Cooperativity in Amide Hydrogen Bonding Chains. Relation between Energy, Position, and H-Bond Chain Length in Peptide and Protein Folding Models

Nadya Kobko and J. J. Dannenberg*

Department of Chemistry, Hunter College and the Graduate School, City University of New York, 695 Park Avenue, New York, New York 10021

Received: August 24, 2003

The individual H-bond energies have been calculated at the B3LYP/D95** level for linear chains of H-bonding formamides containing from 2 to 15 monomeric units. The cooperative effect upon the strongest H-bonds (those nearest the center of the 15-formamide chain) approaches 200% that of the dimer. The cooperative interaction far exceeds that expected for electrostatic interactions. The large variation in the calculated H-bonding enthalpies cannot readily be modeled using pairwise nearest-neighbor potentials. The energetic data obtained from the DFT calculations have been empirically fit using parameters based upon the chain length (n) and the H-bond type (k) which corresponds to the position of the H-bond in the chain ($k = 1$ for terminal, 2 for penultimate H-bonds, etc.). Each type (k) of H-bond asymptotically approaches a limiting interaction energy specific for that type. As k becomes larger, the initial H-bond for that type becomes more stable, but the cooperative contribution to that type becomes less. The results are discussed with respect to their utility for improving the modeling of peptide structure and protein folding.

Introduction

Protein folding is still relatively poorly understood. Many reviews[1−9] including an entire issue of *Accounts of Chemical Research*[10] have appeared. The dynamics of the folding process is typically studied using Monte Carlo techniques on a potential surface defined using the individual nearest-neighbor contacts. While use of pairwise contact potentials has been criticized as inadequate,[11,12] success in using effective two-body potentials to simulate a many-body problem has also been reported.[13]

We have previously reported preliminary results on hydrogen bonding in chains of formamide molecules as depicted in Figure 1.[14] These chains were chosen as models for the hydrogen bonding motifs in protein secondary structures. The previous report presented data that suggest an extraordinarily high level of cooperativity within the hydrogen bonding chains. Since that report has appeared, Kemp has suggested that H-bond cooperativity contributes strongly to the formation of α-helices,[15] while Wu has reported evidence for such cooperativity in helices but not for β-sheets.[16,17] On the other hand, recent NMR evidence suggests that the hydrogen bonds toward the interior of an α-helix are shorter than those closer to the ends.[18]

We have also previously reported substantial cooperative effects for hydrogen bonding chains within molecular crystals. The crystals studied included those of acetic acid,[19] urea,[20] the nitroanilines,[21] and the enol of 1,3-cyclohexanedione.[22,23] There have been several reports of theoretical studies on cooperative H-bonding in amides involving small numbers of molecules[24−30] or using periodic calculations for infinite chains.[31]

In this paper, we present our complete results on density functional theory (DFT) calculations of hydrogen bonds in chains of up to 15 formamides. We originally sought to study planar chains as well as chains that retained the dihedral angles that the formamide entities would have in an α-helix. The data on the chains that were constrained to have dihedrals appropriate for an α-helix did not differ substantially from those for the planar chains when we constrained the individual formamides to have identical optimized structures as in our previous report.[14] For this reason, and because we could not perform vibrational corrections on these helical systems (they are not minima on the potential energy surface), we did not pursue the study of the helically arranged formamides any further. We shall also present an empirical formula that relates the energy of the individual hydrogen bond to the length of the H-bonding chain and the position of the hydrogen bond within the chain.

Methods

Molecular orbital calculations were performed using hybrid DFT methods at the B3LYP/D95(d,p) level. This method combines Becke's three-parameter functional[32] with the nonlocal correlation provided by the correlation functional of Lee, Yang, and Parr,[33] using the Gaussian 98 suite of programs.[34] The geometries were completely optimized with the sole constraints that all the structures were of C_s symmetry (all atoms are coplanar). We used our cluster of Intel and AMD powered computers that are parallelized using LINDA for these calculations. The number of nodes used for each calculation varied with the sizes of the systems studied. The vibrational frequencies were calculated for the planar structures, using the normal harmonic approximations employed in the Gaussian 98 program, to verify the stationary points and calculate the enthalpies of the various species.[34] All frequencies were real except for some very low frequency imaginary vibrations (less than 25 cm^{-1}) that involved out-of-plane twists between pairs of formamides in some of the longer formamide chains. We have previously reported that optimization (CP-OPT) on potential energy surfaces (PESs) that include the counterpoise corrections (CP) for basis set superposition error (BSSE) were performed for planar chains containing up to five formamides.[14]

10.1021/jp0365209 CCC: $25.00
Published on Web 11/07/2003

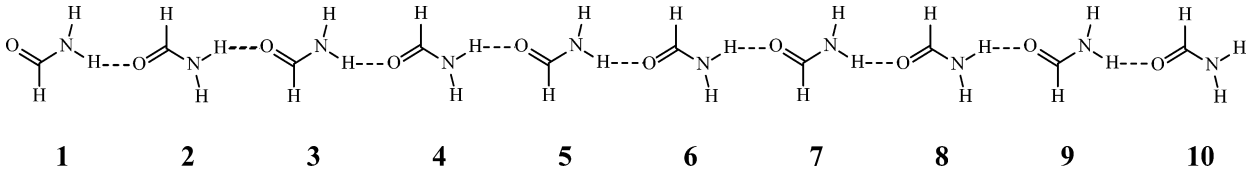

Figure 1. Chains of formamides indicating the numbering convention used in the text. Note that the translational repeating unit contains two formamides.

TABLE 1: B3LYP/D95 Enthalpies (kcal/mol) of Individual H-Bonds as a Function of Chain Length (*n*) and H-Bond Type (*k*)**

	k						
n	1	2	3	4	5	6	7
2	−4.49						
3	−6.23						
4	−6.75	−8.49					
5	−7.16	−9.42					
6	−7.23	−9.90	−10.42				
7	−7.31	−10.05	−10.98				
8	−7.39	−10.20	−11.20	−11.62			
9	−7.44	−10.33	−11.42	−11.90			
10	−7.38	−10.33	−11.49	−12.05	−12.12		
11	−7.50	−10.38	−11.60	−12.23	−12.38		
12	−7.42	−10.42	−11.57	−12.27	−12.49	−12.57	
13	−7.48	−10.41	−11.67	−12.31	−12.59	−12.74	
14	−7.42	−10.41	−11.60	−12.35	−12.57	−12.78	−12.85
15	−7.52	−10.45	−11.71	−12.37	−12.71	−12.86	−12.99

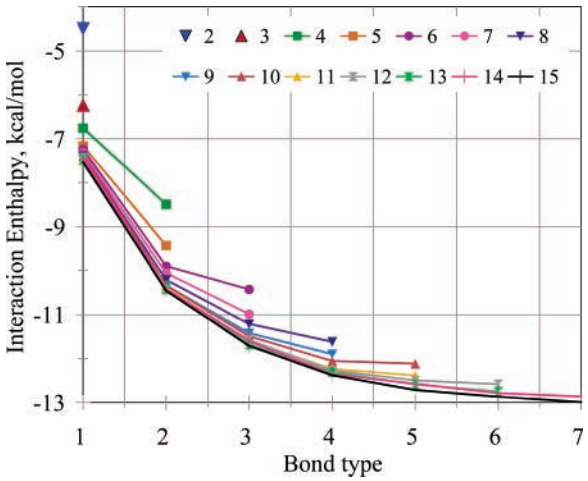

Figure 2. Interaction enthalpies for H-bonds organized by H-bond type (*k*) for chains of the lengths indicated by the symbols. Note that the enthalpies for the first and last H-bonds are the same as are the second and penultimate H-bonds, etc., so that H-bond type varies from 1 to 7.

Results and Discussion

All the formamide chains that we discuss below have been completely optimized with the sole constraint that they be planar. Our earlier communication[14] presented results for chains that were optimized with the constraint that each formamide unit have equivalent molecular geometries. The planar structures are minima on the PESs for the shorter chains. Chains containing more than five formamides had one or more very low frequency imaginary vibrations. The distortions from planarity associated with these imaginary frequencies represent out-of-plane bends. We do not believe that the small imaginary frequencies compromise the results that we present below.

The enthalpies of the H-bonds include correction for basis set superposition error (BSSE) and the appropriate vibrational corrections to 298 K. We determined that the BSSE corrections were found to be reasonably constant at 1.2 kcal/mol per H-bond by performing optimizations of formamide chains containing up to five monomers on a counterpoise corrected PES using a method that has been previously described.[35] The vibrational corrections needed to obtain the enthalpies were calculated using a *complete vibrational analysis* for each chain. In our earlier communication we used a constant additive vibrational correction per H-bond obtained from vibrational analysis of the dimer and tested on chains of up to five formamides.[14]

The data in Table 1 and Figure 2 present the results of DFT calculations on chains containing from 2 to 15 formamides. The energies of the individual hydrogen bonds are calculated by simply taking the difference between a chain containing *n* formamides and subtracting the energies of the two smaller chains (of *i* and *j* formamides where $i + j = n$) which remain after the individual H-bond is broken. Breaking hydrogen bonds that are the same distance from either end of the chain to form the same two smaller fragments will obviously require the same amount of energy (for example, a hexamer can be broken into a dimer and a tetramer by breaking the second or the fourth hydrogen bond). For this reason, Table 1 and Figure 2 contain only entries for breaking hydrogen bonds starting at one end, up to and including the most central hydrogen bond. One can immediately see that the hydrogen bonds become stronger as the chain becomes larger. Furthermore, hydrogen bonds close to the interior of the chain are stronger than those near the ends. The weakest hydrogen bond is that in the formamide dimer (−4.49 kcal/mol), while the strongest is the central bond in the chains containing 15 formamides (−12.99 kcal/mol). Thus, the strongest bond is about 2.9 times stronger than that of the dimer. The data in Table 1 indicate that the enthalpy of the central H-bond has not quite reached an asymptotic limit at 15 formamides. *These results indicate that a cooperative effect of almost 200% of the dimer H-bonding energy operates in long H-bonding formamide chains*. Clearly the strengths of the H-bonds depend on the length of the chain and the position of the H-bond in the chain. We shall refer to the H-bonds at different positions as different *types* of H-bonds (*k*) depending upon their position in a chain. First ($k = 1$) H-bonds exist in all chains starting with the dimer. Second ($k = 2$) H-bonds occur in chains containing four or more formamides. Third ($k = 3$) H-bonds occur in chains of six or more, etc. When the energetic data are presented in as in Figure 3, where each type of H-bond is plotted against chain length, one clearly sees that each type of H-bond asymptotically approaches a distinct limiting enthalpy as the chain length increases. We note that while the pairs of H-bonds that have the same type have the same energy, they do not necessarily have the same H-bond length, as the two termini of the chains are not equivalent (see discussion below). The linear H-bonding chains of formamides contain a repeating unit of two molecules (see Figure 1). The data in Table 1 and Figure 3 show a slight oscillation of H-bond enthalpies between chains containing even and odd numbers of formamides consistent with the two molecule repeating unit. This behavior is most apparent for H-bond types that have approached the asymptotic limit for that type. However, due to the small

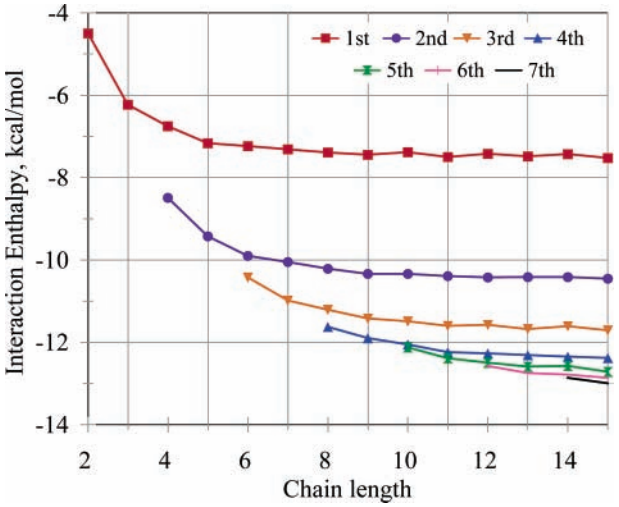

Figure 3. Interaction enthalpies for H-bonds organized by chain length for H-bond types (k) indicated by the symbols. Note that the enthalpies for the first and last H-bonds are the same as are the second and penultimate H-bonds, etc., so that H-bond type varies from 1 to 7.

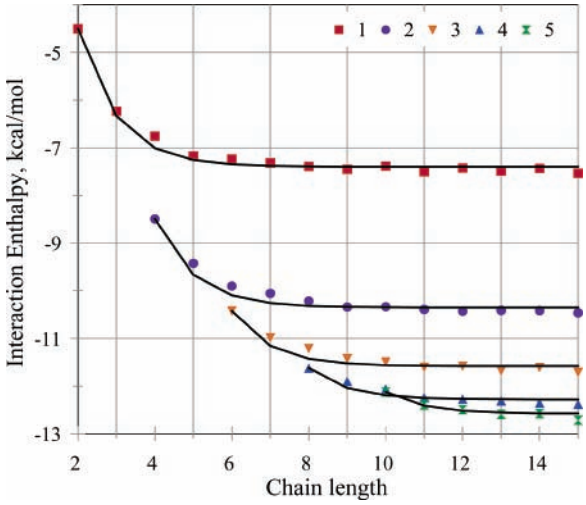

Figure 4. B3LYP/D95** H-bond enthalpies results (dashed line) compared to those fitted by equation (solid line) for the first five H-bond types (k).

magnitude of this oscillation, we have not taken this oscillation specifically in our further analyses.

We shall define the cooperative component of the terminal H-bonds as the additional stabilization over that of the dimer as the chain becomes longer. Based upon the asymptotic relationship for the terminal H-bonds (bond of the first type as displayed in Figure 3), we can imagine that an inverse exponential relationship between the cooperativity and the chain length (n) should be appropriate. We fitted the parameters, a_1 and b_1, of eq 1 to get the best fit for the terminal H-bonds. In this equation, a_1 is set equal to E_{dim}, which represents the H-bond energy of the dimer. When $n = 2$, the calculated energy should be that of the dimer leading to eq 2 for the terminal (type 1) H-bonds.

$$E(n,1) = a_1 + b_1(1 - \mathrm{e}^{(2-n)}) \tag{1}$$

$$E = E_{\mathrm{dim}} + b_1(1 - \mathrm{e}^{(2-n)}) \tag{2}$$

The parameter b_1 should be the difference in energy of the H-bonds of the dimer and the asymptotic limit. Thus b_1 represents the limiting cooperative effect of the terminal H-bonds (eq 1). From Figure 3, one clearly sees that the curvature of the exponential behavior diminishes as the H-bond becomes more removed from the end of the chain. We can consider an equation similar to eq 1 for each type of H-bond (eq 3).

$$E = a_k + b_k(1 - \mathrm{e}^{(m-n)}) \tag{3}$$

In each case, the first term, a_k, should be the energy of the first H-bond for type k. A hexamer would be the smallest chain to have a type 3 H-bond. The limiting cooperative effect for each type would be represented by b_k. The exponent for each equation would change to $(m - n)$, where m represents the number of monomers in the smallest aggregate that contains that type of bond (for $k = 3$, $m = 6$, since the smallest aggregate with a third H-bond is a hexamer). We fit a unique equation similar to eq 1 from the data in Table 1 for each type of H-bond. This leads to a set of similar equations (eq 3), one for each value of k. The first term was always taken to be the energy of the first H-bond of the appropriate type. Only b_k's were varied to best fit the data. As expected, b_k's become smaller in magnitude as the H-bond type becomes farther removed from the termini of the chains. The equations fit the data quite well as can be seen from Figure 4, which compares the B3LYP/D95** data with those obtained from the empirical equations.

Using this series of equations, we can calculate the energy of any H-bond in formamide chains of this type. However, we need to know two parameters, a_k and b_k, for each equation.

$$a_k = a_1 + c(1/\sqrt{k} - 1) \tag{4}$$

$$b_k = b_1 + d(1/\sqrt{k} - 1) \tag{5}$$

Thus, 14 parameters would be needed to calculate all the H-bonds in Table 1. This seems altogether too complex to be useful. However, from Figure 5, which presents the values of a_k and b_k for each type, we see that these values also approach asymptotic limits. Thus, we can fit the values of a_k and b_k to other functions in order to reduce the number of independent parameters needed to fit all the H-bond energies. We fitted those sets of parameters with similar equations (eqs 4 and 5). For b_k's, though, one can get a slightly better fit if eq 6 is used:

$$b_k = b_1 + f \ln k \tag{6}$$

Substituting for a_k and b_k in eq 3, from eqs 4 and 5, we obtain eq 7. If we use eq 6 rather than eq 5, we obtain eq 8 (where a_1 and b_1 are as in eq 1).

$$E(n,k) = a_1 + c(1/\sqrt{k} - 1) + (b_1 + d(1/\sqrt{k} - 1))(1 - \mathrm{e}^{(m-n)}) \tag{7}$$

$$E(n,k) = a_1 + c(1/\sqrt{k} - 1) + (b_1 + f \ln k)(1 - \mathrm{e}^{(m-n)}) \tag{8}$$

Using the data of Table 1 to obtain the best fit to eq 7, we obtain eq 9

$$E(n,k) = -4.493 + 13.789(1/\sqrt{k} - 1) + \{-2.900 - 4.050(1/\sqrt{k} - 1)\}(1 - \mathrm{e}^{(m-n)}) \tag{9}$$

while the best fit to eq 8 leads to eq 10.

$$E(n,k) = -4.493 + 13.789(1/\sqrt{k} - 1) + (-2.900 + 1.516 \ln k)(1 - \mathrm{e}^{(m-n)}) \tag{10}$$

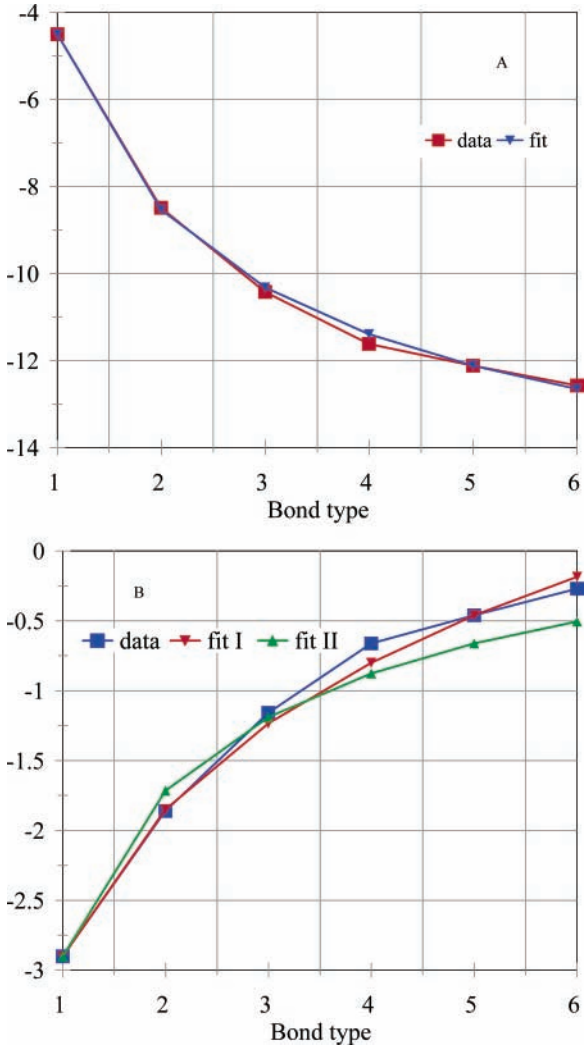

Figure 5. Comparison of calculated results with fitted parameters: (A) a_k and (B) b_k.

TABLE 2: Enthalpies (kcal/mol) of Individual H-Bonds as a Function of Chain Length (n) and H-Bond Type (k) Calculated Using Eq 8

	k						
n	1	2	3	4	5	6	7
2	−4.49						
3	−6.33						
4	−7.00	−8.53					
5	−7.25	−9.70					
6	−7.34	−10.13	−10.32				
7	−7.37	−10.29	−11.10				
8	−7.39	−10.35	−11.39	−11.39			
9	−7.39	−10.37	−11.49	−11.89			
10	−7.39	−10.38	−11.53	−12.08	−12.12		
11	−7.39	−10.38	−11.55	−12.15	−12.41		
12	−7.39	−10.38	−11.55	−12.17	−12.51	−12.65	
13	−7.39	−10.38	−11.55	−12.18	−12.55	−12.77	
14	−7.39	−10.38	−11.56	−12.18	−12.57	−12.81	−13.07
15	−7.39	−10.38	−11.56	−12.19	−12.57	−12.83	−13.04

Tables 2 and 3 present the fitted data from eqs 9 and 10, respectively, while Figure 6 compares the fitted data to the calculated data of Table 1. In these equations, we found the best fit for a_k to be the hyperbola, inverse square root of k. The relationship $f \ln k$ used in eqs 6, 8, and 10 was obtained purely from empirical curve fitting. The parameters c, d, and f come from the empirical fittings discussed above. While eqs 6, 8, and 10 provide a slightly better fit to the data that we have,

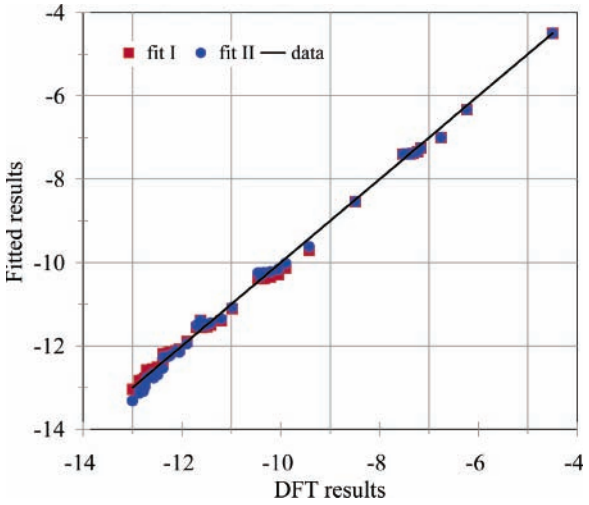

Figure 6. Comparison of calculated DFT with fitted H-bonding enthalpies.

TABLE 3: Enthalpies (kcal/mol) of Individual H-Bonds as a Function of Chain Length (n) and H-Bond Type (k) Calculated Using Eq 10

	k						
n	1	2	3	4	5	6	7
2	−4.49						
3	−6.33						
4	−7.00	−8.53					
5	−7.25	−9.61					
6	−7.34	−10.01	−10.32				
7	−7.37	−10.16	−11.07				
8	−7.39	−10.21	−11.35	−11.39			
9	−7.39	−10.23	−11.45	−11.94			
10	−7.39	−10.24	−11.49	−12.14	−12.12		
11	−7.39	−10.24	−11.50	−12.22	−12.53		
12	−7.39	−10.24	−11.51	−12.25	−12.69	−12.65	
13	−7.39	−10.25	−11.51	−12.26	−12.74	−12.97	
14	−7.39	−10.25	−11.51	−12.26	−12.76	−13.09	−13.07
15	−7.39	−10.25	−11.51	−12.26	−12.77	−13.13	−13.31

they do not extrapolate properly as k goes to infinity. Using these equations would predict that b_k would become infinitely negative (as would the limiting H-bonding enthalpy). Clearly, b_k (and the H-bonding energy) should asymptotically approach some limiting value. Using eqs 5, 7, and 9 preserves this physical principle. Ideally b_k should go to zero (which it clearly does not using eqs 6, 8, and 10), so the physical model implied by eqs 5, 7, and 9 is partially sacrificed to preserve a better fit. We provide both sets of equations as one provides a better fit, the other a somewhat better behaved physical model.

The H-bond lengths in the optimized structures are presented in Table 4 and Figure 7. As expected, the stronger hydrogen bonds tend to be shorter than the weaker ones. Since the formamide chains are not symmetrical (one end terminates with an amino group and the other with the carboxyl group), the H-bond lengths for the pairs of hydrogen bonds that have the same strength within a chain are not exactly the same. Figure 8 shows the relationship between the H-bond strengths and the O···H distances. H-bonding distance is often taken as a linear function of the energies. While this might be approximately true within a certain range of normal H-bond lengths, it cannot hold in general. Clearly, at very long H-bonding distances the interaction energies must approach zero with increase in H-bond length. Equation 11

$$D = 0.0147E + 1.9619 \qquad (11)$$

where D is the O···H distance in angstroms and E is the H-bond

TABLE 4: H-Bond Lengths from B3LYP/D95 Optimized Geometries as a Function of Chain Length (*n*) and Position of H-Bond in Chain (Numbering as in Figure 1)**

	n													
bond	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.920	1.873	1.859	1.853	1.850	1.850	1.849	1.850	1.847	1.847	1.847	1.847	1.847	1.847
2		1.878	1.835	1.813	1.810	1.808	1.809	1.805	1.805	1.807	1.805	1.806	1.805	1.806
3			1.869	1.819	1.806	1.797	1.793	1.792	1.792	1.791	1.790	1.790	1.790	1.790
4				1.862	1.814	1.797	1.791	1.786	1.789	1.785	1.785	1.785	1.785	1.785
5					1.861	1.810	1.794	1.786	1.786	1.783	1.780	1.781	1.781	1.781
6						1.860	1.808	1.793	1.787	1.780	1.781	1.782	1.781	1.782
7							1.859	1.807	1.794	1.784	1.780	1.781	1.781	1.782
8								1.859	1.807	1.791	1.784	1.780	1.780	1.781
9									1.858	1.805	1.790	1.784	1.780	1.781
10										1.858	1.807	1.791	1.784	1.780
11											1.857	1.806	1.791	1.784
12												1.858	1.806	1.791
13													1.857	1.806
14														1.858

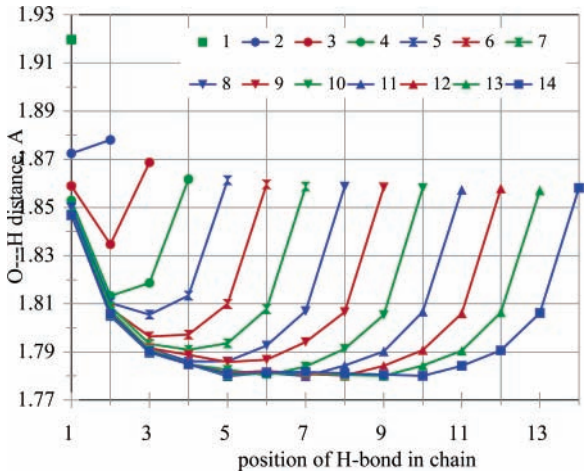

Figure 7. Calculated B3LYP/D95** H-bond lengths organized by bond position in chains of lengths indicated.

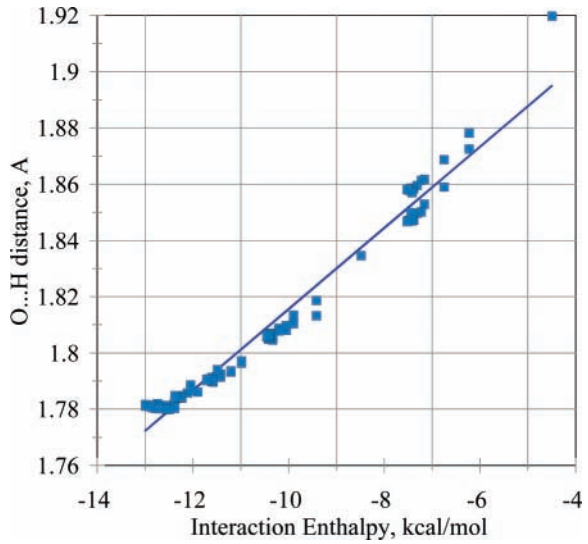

Figure 8. Relation between H-bond length interaction enthalpies for all H-bonds calculated. The best linear fit is plotted: H-bond length (Å) = 0.0147 × interaction enthalpy (kcal/mol) + 1.9619 (R^2 = 0.9765).

interaction enthalpy, gives the best linear fit (R^2 = 0.9765) within the range of the data that we have. Of course, the relationship cannot be linear for complete variation of *D*, as the interaction enthalpy must go to zero for large values of *D* and to infinity as *D* approaches zero. The present data are insufficient for defining the expected curvature in a function that expresses the relationship of *D* with interaction enthalpy for a larger range of *D*. Despite the fact that two energetically equivalent H-bonds within the same chain have different lengths, the correlation between bond strengths and lengths remains reasonably good. Thus, if one can calculate the H-bond energy using the relationship described above, one can obtain a reasonable estimate of the H-bond distance from eq 11.

Origin of the Cooperativity

The unusually large cooperative interactions between H-bonds in formamide chains result form a combination of different kinds of physical interactions. Pairwise interactions suffice to physically model electrostatic interactions. Three-body interactions are necessary to account for polarization, while many-body interactions are required to properly describe mutual polarization and covalent interactions.[36] The electrostatic component can be considered as composed of dipole−dipole interactions. These are proportional to $1/R^{-3}$, where *R* is the distance between the dipoles. If *N* dipoles are linearly aligned and evenly spaced, the 1−3 interactions would be 1/8 of the 1−2 interactions as the 1−3 distances are twice the 1−2 distances. Similarly, the 1−4 interactions would be 1/27, the 1−5 interactions 1/64 of the 1−2 interactions, etc. The maximum cooperative electrostatic interaction would be a sum of all such pairwise interactions. For the strongest bond in the chain of 15 formamides, this interaction would be 1.50 times that of a dimer, while the central H-bond in an infinite chain would be about 1.61 times that of a dimer. The data that we have presented above clearly show this electrostatic model to be inadequate to describe the major component of cooperativity. As the H-bonds studied readily become close to 3 times (200%) larger in magnitude than the prototypical N−H···O interaction of the formamide dimer, the calculated electrostatic cooperative effect of 50% represents only 1/4 of the total cooperativity. Our results agree with the interpretation of Weinhold[11] (for HCN H-bonding chains) rather than with that of Wu, who prefers an electrostatic model.[17]

One must look elsewhere for physical contributions to the cooperativity. Examination of the relationship between the C═O and C−N bond distances (Figures 9 and 10) as a function of the H-bonding chain length and the position of the formamide in the chain clearly illustrates a structural response by each formamide unit to the H-bonding within its chain. As the chains become longer and the formamide monomer becomes more central, the C═O distance increases, while the C−N distance decreases. These relationships further indicate that the cooperativity cannot be purely electrostatic in origin. However, they could be the result of a combination of polarization, mutual

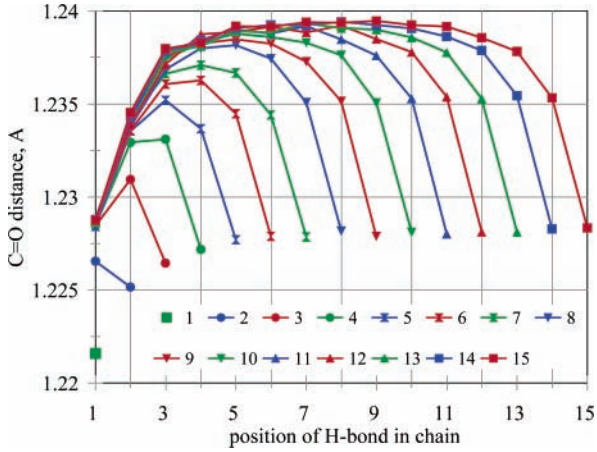

Figure 9. Calculated B3LYP/D95** C═O lengths organized by bond position in chains of lengths indicated.

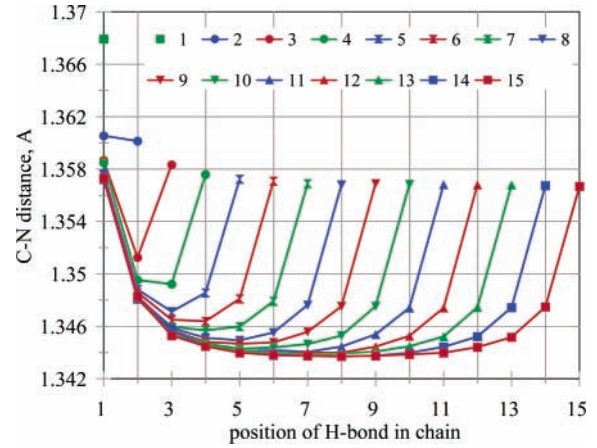

Figure 10. Calculated B3LYP/D95** C−N lengths organized by bond position in chains of lengths indicated.

polarization, and covalent interactions. Gilli has emphasized the importance of resonance assisted hydrogen bonds (RAHBs) in several crystal structures.[37−39] We have addressed this question in a model study where we compared the response of two molecules that form extensive H-bonding chains in their crystal structures to an electric field that reproduces the interaction energy of the molecule in the H-bonding environment.[40] Since only electrostatic interactions and polarization (not mutual polarization or covalent interactions such as resonance assistance) can occur in the model of a molecule in the electric field, we could address the relative importance of the two pairs of effects. The H-bonding environment of one molecule (urea) was somewhat reasonably described by the applied electric field. That of the other molecule (the enol of 1,3-cyclohxanedione), which Gilli considers a model of RAHB, was extremely poorly described by the electric field. H-bonding chains of 1,3-cyclohexane clearly afford a more favorable situation for RAHBs, because forming a covalent O−H in place of the H-bond O···H and vice versa would create an identical H-bonding chain if done for each H-bonding interaction. Thus, the second resonance structure would be very close to the first in energy. They would have the same energy if the O···H and O−H distances were equal. For formamide chains, one needs to form an O−H in place of an O···H and an H···N in place of an H−N. This would provide a second resonance structure of clearly higher energy than the first. We expect RAHB to be an important (but not the only) contributor to the cooperativity that we observe. We address the effects of the H-bonding cooperativity upon the vibrational modes of these formamide chains in detail elsewhere.[41]

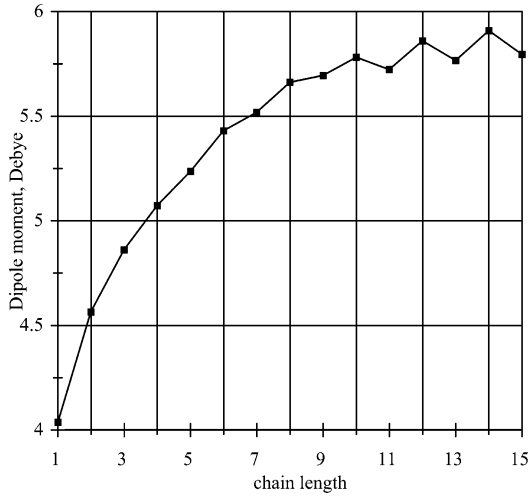

Figure 11. Dipole moments per formamide monomer as a function of chain length. The repeating unit of two formamides in the linear periodic structure gives rise to an alternation of dipoles for chains of even and odd numbers of formamides.

The nonlinear increase in the dipole moment of the formamide chains as they increase in length provides yet another indication of the inadequacy of an electrostatic explanation to the observed cooperativity. Figure 11 illustrates the average dipole moment of a formamide molecule in chains of increasing length. This value increases with chain length, approaching asymptotic limits that are slightly different for chains containing even and odd numbers of formamides. In an electrostatic model, the dipole moment of *N* formamides would simply be the vector sum of the individual *N* dipoles. One cannot determine from the data of Figure 11 how much of the increase in average dipole moment is due to polarization, mutual polarization, and covalent (or RAHB) interactions. All of these effects would tend to increase the dipole moments as observed. This behavior is qualitatively similar to the increasing dipole moments per amino acid residue of α-helical peptides as the helix increases in length.[42] One should note that calculation of the dipole−dipole electrostatic interactions (see above) using the average dipole moment of a formamide in a long chain will give approximately the correct H-bonding interaction for the strongest H-bond in a 15-mer. However, we have noted that an electrostatic model cannot account for the increase in average dipole moment as the chain increases.

Application to Protein Models

Recent reports of *trans*-H-bond $^{13}C-^{15}N$ couplings in the NMR spectra of α-helical peptides confirm that the H-bonds closer to the middle of the helices are shorter,[18] and thus stronger, than those near the ends. Each α-helix contains three H-bonding chains that are similar to the formamide chains that we have studied. Kemp recently explained his experimental demonstration of an enthalpic component to the cooperativity of α-helical peptides as evidence for H-bond cooperativity in these structures.[15] These results are consistent with calculations reported by Wu on helical polyglycines,[17] as well as our recent calculations on five completely optimized α-helical peptides contain 17 amino acid residues.[43] Curiously, Wu has not observed similar cooperativity in *β*-sheets.

Tailoring the relationships between chain length and H-bond type and position in the chain that we have presented above to modeling protein structure does not pose major conceptual or

calculational problems. The determination of the energy of each H-bond using either eq 9 or 10 is straightforward and economical with respect to computer time. Nevertheless, some modifications will be necessary for this (or similar) equations to be directly useful for peptide modeling. Crystallographic H-bond distances in α-helices tend to be longer than those in β-sheets, which are longer than those calculated here for formamide chains.[44,45] These differences are likely to be due to the strain required to form the secondary peptide structures and the fact that most of the C═O's will form another H-bond to water or another H-bonding donor,[44,45] thereby weakening and lengthening the peptide H-bond. The formamide chains are strain free. In principle, the effect of the strain on the H-bond strength can be parametrized in improved peptide models that include the relationships derived here.

H-bonds are the most important interactions between amino acid residues in peptides, as these interactions provide the stability for secondary structures, such as helices or sheets. Most methods currently used to model protein structure treat the H-bonds as pairwise interactions between nearest neighbors. Empirically fitted pairwise potentials do not necessarily represent purely electrostatic interactions. Rather they use whatever functions chosen to reproduce a set of data by appropriately fitting the parameters in the defined functions. Since the energy of an individual H-bond in a formamide chain is seen to vary with chain length and position within the chain, the same pairwise potential cannot be appropriate to all such interactions. However, fitted pairwise interactions could be scaled to account for the position in and size of the H-bonding chain using functions similar to those described in this paper.

Acknowledgment. This work was supported (in part) by grants from the National Institutes of Health (S06GM60654) and PSC-CUNY.

Supporting Information Available: Cartesian coordinates for optimized formamide chains. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.
(2) Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem., Int. Ed.* **1998**, *37*, 868.
(3) Arteca, G. A.; Reimann, C. T.; Tapia, O. *Mass Spectrom. Rev.* **2002**, *20*, 402.
(4) Skolnick, J.; Kolinski, A. *Comput. Sci. Eng.* **2001**, *3*, 40.
(5) Shea, J.-E.; Brooks, C. L., III. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499.
(6) Mirny, L.; Shakhnovich, E. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 361.
(7) Chasse, G. A.; Rodriguez, A. M.; Mak, M. L.; Deretey, E.; Perczel, A.; Sosa, C. P.; Enriz, R. D.; Csizmadia, I. G. *Theochem* **2001**, *537*, 319.
(8) Kaya, H.; Chan, H. S. *Phys. Rev. Lett.* **2000**, *85*, 4823.
(9) Schuster, P.; Wolschann, P. *Monatsh. Chem.* **1999**, *130*, 947.
(10) *Acc. Chem. Res.* **1998**, *31*, entire issue.
(11) King, B. F.; Weinhold, F. *J. Chem. Phys.* **1995**, *103*, 333.
(12) Vendruscolo, M.; Domany, E. *J. Chem. Phys.* **1998**, *109*, 11101.
(13) van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L. I. I. I.; Merz, K. M. J. *J. Phys. Chem. B* **2000**, *104*, 9554.
(14) Kobko, N.; Paraskevas, L.; del Rio, E.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2001**, *123*, 4348.
(15) Kennedy, R. J.; Tsang, K.-Y.; Kemp, D. S. *J. Am. Chem. Soc.* **2002**, *124*, 934.
(16) Zhao, Y.-L.; Wu, Y.-D. *J. Am. Chem. Soc.* **2002**, *124*, 1570.
(17) Wu, Y.-D.; Zhao, Y.-L. *J. Am. Chem. Soc.* **2001**, *123*, 5313.
(18) Jaravine, V. A.; Alexandrescu, A. T.; Grzesiek, S. *Protein Sci.* **2001**, *10*, 943.
(19) Turi, L.; Dannenberg, J. J. *J. Am. Chem. Soc.* **1994**, *116*, 8714.
(20) Masunov, A.; Dannenberg, J. J. *J. Phys. Chem. B* **2000**, *104*, 806.
(21) Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1996**, *100*, 9638.
(22) Turi, L.; Dannenberg, J. J. *Chem. Mater.* **1994**, *6*, 1313.
(23) Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1992**, *96*, 5819.
(24) Van Duijnen, P. T.; Thole, B. T. *Biopolymers* **1982**, *21*, 1748.
(25) Sheridan, R. P.; Lee, R. H.; Peters, N.; Allen, L. C. *Biopolymers* **1979**, *18*, 2451.
(26) Guo, H.; Karplus, M. *J. Phys. Chem.* **1992**, *96*, 7273.
(27) Guo, H.; Karplus, M. *J. Phys. Chem.* **1994**, *98*, 7104.
(28) Guo, H.; Gresh, N.; Roques, B. P.; Salahub, D. R. *J. Phys. Chem. B* **2000**, *104*, 9746.
(29) Ludwig, R. *J. Mol. Liq.* **2000**, *84*, 65.
(30) Cabaleiro-Lago, E. M.; Otero, J. R. *J. Chem. Phys.* **2002**, *117*, 1621.
(31) Suhai, S. *J. Phys. Chem.* **1996**, *100*, 3950.
(32) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(33) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, W. M.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.11, Revision A.10; Gaussian, Inc.: Pittsburgh, PA, 1998−2001.
(35) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024.
(36) Dykstra, C. E. *Chem. Rev.* **1993**, *93*, 2339.
(37) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *Acta Crystallogr., Sect. B: Struct. Sci.* **1998**, *B54*, 50.
(38) Gilli, G.; Bertolasi, V.; Ferretti, V. *Acta Crystallogr., Sect. B: Struct. Sci.* **1993**, *B49*, 564.
(39) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *Chem.—Eur. J.* **1996**, *2*, 925.
(40) Dannenberg, J. J.; Haskamp, L.; Masunov, A. *J. Phys. Chem. A* **1999**, *103*, 7083.
(41) Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6688.
(42) Applequist, J.; Mahr, T. G. *J. Am. Chem. Soc.* **1966**, *88*, 5419.
(43) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124.
(44) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1987**, *44*, 97.
(45) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin, 1991.