FINAL FOR SUBMISSION
ANL/CMB/PP-87153 7/14/95

Progress toward Chemical Accuracy in the Computer Simulation of Condensed Phase Reactions

Paul A. Bash,[†*] L. Lawrence Ho,[†‡] Alexander D. MacKerell, Jr.,[§]
David Levine,[‖] and Philip Hallstrom[‖]

[†] *Center for Mechanistic Biology and Biotechnology*
Argonne National Laboratory
Argonne, IL 60439

[‡] *J.W. Gibbs Laboratory*
Department of Physics
Yale University
New Haven, CT 06511

[§] *Department of Pharmaceutical Sciences*
School of Pharmacy
University of Maryland at Baltimore
Baltimore, MD 21230

[‖] *Mathematics and Computer Science Division*
Argonne National Laboratory
Argonne, IL 60439

* To whom reprint requests should be addressed

Abstract: *A procedure is described for the generation of chemically accurate computer-simulation models to study chemical reactions in the condensed phase. The process involves (i) the use of a coupled semiempirical quantum and classical molecular mechanics method to represent solutes and solvent, respectively; (ii) the optimization of semiempirical quantum mechanics (QM) parameters to produce a computationally efficient and chemically accurate QM model; (iii) the calibration of a quantum/classical microsolvation model using* ab initio *quantum theory; and (iv) the use of statistical mechanical principles and methods to simulate, on massively parallel computers, the thermodynamic properties of chemical reactions in aqueous solution. The utility of this process is demonstrated by the calculation of the enthalpy of reaction in vacuum and free energy change in aqueous solution for a proton transfer involving methanol, methoxide, imidazole, and imidazolium, which are functional groups involved with proton transfers in many biochemical systems. An optimized semiempirical QM model is produced, which results in the calculation of heats of formation of the above chemical species to within 1.0 kcal/mol of experimental values. The use of the calibrated QM and microsolvation QM/MM models for the simulation of a proton transfer in aqueous solution gives a calculated free energy that is within 1.0 kcal/mol (12.2 calculated vs. 12.8 experimental) of a value estimated from experimental pKa's of the reacting species.*

Advances in high-performance computing coupled with computational approaches based on first principles provide the technology to carry out accurate simulations of condensed-phase chemical phenomena. Theoretical methods that represent the noncovalent attributes of chemical systems have become quite sophisticated, and it is now possible to calculate structural and energetic properties of complex biochemical systems, such as proteins and nucleic acids, to reasonable degrees of accuracy (1). However, a realistic depiction of chemical reactions in the condensed phase, where bond-making and -breaking events result in significant changes to the electronic structure of solutes, is much more problematic. The simulation of chemical reactions to experimental accuracy can be attained through the use of computational quantum mechanical (QM) methods, which have been implemented in the form of Gaussian Hartree-Fock and related theories (2). However, the most reliable *ab initio* QM methods are computationally intensive, and chemical accuracy (energies within about 1-2 kcal/mol of experimental values) can be obtained only on small systems consisting of less than 10 heavy (nonhydrogen) atoms (3). Thus, the direct application of high-level *ab initio* QM methods to model condensed phase systems containing thousands of atoms is impractical.

To make the simulation of chemical reactions in the condensed phase feasible, approaches combining QM with classical molecular mechanics (MM) have been devised (4-7). In these hybrid QM/MM implementations, the portion of a chemical system in the condensed phase undergoing changes in electronic structure (typically < 50 atoms) are treated with a QM representation while an MM model is used for the remainder of the system (thousands of atoms), and the interaction between QM and MM portions of the system is based on a QM/MM mixed Hamiltonian. The viability of QM/MM methods in condensed-phase chemical applications is dependent on (i) the computational efficiency of the QM implementation, (ii) the accuracy of the QM method to calculate the electronic structure and energetic properties of a system, (iii) the realistic incorporation of solvation effects represented by the interactions between designated QM (solute) and MM (solvent) components of a system, and (iv) the availability of computational resources to carry out statistical mechanics calculations. An inverse relationship between the computational efficiency and accuracy of QM methods, (i) and (ii), accounts for the major limitation in the practical use of QM/MM

methods, whereas (iii) has been shown to be feasible (8,9), and (iv) is satisfied through the use of massively parallel computers.

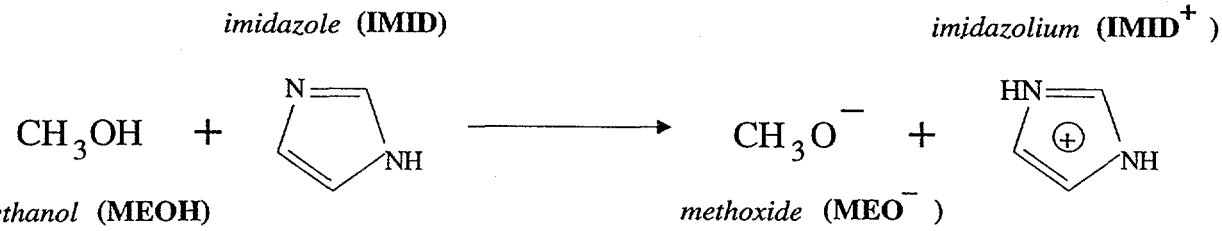
In this paper, we describe a systematic procedure to calibrate and use a QM/MM approach in complex heterogeneous molecular systems, which overcomes current limitations. Central to the success of this approach is the optimization of the QM method to reproduce experimental and/or calculated gas phase physical properties of molecular systems and the generation of QM/MM parameters to insure accurate treatment of interactions between QM and MM portions of the system. The resultant methodology can be used to model the electronic, structural, and energetic properties of condensed-phase molecular systems to levels near chemical accuracy.

We have previously developed a combined QM and MM approach (5,6) for the study of condensed-phase reactions, which consists of the semiempirical QM Austin Model 1 (AM1) (10) and MM (CHARMM) (11) methods. To use this QM/MM approach, a simulation system is partitioned into QM and MM regions. The energies $\left(E_{QM/MM}\right)$ of and the forces $\left(\mathbf{F}_{QM/MM}\right)$ on QM atoms are given by the expectation values of the QM Hamiltonian and its derivative, respectively, and include electrostatic and van der Waals interactions with MM atoms. The determination of $\mathbf{F}_{QM/MM}$ and $\mathbf{F}_{MM}$, the forces on MM atoms (which include effects due to QM atoms), enables energy minimization and classical molecular dynamics to be done in a standard manner (1). The computational efficiency of this semiempirical QM/MM method provides the means to calculate ensemble–averaged thermodynamic quantities such as free energies of reaction for complex condensed-phase chemical reactions (9,12).

Although the AM1 semiempirical QM method is efficient, its accuracy is problem-dependent (13). In most cases, semiempirical QM methods are not able to determine energetic properties of molecular systems to chemical accuracy. To overcome this limitation, we have developed a method that can be used to generate a QM/MM model for specific condensed-phase molecular systems, which enables the calculation of free energies of reaction to within experimental accuracy. Our approach comprises the following steps: (i) the partitioning of a system into QM (solute) and MM (solvent) regions, (ii) the optimization of parameters associated with the semiempirical QM method, (iii) the refinement of van der Waals parameters

on QM atoms, which are required for interactions between QM and MM atoms, and (iv) the use of molecular dynamics and free energy perturbation methods (5,9,12,14) on massively parallel computers to calculate free energies of reaction in the condensed phase.

To demonstrate the above procedure, we studied the energetics of the following proton transfer reaction, **R1**,



in gas phase and aqueous solution. The theoretical and experimental heats of formation (ΔH_f^0) of the various molecular species associated with **R1** are displayed in Table 1. The AM1–derived ΔH_f^0, using standard AM1 parameters, differ by as much as 20 kcal/mol from their experimental counterparts. The AM1 calculated heat of reaction, $\Delta\Delta H_f^0(\mathbf{R1}) = \Delta H_f^0(IMID^+) + \Delta H_f^0(MEO^-) - \Delta H_f^0(IMID) - \Delta H_f^0(MEOH)$ = 164.1 kcal/mol, is in better agreement to the experimental value of 157 kcal/mol, due to a fortuitous cancellation of errors in the heats of formation of the individual molecules of **R1**. The magnitude of these errors are representative of the limitations in the standard AM1 model, which may affect the accuracy attainable in the calculation of the free energies of reaction **R1** in aqueous solution.

The first step in the process to produce a simulation model capable of calculating values in close agreement to condensed-phase experimental data was the optimization of semiempirical QM parameters with respect to constraints derived from gas phase physical properties of the specific molecules under consideration. The problem, which has been outlined by Dewar and coworkers (15), was to find an optimal set of parameters, X_k $(k = 1,\ldots,K)$, by fitting a set of target values, Y_l $(l = 1,\ldots,L)$, of L properties in a basis set of M molecules. This can be done by minimizing the sum (Y) of weighted errors in the calculated values

$$Y = \sum_{l=1}^{L} \left| Y_l(calc) - Y_l'(\exp, ab\ inito) \right| w_l \qquad (1)$$

where w_l is a weighting factor for the quantity Y_l. The values of weighting factors emphasize different types of properties. For the above proton transfer reaction, we used a basis set of molecules consisting of methanol, imidazole, methoxide, and imidazolium. To optimize the AM1 parameters, X_k, for these molecules, the target values, Y_l', were (i) the experimental heats of formation (w_{heat} = 1.0 kcal^{-1}), (ii) experimental or *ab initio* (6-31G(d)) dipole moments (w_{dipole} = 30.0 Debye^{-1}), and (iii) the internal coordinates obtained from MP2/6-31G(d) *ab initio* QM calculations (weights, w_l, used for bond, angle, and dihedral constraints were 1.0 Å^{-1}, 5.0 degree^{-1}, 1.0 degree^{-1}, respectively). Table 1 gives the target experimental heats of formation and experimental/*ab initio* dipole moments used in Equation 1.

To search for an optimal set of parameters that minimize *Y*, we implemented a procedure within the context of our QM/MM method, similar to one described by Rossi and Truhlar (16), that uses a genetic algorithm (GA) to find a set of AM1 System Specific Parameters (AM1-SSP) appropriate for use in the study of reaction **R1**. The GA method we used resembles one described by Goldberg (17) with the addition of features that include the representation of variables as real numbers instead of bit patterns, a uniform distribution of cross-over points instead of one or two (18), and the use of a steady-state algorithm for population replacement (19) instead of the traditional generational replacement (17) genetic algorithm.

To obtain the AM1-SSP values shown in Table 2, we started with the standard AM1 parameters, which included all the types used in the fitting process to generate the original AM1 model except the Slater exponents (ζ_s and ζ_p) (10). A population of 300 chromosomes, each consisting of 42 genes (AM1 parameters), was used, where initial values for specific genes on each chromosome were selected from a random Gaussian distribution having a standard deviation of 10% and centered on the original AM1 value. The crossover and mutation probabilities were 0.7 and 0.01, respectively, and the mutated value for a particular gene was taken from a random Gaussian distribution that was centered on the current value for a gene with a standard deviation of 10%. The GA was run for *15,000* generations with 1% of the population selected for crossover (steady-state population replacement method) in each generation. The geometries of

basis set molecules, in each GA generation, were optimized until the gradient norm of the energy (forces on atoms) was less than 1.0 kcal/Å. The resultant bond lengths, angles, and dihedrals were used as variables (Y_I(calc)) together with calculated heats of formation and dipole moments. The final AM1-SSP parameters are given in Table 2, and the calculated molecular properties are listed in Table 1. The experimental and AM1-SSP calculated heats of formation are within 1.0 kcal/mol and dipole moments agree to within 0.3 Debye of experiment or *ab initio* QM derived values. The AM1-SSP calculated geometries differ from the *ab initio* target values for bonds, angles, and dihedrals by 0.011 ± 0.008 Å, 1.06 ± 0.97 degrees, and 0.19 ± 0.14 degrees, respectively. The result is a chemically accurate and computationally efficient QM model for the compounds in **R1**.

The next step in the process was to generate a QM/MM model, which was calibrated with respect to a microsolvent environment. To represent solvation effects realistically, it was necessary that the QM/MM Hamiltonian accurately depicted the interactions between solute atoms (described by QM) and solvent atoms (described by MM). In our QM/MM method, QM and MM atoms interact through (i) the one-electron Hamiltonian via QM electron and MM "core" partial charges, (ii) QM core positive charges and MM "core" partial charges, and (iii) QM and MM van der Waals interactions. Term (iii) models electronic repulsion and dispersion interactions, which do not exist between QM and MM atoms because MM atoms possess no explicit electrons. Within the framework of this formalism, QM/MM interactions can be calibrated by adjusting the van der Waals parameters of QM atoms such that the interaction energies between QM and MM atoms emulate those determined from *ab initio* QM calculations or experimental data. Specifically, we match the interaction energy between a water molecule (MM) and a functional group (QM) calculated by our QM/MM Hamiltonian to those determined at the HF 6-31G(d) level of theory, which is similar to procedures described previously (8,9,20). The 6-31G(d) basis set has been shown to reproduce interaction energies for hydrogen bonding complexes with good accuracy (2,21), and therefore provided a reasonable standard for QM/MM calibrations. Figure 1 shows solute-water interaction geometries used to generate van der Waals parameters for solute (QM) atoms. Tables 3 and 4 list the minimum interaction energies and geometries between solute and water and the resultant van der Waals

parameters, respectively. Satisfactory agreement between *ab initio* and QM/MM interaction energies and structures was obtained.

These two calibration procedures produced a chemically accurate QM/MM model with respect to gas phase and microsolvation properties for the molecular species of reaction **R1**. In the next step, we calculated the proton transfer free energy for reaction **R1** in aqueous solution using statistical mechanics theory with a QM/MM free energy perturbation method that we have developed (9).

In the free energy perturbation formalism, a system is characterized by a Hamiltonian $H(\mathbf{p},\mathbf{q},\lambda)$ which depends parametrically on a multidimensional coupling parameter λ and is a function of phase space coordinates **p** and **q**. The free energy difference between two states A and B can be calculated by the following procedure. A discrete pathway consisting of N intermediate states λ_i $(1 \le i \le N$ such that $A \equiv \lambda_1 \le \lambda_i \le \lambda_N \equiv B)$ is constructed to connect states A and B in a series of free energy perturbation windows. Free energy differences along the pathway can then be calculated using

$$\Delta G_i(\lambda_i \rightarrow \lambda_i \pm \Delta\lambda) = -RT \ln\left\langle \exp\left\{-\left[H(\lambda_i \pm \Delta\lambda) - H(\lambda_i)\right]/RT\right\}\right\rangle_{\lambda_i}, \qquad (2)$$

which is the free energy difference between a state λ_i and its two neighbors $\lambda_{i\pm1} = \lambda_i \pm \Delta\lambda$ (R: universal gas constant; T: absolute temperature; $\langle\ldots\rangle_\lambda$: ensemble average calculated with the probability density of state λ). The free energy changes of all intermediate perturbation windows along the pathway are summed to obtain the total free energy difference between states A and B, i.e.,

$$\Delta G(A \rightarrow B) = \sum_{i=1}^{N-1} \Delta G_i(\lambda_i \rightarrow \lambda_i + \Delta\lambda), \qquad (3)$$

$$\Delta G(B \rightarrow A) = \sum_{i=1}^{N-1} \Delta G_{N-i+1}(\lambda_{N-i+1} \rightarrow \lambda_{N-i+1} - \Delta\lambda). \qquad (4)$$

In practice, the number of perturbation windows (N) is chosen so that the phase space of the three states represented by λ_i, and $\lambda_i \pm \Delta\lambda$ can be adequately sampled. The protocol employed (see Figure 2) for the free energy simulations described below was identical to one used in a previous study (9).

For each perturbation window, free energies were calculated for the "forward" and "backward" directions (λ_i to $\lambda_i \pm \Delta\lambda$) as defined by Equation 2. The total free energy changes $\Delta G_{proton}(forward)$ and $\Delta G_{proton}(backward)$ were determined using Equations 3 and 4, respectively, and an estimation of the error in the calculated free energy (defined as the hysteresis) is given by

$$\Delta G_{proton}(hysteresis) = \left|\Delta G_{proton}(forward) - \Delta G_{proton}(backward)\right|. \qquad (5)$$

A plot of the free energy changes for the forward and backward directions are shown in Figure 2. The total hysteresis for the proton transfer simulations was less than 0.5 kcal/mol, which suggests that adequate phase space sampling had been achieved along the perturbation pathway.

The energy required to transfer a proton from methanol to imidazole in the gas phase is about 157 kcal/mol (Table 1), whereas the free energy change in water at 298 K is only 12.8 kcal/mol. The free energy of transfer in aqueous solution is calculated from the experimental pK_a values of methanol (22) and imidazole (23), which are 15.5 and 6.05 respectively, using $\Delta G_{solvent}(\mathbf{R1}) = 2.3RT\{pK_a(\mathbf{MEOH}) - pK_a(\mathbf{IMID})\}$. There is a significant energetic stabilization effect due to the solvation of methoxide (charge -1) and imidazolium (charge $+1$) relative to the uncharged methanol and imidazole species. Our simulated proton free energy change of 12.2 kcal/mol (-11.9 for the backward transformation) compares favorably to the experimental value of 12.8 kcal/mol and appears to account for the stabilization effects by the water solution. This close agreement between calculated and experimentally derived proton transfer free energies suggests that our procedure, which produces a system-specific QM/MM model, can result in the determination of condensed phase chemical reaction phenomena to chemical accuracy. These results are due to (i) the ability of the GA optimized AM1 model to accurately reproduce the heats of formation and gas phase dipole moments for the molecular species of reaction **R1**,

(ii) the calibration of QM/MM interaction energies, which lead to a reasonable representation of solvation energetics, and (iii) the adequate sampling of phase space along the pathway chosen for the perturbations as obtained through MD simulations carried out on a massively parallel computer.

The systematic procedure described in this paper can be applied to any condensed-phase system where gas phase experimental data or high-level *ab initio* calculations are available to calibrate a hybrid semiempirical QM and MM method. This new capability provides the means to simulate and analyze, to near chemical accuracy, the structural, energetic, and kinetic properties for a wide variety of chemical reactions in solvents that range from simple homogeneous aqueous solutions to the complex heterogeneous protein environments of enzymes.

References and Notes

1. C. L. Brooks III, M. Karplus, B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics* (Adv. Chem. Phys. LXXI; John Wiley & Sons 1988).
2. W. J. Hehre, L. Radom, P. Scheyer, J. A. Pople, *Ab Initio Molecular Oribital Theory* (John Wiley & Sons, New York,, 1986).
3. L. A. Curtiss and K. Raghavachari, in *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy: Understanding Chemical Reactivity,* F. R. Langhoff, Eds. (Klueer Academic Press, Netherlands, 1995), pp.139.
4. A. Warshel and M. Levitt, *J. Mol. Biol.* **103**, 227 (1976).
5. P. A. Bash, M. J. Field, M. Karplus, *J. Amer. Chem. Soc.* **109**, 8092 (1987).
6. M. J. Field, P. A. Bash, M. Karplus, *J. Comp. Chem.* **11**, 700 (1990).
7. A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solution* (Wiley, New York, 1991).
8. J. Gao, in *Modeling the Hydrogen Bond*, D. A. Smith, Eds. (ACS Symp. Series 569, ACS, 1994).
9. L. L. Ho, A. D. MacKerell, Jr., P. A. Bash, *J. Am. Chem. Soc.* (submitted), (1995).
10. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **107**, 3902 (1985).
11. B. Brooks et al., *J. Comp. Chem.* **4**, 187 (1983).
12. J. Gao and X. Xia, *J. Am. Chem. Soc.* **115**, 9667 (1993).
13. J. J. P. Stewart, *MOPAC manual and references contained therein* (1993).
14. P. A. Bash, U. C. Singh, R. Langridge, P. A. Kollman, *Science* **236**, 574 (1987).
15. M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.* **99**, 4899 (1977).
16. I. Rossi and D. G. Truhlar, *Chem. Phys. Lett.* **233**, 231 (1995).

17. D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, 1989).

18. G. Syswerda and J. Schaffer, *Proceedings of the Third International Conference on Genetic Algorithms* (1989) pp.2.

19. D. Whitley and J. Kauth, *Rocky Mountain Conference on Artificial Intellegence* (1988), pp. 118.

20. A. D. MacKerell, Jr., J. Wiorkiewicz-Kuczera, M. Karplus, *J. Am. Chem. Soc.*, (in press).

21. J. Pranata, S. G. Wierschke, W. L. Jorgensen, *J. Am. Chem. Soc.* **113**, 2810 (1991).

22. H. Maskill, *The Physical Basis of Organic Chemistry* (Oxford University Press, New York, 1989).

23. R. M. C. Dawson, D. C. Elliot, W. H. Elliot, K. M. Jones, *Data for Biochemical Research* (Oxford University Press, New York, Third Edition, 1986), .

24. M. Meot-Ner and L. W. Sieck, *J. Phys. Chem.* **90**, 6687 (1986).

25. A. D. MacKerell, Jr. and M. Karplus, *J. Phys. Chem.* **95**, 10559 (1991).

26. W. L. Jorgensen, J. Chandrasekhar, J. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).

27. C. L. Brooks III and A. Brünger, *Biopolymers* **24**, 843 (1985).

28. C. L. Brooks III and M. Karplus, *J. Mol. Biol.* **208**, 159 (1989).

29. W. F. van Gunsteren and H. J. C. Berendsen, *Mol. Phys.* **34**, 1311 (1977).

31 S. G. Lias et al., *J. Phys. & Chem. Ref. Data* **17**, (1988).

32 A. L. McClellan, *Tables of Experimental Dipole Moments Vol. 2* (Rahara Industry, CA, 1974).

33. The authors gratefully acknowledge use of the Argonne High-Performance Computing Research Facility. The HPCRF is funded principally by the U.S. Department of Energy Office of Scientific Computing. This work was also supported by the U.S. Department of Energy Office of Health and Environmental Research, under Contract no. W-31-109-Eng-38.

Table 1. Experimental and calculated heats of formation and dipole moments of methanol, methoxide, imidazole, and imidazolium.

Experimental and Calculated Physical Observables				
	Molecules			
Physical Observables (method)	**MEOH**	**MEO^-**	**IMID**	**IMID+**
ΔH_f^0 *(Exp)*[a]	-48.30	-33.20 ± 2.40	35.00 ± 0.50	177.00
ΔH_f^0 *(AM1)*	-56.02	-38.52	50.76	196.31
ΔH_f^0 *(AM1–SSP)*	-48.14	-32.59	34.83	177.43
$\lvert\vec{\mu}\rvert$ *(Exp)*[b]	1.70	—	3.80	—
$\lvert\vec{\mu}\rvert$ *(AM1)*	1.62	1.38	3.60	1.63
$\lvert\vec{\mu}\rvert$ [6–31*G*(d)]	1.87	2.16	3.86	1.74
$\lvert\vec{\mu}\rvert$ *(AM1–SSP)*	1.97	2.09	3.69	1.76

ΔH_f^0	heat of formation at 298 K (kcal/mol).
Exp	experimental value.
*AM*1	standard AM1 parameters (10).
*AM*1–*SSP*	AM1 system specific parameters.
6–31*G*(d)	HF 6–31*G*(d) level of *ab initio* theory.
$\lvert\vec{\mu}\rvert$	magnitude of dipole moment vector (Debye).
a	reference (31).
b	reference (32).

Table 2 AM1 (10) and AM1 system specific (AM1–SSP) parameters for hydrogen, carbon, nitrogen, and oxygen. The AM1–SSP parameters are optimized for methanol, methoxide, imidazole, and imidazolium.

Semiempirical AM1 parameters								
	AM1				*AM1–SSP*			
param	**H**	**C**	**N**	**O**	**H**	**C**	**N**	**O**
U_{SS}	-11.396427	-52.028658	-71.860000	-97.830000	-10.690510	-53.180427	-75.073641	-104.113866
U_{PP}		-39.614239	-57.167581	-78.262380		-38.767876	-57.341356	-78.275023
$*\zeta_S$	1.188078	1.808665	2.315410	3.108032	1.188078	1.808665	2.315410	3.108032
$*\zeta_P$		1.685116	2.157940	2.524039		1.685116	2.157940	2.524039
β_S	-6.173887	-15.715783	-20.299110	-29.272773	-6.422484	-16.242199	-20.380228	-33.109001
β_P		-7.719283	-18.238666	-29.272773		-7.272735	-16.121743	-27.876218
α	2.882324	2.648274	2.947286	4.455371	2.893675	2.809818	3.248052	4.235416
K_1	0.122796	0.011355	0.025251	0.280962	0.135194	0.011703	0.026569	0.293314
K_2	0.005090	0.045924	0.028953	0.081430	0.005108	0.042409	0.028418	0.078530
K_3	-0.018336	-0.020061	-0.005806		-0.018117	-0.021220	-0.004842	
K_4		-0.001260				-0.001600		
$*L_1$	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000
$*L_2$	5.000000	5.000000	5.000000	7.000000	5.000000	5.000000	5.000000	7.000000
$*L_3$	2.000000	5.000000	2.000000		2.000000	5.000000	2.000000	
$*L_4$		5.000000				5.000000		
M_1	1.200000	1.600000	1.500000	0.847918	1.108805	1.415666	1.457421	0.907351
M_2	1.800000	1.850000	2.100000	1.445071	1.799327	2.033559	2.067424	1.655841
M_3	2.100000	2.050000	2 400000		2.254171	1.832265	2.455736	
M_4		2.650000				2.408368		

Unit of parameters is in electron volt (eV).
* parameter not optimized.

Table 3. Solute–solvent minimum interaction energies (kcal/mole) and distances (Å) between solute atom "*X*" and water oxygen atom "OW" (Figure 1), and equilibrium angles (Figure 1) calculated by *ab initio* Gaussian HF 6–31G(d) and QM/MM methods. Solute name (e.g. IMID) and water–solute orientation number in column 1 corresponds to the molecular complexes displayed in Figure 1. The experimental interaction energy between water and methoxide has been measured to be –23.9 kcal/mol (24). The HF 6–31G(d) interaction energies have been scaled by a constant factor of 1.16 for uncharged molecules (25).

Solute–Solvent Interaction Energies				
Molecule	**6–31G(d)**		**AM1–SSP/MM**	
(orientation)	*Energy*	*d(OW–X) (angle)*	*Energy*	*d(OW–X) (angle)*
IMID (1)	-2.44	3.54 (CE)	-2.51	3.57 (CE)
IMID (2)	-7.26	3.09 (NE2)	-7.31	2.83 (NE2)
IMID (3)	-0.99	3.75 (CD)	-1.24	3.67 (CD)
IMID (4)	-2.40	3.60 (CG)	-2.36	3.63 (CG)
IMID (5)	-6.65	3.16 (ND)	-6.59	2.87 (ND)
IMID+ (1)	-12.28	3.12 (CE)	-11.21	3.31 (CE)
IMID+ (2)	-9.53	3.22 (CD)	-10.29	3.21 (CD)
IMID+ (3)	-15.94	2.94 (ND)	-16.65	2.77 (ND)
MEO^- (1)	-18.75	2.67 (O2)	-20.30	2.44 (O2)
MEO^- (2)	-20.95	2.67 (O2)	-21.29	2.47 (O2)
MEO^- (3)	-20.95	2.67 (O2)	-21.29	2.47 (O2)
MEO^- (4)	-23.96	2.71 (O2) (θ=115)	-23.97	2.51 (O2) (θ=127)
MEO^- (5)	-6.89	2.44 (C2)	-6.96	2.25 (C2)
MEOH (1)	-5.73	3.01 (O2)	-6.60	2.53 (O2)
MEOH (2)	-3.18	3.01 (O2)	-7.31	2.51 (O2)
MEOH (3)	-3.94	3.02 (O2) (θ=126)	-5.19	2.68 (O2) (θ=110)
MEOH (4)	-6.72	3.01 (O2) (φ=125)	-5.41	2.66 (O2) (φ=127)

Table 4. The van der Waals parameters of solute atoms (molecule and atom names correspond to those in Figure 1) for QM/MM and MM [CHARMM (20)].

van der Waals Parameters				
Molecule	**AM1–SSP/MM**		**MM (CHARMM22)**	
(Atom)	ε (kcal)	r* (Å)	ε (kcal)	r* (Å)
IMID, IMID+ (CE, CD)	-0.1000	2.5000	-0.0500	1.8000
IMID, IMID+ (CG)	-1.0000	1.7000	-0.0500	1.8000
IMID, IMID+ (NE2)	-0.0800	1.9000	-0.2000	1.8500
IMID, IMID+ (ND)	-0.0300	1.9000	-0.2000	1.8500
MEO^-, MEOH (C2)	-0.0800	2.0600	-0.0800	2.0600
MEO^-, MEOH (O2)	-0.0021	2.0000	-0.1200	1.7000

Figure Captions

Figure 1: Solute–solvent complexes for QM/MM interaction energy calibrations.

1 a. IMID – imidazole

1 b. $IMID^+$ – imidazolium

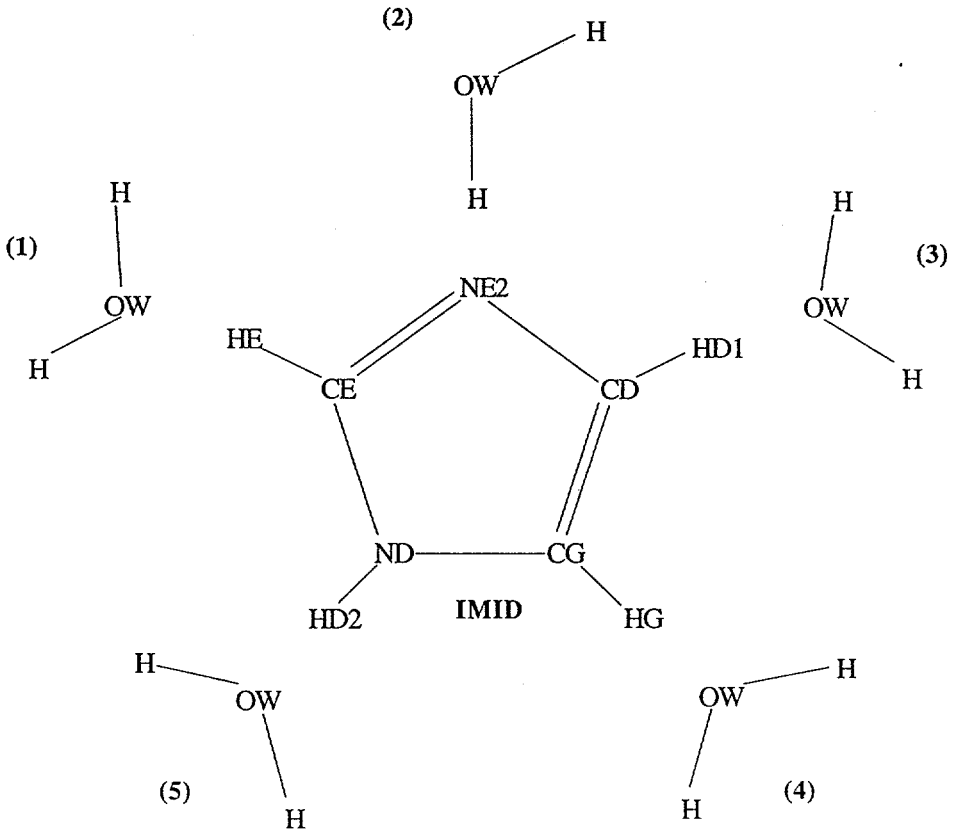
1 c. MEO^- – methoxide

1 d. MEOH – methanol

Figure 2. Free energy "profile" for the proton transfer reaction **R1**. To represent a bulk solution of water, solute molecules are immersed in an 18-Å radius ball of TIP3P water. (26) A deformable stochastic boundary (27,28) with a reaction zone of 16 Å and a buffer region of 2 Å was imposed on the reaction system, and molecular dynamics was used to sample phase space. The reaction coordinates, λ_{proton}, were constrained using SHAKE (29), to specific values along the free energy perturbation pathway. This pathway consists of a series of changes in the two reaction coordinates, which are defined by the distance between the hydroxyl oxygen and proton of methanol and the distance between this proton and the unprotonated nitrogen of the imidazole. The system was transformed from a methanol-imidazole complex to a methoxide-imidazolium complex in increments of 0.05 Å along the defined reaction coordinates. This produced 70 perturbation windows for the transformation between reactants and products. Calculations for equilibration and data collection were done simultaneously on different processors of a massively parallel computer. The points along the ordinate correspond to the reaction coordinates. The abscissa gives the cumulative free energies (kcal/mol) along this reaction coordinate for the forward (——) and reverse directions (– · –). The difference in the free energies between the points A and B are the calculated free energy changes for the proton transfer reaction, which are 12.2 forward and -11.9 backward. Each window

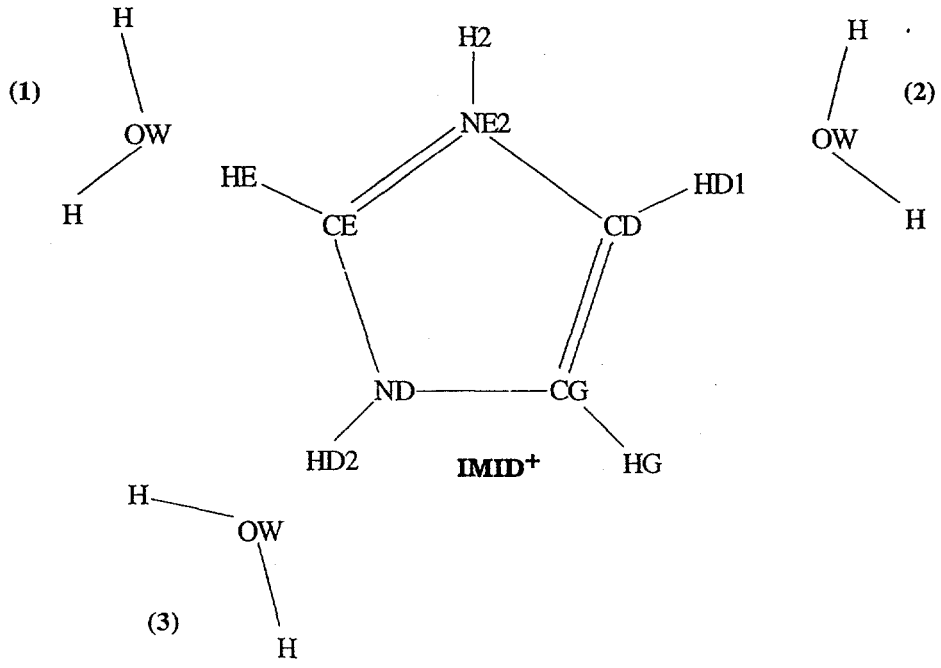
was equilibrated for 20 ps and data collected for 10 ps using 1 fs molecular dynamics time steps for a total simulation of 2.1 ns. The free energy changes after 5 ps of data collection per window were 12.1 and -11.6 for forward and backward directions, respectively (data not shown).

Figure 1 (**1a** to **1d**)



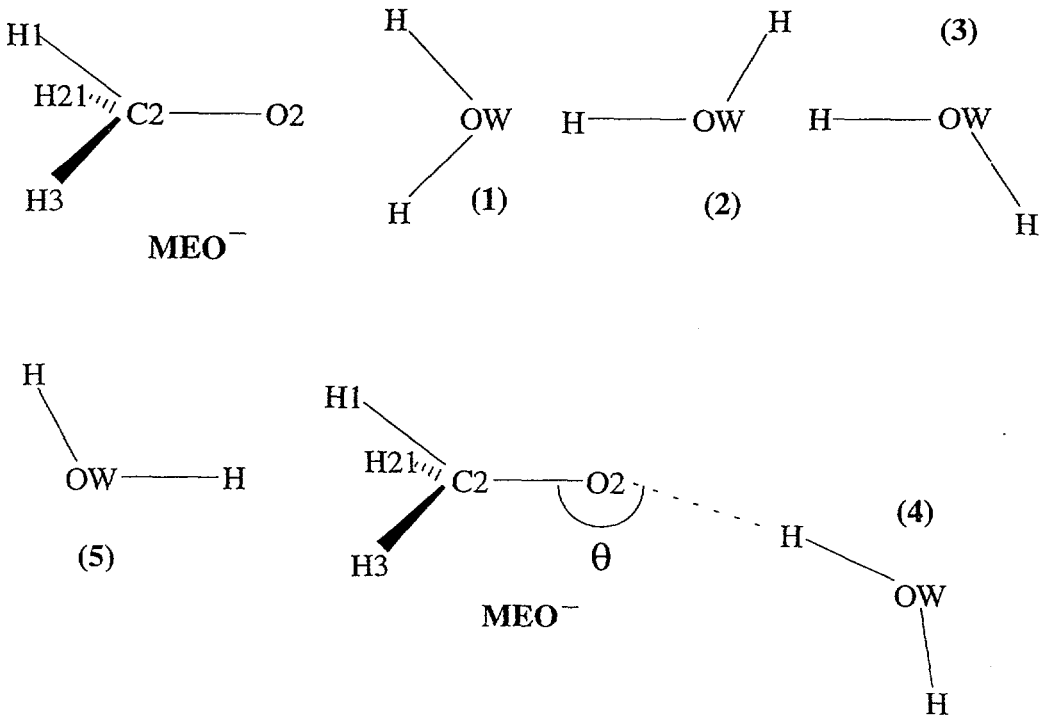
1 a

Figure 1 (**1a** to **1d**)



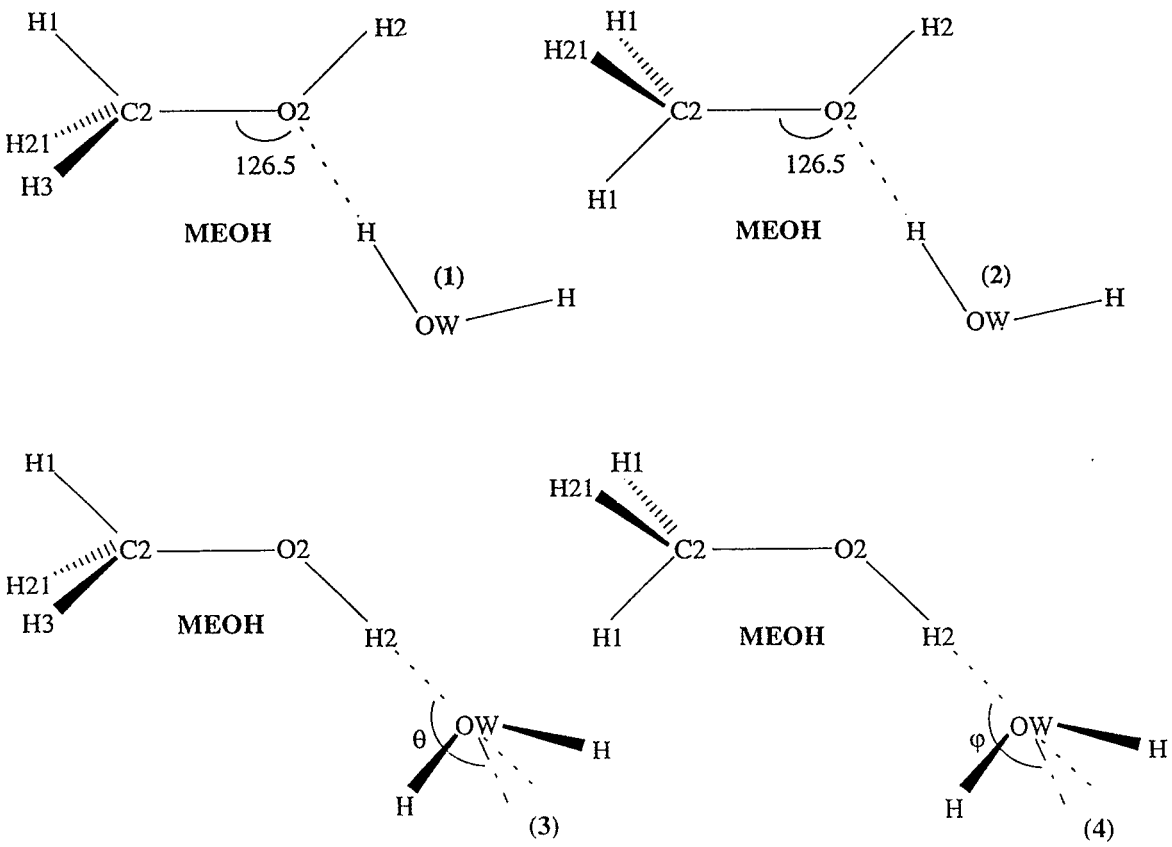
1 b

Figure 1 (1a to 1d)



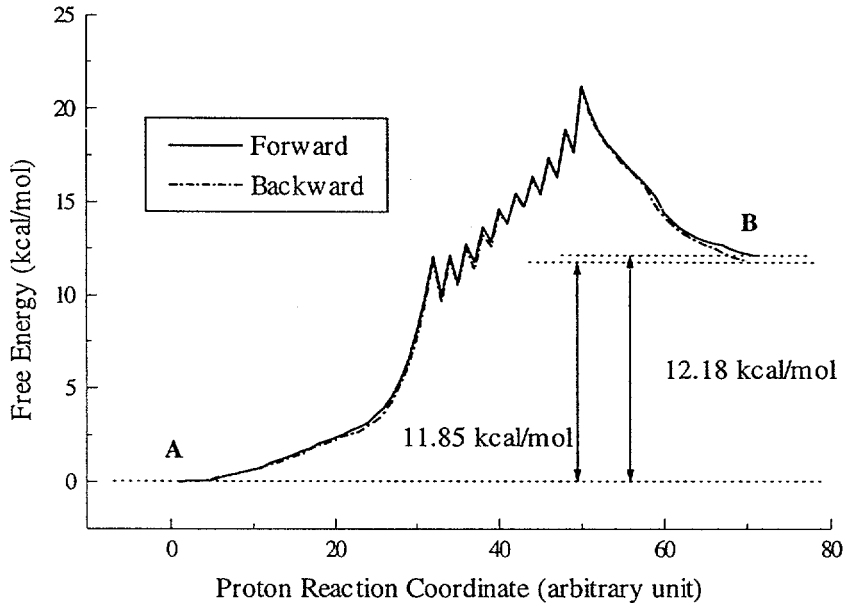
1 c

Figure 1 (1a to 1d)



1 d

Figure 2


25
20
15
10
5
0
Free Energy (kcal/mol)
Forward
Backward
A
B
11.85 kcal/mol
12.18 kcal/mol
0
20
40
60
80
Proton Reaction Coordinate (arbitrary unit)