A statistical potential has been developed to quantitatively describe the CH–O hydrogen bonding interaction at the protein-protein interface. The calculated energies of the CH–O pair interaction show a favorable valley at ~3.3 Å, exhibiting a feature typical of an H-bond and similar to the ab initio quantum calculation result (Scheiner, S., Kar, T., and Gu, Y. (2001) J. Biol. Chem. 276, 9832–9837). The potentials have been applied to a set of 469 protein-protein complexes to calculate the contribution of different types of interactions to each protein complex: the average energy contribution of a conventional H-bond is ~30%; that of a CH–O H-bond is 17%; and that of a hydrophobic interaction is 50%. In some protein-protein complexes, the contribution of the CH–O H-bond can reach as high as ~40–50%, indicating the importance of the CH–O H-bond at the protein interface. At the interfaces of these complexes, C–H–O H-bonds frequently occur between adjacent strands in both parallel and antiparallel orientations, having the obvious structural motif of bifurcated H-bonds. Our study suggests that the weak CH–O H-bond makes an important contribution to the association and stability of protein complexes and needs more attention in protein-protein interaction studies.

The conventional hydrogen bonds of the type X–H–Y (where X and Y = N or O) have been widely found and thoroughly studied in macromolecular structures from both experimental and theoretical perspectives (1, 2, 7–9). On the other hand, close CH–O contacts occur often in protein structures and are considered as hydrogen bonds. It is increasingly recognized that weak CH–O hydrogen bonds play an important role in the stabilization and function of biological macromolecules (3–6).

CH–O contacts are now being increasingly widely accepted as genuine hydrogen bonds (10, 11). Much of the evidence for the CH–O hydrogen bond comes from the observation that short intermolecular CH–O contacts are well established in many small molecule crystals (12, 13). In more recent years, neutron diffraction studies of amino acid crystals (which yield highly accurate positions of hydrogen atoms experimentally) have provided convincing evidence in favor of the ability of the carbon atom to function directly as hydrogen bond donors in CH–O contacts (14). Recently, there have been surveys of high resolution protein structures that reveal the widespread occurrence of weak CH–O hydrogen bonds (15–21, 59, 60). Various studies have reported the existence of a weak C–H–O hydrogen bond between the parallel β-sheets in proteins (17, 59, 60). At the same time, some mutation studies on protein-ligand interactions have reported that weak CH–O bonds stabilize protein-ligand complexes (10, 22, 23). Similar to protein-ligand interfaces, close CH–O contacts abound at protein-protein interfaces. Although CH–O H-bonds are normally weaker than conventional hydrogen bonds, their number cannot be neglected. The CH–O hydrogen bonding interaction is also thought to be important for the interactions of nucleic acids with proteins and drug binding (6, 24, 25); and thus, it would be surprising if CH–O interactions were less important in protein interfaces. In fact, the CH–O hydrogen bonding interaction has been considered to be so important that many crystal refinement programs that treat CH–O contacts as repulsive have been called into question (10, 27, 28).

However, the energy contribution of this kind of interaction to the stability of protein-protein complexes as compared with other forces remains to be explored. Despite the finding of numerous CH–O contacts at protein-protein interfaces, their relative importance in the protein recognition or drug binding process remains unclear. Although some structural analyses have provided a wealth of information about average hydrophobicities and residue compositions of protein-protein complexes (29–33), they did not provide any quantitative information on the strengths of these different sorts of interactions (such as hydrophobic interaction, hydrogen bonding interaction, and CH–O hydrogen bonding interaction); but it is the quantitative magnitude of the interaction that is the most significant in understanding the possible role each of them plays in the association of the protein complex.

In attempts to describe the energy aspects of CH–O interactions, theoretical calculations have been developed to evaluate the strength of CH–O interactions (34–37). Ab initio quantum calculation has shown that a small molecule such as CH2F2 can establish an H-bond with the strength similar to a conventional OH–O interaction (38). Recently, the ab initio quantum calculation for the C–H–O hydrogen bond of some representative amino acid residues (39) demonstrated that the peptide CH group is a potent proton donor and that the CH–O interaction appears to be a true hydrogen bond. Furthermore, the study also showed that some CH–O bonds are even stronger than a conventional OH–O interaction, suggesting that the CH–O hydrogen bonding interactions in proteins need to be paid more attention.

Ab initio quantum calculation (39) gave a series of calculated
binding strengths for the C--H--O contacts of different amino acid residues in an ideal model. However, the energy percentage of the CH--O interaction is still not known (i.e. the ratio of the CH--O interaction in the total binding energy of protein complexes). Thus, the relative importance of the CH--O interaction in protein association remains somewhat an open question.

Here, for the first time, we calculate the average energy percentage of the CH--O interaction in total free binding energy using our established mean-field potential for describing protein-protein complexes (40). The potentials of mean force (PMFs),¹ a beneficial tool in protein fold recognition, generally use the training data base of known protein structures to extract “pseudo-potentials” for predicting unknown structures (41, 42, 45–55). Many applications have demonstrated their usefulness in studies of protein-ligand binding (50–55) and protein-protein associations (56, 57).

Based on the new definition of atom types and our developed method, the distance-dependent potentials have been derived to describe different types of interactions at the protein-protein interface, and the energy aspects of CH--O interactions have been discussed quantitatively. The calculated energies of CH--O pair interactions exhibit a feature typical of a hydrogen bond. A quantitative study of the energy percentage has shown the importance of the CH--O hydrogen bond at the protein interface. The obvious structural motif of a bifurcated hydrogen bond is highlighted in the stereochemical analysis of CH--O contacts in the representative examples with a high CH--O percentage. We expect that this method would be helpful in understanding the interactions at protein-protein interfaces and how they drive protein-protein associations.

MATERIALS AND METHODS

We use the established empirical approach for the description of protein-protein association from an energy aspect. The mean-field potentials were derived from the same training set as that used before (40). Using the methodology as described in our previous work, we implicitly treated solvation and entropic effects and directly estimated total free binding energies of protein-protein complexes without any knowledge of experimental binding affinities and fitting procedures.

New Definition of Five Atom Types

To extract and characterize the CH--O hydrogen bonding interactions, five atom types were defined: hydrogen bond donor (D), hydrogen bond acceptor (A), both donor and acceptor (B), CH type (neutral atom bonded to hydrogen atoms) (CH), and neutral type (neither donor nor acceptor) (N). In the atom type definition, primary and secondary amines are defined as donors; oxygen and nitrogen atoms with no bound hydrogen are defined as acceptors; hydroxyl oxygen, ND, NE of His, and carboxyl oxygen in the C terminus are defined as both donors and acceptors; carbon atoms with hydrogen atoms are defined as the CH type; and carbon atoms with no bound hydrogen and sulfur atoms are defined as neither donors nor acceptors. The atom occupancy in the crystal structure file was used as a weighting factor.

The distance-dependent Helmholtz free energies of protein-protein complexes were extracted from the non-redundant training set (40) in the Brookhaven Protein Data Bank (43, 44). Atoms of the “receptor” part and the “ligand” part are treated differently, so we calculated 25 atom pair interaction terms here (see Table I). In our training set, metal and hetero-atoms were excluded. It is assumed that all crystallized complexes use water as the medium. Water molecules are neglected, as the solvation effects are implicitly treated. Hydrogen atoms were omitted in all analyses.

Statistical Potentials

Pair potentials were derived from the training set using the same methodology described in our previous work (40). Here, we give only a brief description of the method.

According to the reverse Boltzmann relationship, the free energy between the receptor atom of type i and the ligand atom of type j at a distance r can be written as shown in Equation 1,

\[ A_q = -kT \ln(f_q(r)/Z_q) \]  

(Eq. 1)

where \( k \) is the Boltzmann constant and \( T \) is the absolute temperature. \( f_q(r) \) is the frequency of these \( ij \) contacts occurring at distance \( r \). In fact, our statistical potential \( \Delta A_q(r) \) is the difference relative to the reference potential shown in Equation 2,

\[ A_q - \text{reference energy} = \Delta A_q(r) = kT \ln(1 + m_o) - kT \ln(1 + m_o \rho_i \rho_j g_{ij}^p(r)/f(r)) \]  

(Eq. 2)

where \( m_o \) is the total number of contacts between types \( i \) and \( j \), \( g_{ij}^p(r) \) is the distribution of these contacts occurring at distance \( r \), and \( f(r) \) is the distribution of all contacts for all types at distance \( r \). The atom pair distance \( r \) uses a histogram-based representation, and \( r \) refers here to a given bin of width 0.2 Å. We simplified our reference energy as follows: we just imported a large value in a very short distance (i.e. where our statistics are not included) to capture strong van der Waals repulsive potentials in this distance range.

The derived potentials for the interaction of receptor atom type \( i \) and ligand atom type \( j \) are summed up to evaluate the total PMF value in Equation 3,

\[ A = \sum \Delta A_q(r) \times \Delta A_q \]  

(Eq. 3)

where \( \Delta A_q = 0 \) for \( r > r_{\text{cut-off}} \) and \( \Delta A_q = \rho_i \times \rho_j \) for \( r \leq r_{\text{cut-off}} \). For atom presentations, the occupancy ratio \( \rho_i \) was used as a weighting coefficient. The occurrence of the atom pair in a distance \( <8.0 \) Å was recorded. If the total number of atom pairs in the shell of a distance \( r \leq \Delta r (\Delta r = 0.1 \) Å) was <30, the contributions of all atom pairs at the distance interval were ignored because of their statistically insufficient data.

 Contribution of Different Interactions to the Total Free Binding Energy in Each Protein-Protein Complex

We applied the statistical potentials to a larger set of protein-protein complexes. The x-ray and NMR structures of protein-protein complexes with a more flexible threshold of resolution were selected. For the data set, if the calculated binding energy of analyzed protein interfaces was not appropriate (only the calculated energy within the range of −200 to −10 kJ/mol was selected), the data were abandoned in analyses. Metal ions and water molecules were excluded in all analysis. The filtered set included 654 “receptor-protein-ligand protein” pairs from 469 x-ray structures of protein-protein complexes, including antigen-antibody complexes, enzyme-protein inhibitor complexes, heterodimers, and protease-peptide complexes in a wide range of data sets (see the Supplemental Material).

For each protein complex, the contributions of different interaction types to the total binding energy were calculated as shown in Equation 4,

\[ p_z = \sum \Delta A_{i,j} \times \Delta A_{i,j} \]  

(Eq. 4)

where \( p_z \) represents a certain type of pair interaction (see Table I), \( p_z \) is the energy percentage of the interaction of type \( z \) to the total free binding energy of the calculated protein-protein complex. The pair interaction between atoms \( i \) and \( j \) is type \( z \), according to the definition in Table I. The denominator \( A \) represents the total PMF energy calculated by Equation 3.

Analysis of the Stereochemistry of CH--O Hydrogen Bonds

The stereochemical details of CH--O hydrogen bonds in some representative examples with a high energy percentage of CH--O interactions were surveyed. Hydrogen atoms were added to these Protein Data Bank coordinate files using CHARMM (61). Each of the CH--O hydrogen bonds was analyzed using three different geometrical parameters: the C--O distance (\( d \)), the H--O distance (\( d_{\text{H,O}} \)), and the C--H--O angle (\( \theta \)). The definitions of these parameters are given Tables II and III. In the analysis, only those contacts with a \( \theta \) angle >100° were accepted.
Comparison of Energy Minimization Calculations with and without Inclusion of the CH–O Hydrogen Bonding Interaction Energies

Model of the Weak CH–O=C Hydrogen Bond—We have modeled the weak CH–O=C hydrogen bond by adding a distance constraint, as shown in Equation 5,

\[
E_{\text{constraint}} = \begin{cases} 
0.5 k_{\text{min}} (r_{\text{C\#O} = r_{\text{max}}})^2 & \text{if } r_{\text{C\#O} < r_{\text{min}}} \\
0 & \text{if } r_{\text{min}} < r_{\text{C\#O}} < r_{\text{max}} \\
0.5 k_{\text{max}} (r_{\text{C\#O} = r_{\text{max}}})^2 & \text{if } r_{\text{C\#O}} > r_{\text{max}}
\end{cases}
\]  

(Eq. 5)

where \( k_{\text{min}} = 200 \text{ kJ/molÅ}^2 \), \( k_{\text{max}} = 100 \text{ kJ/molÅ}^2 \), \( r_{\text{min}} = 3.2 \text{ Å} \), and \( r_{\text{max}} = 3.4 \text{ Å} \). Because we could not generate realistic binding energies using van der Waals interactions and partial charge electrostatics, a distance constraint was used to force \( d(\text{C\#O}) \) to take on the observed value of \( 3.2-3.4 \text{ Å} \) to model the weak CH–O=C hydrogen bond. According to the potential shape of weak CH–O=C H-bonds, two different force constants of the constraint were used in the close or long distance range, respectively.

The resultant weak CH–O=C interaction potential is \( E_{\text{C\#O}} = E_{\text{elec}} + E_{\text{VdW}} + E_{\text{constraint}} \) and the energy terms of van der Waals \( (E_{\text{VdW}}) \) and electrostatic \( (E_{\text{elec}}) \) interactions were calculated by the values taken from CHARMM (60). The distance constraint was artificially stable and adequately simulated the van der Waals and partial interactions between the carbon and oxygen atoms in the weak CH–O=C hydrogen bond.

Results

Helmholtz Free Energy for General CH–O Hydrogen Bonding Interactions

Based on the definition of the five atom types, the distance-dependent potentials were extracted from the training set. In Table I, there are four atom pairs reflecting the type of CH–O pair interaction. We gave emphasis to four atom pairs: A–CH, B–CH, CH–A, and CH–B. In all cases, the occurrence of CH–O atom pairs is sufficiently high: the number of CH–A (hydrogen bond acceptor) occurrence is 92,026, and that of CH–B (both acceptor and donor) occurrence is 18,101. Fig. 1 (a and b) shows examples of the calculated Helmholtz free energy for these pair interactions, reflecting the characteristics of CH–O hydrogen bonding interactions. Note that the derived pair potentials for the CH–O interaction have a favorable valley in the same distance range (~3.3 Å) in all the figures, which matches well with those from quantum calculations (39).

Fig. 1c shows a comparison of the calculated PMF energies of three representative types of different interactions, including the D–A hydrogen bond as an example of conventional hydrogen bonds, the CH–A hydrogen bond, and the N–N nonpolar pair interaction as an example of hydrophobic interactions. The calculated energies of the three interactions have very different characterizations: the example of conventional hydrogen bonding interactions has a favorable valley in the distance range from 2.4 to 3.0 Å and becomes weakly repulsive in the 3.3–4.2 Å range; the example of CH–O hydrogen bonding interactions has a favorable valley in the distance range from 3.0 to 3.5 Å; and the example of hydrophobic interactions has a repulsive potential at all distances up to 4.5 Å. The calculated energies of conventional hydrogen bonds and CH–O hydrogen bonds have different strengths and optimum distances of maximum bonding strength: the conventional hydrogen bond has a bond strength of ~5.5 kJ/mol and an optimum distance of 2.8 Å, whereas the CH–O hydrogen bond has a binding strength of ~1.9 kJ/mol and an optimum distance of 3.3 Å. The CH–O hydrogen bond has a longer optimum distance, and its strength is approximately one-third of that of the conventional hydrogen bond.

Contribution of Different Interactions at Protein-Protein Interfaces

Based on the new definition of five atom types, different interactions at protein interfaces can be classified easily. We have applied the potentials to the set of 469 protein-protein complexes. The Helmholtz free energies of different types of pair interactions at each protein interface were calculated and compared, including hydrophobic interactions, conventional hydrogen bonds and weak CH–O hydrogen bonds; and their energy contributions were analyzed, focusing on their percentages of total energy in each calculated protein complex (using Equation 4). The mean value of every interaction percentage was then calculated. The mean percentage of conventional hydrogen bonding interactions is ~30%; the mean percentage of hydrophobic interactions is 50%; and the mean contribution of CH–O interactions is 17%. In Fig. 2a, the ratio of the three interactions from the aspect of energy contribution is shown.
Furthermore, we have analyzed the energy percentage of CH–O hydrogen bonding interactions in each protein-protein complex. The distribution is shown in Fig. 2b. It should be noted that the energy percentages of CH–O hydrogen bonding interactions in some examples are up to 50% for protein-protein associations. The steric interaction of atom types such as hy-

**Fig. 1.** PMFs for CH–O pair interactions using five atom types. In all figures, the symbols are merely connected by smooth curves for visualization, and no smoothing procedure was used. As there were <30 observations in the distance range of 0–2.4 Å, the figure is short of statistics in this distance range. To reflect the trends of the potentials in the range, the dashed line is appended. a, the potentials for the hydrogen bond acceptor–CH interactions (A–CH and CH–A) in protein-protein associations are shown. The black curve represents the A–CH interaction and the red curve represents the CH–A interaction. b, the potentials for N and CH interactions (B–CH and CH–B) in protein-protein associations are shown. The black curve is the B–CH interaction, and the red curve is the CH–B interaction. c, the potentials for the three representative types of different interactions are shown. The black curve is the D–A hydrogen bonding interaction, the red curve is the CH–O H-bonding interaction, and the green curve is the non-polar pair interaction between atoms of type N.

**Fig. 2.** Contribution of CH–O hydrogen bonding interactions in protein-protein complexes. a, the average contribution of conventional hydrogen bonding interactions, hydrophobic interactions, and CH–O interactions. b, the distribution of the percentages of CH–O hydrogen bonding interactions in the total binding energy of protein-protein complexes. The y axis represents the number of occurrence. c, the examples of protein complexes with a strong contribution of CH–O interactions. The x axis represents the CH–O interaction percentage in the protein-protein complexes. The protein-protein complexes are denoted by Protein Data Bank codes.
hydrogen bond acceptor-acceptor and donor-donor was also analyzed, with a mean percentage of 3%, which shows that it plays a weaker role at protein interfaces than the three major types of interactions (Fig. 2b).

**Structural Features of CH–O Interactions in Some Protein Complexes with High CH–O Interaction Percentages**

The protein-protein complexes with relatively high percentages of CH–O interactions were analyzed (Fig. 2c). We carefully surveyed the geometrical parameters and structural features of these contacts in some representative examples, focusing on the contacts between adjacent β-strands and α-helices located at protein-protein interfaces.

Adjacent β-Strands in Parallel and Antiparallel Orientations—The stereochemistry of close CH–O contacts between adjacent β-strands at the interface was analyzed in the representative examples with a high contribution of CH–O interactions. The Protein Data Bank codes include 2kin, 2gac, 2bqp, 1pya, 1prt, 1lya, 1kvd, 1fi8, 1dgw, and 1apy. Fig. 2e shows the energy percentages of CH–O interactions in each protein complex, which vary from 28.2 to 49.1%. In these representative examples, the CH–O contacts often occur between adjacent β-strands in parallel and antiparallel orientations. The most common ones are the contacts involving α-carbon and those involving main chain oxygen, especially the C=O–O=C (main chain) contact, which occur frequently between adjacent strands at protein-protein interfaces. A detailed analysis of these CH–O contacts in all the representative examples is given in the Supplemental Material.

Fig. 3 and Table II show representative examples of C=O–O interactions between the β-strands in parallel and antiparallel orientations and list their relevant stereochemical details. These stereochemical details are consistent with the geometry of the C=O–O H-bond reported in different systems (17, 18, 58–60). There is an obvious structural feature in protein-protein interfaces: between adjacent β-strands in both parallel and antiparallel orientations, the formation of the C=O–O H-bond combines with the conventional C=O–HN H-bond to form a bifurcated H-bond, where the carbonyl oxygen atom on one strand forms a H-bond with both the amide hydrogen and the α-carbon hydrogen atom on the other strand. Once the bifurcated H-bond is included, the adjacent β-strands in antiparallel orientation (Fig. 3b) look remarkably like those in parallel orientation (Fig. 3a).

Interhelical CH–O H-bonds in Protein-Protein Interfaces—The geometrical parameters of close CH–O contacts between adjacent α-helices were surveyed in the representative examples with a high contribution of CH–O interactions. The Protein Data Bank entries (including codes 2occ, 1wht, 1aig, 1ryp, and 1qgh) have a high energy percentage of CH–O interactions in each protein complex, varying from 28.8 to 44.2% (Fig. 2c). In these representative examples, the contacts involving polar side chain carbon atoms frequently occur between adjacent α-helices. Fig. 4a and Table II show representative examples of CH–O interactions between antiparallel α-helices and list their relevant stereochemical details. A detailed analysis of these CH–O contacts is given in the Supplemental Material.

The structural features and characterizations of close CH–O contacts between quasi-vertical helix segments are relatively few in number. Their structural features and characterizations are rather esoteric, so no further surveys were possible. Nevertheless, it is still notable that these close CH–O interactions show geometrical parameters very close to those of conventional hydrogen bonds (Fig. 4b and Table II).
tween the β-strands in parallel and antiparallel orientations were compared between the crystal structure and differently minimized structures with and without inclusion of the CH–O hydrogen bonding interaction energies. Table III lists the C–O distances of the CH–O hydrogen bonds between the adjacent β-strands in different structures. The C–O distance shifts in the minimized structures for the selected CH–O hydrogen bonds between adjacent β-strands in the two examples (Protein Data Bank codes 2kin and 1apy) are shown in Fig. 5 (a and b). It is clear that energy minimization without taking the CH–O hydrogen bonding into account will result in longer C–O distances compared with the original crystal structures, implying that the empirical force fields that treat CH–O contacts as repulsive need to be modified for these examples.

### DISCUSSION

**Simple Atom Type Definition**—Biological interfaces of protein-protein complexes contain many specific interactions, including hydrogen bonding, water bridging interactions, and nonspecific interactions (32, 33). In our method, the two atoms of each atom pair at protein-protein interfaces contact through either hydrogen bonding interactions (conventional or CH–O hydrogen bonding interactions) or hydrophobic interactions, which are pertinent on protein interfaces and tend to reflect the forces driving the association of protein complexes (Table I). The occurrence of each atom pair is sufficiently high in all the statistics. Moreover, we omitted the statistics in the distance shells with low atom pair occurrence to reduce the mistakes of insufficient statistics. We believe that the details of the potentials obtained (which is the basis for the comparison of different interactions) are meaningful.

**Is the CH–O Pair Interaction a Real Hydrogen Bond?**—As the favorable potential of the CH–O interaction at protein-protein interfaces has not yet been obtained experimentally, we will discuss the question of whether the CH–O interaction is a real hydrogen bond or one of the nonspecific interactions based on the study here. The reasonable Helmholtz free energies of the CH–O interaction at protein-protein interfaces were extracted using our mean-field potential method, and the calculated energies obviously reflect the characterization of the hydrogen bonding interaction. Recently, the results of ab initio quantum calculation showed that the binding energies of ideal CH–O pair interactions indicate the comparative hydrogen bond energy in the calculation between water molecules and some representative amino acids (39). Here, using our statistical potentials from the training set of real protein-protein complexes, the calculated free energy has a reasonable potential form, which is similar to the quantum calculation of the ideal model.

At the same time, we found that the favorable valley of our calculated CH–O potential is very different from that of nonspecific interactions, which generally reflect the random contacts between atoms. In Fig. 1c, the van der Waals repulsive potential of the N–N atom pair is just one of the nonspecific interactions that represent the distribution between nonbonded and randomly distributed nonpolar atoms. These nonspecific interactions have an even potential close to zero at longer distances and have a rapidly climbing potential at short distances (up to 4.0 Å). More importantly, they have no obviously favorable valleys at all distances. Thus, the favorable valley of the CH–O potential shows that, different from random contacts, the CH–O pair interaction is one of the specific potentials that have the characterization of the hydrogen bonding interaction. Moreover, this shape of potential is similar to that of a conventional hydrogen bond, despite the different strengths and optimum distance. Thus, the CH–O contact at protein-protein interfaces has specific interactions similar to conventional hydrogen bonds. According to the calculation of our PMF at protein-protein interfaces, the CH–O contact should be considered as one of the hydrogen bonding interactions. Our conclusion is supported by the ab initio quantum calculation of Scheiner et al. (38, 39) that CH–O contacts appear to be a true hydrogen bond. Compared with the conventional hydrogen bond, the general CH–O hydrogen bond has rather weaker strength and a longer optimum distance, with its strength near one-third of that of the conventional hydrogen bond and its optimum distance −3.3 Å.

### Table II

| PDB code | Acceptor | C–H donor | Calculated energy | d | dH | θ |
|----------|----------|-----------|------------------|---|----|---|
| 2kin     | O Arg78A | C= Gly80A | −1.9             | 2.26 | 145.6° | --- |
| 2kin     | O Thr78A | C= Thr80A | −1.5             | 2.49 | 129.9° | --- |
| 2kin     | O Thr79A | C= IIe82A | −1.9             | 2.19 | 157.4° | --- |
| 2kin     | O Phe83A | C= IIe86A | −0.2             | 2.44 | 140.2° | --- |
| 2kin     | O Va310B | C= Ala84A | −1.9             | 2.26 | 145.9° | --- |
| 2kin     | O Thr85A | C= IIe88B | −2.0             | 2.37 | 141.0° | --- |

---

The pair energy is recalculated by the potentials.

The donor-acceptor distance of the CH–O hydrogen bond.

The H–O distance, i.e., the CH–O hydrogen bond distance.

The C–H–O angle of the CH–O hydrogen bond.
At the protein-protein interface, the average contribution of the conventional hydrogen bond is \(~30\%\); the percentage of the CH–O hydrogen bond is \(~17\%\); that of the hydrophobic interaction is \(50\%\); and that of the other steric interaction is \(3\%\). Conventional hydrogen bonding and hydrophobic interactions are generally considered to play important roles in protein associations. The largest contribution of hydrophobic interactions indicates that the large number of hydrophobic atom pairs occur at the protein-protein interface because the energy of each hydrophobic atom pair is less than that of the other pair interactions. Also, the \(30\%\) contribution of conventional hydrogen bonds shows the significant involvement of charged or polar residues in protein-protein association that is accepted widely (20, 21).

The most important finding here is that the energy contribution of the weak CH–O hydrogen bonding interaction at protein-protein interfaces cannot be neglected. Close CH–O contacts abound at protein-protein interfaces. Although normally weaker than conventional hydrogen bonds, the number of CH–O contacts cannot be neglected. In our calculation of each protein-protein complex, it should be noted that the energy percentages of CH–O hydrogen bonding interactions in some examples are up to \(~30–50\%\) for protein-protein associations.

**Structural Features of CH–O Hydrogen Bonds at Some Protein Interfaces with a High CH–O Contribution:** Bifurcated H-bond Motif between Adjacent \(\beta\)-Strands—We have analyzed the protein-protein complexes with relatively high percentages of CH–O interactions, focusing on the geometrical parameters of the CH–O hydrogen bonds between adjacent \(\beta\)-strands and \(\alpha\)-helices at the interface. These CH–O H-bonds appear to be a specific interaction favoring the formation and stabilization of the adjacent \(\alpha\)-helices and \(\beta\)-strands (acting as a secondary hydrogen bond between \(\beta\)-strands compared with a conventional C=O–HN hydrogen bond). Especially C\(^{\\prime}\)H–O H-bonds between adjacent strands in parallel and antiparallel orientations have an important contribution at the protein interface. It is interesting to note that the structural motif of a bifurcated hydrogen bond is found between the adjacent strands at some representative examples of real protein interfaces (Fig. 3, a and b), and the weak CH–O hydrogen bond secondarily affects the stabilization of the \(\beta\)-strands. In these examples, the CH–O interaction has a high energy percentage of the total free binding energy, and the bifurcated H-bond motif appears to have an important contribution to the formation and stabilization of the parallel and antiparallel \(\beta\)-strands.

In the protein complexes with strong CH–O interactions, close C\(^{\\prime}\)H–O contacts frequently occur between the parallel and antiparallel \(\beta\)-strands, and the structural motif of bifurcated hydrogen bonds was found between those adjacent strands. The C\(^{\\prime}\)H–O hydrogen bond is indispensable in forming the bifurcated H-bond motif between \(\beta\)-strands. In these complexes, the CH–O hydrogen bonding interaction has a high percentage of binding free energy and thus plays an important role in the formation of the \(\beta\)-strand interactions at protein-protein interfaces.

**Is the CH–O Interaction Important?**—The quantitative PMF calculation of the energy contribution has shown that the energy contribution of the CH–O hydrogen bond has a mean value of 17% in the data set of 469 protein complexes and cannot be neglected at protein-protein interfaces, indicating the importance of the CH–O H-bond. Particularly, it can be up to \(~30–50\%\) in some examples. In these examples, close C\(^{\\prime}\)H–O contacts often occur, and the bifurcated H-bond motif commonly exists between parallel and antiparallel \(\beta\)-strands at the interface, suggesting that they need to be paid more attention.
In fact, close C=H–O contacts often occur, and the bifurcated H-bond motif commonly exists between parallel and antiparallel β-strands in protein interiors and at protein-protein interfaces. There are many surveys of high resolution protein structures that reveal the occurrence of C=H–O hydrogen bonds in both parallel and antiparallel β-sheets (18, 59). Very recently, Ho and Curmi (60) reported that the twist and shear of the β-ribbon can be reproduced through the simulation of a simple model using bifurcated hydrogen bonds. In contrast, isolated β-strands are not twisted in the molecular dynamics simulation, where the CH=O=C pair is not considered as an H-bonded pair (26). This example shows that weak C=H–O=C hydrogen bonds do have a strong impact on protein structures.

Our quantitative PMF calculation has shown the strong energy contribution of the CH–O hydrogen bond between the β-strands in some examples. Furthermore, we have made comparisons of energy minimization calculations with and without inclusion of the CH–O hydrogen bonding interaction energies. Our calculations have shown that the empirical force fields that treat CH–O contacts as non-H-bonded pairs are not suitable in some examples, especially between adjacent β-strands.

Both the study of Ho and Curmi (60) and our own calculation suggest that the C=H–O contact is a specific interaction favoring the formation of β-strands and that the bifurcated H-bond is an obvious structural feature between the adjacent β-strands, regardless of proteins or protein-protein interfaces, and it has an important contribution to the formation and stabilization of β-strands. Moreover, the observed geometrical parameters of all the close contacts are very close to those expected for hydrogen bonds. Recently, stereochemical analysis of close CH–O contacts has reported that these contacts exhibit stereochemical features typical of hydrogen bonds in proteins, membrane proteins, and active sites of proteases (17, 18, 58–60). Adding our observation at protein-protein interfaces, all these studies show that the CH–O H-bond appears to be a specific interaction with a favorable valley regardless of protein interiors, membrane proteins, or protein-protein interfaces. However, popular crystal refinement programs often treat the CH–O contact as a repulsive interaction. The importance of the CH–O interaction (indicated by our quantitative energy calculation) calls for a revision of the refinement programs.

**Conclusion**—An empirical approach to quantitatively describe forces (including CH–O hydrogen bonds) at protein interfaces from the energy aspect is presented here. Our calculated Helmholtz free energies of the CH–O pair have a similar favorable valley exhibiting a feature typical of hydrogen bonds. When applying the scoring function to the set of protein-protein complexes, we found the significant contribution of CH–O hydro.
CH–O Hydrogen Bonds at Protein-Protein Interfaces

droben bonds. The average energy contribution of the CH–O H-bond is 17%, and this value can reach as high as ~40–50% in some complexes. In these complexes, the structural motif of a bifurcated H-bond (combining the C=O–H O-bond with the conventional C–O–HN H-bond) is found between adjacent strands in both parallel and antiparallel orientations at the interface.

In conclusion, the importance of CH–O hydrogen bonding interactions calls for a revision of the point of view that CH–O contacts are repulsive. When studying protein-protein interfaces, the CH–O hydrogen bonds should be taken into appropriate consideration.

Acknowledgments—We thank Dr. Chao Tang for inspiring discussions of the simple model of CH–O hydrogen bonds, Ying Gao for kindly supplying the data set, Hao Chen for helpful assistance in energy minimization, and all other members of the Lai group for helpful discussions.

REFERENCES

1. Jeffrey, G. A., and Saenger, W. (1991) Hydrogen Bonding in Biological Structures, Springer-Verlag, Berlin
2. Scheiner, S. (1997) Hydrogen Bonding: A Theoretical Perspective, Oxford University Press, New York
3. Egli, M., and Gessner, R. V. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 180–184
4. Berger, J., Egli, M., and Rich, A. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 12116–12121
5. Ornstein R. L., and Zheng, Y. J. (1997) J. Biomol. Struct. Dyn. 14, 657–665
6. Takahara, P. M., Frederick, C. A., and Lippard, S. J. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 1143–1159
7. Baker, E. N., and Hubbard, R. E. (1984) Prog. Biophys. Mol. Biol. 44, 97–179
8. Stickle, D. F., Presta, L. G., Dill, K. A., and Rose, G. D. (1992) J. Mol. Biol. 226, 241–252
9. McDonald, R. K., and Thornton, J. M. (1994) J. Mol. Biol. 238, 777–792
10. Wahl, M. C., and Sundaralingam, M. (1997) Trends Biochem. Sci. 22, 97–102
11. Desiraju, G. R., and Steiner, T. (1999) The Weak Hydrogen Bond in Structural Chemistry and Biology, Oxford University Press, New York
12. Taylor, R., and Kenndee, O. (1982) J. Am. Chem. Soc. 104, 5063–5070
13. Steiner, T., and Saenger, W. (1992) J. Am. Chem. Soc. 114, 10146–10154
14. Jeffrey, G. A., and Maluszyńska, H. (1982) Int. J. Biol. Macromol. 4, 173–185
15. Thomas, K. T., Smith, G. M., Thomas, T. B., and Feldmann, R. J. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 4843–4847
16. Steiner, T., and Saenger, W. (1993) J. Am. Chem. Soc. 115, 4540–4547
17. Derewenda, Z. S., Derewenda, U., and Kobos, P. M. (1994) J. Mol. Biol. 241, 83–93
18. Derewenda, Z. S., Lee, L., and Derewenda, U. (1995) J. Mol. Biol. 252, 248–262
19. Bella, J., and Berman, H. M. (1996) J. Mol. Biol. 264, 734–742
20. Chakrabarti, P., and Chakrabarti, S. (1998) J. Mol. Biol. 284, 867–873
21. Ash, E. L., Sudmeier, J. L., Day, R. M., Vincent, M., Torchilin, E. V., Haddad, K. C., Bradshaw, E. M., Sanford, D. G., and Bachovchin, W. W. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 10371–10376
22. Musah, R. A., Jensen, G. M., Rosenfeld, R. J., Melfee, D. E., Goodin, D. B., and Bunte, S. W. (1997) J. Am. Chem. Soc. 119, 9083–9084
23. Bauree, P. W., Wizycki, A., and Beatty, A. M. (2000) Biorg. Med. Chem. 8, 1599–1605
24. Mandel-Gutfreund, Y., Margalit, H., Jernigan, R. L., and Zhurkin, V. B. (1998) J. Mol. Biol. 277, 1129–1140
25. Gray, N. S., Wodicka, L., Thunnissen, A. W. H., Norman, T. C., Kwon, S., Esponzoa, F. H., Morgan, D. O., Barnes, G., Leclerc, S., Meijer, L., Kim, S.-H., Lockhart, D. J., and Schultz, P. G. (1998) Science 281, 533–538
26. Wang, L., O’Connell, T., Trokha, A., and Hermans, J. (1996) J. Mol. Biol. 262, 283–293
27. Desiraju, G. R. (1998) Acc. Chem. Res. 29, 441–449
28. Auffinger, P., and Westhof, E. (1997) J. Mol. Biol. 274, 54–63
29. Chattopadhyay, C., and Jancin, J. (1975) Nature 256, 705–708
30. Janin, J., and Chothia, C. (1990) J. Mol. Biol. 265, 16027–16030
31. Jones, S., and Thornton, J. M. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 13–20
32. Xu, D., Tsai, C., and Nussinov, R. (1997) Protein Eng. 10, 999–1012
33. Scheiner, S. (2000) in Advances in Molecular Structure Research (Hargittai, M., and Hargittai, I., eds) Vol. 6, pp. 159–206, Oxford University Press, New York
34. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, N. T., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) Nucleic Acids Res. 28, 235–242
35. Meyers, E. W., and Miller, W. (1989) Bioch. Biophys. Acta 97, 5–37
36. Sippl, M. J. (1995) Curr. Opin. Struct. Biol. 5, 229–235
37. Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) Nature 358, 86–89
38. Jernigan, R. L., and Bahar, I. (1996) Curr. Opin. Struct. Biol. 6, 195–209
39. Vajda, S., Sippl, M., and Novotny, J. (1997) Curr. Opin. Struct. Biol. 7, 222–228
40. Wallqvist, A., Jernigan, R. L., and Cowell, D. G. (1995) Protein Sci. 4, 181–1903
41. DeWitte, R. S., and Shakhnovich, E. I. (1996) J. Am. Chem. Soc. 118, 11733–11744
42. DeWitte, R. S., and Shakhnovich, E. I. (1997) J. Am. Chem. Soc. 119, 4608–4617
43. Mitchell, J. B. O., Laskowski, R. A., Alex, A., and Thornton, J. M. (1999) J. Comput. Chem. 20, 1165–1176
44. Muegge, I., and Martin, Y. C. (1999) J. Med. Chem. 42, 791–804
45. Ho, B. K., and Curmi, P. M. G. (2002) J. Mol. Biol. 317, 291–308
46. Brook, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1985) J. Comput. Chem. 4, 187–217