QM/MM Methods

Gérald MONARD

Equipe de Chimie et Biochimie Théoriques
UMR 7565 CNRS - Université Henri Poincaré
Faculté des Sciences - B.P. 239
54506 Vandœuvre-les-Nancy Cedex - FRANCE

http://www.monard.info/

Nancy-Université
Université Henri Poincaré

cnrs

Outline

- QM/MM simulations of small solutes in solution: some illustrative examples
  - ✓ IR and VCD spectra of alanine dipeptide in water
  - ✓ Are current semiempirical methods better than force fields?
  - ✓ Coordination and ligand exchange dynamics of solvated metal ions
  - ✓ ONIOM-XS and Adaptative QM/MM simulations
- Reactive trajectories in QM/MM molecular dynamics
  - ✓ ethylene bromination in liquid water
  - ✓ formamide hydrolysis in liquid water
- Where's my proton?
  - ☞ glycine from water to CCl_4, and back
  - ☞ pKa prediction

Kwac et al. J. Chem. Phys. 2008, 128, 105106

THE JOURNAL OF CHEMICAL PHYSICS **128**, 105106 (2008)

Classical and quantum mechanical/molecular mechanical molecular dynamics simulations of alanine dipeptide in water: Comparisons with IR and vibrational circular dichroism spectra

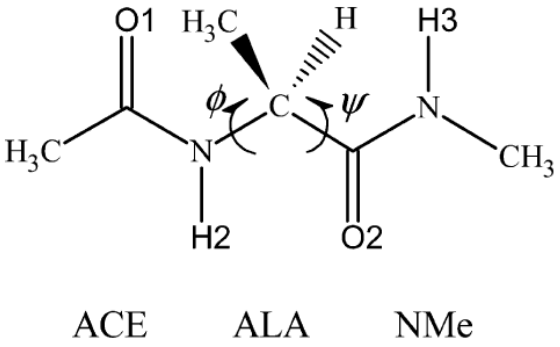
Kijeong Kwac,[1] Kyung-Koo Lee,[2] Jae Bum Han,[2] Kwang-Im Oh,[2] and Minhaeng Cho[2,3,a)]
[1]*Department of Chemistry and Biochemistry, University of Texas at Austin, 1 University Station A5300, Austin, Texas 78712, USA*
[2]*Department of Chemistry, Korea University, Seoul 136-701, Republic of Korea and Center for Multidimensional Spectroscopy, Korea University, Seoul 136-701, Republic of Korea*
[3]*Multidimensional Spectroscopy Laboratory, Korea Basic Science Institute, Seoul 136-713, Republic of Korea*

- ➢ Purpose: test full MM and QM/MM simulations with different QM and MM methods
- ➢ System: alanine dipeptide (Ace-Ala-Nme) in water
- ☞ backbone structure is fully determined by the two dihedral angles ϕ and ψ (Ramachandran plot)
- ➢ How: molecular dynamics + trajectory analysis
- ☞ IR adsorption spectra + Vibrational circular dichroism (VCD)

Kwac et al. J. Chem. Phys. 2008, 128, 105106

- Multiple conformations for alanine dipeptide has been suggested in water:
  - ✓ polyproline II (P_{II})
  - ✓ β-sheet
  - ✓ right-handed α-helix (α_R)

  or in gas phase:
  - ✓ C_5 or C_7: structures having an internal hydrogen bond



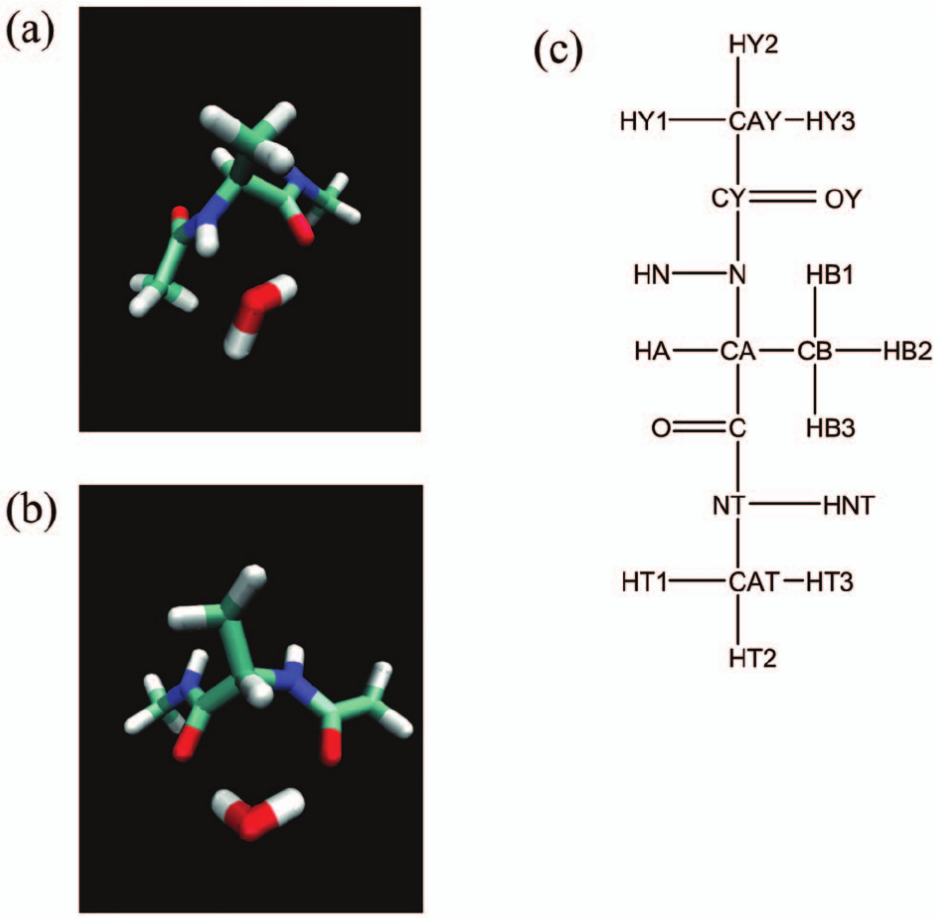


FIG. 1. (Color) (a) and (b) are two snapshot structures of the dipeptide obtained from MD simulation trajectories along with a water molecule which makes two hydrogen bonds with the dipeptide. (c) shows the names of atoms in the alanine dipeptide molecule used in this work.

Kwac et al. J. Chem. Phys. 2008, 128, 105106

Methodology

- 1 alanine dipeptide + 1461 water molecules in a cubic box
- Ten different simulations:

alanine dipeptide	water	alanine dipeptide	water
AM1	TIP3P	AMBER ff03	TIP3P
PM3	TIP3P	AMBER ff03	TIP4P
AMBER ff02	POL3	AMBER ff03	TIP5P
AMBER ff02	TIP3P	AMBER ff02EP	POL3
AMBER ff02	TIP4P	CHARMM CHEQ	

semi-empirical; polarizable and non-polarizable force field

Kwac et al. J. Chem. Phys. 2008, 128, 105106

Radial Distribution Functions (RDF)

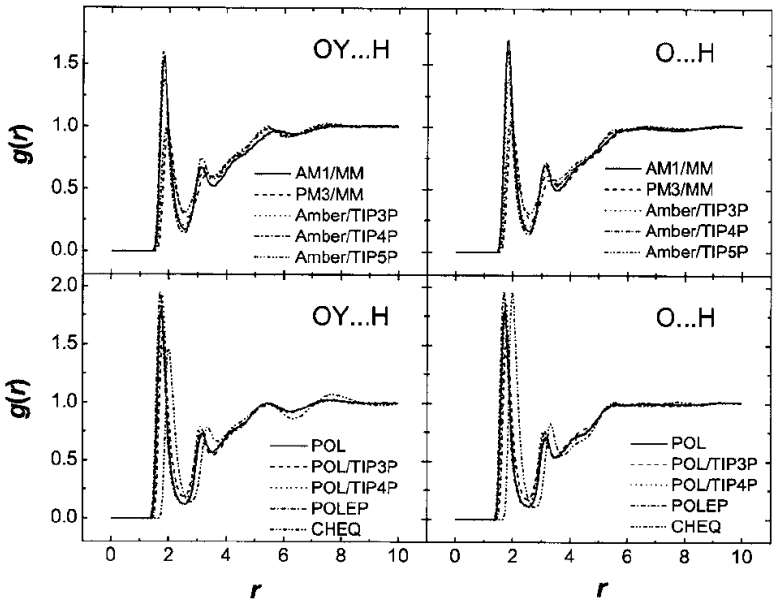


FIG. 2. Radial distribution functions between the carbonyl oxygen atoms and water hydrogen atoms.

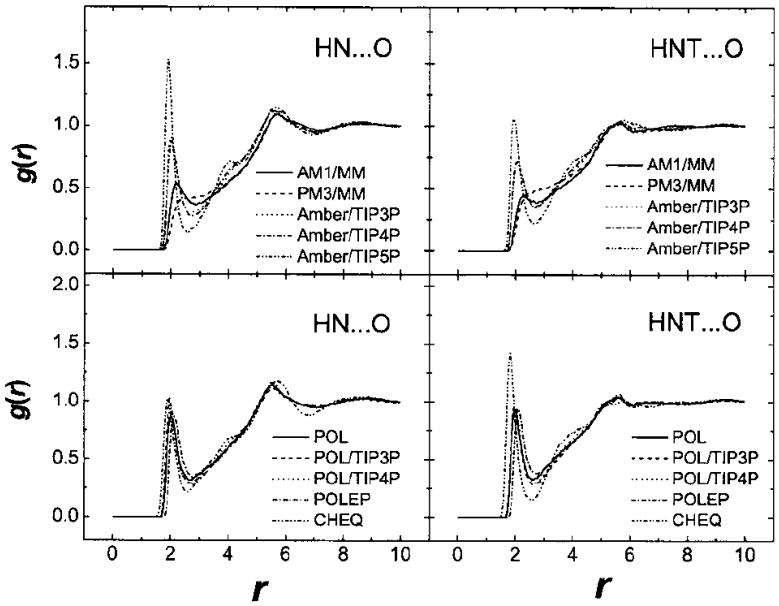
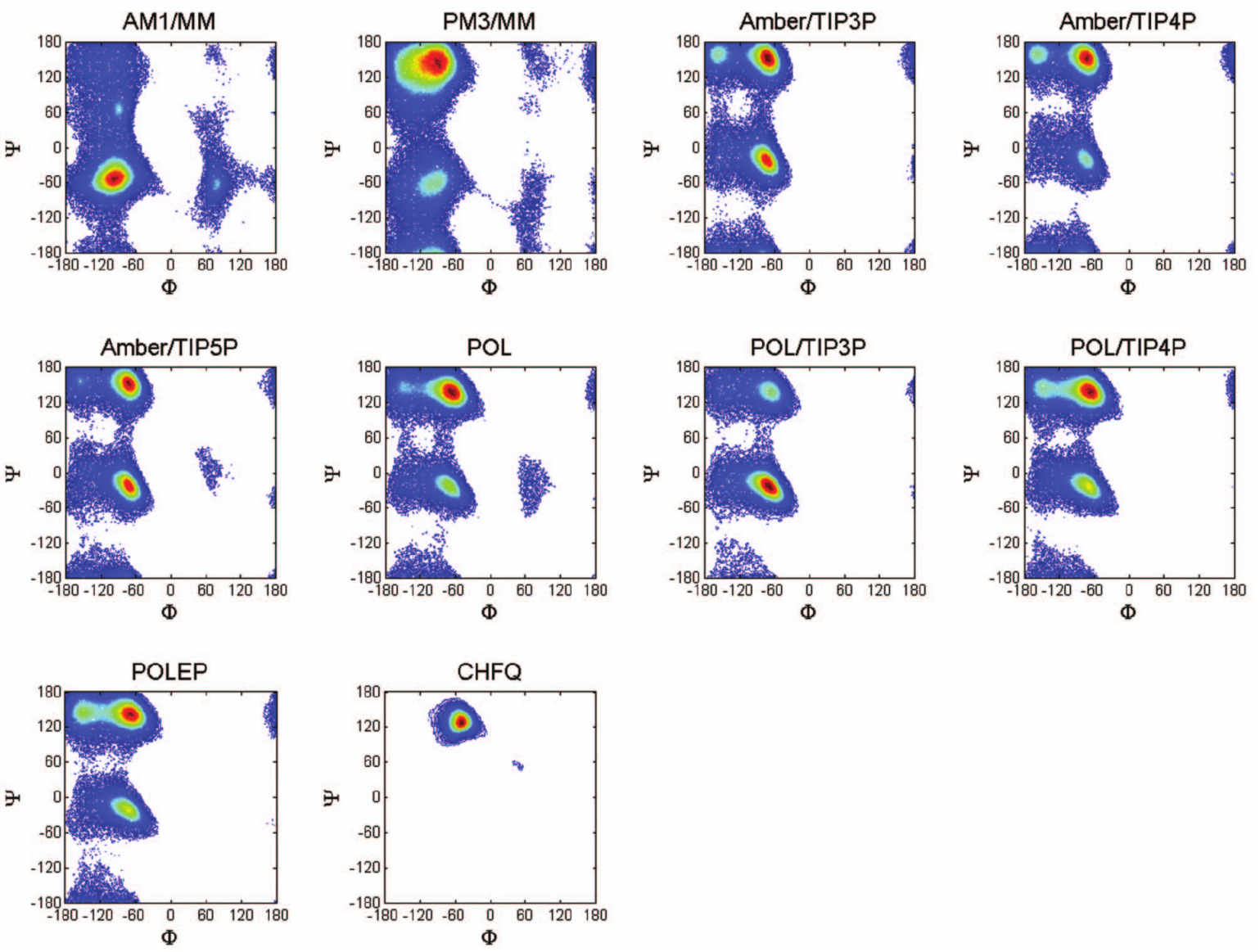


FIG. 3. Radial distribution functions between the amide hydrogen atoms and water oxygen atoms.

[. . .] the difference between the classical MD results and QM/MM MD results is not very conspicuous

[. . .] it is concluded that the solvation structure around the alanine dipeptide varies greatly depending on the force field and simulation methods used to describe the dipeptide molecule as well as the water model employed in the simulation.

Kwac et al. J. Chem. Phys. 2008, 128, 105106



. 4. (Color) Distributions of the two dihedral angles ϕ and ψ of the dipeptide in water for the QM/MM MD and classical MD simulations. (For color res, the population increases from blue to yellow to red.)

Kwac et al. J. Chem. Phys. 2008, 128, 105106

TABLE II. Percent population (P) of the four dihedral angle configurations and the value of the dihedral angle at each local maximum. The four areas on the dihedral angle space are considered and they represent α_R, P_{II}, β, and α_L conformations. The P_{II} conformation corresponds to the case when the dihedral angles of the alanine dipeptide are $-120° \leqslant \phi \leqslant -30°$ and $60° \leqslant \psi \leqslant 180°$. For β, $-180° \leqslant \phi \leqslant -120°$ and $60° \leqslant \psi \leqslant 180°$. For α_L, $0° \leqslant \phi \leqslant 150°$ and $0° \leqslant \psi \leqslant 180°$. For α_R configuration of the two QM/MM MD, $-180° \leqslant \phi \leqslant 0°$ and $-150° \leqslant \psi \leqslant 0°$. For α_R of the classical MD simulations, $-180° \leqslant \phi \leqslant 0°$ and $-115° \leqslant \psi \leqslant 60°$.

	P_{II}		β		α_R		α_L		IV	
	P (%)	$(\phi,\psi)_{max}$	P (%)	$(\phi,\psi)_{max}$	P (%)	$(\phi,\psi)_{max}$	P (%)	$(\phi,\psi)_{max}$	P (%)	$(\phi,\psi)_{max}$
AM1/MM	17.03	(−88, −68)	7.49	(−120, 138)	58.28	(−98, −54)	0.95	(74, 6)	7.86	(80, −64)
PM3/MM	42.44	(−88, 142)	25.82	(−124, 152)	23.89	(−90, −66)	0.36	(86, 134)	0.76	(68, −108)
AMBER/TIP3P	39.12	(−72, 152)	15.22	(−158, 158)	39.45	(−74, −24)	0.01	(144, 164)	⋯	⋯
AMBER/TIP4P	47.97	(−72, 154)	20.24	(−154, 162)	23.61	(−74, −18)	0.02	(148, 166)	⋯	⋯
AMBER/TIP5P	41.39	(−74, 154)	12.55	(−158, 162)	40.72	(−68, −24)	0.35	(64, 6)	0.24	(72, −4)
POL	52.89	(−66, 140)	16.37	(−142, 144)	27.32	(−70, −24)	0.25	(62, 0)	0.94	(74, −34)
POL/TIP3P	28.83	(−68, 136)	10.04	(−142, 144)	59.77	(−70, −20)	0.01	(0, 82)	0.01	(134, −28)
POL/TIP4P	48.99	(−66, 138)	18.81	(−150, 146)	30.41	(−70, −20)	⋯	⋯	⋯	⋯
POLEP	47.72	(−68, 142)	22.90	(−146, 144)	27.16	(−74, −22)	⋯	0	⋯	⋯
CHEQ	96.87	(−48, 126)	0.06	(−120, 146)	⋯	⋯	0.47	(40, 60)	⋯	⋯

[. . .] Except for the cases of AM1/MM and POL/TIP3P, the resultant histograms obtained from the PM3/MM and the other nonpolarizable and polarizable classical MD simulations are similar to each other: the most populated conformation corresponds to the extended structure (upper left region in the Ramachandran plot) rather than the helical conformation (lower left). Furthermore, the results fo QM/MM MD are drastically different between the AM1 and PM3 methods.

Kwac et al. J. Chem. Phys. 2008, 128, 105106

Distribution of electric dipole

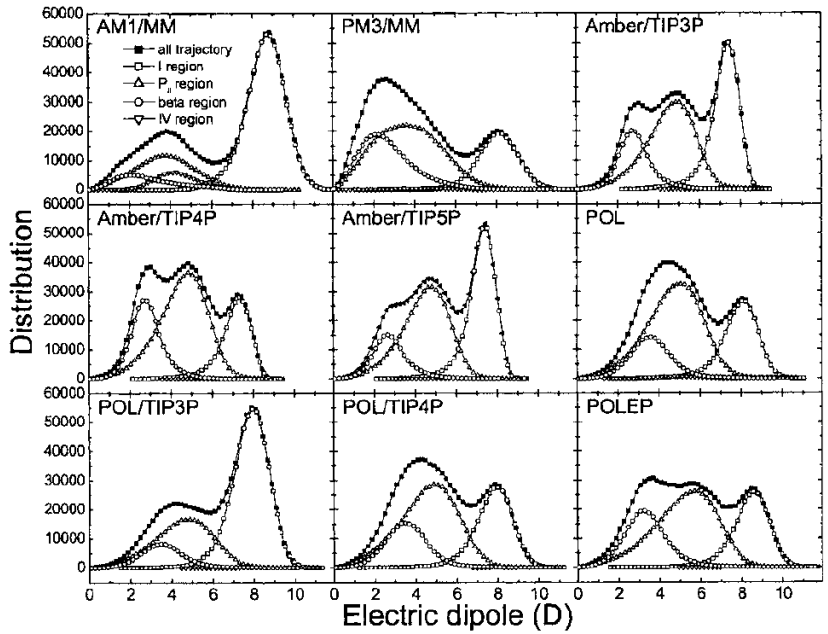


FIG. 7. Distributions of the permanent dipole moment magnitude of the dipeptide.

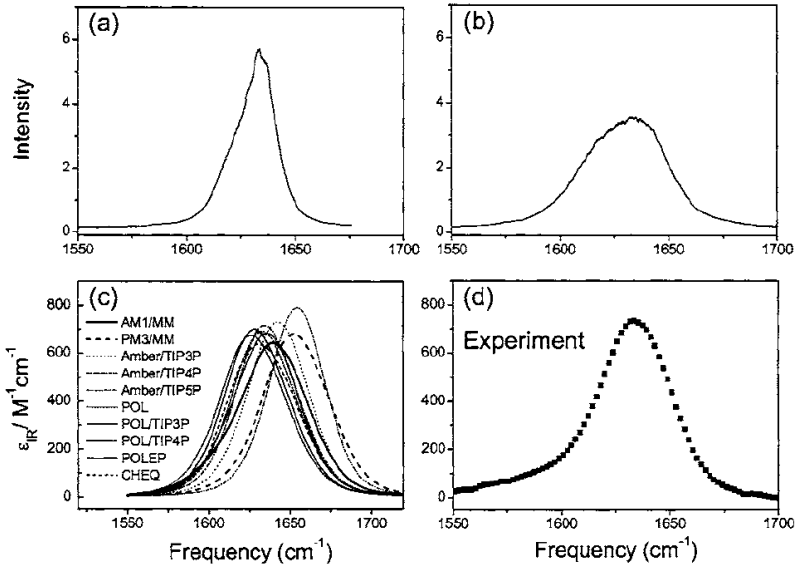
- *the distributions have 2 or 3 peaks, representing different dipeptide conformations*
- $\mu(\alpha_R) \sim 8\ D$; $\mu(P_{II}) \sim 3\ D$; $\mu(\beta) \sim 5\ D$

... *it is concluded that the dipole moment alone cannot explain the preference of the dihedral angle conformation*

... *the permanent point dipole-solvent polarization interaction is not the determining factor for stabilizing a particular dipeptide conformation over the others.*

Kwac et al. J. Chem. Phys. 2008, 128, 105106

IR and VCD spectra



IG. 8. IR absorption spectra of the dipeptide in water calculated from (a) AM1/MM MD using Eq. (1), (b) PM3/MM MD using Eq. (1), and (c) all en simulations using Eq. (5). The experimentally measured amide I IR pectrum is shown in (d).

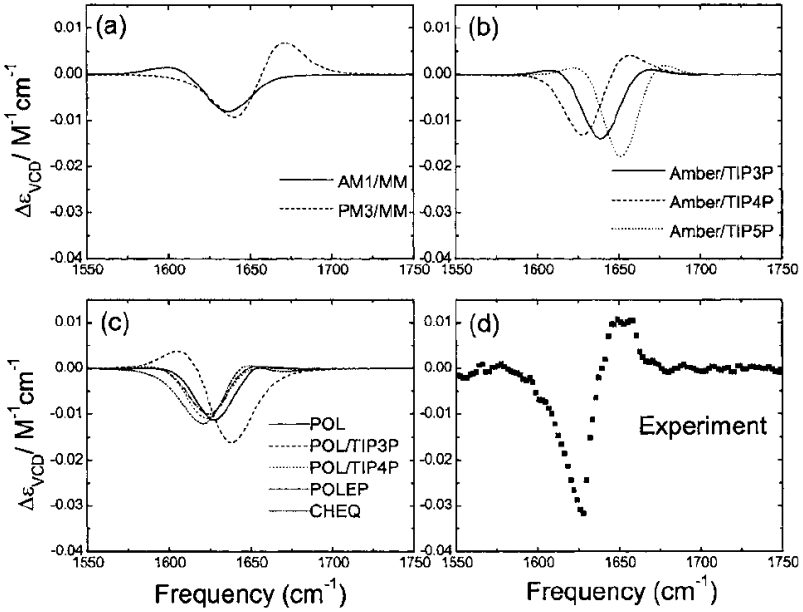


FIG. 9. VCD spectra of the dipeptide calculated by using (a) QM/MM MD, (b) nonpolarizable classical MD, and (c) polarizable classical MD simulation trajectories. The experimentally measured amide I VCD spectrum is shown in (d).

[the simulation and experimental spectra] are broad and featureless so that comparisons of simulated IR spectra with experiment do not provide critical information on which force field calculationsis better than the others.

Kwac et al. J. Chem. Phys. 2008, 128, 105106

IR and VCD spectra

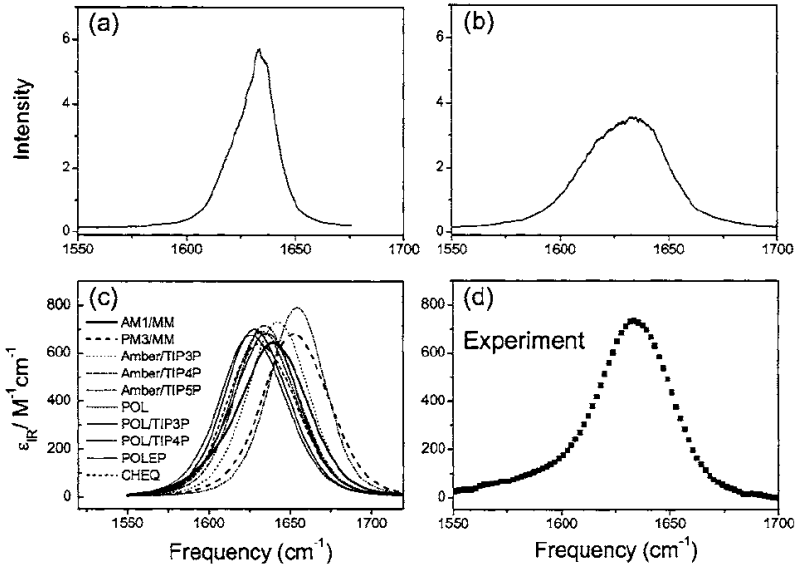


FIG. 8. IR absorption spectra of the dipeptide in water calculated from (a) AM1/MM MD using Eq. (1), (b) PM3/MM MD using Eq. (1), and (c) all atom simulations using Eq. (5). The experimentally measured amide I IR spectrum is shown in (d).

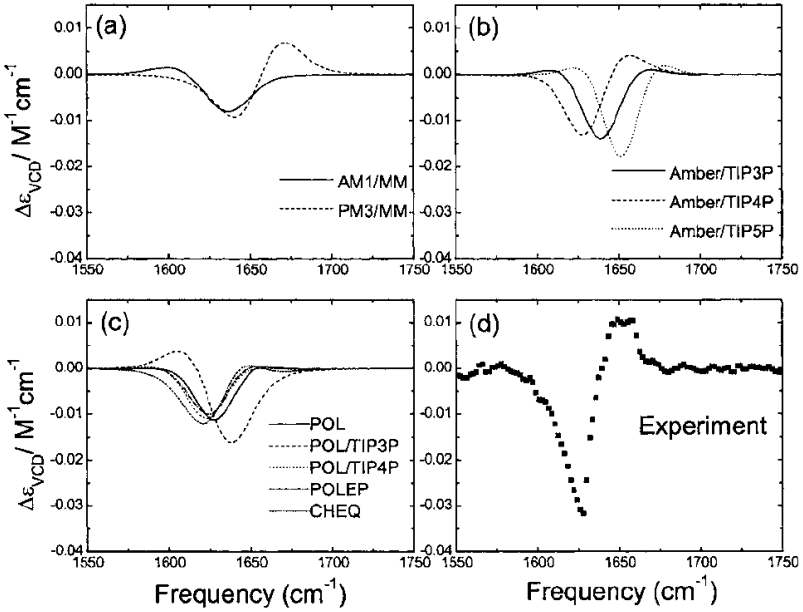


FIG. 9. VCD spectra of the dipeptide calculated by using (a) QM/MM MD, (b) nonpolarizable classical MD, and (c) polarizable classical MD simulation trajectories. The experimentally measured amide I VCD spectrum is shown in (d).

[The VCD xperimental result] shows the negative and positive peaks from low- to high-frequency region. This indicates that the dipeptide structure ressembles the P_{II} conformation. This negative-positive feature is well reproduced by the PM3/MM MD result and the classical MD result with the AMBER/TIP4P force field.

Kwac et al. J. Chem. Phys. 2008, 128, 105106

Some conclusions by the authors...

- *One can conclude that the water model is critical in not only properly determining the dipeptide structure but also correctly simulating the vibrational spectra.*
- *we reached a conclusion that the dipeptide solution structure is close to P_{II} and only those force fields and simulation methods that predict large population of P_{II} are acceptable and reproduce experimental VCD spectrum correctly.*
- *PM3/MM MD method [compared to AM1/MM] is in better agreement with experiment.*
- *the QM/MM MD simulation [...] will be of useful techniques for simulating various linear and nonlinear vibrational spectra in the future, but which QM method is chosen critically determines the dipeptide structure.*

Seabra et al. J. Phys. Chem. A 2009, 113, 11938–11948

11938 *J. Phys. Chem. A* **2009,** *113,* 11938–11948

Are Current Semiempirical Methods Better Than Force Fields? A Study from the Thermodynamics Perspective[†]

Gustavo de M. Seabra,[‡] **Ross C. Walker,**[§] **and Adrian E. Roitberg***[,‡]

Quantum Theory Project and Department of Chemistry, University of Florida, 2234 New Physics Building #92, P.O. Box 118435, Gainesville, Florida 32611-8435, and San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive #0505, La Jolla, California 92093-0505

- ➢ Purpose: test full MM and QM/MM simulations with different semiempirical QM methods
- ➢ System: alanine dipeptide (Ace-Ala-Nme) in water
- ➢ How: molecular dynamics + trajectory analysis
- ☞ free energy surface in the (ϕ, ψ) dihedral angle space
- ☞ $^3J(\mathrm{H}_N,\mathrm{H}_\alpha)$ NMR dipolar coupling constants
- ☞ basin populations
- ☞ peptide-water radial distribution functions

Seabra et al. J. Phys. Chem. A 2009, 113, 11938–11948

Methodology

- 1 alanine dipeptide + 929 TIP3P water molecules
- SE QM: MNDO, AM1, PM3, RM1, PM3/PDDG, MNDO/PDDG + SCC-DFTB (Second-order Self-Consistent-Charge Density Functional Tight Binding)
- MM: Amber force fields (ff94, ff99, ff99SB, ff03)
- Replica Exchange Molecular Dynamics for conformational samplings (32 replicas)
- Dipolar coupling constants obtained from the Karplus relation:

$$^{3}J(H_N, H_\alpha) = a\cos^2(\phi - 60^{\circ}) + b\cos(\phi - 60^{\circ}) + c$$

- Free energy profiles obtained by calculating the (normalized) probability P of finding the alanine dipeptide in a conformation at a particular region ($\Delta G = -RT\ln(P)$)

Seabra et al. J. Phys. Chem. A 2009, 113, 11938–11948

Some experimental results from bibliography

- P_{II} basin is the most populated one
- α-region sampling become significant only for larger peptides
- P_{II} basin population between 60-76%

TABLE 2: $^3J(H_N,H_\alpha)$ NMR Dipolar Couplings for Alanine Dipeptide, in Hz, Calculated as an Average from the 6 ns of MD Simulation[a]

method	$^3J(H_N,H_\alpha)$
ff94	6.20 ± 0.08
ff99	7.80 ± 0.07
ff03	6.69 ± 0.08
ff99sb	7.35 ± 0.08
MNDO	7.67 ± 0.07
AM1	8.25 ± 0.07
PM3	8.14 ± 0.07
RM1	6.77 ± 0.09
PDDG/MNDO	7.76 ± 0.07
PDDG/PM3 (2002)	7.85 ± 0.07
PDDG/PM3 (2008)	8.06 ± 0.07
PM3 + MM correction	8.24 ± 0.07
SCC-DFTB	8.07 ± 0.08
SCC-DFTB + dispersion	8.16 ± 0.08
exp[59]	**6.06 ± 0.05**

[a] The error margin is shown as the 95% confidence intervals calculated using the Student's *t*-value for an infinite number of measurements. The experimental error margin is an estimate based on different parameterizations of the Karplus equation.[59]

TABLE 3: Conformational Distribution of Alanine Dipeptide, Shown as Fractional Populations of the Different Conformational Basins

method	α	β	PP_{II}	other
ff94	0.84	0.04	0.11	0.01
ff99	0.91	0.03	0.01	0.05
ff03	0.45	0.19	0.35	0.01
ff99SB	0.32	0.24	0.40	0.04
MNDO	0.07	0.27	0.63	0.03
AM1	0.57	0.19	0.21	0.04
PM3	0.14	0.51	0.33	0.02
RM1	0.26	0.17	0.53	0.04
PDDG/MNDO	0.33	0.15	0.52	0.01
PDDG/PM3 (2002)	0.16	0.40	0.41	0.02
PDDG/PM3 (2008)	0.08	0.43	0.48	0.01
PM3 + MM correction	0.19	0.38	0.42	0.02
SCC-DFTB	0.40	0.34	0.19	0.06
SCC-DFTB/dispersion	0.48	0.30	0.18	0.05
IR[103]	**0.11**	**0.29**	**0.60**	**0.00**
Raman[103]	**0.18**	**0.06**	**0.76**	**0.00**

Seabra et al. J. Phys. Chem. A 2009, 113, 11938–11948

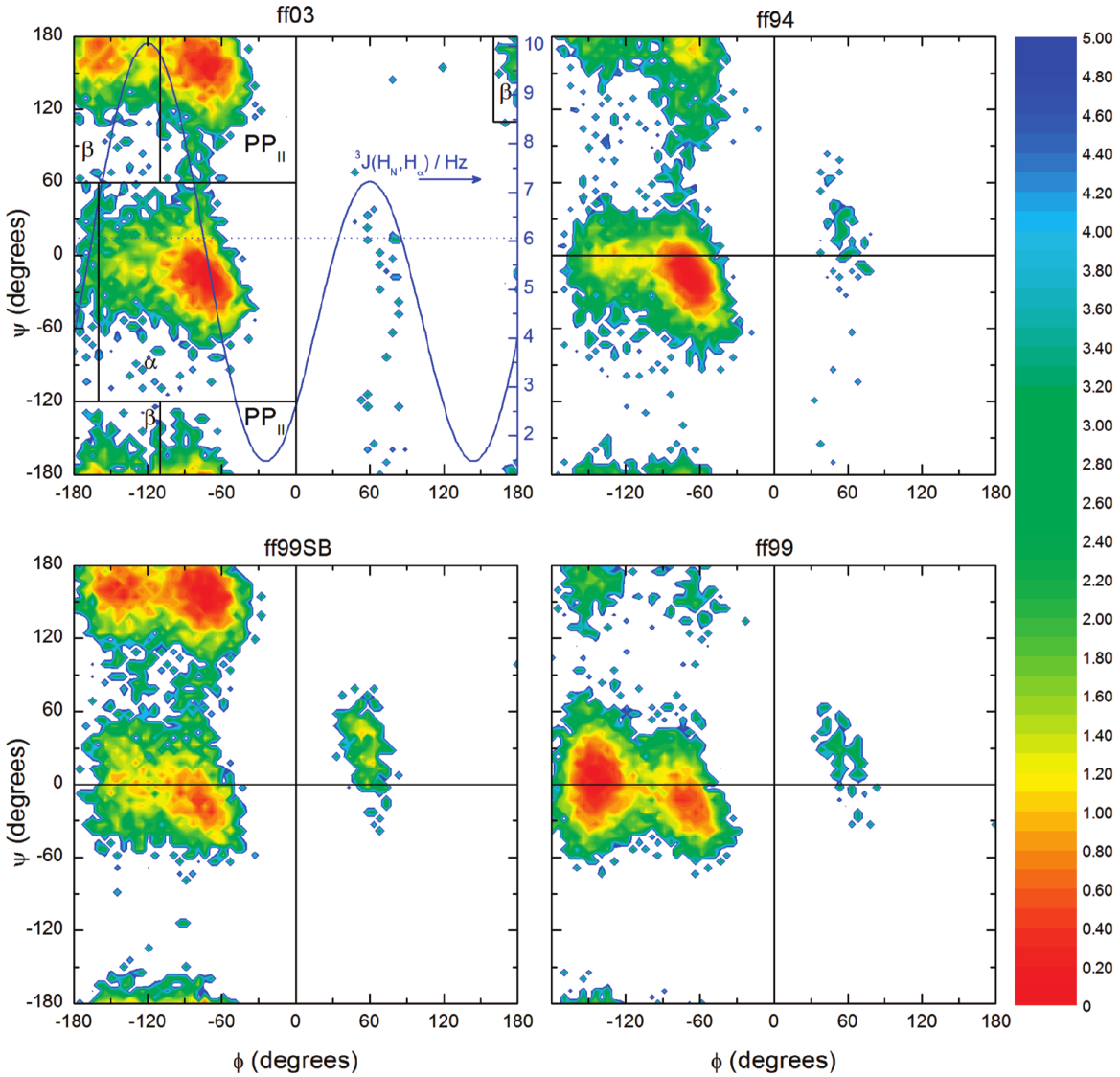


Figure 2. Free energy surfaces obtained from the 300 K replicas of the REMD simulations, for the different force field methods. The graph at the upper left also depicts the basin divisions used in this work for population analysis, and a plot of eq 1 in blue, linked to the right *y*-axis. The dashed red line indicates the experimental dipolar coupling constant.[59]

Seabra et al. J. Phys. Chem. A 2009, 113, 11938–11948

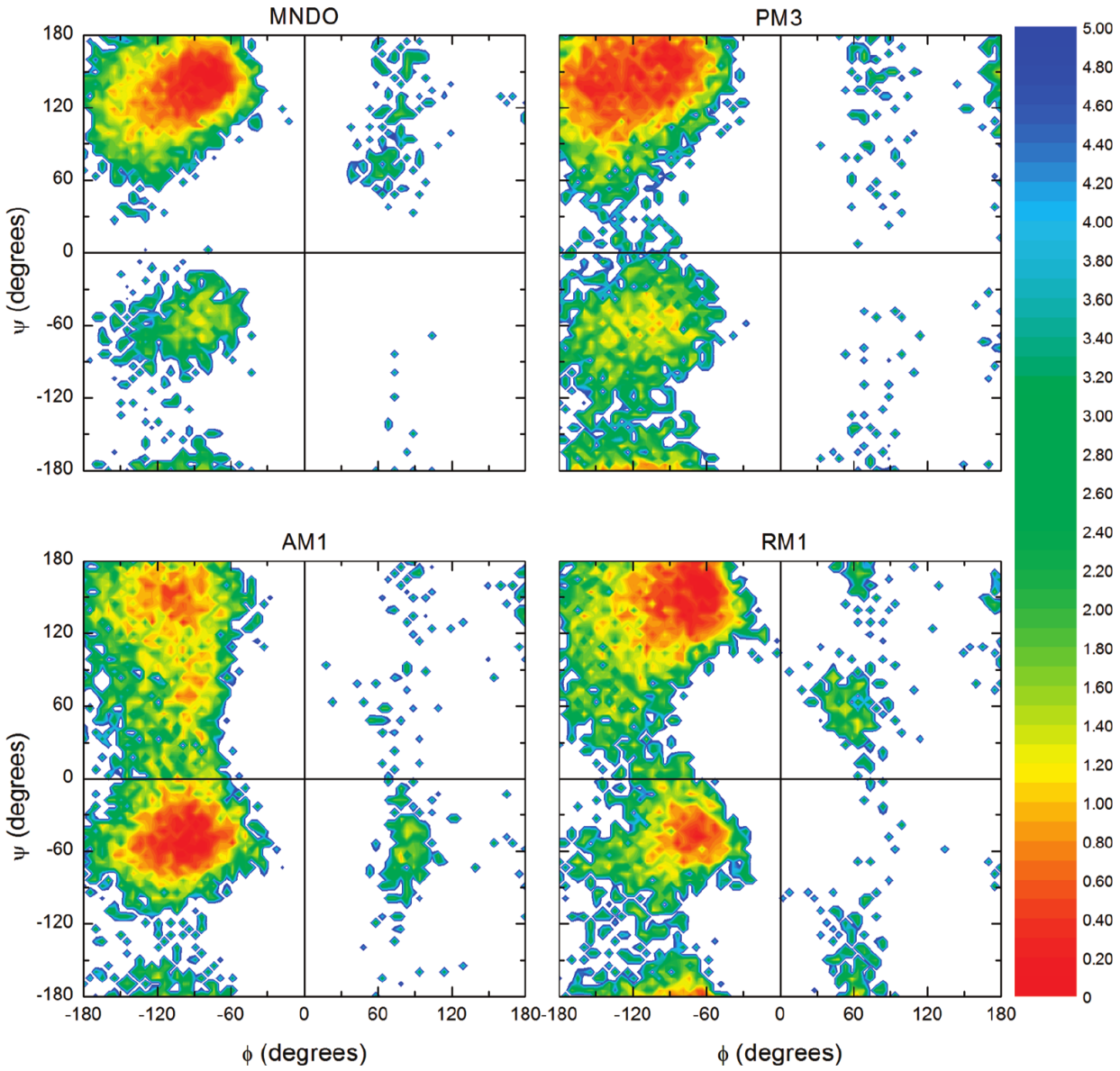


Figure 3. Free energy surfaces obtained from the 300 K replicas of the REMD simulations, for the different MNDO parametrizations.

Seabra et al. J. Phys. Chem. A 2009, 113, 11938–11948

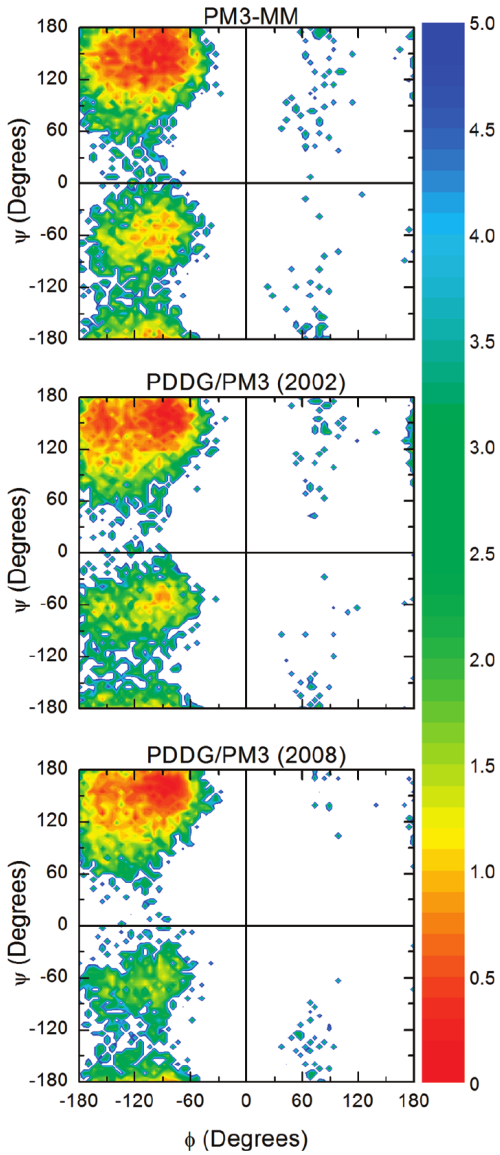


Figure 4. Free energy surfaces obtained from the 300 K replicas of the REMD simulations, for the different variations of the PM3 method.

- *$^3J(H_N, H_\alpha)$ coupling constant alone is incapable of fully distinguishing between basins*
- *The results from the QM methods vary just as much as for the different MM force fields*
- *with the exception of RM1, most QM methods lead to grossly overestimated dipolar coupling constants*
- *Why is it so difficult to reproduce free energies correctly for alanine dipeptide:*

☞ *from exp. data, the free energy of the P_{II} basin should lie only about 0.24–0.67 kcal/mol below other minima.*

Jono et al. J. Comput. Chem. 2010, 31, 1168–1175

- Can QM/MM with ab initio QM do better than semiempirical on alanine dipeptide in water?
- alanine dipeptide immersed in a sphere of 410 TIP3P water molecules
- QM = HF/3-21G
- Conformational sampling with *m*ulticanonical Molecular Dynamics

Jono et al. J. Comput. Chem. 2010, 31, 1168–1175

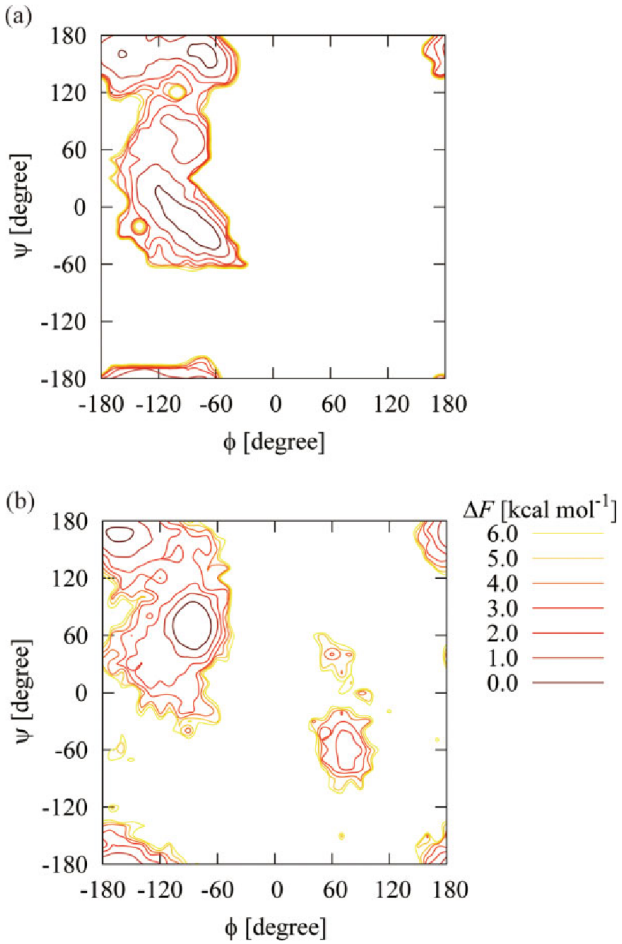


Figure 3. Free-energy maps as a function of the backbone ϕ-ψ di hedral angles at 300 K, calculated from the ensembles generated b (a) the multicanonical QM/MM MD simulation for alanine dipeptid in explicit water and (b) the AIMD simulation for alanine dipeptid in the gas phase.

Table 1. Populations of Conformations Stable in Explicit Water at 300 K.

Conformation	ϕ range (°)	ψ range (°)	Population (%)
C_5	$-240 \leq \phi < -120$	$120 \leq \psi < 240$	15.9
P_{II}	$-120 \leq \phi < 0$	$120 \leq \psi < 240$	19.0
C_{7eq}	$-150 \leq \phi < 0$	$50 \leq \psi < 120$	9.1
α_R	$-240 \leq \phi < 0$	$-120 \leq \psi < 50$	55.9

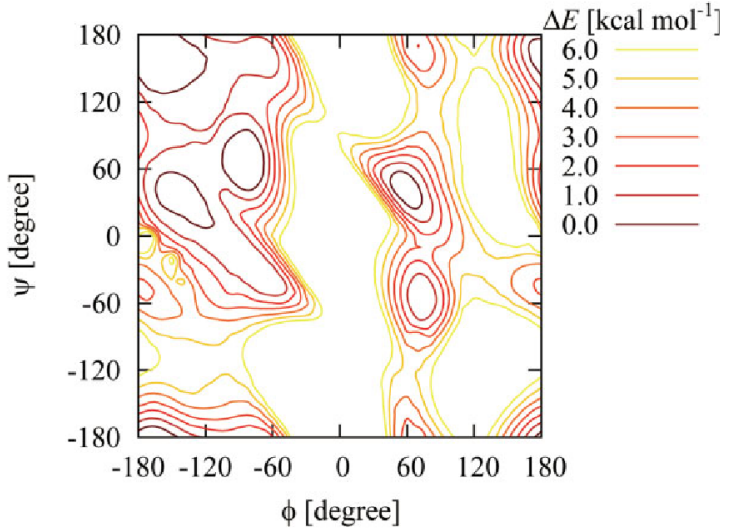


Figure 7. Energy map as a function of the backbone ϕ-ψ dihedral angles calculated for alanine dipeptide at the HF/3-21G level with CPCM (ε = 78.39).

Some conclusions on alanine dipeptide simulations

- QM/MM simulations can correctly model the effect of explicit solvent molecules onto the structure of a QM system
- QM/MM simulations are as good (or as bad!) as MM simulations
- choice of the QM method is of course crucial
- choice of the MM method is also crucial and should be taken care of
- Charge embedding (= QM wavefunction polarization by the MM atomic charges) is essential
- the parameters and equations related to the QM/MM interactions (electrostatic and nonelectrostatic) need special attention

When solvent molecules must be included in the QM part

- The effect of solvent molecules must sometimes be included in the QM part
- when one or many solvent molecules react with the QM solute
- when solvent molecules strongly bind to the QM solute

This is the case for solvated ions:

1. first shell solvent molecules must be included in the QM region to properly describe the electronic structure of the solute ion
2. dynamical behavior of the solvent close to the solute must be conserved (solvent molecule exchange between the first solvation shells and the bulk)

- Works from B. M. Rode et Coll. on *Coordination and ligand exchange dynamics of solvated metal ions*

B. M. Rode et al. Coord. Chem. Rev. 2005, 249, 2993–3006

Kerdcharoen et al. Chem. Phys. 1996, 211, 313–323

Kerdcharoen et al. Chem. Phys. 1996, 211, 313–323

- "Hot Spot" molecular dynamics method

  QM treatment is undertaken to full extent for a selected, chemically relevant spatial region, called "Hot Spot", leaving the rest of the system being treated by classical method.

  For a solvated ion, the "Hot Spot" represents a sphere in space containing the complete first solvation shell.

$$E = \langle \Psi_{in} | H | \Psi_{in} \rangle + E_{out-out} + E_{in-out}$$

first term: ab initio interactions between the particles inside the "Hot Spot"
second term: interactions between bulk particles
third term: interactions between the particles inside and outside the "Hot Spot"

Both latter terms are computed from classical pair potentials (no electrostatic embedding?)

Kerdcharoen et al. Chem. Phys. 1996, 211, 313–323

dynamical exchange of solvent molecules

- Quantum mechanical and pair-potential forces are assigned to the solvent particles by the equation

$$f_i = S_m(r_i) f_{QM} + (1 - S_m(r_i)) f_{PP}$$

f_{QM}	QM forces
f_{PP}	pair-potential forces
r_i	distance of center of mass of solvent molecule i from the center of the spherical "Hot Spot"
$S_m(r_i)$	a smoothing function

$$S_m(r) = \begin{cases} 1 & \text{for} \quad r \leq r_1 \\ \frac{(r_0^2 - r^2)^2 (r_0^2 + 2r^2 - 3r_1^2)}{(r_0^2 - r_1^2)^3} & \text{for} \quad r_1 < r \leq r_0 \\ 0 & \text{for} \quad r > r_0 \end{cases}$$

Kerdcharoen et al. Chem. Phys. 1996, 211, 313–323

Li^+ in liquid ammonia

- 1 Li^+ + 215 NH_3 molecules in a box of length 20.66 Å (experimental density)
- NVT, 235 K, dt=0.2 fs
- "Hot Spot" spherical radius = 8 Å (☞ includes first solvation shell)
- S_m ☞ $r_0 = 4.0$ Å and $r_1 = 3.8$ Å
- QM: HF/DZV+P or MNDO

Table 2
Comparison of ion–N RDF characteristics obtained from various simulations of ion in liquid ammonia. (r_{max}, r_{min} and n_{min} denote distance of the first maximum and the first minimum in Å, and the coordination number of the first shell, respectively)

Ion	r_{max}	r_{min}	n_{min}	T [K]	Ion/solvent	Method	Reference
Li^+	2.15	3.20	4.0	235	1/215	"Hot Spot" (ab initio)	this work
Li^+	2.42	3.14	6.0	235	1/215	"Hot Spot" (semi-empirical)	this work
Li^+	2.15	2.60	6	235	1/215	MD (pair potential)	this work

Rode et al. Coord. Chem. Rev. 2005, 249, 2993–3006

A review

- 1 ion + 499 solvent molecules
- QM region: first (+ second) solvation shell(s)
  $r0 - r1 = 0.2$ Å
- NVT simulations, dt = 0.2 fs
- MM: pair and 3-body potential functions derived from ab initio calculations
- QM: double basis sets plus polarization functions, with effective core potentials (ECP) for heavy atoms
- trajectory analysis:
  - ✓ radial and angular distribution functions
  - ✓ coordination number distributions
  - ✓ exchange rates and mean residence times

Rode et al. Coord. Chem. Rev. 2005, 249, 2993–3006

Table 2
Maxima r_M of the Ion–O radial distribution functions in Å, average coordination numbers CN and mean residence times τ in ps of several main group metal ions

	r_{M1} [a]	r_{M2} [a]	$CN_{av,1}$ [b]	$CN_{av,2}$ [b]	$\tau^{0.5}_{D,1st}$ [c]	$\tau^{0.5}_{D,2nd}$ [c]	Reference
Li(I)	1.95	–	4.2	–	–	–	[52]
Na(I)	2.33	–	5.4	–	2.4	–	[53]
K(I)	2.81	–	8.3	–	2.0	–	[53]
Rb(I)	2.95	–	7.1	–	2.0	–	[54,85]
Cs(I)	3.25	–	7.8	–	1.5	–	[50]
Mg(II)	2.03	4.12	6.0	18.3	–	–	[55]
Ca(II)	2.46	4.78	7.6	19.1	42.6	4.4	[56,45]
Sr(II)	2.70	5.0	9.0	20.4	∼40	5.2	[54]
Ba(II)	2.86	5.0	9.3	23.5	5.5	1.7	[57]
Al(III)	1.86	4.73	6.0	12.2	–	26.4	[58,54]
Ga(III)	1.96	4.3	6.0	13.6	–	–	[54]
Sn(II)	2.51	4.9	8.0	23.7	9.0	3.5	[54]
Pb(II)	2.60	5.0	9.0	24.3	–	5.6	[51,54]

[a] First and second peak maximum of the Ion–O-RDF.
[b] First and second shell coordination number.
[c] First and second shell mean residence time; $t^* = 0.5$ ps.

Rode et al. Coord. Chem. Rev. 2005, 249, 2993–3006

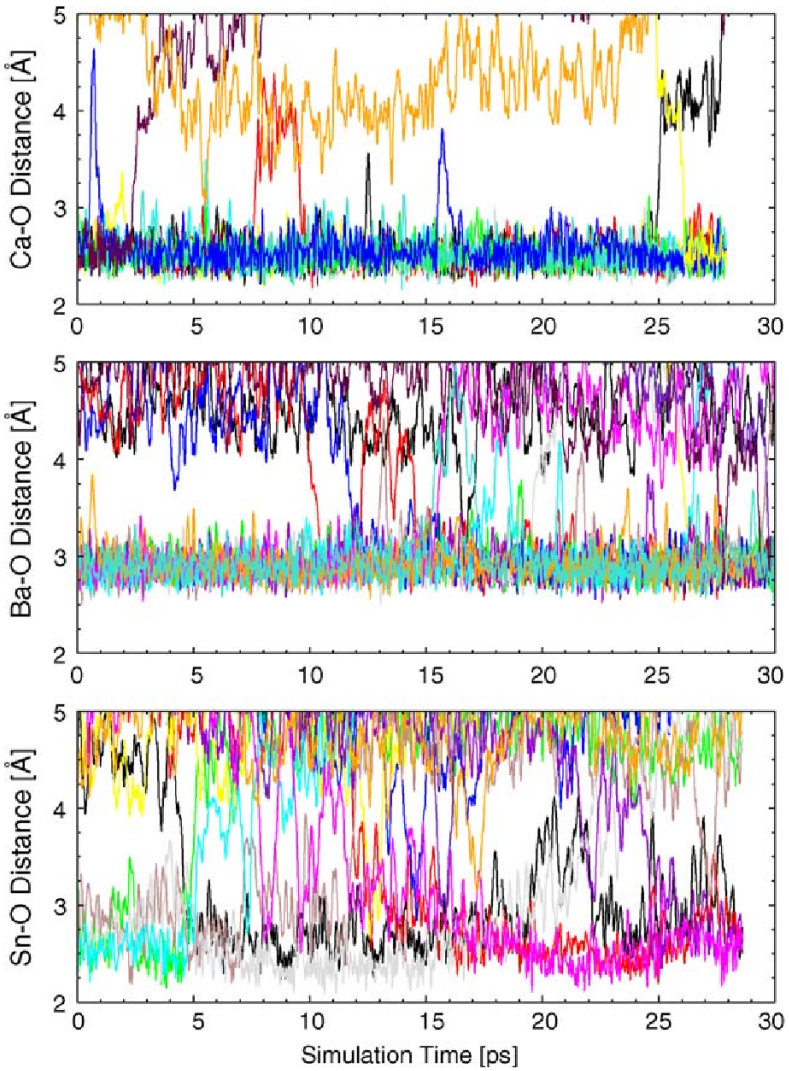


Fig. 3. Distance plots of water molecules showing water exchange processes within the time scale of the QM/MM MD simulation.

Kerdcharoen and Morokuma CPL 2002, 355, 257–262

ONIOM-XS: an extension of the ONIOM method for molecular simulation in condensed phase

- exchange of solvent molecules in the ONIOM framework
- Comments from the authors on B. M. Rode works:

  By employing a switching function, force on an exchanging particle can be smoothed when it changes from QM to MM region or vice versa. However, addition or deletion of a particle in the QM region due to the solvent exchange also effect forces on the remaining QM particles and this problem was not tackled in the previous works. In addition to the abovementioned disadvantage, the original scheme also suffers from the lack of clearly defining appropriate energy expression. Therefore, energy of the integrated system cannot be described during the exchange of particles.

Kerdcharoen and Morokuma CPL 2002, 355, 257–262

A double ONIOM scheme

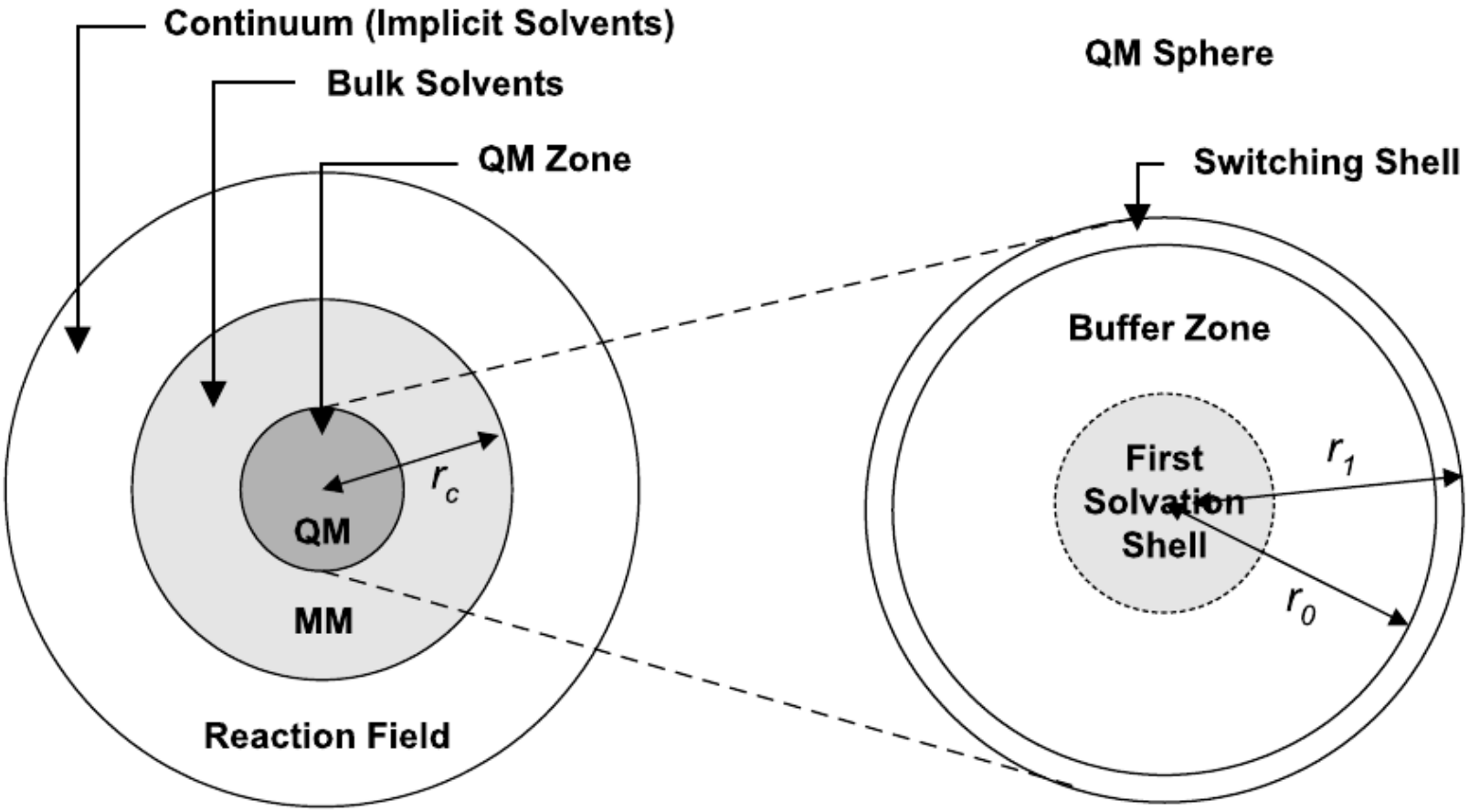


Fig. 1. Schematic diagram of the ONIOM-XS method.

Kerdcharoen and Morokuma CPL 2002, 355, 257–262

A double ONIOM scheme

N particles in the system:

- n_1 in the QM zone;
- l in the switching shell;
- n_2 in the MM
- ☞ $N = n_1 + l + n_2$

$$
\begin{aligned}
E^{\mathrm{ONIOM-XS}}(r_l) &= (1-\bar{s}(\{r_l\})).E^{\mathrm{ONIOM}}(n_1+l;N) \\
&\quad +\bar{s}(\{r_l\}).E^{\mathrm{ONIOM}}(n_1;N) \\
E^{\mathrm{ONIOM}}(n_1+l;N) &= E^{\mathrm{QM}}(n_1+l) - E^{\mathrm{MM}}(n_1+l) + E^{\mathrm{MM}}(N) \\
E^{\mathrm{ONIOM}}(n_1;N) &= E^{\mathrm{QM}}(n_1) \quad - E^{\mathrm{MM}}(n_1) \quad + E^{\mathrm{MM}}(N)
\end{aligned}
$$

Kerdcharoen and Morokuma CPL 2002, 355, 257–262

A double ONIOM scheme

The switching function $\bar{s}(\{r_I\})$ is an average over a set of switching functions for individual particle in the switching shell $s_i(x_i)$

$$\bar{s}(\{r_I\}) = \frac{1}{I}\sum_{i=1}^{I} s_i(x_i)$$

with

$$s_i(x_i) = 6(x_i - \frac{1}{2})^5 - 5(x_i - \frac{1}{2})^3 + \frac{15}{8}(x_i - \frac{1}{2}) + \frac{1}{2}$$

and

$$x_i = \frac{r_i - r_0}{r_1 - r_0}$$

where r_i is the distance between the center of mass of the exchanging particle and the center of the QM sphere.

Kerdcharoen and Morokuma CPL 2002, 355, 257–262

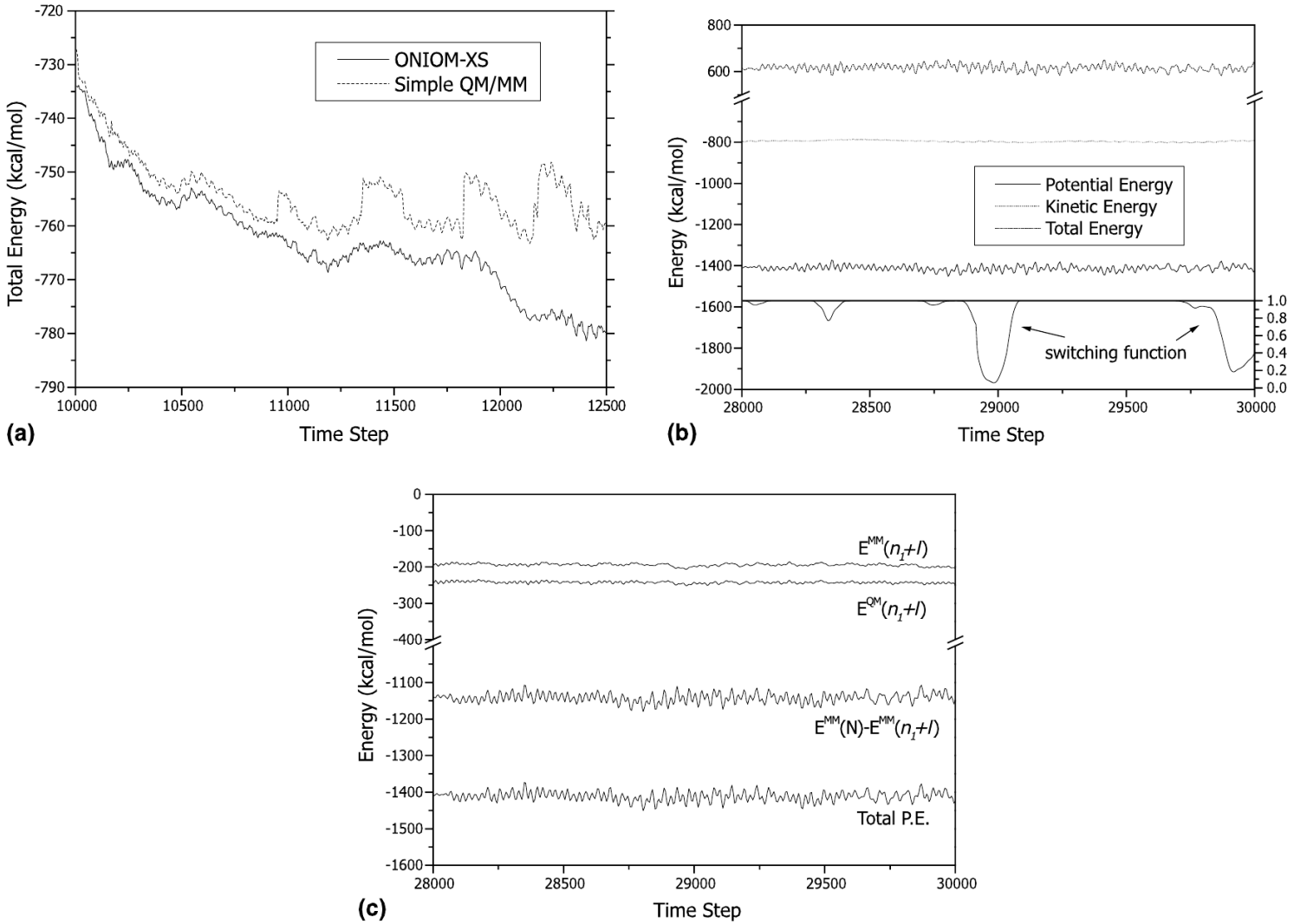


Fig. 2. (a) Total energies obtained from ONIOM-XS MD simulation and simple QM/MM simulation (without smoothing scheme) during equilibration process. (b) Kinetic, potential and total energies during a snapshot period of equilibrium obtained from ONIOM-XS simulation. The inset shows values of the switching function. (c) Components of the smoothed ONIOM potential energy.

Dynamical solvent exchange in QM/MM methods: some conclusions

- A special treatment is needed to account for solvent exchange in the first solvation shells at the QM level
- Kerdcharoen and Rode's proposal: 1 QM calculation/step
- Kerdcharoen and Morokuma's proposal: 2 QM calculations/step
- A new proposal in 2009 by Bulo et al. (JCTC 2009, 5, 2212–2221): up to 4 QM calculations/step to obtain energy conservation ("true" NVE simulations)

Strnad et al. J. Chem. Phys. 1997, 106, 3643–3656

Modeling reactivity in QM/MM simulations

- QM/MM simulation: solute + solvent ☞ (too) many degrees of freedom
- ☞ How to locate transition states?
- The usual mathematical definition of a TS (extremum of the energy with one and only one negative eigenvalue for the hessian) is not useful anymore
- Sampling of the free energy surface is mandatory
- Problem: transition state crossing is a rare event

Strnad et al. J. Chem. Phys. 1997, 106, 3643–3656

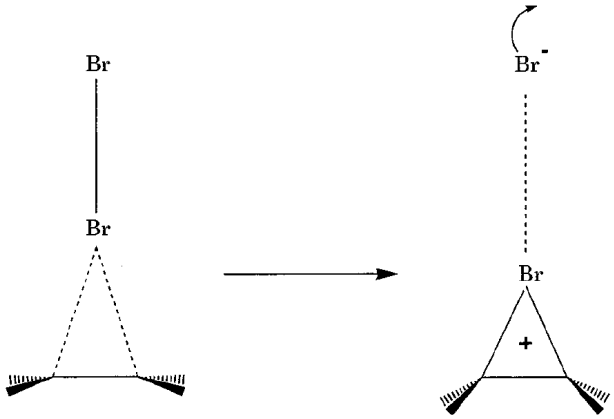
Rare event technique (or how to model reactive trajectories)

1. Define an adequate TS sructure and for such a structure define the pseudo-normal mode of vibration corresponding to the reaction coordinate
2. Perform NVT molecular dynamics simulations for the TS-structure in solution with a frozen reaction coordinate
3. From theses simulations, select a set of independent configurations for the whole system
4. For each initial configuration, define a set of random velocities for the system using a Maxwell-Boltzmann distribution at the requested temperature
5. Integrate the equations of motion forward and backward in time until the chemical system reaches the reactants or the products.
6. Repeat steps 4 and 5 for all the initial configurations so that a statistically representative sample of reactive trajectories is obtained and average properties can be computed

Strnad et al. J. Chem. Phys. 1997, 106, 3643–3656

A test case: first reaction step of bromination of ethylene in water

- ethylene bromination in water: a two step process
- the first rate-limiting step is essentially a charge separation process



Scheme 1

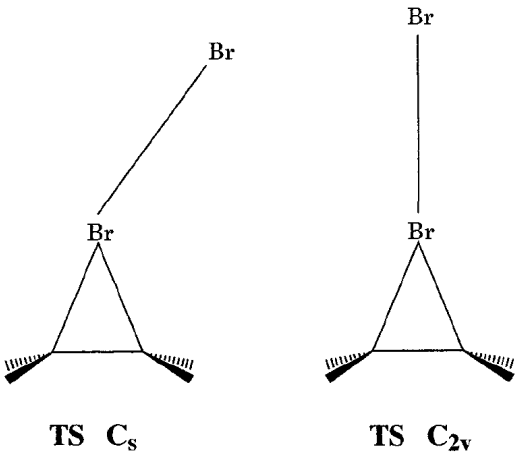


FIG. 1. Schematic structure of the transition state with C_s (left) or C_{2v} symmetry.

charge transfer complex formation two possible TS forms

- TS in gas phase: C_s structure; TS in liquid water: C_{2v} structure

Strnad et al. J. Chem. Phys. 1997, 106, 3643–3656

A test case: first reaction step of bromination of ethylene in water

- System: ethylene + Br_2 + 300 TIP3P water molecules (cubic box of 20.8 Å length)
- QM: DFT from deMon program (LSD+VWN and BP functionals)
- Initial TS structure: located using Nancy Multipole Expansion continuum model (no explicit water molecule)
- 70 ps NVT equilibration with constrained TS structure
- 140 trajectories:
  - ✓ 66% are non-reactive
  - ✓ 34% are reactive
  - ✓ 15% of the reactive trajectories present barrier recrossings

Strnad et al. J. Chem. Phys. 1997, 106, 3643–3656

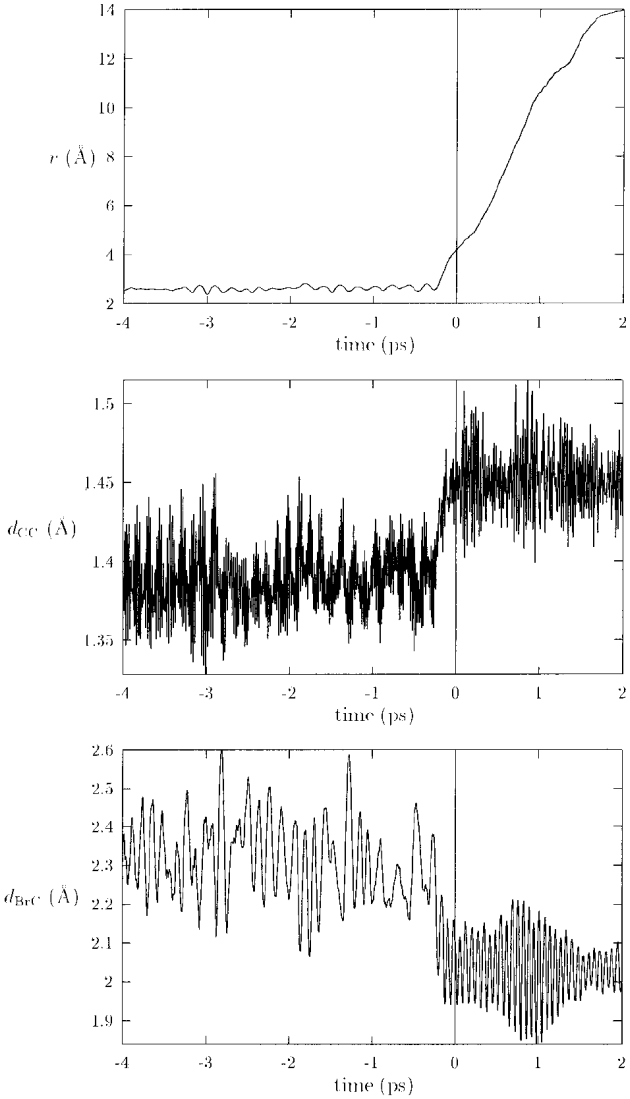


FIG. 5. Evolution of the r coordinate (BrBr distance), CC and BrC bondlengths in a standard Type I reactive trajectory. The simulation is started at $t=0$ forward and backward in time.

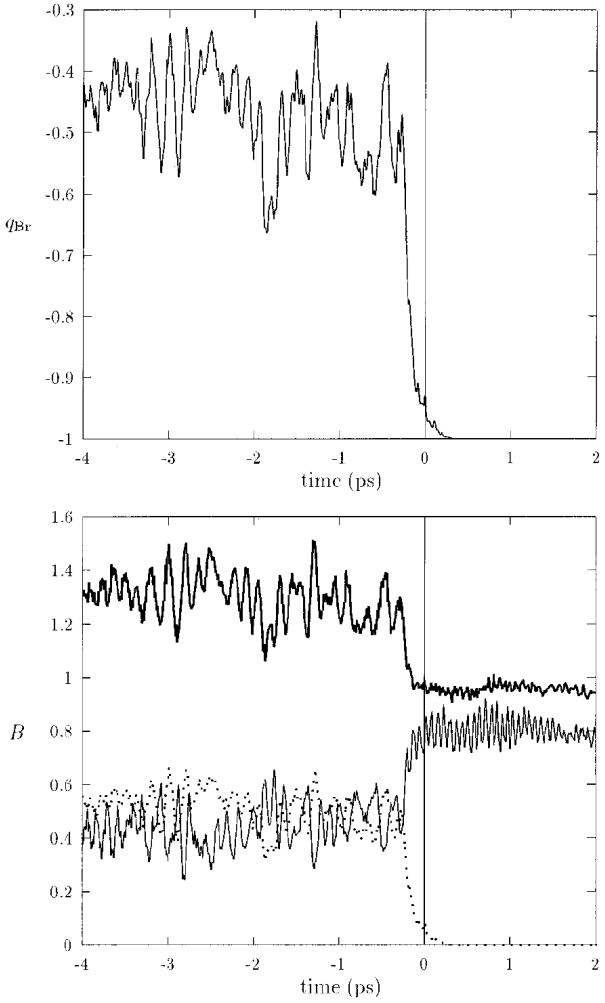


FIG. 7. Evolution of the forming bromide charge and Mayer bond orders B [CC (bold line), BrC (single line) and BrBr (dotted line)] along the trajectory represented in Fig. 5.

Chalmet et al. JPCA 2001, 105, 11574–11581

- *Computer simulation of amide bond formation in aqueous solution*
- DFT/MM simulations with rare event techniques
- if you reverse time:
  Computer simulation of formamide hydrolysis in aqueous solution
- System: NH_3 + HCOOH + 215 TIP3P water molecules
  (cubic box of 18.8Å length)
- 29% reactive trajectories

Where's my proton?

- The proton affinity of a molecule can differ greatly whether it is measured in gas phase, in water, on in an hydrophobic media
- For example, the ionizable properties of an amino acid are different
  - ✓ in gas phase
  - ✓ in water
  - ✓ buried in an enzyme (where solvent is not accessible)
- Classical force field usually models ionizable residue only in their standard state: their protonation state at pH=7
- However, depending on the simulation pH but also on the environnement, the ionizable state of an amino acid can vary greatly
- It is therefore of great importance, when modeling biological systems like peptides or proteins, to analyze the protonation states of the system

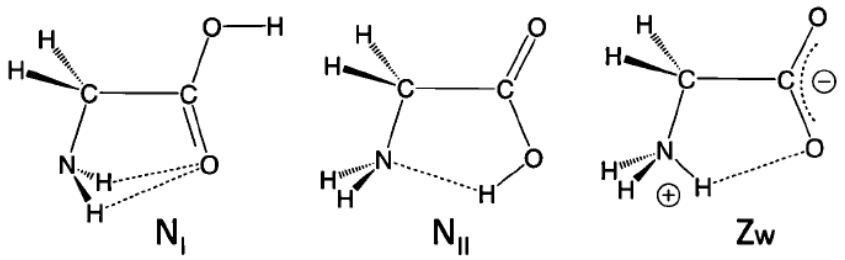
Martins-Costa & Ruiz-López PCCP, 2011[1]

Simulation of amino acid diffusion accross water/hydrophobic interfaces

- Simulation of the diffusion of glycine (Gly) across a water/CCl_4 interface.

$$\begin{array}{ccc} {}^{+}H_3NCH_2COO^{-}_{(water)} & \rightleftarrows & H_2NCH_2COOH_{(water)} \\ \downarrow\uparrow & & \downarrow\uparrow \\ {}^{+}H_3NCH_2COO^{-}_{(CCl_4)} & \rightleftarrows & H_2NCH_2COOH_{(CCl_4)} \end{array}$$

- Gly exists mainly as a zwitterion in water, whereas only neutral tautomers are stable in hydrophobic media



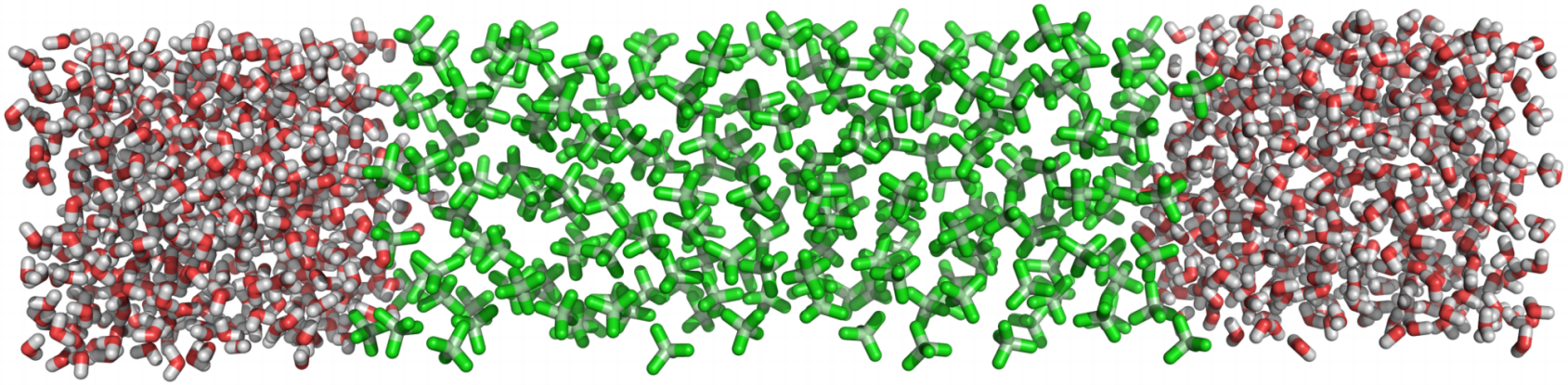
Scheme 1 Main conformations for neutral (N_I and N_{II}) and zwitterionic (Zw) Gly considered in the present work.

[1] Martins-Costa, M. T. C.; Ruiz-Lopez, M. F. *Phys. Chem. Chem. Phys.* **2011**, *13*, 11579–11582

Martins-Costa & Ruiz-López PCCP, 2011

QM/MM Computational methodology

- Gly is described by QM level: B3LYP/6-31G*
- Water: 1000 molecules; CCl_4: 220 molecules
- box size: 24 Å x 24Å x 114 Å



- Simulations start with equilibrated Gly in bulk water or in the organic phase
- A bias harmonic potential is used to gradually push the solute into the opposite phase (each window is 5 to 25 ps)

Martins-Costa & Ruiz-López PCCP, 2011

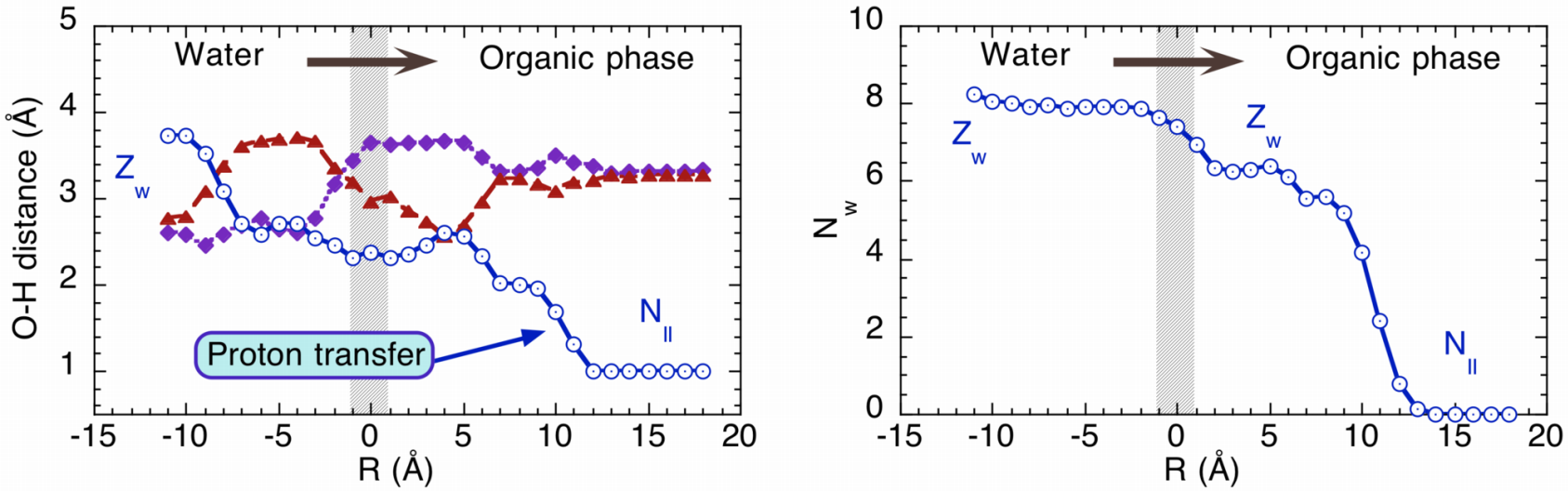
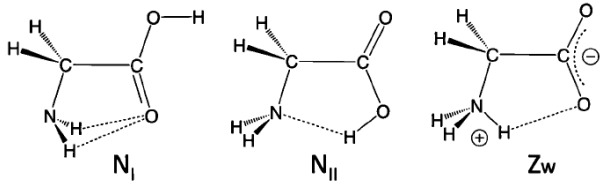
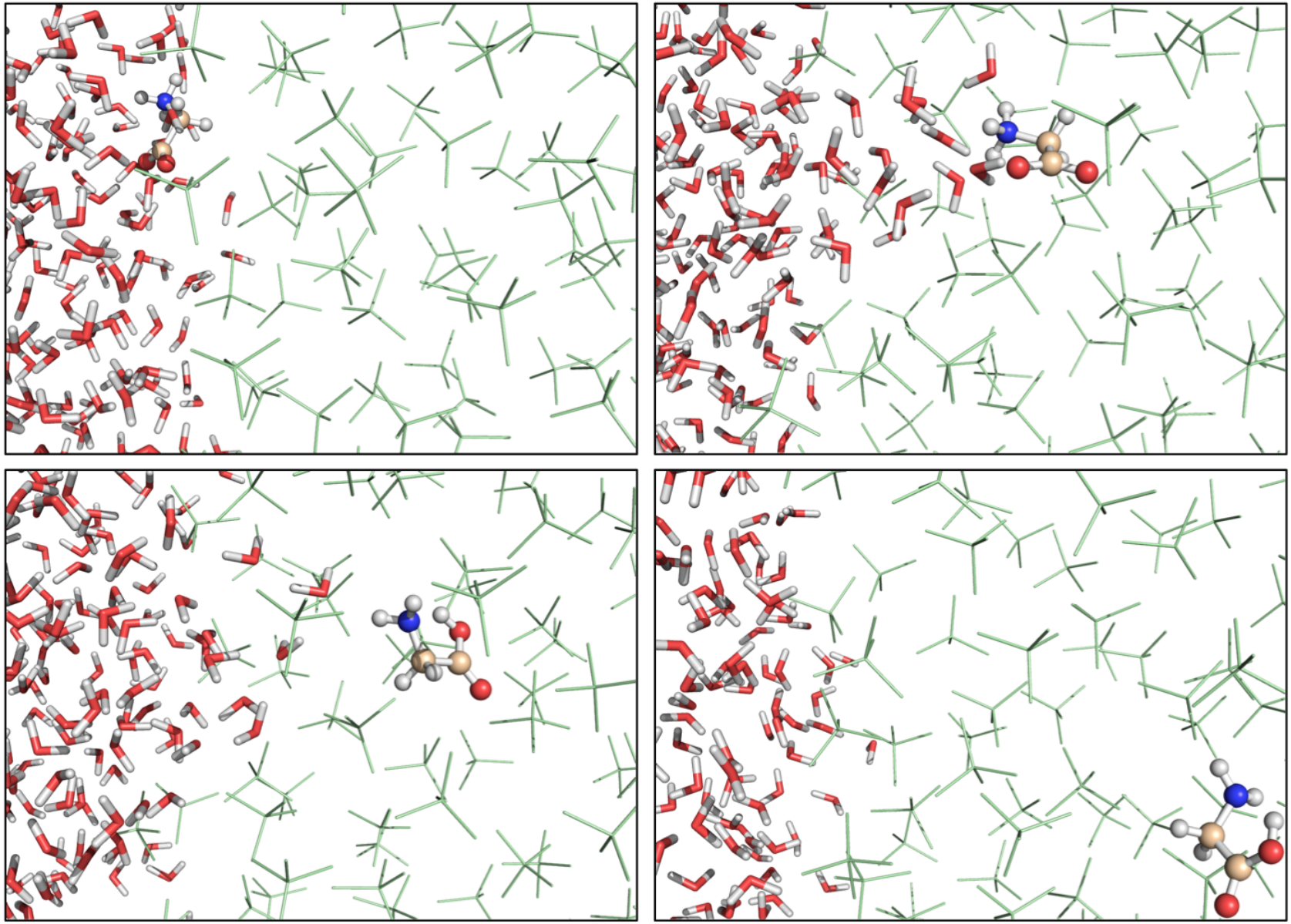


Fig. 2 Average Gly properties from the MD simulation for water to organic phase transfer as a function of the distance to the interface ($R = 0$). Left: O–H distances for H atoms in the Zw NH_3^+ group. Right: number of hydrogen bonds. Values at R_i correspond to averages in the range $R_i \pm 1$ Å. The distance at which internal proton transfer occurs is indicated.



Scheme 1 Main conformations for neutral (N_I and N_{II}) and zwitterionic (Zw) Gly considered in the present work.

Martins-Costa & Ruiz-López PCCP, 2011



Martins-Costa & Ruiz-López PCCP, 2011

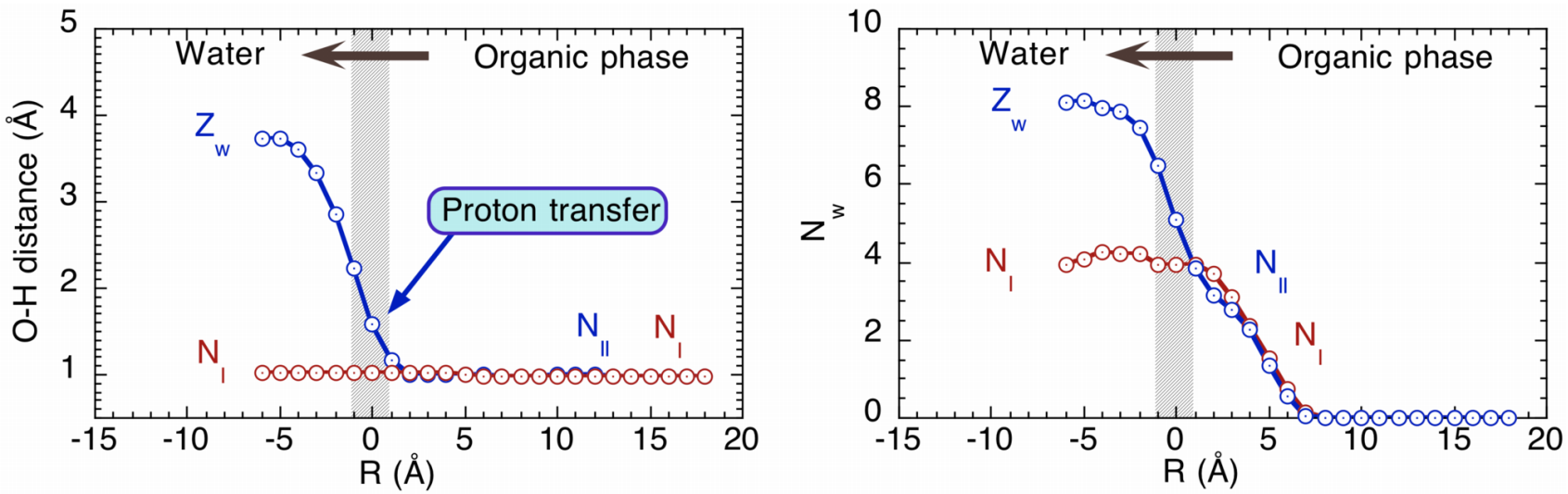
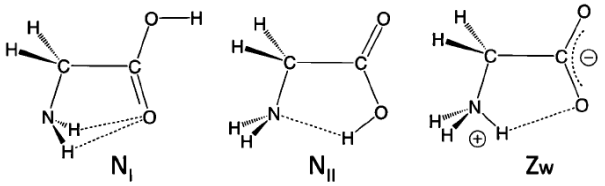


Fig. 4 Average Gly properties computed from the MD simulation for the organic phase to water transfer, as a function of the distance to the interface R. Left: O–H distance. Right: number of hydrogen bonds. Values at R_i correspond to averages in the range $R_i \pm 1$ Å. The distance at which internal proton transfer occurs is indicated.



Scheme 1 Main conformations for neutral (N_I and N_{II}) and zwitterionic (Zw) Gly considered in the present work.

Martins-Costa & Ruiz-López PCCP, 2011

- QM/MM simulations are capable of simulating the proton transfer that can occur at a water/hydrophobic media interface
- Here, the thickness of the boundary between water and CCl_4 is estimated at 1 nm
- At the interface, water molecules can enter the CCl_4 medium to solvate zwiterionic or neutral forms of glycine

Li et al. J. Phys. Chem. B, 106, 3486[2]

- How to evaluate the pKa of an ionizable residue in a protein?
- System: Turkey ovomucoid third domain
- experimental pKa's are known for some ionizable residues
- Idea: residue/protein/water = QM/MM/Continuum

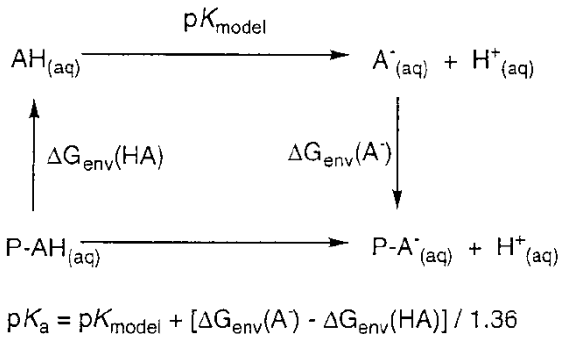


Figure 1. Thermodynamic cycle relating the pK_a of a model compound (pK_{model}) to the pK_a of a protein residue via the environmental energies (ΔG_{env}) of the products and reactants. The value 1.36 corresponds to RT ln(10) at 298 K in kcal/mol.

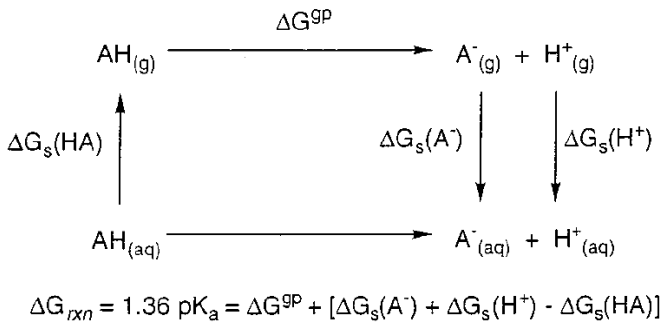


Figure 2. Thermodynamic cycle relating the pK_a to the gas-phase proton basicity (ΔG^{gp}) via the solvation energies (ΔG_s) of the products and reactants. The value 1.36 corresponds to RT ln(10) at 298 K in kcal/mol.

[2] Li, H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H. *J. Phys. Chem.B* **2002**, *106*, 3486–3494

Li et al. J. Phys. Chem. B, 106, 3486

- Benchmark: pKa of small residues in solution

TABLE 1: Computed and Experimental pK_a's of Small Molecules with Functional Groups Found in Amino Acid Residues and the Individual Energy Components Used to Compute the pK_a's (Figure 2 and eq 1) in kcal/mol

			D-PCM/ICOMP = 4		IEF-PCM/ICOMP = 0		
acid	ΔE^{MP2} [a]	ΔG_{trv} [b]	$\Delta\Delta G_s$ [c,e]	pK_a	$\Delta\Delta G_s$ [c]	pK_a	expt
acetic acid	352.55	−13.28	−333.08	4.6	−330.81	6.2	4.8
methylamine	223.28	−13.25	−196.72	9.8	−196.76	9.8	10.6
imidazole	231.26	−12.66	−210.34	6.1	−210.19	6.2	7.0
phenol	354.43	−11.99	−329.20	9.7	−322.88	14.4	10.0
methanethiol	360.71	−10.02	−336.90	10.1	−331.85	13.8	10.3
rmsd [d]				0.6		2.6	
Lys55	249.38	−12.26	−221.78	11.3	−223.61	9.9	11.1

[a] Gas-phase (electronic) deprotonation energy, $\Delta E^{MP2//RHF}$; cf. eq 1. [b] Gas-phase free energy correction using 1 M reference state; sum of the last four terms in eq 1. [c] Change in solvation energy; last three terms in the equation in Figure 2. [d] Root-mean-square deviation from experiment. [e] Calculated using D-PCM-X/ICOMP = 4; cf. eq 3.

Li et al. J. Phys. Chem. B, 106, 3486

➢ Application: pKa of small residues in the protein

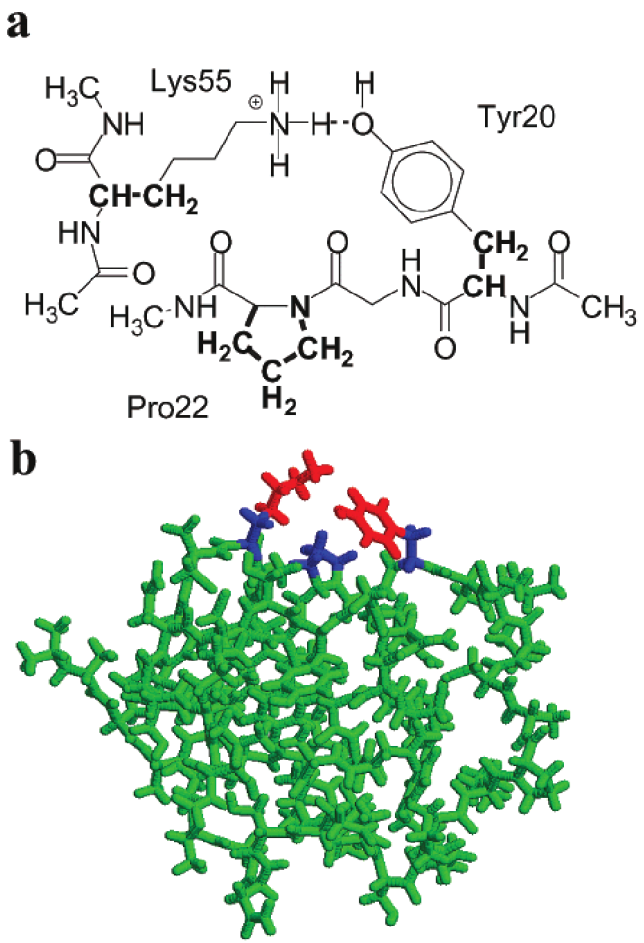


Figure 3. Subsystem of OMTKY3 (a) used to obtain the buffer region (bold) used for (b) ab initio/buffer/EFP regions (red/blue/green) used for the computation of the pK_a of Lys55.

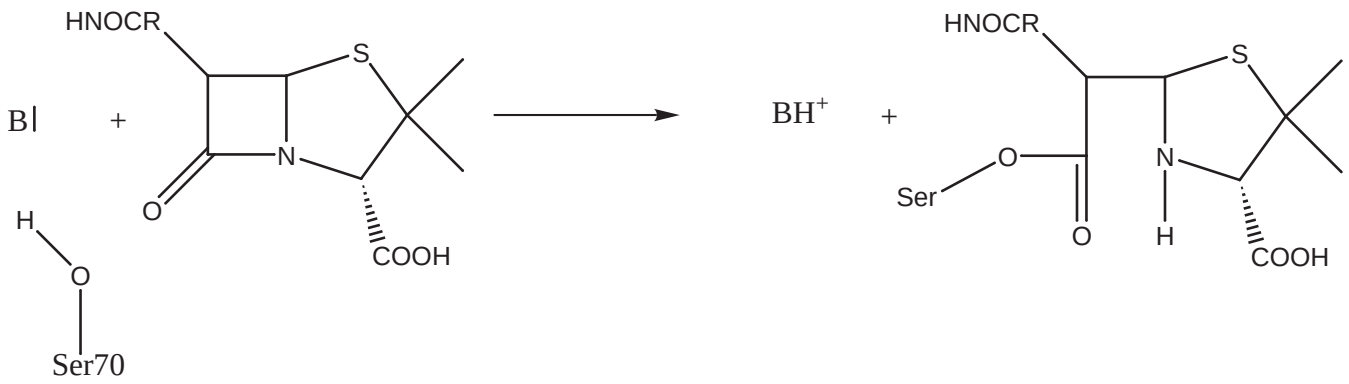
TABLE 3: Experimentally Measured pK_a's and Hill Coefficients (cf. eq 12) for Ovomucoid Third Domain, Taken from Ref 24[a]

residue	exptl pK_a	n	calcd pK_a
Asp27	<2.3	0.85	3.4
CysC56	<2.5	0.87	2.7
Asp7	<2.7	0.72	3.3
Gly19	3.2	1.08	2.8
Glu10	4.2	0.92	3.5
Glu43	4.8	0.95	4.4
His52	7.5	0.93	6.2
LeuN1	8.0	0.88	7.5
Lys13	9.9	0.69	11.2
Lys34	10.1	0.66	11.7
Tyr11	10.2	0.73	10.0
Tyr20	11.1	0.57	9.9
Lys29	11.1	0.87	12.1
Lys55	11.1	0.64	11.3
Tyr31	>12.5		11.2

[a] Standard pK_a values for amino acid residues are as follows: Asp = 4.0; Glu = 4.4; Tyr = 9.6; His = 6.6−7.0; Lys = 10.4; α-carboxyl group = 3.8; α-amino group = 7.5

The acylation step in β-lactamases (1)

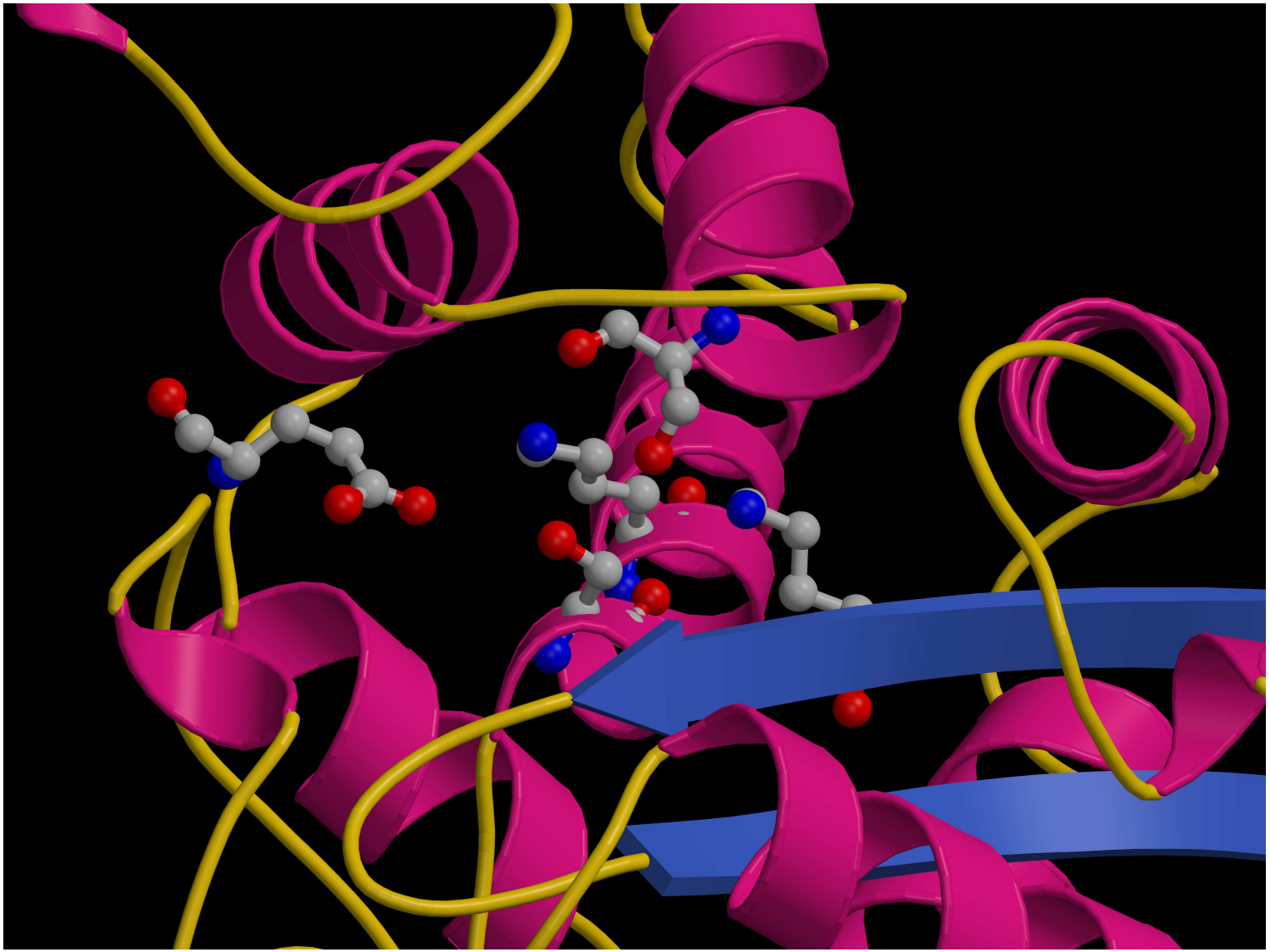
Class A β-lactamases are enzymes that induce bacteria resistance through the degradation of β-lactam derived antibiotics in two consecutive steps: acylation and deacylation.



Active site: S70, S130, K73, K234, E166

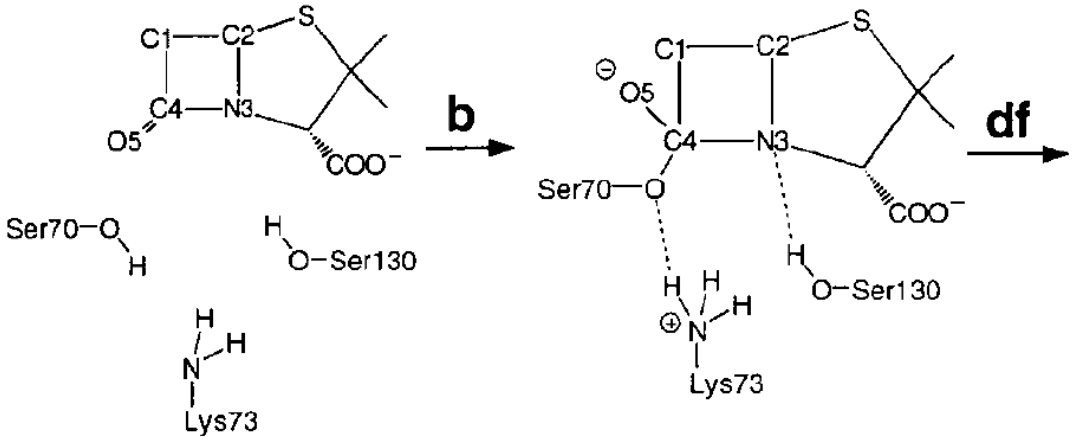
While Glu166 is unambiguously recognised as the general base in the deacylation, there has been much controversy on whether Lys73 or Glu166 acts as a general base to activate Ser70 in the acylation.

The acylation step in β-lactamases (2)



The acylation step in β-lactamases (3)

Two mechanisms in competition



K73 as the general base

E166 as the general base

The acylation step in β-lactamases (4)

Pitarch, J.; Pascual-Ahuir, J.; Silla, E.; Tunon, I. J. Chem. Soc., Perkin Trans. **2000**, *4, 761–767*

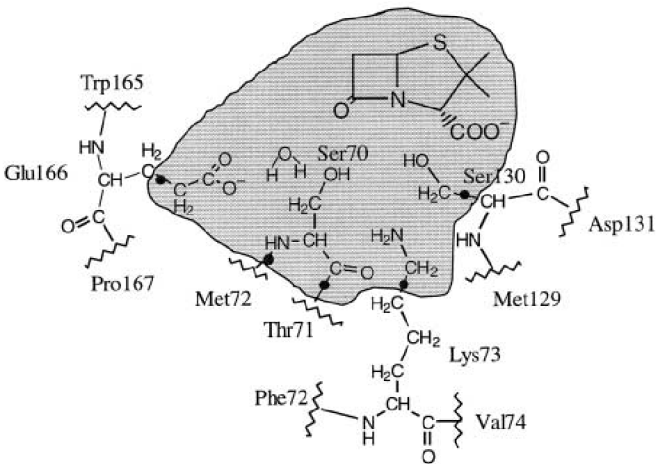


Fig. 2 Schematic representation of the active site: the shaded region corresponds to the QM atoms. The five link atoms are indicated as "●".

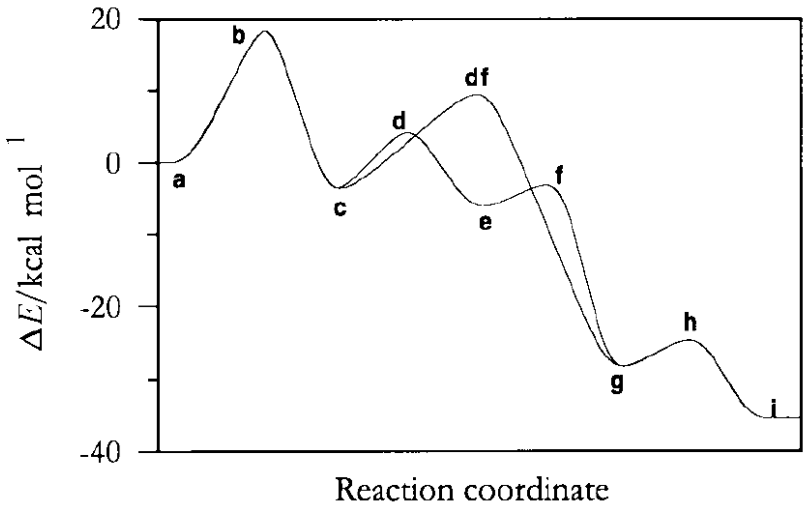


Fig. 3 Energy profile of the acylation process.

Lys73 is the general base (starting from neutral Lys73);

AM1/CHARMM: $\Delta E^{\ddagger} = 18$ kcal/mol

The acylation step in β-lactamases (5)

Hermann, J.; Ridder, L.; Mulholland, A.; Holtje, H. J. Am. Chem. Soc. **2003**, *125, 9590–9591*

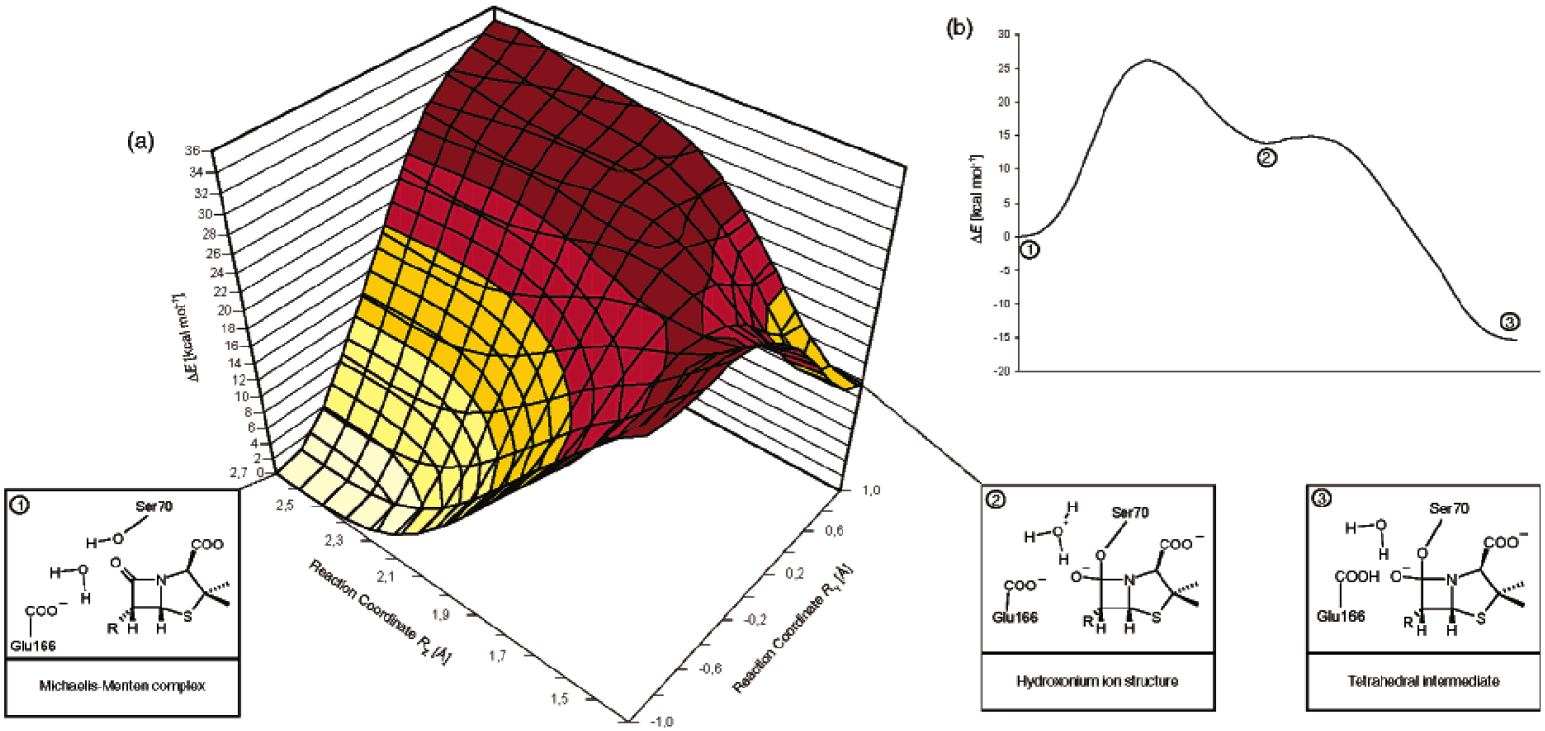


Figure 2. (a) QM/MM potential energy surface of the first step of acylation. (b) Overall reaction energy profile of the formation of the tetrahedral intermediate.

Glu166 as the general base (starting from ionic Lys73);

AM1/CHARMM: $\Delta E^{\ddagger} = 26$ kcal/mol

The acylation step in β-lactamases (6)

Hermann, J. C.; Hensen, C.; Ridder, L.; Mulholland, A. J.; Holtje, H. D. J. Am. Chem. Soc. **2005**, *127, 4454–4465*

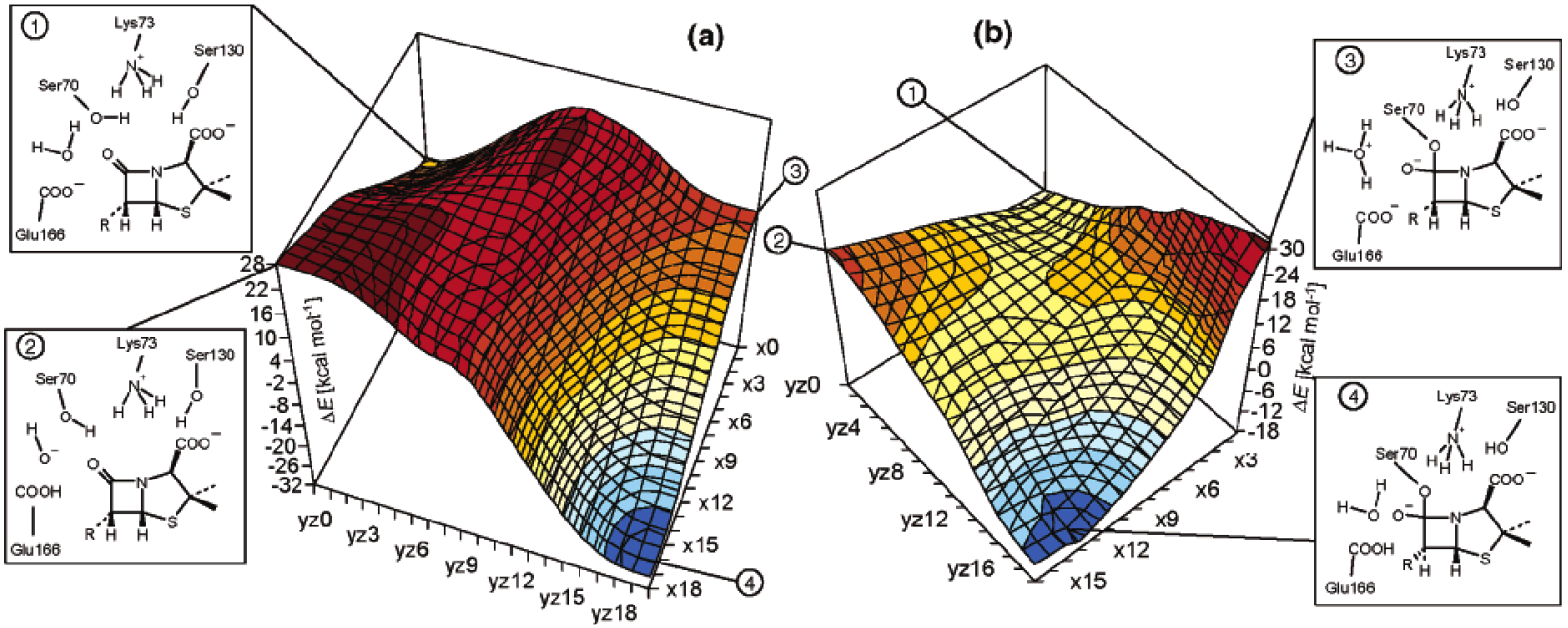
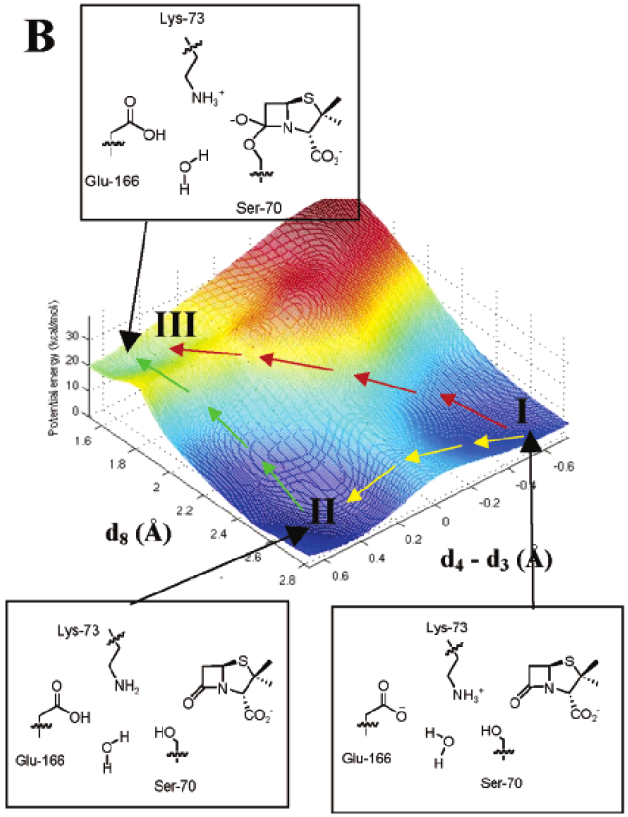


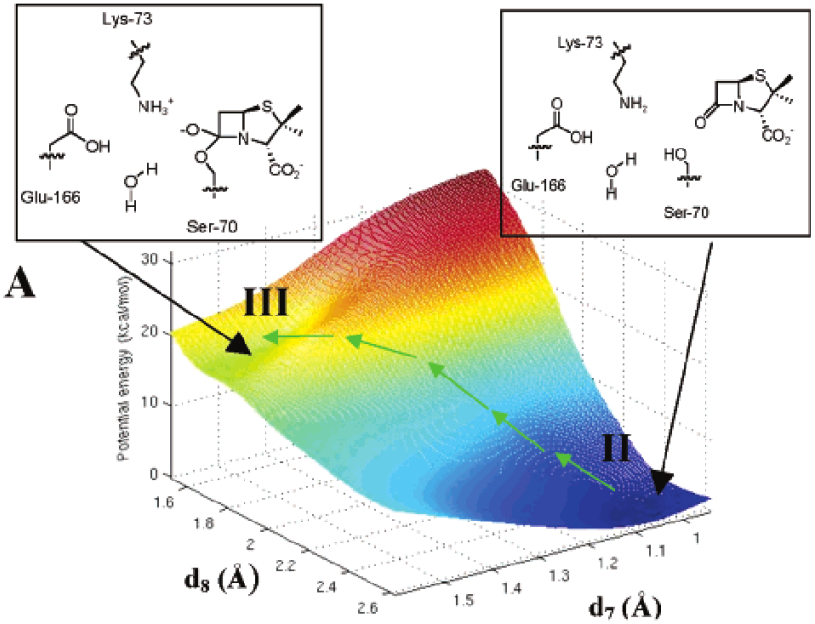
Figure 2. QM/MM potential energy surfaces for the reaction coordinates R_{RX} and R_{YZ}. (a) AM1-CHARMM22-energies, (b) B3LYP/6-31G+(d)//AM1-CHARMM22-energies. (1) is the Michaelis complex; (2) and (3) are unstable structures; and (4) is the tetrahedral intermediate.

Glu166 as the general base (starting from ionic Lys73);
B3LYP/6-31+G(d)/CHARMM//AM1/CHARMM: $\Delta E^{\ddagger} = 9$ kcal/mol

The acylation step in β-lactamases (7)

Meroueh, S.; Fisher, J.; Schlegel, H.; Mobashery, S. J. Am. Chem. Soc. **2005**, *127, 15397–15407*





Lys73 transfers it proton to Glu166, then acts as the general base!
Glu166 as the general base is a competiting pathway

ONIOM:MP2/6-31+G(d)/AMBER: $\Delta E^{\ddagger} = 22$ kcal/mol

The acylation step in β-lactamases (8)

Hermann, J. C.; Pradon, J.; Harvey, J. N.; Mulholland, A. J. J. Phys. Chem. A **2009**, *113, 11984–11994*

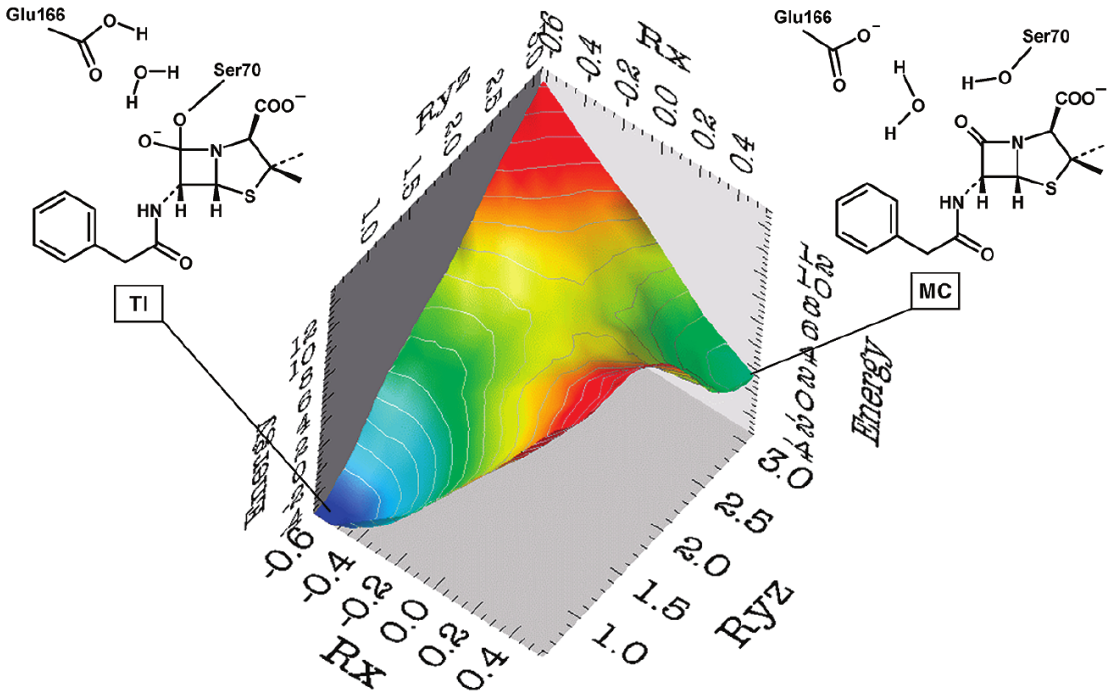


Figure 1. B3LYP/6-31G+(d)/CHARMM27 QM/MM potential energy surface for formation of the tetrahedral intermediate in acylation. MC is the Michaelis (enzyme−substrate) complex; TI is the tetrahedral intermediate. The energy is given in kcal mol^{-1} and the reaction coordinates R_X and R_{YZ} in Å.

Glu166 as the general base (no Lys73 in the model);

MP2/aug-cc-pVTZ//B3LYP/6-31+G(d)/CHARMM: $\Delta E^{\ddagger} = 3 - 12$ kcal/mol

QM/MM applications: selected reviews

- *Solvent effects on organic reactions from QM/MM simulations*
  Avecedo, O.; Jorgensen, W. L.; Elsevier B. V., 2006; Vol. 2 of *Ann. Rep. Comput. Chem.;* chapter 14
- *Chemical accuracy in QM/MM calculations on enzyme-catalysed reactions*
  Mulholland, A. J. *Chem. Cent. J.* **2007**, *1*, 19–23
- *Development and application of ab initio QM/MM methods for mechanistic simulation of reactions in solution and in enzymes*
  Hu, H.; Yang, W. *J. Mol. Struct. (THEOCHEM)* **2009**, *898*, 17–30
- *Advances in Quantum and Molecular Mechanical (QM/MM) Simulations for Organic and Enzymatic Reactions*
  Avecedo, O.; Jorgensen, W. L. *Acc. Chem. Res.* **2010**, *43*, 142–151
- *Investigations of enzyme-catalysed reactions with combined quantum mechanics/molecular mechanics (QM/MM) methods*
  Ranaghan, K. E.; Mulholland, A. J. *Int. Rev. Phys. Chem.* **2010**, *29*, 65–133