

## Development of a Knowledge-Based Potential for Crystals of Small Organic Molecules: Calculation of Energy Surfaces for C=O···H–N Hydrogen Bonds

Bartosz A. Grzybowski,<sup>†</sup> Alexey V. Ishchenko,<sup>†</sup> Robert S. DeWitte, George M. Whitesides, and Eugene I. Shakhnovich\*

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

Received: February 18, 2000; In Final Form: May 1, 2000

This paper describes the derivation of a Knowledge-Based Potential for intermolecular interactions from the statistical information stored in the Cambridge Structural Database. We develop a statistical mechanical method that relates the occurrences of intermolecular contacts in the database to their energies. Our approach allows us to quantify (in the form of energy) the geometrical preferences of interactions. We use our method to construct energy maps for a hydrogen bond between carbonyl oxygen and amino hydrogen. Our results demonstrate high orientational selectivity of this type of hydrogen bonding.

### Introduction

Robust, predictive computational tools for many aspects of chemical design aid experimental programs in drug discovery,<sup>1,2</sup> materials,<sup>3</sup> and self-assembly,<sup>4–6</sup> to name a few areas of current interest. These tools must incorporate both molecular design elements and adequate functional understanding of intermolecular interactions. This functional understanding should be distinguished from fundamental understanding, which, though very desirable, is not necessary. Rather, what is required is a means to quantify interactions between atoms, functional groups, and other, larger structural elements.

Empirical force fields are used in CHARMM,<sup>7,8</sup> AMBER,<sup>9</sup> OPLS,<sup>10,11</sup> MMFF,<sup>12</sup> GROMOS,<sup>13</sup> MM3,<sup>14</sup> and heuristic rule-based methods.<sup>15</sup> These methods have the very well-understood shortcoming of limited transferability, resulting from parametrizations on specific systems. Indeed, it is appropriate to develop potential energy functions tailored to each physical phenomenon in which predictive computational tools are sought. This paper is the first of a series aimed at generating and using a potential energy function for the prediction of crystal structures of small organic molecules.

The prediction of the structures of molecular crystals has proven to be a challenging problem.<sup>16–18</sup> In one common computational method, symmetry operations are used to construct crystal “precursors” (clusters) of several molecules. Their energies are evaluated, and the best ones are chosen. These are expanded into full, three-dimensional crystals using translational search algorithms.<sup>19–21</sup> The main flaw of this approach is that there is no guarantee that a low-energy cluster of a few molecules will represent the low-energy structure in the crystal. Another widely applied strategy is based on the Monte Carlo simulated annealing algorithm. This search method uses, from the beginning, periodic boundary conditions (i.e., a crystal, not a cluster approach). It varies parameters describing the packing in the crystal (unit cell parameters and Euler angles) according to Metropolis criterion.<sup>22–25</sup> The success of this type of approach is highly dependent on the accuracy of the force field used. With common force fields, such as CHARMM, it was possible

to predict crystal structures of only certain classes of simple molecules, but for more complicated cases (e.g., when many polymorphs exist), no consistent computational method is available.

The Cambridge Structural Database (CSD) provides the statistical information describing the occurrence of specific fragments (synthons,<sup>26</sup> polymorphs,<sup>27</sup> or particular intermolecular interactions<sup>28</sup>) in crystals. When combined with computational methods or some phenomenological assumptions (constant packing coefficient,<sup>29</sup> Etter’s postulates<sup>30–32</sup>), the statistics extracted from CSD could be helpful in predicting crystal structures. Although several researchers have collected CSD statistics describing particular types of intermolecular interactions,<sup>33–41</sup> the information stored in CSD has not been expressed in the quantitative form of an empirical potential.

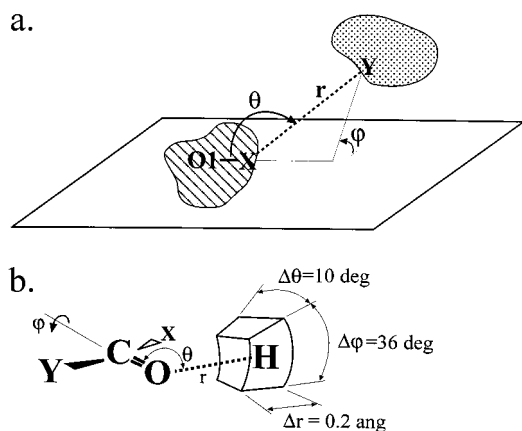
In this work, we present an algorithm that allows construction of an empirical potential function from the statistical information stored in CSD. The approach we use, referred to under several names (Knowledge-Based Potential, Quasi-Chemical Approximation) has been used in studies of protein folding,<sup>42,43</sup> prediction of protein–ligand binding affinity,<sup>44,45</sup> and *de novo* ligand design,<sup>46</sup> and it correlates frequencies of observed geometries to predictive energies by the way of the assumption that the observations are truly representative. The method we develop is capable of quantifying the orientational preferences of intermolecular interactions (e.g., hydrogen bonds). We apply our algorithm to calculate the potential energy maps of C=O···H–N hydrogen bonds—a key feature of protein–ligand interactions, of crystals, and of tertiary structure in biological and nonbiological molecules.<sup>47</sup>

### Theory

In this section, we develop the statistical–mechanical formalism of our knowledge-based potential. We (i) convert the observed frequencies to probability densities, (ii) construct a scoring function that is related to probability densities by a Boltzmann-like relationship, and (iii) show that this scoring function has a meaning of true energy.

We consider a database of  $\chi$  molecular crystals. We specify the cutoff radius for intermolecular contacts  $R$ , such that two

<sup>†</sup> The authors contributed equally to this research..



**Figure 1.** Coordinate system used in describing (a) general interaction  $X\cdots Y$  and (b)  $C=O\cdots H-N$  hydrogen bond. In (a)  $r$  is defined as the distance between atoms  $X$  and  $Y$ ,  $\theta$  is the angle  $O1-X-Y$  (where  $O1$  is an arbitrary reference point), and the dihedral angle  $\varphi$  is taken with respect to a plane containing  $X$  and  $O1$ . In (b),  $X$  and  $Y$  are atoms covalently bonded to the carbonyl carbon ( $X = C$  and  $Y = N$  for amides,  $X = C$  and  $Y = O$  for esters). Volume element ( $\Delta r, \Delta\theta, \Delta\varphi$ ) around the hydrogen atom is also shown.

atoms are said to form an intermolecular contact if their separation is smaller than or equal to  $R$ . The number of intermolecular contacts  $N_C$  in crystal  $C$  is defined as the number of contacts between the atoms of an arbitrary molecule in this crystal with atoms of all other molecules. The total number of intermolecular contacts in the database is  $N\chi = \sum_C N_C$ . Consider an intermolecular interaction  $xy$  between atoms  $x$  and  $y$ . If the number of contacts of this type in the database is  $n_{xy}$ , then the probability of occurrence of such contact in the database is  $p_{xy} = n_{xy}/N\chi$ , independent of the geometry of the interaction.

To investigate the geometric preferences of the  $xy$  contacts, we introduce a local spherical system of coordinates  $r, \theta, \varphi$  (Figure 1). The probability  $p_{xy}(r, \theta)$  of occurrence of the  $xy$  interaction within the volume element  $V(r, \theta)$ :  $r \in (r, r + \Delta r)$  and  $\theta \in (\theta, \theta + \Delta\theta)$ , irrespective of the value of the azimuthal angle  $\varphi$ , is given by eq 1, in which  $p_{xy}(r, \theta|xy)$  is the conditional probability, and  $n_{xy}(r, \theta)$  is the number of interactions in  $V(r, \theta)$ . The corresponding probability density

$$p_{xy}(r, \theta) = p_{xy}(r, \theta|xy) p_{xy} = n_{xy}(r, \theta)/N\chi \quad (1)$$

is  $\rho_{xy}(r, \theta) = n_{xy}(r, \theta)/(N\chi V(r, \theta))$ . We define a scoring function  $E$  ("energy") of interaction, which is related to the probability density in a Boltzmann-like fashion. The normalizing factor  $Z$  is analogous to a partition function over energy states in Boltzmann's statistics (eq 2) and can be eliminated by an appropriate choice of the reference state (eqs 3–5).

$$\rho_{xy}(r, \theta) = Z^{-1} \cdot \exp(-E(r, \theta)/kT);$$

$$Z = \sum_{r, \theta} \Omega(E(r, \theta)) \cdot \exp(-E(r, \theta)/kT) \quad (2)$$

$$E_{\text{ref}}(r, \theta)/kT = -\ln \rho_{\text{ref}}(r, \theta) - \ln Z \quad (3)$$

$$E^*(r, \theta) = E(r, \theta) - E_{\text{ref}}(r, \theta) \quad (4)$$

$$E^*(r, \theta) = -kT \ln \frac{\rho_{xy}(r, \theta)}{\rho_{\text{ref}}(r, \theta)} \quad (5)$$

We took the reference state of constant probability density  $\rho_{\text{ref}}$ , which we set equal to  $n_{xy}/(N\chi \sum_{r, \theta} V(r, \theta))$  to ensure proper normalization of probabilities.

To study the dependence of energy on the azimuthal angle  $\varphi$ , we proceed as before and define the probability of forming a contact  $xy$  within the volume element  $V(r, \theta, \varphi)$ :  $r \in (r, r + \Delta r)$ ,  $\theta \in (\theta, \theta + \Delta\theta)$ , and  $\varphi \in (\varphi, \varphi + \Delta\varphi)$  by  $p_{xy}(r, \theta, \varphi) = p_{xy}(\varphi|r, \theta) p_{xy}(r, \theta) = n_{xy}(r, \theta, \varphi)/N\chi$ , where the conditional probability  $p_{xy}(\varphi|r, \theta) = n_{xy}(r, \theta, \varphi)/n_{xy}(r, \theta)$ . For each volume element  $V(r, \theta, \varphi)$ , we define the corresponding energy  $E^*(r, \theta, \varphi)$  as the sum of  $E^*(r, \theta)$  and the correction energy  $\epsilon^*(\varphi|r, \theta)$ . Because the average value of energy  $E^*(r, \theta, \varphi)$  over the azimuthal angle is equal to  $E^*(r, \theta)$ , the average of  $\epsilon^*(\varphi|r, \theta)$  is zero (eq 6). We introduce coefficients  $\lambda$  (adjusting the zero of energy for each volume element  $V(r, \theta, \varphi)$  with respect to  $E^*(r, \theta)$ ) and  $\beta$  (a small number ensuring finiteness of energy; typically  $\sim 0.0001$ ) and stipulate, as before, the logarithmic dependence of the energy correction on probability density (eq 7). Solving eq 7 for  $\lambda(r, \theta)$ , leads to the final expression for  $E^*(r, \theta, \varphi)$  (eq 8).

$$\sum_{\varphi} p_{xy}(\varphi|r, \theta) \epsilon^*(\varphi|r, \theta) = 0 \quad (6)$$

$$\sum_{\varphi} \rho_{xy}(\varphi|r, \theta) (\ln(\rho_{xy}(\varphi|r, \theta) + \beta) - \lambda(r, \theta)) = 0 \quad (7)$$

$$E^*(r, \theta, \varphi) = \lambda(r, \theta) - kT \ln(\rho_{xy}(r, \theta) (\rho_{xy}(\varphi|r, \theta) + \beta) / \rho_{\text{ref}}(r, \theta)) \quad (8)$$

Because the probabilities of interactions of different types are statistically independent (we assume no significant cross-correlations within the database), our definition of the scoring function ("energy") proportional to the logarithm of the probability density ensures the additivity of pairwise interaction energies. In addition, the energies defined in this way have the meaning of true, physical energies. To show this, we start by investigating how the strength of a particular intermolecular interaction in a molecular crystal influences the entropy of the crystal. Consider a set  $\Omega_M$  of all crystals of small organic molecules, which have the same molecular formula  $M$  (i.e., structural isomers). Because the packing coefficients in crystals of nonionic organic molecules are nearly constant,<sup>48</sup> and because the molecular volumes of structural isomers are similar,<sup>49</sup> the intermolecular energies<sup>50</sup> (in a coarse-grained potential) all have roughly the same value  $E_M$  in all crystals of  $\Omega_M$ . By construction, the sum of energies of pairwise intermolecular interactions (ij) in any of  $\Omega_M$  crystals is approximately equal to  $E_M$ ,  $\sum_{ij} E_{ij} = E_M$ . We select one of the interactions ( $xy$ ) having energy  $E_{xy}$ . The number of  $\Omega_M$  crystals with this interaction present is denoted  $\Omega_M(E_{xy}) = \Omega_M(\sum_{ij \neq xy} E_{ij} \approx E_M - E_{xy})$ . The right-hand side of this equation gives the number of crystals, in which the sum of energies of all interactions other than  $xy$  is equal to  $E_M - E_{xy}$ , and can be related to the number of ways the atoms in a molecule of formula  $M$  can be arranged to give a stable crystal with  $xy$  interaction. Therefore,  $\Omega_M(E_{xy})$  can be viewed as a measure of conformational entropy  $S_M(E_{xy}) = k \ln \Omega_M(E_M - E_{xy})$ . If  $E_M \gg E_{xy}$  (no single interaction dominates the interaction energy; true for organic, nonionic molecules), the logarithm can be expanded in a Taylor series around  $E_M$  (eq 9).

$$\ln \Omega_M(E_M - E_{xy}) = \ln \Omega_M(E_M) - \frac{dS_M}{dE} \Big|_{E=E_M} \cdot \frac{E_{xy}}{k} \quad (9)$$

The derivative of conformational entropy with respect to energy is a constant, which defines the thermodynamic temperature  $T$  of the database (unknown, as in any knowledge-based potential). Because  $\Omega_M(E_{xy})/\Omega_M$  is the conditional probability of the occurrence of crystals with interaction  $E_{xy}$  in  $\Omega_M$ , (9) can be rewritten in the form of eq 10. Finally, the summation over

molecules of different molecular masses, gives the dependence of the probability of interaction  $xy$  on its energy  $E_{xy}$  (eq 11). Note that, despite its mathematical similarity, eq 11 is *not* a

$$p_M(E_{xy}|M) = (\Omega_M(E_M)/\Omega_M)\exp(-E_{xy}/kT) \quad (10)$$

$$p(E_{xy}) = \sum_M p(M)p_M(E_{xy}|M) = (\Omega_M(E_M)/\chi)\exp(-E_{xy}/kT) \quad (11)$$

statement of Boltzmann distribution of a system (database) attaining canonical equilibrium: a database of structures frozen in their minimal energy conformations does not represent a proper canonical ensemble of many conformational states. The equation does, however, relate the probabilities of occurrence of atom–atom interactions in a database, to the true energies of these interactions.

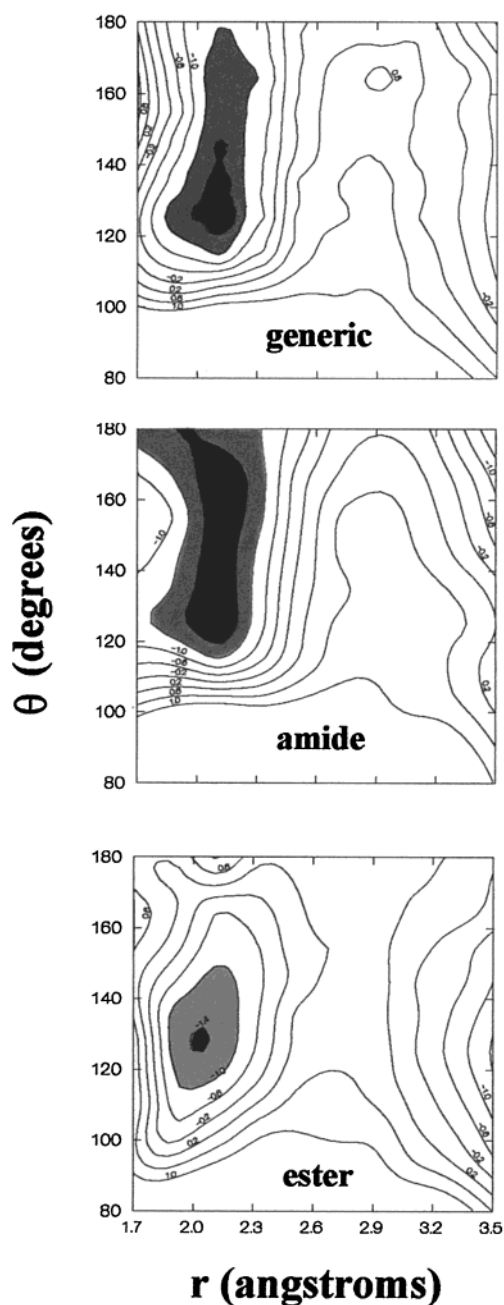
### Computational Methods

We used a hydrogen bond between carbonyl oxygen and amino hydrogen as a model interaction. The geometry of the interaction was defined by three variables  $r, \theta, \varphi$  corresponding to O–H distance, C–O–H angle and X–C–O–H dihedral (azimuthal) angle, respectively (Figure 1b). The cutoff radius was set at 3.5 Å (the value above which the number of hydrogen bonds per unit volume was found to be approximately constant). The  $\theta$  angles considered were between 80 and 180°.

The Cambridge Structural Database was searched for hydrogen bonds of interest. Out of the total of 167 797 entries, only those of very good quality ( $R$ -factor less than 0.05) were considered. This constraint left 27 718 structures subject to further analysis. We further excluded from our searches some classes of compounds in which specific interactions might occlude genuine properties of hydrogen bonding: in particular, ionic substances, metal complexes, and compounds containing diffuse-orbital elements were rejected (Supporting Information).

We found 7424 C=O...H–N hydrogen bonds that fell within these constraints. In 3098 instances, hydrogen bonds were formed by esters (X = C, Y = O) and in 2621 cases by amides (X = C, Y = N). Ureas contributed 949 counts, ketones 374, and remaining classes of compounds 382. It was only for esters and amides that the statistics were abundant enough to guarantee statistical significance. In probing geometric preferences, the  $r, \theta$  plane was divided into 100 volume elements  $V(r, \theta)$  of size 0.2 Å by 10°. A CSD search was performed for every volume element to obtain the number  $n(r, \theta)$  of hydrogen bonds in it. These counts were converted to energies  $E^*(r, \theta)$  using the procedure described in the Theory Section. To study the dihedral-angle preferences of hydrogen bonds, each  $V(r, \theta)$  was subdivided into 10  $V(r, \theta, \varphi)$  volume elements. The statistics were collected as before and subsequently converted to energies  $E^*(r, \theta, \varphi)$ .

In the CHARMM calculation (Figure 4), the atom types and charges for the molecules were assigned by Quanta molecular editor facility. At each  $\theta$ , eight complexes were generated by rotating the donor around H–N axis in 45° intervals. The energy minimization of these complexes was implemented using CHARMM (50 steps of steepest descent followed by adopted basis Newton–Raphson procedure until the convergence of 0.01 kcal/mol was achieved). The interaction value for a given  $\theta$  was calculated by taking the average over these eight complexes (the sterically clashing configurations with the energy greater than 6 kcal/mol were discarded). During the minimization, the positions of hydrogen bonding atoms (H, N, C, O, Y) were

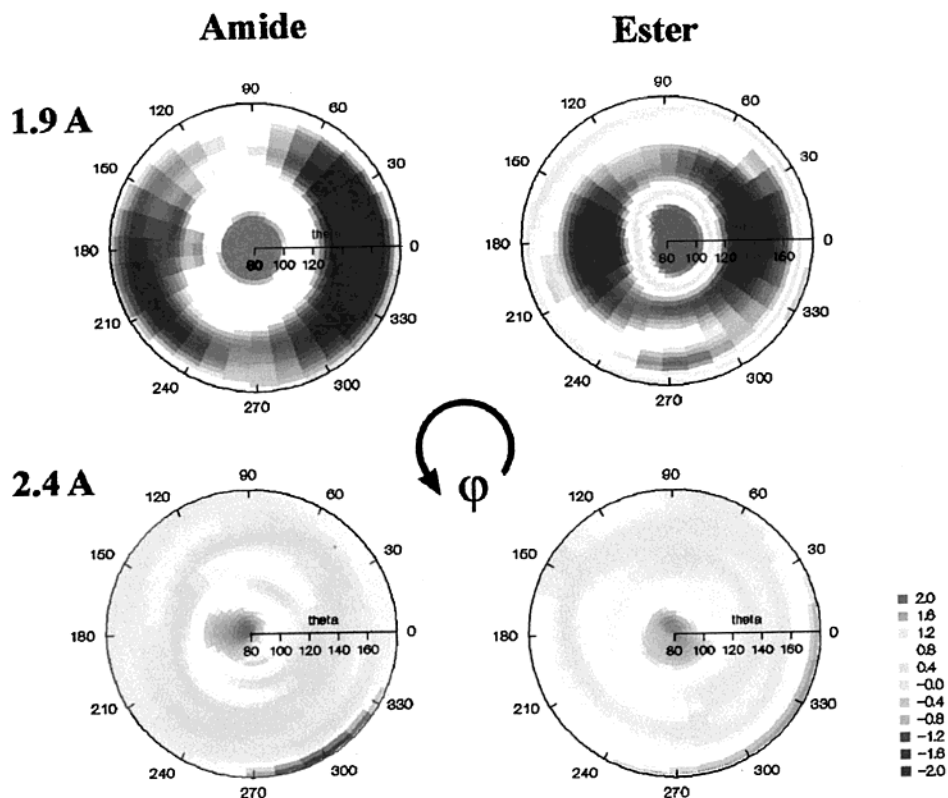


**Figure 2.** Energy maps  $E(r, \theta)$  derived from the CSD for different types of C=O...H–N hydrogen bonds: (a) generic case (X, Y = any atoms), (b) amides (X = C, Y = N), and (c) esters (X = C, Y = O). The energy minima located at distances  $\sim 2$  Å are colored in gray scale. Amides show no angular preference for hydrogen bond in the  $\theta$  range 120–180°. Esters have more localized minimum between  $\theta \sim 120$  and  $\theta \sim 140^\circ$ .

fixed, and the methyl carbons of the acceptors were constrained to the COY plane.

### Results and Discussion

The  $(r, \theta)$  energy maps for C=O...H–N hydrogen bonds are shown in Figure 2. In hydrogen bonds involving a carbonyl group of an arbitrary type (“generic case”; includes amides, esters, etc.), the energy minimum is located between  $r = 1.8$  and 2.2 Å, and  $\theta$  is located in the range  $\sim 115$ – $145^\circ$ , with a “tail” of slightly higher energy extending up to 180°. Interactions corresponding to  $\theta$  below 100° are sterically inaccessible for nearly all distances. No bonds are found for  $r$  smaller than  $\sim 1.5$



**Figure 3.**  $(\theta, \varphi)_r$  maps of the azimuthal energy  $E^*(r, \theta, \varphi)$  for amides and esters. For the O $\cdots$ H distance  $r = 1.9 \text{ \AA}$ , there are two deep minima ( $\varphi = 0^\circ$  and  $\varphi = 180^\circ$ ) corresponding to the donor hydrogen atoms positioned in the plane of the carbonyl. In amides, the  $\varphi = 0^\circ$  minimum is deeper than that at  $\varphi = 180^\circ$ . In esters, the minima have almost equal depth. At larger separation ( $r = 2.4 \text{ \AA}$ ), there is no azimuthal preference for hydrogen bonding and the profiles are almost uniform without distinctive minima.

$\text{\AA}$ . For large  $r$  (above  $\sim 3 \text{ \AA}$ ), there is no angular preference for bonding: the energy surface is uniform in the  $\theta$  direction. Amides show low angular specificity in forming hydrogen bonds: the best bonds are observed at distances  $r$  roughly the same as in generic case, but in a much broader range of  $\theta$  angles ( $\sim 115\text{--}180^\circ$ ). Amides are better-than-average hydrogen binders because the depth of the energy minimum is larger than that found in the generic map. In esters, hydrogen bonding is weaker than average; the energy minimum is more localized than in the generic case and lies between  $\theta \sim 120\text{--}140^\circ$ .

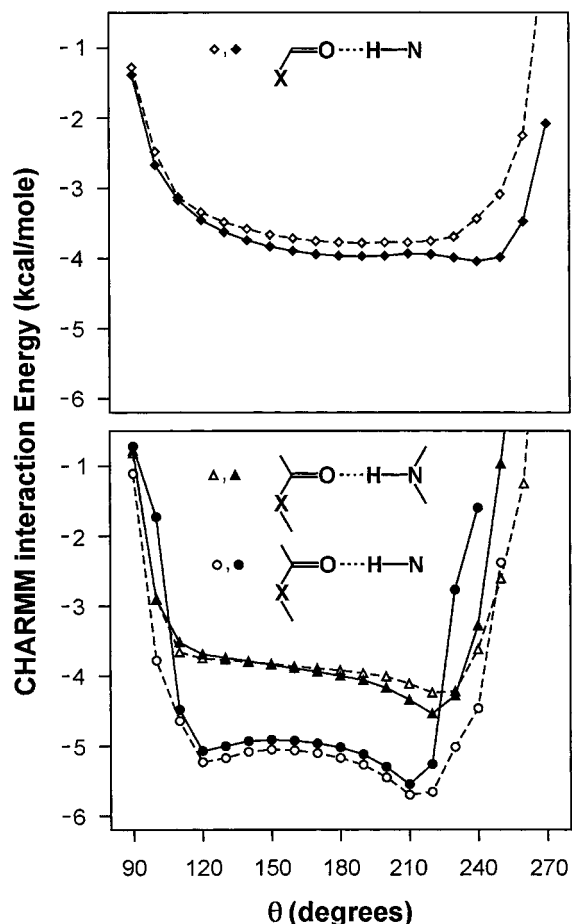
We performed two additional calculations to probe whether the energy maps described above are representative to the entire class of compounds (i.e., esters or amides) and are not skewed by some subgroup of the class. For each class, we constructed a trial set composed of 300 randomly chosen structures and calculated probability densities  $\rho_{\text{trial}}(r, \theta)$  for this set. We correlated probability densities of the trial set with those of the entire class (for the same volume elements). We found statistically significant correlations both for amides, and for esters (correlation coefficients 0.82 and 0.75, respectively), proving the self-consistency of the method.

The dependence of hydrogen bond energy on azimuthal angle  $\varphi$  is shown in Figure 3. For a given range of  $r$ , we constructed  $(\theta, \varphi)_r$  energy maps. We chose two portions of  $(r, \theta)$  conformational space for analysis: a low-energy region between  $r = 1.9 \text{ \AA}$  and  $r = 2.1 \text{ \AA}$  ( $80^\circ < \theta < 180^\circ$ ), and the region of high, roughly constant energy  $2.3 \text{ \AA} < r < 2.5 \text{ \AA}$  ( $80^\circ < \theta < 180^\circ$ ). The  $(\theta, \varphi)_{r=1.9\text{--}2.1 \text{ \AA}}$  energy maps for both amides and esters reveal preference for hydrogen bonding for azimuthal angles near either  $0$  or  $180^\circ$ . For amides, the  $\varphi = 0^\circ$  minimum is slightly deeper than  $\varphi = 180^\circ$ , whereas for esters the opposite is true. At larger distances the azimuthal selectivity for bonding

is absent; there are no distinct minima on the  $(\theta, \varphi)_{r=2.3\text{--}2.5}$  energy maps, either for esters, or for amides.

Our results imply that C=O $\cdots$ H–N hydrogen bonding is a highly distance-dependent and directional interaction, with preference for the hydrogen atom lying in the plane of carbonyl bond and along the direction of oxygen's lone pairs (where the last term is used with due caution<sup>51</sup>). Because the existence (or nonexistence)<sup>52–54</sup> of resonance structures is still a controversial issue, we just briefly notice, that the differences in  $(r, \theta)$  energy maps between amides and esters are compatible with the hypothesis of resonance in amides (development of negative charge on oxygen, resulting in a more electrostatic, less directional character of hydrogen bonding).

In our procedure, we assume that the crystal packing effects do not contribute significantly to the observed frequencies and orientations of individual intermolecular contacts,<sup>55</sup> so that the energies derived from the observed distribution of the CO $\cdots$ NH contacts in the crystal database correspond to the real hydrogen bond energies. In other words, we assume that the tendency of the molecules to adopt the close packing arrangement<sup>29</sup> in the crystal form does not affect the contact statistics strongly enough to make our hydrogen bond potentials significantly different from those in the gas phase. Such an effect would bias the observed hydrogen bond distribution toward one or several particular orientations due to the crystal packing rather than due to pure energetic effects. However, if the database is large and diverse enough as in our case, it is reasonable to assume that the average effect of such biases is small because many different molecular arrangements average out for all atomic contacts and their geometries (unfortunately, the exact proof of this assumption would require computation of higher-order correlations, for which the statistics extracted from the



**Figure 4.** Atomic contributions to the CHARMm interaction energy of the hydrogen bonded dimer (donor, dimethylamine  $(\text{CH}_3)_2\text{NH}$ ; acceptors, methyl acetate  $\text{CH}_3\text{COOCH}_3$ ,  $\text{X} = \text{O}$  and *N*-methylacetamide  $\text{CH}_3\text{CONHCH}_3$ ,  $\text{X} = \text{N}$ ) as a function of  $\theta$  for  $r(\text{O}\cdots\text{H})$  fixed at 1.9 Å. Solid lines and filled symbols correspond to the ester; dashed lines and open symbols correspond to the amide. The atoms included in the calculation of interaction contributions are shown in the figure insert. The energy of the pure hydrogen bond (represented by diamonds) is almost uniform in the  $\theta$  range 120–240°. When the methyl groups are included in the calculation (triangles and circles), the minima in the energy profile appear. The van der Waals attraction between methyl groups and remaining atoms is the strongest for  $\theta = 210^\circ$  and  $\theta = 120^\circ$ . The overall result is the energetic preference of approximately 1 kcal/mol for these dimer configurations. Notice, however, that these minima are *not* due to the pure hydrogen-bond effects.

database are very scarce). Therefore, our method extracts the atomic interaction potentials that have the meaning of “true” energies, in this particular case, the  $\text{CO}\cdots\text{HN}$  hydrogen bond energies.

The ability of our method to account for high directionality of hydrogen bonding is one of its potential advantages over existing force fields. To illustrate this point, we compared our energy maps with those obtained from a commonly used CHARMm potential (Figure 4). We found that CHARMm energy landscapes are virtually flat both for esters and for amides, and they show no angular preferences for contacts. It is only when van der Waals interactions due to the neighboring atoms are included, that directionality appears. Nonetheless, “pure” hydrogen bonds (i.e., only  $\text{C}=\text{O}\cdots\text{H}-\text{N}$ ) calculated by CHARMm are not directional at all.

In our method, we first derive probabilities and energies from the statistics based on only two variables ( $r, \theta$ ), and only later include the dependence on the azimuthal angle. We found this order of calculations advantageous to the (theoretically equiva-

lent) procedure starting with  $(r, \theta, \varphi)$  statistics. If the number of contacts  $n_{xy}$  in a database is moderate, the chances are that there will be no contacts found in some volume elements  $V(r, \theta, \varphi)$ , leading to very high energy values  $E^*(r, \theta, \varphi)$ . On the other hand, if the more abundant  $(r, \theta)$  statistics are collected first, the regions of low and high probability densities can be identified, so that either (i) the volume elements  $V(r, \theta, \varphi)$  can be adjusted accordingly (larger ones for regions of low number of contacts) or (ii) dihedral analysis can be performed only over the high-density  $(r, \theta)$  regions. This approach gives smoother energy surfaces and savings in computer time. Of course, when the statistics for a particular interaction are very scarce in a database, even the analysis starting on the  $(r, \theta)$  level will give meaningless results. For instance, the energy map for ketones (374 hydrogen bonds) showed a rather unexpected minimum at  $1.9 \text{ \AA} < r < 2.1 \text{ \AA}$  and  $170^\circ < \theta < 180^\circ$ . Upon reviewing the raw statistical data, it was found that this result is statistically insignificant because the number of counts in this volume element was only five. To avoid such artifacts, we adopted a criterion for statistical significance stipulating that the relative error ( $n^{-0.5}$ ) be less than 20% (i.e., only volume elements with more than 25 structures are statistically significant).

The hydrogen bonds studied in this paper represent but a small fraction of interactions that need to be incorporated in a complete knowledge-based potential. Thus, our next objective will be to extend the methodology developed here to describe other kinds of hydrogen bonds, as well as nondirectional interactions. Because the energies we extract from CSD using our method have the meaning of true energies, they are in principle transferable to systems other than crystals. In particular, we are interested in incorporating parts of CSD-derived potential (notably, hydrogen bonds) in our SmoG software package<sup>45,46</sup> for protein ligand design. The ultimate test of the accuracy of our potential, would be to compare it with that based on quantum mechanical calculations. Unfortunately, to our knowledge, no high-level precision ab initio orientational potential surfaces have been obtained for model systems that relate directly with our CSD investigations. In the absence of such calculations (which, if properly done would represent a “*tour de force*”), our method might be a complement to existing force fields.

**Acknowledgment.** This work was supported by NIH (Grant No. R0152126).

**Supporting Information Available:** 1. Distance-angle statistics for esters and amides. 2. Angle-dihedral statistics for esters and amides. 3. Distance-angle statistics for different classes of compounds; raw counts. 4. CSD classes excluded from the search. This material is available free of charge via Internet at <http://pubs.acs.org> and [http://www-shakh.harvard.edu/alexey/conh\\_suppl/supplementary.html](http://www-shakh.harvard.edu/alexey/conh_suppl/supplementary.html).

## References and Notes

- (1) Martin, Y. C. *Perspect. Drug. Discovery Des.* **1997**, 7/8, 159.
- (2) Ajay; Murcko, M. A. *J. Med. Chem.* **1995**, 38, 4953.
- (3) Guedes, J. M.; Taylor, J. E. *Struct. Optimization* **1997**, 14, 193.
- (4) Goetz, R.; Gompper, G.; Lipowsky, R. *Phys. Rev. Lett.* **1999**, 82, 221.
- (5) Wennerstrom, H.; Schick, M. S. *Curr. Opin. Colloid Interface Sci.* **1998**, 3, 413.
- (6) Chitanvis, S. M. *Phys. Rev. E* **1998**, 57, 1921.
- (7) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, J.; Karplus, M. *J. Comput. Chem.* **1983**, 4, 187.
- (8) Mackrell, A. D., Jr.; Wiórkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, 117, 11 946.

- (9) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (10) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
- (11) Maxwell, D. S.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1995**, *16*, 984.
- (12) Halgen, T. A. *J. Comput. Chem.* **1996**, *17*, 553.
- (13) Van Gunsteren, W.; Berendsen, H. *Groningen Molecular Simulation Package and Manual (GROMOS)*; Biomos: Groningen, The Netherlands, 1987.
- (14) Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576.
- (15) Böhm, H.-J. *J. Comput.-Aided Mol. Design*, **1992**, *6*, 593.
- (16) Gdanitz, R. J. *Curr. Opin. Solid State Mater. Sci.* **1998**, *3*, 414.
- (17) Dunitz, J. D.; Gavezzotti, A. *Acc. Chem. Res.* **1999**, *32*, 677.
- (18) Chin, D. N.; Zerkowski, J. A.; MacDonald, J. C.; Whitesides, G. M. In *Organised Molecular Assemblies in the Solid State*; Whitesell, J. K., Ed.; Wiley: New York, 1999; Chapter 5.
- (19) Perlstein, J. *J. Am. Chem. Soc.* **1992**, *114*, 1955.
- (20) Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622.
- (21) Gavezzotti, A. *Acc. Chem. Res.* **1994**, *27*, 309.
- (22) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N. *J. Chem. Phys.* **1953**, *21*, 1087.
- (23) Pannetier, J.; Bassasalsina J.; Rodriguez Carvajal, J.; Caignaert, V. *Nature* **1990**, *346*, 343.
- (24) Karfunkel, H. R.; Gdanitz, R. J. *J. Comput. Chem.* **1992**, *13*, 1171.
- (25) Gdanitz, R. J. *Chem. Phys. Lett.* **1992**, *190*, 391.
- (26) Desiraju, R. J. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2311.
- (27) Gavezzotti, A.; Filippini, G. *J. Am. Chem. Soc.* **1995**, *117*, 12 299.
- (28) Desiraju, G. R. *Acc. Chem. Res.* **1996**, *29*, 441.
- (29) Kitaigorodski, A. I. *Organic Chemical Crystallography*; Consultants Bureau: New York, 1961.
- (30) Perlstein, J. *J. Am. Chem. Soc.* **1996**, *118*, 8433.
- (31) Etter, M. C. *J. Phys. Chem.* **1991**, *95*, 4601.
- (32) Etter, M. C. *Acc. Chem. Res.* **1990**, *23*, 120.
- (33) Richardson, T. B.; de Gala, S.; Crabtree, R. H. *J. Am. Chem. Soc.* **1995**, *117*, 12 875.
- (34) Vedani, A.; Dunitz, J. D. *J. Am. Chem. Soc.* **1985**, *107*, 7653.
- (35) Murray-Rust, P.; Glusker, J. P. *J. Am. Chem. Soc.* **1984**, *106*, 1018.
- (36) Braga, D. *J. Am. Chem. Soc.* **1995**, *117*, 3156.
- (37) Gorbitz, C. H.; Etter, M. C. *J. Am. Chem. Soc.* **1992**, *114*, 627.
- (38) Gorbitz, C. H. *Acta Crystallogr.* **1989**, *B45*, 390.
- (39) Taylor, R.; Kennard, O. *Acc. Chem. Res.* **1984**, *17*, 320.
- (40) Allen, F. H.; Baalham C. A.; Lommerse J. P. M.; Raithby P. R.; Sparr, E. *Acta Crystallogr.* **1997**, *B53*, 1017.
- (41) Bruno I. J.; Cole J. C.; Lommerse, J. P. M.; Rowland, R. S.; Taylor, R.; Verdonk, M. L. *J. Comput.-Aided Mol. Design* **1997**, *11*, 525.
- (42) Miyazawa, S.; Jernigan, R. L. *Macromolecules* **1985**, *18*, 534.
- (43) Jernigan, R. L.; Bahar, I. *Curr. Opin. Struct. Biol.* **1996**, *6*, 195, and references therein.
- (44) Wallqvist, A.; Jernigan, R. L.; Covell, D. G. *Protein Science* **1995**, *4*, 1881.
- (45) DeWitte, R. S.; Shakhnovich, E. I. *J. Am. Chem. Soc.* **1996**, *118*, 11 733.
- (46) DeWitte, R. S.; Ishchenko, A. V.; Shakhnovich, E. I. *J. Am. Chem. Soc.* **1997**, *119*, 4608.
- (47) Karle, I. L. *J. Mol. Struct.* **1999**, *474*, 103.
- (48) Alkorta, I.; Rozas, I.; Elguero, J.; Foces-Foces, C.; Cano, F. H. *J. Mol. Struct.* **1996**, *382*, 205.
- (49) Girolami, G. S. *J. Chem. Educ.* **1994**, *71*, 962.
- (50) Intermolecular energy is defined as the energy of interaction of a molecule in a crystal with all other molecules.
- (51) Wiberg, K. B.; Marquez, M.; Castejon, H. *J. Org. Chem.* **1994**, *59*, 6817.
- (52) Wiberg, K. B.; Breneman, C. M. *J. Am. Chem. Soc.* **1992**, *114*, 831.
- (53) Wiberg, K. B. *J. Chem. Educ.* **1996**, *73*, 1089.
- (54) Wong, M. W.; Wiberg, K. B. *J. Phys. Chem.* **1992**, *96*, 668.
- (55) Allen, F. H.; Harris, S. E.; Taylor, R. *J. Comput.-Aided Mol. Design* **1996**, *10*, 247. It was shown that crystal packing effects do not have a strong and systematic effect on distribution of torsions in molecular crystals.