QM/MM methods: Current developments and applications

Walter Thiel

Max-Planck-Institut für Kohlenforschung, Mülheim

- Overview
- Accurate ab initio QM/MM
- Free energy calculations
- Boundary potentials
- Cytochrome P450
- Xanthine oxidase

QM/MM approach: General overview

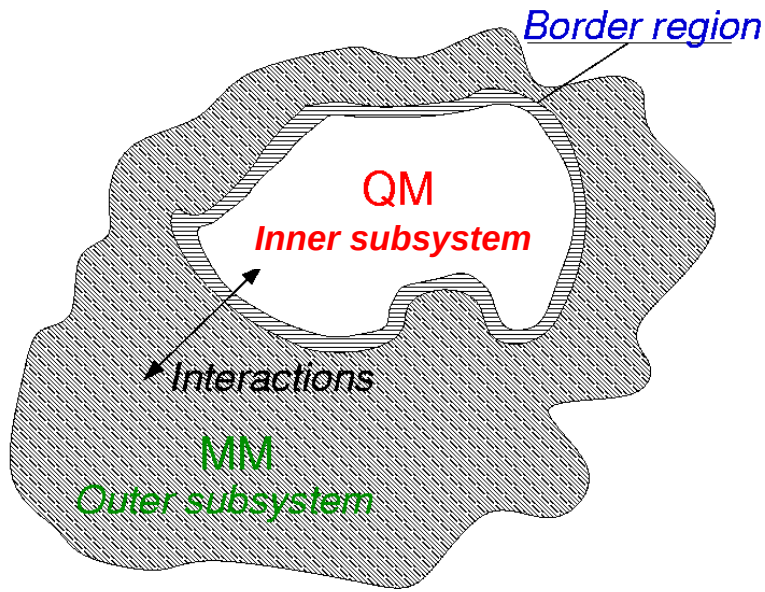
QM: ab initio, DFT, semiempirical

MM: standard force field

QM – MM interactions:

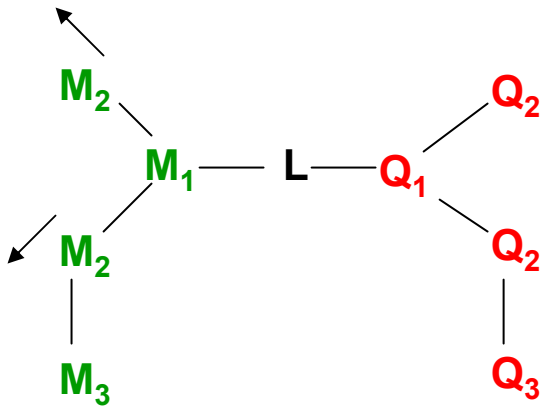
„*electronic embedding*“

$$\hat{H}^{I,O}_{QM-MM} = -\sum_{i,J}\frac{q_J}{r_{iJ}} + \sum_{i,J}\frac{q_J Z_A}{R_{AJ}} + \sum_{A,J}\left(\frac{A_{AJ}}{R^{12}_{AJ}} - \frac{B_{AJ}}{R^{6}_{AJ}}\right)$$



Border region:

- hydrogen link atoms **L**
- charge shift for q(M_1)



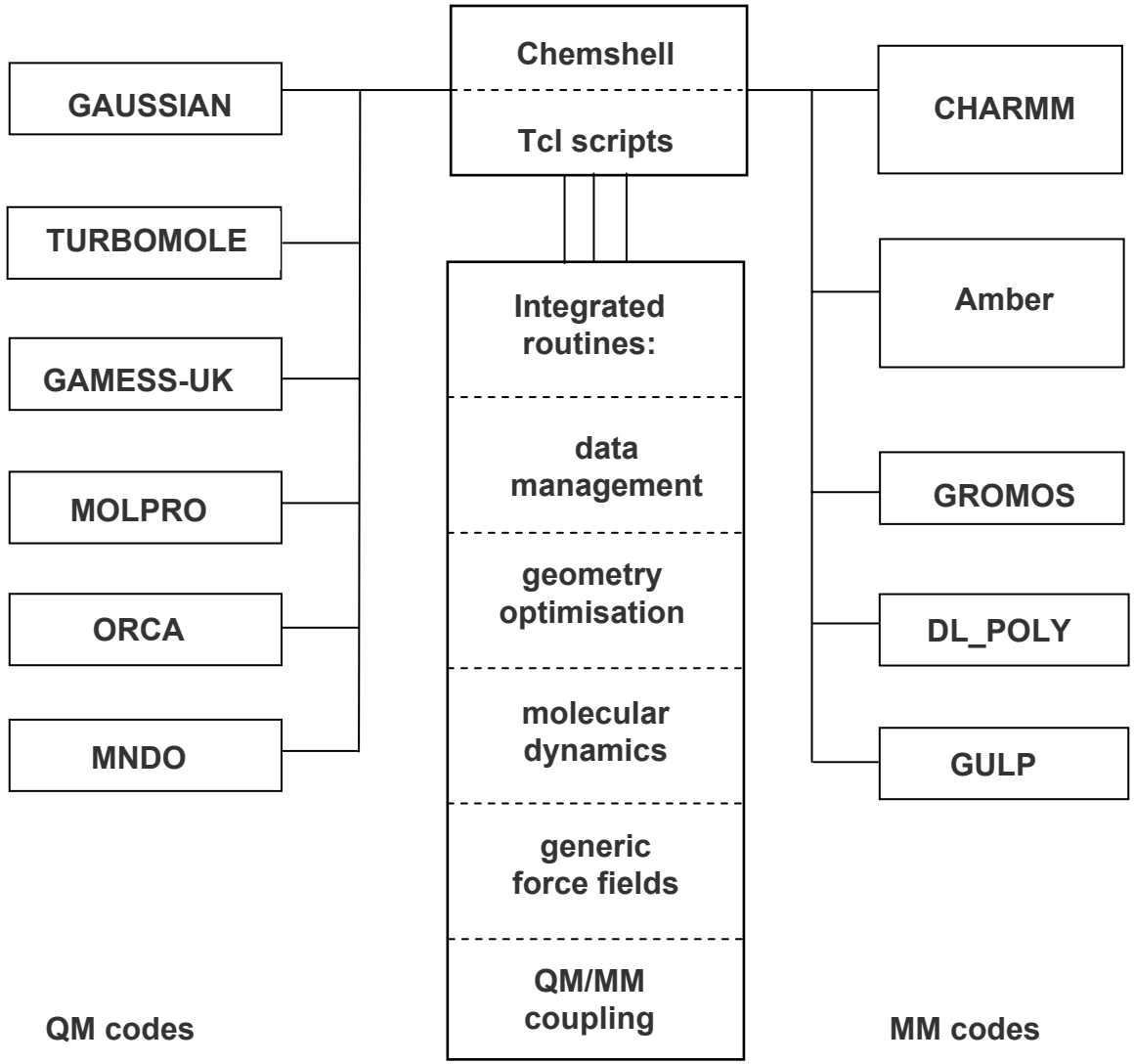
Codes:

ChemShell as control module

Interfaces to standard QM and MM codes

H. M. Senn and W. Thiel, Top. Curr. Chem. **268**, 173-290 (2007).

ChemShell: A modular QM/MM package



P. Sherwood et al, J. Mol. Struct. Theochem **632**, 1-28 (2003).

http://www.chemshell.org

Consensus QM/MM approach to biomolecular modeling

- Total system size of 10000-40000 atoms including solvent
- Active-site QM region of typically 50-100 atoms
- Standard DFT as QM component (typically B3LYP)
- Standard force field as MM component (CHARMM, GROMOS, AMBER, OPLS)
- Electrostatic QM/MM embedding
- QM/MM boundary treated by link atom scheme
- Geometry optimization of a limited number of snapshots
- Computation of reaction paths and energy profiles

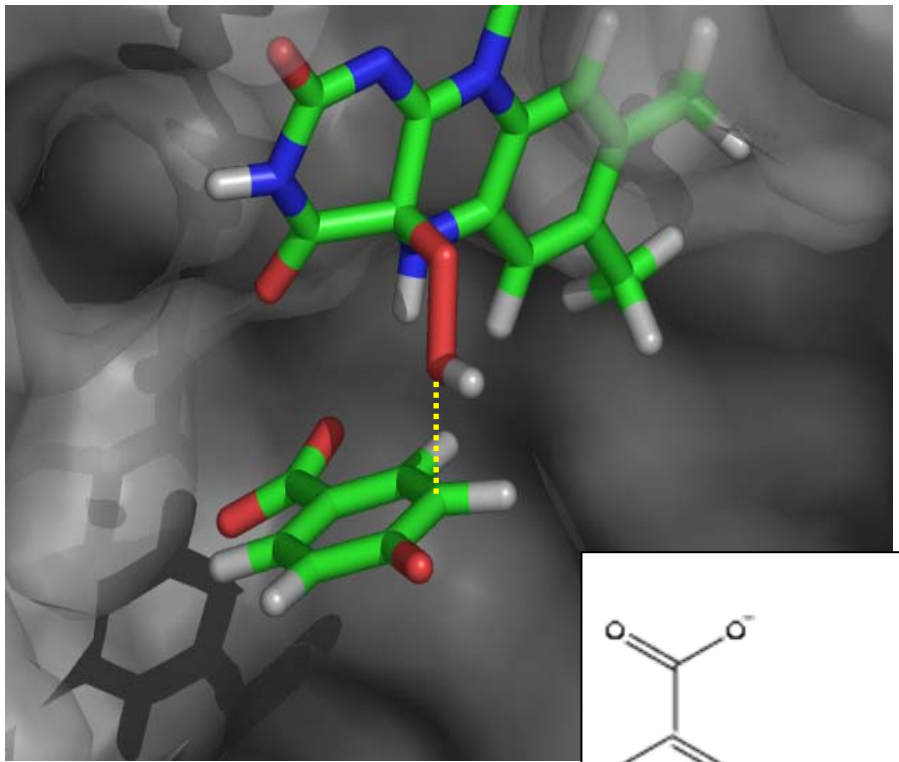
H. M. Senn and W. Thiel, Angew. Chem. Int. Ed. **48**, 1198-1229 (2009).

Possible improvements of the consensus QM/MM approach

- Larger active-site QM regions
- High-level correlated ab initio methods as QM components
- Dispersion corrections for lower-level QM methods
- Polarized force fields as MM component
- Polarized QM/MM embedding
- More refined QM/MM boundary treatments
- Proper sampling through molecular dynamics or Monte Carlo methods
- Computation of free energy profiles
- Adaptive QM/MM partitioning
- Extension to excited-state QM/MM modeling
- Extension to three-layer QM/MM/continuum approaches
- Periodic boundary conditions versus finite model systems

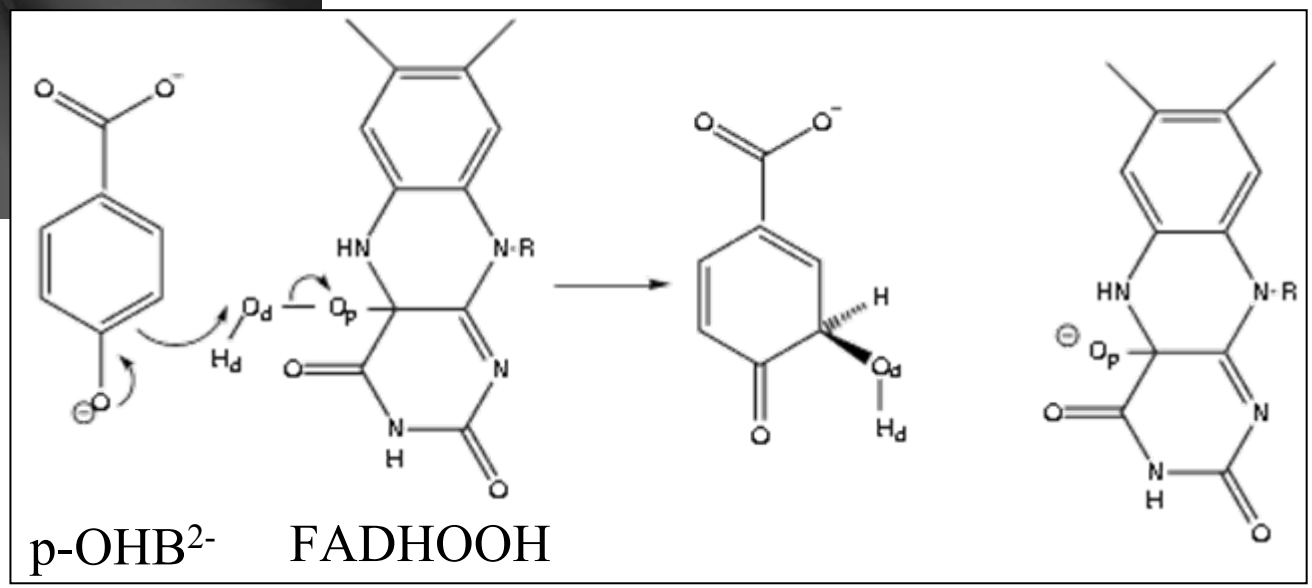
PHBH : p-hydroxybenzoate hydroxylase

Aromatic hydroxylation of p-hydroxybenzoate

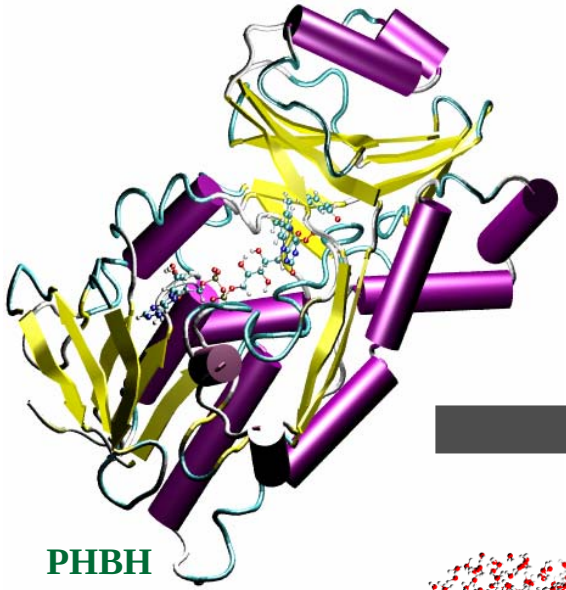


- rate-determining step: oxygen transfer from cofactor FADHOOH to *p*-OHB (FAD: flavin adenine dinucleotide)
- electrophilic aromatic substitution with heterolytic cleavage of the peroxide bond
- activation energy: 12 kcal/mol

Reaction mechanism of key step in PHBH



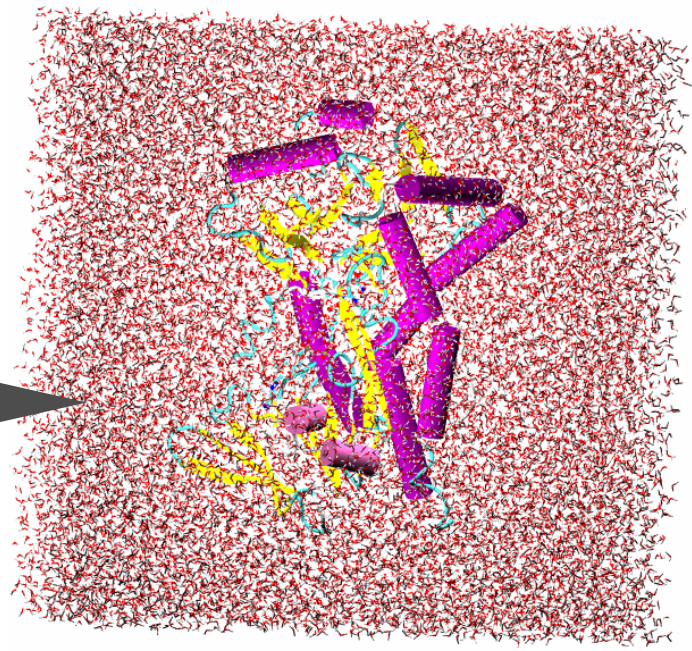
PHBH : General setup



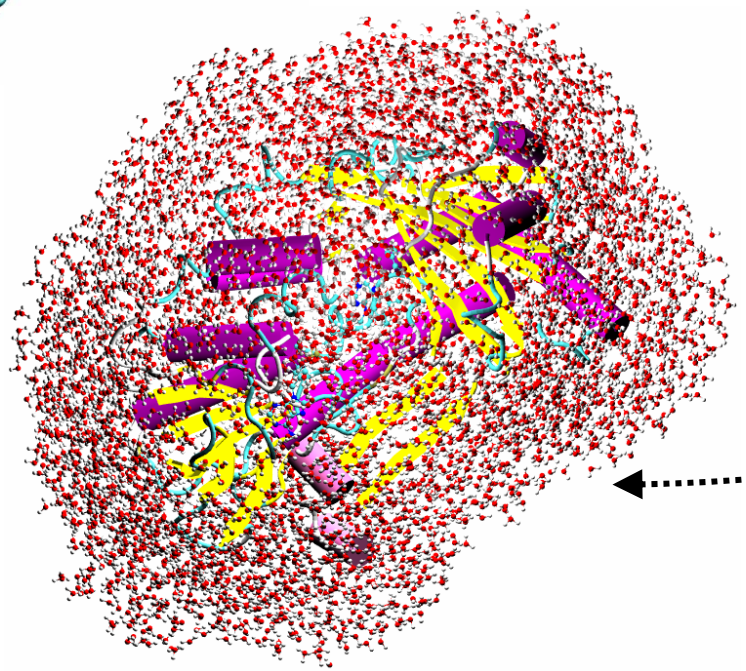
PHBH
7004 atoms

- solvate
- EM – MD cycles
- snapshots from subsequent MD as initial configurations for QM/MM
- substrate, cofactor restrained

solvate



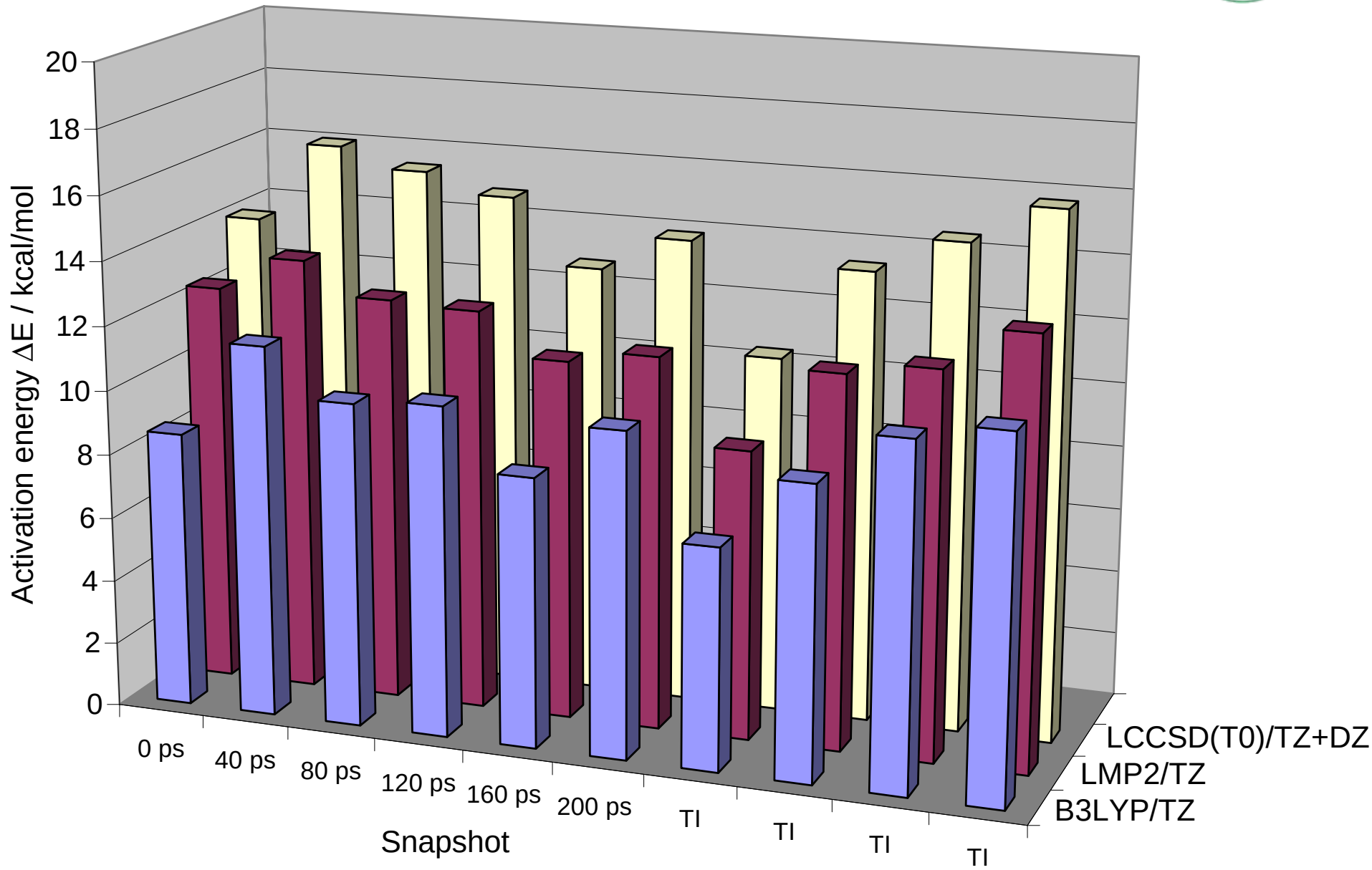
PHBH solvated in cubic box 75000 atoms



PHBH-water cluster 25000 atoms

- cut out PHBH-water aggregate
- outermost water layer fixed

SP LMP2/GROMOS and LCCSD(T0)/GROMOS barriers (TZ basis)



PHBH : QM convergence tests for QM/MM barrier

Basis set: Tested at the DF-MP2 level. The barrier for the chosen basis (TZ=[aug]-cc-pVTZ) is 0.3 (0.4) kcal/mol lower than the aug-cc-pVQZ (CBS) value.

Local approximation: Tested by comparing DF-MP2 and DF-LMP2 barriers for the TZ basis. The LMP2 calculations underestimate the canonical MP2 barriers by 0.3 kcal/mol (average value for 10 snapshots).

Weak-pair approximation: Tested by varying the distance parameters that distinguish between orbital pair types in local coupled cluster calculations (R_c strong/close, R_w close/weak, R_d weak/distant). The barriers are quite intensitive to R_c and R_d (changes of 0.2 kcal/mol or less), but very sensitive to R_w (changes of several kcal/mol). R_w must be chosen large enough (5-7 Bohr) to include alle relevant close pairs in the triples calculation.

R. A. Mata, H.-J. Werner, S. Thiel, and W. Thiel, J. Chem. Phys. **128**, 025104 (2008).

PHBH: Comparison of barriers

QM/GROMOS results (kcal/mol) at B3LYP geometries:

QM method	Range	Average	rms
B3LYP (a)	5.2 - 9.6	7.9	1.3
DF-LMP2 (b)	9.0 - 13.6	12.0	1.3
DF-LCCSD(T0) (b)	11.1 - 16.6	14.6	1.6

a) TZVP basis.
b) cc-pVTZ basis in general, aug-cc-pVTZ for O.

Experimentally derived enthalpy of activation: 12 kcal/mol [1],
from temperature-dependent measurements of the overall rate.
Experimentally derived free enthalpy of activation: 14 - 15 kcal/mol [2],
from measured individual and overall rate constants.

Estimate for the zero-point vibrational and thermal enthalpic corrections to barrier from
AM1 gas-phase calculations of 102-atom QM region: -1.3 kcal/mol (at 300 K)
Resulting LCCSD(T0)-based prediction of activation enthalpy: 13.3 kcal/mol (at 300 K)
Average entropic contribution to barrier from QM/MM TI runs: 0.4 kcal/mol (at 300 K)
Best prediction of free energy barrier: 13.7 kcal/mol (at 300 K)

[1] W. J. H. van Berkel, F. Müller, Eur. J. Biochem. **179**, 307 (1989).
[2] B. Entsch, B. A. Palfey, D. P. Ballou, V. Massey, J. Biol. Chem. **266**, 17341 (1991).

PHBH and CM: Comparison of barriers

PHBH: p-hydroxybenzoate hydroxylase, electrophilic aromatic substitution

CM: chorismate mutase, pericyclic Claisen rearrangement

Computed QM/MM activation enthalpies (kcal/mol)[a]

Method	HF	B3LYP	LMP2	LCCSD	LCCSD(T0)	Experiment
CM	28.3	10.2	9.5	18.7	13.1	12.7
PHBH[b]	36.7	8.4	10.7	20.2	13.3	12.0

(a) Average of 16 (CM) or 10 (PHBH) single-point calculations at B3LYP/MM optimized geometries, zero-point energy and 300 K thermal corrections from QM calculations on cluster models, aug-cc-VTZ basis on oxygen and cc-pVTZ basis on all other atoms, MM=CHARMM for CM and MM=GROMOS for PHBH.

(b) Average AM1/GROMOS values for PHBH: 22.8 kcal/mol

Accurate electronic structure methods and transition state theory describe enzymatic reactions quantitatively.

F. Claeyssens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland, K. E. Ranaghan, M. Schütz, S. Thiel, W. Thiel, and H.-J. Werner, Angew. Chem. **118**, 7010 (2006).

QM/MM free energy calculations

- **Essential for proper treatment of reaction rates**
- **Important for understanding the origin of enzymatic catalysis through comparison of free energy barriers in solution and in the enzyme**
- **Affordable: semiempirical QM/MM**
- **Challenging: ab initio QM/MM**

H. Hu and W. Yang, Annu. Rev. Phys. Chem. **59**, 573-601 (2008).

ChemShell: Simulation techniques

Sampling methods implemented in ChemShell:

-Thermodynamic integration [1]

- Umbrella sampling [2]

- Approximate free-energy perturbation methods [3]

Novel analysis method: **Umbrella integration** [2,4]

Consistent results for free-energy barriers in PHBH [1-4]

from AM1/GROMOS calculations, for example:

24.2 ± 0.5 kcal/mol [1] and 24.3 ± 0.5 kcal/mol. [2,4]

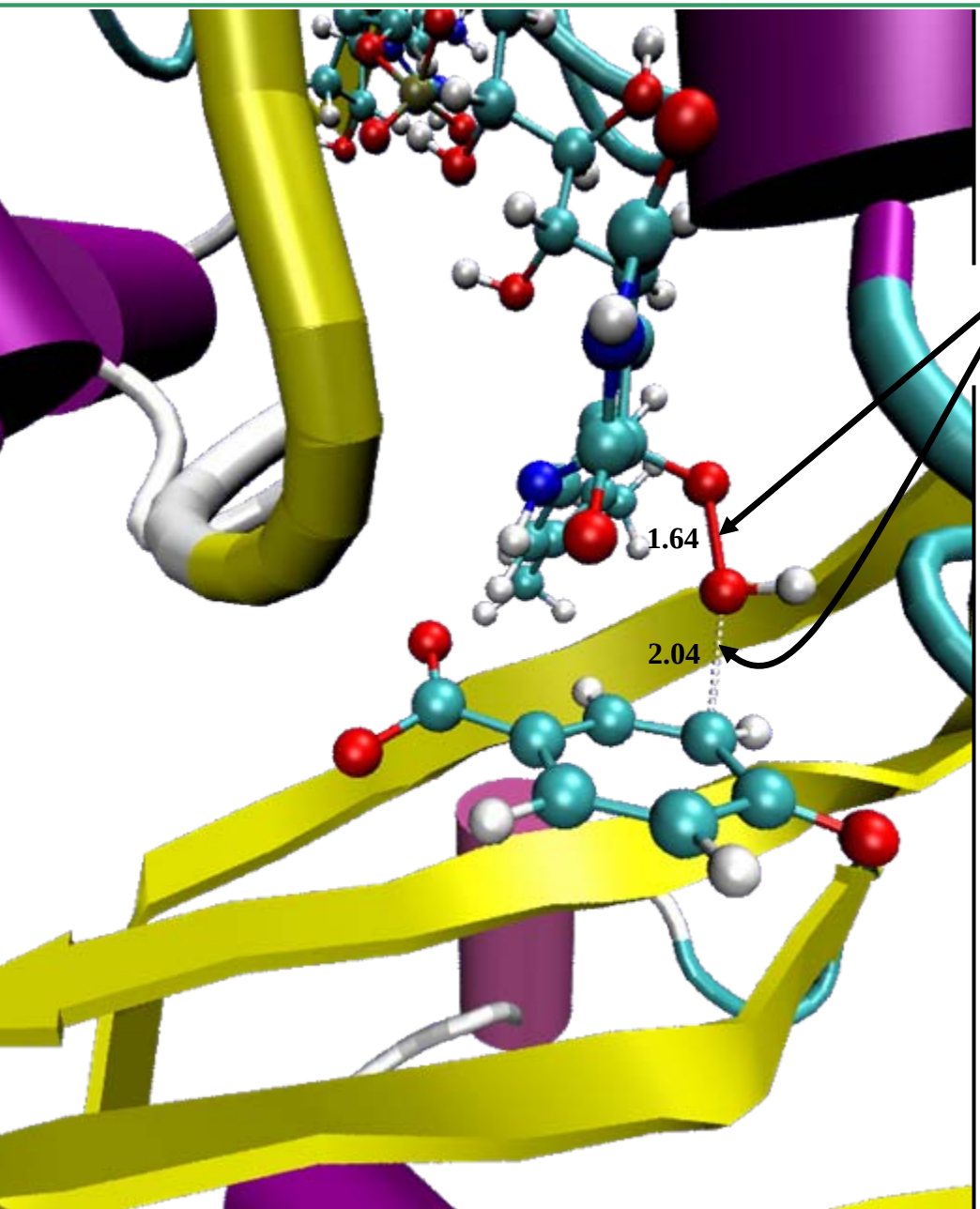
[1] H. M. Senn, S. Thiel and W. Thiel, J. Chem. Theory Comput. **1**, 494 (2005).

[2] J. Kästner and W. Thiel, J. Chem. Phys. **123**, 144104 (2005).

[3] J. Kästner, H. M. Senn, S. Thiel, N. Otte and W. Thiel, J. Chem. Theory Comp. **2**, 452 (2006).

[4] J. Kästner and W. Thiel, J. Chem. Phys. 124, 234106 (2006).

Thermodynamic integration



$$\Delta A_{BA} = A(\lambda_B) - A(\lambda_A) = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

- constrain difference of distances

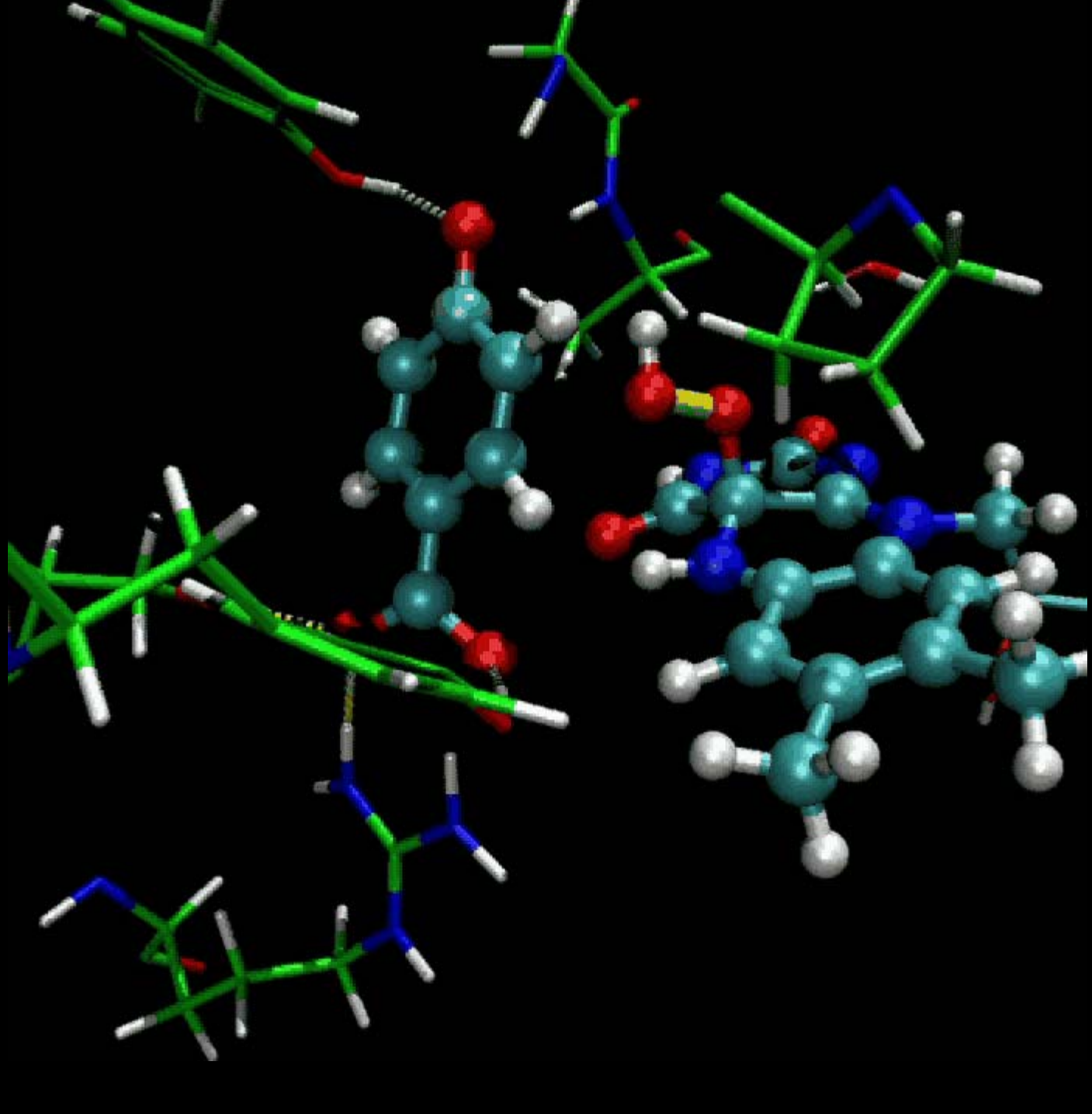
$$\lambda = d(\mathrm{O_d} - \mathrm{C_3}) - d(\mathrm{O_p} - \mathrm{O_d})$$

- implemented into SHAKE algorithm of DL_POLY

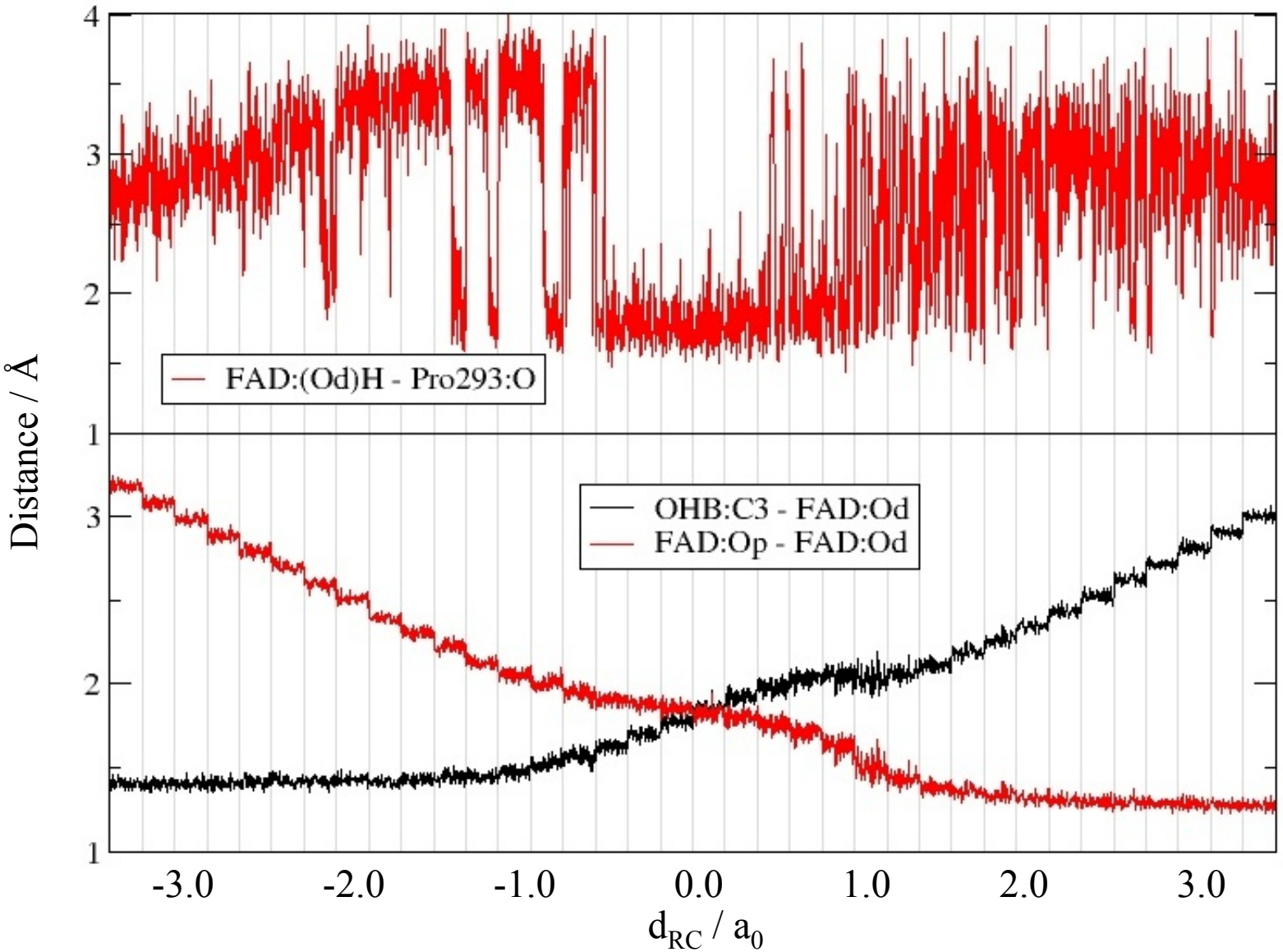
Start: optimised transition states,
optimised reactant states

- 5 ps QM/MM MD
  Berendsen thermostat (300 K)

- 35 ps QM/MM MD
  Nose Hoover thermostat (300 K)



PHBH : Role of Pro293



Chorismate mutase: Entropic contribution to barrier

- Experimental data available for *Bacillus subtilis* chorismate mutase [1]
- SCC-DFTB/CHARMM data from 10 forward and backward reaction paths using umbrella sampling and umbrella integration as implemented in ChemShell [2]

	$\Delta H^{\ddagger}$(kcal/mol)	$-T\Delta S^{\ddagger}$(kcalmol)	S(eu)
Experiment	12.7 ± 0.4	2.7 ± 0.4	-9.1 ± 1.2
QM/MM	6.6 ± 1.3	2.2 ± 0.5	-7.2 ± 1.2

- Note: Large spread in experimental entropies [1] for reaction in water, three different enzymes, and two different catalytic antibodies.
- Note: Related QM/MM free energy studies in other groups (e.g., in Bristol and Tsukuba).

[1] P. Kast, M. Asif-Ullah and D. Hilvert, Tetrahedron Lett. **37**, 2691 (1996).

[2] H. M. Senn, J. Kästner, J. Breidung and W. Thiel, Can. J. Chem. **87**, 1322 (2009).

QM/MM free energy barriers in enzymes

Methods applied:

- Semiempirical QM/MM-MD simulations
- DFT/MM free energy perturbation

Enzymatic reactions studied:

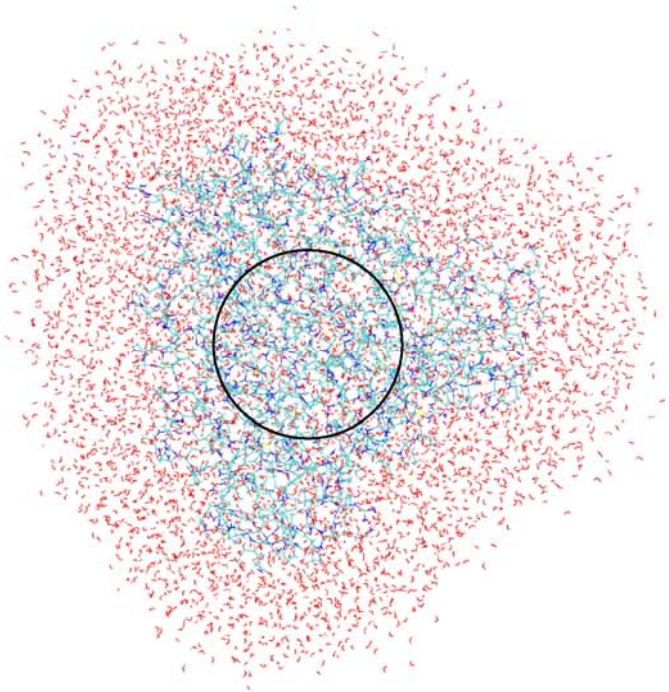
- p-Hydroxybenzoate hydroxylase, electrophilic substitution
- Fluorinase, nucleophilic substitution
- Cytochrome P450cam, hydrogen abstraction by Compound I
- Cystein protease, proton transfer involving His199/Cys29

Results:

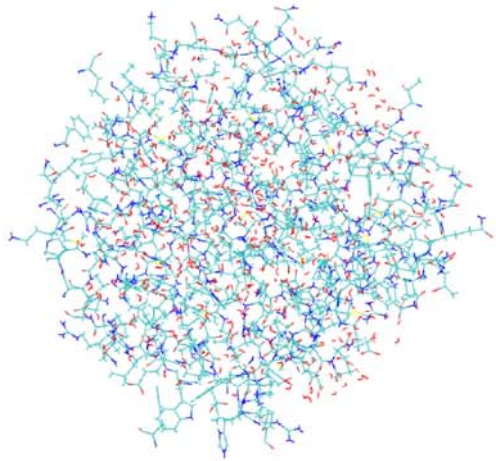
- Barriers and free energy barriers differ by less than 1 kcal/mol
- Similar reaction profiles

H. M. Senn, J. Kästner, J. Breidung and W. Thiel, Can. J. Chem. **87**, 1322 (2009).

Boundary potential approach

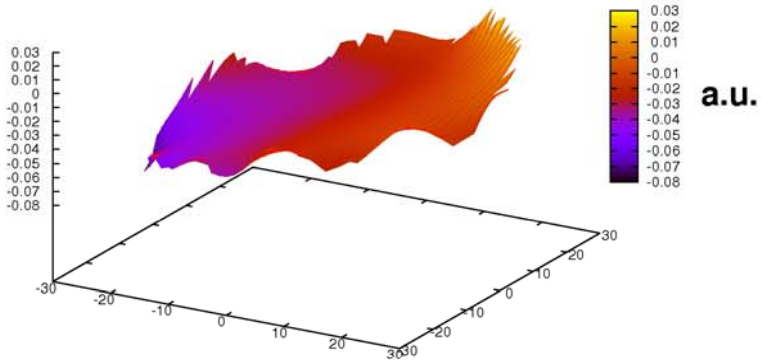


full solvated biomolecule



inner/explicit region

+



boundary potential

2D cut through explicit region

Generalized solvent boundary potential (GSBP)

- **Developed for classical MM simulations in 2001 [1]**
- **extended to QM/MM for SCC-DFTB in 2005 [2]**
- **several successful applications to biomolecular systems**
- **problematic only if macromolecule undergoes major conformational change**

➔ **ChemShell implementation for semiempirical QM/MM methods**

[1] W. Im, S. Bernéche, and B. Roux, J. Chem. Phys. **114**, 2924 (2001).
[2] P. Schaefer, D. Riccardi, and Q. Cui, J. Chem. Phys. **123**, 014905 (2005).

GSBP: Semiempirical QM/MM implementation

- **Implemented in ChemShell and MNDO2004**
- **QM charge density represented by Mulliken charges**
- **QM charge density interacts with GSBP during SCF procedure**
  - ➔ modifications of QM program

$$F_{\mu\nu}^{GSBP} = \frac{\partial \Delta W_{GSBP}}{\partial P_{\mu\nu}}$$

- **improved QM gradient**
  - ➔ coupled Mulliken charge derivatives
  - ➔ solution of CPSCF equations

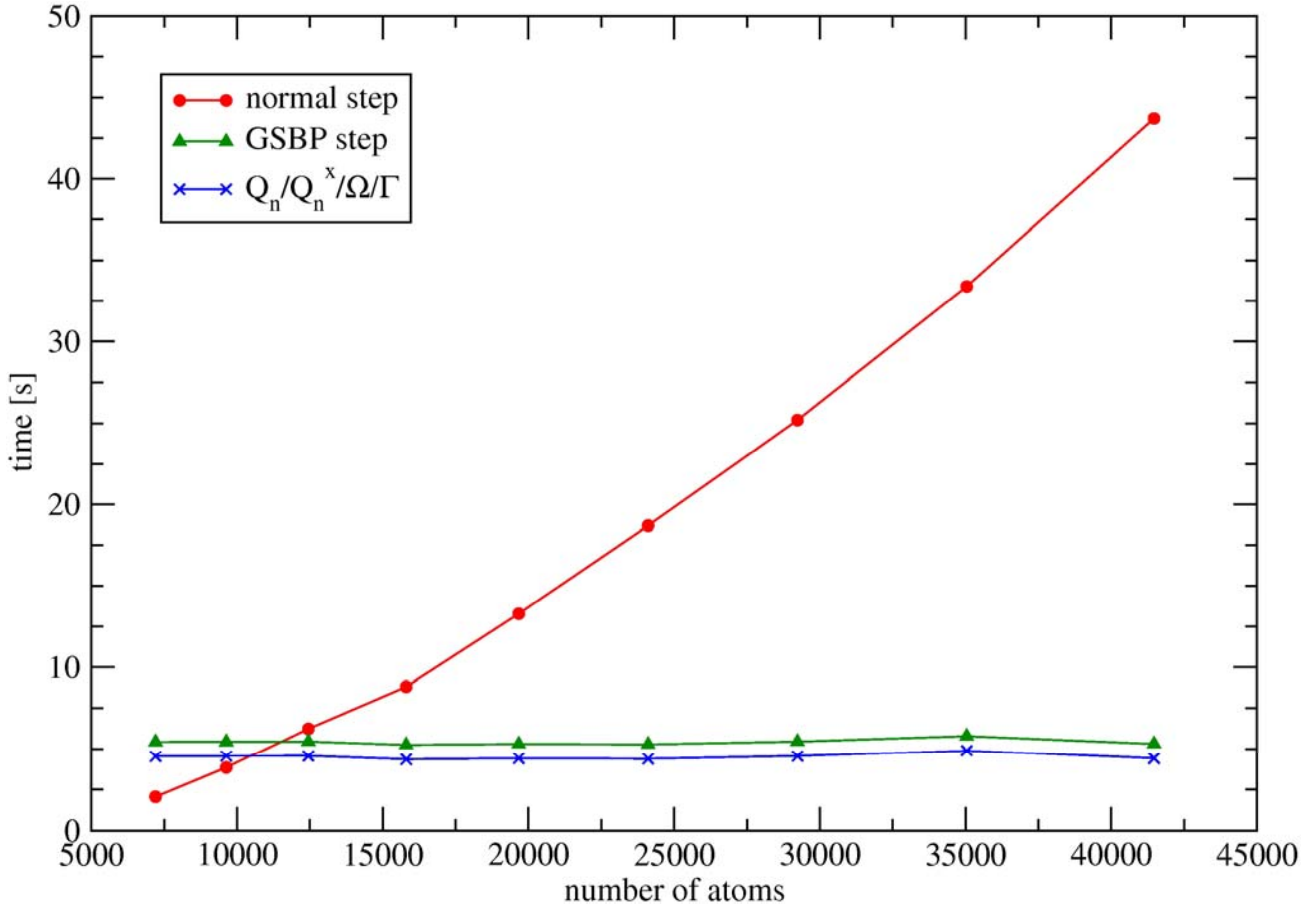
$$\frac{\partial q_A^{Mulliken}}{\partial r_B}$$

- **algorithmic improvements**
  - ➔ reduce overhead by 60%
- **tested accuracy for inherent parameters**

  set of optimal parameters

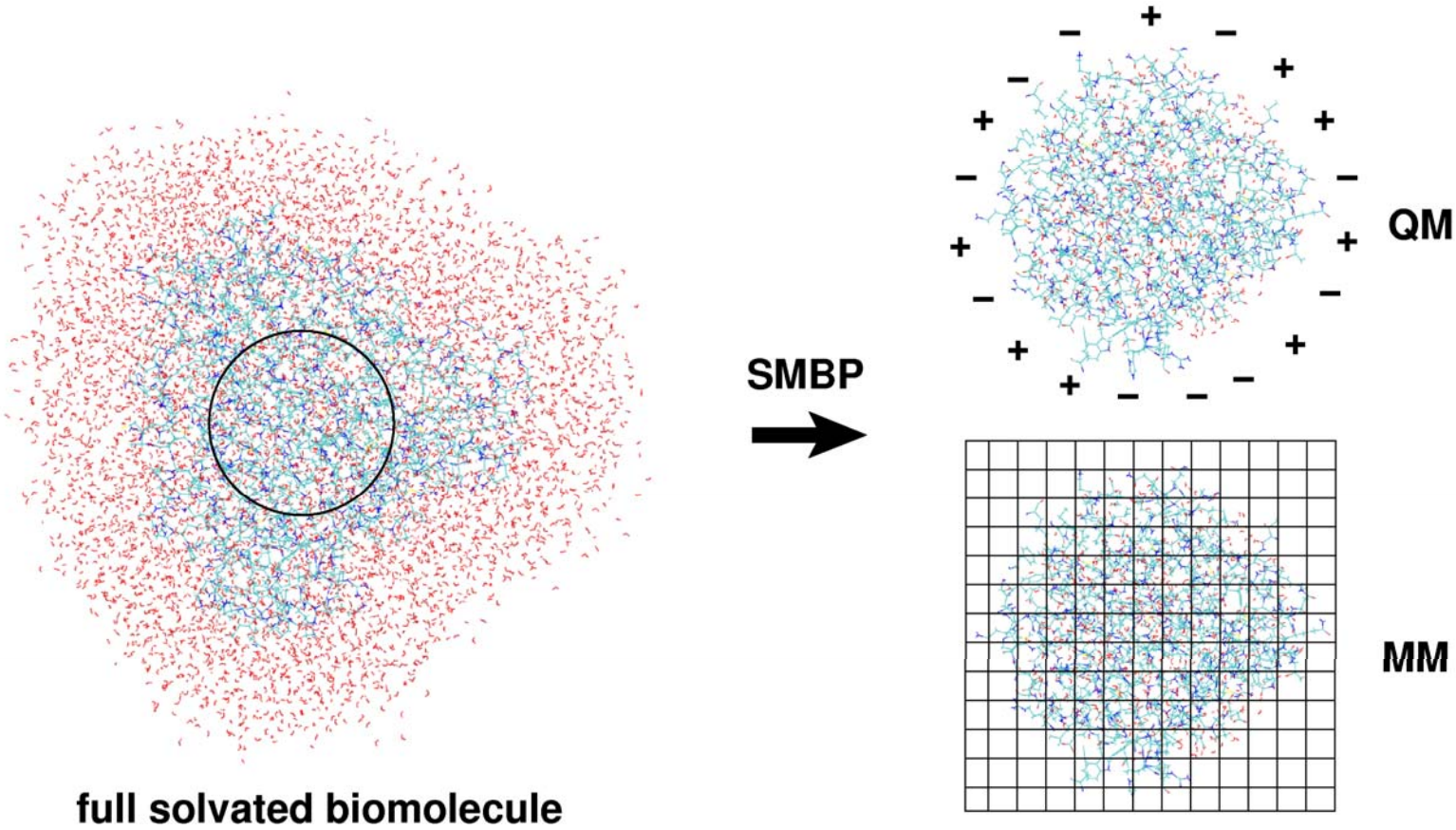
T. Benighaus and W. Thiel, J. Chem. Theory Comp. **4**, 1600 (2008).

GSBP: Threonine in water – computation times (MD)



- **GSBP not efficient for small systems**
- **breakeven point at 12 500 atoms**
- **impressive savings for large systems of 70% and more**

SMBP: Surface charge projection

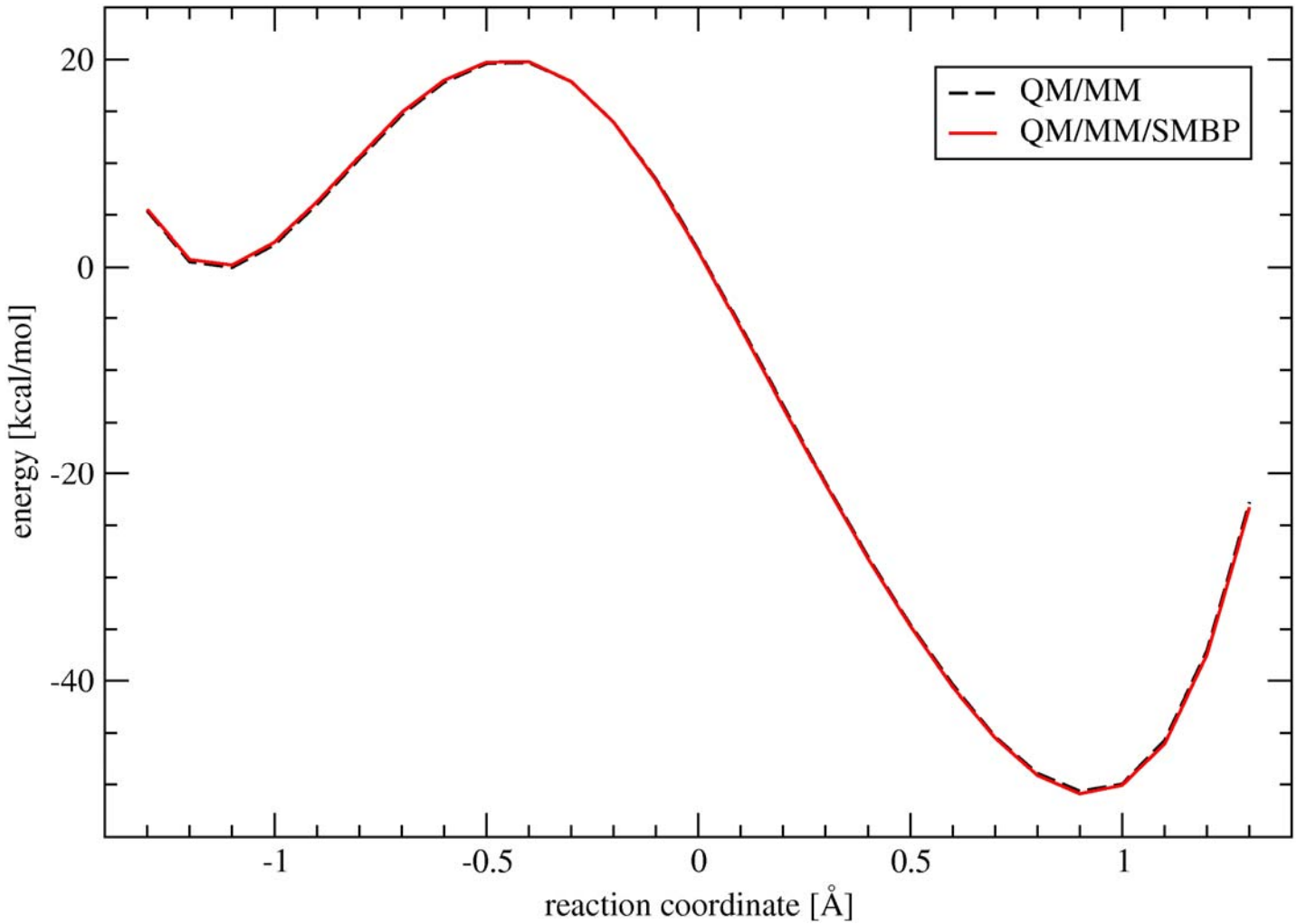


Surface charges are optimized to reproduce ϕ_{tot}^{QM} at the QM atoms

- allows use of any QM code → fits to modular ChemShell philosophy
- solution of CPSCF equations in gradient computation can be avoided

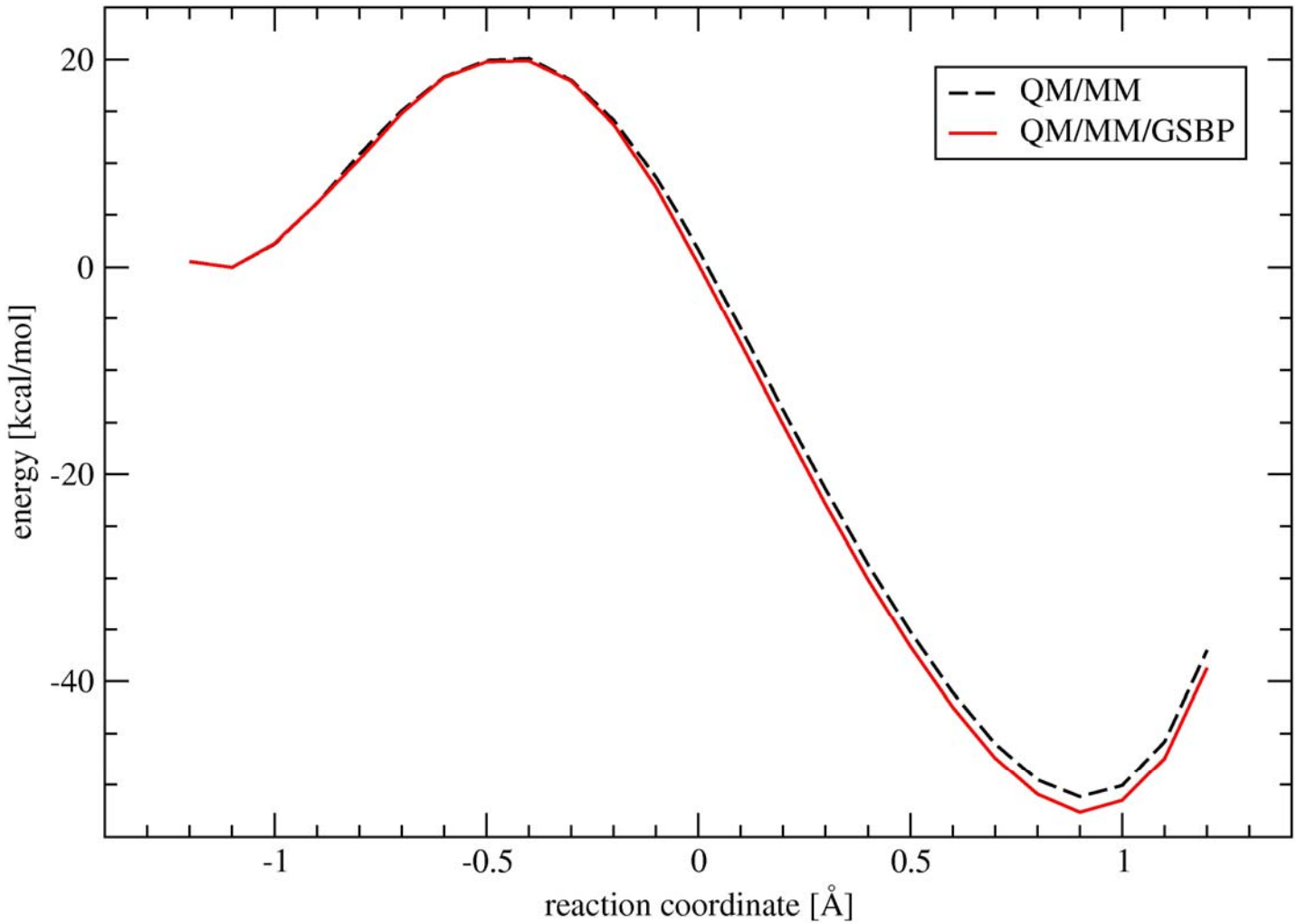
SMBP: PHBH – energy profiles

QM/MM and QM/MM/SMBP potential energy profiles



SMBP: PHBH – free energy profiles

QM/MM and QM/MM/GSBP free energy profiles from FEP calculations



SMBP: PHBH – computation times

Computation times per MD step [s] in FEP simulations

Module	QM/MM	QM/MM/GSBP
MM energy+gradient	69.2	2.6
QM energy+gradient	0.4	0.0
FEP	46.6	2.2
GSBP	-	0.1
Total	116.2	4.9

GSBP reduces computational costs of QM/MM-FEP calculations by at least one order of magnitude.

Conclusions

GSBP:

- **efficient and accurate method for QM/MM MD simulations**
- **currently restricted to semiempirical QM methods**

SMBP:

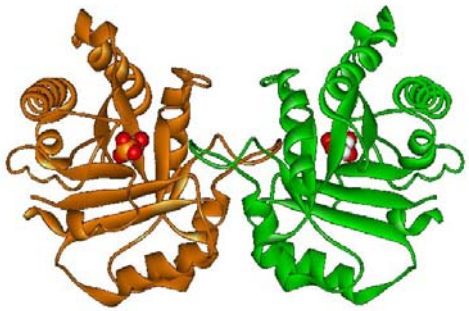
- **efficient method for QM/MM geometry optimizations**
- **accurate for relatively rigid systems (enzymes)**
- **problematic for flexible polar systems (aqueous solution)**
- **allows application of GSBP to sample over MM phase space in FEP calculations (speedup by at least one order of magnitude)**
- **reduces computational costs of standard QM/MM calculations**
- **available for all types of QM methods**

[1] T. Benighaus and W. Thiel, J. Chem. Theory Comput. **4**, 1600 (2008).
[2] T. Benighaus and W. Thiel, J. Chem. Theory Comput. , online, DOI: 10.1021/ct900437b.

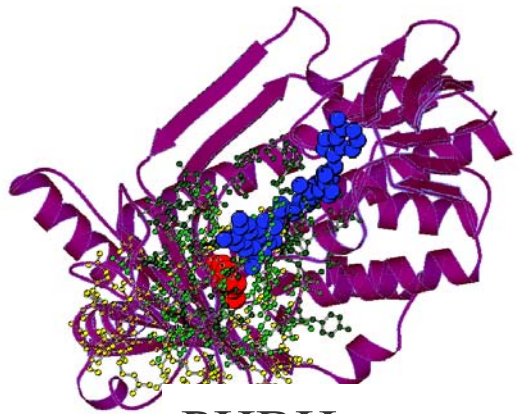
Recent QM/MM applications in our group

Focus on biocatalysis and enzymatic reactions:

- proton transfers in triosephosphate isomerase (TIM),
- oxygenation in p-hydroxybenzoate hydroxylase (PHBH),
- intermediates in the catalytic cycle of cytochrome P450,
- proton transfers and hydroxylation in cytochrome P450,
- enantioselective ester hydrolysis in lipases,
- mechanism of 4-oxalocrotonate tautomerase and synthetic analogues,
- Bergman cyclization of dynemicin A in DNA environment,
- enzymatic C—F bond formation by a fluorination enzyme,
- oxidation reactions of molybdopterin enzymes,
- deprotonation mechanism of the ammonium transporter AmtB.

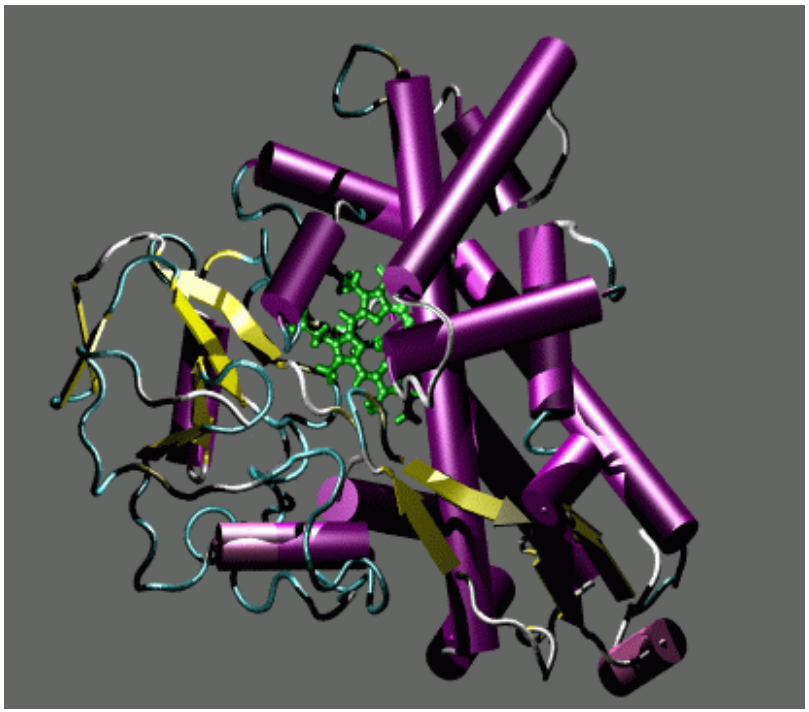


TIM
8331 atoms

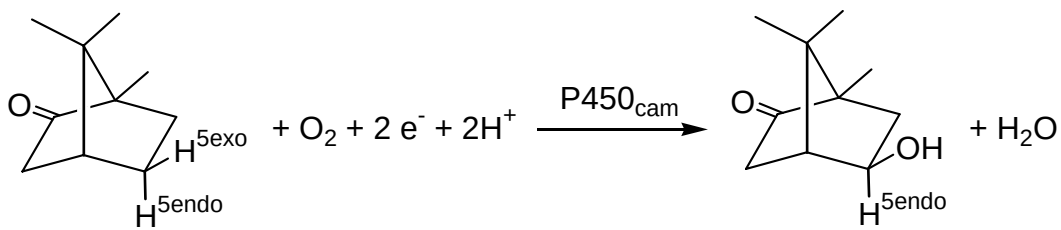


PHBH
7004 atoms

Cytochrome $P450_{Cam}$ (Pseudomonas Putida)

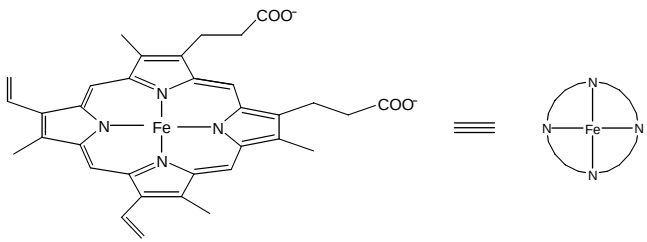


- heme protein, thiolato ligand
- completely buried active site
- soluble - extensively characterized by biochemical / biophysical techniques
- X-ray structures for various intermediates of the catalytic cycle
- natural substrate camphor, also other compounds
- biohydroxylation of non-activated C-H bonds



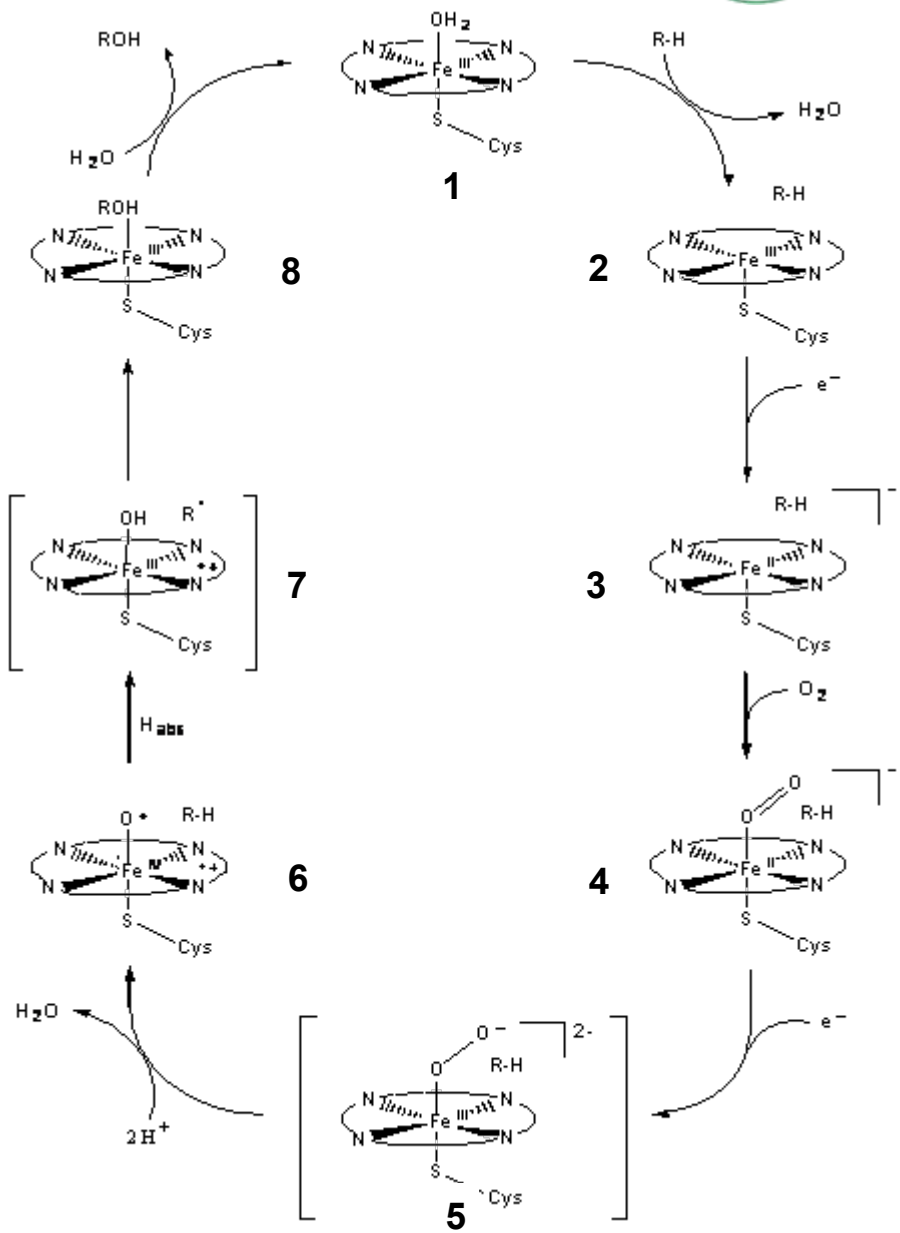
camphor

CYP450: Catalytic cycle



Mechanistic features:

- **electrons from NADPH (2 → 3, 4 → 5)**
- **binding of molecular oxygen (3 → 4)**
- **active species 6 (Compound I) not observed experimentally**
- **hydroxylation mechanism 6 → 8 under dispute (rebound mechanism assumed)**



Cytochrome P450cam: Overview over QM/MM results

Compound I contains a porphyrin cation radical because the protein environment stabilizes electronic charge at the sulfur ligand [1].

P450cam hydroxylation proceeds via a **two-state rebound mechanism** [2].

Inclusion of the protein environment is essential for reliable prediction of the **ESR and Mössbauer spectra** of P450cam Compound I [3].

Fe(V)-oxo species can be ruled out as P450cam intermediates [4], but low-lying **pentaradicaloids** could be involved in P450cam reactivity [3,4].

[1] J. C. Schöneboom et al, J. Am. Chem. Soc. **124**, 8142 (2002).

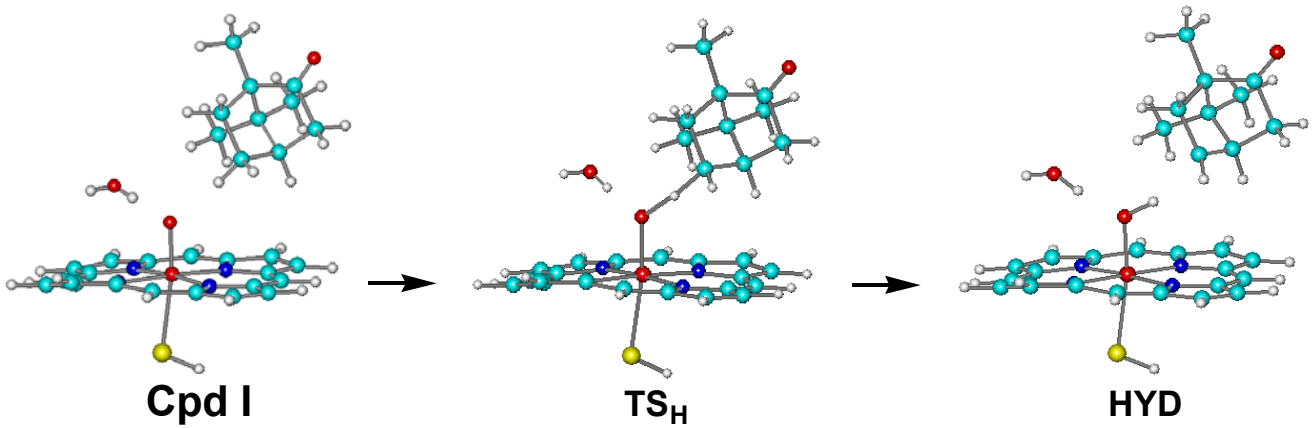
[2] J. C. Schöneboom et al, J. Am. Chem. Soc. **126**, 4017 (2004).

[3] J. C. Schöneboom et al, J. Am. Chem. Soc. **127**, 5840 (2005).

[4] A. Altun et al, J. Am. Chem. Soc. **129**, 8978 (2007).

Catalytic role of water903

- Water903 acts as a **catalyst** for hydrogen abstraction. The stabilization of the transition state arises from favorable electrostatic interactions in **hydrogen bonds** that are stronger in the transition state due to an increasing negative charge at the oxo atom. The computed barrier is lowered by 4 kcal/mol.

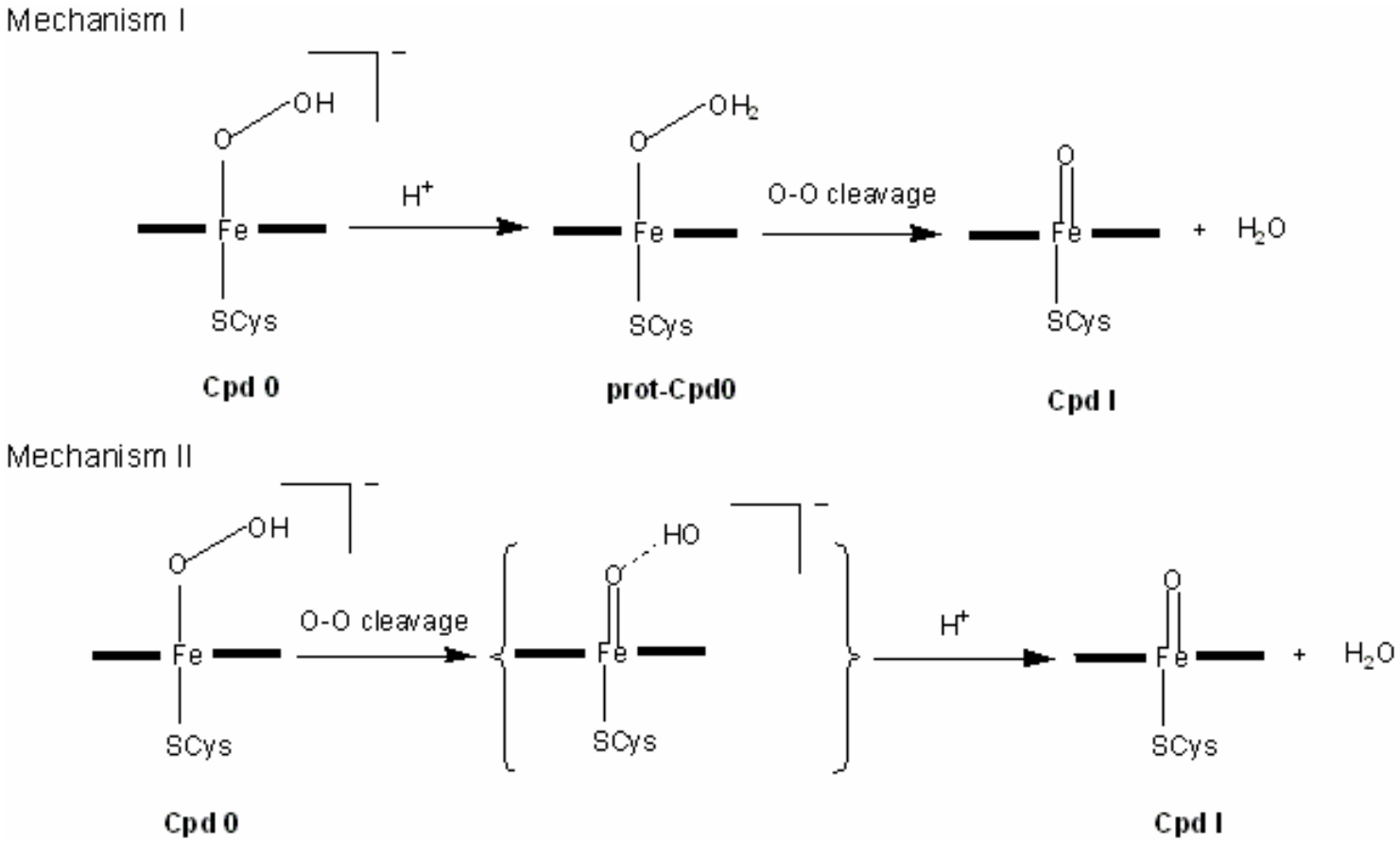


- One water molecule is liberated during the conversion of Cpd 0 to Cpd I :

$$Por(SR)FeOOH^- + H^+ \longrightarrow Por(SR)Fe{=}O + H_2O$$

A. Altun, V. Guallar, R. A. Friesner, S. Shaik, and W. Thiel, J. Am. Chem. Soc. **128**, 3924-3925 (2006).

Conversion of Cpd 0 to Cpd I

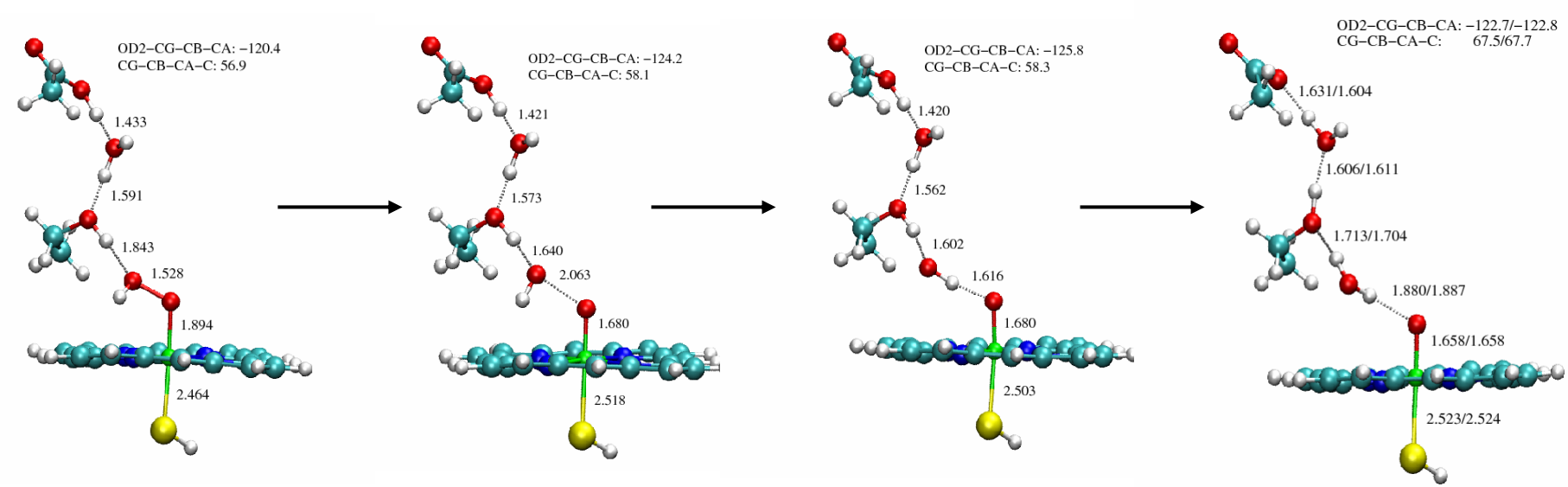


Glu366 channel: protonated Glu366 as proton source
Asp251 channel: protonated Asp251 as proton source

D. Kumar, S. P. de Visser, W. Thiel, S. Shaik et al., J. Phys. Chem. B **109**, 19946-19951 (2005).
J. Zheng, D. Wang, W. Thiel, and S. Shaik, J. Am Chem. Soc. **128**, 13204-13215 (2006).

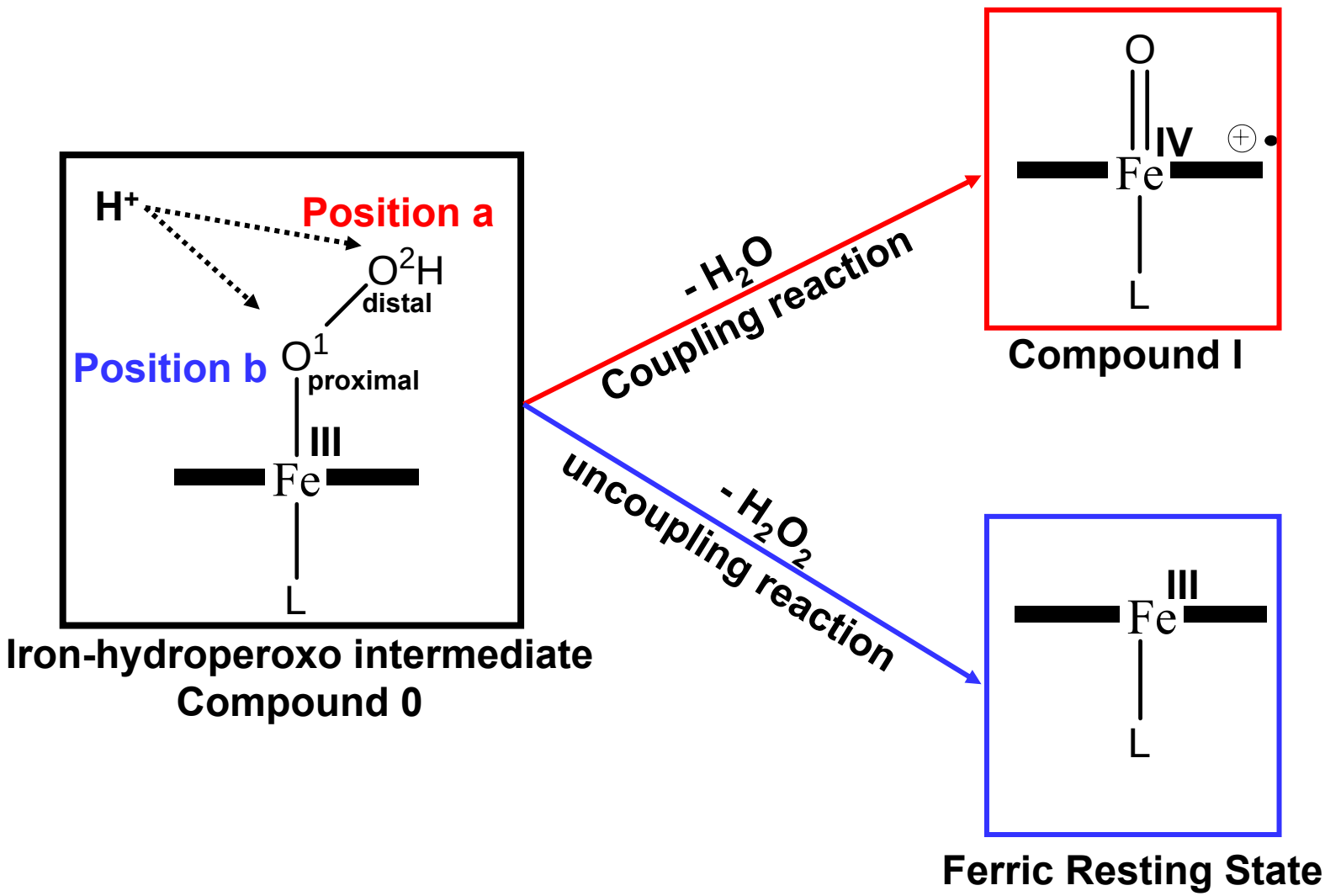
Asp-II mechanism: B3LYP/CHARMM study

Protein solvated by water, 24988 atoms, 62-atom QM region (DQ1) containing Por(SH)FeOOH, Asp251 (CH_3COOH), Thr252 (C_2H_5OH), and water901, extended 79-atom QM region (DQ3) also containing Arg186 ($C_2H_5NHC(NH2)_2$); UB3LYP/B1 data.

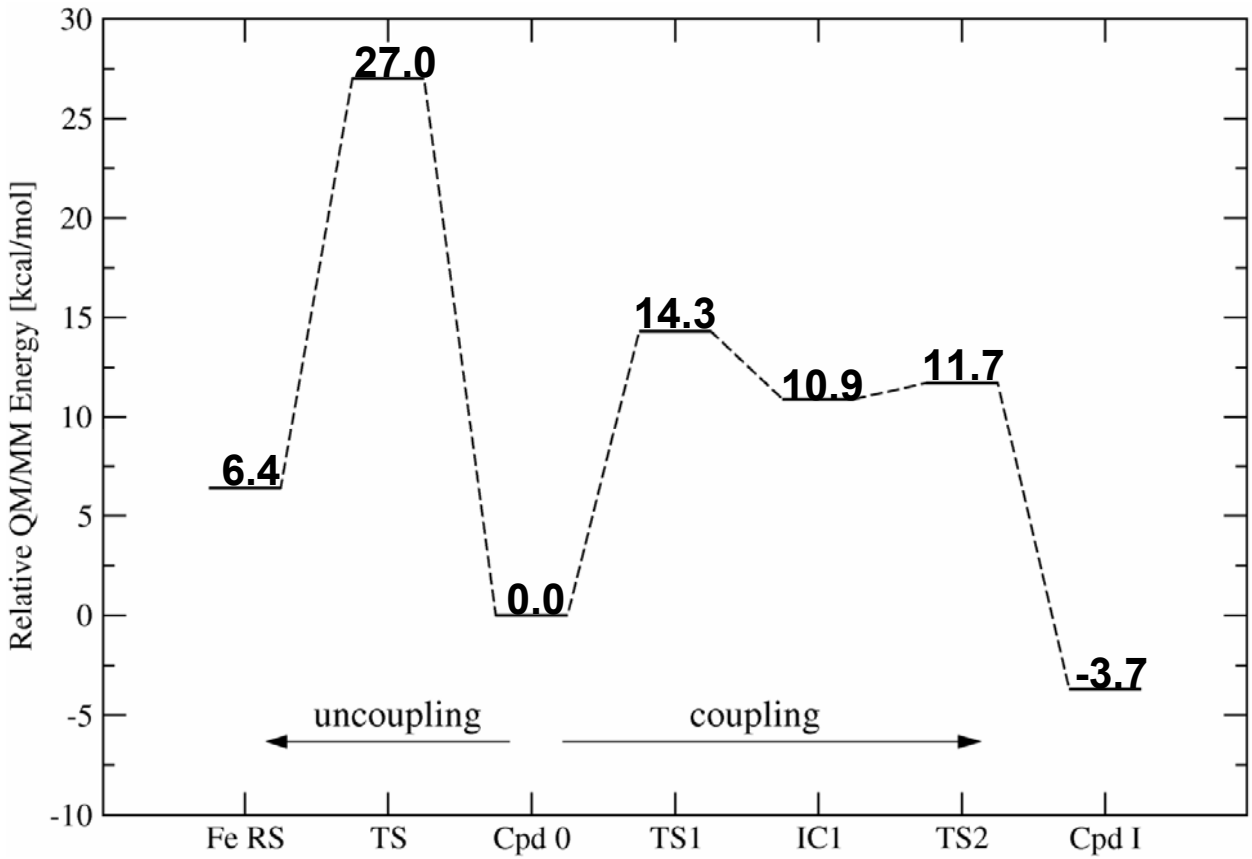


ΔE (kcal/mol)	**0.0**	**14.4**	**11.7**	**(14.2)**	**-0.2**
	Cpd 0	**TS1**	**IC1**	**(scan)**	**Cpd I**

Coupling and uncoupling reaction



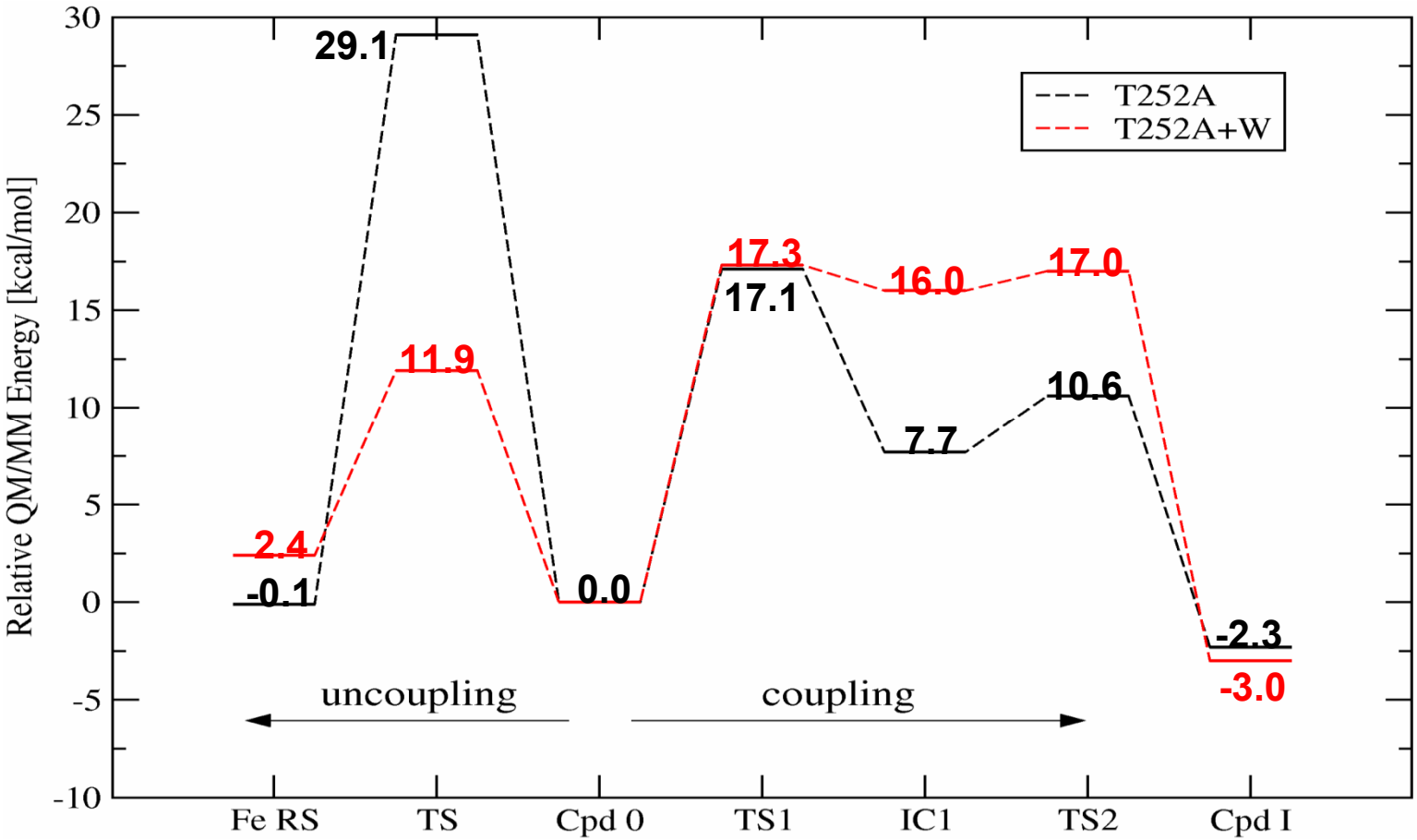
Coupling and uncoupling reactions - WT enzyme



- **The coupling reaction proceeds in two steps. The barrier for initial O-O bond cleavage is around 14 kcal/mol. The second step, proton transfer from Asp251 to OH, is facile (barrier of around 1 kcal/mol).**
- **The uncoupling reaction is concerted (barrier of 27 kcal/mol).**

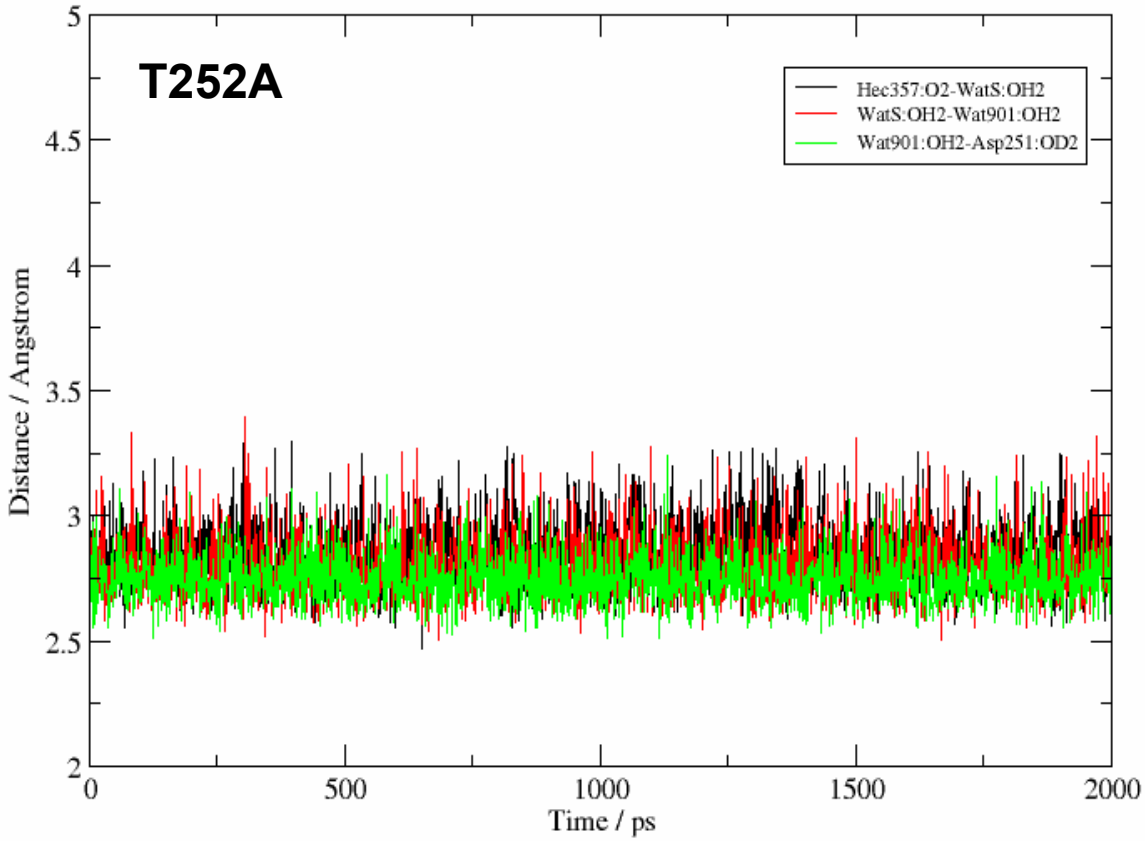
J. Zheng, D. Wang, W. Thiel, and S. Shaik, J. Am. Chem. Soc. **128**, 13204-13215 (2006).
M. Altarsha, T. Benighaus, D. Kumar, and W. Thiel, J. Am. Chem. Soc. **131**, 4755-4763 (2009).

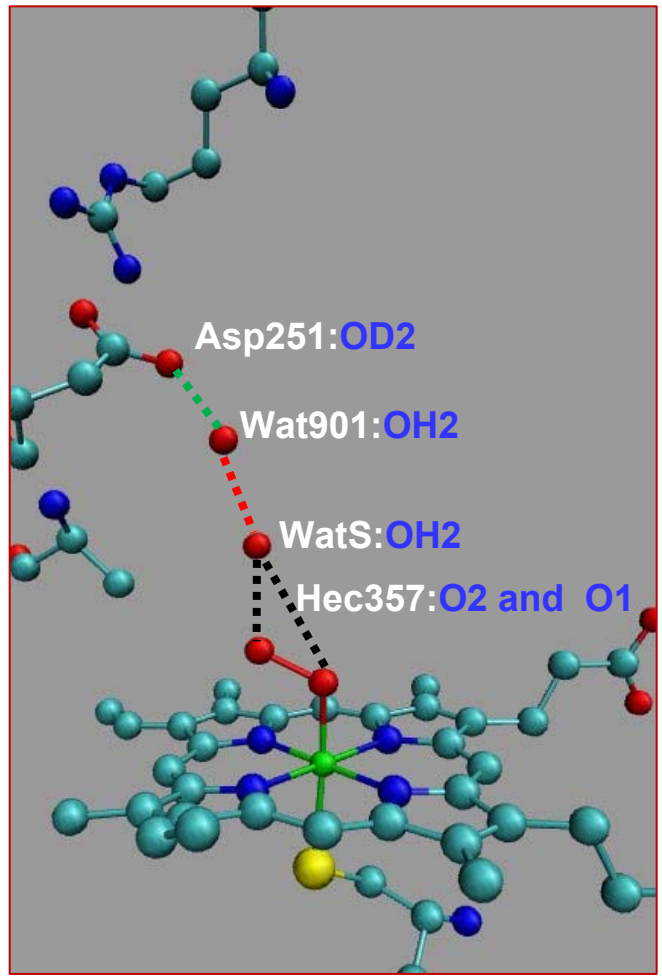
Coupling and uncoupling reactions T252A and T252A+W mutant



- The barrier for uncoupling is lower than that for coupling.
- WatS does not have much effect on the rate-limiting step of the coupling reaction.
- The presence of an extra water significantly reduces the barrier for the uncoupling reaction.

MD results for T252A: mobility of an extra water molecule





- Substitution of threonine by alanine generates some empty space in the distal pocket.
- Classical MD results for the alanine mutant suggest that Wat901 and WatS do not escape from the protein pocket during 2 ns simulation.

Rate limiting barriers

Mutant type	Coupling reaction	Uncoupling reaction
WT	14.3	27.0
T252S	15.6	23.1
T252V	17.1	26.5
T252V + W	18.9	19.5
T252A	17.1	29.1
T252A + W	17.3	11.9

- **If the effect of an additional water molecule is not taken into account, all mutants behave like the WT enzyme.**
- **Insertion of an extra water (WatS) has a significant mechanistic effect in the case of the T252V and T252A mutant.**

M. Altarsha, T. Benighaus, D. Kumar, and W. Thiel, J. Am. Chem. Soc. **131**, 4755-4763 (2009).

QM/MM: Status and perspectives

Methodology

- Methods and tools available (not yet black-box approach)
- Including high-level QM components
- Including sampling techniques

Essential for

- Treatment of electronic events in complex systems
- In particular: Chemical reactions
- In particular: Electronic excitation

Useful for

- Unbiased and reliable treatment of complex systems
- Example: Structural refinement
- Example: Ligand binding

Our research team in July 2004

Our research team in December 2007