Calculation of the Tautomer Ratio of Histamine in Aqueous Solution Using Free Energy Perturbation Methods: An In-Depth Study

Graham A. Worth† and W. Graham Richards*

Contribution from the Physical Chemistry Laboratory and Centre For Molecular Sciences, South Parks Road, Oxford, OX1 3QZ, U.K.

Received April 5, 1993. Revised Manuscript Received September 22, 1993®

Abstract: The free energy perturbation (FEP) method, despite the promise of being able to calculate free energies in the solution phase, has been shown in recent times to have various shortcomings. For chemical systems the major problems are faced when studying a conformationally flexible and/or charged molecule due to the large phase space and long-range forces involved. In this paper we take as a model the well-studied system of histamine monocation in aqueous solution and calculate the tautomerism equilibrium constant. The solution-phase free energy difference between the tautomers is split into intra- and intermolecular parts, and these contributions are calculated separately, the former with *ab initio* methods, the latter with a molecular mechanics potential. This allows the performance of different parts of the calculation to be examined. It is found that a result in good agreement with experiment is obtained if various points are noted. In order to get convergence between independent runs, it was necessary to run over 2 ps of equilibration between windows to ensure solvent relaxation and 8 ps of data collection in each window. This is much longer than the protocol used in earlier calculations. The result is also significantly altered by the use of different atom point charges, and it was found necessary to take into account the conformational species seen during the simulation in the charge set used. Finally, an implementation of the image charge reaction field of Friedman was used to model the long-range part of the Coulombic forces, which is seen to affect the result by around 4 kJ mol^{-1}.

Introduction

In previous work we used a combination of *ab initio* and free energy perturbation (FEP) methods to calculate the equilibrium constant for the tautomerism of histamine and the related 4-(5-)methylimidazole.[1,2] The contrast between the accuracy of the methylimidazole result and the error associated with the histamine values prompted this present study, which focuses on the general problems of the FEP method and how they can be faced for flexible and/or charged systems.

Histamine is a biologically active small molecule that is involved in various response mechanisms of the body. For this reason it has attracted a large amount of research. It has also been well studied theoretically, using both early gas-phase *ab initio* molecular orbital methods and the more recent solution-phase molecular mechanics calculations mentioned above.

Histamine's main feature is three nitrogen atoms available for protonation, two in an imidazole ring and one of the side chain. It can thus exist in three differently protonated states: the uncharged base with only one ring nitrogen protonated; the monocation where the side-chain amino group is also protonated; or a dication where all the nitrogen atoms are protonated. All three species coexist[3] at physiological pH in a ratio of base: monocation:dication of around 1:96:3, respectively. Both the monocation and base are able to undergo tautomerism, with the ring proton on either nitrogen. The convention used for naming these tautomers is shown in Figure 1. This gives five separate species, all able to undergo internal rotations along the side chain to produce a whole host of different conformers, which are differently stabilized by different conditions of acidity, a feature of possible physiological importance.

For one receptor, called the H2 site (responsible for gastric secretions), this tautomerism has been implicated in the mechanism of activation.[2,4,5] It is thus necessary to know the equilibrium data for these two species defined in a particular phase by the constant

$$K_T = \frac{[\text{histamine N}(\tau)\text{H}]}{[\text{histamine N}(\pi)\text{H}]} \quad (1)$$

which is related to the Gibbs free energy difference between the tautomers in the appropriate phase by

$$K_T = \exp(-\Delta G/RT) \quad (2)$$

Background and Method

One way to calculate the equilibrium constant in aqueous solution is to use the free energy perturbation method. If a system has two states A and B connected by a perturbation, A → B, the difference in Gibbs free energy between them can be written[6]

$$\Delta G_{BA} = -kT \ln\langle \exp(-\Delta H_{BA}\beta)\rangle_A \quad (3)$$

where ΔH_{BA} is the perturbation Hamiltonian defined by $H_B = H_A + \Delta H_{BA}$, and $\beta = 1/k_BT$. In theory it is then possible to calculate an ensemble of this perturbation energy, using molecular mechanics (MM) for large systems, over the states of A to obtain the free energy difference. In practice, eq 3 has shown itself to be a very difficult equation to implement. Problems arise if the states are far apart, as is usual when comparing different molecules, from the need to sufficiently sample phase space. For this reason it is necessary to break up the simulation into small

* Author to whom all correspondence should be addressed.
† Present address: European Molecular Biology Laboratory, Meyerhofstrasse 1, 6900 Heidelberg, FRG.
® Abstract published in *Advance ACS Abstracts*, December 15, 1993.
(1) Worth, G. A.; King, P. M.; Richards, W. G. *Biochim. Biophys. Acta* **1989**, *993*, 134.
(2) Worth, G. A.; King, P. M.; Richards, W. G. *Biochim. Biophys. Acta* **1990**, *1036*, 158.
(3) Ganellin, C. R. *J. Pharm. Pharmacol.* **1973**, *25*, 787.
(4) Weinstein, H.; Chou, D.; Johnson, C. L.; Kang, S.; Green, J. P. *Mol. Pharmacol.* **1976**, *12*, 738.
(5) Weinstein, H.; Muzarek, A. P.; Osman, R.; Topiol, S. *Mol. Pharmacol.* **1986**, *29*, 28.
(6) Tembe, B. A.; McCammon, D. A. *Comput. Chem.* **1984**, *8*, 281.

0002-7863/94/1516-0239$04.50/0

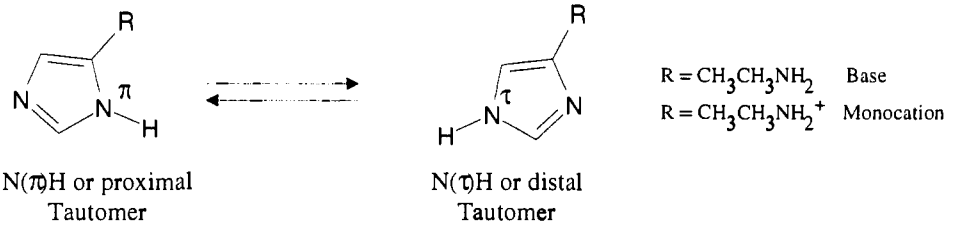


Figure 1. Tautomerism of histamine.

perturbations coupled by a parameter. The change A → B is then the free energy difference at a particular value of this constant which changes from 0 (one molecule) to 1 (second molecule). The total free energy is then obtained by adding up all the parts. This partitioning of a simulation is known as windowing and represents a potentially powerful technique in the study of chemical systems, able to determine the all-important free energy between, in theory, any two systems limited only by available computer time. The full theory of FEP and its many applications to date are well documented elsewhere.[7-11]

Practical problems are also documented. Dividing the simulation into windows introduces the necessity to ensure equilibrium at the start of each window.[12] The need for sufficient sampling also leads to the problem of the slow convergence of results.[13-15] Systems containing rotational isomers present special sampling difficulties.[7,8,16-18] Excluding the problems of parametrization, difficulties are also inherent in the MM method due to the use of nonbonded interaction truncation and inaccuracies in methods of simulating constant temperature and pressure.[19,20] The present work attempts to draw these problems together in the study of one experimentally well characterized system.

One commonly made alteration has been made of eq 3. It was thought desirable to avoid including the intramolecular energies as these terms are large and similar for both states. Hence, inaccuracies in these terms could dominate the result. Excluding bond terms also means that bond constraints such as SHAKE[21] can be applied and so simulations of a reasonable length run. To calculate the difference in free energy of only the intermolecular interactions, the perturbation Hamiltonian must be divided into intramolecular (containing all the solute–solute terms) and intermolecular (containing the solute–solvent and solvent–solvent nonbonded terms) parts. When this is done, eq 3 can be written

$$\Delta G_{BA} = -RT \ln\langle \exp(-\Delta H_{inter}\beta)\exp(-\Delta H_{intra}\beta)\rangle_A \quad (4)$$

Using a force field like that in AMBER,[22] this division is legitimate as there are no polarizability terms in the Hamiltonian and hence the internal energy of a molecule does not depend on its external environment. Naturally, the entire Hamiltonian is used to produce the configurations, it is only in the data collection that it is ignored. Unfortunately,

$$\langle e^x e^y\rangle \neq \langle e^x\rangle\langle e^y\rangle \quad (5)$$

For eq 5 to be an equality, so that $\Delta G_{BA} = \Delta G_{BA,intra} + \Delta G_{BA,inter}$, either x or y must be a constant. In the calculation to be performed, states A and B are the two tautomers of histamine. If the intramolecular potential energy surfaces for the two tautomers were the same shape, the difference in intramolecular Hamiltonian is constant at each point in the perturbation path and the ensemble average can be split to give

$$\Delta G_{BA,inter} = -RT \ln\langle \exp(-\Delta/\mathrm{hsc}_{inter}\beta)\rangle_A \quad (6)$$

Later in this paper it is found that the energy surfaces are different, but due to the sampling of the various conformers during the simulation, the approximation that the intramolecular Hamiltonian is constant is reasonable.

This approximation would be circumvented if the multiconfiguration thermodynamic integration (MCTI)[23] protocol was used in place of windowing.[24] In this formalism the free energy difference is related to the gradient of the Hamiltonian with respect to the coupling parameter rather than the exponential, and so the ensemble average can be split exactly. However, as windowing was used in all the previous calculations, it was also used in this work.

The quantity calculated by eq 6 is the difference in free energy of hydration rather than the difference in free energy between the two solvated tautomers. The free energy difference in aqueous solution can then be completed by using the thermodynamic cycle in Figure 2, where N(x)H stands for the appropriate tautomer.

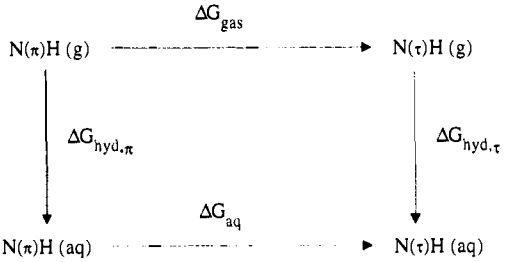


Figure 2. Thermodynamic cycle used in the calculation of histamine tautomer ratios by the FEP method.

The gas-phase difference in Gibbs free energy, ΔG_{gas}, can be calculated absolutely using *ab initio* calculations. This has the advantage that the *ab initio* calculations are equally applicable to all conformations of a molecule, unlike MM calculations for which force fields are parametrized for average properties. FEP calculations can then be used to determine the difference in free energies of hydration of the two tautomers,

$$\Delta\Delta G_{hyd} = \Delta G_{hyd,\tau} - \Delta G_{hyd,\pi} \quad (7)$$

The combination of these two calculations then gives the desired quantity,

$$\Delta G_{aq} = \Delta G_{gas} + \Delta\Delta G_{hyd} \quad (8)$$

Recently, the practice of using thermodynamic cycles to correct

(7) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Chem.* **1989**, *18*, 431.
(8) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407.
(9) *Computer Simulation of Biomolecular Systems*; vanGunsteren, W. F., Weiner, P. K., Eds.; Escom Publishers: Netherlands, 1989; Chapters 1–3.
(10) Reynolds, C. A.; King, P. M.; Richards, W. G. *Mol. Phys.* **1992**, *76*, 251.
(11) McCammon, J. A. *Curr. Opin. Struct. Biol.* **1991**, *1*, 196.
(12) vanGunsteren, W. F.; Mark, A. E. *Eur. J. Biochem.* **1992**, *204*, 947.
(13) Pearlman, D. A.; Kollman, P. A. Chapter 7 In *Computer Simulation of Biomolecular Systems*; vanGunsteren, W. F., Weiner, P. K., Eds.; Escom Publishers: Netherlands, 1989; Chapter 7.
(14) Mitchell, M. J.; McCammon, J. A. *J. Comput. Chem.* **1991**, *12*, 271.
(15) Mark, A. E.; vanGunstern, W. F.; Berendsen, H. J. C. *J. Chem. Phys.* **1991**, *94*, 3808.
(16) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1989**, *90*, 3300.
(17) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1989**, *91*, 3630.
(18) Hodel, A.; Simonson, T.; Fox, R. O.; Brünger, A. T. *J. Phys. Chem.* **1993**, *97*, 3409.
(19) Straatsma, T. P.; Berendsen, H. J. C. *J. Chem. Phys.* **1989**, *89*, 5876.
(20) Guenot, J.; Kollman, P. A. *J. Comput. Chem.* **1993**, *14*, 295.
(21) van Gunsteren, W. F.; Berendsen, H. R. C. *Mol. Phys.* **1977**, *34*, 1311.
(22) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(23) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *95*, 1175.
(24) Tidor, B.; Karplus, M. *Biochemistry* **1991**, *30*, 3217.

Table 1. Results from the Earlier Calculations of K_T for the Tautomeric Shift N(π)H → N(τ)H of Histamine and 4-(5-)Methylimidazole

molecule	ΔG(g,298)[a] (kJ mol^{-1})	K_T(g,298)	$\Delta\Delta G_{hyd}$[b] (kJ mol^{-1})	K_T(aq.298)	K_T(expt)
3-(5-)methylimidazole	−1.37	1.74	0.75 ± 2.93	1.29	0.5–1.2[c]
histamine base	−2.17	2.48	−2.09 ± 8.09	5.75	4[d]
histamine monocation	−49.02	3.87×10^8	33.57 ± 13.49	48.64	2.3–9[e]

[a] This energy difference is from MP2/6-31G**//RHF/6-31G *ab initio* calculations, corrected to a free energy by an AM1 vibrational analysis. [b] The error value is the standard deviation of four independent runs, combined with the average internal error. [c] Taken from refs 48–50. [d] Taken from ref 51. [e] Taken from refs 3, 49, 51, and 52.

for missing terms in a free energy calculation has been critically reviewed.[12,25,26] It is argued that a contribution, such as the intramolecular free energy, will not be the same on either side of the thermodynamic cycle, leading to an error. This is true if the difference in free energy of two systems in various phases is studied, as the conformers sampled in each phase will be different, e.g. when comparing molecules in solution and a protein binding site.

On a closer look at the methodology used in these calculations, it is found that the full thermodynamic cycle is not being evaluated and this error is not present. The free energy has been split into two parts,

$$\Delta G_{BA}(aq) = \Delta G_{BA}(aq,inter) + \Delta G_{BA}(aq,intra) \quad (9)$$

i.e. intra- and intermolecular terms. The quantity that is being given from the evaluation of the intermolecular Hamiltonian during a solution-phase FEP simulation is in fact not the true difference in free energy of hydration between the two end molecules but the difference in free energy of hydration taken over the solution-phase configurations. No account has been taken of the phase space probabilities of the molecules in the gas phase. Similarly, without the presence of polarization terms, the intramolecular free energy difference is equal to the gas-phase free energy difference, but again taken over the solution-phase weighting. Hence, on both horizontal sides of the cycle only the subset of states given by the Boltzmann weights of the solution-phase configurations are being considered.

A second point that has been raised in connection to this problem is the involvement of a bond contribution when bond lengths are different at either end of the perturbation. From a series of simulations, Pearlman and Kollman[25] concluded that even if the intramolecular potential was not included, the free energy calculated did depend on bond energies. For example the perturbations R-methyl → R-propyl and nothing → methane, evaluating only the intermolecular potential, were made. It was discovered that the calculated free energy differs according to the specified lengths of bonds to dummy atoms positioned to become the new atoms, whether these start at the final values or with short bonds that grow to become the final values. These they explain as a forgotten bond contribution and point to an invalidation of the thermodynamic cycle method outlined above as the intramolecular term cannot be split from the intermolecular term. However, these results are open to another interpretation. The reason for starting with short dummy atom bonds is, as they explain at the beginning of the paper and as has been noted elsewhere,[7,13] to prevent "end-point catastrophes". These are spurious energies created by atoms appearing in space occupied by solvent, or disappearing leaving a large solvent cavity, as can occur at the ends of perturbation simulations. This effect could explain the results obtained. The effect of end-point catastrophes is exemplified by a set of tautomer equilibrium calculations made on histamine base, histamine monocation, and 4-(5-)methylimidazole, where the perturbation was made in one step. Every calculation gave the same result, around 40 kJ mol^{-1}. This is presumably the energy of a proton appearing and clashing with the water molecule still in the position where it is coordinated to the unprotonated nitrogen.

From the same work it is also clear that the use of short bonds to dummy atoms produces extra work as the bonds grow during the simulation. This is not just from changing the bond lengths but also the origin position of nonbond parameters (charge and van der Waals sphere) relative to the solute. Hence, this contribution to the free energy must be evaluated even if the intramolecular free energy is not evaluated. This can be done by various methods depending on the FEP formulation used.[25-28] The program used, AMBER3.1,[29] implements the coordinate coupling method of Rao and Singh.[28] This includes the change in bond length between the states A and B in the evaluation of the function ΔH_{BA} needed to implement eq 3, and so the extra work is evaluated. In this calculation, however, the bonds being grown or shrunk are identical in the AMBER force field and so no overall work is being done. Due to the change in geometry, it is also necessary to correct for changing the moment of inertia.[26] This is not likely to be large in this case and has been ignored.

The cycle described above was that used in previous calculations, the results of which are summarized in Table 1. The free energy results, i.e. $\Delta\Delta G_{hyd}$, are the average over four runs. The error is the standard deviation of these run results to give an estimate of the spread of possible values. To this is then added the average internal error for each run, determined by double wide sampling (also known as double-ended sampling, ref 7) during the simulation. At a particular value of the coupling parameter, λ, the free energy difference over the perturbation $\delta\lambda$ is evaluated in both forward and backward directions, i.e. between states $\lambda \rightarrow \lambda + \delta\lambda$ and $\lambda \rightarrow \lambda - \delta\lambda$. If sampling is poor, or the perturbation too large so that the poor sampling makes the calculation nonreversible, this will give different results for a perturbation depending on the direction in which it is evaluated and hence an error. This combination of errors is the method used for all results in this paper when more than one simulation is being considered.

The larger error in the histamine base result is likely to be due to the more complex phase space of these flexible molecules compared to the rigid methylimidazole. The extra error for the monocation must then be due to the additional complication of charge. The variation among these four runs was large: 24.98, 26.57, 30.54, and 52.17 kJ mol^{-1}. The first part of this paper deals with this variance, attempting to obtain a converged value. Later parts then attempt to improve the methodology and with it the result.

Technical Details

All the molecular mechanics calculations were made using AMBER3.1.[29] The system was started with a histamine monocation tautomer in a larger box of waters than previously, to allow for the use of a larger cutoff. To form the box, TIP3P waters[30] were included up to 13 Å away from the solute (previously this had been 11 Å), and there were now 858 waters solvating the solute. This box was minimized in the module BORN using periodic boundary conditions. Standard AMBER all-atom pa-

(25) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1991**, *94*, 4532.
(26) Straatsma, T. P.; Zacharias, M.; McCammon, J. A. *Chem. Phys. Lett.* **1992**, *196*, 297.
(27) Pearlman, D. A. *J. Chem. Phys.* **1993**, *98*, 8946.
(28) Rao, B. G.; Singh, U. C. *J. Am. Chem. Soc.* **1989**, *111*, 3125.
(29) Singh, U. C.; Weiner, P. K.; Caldwell, J. W.; Kollman, P. A. *AMBER (UCSF)*, version 3.1; Department of Pharmaceutical Chemistry, University of San Francisco, 1988.
(30) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.

Table 2. $\Delta\Delta G_{hyd}(298)$ for Histamine Monocation N(π)H Perturbed into N(τ)H, with Varying Amounts of Data Collection at Each λ[a]

run no.[b]	cutoff (Å)[c]	number of equilibration steps (time (ps))	number of data collection steps (time (ps))	average $\Delta\Delta G_{hyd}(298)$ (kJ mol^{-1})[d]
1	9.0	500 (1.0)	1000 (2.0)	48.92 ± 0.77
2	9.0	500 (1.0)	1500 (3.0)	45.42 ± 3.28
3	9.0	500 (1.0)	2000 (4.0)	32.43 ± 0.21
3R	9.0	500 (1.0)	2000 (4.0)	29.45 ± 0.22
4	9.0	500 (1.0)	2000 (4.0)	35.46 ± 0.63
4R	9.0	500 (1.0)	2000 (4.0)	32.60 ± 1.38
5	9.0	500 (1.0)	3000 (6.0)	45.26 ± 1.76
6	8.0	500 (1.0)	2000 (4.0)	32.08 ± 0.53
7	10.0	500 (1.0)	2000 (4.0)	29.18 ± 0.11
8	11.0	500 (1.0)	2000 (4.0)	33.06 ± 1.95
9	9.0 + Cl$^-$	500 (1.0)	500 (1.0)	34.10 ± 0.01
10	9.0 + Cl$^-$	500 (1.0)	1000 (2.0)	35.30 ± 0.47
11	9.0 + Cl$^-$	500 (1.0)	2000 (4.0)	31.80 ± 0.48
11R	9.0 + Cl$^-$	500 (1.0)	2000 (4.0)	36.68 ± 0.33

[a] Results are given for the change N(π)H → N(τ)H. [b] Runs denoted by *n*R are started from the final configuration of run *n*. [c] 9.0 + Cl$^-$ indicates the presence of a chloride counterion. [d] Errors are from the double-wide sampling during the simulation.

rameters[31] were used throughout the calculations, taking histamine atom types from the standard residue histidine. The exception was the atom-centered charges, which were initially fitted to a molecular electrostatic potential (MEP) from a 6–31G wave function. The cutoff was 9 Å, and a constant dielectric of 1 was used. Initially, 20 steps of steepest descent were used to remove any bad contacts formed by the solvation process. This was followed by conjugate gradient minimization.

After minimization, the system was put into the NEWTON module for 8 ps of dynamics to equilibrate the system, starting at 0.2 K and finishing at 298 K and a constant pressure of 1 atm. In addition to the conditions applied in BORN, SHAKE was used to constrain all bonds and a time step of 0.002 ps was used. The center of mass motion was also removed every 50 steps to keep the solute in the center of the box.

The GIBBS module was then used to carry out the perturbation. As before, the run was divided up into 21 windows. The coupling parameter, λ, started at a value of 1 and decreased in each window by 0.05 until in the last window it had a value of 0. To prevent the problem of end-point catastrophes, the initial dummy bond lengths were started at a value of 0.4 Å and scaled with λ to be finally the desired N–H length. The center of mass motion was removed only at the beginning of each window, otherwise the conditions were the same as for the dynamics. This was found to be necessary as otherwise, due to a problem in AMBER3.1, it was possible for the solute to leave the solvent box during a long simulation. Providing there is enough equilibration before the data collection begins, this should have no effect. The problem of equilibration will be returned to below.

Data Convergence: The Effect of Phase Space Sampling

Convergence problems are usually due to poor sampling, i.e. insufficient data collection. This sampling could be aggravated by the implementation of a simple nonbonded cutoff, especially in the case of a charged molecule. In order to test these hypotheses, it was decided simply to run the FEP calculations with various numbers of data collection steps and different cutoffs. It would be hoped that as the ensemble size and cutoff increased, the result should converge. The N(π)H to N(τ)H perturbation was studied in more depth, as this run gave the result most different from the mean.

Five runs were made using different amounts of data collection. Initially 3 ps more dynamics were run to ensure equilibration in this module. Then at each value of λ, the system was equilibrated for 1 ps before data collection commenced. The results for these runs are given in Table 2 (runs 1–5).

With 2000 data collection steps, the result seems to have converged and the $\Delta\Delta G_{hyd}(298)$ for this perturbation is now very similar to the three others calculated in the original study. These other simulations were thus repeated using these present conditions, i.e. a large box of waters and 4 ps of data collection. Firstly, the above run, run 3, was performed in reverse, i.e. starting at λ = 0 with the N(τ)H tautomer in the conformation left at the end of the run and perturbing the ring to give the N(π)H tautomer when λ = 1 (run 3R). An independent run, starting as the N(τ)H tautomer minimized and equilibrated using the above conditions, was also set up and run in both directions, i.e., starting at λ = 1 (run 4) and then, from the final configuration of this run as N(π)H, at λ = 0 (run 4R). These results are all very similar, and the internal errors are small. Unfortunately, an increase of data collection to 3000 steps moved the value back to its original high value, with an increase in hysteresis. This result meant that either the ensemble average was still a long way from convergence, even with 6 ps of data collection per window, or something else was not being considered.

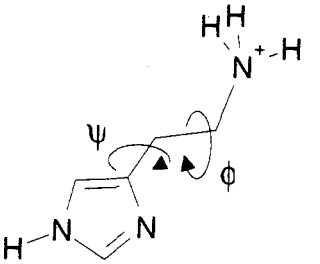


Figure 3. Torsion angles in histamine monocation that produce changes in conformation.

The next step was to vary the nonbond cutoff distance and see the effect this had on the calculation. Again starting from the 8 ps equilibrated N(π)H configuration, simulations were run in the same way as run 3, 3 ps of equilibration followed by 21 windows with 1 ps of equilibration at each λ and 2000 steps (4 ps) of data collection. The results for these simulations are also in Table 2 (runs 6–8).

While these values using different cutoffs do not seem to be very different, it is noticeable both that there is no convergence and also that with the largest cutoff used the internal error increases, indicating that the configurations sampled by successive λ are different. Thus, while a cutoff in the region of 8–11 Å does not seem to have much direct effect on the result, it does affect the configurations seen, and this could be significant if interactions out to, say 20 Å, are included.

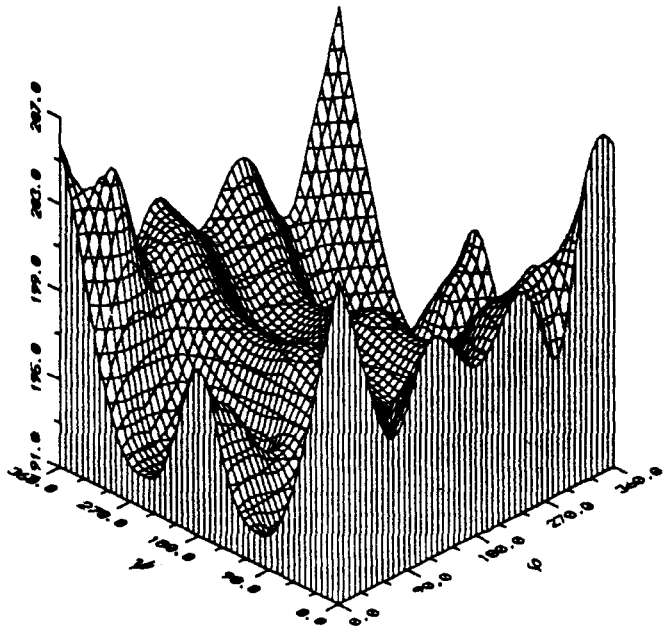
To damp out the long-range forces, a counterion was then included. It was argued that if a 1– ion is present in close proximity to the 1+ ion, the effective field will drop faster to 0, effectively modeling the long-range interaction of the histamine monocation by the interaction with the counterion. A chloride ion was added 2 Å from the N^+H_3 group of the histamine side chain. The parameters for chloride were not in the standard AMBER parameter set but were taken from Cieplak *et al.*[32] The minimization and equilibration stages were repeated, during which it was noted that the counterion did not travel more than 6 Å from the quaternary nitrogen atom.

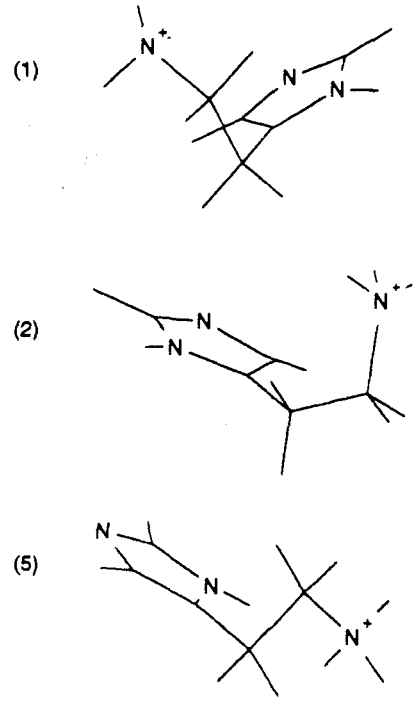
Three simulations (run 9–11 in Table 2) were run using various numbers of data collection steps to see if the $\Delta\Delta G_{hyd}$ value converged more quickly when the long-range forces were damped. It can be seen that the results are much less dependent on the data collection size. This could mean either that eliminating the long-range forces prevents the problems seen earlier with sampling or that the bigger electrostatic potential in the vicinity of the solute means that there are fewer configurations available to be sampled due to an ordering of the solvent, or even that the free energy surface of the histamine has been altered, preventing some conformers from being sampled.

To remove this possible biasing, it is necessary to model the long-range forces explicitly without the present of a counterion and all the following calculations were made with the histamine

(31) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.

(32) Cieplak; Lybrand; Kollman, P. A. *J. Chem. Phys.* **1987**, *86*, 6393.





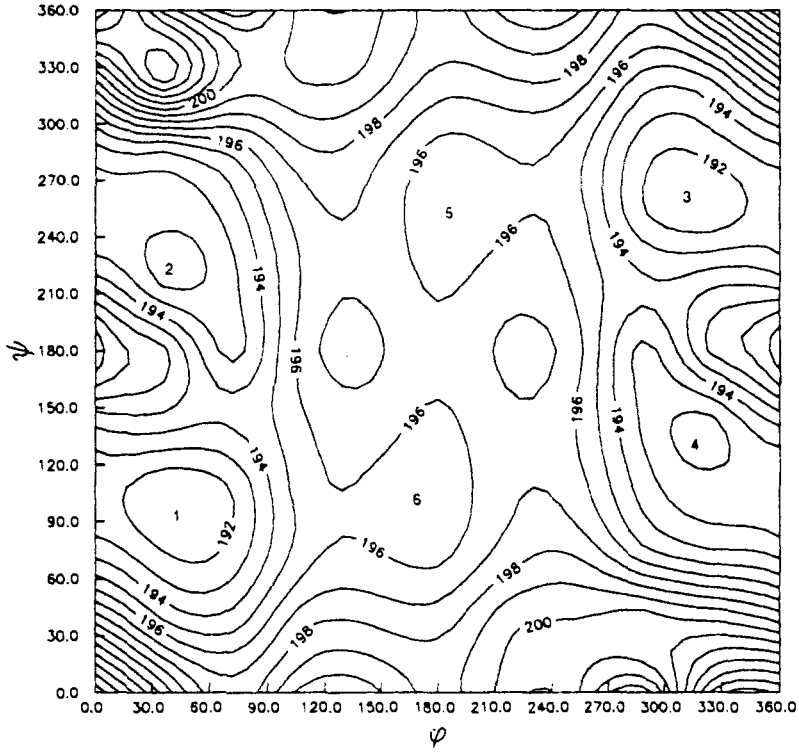


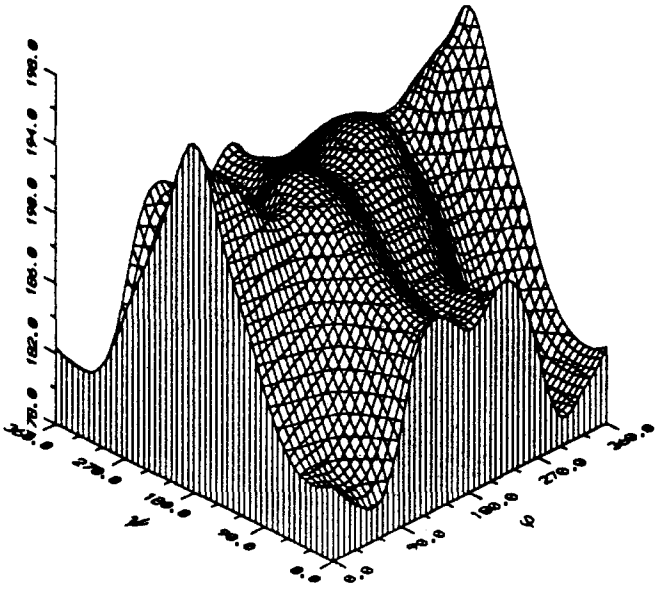
Figure 4a. Energy surface and corresponding contour plot for the variation in energy of histamine monocation N(π)H with the change in torsion angles φ and ψ. The minima are numbered and relevant conformers shown. The conformers of minima 3, 4, and 6 are found by inverting structures 1, 2, and 5 with respect to the plane of the ring. Contours are marked in kilocalories/mole.

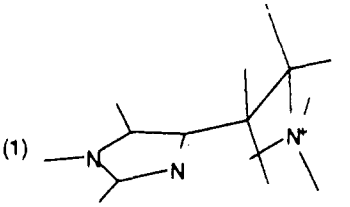
monocation along in solution. Before this is done, we shall examine in more depth the problems associated with a system of many possible conformations.

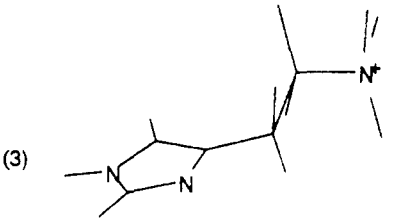
Effect of Conformational Flexibility on the Tautomer Gas-Phase Free Energy

Previously, only the "crystal structure" was used to calculate both the atomic charges for the simulations and the gas-phase energies. It was not determined whether these were the global minima or how the equilibrium is affected by the possibility of histamine adopting different conformations. If conformations other than that in the crystal are important, ΔG_{gas} would have to be modified accordingly. A method would also be needed that can better model the changing charge distribution of the molecule as a function of conformation.

In histamine three degrees of torsional freedom exist along the side chain. The torsion involving the quaternary hydrogens can be ignored as long as they are free to rotate and follow the conformer into its minima. The important degrees of freedom are the torsion angles shown in Figure 3.







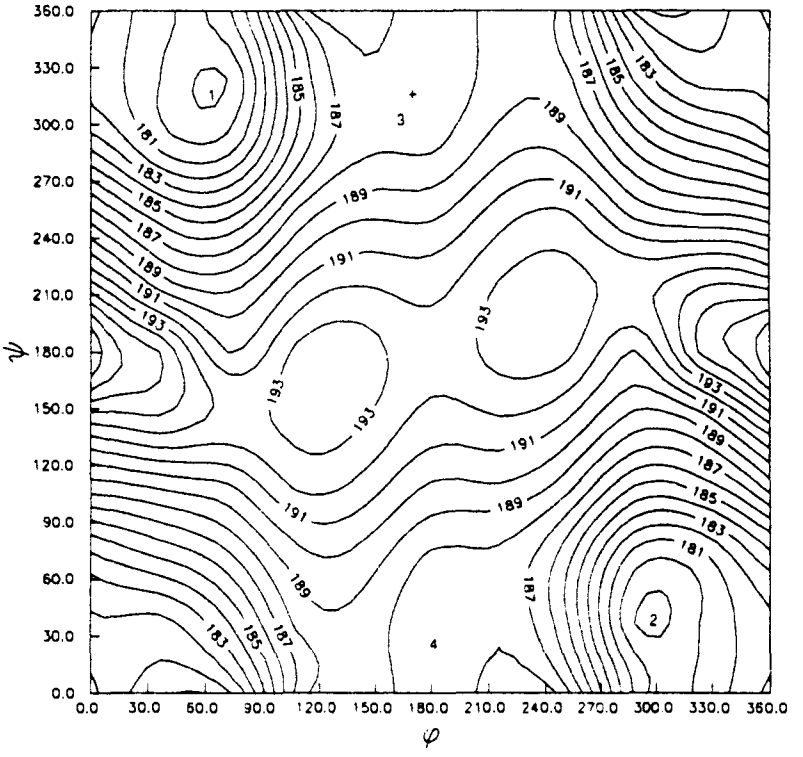


Figure 4b. Energy surface and corresponding contour plot for the variation in energy of histamine monocation N(τ)H with the change in torsion angles φ and ψ. The minima are numbered and the relevant conformers shown. The conformers of minima 2 and 4 are found by inverting structures 1 and 3 with respect to the plane of the ring. The crystal structure geometry is denoted by a cross.

MOPAC5[33] was used to calculate a potential energy surface for this torsional space using an AM1 Hamiltonian.[34] This calculation method was used in the original calculations of K_T-(histamine, aq), and it was decided that it performed badly.[2] However, on reevaluating this data it was discovered that a mathematical error had occurred and in fact the AM1 result was of the same order of magnitude as the full *ab initio* calculation. Grid points were placed at 36° intervals, and the geometry, excluding the torsions φ and ψ, was optimized at each point. The resulting surfaces are shown in Figure 4.

Each minimum is a stable conformation and can be thought of as different species which are in equilibrium. Free energy is a state function, and so

$$G^\circ_m = \sum_i n_i G^\circ_{m,i} \qquad (10)$$

where n_i is the number of moles of species i with standard molar Gibbs free energy $G^\circ_{m,i}$. From statistical mechanisms it is known that, for the equilibrium A → B, the ratio between two species is given by

(33) Stewart. J. P. P. *MOPAC 5.0*; Frank J. Seiler Research Lab., U.S. Air Force Academy, Colorado 80840.

(34) Dewar, M. J. S.; Zoebisch, E. G.; Healey, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

Table 3. Geometries and Energies of the Lowest Energy Conformers of Histamine Monocation Conformers from MOPAC AM1 Potential Surfaces, Calculated Using MOPAC and GAUSSIAN88

minima[a]	optimized ϕ^g 3-21G (AM1)	optimized φ^g 3-21G (AM1)	AM1 $\Delta H_f(0)$ (kcal mol^{-1})	q_{tot}^e ($\times 10^{-34}$)	n_i^f	RHF/3-21G//b (au's)	RHF/6-31G*//b (au's)
			(a) Histamine Monocation N(π)H				
1	56.34 (50.67)	80.25 (96.87)	190.75	2.68	0.428	−356.288 55	−358.277 31
2	55.71 (49.27)	208.57 (213.64)	191.23	2.42	0.320	−356.286 53	−358.276 01
5[c]	181.73 (186.56)	279.49 (258.13)	195.265	10.57	0.252	−356.278 94	−358.269 93[d]
			(b) Histamine Monocation N(τ)H				
1	58.86 (61.57)	319.63 (324.87)	178.54	2.24	0.896	−356.327 20	−358.305 86
3[c]	168.73 (180.28)	318.77 (326.21)	187.15	9.24	0.104	−356.299 42	−358.288 19[d]

[a] Minima numbered as on contour plots in Figure 4. [b] Energy calculated on RHF/3-21G-optimized geometry. [c] These structures were started from the earlier 6-31G-optimized structures based on the N(τ)H crystal structure, with the ring proton appropriately moved. The crystal angles are ϕ = 170.22° and φ = 317.05°. [d] These energies were actually calculated on the 6-31G-optimized geometry, for which the relevant dihedrals vary by about a degree from the 3-21G angles. [e] q_{tot} is the total partition function for the various conformers, calculated from the AM1 vibrational analysis. [f] n_i is the number of moles of conformer calculated to be present in 1 mol of species, calculated from eq 10. [g] The values, in degrees, of the conformer torsion angles after optimization by *ab initio* calculations with a basis set of 3-21G. In parentheses are the values after AM1 optimization.

$$K = \frac{q_B}{q_A} \exp(-\beta \Delta E_0) \qquad (11)$$

where ΔE_0 is the difference in ground-state energies between A and B and q_A, q_B, their molecular partition functions. We now need to calculate these quantities for the various histamine rotamers.

What are the species involved? The energy surfaces show that N(π)H has four deep gauche–gauche minima (1–4) and two flat high-energy gauche–trans minima (5 and 6). N(τ)H has two deep gauche–gauche minima (1 and 2) and again two gauche–trans minima on the high plateau (3 and 4). The symmetry of the system can also be seen. The imidazole ring lies in a plane of reflection, and so each minimum has a mirror image. The crystal geometry[35] lies in minimum 3 of the N(τ)H plot. The corresponding N(π)H geometry used in the earlier calculations lies in minimum 5. For the N(π)H tautomer this is ≅5 kcal mol^{-1} above the lowest minimum, while for N(τ)H this height is ≅9 kcal mol^{-1}. Of the six minima for N(π)H, the molecular symmetry means that only three are unique. Similarly, for N(τ)H, only two minima are unique. Thus, only five conformations need to be studied in depth, stick plots of which are shown alongside the surfaces.

These conformers were taken and fully optimized using AM1 and the *ab initio* program GAUSSIAN88[36] with a basis set of 3–21G.[37] 6–31G*[38] single-point energies were then calculated using these 3-21G geometries. A normal mode analysis was then performed using AM1 on the AM1-optimized structure to give the data needed to form the partition functions. The geometries and energies are given in Table 3, along with the number of molecules calculated to be present in 1 mol of histamine tautomer in the gas phase at 298 K.

The higher partition functions for the gauche–trans conformations are due to the existence of many more accessible vibrational levels than for the gauche–gauche absolute minimum conformation. Presumably, the intramolecular bonding that occurs in these later structures makes them more rigid, pushing up the vibrational energy and so widening the separation of the levels. Also, the slightly higher moment of inertia of the gauche–trans side chain gives closer rotational levels. Because of this, the gauche–trans conformation is present in appreciable amounts in the N(π)H system and in small amounts in the N(τ)H system, despite it being several kilocalories/mole above the absolute minimum.

Table 4. Difference in Free Energy and the Equilibrium Constant between Histamine Monocation Tautomers in the Gas Phase

basis set	ΔG(g,298) (kJ mol^{-1})	K_T(g,298)
AM1	−51.51	1×10^9
3-21G	−100.41	4×10^{17}
6-31G*	−74.58	1×10^{13}

The various thermodynamic quantities, $\Delta H_{0\to298}$, $\Delta S_{0\to298}$ (the change in enthalpy and entropy on going from 0 to 298 K), and zero-point energies, are also produced by MOPAC5 from the vibrational analysis. The gas-phase energies can now be corrected to Gibbs free energies at 298 K and so the gas-phase equilibrium constant between the tautomers calculated. The results are given in Table 4. The difference between these results and that in Table 1 shows that there is a significant change in the equilibrium when all conformers are taken into account. The next question is how does this problem affect the solution-phase calculation?

Using Multi-Conformational Atom-Centered Charges for Conformationally Flexible Systems

Returning to the molecular mechanics calculations, one source of error that has not yet been examined is whether the charges used in the simulations bias the conformations sampled. In the AMBER force field, the electrostatic interaction between two atoms is modeled with a Coulombic potential where the relevant charges are atom-centered partial charges calculated from fitting to *ab initio* molecular electrostatic potentials (MEPs) at random points on surfaces beyond the van der Waals surface of the molecule. In all the previous simulations the atom-centered charges had been fitted to the MEP from a 6–31G[39] wave function calculated on a 6–31G-optimized structure. As mentioned above, this structure was, for both tautomers, based on that found in the histamine hydrochloride crystal structure,[35] which is the monocation in the gauche–trans N(τ)H form.

The charges calculated at one geometry fit the dipole moment of that geometry very well and so are thought to model the electrostatic properties of the molecule adequately. In a different conformation from that for which the charges are calculated, the charges may no longer be able to reproduce the relevant dipole. This has been found to be a problem with simulations of torsionally flexible molecules. In the simulations of propanol for example it was found that the charges calculated on just one conformation prevented the molecule from accessing all its conformations.[40] This effect could also be present in this case.

A method has been developed to try and take this conforma-

(35) Prout, K.; Critchley, S.; Ganellin, C. R. *Acta Crystallogr. Sect. B* **1974**, *30*, 2884.
(36) Frisch, M. J.; Head-Gordon, M.; Schlegel, H. B.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Defrees, D. F.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, R.; Kahn, L. R.; Stewart, J. P. P.; Fluder, E. M.; Topiol, S.; Pople, J. A. *GAUSSIAN88*; Gaussian Inc.: Pittsburgh, PA, 1988.
(37) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.
(38) Binkley, J. S.; Pople, J. A. *J. Chem. Phys.* **1977**, *66*, 879.
(39) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.
(40) Reynolds, C. A.; Essex, J. W.; Richards, W. G. *Chem. Phys. Lett.* **1992**, *199*, 257.

Table 5. Charges Calculated for the Various Conformers of Histamine and Fitted to the Boltzmann Averaged MEP[a]

(a) Histamine Monocation N(π)H

		6-31G*//a geom 1	6-31G*//a geom 2	6-31G*//b geom 5	weighted[d]	6-31G//b geom 5	QM[g]
HN11		0.361	0.376	0.321	0.351	0.345	
N		–0.593	–0.587	–0.379	–0.496	–0.459	
HN12		0.361	0.376	0.321	0.354	0.345	
HN13		0.361	0.376	0.321	0.337	0.345	
CA		0.399	0.088	0.344	0.260	0.354	
HA1		0.026	0.108	0.034	0.066	0.039	
HA2		0.022	0.139	0.017	0.075	0.018	
CB		–0.458	–0.497	–0.363	–0.502	–0.426	
HB1		0.190	0.208	0.153	0.181	0.184	
HB2		0.146	0.202	0.120	0.208	0.138	
CG		0.226	0.270	0.096	0.241	0.093	
ND2		–0.523	–0.583	–0.472	–0.501	–0.526	
HD2		0.395	0.414	0.382	0.396	0.390	
CE2		0.287	0.354	0.262	0.266	0.378	
HE2		0.153	0.138	0.149	0.154	0.128	
CD1		0.057	0.050	0.071	–0.056	0.194	
HD1		0.129	0.151	0.146	0.162	0.116	
NE1		–0.541	–0.583	–0.525	–0.496	–0.657	
dipole moment[f]							
geom 1	*x*	9.709	9.893	9.976	9.854		9.716
	y	–0.547	–0.379	0.463	–0.559		–0.566
	z	–2.753	–2.369	–2.817	–2.798		–2.693
geom 2	*x*	–9.357	–9.510	–9.659	–9.459		–9.421
	y	1.092	1.308	0.410	1.206		1.335
	z	3.374	3.549	3.913	3.583		3.517
geom 5	*x*	–13.769	–13.771	–14.807	–14.301		0.000
	y	–2.939	–2.809	–2.999	–3.103		0.001
	z	0.430	0.566	–0.084	0.157		0.013

(b) Histamine Monocation N(τ)H

		6-31G*//a geom 1	6-31G*//b geom 3	weighted[e]	6-31G//b geom 3	QM[g]
HN11		0.295	0.325	0.350	0.363	
N		–0.414	–0.425	–0.551	–0.568	
HN12		0.295	0.325	0.358	0.363	
HN13		0.295	0.325	0.331	0.363	
CA		0.302	0.551	0.494	0.567	
HA1		0.057	–0.026	0.018	–0.021	
HA2		0.028	–0.028	–0.007	–0.008	
CB		–0.448	–0.615	–0.634	–0.632	
HB1		0.168	0.178	0.206	0.199	
HB2		0.166	0.159	0.191	0.162	
CG		0.357	0.522	0.491	0.497	
ND2		–0.389	–0.575	–0.619	–0.653	
CE2		0.001	0.146	0.198	0.266	
HE2		0.219	0.178	0.178	0.146	
CD1		–0.388	–0.410	–0.370	–0.339	
HD1		0.269	0.258	0.264	0.247	
NE1		–0.167	–0.256	–0.271	–0.326	
HE1		0.354	0.368	0.374	0.375	
diple moment[f]						
geom 1	*x*	3.771	4.708	3.813		3.740
	y	–1.448	–2.230	–1.466		–1.455
	z	0.430	0.393	0.444		0.394
geom 3	*x*	–6.849	–9.043	–7.871		0.000
	y	1.286	1.303	1.523		0.067
	z	–1.210	–0.939	–0.860		0.311

[a] "Geom number" designates the minima conformer used in the calculation, taken from the contour plots in Figure 4. The ratio in the weighted MEP column shows the weighting of the various geometries. The final column in both tables is the 6-31G charges used in the original simulations. All MEPs were calculated with a RHF wave function. [b] 3-21G-optimized geometry. [c] 6-31G optimized geometry. [d] Weighted MEP 0.43:0.32:0.25 (geom 1:geom 2:geom 5). [e] WeightedMEP 0.90:0.10 (geom1:geom 3). [f] Dipole moment along the Cartesian axes, calculated from the various charge sets on the various geometries. [g] Dipole moment from the quantum mechanical calculation.

tional flexibility into account.[41] The MEPs are calculated for all the important conformations of the molecule, i.e. those that would be present under the conditions of the simulation. A Boltzmann weighted average of all these MEPs is then taken, and the charges are fitted to this. So far it has been found that these multiconfiguration charges reproduce the dipole at all the included geometries better than any single set.

(41) Essex, J. W.; Reynolds, C. A.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 3634.

The minima conformations have already been optimized to 3-21G. Wave functions at 6-31G* have then been calculated for these structures. Charges obtained by fitting to MEPs from these wave functions are thought to be the best that can be practically obtained. This was therefore done. The results are given in Table 5.

Dipole moments along Cartesian axes calculated with these charge sets at the various geometries are also given. Inexplicably, the quantum mechanical calculations on the gauche–trans forms

Table 6. $\Delta\Delta G_{hyd}(298)$ for the Histamine Monocation N(π)H Perturbed into N(τ)H, Using Conformationally Weighted Charges, with Varying Amounts of Data Collection and Equilibration at Each λ[a]

run no.	charge set	number of equilibration steps (time (ps))	number of data collection steps (time ps))	average $\Delta\Delta G_{hyd}(298)$ (kJ mol^{-1})[e]
12	weighted[b]	500 (1.0)	1000 (2.0)	49.60 ± 0.22
13	weighted[b]	1000 (2.0)	1000 (2.0)	39.98 ± 1.46
13R	weighted[b]	1000 (2.0)	1000 (2.0)	43.64 ± 0.36
14	weighted[b]	2000 (4.0)	2000 (4.0)	39.22 ± 1.12
15	weighted[b]	4000 (8.0)	4000 (8.0)	38.64 ± 0.24
15[d]	weighted[b]	4000 (8.0)	4000 (8.0)	−55.38 ± 0.28
15R	weighted[b]	4000 (8.0)	4000 (8.0)	43.55 ± 4.88
16	weighted[b]	4000 (8.0)	4000 (8.0)	42.44 ± 3.49
16R	weighted[b]	4000 (8.0)	4000 (8.0)	41.08 ± 0.14
17	6-31G* [c]	4000 (8.0)	4000 (8.0)	36.13 ± 0.32
17[d]	6-31G* [c]	4000 (8.0)	4000 (8.0)	−99.06 ± 1.03
18	6-31G* [c]	4000 (8.0)	4000 (8.0)	35.28 ± 1.99

[a] Results given for N(π)H → N(τ)H. [b] Charge set from the weighted MEPs in Table 5. [c] Charge set from the 6-31G* charges on the gauche–trans geometries. [d] Evaluating intra- and intermolecular energy terms. [e] Errors are from the double-wide sampling during the simulation.

return a dipole moment of 0 along the x axis. All the molecules used the same set of axes (and comparable molecules identical coordinates), with the charged amino group at the origin and the rest of the structure in the positive quadrant, with the ring over the x axis, hence the large negative dipole in this direction in all other cases. With this exception, it can be seen that the weighted charge sets are indeed able to reproduce the dipole moments at all the geometries better than any one set alone. Hence, if the molecule adopts all these conformers, these charges should be the ones used.

Using the new conformationally adjusted charges, perturbations were then carried out as before, using 21 windows. Again, runs with various amounts of data collection per window were run to see how the new charges affect the convergence. These results are in Table 6. Runs 12–15 all start with the N(π)H tautomer at $\lambda = 1$, while run 16 starts with N(τ)H.

The result for the shortest run, number 12, is significantly higher than the others. The value appears to have converged by the time 8 ps of equilibration and 8 ps of dynamics are used at each window. However, this converged result is about 8 kJ mol^{-1} higher than that using the single geometry charges (the results in Table 2). The final value, taken from the average of the 8 ps equilibration/8 ps data collection runs, is therefore $\Delta\Delta G_{hyd}$ = 41.43 ± 6.53.

To check that this difference is not just an artifact of using a larger basis set, two further runs (17 and 18 in Table 6) were made using the 6–31G* charges calculated on the gauche–trans geometries given in Table 5. Run 17 starts with the N(π)H tautomer at $\lambda = 1$, while run 18 starts with N(τ)H. These are very similar to the runs made using the 6–31G charges and appear to have converged on the result given by runs 3 and 4 in Table 2.

Interestingly, it can be seen in Table 6 that the hysteresis increases as the amount of both data collection and initial equilibration is increased. After 8-ps MD/8-ps DC, the result appears to have converged on an average, but the sampling appears to be much worse than with the early runs. This means that the system was not being allowed to relax properly at each value of λ. When collecting data for the ensemble average, it is important that the system is at equilibrium at that value of λ. Initially it was believed that water relaxes around a solvent in under 1 ps; hence, this was the value used in the early FEP protocols. A major problem is that the solvation energy could be affected as the relaxation energy will also contribute. This is demonstrated by the histamine system. In Table 6 the difference in calculated free energy is 49.60 kJ mol^{-1} with 1 ps of equilibration but 39.98 kJ mol^{-1} with 2 ps. Assuming that the extra time does not mean that significantly different conformations are sampled, this difference is due entirely to the extra relaxation.

A more subtle effect is also produced. If full relaxation does not occur between windows, the system at $\lambda + \delta\lambda$ will be more like that at λ than it should be. Hence, running over a window from both directions will automatically be very similar. This leads to a small hysteresis on the double-wide sampling and a spurious belief that the calculation is accurate. This effect of this Hamiltonian lag has been discussed by Pearlman and Kollman in a different situation[42] and can also be clearly seen with the histamine results above. The hysteresis rises from 0.2 kJ mol^{-1} in run 12 to 1.5 kJ mol^{-1} in run 13 before dropping back to around 0.2 kJ mol^{-1} in runs 15 and 16R. When more equilibration steps are run between windows, allowing the system to relax properly, it becomes apparent that sampling in the shorter runs was very poor.

The water thus seems to need on the order of 2–3 ps to relax, not the 1 ps as thought earlier. It is possible of course that the relaxation has been extended due to the removal of solute center of mass motion at the start of the window. Even taking this into account, it is likely that longer than 1 ps would be needed. The reason for this mistake is well demonstrated by the work of Maroncelli and Fleming.[43,44] Using various models for solvation, it is found that, on average, the solvation energy response decayed after around 1 ps. However, if this was broken down into the separate solvation shells, the first-shell response had only decayed to half its initial value by this time.

The reason for the systematic large hysteresis error in all the runs starting with the N(τ)H tautomer and using the weighted charges is not obvious. However, it may be due to the fact that, as noted below, the weighted charges are a poor model for this tautomer.

Inclusion of a Reaction Field to Model the Long-Range Forces

As a final calculation, a reaction field method was applied to include all the long-range forces. The method that was chosen was the image point charge model of Friedman.[45] This has the advantage of simplicity of implementation, but the disadvantage that it is valid for a spherical cavity in a dielectric, not a cubic system with periodic boundary conditions.

The approximation that is made is that in a FEP calculation only the energy difference between the end points for the perturbed molecule is important. The reaction field is thus only applied to this molecule, the solute. For example for the histamine system, the total energy can be broken down into

$$V = V_{\text{solute–solute}} + V_{\text{solute–solvent}} + V_{\text{solvent–solvent}} \quad (12)$$

It is assumed that the solvent–solvent potential is fairly well described with an 8-Å cutoff. Simulations show that the loss of long-range dipolar coupling prevents a good description of the dielectric properties of water, but do not qualitatively affect calculated thermodynamic and structural values.[46] Hence, a solvent–solvent cutoff should not change a FEP calculation result. The solute–solute interaction is completely described within the cutoff, as the solute is smaller than this distance. Thus, the only term that needs modifying is the solute–solvent interaction. This is broken down further into

$$V_{\text{solute–solvent}} = V_{\text{short-range}} + V_{\text{long-range}} \quad (13)$$

The idea is now to put the solute inside a sphere of solvent inside the solvent. All interactions inside this sphere are calculated explicitly. They form the short-range part of the potential. A

(42) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1989**, *91*, 7831.
(43) Maroncelli, M.; Fleming, G. R. *J. Chem. Phys.* **1988**, *89*, 5044.
(44) Castner, E. W.; Fleming, G. R.; Baghi, B.; Maroncelli, M. *J. Chem. Phys.* **1988**, *89*, 3519.
(45) Friedman, H. L. *Mol. Phys.* **1975**, *29*, 1533.
(46) Barker, J. A.; Watts, R. O. *Mol. Phys.* **1973**, *26*, 789.

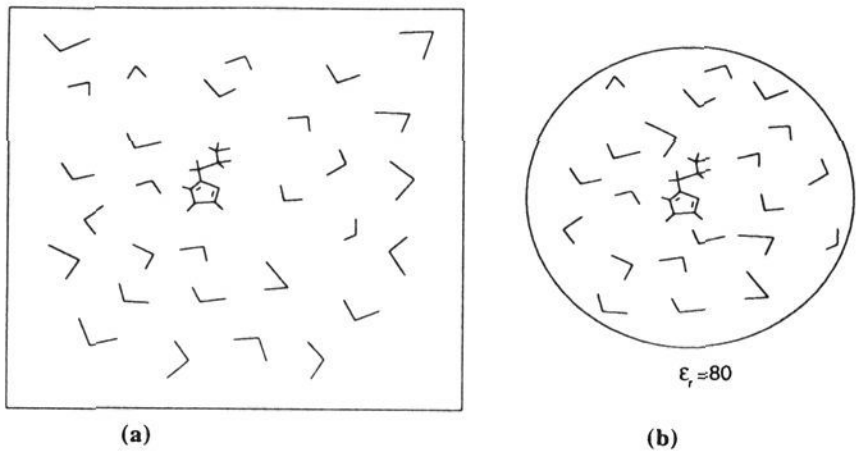


Figure 5. Changing a periodic system to allow the use of a reaction field: (a) the total system, a solute atom in a periodic box of solvent molecules; (b) how the system seems to the solute, a sphere of solvent molecules inside a continuum of dielectric 80.

reaction field is then included for the long-range part. All the solvent behaves as normal, and so boundary conditions are possible. What has changed is that the charged solute now feels a greater potential, that due to the explicit solvent and that due to the continuum model of the distant solvent. The diagram Figure 5 sums up this implementation.

The image charges modeling the reaction field, under the approximation that the dielectric outside the cavity, ϵ, is much larger than unity, are then given by simple relationships to the solute charges,

$$q_{\text{im}} = -\frac{\epsilon - 1}{\epsilon + 1}\frac{a}{r}q \qquad (14)$$

and position,

$$\vec{r}_{\text{im}} = \frac{xa^2}{r^2}\hat{i} + \frac{ya^2}{r^2}\hat{j} + \frac{za^2}{r^2}\hat{k} \qquad (15)$$

where the source charge is at (x,y,z) a distance r from the center of the cavity which has radius a.

From this, all that must be known is the magnitude of the source charge and its position. The appropriate scaling factors then give the model of the reaction field. This was added into GIBBS3.1 as follows. The cutoff for the solute was changed from one which is atom based to one which is residue based, with its origin at the center of mass. Every time the nonbonded pair list is calculated (every 100 steps of dynamics), the image charge for each solute atom is calculated according to the above eqs 14 and 15. These are then added into the pair list for the solute Coulombic potential energy, adding an extra n^2 interactions to the calculation, where n is the number of solute atoms, an insignificant number for a 40-atom solute in a box of solvent. The forces due to the long-range solute–solvent interactions are also added onto the solute as these are calculated from the gradient of the potential.

This method was then used to recalculate the FEP energies calculated above for the histamine tautomerism, using both charge sets, and these results are given in Table 7. Runs 20 and 22 start with the N(π)H tautomer, while runs 21 and 23 with the N(τ)H.

It can be seen that the reaction field has changed the FEP result by around 4 kJ mol^{-1}. With the conformationally averaged charges this change favors the N(π)H tautomer, whereas with the single geometry charges this is reversed.

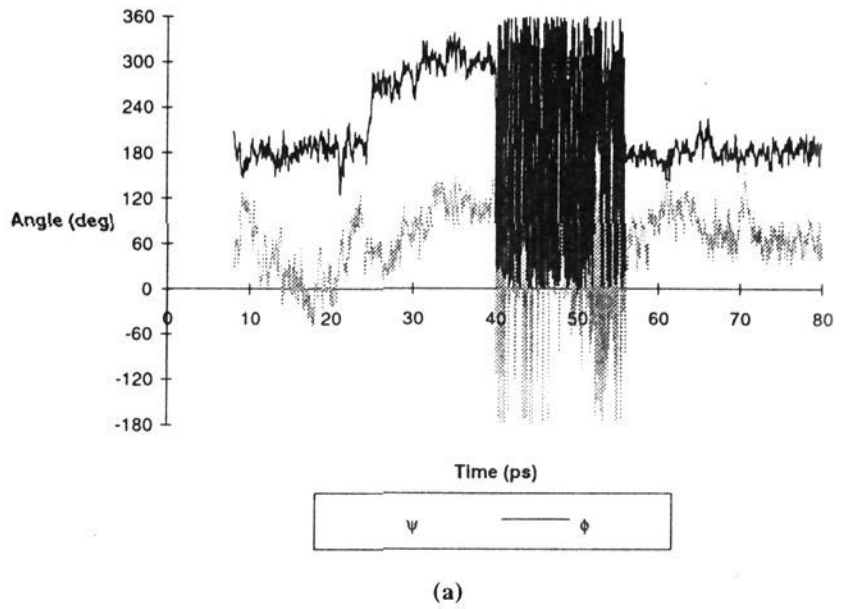
Calculating the Tautomeric Difference in Free Energy in Solution

We now have converged values for $\Delta\Delta G_{\text{hyd}}$ that have to be corrected to ΔG_{aq} to calculate possible equilibrium constants. It might be thought that from eq 4 simply adding the total ΔG_{gas}, shown in Table 4, to the FEP result gives the desired answer. However, as discussed in the Background and Method section, this would ignore the nature of the thermodynamic cycle which is being applied. Due to the calculation method, the required ΔG_{gas} should be taken over the conformers seen in solution, not in the gas phase.

Table 7. $\Delta\Delta G_{\text{hyd}}(298)$ for the Histamine Monocation N(π)H Perturbed into N(τ)H, Using Different Charge Sets, with the Inclusion of the Image Charge Reaction Field

run no.	charge set[a]	number of equilibration steps (time (ps))	number of data collection steps (time (ps))	average $\Delta\Delta G_{\text{hyd}}(298)$ (kJ mol^{-1})[b]
20	weighted	4000 (8.0)	4000 (8.0)	42.80 ± 2.77
20R	weighted	4000 (8.0)	4000 (8.0)	45.44 ± 3.36
21	weighted	4000 (8.0)	4000 (8.0)	45.31 ± 3.63
21R	weighted	4000 (8.0)	4000 (8.0)	42.83 ± 0.13
22	6-31G*	4000 (8.0)	4000 (8.0)	37.09 ± 0.42
23	6-31G*	4000 (8.0)	4000 (8.0)	30.29 ± 0.52

[a] Charge set denoted as in Table 5. [b] Errors are from the double-wide sampling during the simulation.



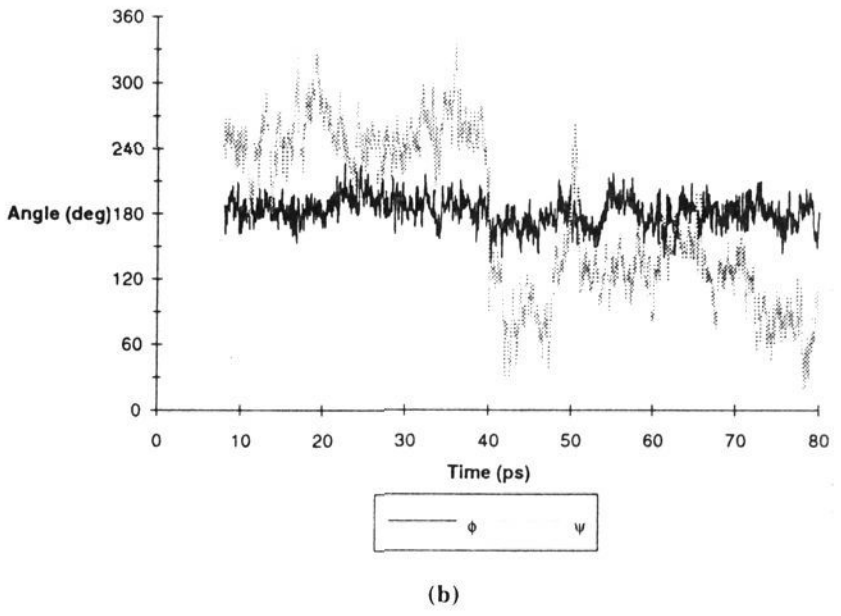


Figure 6. Variation of histamine monocation torsions with time for the (a) N(π)H tautomer and (b) N(τ) tautomer.

The question to be asked now is exactly which conformers are being seen in solution? To give an idea, 80 ps of dynamics were run using both tautomers. Coordinates of the histamine were dumped every 16 steps (0.032 ps). The torsion angles ϕ and ψ at each point were then calculated, and this variation is shown in Figure 6.

These plots show that conformational changes seem to occur on roughly 20 ps cycles. The N(π)H tautomer started in minimum 6 and then moved into minimum 4 after 20 ps. After 40 ps a large fluctuation between the conformers of minima 2 and 4 occurs for around 15 ps before the molecule returns to minimum 6. In contrast, the N(τ)H tautomer only sampled the gauche–trans

Table 8. Difference in Free Energy and the Equilibrium Constant between Histamine Monocation Tautomers in the Gas and Solution Phases[f]

		AM1[a]		3-21G[a]		6-31G*[a]	
	RF?[d]	$\Delta G(298)$ (kJ mol^{-1})[e]	$K_T(298)$	$\Delta G(298)$ (kJ mol^{-1})[e]	$K_T(298)$	$\Delta G(298)$ (kJ mol^{-1})[e]	$K_T(298)$
gas phase[b]		−24.25	2×10^5	−42.29	3×10^8	−38.41	5×10^7
charge set[c]							
weighted	no	17.18 ± 6.53	1×10^{-3}	−0.86 ± 6.53	1.4	3.02 ± 6.53	0.3
weighted	yes	19.85 ± 7.95	3×10^{-4}	1.87 ± 7.95	0.5	5.69 ± 7.95	0.1
6-31*G**	no	10.46 ± 4.19	1×10^{-2}	−6.58 ± 4.19	14.3	−2.70 ± 4.19	3.0
6-31G*	yes	9.44 ± 5.28	2×10^{-2}	−8.60 ± 5.28	32.2	−4.72 ± 5.28	6.7

[a] Basis set used in the *ab initio* calculations. [b] Gas-phase values calculated over the conformers seen in the solution-phase simulations. [c] Charge set used in the solution-phase calculation, as described in Table 6. [d] Was the reaction field correction to long-range forces included? [e] The error value is the standard deviation of four independent runs, combined with the average internal error. [f] K_T(expt) ranges from around 4 to 9; see Table 1.

conformation during the whole time, with ϕ staying around 180°. The only change was a move from minimum 3 to minimum 4 after 20 ps.

Only the gauche–trans N(τ)H conformer is seen, and so the gas-phase free energy required is just for this conformer. Similarly, G_{gas}(N(π)H) should be calculated using the weighting of conformers as seen in solution. For the simulations run above, this appears to be around 50% for minimum 2 and 50% for minimum 5.

More accurate populations could be obtained from longer simulations or calculations of transition rates between the conformers. However, this extra calculation is not warranted, and these results are adequate for the present purpose. What is important is that the transitions are slow. At no time when the simulations were checked, i.e. at the end of each window, was the N(τ)H tautomer in the gauche–gauche form. If the transitions were fast, these conformers would probably be sampled but would not necessarily be seen at these snapshot intervals. More evidence that it was not sampled is given by the convergence of the four runs. It is likely that due to the intramolecular hydrogen bond, the gauche–gauche form would have a much smaller solvation energy than its extended counterpart. If one of the runs moved in this minimum enough to be significant, that run would have a very different calculated energy. Hence, it is likely that this conformer is never seen during the simulations. This is important for the calculated intramolecular free energy from the gas-phase calculations, as this form has a significantly different internal energy compared to the trans–gauche rotamer. In contrast, the exact proportions of N(π)H conformers are not so essential as due to the higher partition function of the conformer in minima 5, the free energies are within 4 kJ mol^{-1} of each other.

Table 8 shows the final results. The gas-phase free energy differences are obtained in the same way as those in Table 4, i.e. the internal energy difference between the tautomers as calculated *ab initio* using the basis set shown at the top of the table, corrected for the zero point energy, entropy, and enthalpy contributions at 298 K as obtained from an AM1 vibrational analysis. These values differ from Table 4 in that the tautomer rotomers have been Boltzmann weighted according to the populations seen during the solution-phase simulations described above, rather than the gas-phase weightings. To these values have been added the solvation free energy difference, averages of the FEP values calculated using the different charge sets with or without the inclusion of the reaction field correction, as given in Tables 6 and 7. Errors are calculated as described in the Background and Method section.

All the methods now give a result that agrees, within the error bounds of the simulations, with the experimental value. The values using 6–31G* MEP-fitted charges taken from only the gauche–trans conformers are the closest. This is not surprising as these were the species predominant in the calculation. It can also be seen that the use of the reaction field makes a significant difference to the results.

To compare the approach used with a full molecular mechanics FEP run, during runs 15 and 17, the intramolecular term was also evaluated. The results are given in Table 6. The values are much more negative than the final values in Table 8. This difference might be due to nonconvergence of the simulations, but given the length of the simulations and the large size of the divergence, this is unlikely. Noting the large difference between runs 15 and 17, the error is more likely to be due to the electrostatic intramolecular energy as these runs used different charge sets.

Later experience seems to indicate that atom-centered charges derived from 6–31G* MEPs are on average larger than standard AMBER charges, which were derived from STO-3G calculations. This may make them incompatible with the AMBER Lennard–Jones parameters which were optimized with the smaller charges. For example the amino nitrogen in protonated lysine, from which the atom type was taken, has a charge of −0.14 compared to values of around −0.5 used here. A similar difference is found for the β-carbon from histidine, −0.1 (standard) compared to −0.5 (used here). The electrostatic energies produced by these charges over the short distances inside the solute could make a significant difference to the intramolecular energies. Perhaps crucially, there is also a difference for the δ-nitrogen in the N(π)H tautomer, −0.15 (standard) against −0.5 (used here), which is not so pronounced for the N(τ)H tautomer. This large negative charge near the negative β-carbon and amino nitrogen atoms could lead to the destabilization of the N(π)H tautomer seen in the simulations when the intramolecular energy is included.

The charges used do however reproduce the 6-31G* dipole moment, and so the interaction with the solvent will be correctly simulated. For this reason the results ignoring the intramolecular terms are closer to experimental values. This points to the necessity of correctly determining electrostatic parameters, not only to describe electrostatic properties of a molecule but also to fit in with other nonbonded parameters used. More work is needed to clarify the situation.

Discussion

This work set out to investigate the problems involved in using the FEP method to calculate the free energy difference between two charged, flexible molecules, in this case the tautomers of histamine monocation. In agreement with other authors, the major effects are found to be due to relaxation between windows, the modeling of electrostatics and long-range interactions, and conformational sampling.

It is essential that enough molecular dynamics steps are run before data collection begins to ensure a relaxed and equilibrated system. In these calculations this was on the order of 2–3 ps; it could possibly be shorter if solute center of mass motion had not been removed at the start of each window. Lack of relaxation leads to incorrect sampling and poor results, but with low statistical errors calculated within the run. This problem is easily seen by running simulations with different amounts of equilibration. If the internal error increases with increasing equilibration, relaxation is a problem and the equilibration time should be increased until the error drops to a suitable value.

Similarly, it is important to run enough data collection steps to produce a representative ensemble. This can again be seen by using multiple runs of varying length, and enough data collection

should be used to give converged values from independent simulations. This was found to be on the order of 8 ps for this system.

In this system, the inclusion of long-range forces is also important. The use of a simple point charge reaction field to model solute–solvent interaction out to infinity was adapted to apply in a periodic water simulation. This produced around a 4 kJ mol^{-1} difference in calculated free energy of hydration.

Also critical is the choice of electrostatic parameters. Different charge sets, calculated using the same *ab initio* basis set but on different conformers, also produced around a 4 kJ mol^{-1} difference in calculated free energy of hydration. Interestingly, this difference was greater when combined with the use of the reaction field. The essential factor is that the charges must reflect the solute structure; when different conformers are possible, this must be accounted for. The best method would obviously be dynamic charge sets that follow the molecule as it changes shape. This however would be difficult to computationally implement and to our knowledge has not yet been attempted. If static charges are to be used, one possible method is to produce them averaged over the possible conformers. Here, gas-phase conformer ratios were used to produce a set of charges which modeled the quantum mechanical dipole in all states. Conformer states during a simulation are then unbiased by the atom point charges. It is then possible to evaluate which conformers are actually present during the simulation, and charges that accurately represent these should be used. In this case the 6–31G* charges produced from the most populated trans–gauche forms were found to be relevant.

As well as affecting the charge set to be used, conformer sampling also affects the simulation result. In the case of histamine, the torsional behavior is a serious difficulty for searching phase space. Even after 80 ps the N(π)H tautomer has not sampled the minima 1 and 3 at all, and N(τ)H does not at any time go into minima 1 or 2. Using proton NMR, Ganellin *et al.*[47] deduced that only 45% of histamine in solution is in the trans form, i.e. minimum 3 for N(π)H and minimum 5 for N(τ)H. Combined with the known ratio of the tautomers, a minimum of 20% of the N(τ)H tautomer must be in the gauche–gauche form. This indicates that in fact all the species are seen in solution, and so the simulation is not sampling the full histamine monocation phase space. Various methods can be used to correct for this problem, either long MD simulations, umbrella sampling transition rate calculations, or using FEP between restricted conformational space.

In an FEP simulation however, it is not necessary to sample the full phase space. Only a representative sample of configurations of energy differences between the two end states of the perturbation is needed. If all configurations have the same perturbation energy difference, one alone would give the desired answer. In the histamine system, it appears that the various conformers of each tautomer have similar free energies in solution. This can be seen qualitatively from the *ab initio* calculations. The N(π)H tautomer with the most positive internal free energy (geometry 5) has the highest dipole moment due to its longer length and is therefore likely to have a higher solvation free energy. The same is true for the N(τ)H tautomer (geometry 3), with the added effect that the intramolecular hydrogen bond in the lowest energy conformer would further increase this relative stabilization. The gas-phase free energies of the conformers for each tautomer would thus become more similar on conversion to solution phase and conformer sampling correspondingly less important.

The method used to calculate the free energy, i.e. calculating the intramolecular part using *ab initio* calculations and the intermolecular part by molecular mechanics, cannot be endorsed on methodological grounds. It does not guarantee that the correct behavior of the intramolecular sampling with respect to the coupling parameter is reproduced. However, static gas-phase *ab initio* calculations can give accurate internal free energies, usually at no extra computational cost as these calculations are routinely used to evaluate the atom-centered charges used in the simulation. The FEP is then used as a solvation correction to these calculations. This has the advantage of avoiding the need for a highly optimized intramolecular force field. For example in these calculations, parameters developed for histidine were used. These should give the correct functional form for the molecular motion when applied to histamine monocation, but may not be quantitatively correct. The results obtained here seem to indicate that the solvation energy is indeed being adequately represented, but the intramolecular potential, possibly due to the incorrect application of charges, needs further optimization.

In summary, this paper makes various recommendations for the use of the FEP method to calculate free energy differences. Firstly, it is essential to obtain converged values in different simulations, free of relaxation effects. Secondly, it is important to model the electrostatic properties of all the conformers present during the simulation; it may also be necessary to include long-range forces. The final problem is then to see if all the possible conformers have been sampled and if not, whether this is important. The use of *ab initio* calculations to evaluate the intramolecular terms is also seen to be useful, allowing the performance of the empirical potential to be assessed.

A value for the equilibrium constant between the histamine monocation tautomers has been calculated as 6.7, in good agreement with experimentally determined values. This demonstrates both the power and pitfalls of FEP. Although a large variation of results with different simulation conditions is possible, by taking into account the essential features of a system necessary for its energetic description, quantitatively correct values can be obtained.

(47) Ganellin, C. R.; Pepper, E. S.; Port, G. N. J.; Richards, W. G. *J. Med. Chem.* **1973**, *16*, 610.
(48) Pyman, F. L. *J. Chem. Soc.* **1922**, *121*, 2616.
(49) Charton, M. *J. Org. Chem.* **1965**, *30*, 3346.
(50) Ganellin, C. R. *Jerusalem Symp. Quantum Chem. Biochem.* **1974**, 43.
(51) Reynolds, W. F.; Tzeng, C. W. *Can. J. Biochem.* **1977**, *55*, 576.
(52) Wasylishen, R. E.; Tomlinson, G. *Can. J. Biochem.* **1977**, *55*, 579.