Ab Initio Molecular Dynamics Study of the Reaction of Water with Formaldehyde in Sulfuric Acid Solution

Evert Jan Meijer* and Michiel Sprik†

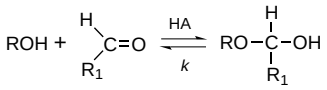
Contribution from the IBM Research Division, Zurich Research Laboratory, Säumerstrasse 4, CH-8803 Rüschlikon, Switzerland

Received August 20, 1997. Revised Manuscript Received April 28, 1998

Abstract: Ab initio molecular dynamics methods have been used to study the reaction mechanism of acid-catalyzed addition of water to formaldehyde in a model system of an aqueous solution of sulfuric acid. Using the method of constraints we find that an H_2O molecule can be added to formaldehyde by a controlled transfer of a catalytic proton from a hydronium ion in acid solution to the carbonyl oxygen. The formation of the CO bond between the carbonyl carbon and the water oxygen occurs at a stage midway in the proton transfer process. The process can be reversed by removing the H^+ from the protonated product diol, leading to CO bond breaking at approximately the same stage of proton transfer. This suggests that the kinetics of the acid-catalyzed reaction is governed by a concerted protonation and addition.

1. Introduction

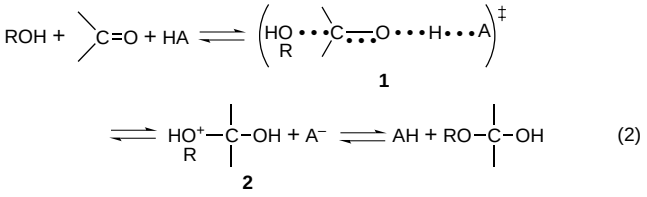
The addition of nucleophiles to the carbonyl group is a fundamental reaction in organic chemistry. Many of these reactions are reversible and are catalyzed in both directions by acids and bases. In particular, acid catalysis of reactions involving aldehydes serves as a primary example of the chemistry induced by the exchange of protons in an aqueous environment (see, e.g., refs 1 and 2). Reaction 1 specifies the reactants and products for acid-catalyzed addition of water and alcohols to aldehyde carbonyl groups.



(1)

Research on the addition reaction 1 and the reverse decomposition of alcohols has a long tradition. Some of the key contributions can be found in refs 3−7. The status of reaction 1 as a popular model system in physical organic chemistry is related to the large amount of information that can be obtained from structure−reactivity studies. The reason is that reaction 1 is a case of general acid catalysis, meaning that not only does substitution of the R and R_1 groups have an effect on the rate coefficient k but also the nature of the conjugate base A^- of acid HA, which supplies the catalyzing H^+ ion. In fact, the dependence of log(k) on the pK_a of HA is approximately linear and the slope α of these Bronsted plots is an important parameter characterizing reaction 1.[1,2] The value of α can be correlated with a similar linear coefficient β for the dependence of log(k) on the pK of the leaving group ROH in the decomposition reaction, thus providing further insight in the reaction mechanism.

Interpreting their experimental data in terms of these correlations, Jencks and co-workers[6,7] proposed a detailed mechanism for the general acid catalysis of reaction 1. Their conclusion was that transfer of the catalytic proton and nucleophilic attack are concerted processes in which the anion A^- plays an active role. Reaction 2 summarizes the main reaction steps as inferred in refs 1 and 2. For convenience, explicit representation of the H atom and R_1 group bonded to the carbon has been omitted.



(2)

Complex **1** is an unstable transition state whereas **2** is considered a metastable tetrahedral intermediate.

Details such as the structure of the transition state **1** in reaction 2 are somewhat hypothetical. They have been introduced as explanations for the observed kinetic behavior but are not so easily accessible to direct verification by structural and dynamical experimental probes. Therefore, computational chemistry methods, which provide a microscopic basis for kinetic pathways, could be a useful complement to experiment by resolving issues concerning reaction mechanisms. Computer simulation of chemical reactions in aqueous solutions, however, requires a realistic treatment of the changes in electronic states leading to bond making and breaking as well as the finite temperature dynamics of a condensed system in the liquid phase. These conditions are met by numerical methods that apply the latest advances in the density functional theory (DFT) approach[8] to

* Present address: Department of Physical and Theoretical Chemistry, Vrije Universiteit, NL-1081 HV Amsterdam, The Netherlands.

† Present address: Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

(1) Stewart, R. *The Proton: Applications to Organic Chemistry*; Academic Press: Orlando, 1985.

(2) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; Dover: New York, 1987.

(3) Bell, R. P.; Rand M. H.; Wynne-Jones, K. M. A. *Trans. Faraday Soc.* **1956**, *52*, 1093.

(4) Pocker, Y. *Proc. Chem. Soc.* **1960**, 17.

(5) Critchlow, J. E. *J. Chem. Soc., Faraday Trans. 1* **1972**, *68*, 1774.

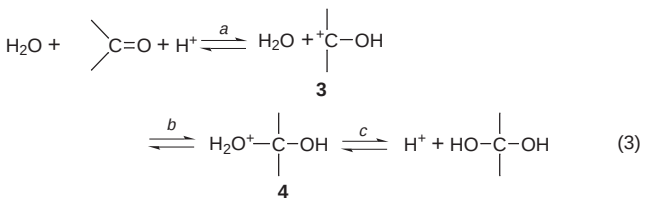
(6) Funderburk, L. H.; Aldwin, L.; Jencks, W. P. *J. Am. Chem. Soc.* **1978**, *100*, 5444.

(7) Sørensen, P. E.; Jencks, W. P. *J. Am. Chem. Soc.* **1987**, *109*, 4675.

S0002-7863(97)02935-1 CCC: $15.00
Published on Web 06/14/1998

electronic structure calculation and the ab initio molecular dynamics (MD) technique.[9] The development of efficient and accurate generalized gradient approximations (GGA) for the exchange[10] and correlation[11] energy functional was a crucial step in this context (for reviews on the GGA see, e.g., refs 8, 12, and 13). These functionals are used in the continuously updated electronic structure calculation from which the forces for the MD propagation are determined.[9,14] Previous applications of DFT/ab initio MD to aqueous systems have verified that this approach can be used to treat hydrogen bonding in the condensed phase. Ab initio MD studies[15,16] have demonstrated that the structure and dynamics of pure liquid water can be reproduced if a suitable GGA is used. In refs 17−19 these investigations were extended to simple aqueous solutions of nonneutral pH. The results of these tests confirmed that the range of time scales and system sizes accessible to ab initio MD is adequate for observation of spontaneous proton exchange between solute and solvent molecules. The first application to organic chemistry in the condensed phase was a study of a reaction in a nonaqueous medium, namely the proton-catalyzed chemistry in a liquid mixture of formaldehyde and trioxane.[20] In this work also the use of constraint methods[21] to calculate relative free energies along the reaction path defined by a constrained reaction coordinate was explored. The reaction studied in the present work is activated. For example, in pure water the activation energy is measured to be $E_a = 67$ kJ mol^{-1}.[22] A successful reaction is therefore a rare event and outside the ab initio MD time scale. To study such reactions by MD a reactive encounter must be forced on the system by some form of microscopic control of a suitable reaction coordinate. This, in fact, is the main purpose of applying constraint methods in MD simulations of activated processes[21] and is also the approach adopted here.

In the present ab initio MD study we examine a simple instance of reaction 1, namely the addition of water to formaldehyde (R = R_1 = H) in a 8.5 mol % H_2SO_4 aqueous solution. In reaction 3 we have resolved this reaction in three elementary steps: protonation of formaldehyde (a), nucleophilic attack (b), and deprotonation yielding the diol product (c).



(3)

Sequence 3 is a schematic representation. Whether the steps are cooperative or the carbo cation **3** and oxonium ion **4** correspond to real stable species is the subject of our investigation. A further simplification in reaction 3 is the omission of the conjugate base of the catalytic proton [compare with reaction 2]. This important issue is closely related to our choice of a controlled degree of freedom used to overcome the activation barrier in the simulation, discussed below.

The outline of this paper is as follows. Details concerning the computation of the electronic structure, the MD method, and the definition of the constrained reaction coordinate are summarized in section 2. Gas-phase geometries and energies of the species participating in the reaction are discussed in section 3. The purpose of including gas-phase data is to establish the accuracy of our DFT calculations by comparing them to results obtained by other computational methods and experiment. The definition and preparation of the MD system modeling the reactive solution are outlined in section 4. In section 5 the results of the condensed phase simulation are presented. We conclude with a critical discussion. The appendix contains technical comments on the calculation of free energies in the modification of the method of constraints used here.

2. Theory and Numerical Techniques

1. Density Functional Method. Electronic structures in the ab initio MD method[9,14] are computed with DFT in the Kohn−Sham formulation (see, e.g., ref 8). The Kohn−Sham orbitals are expanded in a plane-wave basis set and fully occupied with a pair of electrons of opposite spin. Only valence electrons are considered explicitly, and in the present application, semilocal norm-conserving pseudopotentials were used to account for interactions between the valence electrons and the ionic cores. The calculations were performed with the ab initio MD package of ref 23.

For hydrogen we employed a simple analytic pseudopotential, which is essentially a softened Coulomb potential (error function) with a pseudization radius of r_c= 0.25 a_0. The pseudopotentials for C, O, and S were generated with the Troullier−Martins procedure.[24] For a given element, the pseudization radius r_c is identical for all orbital components. For carbon and oxygen we used a potential consisting of $l = s$ and $l = p$ terms with $r_c = 1.23$ a_0 and $r_c = 1.11$ a_0, respectively. For sulfur a $l = d$ component was also included and r_c was set to 1.34 a_0. The potentials for C and O are derived from the atomic ground-state configuration. To account for the $l = d$ anisotropy for S, we used as the reference state for the S potential an excited configuration with a fractional occupation of 1.75 for the 3*s* orbital, 3.25 for 3*p*, and 0.5 for 3*d*. For application in the ab initio MD code, the potentials were transformed into Kleinman−Bylander form[25] by taking the maximum *l* term to be the local part.

The plane-wave basis set employed consists of all waves with a maximum kinetic energy of $E_{cut} = 70$ Ry and smaller. For a validation of the use of a plane-wave basis set of this size for the H, C, and O pseudopotentials, we refer to previous work.[16,20,26] The parametrization for the S pseudopotential was

(8) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University: New York, 1989.
(9) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.
(10) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(11) Lee, C.; Yang, W.; Parr, R. *Phys. Rev. B* **1988**, *37*, 785.
(12) Perdew, J. P.; Burke, K. *Int. J. Quantum Chem.* **1996**, *57*, 309.
(13) Neumann, R.; Nobes, R. H.; Handy, N. C. *Mol. Phys.* **1996**, *87*, 1.
(14) Galli, G.; Pasquarello, A. In *Computer Simulation in Chemical Physics*; Allen, M. P., Tildesley, D. J., Eds.; NATO ASI Series C; Kluwer: Dordrecht, 1993; p 261.
(15) Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. *J. Chem. Phys.* **1993**, *99*, 9080.
(16) Sprik, M.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1996**, *105*, 1142.
(17) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. *J. Phys. Chem.* **1995**, *99*, 5749. Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. *J. Chem. Phys.* **1995**, *103*, 150.
(18) Laasonen, K. E.; Klein, M. L. *J. Phys. Chem. A* **1997**, *101*, 98.
(19) Meijer, E. J.; Sprik, M. *J. Phys. Chem. A* **1998**, *102*, 2893.
(20) Curioni, A.; Sprik, M.; Andreoni, W.; Schiffer, H.; Hutter, J.; Parrinello, M. *J. Am. Chem. Soc.* **1997**, *119*, 7218.
(21) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472.
(22) Williams, I. H.; Spangler, D.; Femec, D. A.; Maggiora, G. M.; Schowen R. L. *J. Am. Chem. Soc.* **1983**, *105*, 31.
(23) Hutter, J.; Parrinello, M. Car-Parrinello MD code (CPMD) developed by J. Hutter and the group of numerical intensive computing at IBM Zurich Research laboratory and the group of M. Parrinello at MPI in Stuttgart.
(24) Troullier, N.; Martins, J. *Phys. Rev. B* **1991**, *43*, 1993.
(25) Kleinman, L.; Bylander, D. *Phys. Rev. Lett.* **1982**, *48*, 1425.
(26) Meijer, E. J.; Sprik, M. *J. Chem. Phys.* **1996**, *105*, 8684.

subjected to extensive accuracy and transferability tests in an investigation of a number of small sulfur−sulfur, carbon−sulfur, and oxygen−sulfur compounds. The results of these tests will be included in a separate, more technical publication.[27] Details of the structure and energetics of the SO_3, H_2SO_4, and complexes with H_2O obtained with the same H, O, and S pseudopotentials and basis-set size can be found in ref 19.

We applied a gradient-corrected density functional, which is widely used in DFT calculations of molecular properties.[12,13] The exchange energy density is approximated by the local Slater−Dirac function, extended with terms dependent on the gradient of electronic density according to Becke.[10] The correlation energy is described by a functional developed by Lee, Yang, and Parr.[11] The literature evaluating this combination of gradient-corrected exchange and correlation functionals, usually indicated by the acronym BLYP, is rapidly growing. The accumulated experience shows that, on average, the accuracy of molecular geometries and chemical bonding energies is at the level of standard correlated Hartree−Fock-based methods (see, e.g., refs 28−31). Moreover, on the more subtle scale of energy differences of hydrogen-bonded interactions in condensed aqueous systems, the performance of the BLYP is also satisfactory[15−18] (for van der Waals systems see, e.g., ref 26).

2. Molecular Dynamics. We performed standard ab initio MD simulations at fixed volume and constant temperature. Such a simulation generates a finite temperature trajectory of the configuration of atoms under adiabatic ground-state conditions for the electrons. The characteristic feature of this approach is that the electronic wave function, i.e., the coefficients of the plane-wave basis set, are dynamically optimized to be consistent with the changing positions of the atomic nuclei. The actual implementation involves the numerical integration of the equations of motion of second-order Newtonian dynamics. A crucial parameter in this scheme is the fictitious mass m_e associated with the electronic degrees of freedom. In practice, the value of m_e is determined by the balance of two contradicting requirements. It should be sufficiently small to keep the electronic wave function close to the ground state and sufficiently large to have a workable upper bound of the time step used in the numerical integration. In the present work, the value of m_e is set to 1000 au $= 1000 \times 9.1093910^{-31}$ kg, which limits the time step to $\delta t = 0.17$ fs. All atoms, including H, have the mass of their most abundant isotope. All nuclear particles are treated in a classical manner. The equations of motion are integrated by using the velocity Verlet algorithm. Atomic temperature is controlled by a Nosé thermostat to converge to an average of $T = 300$ K. A thorough discussion of thermostats, the choice of m_e, and other issues of ab initio MD can be found in ref 32.

3. Reaction Control by Constraints. In a condensed-phase environment the true reaction coordinate can be a complicated quantity involving a combination of a large set of atomic coordinates including solvent, which is generally not known in advance. Therefore, the definition of a degree of freedom for microscopic control of the reaction is not obvious and often not unique. Moreover, implicit in the choice of a constrained reaction coordinate are certain general a priori assumptions about the course of the reaction. In this section we define the variable that we have constrained in our study of formaldehyde hydration and diol dehydration in acidic solution, and discuss the limitations imposed by this procedure on the chemistry.

The first step of reaction 3 in section 1 (step a) is to transfer the catalytic proton from the acidic solution to the carbonyl O, creating a (hypothetical) carbocation **3**. This suggests that controlling this transfer is one of the ways to initiate the addition reaction. Anticipating the results discussed in section 5, we found that forcing protonation of the carbonyl O is, in fact, sufficient to induce also a nucleophilic attack on the carbon by a nearby solvent H_2O molecule [step b in reaction 3]. We refer to the carbonyl O as the acceptor oxygen O_A. The proton H^+ that is to be attached to O_A is removed from either a donor oxygen atom O_D of a hydronium ion $H_3O_D^+$ or an acid sulfate oxygen atom HO_DSO_3H. A convenient transfer reaction coordinate Q_{DA} can then be defined as the difference in distance of H^+ with respect to O_D and O_A

$$Q_{DA} = R_{H^+O_D} - R_{H^+O_A} \tag{4}$$

with $R_{AB} = |\mathbf{r}_A - \mathbf{r}_B|$. Note that the sign of Q_{DA} conveys information about the phase of transfer; if $Q_{DA} < 0$ the proton is on the donor side, whereas for $Q_{DA} > 0$ the proton is closer to the acceptor oxygen. An important consideration in the definition of eq 4 is that any restriction on the value of the donor−acceptor distance $R_{O_DO_A}$ is avoided even when a rigid constraint is imposed on Q_{DA}. Similarly the $O_DH^+O_A$ angle is free to adapt or fluctuate.

The definition of Q_{DA} of eq 4 is sufficiently flexible for the study of proton exchange involving a gas-phase dimer (see section 3). However, in an acidic solution containing a large number of equivalent protons and donor oxygens, the precise specification concerning which of the H^+ and O_D is to be constrained is a serious limitation in terms of statistical mechanics. Moreover, such a constraint also interferes with the chemistry because selection of the acid species H^+B (H_3O^+ or H_2SO_4) as the starting point of the transfer preconditions the reaction. For this first exploration of a reactive acid aqueous solution we have opted for the simplest approach taking B to be a water molecule. Owing to this restriction we will not be able to address questions of particular interest concerning general acid catalysis, for example how a weak, undissociated acid molecule AH can act as a catalyst, see reaction 2.

The crucial quantity to be calculated is the change in free energy of the constrained system along the reaction coordinate. This can be determined by integrating the mean constraint force along the reaction path. Key elements of the statistical-mechanics derivation of the calculations of the free-energy profile are given in the appendix.

3. Gas-Phase Structure and Energetics

Assuming that the catalytic proton in reaction 3 is delivered and re-absorbed by a water molecule (see section 2.3), the molecular species participating in the reaction are H_2O, H_2CO, $H_2C(OH)_2$, and their protonated derivatives H_3O^+, H_2C^+OH, and $(H_2COH)O^+H_2$. We have evaluated the accuracy and performance of the DFT−BLYP-based computational approach of section 2.1 by a series of gas-phase calculations for these compounds and their dimers. These calculations were performed in a large cubic box ($L = 12$ Å). Interactions with periodic images created by the plane-wave basis set were eliminated by using a numerical screening technique similar to the method proposed in ref 33.

(27) Meijer, E. J.; Sprik, M. Manuscript in preparation.
(28) Andzelm, J.; Wimmer, E. *J. Chem. Phys.* **1992**, *96*, 1280.
(29) Johnson, B.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 5612.
(30) Lee, Ch.; Sosa, C. *J. Chem. Phys.* **1994**, *100*, 9018.
(31) Proynov, E. I.; Ruiz, E.; Velo, A.; Salahub, D. R. *Int. J. Quantum Chem.* **1995**, *29*, 61.

Table 1. Reaction Energies (in kJ mol^{-1})[a]

	BLYP/70	BLYP/150	BLYP/G	BLYP/ADF	MP2	MP4	exptl
$H_2O + H^+ \rightarrow H_3O^+$	676	679	681[b]		684[b]		690[c]
$H_2CO^+ H^+ \rightarrow H_2C^+OH$	701	704	704[b]		702[b]		713[c]
$H_2CO^+ H_2O \rightarrow H_2C(OH)_2$	8				36,[d] 42[e]		
	(34)			(36[h])	(61,[d] 62,[e] 60[f])	(50[g])	

[a] Zero-point energy corrections have been included except for the numbers in parentheses. They are adopted from the values of ref 36 obtained from frequencies calculated at the MP2/6-31G* level. The data designated BLYP/70 and BLYP/150 have been obtained by using the ab initio MD code with a reciprocal space cutoff of 70 and 150 Ry, respectively. [b] Reference 34, using 6-311+G(3df,2p). Note that for the BLYP/G and MP2 calculation of proton affinity in ref 34 equal basis sets were used. [c] Reference 35. [d] Reference 36, using 6-31G*. [e] Reference 39, using 6-31G**. [f] Reference 40, using 6-311G**. [g] Reference 40, using 6-311++G**/MP4SDTQ. [h] References 37 and 38.

1. Monomers. For the monomers and their protonated derivatives, the calculated structural characteristics are consistent with the DFT data available in the literature.[28−31] Various evaluation studies[28−31] comparing high-level quantum chemistry calculations and experiment have revealed a systematic trend for generalized gradient corrections, namely a bias toward overextended bond lengths. The discrepancies, however, are not very large and the overall agreement with experiment and accurate self-consistent field (SCF)-based methods is satisfactory. Of greater interest in the present context are the protonation energies. Table 1 compares our BLYP results for H_2O and H_2CO to theoretical values from other sources and experiment. The high-level DFT−BLYP values in ref 34, obtained with a large basis set including a correction for basis-set superposition errors, have been listed to validate our plane-wave approach involving pseudopotentials. The excellent agreement for the proton affinities is evidence that the perturbations introduced by the pseudopotentials are negligible for this quantity. Note, however, that the small relative difference of 1.4% (≈10 kJ mol^{-1}) with respect to experiment[35] is still significant on the scale of the energies relevant for chemistry at ambient temperatures. Fortunately, the difference between the proton affinities of formaldehyde and water (25 kJ mol^{-1}) is considerably closer to the experimental value of 23 kJ mol^{-1}.

Table 1 also lists our estimate for the reaction energy of the hydration of formaldehyde in a vacuum. The zero-point energies are adopted from the values of ref 36 obtained from frequencies calculated at the MP2/6-31G* level. We have not been able to find experimental reference data for the gas-phase reaction. Comparison of our result with the value obtained with the ADF package[37] by using a large basis set[38] and the BLYP functional confirms the good accuracy of the pseudopotential approach. However, the reaction energy we obtained is substantially smaller than the values of MP2 and MP4 calculations.[36,39,40] This is reason for some concern. Previous experience with related reactions[19,20] suggests that BLYP has a tendency to underestimate the energy gained by converting a bond of high bond order, such as a carbonyl bond, to two bonds of lower bond order. This is exactly what happens in addition reactions (for more examples and further discussion see also ref 31). On the basis of these comparisons, the energy produced by diol formation in our approach can be expected to be significantly underestimated by BLYP.

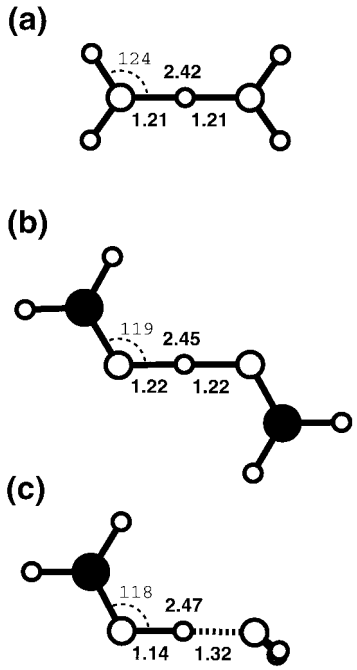


Figure 1. Equilibrium structure of protonated dimers calculated with DFT-BLYP: (a) water dimer, (b) formaldehyde dimer, and (c) mixed formaldehyde−water dimer. Oxygen atoms are pictured gray, carbon atoms black, and hydrogen atoms white. Dotted lines represent hydrogen bonds. The bold numbers indicate the OH and OO distances in the O−H$^+$−O bond that stabilizes the dimers. Lengths are in angstroms. Selected bond angles are also indicated (in degrees). Binding energies can be found in Table 2.

2. Dimers. For many molecular solutes the acidities in aqueous solution and relative proton affinities in the gas phase show little or no correlation. Indeed, protonated aldehydes and ketones (if stable) tend to have p*K* values that are lower than the p*K* for H_3O^+ despite having greater gas-phase proton affinities.[41] To verify that H_2C^+OH, i.e., the unstable carbocation **3** formed by protonation of formaldehyde in step a of reaction 3, conforms to this trend, we have investigated the effect of hydration of protonated formaldehyde−water dimers in a vacuum. Figure 1 shows the geometry of the O−H$^+$−O hydrogen bond in the protonated H_2O and H_2CO dimers and the mixed H_2O−H_2CO complex. The OO distance R_{OO} is virtually independent of monomer composition. The value of $R_{OO} = 2.45$ Å is characteristic of strong hydrogen bonds formed by the H^+ ion and also found in solution.[17] For both H_2O−

(32) Tuckerman, M. E.; Parrinello, M. *J. Chem. Phys.* **1994**, *101*, 1302, 1316.
(33) Barnett, R. N.; Landman, U. *Phys. Rev. B* **1993**, *48*, 2081.
(34) Smith, B. J.; Radom, L. *Chem. Phys. Lett.* **1994**, *231*, 345.
(35) Hay, P. J.; Wadt, J. *J. Chem. Phys.* **1985**, *82*, 270.
(36) Wolfe, S.; Kim, C.-K.; Yang, K.; Weinberg, N.; Shi, Z. *J. Am. Chem. Soc.* **1995**, *117*, 4240.
(37) ADF 2.2.2, *Theoretical Chemistry*; Vrije Universiteit, Amsterdam. See also: Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 42. te Velde, G.; Baerends, E. J. *J. Comput. Phys.* **1992**, *99*, 84. Fonseca Guerra, C. et al. In *Methods and Techniques in Computational Chemistry*; Clementi, E., Corongiu, G., Eds.; STEF: Cagliari, 1995; Chapter 8, pp 305−395.
(38) Molecular orbitals were expanded in an uncontracted triple-ζ Slater-type basis set augmented with 2p and 3d polarization functions for H, and with 3d and 4f polarization functions for C and O. The cores were kept frozen.
(39) Böhm, S.; Anipova, D.; Kuthan, J. *Int. J. Quantum Chem.* **1996**, *58*, 47.
(40) Francisco, J. S.; Wiliams, J. H. *J. Am. Chem. Soc.* **1993**, *115*, 3746.
(41) March, J. *Advanced Organic Chemistry*; Wiley: New York, 1992.

Table 2. Dimerization Energies ΔE_{D^+} (in kJ mol^{-1})[a]

	BLYP	HF, MP, CC	exptl
$H_2O-H^+-OH_2$	138	141,[b] 138,[c] 144[d]	138,[e] 132,[f] 133[g]
$H_2CO-H^+-OCH_2$	136	115[h]	
$H_2O-H^+-OCH_2$	156, 128 (+28)		23[i]

[a] The value of ΔE_{D^+} is defined as the energy needed to separate the dimer into a neutral and a protonated monomer. The energies are given without zero-point energy corrections. The zero-point energy correction for $H_5O_2^+$ is estimated in ref 42 to be −4.5 kJ mol^{-1}. The two values of ΔE_{D^+} given for the mixed dimer are for separation in H_2O, H_2C^+OH and H_3O^+, H_2CO, respectively. The number in parentheses is the difference between these two energies and should be compared to the experimental difference in proton affinity of the monomers. [b] Reference 42, using TZ2P/CCSD(T). [c] Reference 43, using 6-31+G(d,p)/MP4(SDQ). [d] Reference 44, using pVTZ+/MP2. [e] Reference 45. [f] Reference 46. [g] Reference 47. [h] Reference 48, using 4-31G*/HF (including BSSE correction). [i] Proton affinity H_2CO relative to H_2O according to ref 35.

H^+-OH_2 and $H_2CO-H^+-OCH_2$, the OH^+O bond is symmetric. In the mixed dimer the hydrogen bond is asymmetric, with the H^+ displaced toward the O atom of the formaldehyde. This is consistent with the larger proton affinity of this monomer (see Table 1).

The formation energies of the protonated dimers ΔE_{D^+} are listed in Table 2. For $H_5O_2^+$ Table 2 also gives a selection of the many accurate quantum chemistry calculations available in the literature for this species[42−44] and experiment.[45−47] Our BLYP result for ΔE_{D^+} of $H_5O_2^+$ is in good agreement with these data. The $H_2CO-H^+-OCH_2$ dimer is far less documented. In spite of a rather different hybridization of the O atoms in the free monomers, we find a binding energy that is very similar to the $H_2O-H^+-OH_2$ value. Compared to the MP2/4-31G* calculation in ref 48 we find a significantly larger stabilization of the protonated dimer, by as much as 20 kJ mol^{-1}. A further difference between the two calculations is that, whereas the BLYP dimer in Figure 1 has a centrosymmetric $O-H^+-O$ bond, the HF/4-31G* geometry found in ref 48 is asymmetric. We are inclined to attribute these discrepancies to insufficient treatment in ref 48 of correlations and/or basis-set convergence. This presumption is supported by the fact that for $H_5O_2^+$ the HF level geometries are also asymmetric. More accurate correlated calculations restore the symmetry of the $O-H^+-O$ bond (see, e.g., ref 42). For the ΔE_{D^+} of the mixed dimer Table 2 lists two values. The larger value is for separation in H_2C^+-OH and H_2O, the smaller for H_2CO and H_3O^+ dissociation. The difference of these two energies corresponds to the proton affinity of H_2CO relative to that of H_2O, as can be checked from the values listed in Table 1.

3. Proton Transfer. Hydration of the water molecule in the protonated H_2O-H_2CO dimer has a clear effect on the $O-H^+-O$ hydrogen bond structure. Figure 2 shows that adding only a single H_2O molecule shifts the H^+ away from the H_2CO toward the water O. Linking also the second H atom of the water monomer with a H_2O ligand yields a stable hydronium H_3O^+ ion with almost equivalent hydrogen bonds to the two H_2O and the H_2CO neighbor. The response to hydration can be interpreted as the effect of an increase in proton affinity of small water clusters compared to monomers. This is an indication of what will occur in solution. To obtain a quantitative description of the behavior of the $O-H^+-O$ hydrogen bond, we used the method of constraints described in section 2.3 and in the appendix. Taking in eq 4 for O_D the water oxygen (D = W) and for O_A the carbonyl oxygen (A = F) we determined the effective potential for displacement of the excess proton along the hydrogen bond by constraining the transfer coordinate Q_{WF}. In Figure 3a we have plotted the force of constraint f_c as a function of Q_{WF} for hydration numbers $n = 0$ and $n = 2$. These systems correspond to the complexes in Figure 2a and 2c, respectively. The forces in Figure 3a are averages over constrained MD trajectories of equilibrium dimer structures at low temperature ($T = 10$ K). The length of the trajectories varied between 0.5 and 1.5 ps.

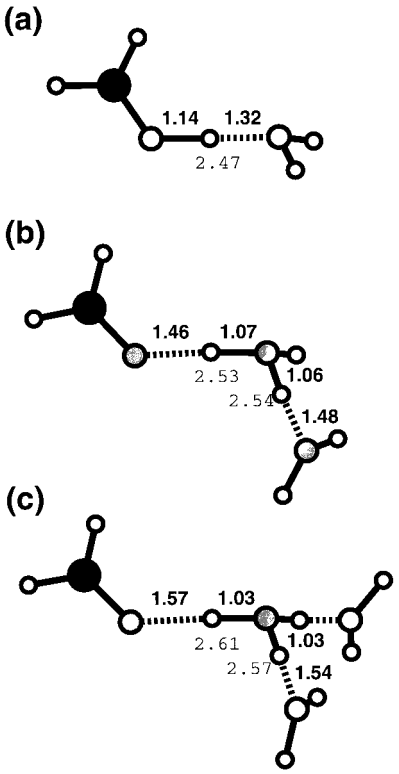


Figure 2. Equilibrium structures of hydration of protonated formaldehyde−water dimer, calculated with DFT-BLYP. Oxygen atoms are pictured gray, carbon atoms black, and hydrogen atoms white. Dotted lines represent hydrogen bonds. (a) Unhydrated mixed dimer (same as in Figure 1c). The effect of the formation of the hydrogen bonds to the H_2O in the dimer by one and two additional H_2O molecules is shown in (b) and (c), respectively. The bold numbers indicate the OH distances in the dimer of the $O-H^+-O$ bonds. The regular typeface numbers are the corresponding OO distances. Lengths are in angstroms.

In our convention, a positive value of f_c indicates that H^+ is pulled in the direction of the formaldehyde ($Q_{WF} > 0$) and a negative value indicates that H^+ is attracted by the water ($Q_{WF} < 0$). The $f_c = 0$ point corresponds to the equilibrium geometry. In a first approximation f_c can be regarded as a harmonic restoring force. Integrating the f_c curves with the $f_c = 0$ equilibrium point as a reference yields the relative energies ΔE_{H^+} for proton transfer. These are displayed in Figure 3b. From these curves we can see that the energies involved in solvation-induced transfer are substantial. For example, ≈30 kJ mol^{-1} is required to distort the hydrogen bond in the $n = 2$ dimer to a configuration similar to that of the $n = 0$ dimer, where the proton is closer to the carbonyl O.

To properly appreciate the results in Figure 3b it should be recalled that Q_{WF} is the only coordinate that is controlled. All

(42) Xie, Y.; Remington, R. B.; Schaefer, H. F., III *J. Chem. Phys.* **1994**, *101*, 4878.
(43) Del Bene, J. E.; Frisch, M. J.; Pople, J. A. *J. Phys. Chem.* **1985**, *89*, 3669.
(44) Ojamae, L.; Shavitt, I.; Singer, S. J. *Int. J. Quantum Chem.* **1995**, *S29*, 657.
(45) Meot-Ner, M.; Field, F. H. *J. Am. Chem. Soc.* **1977**, *99*, 998.
(46) Cunningham, A. J.; Payzant, T. D.; Kebarle P. *J. Am. Chem. Soc.* **1972**, *94*, 7627.
(47) Meot-Ner, M.; Speller, C. V. *J. Phys. Chem.* **1986**, *90*, 6616.

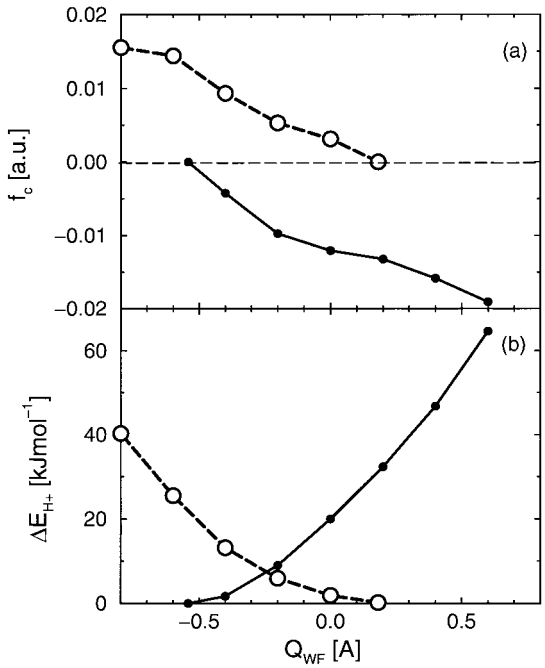


Figure 3. Proton transfer along the hydrogen bond $O_W-H^+-O_F$ in the protonated water (W)−formaldehyde (F) dimer. (a) Force of constraint and (b) energy for displacement of the excess proton H^+ as a function of the transfer coordinate Q_{WF} [for a definition see eq 4]. The large open circles connected by dashed lines indicate the result for the unhydrated dimer (Figure 2a). The small filled circles connected by solid lines are the corresponding data for the 2-fold hydrated dimer (Figure 2c). Energies are relative to the equilibrium geometry ($f_c = 0$).

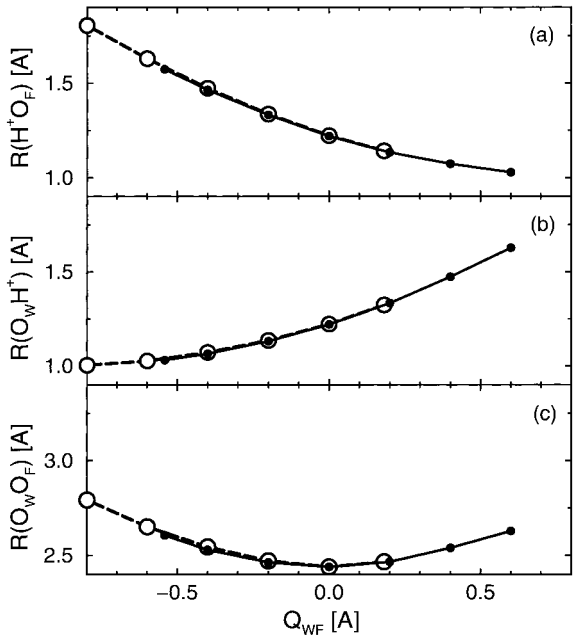


Figure 4. Geometry of the hydrogen bond $O_W-H^+-O_F$ in the protonated water (W)−formaldehyde (F) dimer as a function of the transfer coordinate Q_{WF}. (a) Distance of the excess proton to the carbonyl oxygen O_F. (b) Distance of the excess proton to the water oxygen O_W. (c) O_W-O_F distance. See Figure 3 caption for a definition of the symbols.

other degrees of freedom have been fully relaxed. Hence, as can be seen in Figure 4, not only the distance of H^+ with respect to O_W and O_F changes when the constraint Q_{WF} is varied, but also the O_W-O_F distance is adjusted. This distinguishes our approach from the majority of computational studies of energy profiles for proton transfer in gas-phase complexes.[48−51] In these calculations $R_{O_DO_A}$ is held fixed or is restrained by a potential. Figure 4 also shows that, for the same value of the transfer coordinate Q_{WF}, the geometry of the $O_W-H^+-O_F$ hydrogen bond is virtually independent of the hydration number. The effective invariance of structure with hydration number for given Q_{WF} is perhaps somewhat surprising in view of the appreciable shift of the Q_{WF} value where the energy is minimal (Figure 3a).

In the present calculations H^+ is treated as a classical particle. Here, we argue that the quantum effects are of minor importance in the present work: One of the conclusions of the theoretical work on transfer[48−51] is that the potential for H^+ changes shape over the range of $R_{O_DO_A}$ values typical for hydrogen bond lengths in these systems (2.4−2.7 Å). In weak hydrogen bonds ($r_{OO} \approx 2.6-2.7$ Å) the proton moves in a double well. It has recently been shown by a path-integral ab initio MD simulation[50] of $H_3O_2^-$ (using BLYP) that the effect of tunneling and zero-point motion is significant for this type of hydrogen bond. In contrast, for strong hydrogen bonds, such as those in $H_5O_2^+$ ($r_{OO} \approx 2.4$ Å), the potential for H^+ exhibits a single well. The same path-integral study[50] showed that in $H_5O_2^+$ the quantum effect is of minor importance. The hydrogen bond in the present work, characterized by the $O_W-H^+-O_F$ bond in the $n = 2$ complex, has a length of 2.61 Å (see Figure 2 or 4c) and is relatively weak. However, because of the chemical inequivalence of donor and acceptor, the minima of a possible double-well potential are not symmetric. Such a potential resembles in effect a single well. This argument gives us some justification for neglecting proton tunneling and zero-point motion in the simulation of the addition reaction in solution.

4. Model System

1. System Size and Composition. In this section we discuss the setup of the model system for the ab initio MD study of the hydration of formaldehyde in an aqueous solution of sulfuric acid. In MD simulations of bulk systems, the system size is a critical parameter. It is essentially specified by the dimension L of the periodic MD cell. In the present work we used a simple cubic cell. The selection of a value for L is a compromise between the length scales in the system and the minimum time required for equilibration and sampling. Given the current stage of technical development, a solution consisting of from 50 to 100 heavy atoms (i.e., not counting H atoms) can be followed by ab initio MD for 30 ps. With these rather severe limitations in mind we opted for solutions consisting of ≈30 water molecules.

We considered two systems. The first system constituted a 2.5 mol % sulfuric acid solution of one H_2SO_4 molecule and 39 water molecules, and was used to test the behavior of the solvated H_2SO_4 molecule, in particular its acidity. The second system consisted of one H_2CO molecule in a 8.5 mol % sulfuric acid solution of three H_2SO_4 and 32 water molecules. The volumes for these systems were obtained from the experimental thermodynamic data in ref 52 for the dependence of density on molarity for a pure H_2SO_4 solution. Converting these macroscopic data to the microscopic dimension of our MD system, we obtain for the 2.5 mol % sulfuric acid solution a cubic cell size of length $L = 10.71$ Å. For the 8.5 mol % solution without

(48) Chu, C.-H.; Ho, J.-J. *J. Phys. Chem.* **1995**, *99*, 1151, 16591.
(49) Latajka, Z.; Scheiner, S. *J. Mol. Struct. (THEOCHEM)* **1991**, *234*, 373.
(50) Tuckerman, M. E.; Marx, D.; Klein, M. L.; Parrinello, M. *Science* **1997**, *275*, 817.
(51) Tapas, K.; Scheiner, S. *J. Am. Chem. Soc.* **1995**, *117*, 1344.

the H_2CO molecule a cell of size $L = 10.52$ Å was calculated. Adding the one formaldehyde solute molecule in this small volume could increase the pressure substantially. To compensate for this effect the cell volume was increased by the effective solvation volume of H_2CO. For a very approximate estimate of this number we used the volume per molecule, $V_m = 60$ Å^3, obtained from the density of liquid formaldehyde at -20 °C (for comparison, the effective volume of a water molecule in liquid water is $V_m \approx 30$ Å^3). With this correction the cell length for the 8.5 mol % solution becomes $L = 10.69$ Å.

It should be noted that the present system is relatively small. However, we believe that the finite-size corrections for this system, while certainly not negligible, will be of secondary importance. For pure water this presumption is supported by results of the ab initio MD study of ref 16, which showed that both structural and dynamical properties can be reproduced with reasonable accuracy with similar sized systems. For the inhomogeneous solvent structures around solutes boundary effects are more critical. The size of the system used here is large enough to include up to the second solvation shell of water molecules. The first two coordination shells are vital for stabilization of the specific hydrogen bonding to solutes, which is the most important characteristic of aqueous solvents. These local solvation properties can therefore be accounted for by the present system size. On the other hand, our system will be clearly too small to describe long-range effects, such as the outer shell contributions to the solvation energy. This will affect the accuracy of the energetics in our simulation. The uncertainty in relative energies as a result of boundary effects is estimated to be in the order of 30%. Unfortunately, the computational effort required to perform a similar study of systems of a significantly larger size is simply not feasible with present-day computers.

2. Equilibration of Acid Solution and pH. For both the 2.5 and 8.5 mol % sulfuric acid solutions, the initial state of the system was prepared in two steps. First a sample of a solution with molecular (undissociated) sulfuric acid was set up by using classical MD simulation and a conventional force field model. From this configuration the ab initio MD was started. The two systems responded in similar ways to the switch to ab initio MD. Within 0.5 ps the first H^+ ion was observed to exchange bonding sites from a sulfate O to an O atom of a hydrogen-bonded water molecule, forming a hydronium ion (H_3O^+). Other proton-transfer events followed, releasing further H^+ to the solvent or passing protons from H_3O^+ ions on to the next solvent molecule or occasionally back to a HSO_4^- ion.

We accumulated a 10 ps trajectory of the 2.5 mol % solution in order to obtain an estimate of the acidity of a sulfuric acid molecule in bulk solution. This turned out not to be an easy task because of difficulties in defining an appropriate configurational function measuring pH. If the concentration of H^+ is calculated from the fraction of dissociated sulfuric acid molecules, the result is a negative pH value. On the other hand, if only the H^+ ions separated from the (bi)sulfate ion by at least one water molecule are counted, we obtain a pH value of 0.4 for the 39 H_2O + 1 H_2SO_4 composition. The details of this calculation will be reported elsewhere.[27] A pH value of 0.4 for a 2.5 mol % sulfuric acid solution is somewhat high compared to the experimental value of -0.5.[53] This disagreement may have a number of origins. Insufficient screening of ions due to the limited size of our samples is one source of error. Furthermore, our ad hoc definition of pH is also open to criticism (see below).

The equilibration of the initial configuration of the 8.5 mol % sulfuric acid solution plus formaldehyde was continued for 10 ps. The observed frequency of proton exchange events is approximately once per picosecond (see also ref 17). Under these conditions a period of 10 ps is sufficient to establish an equilibrium distribution of H^+ ions. To a good approximation the system was found to contain three H^+ and an equal number of HSO_4^- ions. Hence, on average, each of the sulfuric acid molecules has released one of its protons to the solvent. Translating these numbers directly in an estimate of the concentration of free H^+ ions yields, as for the 2.5 mol % solution, an unrealistically low (negative) pH value. However, determining the pH by classifying acid H^+ ions by using solvation structure also fails for the 8.5 mol % solution: At this high concentration, the majority of H^+ ions are part of a contact ion pair and, therefore, excluded from contributing to the pH. Clearly, the question of relating a microscopic quantity to the experimental pH is more involved and most likely will require a definition in terms of thermodynamic potentials. For the purpose of modeling an acid-reactive medium, however, the observed dissociation of one proton per sulfuric acid molecule seems acceptable.

The highly viscous nature of high-concentration sulfuric acid solutions is a further reason for concern. We verified that the solvent (H_2O) dynamics is still in the diffusive regime. The water molecules are sufficiently mobile to set up an equilibrated network of hydrogen bonds, solvate the ions present in the solution, and respond to changes in the charge distribution. On the other hand, for a proper relaxation of the distribution of the very sluggish bi-sulfate ions, the MD runs are too short. In fact, it was not possible to ascertain that the diffusion coefficient of the anions is finite in our system.

5. Reaction in Solution

1. Protonation of Carbonyl Oxygen. The presumed first step in the acid-catalyzed addition of water to formaldehyde is the protonation of the carbonyl oxygen, step a in reaction 3. From the results on hydrated dimers in section 3 it can be expected that, also in solution, spontaneous attack by a H^+ ion is a highly unlikely event on the time scale of ab initio MD. Therefore, we initiated the reaction by a forced transfer of a H^+ from a H_3O^+ cation to the carbonyl O using the method of constraints, as explained in section 2.3 and the appendix. To prepare the system for this operation, the last few picoseconds of the 10-ps equilibration period (see section 4.2) were performed with a H_3O^+ hydrogen bonded to the carbonyl O. The hydronium cation was stabilized by fixing the length of its three OH bonds, thus preventing a possible escape of the excess proton.

Subsequently, the constraint on the three OH bonds was released and replaced by a constraint on the transfer coordinate Q_{WF}, defined in eq 4. The value of the transfer coordinate at that time was $Q_{WF} = -0.532$ Å. As in section 3.3, proton donor and acceptor species are indicated by $H_3O_W^+$ and CH_2O_F, respectively. In the following, the proton is extracted from the hydronium O_W by increasing Q_{WF} in steps until, at positive values of Q_{WF}, it is bonded to the carbonyl O_F. A total of seven Q_{WF} points were used for the controlled protonation of O_F. The forces of constraints f_c were averaged over constrained MD trajectories according to eq 6, see appendix, using eqs 10 and 11. The length of these runs varied between 1.5 and 2.5 ps for each of these Q_{WF} points, making it a total of 13.5 ps. The results of the averaged forces of constraint are the mean forces f_s and are shown in Figure 5a.

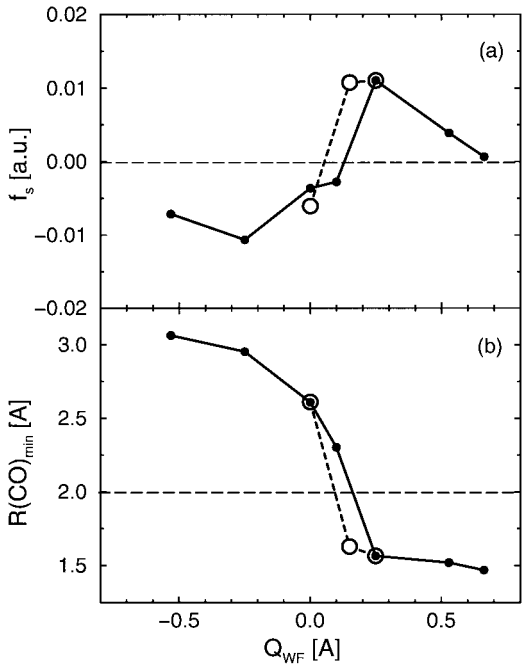


Figure 5. Addition of water to formaldehyde in acidic solution controlled by forced protonation of the formaldehyde oxygen atom (O_F) by using the method of constraints. The proton is transferred from the oxygen atom (O_W) of a hydronium cation. (a) Mean force f_s as a function of the reaction coordinate Q_{WF}. (b) Nucleophilic attack monitored by the average distance R_{CO}^{min} of the carbonyl C to the nearest solvent oxygen atom, as a function of Q_{WF}. Filled circles connected by solid lines correspond to the forward proton transfer (W → F) inducing formaldehyde hydration. Open circles connected by dashed lines correspond to the reverse transfer (W ← F) inducing diol dehydration.

2. Nucleophilic Attack and Deprotonation. As can be seen in Figure 5a, the mean force in the initial phase of the proton transfer is negative. Similar to the response of the proton in the 2-fold hydrated $H_2O-H^+-OCH_2$ dimer in Figure 3, the negative sign must be interpreted as a restoring force that pulls the H^+ back to the H_2O_W molecule, consistent with the larger proton affinity of water clusters relative to formaldehyde. However, unlike in Figure 3 the force curve is not monotonic. With increasing Q_{WF} the value of f_s turns around and changes sign approximately midway in the transfer at $Q_{WF} \approx 0.1$ Å. From this point on, the H^+ is attracted by the carbonyl group. After a steep rise, f_s relaxes to zero again at $Q_{WF} \approx 0.67$ Å, signifying completion of the protonation resulting in an equilibrium HO_F bond. The explanation of the stationary point for f_s becomes immediately evident when the MD evolution is visualized by snapshots of configurations. For the trajectory with $Q_{WF} = 0.0$ Å, the animation shows several successive events of H_2O molecules approaching the carbonyl C in, unsuccessful, attempts to form a CO bond (solid line in Figure 6). At $Q_{WF} = 0.25$ Å, i.e., the first system of our set having a positive value of f_s, nucleophilic attack has been successful and the system has been transformed into the oxonium molecule **4** of reaction 3, i.e., a protonated diol, with one stretched hydroxyl group hydrogen bonded to a water molecule. When we increase Q_{WF} further, we observe repeated transfers of the protons of the added O^+H_2 group to solvent molecules and back again to the diol. Finally, at $Q_{WF} = 0.67$ Å, corresponding to a hydroxyl bond with equilibrium geometry, the excess proton is passed on to second nearest-neighbor H_2O molecules or bi-sulfate ions. The catalytic H^+ has been returned to the solvent and the neutral diol product remains. By continuing the simulation without any constraints for 2 ps, we verified that, on a time scale comparable to that of the constrained runs, the diol in acidic solution is stable.

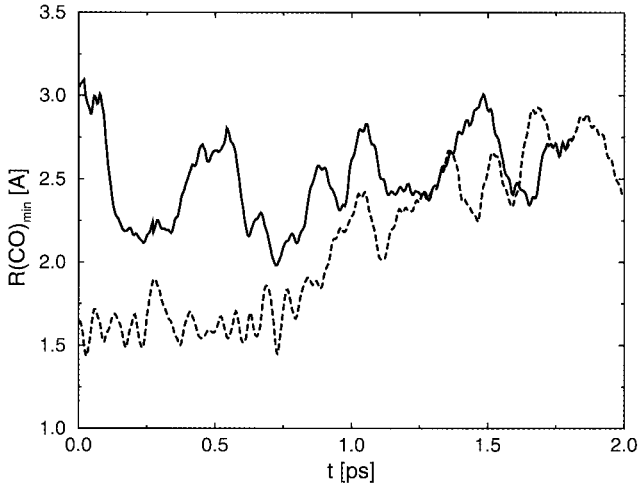


Figure 6. Distance of carbonyl carbon to the nearest solvent oxygen as a function of time. Shown are two trajectories at $Q_{WF} = 0.0$ Å. The solid curve corresponds to the forward transfer (W → F) inducing formaldehyde hydration, and the dashed line corresponds to the reverse transfer (W ← F) forcing diol dehydration. The solid curve fluctuates between values of ≈2.8 Å, characteristic of a nonbonded C−O interaction and values of ≈2.0 Å indicating, unsuccessfull, attempts to form a CO bond. The dashed curve shows a conversion of a CO bond distance to distances similar to the solid line. These curves suggest that the nonbonded C−O distance is the equilibrium state for $Q_{WF} = 0.0$ Å.

The correlation between proton transfer and CO bond formation can be quantified by monitoring the time dependence of the distance R_{CO}^{min} of the solvent oxygen closest to the carbonyl C. The identity of a nearest O atom is well-defined in every configuration but may change from one atom to another in the course of time. Figure 6 shows two examples of such curves. Time averages of R_{CO}^{min} are plotted as a function of Q_{WF} in Figure 5b. The mean force in Figure 5a changes sign at the same point ($Q_{WF} \approx +0.15$ Å) as where, in Figure 5b, the average of R_{CO}^{min} crosses from values characteristic for nonbonded to those typical for bonded C−O interactions.

3. Reversibility and Thermochemistry. The response of the system to a stepwise variation of a constrained coordinate is of dubious significance for the understanding of spontaneous behavior if this operation has not been performed under equilibrium conditions. This question is particularly critical in ab initio MD in view of the very limited length of the trajectories. A severe test for our procedure is the reversal of the constrained reaction path. If the mean forces can be reproduced within the margin of a small amount of hysteresis in the constrained reaction coordinate we can be confident that the results are meaningful and also reliable from a quantitative point of view.

The second set of data in Figure 5 (open circles) are results obtained by decreasing Q_{WF} starting from the $Q_{WF} = 0.25$ Å step in the protonation reaction. Hence, the initial configuration for these constrained runs of the backward (decomposition) reaction corresponds to the situation in the forward (addition) reaction for which a H_2O molecule remained for the first time firmly attached by a chemical bond to the C atom. As is evident in Figure 5, in the backward reaction the breaking of the CO bond has already occurred when Q_{WF} has been reduced to 0.0

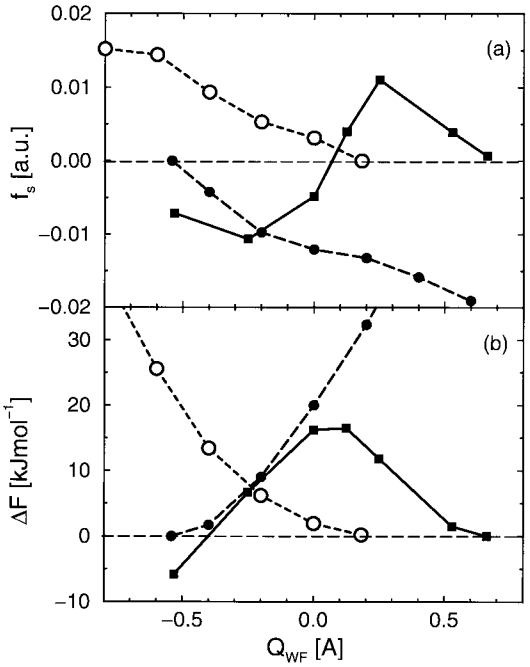


Figure 7. (a) Mean force as a function of the constrained proton-transfer coordinate Q_{WF}. (b) Relative free energy obtained by integration of the mean force of (a). The data for the addition reaction in acidic solution are indicated by squares connected by solid lines. In (a) these data are averages over the forward and reverse reaction data plotted in Figure 5. The open and filled circles connected by dashed lines are the forces of constraint and corresponding energies for the unhydrated and 2-fold hydrated protonated water−formaldehyde dimer, respectively. These data have been copied from Figure 3.

Å. In fact, a single dehydration reaction was observed, which is manifested by the sudden increase in R_{CO}^{min} in the second curve in Figure 6. Averages of f_s and R_{CO}^{min} over the time interval following this event are in good agreement with those obtained at $Q_{WF} = 0.0$ Å in the forward reaction. For $Q_{WF} = 0.1$ Å we find the system still in the product (protonated diol) state and, hence, the two points in Figure 5 at $Q_{WF} = 0.1$ Å must be considered branches of a hysteresis loop. The width of at most 0.2 Å is small, confirming that the MD system has been given enough time to equilibrate and to adapt to each of the steps of the controlled proton transfer. Figure 7a shows the same data for the mean force as in Figure 5a averaged over the forward and reverse reactions. In Figure 7b we plot the relative free energies obtained by numerically integrating the mean force with the equilibrium point at $Q_{WF} = 0.67$ Å as a reference. For a first analysis of these results it is instructive to compare them with the equivalent force and energy curves in Figure 3 for the $n = 0$ and 2 protonated gas-phase dimers. For negative values of Q_{WF}, when the proton is bonded in the hydronium ion, the mean force in solution and the restoring force in the $(H_2O)_2H_2O-H^+-OCH_2$ complex are similar. This confirms that the effect of full solvation of the reactant complex can be reasonable well approximated by a gas-phase hydration model. In contrast, the mean forces at positive Q_{WF} exhibit little resemblance to the $H_2O-H^+-OCH_2$ dimer. In the $n = 0$ complex as well as in solution the proton is bonded to the (former) carbonyl O. The difference in attractive force is evidence of the change of the product of proton transfer from a formaldehyde carbocation in a vacuum to a protonated diol in solution.

A further distinction with proton transfer in a vacuum is that at $Q_{WF} = -0.53$ Å, where the $n = 2$ dimer is stable ($f_c = 0$), the reactant complex in solution still yields negative mean forces. Only for the product states can we unambiguously identify in Figure 7a a stable point ($Q_{WF} = 0.67$ Å). This means that it is only possible to determine from Figure 7b the activation energy needed to decompose the diol. The estimate we obtain is 16.5 kJ mol^{-1}. However, for the reaction in the opposite direction, i.e., addition, the barrier is ill-defined. With a true $f_s = 0$ reactant state missing, it is also difficult to obtain a definite number for the free energy of the reaction. Extending the set of Q_{WF} constraints to lower values in search of an $f_s = 0$ state would be the obvious solution for these problems. Unfortunately, however, in this range of Q_{WF} values we were already confronted with the complication mentioned in section 2.3, namely the equivalence of protons in solution. Controlling the position of one of the protons in $H_3O_W^+$ by constraining the transfer coordinate Q_{WF} is no longer sufficient to prevent the escape of one of the other two protons, which leads to neutralization of the ion.

6. Summary and Discussion

By ab initio MD calculations we have demonstrated that, in acid solution, an H_2O molecule can be reversibly added to formaldehyde by a controlled transfer of a catalytic proton from a hydronium ion to the carbonyl oxygen. CO bond formation occurs at a stage midway in the transfer process where the mean force for the transfer is zero and the relative free energy is maximum. This state is the transition state for not only the protonation, step a in reaction 3, but also the addition (step b). The simulation suggests, therefore, that in the true reaction dynamics that occurs on a much longer time scales, protonation and addition are concerted reactions, and the carbocation intermediate **3** in the hypothetical reaction 3 is not stable. The protonated diol returns the catalytic proton to the solution before or upon completion of the forced proton transfer. Hence, the intermediate **4** is a transient species that is also unstable in solution. These correlations between successive steps of the catalytic addition are consistent with the interpretation of data of general acid catalysis experiments in refs 6 and 7. In fact, the transition state we find is very similar to the transition state **1**[6,7] in reaction 2 with $A = H_2O$ (specific acid catalysis). Our results are also a good illustration of the application of the method of constraints to the study of chemical reactions. The MD system is carried through an activated chemical transformation by controlling a reaction coordinate. In view of the cooperative nature of proton transfer and nucleophilic attack, the proton transfer coordinate we chose to constrain is most likely not the true reaction coordinate. It is a component, defined by a projection of the minimum free energy path. However, knowledge of the true reaction coordinate is, in principle, not necessary. Control of only a component of the activated degree of freedom is sufficient to produce an unbiased picture of the reaction, provided the system is in equilibrium on the time scale of the simulation. If these conditions are satisfied for the set of states we impose by constraints, all the crucial events can be generated by spontaneous fluctuations. As can be concluded from the discussion of reversibility in section 5.3, thermal equilibrium is effectively realized for the simple reaction model studied here.

The experimental literature available to us contained no information on activation and reaction energies in high-concentration sulfuric acid solutions similar to the model system used here. Lacking these data, we will compare our results with the thermochemistry of the uncatalyzed addition in neutral water. The estimate for the activation energy derived in ref 22 from

the rate constant in pure water is 67 kJ mol^{-1}. The barrier in Figure 7b being a factor of 4 lower is consistent with an appreciable catalytic effect. In ref 54 a free energy of -19 kJ mol^{-1} is obtained from the experimental equilibrium constant for uncatalyzed addition. The net change in free energy obtained in the present work (see Figure 7b) is of opposite sign, which favors decomposition rather than addition. This discrepancy is perhaps too large to be attributed solely to the strongly ionic environment shifting the reaction equilibrium to the reactant side. We recall that, according to Table 1, the computational method applied to determine electronic structure underestimates the formation energy of $H_2C(OH)_2$ in a vacuum by an amount comparable to the reaction energy of $\approx$20 kJ mol^{-1} quoted in ref 54. With these large errors in energetics, a quantitative interpretation of the thermochemistry of Figure 7b becomes somewhat hazardous.

Indeed, quantitative comparison with experiment requires some care. First of all, the error introduced by using an approximate DFT functional will give rise to deviations. We have already seen that the reaction energy for addition of water to formaldehyde is underestimated by the BLYP functional. This holds for the reaction in the gas phase as well as solution. Similarly, with the growing number of DFT studies of chemical systems, the evidence is increasing that activation energies are underestimated also by BLYP, in particular barrier heights for proton-transfer reactions in strong hydrogen bonds. For example, for the $(F\cdots H\cdots F)^-$ system the barrier may be too low by as much as 10$-$20 kJ mol^{-1}.[55] This might also apply to the acid-catalyzed hydration of formaldehyde, where proton transfer is part of the reaction mechanism. Second, as noted earlier, the small system size of $\approx 10^3$ Å^3 will introduce errors related to the approximate treatment of long-range electrostatic effects present in infinite bulk solution. Moreover, since the volume of our MD cell is fixed, the uncertainty in the pressure and the constraints on density fluctuations in general give reason for concern. Of these two restrictions, the boundary effect on the balance of electrostatic forces is most likely the more serious (see the discussion in section 4.1) and future studies should reveal its quantitative effect on processes such as proton transfer in aqueous solution. Finally, the trajectory lengths of $\approx$10 ps will only allow for the relaxation of relatively fast processes. Therefore, calculated properties will lack the equilibrated contribution of slower processes. As mentioned above, for the present study, 10 ps is sufficiently long to ensure equilibration of the solvent (H_2O), but does not allow for a proper relaxation of the spatial distribution of HSO_4^- ions. The latter might have a quantitative effect on the reaction energy and barrier, but is of secondary importance, as the well-equilibrated motion of protons and water molecules is the major factor driving the reaction mechanism of the process investigated here. These critical remarks indicate that there is substantial reason for quantitative improvement but, in our view, do not alter the qualitative picture provided by the present study.

In the final evaluation, we conclude that this study confirms that the ab initio MD approach is capable of revealing microscopic details of reaction mechanisms in solution that are directly relevant to the understanding of experimental results. For quantitative comparisons, however, the methodology for electronic structure calculation that we have applied, in particular the DFT functionals, need further improvement.

Acknowledgment. We thank Alessandro Curioni for introducing us to the chemistry of aldehydes and Giovanni Ciccotti for helping us with the subtleties of free-energy calculation from constraint forces.

Appendix

The dynamical information obtained by constraining reaction coordinates is of little use without a reliable estimate of the increase in free energy related to the restrictions imposed on the system. Relative free energies can be determined by integrating the mean or solvent averaged force f_s with respect to the reaction coordinate Q,

$$\Delta F(Q) = -\int_{Q_o}^{Q} \mathrm{d}Q' f_s(Q') \tag{5}$$

In ref 21 it is shown that f_s can be obtained from a properly reweighted average over the constrained trajectory of a generalized force. This generalized force is a function only of configurational degrees of freedom and contains a temperature-dependent term. The temperature dependence and the reweighting are the result of kinetic contributions to f_s. In the context of ab initio MD it is more convenient to use an alternative method that relates the mean force f_s to the average force of constraint f_c.[56] This approach is particularly straightforward for a reaction coordinate consisting of the distance between a pair of reactive atoms.[20] In this case, the mean force is simply equal to the time-averaged force of constraint without any reweighting or additional terms.

For more general reaction coordinates, such as bond angles and the transfer coordinate in eq 4, the procedure to obtain f_s from f_c is more involved. For a system of particles with Cartesian position vectors $\boldsymbol{r} = \{\boldsymbol{r}_i\}$ subject to a constraint $Q(\boldsymbol{r}) = Q'$, the correct expression for the mean force is

$$f_s(Q') = \frac{\langle Z^{-1/2}[\lambda - k_B T G]\rangle_{Q'}}{\langle Z^{-1/2}\rangle_{Q'}} \tag{6}$$

Here, T denotes the temperature and k_B denotes the Bolztmann constant. The brackets with subscript Q' denote an average over the constrained ensemble defined by setting the reaction coordinate $Q(\boldsymbol{r})$ to the fixed value Q' and its time derivative to zero, $\dot{Q} = 0$. The term λ is the Lagrange parameter associated with the force of constraint. Formally, it appears in the definition of $f_{c,i}$, the force of constraint on particle i:

$$f_{c,i} = -\lambda \frac{\partial Q}{\partial \mathbf{r}_i} \tag{7}$$

As in ref 21 the bias introduced by the constraint in momentum space ($\dot{Q} = 0$) is eliminated by multiplication by the factor Z, defined by

$$Z = \sum_i \frac{1}{m_i}\left(\frac{\partial Q}{\partial \mathbf{r}_i}\right)^2 \tag{8}$$

where Z is identical to the function derived in ref 21. Note, however, that eq 6 is a generalization of the configurational

(52) Carslaw, K. S.; Clegg, S. L.; Brimblecombe, P. *J. Phys. Chem.* **1995**, *99*, 11557.
(53) Johnson, C. D.; Katrizky, A. R.; Shapiro, S. A. *J. Am. Chem. Soc.* **1969**, *91*, 6654.
(54) Zavistas, A. A.; Coffiner, M.; Wiseman, T.; Zavistas, L. R. J. *J. Phys. Chem.* **1970**, *74*, 2746; **1995**, *99*, 16590.
(55) Latajka, Z.; Bouteiller, Y.; Scheiner, S. *Chem. Phys. Lett.* **1995**, *234*, 159.
(56) Mülders, T.; Krüger, P.; Swegat, W.; Schlitter, J. *J. Chem. Phys.* **1996**, *104*, 4869.

space formalism of ref 21 because λ also depends on momenta through the kinetic forces. Finally, the correction term G is defined as

$$G = \frac{1}{Z^2}\sum_{i,j}\frac{1}{m_i m_j}\frac{\partial Q}{\partial \mathbf{r}_i}\frac{\partial^2 Q}{\partial \mathbf{r}_i \partial \mathbf{r}_j}\frac{\partial Q}{\partial \mathbf{r}_j} \tag{9}$$

The proof of eqs 6−9 will be given in a forthcoming publication.[57] Substituting the definition for the transfer reaction coordinate of eq 4 we find for Z

$$Z = \frac{2}{m_o} + \frac{4}{m_H}\sin^2\frac{\theta}{2} \tag{10}$$

where m_O and m_H are the masses for an oxygen and hydrogen atom, respectively. The factor θ is the angle between the $\mathbf{r}_{O_WH^+}$ and $\mathbf{r}_{O_FH^+}$ vectors, i.e., the hydrogen bond angle. For G we have

$$G = -\frac{1}{Z^2 m_H{}^2}\frac{Q_{WF}}{\mathbf{r}_{O_WH^+}\mathbf{r}_{O_WH^+}}\sin^2\theta \tag{11}$$

Corrections to the bare force of constraint in eq 6 are generally insignificant. Moreover, for a linear $O_W-H^+-O_F$ hydrogen bond ($\theta = \pi$) the factor Z is a constant and $G = 0$.

JA972935U

(57) Sprik, M.; Ciccotti, G. *J. Chem. Phys.* In press.