Hydrogen Bond Energetics: A Simulation and Statistical Analysis of *N*-Methyl Acetamide (NMA), Water, and Human Lysozyme†

Matthias Buck‡,§,|| and Martin Karplus*,‡,§

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, and Laboratoire de Chimie Biophysique, Institut le Bel, Universite Louis Pasteur, 4, rue Blaise Pascal, 67000 Strasbourg, France

Received: March 20, 2001; In Final Form: July 31, 2001

Energy minimization and molecular dynamics simulations have been used to study hydrogen bond interactions in dimers of *N*-methylacetamide (NMA), in NMA−water complexes, and in human lysozyme. The potential energy surface is found to be determined by the interactions of entire peptide groups ($O{=}C_{i-1}{-}N_i{-}H$) or water molecules rather than by single donor and acceptor groups. The contact distance between the donor hydrogen and the acceptor as well as the angle of the bond at the donor hydrogen are the principal geometric parameters that describe the hydrogen bond. Potential energy surfaces were also examined in the presence and absence of explicit solvent molecules. The results suggest that both competing hydrogen bond interactions and the thermal motion of atoms broaden the distribution of low energy donor−acceptor contacts. Comparisons are made with a statistical analysis of mainchain hydrogen bond donor and acceptor contacts in high-resolution crystal structures of nonhomologous proteins. Interaction energies and geometries of the NMA model system mimic those found in folded polypeptide chains. All systems are characterized by a minimum in the population of donor−acceptor contacts at interaction distances of 2.4−2.6 Å. This minimum originates from spatial constraints that are enhanced by electrostatic interactions in environments that are characterized by competition for hydrogen bonding interactions. The presence of such a minimum in the distribution of donor−acceptor contacts supports the definition of hydrogen bonds by geometrical cutoff criteria with a donor−hydrogen acceptor distance of less than 2.5 Å and an angle of deviation not more than 90° from linearity of the donor, donor−hydrogen and acceptor atoms.

Introduction

The concept of hydrogen bonding, based originally on the observation of the preferred spatial arrangements of polar groups involving donor and acceptor atoms, is essential to our understanding of systems as diverse as ice, organic crystals, proteins, and nucleic acids.[1−5] For proteins in their native state, the large majority of the mainchain polar groups are involved in hydrogen bonding interactions. When buried inside a globular protein, about 90% of the polar groups of the mainchain participate in hydrogen bonds,[6] whereas surface polar groups interact mainly with the surrounding solvent. In fact, the proposal of Pauling and Corey that α-helices and *β*-sheets play an important role in proteins, made before globular protein structures were available, postulated the importance of saturating the hydrogen bonding capacities of the polypeptide chain.[7,8] Analyses of crystallographic data has recently pointed to the importance of understanding hydrogen bonding and the role of water in protein−ligand and protein−protein interactions.[9,10]

Many studies concerning the features of hydrogen bonding have been based on the statistics of their occurrence in protein structures.[6,11−17] Hydrogen bonds involving protein−water interactions, although less tractable experimentally[18,21] have also been analyzed in this way.[12,22] In such analyses, the criterion for the existence of a hydrogen bond plays an essential role. Criteria based on geometrical considerations (donor−acceptor distances and donor−acceptor angles) are widely used, but specific justification of the choices of the cutoff values are difficult.[15,23,24] "Using geometric criteria [to define hydrogen bonds] is highly unusual. However, an energy based alternative would be little better than guesswork" is the conclusion of a review by McDonald and Thornton.[14] The recent results concerning the possibility of using scalar coupling constants to determine the existence of hydrogen bonds by NMR spectroscopy[25−27] do not solve the problem of selecting the most useful physical criterion, because positional data from a structural model (e.g., a high-resolution X-ray structure) of the protein is required to interpret the NMR results.

In this study, we use energy-based calculations, molecular dynamics simulations, and statistical analyses of structural data to obtain a better description of the energy and free energy surfaces arising from the hydrogen bonding interaction. For the energetic calculations, the CHARMM22 empirical energy function[28,29] is used. It does not include an explicit hydrogen bonding term but aims to reproduce hydrogen bonding geometries and propensities based on bonded and nonbonded parameters (principally, van der Waals radii and partial charges).

† This work was supported by grants from NIH and NSF. M.B. was a fellow of the International Human Frontiers Science Program Organization and a Marie Curie research fellow (Biotechnology) of the European Commission. He is currently funded by an NIH postdoctoral training grant.

* To whom correspondence should be addressed. E-mail: marci@tammy.harvard.edu; fax: 617-496-3204.

‡ Harvard University.

§ Universite Louis Pasteur.

|| Present address: Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021.

10.1021/jp011042s CCC: $20.00
Published on Web 10/12/2001

These have been chosen to achieve the correct balance for the energetics of protein−protein, protein−solvent, and solvent−solvent hydrogen bonds, in correspondence with ab inito and experimental data.[29−31] The CHARMM energy function has been employed for many simulations,[32] including studies of protein and polypeptide (un)folding. We compare the results obtained from calculations for model systems with those of a complete protein. We consider residue and structurally averaged results that correspond to the most readily obtainable experimental data and compare the energetics and geometry of protein−protein and protein−solvent polar contacts with those seen in a model system (*N*-methylacetamide) for the peptide group. The results are used to obtained a more fundamental basis for the hydrogen bonding criteria in common use.

Methods. The interaction energies and the geometric distribution of polar donor and acceptor groups are studied for a pair of *N*-methylacetamide (NMA) molecules and a single NMA molecule complexed with one to three TIP3P water molecules. Both systems are also studied when solvated by a sphere of water molecules. A complementary analysis is made of the hydrogen bonding interactions in a simulation of the protein human lysozyme in the presence of explicit solvent. Details of the methods used for the different systems are given below. The CHARMM program[33] was used with the all-atom CHARMM22 potential energy function.[29] A force switched cutoff between 9.0 and 11.5 Å[31] and a constant dielectric of 1 were employed.

(a) Sampling of Hydrogen Bond Conformational Space with Small Molecule Model Systems Using Minimization. Two NMA molecules, two TIP3P water molecules, or one NMA and one TIP3P molecule were placed initially near their energetically optimal conformations for complexation that have been described previously[35,36] (see Figure 1a). Figure 1b shows the parameters that were varied in a grid search of the energetics of the two NMA molecules. The hydrogen donor−acceptor distance, r(H..A) was varied from 1.5 to 3.8 Å in 0.1 Å, steps and the principal angle of deviation of the hydrogen bond θ-(D−H...A) from linearity at the donor hydrogen (H) was sampled in 10° steps by clockwise rotation from −180° to +180° about the donor hydrogen. Deviation (Ω) from bond linearity at the acceptor (A) relative to the donor hydrogen (H) and the acceptor antecedent (AA) was explored by rotation around any one of two axes. For the other complexes, only r(H...A) and θ(D−H...A) were examined. The geometry of the donor heavy atom, the hydrogen atom, and the acceptor atom was then held fixed by application of a harmonic constraint (5.0×10^4 kcal/mol Å) applied to the three atoms while the remaining atoms were minimized for 25 steps of steepest descent followed by Newton Raphson minimization until convergence of the total system energy had reached 10^{-4} kcal/mol or to a maximum of 1000 steps. Minimization allows adjustment of the internal coordinates (such as methyl group orientation) and of the remaining degrees of freedom of the two molecules relative to one another for each change in hydrogen bond geometry. The energies were typically lowered by less than 1% for structures that are close to the energy minimum for a set of hydrogen bond parameters. In other cases, minimization decreased energies by tens of kcal/mol, in primarily relieving bad van der Waals contacts.[37]

(b) Sampling of Hydrogen Bond Conformational Space Using Molecular Dynamics. The NMA, NMA{NH-}TIP3P, and NMA{C═O}TIP3P dimer complexes were initially positioned in their optimal configurations as found by the grid search (Figure 1a). Also, a single NMA molecule was studied in complexes with two and three TIP3P molecules[35] (see Figure 3). One nanosecond vacuum simulations were carried out on the complexes in absence of restraints, except for SHAKE,[38] which was applied to all covalent bonds involving hydrogen. The kinetic energy was controlled by coupling to a NOSE-HOOVER heatbath at 300 K,[39,40] as implemented in CHARMM (M. Watanabe and M. Karplus, unpublished). A solvated system was constructed by placing the NMA near the center of a 14 Å sphere of TIP3P solvent. The integrity of the latter was maintained by a deformable boundary with a force constant of about 12 kcal/Å mol applied to the TIP3P oxygen atom if located more than 13.5 Å from the center of the sphere.[32] Control calculations were carried out in which the NMA dimer was placed at the center of a box with a sidelength of 24 Å, which was subject to periodic boundary conditions. The data are nearly identical, demonstrating that the size of the system and choice of boundary conditions do not significantly affect the results.[41] Stochastic boundary conditions were used in the simulations of human lyosozyme at different temperatures and in unfolding simulations (Buck and Karplus, manuscript in preparation). The temperature was maintained at 300 K by weak coupling to a Berendsen type heatbath[42] for the fully solvated system. Again, SHAKE[38] was used to restrain all bonds to hydrogen. Atomic coordinates were propagated for 1 ns with the leapfrog dynamics integrator at a simulation time step of 2 fs. In presence of solvent only sporadic interactions were observed between the two free NMA molecules, and therefore a NOE-type restraint of 50 kcal/Å mol was applied for distances greater than 3.8 Å (R_{on}) between the amide hydrogen and carbonyl oxygen. This does not allow the NMA molecules to move apart but also does not alter their distribution for distances less than 3.8 Å. Radial distribution functions of donors and acceptors were calculated using

$$g_{ab}(r) = n_{ab}(r)/4\pi r^2\rho$$

where n_{ab} is the number of particles b at a distance r to $r + dr$ from particle a, and the number ρ is the density $= N/V$, where N is the total number of particles and V is the volume of the system. Values for $n_{ab}(r)$ are taken from a histogram, which is normalized with respect to the number of sampling steps.[43] The potential of mean force was obtained from the radial distribution function as

$$w(r) = -kT \ln [g_{ab}(r)]^{44}$$

where k is the gas constant, and T is the absolute temperature (here 300 K). The radial distribution functions and potentials of mean force were calculated for 125 TIP3P or TIP3P^{e0} (van der Waals only) water molecules contained in a box of 15 Å length with periodic boundary conditions from analysis of a 4 ps simulation at 300 K after 20 ps of equilibration.

(c) Protein Simulations. The crystal structure of human lysozyme, refined to 1.5 Å by Artymiuk and Blake[45] served as a starting point for a 200 ps simulation carried out at 300 K. The protein was simulated with all titratable groups protonated (low pH simulations were carried out to study protein unfolding at high temperatures to be presented elsewhere) as part of a solvated system at the center of a 34 Å sphere of TIP3P water molecules. After addition of hydrogen atoms to the protein,[46] the system was subjected to several short cycles of minimization (100 steps SD), followed by heating to 300 K and a short (14 ps) equilibration period. The mainchain rmsd was 0.85 Å after 200 ps of simulation

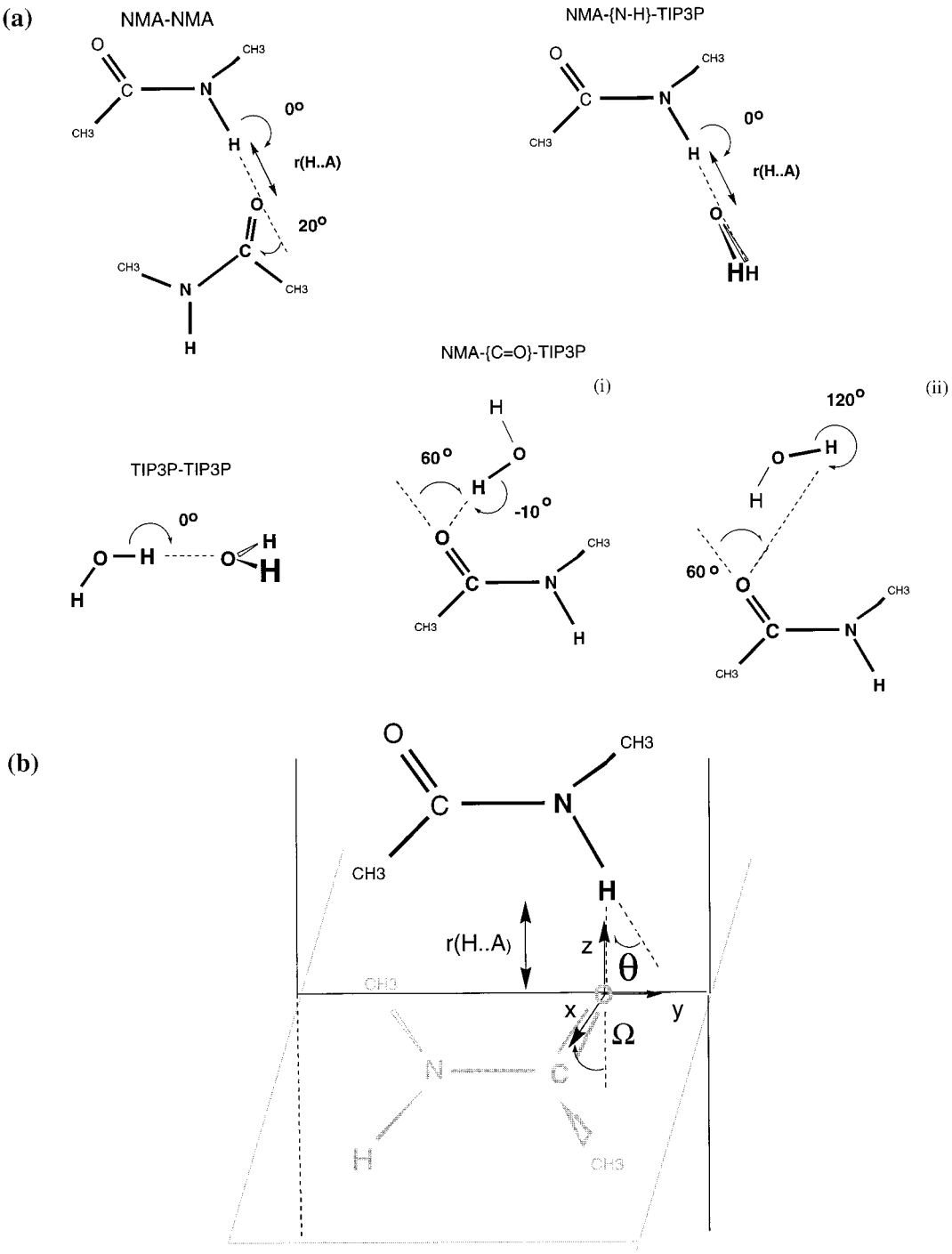

Figure 1. (a) Minimum energy configurations for the NMA−NMA, NMA{N−H}−TIP3P, NMA{C═O}−TIP3P, and TIP3P−TIP3P complex obtained with the CHARMM22 potential function[35] used as the starting coordinates for the grid search. In NMA{N−H}−TIP3P and the TIP3P dimers, the TIP3P is in a plane perpendicular to that of the other molecule. The NMA{C═O}−TIP3P complex has a second minimum (ii) as shown. (b) Two NMA molecules in the trans conformation positioned as a nonplanar antiparallel complex. The molecules may be in separate planes intersecting at the oxygen about a line (axis *y* is perpendicular to the plane formed by atoms H, C, and O). Hydrogen parameters are *r*(H...A), the hydrogen bond length, defined as the distance between donor hydrogen (H) and acceptor oxygen (A); the angle (θ) defined as deviation from linearity of the N−H...O hydrogen bond; and the angle (Ω) defined as the deviation from linearity of the H...O═C bond. Ω was explored by in or out of plane tilting of the second molecule (in gray) around the axes *z* and *y*, respectively (*z* is an axis through O perpendicular to the plane formed by atoms O, C, and N of the second molecule). Rotation around the acceptor site about a third axis *x* (along the C═O bond) does not lead to changes in θ or Ω and has little influence on the total potential energy. Rotation around the acceptor (O as the site of plane intersection) about the *y*, *z* axes is possible from −50° to +30° and −5° to +50°, respectively, without severe van der Waals clashes in the NMA dimer.

The sum of pairwise interaction energies between the NMA/protein mainchain donor and acceptor groups (N−H and C═O, respectively) and between the NMA/protein mainchain and water molecules were determined together with their interaction geometries from coordinate frames saved every 0.2 ps. The following types of donor−acceptor interactions were excluded from the analysis: mainchain-side chain, intrapeptide group ($C{=}O_{i-1}$ to $N{-}H_i$) as well as neighboring peptide group ($N{-}H_i$ to $C{=}O_i$) interactions. The latter are excluded because although they are within hydrogen bonding distance in some β-sheet conformations the hydrogen bonds would deviate by more than 90° from bond linearity. Figures 2 and 4 show that

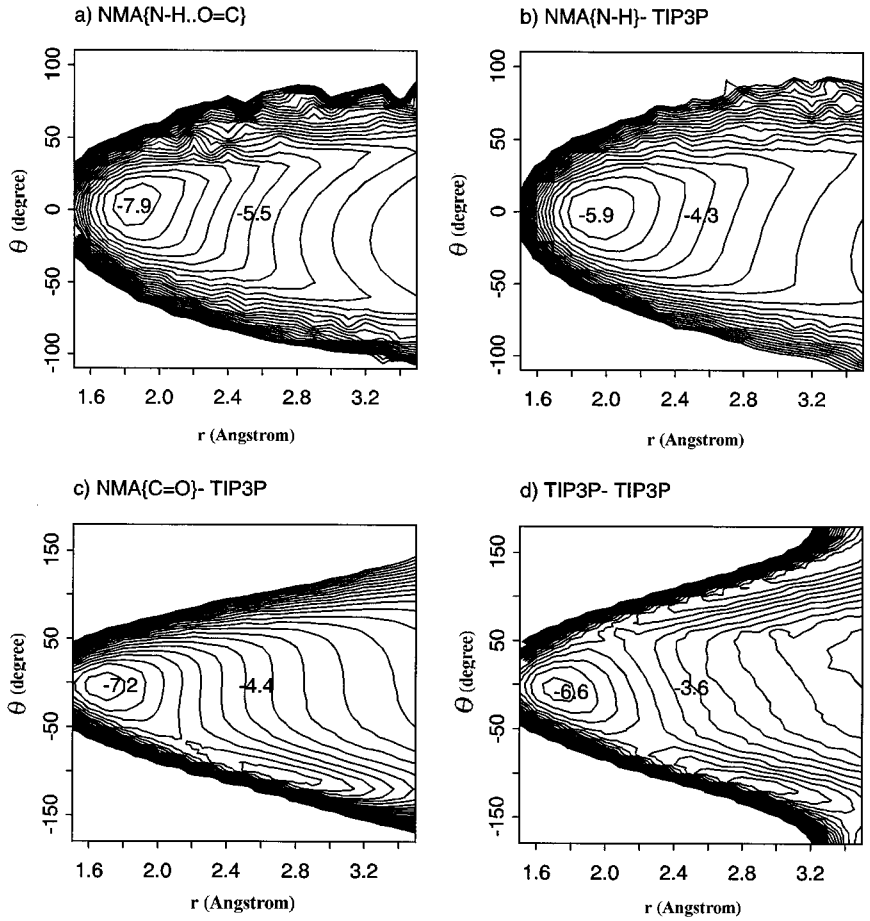

Figure 2. Total interaction energy as a function of r and θ for minimized structures (see Methods) of (a) two antiparallel NMA molecules; (b) and (c) the NMA−TIP3P complex with a water molecule bound to the NMA amide and carbonyl oxygen, respectively; and (d) the TIP3P dimer. Contour levels are plotted at intervals of 0.5 kcal/mol. Energies are indicated for the energy minimum and at $r = 2.5$ Å, $\theta = 0°$. (See also Table 1). Starting conformations for the grid search are shown in Figure 1. A valley in the energy surface that extends toward greater hydrogen bond deviation from linearity as a function of increased donor,acceptor distance and merges with a second minimum is seen in the potential surface of NMA{C═O}−TIP3P at 2.8 Å and −105°. This minimum arises because the second hydrogen of water can hydrogen bond to the C═O of NMA on rotating the water molecule by approximately 105° and is an inevitable effect in our grid search procedure. In terms of the hydrogen bond angle measured, this places the first hydrogen of water away from the acceptor while giving a similar total energy (this is shown in Figure 1a). For the TIP3P dimer, there is additional symmetry.

the interaction energies between the N−H and O═C dipoles are always unfavorable in this case. Distances and angles between acceptor atoms and donor hydrogens were calculated using the COOR DIST, COOR AXIS, and HBOND function in CHARMM. The interaction energies for the configurations were then averaged over molecule pairs/interacting peptide groups that have similar geometric hydrogen bonding parameters (r and θ) by binning them into ranges of 0.1 Å and 10° and thus creating a potential energy surface. In addition, interaction energies and values for the deviation of hydrogen bonding contacts around the acceptor atom (Ω) were extracted from the trajectory for complexes that populate particular regions of r, θ space. To obtain the population density at a given distance range between R_{low} to R_{high} (i.e., the number of hydrogen bonding contacts per Å^3) we divide the population of donor−acceptor contacts by a volume term $(2/3\pi(\cos\theta_{high} - \cos\theta_{low})(R^3_{high} - R^3_{low})$, arising from the Jacobian. Both the population and the population density are examined; the latter can be compared directly with the results of McDonald and Thornton.[15]

Results and Discussion

The energetic and geometrical properties of hydrogen bonds are assessed in several ways.

Two types of simulations were employed to study the interaction between two *N*-methylacetamide (NMA) molecules and between NMA and TIP3P water. The first samples conformational space of small molecule models using grid searches with minimization, and the second uses molecular dynamics for the same models and for the models in a sphere of water. The results are then compared with data from hydrogen bonding in proteins obtained from a molecular dynamics simulation and from protein crystal structures. Finally, the results are used to evaluate different hydrogen bonding criteria in common use.

(1) Sampling via Grid Seach with Minimization. Two molecules were complexed and the hydrogen bond distance r(H...A) between donor and acceptor, and the deviation from linearity at the donor hydrogen θ(D−H...A) were set to certain values (see Materials and Methods and Figure 1b). The remaining degrees of freedom of the conformation were then minimized. Figure 2a−d shows adiabatic potential energy surfaces as a function of these two hydrogen bond parameters for the total interaction energies between two NMA molecules, between one NMA and a TIP3P water (bound at the carbonyl oxygen and separately bound to the amide hydrogen) and between two TIP3P molecules. There is a slight energetic preference for the antiparallel configuration of the dimer over a parallel one (data not shown). A third parameter, Ω (Figure

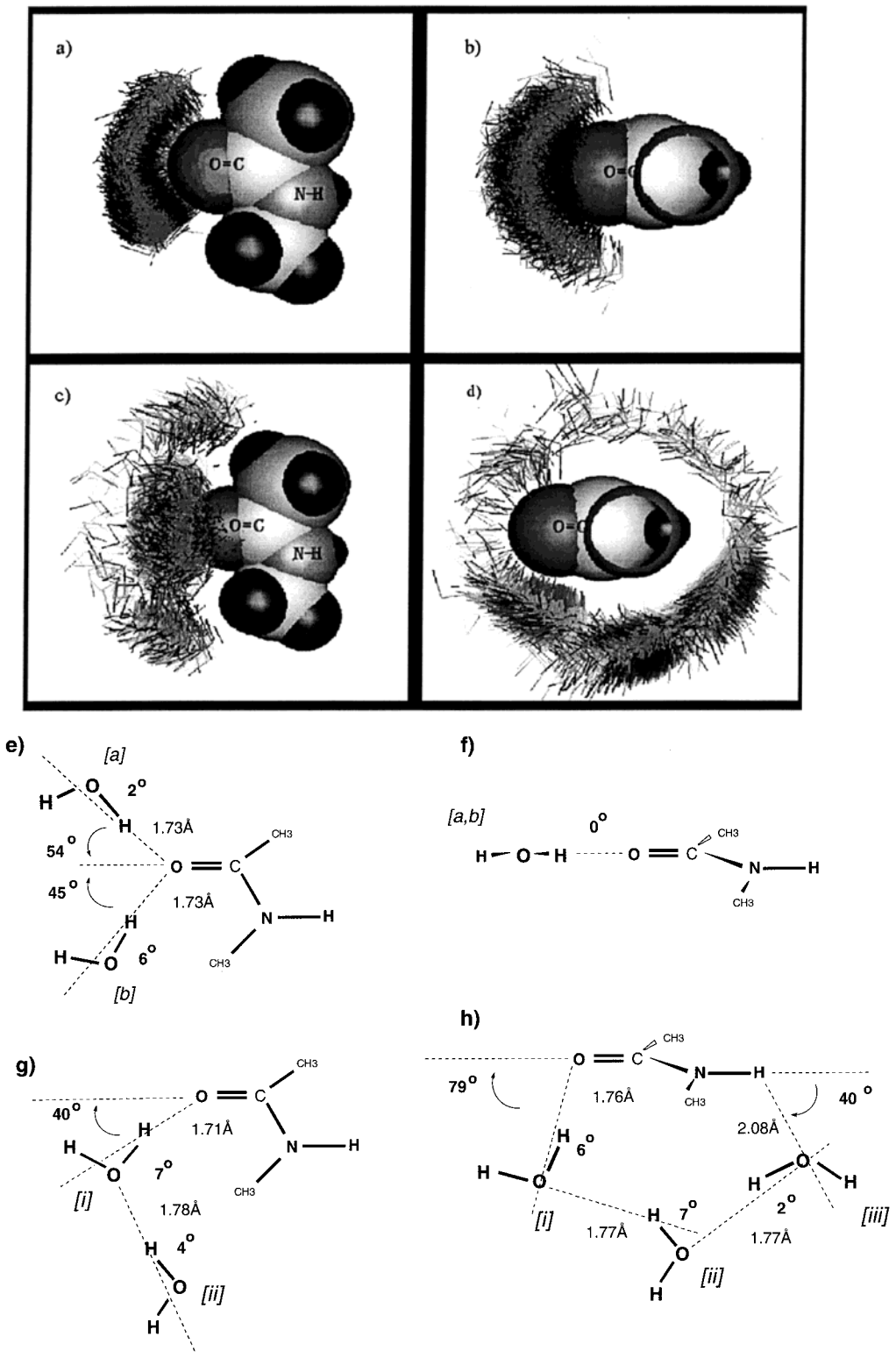

Figure 3. (a and b) TIP3P water bound to the NMA molecule acceptor in the molecular dynamics simulation. NMA is shown as a spacefilling model based on van der Waals radii, and the water molecule is in stick representation. The pictures were generated by rms superposition of trajectory frames on the peptide group of the NMA molecules. The oxygen of NMA points to the left in all cases. Water molecules whose oxygen atoms are located within a section of 2 Å below or above a plane through or perpendicular to the NMA molecule are included. (a) in-plane population of a single TIP3P bound to NMA{C═O}; (b) as panel a, but showing the out-of-plane population for the NMA molecule viewed with the methyl group seen end on; (c) in-peptide-plane population of two TIP3P molecules around NMA{C═O}; (d) out-of-plane population of three TIP3P molecules around NMA (viewed as in panel b); atoms colored black are hydrogens. Some of the van der Waals spheres are cut for clarity in panels b and d. (e−h) The distribution of configurations shown in panels a−d were minimized and one, or, in the case of NMA{C═O}−TIP3P, several minima were obtained, denoted [a] and [b]. These configurations are shown in the same orientation as panels a−d together with the principal bond angles and bond lengths. Interaction energies and bond lengths are given in Table 2.

1b), explores the deviation of the hydrogen bond from linearity about the oxygen acceptor and was sampled in a systematic manner by rotating one molecule relative to the other about the acceptor atom. The variation of the potential energy is relatively

TABLE 1: Minimum Energies (in kcal/mol) for Small Molecule Models and Their Hydrogen Bonding Geometries [Interaction Energy of Dimer at (a) the Minimum of the Energy Surface and (b) at the Commonly Used Cutoff]

model system	NMA−NMA		NMA{C═O}−TIP3P		NMA{N−H}−TIP3P		TIP3P−TIP3P	
conformation	antiparallel							
r, θ, Ω	(a) 1.85 Å, 0°, 18° (b) 2.50 Å, 0°, 45°		1.80 Å, −5°, 60° 2.50 Å, 0°, 60°		1.90 Å, 0° 2.50 Å, 0°		1.70 Å, 0° 2.50 Å, 0°	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
VDW	−0.93	−1.57	0.56	−0.31	0.59	−0.61	1.70	−0.61
elec	−7.03	−3.90	−7.80	−4.04	−6.52	−3.64	−8.34	−2.95
total	−7.96	−5.47	−7.24	−4.35	−5.93	−4.25	−6.64	−3.56
VDW[a]	−0.44		0.93		0.49		1.50	
total[a]	−7.75		−7.60		−6.30		−6.55	
	Interaction Energy of Nonmethyl Atoms							
VDW	0.85	−0.50	0.78	−0.51	0.97	−0.31		
elec	−6.69	−3.32	−7.78	−3.86	−5.70	−2.83		
total	−5.84	−3.82	−7.00	−4.37	−4.73	−3.14		
	Interaction Energy of Two Dipoles (C═O or H^1−OH^2 with N−H or OH^2-H^1)							
VDW	0.83	−0.28	0.85	−0.27	1.14	−0.21		
elec	−2.87	−1.00	−1.10	0.56	−4.01	−1.36		
total	−2.04	−1.28	−0.05	0.29	−2.87	−1.57		

[a] MacKerell and Karplus, 1991.

small, except for the van der Waals clashes which limit the accessibility[30] (legend to Figure 1, see also Adalsteinsson et al.[47]).

The optimal configurations and associated interaction energies for these pairs of molecules are very similar to those calculated on the basis of Monte Carlo simulations by MacKerell and Karplus[35] (see Table 1), which in turn are in good agreement with ab initio quantum mechanical results (e.g., Guo and Kaplus;[36,48] Gao and Freindorf[49]); for another empirical approach; see Mitchell and Price.[31] The differences in the minimum value of the all-atom interaction energies between the complexes is less than 1.5 kcal/mol (or 25% of the total value) with energies in order of NMA−NMA < NMA{C═O}−TIP3P < TIP3P−TIP3P < NMA{N−H}−TIP3P. This suggests that the molecules have a similar propensity to form hydrogen bond interactions given the CHARMM param22 potential function, as far as the energetic contribution to the free energy of dimerization is concerned. A similar ordering of the interactions energies have been found in ab initio simulations[30,50] and with the CHARMM Param19 potential function.[30] We show below that this ordering is subtly affected by the contribution of the NMA methyl groups and by averaging via dynamics, both in the presence and absence of explicit TIP3P solvation.

NMA Dimer Complex. Figure 2 shows the variation of the potential energy as a function of hydrogen bond parameters r and θ. The shape of the surface of the NMA dimer (Figure 2a) and of NMA-amide hydrogen bonded TIP3P (Figure 2b) is similar. They have minima at a similar donor acceptor distance (1.85−1.90 Å) with negligible deviation from hydrogen bond linearity (angle θ) at the donor hydrogen. However, a valley in the energy surface extends asymptotically from this minimum, approaching a maximum deviation at a donor angle of around 45° from hydrogen bond linearity with increasing hydrogen bond distance (the origin of this behavior is discussed below). Deviations from hydrogen bond linearity at the hydrogen bond acceptor (Ω) are small at the potential energy minimum of the r, θ surface (18° and 0°, respectively, for antiparallel and parallel configurations). However, different values for Ω (e.g., ± 20° for NMA{C═O}−TIP3P) are obtained as r and θ increase. This is in agreement with other studies.[30,47]

The analysis of the components of the energy surface allows us to discuss two of its features: its range (i.e., absolute energy values) and its shape. It is apparent that the stabilizing electrostatic energy increases in magnitude as the hydrogen bond distance decreases. The electrostatic interactions are relatively long range. In the simple case of two dipoles (C═O...H−N), the interaction energy falls off proportionally to $1/R^3$ where R is the distance between the dipole centers. The van der Waals interactions, by contrast, are extremely short range, i.e., when atoms approach each other at distances close to the sum of their radii, the potential energy function rises steeply. Thus, the van der Waals and the total energy surface is bounded by steric constraints between the interacting molecules. At distances just outside the sum of the van der Waals radii, the interactions are favorable (up to −0.5 kcal/mol). The energy surface is dominated by electrostatic interactions in the regions not affected by unfavorable van der Waals interactions.

Grzybowski et al.[53] have recently used the crystal structures of small organic molecules to determine a knowledge based potential for C═O...H−N hydrogen bonds. In their analysis, they find an effective potential that aims to describe the angular and distance dependence of the observed hydrogen bonds in terms of the interactions of these four atoms and ignores the contributions of other atoms present in the compounds used for the analysis. However, a significant contribution to the angular dependence arises from interactions other than those of the four atoms that "make" the hydrogen bond. Thus, although their statement (p 7297) that the C═O...H−N energy surface from CHARMM are not directional is true, per se, it is not germane to actual calculations, because, as they point out, the correct angular dependence is obtained when other atoms in the molecule are included. This was the reason for removing the angular-dependent hydrogen bonding term from the CHARMM force field.[36] Ascribing all of the angular dependence to an effective potential between the four atoms directly involved in the hydrogen bonds may be useful for certain types of modeling but is not a physically complete description of the interactions.

H_2O−Donor Complex. The total energy surface of the NMA{C═O}−TIP3P complex and that of the TIP3P dimer (Figure 2c,d) is somewhat broader in terms of the hydrogen bond angle around the donor hydrogen (θ) as compared to that of the NMA dimer. The size of the interacting molecules and the accessibility of the donor and acceptor groups within them appears to determine the extent of conformational space, that is, it is primarily the van der Waals contacts that determine the accessible region. Because of this, the size of the accessible low energy region is in the order of NMA−NMA < NMA-

{N−H}−TIP3P < NMA{C═O}−TIP3P < TIP3P−TIP3P (Figure 2). Similarly to the NMA dimer above, both the groups involved in the hydrogen bond and those which are not (e.g., $N-H^2...O-H$ for H_2O bound at $C{=}O^1$) contribute to the potential energy surface.[51] Hydrogen bond parameter space for more complicated complexes, for example, for NMA with two TIP3P molecules bound to the carbonyl oxygen,[36] have not been considered in the grid search but have been examined by molecular dynamics.

(2) Sampling of Hydrogen Bond Conformational Space with Simple Models Using Molecular Dynamics. Coordinate frames were analyzed from a 1 ns simulation of a NMA dimer in a vacuum, of NMA−TIP3P complexes in a vacuum, and of a NMA dimer at the center of a 14 Å sphere of TIP3P solvent at 300 K. These simulations take into account conformational disorder introduced by the dynamics of the system at finite temperature and for the fully solvated model, of competition for hydrogen bonding with the surrounding solvent.

Molecular Dynamics of NMA−Water Complexes. Molecular dynamics simulations of NMA with one, two, or three water molecules were performed at 300 K (see Materials and Methods). Figure 3a,b depicts a "slice" of coordinate space around the plane of the NMA molecule and shows the distribution of a single water molecule around the NMA oxygen as it explores the conformational space over the course of a 1ns simulation at 300 K (see Figure 3 legend for details). Consistent with the different contributions to the interaction energy between the two molecules mentioned above (e.g., Table 1), additional interactions between the polar groups (such as the TIP3P {O-}H with the N{-H} of NMA) cause a preference for water to bind to NMA{C═O} on the side of the NMA amide group. This is manifested in Figure 3a by an increased number of waters on this side of the NMA molecule (i.e., the lower side of distribution in Figure 3a is "fatter" and more dense). Figure 3b shows the out-of-plane positions of water molecules, where there is no preference. Because there is no hindrance from methyl groups, a greater deviation from bond linearity is allowed out off the plane (± 80°) than in the plane (± 45°) of the NMA molecules.

Fifteen coordinate frames were taken from the simulation and the NMA−water complex was minimized to convergence as above (see Methods). Two minima were found (see Tables 1 and 2, and Figure 3e) with one closer to the amide group of NMA (4.0 vs 4.9 Å for the TIP3P-oxygen to N distance) having a slightly lower energy and higher population. The minima have the TIP3P water in the plane of the NMA molecule (Figure 3f).

In the case of two TIP3P molecules binding to the NMA acceptor group, the individual water−NMA interaction energies are less favorable than for the NMA interacting with a single TIP3P during dynamics or the grid search and minimization (Table 1). Figure 3c shows that when two TIP3P molecules are bound to the carbonyl group of the NMA molecule, two regions of density result, one essentially in-line as described above and one at a larger distance and larger angle with respect to the acceptor. Examination of individual frames shows that the two TIP3P molecules are predominantly arranged in this way with respect to the NMA acceptor. The donor acceptor distance of one water is significantly longer than that of the other (the values are 1.71 versus 1.97 Å and beyond; see Figure 3g and Table 2 for configurations that result from minimization). It is therefore appropriate to refer to the two TIP3P waters separately as first and second "layer" according to their distance relative to the NMA acceptor and to consider their interaction energies individually. Table 2 lists interaction energies and geometries for the NMA−TIP3P water complexes (also after minimization of the structures). The interaction of the water molecules with the NMA{C═O} is not optimal because interactions between the two TIP3P waters also contribute. The two waters interact with each other with an interaction energy of around −6.5 kcal/mol for more than 80% of the time. This average arrangement found in the simulation contrasts with one in which both waters are in the plane, are equidistant from the acceptor and form hydrogen bonds to it with considerable deviation from linearity at the carbonyl oxygen (Figure 2e of Guo and Karplus[36]). One such arrangement was also sampled in our minimization of 15 structures. The latter is a minimum energy conformation found by ab initio calculations[36,54] and from molecular mechanics calculations.[55] These investigators constrained the NMA and waters to remain in the same plane. The peptide group is distorted by 6° from its planar configuration in this case. Recently, Demetropoulos et al.[56] relaxed this restraint and found an arrangement of two waters bound to NMA{C═O} similar to that of Figure 3c. However, when the NMA molecule is solvated more completely, the oxygen acceptor tends to interact with two almost equidistant TIP3P molecules as hydrogen bond acceptors. This has been observed in small molecule crystal structures[24] and is discussed further below. We find that such configurations exist for short times during the simulation of NMA with only two TIP3P molecules, but that overall energetically more favorable interactions are achieved by optimization of the TIP3P interaction with the other water molecule, which also interacts favorably with the N-terminal methyl group of NMA (see Table 1 and Figure 3g).

TABLE 2: Interaction Energies (of Non-Methyl Atoms) and Geometries after Minimization of 15 Structures Taken from the Vacuum Dynamics of NMA Complexes[a]

model system	geometry (r, θ, Ω)	total energy (vdW)
NMA−NMA	1.84 Å, 12°, 39°	−5.8 (0.4)
NMA{C═O}−TIP3P; 2 in [a]:	1.73 Å, 2°, 54°	−7.5 (1.2)
13 in [b]:	1.73 Å, 6°, 45°	−7.6 (1.1)
NMA{N−H}−TIP3P	1.90 Å, 1°, 0°	−5.2 (0.9)
TIP3P−TIP3P	1.78 Å, 3°	−6.9 (1.6)

NMA{C═O}−2TIP3P		geometry (r, θ, Ω)	total energy (vdW)
6 in [a]:	NMA{C═O}−TIP3P[i]	1.71 Å, 7°, 40°	−7.7 (1.4)
	NMA{C═O}−TIP3P[ii]	> 3.8 Å	−1.1 (−0.2)
	TIP3P[i]−TIP3P[ii]	1.76 Å, 4°	−7.2 (1.9)
8 in [b]:	NMA{C═O}−TIP3P[i]	1.73 Å, 7°, 73°	−7.1 (1.3)
	NMA{C═O}−TIP3P[ii]	3.25 Å, 56°, 108°	−2.7 (−0.4)
	TIP3P[i]−TIP3P[ii]	1.78 Å, 3°	−6.9 (1.6)
1 in [c]:	NMA{C═O}−TIP3P[i]	1.76 Å, 14°, 40°	−6.1 (1.1)
	NMA{C═O}−TIP3P[ii]	1.97 Å, 28°, 64°	−7.1 (0.1)
	TIP3P[i]−TIP3P[ii]	1.99 Å, 38°	−2.4 (1.0)

NMA {C═O/N−H}−3TIP3P		geometry (r, θ, Ω)	energy total (and vdW)
all:	NMA{C═O}−TIP3P[i]	1.76 Å, 6°, 79°	−6.9 (1.2)
	TIP3P[i]−TIP3P[ii]	1.77 Å, 7°	−7.2 (1.7)
	TIP3P[ii]−TIP3P[iii]	1.77 Å, 2°	−7.2 (1.8)
	NMA{N−H}−TIP3P[iii]	2.08 Å, 41°, 92°	−4.2 (0.8)

[a] Good convergence to a single or to two minima (e.g., for NMA{C═O}−TIP3P denoted by [a] and [b]) was found to all complexes except for NMA{C═O}−2 TIP3P, where the data suggest the possibility of three configurations with an energy minimum. Populations are indicated (refer also to Figure 4e−h).

Figure 3d shows an end-on view of the NMA molecule for the case of three water molecules. Two of the TIP3P waters are close to the NMA acceptor and donor groups, but one of them is not in a favorable hydrogen bonding position with respect to either group but forms a bridge between the other two waters (see Figure 3h). All three waters are either below or above the NMA plane. This asymmetric distribution suggests that sampling is incomplete in this calculation and that a significant activation barrier needs to be overcome to relocate all three waters on the opposite side of the NMA plane. This results in a significant deviation of the N−H bond from the plane defined by the C, O, and N atoms of the NMA molecule (an average of 7.4 ± 5.4° toward the TIP3P complex). The deviation is significantly larger than the 1−3° deviation typically seen in protein or solvent exposed peptide groups.[57] The simulation results for more than one donor and acceptor group demonstrate the competitive nature of the interactions, which are optimized not with respect to the individual groups but for the entire system.

NMA Monomer and Dimer in a Water Sphere. The effect of multiple competing donor and acceptor groups on hydrogen bonding was examined further by placing a single NMA molecule and the NMA dimer at the center of a 14 Å radius sphere containing 388 and 384 TIP3P water molecules, respectively. Conformations were sampled every 0.2 ps in a 1 ns molecular dynamics simulation at 300 K. Because the NMA−NMA interaction energy is slightly unfavorable as compared to the NMA−solvent interactions (the free energy of dimerization has been calculated to be positive by about 1.1 kcal/mol[52,58]), a restraint was applied in the dimer simulation to prevent separation of the $N-H^1$ and $^2C{=}O$ groups by more than 3.8 Å. In the 1 ns simulation, 238 instances were recorded during which the distance between donor and acceptor were less than 2.5 Å. If these configurations are defined as the dimer (the minimum energy distance is 1.8 Å), such dimerization lasted 0.72 ps on average and the ratio of dimer/monomer configurations is 0.21, i.e., the monomer configuration is more stable than the dimer, even in this constrained space. However, the statistics are not good enough to determine an equilibrium constant and correct for the restraint on the system.[59]

The location of the minima of the potential energy surface and the magnitude of the energies are similar to those found in the grid search and minimization in a vacuum, as can be seen by comparing with Figure 2a−c (minimization) and with Figure 4a−c (dynamics), as well as with Tables 1 and 3. The values for the energy at the minimum of the potential energy surfaces of the NMA−NMA and of the NMA−TIP3P interactions in the solvent sphere are between those obtained from dynamics simulation and minimization in a vacuum. The differences arise from possible sampling errors, the fact that the results are at a finite temperature and that a reduced coordinate set is used in the grid search (Figure 2). This means that the distribution of the other degrees of freedom are weighted differently in the presence of solvent due to competing interactions.

One additional degree of freedom is the angle Ω (for a definition see Figure 1b). Figure 5a,i shows the distribution of the potential energy as a function of the observed deviation from hydrogen bond linearity at the acceptor (Ω) for NMA−NMA contacts found in the simulation of the NMA dimer in a solvent sphere. A donor−hydrogen acceptor distance range of 1.75−1.85 Å and donor angle, θ, of 13−22° is used. This region of r, θ space is highly populated because it corresponds to strong short-range hydrogen bonds. By contrast to the results on the NMA dimer from the grid search and minimization (Figure 2), a range of angles (Ω from grid search ± 30°) and of potential interaction energies are obtained. A slightly greater range occurs at the larger distance, where the center of the distribution has shifted to greater Ω values (see Table 4). At 1.75−1.85 Å, the total (non-methyl) interaction energy increases by 0.16 kcal/mol for a 10° change in Ω, confirming that the total energy is relatively insensitive to changes in Ω over a broad range of this angle. This insensitivity is responsible for the large spread in Ω for a given range of r, θ space. Moreover, since r, θ, and Ω do not define the positions of all eight atoms that contribute to the interaction energy, different values are found for the system having the same r, θ, and Ω parameters.

Figure 4 shows that the potential surface of TIP3P bonding to the amide hydrogen is steeper, but less deep than that of the other interactions (minimum at −3.9 kcal/mol for NMA{N−H}−TIP3P as compared to −5.2 and −6.3 kcal/mol for NMA−NMA and NMA{C═O}−TIP3P, respectively). The differences in the interaction energies of the solvated complexes are slightly greater than those in the vacuum minimization (compare Tables 1 and 3). The considerable preference for hydrogen bonding at the acceptor (NMA−{C═O}...TIP3P, Figure 4c), manifest by the stronger interaction energy as compared to that of waters bound to NMA via its amide group, is in agreement with experimental data on the frequency of such interactions in protein crystal structures.[6,24]

Probability Distribution in Hydrogen Bond Parameter Space. Figure 6a−c shows a contour plot of the number of NMA−NMA and NMA−TIP3P as a function of the hydrogen bonding parameters, r and θ, used in the potential energy plots (Figure 2). There is a considerable preference for hydrogen bonding contacts to populate nonlinear arrangements with increased donor acceptor separation (Figure 6). For example, the maximum number of N−H...O═C and C═O...TIP3P contacts occur with a donor−donor−hydrogen−acceptor angle (θ) deviating nearly 20° from linearity (approximately 10° for N−H...TIP3P). Similar trends have been found in the Cambridge database of small molecules.[26] This is in accord with the fact that the potential energy difference between linear and nonlinear hydrogen bond arrangements is small (around 0.5 kcal/mol in Figure 2) and even slightly less in the solvated system (Figure 4). Also, even for a flat potential energy surface, nonlinear configurations are more probable than a linear arrangement by a factor of sin θ.

Although the potential energy of the interaction between NMA molecules at a distance r(H...A) of 3.4 Å and a deviation θ(D−H...A) of 100° from bond linearity is only marginally favorable (Figure 4a), a large population of hydrogen bonding contacts is seen in this region in Figure 6a. This arises because solvent molecules compete with the hydrogen bonding interaction, resulting in non-optimal interaction geometries between the two NMA molecules. The result is in accord with the fact that the hydrogen bond geometries for two molecules are affected by the presence of other molecules (see section 2 and Figure 3).

(3) Hydrogen Bond Interactions in a Protein. Interaction energies and probability distributions were also calculated for the C═O and N−H peptide groups belonging to the mainchain of human lysozyme from coordinate frames of a 200 ps simulation at 300 K (see Methods). The potential energy surfaces are plotted as a function of hydrogen bond parameters (Figure 4e−f). The shapes of the potential energy surfaces that are derived from the protein are similar to those of the NMA dimer and NMA−TIP3P complexes. Empirical interaction parameters, which have been derived for the NMA model system (ref 30;

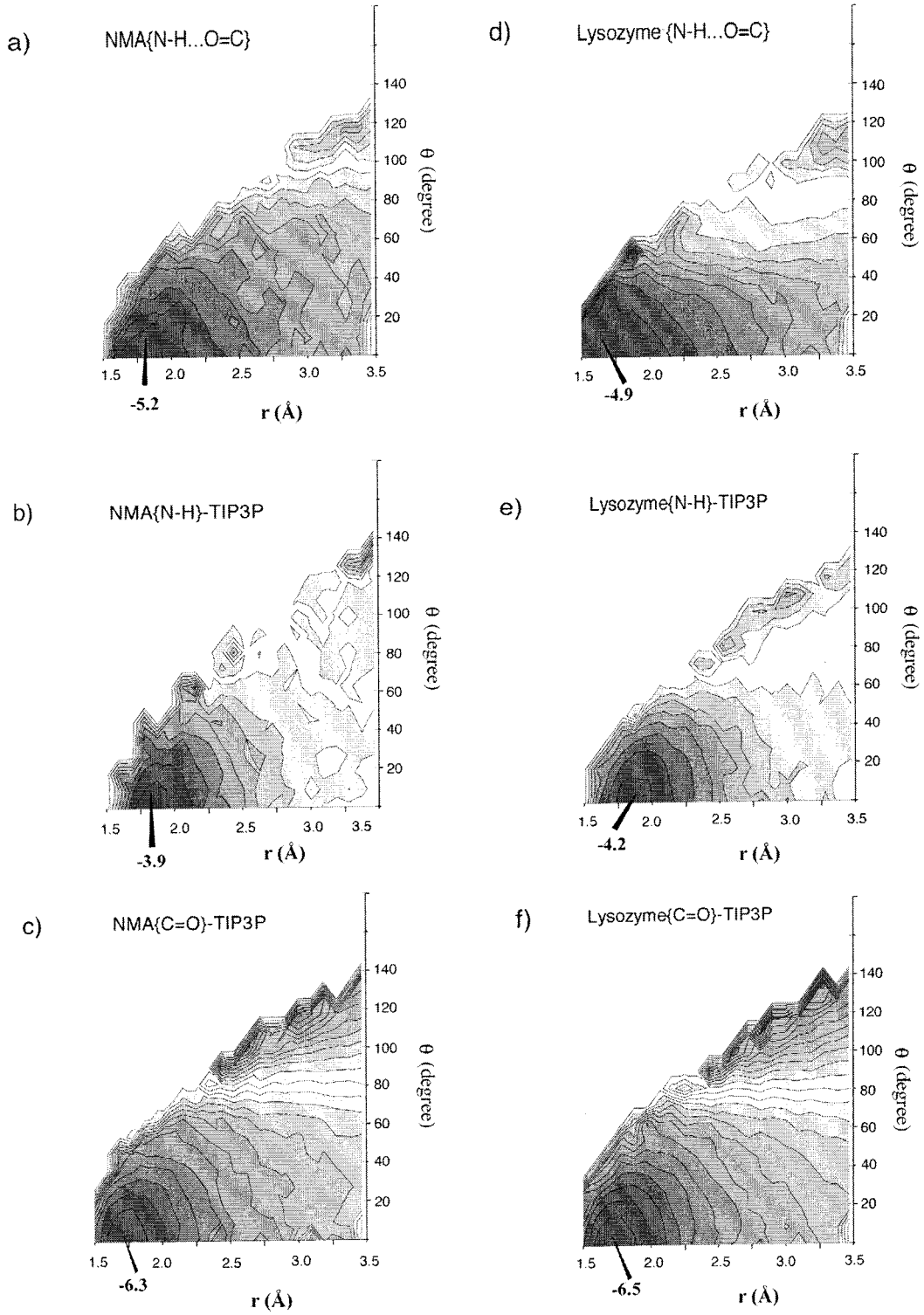

Figure 4. Potential energy surfaces for solvated NMA complexes and human lysozyme sampled by molecular dynamics. (a) NMA dimer; (b) and (c) TIP3P interacting with an NMA amide hydrogen, and carbonyl oxygen, respectively; (d) mainchain−mainchain interactions in human lysozyme; (e) and (f) interactions of mainchain amides and oxygens with TIP3P, respectively. The surfaces were generated by averaging the energies for all conformations with common hydrogen bond parameters; r(H...A) and θ(D−H...A) (see Methods). The contour levels are plotted at intervals of 0.5 kcal/mol; lighter gray shade at higher energy. Refer to Table 3 for the minima of the surfaces.

Kuchnir and Karplus, unpublished), form the basis of parameters for the protein peptide group in the CHARMM potential function param22.[31] Since we calculate the interaction energy between two groups of atoms, the effects of other atoms that are present in the surroundings do not appear in the energy surface. It follows that common configurations of the NMA, TIP3P, and the protein peptide groups have similar energies. The similar shape of the potential energy surface implies that the mainchain donors and acceptors assume configurations in the context of a complex folded protein that are similar to those found in the NMA dimer. Their relative probabilities are described below.

The range of potential energies and hydrogen, acceptor,acceptor−antecedent angles (Ω) that are possible for a given range of r and θ, are similar for the NMA model and the protein (Figure 5; Table 4). A slightly larger range of Ω angles is sampled in the NMA complexes as compared to the protein

TABLE 3: (Non-Methyl) Interaction Energies in kcal/mol, Geometries (r, θ, Ω) from MD Simulation (Solvated Systems)

	NMA−NMA		NMA{C═O}−TIP3P		NMA{N−H}−TIP3P	
(a) min	1.8 Å, 0°, 44°		1.8 Å, 10°, 27°		1.9 Å, 10°	
(b) ave	1.8 Å, 0°, 58°		1.7 Å, 0°, 49°		1.9 Å, 0°	
	(a)	(b)	(a)	(b)	(a)	(b)
VDW	0.93	1.07	0.55	1.22	0.46	0.58
TOT	−6.06	−5.19	−8.02	−6.32	−5.07	−3.89

human lysozme	N−H...O═C mc-mc		mc. C═O-TIP3P		mc. NH-TIP3P	
(a) min (polar)	1.9 Å, 10°, 50°		1.7 Å, 10°, 38°		1.8 Å, 0°	
(b) all polar ave	1.7 Å, 0°, 23°		1.7 Å, 0°, 45°		1.9 Å, 0°	
	(a)	(b)	(a)	(b)	(a)	(b)
VDW	0.47	0.54	1.01	1.21	1.17	0.64
TOT	−6.18	−4.91	−7.61	−6.46	−5.26	−4.17

and the average energy is slightly lower. The results indicate that the N−H...O═C and C═O...TIP3P hydrogen bonds are confined to a similar range of deviations from linearity at the acceptor (Ω) and that these conformations have similar energies for NMA and the protein.

Figure 6 shows the probability distributions for donor and acceptor contacts as a function of r and θ for the NMA complexes and for the protein. There is a good correspondence between these surfaces. The similarities include the relative scarcity of conformations with near linear hydrogen bonding (<20−30° about the donor hydrogen) at large distances of donor−acceptor separation (greater than 2.4 Å), shown in Figure 6d−f (and see below for scaling). The location of the maxima in the population of donor and acceptor contacts in the r, θ parameter space correspond closely to those seen in solvated NMA, both for the NMA dimer versus the protein, and for NMA−water versus protein−water. The relative populations are also similar for the latter two, but significantly different for the N−H...O═C hydrogen bonds.

The relative population of NMA dimers which are tightly hydrogen bonded (r < 2.5 Å, θ < 45°) is much smaller than that for strong mainchain−mainchain hydrogen bonds in the protein (44 in Figure 6a versus 1550 in Figure 6d). This difference is due to the cooperative effect of secondary structure in the protein. The overall loss of entropy in bringing donor and acceptor groups together is shared by several hydrogen bonding sites in proteins, which stabilizes close donor−acceptor interactions relative to those of a single hydrogen bond that binds two NMA molecules together. (In this regard, we note that the free energy of association of the two NMA molecues has been shown to be unfavorable.[52,58])

We showed above that TIP3P binding to NMA{N−H} is weaker than interactions of TIP3P with other waters or C═O groups. This is in accord with the results of Mitchell and Price.[31] In addition, comparison of the NMA{N−H}−TIP3P and the lysozyme{N−H}−TIP3P contact distributions suggest that water molecules bind less frequently to the NMA amide group than to protein amide groups (Figure 6b vs 6e), even though only 20 of the 127 mainchain N−H groups are (>30%) solvent accessible in the protein. This result suggests that N−H...TIP3P hydrogen bonding could be more favorable near the protein surface as compared to that in the NMA molecule. Partial shielding the TIP3P group from interactions with other groups (esp. other waters) may allow a better interaction geometry with the amide and interaction of a TIP3P molecule with other groups of the protein prior to interacting with the amide group may reduce some of the entropy cost of hydrogen bonding.

The probability surfaces for the distribution of NMA{C═O}−TIP3P and lysozyme{C═O}−TIP3P contacts are very similar (Figures 5c,d and 6c,f), as mentioned above. An examination of trajectory frames shows that two water molecules are preferentially bound to the NMA oxygen at near equal distance with nonlinear hydrogen bonds. This contrasts with that seen in the vacuum dynamics simulation (Figure 3c) where two water molecules arranged themselves in a "layered" configuration to optimize interactions between them, with only one water interacting with the NMA hydrogen bond acceptor in a near linear arrangement. Thus, in the presence of other protein groups and additional water molecules, optimization of relatively distant pairwise interactions no longer leads to the most favorable configuration because of more proximate competing interactions. Indeed, 74% of the donor and 87% of the acceptor groups in human lysozyme that have a hydrogen bonding partner at between 2.5 and 3.5 Å have at least one hydrogen bonding partner in closer proximity and with more optimal hydrogen bond geometry.

Statistical Distribution of Protein−Protein Hydrogen Bonds. Figure 7 shows the distribution of mainchain protein−protein donor−hydrogen acceptor pairs over the course of the human lysozyme simulation at 300 K (same as Figure 6d) in comparison with those seen in 415 high-resolution nonhomologous crystal structures. Classical, strong hydrogen bonds are seen as a population maximum at a distance near 1.9 Å and at an angle of 15−20° from linearity. Donor acceptor contacts that are medium range (2.5−3.5 Å) are also significantly populated. A pronounced clustering of these interactions is found in a statistical survey of protein crystal structures (labeled peaks in Figure 7b). The plots show that the density distribution of donor−acceptor contacts is different from that in lysozyme, with a maximum due to bifurcated hydrogen bonds in helical structures, (i.e., forming $i \rightarrow i + 3$ and $i \rightarrow i + 4$ contacts) being more pronounced in the database. This suggests that individual proteins may have significant deviations from the average over the crystal structures in the protein database. The latter results are similar to those described previously.[6,14,15]

It is of interest to compare the results shown so far with the population density (population per Å^3; see Methods) as a function of r and θ. In their survey of hydrogen bonds in a database of high-resolution crystal structure, McDonald and Thornton[14] found that the scaled population density of hydrogen donor−acceptor contacts is small beyond 2.3 Å and when deviations from hydrogen bond linearity are greater than 60°. This is in accord with our analysis of 415 nonhomologous protein structures (see positions of the minima in the density distributions in Figure 6). The number of short-range hydrogen bond contacts (peak at 1.8 Å, 0°) per unit volume are approximately 40- and 25-fold greater than hydrogen bond contacts populated at greater interaction distances (range 2.85−2.95 Å, 75−85°) in lysozyme and the protein databank, respectively. The minimum in the population of hydrogen bonding contacts present in the surfaces of Figures 6 and 7 near 2.5 Å donor hydrogen acceptor separation is also seen in the

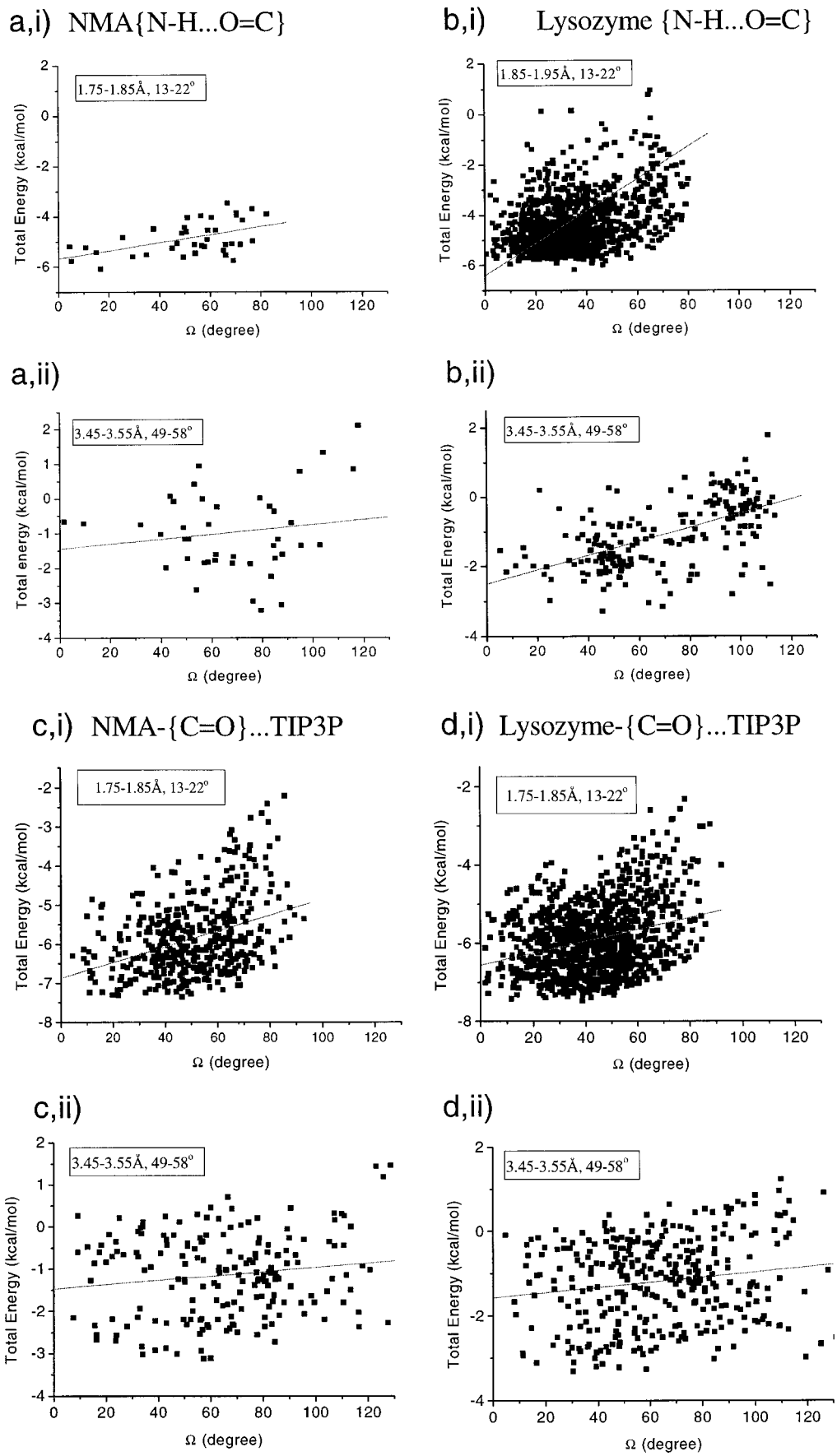

Figure 5. Total energy as a function of angle Ω for a series of hydrogen bonding interactions; (a) NMA dimer; (b) lysozyme mainchain−mainchain; (c) NMA-{C═O}...TIP3P; and (d) lysozyme-{C═O}...TIP3P. The plots show the interactions that occur during the simulation in a particular region of r, θ space. In series (i) the r, θ space contains strong interactions between the donor and acceptor (1.75−1.85 Å and 13−22° for r and θ, respectively), whereas series (ii) shows a region with weaker interactions, typically bifurcated hydrogen bonding (3.45−3.55 Å and 49−58° for r and θ). A least-squares fit though the points is shown.

TABLE 4: Averages and Standard Deviations for Ω (degrees) and Potential Energies (kcal) Shown in Figure 5

	NMA−NMA			lysozyme−{C═O...H−N}	
figure	ave Ω ± std(°)	ave E ± std (kcal/mol)	figure	ave Ω ± std	ave E ± std
a,i	51.6 ± 20.1	−4.9 ± 0.7	b,i	32.5 ± 16.0	−4.2 ± 2.5
a,ii	67.7 ± 25.4	−1.0 ± 1.2	b,ii	67.8 ± 26.9	−1.1 ± 1.0
	NMA−{C═O}...TIP3P			lysozyme-{C═O}...TIP3P	
figure	ave Ω ± std(°)	ave E ± std (kcal/mol)	figure	ave Ω ± std	ave E ± std
c,i	50.4 ± 17.4	−5.9 ± 1.0	d,i	43.7 ± 17.2	−5.9 ± 0.9
c,ii	65.8 ± 28.3	−1.1 ± 1.0	d,ii	62.9 ± 26.1	−1.2 ± 1.0

surface after normalization but is a less prominent feature. The ratio of the number of contacts in the region 2.85−2.95 Å, 75−85° vs 2.45−2.55 Å, 55−65° is equal to 1.1 and 1.4 per unit volume for lysozyme and the protein databank, respectively. These numbers compare with 1.7 and 2.5 for the unscaled distributions for lysozyme and the protein databank, respectively.

In the lysozyme molecular dynamics simulation the population maxima are slightly broadened compared to the crystallographic data (see Figure 7a). This relative broadening is a consequence of the proteins flexibility at a simulation temperature of 300 K and suggests that the statistical results, which correspond to the best fit of a single structure to the X-ray data do not give a "free energy" surface, although B-factors could be used to correct for this. In the following section, we suggest that the minimum in the population of hydrogen bonding contacts at 2.5 Å has a physical origin that supports the use of a geometric criterion for the definition of hydrogen bonding.

(4) Criteria for Hydrogen Bonding. The usual analysis of molecular dynamics trajectories and of protein structures assumes implicitly that hydrogen bonds are either present or absent. Hydrogen bonds have been traditionally defined either by geometric criteria (e.g., MacDonald and Thornton[15]) or by energy-based criteria based on the interaction of single N−H and C═O dipole vectors.[11,23] The present simulations provide results for the interaction as of two complete peptide groups (i.e., two N−H−C═O groups) or a peptide group and water. The analysis in terms of energy and probability (free energy) surfaces, makes possible a determination of whether the standard criteria are physically meaningful or arbitrary, but useful.

As we demonstrated in section 1, the potential energy surfaces resulting from mapping of all-atom interaction energies or interaction energies between two N−H, C═O pairs for the mainchain imply that an energy based cutoff is not suitable. The decrease in the all-atom interaction energy with increasing hydrogen bond distance in the NMA dimer is more gradual than the "hydrogen bond energy", calculated between the C═O,[1] H−N[2] dipoles alone. Thus, for example, the all-atom and the C═O,[1,2] N−H[2,1] pair interaction energy for dimerization decreases by only 30% and 35%, respectively, between the minimum for r_{O-H} near 1.9 Å and a distance of 2.5 Å. Thus, an energy bound cutoff is complicated by the nature of the rather flat potential surface as a function of donor−acceptor distance and by the significant contributions from groups, in addition to those directly involved in the hydrogen bond.[51] This conclusion is in accord with other results in the literature,[59−61] suggesting that an energy based description of donor and acceptor group interactions in proteins can be complicated by their dependence on a specific environment.

Examination of the distributions of the population of donor−acceptor contacts in Figure 6 and Figure 7a,b shows clearly that the choice of the cutoff criterion has a direct effect on the number of interactions that are included as hydrogen bonds, particularly the multicenter hydrogen bonds (i.e., donors and acceptors with more than one hydrogen bond partner within the cutoff). In fact, more than 75% of the mainchain amide donors and more than 90% of mainchain carbonyl oxygen acceptors interact with more than one acceptor or donor site, respectively, in the human lysozyme simulation when an acceptor−donor hydrogen distance cutoff of 3.5 Å is used to identify hydrogen bonds in the unscaled distributions. When water molecules are taken into account, this is true for all but two of the mainchain oxygens and for 85% of the mainchain amides. If a cutoff of 2.5 Å and an angle cutoff of 90° is used, a second hydrogen bond partner is found for only 30% of the mainchain hydrogen bond acceptors and for less than 5% of mainchain donors. The latter numbers are in agreement with those reported in surveys of protein mainchain solvation and protein−protein hydrogen bond interactions in crystal structures.[6,15,24]

The population of donor acceptor contacts at a distance of 2.4 to 2.6 Å with a deviation of around 50° from donor−hydrogen acceptor linearity is significantly smaller than for distances < 2.4 Å, angles < 40°, and distances > 2.6 Å and deviation > 50° in both the simulation and in protein crystal structures (Figure 7). Radial distribution functions provide an alternative way of analyzing the population of donor−acceptor contacts. The radial distributions give rise to potentials of mean force (Figure 8 and methods) and have been used extensively for the examination of structure in water,[63] solvated NMA,[64] simple peptide model systems[65] and proteins.[66] All donor−acceptor distribution functions are characterized by a minimum in the population of hydrogen bonding contacts near 2.5 Å. To determine the physical origin of this phenomenon, we use donor−acceptor contacts in TIP3P water as an example and calculated radial distribution functions for a box of water molecules (Figure 8a). We also used a box of modified water molecules (Figure 8b), denoted TIP3P^{e0}, for which all partial charges were set to zero (compare Lazaridis[67]). The latter only has a hydrogen and an oxygen atom van der Waals term without electrostatic terms. A deep minimum is seen at 2.5 Å for r_{O-H} in the radial distribution function of g(OH) of TIP3P, whereas the TIP3P^{e0} system, with the same density, shows a more shallow minimum at $r_{O...H}$ of 4.5 Å, very similar to the position of the minimum seen in a Lennard-Jones liquid with particles of 2.8 Å diameter.[45] Such minima in radial distribution functions are anticipated in any dense system (e.g., spherical particles packing around a central one) purely because of spatial restraints even if there are no preferential interactions.

Weeks et al.[68] used a simple van der Waals solvent to show that ordering of liquids was due primarily to repulsive forces. Preferential (molecular shape dependent) or directional interactions, such as partial atomic charges, can introduce additional order in molecular systems. Specifically, in the case of TIP3P water the first peak at 1.9 Å for g(OH) arises from two hydrogen bonded water molecules (as does g(OO) at 2.8 Å). The second peak of g(OH) at 3.2 Å is related to the tetrahedral structure of

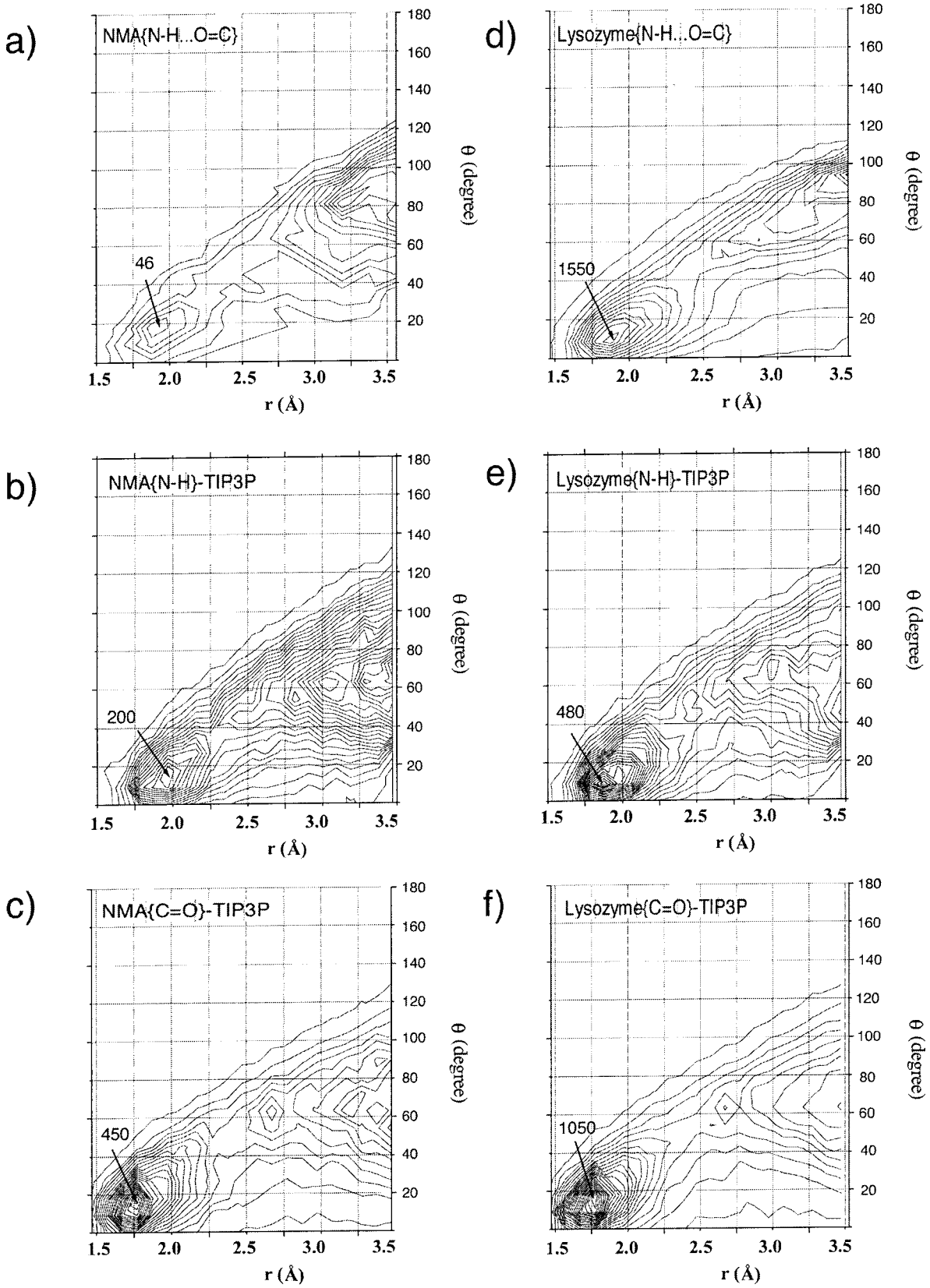

Figure 6. Probability distribution of the hydrogen bonding interactions as a function of r and θ for (a) the NMA-dimer; (b) and (c) TIP3P interacting with in NMA amide hydrogen and carbonyl oxygen, respectively; (d) mainchain-mainchain interactions in human lysozyme; (e) and (f) interactions of mainchain amides and oxygens with TIP3P, respectively. The 15 and 13 contours in plots panels a and d and the 24 contours in the other panels are spaced uniformly, with maxima indicated, allowing a comparison of the distributions with each other and with the potential energy surfaces (Figure 5). The data were derived from 5000 coordinates sets of a 1 ns simulation of solvated NMA molecules and from 500 coordinates of a 200 ps simulation of human lysozyme.

near neighbors and corresponds to the donor−acceptor distance between two water molecules that are hydrogen bonded to a third, bridging water molecule (i.e., the accepting atom belongs to a water molecule in the second solvation sphere, and typically

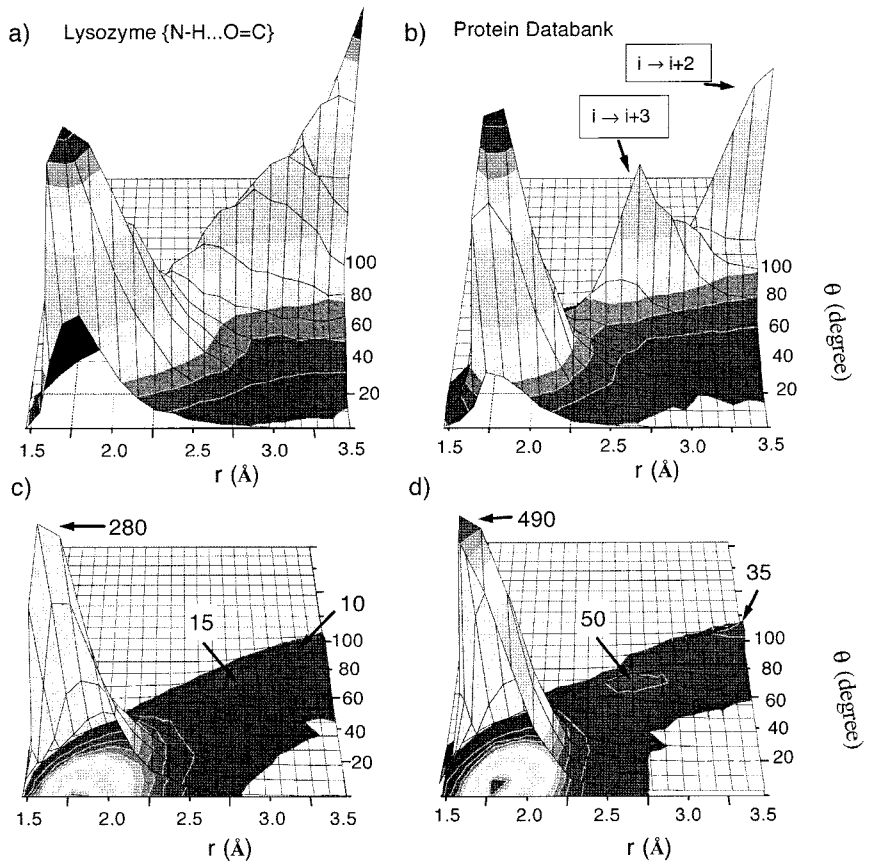

Figure 7. (a) Probability distribution of protein−protein mainchain hydrogen bonding interactions as a function of r and θ over the course of the 300 K trajectory of human lysozyme sampled at 0.4 ps intervals over 200 ps (contour plot is spaced linearly at intervals of 20 donor and acceptor pairs occurring in an area of 0.1 Å by 10°. (b) As in panel a but derived from 415 high-resolution nonhomologous structures of the protein crystal structure data bank with a resolution < 2.0 Å and R-factor < 20%. The maxima of the distributions are labeled and correspond to those identified by McDonald and Thornton[15] (see text). Panels c and d show the distributions of panels a and b after rescaling using a volume density element (see Methods).

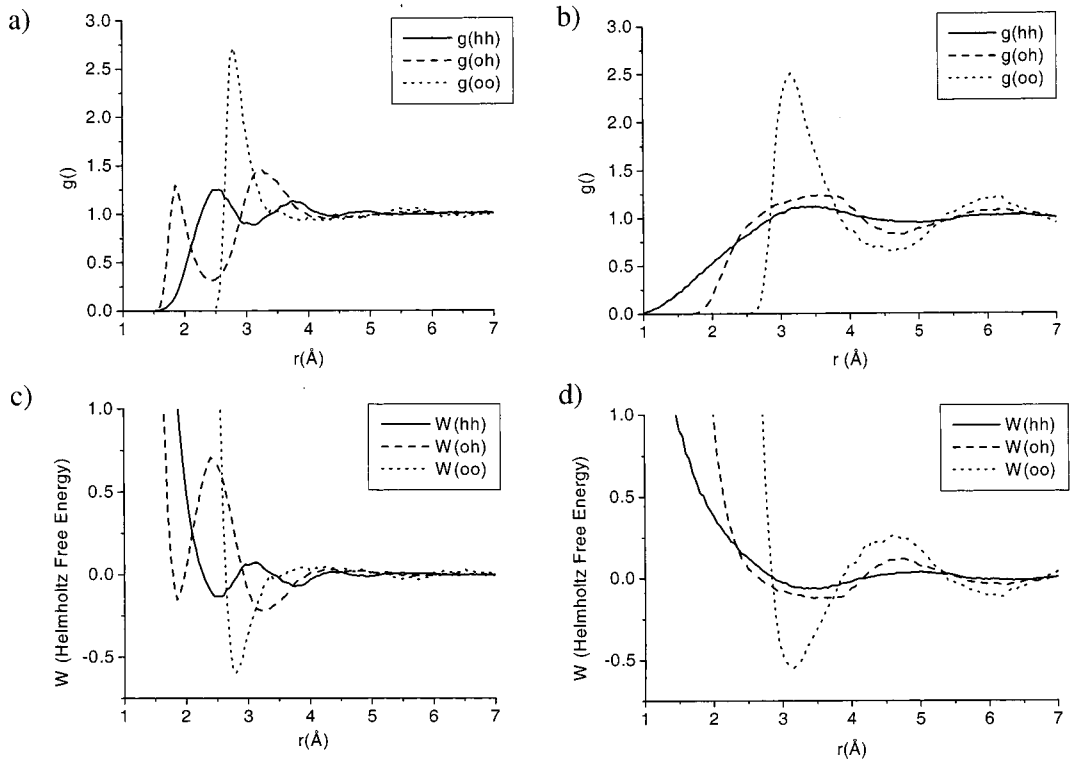

Figure 8. Radial distribution functions (a) TIP3P water (b) TIP3P^{e0} water without partial charges, van der Waals only model (see text). (c) Potential of mean force (in kcal/mol) for TIP3P and (d) for TIP3P^{e0} water.

it has more than one hydrogen bonding partner, although non-hydrogen bonded configurations also contribute). Thus, the minimum at 2.5 Å for g(OH) relates to a region that spatially excludes donor−acceptor interactions in this ordered environment. Indeed, as thermal motion is increased (i.e., the temperature is raised) the depth of this minimum, as well as the height of the surrounding peaks in g(OH) are decreased.[68,69] Minima in the distribution of donor−acceptor contacts in solvated NMA and lysozyme are related to similar ordering that results from hydrogen bonding and show a corresponding temperature dependence (Buck and Karplus, in preparation). Integration of the area under the radial distribution function to 2.5 Å gives the number of water molecules that are hydrogen bonded in the first solvation shell as 3.8 (close to 4 for a perfect tetrahedral lattice, see Lazaridis and Karplus[69]), and therefore the existence of a minimum in the distribution function provides a natural criterion for the definition of hydrogen bonding by a geometric cutoff.

Concluding Discussion

The present study has examined the energy surfaces arising from hydrogen bonding interactions and the distribution of donor and acceptor groups obtained by minimization and molecular dynamics simulation, as well as by statistical methods based on known protein structures. We use the donor hydrogen−acceptor distance, r, and the angle of deviation of the donor, donor hydrogen and acceptor atom from a linear arrangement, θ, as the principal geometric parameters for describing hydrogen bonds. The potential energy surfaces for hydrogen bonds as a function of these two parameters are shown to be very similar for models involving NMA dimers and NMA−TIP3P complexes and correspond closely with those derived from hydrogen bonding interactions in a typical protein. This similarity is in part a consequence of the fact that we calculate the energetic contributions of distinct donor and acceptor groups and neglect the influence of other closeby groups. The latter appear not to be important because the probability distributions of donor−acceptor contacts for the NMA dimer model system in solution and for a solvated protein at 300 K are very similar. This suggests that hydrogen bonding interactions are dominated by the local energy, not by spacial ordering constraints provided by the protein environment. The minima in the population distribution of donor acceptor separation arise from volume exclusion effects as modified by electrostatic interactions. The shape of the population distribution (or free energy) surface for hydrogen bonds reflects the underlying energetics of the groups considered, but at longer distance there are perturbations due to the competition with more proximate groups for good hydrogen bond geometry and energetics.

The consideration of the interaction of entire peptide groups and of fully solvated systems made here allow us to relate the cutoff distance to the minimum in the distribution of donor−acceptor interactions in the NMA model system and in proteins. The value obtained is 2.5 Å, in agreement with the traditional value.

Acknowledgment. We thank Prof. Andrej Sali and Dr. John-Marc Chandonnia for help with the construction of the database of 415 nonhomologous protein crystal structures and Dr. Christian Bartels for suggesting the TIP3P^{e0} model. Some of the calculations were done at the Pittsburgh Supercomputer Center.

References and Notes

(1) Pauling, L. C. *The Nature of the Chemical Bond*. Cornell University Press, 1960.
(2) Eisenberg, D.; Kauzmann, W. *The Structure and Properties of Water*. Oxford University Press: New York, 1960.
(3) Franks, F.; Ed. Water: A comprehensive treatise. Plenum Press: New York, 1972.
(4) Schuster, P.; Zundel, G.; Sandorfy, A. The Hydrogen Bond. Elsevier Pub. Co.: New York, 1979.
(5) Glusker, J. P. *Top. Curr. Chem.* **1998**, *198*, 2−56.
(6) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97−179.
(7) Mirsky, A. E.; Pauling, L. *Proc. Natl. Acad. Sci. U.S.A.* **1936**, *22*, 439−447.
(8) Pauling, L. C.; Corey, R. B. *Proc. R. Soc. London, Ser. B* **1953**, *141*, 21−33.
(9) Karplus, P. A.; Faerman, C. *Curr. Opin. Struct. Biol.* **1994**, *4*, 770−776.
(10) Covell, D. G.; Wallquist, A. *J. Mol. Biol.* **1997**, *269*, 281−297.
(11) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.
(12) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Proteins*. Chapter 19 from Hydrogen Bonding in Biological Structures; Springer, New York, 1991
(13) Stickle, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D. *J. Mol. Biol.* **1992**, *226*, 1143−59.
(14) McDonald, I. K.; Thornton, J. M. Hydrogen bonding in globular proteins and Atlas of side chain and mainchain hydrogen bonding in proteins. Material supplementary to McDonald and Thornton, 1994.
(15) McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* **1994**, *238*, 777−793.
(16) Bordo, D.; Argos, P. *J. Mol. Biol.* **1994**, *243*, 504−519.
(17) Xu, D.; Tsai, C.-J.; Nussinov, R. *Protein Eng.* **1997**, *10*, 999−1012.
(18) Saenger, W. *Annu. Rev. Biophys. Biophys. Chem.* **1987**, *16*, 93−114.
(19) Teeter, M. M. *Annu. Rev. Biophys. Chem.* **1991**, *20*, 577−600.
(20) Levitt, M.; Park, B. H. *Structure* **1993**, *1*, 223−226.
(21) Sidhu, K. S.; Goodfellow, J. M.; Turner, J. Z. *J. Chem. Phys.* **1999**, *110*, 7943−7950.
(22) Thanki, N.; Thornton, J. M.; Goodfellow, J. M. *J. Mol. Biol.* **1991**, *221*, 669−691.
(23) Hooft, R. W. W.; Sander, C.; Vriend, G. *Proteins: Struct. Funct., Genet.* **1996**, *26*, 363−376.
(24) Mills, J. E. J.; Dean, P. M. J. *Comput. Aided Mol. Design* **1996**, *10*, 607−622.
(25) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc.* **1998**, *120*, 8293−8297.
(26) Cornilescu, G.; Ramirez, B. E.; Frank, M. K.; Clore, G. M.; Gronenborn, A. M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *121*, 6275−6279.
(27) Harris, T. K.; Mildvan, A. S. *Proteins: Struct. Funct., Genet.* **1999**, *35*, 275−282.
(28) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(29) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(30) Reiher, W. D. Theoretical studies of hydrogen bonding. Ph.D. Thesis, Harvard University, 1985.
(31) Mitchell, J. B. O.; Price, S. L. *J. Phys. Lett.* **1991**, *180*, 517−523.
(32) Brooks, C. L.; Karplus, M.; Pettitt, B. M. Proteins: A theoretical perspective of dynamics, structure and thermodynamics. *Adv. Chem. Phys. LXXI*; Wiley & Sons: N. Y., 1988.
(33) Brooks, C. L.; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312−6325.
(34) Loncharich, R. J.; Brooks, B. R. *Proteins: Struct. Funct., Genet.* **1989**, *6*, 32−45.
(35) MacKerell, A. D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 10559−10560.
(36) Guo, H.; Karplus, M. *J. Phys. Chem.* **1992**, *96*, 7273−7287.
(37) Gelin, B. R.; Karplus, M. *J. Am. Chem. Soc.* **1975**, *97*, 6996−7006.
(38) Rychaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(39) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(40) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511−519.
(41) Brünger, A.; Brooks, C. L.; Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495−500.
(42) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(43) Allan; M. P.; Tildesley, D. J. Computer simulations in liquids. pp 181−185. Oxford University Press: New York, 1987
(44) McQuarrie, D. A. Statistical Mechanics, Harper Collins Pub. Macmillan Press: London, 1976.

(45) Artymiuk, P. J.; Blake, C. C. F. *J. Mol. Biol.* **1981**, *152*, 737−762.
(46) Brünger, A. T.; Karplus, M. *Proteins: Struct. Funct., Genet.* **1988**, *4*, 148−156.
(47) Adalsteinsson, H.; Maulitz, A. H.; Bruice, T. C. *J. Am. Chem. Soc.* **1996**, *118*, 7689−7693.
(48) Guo, H.; Karplus, M. *J. Phys. Chem.* **1994**, *98*, 7104−7105.
(49) Gao, J.; Freindorf, M. *J. Phys. Chem A* **1997**, *101*, 3182−3188.
(50) Dixon, D. A.; Dobbs, K. D.; Valentini, J. J. *J. Phys. Chem.* **1994**, *98*, 13435−13439.
(51) Lazaridis, T.; Archontis, G.; Karplus, M. *Adv. Protein Chem.* **1995**, *47*, 262−344.
(52) Ben-Tal, N.; Sitkoff, D.; Topol, I. A.; Yang, A.-S.; Burt, S. K.; Honig, B. *J. Phys. Chem. B* **1997**, *101*, 450−457.
(53) Grzybowski, B. A.; Ishchenko, A. V.; DeWitte, R. S.; Whitesides, G. M.; Shaknovich, E. I. *J. Phys. Chem. B* **2000**, *104*, 7293−7298.
(54) Markham, L. M.; Hudson, B. S. *J. Phys. Chem.* **1996**, *100*, 2731−2737.
(55) Baudry, J.; Smith, J. C. *J. Mol. Struct. (Theochem.)* **1994**, *308*, 103−113.
(56) Demetropoulos, I. N.; Gerothanassis, I. P.; Vakka, C.; Kakavas, C. *J. Chem. Soc. (Faraday Trans.)* **1996**, *92*, 921−931.
(57) Buck, M.; Karplus, M. *J. Am. Chem. Soc.* **1999**, *121*, 9645−9658.
(58) Jorgensen, W. L. *J. Am. Chem. Soc.* **1989**, *111*, 3770−3771.
(59) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707−2714.
(60) Apaya, R. P.; Bondi, M.; Price, S. L. *J. Comput. Aided Mol. Design* **1997**, *11*, 479−490.
(61) Nielsen, J. E.; Andersen, K. V.; Honig, B.; Hooft, R. W. W.; Klebe, G.; Vriend, G.; Wade, R. C. *Protein Eng.* **1999**, *12*, 657−662.
(62) Oliva, M. T.; Moult, J. *Protein Eng.* **1999**, *12*, 727−735.
(63) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(64) Yu, H.-A.; Pettitt, M.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 2425−2434.
(65) Rossky, P. J.; Karplus, M. *J. Am. Chem. Soc.* **1979**, *101*, 1913−1937.
(66) Sippl, M. J. *J. Mol. Biol.* **1996**, *260*, 644−648.
(67) Lazaridis, T. *J. Phys. Chem. B* **1998**, *102*, 3542−3550.
(68) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237−5244.
(69) Lazaridis, T.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 4294−4315.