Ectoine synthase (EctC) - Fe(II) binding site parametrization

MSc Justyna Sonia Andrys
supervisor: prof. dr hab. Tomasz Borowski

Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences

Outline

1. Brief introduction to the enzyme and catalyzed reaction
2. Motivation for parameterization
3. DFT optimization
4. ADMP
5. Methods description
6. Parameters & their evaluation
7. Conclusions

Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences

Ectoine synthase and Ectoine

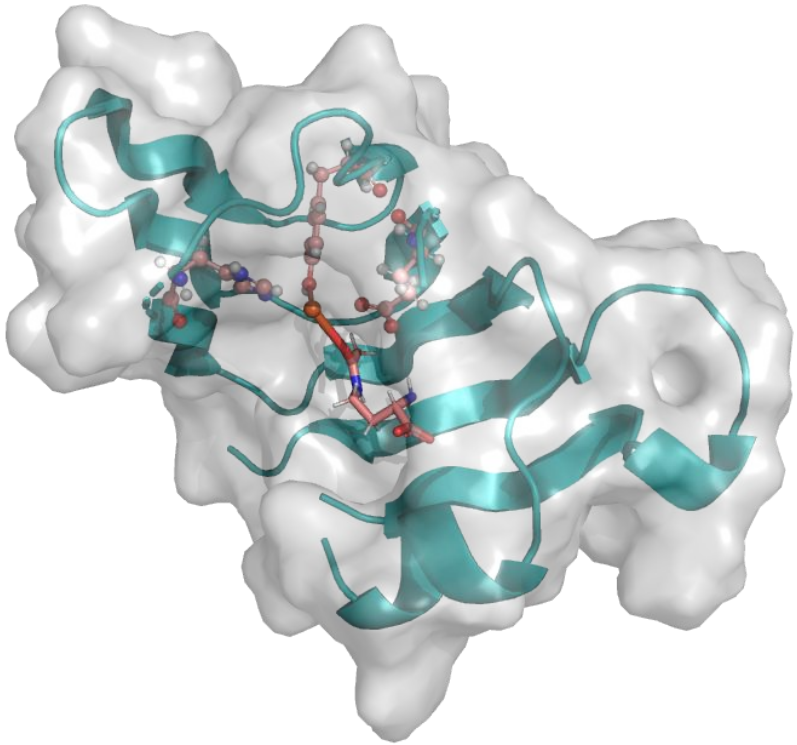


Fig. 1 *P. laucus* EctC crystal structure with catalytic core exposed. Fe(II) ion with ligands (Glu57, Tyr84, His92 and N-γ-acetyl-2,4-diaminobutyrate) are visible inside the pocket. PDB: 5ONN[1,2]

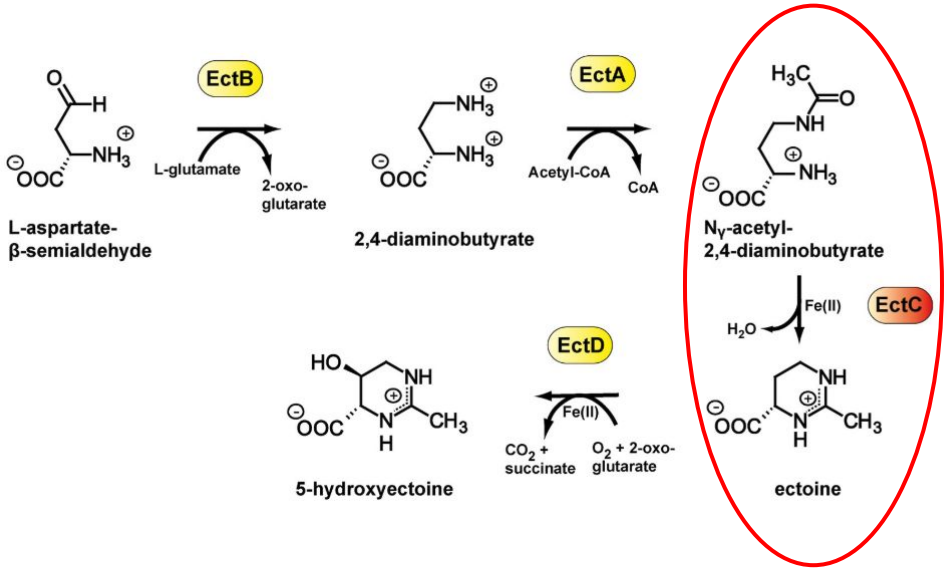
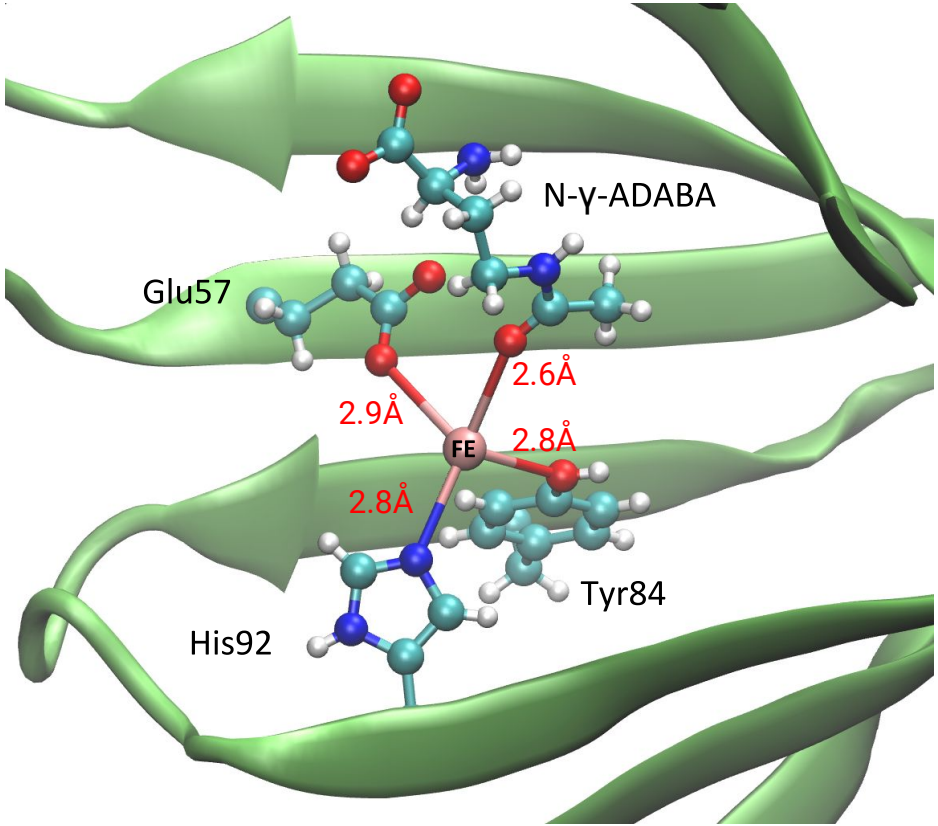


Fig. 2 Pathway of the ectoine biosynthesis. Cyclocondensation reaction of N-γ-ADABA marked with red circle[3].

Motivation for the parametrization



- Fe - L bonds too long
- geometry refinement needed before QM and QM/MM calculations
- reliable parameters necessary for MD simulation

Fig. 3 Model of Fe(II) binding site cut from 5ONN[1] crystal structure. Fe - L distances shown in red.

DFT/B3LYP optimization

E_5+ 36.7 kcal/mol

E_5+ 18.3 kcal/mol

E_5- reference

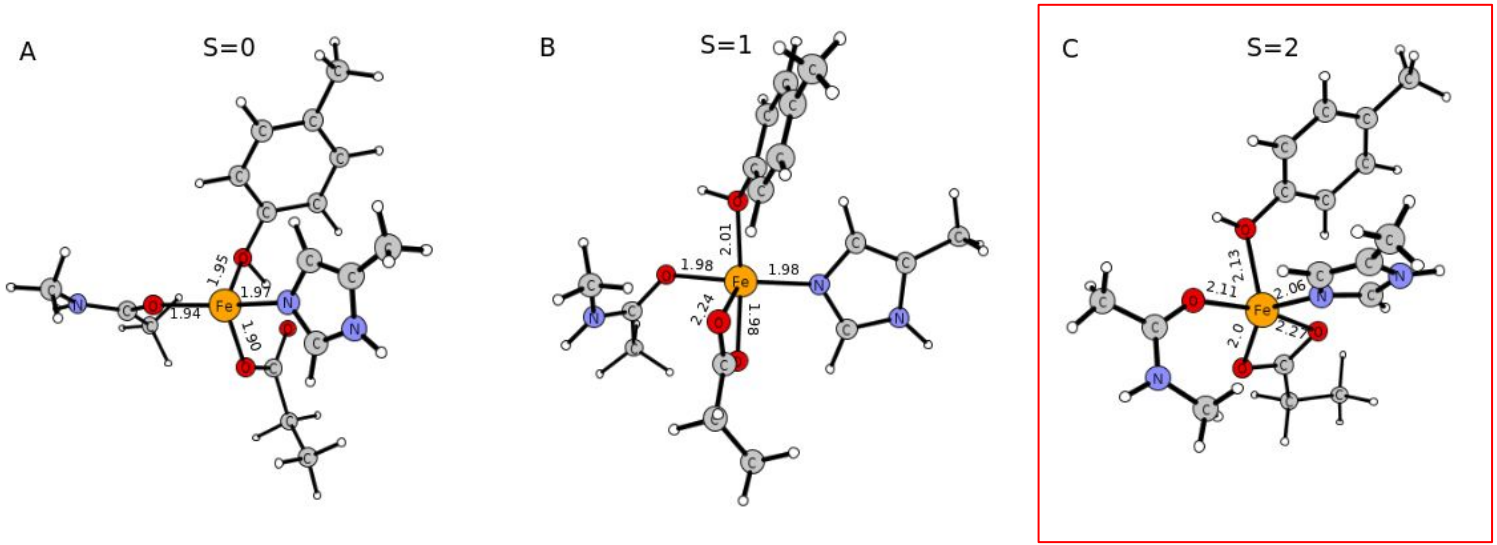


Fig. 4 Three optimized geometries with Fe(II) in different spin-state: A) singlet, B) triplet i C) quintet. Calculated with DFT: B3LYP-D3/Def2TZVP.

ADMP - quintet model

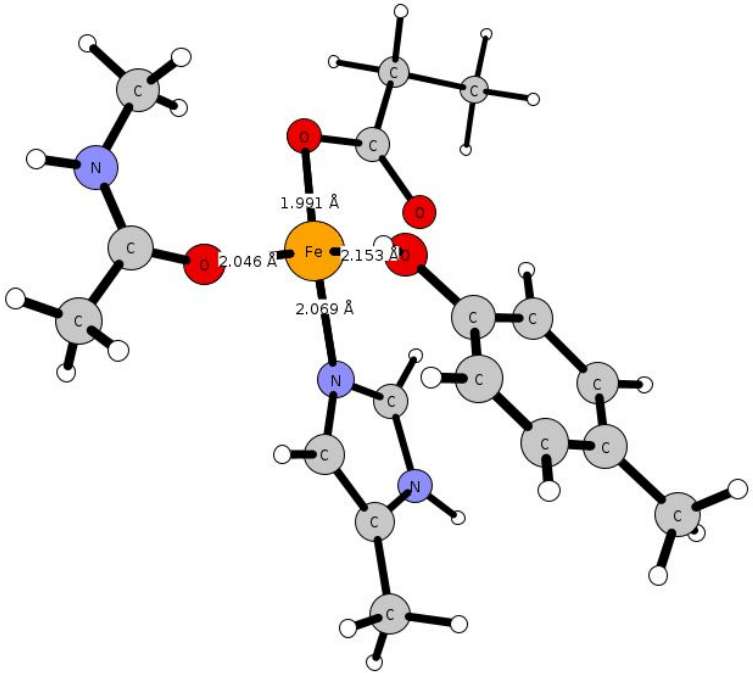


Fig. 5 Reference trajectory (149 frames out of 23,000)

ADMP - *atom-centered density matrix propagation*; combination of QM with MD (*Gaussian16*)

- ADMP trajectory: ~23,000 frames
- 149 frames cut from original trajectory for parameterization purposes (in average 1 frame per 154)
- Each of geometry from the short trajectory optimized (with constraints on Fe - X bonds and X-Fe-Y and X-Y-Fe angles)
- QM energies calculated with UB3LYP-D3 functional and Def2TZVP basis set

SEMINARIO[4]

- the easiest of investigated methods
- uses Hessian matrix projection for MM parameters derivation
- what do we need?
  - *.fchk file generated from *.chk file obtained after freq calculation with use of the formchk program (Gaussian)
  - xyzviewer for force constants reading

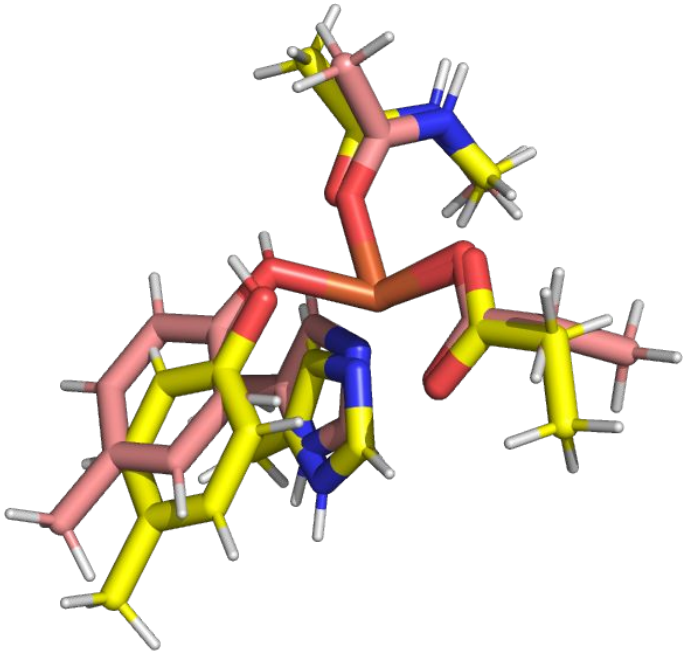


Fig. 6 Optimized geometries alignment: **QM** (salmon), **seminario** (yellow)

PARMHESS-1[5,6]

- PHF - partial hessian fitting
- what do we need?
  - python3 environment
  - Gaussian09 (local access)
  - AmberTools (resp, antechamber)
  - *.log file from freq calculations
  - *.fchk file from freq calculations
  - Gaussian MM input that contains geometry, charges, connectivity section and frcmod parameters including nonbon and VDW
    - HrmBnd1 Fe OH HO XXXXXX 121.0489

- **tsubasa.py** program from the *parmhess package* automatically generates input for **parmhess.py** program: calculates RESP, MK charges, frequencies
- **parmhess.py** generates G09 inputs for optimization: ihf, phf, fhf

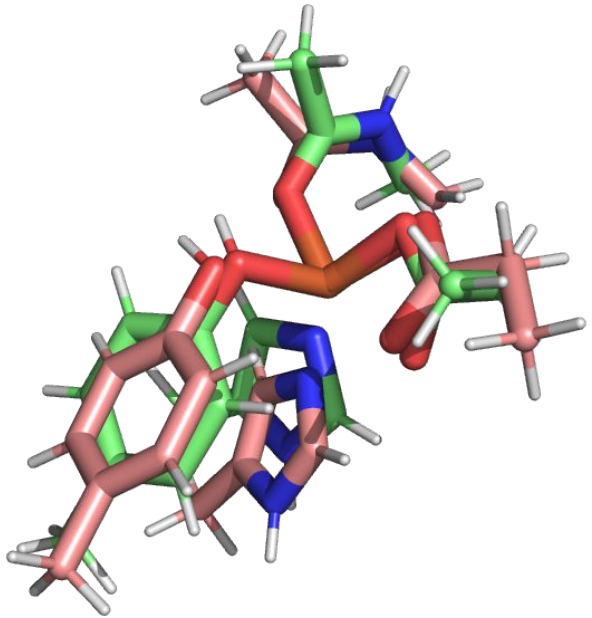


Fig. 7 Optimized geometries alignment: **QM** (salmon), **phf** (green)

PARMHESS-2[5,6]

- PHF - **katachi**
- what do we need?
  - **parmhess.py**-generated input (ihf, fhf or phf*.com)
- **katachi.py** improves input parameters and generates new G09 input for optimization

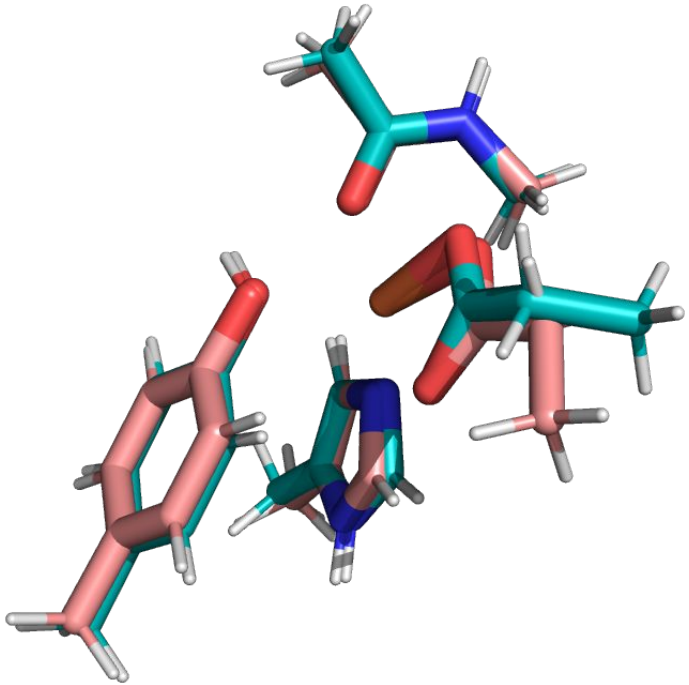


Fig. 8 Optimized geometries alignment: **QM** (salmon), **phf-katachi** (blue)

SEMINARIO-KATACHI

- **Refinement of seminario-generated parameters with katachi.py software**
- what do we need?
  - katachi.py input (as previous), this time with seminario parameters

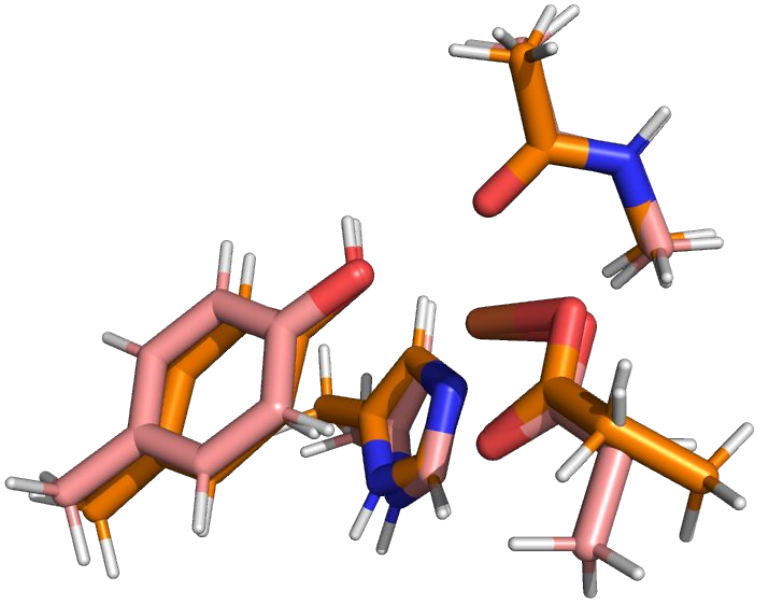


Fig. 9 Optimized geometries alignment: **QM** (salmon), **seminario-katachi** (orange)

PARAMFIT[7]

- the most time-consuming method
- available in AmberTools package
- what do we need?
  - pdb-cut model and *.in files as input for the tleap program (each amino acid described separately), with resp charges
  - trajectory (149 frames)
  - QM energies for each frame
- paramfit
  - generates Gaussian inputs (calculated on prometheus),
  - generates *.paramfit file that contains data on fitted parameters (*Fe-X, X-Fe-Y, X-Y-Fe*)
  - does parameters fitting on the basis of ADMP trajectory (149 frames long) and QM energies
  - user-friendly for tuning purposes

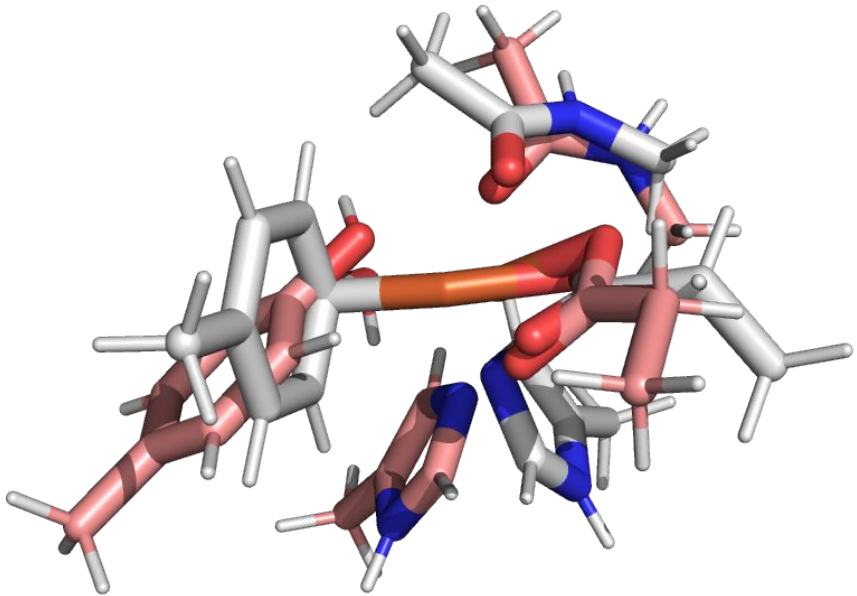


Fig. 10 Optimized geometries alignment: **QM** (salmon), **paramfit** (silver)

PARAMFIT-TUNED

- paramfit parameters with only one K and one th_{eq} changed (O2-FE-O parameters taken from the seminario method)

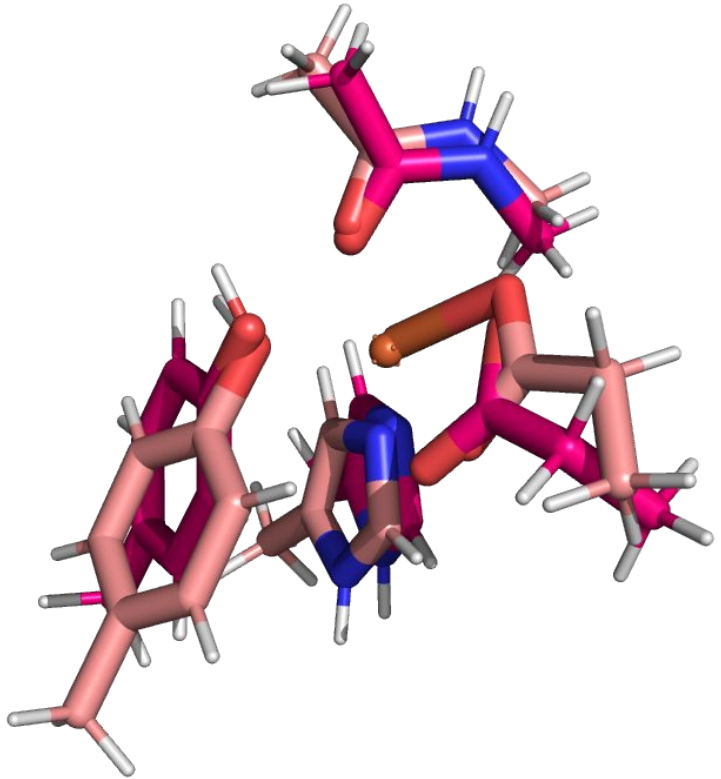
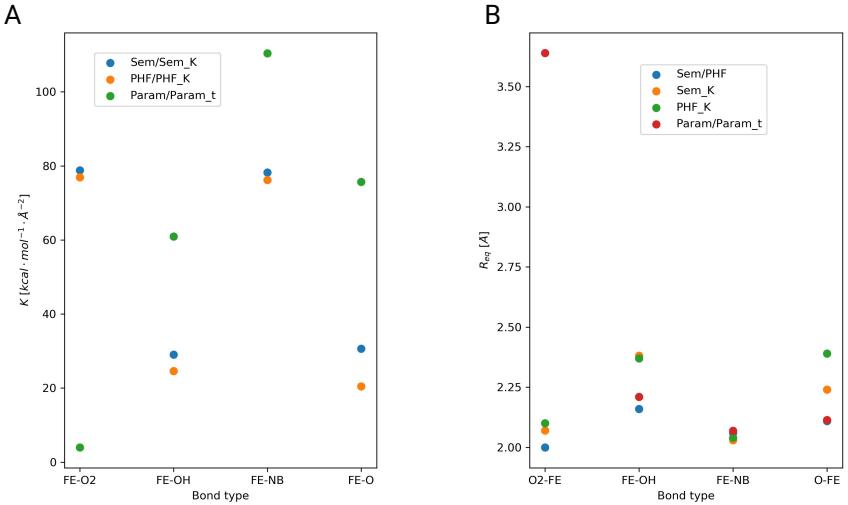


Fig. 11 Optimized geometries alignment: **QM** (salmon), **paramfit-tuned** (magenta)

PARAMETERS



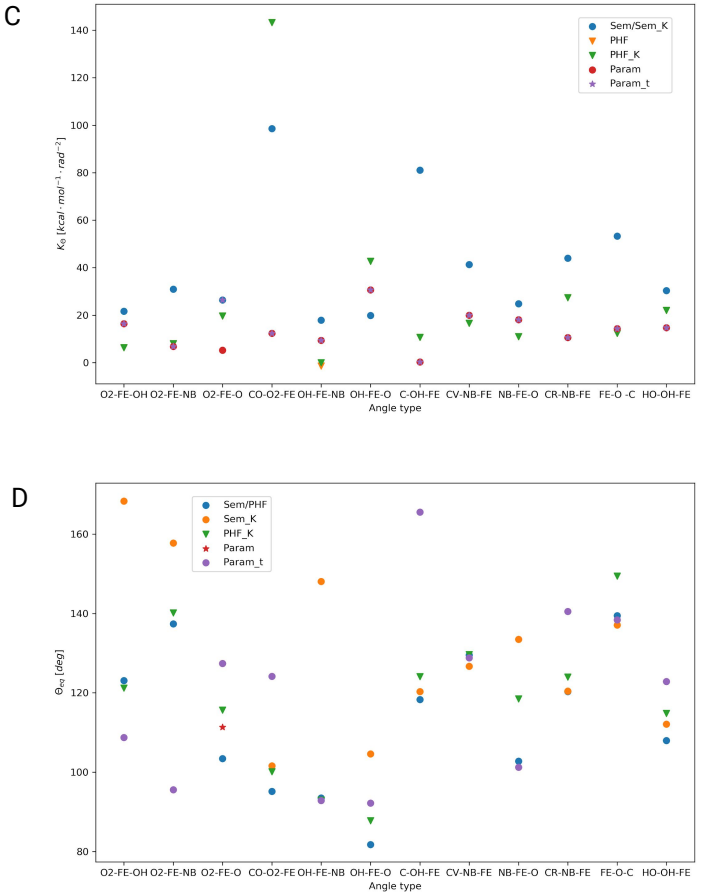


Fig. 12 Plots presenting derived parameters for each method: A) K_{bond}, B) R_{eq}, C) K_{angle}, D) Th_{eq}

Quality assessment criteria

quick sander MD simulation was run for each set of parameters in order to obtain angle/bond values distributions and calculate means

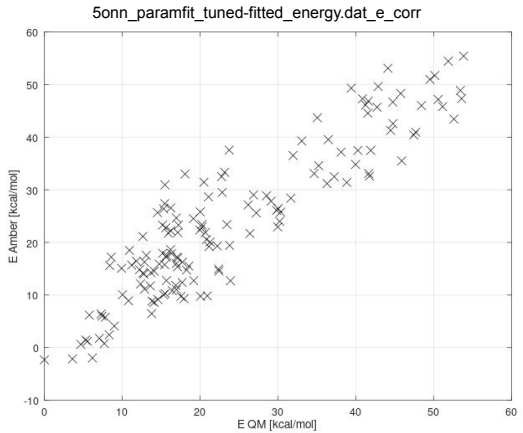
1. Convergence of the optimized geometry with the referential one (graphics) - based on differences between bond/angle values
2. QM and AMBER (fitted) energies convergence (R^2 coefficient) calculated with **paramfit**
3. Consistency of mean angle/bond values with means calculated for the **ADMP** (~23,000 frames)

Quality assessment

	SEMINARIO	PARMHESS-1	PARMHESS-2	SEMINARIO-KATACHI	PARAMFIT	PARAMFIT-TUNED
R^2	0.42	0.46	0.46	**0.3**	**0.94**	**0.83**

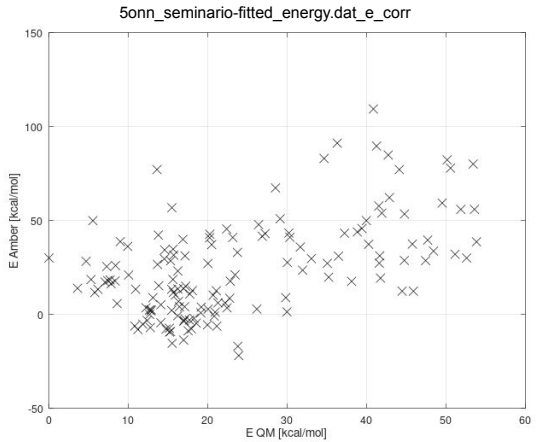
Tab. 1 Energy correlation coefficients calculated with paramfit for each method

A



$R^2 = 0.83$

B

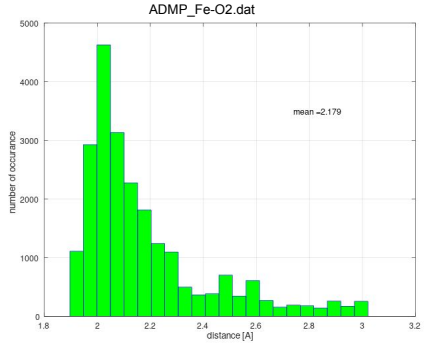


$R^2 = 0.3$

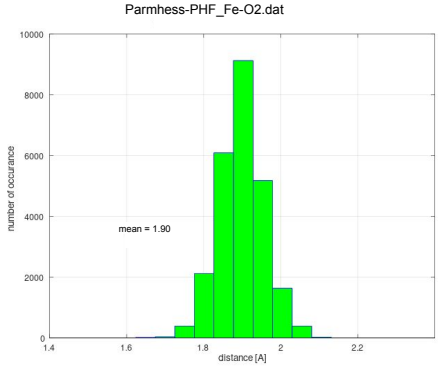
Fig. 13 Plots E_{Amber} vs E_{QM} generated for A) PARAMFIT-TUNED and B) SEMINARIO-KATACHI

Quality assessment

A



B



C

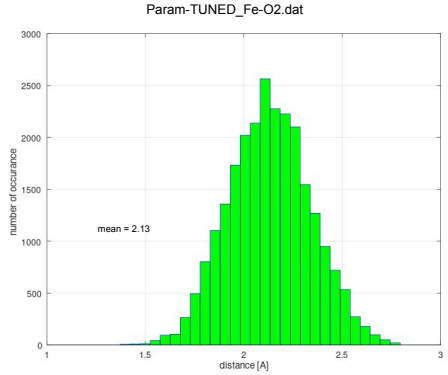
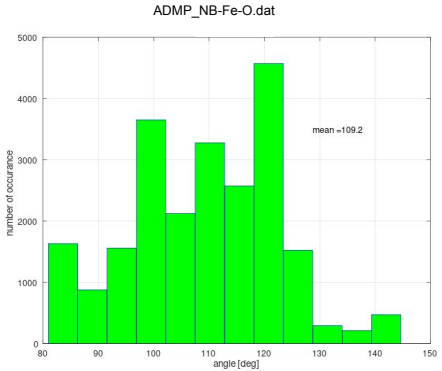
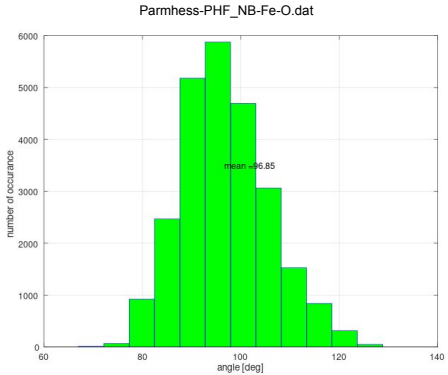


Fig. 14 Histograms for bond Fe-O2 distributions: A) ADMP B) Parmhess-PHF C) Paramfit-TUNED

A



B



C

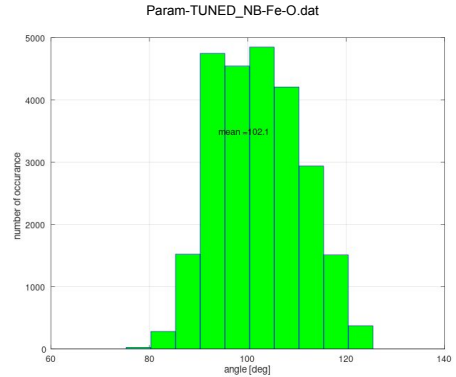


Fig. 15 Histograms for NB-Fe-O angle distributions: A) ADMP (reference) B) Parmhess-PHF C) Paramfit-TUNED

Conclusions

1. Three methods with variations were tested
2. One criterion is not enough to assess method's quality properly
3. Best optimized geometry was obtained with ***parmhess-phf*** method
4. Best correlation between fitted energy and QM energy (represented by R^2) was obtained with ***paramfit*** method
5. Best results for angle/bond means from MD were assigned to the ***paramfit_tuned*** method (smallest errors)
6. Overall, ***paramfit_tuned*** approach, where parameters describing just one angle were switch with **seminario** parameters, was chosen as the method providing the best results for investigated model

Thank you for the attention :)

Special thanks to Tomek

Calculations financed by IKiFP and group funds.

Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences

Sources

1. https://www.rcsb.org/structure/5ONN
2. Czech L. et al. *Illuminating the catalytic core of ectoine synthase through structural and biochemical analysis,* Scientific Reports, 2019
3. Widderich N. et al. *Biochemistry and Crystal Structure of Ectoine Synthase: A Metal-Containing Member of the Cupin Superfamily,* Plos One, 2016
4. J. M. Seminario, *Calculation of intramolecular force fields from second-derivative tensors*, Int. J. Quantum Chem, 1996.
5. R. Wang et al. *Analytical Hessian Fitting Schemes for Efficient Determination of Force-Constant Parameters in Molecular Mechanics*, J.Comput. Chem. 2017,
6. R. Wang et al *Partial Hessian Fitting for Determining Force Constant Parameters in Molecular Mechanics* J.Comput. Chem. 2016,
7. Betz R.M. & Walker R.*C, Paramfit: automated optimization of force field parameters for molecular dynamics simulations,* Journal of Computational Chemistry, 2014