Journal of Computer-Aided Molecular Design, 11 (1997) 479–490.
KLUWER/ESCOM

J-CAMD 411

The orientation of N-H···O=C and N-H···N hydrogen bonds in biological systems: How good is a point charge as a model for a hydrogen bonding atom?

Robert P. Apaya*, Maria Bondí and Sarah L. Price**

Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, U.K.

Received 12 March 1997
Accepted 19 May 1997

Keywords: Hydrogen bonds; Distributed multipoles; Electrostatic potential extrema

Summary

In order to design new ligands for protein-binding sites of unknown structure, it would be useful to predict the likely sites of hydrogen bonding of an unknown protein fragment to a known molecule. The positions of maxima and minima in the electrostatic potential at appropriate distances from the van der Waals surface were calculated for various small molecules, nucleic acid bases, peptide units and amino acid side chains containing groups which can form the biologically important N-H···O=C and N-H···N hydrogen bonds. Their ability to predict the positions of H and O/N in hydrogen bonded complexes, as predicted by optimising the electrostatic interactions of pairs of such molecules constrained by the molecular shapes, was assessed. It is shown that extrema in the electrostatic potential around the isolated molecules give worthwhile predictions for the locations of hydrogen bonding partners. For molecules bound by a single N-H···O=C hydrogen bond, the electrostatic maximum associated with the H is usually less than 1 Å from an acceptor atom, while a C=O electrostatic minimum is generally less than 1.5 Å from the hydrogen bond proton. However, a significant number of hydrogen bonds form to the opposite lone pair from the electrostatic minimum, in which case the separation is up to 3.3 Å. This reflects the broad electrostatic potential well around a carbonyl oxygen between the lone pair directions. The model predicts when neighbouring atoms drastically change the hydrogen bonding characteristics of an N-H or C=O group. Although the geometries of hydrogen bonded complexes are influenced by the other van der Waals contacts between the molecules, particularly multiple hydrogen bonds, these influences are constant when considering hydrogen bonding to a specific uncharacterised binding site. Hence, the consideration of sterically accessible electrostatic extrema will be useful in the design of new ligands.

Introduction

The question of where the hydrogen bond donor of an unknown molecule is likely to be relative to the acceptor of a known molecule, and vice versa, is important in our understanding of biochemical interactions. It is practically important for the design of new molecules which will have similar binding properties to a protein or DNA site of unknown structure as the substrate or a lead compound. The receptor site may be thought of as an unknown three-dimensional (3D) array of hydrogen bond donor and acceptor sites, and a suitable ligand will consist of a complementary set of acceptor and donor sites [1], which satisfy the distance and directional requirements to form strong hydrogen bonds. The modeller wishes to know whether two ligands are sufficiently similar in the 3D disposition of their hydrogen bonding sites so that both can complement the receptor and bind strongly. This depends on the distance and directional requirements of the specific hydrogen bonds that can be made by the functional groups. The N-H···O=C hydrogen bond is the most commonly found in proteins and complexes with, or

*Present address: Drug Design Group, Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, U.K.
**To whom correspondence should be addressed.

between, nucleic acid bases, so this is the main focus of the work, though the N-H···N hydrogen bonds which arise are also considered.

The main source of information about the preferred geometry of hydrogen bonds comes from the many surveys of organic and protein crystal structures [2,3]. Statistical surveys of the orientational preferences for N-H···O=C hydrogen bonds in small molecule crystal structures show a preference for hydrogen bonding in the lone pair direction ($\angle CO{\cdot}H \approx 120°$), but with a large scatter of orientations [4]. More specific surveys of this bond within protein main chain and side chains [5], or for specific side chains such as arginine, aspartate, asparagine and histidine [6], lead to the observation that clearer, and sometimes different, results can be obtained when the functional groups are defined more specifically, by a larger fragment or even conformation. This is expected, since it gives smaller variations in the electronic properties of the donor and acceptor, and in the steric requirements of molecular fragments, including the possibility of C=O accepting two hydrogen bonds. The question arises as to whether the geometry of possible hydrogen bonds can be predicted more accurately when the detailed electron distribution of only one of the molecular fragments is known. Do the nature and relative positions of other heteroatoms in a molecule have sufficient effect on the hydrogen bond geometry that this can be used to predict the hydrogen bond geometry more accurately than the range found in general crystallographic surveys? Do the different hydrogen bond donor and acceptor groups in nucleic acid bases affect the hydrogen bonding properties of each other in a predictable way? If this also occurs in proteins, can the variation in likely hydrogen bonding geometries with conformation be predicted?

There is a considerable body of work linking the electrostatic potential around hydrogen bond donors and acceptors to their hydrogen bonding ability [7]. For example, Politzer and co-workers [8] have shown that the electrostatic potential of a gas phase molecule can be quantitatively related to its ability in solution to donate or accept protons in solvent-to-solute hydrogen bonds for a series of nitrogen atoms in primary amines and azines, and oxygen atoms within alkyl ethers and molecules containing carbonyl oxygens [9], and for a series of alcohols, ethers and molecules in which alkyl groups were the proton donors [10]. Recently, Kenny [11] has shown correlations for a range of heterocycles between the hydrogen bond basicities and both electrostatic potential minima and the gradient of the molecular electrostatic potential (i.e. the electrostatic field) further out from the molecule.

These observations are consistent with Pauling's [12] early view of the hydrogen bond as electrostatic in nature. If the unknown hydrogen bond donor could be represented by just a partial positive charge on the H atom, $H^{\delta+}$, then the hydrogen bond would form with the proton at the minimum in the electrostatic potential at the sterically constrained distance of approximately 1.9 Å from the oxygen of the known molecule. Similarly, a partial negative charge model for the unknown hydrogen bond acceptor, $O^{\delta-}$, would result in the oxygen atom being at the site of the potential maximum approximately 2.9 Å from the nitrogen donor of the known molecule. Thus, the maxima and minima in the potential, constrained to these distances, give a simple model for predicting the likely hydrogen bonding sites around a molecule, which can be tested against the geometries of the hydrogen bonded complexes of the molecule with various common biological molecular fragments.

Such a detailed testing of the electrostatic model for predicting hydrogen bonding geometries for biological hydrogen bonds is hampered by the lack of structures of van der Waals complexes of suitable model molecules. However, there is considerable evidence that the actual hydrogen bonded structure of a gas phase complex of two molecules can generally be predicted by optimising their electrostatic interaction energy within accessible orientations, provided the electrostatic forces were calculated from a distributed multipole analysis [13,14] (DMA) of a reasonable quality ab initio charge density. This model has successfully been used to predict structures of a wide range of van der Waals dimers of small polyatomics [15,16], and to rationalise the observed preferred orientations of N-H···O=C hydrogen bonds [17], phenylalanine, carboxylate and arginine side chain interactions [18], N-H aromatic hydrogen bonds [19] and phenylalanine hydration [20] in proteins. It also predicts the many hydrogen bonded dimers of the nucleic acid bases [21], in reasonable agreement with self-consistent field (SCF) supermolecule calculations [22]. It is expected to be a good guide to the likely hydrogen bonding sites for the molecular complexes, because the electrostatic interaction generally dominates the orientation dependence of the intermolecular potential [23].

In this paper, we establish the positions of the minima and maxima in the electrostatic potential subject to the steric constraints applicable to a hydrogen bonding proton or oxygen acceptor, respectively, for a variety of model peptide units, nucleic acid bases and protein side chain models. This provides an indication of how much effect the rest of the molecule has on the electrostatic potential around N-H donors and C=O acceptors. To determine whether this variation is reflected in the geometries of the hydrogen bonds formed to the molecule, the positions of the extrema are compared with the positions of the proton donors and acceptors in several complexes found by optimising the electrostatic interaction between the molecule and other biochemical model molecules. Although this strictly studies the correlation between the electrostatic properties of an isolated molecule and its electrostatic interactions in complexes subject to steric

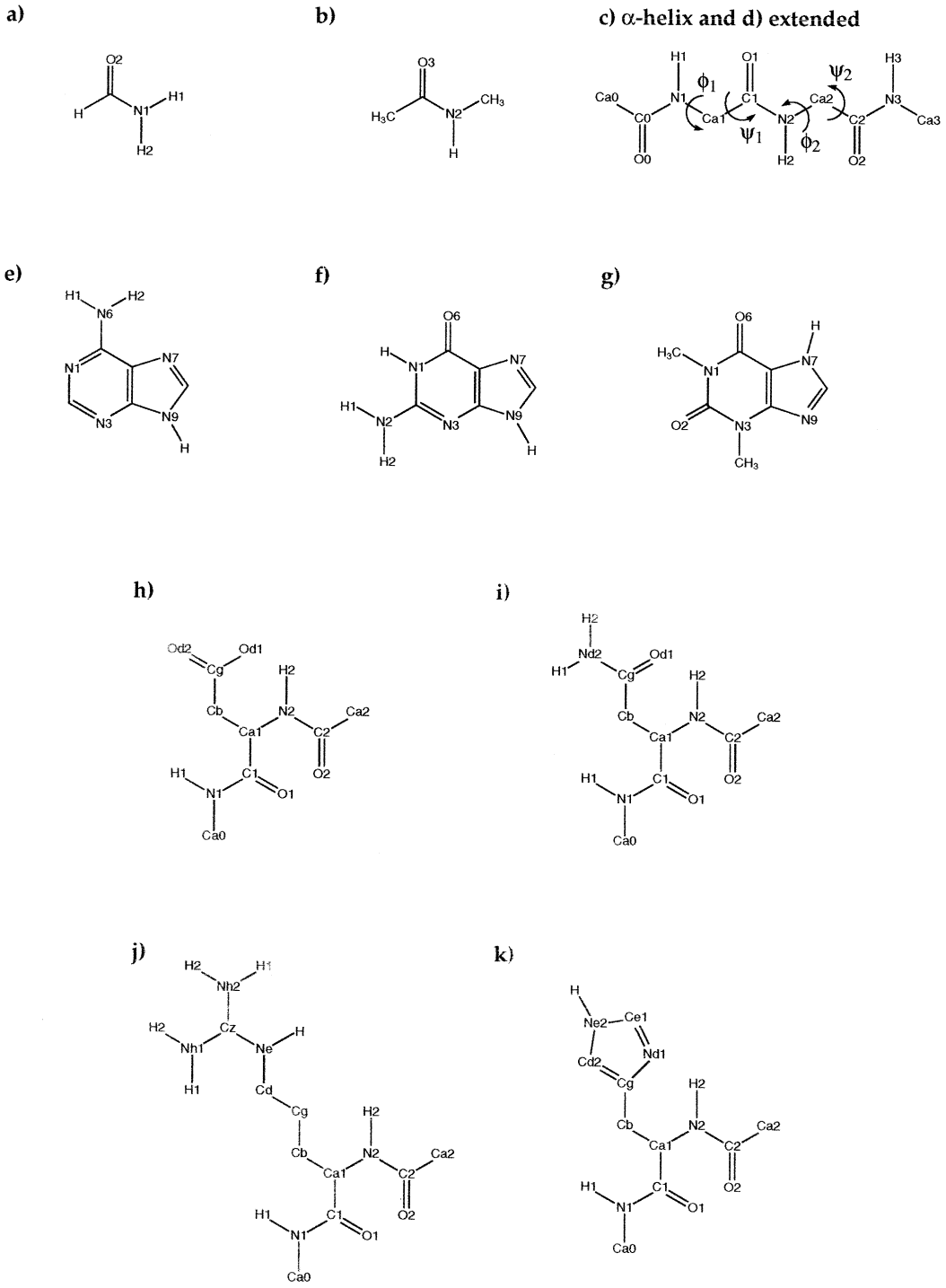

Fig. 1. Molecules used in this study, defining the numbering system for (a) formamide, (b) methylacetamide, *N*-acetyl, *N*′-methylamide blocked diglycine in (c) an α-helical conformation ($\phi_1 = \phi_2 = -57°$, $\psi_1 = \psi_2 = -47°$) and (d) an extended conformation ($\phi_1 = \phi_2 = 180°$, $\psi_1 = \psi_2 = 180°$), (e) adenine, (f) guanine, (e) theophylline, and *N*-acetyl, *N*′-methylamide blocked residues (with $\phi = -57°$, $\psi = -47°$) for (h) aspartate ($\chi_1 = -68°$, $\chi_2 = -26°$), (i) arginine ($\chi_1 = 307°$, $\chi_2 = 183°$, $\chi_3 = 189°$, $\chi_4 = 175°$), (j) asparagine ($\chi_1 = -68°$, $\chi_2 = -37°$) and (k) histidine ($\phi = -71.6°$, $\psi = 155.6°$, $\chi_1 = -58.5°$, $\chi_2 = -55.5°$).

constraints, the latter is expected to give a reasonable prediction of the hydrogen bond geometries.

This study of how closely the electrostatic potential outside hydrogen bonding groups determines the likely positions of hydrogen bond donors is directly relevant to a recently proposed method of determining the relative binding orientation of a set of structurally diverse ligands for a binding site of unknown 3D structure. It has been shown for a structurally diverse range of phosphodiesterase III inhibitors [24], agonists and antagonists for the

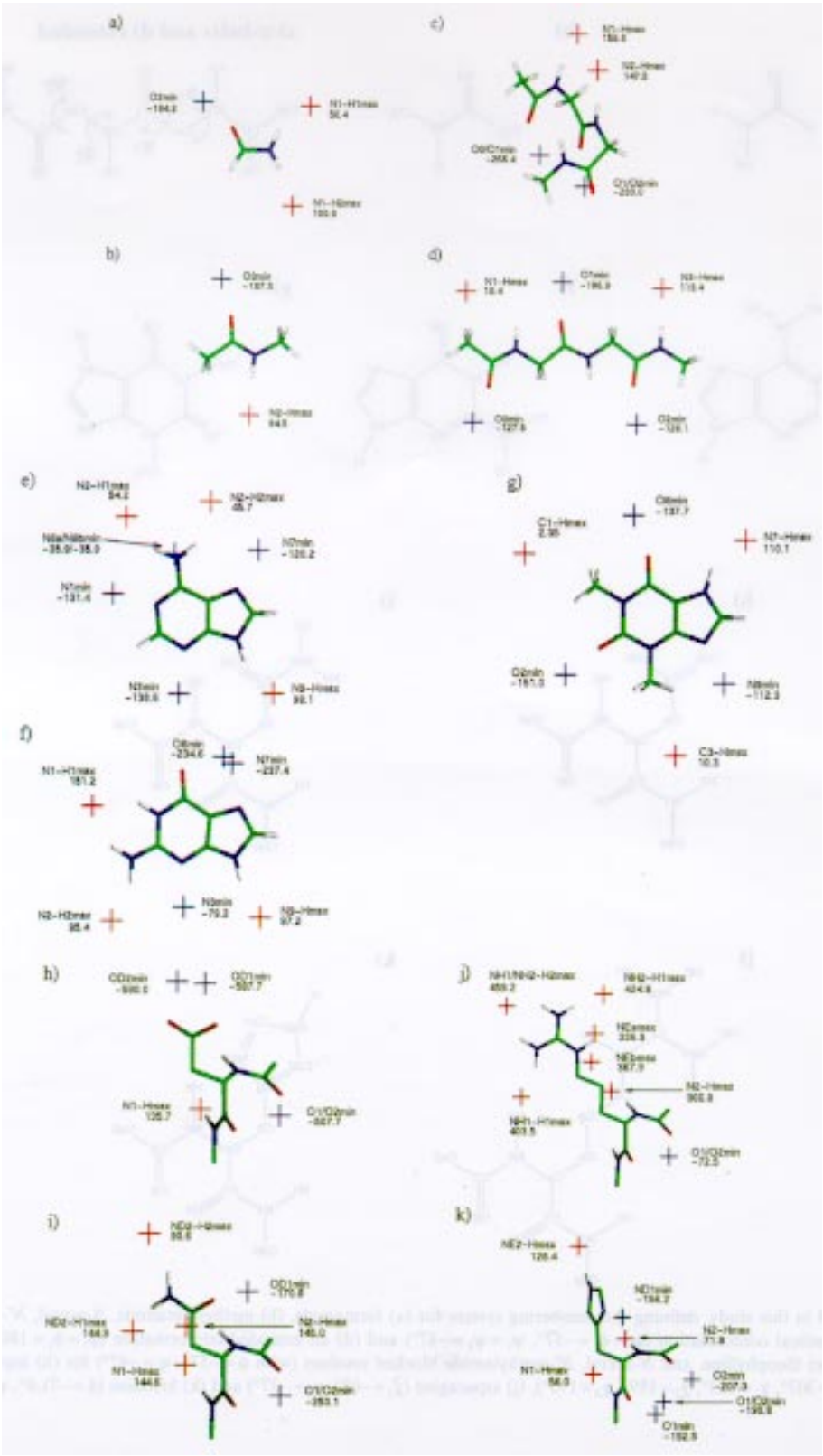

Fig. 2. Electrostatic potential maxima (minima) calculated by interaction with a single negative (positive) point charge of radius 1.4 Å (0.5 Å) for (a) formamide, (b) methylacetamide, *N*-acetyl, *N′*-methylacetamide blocked diglycine in (c) an α-helical conformation and (d) an extended conformation, (e) adenine, (f) guanine, (g) theophylline, and *N*-acetyl, *N′*-methylacetamide blocked residues for (h) aspartate, (i) asparagine, (j) arginine and (k) histidine. The maxima and minima are in the plane except for ADMα: O0, O1 and N1-H; ADMβ: N1-H; Asp: NH1-H1; Arg: N2-H2; His: N2-H2; N1-H1; Asn: N2-H2. Negligible deviations from the plane are found for Asp: OD1 and OD2; Asn: ND-H1, ND-H2; and Asn: N1-H1. Key: electrostatic potential maxima, red +; electrostatic potential minima, blue +. Values in kJ mol^{-1}.

adenosine A_1 receptor [25], and a range of lactam compounds [26] that optimising the overlay of three or four corresponding electrostatic extrema between the ligand and the natural substrate produces a sterically and electrostatically plausible relative binding orientation. The method ensures that there is good matching of at least the sign of the electrostatic potential in the regions of space where both molecules could interact strongly with the binding site. However, since most of the electrostatic extrema are associated with hydrogen bonding groups, and the separations of the overlaid extrema are often remarkably small (rms errors of the order of 1 Å), the method may well be identifying which groups are hydrogen bonded to the unknown binding site. The current study aims to establish how close two overlaid extrema (and thereby two hydrogen bonding functional groups) have to be in a relative binding orientation for it to be likely that both could form strong hydrogen bonds to a rigid binding site.

Methods

A representative but arbitrary range of model molecules for biologically important fragments (Fig. 1) were studied: formamide, methylacetamide, adenine, guanine and theophylline, four amino acid side chains (arginine, aspartate, asparagine and histidine), and two conformations of a dipeptide chain fragment. The peptide residue models were blocked using *N*-acetyl and *N*′-methylamide terminal groups to approximate the charge distribution of the peptide backbone [27], using an α-helical main chain conformation ($\phi=-57°$, $\psi=-47°$). The *N*-acetyl, *N*′-methylamide blocked diglycine peptide was considered in two different conformations: a helical conformation ($\phi_1=\phi_2=-57°$, $\psi_1=\psi_2=-47°$, labelled ADMα) and an extended conformation ($\phi_1=\phi_2=180°$, $\psi_1=\psi_2=180°$, labelled ADMβ). The molecular structures were built using INSIGHT II [28] and optimised using the default CVFF force field [29], with the torsion angles of the peptide residues fixed at the typical values given in Fig. 1.

SCF wave functions were obtained for each molecule, using the CADPAC [30] suite of ab initio programs, using a 3-21G [31] basis set in each case. The DMA [13,14] of each wave function was calculated to represent the charge density by a charge, dipole, quadrupole, octupole and hexadecapole moment at every nuclear position.

Using the program ORIENT3 [32], the positions and strengths of the sterically accessible maxima (minima) in the electrostatic potential energy were determined by minimising the interaction energy of a single negative (positive) point charge of radius 1.4 Å (0.5 Å) with the hard-sphere molecule under examination. This locates the extrema in the electrostatic potential in the region outside the molecular van der Waals surface which could be occupied by hydrogen bond acceptor and donor groups, respectively. Pseudo-hard-sphere repulsion was used between all sites with nonzero van der Waals radii. The molecular van der Waals volume was defined using the Pauling radii, i.e. C, 2.0 Å; N, 1.5 Å; O, 1.4 Å. Polar (N-H) hydrogens were treated as having a zero van der Waals radius, as hydrogen bond separations are typically the sum of the donor and acceptor atom radii [33]. Nonpolar hydrogens, such as methyl hydrogens, are included in a 'united atom' carbon radius.

We tested for probable hydrogen bonded structures by considering four categories of complexes comprised of the molecules in Fig. 1: three pairs of small amides (formamide dimer, methylacetamide dimer, and formamide with

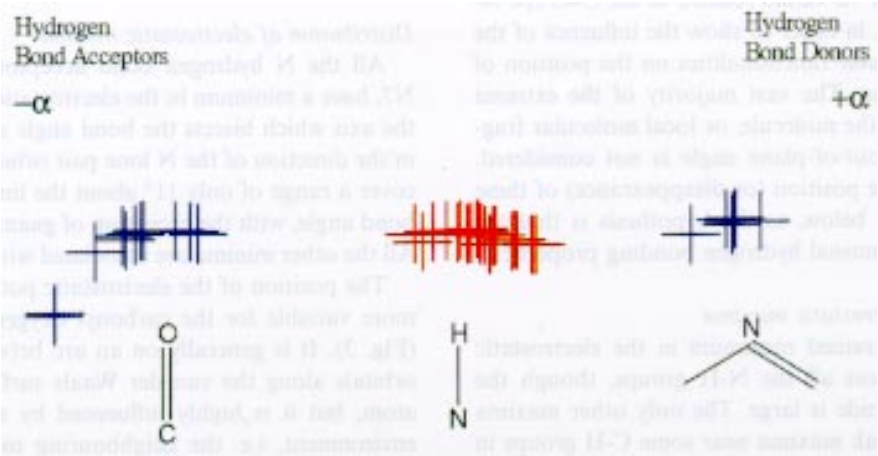

Fig. 3. The distribution of oxygen C=O, and sp^2 nitrogen electrostatic potential minima (at 0.5 Å from the molecular van der Waals surface) and sp^2 nitrogen N-H electrostatic maxima (at 1.4 Å from the molecular van der Waals surface) as a function of α. The electrostatic potential minima (maxima) are shown transformed onto a common orientation relative to a reference C=O, N or N-H group. The sign of the angle α, which defines the position of the electrostatic minima ($\alpha=\angle$CO(Min) or $\angle$N(Min)) or maxima ($\alpha=\angle$NH(Max)), is defined as negative in the direction of neighbouring hydrogen bond acceptors and positive for hydrogen bond donors within the molecule. This angle is still used in the few cases where the extremum is not in the plane of the molecular fragment.

methylacetamide); three pairs of biologically important heterocycles (adenine with theophylline, adenine with guanine, and guanine with theophylline); the 10 combinations of *N*-acetyl and *N′*-methylamide blocked diglycine in both the α-helical and extended conformations each with formamide, methylacetamide, adenine, guanine and theophylline; and the 12 combinations of an amino acid side chain (arginine/asparagine/aspartate/histidine) with a heterocycle (adenine/guanine/theophylline). The minimum energy structures for pairs of molecules were found by optimising their electrostatic interaction energy, calculated from the distributed multipoles, using the pseudo-hard-sphere repulsion to define sterically accessible conformations, within ORIENT3. Minima in the electrostatic interaction energy between the molecules were located from dozens of varied starting positions, chosen on a sphere surrounding the fixed molecule, sufficient to be reasonably certain that all the significant minima had been located. A few minima which were considered doubtful on the convergence criteria, and three very weakly bound minima where a C-H···N/O contact was the dominant interaction, were excluded from the results.

Results

The sterically accessible electrostatic maxima and minima obtained for each molecule are shown in Fig. 2. For comparison, the electrostatic minima (maxima) were transformed onto a common orientation, and are shown relative to a 'reference' C=O (or N-H) group in Fig. 3, with the angle α being the deviation of the extremum from collinearity with the C=O or N-H bond or the bisector of the angle at the N donor. The C=O (N, N-H) groups were overlaid so that neighbouring hydrogen bond acceptor groups are at −α values, and hydrogen bond donor groups are at +α values relative to the C=O (N, N-H) reference group, in order to show the influence of the surrounding molecular functionalities on the position of electrostatic extrema. The vast majority of the extrema are in the plane of the molecule, or local molecular fragment, and so the out-of-plane angle is not considered. The variation in the position (or disappearance) of these extrema is detailed below, as the hypothesis is that the outliers will have unusual hydrogen bonding properties.

Distribution of electrostatic maxima

There is a constrained maximum in the electrostatic potential near almost all the N-H groups, though the variation in magnitude is large. The only other maxima located are very weak maxima near some C-H groups in theophylline. The electrostatic maxima are normally found along the axis of the N-H donor bond, within a narrow range (45°) for the angle at the nitrogen (Fig. 3). An example of a deviation from linearity is theophylline N7-H_{max} (Fig. 2g), which is significantly displaced by 29° from the N-H bond axis due to the presence of the neighbouring C6=O6 group. The extreme deviations are guanine N2-$H1_{max}$, which is between N1-H1 and N2-H (Fig. 2f), and arginine NH1/NH2-H_{max}, which is between NH1-H2 and NH2-H2 (Fig. 2j). These are only two examples among the molecules studied here of N-H donor groups separated by two intermediate bonds, so that the N-H groups are parallel and coplanar.

In contrast, no electrostatic maxima were found for N3-H in ADMα (Fig. 2c) and for N2-H in the middle of the extended ADMβ (Fig. 2d). The ADMβ donors N1-H and N2-H are both parallel and coplanar with neighbouring C=O groups, separated by two intermediate bonds. This has led to a cancellation of potentials of opposite magnitude. The electrostatic potential across the region outside the van der Waals surface of these two donor groups is relatively featureless and only changes by a small magnitude. Hence, only a shallow displaced maximum close to the neighbouring methyl group was found for ADMβ N1-H, and no maximum was located for N2-H. The lack of a maximum in positive potential near the ADMα N3-H is a consequence of the α-helical conformation which presents two distinct hydrogen bonding aspects to a potential second molecule, characterised by N-H donor groups along the 'top' edge of the helical backbone and C=O acceptors along the 'bottom'. As a result of the helical turn, N3-H and C0=O0 are brought into close proximity with each other, producing a cancellation of potentials. The sensitivity of the electrostatic potential around the hydrogen bonding groups with conformation is not purely a geometric effect, as changes in the atomic charge distribution also contribute [34]. There is also no maximum associated with the N-H group in aspartate (Fig. 2h) because of the proximity of the charged CO_2^- group.

Distribution of electrostatic minima

All the N hydrogen bond acceptors, except guanine N7, have a minimum in the electrostatic potential close to the axis which bisects the bond angle at the sp^2 nitrogen, in the direction of the N lone pair orbital. The N minima cover a range of only 11° about the line bisecting the sp^2 bond angle, with the exception of guanine N7 at $\alpha = -32°$. All the other minima are associated with the C=O groups.

The position of the electrostatic potential minimum is more variable for the carbonyl oxygen acceptor groups (Fig. 3). It is generally on an arc between the lone pair orbitals along the van der Waals surface of the oxygen atom, but it is highly influenced by the local chemical environment, i.e. the neighbouring molecular functionalities. This is particularly apparent in the case of guanine C6=O6, where the minimum is almost perpendicular to the C=O axis. Although a distinct minimum was located, its position is closely related to the nearby N7 minimum. The proximity of these two hydrogen bond acceptors to

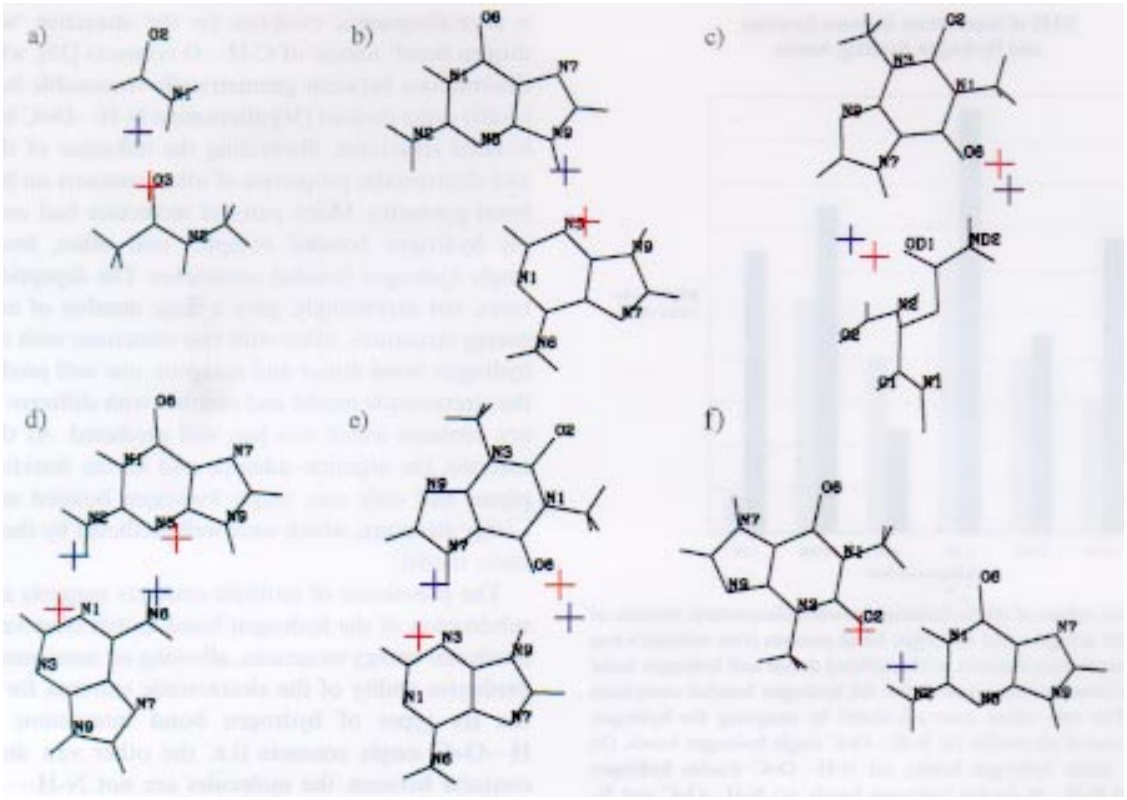

Fig. 4. Examples of complexes containing the six hydrogen bond motifs illustrated for (a) formamide–methylacetamide (N-H···O=C single hydrogen bond), (b) guanine–adenine (N-H···N single hydrogen bond), (c) asparagine–theophylline (N-H···O=C double hydrogen bonds), (d) guanine–adenine (N-H···N double hydrogen bonds), (e) adenine–theophylline (N-H···O=C, N-H···N double hydrogen bonds) and (f) guanine–theophylline (bifurcated hydrogen bond). For each complex, the relevant electrostatic maxima and minima of the isolated molecules (from Fig. 2) are overlaid. Key: electrostatic potential maxima, red +; electrostatic potential minima, blue +.

each other creates a broad region of negative electrostatic potential between them. If the positions of the 'shared' minima (guanine $O6_{min}$, aspartate $OD1_{min}$ and $OD2_{min}$) are excluded, the C=O minimum positions, the positions of minima relative to carbonyl oxygen acceptors, are clustered well within the arc defined by the lone pair orbitals, with a maximum deviation from linearity with the carbonyl C=O bond of 42° in the direction of the hydrogen bond acceptors in the same molecule, and only 18° in the direction of hydrogen bond donors.

Analysis of hydrogen bonding motifs to determine predictive ability of the electrostatic extrema

The 28 combinations of model molecules studied produced a total of 116 minimum energy structures containing N-H···O=C or N-H···N hydrogen bonds. The extent to which the sterically accessible electrostatic extrema of the isolated molecules predict the structure of the bimolecular complex can be quantified by superimposing the electrostatic extrema of the isolated molecules (from Fig. 2) onto the minimum energy structure of the complex, as shown for some examples of different types of hydrogen bonded minima in Fig. 4. The separation between the proton of one molecule and the nearest electrostatic minimum of the other is denoted by *a* and the separation between the hydrogen bond acceptor of one molecule and the electrostatic maximum of the other is denoted by *b* (see Fig. 5). The hydrogen bonding interactions involved

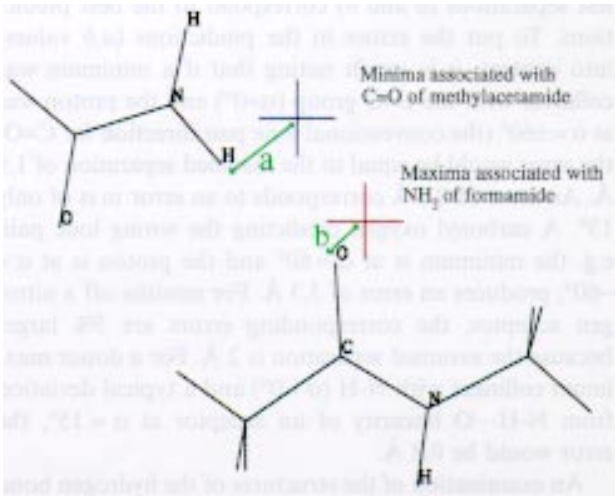

Fig. 5. Correspondence between the positions of electrostatic extrema of the isolated molecules, and the hydrogen bond geometry of the minimum energy complexes, illustrated for formamide and methylacetamide. The quality of the prediction is considered in terms of *a* and *b*. Key: $N1\text{-}H1_{Form}$ electrostatic maximum, red +; $O3_{Meth}$ electrostatic minimum, blue +.

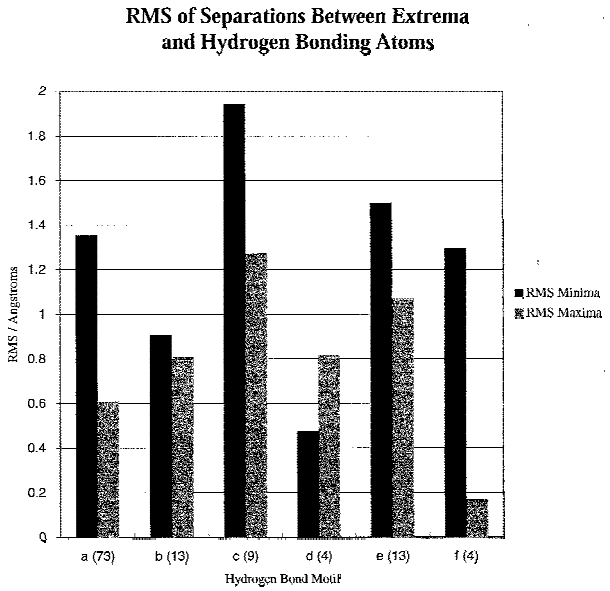

Fig. 6. Rms values of the separations between electrostatic minima of the isolated acceptor and hydrogen bond protons (rms minima = rms *a*), and electrostatic maxima of the isolated donor and hydrogen bond acceptors (rms maxima = rms *b*) for the hydrogen bonded complexes studied. The rms values were calculated by assigning the hydrogen bonds to one of six motifs: (a) N-H···O=C single hydrogen bonds, (b) N-H···N single hydrogen bonds, (c) N-H···O=C double hydrogen bonds, (d) N-H···N double hydrogen bonds, (e) N-H···O=C and N-H···N double hydrogen bonds and (f) bifurcated hydrogen bonds. The number of examples of each motif is given in parentheses.

in each of the minimum energy structures, along with the corresponding values of *a* and *b*, are given in supplementary data.

Since we propose that locating electrostatic maxima and minima identifies the positions of hydrogen bond acceptors and donors in a bimolecular complex, the smallest separations (*a* and *b*) correspond to the best predictions. To put the errors in the predictions (*a*,*b* values) into context, it is worth noting that if a minimum was collinear with the C=O group (α=0°) and the proton was at $\alpha = \pm 60°$ (the conventional lone pair direction for C=O) the error would be equal to the assumed separation of 1.9 Å. An error of 0.5 Å corresponds to an error in α of only 15°. A carbonyl oxygen predicting the wrong lone pair, e.g. the minimum is at $\alpha = 60°$ and the proton is at $\alpha = -60°$, produces an error of 3.3 Å. For minima off a nitrogen acceptor, the corresponding errors are 5% larger because the assumed separation is 2 Å. For a donor maximum collinear with N-H ($\alpha = 0°$) and a typical deviation from N-H···O linearity of an acceptor at $\alpha = 15°$, the error would be 0.8 Å.

An examination of the structures of the hydrogen bond complexes revealed that most of the minimum energy structures, for all but the smallest molecules, had at least two van der Waals contacts between the molecules. The minimum energy structures containing a single hydrogen bond often had an additional C-H···O=C contact. There is crystallographic evidence for the attractive 'weak hydrogen bond' nature of C-H···O contacts [35], which can discriminate between geometrically reasonable but structurally quite distinct [36] alternative N-H···O=C hydrogen bonded structures, illustrating the influence of the steric and electrostatic properties of other contacts on hydrogen bond geometry. Many pairs of molecules had one multiply hydrogen bonded complex and other, less stable, singly hydrogen bonded complexes. The dipeptide structures, not surprisingly, gave a large number of minimum energy structures, often with two structures with the same hydrogen bond donor and acceptor, one well predicted by the electrostatic model and another with different secondary contacts which was less well predicted. At the other extreme, the arginine–adenine and all the histidine complexes had only one, singly hydrogen bonded minimum energy structure, which were well predicted by the electrostatic model.

The prevalence of multiple contacts suggests a natural subdivision of the hydrogen bond motifs observed in the minimum energy structures, allowing an assessment of the predictive ability of the electrostatic extrema for each of the six types of hydrogen bond interaction: (a) N-H···O=C single contacts (i.e. the other van der Waals contacts between the molecules are not N-H··· O=C or N-H···N); (b) N-H···N single contacts; (c) N-H···O=C double contacts; (d) N-H···N double contacts; (e) N-H···O=C, N-H···N double contacts; and (f) bifurcated hydrogen bonds. An example of each type of hydrogen bonding motif, showing the appropriate electrostatic maxima and minima, is given in Fig. 4. The predictive ability of the electrostatic extrema for each type of hydrogen bond can be assessed by rms values (Fig. 6) for the separation between electrostatic maxima and hydrogen bond acceptor atoms, and electrostatic minima and hydrogen bond protons (*a* and *b*, as defined in Fig. 5), as a rough summary of the geometries found. There is considerable variation in the predictive ability of the electrostatic maxima and minima of the isolated molecules to predict the hydrogen bond geometry, ranging from good to fair, according to this classification, so each are discussed below.

Single hydrogen bonds

The N-H···N single hydrogen bond structures (Fig. 6b) are well predicted by the electrostatic extrema of the isolated molecule, as is the position of the oxygen in the single N-H···O=C contacts. However, the positions of the protons in the single N-H···O=C contacts (Fig. 6a), which are the most frequently observed hydrogen bonds in this study, appear to be predicted considerably less reliably by the electrostatic model. The reason is apparent from the histograms (Fig. 7) of the distances between the electrostatic extrema and the actual positions of the hydrogen bonding H or O/N atoms. The most common separation

of the acceptor O from the maximum (Fig. 7a) is very small (0.1–0.2 Å), but any separation up to 1 Å is quite common, with only two markedly larger errors, giving good overall predictions. However, the hydrogen bonded proton, although usually within 1.5 Å of the minimum (Fig. 7b), is rarely very close, and there are a significant

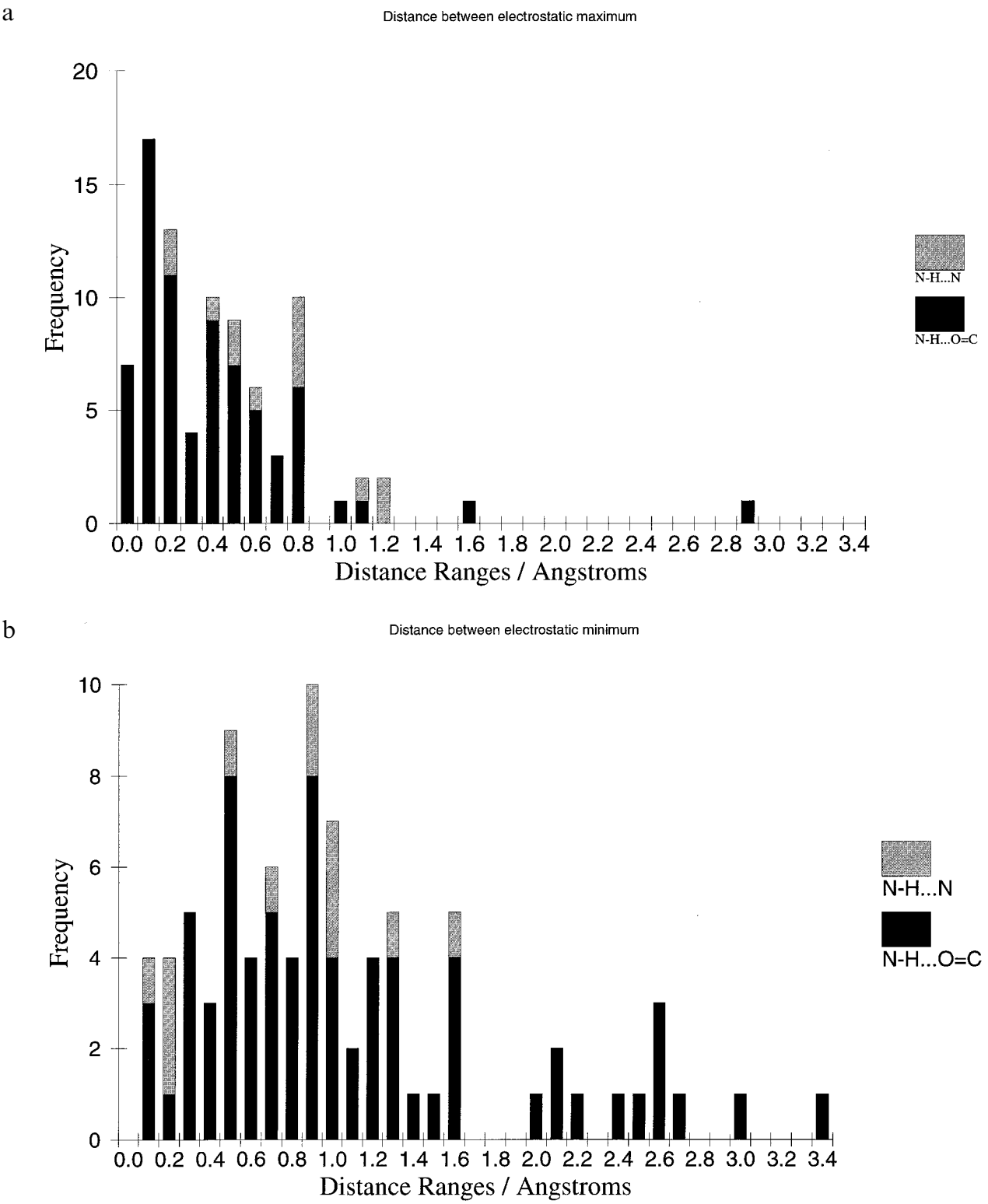

Fig. 7. Histograms of the number of singly N-H···O=C hydrogen bonded complexes in Tables 1–5 of the supplementary material whose hydrogen bond geometry is estimated to a given accuracy by the electrostatic extrema of the isolated molecules. (a) Distance of the O atom of the acceptor from the electrostatic maximum of the N-H. (b) Distance of the H atom of the donor from the electrostatic minimum of the O=C.

number of separations over 2.5 Å corresponding to the proton being located at the opposite lone pair to the minimum. Thus, the large rms error in the minima predictions (Fig. 6a) is because the minima are only being moderately successful at predicting to which side of their associated oxygen atoms the hydrogen bonds form.

Multiple contacts

The opportunity of forming a second hydrogen bond dominates the hydrogen bond geometry of a complex [17], and so the success of the electrostatic extrema model in predicting the component hydrogen bond geometries varies according to whether it gives the same or conflicting geometries. The N-H···N hydrogen bonds generally have nearly collinear electrostatic extrema with only slight deviations caused by the neighbouring functional groups. This matches the geometric requirements of the doubly hydrogen bonded structures (e.g. Fig. 4d), which are therefore generally compatible with the position of the electrostatic extrema. Thus, the minima predictions for doubly N-H···N hydrogen bonded structures (Fig. 6d) are better than for singly hydrogen bonded structures (Fig. 6b). In contrast, the double N-H···O=C hydrogen bonds are poorly predicted because the electrostatic extrema tend to be displaced to outside the amide group whilst the double hydrogen bonds form inside (see Fig. 4c). For example, in order to achieve the most stable antiparallel pair of hydrogen bonds in the formamide dimer, the protons must be located on the opposite side of the acceptor atoms to the potential minima, which are located towards the lone pair furthest from the N-H groups. This results in a large separation of 2.15 Å between the minima and the hydrogen bonding protons, and 1.15 Å between the maxima and the oxygen acceptors.

The most commonly observed multiple contacts, N-H···O=C and N-H···N, as found in many nucleic acid complexes, generally have the minima of the N much closer to the protons than the minima near O, and so the predictions (Fig. 6e) are intermediate between those of the other doubly hydrogen bonded structures (Figs. 6c and d). It should be noted that the features of the molecular structure which allow the formation of a pair of antiparallel hydrogen bonds (e.g. amide group) also produce extrema which deviate away from the hydrogen bonding region, with this deviation being most pronounced for the oxygen acceptors.

Bifurcated hydrogen bonds

A success, and consequent limitation, of the electrostatic extrema model is very clearly demonstrated by the four examples of bifurcated hydrogen bonds (Fig. 6f), in which an acceptor C=O group is shared between two N-H donors. For these parallel, coplanar N-H donors, the electrostatic maximum is located between the donors, and consequently gives an excellent prediction of the oxygen acceptor atom position (Fig. 6f, maxima). However, there is only one electrostatic minimum related to the oxygen acceptor. Since this minimum must be 'shared' between the two protons, it does not lie particularly close to either of the hydrogen donor atoms, and the predictions (Fig. 6f, minima) are necessarily very poor. The shared maxima also give poor predictions of the other complexes of guanine and arginine, in which only one of the two distinct hydrogen bonds forms to one of the parallel N-H groups. Hence, the displacement of the maximum from collinearity with the N-H group caused by parallel N-H groups predicts the possibility of a bifurcated hydrogen bond, but gives poor predictions for other structures.

No minima results

Another important success of the electrostatic model, which is not reflected in the above analysis, is its ability to predict when hydrogen bonds are unlikely to form. It was noted earlier that, as a result of the conformation of the *N*-acetyl, *N*′-methylamide blocked diglycine peptide, there is no potential maximum associated with N3-H in the α-helical conformation or with N2-H in the extended conformation. The only hydrogen bonds found involving these protons were either a very weak bond, for ADMα-formamide, or a second hydrogen bond in a double hydrogen bonded motif which had an energy only marginally stronger than a single hydrogen bond (ADMβ-formamide). Thus, the absence of a maximum in the positive electrostatic potential around an N-H group does appear to predict poor hydrogen bonding properties.

Discussion

Figure 6 shows that most of the complexes are being predicted better than by the usual ideas of ranges of hydrogen bonding directionality, though there are exceptions as previously highlighted. From Fig. 6, it can be seen that the positions of hydrogen bond acceptor atoms are generally predicted reasonably well by the electrostatic maxima, though this deteriorates for double bonds involving at least one N-H···O=C. The minima give much poorer predictions of the position of the proton hydrogen bonded to C=O, as would be expected from the common occurrence of hydrogen bonded structures with the proton around the opposite lone pair direction from the minimum.

Variable position of oxygen electrostatic minima

An examination of the rms values of the separations between hydrogen bond donors and oxygen or nitrogen minima has highlighted the poorer predictive ability of oxygen electrostatic potential minima. This contrasts with the more consistent position of the nitrogen electrostatic minima in relation to the nitrogen atom (for N minima the range of α values is 21.5°; for O minima the range of

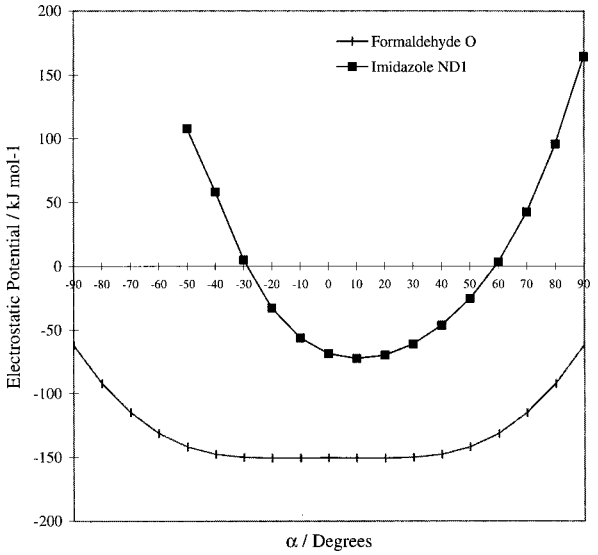

Fig. 8. Comparison of the electrostatic potential energy as a function of α for the formaldehyde oxygen and histidine imidazole ND1 atoms. The electrostatic potential energy between a unit positive charge of radius 0.5 Å and each molecule was calculated in the molecular plane of formaldehyde and in the imidazole ring plane for histidine from the distributed multipole representation. For the imidazole ND1 plot, points from –60° to –90° lie within the van der Waals volume of the *N*-acetyl, *N'*-methylacetamide blocked histidine main chain atoms.

α is 59.4°). Figure 8 compares the electrostatic potential surface around the carbonyl oxygen of formaldehyde and the sp^2 nitrogen ND1 of histidine. Although the oxygen electrostatic potential well is much deeper than that of nitrogen, the potential across the region outside the van der Waals surface of the nitrogen acceptor shows a greater change in the magnitude of the potential, changing rapidly across the arc as α varies. Therefore, the nitrogen minimum is more sharply defined (in the direction of the sp^2 ring nitrogen lone pair), and its relative position is consequently more consistent. The oxygen electrostatic potential is fairly insensitive to α, producing a broad featureless minimum. Thus, the actual location of the minimum will be very easily perturbed by neighbouring function groups, and the electrostatic interactions for a wide range of geometries displaced from the minimum will be only negligibly smaller.

The flatness of the electrostatic potential around the C=O group for an arc in the lone pair plane will generally imply a flat total intermolecular potential surface for hydrogen bonding, allowing a wide range of angular geometries, so that the actual minima are determined by the other interactions. The flatness of the total potential for a fixed length, N-H···O linear hydrogen bond in the lone pair plane has been demonstrated [37] by intermolecular perturbation theory calculations on the hydrogen bonded *trans*-formamide/formaldehyde complex, where the total potential varied by less than 1 kJ mol^{-1}. Thus, the electrostatic minima could only reliably predict the hydrogen bonding proton position on the arc between the lone pairs if the other groups in the molecule were able to distort the total intermolecular potential surface sufficiently to give only one sharp minimum.

Conclusions

These results demonstrate that there is a good correlation between the positions of electrostatic extrema of the isolated molecules and the location of hydrogen bonded atoms in their complexes for N-H donors and N acceptors, where the potential has a well-defined minimum. The correlation is worse for C=O acceptors because the electrostatic potential is so flat, and the carbonyl group can accept hydrogen bonds in a wide arc in the lone pair plane, and may well accept more than one hydrogen bond. Thus, the actual geometry of a hydrogen bonded complex involving a C=O group is much more sensitive to the possibilities for other favourable intermolecular contacts between the molecules involved.

For all three hydrogen bonding groups, the nature of the rest of the molecule can sometimes significantly change the electrostatic potential in their environment. Marked displacements (and even the disappearance) of the extrema are certainly indicative of significant changes in the hydrogen bonding properties.

Hence, considering the electrostatic properties of a specific molecule does provide more information about its likely hydrogen bonds than just the statistical results for the donors and acceptors involved. Preliminary studies with a point dipole molecule suggested that this was unlikely to be more successful than the point charge model for the proton and acceptor atom of the unknown hydrogen bonding molecule. This study has demonstrated that the electrostatic extrema of a molecule can predict the geometries of its hydrogen bonded complexes with a range of other molecules, but the accuracy depends on other intermolecular contacts that are formed. These are determined by the structure of the unknown molecule. In applications of this model, the unknown molecule will be a protein-binding site, and so the atoms adjacent to its hydrogen bonding groups, although unknown, will be constant for interactions with a series of possible ligands. Thus, the other intermolecular contacts that can be formed to the binding site, such as a second hydrogen bond, will be constant, giving a systematic deviation of the hydrogen bonds of the bound ligand from those predicted by the electrostatic extrema. Hence, the extrema are likely to be even more reliable predictors of relative hydrogen bond geometries in comparing the hydrogen bonding capabilities of different ligands for the same uncharacterised binding site.

The results indicate that an oxygen or nitrogen acceptor is likely to be less than 1 Å from a potential maximum (Fig. 7a), and a hydrogen bonding proton is likely

to be within 1 Å of the potential minimum associated with an N acceptor (Fig. 7b), though these deviations can be larger, particularly when there are multiple hydrogen bonds. Hydrogen bonds to oxygen acceptors are often moderately well predicted (<1.5 Å, Fig. 7b) by the potential minima, except in the relatively common situation when the hydrogen bond forms to the opposite lone pair direction. These results can be used to interpret proposed relative binding orientations of ligands to an unknown binding site. If a maximum for each molecule is within about 2 Å, or a minimum is within 3 Å, it is highly likely that both would form hydrogen bonds to the same functional group in the binding site. The empirical observations of matching extrema suggest that the method mainly works by matching possible hydrogen binding sites. Thus, the method could be refined to weight different types of acceptors and donors according to the nature of their associated electrostatic potential, and to provide a refinement on purely geometric hydrogen bonding criteria in the design of new ligands.

Supplementary material

Supplementary material giving details of the hydrogen-bonded complexes is available from the authors.

Acknowledgements

R.P.A. would like to thank the EPSRC for financial support in the form of a studentship. M.B. worked at UCL as an ERASMUS program student.

References

1 Dean, P.M., Molecular Foundations of Drug–Receptor Interaction, Cambridge University Press, Cambridge, U.K., 1987.
2 Jeffrey, G.A. and Saenger, W., Hydrogen Bonding in Biological Structures, Springer, Berlin, Germany, 1991.
3 Mills, J.E.J. and Dean, P.M., J. Comput.-Aided Mol. Design, 10 (1996) 607.
4 Taylor, R., Kennard, O. and Versichel, W., J. Am. Chem. Soc., 105 (1983) 5761.
5 Baker, E.N. and Hubbard, R.E., Prog. Biophys. Mol. Biol., 44 (1983) 5761.
6 Ippolito, J.A., Alexander, R.S. and Christianson, D.W., J. Mol. Biol., 215 (1990) 457.
7 Náray-Szabó, G. and Ferenczy, G.G., Chem. Rev., 95 (1995) 829.
8 Murray, J.S., Lane, P. and Politzer, P., J. Mol. Struct. (THEOCHEM), 209 (1990) 163.
9 Murray, J.S., Ranganathan, S. and Politzer, P., J. Org. Chem., 56 (1991) 3734.
10 Murray, J.S. and Politzer, P., J. Org. Chem., 56 (1991) 6715.
11 Kenny, P.W., J. Chem. Soc. Perkin Trans., 2 (1994) 199.
12 Pauling, L., Proc. Natl. Acad. Sci. USA, 14 (1928) 359.
13 Stone, A.J., Chem. Phys. Lett., 83 (1981) 233.
14 Stone, A.J. and Alderton, M., Mol. Phys., 56 (1985) 1047.
15 Buckingham, A.D. and Fowler, P.W., J. Chem. Phys., 79 (1983) 6426.
16 Buckingham, A.D. and Fowler, P.W., Can. J. Chem., 63 (1985) 2018.
17 Mitchell, J.B.O. and Price, S.L., Chem. Phys. Lett., 154 (1989) 267.
18 Mitchell, J.B.O., Nandi, C.L., Thornton, J.M., Price, S.L., Singh, J. and Snarey, M., J. Chem. Soc. Faraday Trans., 89 (1993) 2619.
19 Mitchell, J.B.O., Nandi, C.L., McDonald, I.K., Thornton, J.M. and Price, S.L., J. Mol. Biol., 239 (1994) 315.
20 Flanagan, K., Walshaw, J., Price, S.L. and Goodfellow, J.M., Protein Eng., 8 (1995) 109.
21 Price, S.L., Lo Celso, F., Treichel, J.A., Goodfellow, J.M. and Umrania, Y., J. Chem. Soc. Faraday Trans., 89 (1993) 3407.
22 Hobza, P. and Sandorfy, C., J. Am. Chem. Soc., 109 (1987) 1302.
23 Hurst, G.J.B., Fowler, P.W., Stone, A.J. and Buckingham, A.D., Int. J. Quantum Chem., 29 (1986) 1223.
24 Apaya, R.P., Lucchese, B., Price, S.L. and Vinter, J.G., J. Comput.-Aided Mol. Design, 9 (1995) 33.
25 Van der Wenden, E.M., Price, S.L., Apaya, R.P., IJzerman, A.P. and Soudijn, W., J. Comput.-Aided Mol. Design, 9 (1995) 44.
26 Frau, J. and Price, S.L., J. Comput.-Aided Mol. Design, 10 (1996) 107.
27 Faerman, C.H. and Price, S.L., J. Am. Chem. Soc., 112 (1990) 4915.
28 INSIGHT II, v. 2.3.0, Biosym Technologies, San Diego, CA, U.S.A., 1993.
29 Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M. and Hagler, A.T., Proteins Struct. Funct. Genet., 4 (1988) 31.
30 CADPAC5: The Cambridge Analytical Derivatives Package, Issue 5.0, 1992. A suite of quantum chemistry programs developed by R.D. Amos with contributions from I.L. Alberts, J.S. Andrews, S.M. Colwell, N.C. Handy, D. Jayatilaka, P.J. Knowles, R. Kobayashi, N. Koga, K.E. Laidig, P.E. Malsen, C.W. Murray, J.E. Rice, J. Sanz, D. Simandiras, A.J. Stone and M.-D. Su.
31 Binkley, J.S., Pople, J.A. and Hehre, W.J., J. Am. Chem. Soc., 102 (1980) 939.
32 Stone, A.J., Popelier, P.L.A. and Wales, D.J., ORIENT: A program for calculating the electrostatic interactions between molecules, v. 3.0, University of Cambridge, Cambridge, U.K., 1994.
33 Spackman, M.A., J. Chem. Phys., 85 (1986) 6587.
34 Price, S.L. and Stone, A.J., J. Chem. Soc. Faraday. Trans., 88 (1992) 1755.
35 Taylor, R. and Kennard, O., J. Am. Chem. Soc., 104 (1982) 5063.
36 Desiraju, G.R., Acc. Chem. Res., 24 (1991) 290.
37 Mitchell, J.B.O. and Price, S.L., J. Comput. Chem., 11 (1990) 1217.