force field parameters, one tries to determine individual terms separately. Hence, if one uses within the amber force field bond and angle force constants derived with the Seminario method together with atomic partial charges and atomic vdW (Lennard–Jones) parameters, a potential problem of double-counting of non-bonded interactions arises. To address this issue Wang et. al. developed three Hessian-based methods, whereby before the fitting of MM-based Hessian to the QM counterpart, the non-bonded contributions to the MM Hessian are subtracted from the QM one. The free methods, implemented in the ParmHess program, differ in the details of the fitting procedure and are called: partial Hessian fitting (PHF), full Hessian fitting (FHF) and internal Hessian fitting (IHF) [4, 5]. Here, we have used ParmHess to derive force constants for bonds and angles involving Fe. The same QM Hessian as used with the Seminario methods was employed here (B3LYP-D3/Def2TZVP). For our system, only the PHF method gave satisfactory results, i.e., no negative force constant, hence only these parameters are presented and discussed beneath.

2.3.3 Katachi amendment

The bonded force constants derived from any Hessian-based method, be it Seminario or PHF/FHF/IHF, are usually combined with reference bond lengths and angle values taken directly from the QM-optimized geometry. For systems with soft bonds where non-bonded interactions (mainly electrostatic) are significant, this procedure often leads to significant deviations between QM- and MM-optimized geometries. This problem can be solved with the Katachi amendment procedure proposed by Wang et al., whereby

bond and angle reference values (r_{eq} , Θ_{eq}) are iteratively changed until the MM-optimized geometry reproduces bond and angle values from the QM structure [5]. The Katachi procedure was used here with 100 iterations limit and it was applied to Seminario, PHF and ParamFit derived sets of parameters. Thus, obtained parameter sets are labeled with the “_K” suffix.

2.3.4 Paramfit

Paramfit program, which is part of the AmberTools package [33], fits the bonded force field parameters using a set of geometries for which it tries to minimize differences between MM and QM energies by using the least square method [6]. To obtain a representative set of geometries that would allow reliable fitting of bond and angle parameters involving Fe, we used the following procedure. First, around 150 evenly spaced snapshots from the ADMP trajectory (151 for model 1 and 149 for model 2) were selected and subjected to constrained minimization at the B3LYP-D3/def2SVP level with all bond lengths and valence angles involving Fe constrained to the values as in the given snapshot. The purpose of this minimization is to reduce the noise in the data due to energy changes caused by variation of internal coordinates for which the parameters are not going to be fitted. For the minimized structures, single point QM energy values were computed at the B3LYP-D3/def2TZVP level, and these values were used for fitting. QM input preparation and parameter fitting were done with the paramfit program. During the fitting procedure, one first needs to calculate the K constant, which is an intrinsic discrepancy between QM and MM energies: $E_{\text{MM}} - E_{\text{QM}} + K = 0$. Then, the K value is used as one of input options for actual parameters fitting. Parameter sets derived with this method are labeled “Param” throughout the manuscript. Sets with a few manually adjusted parameters are labeled “Param_t” (tuned) throughout the manuscript.

Paramfit was also used, with the same QM reference data points, to test how well different parameter sets reproduce QM energy variations with the geometry changes around the Fe(II) ion.

2.3.5 Tuning paramfit parameters

In the case of model 1, param force constants for 2 bonds and 3 angles with MD averages most deviating from the reference ADMP values were iteratively changed and newly obtained MD averages were used to test if the modified set gave better or worse description of dynamics of the cofactor. Five iterations gave a parameter set, labeled as param_t, which yielded satisfactory results. In the case of the model 2, param_t parameters were obtained by substituting parameters (both K and Θ_{eq}) describing only one angle, namely

O2–Fe–O, which showed the biggest deviation from the ADMP averages with those taken from the first seminario_katachi run. This adjustment improved rather poor initial paramfit-based geometries, both minimized and MD averages.

2.3.6 Non-bonded force field parameters

Lennard–Jones parameters used for Fe^{2+} ($R^* = 1.456 \text{ \AA}$, $\epsilon = 0.013 \text{ kcal/mol}$) were taken from the UFF force field [34]. Atomic partial charges were fitted with the RESP program, which is part of the AmberTools package. For model 1 and model 2, electrostatic potential was calculated at the B3LYP-D3/def2TZVP level and it was used to fit RESP atomic charges, which were used in MM optimization and MD simulations in vacuum for model 1 and model 2. To derive atomic charges for whole residues coordinated to Fe^{2+} , another, larger model, which includes parts of the protein backbone, was optimized at the B3LYP-D3/def2SVP level with constrained coordinates of the backbone heavy atoms (see Fig. S2). For this large model electrostatic potential was then computed at the UHF/6-31G(d) level, which is consistent with the ff14SB amber protein force field, and it was used for RESP fitting. In the latter charges on CA, HA, N and C atoms of the backbones of Tyr, His and Glu and all atoms of NGA that were not part of model 1 or model 2 were fitted, whereas charges on all other atoms were kept fixed to their values taken either from the previous RESP fit for model 1 or model 2 (core region) or ff14SB force field (for backbone atoms of Tyr, His and Glu). Thus fitted charges for model 1 sum to the following residue values: Fe (0.9719), His (0.0788), Glu (−0.6368), Tyr(−0.6185), NGA (−0.7955), whereas for model 2 their totals are: Fe (0.9719), His (0.2195), Glu (−0.5391), Tyr(0.094), NGA (−0.7955).

2.4 Classical MD simulations

The classical MD simulations were performed for model 1 and model 2 in vacuum and also for the EctC-NGA complex in explicit water under periodic boundary conditions. In vacuum, simulations were done with sander program, which is part of the AmberTools package [33], using Langevin dynamics, $T = 300 \text{ K}$, time step of 0.5 fs, total simulation time was 25 ps and snapshots were saved every 20 steps (2500 snapshot in total). To prepare the EctC system for simulations, the protein was placed in a cuboid filled with TIP3P water with the cuboid faces at least 10 Å from the protein surface in each direction. To neutralize the charge of the system and also mimic the ionic strength of physiological conditions ($I = 0.15 \text{ M}$), 33 Na^+ and 23 Cl^- ions were added to the system. The system was subsequently minimized and then heated from 0 to 300 K during

50 ps NV Langevin dynamics. The next step was 0.5 ns NPT ($T=300$ K, $p=1$ bar) density equilibration dynamics. During these two initial MD runs, coordinates of the protein were restrained with 1 kcal/mol \AA^2 force constant to their values in the minimized structure. Subsequently, 100 ns NPT ($T=300$ K, $p=1$ bar) dynamics was simulated with no restraints. The time step was 2 fs and the SHAKE algorithm was used to constraint bond lengths and valence angles involving hydrogen atoms. Snapshots were saved every 5000 steps (every 10 ps; 10 000 snapshots in total). Simulations were done with the use of sander and parmed programs [33].

2.5 Metrics used to validate FF parameter sets

In order to perform quality assessment of a given set of tested parameters, several different metrics were used.

- (1) Discrepancy between bond lengths [\AA] and valence angles [$^\circ$] involving Fe between MM and DFT optimized geometries. Furthermore, mean absolute error and max signed error were calculated separately for bonds and angles.
- (2) Discrepancy between mean values of bond lengths [\AA] and angles [$^\circ$] that involve iron ion between the ADMP and MM MD trajectories. Furthermore, mean absolute error and max signed error were calculated separately for bonds and angles.
- (3) Dissimilarity index (D_{QF}) computed for bond or angle histograms, which is based on a quadratic form (QF) [35]. Before computing the value of D_{QF} , the histograms were normalized to 1 and one of them was shifted, so that the mean values of the two compared histograms are the same. This shift guarantees that the D_{QF} measures dissimilarity between the shape of the two histograms. The value of D_{QF} for a pair of histograms \mathbf{h} and \mathbf{f} , where \mathbf{h} and \mathbf{f} are vectors with bin counts, was calculated using the formula: $(D_{\text{QF}})^2 = (\mathbf{h}-\mathbf{f})^T \mathbf{A} (\mathbf{h}-\mathbf{f})$. The similarity matrix elements were computed as: $A_{ij} = 1 - |i-j|/d_{\text{max}}$, where d_{max} is the maximum distance (in number of bins) between bins of the two compared histograms. For bond lengths, the bin width was 0.02 \AA , whereas for angles the bin width was 1° . Procedures to compute D_{QF} for bond and angle histograms were implemented as Octave scripts (see Code availability) [36].
- (4) Correlation between QM and MM energy values calculated for the set of ~ 150 relaxed geometries generated from the ADMP trajectory as described above. R^2 coefficient was computed with paramfit for all sets of tested FF parameters.

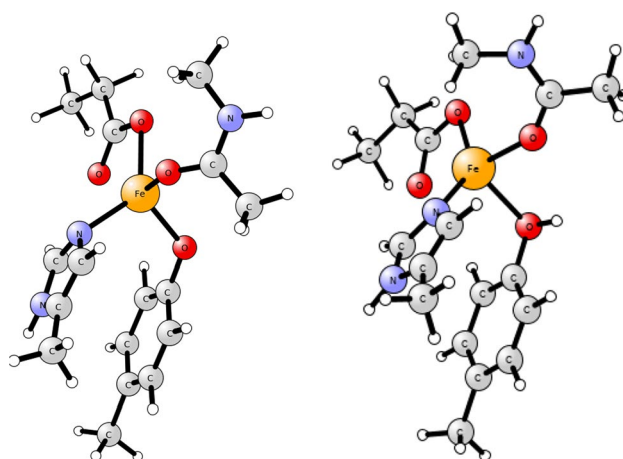


Fig. 3 Optimized (B3LYP-D3/def2-TZVP) structure of model 1 (left) and model 2 (right) in the quintet ground state

Table 2 Fe–X bond lengths [\AA] for the Fe(II) cofactor

Atom pair	Model 1 ($S=2$)	Model 2 ($S=2$)	PDB: 5ONN
Fe–O2 (Glu57)	2.07	2.00	2.93
Fe–OH (Tyr84)	1.92	2.16	2.75
Fe–NB (His92)	2.11	2.06	2.82
Fe–O (N- γ -ADABA)	2.18	2.11	2.78
Fe–O2 (Glu57)	2.50	2.18	4.10

3 Results and discussion

3.1 Geometry of the cofactor

Overall geometry of the cofactor optimized in vacuum at the DFT level (in the quintet ground state) resembles the geometry from the crystal structure (PDB: 5ONN), the coordination geometry remains tetrahedral (cf. Figures 2 and 3), but the bond lengths are significantly shorter (see Table 2).

3.2 Spin state energetics probed with DFT and post-HF methods

In order to investigate the spin states energy ladder of the cofactor, we have optimized model 1 and model 2 in three spin states: quintet ($S=2$), triplet ($S=1$) and closed-shell singlet ($S=0$). Since the singlet state was computed to lie very high in energy relative to the quintet, i.e., 44.28 and 36.66 kcal/mol for model 1 and 2, respectively, only quintet and triplet spin states were considered further. Using the B3LYP-D3/def2TZVP optimized geometries, single

Table 3 Relative energies [kcal/mol] computed for lowest lying spin-states of the Fe(II) cofactor

Method	Model 1		Model 2	
	S=2	S=1	S=2	S=1
B3LYP-D3	0.00	18.42	0.00	18.27
CAM-B3LYP-D3	0.00	19.26	0.00	19.16
MN15	0.00	19.04	0.00	14.41
ONIOM(CCSD(T), B3LYP-D3)	0.00	31.94	0.00	n.d
DLPNO-CCSD(T) (based on ROHF)	0.00	35.69	0.00	29.62
DLPNO-CCSD(T) (based on UB3LYP)	0.00	34.39	0.00	31.26
NEVPT2	0.00	35.90	0.00	28.78

point energy values were computed for quintet and triplet with a range of DFT and post-HF methods (Table 3). All of the methods used predicted the quintet to be the ground state and the quintet—triplet adiabatic energy difference ranged from 14.4 to 35.9 kcal/mol. This energy difference is sufficiently large to justify focusing only on the quintet state when deriving FF parameters for the cofactor. Interestingly, DFT methods consistently give a smaller energy gap compared to post-HF methods. Moreover, protonation of the tyrosine ligand lowers the gap by 3–7 kcal/mol according to post-HF methods and the MN15 functional, whereas B3LYP-D3 and CAM-B3LYP-D3 predict much smaller decrease, by 0.1–0.2 kcal/mol.

3.3 Dynamics of the cofactor probed with the ADMP method

The 300 K ADMP trajectories obtained for model 1 and 2 provide valuable insights into the dynamics and plasticity of the cofactor (for movies of these trajectories, see Supporting Information). Analysis of ADMP trajectories reveals that both bidentate and monodentate binding modes of Glu57 are present throughout the simulated time span. Classifying snapshots with two Fe–O bonds < 2.2 Å as bidentate and snapshots with at least one Fe–O bond > 2.4 Å as monodentate, for model 1 51% of frames were assigned to the monodentate category while only 9% to the bidentate one. For model 2 both values are 24%. Therefore, we can conclude that for model 1 bidentate mode is present in a small minority of structures, whereas for model 2, the situation is more even. Since the Fe-liganding oxygen atom of Glu57 exchanged several times with its neighboring carboxylic oxygen during the simulated time span, atomic labels in the obtained ADMP trajectories were re-ordered to give the same label (and number) to the oxygen atom of Glu57 that is closest to Fe throughout the whole trajectory. This relabeling enabled straightforward preparation of geometries for MM calculations (where explicit bonds between bonded atoms need to be specified) and also for direct comparison

of histograms for bonds and angles generated for ADMP and MM trajectories. Histograms were generated for all bonds and angles involving Fe, and they are shown in Figs. 4, 5 and 6 for model 1 (for histograms for model 2, see Fig. S4, S5 and S6). Average bond lengths and angles are gathered for model 1 in Table 4, where they are compared to averages calculated for MM MD trajectories (for data for model 2 see Tab. S3).

Analysis of bond-length histograms (Fig. 4) reveals that they are notably narrower for Fe–NB (His) and Fe–OH (Tyr) compared to those for Fe–O2 (Glu) and Fe–O (N- γ -ADABA), which suggests stronger bonds for the former group. Indeed, this finds confirmation in values of bond force constants (vide infra). The histogram for the ‘non-bonded’ second oxygen atom of Glu (Fig. 5) has a very asymmetric shape, which reflects multiple stable positions this atom can reach. Histograms for angles (Fig. 6) are all very broad suggesting low force constants; those for Fe–X–Y angles seem on average more symmetric than those for X–Fe–Y angles. The shape of the histograms for (NB)NE2–Fe–O2 and OH–Fe–NE2(NB) suggests that these angles might have two preferred optimal ranges.

3.4 Amber force field parameters and their performance in vacuum

Force field parameters for bonds and angles involving Fe, derived with each method tested here, are gathered in Table 5 and they are also presented graphically on Figs. S7, S8 and S11, S13 (for data for model 2, see Tab. S1 and Fig. S7, S8, S12, S14). A quick survey of the values presented in Table 5 reveals that both force constants and reference values can differ very significantly between the FF sets.

These bonded parameters for the Fe²⁺ ion and its surrounding were combined with RESP atomic charges and standard ff14SB amber force field parameters (for bonded and non-bonded terms involving other atoms) and applied to model 1 and model 2. Molecular mechanics geometry minimization was carried out with Gaussian 16, whereas a short (6 ps) $T=300$ K MD simulation in vacuum was done with the sander program. Equilibrium bond lengths and valence angle values are gathered in Table 6 (for data for model 2, see Tab. S2), whereas average values from trajectories are reported in Tables 4, 7 (Tab. S3 for model 2).

Concerning stationary geometries, Sem_K and PHF_K produced the best geometries for model 1 and model 2, respectively, as measured by RMSD computed for the 1st shell of Fe ion. Consistently, these parameter sets also gave lowest values of errors for bond lengths and valence angles involving Fe. Analysis of the error values shows that Sem_K and PHF_K perform rather similarly and in both cases application of the Katachi amendment brings about a very significant improvement of the stationary geometry. On the

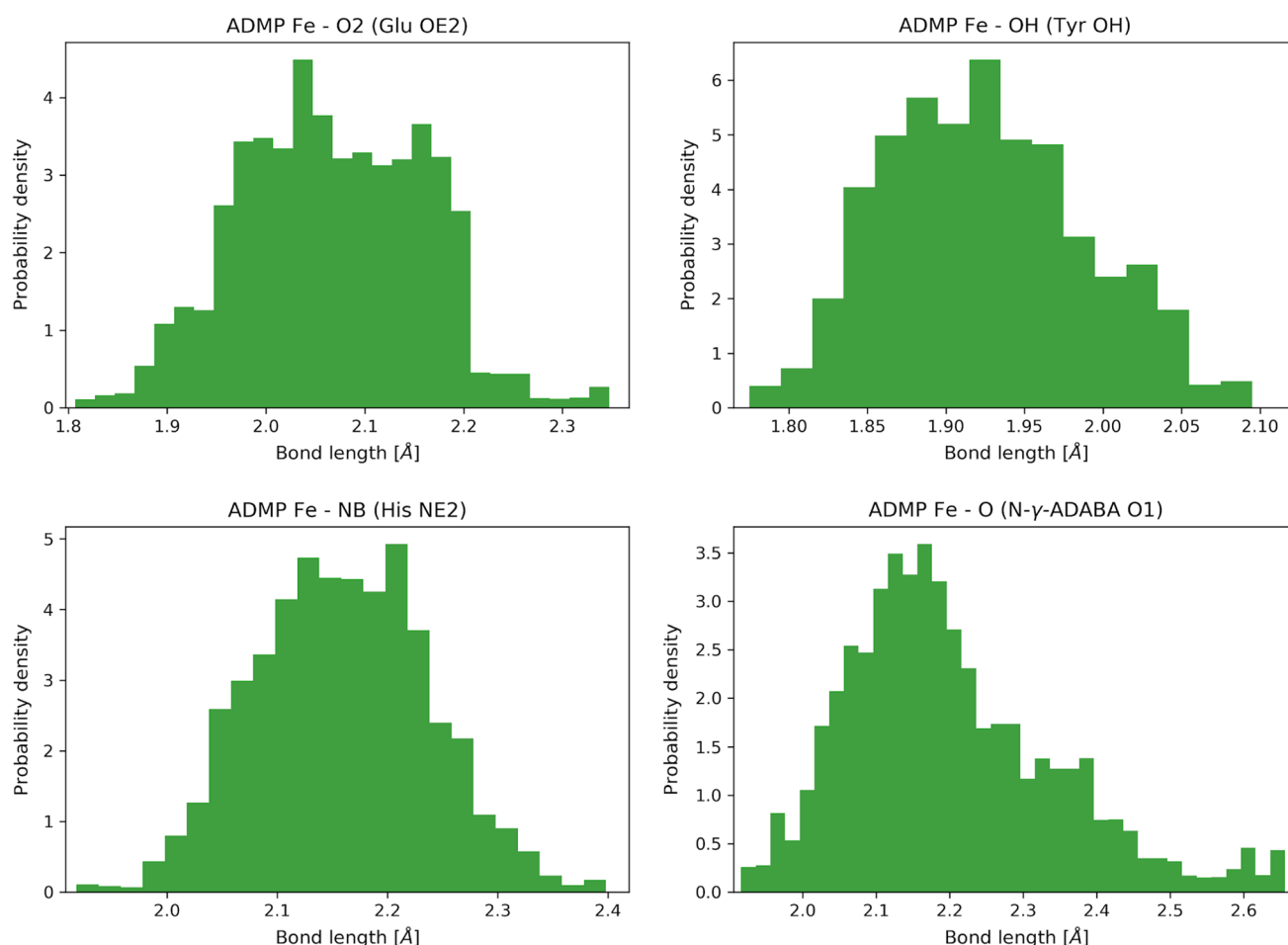


Fig. 4 Histograms for Fe–X bonds computed for ADMP trajectory for model 1

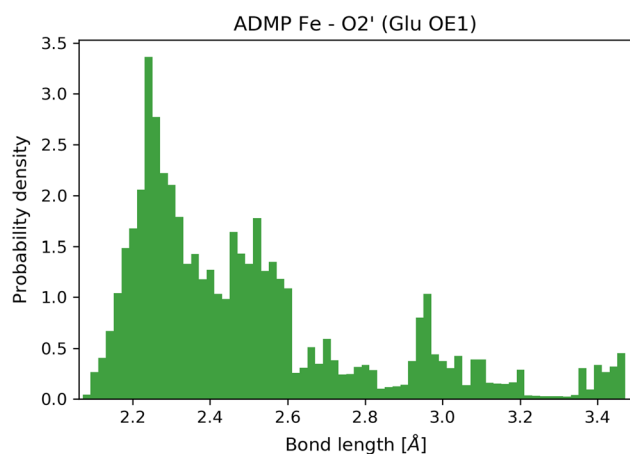


Fig. 5 Histogram for Fe – O2'(OE1) distances computed for ADMP trajectory for model 1

other hand, application of the Katachi method to the Param set of parameters improves stationary geometry either only slightly (model 1) or not at all (model 2). The original Param

sets of parameters yielded a rather poor geometry, i.e., 1st shell RMSD of 0.556 and 0.985 Å, bond MaxE of -0.240 and -0.209 Å and angle MaxE of -57.60° and -69.77° , for model 1 and model 2, respectively. However, manual adjustment of a few parameters for bonds and/or angles most deviating from the reference structure, gave us the Param_t sets that yielded an acceptable (in our subjective view) stationary geometry.

Focusing on the average bond and angle values sampled during 300 K MD simulations (Table 4 and S3, as well Fig. S10 and S17, S18), performance of Sem_K, PHF_K and Param_t sets is very similar, whereas an analysis of quadratic form distance (D_{QF}) between bond and angle histograms (see Tab. S4, S5) suggests that Param_t sets are marginally better than the other sets (superposition of ADMP and Sem_K/Param_t original histograms is shown in Fig. S32–S41).

Finally, focusing on how well a given parameter set can reproduce QM energy values for a set of ca. 150 geometries, one can infer already from the R^2 values reported in Table 7 and S3 that Param and Param_t sets are significantly better performing than any Hessian-based set. Analysis of correlation

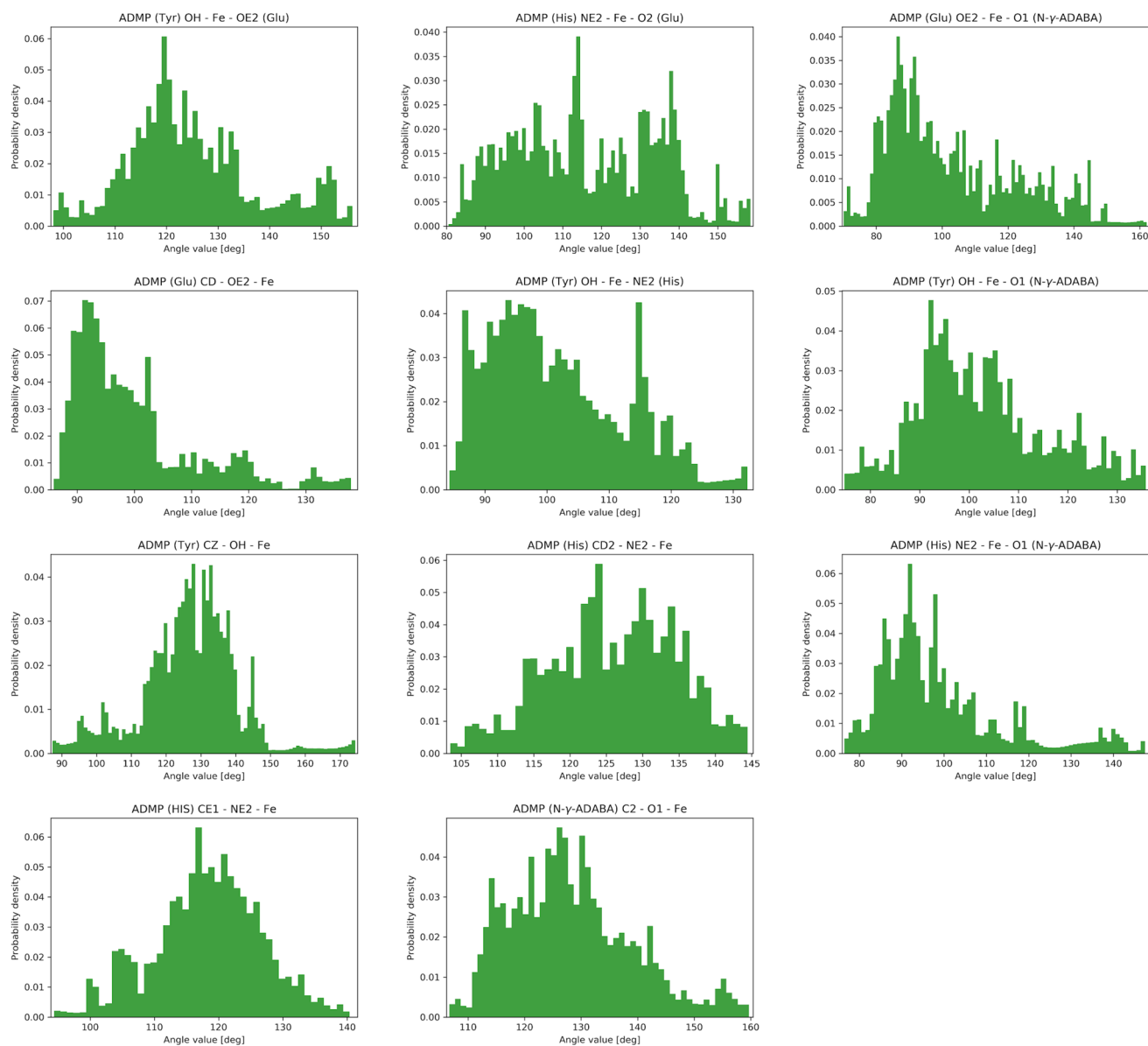


Fig. 6 Histograms for Fe–X–Y and X–Fe–Y valence angles computed for ADMP trajectory for model 1

plots and energy residues, presented in Fig. S21–S27, clearly shows that only the Param and Param_t sets give good correspondence between QM and MM energies with energy residues of reasonable magnitudes (up to 10–15 kcal/mol). In contrast, the MM–QM energy differences obtained from the other parameter sets can reach values in the range 60–150 kcal/mol, which means that if one uses these sets for MD simulations, some regions of configurational space will be much less frequently probed than they should.

3.5 Testing the parameters for the cofactor within the protein

Four test simulations were also performed for the whole EctC protein complexed with the substrate. Both protonation states of Tyr84 were considered, which corresponds to model 1 and model 2, and for each of them two parameter sets were employed: Sem_K and Param_t. Apart from Tyr84, the rest of tyrosines were in charge-neutral,

Table 4 Averaged bond length [Å] and valence angle [°] for model 1 derived from MD simulations in vacuum at 300 K performed at a specified level of theory

Bond / angle avg.	ADMP	Sem	Sem_K	PHF	PHF_K	Param	Param_K	Param_t
Fe-O2	2.067	1.938	2.083	1.917	2.073	1.845	2.193	1.994
Fe-OH	1.925	1.870	1.921	1.871	1.914	1.982	2.009	1.968
Fe-NB	2.159	2.128	2.129	2.127	2.125	2.162	2.163	2.154
Fe-O	2.202	1.842	2.311	1.885	2.252	2.079	2.415	2.137
Fe---O2 (not bonded)	2.500	2.145	2.377	2.148	2.343	2.374	2.438	2.588
MAE	-	0.186	0.060	0.181	0.050	0.106	0.098	0.050
MaxE	-	-0.360	-0.123	-0.352	-0.157	-0.222	0.213	0.090
O2-Fe-OH	124.90	128.30	124.20	122.60	113.60	133.40	130.30	129.60
O2-Fe-NB	115.90	129.80	128.10	133.60	139.50	115.60	130.70	96.55
O2-Fe-O	103.00	92.40	92.88	91.55	95.83	63.26	63.21	110.00
OH-Fe-NB	101.70	98.65	95.95	97.26	88.87	98.06	97.18	95.27
O-Fe-NB	99.15	79.69	99.96	80.29	104.60	74.54	83.03	96.87
OH-Fe-O	102.40	81.57	103.30	74.73	103.30	101.80	133.40	110.00
CO-O2-Fe	99.96	89.04	95.35	89.93	94.95	99.94	96.35	103.60
C-OH-Fe	126.70	115.80	123.10	105.20	123.70	72.36	142.80	126.90
CV-NB-Fe	125.90	124.90	124.70	127.00	126.60	130.70	126.20	129.10
CR-NB-Fe	118.30	112.80	117.90	111.30	118.80	109.30	120.80	111.80
C-O-Fe	128.40	118.40	120.00	119.50	123.90	125.20	133.40	118.40
MAE	-	9.96	4.43	11.90	6.81	13.50	12.65	6.45
MaxE	-	-20.80	12.20	-27.70	23.60	-54.30	-39.79	-19.40

MAE—mean absolute error, MaxE—maximum signed error; errors given separately for bonds and angles; all errors are calculated with respect to the ADMP average values. R_E^2 —correlation coefficient between energies computed at the DFT level and a given FF

whereas all lysines and arginines were in their cationic forms. Histidine residues were neutral, with protonated δ -nitrogen (His55, His89, His92) or ϵ -nitrogen (His5, His116, His136), while acidic amino acids (Glu, Asp) were all ionized. The protonation states of all amino acid residues were based on the results of the propKa3.1 program for pH 7.0 [37, 38]. We generated 100 ns-long

trajectories with typical settings for a protein in explicit solvent MD run (e.g., 2 fs time step, SHAKE algorithm used for hydrogen atoms) to serve as tests if these parameter sets cause any instabilities during the simulations or not. All four runs went smoothly, RMSD calculated for the whole main chain reached plateaus at around 3.9, and 4.8 Å for model 1 with Param_t and Sem_K sets,

Table 5 Derived amber force field bonded parameters for model 1

Bond / angle	Sem		Sem_K	PHF		PHF_K	Param		Param_K	Param_t
	r_{eq}/θ_{eq}	K	r_{eq}/θ_{eq}	r_{eq}/θ_{eq}	K	r_{eq}/θ_{eq}	r_{eq}/θ_{eq}	K	r_{eq}/θ_{eq}	K
Fe-O2	2.071	52.8	2.242	2.071	46.2	2.264	2.341	16.3	2.730	24
Fe-OH	1.923	105.0	1.968	1.923	104.7	1.962	2.044	71.3	2.078	71.3
Fe-NB	2.109	52.5	2.115	2.109	51.8	2.114	2.156	75.7	2.167	75.7
Fe-O	2.182	10.6	2.907	2.182	11.6	2.804	2.410	16.1	2.718	24
O2-Fe-OH	130.94	16.7	129.60	130.94	9.6	131.19	139.57	13.1	154.05	13.1
O2-Fe-NB	128.01	23.3	122.96	128.01	5.6	119.37	123.16	9.3	158.58	9.3
O2-Fe-O	93.46	27.3	102.71	93.46	23.8	105.66	123.72	7.4	109.05	12
OH-Fe-NB	98.42	14.4	99.38	98.42	7.1	102.47	107.17	21.0	117.08	21.0
O-Fe-NB	95.50	9.4	111.81	95.50	11.1	112.75	166.47	2.0	172.91	3.0
OH-Fe-O	97.31	19.3	117.29	97.31	14.5	120.58	120.72	20.1	143.79	20.1
CO-O2-Fe	94.83	102.3	102.12	94.83	126.1	100.81	136.58	18.1	141.37	18.1
C-OH-Fe	121.02	38.9	127.22	121.02	15.9	133.64	145.37	4.5	177.69	9.0
CV-NB-Fe	123.26	50.3	123.87	123.26	16.7	124.37	132.60	3.4	144.88	3.4
CR-NB-Fe	115.16	43.8	119.34	115.16	21.3	121.96	126.40	3.5	168.13	3.5
C-O-Fe	118.18	51.0	125.16	118.18	12.1	145.20	138.13	14.7	143.80	14.7

Bond length reference values in Å, bond angle reference values in degrees, bond force constants in kcal mol⁻¹ Å⁻², angle force constants in kcal mol⁻¹ rad⁻²

respectively, and 4.4, 4.7 Å for model 2 with Param_t and Sem_K sets. When the mobile and unstructured C-terminal fragment (aa: 126–138) was excluded, the RMSD values reached plateau at: 1.9 and 2.5 Å for model 1 with Param_t and Sem_K sets, respectively, 2.3 and 2.5 Å for model 2 with Param_t and Sem_K sets, respectively. Detailed results of MD simulations will be published elsewhere, but here we want to mention that the choice of the parameter set representing the metal cofactor seems to have non-negligible impact on the conformational freedom of the N-γ-ADABA substrate. More specifically, we observed how many snapshots of MD simulations yielded distances of less than 3.5 Å between the amino nitrogen and carbonyl carbon of N-γ-ADABA, which may be assumed as a limit for enabling the cyclisation reaction of EctC. Defining these situations provisionally as “near attack conformations” (NAC), we observed 88 and 17 NAC snapshots out of 10 000 for model 1, using the Sem_K and Param_t sets, respectively. For model 2, 2 (Sem_K) and 4 (Param_t) NAC snapshots were observed. It remains open if this difference is statistically

significant, yet we consider it as a warning that meaningful modeling of the reaction can only be achieved with an appropriate parametrization of the Fe cofactor.

4 Conclusions

The results presented above show that very accurate minimum energy geometries could be obtained with force constants derived from Hessian (Seminario or PHF methods) when combined with reference bond length and valence angle values refined with the Katachi amendment (Sem_K and PHF_K sets). Unfortunately, such parameter sets do not allow for accurate reproduction of energy variation with geometrical changes brought about by thermal fluctuations ($T=300$ K), as for these sets MM-QM energy differences can be as high as 150 kcal/mol. On the other hand, parameters derived from the energy fitting procedure (Paramfit) gave much, much better performance in terms of energy, yet stationary geometry and, to a lesser extent, average dynamics geometry are less accurate. As for Param sets only a

Table 6 Bond length [Å] and valence angle [°] for model 1 optimized in vacuum at a specified level of theory

Bond / angle	DFT	Sem	Sem_K	PHF	PHF_K	Param	Param_K	Param_t
Fe-O2	2.071	1.931	2.065	1.905	2.064	1.831	2.135	1.968
Fe-OH	1.923	1.864	1.920	1.864	1.918	1.986	2.000	1.977
Fe-NB	2.109	2.116	2.108	2.166	2.109	2.153	2.160	2.157
Fe-O	2.182	1.767	2.162	1.821	2.149	2.049	2.353	2.083
Fe---O2 (not bonded)	2.311	2.120	2.274	2.133	2.285	2.276	2.209	2.272
MAE	-	0.162	0.013	0.164	0.014	0.103	0.093	0.069
MaxE	-	-0.416	-0.037	-0.361	-0.033	-0.240	0.171	-0.103
O2-Fe-OH	130.94	128.40	130.62	118.96	125.82	130.24	131.65	133.11
O2-Fe-NB	128.01	129.17	127.85	131.73	133.65	115.69	129.57	121.86
O2-Fe-O	93.47	94.79	93.44	96.09	94.48	59.58	57.40	95.74
OH-Fe-NB	98.42	100.81	99.48	106.17	98.02	95.56	98.57	102.95
O-Fe-NB	95.50	80.15	93.76	81.09	96.47	70.39	84.48	88.78
OH-Fe-O	97.31	80.31	97.00	72.58	94.46	101.51	136.92	98.44
CO-O2-Fe	94.83	88.39	94.27	89.16	94.52	96.87	91.03	95.75
C-OH-Fe	121.02	114.71	121.24	104.90	120.81	63.40	114.07	122.24
CV-NB-Fe	123.26	123.93	123.25	125.23	123.10	133.66	128.72	125.12
CR-NB-Fe	115.16	110.98	115.50	108.05	115.02	111.67	109.66	105.15
C-O-Fe	118.18	117.60	117.94	115.37	116.43	124.01	129.30	111.92
MAE	-	5.27	0.45	8.99	1.69	14.40	11.09	3.930
MaxE	-	-17.00	-1.74	-24.7	5.65	-57.60	39.61	-10.00
1st shell RMSD	-	0.303	0.026	0.352	0.060	0.556	0.595	0.132

MAE—mean absolute error, MaxE—maximum signed error; errors given separately for bonds and angles, 1st shell RMSD [Å]—root mean square displacement for Fe and atoms bonded directly with it. All errors are calculated with respect to the DFT optimized structure

limited number of bonds and/or angles showed large discrepancies from the reference values, manual adjustment of a few parameters gave us Param_t sets that are still very good in terms of energy and they well reproduce averaged bond lengths and angle values as well as their distributions (histograms). Most likely the procedure of amendment of

Param sets can be automated via adding geometry terms to the minimized penalty function, as previously demonstrated by Norrby and Liljefors [8], yet this requires code development, optimization of weights for different terms and validation of the method. Hence, for the moment we are of the opinion that the Param_t sets derived here offer the best

Table 7 MAE and MaxE values calculated for energy values [kcal/mol]

	DFT	Sem	Sem_K	PHF	PHF_K	Param	Param_K	Param_t
<i>MAE</i>	-	15.42	10.48	12.53	8.33	2.24	5.22	2.38
<i>MaxE</i>	-	68.35	49.59	74.07	59.79	9.74	16.37	9.09
R_E^2	-	0.474	0.506	0.406	0.399	0.588	0.247	0.582

compromise between energy and geometry accuracy, and they will be used in our ongoing computational studies on structure and function of EctC.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00214-021-02796-z>.

Authors' contribution J.A. Methodology, Investigation, Data curation, Writing—original draft, Writing—review and editing, Visualization. J.H. Conceptualization, Data curation, Writing—original draft, Writing—review and editing. T.B. Conceptualization, Methodology, Investigation, Data curation, Validation, Writing—original draft, Writing—review and editing. Visualization, Supervision, Project administration, Funding acquisition.

Funding This research was funded by the statutory research fund of ICSC PAS. This research was supported in part by PL-Grid Infrastructure. Computations were performed at Academic Computer Centre Cyfronet AGH.

Availability of data and material DFT optimized structures and ADMP trajectories (xyz files) as well as movies of ADMP trajectories (MPEG-2 files) are available at Mendeley Data: <http://dx.doi.org/10.17632/5cy5ph4ssj.1>.

Code availability Octave scripts used to compute quadratic form based distance between histograms as well as python3 scripts used to generate superposed histogram plots, together with underlying data, are available at Mendeley Data: <http://dx.doi.org/10.17632/5cy5ph4ssj.1>.

Declarations

Conflict of interest Authors declare no conflicts of interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Czech L et al (2019) Illuminating the catalytic core of ectoine synthase through structural and biochemical analysis. *Sci Rep* 9(1):364
- Widderich N et al (2016) Strangers in the archaeal world: osmo-stress-responsive biosynthesis of ectoine and hydroxyectoine by the marine thaumarchaeon *Nitrosopumilus maritimus*. *Environ Microbiol* 18(4):1227–1248
- Seminario JM (1996) Calculation of intramolecular force fields from second-derivative tensors. *Int J Quantum Chem* 60(7):1271–1277
- Wang R, Ozhgibesov M, Hirao H (2018) Analytical hessian fitting schemes for efficient determination of force-constant parameters in molecular mechanics. *J Comput Chem* 39(6):307–318
- Wang R, Ozhgibesov M, Hirao H (2016) Partial hessian fitting for determining force constant parameters in molecular mechanics. *J Comput Chem* 37(26):2349–2359
- Betz RM, Walker RC (2015) Paramfit: Automated optimization of force field parameters for molecular dynamics simulations. *J Comput Chem* 36(2):79–87
- Hu L, Ryde U (2011) Comparison of methods to obtain force-field parameters for metal sites. *J Chem Theory Comput* 7:2452–2463
- Norrby P, Liljefors T (1998) Automated molecular mechanics parameterization with simultaneous utilization of experimental and quantum mechanical data. *J Comput Chem* 19(10):1146–1166
- Li P, Merz KM Jr (2017) Metal ion modeling using classical mechanics. *Chem Rev* 117:1564–1686
- Schwibbert K et al (2011) A blueprint of ectoine metabolism from the genome of the industrial producer *Halomonas elongata* DSM 2581 T. *Environ Microbiol* 13(8):1973–1994
- Goraj W, Stępniewska Z, Szafranek-Nakoneczna A (2019) Biosynthesis and the possibility of using ectoine and hydroxyectoine in health care. *Postępy Mikrobiol - Adv Microbiol* 58(3):339–349
- Zheng H et al (2017) Data mining of iron(II) and iron(III) bond-valence parameters, and their relevance for macromolecular crystallography. *Acta Crystallogr Sect D Struct Biol* 73(4):316–325
- Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98(7):5648–5652
- Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion

- correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 132(15):154104
15. Weigend F, Ahlrichs R (2005) Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys Chem Chem Phys* 7(18):3297–3305
 16. Yanai T, Tew DP, Handy NC (2004) A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem Phys Lett* 393(1–3):51–57
 17. Yu HS, He X, Li SL, Truhlar DG (2016) MN15: A Kohn-Sham global-hybrid exchange-correlation density functional with broad accuracy for multi-reference and single-reference systems and noncovalent interactions. *Chem Sci* 7(8):5032–5051
 18. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery JA, Jr, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ (2016) Gaussian 16, Revision C.01
 19. Schlegel HB et al (2001) Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. *J Chem Phys* 114(22):9758–9763
 20. Iyengar SS, Schlegel HB, Millam JM, Voth GA, Scuseria GE, Frisch MJ (2001) Ab initio molecular dynamics: propagating the density matrix with Gaussian orbitals. II. Generalizations based on mass-weighting, idempotency, energy conservation and choice of initial conditions. *J Chem Phys* 115(22):10291
 21. Schlegel HB et al (2002) Ab initio molecular dynamics: propagating the density matrix with gaussian orbitals. III. Comparison with born-oppenheimer dynamics. *J Chem Phys* 117(19):8694–8704
 22. Iyengar SS, Schlegel HB, Voth GA, Millam JM, Scuseria GE, Frisch MJ (2002) Ab initio molecular dynamics: propagating the density matrix with gaussian orbitals IV Formal analysis of the deviations from born-oppenheimer dynamics. *Isr J Chem* 42(2–3):191–202
 23. Guo Y, Riplinger C, Liakos DG, Becker U, Saitow M, Neese F (2020) Linear scaling perturbative triples correction approximations for open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory [DLPNO-CCSD(T0/T)]. *J Chem Phys* 152(2):0–13
 24. Guo Y et al (2018) Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]. *J Chem Phys* 10(1063/1):5011798
 25. Riplinger C, Sandhoefer B, Hansen A, Neese F (2013) Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J Chem Phys* 139(13):134101
 26. Riplinger C, Neese F (2001) An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J Chem Phys* 113(3):034106
 27. Neese F (2017) Software update: the ORCA program system, version 4.0. *WIREs Comput Mol Sci* 8(1):e1327
 28. Angeli C, Cimiraglia R, Evangelisti S, Leininger T, Malrieu JP (2001) Introduction of n-electron valence states for multireference perturbation theory. *J Chem Phys* 114(23):10252–10264
 29. Angeli C, Cimiraglia R, Malrieu JP (2001) N-electron valence state perturbation theory: a fast implementation of the strongly contracted variant. *Chem Phys Lett* 350(3–4):297–305
 30. Angeli C, Cimiraglia R, Malrieu JP (2002) N-electron valence state perturbation theory: a spinless formulation and an efficient implementation of the strongly contracted and of the partially contracted variants. *J Chem Phys* 117(20):9138–9153
 31. Havenith RWA, Taylor PR, Angeli C, Cimiraglia R, Ruud K (2004) Calibration of the n-electron valence state perturbation theory approach. *J Chem Phys* 120(10):4619–4625
 32. Schapiro I, Sivalingam K, Neese F (2013) Assessment of n-electron valence state perturbation theory for vertical excitation energies. *J Chem Theory Comput* 9(8):3567–3580
 33. Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham III TE, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Gilson MK, Gohlke H, Goetz AW, Greene D, Harris R, Homeyer N, Huang Y, Izadi S, Kovalenko A, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein DJ, Merz KM, Miao Y, Monard G, Nguyen C, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shen J, Simmerling CL, Smith J, Salomon-Ferrer R, Swails J, Walker RC, Wang J, Wei H, Wolf RM, Wu X, Xiao L, York DM and Kollman PA (2018), AMBER 2018, University of California, San Francisco
 34. Rappé AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM (1992) UFF, a full period table force field for molecular mechanics and molecular dynamics simulations. *J Am Chem Soc* 114(25):10024–10035
 35. Bernas T, Asem EK, Robinson JP, Rajwa B (2008) Quadratic form: A robust metric for quantitative comparison of flow cytometric histograms. *Cytom Part A* 73(8):715–726
 36. Eaton JW, Bateman D, Hauberg S, Wehbring R (2021) GNU Octave: Free Your Numbers. A high-level interactive language for numerical computations. Edition 6 for Octave version 6.2.0. URL <https://octave.org/doc/v6.2.0/>
 37. Sondergaard CR, Olsson M, Rostkowski M, Jensen JH (2011) Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J Chem Theory Comput* 7:2284–2295
 38. Olsson M, Sondergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput* 7(2):525–537

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.