7440-57-5; $CH_3(CH_2)_nSH$ (n = 3), 513-53-1; $CH_3(CH_2)_nSH$ (n = 4), 110-66-7; $CH_3(CH_2)_nSH$ (n = 5), 111-31-9; $CH_3(CH_2)_nSH$ (n = 6), 1639-09-4; $CH_3(CH_2)_nSH$ (n = 7), 111-88-6; $CH_3(CH_2)_nSH$ (n = 8), 1455-21-6; $CH_3(CH_2)_nSH$ (n = 9), 143-10-2; $CH_3(CH_2)_nSH$ (n = 10), 5332-52-5; $CH_3(CH_2)_nSH$ (n = 11), 112-55-0; $CH_3(CH_2)_nSH$ (n = 12), 19484-26-5; $CH_3(CH_2)_nSH$ (n = 13), 2079-95-0; $CH_3(CH_2)_nSH$ (n = 14), 25276-70-4; $CH_3(CH_2)_nSH$ (n = 15), 2917-26-2; $CH_3(CH_2)_nSH$ (n = 16), 53193-22-9; $CH_3(CH_2)_nSH$ (n = 17), 2885-00-9; $CH_3(CH_2)_nSH$ (n = 18), 53193-23-0; $CH_3(CH_2)_nSH$ (n = 19), 13373-97-2; $CH_3(CH_2)_nSH$ (n = 20), 66326-17-8; $CH_3(CH_2)_nSH$ (n = 21), 7773-83-3.

Supplementary Material Available: Table of wetting properties of monolayers prepared from different solvents, IR spectra of low-quality monolayers on copper, plot of intensity of CH_2 modes with chain length, and discussion of the methods used in simulating IR spectra (11 pages). Ordering information is given on any current masthead page.

Evaluating the Assumptions Underlying Force Field Development and Application Using Free Energy Conformational Maps for Nucleosides

David A. Pearlman† and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446. *Received February 7, 1991*

Abstract: A recently developed method for applying constraints to internal coordinates during molecular dynamics is utilized to test a fundamental assumption in empirically based force field development: that it is acceptable to derive parameters by optimizing agreement between experiment and potential (rather than free) energy predictions. This new method is used in conjunction with free energy perturbation techniques to generate free energy contour maps in the γ-χ torsion plane for adenosine and deoxyadenosine nuclesides. These maps are compared with analogous potential energy maps, generated by using standard minimization techniques. While qualitatively similar, the maps display significant quantitative differences, calling into question the general validity of the fundamental assumption. The dependencies of the maps on the level of atomic representation (all atom versus united atom) and on the solvent model used (explicit versus implicit) are also examined. These comparisons make clear the large effects that even small changes in the representation can have. The calculations presented herein suggest that we can now advance to a new level of sophistication in our ability to incorporate experimental solvation and free energy results into force field development.

Introduction

The past two and a half decades have seen the rapid growth and evolution of a new field of chemistry: molecular modeling.[1,2] In molecular modeling, an empirically derived penalty ("energy") function is used in conjunction with various theoretical methods to predict molecular behavior.

Originally, this field was limited to molecular mechanics (minimization, systematic search) studies of small systems.[3] But subsequent leaps in computer power, coupled with algorithmic developments, have made possible more complex simulations. Among these recent advances are molecular dynamics (MD),[4,5] from which a dynamic picture of molecular behavior over time can be extracted, free energy perturbation (FEP),[6,7] which allows determination of the free energy difference between two states, and joint molecular dynamics-NMR or -X-ray refinement,[8,9] which provide a means to more efficient and reliable refinement of experimental data. These methods can be practically applied, in principal, to a broad range of systems, from small molecules to moderate-sized proteins.[10] Attempts have even been made to model systems as complex as membranes and ion channels.[11]

One feature common to all the aforementioned techniques is a reliance on a classical, primarily empirically based, potential energy force field (CPFF). Such a force field relates any system configuration to an analytically defined potential energy.[12] An example of a frequently used macromolecular CPFF that typifies those in use is[13,14]

$$V_{\text{Total}}(\mathbf{x}) = \sum_{\text{bonds}} K_r(r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left\{ \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right] + q_i q_j / \epsilon R_{ij} \right\} + \sum_{\text{h-bonds}} \left\{ \frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right\} \quad (1)$$

V_{Total} is the potential energy of the system, K_r and r_{eq} are the bond stretching constant and the equilibrium bond distance, K_θ and θ_{eq}

† Current address: Vertex Pharmaceuticals Incorporated, 40 Allston Street, Cambridge, MA 02139-4211.

(1) Kollman, P. *Annu. Rev. Phys. Chem.* **1987**, *38*, 303-316.
(2) Ramachandran, G. N.; Sasisekharan, V. *Adv. Prot. Chem.* **1968**, *23*, 283.
(3) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; ACS Monograph Series; American Chemical Society: Washington, D.C., 1982; Vol. 177.
(4) McCammon, J. A.; Gelin, B.; Karplus, M. *Nature* **1976**, *267*, 585.
(5) van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992-1027.
(6) Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. *Faraday Symp. Chem. Soc.* **1982**, *17*, 55-67.
(7) Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281.
(8) Kaptein, R.; Zuiderweg, E. P. R.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. F. *J. Mol. Biol.* **1985**, *182*, 179-182.
(9) Brunger, A. T.; Kuriyan, J.; Karplus, M. *Science* **1987**, *235*, 458-460.
(10) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
(11) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: New York, 1987.
(12) Rasmussen, K. In *Potential Energy Functions in Conformational Analysis*; Berthier, G., Dewar, M. J. S., Fischer, H., Fukui, K., Hall, G. G., Minze, J., Jaffe, H. H., Jortner, J., Kutzelnigg, W., Ruedenberg, K., Eds.; *Lecture Notes in Chemistry*; Springer-Verlag: New York, 1984; Vol. 37.

0002-7863/91/1513-7167$02.50/0

are the bond angle stretching constant and the equilibrium bond angle, V_n, n, and γ are the torsional force constant, the periodicity of the torsional term, and the phase angle, A_{ij} and B_{ij} are the nonbond (Lennard-Jones) repulsion and attraction coefficients, R_{ij} is the interatomic distance between atoms i and j, q_i and q_j are the atomic partial charges on atoms i and j, and ϵ is the effective dielectric constant.

The reliability of such a force field depends on the validity of both the analytic form and the parametrization of the equation. Considerable discussion and justification of the analytic form has appeared in the literature,[15] and for most of the terms there is consensus—for macromolecular applications—that the form shown in eq 1 represents a reasonable compromise between accuracy and practicability.

Parametrization is another matter entirely. If we are to obtain meaningful results with molecular modeling methods, the adjustable parameters of the CPFF must be carefully derived and tested. Clearly, these parameters should be derived, whenever possible, from experimental data. But from what experimental data and how? For many systems, there is a vast body of such data available, some of it conflicting. More importantly, once the individual parameters have been derived, the entire parameter set must be "massaged" in order to obtain a self-consistent force field that correctly reproduces experimentally observed conformational preferences.[12] This need arises for a variety of reasons, including the fact that the analytic form (1) is only an approximation (albeit a good one, in many cases) to the true quantum nature of the system, the need to scale all terms to a common zero-point energy, and experimental errors.

The difficulty in deriving a reliable set of macromolecular parameters is underscored by the large number of differing force fields that have been published for the proteins and nucleic acids,[13,14,16-25] and the fact that none of these has emerged as the undisputed choice among researchers.

The focus of this work is to examine one of the fundamental assumptions made in deriving all these force fields: that if we aim to optimize agreement between the potential energy predictions of the CPFF and experimentally observed conformational behavior, we will obtain a reliable energy function. To this end, we utilize a recently developed method[26] for calculating relative free energy differences between conformations. Contour maps relating free energy to conformation for a couple of common nucleosides are generated and compared to the analogous potential energy maps.

Our simulations also enable us to evaluate two force field simplifications frequently employed to reduce the computer time required for simulations. The first of these, the "united atom model",[14,21,27] allows molecules to be represented by fewer explicit particles. In the second, the solvent environment of a system is modeled implicitly through use of some nonunity effective dielectric function ϵ (cf. eq 1).

Background

As noted, in the "massaging" phase of force field development, all of the aforementioned force fields shared a common criterion: to optimize agreement between the potential energy predictions of the force field and experimentally observed conformational behavior. Unfortunately, the actual behavior of molecules is governed by the nature of the *free* energy hyperspace, not that for the potential energy. Thus, in any case where there is a nonnegligible entropic contribution to the free energy, it is quite possible that predictions based on the CPFF alone will differ significantly from those that would be made if the free energy surface were known. Until recently, however, there existed no practical means for reliably calculating relative free energies for systems of any substantial size.

Fortunately, techniques have recently been elaborated that, given advances in computing power, allow the *free* energy differences between molecular conformations to be determined.[28] Briefly, these methods, termed free energy perturbation (FEP), rely on the following connection formula, which is readily derivable from classical statistical mechanics:

$$\Delta G = G_B - G_A = -RT \ln \langle e^{-\Delta H/RT} \rangle_A \quad (2)$$

$$\Delta H = H_B(\mathbf{r}) - H_A(\mathbf{r}) \quad (3)$$

ΔG is the free energy difference between two states, B and A. ΔH is the difference between the Hamiltonians representative of states B and A, both evaluated with the same set of coordinates $\mathbf{r}$. R is the gas constant and T is the temperature. In practice, ΔH is replaced by ΔV, the difference in the CPFF's for states B and A. $\langle\rangle_A$ means we need the ensemble average of the quantity in these brackets, generated from the potential function representative of state A. The requisite ensemble is generated by using either a molecular dynamics or Monte Carlo[29] simulation. It is the compute-intensive nature of these procedures that hindered the utilization of eq 2 until recently. In practice, determining an accurate value can be problematic, but techniques have been developed that help in this regard.[30]

The "states" defined as A and B in eq 2 can be either compositional or conformational. If they are defined as compositional, then we can use this equation to determine free energy differences for systems that differ in their molecular makeup. Most applications of FEP (e.g. calculation of solvation free energies,[6,31] determination of relative binding constants,[32] etc.) have been of this type.

Alternatively, eq 2 can be used to evaluate free energy differences for molecular systems differing in conformation: so-called potential of mean force (PMF) calculations.[28] In this type of simulation, for example, states A and B might refer to two different rotamers of a particular torsion angle. PMF applications using eq 2 have been much less common than compositional FEP simulations. The primary reason for this has been that in order to use eq 2, we must generate a representative Boltzmann ensemble with the appropriate degrees of freedom representative of conformers A and B fixed at prechosen values. While "umbrella" techniques[33] have been applied for this purpose,[34-37] wherein restraining potential terms are applied during dynamics and the

(13) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230–252.
(14) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
(15) Cornell, W. D.; Howard, A. E.; Kollman, P. A. *Current Opinion in Structural Biology*; in press, 1991.
(16) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(17) Nilsson, L.; Karplus, M. *J. Comput. Chem.* **1986**, *7*, 591–616.
(18) Momany, F. A.; McGuire, R. F.; Burgers, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361–2381.
(19) Nemethy, G.; Pottle, M.; Scheraga, H. *J. Phys. Chem.* **1983**, *87*, 1883.
(20) Sippl, M.; Nemethy, G.; Scheraga, H. *J. Phys. Chem.* **1983**, *88*, 6231.
(21) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
(22) Hagler, A.; Huber, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319.
(23) Olson, W. K.; Sussman, J. L. *J. Am. Chem. Soc.* **1982**, *104*, 278–286.
(24) Levitt, M. *J. Mol. Biol.* **1983**, *168*, 595.
(25) Clark, M.; Cramer, R. D., III, Opdenbosch, N. V. *J. Comput. Chem.* **1989**, *10*, 982–1012.
(26) Tobias, D. J.; Brooks, C. L. *J. Chem. Phys.* **1988**, *89*, 5115–5127.
(27) Scheraga, H. *Adv. Phys. Org. Chem.* **1968**, *6*, 103.
(28) Beveridge, D. L.; Di Capua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492.
(29) Jorgensen, W. L. *J. Phys. Chem.* **1983**, *87*, 5304–5314.
(30) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1989**, *90*, 2460–2470.
(31) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564–568.
(32) Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108*, 3830.
(33) Valleau, J. P.; Torrie, G. M. In *Modern Theoretical Chemistry*; Berne, B., Ed.; Plenum: New York, 1977; Vol. 5, p 137.
(34) Northrup, S. H.; Pear, M. R.; Lee, C. Y.; McCammon, J. A.; Karplus, M. *Proc. Natl. Acad. Scid. USA* **1982**, *79*, 4035.
(35) Mezei, M.; Mehrotra, P. K.; Beveridge, D. L. *J. Am. Chem. Soc.* **1985**, *107*, 2239.
(36) Bell, C. D.; Harvey, S. C. *J. Phys. Chem.* **1986**, *90*, 6595–6597.
(37) Anderson, A. G.; Hermans, J. *Proteins* **1988**, *3*, 262–265.

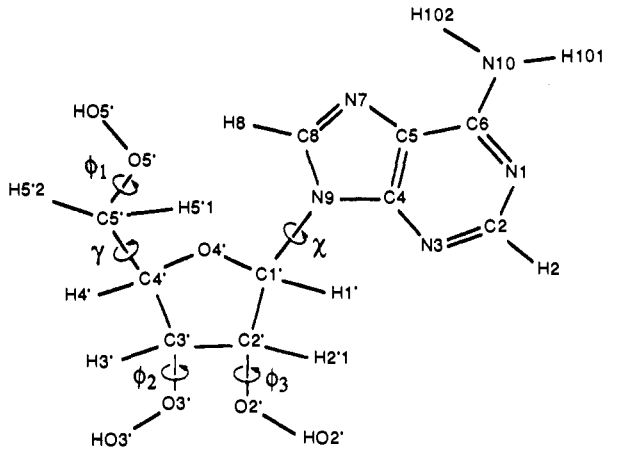


Figure 1. Schematic representation of an adenosine nucleoside. The five significant free rotatble torsional degrees of freedom are denoted γ, χ, ϕ_1, ϕ_2, and ϕ_3, where ϕ_1–ϕ_3 refer to torsions that specify only the orientation of a hydroxyl group. The sugar can also undergo conformational interconversion. The sugar conformation can be described by *P*, the phase angle of pseudorotation. A deoxyadenosine nucleoside differs from the molecule shown in the O2′–HO2′ group is replaced by H2′2, and there is no corresponding ϕ_3 torsion. The all-atom (AA) representation is shown.

ensemble so-generated is then corrected for its non-Boltzmann character, determining the required correction is often problematic.

An alternative to the umbrella methods is the use of holonomic constraints, a series of simultaneous equations that can be solved iteratively to fix chosen internal degrees of freedom without otherwise adversely affecting the dynamical motion of the system. The application of such constraints was first demonstrated for macromolecular systems by Ryckaert et al.,[38] who applied them to constrain bond lengths during dynamics. This is the well-known SHAKE algorithm, and has been widely used to allow longer integration time steps in molecular dynamics.

More recently, Tobias and Brooks[26] have derived equations that allow holonomic constraints to be applied to other "natural" internal coordinates: valence and dihedral angles. This is exciting, since these internal coordinates, particularly dihedrals, account for the bulk of morphological conformational variability in many molecules.

Briefly, in these constraint methods, a series of equations for atomic positions are formulated that depend both on the positions from an unconstrained MD step, and on terms that include undetermined multipliers. A set of multipliers is sought that will affect the forces necessary to concurrently keep all the chosen internal coordinates fixed at specified values. The actual solution is determined by iterating cyclically through the constraints, adjusting each multiplier in turn to effect the relevant constraint. This continues until all constraints are met to the desired degree of accuracy. We have implemented these equations in a new version of the AMBER/GIBBS program.[39]

With this new version of GIBBS, we are now in a position to reliably compare potential energy and free energy predictions, and thus determine the appropriateness of the former in force field development. To this end, we chose two nucleoside model systems: deoxyadenosine and (ribo)adenosine (Figure 1). These nucleosides present excellent model systems both because they are neither trivially simple nor prohibitively complex, and because they are biologically significant, and as a result their conformational behavior has been well-characterized experimentally.[40,41]

The nucleoside conformation can be satisfactorily described by only five or six parameters: the torsion angles denoted as γ, χ, and ϕ_1–ϕ_3 in Figure 1, plus the phase angle of pseudorotation *P*, which ranges from 0 to 360° and describes the conformation of the furanose sugar.[42] In fact, a morphological description of these nucleoside systems can be generated with even fewer parameters: the ϕ torsions only describe the rotations of the hydroxyl groups, which do not affect the gross conformation.

And in any scheme to sample the available conformational spaces for the nucleosides, things are simpler still. As is well-known, the range of observable sugar conformations is usually limited to a subrange of the conformations theoretically available (namely, −20 to 200° in *P*).[43] This subrange is roughly bounded by minima at the so-called C3′-endo and C2′-endo conformations characteristic of canonical A-form and B-form DNA, respectively. The two minima in this subrange are separated by a barrier of no more than a few kilocalories/mole. The consequences of this are that (A) for minimization, we generally need only start with the C3′-endo, C2′-endo, and an intermediate conformation (O4′-endo) to reasonably sample available conformations, and (B) in molecular dynamics simulations, interconversion between the two sugar minima is quite facile for nucleosides, allowing us to generate, in a practical amount of time, an ensemble that can be used to reliably evaluate the average in eq 2 for a PMF simulation where the remaining torsion variables γ and χ are constrained.[44]

We chose to examine nucleoside energetics as a function of both of the torsion angles γ and χ, for both minimized potential energies and free energies. Energies were evaluated at 18° increments in both γ and χ (resulting in a 21 × 21 grid) from the Weiner et al. all-atom force field.[13] For some simulations (denoted UAOH), the hydroxyl groups were represented by a united-atom oxygen, where the charge of the hydrogen was incorporated into that of the oxygen, and van der Waals interactions with the hydrogen were set to zero. In one set of simulations (denoted UA), we used the "united-atom" force field of Weiner et al.[14] again with hydroxyl groups replaced by united-atom oxygens. In this united-atom force representation, all aliphatic hydrogens are collapsed into their attached heavy atoms. Except where noted, all simulations were carried out in a pseudovacuum, with a distance-dependent dielectric constant to model solvent effects. To produce the desired data, different protocols were required for minimization and FEP.

Minimization. At each grid point, three two-step minimizations were carried out. First, minimization was carried out with 10 kcal/mol harmonic restraints to maintain the appropriate values of γ, χ, and the sugar conformation, which was restrained to either the C3′-endo, C2′-endo, or O4′-endo conformation. Then the restraints on the sugar conformation were removed, and the structure reminimized to an rms energy gradient of 0.001 kcal/mol·Å. For the all-atom simulations, the hydroxyl groups were started either from arbitrary reasonable positions or considered explicitly, starting separately in each of the classic staggered positions (60, 180, and 300°, resulting in nine two-step minimizations per grid point), depending on the simulation.

FEP. Due to the high computer cost of generating an ensemble that yields a reliable value of the average in eq 2—even without explicit solvent—it was imperative that we minimize the number of simulations required to characterize the grid. The scheme we

(38) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(39) D. A. Pearlman developed the version of Gibbs used for the calculations described herein, starting with AMBER/Gibbs version 3.0. Among the many features new to this version are the ability to perform PMF simulations for any chosen internal coordinate(s) in any molecule, the ability to automatically generate for free energy maps in the manner described in Figure 2, dynamically modified windows, and residue-based periodic imaging. AMBER(UCSF) 3.0 is by U. C. Singh, P. K. Weiner, J. W. Caldwell, and P. A. Kollman (Department of Pharmaceutical Chemistry, University of California, San Francisco, 1986).

(40) Davies, D. B. *Prog. Nucl. Magn. Reson. Spectrosc.* **1978**, *12*, 135–225.

(41) de Leeuw, H. P. M.; Haasnoot, C. A. G.; Altona, C. *Isr. J. Chem.* **1980**, *20*, 108–126.

(42) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.

(43) Pearlman, D. A.; Kim, S.-H. *J. Biomol. Struct. Dyn.* **1985**, *3*, 99–125.

(44) For a free nucleoside subject to stochastic disturbances (such as collisions with explicit solvent or periodic removal of center of mass motion), the mean time to sugar repuckering for both adenosine and deoxyadenosine (using MD at a temperature of 300 K) is between 3 and 4 ps. For the simulations used to generate the FEP maps described herein, the overall mean time to repuckering ranged from 9 to 17 ps. Forty picoseconds of sampling were carried out at every "window".

Table I. Average Error/"Window" for FEP Maps

system	atomic representation[b]	average error/"window"[a] (kcal/mol) χ	γ
adenosine	AA	0.77	0.35
adenosine	UAOH	0.18	0.07
adenosine	UA	0.17	0.05
deoxyadenosine	AA	0.71	0.37
deoxyadenosine	UAOH	0.07	0.04

[a] Errors are estimated from values of ζ, the uncorrected free energy difference between the first (0°) and last (360°) value for each row/column in the maps. If there were no errors, these values would be the same. A "window" is the interval used for each incremental FEP calculation (e.g. each of the simulations represented by arrows in Figure 1). In these cases, each window has a width of 18°. The average error/window is calculated from $(N \text{ av err}^2)^{1/2} = \langle\zeta\rangle$, where N is the number of intervals across each row/column (here 20). [b] Abbreviations: AA = all-atom; UAOH = united-atom representation for hydroxyl groups only; UA = united-atom representation for hydroxyl groups and al aliphatic carbons. See text.

used is described in Figure 2. Note that we calculate several grid points from each MD simulation point. For the map-generating FEP calculations, at each point we used an equilibration phase of 10 ps, and a data collection phase of 30 ps (a series of MD simulations totaling 4.4 ns for each map). Constant temperature molecular dynamics were run with a target temperature of 300 K. Center of mass motion was not removed. In all simulations, a time step of 2 ps was used.

To determine the effects of inclusion of solvent molecules in the model, we also carried out a couple of simulations where the nucleoside was "immersed" in a periodic system of explicitly included water. The TIP3P water model[45] was used, with a central "box" of 125 waters surrounding the solute and a 7-Å nonbonded cutoff. These simulations were carried out at constant temperature (300 K) and pressure (1 atm). The dynamically modified windows protocol was employed,[30] with a target free energy change/window of 0.03 kcal/mol, and with equilibration and data-collection phases of 2 ps each per window. The PMF trajectory is determined as

$$G(\tau) = G(\tau_0) + \sum_{i=0}^{N-1} \Delta G(\tau = \tau_i \rightarrow \tau_{i+1}) \qquad (4)$$

where τ is the torsion angle of interest and the individual ΔG's are evaluated as given by eqs 2 and 3. By use of dynamically modified windows, the spacing between the values τ_i is continually modified during the run (based on the recent $\partial G/\partial\tau$ slope) to keep the individual values of ΔG approximately constant. This appears to improve sampling efficiency.[30] Because we carry out the FEP/PMF simulations through a complete 360° period, we can extract an excellent bound on the errors for these simulations: the difference between the free energy values at 0° and 360°. In the absence of errors, these values would necessarily be the same. Note that this error estimate is probably better than that provided in standard FEP simulations, where the error is derived from the spread among redundant determinations of the same value. As has been noted, this latter method often leads to misleadingly low error estimates.[46,47]

Results

Determination of a Reasonable and Amenable Model. In any systematic examination of conformational space, one attempts to balance the inherent costs of a more accurate/exact model with an acceptable level of accuracy for the results. In the current study, this amounts to choosing the solvent model and the atomic representation of the solute. For most of our studies we used a pseudovacuum model with a distance-dependent dielectric constant to represent solvent effects. (This is a very commonly used approximation and is discussed in more detail later.)

This leaves the question of what atomic model to use for the nucleoside solute. As noted above, there are three torsional degrees

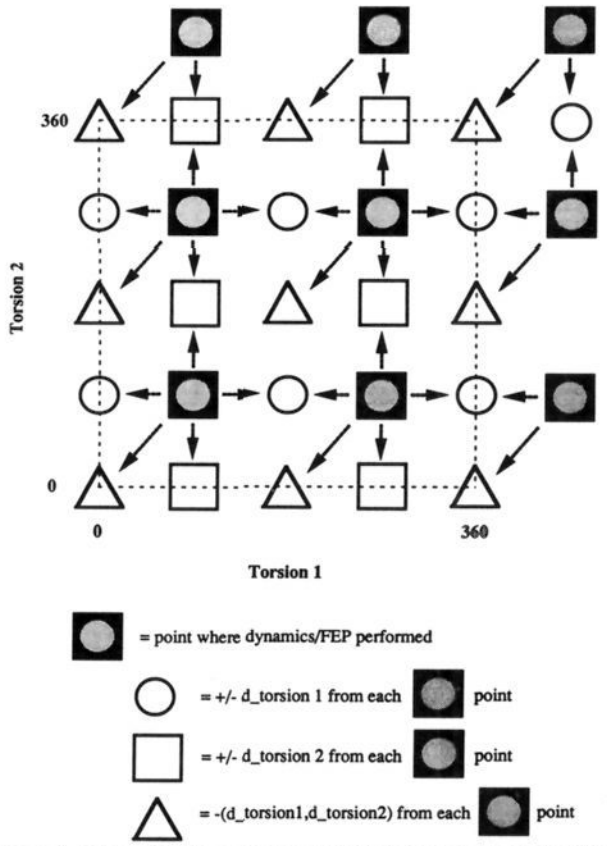


Figure 2. Schematic diagram representing the scheme used in calculating the free energy maps. This scheme was chosen to allow the map to be calculated in the near-minimum amount of computer time. For the map shown here, the size of each "window" (the distance between adjacent points in the horizontal and vertical directions) would be 90°. For the maps actually calculated, each "window" was 18°, with correspondingly more points (21) in both directions. "Torsion 1" corresponds to χ and "Torsion 2" corresponds to γ. As can be seen, at each "mother" point where molecular dynamics sampling is carried out, five different free energy differences are determined, and the "mother" points are chosen so that there is no redundancy in the free energies determined among them. This minimizes the amount of computation required to carry out the full calculation, while allowing the simulation to be performed in a general and automated manner. A final set of "mother" points outside the bounds of the map are chosen to fill in the missing "triangle" points at the north and east edges of the map (note that these outside-bounds calculations could be done in a slightly more efficient manner than illustrated here; the manner used here was chosen for consistency and convenience in a general implementation). After the individual free energy differences indicated have been determined, they are converted into a set of consistent energies for the entire map. Errors in the calculated free energies will result in differing values at 0° and 360° along each row/column. These are rationalized as follows: (1) Define two sets of points, one that corresponds to the "rows", and one that corresponds to the "columns". At the onset, these are the same. Linearly scale the points along each row, so that the 0° and 360° endpoints have the same value. Do the same for the columns. (2) Average the outer (0° and 360°) rows together. Do the same for the columns. (3) Scale all the rows so that the (0°,x) endpoint is coincident with the intersecting (x,0°) point in the first column. Do the same for all the columns with points in the first row. (4) Average the "row" and "column" values at each interaction. (5) Scale all the averaged values so that the minimum point in the map is assigned 0.0.

(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(46) Pearlman, D. A.; Kollman, P. A. In *Computer Simulation of Biomolecular Systems; Theoretical and Experimental Applications*; van Gunsteren, W. F.; Weiner, P. K., Eds.; Escom Science Publishers: The Netherlands, 1989; 101–119.
(47) van Gunsteren, W. F. In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*; van Gunsteren, W. F.; Weiner, P. K., Eds.; Escom Science Publishers: The Netherlands, 1989; 27–59.

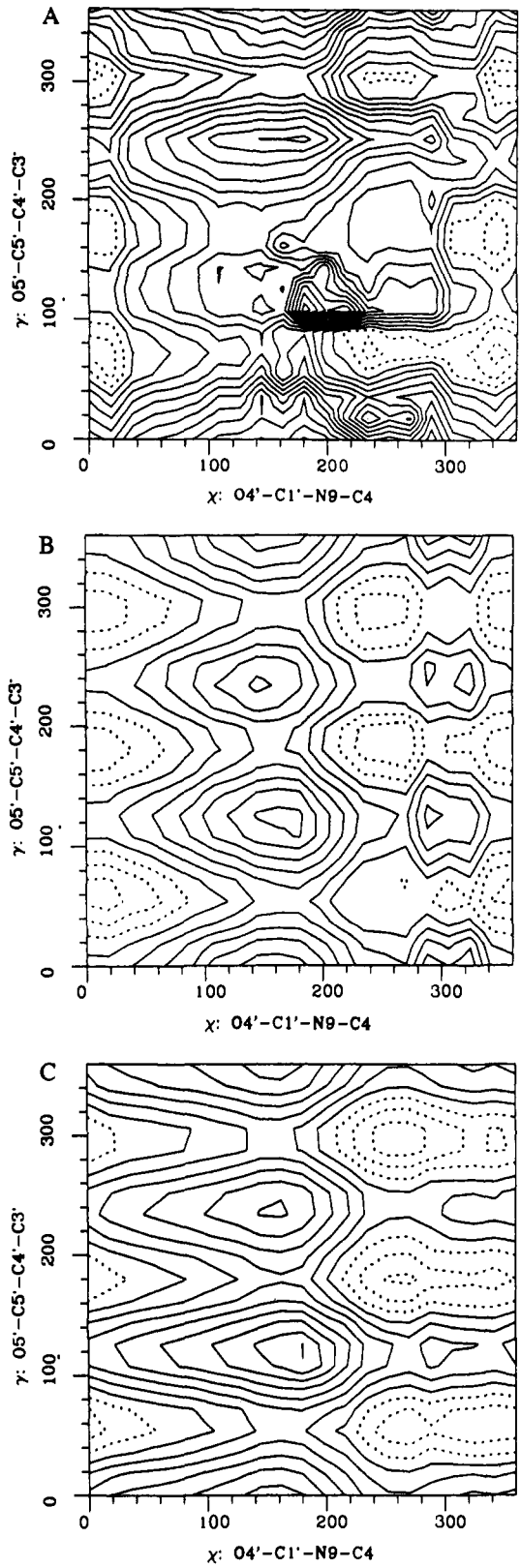


Figure 3. Free energy χ–γ maps for adenosine. Each map was calculated as described in Figure 2 and the text. Constant energy contours are plotted from 0.5 to 19.5 kcal/mol, in increments of 1.0 kcal/mol. The three lowest energy contours (0.5, 1.5, and 2.5 kcal/mol) are plotted with dashed lines. The energies in each map are scaled so the global minimum corresponds to 0.0 kcal/mol. (A) All-atom (AA) representation. (B) United-atom hydroxyl (UAOH) representation. (C) United-atom carbons plus united-atom hydroxyl (UA) representation.

of freedom (ϕ_1–ϕ_3) in the all-atom representation of adenosine (two for deoxyadenosine), which can be effectively eliminated if we use united-atom hydroxyl groups. A further reduction in

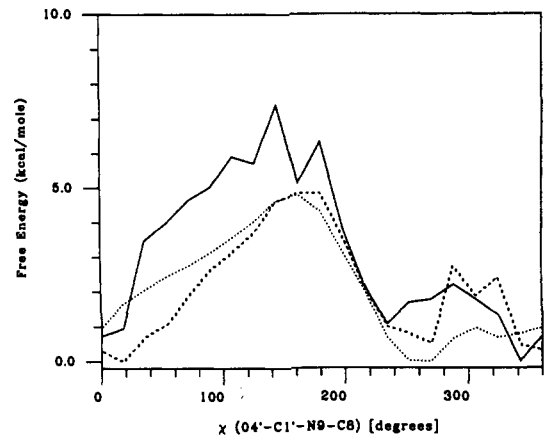


Figure 4. Profiles relating the free energy to the value of χ, derived from the maps in Figure 3. At each point is plotted the Boltzmann-weighted (300 K) energy average for all values of γ corresponding to the appropriate value of χ. Solid line, AA representation; dashed line, UAOH representation; dotted line, UA representation.

complexity can be effected by using united-atom representations of all aliphatic carbons. However, it has been observed that such modifications may significantly affect the free energy hyperspace.[36] The resulting free energy maps for all three models are shown in Figure 3. Figure 4 shows the energy vs angle profiles for χ, extracted from each of these contour plots. The estimated errors for the various simulations, based on the difference in free energy between 0° and 360° for each row (γ) or column (χ), are given in Table I.

As can be seen, the all-atom model results in by far the most complicated potential surface. But this model also results in by far the largest errors (Table I). This is a result of insufficient sampling of rotamers about the hydroxyl bonds; the large positive charge (0.3 *e*) on the hydroxyl hydrogen gets attracted to a nearby negative charge, and is subsequently "stuck" in this conformation for a long period of time. This is a result of the somewhat artifical pseudovacuum model being used; though we use a distance-dependent dielectric, this model does not allow the conformation-changing stochastic collisions and competing solvent–hydroxyl interactions of a full solvent model. To show that more complete sampling about the hydroxyls would result in smoother maps, we compared two minimized maps for all-atom deoxyadenosine. In the first, the initial positions of the hydroxyl groups were arbitrary. In the second, we considered the effects of these hydroxyl rotations on the energy by explicitly sampling at the three rotamers observed experimentally: 60°, 180°, and 300° (see above). The resulting maps are shown in Figure 5, and demonstrate that the map becomes considerably more smooth when we explicitly search the conformational space about the hydroxyls.

In contrast to the all-atom representation, the errors associated with the united-atom hydroxyl (UAOH) are acceptably small. Just as importantly, the free energy profile is in qualitatively reasonable agreement with experiment:[40,41] For χ (Figure 4), there are two major minima, the global minimum at around 0° ("anti") and a second minimum around 260° ("syn"), and these minima are separated by barriers of a few kcal/mol; for γ, there are three staggered minima at 60°, 180°, and 300°, separated by moderate barriers, as qualitatively expected (Figure 8).

The commonly used all united-atom (UA) approximation[14,21,27] also results in reasonable errors. However, the resulting χ versus energy profile compares more poorly with experiment; there is now only one significant minimum, corresponding to the syn conformer, plus a low-energy trough extending to the low anti region. This conflicts with the bimodal distribution of conformers observed experimentally, and implies that in the standard UA model, certain steric barriers have been underestimated, allowing for rotation about χ, which is too free in the 260–360° range. In summary, the UAOH model appears to provide the best compromise between accuracy and reasonable conformational sampling for this study.

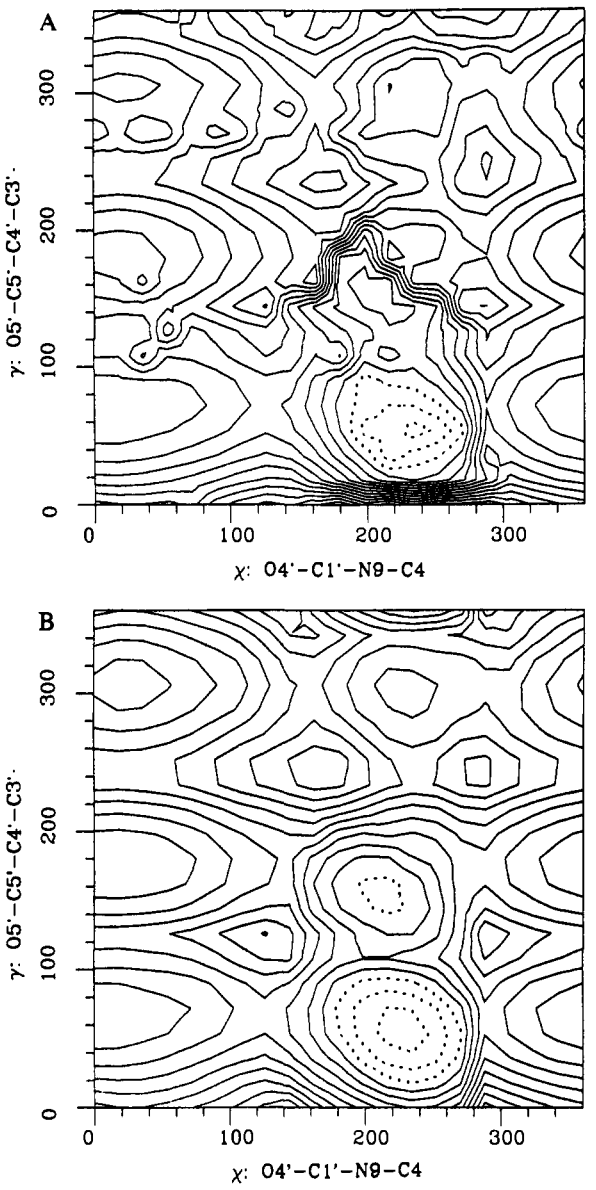


Figure 5. Minimized potential energy χ–γ maps for deoxyadenosine. Each map was calculated as described in the text. Contour levels are as described for Figure 3. (A) AA representation with hydoxyls started in arbitrary positions. (B) AA representation with systematic searching of the torsional spaces for the two hydroxyl groups (ϕ_1 and ϕ_2; cf. Figure 2). For (B), the energy at each point corresponds to the values of ϕ_i that result in the lowest energy for the given χ and γ.

Comparison of Free Energy and Minimization Results. The minimized potential energy and corresponding free energy maps for deoxyadenosine and adenosine are presented in Figures 6 and 7. As can be seen, at the qualitative level these maps are quite similar; all minima and maxima appear in both maps, in approximately the same locations. For γ, minima at the classical staggered positions (60°, 180°, and 300°)) are observed; for χ, two minima are seen, the global minimum at anti and a higher energy minimum at syn. Consistent with experimental observation,[40,41] the nature of the sugar (deoxyribo vs ribo) makes little *qualitative* difference.

There *are* significant differences between the adenosine and deoxyadenosine maps at the quantitative level, as the energy vs angle plots in Figures 8 and 9 make clear. Notably, the free energy well depth for the syn conformer is much deeper for adenosine than for deoxyadenosine; and the relative depth of the g^- (300°) minimum for γ is considerably higher for adenosine than deoxyadenosine.

More important, though, are the quantitative differences between the respective potential energy and free energy curves for

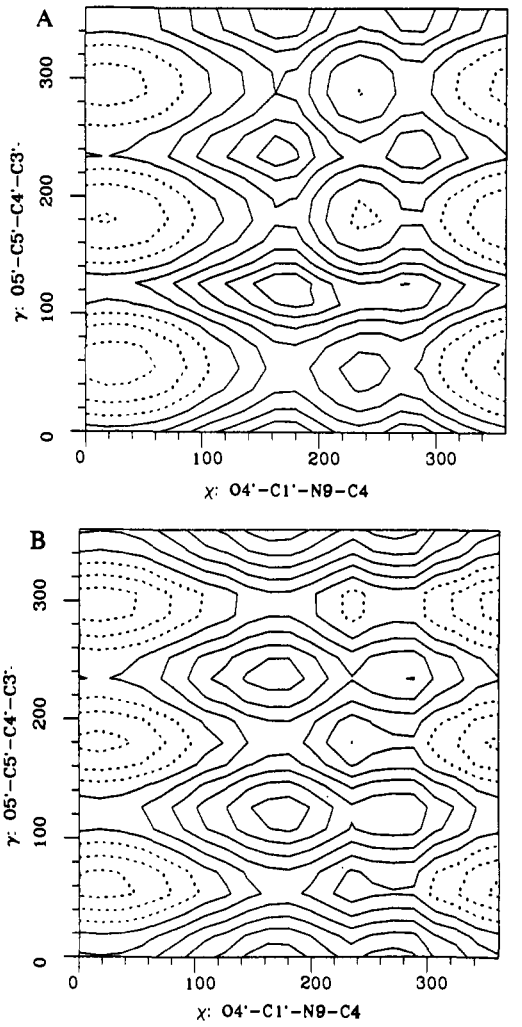


Figure 6. Minimized potential energy and free energy χ–γ maps for deoxyadenosine using the UAOH atomic representation. Contour levels are as described for Figure 3. (A) Minimized potential energy map. (B) Free energy map.

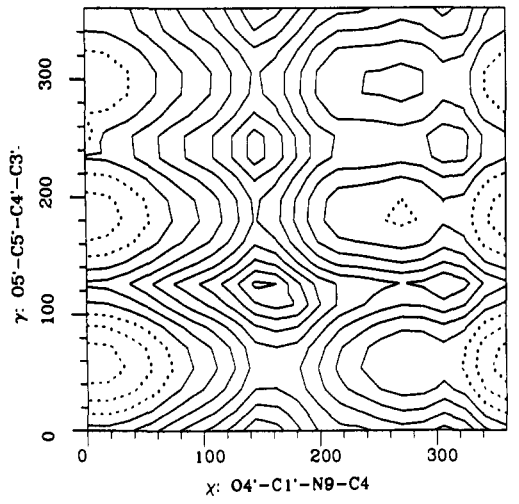


Figure 7. Minimized potential energy χ–γ map for adenosine using the UAOH atomic representation. The corresponding free energy map is Figure 3B. Contour levels are as described for Figure 3.

the same nucleoside. Of particular note are the relative differences in the depths and predicted populations for the γ minima (Figures 10 and 11; see also Table II). Experimentally, γ typically assumes either the g^+ (60°) or t (180°) conformation in nucleosides; the g^- (300°) conformation is observed much less frequently (Table II). This is reasonably consistent with the predictions of the

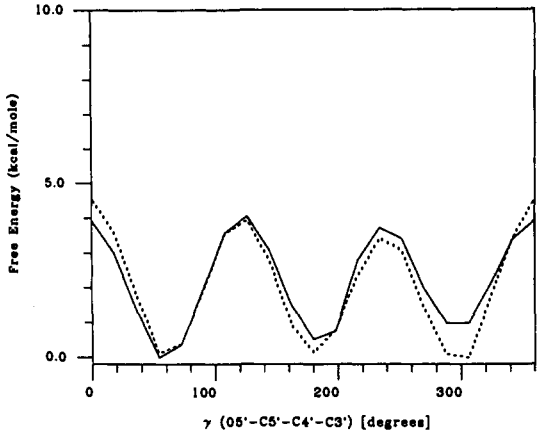


Figure 8. Profiles relating the free energy to the value of γ for adenosine and deoxyadenosine. These profiles correspond to the UAOH representation maps in Figures 3B and 6B. At each point is plotted the Boltzmann-weighted (300 K) energy average for all values of χ corresponding to the appropriate value of γ. Solid line, adenosine; dashed line, deoxyadenosine.

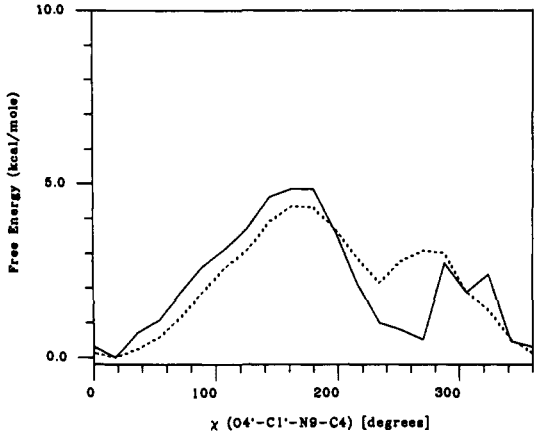


Figure 9. Profiles relating the free energy to the value of χ for adenosine and deoxyadenosine. These profiles correspond to the UAOH representation maps in Figures 3B and 6B. At each point is plotted the Boltzmann-weighted (300 K) energy average for all values of γ corresponding to the appropriate value of χ. Solid line, adenosine; dashed line, deoxyadenosine.

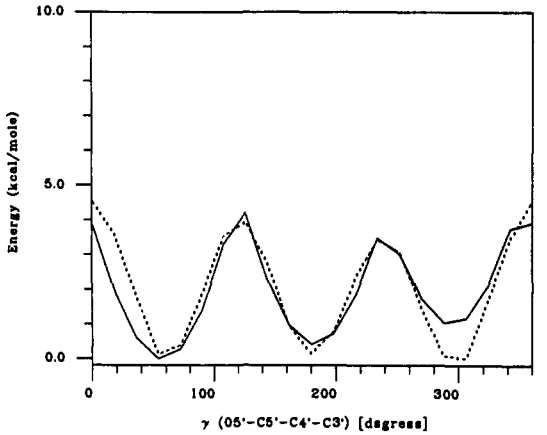


Figure 10. Profiles relating the minimized potential and free energies to the value of γ for deoxyadenosine. These profiles correspond to the UAOH representation maps in Figure 6A,B. At each point is plotted the Boltzmann-weighted (300 K) energy average for all values of χ corresponding to the appropriate value of γ. Solid line, minimized energy; dashed line, free energy.

potential energy calculations, where $E_{g^+} < E_t$, and E_{g^-} is more than 1 kcal/mol higher than E_{g^+}. But it is in direct contradiction of the free energy results for deoxyadenosine, where E_{g^-} is lower then E_{g^+} or E_t. Likewise, the free energy results for adenosine

Table II. Computed Energies and Populations for the γ Torsion Minima

	deoxyadenosine			adenosine		
	g^+	t	g^-	g^+	t	g^-
Calculated Energies[a]						
potential energy	0.00	0.45	1.10	0.00	1.03	1.93
free energy	0.10	0.11	0.00	0.00	0.44	0.78
normal modes analysis[b]	0.00	0.06	0.65	0.00	0.30	1.11
free energy (solvent) $\chi = 18$[c]	1.83	0.00	2.81	0.00	0.79	2.59
free energy (solvent) $\chi = 72$[c]	1.25	0.00	3.42			
Calculated Population Percentages[d]						
potential energy	58.3	29.7	12.0	76.4	18.1	5.5
free energy	30.0	27.4	42.6	48.9	28.0	23.1
normal modes analysis[b]	41.5	44.1	14.4	60.3	29.5	10.2
free energy (solvent) $\chi = 18$[c]	5.2	94.2	0.6	79.7	19.3	1.1
free energy (solvent) $\chi = 72$[c]	8.2	91.6	0.2			

	deoxyribopurines			ribopurines		
	g^+	t	g^-	g^+	t	g^-
Experimental Population Percentages						
x-ray[e]	20.0	80.0	0.0	66.7	20.0	13.3
NMR[f]	53	29	18	61	25	14

[a] All energies are relative to 0 for the global minimum for the simulation in which they were derived, and are in kcal/mol. "g^+", "t", and "g^-" refer to the minima centered around the values of 60°, 180°, and 300°, respectively. With the one noted exception, the values reported here are for pseudovacuum simulations using the UAOH (united-atom hydroxyl) model. [b] Normal mode analysis was carried out by minimizing structures starting with all possible combinations of the following conformational parameters: $\gamma = g^+, t, g^-$; χ = syn, anti; sugar = C3'-endo, C2'-endo, O4'-endo. The values reported here correspond to the lowest energies for structures with the relevant values of γ. [c] Energies for this simulation correspond to the all-atom explicit solvent simulation described in the text. In this explicit solvent case, a free energy profile was generated for γ, with χ fixed at 18° or 72°, as noted. The average errors per 18° interval (to be comparable to the values in Table I) were derived as described in the legend to Table I from the 0° → 360° hysteresis: dA, χ = 18; 0.04 kcal/mol; dA, χ = 72, 0.71 kcal/mol; rA, χ = 18, 0.18 kcal/mol. [d] For the potential and free energy maps, populations were determined by determining the Boltzmann factor corresponding to every point in each map, and binning them as $g^+ = 0 < \gamma \leq 120°$, $t = 120 < \gamma \leq 240°$, and $g^- = 240 < \gamma \leq 360°$. For the normal modes calculations, populations were determined from a Boltzmann summation over the 12 states that remained unique after minimization. For the explicit solvent free energy calculations, the populations were determined from a Boltzmann summation over the γ profile with either χ = 18° or χ = 72°, as appropriate. [e] Distributions based on the compendium of nucleoside crystal structures by de Leeuw et al.[41] There are 30 relevant purine ribonucleoside and 5 relevant purine deoxyribonucleoside structures in this compilation. [f] NMR distributions are from the approximate analysis of Davies[40] for anti-base-conformer purine nucleosides.

are in poorer agreement with experiment than are the potential energy results, though not as strikingly so as for deoxyadenosine.

By reference to Figures 10 and 11, it is clear that the discrepancy arises because the energy level of the g^- minimum has been lowered in the free energy simulations, relative to the potential energy ones. Physically, this reflects the entropic contribution due to the greater conformational freedom allowed the O5' hydroxyl group in the g^- conformation, where it points away from the base/sugar; in the g^+/t conformations it points toward these moieties. Since these simulations were run without explicit solvent, we do not see any of the mediating entropic effects due to solvent reordering that would be required for the g^- conformation in solvent.

We have seen that the potential energy results for the γ rotational profiles are in excellent agreement with experiment, while the more experimentally relevant free energy results are not. This disparity can be rationalized by examining the force field and its development. There are two dominant factors that determine the

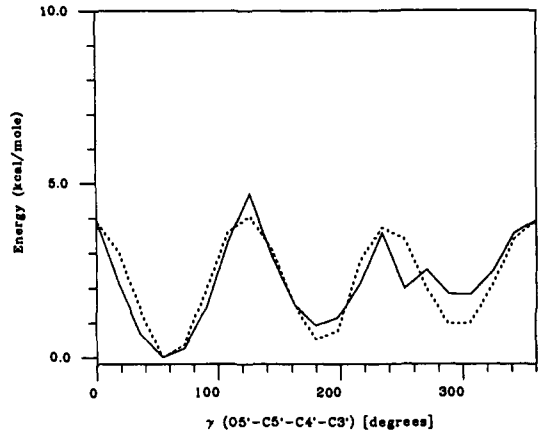


Figure 11. Profiles relating the minimized potential and free energies to the value of γ for adenosine. These profiles correspond to the UAOH representation maps in Figures 7 and 3B. At each point is plotted the Boltzmann-weighted (300 K) energy average for all values of χ corresponding to the appropriate value of γ. Solid line, minimized energy; dashed line, free energy.

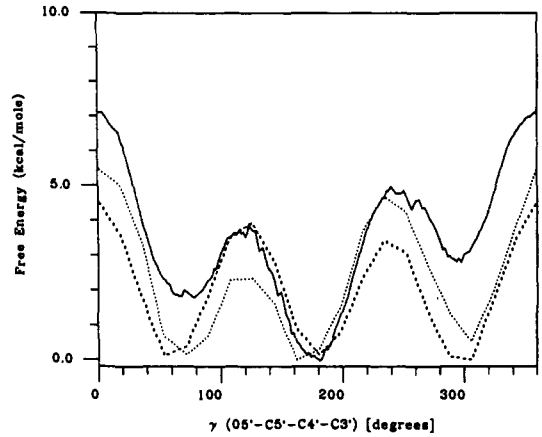


Figure 12. Profiles relating the free energy to the value of γ for deoxyadenosine and various atomic/solvent representations. These profiles correspond to a fixed value of $\chi = 18°$, and are arbitrarily scaled so that each is relative to a minimum of 0.0 kcal/mol. Solid line, AA and explicit solvent representation; dashed line, AA pseudovacuum representation; dotted line, UAOH pseudovacuum representation.

conformational preferences of γ: changes in the nonbonded interactions as rotation occurs, and contributions from the torsional terms (eq 1). For the torsions centered on the C4′–C5′ bond, a standard 3-fold torsional barrier (V_3) of magnitude 3 kcal/mol would lead to minima for γ at 60°, 180°, and 300°. But Marsh et al.[48] showed that torsions centered on a C–C bond with an electronegative atom on each carbon will prefer to have the electronegative substituents gauche rather than trans. Experimentally, this conformational preference has been clearly shown for the FCCF torsion of 1,2-difluoroethane.[49] The most compelling explanation of this effect emphasizes the role of bond–antibond interactions.[50]

This "gauche" effect can be included in molecular mechanics force fields by concurrently using both 2-fold (V_2) and 3-fold (V_3) torsional barriers for appropriate torsions (such as OCCO), which leads to preferential stabilization of the gauche over the trans conformation.[23,48] Weiner et al.[13,14] did this in the development of the nucleic acid force field used here, taking the magnitude of $V_2/2$ = 0.5 kcal/mol for OCCO directly from the study of Marsh et al.[48] without modification. The consequences of this choice were (A) reasonable potential energy vs sugar pucker profiles for deoxyribonucleosides (where such a V_2 term is critical in reproducing the observed preference for C2′-endo conformation over C3′-endo), and (B) the relative energy of the g^+ conformer of γ is more stable than t or g^- (in vacuo, with $\epsilon = r_{ij}$, as in Table I). Thus, this choice was reasonable. But no analysis was done to determine the influence of free energies (nor the roles of solvent effect or of dielectric model) on conformational preferences.

Given the large differences seen here between potential energy and free energy predictions (for relatively simple systems), we may be building in spurious biases when deriving our CPFF's solely from comparisons between potential energy predictions and experiments.

Normal Mode Analysis: A Compromise? In actuality, some free energy information *is* used in force field development, albeit indirectly. Normal mode analysis (NMA)[51,52] is often used to calculate the vibrational normal modes, and the force field parameters can be modified to improve the fit of these calculated normal modes to those determined from experimental spectra. Since the normal mode frequencies can also be related to thermodynamic quantities such as entropy and free energy via equations of classical statistical mechanics,[53,54] frequency matching implicitly includes such thermodynamic information. Unfortunately, normal mode analysis is only applicable to conformations that correspond to local minima, and assumes harmonic motion about these minimum energy conformations, which is unlikely to be an accurate assumption for any but the simplest molecules, not to mention the thorny question of whether one can even *find* an appropriate minimum. Nonetheless, NMA can often give an at least qualitatively better approximation to the free energy than minimization itself. And, for small-to-moderate sized systems, NMA is substantially computationally cheaper than free energy perturbation (as much as 100 times cheaper for small molecules).

In Table II, we have listed the results from NMA along with the results from energy minimization and FEP. As can be seen, the results from NMA are intermediate between the other two sets; NMA picks up some but not all of the stabilizing entropic contribution to the g^- conformer predicted by FEP.

Effects of the Solvent Model. Until now, the calculations we have described have relied on a pseudovacuum model, where we include the effects of the solvent environment implicitly through the distance-dependent effective dielectric constant, $\epsilon = r_{ij}$. This is a very commonly used model,[11] particularly for simulations involving large solutes (such as proteins), and for studies that attempt to systematically sample conformational space (such as those here). We know this model is only moderately good, at best. But better models, such as explicit inclusion of a periodic system of water molecules, are much more computationally expensive. For FEP simulations, this increased cost results not only from the increased number of nonbonded interactions that must be calculated for such a system but also from the increased amount of sampling that can be required to evaluate the ensemble average in eq 2, when we must allow both for the damping effects of the dynamics of the solute and for solvent relaxation.

These considerations made it impractical for us to determine complete free energy maps analogous to those described above but with explicit inclusion of solvent. However, it *was* computationally feasible to determine energy profiles for one torsion (i.e. γ *or* χ), holding the other torsion fixed. Comparing the resulting profiles to the analogous curves for the pseudovacuum model will allow us insights into the reliability of the latter, simpler model.

We carried out simulations where γ was varied from 0° to 360°, while χ was fixed at 18° for both deoxyadenosine and adenosine. This value of χ was chosen because it corresponds to the global minimum for the UAOH pseudovacuum case (Figure 4). (A

(48) Marsh, F. J.; Weiner, P.; Douglas, J. E.; Kollman, P. A.; Kenyon, G. L.; Gerlt, J. A. *J. Am. Chem. Soc.* **1980**, *102*, 1660–1665.
(49) Butcher, S. S.; Cohen, R. A.; Rounds, T. C. *J. Chem. Phys.* **1971**, *54*, 4123.
(50) Brunck, T. K.; Weinhold, F. *J. Am. Chem. Soc.* **1979**, *101*, 1700–1709.
(51) Lifson, S.; Warshel, A. *J. Chem. Phys.* **1968**, *49*, 5116.
(52) Kottalam, J.; Case, D. A. *Biopolymers* **1990**, *29*, 1409–1421.
(53) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1976.
(54) Hagler, A. T.; Stern, P. S.; Sharon, R.; Becker, J. M.; Naider, F. *J. Am. Chem. Soc.* **1979**, *101*, 6842–6852.

second simulation was carried out for deoxyadenosine with χ = 72°; the results are effectively the same as those described below; cf. Table II). The dynamically modified windows protocol[30] was used, where the free energy path is broken into segments, the width of each chosen to keep the free energy change per segment approximately constant (here 0.03 kcal/mol). For each segment, we used 2 ps of equilibration and 2 ps of data collection, with resulting total simulation times of 726–827 ps. An all-atom (AA) model was used in the simulations. Each simulation took around 25 cpu hours on a Cray YMP supercomputer.

The results of the explicit solvent (ES) run for deoxyadenosine (χ = 18°) are shown in Figure 12. Also plotted are the analogous profiles for the UAOH and all-atom pseudovacuum simulations (corresponding to a fixed value of χ = 18°). As can be seen, the ES results are significantly different from those from either other model; for the ES run, the trans conformer is most favorable, almost 2 kcal/mol lower in energy than the g^+ conformer, and almost 3 kcal/mol lower in energy than the g^- conformer.

While the ES results for deoxyadenosine differ from those of the pseudovacuum simulation, they still do not completely agree with the estimated experimental distributions (Table II). It is worth noting, though, that if one only considers the purine deoxyribonucleosides in the compilation of nucleoside crystal data,[41] four of the five relevant structures are in the $\gamma = t$ conformation, and only one adopts the $\gamma = g^+$ conformation, in acceptable agreement with the ES predictions here. It would be unwise to draw any far-reaching conclusions from this limited experimental data set, though, particularly in light of the conflicting NMR population estimates (Table II).

As noted, we used an all-atom model for the explicit solvent simulations. Stochastic collisions between the solvent and the hydroxyls, and the large number of potential electrostatic interactions between these moieties and solvent should help avoid the sampling problems inherent to the all-atom model in vacuo. The pseudovacuum all-atom results presented here are of questionable accuracy, for reasons discussed earlier. But we have included the all-atom curve in Figure 12, to demonstrate that the differences between the all-atom explicit solvent model and the UAOH pseudovacuum model are not primarily due to differences in the all-atom vs united-atom models.

For adenosine (Table II), the difference between the free energy predictions of the UAOH pseudovacuum and ES models is not nearly as striking as for deoxyadenosine, but the two models still lead to quantitatively different predictions. Surprisingly (and certainly somewhat fortuitously), in this case the minimized potential energy results lead to predictions that are quite close to those extracted from the free energies calculated with the ES model. In this case, the results using the ES model are in much better agreement with those from NMR (Table II).

At any rate, it is very clear that for free energy analysis the pseudovacuum models, particularly for deoxyadenosine, do not do a good job of reproducing the exact *quantitative* preferences of the ES models, although for some applications they are probably reasonably close.

Discussion

As we have shown, even for a relatively simple system like a nucleoside, entropy can play a significant role in stabilizing various conformers, making potential energy alone an unreliable criterion for deriving force field parameters. A similar conclusion was reached in an earlier, more limited study of the rotational barrier for *n*-butane using umbrella sampling.[36] Comparison of computed normal modes with experimental spectra has played some part in the development of most modern force fields, and this indirectly adds some free energy information to the fitting procedure, but most likely not enough. Hopefully, with the widespread proliferation of faster computers, the development of reliable FEP procedures and the concurrent improvement in other free energy techniques,[55,56] the next generation of force fields will more fully reflect free energy considerations.

We have also shown how small changes made in the force field can have significant effects on the character of the resultant potential surface. It is common practice, especially when simulating large systems such as proteins, to reduce the computational complexity of the problem by modifying the atomic representation. The *most modest* such change is typically to change from an all-atom model to a united-atom model. While there is nothing intrinsically wrong with such a simplification, and a carefully parameterized united atom model[57] will probably be better than a mediocre all-atom model, the results here emphasize the fact that *any* change in describing the system will affect the potential surface, often significantly. This should serve as a caveat to those adopting ad hoc simplifications to make their simulations more manageable.

On a similar note, we saw how even conformational variables that are typically considered to be of secondary importance (the hydroxyl rotations) can affect the energetics of the system. The presence of such variables, if their effects are not methodically explored, greatly increases the inherent error in any modeling simulation. In the nucleosides considered, there were only two to three such variables. One would imagine that in large systems, such as proteins and nucleic acids, where there are many more such variables (and no computationally feasible way of surveying them all), the resultant uncertainties in derived quantities will be considerably larger.

Finally, it was shown how the use of a more accurate, explicit solvent (ES) model during a FEP simulation leads to yet another strikingly different prediction. It is interesting that this ES model is in much better agreement with the limited relevant crystal data base than the pseudovacuum model (although it is in poorer agreement with the NMR population estimates). This is somewhat encouraging, since it implies that the existing potential model may be better than was implied by the less exact pseudovacuum simulation. At any rate, it certainly points out the importance of using an explicit solvent model when trying to derive quantitative free energy estimates.

Two other questions are raised by the results in Table II. First, why is it that for adenosine the populations predicted on the basis of potential and free energies agree qualitatively both with each other and with experiment, but for deoxyadenosine only the populations derived from potential energy agree with experiment? The empirical reason lies in the magnitudes of the potential energy differences between g^+ and t, and between g^+ and g^- for both molecules. In adenosine, these differences are 1.0 and 1.9 kcal/mol, respectively, while in deoxyadenosine they are 0.5 and 1.1 kcal/mol, respectively. Consideration of free energy effects in either molecule lowers t relative to g^+ by about 0.5 kcal/mol, and lowers g^- relative to g^+ by 1.1–1.2 kcal/mol. As discussed in the text, these results make qualitative sense because the O5′ atom is most constrained by nonbonded interactions in the g^+ state (where it hangs over the sugar ring), and successively less constrained by such interactions in the t and g^- states. At any rate, because the potential energy of the g^- state is lower for deoxyadenosine than adenosine and the free energy effects are comparable, the resulting relative free energy for deoxyadenosine appears spuriously low. One could "correct" this defect for deoxyadenosine by increasing the V_2 (OCCO) torsional potential to a somewhat larger value, one large enough to significantly disfavor g^-. But it is not obvious how to modify the force field to further stabilize g^+ over t. It is not clear *why* the relative potential energies of g^+, t, and g^- differ so much for adenosine and deoxyadenosine, nor to what extent this calculated difference is an artifact of the electrostatic model, particularly our treatment of the hydroxyls as united atoms.

We have determined for deoxyadenosine the energy versus rotation profile for χ, where the lowest energy γ and hydroxyl (for the AA model) orientations were chosen for each value of χ for both the UAOH and AA models (Figure 13). The AA

(55) Chiles, R. A.; Rossky, P. J. *J. Am. Chem. Soc.* **1984**, *106*, 6867.
(56) Pettitt, B. M.; Karplus, M. *J. Phys. Chem.* **1988**, *92*, 3994–3997.
(57) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638–6646.

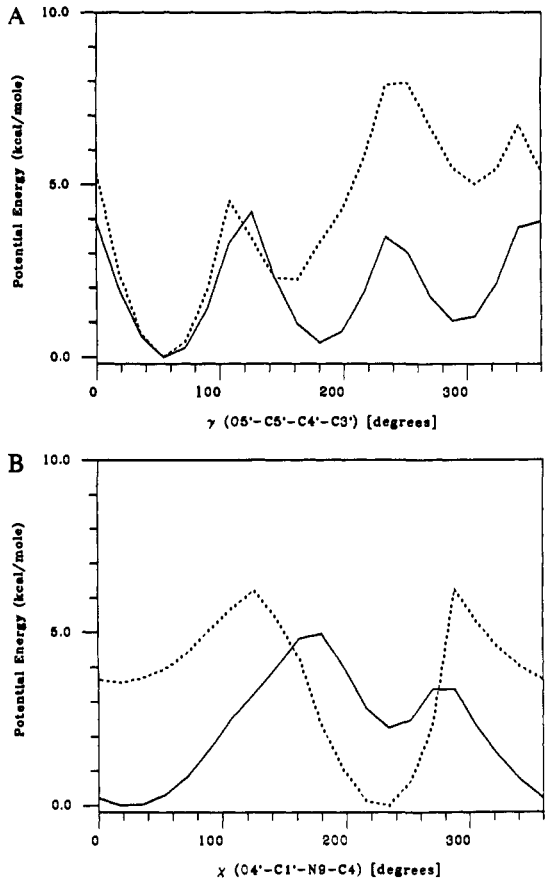


Figure 13. Profiles relating the minimized potential energy for deoxyadenosine to the values of γ and χ for two atomic representations. Solid lines correspond to the UAOH model (from the contour map in Figure 6A). Dashed lines correspond to the AA model where the torsional spaces of the hydroxyls were systematically searched (from the contour map in Figure 5B). For the γ profiles, each energy corresponds to a Boltzmann-weighted (300 K) average for all values of χ (and of ϕ_1 and ϕ_2 for the AA model) corresponding to the given value of γ. The profiles for χ are analogously defined in terms of Boltzmann-weighted averages of energy rotomers of γ (and ϕ_1, ϕ_2). (A) Potential energy versus γ. (B) Potential energy versus χ.

model predicts the χ = syn conformation to be 3.6 kcal/mol more stable than the anti conformation, whereas the UAOH model finds the anti conformation 2.2 kcal/mol more stable than syn. The UAOH model prediction is thus much more consistent with experiment.[40,41] We have performed analogous calculations to evaluate and compare the γ rotation profiles for the AA and UAOH models (Figure 13). In this case, both models lead to the same predicted order of relative energies: $g^+ < t < g^-$. But the relative energies of t and g^- are only 0.5 and 1.1 kcal/mol, respectively, for the UAOH model (Table II), while for the AAOH model they are 2.4 and 5.1 kcal/mol, respectively. In light of experimental observations, the latter values appear unreasonably large. In actuality, the "truth" may lie somewhere in between. As noted above, we would like to increase the relative potential energies for t and g^- from their UAOH values to effect in vacuo free energies qualitatively consistent with those found for adenosine and found experimentally. In effect then, we are suggesting that the UAOH model, although certainly more realistic than the AA model for the reasons we have described herein and capable of good qualitative predictions, may still lead to certain artifactual results at a quantitative level. In such a case, the all-atom explicit solvent model may be the most reliable.

The results in Table II also raise the question of why the relative free energies are so different in solution and in vacuo for *deoxy*-adenosine. The V_2 torsional force constant for OCCO stabilizes $\gamma = g^+$ and t over g^- by about 1 kcal/mol, and one could imagine that solvation of the O5′–HO5′ hydroxyl group (and perhaps N7 of the base) would be least favorable in the g^+ conformation, where the O5′–HO5′ group is over the sugar ring and closest to the base. But the results in Table II indicate solvent interactions are significantly more favorable for the t conformation than for g^-, and it is hard to rationalize why the t and g^- conformations about γ should not be equally well solvated. One could speculate that this is due to the t conformation allowing synergistic solvation of O5′–HO5′ and O4′ in the sugar, whereas in the g^- conformation the relative orientation of the O5′–HO5′ and O3′–HO3′ groups reduces the effective solvation of both. This hypothesis could be tested by studies of sugar analogues, e.g. cyclopentane and 3′-deoxyanalogues of furanose.

As important as the above observations are, perhaps the greatest significance of this paper is the affirmation that the software tools and computational capabilities are finally at hand to allow one to look at conformational free energies in fragments of relevance to the study of nucleic acids and proteins in solution.[34-37] The fundamental work on the solvent dependences of the conformations of *n*-butane and 1,2-dichloroethane[29,36,58] has been of great importance, and we feel this work is the next step, carrying such approaches to significantly more complex systems, where free energies as a function of two conformational variables in vacuo can be studied with implicit sampling of a third (in these studies sugar pucker). In solution, one can study, for example, the free energy as a function of γ with selected χ and again implicit sampling of sugar pucker. There are still important limitations: for example, the sampling of three ϕ angles involving the OH groups in vacuo, and the large computational burden of such conformational studies in solution. Nonetheless, we expect that as computer performance/price ratios continue to increase and longer MD sampling trajectories become routine, studies of conformational profiles, such as those reported here, will be greatly facilitated because more complete sampling of "implicit" degrees of freedom can be done.

From a pedagogical point of view, the dramatic differences in the γ preferences in the gas phase and in solution is a phenomenon that must be analyzed if we are to fully understand aqueous solvation effects in complex systems. We have already shown the qualitative importance of solvent effects for other nucleic acid systems, e.g. the B to Z equilibrium in DNA.[59] An understanding of solvent effects on conformational preferences in systems such as deoxyadenosine or adenosine will be more difficult to achieve than in earlier free energy studies of model compounds such as *n*-butane or 1,2-dichloroethane, since they cannot be simply rationalized by solvent exposure or relative molecular dipole moment. But we stress that the computational tools are now at hand to begin to develop such an understanding.

What role should studies of the type presented here play in force field development? It is clear that it is currently impractical to "fine tune" each conformational parameter by using as computationally intensive an approach as we have employed here. But what is exciting is that we have reached the point where we can directly compare selected NMR-determined conformational preferences with those derived from calculated free energies and use this information in the evaluation of force fields.

An alternative view of parameter development holds that we may need to use different parameters for varying applications and conditions, e.g. energy minimization (T = 0) vs molecular dynamics (T = 300), or with and without explicit solvent. We do not fully agree with this view. Obviously, parameters derived by optimizing the fit between potential energy predictions and experimental observations will not allow free energies to be calculated as accurately as parameters for which free energy calculations were explicitly considered during the derivation process. But it is not yet practical to derive relative free energies in solution with free energy calculations for every molecule of interest. One should aim to do so whenever possible, and as computer power increases, one can expect increasingly greater numbers of free energy calculations to be incorporated in the parametrization process.

(58) Rebertus, D. W.; Berne, B. J.; Chandler, D. *J. Chem. Phys.* **1979**, *70*, 3395–3400.
(59) Pearlman, D. A.; Kollman, P. A. *Biopolymers* **1990**, *29*, 1193–1209.

The assumption in the Weiner et al. force fields[13,14] has been that the use of a constant dielectric $\epsilon = 1$ is appropriate with full inclusion of solvent, and that a distance-dependent dielectric $\epsilon = r_{ij}$ is appropriate for implicitly mimicking solvent effects, with no other required changes in the parameters. Obviously, the use of $\epsilon = r_{ij}$ is a very rough approximation used to modulate long-range electrostatic interactions. But it appears to be a reasonable one, since local hydrogen-bonding effects are shown to be similar with both models. In our opinion, it is not appropriate or necessary to derive different force fields in different solvent conditions, as long as one (A) parametrizes to fit experimental data in a given solvent condition by carrying out the calibration free energy calculations in that solvent, and (B) represents the environmental effects with the reasonable accuracy allowed by commonly used liquid solvent models.

In total, it is clear that the choice of force fields model is an important one for quantitative work, and that as the focus of modeling work continues to become more quantitative, we must develop better, and better characterized, force fields by using tools such as those presented here.

Acknowledgment. We thank the NSF (DMB-87-14775) and the San Diego Supercomputing Center for supercomputing time. We also thank the NIH, the National Cancer Institute (CA-25644), and Merck Sharp and Dohme for research support. Finally, we thank reviewer one, whose detailed comments on an alternative philosophy of force field development led us to expand our discussion.

Registry No. Adenosine, 58-61-7; deoxyadenosine, 958-09-8.

π-Stacking and the Platinum-Catalyzed Asymmetric Hydroformylation Reaction: A Molecular Modeling Study

L. A. Castonguay,[†] A. K. Rappé,*[,†] and C. J. Casewit[‡]

Contribution from the Department of Chemistry, Colorado State University, Fort Collins, Colorado 80523, and Calleo Scientific Software Publishers, 1300 Miramont Drive, Fort Collins, Colorado 80524. *Received February 22, 1991*

Abstract: The importance of steric factors in determining the regioselectivity of a class of platinum(II) hydroformylation catalysts ($L_2Pt(CO)X$) has been theoretically examined by using an augmented Dreiding force field. This paper characterizes the differential stabilization of the initially formed platinum alkyl complexes resulting from styrene insertion into a Pt-H bond to form either the primary phenethyl product (leading to normal or linear aldehyde) or the secondary α-methyl styryl intermediate (leading to branched aldehyde). The α-methyl styryl products displayed a pronounced π-stacking interaction between the aryl ligand and an aromatic ring of the phosphine ligand, whereas this interaction was absent in the phenethyl intermediate. When pentene is substituted for styrene, the normal intermediate is favored over the branched intermediate.

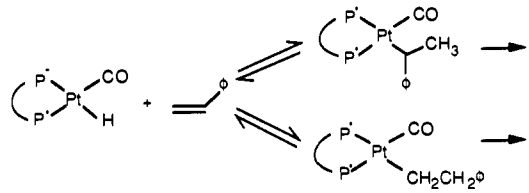
Introduction

The hydroformylation reaction or Oxo process produces aldehydes from the reaction of alkenes with synthesis gas ($CO + H_2$). The process was first discovered for a cobalt catalyst,[1] though all group 8, 9, and 10 transition metals show catalytic activity.[1,2] Catalytic asymmetric hydroformylation reactions have been used for the synthesis of optically pure organic products using rhodium and platinum catalysts with optically active ligands.[3,4,5]

Platinum complexes have not been used extensively in asymmetric hydroformylation due to the generally lower reaction rates in comparison with rhodium catalysts, and the tendency for the substrate olefin to undergo competitive hydrogenation.[1] In addition, relatively low branched to normal ratios (b/n ratio) have been observed in the hydroformylation of monosubstituted olefins. Recent work,[3] however, on $L_2PtCl_2/SnCl_2$ catalyst precursors and preformed $L_2PtClSnCl_3$ reports enantiomeric excesses at nearly an acceptable level ($\sim$80% ee). Modest increases in the obtained enantiomeric excess and increased regioselectivity would establish this reaction as being viable for asymmetric synthesis.

Successful asymmetric hydroformylation of olefins requires the control of regiochemistry (b/n ratio) as well as control of absolute stereochemistry (ee) of the branched product. Experimental investigations of asymmetric hydroformylation have focused on developing an understanding of the chiral induction step; factors affecting the regioselectivity have rarely been studied for asymmetric catalysis, and are the focus of the present work.

The intermediacy of a platinum alkyl complex in the Pt(II)-catalyzed hydroformylation reaction has been established by Scrivanti and co-workers.[6] The alkyl is formed from the insertion of an olefin into a Pt-H bond. Olefin insertion/β-elimination reaction pairs are in general reversible,[1] establishing the pair of competing equilibria prior to the rate-limiting steps of each pathway, as shown in eq 1 for a Pt complex.



(1)

In the case of asymmetric olefin hydrogenation, those factors favoring the major isomer in the set of equilibria prior to the rate-determining step ultimately adversely affect the rate-deter-

[†] Colorado State University.
[‡] Calleo Scientific Software Publishers.

(1) Parshall, G. W. *Homogeneous Catalysis: The Applications and Chemistry of Catalysis by Soluble Transition Metal Complexes*; Wiley: New York, 1980. Collman, J. P.; Hegedus, L. S.; Norton, J. R.; Finke, R. G. *Principles and Applications of Organotransition Metal Chemistry*; University Science Books: Mill Valley, CA, 1987.
(2) Pruett, R. L. *Adv. Organomet. Chem.* **1979**, *17*, 1-60. Gates, B. C.; Katzer, J. R.; Schuit, G. C. A. *Chemistry of Catalytic Processes*; McGraw-Hill: New York, 1979. Pines, H. *The Chemistry of Catalytic Hydrocarbon Conversions*; Academic Press: New York, 1981.
(3) Stille, J. K.; Parrinello, G. *J. Mol. Catal.* **1983**, *21*, 203. Parrinello, G.; Deschenaux, R.; Stille, J. K. *J. Org. Chem.* **1986**, *51*, 4189. Parrinello, G.; Stille, J. K. *J. Am. Chem. Soc.* **1987**, *109*, 7122. Stille, J. K.; Su, H.; Brechot, P.; Parrinello, G.; Hegedus, L. S. *Organometallics* **1991**, *10*, 1183.
(4) Consiglio, G.; Pino, P. *Top. Curr. Chem.* **1982**, *105*, 77-123. Consiglio, G.; Morandini, R.; Scalone, M.; Pino, P. *J. Organomet. Chem.* **1985**, *279*, 193.
(5) Kawabata, V.; Suzuki, T. M.; Ogata, I. *Chem. Lett.* **1978**, 361.
(6) Scrivanti, A.; Paganelli, S.; Matteoli, U.; Botteghi, C. *J. Organomet. Chem.* **1990**, *385*, 439 and references therein.

0002-7863/91/1513-7177$02.50/0