A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins

Scott J. Weiner, Peter A. Kollman,* David A. Case,† U. Chandra Singh, Caterina Ghio,‡ Giuliano Alagona,‡ Salvatore Profeta, Jr.,§ and Paul Weiner⊥

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143. *Received April 28, 1983*

Abstract: We present the development of a force field for simulation of nucleic acids and proteins. Our approach began by obtaining equilibrium bond lengths and angles from microwave, neutron diffraction, and prior molecular mechanical calculations, torsional constants from microwave, NMR, and molecular mechanical studies, nonbonded parameters from crystal packing calculations, and atomic charges from the fit of a partial charge model to electrostatic potentials calculated by ab initio quantum mechanical theory. The parameters were then refined with molecular mechanical studies on the structures and energies of model compounds. For nucleic acids, we focused on methyl ethyl ether, tetrahydrofuran, deoxyadenosine, dimethyl phosphate, 9-methylguanine–1-methylcytosine hydrogen-bonded complex, 9-methyladenine–1-methylthymine hydrogen-bonded complex, and 1,3-dimethyluracil base-stacked dimer. Bond, angle, torsional, nonbonded, and hydrogen-bond parameters were varied to optimize the agreement between calculated and experimental values for sugar pucker energies and structures, vibrational frequencies of dimethyl phosphate and tetrahydrofuran, and energies for base pairing and base stacking. For proteins, we focused on Φ,Ψ maps of glycyl and alanyl dipeptides, hydrogen-bonding interactions involving the various protein polar groups, and energy refinement calculations on insulin. Unlike the models for hydrogen bonding involving nitrogen and oxygen electron donors, an adequate description of sulfur hydrogen bonding required explicit inclusion of lone pairs.

There are two fundamental problems in simulating the structural and energetic properties of molecules: the first is how to choose an analytical been placed $E(\vec{R})$ which correctly describes the energy of the system in terms of its $3N$ degrees of freedom. The second is how the simulation can search or span conformational space ($\vec{R}$) in order to answer questions posed by the scientist interested in the properties of the system.

For complex systems, solution to the first problem are an essential first step in attacking the second problem, and thus, considerable effort has been placed in developing analytical functions that are simple enough to allow one to simulate the properties of complex molecules yet accurate enough to obtain meaningful estimates for structures and energies.

In the case of the structures and thermodynamic stabilities of saturated hydrocarbons in inert solvents or the gas phase, the first problem has been essentially solved by molecular mechanics approaches of Allinger,[1] Ermer and Lifson,[2] and their co-workers. However, for polar and ionic molecules in condensed phases, unsolved questions remain as to the best form of the analytical function $E(\vec{R})$. In the area of proteins and peptides, seminal work has come from the Scheraga[3] and Lifson[4] schools. The Scheraga group has used both crystal packing (intermolecular) and conformational properties of peptides to arrive at force fields ECEPP, UNECEPP, and EPEN for modeling structural and thermodynamic properties of peptides and proteins. Levitt, using the energy refinement software developed in the Lifson group, has proposed a force field for proteins based on calculations on lysozyme,[5] and Gelin and Karplus have adapted this software along with many parameters from the Scheraga studies to do molecular dynamics simulations of proteins.[6] Dauber and Hagler[7] have clearly demonstrated the usefulness of crystal packing energies and structures in force-field development. Hermans et al.[8] have taken another approach, which combines the Scheraga nonbonded parameters with quadratic stretching and bending functions for use with X-ray data or in a stand-alone mode to refine protein structures.

In the area of nucleic acids, the work of Sasisekharan[9] and Olson and Flory[10] was pioneering in the development of force fields, but significant contributions have been made as well by Rein et al.[11] and Pullman and Pullman.[12] Levitt has adapted his protein force field to nucleic acids and has carried out some important molecular mechanics and dynamics simulations on DNA fragments.[13,14]

Our approach has been to use the powerful cartesian-coordinate energy refinement of Lifson and Warshel[15] and to develop empirical force fields within this context. Our original parameter set was the first published nucleic acid force field in which complete geometry optimization of all atomic degrees of freedom could be carried out.[16] Our related force field for proteins[17] was similar but contained only modest modifications of the parameters used by Gelin and Karplus.[6] The most important changes concerned the explicit inclusion of H-bonding hydrogens and the use of partial charges taken from Mulliken populations of ab initio calculations.

Although many of the results of our simulations of proteins and nucleic acids with these force fields were encouraging, there were a few places where it was clear that improvements could be made.[17,18] The areas in most need of refinement involved the

† University of California, Davis, Department of Chemistry, Davis, CA 95616.
‡ Istituto di Chimica Quantistica, Consiglio Nazionale delle Ricerche, Pisa.
§ Louisiana State University.
⊥ Squibb and Co.

(1) Allinger, N. *J. Am. Chem. Soc.* **1977**, *99*, 8127.
(2) Ermer, O.; Lifson, S. *J. Am. Chem. Soc.* **1973**, *95*, 4121.
(3) Momany, F.; McGuire, R.; Burgess, A.; Scheraga, H. *J. Phys. Chem.* **1975**, *79*, 2361.
(4) Warshel, A.; Levitt, M.; Lifson, S. *J. Mol. Spectrosc.* **1970**, *33*, 84.
(5) Levitt, M. *J. Mol. Biol.* **1974**, *82*, 93.
(6) Gelin, B.; Karplus, M. *Biochemistry* **1979**, *18*, 1256.
(7) Dauber, P.; Hagler, A. *Acc. Chem. Res.* **1980**, *13*, 105.
(8) Hermans, J.; Ferro, D.; McQueen, J.; Wei, S. In "Environmental Effects on Molecular Structure and Properties"; Pullman, B., Ed.; Reidel: Dordrecht, Holland, 1976.
(9) Sasisekharan, V. In "Conformation of Biological Molecules and Polymers"; Bergman, E., Pullman, B., Eds.; Jerusalem, 1973. Lakshiminarayanan, A.; Sasisekharan, V. *Biopolymers* **1969**, *8*, 475, 489, 505.
(10) Olson, W.; Flory, P. *Biopolymers* **1972**, *11*, 25.
(11) Rein, R.; Goel, N.; Fukida, N.; Pollack, M.; Claverie, P. *Ann. N. Y. Acad. Sci.* **1969**, *153*, 805–814. Ornstein, R.; Rein, R. *Biopolymers* **1979**, *18*, 2821–2847.
(12) Pullman, B.; Pullman, A. *Prog. Nucleic Acid Res. Mol. Biol.* **1969**, *9*, 327.
(13) Levitt, M. *Proc. Nat. Acad. Sci. U.S.A.* **1978**, *75*, 640.
(14) Levitt, M., presented in part at the Cold Spring Harbor Symposium on DNA Structure, 1982.
(15) Lifson, S.; Warshel, A. *J. Chem. Phys.* **1968**, *49*, 5116.
(16) Kollman, P.; Weiner, P.; Dearing, A. *Biopolymers* **1981**, *20*, 2583.
(17) Blaney, J.; Weiner, P.; Dearing, A.; Kollman, P.; Jorgensen, E.; Oatley, S.; Burridge, J.; Blake, C. *J. Am. Chem. Soc.* **1982**, *104*, 6424. Wipff, G.; Dearing, A.; Weiner, P.; Blaney, J.; Kollman, P. *Ibid.* **1983**, *105*, 997.
(18) Singh, U.C.; Kollman, P. *J. Comput., Chem.*, in press.

0002-7863/84/1506-0765$01.50/0

nonbonded (Lennard-Jones) and electrostatic parameters. In view of the apparent power of our general approach for determining partial charges for complex molecules based on analysis of quantum mechanical electrostatic potentials,[18] it seemed a propitious time to develop a second generation force field. Thus, in this paper, we present the development and, in Tables XIV–XIX and Figures 3 and 4, the results of a force field for proteins and nucleic acids.

General Perspective

The basic equation for the force field is the same as that used earlier[16,17] with the addition of a weak 10–12 hydrogen bond term between hydrogen-bonding hydrogens and H-bond acceptor atoms (eq 1). In the previous force field, the 10–12 coefficients C and

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

D were set equal to zero for *hydrogen atom–hydrogen acceptor* interactions, following the results of Hagler et al.[19] This new force field contains a 10–12 function for two reasons. First, for strong H-bonds, it is clear that some repulsive term is required to prevent the occurrence of unrealistically short H-bonds[17] during energy refinement. Second, such a 10–12 function allows one to "fine tune" the H-bond distances and energies to desired values.

The bond stretching and bending functions are quadratic, which allows an adequate description of the structure and energies for relatively unstrained proteins and nucleic acids. A Fourier series approach to the torsional energy (i.e., more than one value of n may be used per dihedral angle in eq 1) allows rather accurate simulation of conformational preferences in simple and complex molecules. For computational speed, a 6–12 function is used for the nonbonded parameters even though a 6-exponential is likely to be a better simple functional form.[20] Hagler et al.[19] compared 6–12 and 6–9 nonbonded functions in crystal packing calculations and found neither to be clearly superior. As long as the interatomic distance is not well below the sum of the van der Waals radii, the 6–12 form should be adequate.

We retain the atom-centered monopole approach to the electrostatic energies (with the exception of sulfur, where lone pairs are also included). This approach appears to do a satisfactory job in simulating molecular electrostatic interaction energies, provided the charges are chosen in a reasonable fashion. We feel that the fit of the potential charge model to quantum mechanically calculated electrostatic potentials is a superior method for determining the point charges.

We use a distance-dependent dielectric, $\epsilon = R_{ij}$, for the electrostatic energies, although we demonstrate in hydrogen-bonding cases of model systems that results with a constant dielectric constant, $\epsilon = 1$, are very similar to those found with $\epsilon = R_{ij}$. A rationale for using a distance-dependent dielectric constant is that it mimics the polarization effect in attractive interactions, with closer interactions weighted more heavily. Second, it helps compensate for the lack of explicit solvation by implicitly damping longer range charge interactions more than shorter range ones. There is empirical and computational support for such a model,[17] given that solvent (water) is not explicitly included in the calculation. However, when water is explicitly included, a constant dielectric constant is probably more appropriate, and, as noted below, this can be used with the same set of charges and at most small modifications in the 10–12 H-bond parameters.

In the "united-atom" force field presented here, we include all atoms explicitly with the exception of hydrogens bonded to carbon. We should stress that this is merely for computational efficiency in simulations of large proteins and nucleic acids. Below we present all-atom simulations on a sugar-puckering model of nucleosides and on the Φ,Ψ maps of alanyl and glycyl dipeptides. In both cases, the united-atom representation gives results quite similar to the all-atom model. In a previous study,[17] we used a united-atom approximation for prealbumin but included the aromatic hydrogens of thyroxine explicitly, in order to correctly reproduce the Φ_1,Φ_2 conformational energies of thyroxine (these energies are strongly influenced by H···I nonbonded interactions). Thus, there will be cases where a hybrid force field is appropriate, and, as noted below, this is straightforward to implement.

In the development of force-field parameters, we used the following approach: we began with an initial set of parameters and then carried out simulations on a number of model systems, relevant to proteins and nucleic acids, to test these parameters or to determine some from scratch. One of the frustrating aspects of force-field development is the dependence of the final results on "the pathway" or choice of model systems. It is thus incumbent on the developer to elucidate his pathway as clearly as possible, to ensure that further work need not start from scratch. This methodology is carried out below by use of the AMBER molecular mechanics program.[21]

Force-Field Development

Atom Types. The basis of a force field is the choice of atom types; i.e., the selection of atoms which are enough alike, both chemically and physically, to be treated identically in the molecular mechanics refinement. In the case of a quantum mechanical calculation, one needs only a single atom "type" per atom; i.e., only the number of electrons is relevant. The decisions on atom types are inevitably compromises between possessing the most accurate representation of many molecules and having a manageable number of types. We list the types and their characteristics in Table I, so only a single comment is in order. The sp^3 atom types are fairly typical, but we have included more sp^2 types than earlier force fields to ensure increased geometrical precision for such ring systems as purines, pyrimidines, indoles, and imidazoles.

Sources of Parameters

Nonbonded Parameters. The most difficult set of parameters to derive a priori are the nonbonded ones. We used as our starting point for sp^2 atoms the 6–12 and 6–9 parameters, derived by Hagler et al.[19] from a fit of the lattice energies and crystal structures in amides. The significant difference between the 6–9 and 6–12 values of R^* (van der Waals minimum) and ϵ^* (van der Waals well depth) for a given atom caused us not to take these parameters directly. For example, the carbonyl carbon van der Waals radius *increased* from $R^* = 1.81$ Å, $\epsilon = 0.184$ kcal/mol in the 6–9 force field to $R^* = 2.175$ Å, $\epsilon = 0.039$ kcal/mol in the 6–12, whereas the aliphatic hydrogen *decreased* from $R^* = 1.77$ Å, $\epsilon = 0.0025$ kcal/mol in the 6–9 potential to $R^* = 1.375$ Å, $\epsilon = 0.038$ kcal/mol in the 6–12. In the 6–9 potential, oxygen and carbon had nearly the same size, with nitrogen 0.4 Å larger, whereas in the 6–12, the sizes varied in a smooth fashion from $R^* = 1.6$ (oxygen), 1.95 (nitrogen), and 2.2 Å (carbon).

To avoid these inconsistencies, we began with the Hagler et al.[19] 6–12 oxygen parameters since oxygen (of C, N, and O) has the most direct contact with neighboring molecules in the amides. This led us to take $R^* = 1.6$ Å and $\epsilon = 0.20$ kcal/mol for oxygen. We expect nitrogen to have a larger R^* than oxygen, in the range of 0.1–0.2 Å larger, given the standard van der Waals radii of the atoms determined by observed atom–atom contacts in crystals.[22,23] The well depth of nitrogen was consistently 0.04–0.06 kcal/mol less than oxygen in the Hagler et al. study, leading us to settle on the compromise parameters $R^* = 1.75$ Å, $\epsilon = 0.16$ kcal/mol. We then estimated the parameters for sp^2 carbons in an analogous fashion, obtaining $R^* = 1.85$ Å, $\epsilon = 0.12$ kcal/mol.

(19) Hagler, A.; Euler, E.; Lifson, S. *J. Aam. Chem. Soc.* **1974**, *96*, 5319.
(20) Margenau, H.; Kestner, N. "The Theory of Intermolecular Forces"; Pergammon Press: Oxford, 1970.
(21) Weiner, P.; Kollman, P. *J. Comput. Chem.* **1981**, *2*, 287.
(22) Pauling, L. "The Nature of the Chemical Bond", 3rd ed.; New York, 1960.
(23) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.

Table I. List of Atom Types

atom	type
carbons	
united[a]	
C2	sp^3 carbon with two hydrogens
C3	sp^3 carbon with three hydrogens
CD	sp^2 aromatic carbon in six-membered ring with one hydrogen
CE[c]	sp^2 aromatic carbon in five-membered ring between two nitrogens with one hydrogen (in purines)
CF	sp^2 aromatic carbon in five-membered ring next to a nitrogen without a hydrogen (e.g., $C_\delta-N_\epsilon=C_\epsilon$ in histidine)
CG	sp^2 aromatic carbon in five membered ring next to a N–H (e.g., $C_\delta-N_\epsilon=C_\epsilon$ in histidine)
CH	sp^3 carbon with one hydrogen
CI	sp^2 carbon in six-membered ring of purines between two "NC" nitrogens
CJ	sp^2 carbon in pyrimidines at positions 5 and 6 (more pure double bond than aromatic) with one hydrogen
CP[c]	sp^2 aromatic carbon in five-membered ring between two nitrogens with one hydrogen (in His)
all atom[b]	
C	sp^2 carbonyl carbon and aromatic carbon with hydroxyl substituent in tyrosine
C*	sp^2 aromatic carbon in five-membered ring with one substituent (e.g., CE_γ in Trp)
CA	sp^2 aromatic carbon in six-membered ring with one substituent
CB[d]	sp^2 aromatic carbon at junction between five- and six-membered rings (e.g., CE_δ in Trp, C4 and C5 in purines)
CC	sp^2 aromatic carbon in five-membered ring with one substituent and next to a nitrogen group (e.g., C_γ in His)
CM	sp^2 same as CJ but one substituent
CN[d]	sp^2 aromatic junction carbon in between five- and six-membered rings (e.g., C_ϵ in Trp)
CT	sp^3 carbon with four explicit substituents
nitrogens	
NC	sp^2 nitrogen in six-membered ring with lone pairs (e.g., N3 in adenine)
NA	sp^2 nitrogen in five-membered ring with hydrogen attached (e.g., protonated His)
NB	sp^2 nitrogen in five-membered ring with lone pairs (e.g., N7 in purines)
N*	sp^2 nitrogen in purines and pyrimidines with alkyl group attached (N9 in purines, N1 in pyrimidines)
N	sp^2 nitrogen in amide groups
N2	sp^2 nitrogen in base NH_2 groups and arginine NH_2
N3	sp^3 nitrogen with four substituents (e.g., Lys N_ζ)
NT	sp^3 nitrogen with three substituents (e.g., unprotonated amines)
oxygens	
O	carbonyl oxygen
O2	carboxyl and phosphate nonbonded oxygens
OS	ether and ester oxygens
OH	alcohol oxygens
hydrogens	
H3	hydrogens of lysine and arginine (positively charged)
H2	amino hydrogens from NH_2 in purines and pyrimidines
HC	explicit hydrogen attached to carbon
H	amide and imino hydrogens
HO	hydrogen on hydroxyl or water oxygen
HS	hydrogen attached to sulfur
sulfurs	
S	sulfurs in disulfide linkages and mentioning
SH	sulfur in cystine
phosphorus	
P	phosphorus in phosphate groups
lone pair	
LP	lone pairs

[a] United-atom carbons with implicit inclusion of hydrogens. [b] Nonhydrogen containing carbons. [c] Structural differences in the internal angles of the five-membered rings are the reason why these atoms, which appear in the same environment by definition, are assigned different atom types. [d] Neutron diffraction studies on tryptophan show that $\vartheta_{eq}(C_\epsilon',C_\delta',C_\delta) = 116.2°$ while $\vartheta_{eq}(C_\zeta,C_\epsilon,C_\delta') = 122.7°$. Due to this structural variation, we opted for two atom types at the junction carbons in Trp.

Since all of these values are close to 0.2 Å larger than the "standard" van der Waals contact radii from crystal packing data,[22,23] we also selected "larger" values for both phosphorus (R^* = 2.10 Å, ϵ = 0.20 kcal/mol) and sulfur (R^* = 2.00 Å, ϵ = 0.20 kcal/mol) to be consistent within this framework. These P and S values are similar to those found in MM2,[1] although a 6-exponential is used there.

For aliphatic CH, CH_2, and CH_3 groups (atom types CH, C2, and C3) there are two papers in the literature that suggest appropriate van der Waals 6–12 parameters for these extended atoms. Dunfield et al.[24] have determined values based on crystal packing calculations of hydrocarbons, and Jorgensen[25] has calculated the parameters from Monte Carlo liquid simulations of ethers and alcohols. These two parameter sets are very similar, suggesting appropriate values to use for united atoms. However, in our simulations of the conformational profile of methyl ethyl ether, *n*-butane, and deoxyadenosine (and in our earlier study of base-paired dinucleoside phosphates[16]), we found that the use of van der Waals parameters with R^* as large as Jorgensen's or Dunfield's (for example, for a CH group R^* = 2.385 Å and ϵ = 0.049 kcal/mol) gave a significantly poorer representation of *intramolecular* energies and structures than did smaller values.

With methyl ethyl ether, *n*-butane, and deoxyadenosine as model systems (described in detail below) we settled on compromise values of R^* = 2.00 Å, ϵ = 0.15 kcal/mol, for C3, R^* = 1.925 Å, ϵ = 0.12 kcal/mol, for C2, R^* = 1.85 Å, ϵ = 0.09 kcal/mol, for CH, and R^* = 1.8 Å, ϵ = 0.06 kcal/mol, for CT (sp^3 carbon without hydrogens). This last value for CT was a compromise between our expectation that CT should be smaller than CH but larger than an sp^2 carbon. In view of our expectation that the sp^3 atoms should be somewhat larger and less polarizable than sp^2, we used R^* = 1.85 Å, ϵ = 0.12 kcal/mol, for sp^3 nitrogen and R^* = 1.65 Å, ϵ = 0.15 kcal/mol, for sp^3 oxygens in alcohols and ethers. The aliphatic hydrogen parameter (R^* = 1.37 Å, ϵ = 0.038 kcal/mol) was taken from the Hagler et al. study.[19] We used a significantly smaller (R^* = 1.00 Å, ϵ = 0.020 kcal/mol) value for potentially H-bonding hydrogens (N–H, O–H, S–H), in view of the fact that these atoms have significant parts of their density shifted to the heteroatom to which they are attached. Our lone-pair van der Waal parameters (used only for sulfur) come directly from MM2.[1]

In this force field, for interactions involving hydrogen-bonding hydrogens and heteroatoms, we replace the 6–12 parameters with 10–12 parameters of well depth 0.5 kcal/mol except in one case noted below. The retention of the 6–12 parameters in these cases would lead to much too long H-bond distances while using no H-bond parameter (as done by Hagler et al.[18]) leads to H-bonds, in some cases, which are too short.[17]

(24) Dunfield, L.; Burgess, A.; Scheraga, H. *J. Phys. Chem.* **1978**, *82*, 2609.
(25) Jorgensen, W. *J. Am. Chem. Soc.* **1981**, *103*, 335.

In the original Hagler et al. study,[19] all the nonbonded H···X (hydrogen bound to N or O; X = any atom) van der Waals parameters were taken to be the geometric mean of the H and X parameters. Since their study found that this approach gave good H-bond distances, they used $A = 0$ and $B = 0$ for these specific hydrogen-nonbonded parameters. However, such an approach can lead to artifacts in which H···C distances become arbitrarily short. In our previous calculations,[16,17] we had set 6–12 nonbonded parameters for *H-atom–H-bond acceptor* atoms equal to zero and used the geometrical mean values $A_{ij} = (A_iA_j)^{1/2}$ and $B_{ij} = (B_iB_j)^{1/2}$ for all other heteroatomic interactions involving these hydrogens. Here we employ the same approach, except for *H-atom–H-bond acceptor* where a 10–12 function is used.

Electrostatic Parameters. We have used quantum mechanical calculations of the electrostatic potential to derive charges for atoms in salient molecules, as described in detail elsewhere.[18] This method uses quantum mechanically calculated electrostatic potentials to numerically fit atomic charge models. While we feel that this is a superior method for determining such charges, the process is still subject to three uncertainties. First, while the charges may depend upon the conformation of the molecule used, it is impractical to possess a separate set of charges for every conformation. Second, it is only practical to do quantum mechanical calculations for fragments of polymers and then to "patch" these together. Finally, the charges will differ depending on the basis set chosen. Elsewhere,[18] we have analyzed the error due to the first problem by carrying out calculations on C3′ endo and C2′ endo conformations of a deoxyribose model. The polar group charges derived by a fit to the potential in a C2′ endo conformation were qualitatively similar to those found with C3′ endo (within 5–10%). We have analyzed the second problem by comparing the charges derived by fitting the electrostatic potential of 1-methylcytosine, 1-aminodeoxyribose, and dimethyl phosphate and "averaging" the charges at the linking atoms with the charges derived from a fit of the electrostatic potential for cytosine 3′-phosphate. Not only is the agreement between the derived charges generally quite good (within 0–10%), but the charges for atoms in the linker regions (C1′ and O3′ in the sugar) are similar (0.500 and −0.535 on the basis of the truncated models and 0.547 and −0.514 for cytosine 3′-phosphate). The final problem, the basis set dependence of the charges, is one that is crucial to deal with. We have followed the approach of Cox and Williams[26] (described below) and have checked the H-bond energies derived with these charges for consistency with experiment and/or accurate ab initio calculations. The use a 10–12 H-bond function, with a well depth of 0.5 kcal/mol, enabled us to "fine tune" hydrogen distances by varying the repulsive R^{-12} H-bond coefficient. In two cases, a small change in the point charges was necessary to ensure accurate H-bond energies and geometries (see below).

Bond Length and Bond Angle Parameters. We took the parameters for equilibrium bond length, r_{eq}, and bond angle, ϑ_{eq}, from microwave and X-ray data on appropriate compounds. For example, ϑ_{eq}(C3–C2–C3) came from the microwave structure of propane,[27] this being the most appropriate source for it. We made efforts to select the highest quality data on a reasonable reference compound, rather than less accurate data on a particular molecule that might more closely resemble the fragment considered.

We were able to find suitable values for all r_{eq} parameters in the experimental literature. Many of the K_r came from normal-mode calculations, in which the K_r values were varied to give the best fit to experimental frequencies of tetrahydrofuran, dimethyl phosphate, *N*-methylacetamide, methanol, methanethiol, dimethyl sulfide, dimethyl disulfide, and benzene. We used a linear interpolation model for the remaining bond stretching force constants involving partial double bonded sp^2 atom–sp^2 atom (described in the next paragraph). In accord with our harmonic approximation, we assumed that any stretching force constant can be calculated via a direct linear interpolation between the "pure" C–C single bond (r_{eq} = 1.507 Å, K_r = 337 kcal/mol Å^2) and "pure" double bond (r_{eq} = 1.336 Å, K_r = 570 kcal/mol Å^2). The MM2 force constant for a *single* bond was taken as a fixed reference point. The pure *double* bond came from the analogous carbonyl stretching K_r (since C=O and C=C possess similar stretching frequencies) calculated from our normal mode analysis of *N*-methylacetamide (described below). The structural parameters were selected from microwave data on propane and propene, respectively. This algorithm (Table II) was applied to all remaining carbon–carbon bonds, regardless of specific atom type, in building a consistent family of stretching force constants. Wherever applicable, we checked calculated force constants from our normal modes analysis with predicted scaled values. For example, our interpolation algorithm predicted a stretching force constant for benzene (r_{eq} = 1.40 Å) of 475 kcal/mol Å^2, and the calculated value, which gave the best fit to the experimental frequencies, was 469 kcal/mol Å^2.

Table II. Standardized Parameters for Scaling Algorithms

bond	r_{eq}[a]	K_r[b]
pure C–C	1.507[c]	317[d]
pure C=C	1.336[e]	570[f]
pure C–N	1.449[g]	337[h]
pure C=N	1.273[i]	570[j]
torsion	r_{eq}[a]	$V_2/2$[k]
pure X–C–C–X	1.507[c]	0.0[l]
partial X–C=C–X	1.397[m]	5.5[n]
pure X–C=C–X	1.336[e]	30.0[o]
pure X–C–N–X	1.449[g]	0.0[p]
partial X–C=N–X	1.335[q]	10.0[r]
pure X–C=N–X	1.273[i]	30.0[s]

[a] In Å. [b] In kcal/mol Å^2. [c] Microwave data from acetone.[27] [d] Value taken from MM2 ref 1. [e] Microwave data from propene.[26] [f] Default from NMA normal mode analysis for carbonyl force constant. [g] Benedetti structural data (ref 28). [h] Value derived from normal mode analysis on NMA. [i] Microwave data from methylenimine (ref 27). [j] Default value. [k] In kcal/mol. [l] Assumed free rotation about pure C–C single bond. [m] Structural data from benzene (ref 27). [n] From normal modes analysis of benzene. [o] Approximate rotational barrier of ethylene is ≈60 kcal/mol (see ref 32). [p] Assumed free rotation about a pure single C–N bond. [q] Benedetti structural data (ref 28). [r] Reference 3. [s] Calculated rotational barrier in methylene imine is 57.5 kcal/mol (see ref 33).

Similarly, an analogous interpolation scheme was employed for all carbon–nitrogen bonds. The "pure" single bond N–C equilibrium distance of 1.449 Å came from Benedetti's[28] structural parameter for the N–C_α bond while the K_r = 317 kcal/mol Å^2 was taken from MM2. For the pure double bond N=C force constant, we selected Harmony's[27] microwave data on methylenimine (r_{eq} = 1.273 Å) and used the default value of 570 kcal/mol Å^2 for K_r. Our algorithm predicted K_r = 490 kcal/mol Å^2 for the *partial* C=N bond of atom types found in the amide linkage, while the value which gave the best fit to experimental frequencies was 488 kcal/mol Å^2. These close correlations support the usefulness of our interpolation method for derivation of approximate stretching force constants. In this fashion we were able to derive K_r values for all the bonds in our force field.

We should note that this choice of 570 kcal/mol gives approximately 100–200 cm^{-1} too low frequency for pure C=O and C=C double bonds, in such systems as acetone and 2-butene, where a value of $K_r \approx$ 700 kcal/mol Å^2 is required to fit the vibrational frequencies. However, the use of such a pure C=C force constant gives a much poorer K_r for benzene. Since proteins and nucleic acids have more "aromatic" character than pure double bond C=C or C=N, we chose to use the 570 kcal/mol Å^2.

The development of bond angle parameters followed a similar route. We chose initial ϑ_{eq} values from experimental data on

(26) Cox, S.; Williams, D. *J. Comput. Chem.* **1981**, *2*, 304.

(27) Harmony, M.; Laurie, V.; Kuezkowski, R.; Schwendeman, R.; Ramsay, D.; Lovas, F.; Lafferty, W.; Maki, A. *J. Phys. Chem. Ref. Data* **1979**, *8*, 619.

(28) Benedetti, E. In "Peptides-Proceedings of the 5th American Peptide Symposium"; Goddman, M., Meienhofer, J., Eds.; New York, 1977; p 257.

appropriate reference compounds; e.g., ϑ_{eq} for C3–C2–C3 (generally CX–C2–CX) came from the respective angle in propane. We note that such a choice for X–C2–X indirectly corrects for the absence of explicit hydrogens on C2. Initial values of K_ϑ for typical Xsp^3–Xsp^3–Xsp^3 values came from MM2, but both K_ϑ and ϑ_{eq} values were altered in our model calculations on THF, described below. Our normal mode calculations also played a large role in our choice of K_ϑ values. For example, we made the assumption that all K_ϑ(C2–C2–X), where X = OS, OH, N*, or other electronegative atoms, was the same as the K_ϑ(C2–C2–OS) derived for THF. In the above fashion we were able to derive reasonable and internally consistent values for all the K_r, K_ϑ, r_{eq}, and ϑ_{eq} parameters in our force field.

Torsional Parameters. Our torsional parameters initially came from experimental data on conformational equilibria of molecules. However, since the nonbonded and torsional terms are highly coupled, many were modified during the test case studied described below.

The torsional parameters we use are divided into three types: *general*, *specific*, and *improper*. Examples should suffice to illustrate each. The *general* torsional parameter for X–CH–CH–X has $V_3/2$ = 2.0 kcal/mol, number of bonds = 4, and phase (γ) = 0°. Recall that each CH atom has two non-hydrogen attachments, implying that there are four X–CH–CH–X associated with a given bond. Each of these are assigned a torsional potential of magnitude 0.5 kcal/mol with γ = 0°, to be consistent with $V_3/2$ = 2.0 kcal/mol (0.5 × 4 = 2.0). Such a phase leads to a preference for staggered over eclipsed bonds.

Any *specific* parameter, such as OS–CH–CH–OS, overrides any *general* parameter. For this example we use both the $V_3/2$ = 0.5 kcal/mol and a $V_2/2$ = 0.5 kcal/mol to ensure the correct gauche tendency of O–C–C–O units.[29,30] In the case of the peptide bond, we employ a $V_1/2$ = 0.65 kcal/mol Fourier component to correctly reproduce the cis/trans energy difference for *N*-methylacetamide of 2.1 kcal/mol.[31]

Improper torsions such as X–X–N–H (e.g., C′–C_α–N–H for a peptide) possess four atoms *not* bonded in a successive fashion to one another and serve two purposes. The twofold terms ensure the correct planar tendency of sp^2 atoms while the threefold terms (e.g., X–X–CH–C2) keep asymmetric centers, such as the C_α of amino acids, from racemizing when one uses the united-atom approximation. The magnitude of the V_2 terms has been determined from normal mode calculations on *N*-methylacetamide (described below); $V_3/2$ is assigned an arbitrarily large (14 kcal/mol) value.

In the case of proteins, there are a number of cases where normal twofold torsional parameters were used, for example, in the amide bond to reprodue a rotational barrier of 20.0 kcal/mol (ref 3), in X–S–S–X bonds, to ensure a gauche tendency, and in tyrosine O–H to maintain a barrier of 4.0 kcal/mol about the C–O torsion. We carried out model calculations on single amino acid systems, to ensure that our V_2 values lead to net barriers in reasonable agreement with experiment. Our torsional parameters in those cases are very similar to the ones in ref 3.

For the many torsions where no detailed experimental data exists, we initially employed a simple linear interpolation method to derive V_2(X–C_{sp^2}–C_{sp^2}–X) and V_2(X–C_{sp^2}–N_{sp^2}–X), under the assumption that the relative bond lengths was inversely proportional to V_2. We then compared the calculated values for benzene (r_{eq} = 1.397 Å) and *N*-methylacetamide (r_{eq} = 1.335 Å) and found that a simple linear interpolation produced a significantly higher V_2 value than that which gave a good fit to the out of plane vibrational frequency. For this reason, we employed a dual linear scaling method using a *pure* single, *pure* double, and *partial* double bond (i.e., NMA) as reference points (Table II). An example here will suffice to exhibit this algorithm. The *pure* double bond had $V_2/2$ = 30.0 kcal/mol from methylenimine, the *partial* double bond was $V_2/2$ = 10.0 kcal/mol as fit from our normal mode analysis of NMA, and the barrier to rotation for *pure* single bond was assumed to be 0.0 kcal/mol. The structural parameters came from microwave experimental data on the NMA (CH_3—N), NMR (C=N), and methylenimine (C=N) for single, partial double, and pure double bond character, respectively. All C—N bond lengths between 1.335 and 1.449 Å were scaled between 10.0 and 0.0 kcal/mol, while C—N bonds between 1.273 and 1.335 Å were scaled between 30.0 and 10.0 kcal/mol. Our C—C torsional parameters were analyzed in an identical fashion (Table II). We found that the $V_2/2$ value which fit the lowest out of plane frequency for benzene was 5.5 kcal/mol. This value then represented the partial double bond character for C=C bonds and was used as an intermediate in the scaling algorithm for C—C torsions.

Nucleic Acid Test Cases

Tetrahydrofuran and Methyl Ethyl Ether. We began our analysis of model systems with a study of tetrahydrofuran (THF) and methyl ethyl ether (MEE). THF was selected to give us appropriate values for C–C–O, C–O–C, and C–C–C bending force constants as well as C–C and C–O torsional potentials. The results of calculations on MEE suggest which C3···C3 van der Waals parameters to select.

THF, in the united-atom approximation, is a five-atom system OS–C2–C2–C2–C2. We began with microwave data on small models to determine r_{eq} and ϑ_{eq}, except for r_{eq}(C–O), where we used a standard value of 1.425 Å instead of the MM2 value of 1.41 Å (the length in methanol and closer to the average of the Cambridge crystal file average for phosphate ethers of 1.422 Å, rather than the 1.410 Å value for dimethyl ether). MM2[1] values were taken for the bond stretching parameters K_r for C–C and C–O bonds and bond bending parameters K_ϑ for C–C–C, C–O–C, and C–C–O angles. The torsional parameters, $V_3/2$ = 1.5 kcal/mol for X–C–C–X and $V_3/2$ = 1.0 kcal/mol for X–C–O–X bonds, roughly reproduce typical rotational barriers in alkanes and ethers. We set as our goal the reproduction of the following experimental data for THF (Table III): the energy difference between C_2 and C_S geometries of 0.1 kcal/mol, the barrier to planarity $\Delta E(C_{2v}$–$C_S)$ = 3.7 kcal/mol, the sugar pucker parameter $q \approx 0.4$ Å, and the bond angles C–C–C, C–O–C, and C–C–O for the C_2 and C_{2v} conformations.

To reach this goal, we varied K_ϑ and ϑ_{eq} as well as the torsional parameters. The MM2[1] values for the bending parameters are 32, 50, and 55 kcal/mol rad^2 for C2–C2–C2, C2–C2–OS, and C2–OS–C2, respectively. We only altered the C2–C2–C2 value to 40 kcal/mol rad^2 to obtain a good fit with the various experimental data, with $V_3/2$(X–C2–C2–X) = 1.45 kcal/mol, $V_3/2$(X–C2–OS–X) = 1.05 kcal/mol, and $V_2/2$(C2–C2–OS–C2) = 0.1 kcal/mol. (The latter torsional parameters is only a small perturbation used to refine the gauche–trans energy difference in methyl ethyl ether, but since it makes physical sense, we retain it. One would expect from the work of Brunck and Weinhold[42] that the oxygen lone pairs would prefer to be trans to C–C rather than C–H bonds and thus C–O–C–C would have a small gauche tendency from electronic through-bond effects.) The increase from the MM2 K_ϑ(C–C–C) is also sensible since that value refers only to C–C–C bending while our force field should have implicit contribution from H–C–H and H–C–C angle distortions. The

(29) Hayes, D.; Rothenberg, S.; Kollman, P.; *J. Am. Chem. Soc.* **1977**, *99*, 2150. Marsh, F.; Weiner, P.; Douglas, J.; Kollman, P.; Kenyon, G.; Gerlt, J. *Ibid.* **1980**, *102*, 1660. Govil, G. *Biopolymers* **1976**, *15*, 2303.
(30) Olson, W. *J. Am. Chem. Soc.* **1982**, *104*, 278.
(31) Barker, R.; Boudreaux, G. *Spectrochim. Acta, Part A* **1967**, *23A*, 727.
(32) Douglas, J.; Rabinovich, ; Looney, F. *J. Chem. Phys.* **1955**, *28*, 315.
(33) Lehn, J. *Theor. Chim. Acta* **1970**, *16*, 351.
(34) Engelsholm, G.; Luntz, A.; Gwinn, W.; Harris, D. *J. Chem. Phys.* **1969**, *50*, 2446.
(35) Cremer, D.; Pople, J. *J. Am. Chem. Soc.* **1975**, *97*, 1354.
(36) Almenningen, A.; Seip, H.; Walladsen, *Acta. Chim. Scand.* **1969**, *23*, 2748.
(37) Geise, H.; Adams, W.; Bartell, L. *Tetrahedron* **1969**, *25*, 3045.
(38) (a) Eyster, J.; Prohofsky, E. *Spectrochim. Acta, Part A* **1974**, *30A*, 2041. (b) Deroualt, J.; Forel, M.; Maraval, P. *Can. J. Spect.* **1978**, *23*, 67.
(39) Profeta, S.; Jr.; Kollman, P., unpublished MM2 results on methyl ethyl ether.
(40) Kitayawa, T.; Miyazawa, T. *Bull. Chem. Soc. Jpn.* **1968**, *41*, 1976.
(41) Jorgensen, W.; Ibrahim, M. *J. Am. Chem. Soc.* **1981**, *103*, 3976.
(42) Brunck, T.; Weinhold, F. *J. Am. Chem. Soc.* **1979**, *100*, 1700.

Table III. Comparison of Two Model Force Fields for THF and MEE

parameter	FF1[a]	FF2[b]	exptl[c]
K_ϑ(C–C–O)[d]	50	100	
K_ϑ(C–O–C)[d]	55	80	
K_ϑ(C–C–C)[d]	40	63	
$V_3/2$(C–C)[e]	1.45	2.0	
$V_3/2$(C–O)[e]	1.05	1.45	
Results for Tetrahydrofuran			
$\Delta E(C_2-C_s)$[f]	0.1	0.2	≈0.1[h]
$\Delta E(C_2-C_{2v})$[g]	3.44	3.7	3.52[h]
structural parameters			
C_2 conformation			
q[i]	0.43	0.40	0.38[j]
ϑ(C–O–C)	109.7	109.9	110.5[j]
ϑ(C–C–O)	105.1	105.2	106.5[j]
ϑ(C–C–C)	100.6	101.3	101.8[j]
C_s conformation			
q[i]	0.40	0.37	0.364,[k] 0.38[k]
ϑ(C–O–C)	106.0	107.0	106.2[k]
ϑ(C–C–O)	103.4	103.7	105.0[k]
ϑ(C–C–C)	104.2	104.2	104.1[k]
vibrational frequencies[l]			
B; pseudorotation	43	39	(≈0)[m]
A; out of plane torsion	301	356	286[n]
B; ring bending	441	543	581[n]
A; ring bending	509	620	655[n]
A; ring stretching	840	839	888, 913[n]
B; ring stretching	955	976	909, 964[n]
A; ring stretching	970	991	918, 1071[n]
A; ring stretching	1069	1116	1030[n]
B; ring stretching	1130	1148	1070, 1241[n]
Results for Methyl Ethyl Ether			
ΔE(g–t)	1.4	1.6	1.4 ± 0.2[o]
ΔE(c–t)	7.7	9.4	(5.9)[p]
ΔE(120°–t)	2.1	2.8	(2.9)[p]
structural parameters			
gauche			
Φ	81	77	72, 85[q]
ϑ(C–O–C)	112.9	113.2	(113.2)[p]
ϑ(C–C–C)	112.2	111.4	(112.2)[p]
cis			
ϑ(C–O–C)	119.5	117.6	(116.3)[q]
ϑ(C–C–C)	119.7	118.1	(117.3)[q]

[a] Force field developed initially. [b] Force field developed with rger bond angle bending terms. [c] Experimental values. [d] Bond ending force constant in kcal/mol rad^2. [e] Torsional parameter in cal/mol. [f] Difference in energy between C_2 and C_s conformations of THF (kcal/mol). [g] Difference in energy between C_2 and anar C_{2v} conformations of THF (kcal/mol). [h] Reference 34. Mean out of plane distance of ring, as defined in ref 35. [j] Reference 36. [k] Reference 37. [l] Vibrational frequencies of THF in n^{-1}. The symmetry is C_2. [m] Pseudorotation mode (see ref 34). Experimental frequencies from ref 38a and 38b. The two sets assignments in the two references are given when they are not agreement. [o] References 39–41. [p] MM2 calculations ref 39. See ref 41 for discussions on these parameters.

sults are summarized in Table III under the heading FF1 and dicate that we have done a reasonable job of reproducing the perimental data for THF noted above. After this had been mpleted, the normal mode calculational facility became available AMBER,[21,43] and we carried out such calculations on dimethyl her and dimethyl phosphate (described below).

We found that the bond bending force constants in the range 40–55 kcal/mol rad^2 could not qualitatively reproduce the ending frequencies of THF. We confirmed that this was not an tifact of the united-atom approximation by carrying out parallel lculations with an all-atom model. Hence, we began to search r a set of bending parameters which could more closely reproduce the structures, energies, and vibrational frequencies of THF. Table III contains the end results of our calculations. To reach the goal of reproducing the experimental frequencies, it was necessary to increase the bending force constants by 50–100%, which caused significant flattening of the ring. Increasing the torsional parameters compensated for thus by raising the barrier to planarity, resulting in significantly larger but still qualitatively reasonable values of $V_3/2$(X–C2–C2–X) = 2.0 kcal/mol and $V_3/2$(X–C2–OS–X) = 1.45 kcal/mol. These changes in the torsional barrier worsened the agreement with experiment for the lowest non-pseudo rotation mode (out of plane torsion in Table III), but the sum of the errors between calculated and observed frequencies, for the torsion and bending modes, is significantly reduced. It was clear from our vibrational analysis calculations on all-atom and united-atom THF that the separation of C–C and C–O models from H–C–C bending would be very difficult. This fact, plus the ambiguity in the assignments of C–C and C–O stretching modes from THF vibrational analyses in the literature, caused us to use the K_r(C–C) and K_r(C–O) derived below for diethyl phosphate.

In parallel with the THF calculations, we had been carrying out calculations on methyl ethyl ether using the bond, angle, and torsion parameters from THF and varying the C3···C3 van der Waals parameter as a means of ascertaining which values could best reproduce the experimental data on MEE. We found that, with the use of Dunfield's[24] or Jorgensen's[25] van der Waals parametes, we could not get qualitatively reasonable barriers of rotation about the C–O bond for methyl ethyl ether. We reverted to C3 van der Waals parameters of about the same R^* and somewhat smaller ϵ as found in our previous force field.[16] We felt that the previous force field has possessed too deep van der Waals well depths and this had helped lead to excessively attractive energies for ligand–protein binding.[17] However, we were concered that our small R^* (compared to the Dunfield and Jorgensen values) might lead to a significant "collapse" during refinement. Thus, it was necessary to use as large a C3 R^* as possible and still get reasonable properties for MEE.

We examined the effect of "scaling down" the 1–4 van der Waals (van der Waals interactions separated by only three bonds) by an empirical factor, as had precedent in the work of Dunfield.[3] Hagler et al.,[44] in studies of peptides, argued that if one allows bond angle relaxation, the necessity of reducing 1–4 van der Waals parameters disappears. However, there are a number of compelling reasons for choosing to scale down 1–4 van der Waals interactions. First, a R^{-12} repulsion term is too steep compared with the more correct exponential term, and the error from this should be largest for the close 1–4 interactions. Second, there is likely to be more significant charge redistribution during close 1–4 interactions than during corresponding intermolecular contacts, which will have the effect of reducing the net repulsion. Third, this reduction will enable us to use R^* values closer to those of Dunfield's[24] or Jorgensen's[25] and still calculate reasonable properties for MEE. Below, we give further empirical examples of how reducing the 1–4 van der Waals interactions by a factor of 0.5 improves results of calculations on nucleosides and peptides. Such calculations led us to settle on an R^* value of 2.00 Å and ϵ = 0.15 kcal/mol for C3. This value was based on calculations with FF1 (Table III). When we derived FF2, the calculation was repeated on MEE. The agreement with experiment was better for the 120°–trans difference, not so favorable for the gauche–trans difference, and significantly worse for the cis–trans energy difference. However, the cis conformation is sufficiently high in energy that, with a simple united-atom force field, one cannot expect to reproduce the values as accurately as for low-energy structures.

We wished to assess how much the effect of using this van der Waals parameter (plus the value of R^* = 1.65 Å, ϵ = 0.15 kcal/mol, for OS) would have on the calculated density and vaporization enthalpy of dimethyl ether (DME) previously studied by Jorgensen in Monte Carlo simulations.[45] Jorgensen had found

(43) The code to derive the analytical second derivatives for torsional ergies which avoided singularities was provided to us by B. Brooks and M. arplus of Harvard University.

(44) Hagler, A.; Stern, P.; Sharon, R.; Becker, J.; Naider, F. *J. Am. Chem. Soc.* **1979**, *101*, 6842.

Table IV. Dimethyl Phosphate Energies and Geometries[a]

	FF2[b]	FF0[c]	QM[d]	X-Ray[e]
Relative Energies of Conformations[f]				
g,g[g]	0	0	0	45
g,t[h]	0.8	0.9	0.14	6
t,t[i]	1.7	1.9	0.68	
Geometrical Parameters (g,g)[j]				
Φ_1[k]	70	63	68	73
Φ_2[l]	70	63	68	73
ϑ(C–O–P)[m]	120.6	123.0	112.4	121.7
ϑ(O–P–O)[n]	103.0	103.6	98.8	104.8
ϑ(O'–P–O')[o]	119.8	118.6	125.7	119.7
ϑ(O'–P–O)[p]	108.3	109.5		110.6
ϑ(O'–P–O)[p]	108.1	107.4		105.6
Geometrical Parameters (g,t)[q]				
Φ_1[k]	69	63	75	74
Φ_2[l]	179	179	179	169
ϑ(O–P–O)[n]	102.8	102.4	94.9	99.3
ϑ(O'–P–O')[o]	119.8	119.2	124.3	*e*
ϑ(O'–P–O)[p]	108.4	109.6		*e*
ϑ(O'–P–O)[p]	108.0	106.6		*e*
Geometrical Parameters (t,t)[r]				
ϑ(O–P–O)[n]	102.6	101.2	91.0	*e*
ϑ(O'–P–O)[o]	119.9	119.7	123.1	*e*
ϑ(O'–P–O)[p]	108.3	108.6		*e*

[a] Energies in kcal/mol, angles in degrees. [b] This study, FF2 as described in Table III, with scale factor of 0.5 for 1-4 van der waals and electrostatic interactions. [c] Previous force field (see ref 16). [d] Gorenstein et al. (ref 51). [e] Unpublished results by Peter Murray-Rust using the Cambridge crystal data file. The geometrical parameters are averages taken from the 10 structures which have $R < 9\%$ and no atom larger than Br. Of these, seven are g,g and three are g,t. Except for ϑ(O–P–O), there is no statistically significant difference between the ϑ's, so the values reported under the g,g average are the average for all 10 structures. [f] For X-ray structures, the numbers reported are the number of structures of type R-O-P-O_2'-O-R'- with each conformation. [g] Gauche,gauche conformation. [h] Gauche,trans conformation. [i] Trans,trans conformation. [j] Selected geometrical parameters for the g,g conformation. [k] Φ_1 = dihedral angle C_1–O_1–P–O_2. [l] Φ_2 = dihedral angle O_1–P–O_2–C_2. [m] C–O–P angle (or its average). [n] O–P–O angle (or its average). [o] O'–P–O' angle (involving anionic oxygens). [p] O'–P–O angle; there are four such angles and we report only the largest and smallest. [q] Selected geometrical parameters for the g,t conformation. [r] Selected geometrical parameters for the t,t conformation.

a density 3% too low and an enthalpy of sublimation 5% too high for DME. Since our van der Waals parameters allow closer contacts, but have shallower well depths, we expected that a Monte Carlo simulation of DME with our van der Waals parameters would lead to errors in the opposite direction as those of Jorgensen. We carried out[46] such a simulation of DME using the same electrostatic parameters as his and found, indeed, that our density was 10% too high and the enthalpy of sublimation 10% too low.

We also tested these parameters on *n*-butane and calculated a gauche–trans energy difference of 0.91 kcal/mol, a dihedral angle in the gauche confomation of 66°, and a C–C–C angle in the gauche conformation of 114.1°, all in quite good agreement with the experimental values of 0.97 kcal/mol,[47] 67.5°, and 113.5°,[48] respectively. As in the case of MEE, we did, however, overestimate the cis barrier (calculated at 6.9 kcal/mol, with HFSCF + CI results of 4.5–4.7 kcal/mol).[48]

Dimethyl Phosphate. The geometrical parameters r_{eq} and ϑ_{eq} for the phosphate group of dimethyl phosphate (DMP) were taken from the older "standard" values for the $ROPO_2OR'$ group suggested by Newton.[49] A more recent critical analysis using the Cambridge crystal data bank[50] found parameters within a standard deviation of these values, leading us not to change the earlier parameters. We also left our torsional parameters as in the old force field, with both $V_2/2$ and $V_3/2$ terms at 0.75 kcal/mol. The results of calculations using our old and new parameters are compared with the best quantum mechanical values and the average experimental values in Table IV. The calculated energy as a function of conformation is qualitatively reasonable, although the gap between the molecular mechanical energies of various conformations is still significantly larger than that found quantum mechanically. However, the quantum mechanical values came from calculations with a minimal basis set, and some of the OS–P–OS and O2–P–O2 angles calculated with such an approach are outside the range of experimental values.[50,51] Hence, we have not required that the molecular mechanical calculations give relative energies in precise agreement with the quantum mechanical results. It is clear that one could make the relative molecular mechanical conformational energies for g,g and g,t agree more closely with the quantum mechanical difference of 0.2 kcal/mol by merely reducing the twofold torsional parameter, $V_2/2$, for the C–O–P–O linkage.

There are two other interesting structural aspects of DMP that deserve comment. Gorenstein et al.[51] noted that both quantum mechanical calculations and X-ray structures suggest that ϑ(O–P–O) is strongly dependent on phosphate conformation, with X-ray structure averages $\vartheta_{eq} = 104.8°$ for g,g conformation and 99.3° for g,t. Second, ϑ_{eq}(O–P–O) angles differ by ≈5° when there are gauche ROPO linkages, since one anionic oxygen is gauche to the methyl group and the other is trans (ϑ_{eq}(OPO' gauche) = 110° and ϑ_{eq}(OPO' trans) = 105°). Both the new and old force fields show evidence of these two effects, although in FF2 they are an order of magnitude too small. In the old force field, the magnitudes of the effects are 20–60% of the observed ones, most likely due to the fact that the 1–4 van der Waals were not scaled by a factor of 0.5. Again this is a situation where one must compromise; we deem the reproduction of the structure and energies of MEE and adenosine more important than the reproduction of the magnitude of the DMP angle differences. Our calculated ϑ(O–P–O) is very close to both the average of the angle for the g,g and g,t conformations as well as the X-ray crystallo-

Table V. Calculated and Observed Frequencies for Diethyl Phosphate (cm^{-1})

mode	this work	Brown and Peticolas[a]	exptl[b]
ν_7	54		
ν_8	105		
ν_9	107		
ν_{10}	136		
ν_{11}	200	187	195
ν_{12}	211	201	210
ν_{13}	287	290	321
ν_{14}	290	329	345
ν_{15}	367	333	357
ν_{16}	418	398	393
ν_{17}	458	492	503
ν_{18}	524	565	551
ν_{19}	580	578	569
ν_{20}	741	775	763
ν_{21}	841	814	812
ν_{22}	983	941	945
ν_{23}	983	954	945
ν_{24}	1036	1051	1053
ν_{25}	1088	1064	1053
ν_{26}	1102	1080	1077
ν_{27}	1235	1225	1215

[a] Reference 52. [b] Reference 53.

(45) Jorgensen, W.; Ibrahim, M. *J. Am. Chem. Soc.* **1981**, *103*, 3976.
(46) We first reproduced values for internal energy and volume reported in ref 41 and then repeated the simulation with the altered parameters.
(47) Verma, A.; Murphy, W.; Bernstein, H. *J. Chem. Phys.* **1974**, *69*, 1540. Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1959**, *32*, 748.
(48) This is an MM2 value from: Allinger, N.; Profeta, S., Jr. *J. Comput. Chem.* **1980**, *1*, 181.
(49) Newton, M. *J. Am. Chem. Soc.* **1973**, *95*, 256.
(50) Unpublished analysis of $ROPO_2R^-$ structures by Peter Murray-Rust.
(51) Gorenstein, D.; Findlay, J.; Luxon, B.; Kar, D. *J. Am. Chem. Soc.* **1977**, *99*, 3473.

graphic value. Thus we do not expect that large errors will result from the use of these average values.

We determined the stretching and bending force constants for the phosphate group by carrying out normal mode calculations on diethyl phosphate and comparing these with results from a typical vibrational analysis calculation by Brown and Peticolas[52] and with experiment (Table V). The fit between our calculations and the more complete force-field calculations/exepriment is reasonable. The C–C stretching force constant of 260 kcal/mol Å^2 is similar to that found by Karplus and Kuschick[54] on *n*-butane while the C–O K_r = 310 kcal/mol Å^2 fits the C–O stretching frequency in dimethyl ether[55] very well.

Calculations on Deoxyadenosine and a Model Deoxyribose. When we turn to a more complex molecule such as deoxyadenosine or DNA in general, there are a large number of empirical parameters to be determined. The sources for the initial sets of parameters for the nucleic acid bases adenine, guanine, cytosine, thymine, and uracil were as follows. The bond length and angle equilibrium parameters for the nucleic acid bases were taken from X-ray structures. The torsional potentials were all assumed to be twofold, as in typical double or partially double bonded molecules. The interpolation approach (Table II) was used for determining the stretching force constants.

The sp^2 angle bending parameters remained undetermined, and, in the absence of definitive values, we used analogies from our *N*-methylacetamide (NMA) normal mode calculations (described below). The default value of K_ϑ = 70 kcal/mol rad^2 was used for the bases, and this value was modified on the basis of normal mode calculations of NMA and benzene. For example, in NMA, the X—C=O bending parameter was 80 kcal/mol rad^2, and this value was used for such groupings in the nucleic acid bases. For NMA, the X–N–H value was somewhat smaller, 35 kcal/mol rad^2, causing us to use this X–N–H parameter in the bases. The C_{sp^2}–C_{sp^2}–C_{sp^2} value of 85 kcal/mol rad^2 came from the benzene normal mode calculations.

Some additional parameters needed to be determined for the sugar phosphate backbone of the nucleic acids. For all C_{sp^3}–C_{sp^3}–C_{sp^3} we used the same bending force constant as the value for C2–C2–C2 derived from THF. However, ϑ(X–CH–Y) (X,Y sp^3 carbons) should not be the same as ϑ(X–C2–Y), since propane is no longer an appropriate reference. The K_ϑ for C_{sp^3}–OH–HO came from normal mode analysis calculations on methanol. The remaining parameters were taken from appropriate analogies (for example, we took all K_r and r_{eq} for C_{sp^3}–C_{sp^3} from C2–C2).

The deoxyadenosine calculation was the first of these model calculations in which the electrostatic term has been included. The partial atomic charges employed here are listed in Figure 3 in the appendix. By use of a distance-dependent dielectric constant, $\epsilon = R_{ij}$, the C5′–O5′ bond in the g,g range, and the HO5′–O5′–C5′–C4′ and C4′–C3′–O3′–HO3′ in the trans ranges, complete energy refinement of deoxyadenosine was carried out. The torsional angle C4′–C3′–C2′–C1′ was constrained to various values as a means of evaluating the energy as a function of sugar puckering. Only two local minima are found, with C2′ endo and C3′ endo conformations; the lowest barrier between them occurs in the O1′ endo region. In good agreement with experiment, the C2′ endo conformation is more stable than C3′ endo by 0.6 kcal/mol, with the barrier between C2′ endo and C3′ endo being 1.3 kcal/mol (Table VI). The sugar pucker pseudorotation *W* values for the C3′ endo and C2′ endo conformations are 5° and 152°, respectively. They occur near the middle of the range of observed values, although both are somewhat smaller than the center of the C3′ endo and C2′ endo ranges (18° and 162°, respectively). The use of nonbonded parameters similar to Dunfield's[24] or Jorgensen's[25] significantly reduces the *W* values for the C2′ endo minimum near the lower end of the observed sugar pucker values (≈140°) and reduces the C2′ endo/C3′ energy difference to near 0.0 kcal/mol. This also occurs with a scale factor of 1.0 rather than 0.5 for 1–4 van der Waals interactions.

Olson[30] has recently carried out theoretical studies on a model for a deoxyribose ring, with substituents 1′-NH_2, 5′-CH_3, and 3′-OH, studying the pseudorotation profile of this model with a fixed out of plane q = 0.38 Å and constrained bond lengths. The 1′, 3′, 5′ substituents were treated as united atoms, but all H's on the ring were included in the calculation. Only endocyclic bond angles were energy refined at each *W*. We repeated these calculations (Table VI) with complete energy refinement using a standard bond-stretching force constant of 300 kcal/mol Å^2. All bonds were assigned the Olson values for r_{eq}, while ϑ_{eq} (H–C–H) and ϑ_{eq} (C–C–H) were tetrahedral with a force constant of 40 kcal/mol rad.[2] The agreement between the C2′ endo/C3′ endo energy difference and the *W* and *q* values for this model and our

Table VI. Comparison of Calculations on Deoxyadenosine and a Simplified Model

	FF1[a]	Olson model[b]
E(C2′ endo)[c]	0.0	0.0
W[d]	152	159
q[e]	0.40	0.41
E(C3′ endo)[c]	0.56	0.46
W[d]	5	18
q[e]	0.40	0.42
E(O1′ endo)[c]	1.29	1.27
W[d]	75	86
q[e]	0.40	0.42

[a] This work, with force constants as in Table III; calculations on deoxyadenosine with base 5′ and 3′ substituents all included (CH united atoms for CH carbons). Scale factor of 0.5 for 1–4 nonbonded and electrostatic interactions, $\epsilon = R_{ij}$. [b] Using the standard model of Olson (ref 30), with explicit hydrogens on the sugar, but united atoms at the 5′ (CH_3), 3′ (OH), and 1′ (NH_2) positions and using her torsional, bending, van der Waals, and electrostatic parameters with complete minimization (all K_r = 300 kcal/mol Å^2, ϵ = 4). Scale factor of 1.0 for 1–4 interactions. [c] Relative energy in kcal/mol of given conformation. [d] Pseudorotation angle (see ref 35) of C2′ endo or C3′ endo conformation; value at the top of the potential curve for O1′ endo. [e] Mean out of plane distance for furanose atoms (in Å) (see ref 35).

Table VII. Calculations on Adenosine and Deoxyadenosine

	FF1[a]	FF2[b]	FF2′[c]	FF2″[d]	exptl
			Deoxyadenosine		
ΔE(C3′endo-C2′endo)[f]	0.56	0.53	0.66	0.52	0.66[e]
ΔE(O1′endo-C2′endo)[g]	1.29	2.00	1.80	1.80	
W(C2′endo)[h]	152	152	170	155	165[e]
q(C2′endo)[i]	0.40	0.38	0.36	0.38	(0.35–0.41)[e]
W(C3′endo)[h]	5	3	1	5	(2–20)[e]
q(C3′endo)[i]	0.40	0.37	0.37	0.38	(0.35–0.41)[e]
			Adenosine		
ΔE(C3′endo-C2′endo)[f]		−0.68	1.46	0.21	(0.19–0.42)
W(C2′endo)[h]		174	178	170	(150–170)
q(C2′endo)[i]		0.37	0.35	0.37	(0.35–0.41)
W(C3′endo)[h]		3	359	3	(2–20)
q(C3′endo)[i]		0.39	0.34	0.38	(0.35–0.41)

[a] Same force field as in Table III. [b] Same force field as in Table III; V_Φ are larger than force field 1 (FF1). [c] Same as FF2 with 1–4 electrostatic tems scaled by 0.5. [d] Same as FF2′ with ϵ = $4R_{ij}$. [e] Experimental data from Davies (ref 56) and Altona and Sundaralingham (ref 57). [f] Energy difference between energy minimized C3′endo and C2′endo conformations. [g] Energy difference between O1′endo and C2′endo conformations. [h] Energy refined pseudorotation angle (ref 35) for given conformation. [i] Energy refined mean out of plane sugar distance (see ref 35) for sugar ring.

(52) Brown, E.; Peticolas, W. *Biopolymers* **1975**, *14*, 1259.
(53) Reference 58 and: Shimanouchi, T.; Tsuboi, M.; Kyogoku. Y. *Adv. Chem. Phys.* **1964**, *3*, 435.
(54) Karplus, M.; Kushick, J. *Macromolecules* **1981**, *14*, 325.
(55) Shimanouchi, T. "Tables of Molecular Vibrational Frequencies"; National Standard Reference Data Series-National Bureau of Standards: Washington, D.C., 1967; parts 1–3.
(56) Davies, D. *Prog. Nucl. Magn. Reson. Spectrosc.* **1978**, *12*, 135.
(57) Altona, C.; Sundaralingham, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205.

deoxyadenosine calculations was encouraging and supported the reasonableness of our calculated 1.3 kcal/mol C2′ endo–C3′ endo barrier, with a maximum near the O1′ endo conformation. In the earlier study[30] and in our previous force field,[16] a barrier of 2.0 kcal/mol had been found.

After our normal mode analysis of THF, we returned to the sugar conformational profile of deoxyadenosine. The results of that study are summarized in Table VII. The largest difference between FF1 (with sp^3 bending force constants ≈40–55 kcal/mol Å^2) and FF2 (bending force constants ≈60–100 kcal/mol Å^2) is found mainly in the C2′ endo–O1′ endo energy difference, which had been increased from 1.3 to 2.0 kcal/mol. FF2′ differs from FF2 only in that the 1–4 electrostatic interactions have also been reduced by the 0.5 scale factor. This leads to a minimum energy W closer to the experimental value, but a somewhat smaller q (although still within the experimental range). We also studied conformations in the range of O1′ exo ($W \approx 270°$). The barrier for the C2′ endo–O1′exo–C3′ endo transition was calculated to be 3.4 kcal/mol, with the minimum energy $q = 0.15$ Å. The ring flattens considerably in this conformation, presumably to relieve the base–C5′,O5′ repulsions in the O1′ exo conformation.

Model calculations on the ribonucleoside adenosine led to the same two local minima (C2′ endo and C3′ endo), with the O1′ endo barrier in the range of 3.0 kcal/mol. However, the C2′ endo/C3′ endo energy difference depended on the electrostatic energy and the orientation of the 2′OH, which began for each refinement in a conformation O2′–HO2′ eclipsing the C3′–C2′ bond. Since our distance-dependent dielectric model, $\epsilon = R_{ij}$, still allows for rather strong intramolecular electrostatic interactions compared to what would occur for adenosine in aqueous solution (where presumably all the H-bonding sites would be occupied by H_2O molecules), we examined the effect of using a larger effective dielectric constant, $\epsilon = 4R_{ij}$. This is the calculation labeled FF2″ in Table VII, in which the agreement with both experimental structures and energy differences is quite satisfactorily represented for deoxyadenosine and riboadenosine.

One of the major points in the Olson[30] work was the fact that sugar pseudorotations (between C2′ endo and C3′ endo conformations in furanose rings) were not "nearly free[44]" but required surmounting a 2.0 kcal/mol barrier at the O1′ endo conformation. Our more complete refinement using the Olson parameters led to a barrier of 1.3 kcal/mol, both for her model and deoxyadenosine. Changing to larger K_ϑ, V_n raised the barrier to 2.0 kcal/mol, the value also found with our old force field.[16] Thus, the best available theoretical estimates for this barrier suggest it to be in the range of 1.3–2.0 kcal/mol.

Calculations on Base Pairing, Stacking, and Sequence-Dependent Stabilities. Recently, gas-phase experiments have determined interaction energies for nucleic acid bases giving both hydrogen-bonding and stacking energies. To provide a check of our van der Waals parameters for sp^2 atoms and on our method for deriving partial charges for atoms, we model built and energy refined Watson–Crick H-bonded complexes between 9-methylguanine and 1-methylcytosine, 9-methyladenine and 1-methylthymine, and Hoogsteen base-paired models between 9-methyladenine and 1-methylthymine. Stacking between two 1,3-dimethyluracil molecules was also studied. Such calculations were carried out with dielectric models $\epsilon = 1$ and $\epsilon = R_{ij}$ and are compared with the more elaborate calculations by Langlet et al.[58] and the gas-phase mass spectrometric experiments of Yanson et al.[59] (Table VIII). The agreement between the calculated and experimental values for H-bonding with either dielectric model is quite good, and our calculations, with $\epsilon = R_{ij}$, agree with the findings of Langlet et al. that the Hoogsteen base pairing for AT is better than Watson–Crick.

In the 1,3-dimethyluracil stacking calculations, Langlet et al. have noted that one must extrapolate the experimental data to zero field (which was not done by Yanson et al., who found a base-stcking enthalpy of −3.6 kcal/mol). Since the Langlet et al. calculations find a similar enthalpy at the electric field used by Yanson et al., we feel that their calculated values at zero field ($\Delta H = -9.1$ kcal/mol) are a good estimate for the "true" experimental enthalpy. Our calculations are in satisfactory agreement with their results with a distance between base planes of 3.43 Å). (In our previous force field, the minimum energy distance is calculated to be 3.30 Å.) We also carried out such calculations in the all-atom representation with the van der Waals parameters exactly as used by Hagler et al.;[19] these results were similar to those we found in the united-atom representation with our van der Waals parameters.

In a previous study of dinucleoside phosphates,[16] we had found that observed sequence-dependent stabilities in DNA melting could be rationalized with a simple dinucleoside model, leading us to carry out complete energy refinements on the 10 base-paired dinucleoside phosphates studied earlier with our previous force

Table VIII. Hydrogen Bonding and Stacking for Base Pairs

complex	ΔE- ($\epsilon = 1$)[a]	ΔE- ($\epsilon = R$)[b]	ΔE- (Langlet)[c]	ΔH- (exptl)[d]
GC Watson-Crick[e]	−21.2	−21.6	−23.7	−21.0
AT Watson-Crick[f]	−11.3	−12.9	−12.9	−13.0
AT Hoogsteen[g]	−11.8	−13.5	−13.6	−13.0
1,3-dimethyluracil stack[h]	−9.8[i]	−9.3	−9.1	(−9.1)

[a] Energy of complex formation with $\epsilon = 1$ in kcal/mol. [b] Energy of complex formation with $\epsilon = R_{ij}$ in kcal/mol. [c] Energy calculated by Langlet et al. (ref 58). [d] Experimental value for associations inferred from the experiments by Yanson et al. (ref 59). In the case of the 1,3-dimethyluracil stacking, the value in parentheses is the value calculated by Langlet et al. (ref 58), since these authors showed that there was an important electric field dependence in the experiments by Yanson. [e] Watson and Crick H-bonded structure of 9-methylguanine and 1-methylcytosine. Model built using computer graphics and then energy refined. [f] Watson and Crick H-bonded structure of 9-methyladenine and 1-methylthymine. Model built using computer graphics and then refined. [g] Hoogsteen H-bonded structure of 9-methyladenine and 1-methylthymine. Model built using computer graphics and then energy refined. [h] Stacked complex of 1,3-dimethyluracil model built using Figure 11 A1 in the paper by Langlet et al. (ref 58) and energy refined, base-base minimum energy distance = 3.43 Å. [i] Calculation using explicit representation of C–H groups and the Hagler et al. nonbonded parameters (ref 19) base-base minimum energy distance ≈ 3.71 Å.

Table **IX.** Calculations on Sequence Specificity of Melting Temperature of Nucleotides

	$-\Delta t_m$[a]	ΔE- ($\epsilon = 1$)[b]	ΔE- ($\epsilon = R_{ij}$)[c]
DNA polymer			
1. poly[d(G-C)]·poly[d(G-C)]		0.0	0.0
2. poly d(G)·poly d(C)	13	5.4	2.2
3. poly[d(A-T)]·poly[d(A-T)]		0.0	0.0
4. poly d(A)·poly d(T)	−9	−0.9	−1.2
5. poly[d(T-G)]·poly[d(C-A)]		0.0	0.0
6. poly[d(T-C)]·poly[d(G-A)]	7	1.7	0.8
7. poly[d(A-T-C)]·poly[d(G-A-T)]		1.1	0.9
8. poly[d(T-T-G)]·poly[d(C-A-A)]	2	0.0	0.0
9. poly[d(T-A-C)]·poly[d(G-T-A)]	5	0.5	0.5
10. poly[d(T-T-C)]·poly[d(G-A-A)]	8	1.2	0.5

[a] Difference in melting temperature between isomers in °C; a positive value in the case of 1 vs. 2; 3 vs. 4; 5 vs. 6; means the heteropolymer (1, 3, or 5) melts higher (is more stable). In the case of polymers 7–10; the highest melting polymer is 7, followed in order by 8, 9, 10; see ref 16 for discussion; experimental data from ref 60. [b] Difference in calculated energy (kcal/mol) between polymers. For example, in comparing 1 vs. 2, we compare the energy of $d(CG)_2$ and $d(GC)_2$ to the energy of $d(G_2)\cdot d(G_2)$. [c] Same as *b* for calculations with dielectric constant $\epsilon = R_{ij}$.

(58) Langlet, J.; Claverie, P.; Caron, F. "Intermolecular Forces"; Pullman, R., Ed.; 14th Jerusalem Symposium; Reidel: Dordrecht, 1981.
(59) Yanson, I.; Teplitsky, A.; Sukhodur, L. *Biopolymers* **1979**, *18*, 1149.
(60) Wells, R.; Larson, J.; Grant, R.; Shortle, B.; Cantor, C. *J. Mol. Biol.* **1970**, *54*, 465.

field (Table IX). The relative calculated energies (especially with $\epsilon = R_{ij}$) are a significant improvement over the previous model calculations in comparing the homo- and heteropolymers with a mononucleotide or dinucleotide repeat. In particular, the new force field qualitatively reproduces the relative melting temperature independent of dielectric model whereas the old force field was only successful in doing this with $\epsilon = 1$. In addition, the relative magntidues of the energy differences with the new force field, $\epsilon = R_{ij}$, are in the same order as the magnitudes of the Δt_m. Neither the old nor the new force field had been able to successfully calculate the relative energies for the trinucleotide repeat models, but these may require calculations on tetranucleosides, rather than dinucleosides.

The "improvement" of our new calculated relative energies over the previous ones is encouraging, but we must stress that the relation between our calculated relative energies and the experimental relative melting temperature is very indirect. First, we are assuming that the melting temperature differences are due mainly to differences in the energies of the double-stranded forms. Second, we are assuming that our energy-refined geometries for the base-paired dinucleoside phosphates here and in ref 16 are good representations of the double-stranded geometries in longer DNA double helicies. Although we have concluded that such approximations are likely to be reasonable, we cannot prove this.

Protein Test Cases

Determination of the Peptide Backbone Parameters. The first goal in developing the protein segment of our force field was to derive a consistent set of charges for the hydrogen, nitrogen, carbon, and oxygen atoms contained in the amide segment of the peptide chain. Since it would be impractical to generate charges for various sizes of oligopeptide strands possessing all side chain combinations, it was necessary for us to select a model system that we felt could best represent the backbone. Our choice consisted of the dipeptides *N*-acetyl-*N'*-methylglycinamide and *N*-acetyl-*N'*-methylanalinamide for which numerous theoretical studies have previously been performed (Maigret et al.,[61] Schafer et al.[62]). We hoped to develop a force field that best reproduced structures and energies of the dipeptides for the local minima of PCILO and all-atom molecular mechanical calculations.

We initially generated the electrostatic potential surface at the 6-31G level for *N*-methylacetamide and fit the surface to a 12 point charge model, where all hydrogens were included. It must be pointed out that, at the time, our limited disk storage space made it impossible for us to calculate an initial charge set at the 6-31G basis level for even a glycyl dipeptide; hence the rationale for our choice of NMA.

Due to the inherent dependence of derived charges upon basis set, we felt it paramount to employ the most accurate charge model. Cox and Williams[26] found, in a study on small molecules, that electrostatic potential derived charges from a 6-31G basis set differed from the "optimal" 6-31G** by a ratio of 0.82:1. They also noted that a scaling of 0.91 is needed to adjust the 6-31G** calculated dipole moments to fit the experimental values. To be most consistent within this framework, we decided on a scaling factor of 0.75 (0.82 × 0.91) for our 6-31G derived NMA partial charges.

The bond and angle parameters for *N*-methylacetamide were taken from crystallographic data of the peptide linkage presented by Benedetti.[28] The twofold barrier to rotation about the C–N bond was $V_2/2 = 10.0$ kcal/mol (ref 3). To best represent the experimental energy difference of 2.1 kcal/mol between the cis and trans conformations of NMA,[31] we included $V_1/2 = 0.65$ kcal/mol. Using this updated parameter set we energy refined the all-atom glycyl and alanyl dipeptides using the 0.75 scaled charges. The dipeptides were constrained about their Φ and Ψ angles in 60° intervals, and an energy map of these 36 regions was constructed. Each low-energy area was searched for a local minimum by relaxing all degrees of freedom and further refining.

(61) Maigret, B.; Pullman, B.; Perahia, D. *J. Theor. Biol.* **1971**, *31*, 269.
(62) Schafer, L.; Alsenoy, C.; Scarsdale, J. *J. Chem. Phys.* **1982**, *76*, 1439.

Table X. Geometries and Local Minima for Glycyl and Alanyl Dipeptides

local minima	Φ,[a] deg	Ψ,[b] deg	ΔE[c]
Glycyl Dipeptide All-Atom Representation			
1-7 H-bonded	75	−65	0.0
1-7 H-bonded	−75	65	0.0
extended	180	180	3.3
1-10 H-bonded	60	39	4.1
1-10 H-bonded	−60	−39	4.1
Glycyl Dipeptide United-Atom Representation			
1-7 H-bonded	77	−64	0.0
1-7 H-bonded	−77	64	0.0
extended	180	180	3.2
1-10 H-bonded	66	35	4.1
1-10 H-bonded	−66	−35	4.1
Alanyl Dipeptide All-Atom Representation			
1-7 H-bonded	−76	66	0.0
1-7 H-bonded	69	−64	0.6
extended	−161	169	3.2
1-10 H-bonded	−61	−41	3.6
1-10 H-bonded	54	42	4.3
Alanyl Dipeptide United-Atom Representation			
1-7 H-bonded	−79	69	0.0
1-7 H-bonded	68	−58	0.8
extended	−150	154	2.3
1-10 H-bonded	−69	−29	3.0
1-10 H-bonded	55	35	4.6

[a] Φ convention appears in ref 69. [b] Ψ convention appears in ref 69. [c] Relative energy in kcal/mol.

Finally, "true" local minima were confirmed by using the Newton–Raphson second derivative routine.[43]

The next step was to best "fit" both the structures and energies obtained above for our all-atom model dipeptides to their respective united-atom representations. This time, our charges were derived by fitting the electrostatic potential points to a six-point model of NMA. The structures were subjected to an indentical grid search and local minima analysis as done above.

In the all-atom model for alanyl dipeptide we located five local minima. However, if the 1–4 electrostatic energies were not scaled down in the united-atom case, only three "true" local minima were found. Even with the initial Φ and Ψ torsions constrained to "force" the dipeptide into these regions, the extended structure and the highest energy conformation were not local minima along the potential energy surface. For this reason, and from results mentioned above on adenosine, we decided to empirically scale the 1–4 electrostatic interactions by a factor of 0.5. Unlike the united-atom model, the all-atom relative energies were very insensitive to the inclusion of a scale factor for 1–4 van der Waals energies, and somewhat more sensitive to scaling 1–4 electrostatic.

The alanyl local minima geometries and energies were then recalculated with the appropriate scale factor for both the united- and all-atom models. For the alanyl dipeptide, we located five local minima. The two lowest in energy correspond to 1–7 hydrogen-bonded structures (where the right-handed system is 0.9 kcal/mol more stable than the left handed). An extended structure, forming 1–5 H-bonds, lies 3.2 kcal/mol above the global minimum. The right- and left-handed helicies, corresponding to 1–10 H-bonded conformations, occur at 3.6 and 4.5 kcal/mol above the most stable structure.

Five local minima are found for the glycyl dipeptide. Two isoenergetic global minima occur, as nonsuperimposable mirror images of themselves, for the 1–7 hydrogen-bonded structures. Lying 3.1 kcal/mol above these systems is the extended conformation $\Phi = 180°$ and $\psi = 180°$, with two more isoenergetic helical structures at 4.0 kcal/mol.

It was found that by scaling the 6-31G united-atom charges by 0.81, we achieved good agreement for both the alanyl and glycyl dipeptides, with an average error less than 0.5 kcal/mol between the united- and all-atom representations. The full results appear in Table X. We have shown that an empirical scale factor of 0.5 in the 1–4 electrostatics is necessary to best fit the local minima

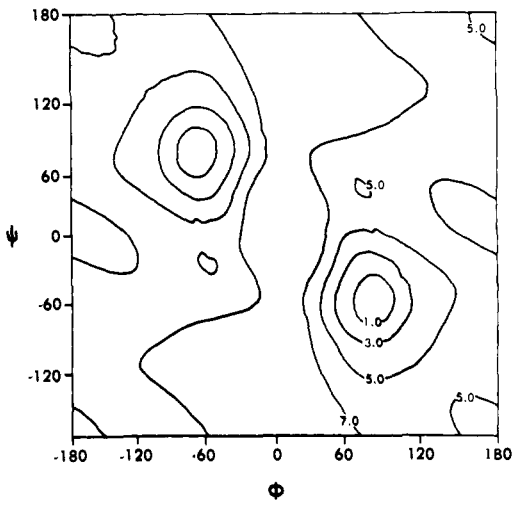


Figure 1. Energy contour map for *N*-acetyl-*N'*-methylglycinamide. Contours are in kcal/mol. The usual IUPAC convention for Φ, Ψ is used.

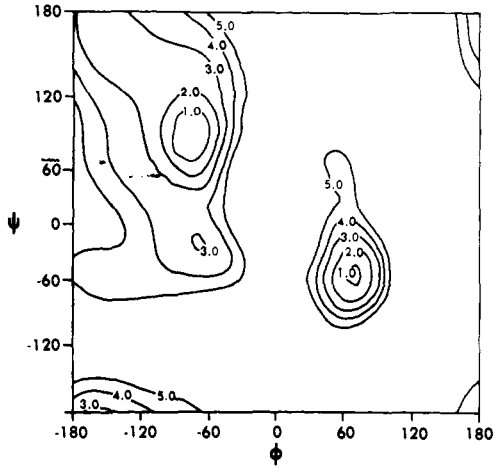


Figure 2. Energy contour map for *N*-acetyl-*N'*-methylalaninamide. Contours are in kcal/mol.

structures and energies for alanyl dipeptide.

To best illustrate the low-energy regions corresponding to these local minima structures, we constructed Φ, Ψ maps consisting of 1296 points derived from energy minimization with all geometric degrees of freedom relaxed (Figures 1 and 2). The alanyl dipeptide maps of Ramachandran and Sasisekharan[63] and Brant et al.[64] are qualitatively quite similar to our plot in the left half of the map (Φ between −180° and 0°). However, their plots fail to locate a low-energy contour in the right half region centered about Φ = 60° and ranging from Ψ ≈ −90° to Ψ ≈ 60°. The appearance of this region is a manifestation of our methodology for generating the structures, in which we allowed all geometric degrees of freedom to relax during the minimization process. The occurrence of additional regions, in refinements employing relaxed geometries compared with constrained minimizations, has been shown by Gibson and Scheraga[65] in the alanyl dipeptide and Gelin and Karplus[66] in a study on β-methylacetylcholine. Furthermore, infrared data by Cung et al.[67] and, Avignon and Lascombe[68] support the existence of both the axial and equatorial C_7 structures (Φ ≈ −80°, Ψ ≈ 70° and Φ ≈ 70°, Ψ ≈ −60°, respectively) for alanyl dipeptide. The 1–13 stabilizing hydrogen bonds which would be formed in the left-handed α helical structure (Φ ≈ 60°, Ψ ≈ 60°) cannot exist in a simple dipeptide model, causing this region to be 5 kcal/mol higher in energy in our map, compared to the global minimum. Thus, we would expect this area to become more stable relative to the right-handed 1–7 system as the dipeptide model is extended to a tetrapeptide structure, and we are currently addressing this question. Additionally, our map failed to exhibit the ubiquitous low energy "finger" contour extending down from the upper right quadrant. This left-handed "finger" region (Φ ≈ 60°, 90° < Ψ < 180°) is present as a low-energy region (<5 kcal/mol) in the other Φ, Ψ maps[63-65] but is 7–8 kcal/mol higher in ours. We should point out that Richardson[69] has analyzed the conformation of 1000 non-glycine residues in globular proteins and has found none in this region.

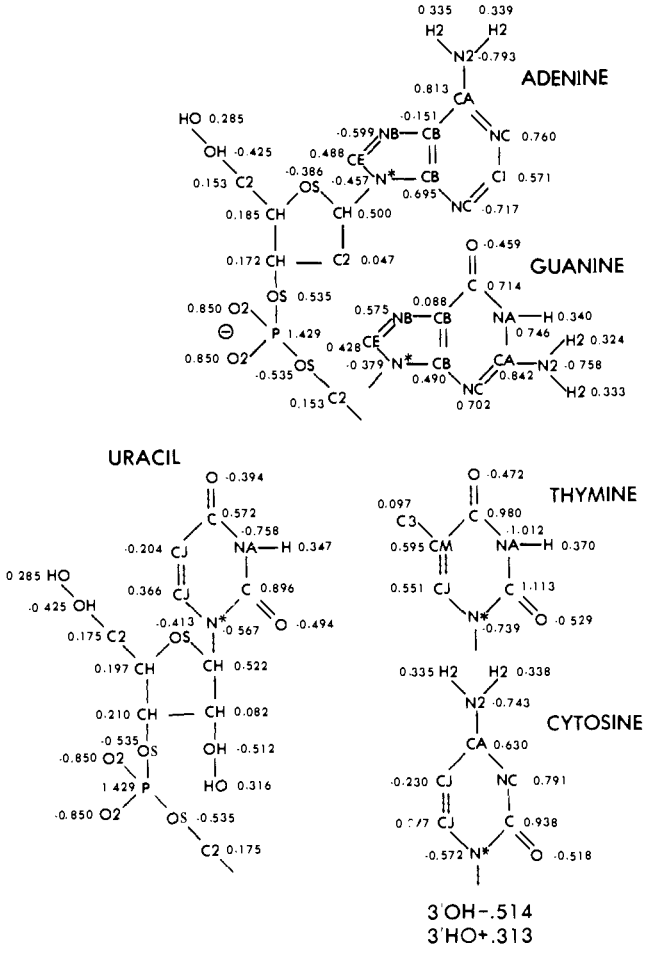


Figure 3. Net atomic charges for the nucleic acids in units of electron charge.

Upon the implementation of software within AMBER[21,43] designed for generating a complete vibrational normal modes analysis, it became possible to "fine tune" both stretching and bending force constants for small molecules. The goal was to obtain a good fit of calculated normal mode frequencies with the best experimental data available for *N*-methylacetamide. Our initial starting parameter set consisted of the Benedetti structural terms, one- and twofold rotational values as mentioned above, and our 6-31G NMA charges scaled by the appropriate value of 0.81. The stretching and bending force constants were assigned standard default values taken from our original parameters set (e.g., 50 kcal/mol rad^2 for bending terms).

The methodology for our normal mode analysis employed an iterative process where initial calculations were run on an all-atom model. Results from this simulation gave us insight into which

(63) Ramachandran, G.; Sasisekharan, *Adv. Protein Chem.* **1968**, *23*, 283.
(64) Brant, D.; Miller, W.; Flory, P. *J. Mol. Biol.* **1967**, *23*, 47.
(65) Gibson, K.; Scheraga, H. *Biopolymers* **1966**, *4*, 709.
(66) Gelin, B.; Karplus, M. *J. Am. Chem. Soc.* **1975**, *97*, 6996.
(67) Cung, M.; Marraud, M.; Neel, J. In "The Jerusalem Symposia on Quantum Chemistry and Biochemistry-Conformation of Biological Molecules and Polymers"; Bergmann, E., Pullman, B., Ed.; Jerusalem, 1973; p 69.
(68) Avignon, M.; Lascombe, J. In ref 67, p 97.
(69) Richardson, J. *Adv. Protein Chem.* **1981**, *34*, 167.
(70) Rey-Lafon, H.; Forel, M. T. *Spectrochim. Acta*; *Part A* **1973**, *29A*, 471.
(71) Scott, D.; McCullough, J. *J. Am. Chem. Soc.* **1958**, *80*, 3554.

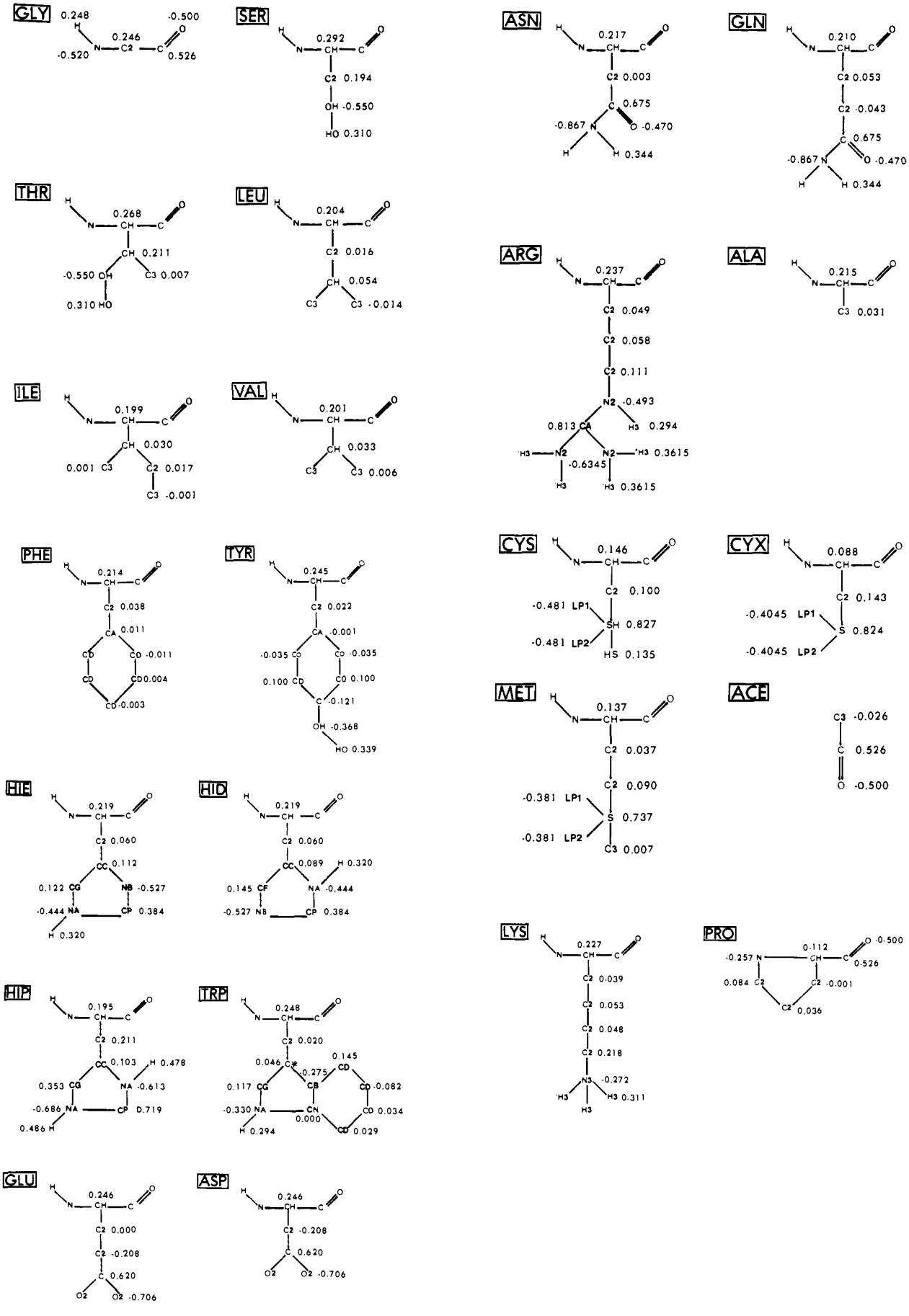


Figure 4. Net atomic charges for the amino acid side chains and backbone atoms in units of electron charge.

Table XI. Normal Modes Analysis

assignment	exptl ν^b	calcd ν^c
Benzene[d]		
Out of Plane		
E_{2u}	410	415
B_{2g}	703	703
In Plane		
E_{2g}	606	616
A_{1g}	992	949
B_{1u}	1010	972
E_{1u}	1038/1486	1182
B_{2u}	1310	1596
E_{2g}	1596	1544
N-methylacetamide[e]		
A" Out of Plane		
peptide torsion	192	194
C=O out of plane wag	600	598
N–H out of plane wag	725	718
A' In Plane		
C–N–CH_3 bend	289	315
CH_3 bend	439	452
amide IV	628	591
CH_3–C stretch	883	838
C–N stretch	1120	1023
amide III	1300	1295
amide II	1569	1588
amide I	1660	1667
N–H stretch	3306	3304
Methanol[f]		
C–O–H bend	1033	1040
C–O stretch	1345	1300
O–H stretch	3681	3709
Methanethiol[g]		
C–S–H bend	708	701
C–S stretch	803	813
S–H stretch	2573	2571
Dimethyl Sulfide[h]		
C–S–C bend	282	284
C–S stretch	691	705
C–S stretch	741	733
Dimethyl Disulfide[i]		
C–S–S–C torsion	102	104
C–S–S bend	239	234
C–S–S bend	272	272
S–S stretch	509	509
C–S stretch	689	718

[a] Relative assignment number. [b] In cm^{-1}. [c] Calculated with the second derivative routine in AMBER, in cm^{-1}. [d] Reference 54. [e] Reference 69. [f] Reference 54. [g] Reference 70. [h] Reference 70. [i] Reference 70.

modes were most coupled to hydrogen motions. The remaining non-hydrogen modes were the ones used in our analysis of NMA.

A united-atom model for NMA possesses six atomic centers which generate three out of plane and nine in plane motions. The out of plane bending modes are highly dependent upon the improper torsional parameters. With X–X–N–H = 1.0 kcal/mol and X–X–C–O = 10.5 kcal/mol, all three out of plane experimental frequencies were fit with an average error less than 4.0 cm^{-1} (see Table XI).

The highest frequency normal mode is almost entirely due to N–H stretching and was calculated by adjusting its respective force constant. The next two high-energy modes, amide I and amide II, are strongly mixed with carbonyl stretching and were derived accordingly, giving us C=O K_r = 570 kcal/mol Å^2. Since the remaining in plane modes are highly coupled to each other, the six bond-bending force constants were varied in a cyclic fashion until a reasonable fit with experiment was achieved (see Table XI). From our results obtained on glycyl dipeptide, alanyl dipeptide, and *N*-methylacetamide, we now possessed a complete parameter set for the amide linkage segment of the protein backbone.

Charge Derivation. At this point in the force-field development we possessed a reasonable set of charges for the atoms of the peptide *backbone*. The next step was to generate a set of electrostatic potential derived charges for the side chains of peptides. The protein residues were broken down into two structural units, the *bridge* containing both the α and β carbons (where applicable), and the *chromophore* possessing the remaining side chain atoms. Charges were computed at the STO-3G level for representative molecules of hydrogen-bonding peptide side chains. For example, phenol and imidazole were used as the chromophores for tyrosine and histidine, respectively. Again, due to basis set dependence of the calculated charges, we wished to represent the charge distribution in a manner consistent with the results of Cox and Williams.[26] Their study found that the STO-3G basis set derived charges could be best fit to the "optimal" 6-31G** by a ratio of 1.12:1. Due to this fact, coupled along with the 0.91 scaling needed to bring the 6-31G** charges in line with experimental dipoles, we decided upon a scaling factor of 1.0 (1.12 × 0.91) for all STO-3G derived charges.

The AMBER protein data base consists of the 20 amino acids, plus histidine protonated at both the δ and ϵ positions and a special residue for forming disulfide linkages. Each α and β carbon, or bridge atom, can be thought of as a buffer for absorbing the remaining charge distribution from the chromophoric and backbone segments, needed to achieve neutrality or an ionic state. The overall charge for the backbone atoms is −0.246, while the chromophores are neutral, singly protonated, or singly anionic. This excess charge was ratioed between the two carbon atoms by the same proportion as existed in our previous force field,[16,17] where the charges came from Mulliken populations. The rationale for piecing together segments of molecules to form an overall partial charge distribution for larger molecular units appears above. For the special case of hydrocarbon side chains where hydrogen bonding is not important (e.g., alanine, valine, leucine, isoleucine, and phenylalanine), no chromophore was used and the 0.246 was ratioed in direct accord with the Mulliken populations found in the previous force field.

As a final test for our charge selection, we examined water/chromophore hydrogen-bonding systems for bonding energy, overall structure, and hydrogen-bonding distance. All molecular mechanical calculations were carried out with a distance-dependent dielectric, $\epsilon = R_{ij}$. The water charges were chosen to give us reasonable results for stabilization energy and H···O distance for the H_2O dimer system. Using the partial charges oxygen = −0.66 and hydrogen = 0.33, we calculated a water dimer H···O distance of 1.79 Å and ΔE = −6.4 kcal/mol, in qualitative agreement with experimental values.[72]

Next, we employed these water charges, with the electrostatic potential derived ones from the chromophores for the remaining dimer calculations (Table XII). The 10–12 hydrogen-bonding term in our potential energy function allowed us another degree of freedom for "fine tuning" the hydrogen-bond distance. We found that a standard well depth of 0.5 kcal/mol and an equilibrium distance of 1.95 Å produced hydrogen/heteroatom distances of the magnitude 1.80 Å (see Table XII). However, r_{eq} = 1.95 Å produced a too long and not strong enough hydrogen bond for ammonium/water (which was selected as a model for our charged chromophores). We found that a more reasonable distance, 1.66 Å, could be achieved with a r_{eq} = 1.85 Å and employed this value for all cationic and anionic/water interactions.

Finally, we addressed the more common hydrogen-bonding situation in protein environments, with carbonyl acting as the proton acceptor and N–H as the donor. For the model system *N*-methylacetamide dimer, we calculated an *H*-bond distance of 1.82 Å and a stabilization energy ΔE = −7.0 kcal/mol. As is shown in Table XII, rather similar H-bond energies are achieved, with the electrostatic potential derived charges, for both calcu-

(72) Dyke, T.; Muenter, J. *J. Chem. Phys.* **1974**, *60*, 2929. Curtiss, L.; Frurip, D.; Blander, M. *Ibid.* **1979**, *71*, 2703.
(73) Johansson, A.; Kollman, P.; Rothenberg, S.; McKelvey, J. *J. Am. Chem. Soc.* **1974**, *96*, 3794.

Table XII. Water/Chromophore Hydrogen-Bonding Distances and Energies

chromophore	H_2O as proton acceptor distance[a]	ΔE[b]	H_2O as proton donor distance[a]	ΔE[b]
	$\epsilon = R_{ij}$			
NMA[c]	1.82	−5.8	1.79	−7.1
methanol[d]	1.80	−5.8	1.80	−6.0
imidazole[e]	1.78	−7.6	1.81	−6.5
phenol[f]	1.77	−7.5	1.85	−4.5
imidazolium[g]	1.72	−14.0		
methylguanidinium[g]	1.68	−11.8		
methylammonium[h]	1.66	−14.0		
indole[i]	1.79	−6.9		
water[j]	1.79	−6.4		
acetate[k]			1.73	−19.8
	$\epsilon = 1$			
NMA	1.87	−5.2	1.83	−6.7
methanol	1.85	−4.6	1.86	−5.0
imidazole	1.83	−6.7	1.87	−5.8
phenol	1.83	−6.1	1.89	−4.1
imidazolium	1.77	−15.3		
methylguanidinium	1.73	−13.5		
methylammonium	1.70	−17.0		
indole	1.84	−5.8		
water	1.84	−5.1		
acetate			1.80	−19.6

[a] In Å. [b] In kcal/mol. [c] Ab initio 4-31G calculations (ref 73) find formamide···HOH to give a $\Delta E = -9.2$ kcal/mol and formamide···OH_2 to give a $\Delta E = -6.8$ kcal/mol. Given the usual overestimation of H-bond energies by 4-31G, these ab initio energies are probably upper bounds. [d] References 74 and 75. [e] Del Bene finds $\Delta E = -5.6$ kcal/mol for Im···HOH and a $\Delta E = -9.0$ kcal/mol for Im···OH_2 (ref 76). [f] No experimental data but should be a better proton donor than H_2O and a worse proton acceptor. [g] No experimental data but the right order of magnitude (ref 77). [h] Kebarle suggests $\Delta H \approx -16.0$ kcal/mol (ref 77). [i] No experimental data but ΔE should be similar to imidazole. [j] Experimental data (ref 78) suggests a $\Delta E = -5.5$ kcal/mol. [k] Very accurate 6-31G**/MP2 calculations (ref 79) on $HCOO^-\cdots H_2O$ suggest a $\Delta E = -21$ kcal/mol.

lations where $\epsilon = 1$ and $\epsilon = R_{ij}$. We used STO-3G basis set derived charges for all chromophores with the exception of two special cases where slight alterations in the charges were necessary to give good agreement with quantum mechanical calculations.

Our first special case involves the methanol/water hydrogen-bonding model, which we used as our representation of both serine and threonine hydrogen-bonding interactions. This system shows a direct basis set dependence upon overall stabilization energies for the two possible structures $CH_3OH\cdots OH_2$ and $HOH\cdots OHCH_3$. Work done by Del Bene[74] using an STO-3G basis set revealed that methanol as a proton donor was energetically more stable than water, while theoretical results by Tse et al.,[75] employing a 6-31G* basis set, showed methanol to be a better proton acceptor by 0.20 kcal/mol. To be consistent within our framework of using data from theoretical calculations employing "optimal" basis sets (6-31G*, 6-31G**), our goal was to best reproduce the relative energies found by Tse et al. We found that good agreement with quantum mechanical theory could be achieved by adding −0.058 unit of charge to the oxygen, while placing the residual on the carbon, for an overall distribution of O = −0.550, H = 0.310, and C = 0.240. Aside from a reasonable correlation with theory, the altered charges imply a dipole moment of 1.81 D for methanol, as compared with the experimental value of 1.70 D.

The second special case we considered was the energy and geometry of hydrogen-bonded systems involving sulfur, which are of relevance in developing charges for methionine, cysteine, and cystine. Our first set of model calculations were on the H-bonded complexes $HSH\cdots OH_2$ and $H_2S\cdots HOH$. Specifically, we were interested in the angle above the plane (AAP) formed between the bisector of the sulfur hydrogens and the vector formed with the water hydrogen (involved in the H-bond) and the sulfur in the $H_2S\cdots HOH$ complex. Theoretical calculations by Kollman et al.[80] suggest that dimers formed by second-row hydrides and H_2O possess much greater AAP than the corresponding first-row hydrides. Reoptimizing this $H_2S\cdots HOH$ complex with a 4-31G basis set, we calculated an AAP of 78° and $\Delta E_{stabilization} = -3.88$ kcal/mol for the $H_2S\cdots HOH$ "linear" structure. With atom-centered partial charges, the AMBER local minimum corresponds to a nearly "bifurcated" structure possessing an AAP of 15°. With a 4-31G basis set and the AMBER geometry, we calculated an energy 1.40 kcal/mol higher than the lowest energetic quantum mechanical structure. To place these values for second-row hydride electron donors into a proper perspective, we should not that Umeyama and Morokuma[81] found *only* a 0.5 kcal/mol difference in energy, $\Delta E_{stabilization} = -7.8$ kcal/mol, for $H_2O\cdots HOH$ with AAP = 45°, compared to the geometry possessing an AAP = 0°. We concluded that to achieve reasonable qualitative agreement with the quantum mechanically calculated structures for second-row elements, it was necessary to place explicit lone pairs on all sulfur atom types in our new force field. Our goal was to best fit the structures and energies of the water/H_2S system to theoretical calculations carried out with a 4-31G basis set and then to extrapolate these results to determine charges for sulfur-containing amino acids.

The quantum mechanical electrostatic potential was calculated for the five atom centered model at the STO-3G level. To keep the lone pairs from "fusing" into the sulfur, we froze the sulfur charge and optimized the lone-pair distances using a four-point model. In this case, we found we had to vary both the charges and the 10–12 parameters to achieve optimal agreement with quantum mechanical theory. Our final molecular mechanical calculated structure had an AAP of 64° and possessed a quantum mechanical energy only 0.16 kcal/mol above the 4-31G optimized structure.

Proceeding with the concept of explicitly including lone pairs on sulfur atoms, we selected methyl sulfide as our charge model for cysteine residues. An initial charge set was generated by fitting the quantum mechanical electrostatic potential at the STO-3G level. Our degrees of freedom consisted of the 10–12 H-bonded parameters, the partial charges and the LP–S–LP angle. We found that with $R^* = 3.0$ Å and $\epsilon = 0.1$ kcal/mol and by empirically placing the lone pairs at ϑ(LP–S–LP) = 160° about the sulfur we calculated an AAP of 79° and an O···S distance of 3.37 Å for $CH_3HS\cdots HOH$. Consistent with the Kollman et al.[80] results on hydrogen sulfide, we felt it necessary to fit the relative energies for H_2S as a proton donor and acceptor. To achieve very nearly isoenergetic states for this system, we varied the charges by adding 0.117 to hydrogen, −0.100 to the lone pairs, 0.090 on the sulfur, and finally −0.007 to the carbon for neutrality. The final results gave $\Delta E_{stabilization} = -3.1$ kcal/mol for $CH_3HS\cdots HOH$ and $\Delta E_{stabilization} = -3.2$ kcal/mol for $CH_3SH\cdots OH_2$. These methyl sulfide charges are actually quite reasonable as they bring the dipole moment into a better agreement with the experimental value of 1.51 D. The new charges lead to a dipole moment of 1.82 D compared with 1.02 D if we used those directly fit to the electrostatic potential. Hence, we have shown that empirically placing the lone pairs at 160° about the sulfur gives the proper directionality needed to be consistent with quantum mechanical results.

Insulin Refinement. Thus far, we have parameterized the force field for small molecular subunits with the ultimate goal of applying the new parameters, with both generality and transferability,

(74) Del Bene, J. *J. Chem. Phys.* **1971**, *55*, 4633.
(75) Tse, Y. C.; Newton, M.; Allen, L. *Chem. Phys. Lett.* **1980**, *76*, 350.
(76) Del Bene, J. *J. Am. Chem. Soc.* **1978**, *100*, 5285.
(77) Kebarle, P. In "Environmental Effects on Molecular Structure and Properties"; Pullman, B., Ed.; Reidel: Dordrecht, Holland, 1976; p 81.
(78) Curtis, L.; Frurip, D.; Blander, M. *J. Am. Chem. Soc.* **1978**, *71*, 2703.
(79) Alagona, G.; Ghio, C.; Kollman, P. *J. Am. Chem. Soc.* **1983**, *105*, 5226.
(80) Kollman, P.; McKelvey, J.; Johansson, A.; Rothenberg, S. *J. Am. Chem. Soc.* **1975**, *97*, 955.
(81) Umeyama, H.; Morokuma, K. *J. Am. Chem. Soc.* **1977**, *99*, 1316.

Table XIII. Refinement of Insulin

	energy evaluations	rms gradient[a]	rms backbone[b]	rms all atoms[c]	compaction/expansion,[d] %
cutoff[e] = 9.0 Å, $\epsilon = R_{ij}$	1318	0.14	0.28	0.43	−1.6
cutoff = 9.0 Å, $\epsilon = 4R_{ij}$	3642	0.30	0.51	0.66	−9.4
cutoff = 9.0 Å, $\epsilon = R_{ij}$ (Jorgensen nonbonded)[f]	1080	0.20	0.26	0.41	+0.2
cutoff = 9.0 Å, $\epsilon = R_{ij}$ (old parameter set)[g]	1100	0.06	0.79	1.01	−16.0
cutoff = 12.0 Å, $\epsilon = R_{ij}$	3937	0.09	0.56	0.72	−7.0

[a] Root mean square (rms) gradient, in units of kcal/Å, calculated at end of the energy refinement. [b] Root mean square fit in Å for the minimized insulin backbone atoms, compared with coordinates from the starting crystal structure. [c] Root mean square fit in Å for all the minimized insulin atoms, compared with coordinates from the starting crystal structure. [d] These values represent the ratio of the minimized volume to initial volume. The specific volumes were generated using the radius of gyration calculated from all insulin backbone atoms. [e] Cutoff is the distance from which all nonbonded interactions will be evaluated. [f] Reference 24. [g] This run incorporated our old parameter set (ref 17) except that we used the new hydrogen-bond 10-12 potentials. The 10-12 parameters were necessary to keep atoms from "fusing", which we found occurred in our initial run.

to larger systems of nucleic acids and proteins. Within the context of proteins, we have shown that the force field produced quite reasonable results for hydrogen-bonding structures and energies of water/chromophore interactions, local energy minima of alanyl and glycyl dipeptides, and vibrational frequencies of *N*-methylacetamide. As the first direct application of the new force field to an entire protein, we used our new parameter set to energy refine insulin by conjugate gradient optimization. We selected insulin for two reasons: first, its relatively small size, 500 atoms (404 crystallographically located heavy atoms and 96 added hydrogens), enabled us to refine the energy of the entire protein within a reasonable time frame; second, due to the high resolution of the crystal structure (resolved to 1.5 Å), any large-scale motions within the minimized structure, relative to the starting one, would reveal areas where reparameterization of our force field might be necessary. Specifically, we were interested in "quality" of the intramolecular hydrogen bonds, the overall degree of inward compactness of the entire minimized protein, and the general question of how a distance cutoff for evaluating the nonbonded interactions affects the final structure.

The starting coordinates for porcine insulin by Dodson et al.[82] were taken from the Brookhaven Protein Data Bank.[83] A subroutine within AMBER placed all potential H-bonding hydrogens on each respective nitrogen, oxygen, and sulfur with standard bond lengths and angles. Our first minimization employed a distance-dependent dielectric, $\epsilon = R_{ij}$, and a 9.0-Å cutoff distance for evaluating the nonbonded interactions. All nonbonded interactions between 8.0 and 9.0 Å were multiplied by a cubic equation whose value varies from 1 at 8.0 Å to 0 at 9.0 Å. Consistent with our rationale explained above, all 1–4 van der Waals and electrostatic interactions were scaled by 0.5 and the method of neutral spheres[21] was used.

After 1318 energy evaluations, the rms energy gradient reached 0.14 kcal/Å. An rms fit of the initial and final structures was 0.28 Å for the backbone atoms and 0.43 Å for all insulin atoms. Using the radius of gyration $((\sum_{i=1}^{n} R_i^2/n)^{1/2}$, where R_i is the distance of atom i to the center of mass and n is the number of atoms), we generated a sphere to represent the "volume" of insulin. Although many elaborate methods exist for generating the volume of a protein, a simple spherical representation will suffice to give a feel for the extent of inward compactness of insulin. In this specific case, we calculated an overall volume compaction of only 1.5%. This value is most likely an overestimation since our model does not explicitly include solvation, which would tend to "pull" solvent-facing side chains and backbone atoms outward from the center of mass.

Table XIV. Bond Parameters

bond	K_r	r_{eq}	bond	K_r	r_{eq}
C–C2	317	1.522	CC–C2	317	1.504
C–C3	317	1.522	CC–CF	512	1.375
C–CB	447	1.419	CC–CG	518	1.371
C–CD	469	1.40	CC–NA	422	1.385
C–CH	317	1.522	CC–NB	410	1.394
C–CJ	410	1.444	CD–CD	469	1.40
C–CM	410	1.444	CD–CN	469	1.40
C–CT	317	1.522	CE–N*	440	1.371
C–N	490	1.335	CE–NB	529	1.304
C–N*	424	1.383	CF–NB	410	1.394
C–NA	418	1.388	CG–C*	546	1.352
C–NC	457	1.358	CG–NA	427	1.381
C–O	570	1.229	CH–C2	260	1.526
C–O2	656	1.25	CH–C3	260	1.526
C–OH	450	1.364	CH–CH	260	1.526
C2–C*	317	1.495	CH–N	337	1.449
C2–C2	260	1.526	CH–N*	337	1.475
C2–C3	260	1.526	CH–NT	367	1.471
C2–N	337	1.449	CH–OH	386	1.425
C2–N2	337	1.463	CH–OS	320	1.425
C2–N3	367	1.471	CI–NC	502	1.324
C2–NT	367	1.471	CJ–CJ	549	1.350
C2–OH	386	1.425	CJ–CM	560	1.343
C2–OS	320	1.425	CJ–N*	448	1.365
C2–S	222	1.81	CN–NA	428	1.38
C2–SH	222	1.81	CP–NA	477	1.343
C3–CM	317	1.51	CP–NB	488	1.335
C3–N	337	1.449	CT–CT	310	1.526
C3–N*	337	1.475	CT–HC	331	1.09
C3–N2	337	1.463	CT–N	337	1.449
C3–N3	367	1.471	H–N	434	1.01
C3–OS	320	1.425	H–N2	434	1.01
C3–S	222	1.81	H–NA	434	1.01
CA–C2	317	1.51	H2–N	434	1.01
CA–CB	469	1.404	H2–N2	434	1.01
CA–CD	469	1.40	H2–NT	434	1.01
CA–CJ	427	1.433	H3–N2	434	1.01
CA–N2	481	1.340	H3–N3	434	1.01
CA–NA	427	1.381	HO–OH	553	0.96
CA–NC	483	1.339	HO–OS	553	0.96
CB–C*	388	1.459	HS–SH	274	1.336
CB–CB	520	1.370	LP–S	600	0.679
CB–CD	469	1.40	LP–SH	600	0.679
CB–CN	447	1.419	O2–P	525	1.48
CB–N*	436	1.374	OH–P	230	1.61
CB–NB	414	1.391	OS–P	230	1.61
CB–NC	461	1.354	S–S	166	2.038

The second test run used the identical criterion as our first with the exception that a nonbonded cutoff of 12.0 Å was used. It should be pointed out that since the calculation of the van der Waals interactions is the rate-limiting computational step in all energy minimizations, it is important to find the smallest nonbonded cutoff that still gives reasonable results. After 3937 energy evaluations the rms gradient had reached 0.09 kcal/Å. It appears, from calculations on insulin and our unpublished results on papain, that the overall number of energy evaluations and the rms gradient (after the system has reached a minimum by conjugate gradient techniques) are a function of nonbonded cutoff. We are currently addressing this question as a means of assessing the best methodology to undertake for subsequent energy refinements of proteins and nucleic acids.

We carried out five different refinements on insulin with the results appearing in Table XIII. It is interesting to note that nearly identical results are attained whether we use Jorgensen's or our new nonbonded parameters for the united-atom carbons. Finally, the most encouraging result, from the standpoint of displaying the improvement of the new force field over the old one, appears for the refinement using our old parameter set (which possessed similar R^*'s but smaller ϵ's for the nonbonded terms,

(82) Dodson, G.; Dodson, E.; Hodgkin, D.; Reynolds, C. *Can. J. Biochem.* **1979**, *57*, 469.

(83) Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973.

Table XV. Angle Parameters

angle	K_ϑ	ϑ_{eq}	angle	K_ϑ	ϑ_{eq}	angle	K_ϑ	ϑ_{eq}	angle	K_ϑ	ϑ_{eq}
C*-CG-NA	70.0	108.7	C2-S-C3	62.0	98.9	CD-C-OH	70.0	120.0	CN-CB-C*	85.0	108.8
C-C2-C2	63.0	112.4	C2-S-S	68.0	103.7	CD-CA-C2	70.0	120.0	CN-NA-CG	70.0	111.6
C-C2-CH	63.0	112.4	C2-SH-HS	44.0	96.0	CD-CA-CD	85.0	120.0	CN-NA-H	35.0	124.2
C-C2-N	80.0	110.3	C3-C-N	70.0	116.6	CD-CB-C*	85.0	134.9	CP-NA-CG	70.0	107.3
C-CB-CB	85.0	119.2	C3-C-O	80.0	120.4	CD-CB-CN	85.0	116.2	CP-NA-H	35.0	126.35
C-CB-NB	70.0	130.0	C3-C-O2	70.0	117.0	CD-CD-C	85.0	120.0	CT-C-N	70.0	116.6
C-CH-C2	63.0	111.1	C3-C2-OS	80.0	109.5	CD-CD-CD	85.0	120.0	CT-C-O	80.0	120.4
C-CH-C3	63.0	111.1	C3-CH-C3	63.0	111.5	CD-CD-CN	85.0	120.0	CT-N-H	38.0	118.4
C-CH-CH	63.0	111.1	C3-CH-N	80.0	109.5	CD-CN-NA	70.0	132.8	H-N-H	35.0	120.0
C-CH-N	63.0	110.1	C3-CH-OH	80.0	109.5	CE-N*-C3	70.0	128.8	H2-N2-H2	35.0	120.0
C-CJ-CJ	85.0	120.7	C3-N-H	38.0	118.4	CE-N*-CH	70.0	128.8	H3-N2-H3	35.0	120.0
C-CM-C3	85.0	119.7	C3-N2-H2	35.0	118.4	CF-CC-NA	70.0	105.9	H2-NT-H2	35.0	109.5
C-CM-CJ	85.0	120.7	C3-N3-H3	35.0	109.5	CF-NB-CP	70.0	105.3	H3-N-H3	35.0	120.0
C-N*-CH	70.0	117.6	C3-OH-HO	55.0	108.5	CG-C*-C2	70.0	125.0	H3-N3-H3	35.0	109.5
C-N*-CJ	70.0	121.6	C3-OS-P	100.0	120.5	CG-CC-NA	70.0	108.75	HO-OH-HO	47.0	104.5
C-N-C2	50.0	121.9	C3-S-S	68.0	103.7	CG-CC-NB	70.0	109.9	HO-OH-P	45.0	108.5
C-N-C3	50.0	121.9	C3-SH-HS	44.0	96.0	CG-NA-H	35.0	126.35	LP-S-C2	600.0	96.7
C-N-CH	50.0	121.9	CA-CB-CB	85.0	117.3	CH-C-N	70.0	116.6	LP-S-C3	600.0	96.7
C-N-CT	50.0	121.9	CA-CB-NB	70.0	132.4	CH-C-O	80.0	120.4	LP-S-LP	600.0	160.0
C-N-H	35.0	119.8	CA-CD-CD	85.0	120.0	CH-C-O2	65.0	117.0	LP-S-S	600.0	96.7
C-N-H2	35.0	120.0	CA-CJ-CJ	85.0	117.0	CH-C-OH	70.0	115.0	LP-SH-C2	600.0	96.7
C-NA-C	70.0	126.4	CA-N2-C2	50.0	123.2	CH-C2-C*	63.0	115.6	LP-SH-HS	600.0	96.7
C-NA-CA	70.0	125.2	CA-N2-C3	50.0	123.2	CH-C2-C2	63.0	112.4	LP-SH-LP	600.0	160.0
C-NA-H	35.0	116.8	CA-N2-H	35.0	120.0	CH-C2-C3	63.0	112.4	N*-C-O	80.0	120.9
C-NC-CA	70.0	120.5	CA-N2-H2	35.0	120.0	CH-C2-CA	63.0	114.0	N*-CE-NB	70.0	113.9
C-OH-HO	35.0	113.0	CA-N2-H3	35.0	120.0	CH-C2-CH	63.0	112.4	N*-CH-OS	80.0	109.5
C2-C-N	70.0	116.6	CA-NA-H	35.0	118.0	CH-C2-OH	80.0	109.5	N-C-O	80.0	122.9
C2-C-O	80.0	120.4	CA-NC-CB	70.0	112.2	CH-C2-OS	80.0	109.5	N-CH-C2	80.0	109.7
C2-C-O2	70.0	117.0	CA-NC-CI	70.0	118.6	CH-C2-S	50.0	114.7	N-CH-CH	80.0	109.7
C2-C2-C2	63.0	112.4	CB-C*-C2	70.0	128.6	CH-C2-SH	50.0	108.6	N2-CA-N2	70.0	120.0
C2-C2-N	80.0	111.2	CB-C*-CG	85.0	106.4	CH-CH-C2	63.0	111.5	NA-C-N*	70.0	115.4
C2-C2-N2	80.0	111.2	CB-C-NA	70.0	111.3	CH-CH-C3	63.0	111.5	NA-C-O	80.0	120.6
C2-C2-N3	80.0	111.2	CB-C-O	80.0	128.8	CH-CH-CH	63.0	111.5	NA-CA-N2	70.0	116.0
C2-C2-OS	80.0	109.5	CB-CA-N2	70.0	123.5	CH-CH-N*	80.0	109.5	NA-CA-NC	70.0	123.3
C2-C2-S	50.0	114.7	CB-CA-NC	70.0	117.3	CH-CH-OH	80.0	109.5	NA-CP-NA	70.0	110.75
C2-CC-CF	70.0	131.9	CB-CB-N*	70.0	106.2	CH-CH-OS	80.0	109.5	NB-CP-NA	70.0	111.6
C2-CC-CG	70.0	129.05	CB-CB-NB	70.0	110.4	CH-N-C2	50.0	118.0	NC-C-N*	70.0	118.6
C2-CC-NA	70.0	122.2	CB-CB-NC	70.0	127.7	CH-N-H	38.0	118.4	NC-C-O	80.0	122.5
C2-CC-NB	70.0	121.05	CB-CD-CD	85.0	120.0	CH-NT-H2	35.0	109.5	NC-CA-N2	70.0	119.8
C2-CH-C3	63.0	111.5	CB-CN-CD	85.0	122.7	CH-OH-HO	55.0	108.5	NC-CB-N*	70.0	126.0
C2-CH-N*	80.0	109.5	CB-CN-NA	70.0	104.4	CH-OS-CH	100.0	111.8	NC-CI-NC	70.0	129.1
C2-CH-OH	80.0	109.5	CB-N*-C3	70.0	125.8	CH-OS-HO	55.0	108.5	NT-C2-C	80.0	111.2
C2-CH-OS	80.0	109.5	CB-N*-CE	70.0	105.4	CH-OS-P	100.0	120.5	NT-C2-C2	80.0	111.2
C2-N-H	38.0	118.4	CB-N*-CH	70.0	125.8	CJ-C-NA	70.0	114.1	NT-CH-C	80.0	109.7
C2-N2-H2	35.0	118.4	CB-NB-CE	70.0	103.8	CJ-C-O	80.0	125.3	NT-CH-C2	80.0	109.7
C2-N2-H3	35.0	118.4	CB-NC-CI	70.0	111.0	CJ-CA-N2	70.0	120.1	NT-CH-CH	80.0	109.7
C2-N3-H3	35.0	109.5	CC-C2-CH	63.0	113.1	CJ-CA-NC	70.0	121.5	O2-C-O2	80.0	126.0
C2-NT-H2	35.0	109.5	CC-CF-NB	70.0	109.9	CJ-CJ-N*	70.0	121.2	O2-P-O2	140.0	119.9
C2-OH-HO	55.0	108.5	CC-CG-NA	70.0	105.9	CJ-CM-C3	85.0	119.7	OH-P-O2	45.0	108.2
C2-OS-C2	100.0	111.8	CC-NA-CP	70.0	107.3	CJ-N*-CH	70.0	121.2	OS-P-O2	100.0	108.2
C2-OS-C3	100.0	111.8	CC-NA-H	35.0	126.35	CM-C-NA	70.0	114.1	OS-P-OH	45.0	102.6
C2-OS-HO	55.0	108.5	CC-NB-CP	70.0	105.3	CM-C-O	80.0	125.3	OS-P-OS	45.0	102.6
C2-OS-P	100.0	120.5	CD-C-CD	85.0	120.0	CM-CJ-N*	70.0	121.2			

as compared to the new force field) where we calculated 16% compaction for the backbone atoms and 1.01 Å rms movement for all insulin atoms.

Hartmann et al.[84] have shown that there is a definite decrease in the volume of the crystal structure of myoglobin at 80 K compared with 300 K. Our insulin energy refinements are equivalent to simulations carried out at 0 K; hence, the compaction which we observe is consistent within this framework. However, in view of the lack of inclusion of explicit water molecules or crystal symmetry in our minimization, more precise comparisons are not possible at this time. The nature and extent of protein compaction expected from an energy refinement are still unclear, and we are analyzing this question as a means of obtaining a greater understanding of the forces involved in protein structure.

Discussion

The approach we have taken to developing the force field presented here has several unique aspects, but, in general, the various steps have precedent in the literature. The unique aspect in this case is that this is the first complete "consistent" force field developed for both proteins and nucleic acids. However, force fields are constantly evolving objects, and, although we may have reached a plateau, we anticipate further development in the future. Below we critically analyze the results of our study: intramolecular parameters (bond stretching, bending, and torsion), nonbonded parameters (Lennard-Jones), and electrostatic charges.

We have used spectroscopic structural data to determine the K_r, r_{eq}, K_ϑ, and ϑ_{eq} parameters for use in our basic energy expression (eq 1). After energy refinement of model systems, the calculated bond lengths, because of the generally large K_r values and the fact that our molecules are relatively unstrained, remain very near r_{eq}. This is also true for the bond angles for noncyclic parts of the structure. We note that our method of interpolating K_r for sp^2 atoms in planar rings, possessing bond lengths between pure single and double bond values, works adequately.

In the furanose ring of tetrahydrofuran, it is clear that the value of K_ϑ's derived from fit to energies and structure (40–50 kcal/mol rad^2 here and in the MM2 force field) are significantly smaller

(84) Hartmann, H.; Parak, F.; Steigemann, W.; Petsko, G.; Ponzi, D.; Frauenfelder, H. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4967.

Table XVI. Torsional Parameters

torsion	$V_n/2$	γ	n	torsion	$V_n/2$	γ	n
X-C*-C2-X	0.0	0	2	C2-C2-S-LP	0.0	0	3
X-C*-CB-X	2.4	180	2	C2-OS-C2-C2	0.10	0	2
X-C*-CG-X	23.6	180	2	C2-OS-C2-C2	1.45	0	3
X-C-CB-X	4.4	180	2	C2-OS-C2-C3	0.1	0	2
X-C-CD-X	5.3	180	2	C2-OS-C2-C3	0.725	0	3
X-C-CH-X	0.0	0	2	C2-OS-CH-C2	0.1	0	2
X-C-CJ-X	3.1	180	2	C2-OS-CH-C2	0.725	0	3
X-C-CM-X	3.1	180	2	C2-OS-CH-C3	0.1	0	2
X-C-CT-X	0.0	0	2	C2-OS-CH-C3	0.725	0	3
X-C-N*-X	5.8	180	2	C3-OS-C2-C3	0.10	0	2
X-C-N-X	10.0	180	2	C3-OS-C2-C3	1.45	0	3
X-C-NA-X	5.4	180	2	C3-OS-CH-C3	0.1	0	2
X-C-NC-X	8.0	180	2	C3-OS-CH-C3	0.725	0	3
X-C-OH-X	1.8	180	2	C2-S-S-C2	0.6	0	3
X-C2-C-X	0.0	180	3	C2-S-S-C2	3.5	0	2
X-C2-C2-X	2.0	0	3	CH-C2-SH-LP	0.0	0	3
X-C2-CC-X	0.0	0	2	CH-OS-CH-C2	0.1	0	2
X-C2-N-X	0.0	0	3	CH-OS-CH-C2	0.725	0	3
X-C2-N2-X	0.0	0	3	CH-OS-CH-CH	0.1	0	2
X-C2-N3-X	1.4	0	3	CH-OS-CH-CH	0.725	0	3
X-C2-OH-X	0.5	0	3	CH-OS-CH-N*	0.0	0	2
X-C2-OS-X	1.45	0	3	CH-OS-CH-N*	0.725	0	3
X-C2-S-X	1.0	0	3	CT-CT-C-O	0.067	180	3
X-C2-SH-X	0.75	0	3	H-N-C-O	0.65	0	1
X-CA-C2-X	0.0	0	2	H-N-C-O	2.5	180	2
X-CA-CB-X	5.1	180	2	HC-CT-C-O	0.067	180	3
X-CA-CD-X	5.3	180	2	LP-S-S-C2	0.0	0	3
X-CA-CJ-X	3.7	180	2	LP-S-S-LP	0.0	0	3
X-CA-N2-X	6.8	180	2	N-CT-C-O	0.067	180	3
X-CA-NA-X	6.0	180	2	O-C-C2-N	0.2	180	3
X-CA-NC-X	9.6	180	2	O-C-CH-C2	0.1	180	3
X-CB-CB-X	16.3	180	2	O-C-CH-CH	0.1	180	3
X-CB-CD-X	5.3	180	2	O-C-CH-N	0.1	180	3
X-CB-CN-X	4.4	180	2	OH-C2-C2-OH	0.5	0	2
X-CB-N*-X	6.6	180	2	OH-C2-C2-OH	2.0	0	3
X-CB-NB-X	5.1	180	2	OH-C2-CH-OH	0.5	0	2
X-CB-NC-X	8.3	180	2	OH-C2-CH-OH	1.0	0	3
X-CC-CF-X	14.3	180	2	OH-CH-CH-OH	0.5	0	3
X-CC-CG-X	15.9	180	2	OH-CH-CH-OH	0.5	0	2
X-CC-NA-X	5.6	180	2	OH-P-OS-C2	0.25	0	3
X-CC-NB-X	4.8	180	2	OH-P-OS-C2	0.75	0	2
X-CD-CD-X	5.3	180	2	OH-P-OS-C3	0.25	0	3
X-CD-CN-X	5.3	180	2	OH-P-OS-C3	0.75	0	2
X-CE-NB-X	20.0	180	2	OH-P-OS-CH	0.25	0	3
X-CF-NB-X	4.8	180	2	OH-P-OS-CH	0.75	0	2
X-CG-NA-X	6.0	180	2	OS-C2-C2-OH	0.5	0	2
X-CH-C2-X	2.0	0	3	OS-C2-C2-OH	2.0	0	3
X-CH-CH-X	2.0	0	3	OS-C2-C2-OS	0.5	0	2
X-CH-N*-X	0.0	0	2	OS-C2-C2-OS	2.0	0	3
X-CH-N-X	0.0	0	3	OS-C2-CH-OH	0.5	0	2
X-CH-OH-X	0.5	0	3	OS-C2-CH-OH	1.0	0	3
X-CH-OS-X	1.45	0	3	OS-C2-CH-OS	0.5	0	2
X-CI-NC-X	13.5	180	2	OS-C2-CH-OS	1.0	0	3
X-CJ-CJ-X	24.4	180	2	OS-CH-C2-OH	0.5	0	2
X-CJ-CM-X	27.2	180	2	OS-CH-C2-OH	1.0	0	3
X-CN-NA-X	6.1	180	2	OS-CH-CH-OH	0.5	0	3
X-CP-NA-X	9.3	180	2	OS-CH-CH-OH	0.5	0	2
X-CP-NB-X	10.0	180	2	OS-CH-CH-OS	0.5	0	3
X-CT-CT-X	1.6	0	3	OS-CH-CH-OS	0.5	0	2
X-CT-N-X	0.0	0	3	OS-P-OS-C2	0.25	0	3
X-N*-CE-X	6.7	180	2	OS-P-OS-C2	0.75	0	2
X-N*-CJ-X	7.4	180	2	OS-P-OS-C3	0.25	0	3
X-NT-C2-X	1.0	0	3	OS-P-OS-C3	0.75	0	2
X-OH-P-X	0.75	0	3	OS-P-OS-CH	0.25	0	3
X-OS-P-X	0.75	0	3	OS-P-OS-CH	0.75	0	2

than the values (60–100 kcal/mol rad^2) needed to reproduce angle bending frequencies in both cyclic and noncyclic systems with an sp^3 carbon or oxygen at the apex of the bond. In order to reasonably fit frequencies, energy, and the structure of THF, we needed to increase the torsional parameters V_n for rotation around the C–O and C–C bond. When we critically compare the results of calculations using these two choices of K_ϑ and V_n in fitting the structures, energies, and frequencies of THF and MEE, we find little to choose from, with both approaches giving fair agreement to experiment. However, since we have used vibrational frequencies of amides and other model systems to determine many K_ϑ and K_τ we slightly favor the parameter set containing the larger K_ϑ, V_n values. It should be emphasized, however, that it is a simple

Table XVII. Improper Torsional Parameters

torsion	$V_n/2$	γ	n
X-X-C-O	10.5	180.0	2
X-X-N-H	1.0	180.0	2
X-X-NA-H	1.0	180.0	2
X-C2-CH-X	14.0	180.0	3
X-CH-CH-X	14.0	180.0	3
X-CH-N-C2	1.0	180.0	2
X-CH-N-C	14.0	180.0	3
X-H2-N-H2	1.0	180.0	2
X-N2-CA-N2	10.5	180.0	2
X-O2-C-O2	10.5	180.0	2
C2-CH-C-N3	7.0	180.0	3
C3-CH-CA-C3	7.0	180.0	3
CH-CH-C-N3	7.0	180.0	3

Table XVIII. Hydrogen-Bond Parameters

acceptor	donor	C	D
H	NB	7557	2385
H	NC	10238	3071
H	O2	4019	1409
H	O	7557	2385
H	OH	7557	2385
H	S	265720	35429
H	SH	265720	35429
HO	NB	7557	2385
HO	O2	4019	1409
HO	O	7557	2385
HO	OH	7557	2385
HO	S	265720	35429
HO	SH	265720	35429
H2	NB	4019	1409
H2	O2	4019	1409
H2	O	10238	3071
H2	OH	4019	1409
H2	S	265720	35429
H2	SH	265720	35429
H3	NB	4019	1409
H3	O2	4019	1409
H3	O	4019	1409
H3	OH	4019	1409
H3	S	7557	2385
H3	SH	7557	2385
HS	NB	14184	3082
HS	O2	4019	1409
HS	O	14184	3082
HS	OH	14184	3082
HS	S	265720	35429
HS	SH	265720	35429

Table XIX. Nonbonded Parameters

atom	R^*	ϵ	atom	R^*	ϵ
C	1.85	0.12	H3	1.00	0.02
C*	1.85	0.12	HC	1.375	0.038
C2	1.925	0.12	HO	1.00	0.02
C3	2.00	0.15	HS	1.00	0.02
CA	1.85	0.12	LP	1.20	0.016
CB	1.85	0.12	N	1.75	0.16
CC	1.85	0.12	N*	1.75	0.16
CD	1.85	0.12	N2	1.75	0.16
CE	1.85	0.12	N3	1.85	0.08
CF	1.85	0.12	NA	1.75	0.16
CG	1.85	0.12	NB	1.75	0.16
CH	1.85	0.09	NC	1.75	0.16
CI	1.85	0.12	NT	1.85	0.12
CJ	1.85	0.12	O	1.60	0.20
CM	1.85	0.12	O2	1.60	0.20
CN	1.85	0.12	OH	1.65	0.15
CP	1.85	0.12	OS	1.65	0.15
CT	1.80	0.06	P	2.10	0.20
H	1.00	0.02	S	2.00	0.20
H2	1.00	0.02	SH	2.00	0.20

matter to employ the smaller parameters by using the set noted in Table III (FF1).

The torsional parameters we have used come mainly from literature values for proteins and our previous nucleic acid force field. Our use of V_2 values for O–P–O–C, O–C–C–O, and O–C–C–C groups allows us to "fine tune" conformational energy differences. We have used an interpolation method for V_2 values for partially double bonded sp^2 atom–sp^2 atom parameters which appears to give good agreement with experimental nucleic acid base frequencies.[56] A purely empirical fit of the out of plane "improper" V_2 torsional parameters for the C=O and N—H groups in amides has resulted in an excellent correlation to the out of plane experimental frequencies.

There are two straightforward ways in which these intramolecular aspects of our force field can be improved: the first is simply to abandon the united-atom description of C–H groups, and the second is to use a more elaborate description of the intramolecular energy, including many more coupling and anharmonic terms in the energy expression. Our comparison of all-atom vibrational calculations with the united-atom ones suggests that we can adequately represent low frequency ($<700\ cm^{-1}$) modes. However, extensive coupling between various hydrogen-including modes in the 800–1500-cm^{-1} range precludes a more precise representation of these modes in this study. The use of a more elaborate energy function for intramolecular interactions has much precedence in the literature,[44] and, if our main goal was a more precise description of all the virbational properties of macromolecules, we should certainly switch immediately to a more complex function. However, we have evidence that the low-frequency vibrations that contribute most to thermodynamic properties are relatively insensitive to the inclusion of anharmonic bond/angle and coupling terms.[85] In addition, the analysis of the vibrational spectra and its fit to many more empirical parameters would need to be done for the large number of atom types in proteins and nucleic acids. This is a major undertaking and worth doing (we are currently engaged in such an analysis, still using eq 1, for the four nucleic acid bases),[86] but the other issues, discussed below, are much more critical for the development and understanding of the structures and energies of a macromolecular system.

One of the most difficult problems in the development of this force field was the choice of nonbonded parameters and the way to handle them for 1–4 interactions. For sp^2 atoms, we used values very similar to those of Hagler et al.,[19] and our model calculations on 1,3-dimethyluracil stacking (Table III) suggest that our stacking energy values are very similar to those calculated by using the actual Hagler et al. parameters. The base–base separation, interestingly enough, is somewhat larger (3.71 Å) with the Hagler et al. values than with ours (3.43 Å), but this likely reflects the fact that we explicitly include the C–H hydrogens in the former case. However, this difference in structure may be significant in its relevance to our choice of nonbonded parameters for sp^3 atoms. Use of nonbonded parameters, as have been determined to give good crystal packing parameters and energies for hydrocarbons by Dunfield et al.,[24] and a very similar set of values which gives good agreement with liquid-state energies and densities for ethers,[25] results in a much poorer representation of sugar conformational properties than if smaller values are used in our force field. Above, we have argued why the use of a scale factor (completely empirical) for the 1–4 nonbonded interactions makes some sense, particularly because a 6–12 function would be too repulsive for shorter nonbonded interactions, compared to the more realistic 6-exponential form.

A 1–4 nonbonded scale factor also allows us to use van der Waals radii for CH, C2, and C3 atoms somewhat closer to Jorgensen's.[25] Above we have shown that such parameters for C3 lead to 10% errors in energy and density in Monte Carlo simulations of DME, compared to errors of 3–5% found by Jorgensen. However, since proteins and nucleic acids contain a considerable fraction of sp^2 atoms, our van der Waals parameters for CH, C2, and C3 should not lead to very large errors. The refinement of insulin suggests that our nonbonded parameters are reasonable.

It is possible that all these "problems" could be simply solved by abandoning the united-atom representation for CH, C2, and C3 (and corresponding sp^2) groups. However, one should not do this "lightly", given the sensitivity of simulations on large molecules to nonbonded cutoffs (above discussions on insulin and our unpublished refinements of papain) and the fact that nonbonded function evaluation is the rate-limiting step in such simulations. (C–H hydrogens can make up to 50% of the atoms in proteins and 30% for nucleic acids.) However, such a step may be appropriate in some cases (our simulations on thyroxine and its protein binding),[17] and we have almost completed extending our current force field in such a way, since our calculations for partial charges have been done both in the united- and all-atom representations. Switching the nonbonded parameters to use a 6-exponential rather than 6–12 functional form is simple enough to do, but also would slow the most time consuming part of the calculation. However, as computing power increases, this also seems a likely refinement for the near future.

One of the most useful results from this study has been the generalization of the initial studies by Momany et al.,[87] Smit et al.,[88] and Cox and Williams[26] to using electrostatic potentials for determining the appropriate atomic partial charges used in evaluating the electrostatic term in eq 1. Together with 10–12 parameters, these charges lead to H-bond energies and structures in reasonable agreement with available ab initio calculations and experiments. The two cases where the charges were altered from the electrostatic potential determined values are instructive in this regard: polarization effects clearly play a role in whether $CH_3OH\cdots OH_2$ or $HOH\cdots OHCH_3$ is the lower energy structure, and these effects are also of greater importance for the hydrogen bonding involving sulfur, rather than oxygen, as an electron donor. Thus, our simple representation of H-bonding in eq 1 is clearly an area for future improvement of the force field. Ultimately, a more complete description of the charge distribution and its effect on hydrogen bonding can be considered by a modification of eq 1 (this is subject to the time consuming nature of evaluating nonbonded interactions).

We have developed this force field using a distance-dependent dielectric constant ($\epsilon = R_{ij}$) since such an approach is a way to qualitatively simulate the fact that the system is in water and the intramolecular electrostatic interactions should die off more rapidly with distance than in the gas phase. However, we have shown for H-bonding in nucleic acid and protein models that the use of $\epsilon = 1$ gives quite similar H-bond energies and structures as that with $\epsilon = R_{ij}$. Thus, our force field may also be well suited to simulations with explicit inclusion of water, and we are currently developing approaches to do just that. However, we emphasize that the use of larger distance dependence ($\epsilon = 4R_{ij}$) was required to reproduce the relative C2′ endo/C3′ endo energy in adenosine, where the intramolecular electrostatic effects would be expected to be much more damped in solution than, for example, stacked base pairs. Also, the decision (purely empirical) to scale the 1–4 electrostatic interactions might be alleviated by the more complex electrostatic energy function, but it is not completely clear how to proceed along such lines. We emphasize that the single most crude aspect in the application of the force field is the way solvation effects are modeled, and this is the area which deserves the most effort for refinement in the near future.

At this point, it is worthwhile to make a brief comparison of our parameter set with other force fields. First we will consider DNA. Elsewhere, we have shown that our electrostatic charges for the nucleic acid bases gave more accurate gas-phase energies than others.[18] Our nucleic acid backbone charges are somewhat larger in magnitude but qualitatively similar to others in the

(85) Kollman, P.; Case, D.; Profeta, S., Jr.; Murray-Rust, Peter, unpublished normal mode calculations on 17-OH progesterone.
(86) Case, D.; Nguyen, D., unpublished results.
(87) Momany, F. *J. Phys. Chem.* **1978**, *82*, 592.
(88) Smit, P.; Derissen, J.; van Duijneveldt, F. B. *Mol. Phys.* **1979**, *37*, 521.

literature.[9,11,16] For simulations without explicit inclusion of water, it is not clear that our more accurately determined charges will water–solute an advantage over the others but, if water is included, they should be able to give an accurate representation of relative water–water and water–solute interactions. The question of the magnitude of the phosphate charge is appropriate to note here, since some other force fields[9,89] use less than a unit negative charge. Again, without explicit counterions in the calculation, this seems reasonable, but we have shown that complete neutralization appears too extreme an approximation.[90] A correct representation of intra- and interstrand phosphate repulsion is important, and at this point, it is not clear how best to handle this, since the net charges and the dielectric constant used in eq 1 are so interdependent. Further work is needed to sort out this point, since it may be that a range of net phosphate charges and effective dielectric constants would be capable of reasonably representing the hydrodynamics as well as the local conformational energies of nucleic acids.[10,91]

The van der Waals parameters employed here are similar to those in the recent literature[10] in that the radius is typically ≈0.2 Å larger than the standard crystallographic van der Waals radius.

Our intramolecular force field (FF2, Table III) has larger K_ϑ, V_n than earlier force fields because of our desire to better fit bending vibrational frequencies. However, sugar puckering profiles and conformational energies are similar to both FF1 (Table III) and other force fields. The take-home message here is that all molecular mechanics studies on deoxyfuranoses have led to two local minima structures (C2′ endo and C3′ endo), with a smaller O1′ endo than O1′ exo barrier connecting them. All the calculations, except the Levitt and Warshel,[92] suggest an O1′ barrier of 1.3–2.0 kcal/mol, significantly larger than thermal energies. It appears that the reason for the discrepancy is merely an inappropriate choice for ϑ_{eq}(C–O–C) in the Levitt and Warshel paper (where $\vartheta_0 = 120°$). However, FF1, FF2, and the Olson study[30] found C2′ endo more stable, whereas our earlier force field (FF0)[18] and that of Hingerty and Broyde[89] found C3′ endo more stable. NMR data support the greater stability of C2′ endo. Olson attributed C2′ endo stabilization to the presence of an O–C–C–O gauche torsional term. The fact that even the force fields that find C3′ endo more stable contain a gauche O–C–C–O term suggests that the reason why deoxyfuranose prefer C2′ endo to C3′ endo may be more subtle.

We disagree with the suggestion by Olson[93] that force-field parameters appropriate for proteins and hydrocarbons are not per se transferable to nucleic acids. Her argument was based on two facts. First, she pointed out the inadequacy of very early protein van der Waals parameters in representing base stacking. As we have demonstrated here, the Hagler et al.[19] amide parameters do not have this flaw. Second, she cited the poor representation of the O1′ endo sugar puckering barrier by Levitt and Warshel.[92] As noted above, this was probably due to an incorrect choice of parameters. Of course, one could imagine that "fine tuning" parameters separately for proteins or nucleic acids might lead to quantitatively better individual force fields than compromise efforts like this one, but we suggest that there will be no major flaws.

At this point, it is worth briefly comparing our force field with some of the other protein force fields in the literature. It is difficult to compare parameters in detail due to the different methodologies employed in the application of the force fields. For example, our force field uses united atoms only for C–H's but considers complete energy refinement, Gelin and Karplus[6] use united atoms for all hydrogens with complete energy refinement, and Momany's ECEPP[3] uses rigid bond lengths and angles but includes all hydrogens explicitly. Although there are some nontrivial differences in the relevant bond length, angle, and dihedral parameters, it appears that these will not result in great differences in predicted conformational energies and structures. To our knowledge, ours is the first presentation in the literature in which the derivation of K_r, R_{eq}, K_ϑ, ϑ_{eq}, V_n, and γ have been given in detail, although Momany et al.[3] have analyzed in some detail their derivation of V_n and γ.[3] Although our use of vibrational calculations and a scaling algorithm has enabled a reasonably consistent set of parameters, we have noted the inherent limitations of the united-atom (C–H) approximation and the simple harmonic potential function (eq 1) in deriving more quantitative agreement with experimental frequencies.

The most important difference in the force fields resides in the nonbonded (electrostatic and van der Waals) parameters. It is likely that our electrostatic potential derived charges are a more accurate representation of the nature of electrostatic interactions than the CNDO/2 Mulliken charges used in ref 3 and 6, thus allowing smaller well depths for our 10–12 parameters than theirs. Our van der Waals radii are similar to those of Gelin and Karplus,[6] but our well depths are somewhat smaller. As discussed above, our van der Waals radii are smaller than those in the CH united-atom force field of Dunfield.[25] Our representation of the electrostatic term also differs from those used in the extensively parameterized force fields of Allinger[1] and Oie,[94] where a bond dipole model of electrostatics is used. We feel our approach is the more general, particularly since the ionic systems considered here would be difficult to adequately represent with bond moments. However, we stress that one could equally well use our methodology for deriving empirical electrostatic models from quantum mechanical calculations[18] to fit to either bond dipoles or partial charge models or a combination of the two.

Conclusions

We have presented an approach and the results of the development of a molecular mechanical force field. To our knowledge, this is the first time that such a general force field has been developed in a consistent way for both proteins and nucleic acids. Although we have done only a limited number of detailed calculations to test the parameter set, the results of calculations on furanose sugar puckering, base stacking, and hydrogen bonding, base-paired dinucleoside phosphate refinement, Φ, Ψ energy contours for dipeptide models, H-bonding calculations on protein polar and ionic groups, and refinement of insulin all suggest that the model contains no major flaws. However, we have also delineated areas for future improvement of such force fields, and we feel that the results presented here are a reasonable starting point for such development.

Note Added in Proof. To further address the question of protein compaction discussed above, we have carried out a number of unrestrained refinements on myoglobin, which has some hydrophobic internal cavities. We have found in these refinements that using the Jorgensen (ref 25) and van der Waals parameters for CH, C2, C3 atoms along with a scale factor of 8.0 for the 1–4 van der Waals interactions that the internal cavities of the molecule are better preserved than with the parameters as presented here (scale factor of 2.0 and smaller CH, C2, C3 van der Waals radius of Table XIX). This scale factor also leads to reasonable conformational profiles for *n*-butane and methyl ethyl ether when used with the larger van der Waals parameters. However, such a modification is not appropriate for nucleic acids. An all-atom version of the force field presented her is nearly complete. Hall and Pavitt (*J. Compt. Chem.*, in press) have tested the force field presented here on crystal simulations of cyclic peptides and have found it to be effective.

Acknowledgment. We are pleased to acknowledge the support of the NIH (GM-29072 and CA-25644) in this study. D.A.C is an Alfred P. Sloan Foundation Fellow. Our development of the torsional second derivative part of AMBER was based upon a

(89) Hingerty, B.; Broyde, S. *Biochemistry* **1982**, *21*, 3243.
(90) Tilton, R.; Weiner, P.; Kollman, P. *Biopolymers* **1983**, *22*, 969.
(91) Olson, W. *Biopolymers* **1975**, *14*, 1775, 1797.
(92) Levitt, M.; Warshel, A. *J. Am. Chem. Soc.* **1979**, *100*, 2607.
(93) Olson, W. In "Topics in Nucleic Acid Structure"; Neidle, S., Ed.; McMullian: London, 1982, Vol. II.
(94) Oie, T.; Maggiora, G.; Christoffersen, R. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1981**, 1.

formulation by B. Brooks and M. Karplus of Harvard University. The curve plotting software was provided to us by N. Pattabiramin at U.C.S.F. Unpublished normal modes analyses for the nucleic acids were done by D. Nguyen of U.C. Davis. A special thanks goes to Shoshana Wodak for bringing the question of protein compaction in insulin refinements to our attention and Peter Murray-Rust for carrying out the Cambridge crystal file search for phosphate geometries.

Registry No. *N*-acetyl-*N'*-methylglycinamide, 7606-79-3; *N*-acetyl-*N'*-methylalaninamide, 19701-83-8; THF, 109-99-9; MEE, 540-67-0; dimethyl phosphate, 813-78-5; diethyl phosphate, 598-02-7; deoxyadenosine, 958-09-8; adenosine, 58-61-7; 9-methylguanine, 5502-78-3; 1-methylcytosine, 1122-47-0; 9-methyladenine, 700-00-5; 1-methylthymine, 4160-72-9; 1,3-dimethyluracil dimer, 40037-86-3; poly[d(G-C)], 36786-90-0; poly d(G)·poly d(c), 25512-84-9; poly[d(A-T)], 26966-61-0; poly d(A)·poly d(T), 24939-09-1; poly[d(T-G)]·poly[d(C-A)], 27732-52-1; poly[d(T-C)]·poly[d(G-A)], 29627-66-5; poly[d(A-T-C)]·poly[d(G-A-T)], 24939-08-0; poly[d(T-T-G)]·poly[d(C-A-A)], 27902-32-5; poly[d(T-A-C)]·poly[d(G-T-A)], 57473-66-2; poly[d(T-T-C)]·poly[d(G-A-A)], 27861-08-1; glycylglycine, 556-50-3; alanylalanine, 1948-31-8; benzene, 71-43-2; *N*-methylacetamide, 79-16-3; methanol, 67-56-1; methanethiol, 74-93-1; dimethyl sulfide, 75-18-3; dimethyl disulfide, 624-92-0; insulin, 9004-10-8; Gly, 56-40-6; Thr, 72-19-5; Ile, 73-32-5; Ser, 56-45-1; Leu, 61-90-5; Val, 72-18-4; Asn, 70-47-3; Arg, 74-79-3; Gln, 56-85-9; Ala, 56-41-7; Phe, 63-91-2; Tyr, 60-18-4; Cys, 56-89-3; Met, 63-68-3; Hie, 71-00-1; Hip, 70805-60-6; Gln, 56-86-0; Trp, 73-22-3; Asp, 56-84-8; Lys, 56-87-1; Pro, 147-85-3; adenine, 73-24-5; guanine, 73-40-5; thymine, 65-71-4; cytosine, 71-30-7; uracil, 66-22-8.

Communications to the Editor

Biosynthetic Origin of the Carbon Skeleton and Oxygen Atoms of Nargenicin A_1

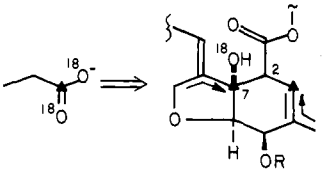
David E. Cane*[1] and Chi-Ching Yang

Department of Chemistry, Brown University
Providence, Rhode Island 02912

Received September 21, 1983

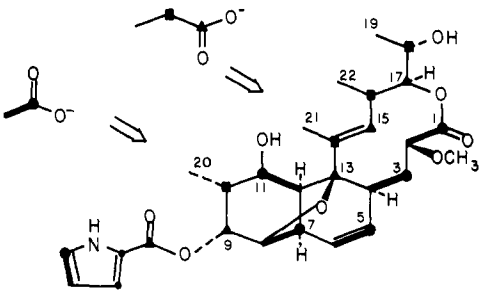
It is well accepted that in the biosynthesis of aromatic polyketides the fundamental chain-building and ring-forming reactions take place by similar carbonyl condensation mechanisms.[2] An analogous, albeit far more complex, sequence of reactions is also believed to be responsible for the formation of reduced linear polyketides typified by the macrolide[3] and polyether[4] antibiotics. Thus recent studies of erythromycin,[5] monensin,[6] and lasalocid[7] biosynthesis have supported the notion that these polyoxygenated, branched-chain fatty acids are assembled by a sequence of condensation, reduction, dehydration, and reduction reactions closely related to those leading to classical saturated fatty acids. The biosynthesis of reduced, carbocyclic polyketides, on the other hand, is far less well understood. Although the parent polyketide chains, in all cases examined to date, have been shown to be derived largely from the common precursors acetate and propionate, little evidence is available to allow a distinction among plausible carbonyl condensation, electrophilic polyolefin cyclization, and Diels–Alder sequences that might account for the formation of the characteristic carbocyclic ring systems.[8] Recently, as part of a study of avermectin (**1**) biosynthesis, we presented evidence suggesting that the C–2,7 bond of the constituent cyclohexene ring is probably generated by condensation of a C-7 carbonyl group with a carbonyl-stabilized anion.[9] Specifically, the observed derivation of the C-7 hydroxyl group from the carboxylate oxygens of the propionate precursor ruled out the alternative cyclization of a polyolefinic intermediate (Scheme I). We have now extended our studies to an examination of the biosynthesis of the saturated cyclic polyketide nargenicin A_1 (**2**) an antibiotic active against *Staphylococcus aureus* and containing a novel octalin ring system.[10,11]

Scheme I



1

Scheme II



2

(1) National Institutes of Health Research Career Development Award, 1978–1983.

(2) Weiss, U.; Edward, J. M. "The Biosynthesis of Aromatic Compounds"; Wiley: New York, 1980; pp 326–428. Herbert, R. B. "The Biosynthesis of Secondary Metabolites"; Chapman and Hall: London, 1981; pp 28–49.

(3) Corcoran, J. W. In "Antibiotics IV. Biosynthesis"; Corcoran, J. W., Ed.; Springer-Verlag: New York, 1981; pp 132–174. Omura, S.; Nakagawa, A. *Ibid.* pp 175–192.

(4) Westley, J. W., ref 3, pp 41–73. Liu, C.-M. In "Polyether Antibiotics. Naturally Occurring Acid Ionophores"; Westley, J. W., Ed.; Marcel Dekker: New York, 1982; Vol 1, pp 43–102.

(5) Cane, D. E.; Hasler, H.; Liang, T.-C *J. Am. Chem. Soc.* **1981**, *103*, 5960. Cane, D. E.; Hasler, H.; Taylor, P. B.; Liang, T.-C. *Tetrahedron* **1983**, *39*, 3449.

(6) Cane, D. E.; Liang, T.-C; Hasler, H. *J. Am. Chem. Soc.* **1981**, *103*, 5962; **1982**, *104*, 7274. Ajaz, A. A.; Robinson, J. A. *J. Chem. Soc., Chem. Commun.* **1983**, 645.

(7) Hutchinson, C. R.; Sherman, M. M.; Vederas, J. C.; Nakashima, T. T. *J. Am. Chem. Soc.* **1981**, *103*, 5953. Hutchinson, C. R.; Sherman, M. M.; McInnes, A. G.; Walter, J. A.; Vederas, J. C. *Ibid.* **1981**, *103*, 5956.

(8) Besides the bicyclic octalin metabolites discussed in this report, members off the class of reduced carbocyclic polyketides include, inter alia, monocyclic cyclopentanes (e.g., brefeldin A: Haerri, E.; Loeffler, W.; Sigg, H. P.; Staehlin, H.; Tamm, Ch. *Helv. Chim. Acta* **1963**, *46*, 1235. Sigg, H. P. *Ibid.* **1964**, *47*, 1401), monocyclic cyclohexanes (palitantin: Bowden, K.; Lythgoe, B.; Marsden, D. J. S. *J. Chem. Soc.* **1959**, 1662), bicyclic hydrindanes (antibiotic X-14547A: Westley, J. W.; Evans, R. H.; Liu, C.-M; Hermann, T.; Blount, J. F. *J. Am. Chem. Soc.* **1978**, *100*, 6784), and *as*-hydrinacenes (ikarugamycin: Ito, S.; Hirata, Y. *Bull. Chem. Soc. Jpn.* **1977**, *50*, 1813.)

(9) Cane, D. E.; Liang, T.-C.; Kaplan, L.; Nallin, M. K.; Schulman, M. D.; Hensens, O. D.; Douglas, A. W.; Albers-Schoenberg, G. *J. Am. Chem. Soc.* **1983**, *105*, 4110.

0002-7863/84/1506-0784$01.50/0