BIOINFORMATICS

Vol. 00 no. 00 2005
Pages 1–7

Analysing the Ability to Retain Sidechain Hydrogen-bonds in Mutant Proteins

Alison L. Cuff[1,†], Robert W. Janes[2], and Andrew C.R. Martin[3,*,‡]

[1]School of Animal & Microbial Sciences, University of Reading, Whiteknights, P.O. Box 228, Reading RG6 6AJ, U.K.; [2]School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, U.K.; [3]Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, U.K.

ABSTRACT

Motivation: Hydrogen bonds are one of the most important inter-atomic interactions in biology. Previous experimental, theoretical and bioinformatics analyses have shown that the hydrogen bonding potential of amino acids is generally satisfied and that buried unsatisfied hydrogen-bond-capable residues are destabilizing. When studying mutant proteins, or introducing mutations to residues involved in hydrogen bonding, one needs to know whether a hydrogen bond can be maintained. Our aim, therefore, was to develop a rapid method to evaluate whether a sidechain can form a hydrogen-bond.

Results: A novel knowledge-based approach was developed in which the conformations accessible to the residues involved are taken into account. Residues involved in hydrogen bonds in a set of high resolution crystal structures were analyzed and this analysis is then applied to a given protein. The program was applied to assessment of mutations in the tumour-suppressor protein, p53. This raised the number of distinct mutations identified as disrupting sidechain-sidechain hydrogen bonding from 181 in our previous analysis to 202 in this analysis.

Availability: http://www.bioinf.org.uk/hbonds/

Contact: andrew@bioinf.org.uk –or– martin@biochem.ucl.ac.uk

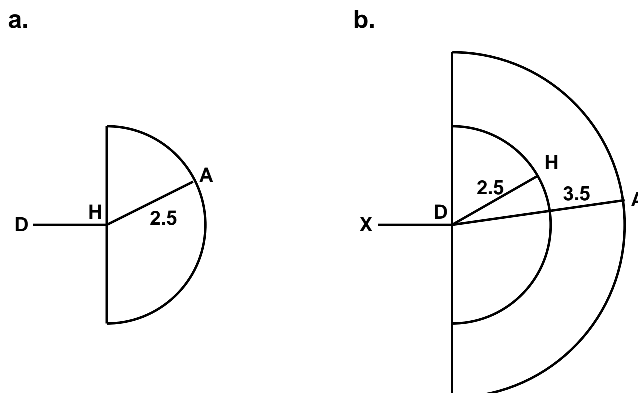

Fig. 1. Geometrical criteria for identifying hydrogen bonds (Baker and Hubbard, 1984). a) In cases where hydrogen positions can be calculated, the H...A distance is ≤ 2.5Å and the angle at the hydrogen is from 90–180°. b) In cases where hydrogen positions cannot be calculated, the D...A distance is ≤ 3.5Å and the angle between the donor antecedent (X), donor and acceptor 90–180°.

INTRODUCTION

Baker and Hubbard (1984) described the hydrogen bond as involving 'the interaction of a proton, carrying a partial positive charge, on a donor group, with the electron density on an acceptor atom'. It wasn't until Pauling *et al* published two papers proposing repetitive secondary structure elements, the α-helix (Pauling *et al.*, 1951) and both parallel and anti-parallel β-sheets (Pauling and Corey, 1951), that the important rôle backbone hydrogen-bonds played in peptide and protein structures was realized. Important conserved hydrogen-bonds also occur in tight turns (Sibanda *et al.*, 1989) and the hydrogen bond is considered to be one of the most important inter-atomic interactions in both biology and chemistry.

Backbone hydrogen bonds represent an average of 68% of the hydrogen-bonds in globular proteins (Stickle *et al.*, 1992) and are largely responsible for the conformation of conserved secondary structure elements. The remaining hydrogen-bonds in globular proteins are sidechain-sidechain and sidechain-mainchain and occur in roughly equal proportions (Stickle *et al.*, 1992). Sidechain-mainchain hydrogen bonds are often observed in turns and at the termini of helices and it has been suggested that, as well as contributing to protein stability, they are involved in the initiation of helix and turn formation (Baker and Hubbard, 1984). Finally, polar sidechains located at the protein surface and therefore exposed to bulk solvent, will often hydrogen bond to water molecules, as will those lining a solvated cavity in the interior of a protein.

The geometry of the hydrogen bond depends upon whether the hydrogen is bound to sp^2 or sp^3 hybridized atoms. Sp^2 hybridized atoms possess three orbitals, in a trigonal planar arrangement, the angle between any two orbitals being approximately 120°. Sp^3 hybridized hydrogen atoms have four orbitals in a tetrahedral arrangement, the angle between any two orbitals being approximately 109.5°. Baker and Hubbard (1984) defined relatively generous geometric criteria for identifying hydrogen bonds so as to include all reasonable hydrogen bonds including weak bifurcated bonds. They defined the minimum D-H...A and P-A...D angles as 90° and the maximum H...A distance as 2.5Å, where 'D' is the donor atom,

* To whom correspondence should be addressed.

† Present address: School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, U.K.

‡ This work was initiated while working at: School of Animal & Microbial Sciences, University of Reading, Whiteknights, P.O. Box 228, Reading RG6 6AJ, U.K.

© Oxford University Press 2005.

'A' is the acceptor atom (both are oxygen or nitrogen in proteins) and 'P' is the antecedent atom (normally a carbon) to which the acceptor is bound. Since crystallographic structures do not normally define the positions of hydrogens (except when solved at extremely high resolution), these must be calculated by application of standard geometric rules. For those hydrogen bonds involving a hydrogen bound to a terminal sp^3 atom, such as a hydroxyl oxygen, the position of the hydrogen atom cannot be determined. In these cases, they defined the minimum value for the angle X-D...A as 90° and the maximum distance for D...A as 3.50Å (where 'X' is the donor antecedent atom). See Figure 1.

Main chain NH groups can donate a single hydrogen bond whilst the C=O group can accept two via two lone pair electrons on the sp^2-hybridized oxygen. All but one of the polar sidechains can contribute to at least two hydrogen bonds, the exception being Trp, which can only donate a single hydrogen bond. Baker and Hubbard (1984), and McDonald and Thornton (1994) showed that the vast majority of mainchain NH and C=O groups are involved in hydrogen bonding. They and others have also shown that nearly all sidechains capable of being involved in a hydrogen bond are indeed involved in at least one hydrogen bond (Baker and Hubbard, 1984; McDonald and Thornton, 1994; Stickle *et al.*, 1992). It has been suggested that, for those residues that are not hydrogen bonded, some disorder of protein atoms, disorder of water atoms, or steric constraints are responsible (Baker and Hubbard, 1984). McDonald and Thornton also found that residues able both to donate and accept hydrogen bonds will often donate, but not accept.

Experimental studies have supported the notion that an unsatisfied donor or acceptor atom has a detrimental effect on protein stability. Alber *et al* (1987), for example, found that replacing threonine with other residues not capable of contributing to a hydrogen-bond resulted in the destabilization of the protein. Vogt and Argos (1997) found that when comparing thermostable and mesostable proteins from the same family, 80% of the thermostable proteins had more hydrogen bonds than their mesostable counterparts. Chen *et al* (1993) investigated the contribution of hydrogen bonds to protein stability by studying two barnase mutants both of which had lost a hydrogen bond and were found to be less stable than the native structure. Crystal structures confirmed there was no disruption of the structure, overall loss of the hydrogen bonding pattern, or introduction of any new interactions. This suggested that, in both cases, the loss of a single hydrogen bond was the primary cause of the resulting loss in protein stability.

Takano *et al* (2001) showed that replacing a non-polar sidechain with a polar sidechain in a number of different lysozyme mutants contributed to protein stability through the creation of hydrogen bonds. Pace *et al* (2001) replaced tyrosine with phenylalanine in Ribonuclease Sa (RNAse Sa) and RNAse Sa3 and found the tyrosine hydroxyl groups made a positive contribution to protein stability. Takano *et al* (2003) studied threonine-to-valine mutations in RNAse Sa and found the mutants were less stable than the wild-type protein by an average 1.3±0.9 kcal/mol. In addition, they created four valine-to-threonine mutants in RNAse Sa and found that these were also less stable than wild type, by an average of 1.8±1.1 kcal/mol. These studies all support the notion that unpaired hydrogen-bonding residues are unfavourable.

Some theoretical studies contradict these observations. Yang and Honig, for example, stated that the contribution of hydrogen bonds to protein stability was zero, and that they even have a detrimental effect, based on determining the free energy balance of hydrogen bonds in α-helices (Yang and Honig, 1995a) and anti-parallel β-sheets (Yang and Honig, 1995b). Sippl *et al* (1996), came to similar conclusions by alternative means.

In summary, the net stabilization of a hydrogen bond is probably small since, in the unfolded state, hydrogen bonds can be formed with water. However, unsatisfied hydrogen-bonding residues are unfavourable and a mutation leaving behind an unsatisfied hydrogen bond donor, or acceptor, is likely to cause structural destabilization.

Therefore, when studying mutant proteins, or when designing experiments in which substitutions are to be made to residues involved in forming sidechain hydrogen bonds, one wishes to know whether the hydrogen bond can be maintained. Clearly if a sidechain involved in a hydrogen bond is mutated to a non-polar residue, then the answer is 'no', but if, for example, an asparagine is substituted by a serine, can the hydrogen-bond the asparagine was making be maintained?

METHODS

Our aim was to develop a method able to evaluate whether a pair of sidechains could adopt conformations in which a hydrogen-bond could be formed. A conformational search using a program such as CONGEN (Bruccoleri and Karplus, 1987) could be used for this purpose, but searches on longer sidechains using a fine search grid (say 5°) and off-grid relaxation using a few cycles of energy minimization take a few hours on a 2GHz PC. Thus a novel knowledge-based approach was adopted in which the geometry of the residue replacing a native residue involved in a hydrogen bond is taken into account. The method was later adapted to allow sidechain-mainchain hydrogen bonds to be assessed.

The algorithm for hydrogen bond assessment is performed in two phases. The first phase is an analysis of residues involved in hydrogen bonds in a set of high resolution crystal structures. In the second phase, this analysis is applied to the protein in question. Donor sidechains are defined as Arg, Asn, Gln, His, Lys, Ser, Thr, Trp and Tyr. Acceptor sidechains are Asn, Asp, Glu, Gln, His (uncharged), Ser, Thr and Tyr. Remaining sidechains (Ala, Cys, Phe, Gly, Ile, Leu, Met, Pro, Val) are unable to participate in hydrogen bonds. The mainchain of all amino acids is able to accept a hydrogen bond via the backbone carbonyl oxygen while the mainchain of all amino acids other than proline (strictly an imino acid) is able to donate a hydrogen bond via the backbone nitrogen.

The first phase is implemented in the program *hydrogen_matrices*. Input to the program is a set of protein domains from the CATH database (Orengo *et al.*, 1999) in which the resolution is $\leq$2.5Å. For each amino acid type capable of acting as a sidechain hydrogen bond donor, two 3D matrices (or 'grids'), 'DONOR' and 'PARTNERTODONOR' are initialized with all values set to FALSE. Similarly 'ACCEPTOR' and 'PARTNERTOACCEPT' grids are created for each amino acid capable of acting as an acceptor. Mainchain donor nitrogen and acceptor oxygen hydrogen bonds are treated in the same way, but separately.

Each sidechain capable of acting as a hydrogen bond donor (termed a 'key' residue) is analyzed in turn. The protein is translated such that the Cα of the key residue is at the origin and rotated such that its Cβ is on the xy-plane and its N is on the positive x-axis. This is referred to as the 'standard orientation'. The location of each donor atom of the key residue is recorded and converted to a grid point. The flag for this location in the DONOR grid is then set to TRUE. Hydrogen bonds are identified using the simple geometric criteria of Baker and Hubbard (see Figure 1) and the locations of any acceptor atoms forming hydrogen bonds with these donor atoms are converted to grid points in the same way. The flag for their locations in the PARTNERTODONOR matrix for the key residue is then set to TRUE.

Having analyzed all the donor residues, the procedure is repeated using each residue capable of acting as a hydrogen bond acceptor as the key residue

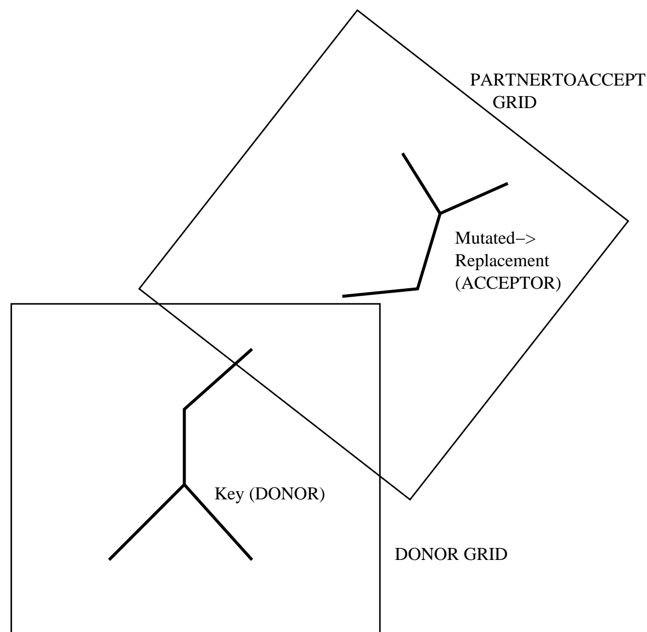

Fig. 2. The PARTNERTOACCEPT grid is rotated such that it is centred around the partner residue. If a hydrogen bond can be formed between the residues then a populated cell in the DONOR grid will be coincident with a populated cell in the PARTNERTOACCEPT grid.

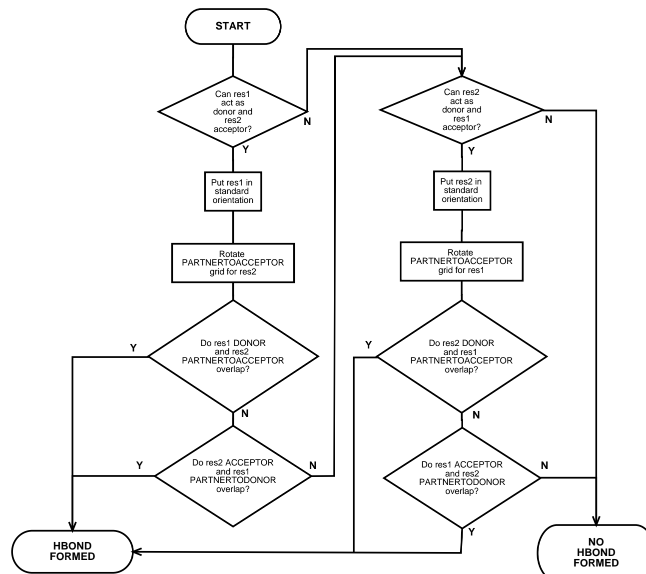

Fig. 3. Flowchart of the overall process of the analysis program.

and populating the ACCEPTOR and PARTNERTOACCEPT matrices. Note that since the grids are purely Boolean, there is no requirement for the dataset to be non-redundant since counts are not made.

For mainchain/sidechain hydrogen bonds, the procedure is similar, but the sets of key atoms used to define the standard orientation are different for the mainchain atoms. If the same set of atoms (N, Cα, Cβ) was used then the location of the hydrogen on the nitrogen would depend on the backbone ϕ angle while the location of the carbonyl oxygen would depend on the backbone ψ angle. Thus for the donor nitrogen, atoms C$'$, N and Cα are used (where C$'$ indicates the carbonyl carbon of the previous residue) and for the acceptor carbon, atoms Cα, C and O are used. Clearly this fixes the position of the backbone nitrogen and carbonyl oxygen in a single location in the mainchain DONOR and ACCEPTOR grids respectively, but atoms will be distributed around the PARTNERTODONOR and PARTNERTOACCEPT grids.

The second phase, implemented in the program *checkhbond*, applies this analysis to a pair of residues to determine whether they are able to form a hydrogen bond. Input to the program is a set of grids generated in the first phase, the protein to be tested, two residue identifiers to indicate the locations of the residues in question, and the identity of the amino acids to be tested at the second locations. (The identity of the amino acid at the first location is unchanged from what is present in the PDB file.)

In summary, one residue is chosen as the key residue and here is assumed to be donor. The protein is moved such that this residue is at the origin in the standard orientation as in the first phase; the other residue is the acceptor. The PARTNERTOACCEPT grid, which stores the location of donor atoms able to hydrogen bond with this acceptor, is then translated and rotated from the origin to be coincident with the acceptor residue. If the backbones of two amino acids are in positions where the sidechains can rotate to interact to form a hydrogen bond, then a populated cell in the DONOR grid of the key residue should be coincident with a populated cell in the PARTNERTOACCEPT grid of the acceptor residue. Equally a populated cell in the PARTNERTODONOR grid of the key residue should be coincident with a populated cell in the ACCEPTOR grid of the acceptor residue. This procedure is illustrated in Figure 2. If no hydrogen bond is identified, then if the two residues are both able to donate and accept hydrogen bonds, the nature of the two residues is swapped such that the key residue is assumed to be acceptor and its partner is donor. The analysis is then repeated. The overall process is illustrated in Figure 3. An equivalent procedure is followed for mainchain/sidechain hydrogen bonds.

In practice, the grids are only conceptually rotated — in reality the key residue is fitted to the partner residue by superimposing the N, Cα, Cβ atoms using the McLachlan fitting algorithm (McLachlan, 1979) as modified by Mike Sutcliffe (personal communication) and implemented locally in C. This provides a rotation matrix and translation vector which enables a point on the partner residue's grid (centred around the origin in the standard orientation) to be projected into the position it would occupy if the grid were moved to be coincident with the partner residue. Thus by applying this manipulation to each populated grid point in the partner residue matrix, one can then check that location in the key residue matrix to determine whether there is a match.

Prior to comparing the matrices, they are culled to remove locations which would lead to steric clashes. Any grid cells within two radii of another atom in the structure are set to a value of FALSE such that these locations are not allowed for hydrogen bonding.

In addition, some plasticity is allowed in the matching. The quantization of space performed when populating the grids is compounded by rounding errors during conversion between integer grid points and real coordinates and manipulation of coordinates in real space. Thus if no match is found between cells, the program searches for populated grid squares that are within a specified cut-off distance from each other (default, 0.25Å).

A further modification was made to the programs to enable the quality of hydrogen bonds to be assessed. Here, by using a non-redundant set of domains from CATH (the SREPs), counts rather than Boolean values were stored in the matrices. SREPs are 'sequence representatives' of each CATH homologous family with each pair of SREPs having a sequence identity $< 35\%$. The counts can then be converted to probabilities of observing a hydrogen bonding atom in one of the grid cells and, by applying a simplified form of the inverse Boltzmann equation, one can simply sum the log of the probabilities in a pair of overlapping cells to obtain a pseudo-energy for the interaction:

$$E = -(\log(P_1) + \log(P_2))$$

where $P_i = \frac{n_i}{N_i}$ and n_i is the count of hydrogen-bonding atoms observed in this cell and N_i is the total count of hydrogen-bonding atoms observed across the matrix.

When this option is chosen, the program searches for the best pseudo-energy for an interaction between a pair of residues rather than just determining that an interaction occurs.

RESULTS

The program was applied to the assessment of mutations in the core domain of the tumour-suppressor protein, p53 (Cho *et al.*, 1994, PDB file: 1tsr). p53 is mutated in more than 50% of human cancers and in a recent study by Martin *et al* (2002), an automated protocol was developed to classify the effects of mutations on the p53 core domain according to their likely effects on the local structure of the protein. One of the steps performed in this protocol was to assess the effects of mutations to residues involved in hydrogen bonding. The protocol took a very conservative approach to classifying mutations as disrupting hydrogen bonding. It was assumed that if a residue donating a hydrogen bond was replaced by another hydrogen bond donor, then the hydrogen bond would be maintained. In reality, geometric constraints may prevent the hydrogen bond from forming, allowing additional mutations to be classified as disrupting hydrogen bonding.

In their analysis, Martin *et al* (2002) identified 181 distinct mutations that were classified as disrupting hydrogen bonding. Consideration of geometric constraints on sidechain-sidechain hydrogen bonds by application of the algorithm described here now identifies 202 distinct mutations likely to disrupt hydrogen bonds.

Figure 4 illustrates one example from the analysis of p53 in which Tyr126, which hydrogen bonds with Asn131, is mutated to Asp. The original assessment assumed that this hydrogen bond would be preserved (Tyr126 was acting as an acceptor and Asp is capable of acting as an acceptor); one can now ask whether the geometry is such that Asp126 and Asn131 can form a sidechain-sidechain hydrogen bond. The illustration suggests that these residues may not be able to interact with one another, but it is possible that rotation of the χ_1 angle of Asn131 might allow an interaction to occur. However, analysis performed using the algorithm described here shows that no such interaction can occur.

Comparing hydrogen bond pseudo-energy with calculated enthalpy

Figure 5 shows the distribution of pseudo-energies for sidechain-sidechain hydrogen bonds. It can be seen that the distribution is approximately normal with a mean of 10.4 and standard deviation of 2.04. 99.5% have a pseudo-energy < 15 corrsponding to a Z-score of ~ 2.25. For sidechain-backbone-donor hydrogen bonds, values are $\bar{x} \approx 7.0, \sigma \approx 1.21$ with 95.8% having a pseudo-energy < 10 corrsponding to a Z-score of ~ 2.48. For sidechain-backbone-acceptor hydrogen bonds, values are $\bar{x} \approx 13.7, \sigma \approx 3.33$. with 90% having a pseudo-energy < 18 corresponding to a Z-score of ~ 1.29.

Pseudo-energies for sidechain-sidechain hydrogen bonds were compared with enthalpies calculated using the CHARMm potential (Brooks *et al.*, 1983) implemented in the program *ECalc* (Martin, unpublished). *ECalc* allows calculations to be restricted to a region of interest and to use only parts of the full CHARMm potential. ECalc also implements a variation of the SHAKE algorithm (Ryckaert *et al.*, 1977) to relax bad van der Waals contacts.

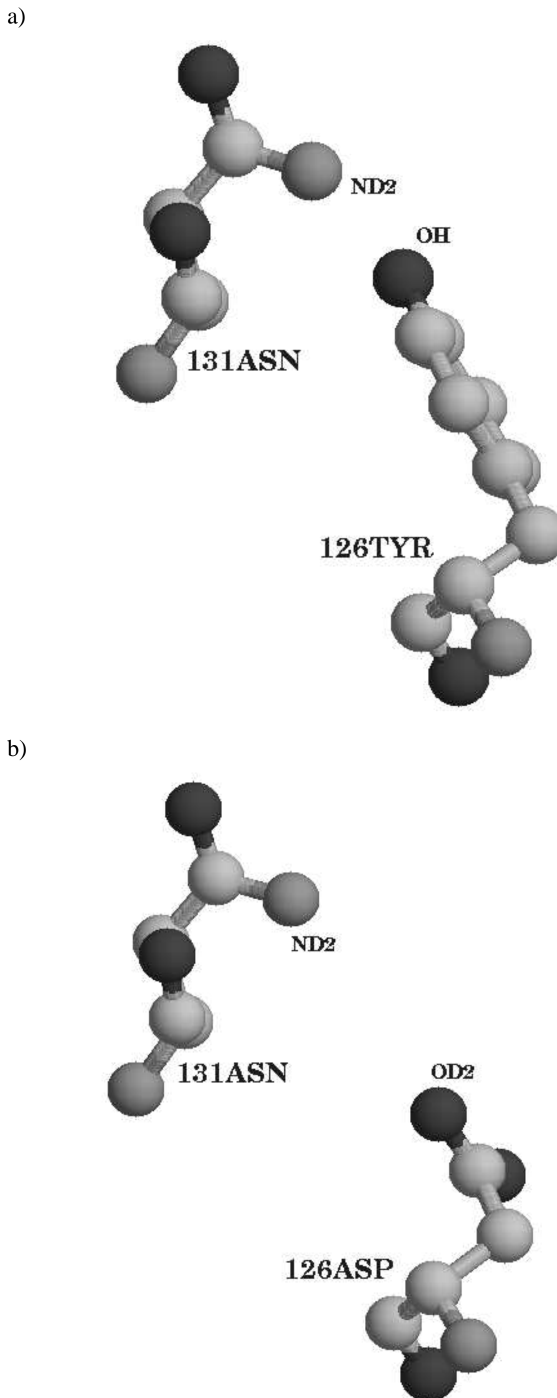

Fig. 4. a) The native Tyr126 and Asn131 residues viewed in relation to one another in the p53 core domain. They interact with each other via a hydrogen bond in which Asn131 donates a hydrogen and Tyr126 accepts it. b) After mutation of Tyr126 to Asp, can the hydrogen bond with Asn131 still be formed? Clearly the Asp residue is much smaller than Tyr, but sidechain torsion angle rotations might allow the residues to interact. The algorithm described here shows that the residues are not able to adopt a conformation in which a hydrogen bond can be formed.

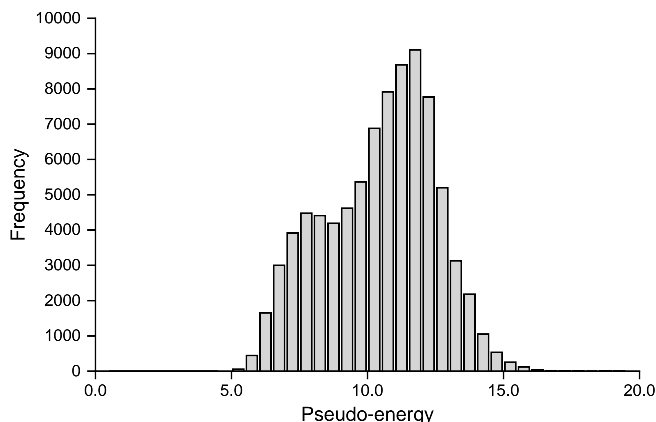

Fig. 5. Distribution of sidechain-sidechain hydrogen bond pseudo-energies calculated for SReps from CATH V2.6.0

Where atoms clash, they are moved apart along the vector between them in inverse proportion to their masses. This iterates to convergence and is followed by a standard SHAKE algorithm to optimize bond lengths. All SReps from CATH v2.6.0 were analyzed. Using the list of sidechain-sidechain hydrogen bonds identified by *hydrogen_matrices*, for each hydrogen bond, the enthalpy was calculated using *ECalc* and the pseudo-energy was calculated using *checkhbond*.

Three variations were used in which (1) the full enthalpy for the pair of residues making the hydrogen bond was calculated, (2) the full enthalpy was calculated after relaxing van der Waals contacts, (3) only the hydrogen bonding potential was calculated. Results are shown in Figure 6.

The graphs show that there is no correlation between pseudo-energy and enthalpy. The lack of correlation is perhaps not surprising. While the enthalpic calculations take account only of purely local energetic criteria in a particular instance of an interaction between two amino acids, the pseudo-energy calculation includes preferences for sidechain conformations. In addition, with the CHARMM potential we used for comparison, the hydrogen bonding potential only contains a component based on the angle between the atoms when the position of the hydrogen can be defined precisely from the non-hydrogen atom geometry; in other cases hydrogen bonds are treated as essentially elctrostatic interactions providing the angle between the non-hydrogen atoms is within a specified range. Thus the pseudo-energy, which implicitly accounts for angular information, even where the hydrogen cannot be placed unambiguously, may be a more realistic measure of hydrogen bond quality. When comparing pseudo-energy and enthalpy, the only clear trend is that when the hydrogen bond energy is considered alone, where the pseudo-energy is low, the enthalpy of the hydrogen bond energy is also low, but high pseudo-energies can also occur when the enthalpy of the hydrogen bond is low. In particular, a pseudo-energy $\leq$ 8 (corresponding to a Z-score of -1.18) for sidechain-sidechain hydrogen bonds is indicative of a low hydrogen bond energy.

Using the pseudo-energy, another hotspot mutation, Arg175His (very commonly observed in cancer), while still capable of forming a sidechain-sidechain hydrogen bond with Ser183, does so with a

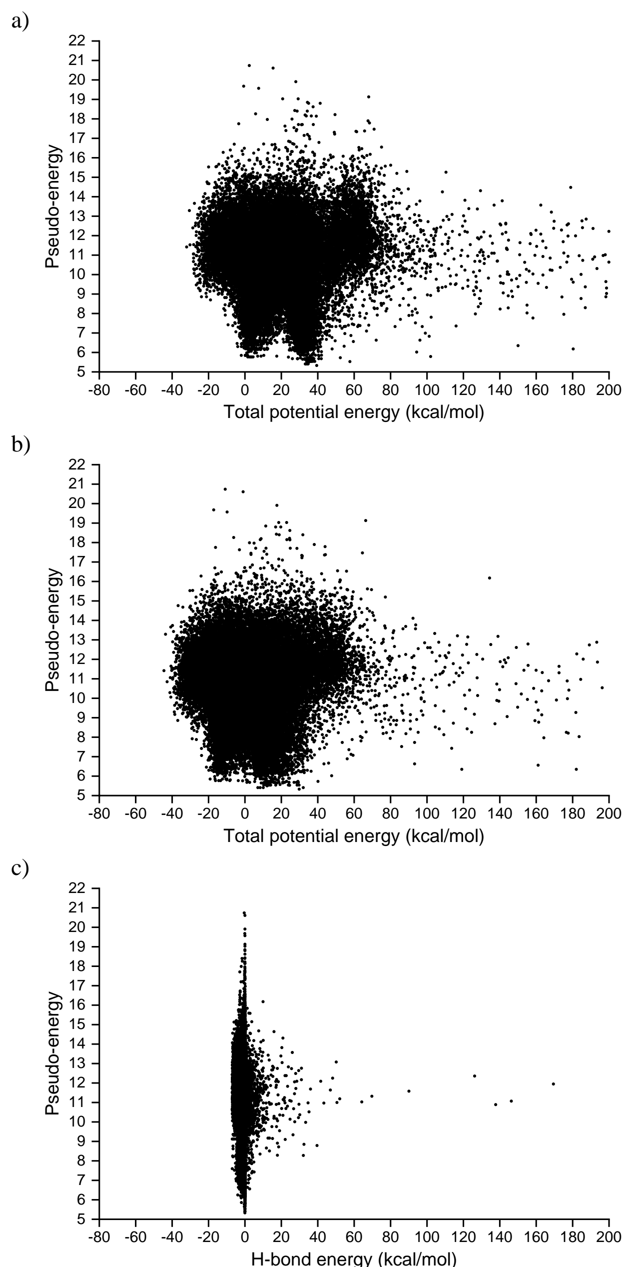

Fig. 6. Graphs of pseudo-energy plotted against enthalpy using the CHARMm potential calculated using ECalc for a) full potential, b) full potential after relaxation, c) hydrogen bonding energy only. In all cases, only datapoints with enthalpies less than 200kcal/mol are shown and Pearson's r was less than 0.05

much poorer energy. The native bond forms with a pseudo-energy of 12.09 (Z-score $= +0.81$), while the mutant is 13.96 (Z-score $= +1.73$).

DISCUSSION

The knowledge-based nature of the method we have developed automatically considers all likely conformations for both the key and

partner sidechains in a very rapid manner compared with conformational search. Evaluation of a hydrogen bond takes no more than a couple of seconds. The grids contain information for all observed conformations of a sidechain and locations of partner atoms and therefore consider the likely orientations in which hydrogen bonds will be formed. When applied to analysis of mutations in the p53 tumour suppressor protein, the method identified disruptions to a number of hydrogen bonds considered to be conserved in our older simple analysis scheme (Martin *et al.*, 2002).

In some cases, in order for a sidechain-sidechain hydrogen bond to be maintained after a mutation, the direction of the hydrogen bond may be changed. The old hydrogen bond assessment method was not able to detect conservation of such hydrogen bonds as it only accounted for the hydrogen bond capabilities of the mutated residue. Thus if a hydrogen bond donor is replaced by a residue only capable of accepting a hydrogen bond, but the partner residue can both donate and accept hydrogen bonds, the new method may be able to identify a hydrogen bond between the residues whereas the old method would not. For example, in p53, considering the hydrogen bonded residue pair Asn131-Tyr126, Asn131 is the donor and Tyr126 the acceptor. If Asn131 is mutated to Asp which can only accept hydrogen bonds, the old method would suggest that the hydrogen bond had been disrupted, while the new method suggests that a hydrogen bond can be maintained with Tyr126 as donor and Asp131 as acceptor.

While this is a clear improvement in the methodology, there may still be problems if a residue is involved in a hydrogen-bond network. Assume residue A is at the centre of a hydrogen-bond network in which it acts as a donor to residue B (which is an acceptor) and as an acceptor to residue C (which is a donor). If residue C is replaced by an acceptor-only residue, then, as in the example above, the method may suggest that a hydrogen bond can be maintained between A and C if A becomes the donor and C the acceptor. However, the donor potential of A is required to maintain its hydrogen bond with B.

While we intend to address this problem in a future version of the program, its impact is probably small. Although most hydrogen-bond-capable atoms will contribute to at least one hydrogen bond, the frequency with which they contribute to two or more hydrogen bonds varies considerably. Charged groups, and in particular Lys NH_3^+, Arg NH_2^+ and Asp COO^- are much more likely to contribute to all their possible hydrogen bonds than uncharged groups. Buried Glu COO^- oxygen atoms can contribute to two hydrogen bonds and do so around 50% of the time. However, the hydrogen bond capable atoms of Ser, Thr, Cys and Tyr rarely contribute to more than one hydrogen bond (McDonald and Thornton, 1994). Stickle *et al* (1992) found that there were only slightly more hydrogen bonds in a protein than there were hydrogen donor/acceptor pairs (1.08 hydrogen bonds per pair). Thus, despite the fact that all but one of the hydrogen-bond-capable residues can be involved in more than one hydrogen bond, only a small fraction of such residues actually participate in hydrogen bonding networks.

In the example of the Arg175His mutation given above, in addition to the sidechain/sidechain hydrogen bond with Ser183, the arginine sidechain acts as a donor in interactions with the mainchain carbonyl oxygens of Pro191, Gln192 and Met237. Separate evaluation of these hydrogen bonds when the histidine is substituted for arginine shows that all can be maintained (two with poorer pseudo-energies, one with a better pseudo-energy). However, it would not be possible for the histidine to satisfy all four hydrogen bonds simultaneously. One 'hotspot' mutation, Arg249Ser (very commonly observed in cancer), was previously classified as not disrupting hydrogen-bonding. Arg249 forms sidechain/sidechain hydrogen bonds with Tyr163 and His168 and acts as a donor in hydrogen bonds with the backbone carbonly oxygen of residues 245 and 246. When geometry is considered, the individual sidechain/sidechain hydrogen bonds can be maintained when Arg249 is mutated to serine, one of the sidechain/mainchain-acceptor hydrogen bonds can be maintained, but the other is lost. It would not be possible for serine to satisfy more than one of these hydrogen bonds simultaneously.

Histidine is also a special case and requires further consideration. The two imidazole nitrogens, ND1 and NE2, cannot donate and accept a hydrogen bond simultaneously, so each nitrogen contributes to only one hydrogen bond (McDonald and Thornton, 1994). When uncharged, the hydrogen may be located at either of the nitrogens such that one becomes a donor and the other an acceptor. However, when charged, a hydrogen is located at both nitrogens, making them both donors.

In assessing the impact of breaking a hydrogen bond, we also intend to consider surface residues explicitly. Suppose residue A which is partially solvent exposed, is involved in a hydrogen bond with residue B. When residue B is mutated such that the hydrogen-bond can no longer be formed, residue A may be able to rotate to hydrogen-bond with water. In the analysis of p53, out of the 202 distinct mutations identified as affecting hydrogen bonding, 147 affected residues with relative solvent accessibility $\geq 10\%$. Of these, the hydrogen bonding partners also had relative solvent accessibility $\geq 10\%$ in 102 cases where it may be possible for these residues to satisfy their hydrogen bonding potential through interaction with water.

In summary, we have developed a new method for evaluating the impact of mutations to residues involved in hydrogen bonds. The method is very rapid compared with conformational search or dynamics techniques and identifies only realistic hydrogen bonds through use of a knowledge-based distribution of potential hydrogen-bonding sites around residues. We have applied the method to the analysis of mutations in the p53 tumour suppressor protein substantially increasing our ability to offer suggestions of mutations which may disrupt the stability of the protein.

A server allowing the user to upload a PDB file and specify two residues together with the amino acid present at one of the locations has been implemented and is available over the web at `http://www.bioinf.org.uk/hbonds/` or `http://acrmwww.biochem.ucl.ac.uk/hbonds/`. The server can assess sidechain/sidechain or sidechain/mainchain hydrogen bonds.

ACKNOWLEDGMENTS

ALC was funded through an MRC Priority Studentship in Bioinformatics.

REFERENCES

Alber, T., Dao-pin, S., Wilson, K., Wozniak, J. A., Cook, S. P. and Matthews, B. W. (1987) Contributions of hydrogen bonds of Thr157 to the thermodynamic stability of phage T4 lysozyme, *Nature (London)*, **330**, 41–46.

Baker, E. N. and Hubbard, R. E. (1984) Hydrogen bonding in globular proteins, *Progr. Biophy. Molec. Biol*, **44**, 97–179.

Brooks, R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and M., K. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J. Comp. Chem.*, **4**, 187–217.

Bruccoleri, R. E. and Karplus, M. (1987) Prediction of the folding of short polypeptide segments by uniform conformational sampling, *Biopolymers*, **26**, 137–168.

Chen, Y. W., Fersht, A. R. and Henrick, K. (1993) Contribution of buried hydrogen bonds to protein stability. The crystal structures of two barnase mutants, *J. Mol. Biol.*, **234**, 1158–1170.

Cho, Y., Gorina, S., Jeffrey, P. D. and Pavletich, N. P. (1994) Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations, *Science*, **265**, 346–355.

Martin, A. C. R., Facchiano, A. M., Cuff, A. L., Hernandez-Boussard, T., Olivier, M., Hainaut, P. and Thornton, J. M. (2002) Integrating mutation data and structural analysis of the TP53 tumour-suppressor protein, *Human Mutation*, **19**, 149–164.

McDonald, I. K. and Thornton, J. M. (1994) Satisfying hydrogen-bonding potential in proteins, *J. Mol. Biol.*, **238**, 777–793.

McLachlan, A. D. (1979) Least squares fitting of two structures, *J. Mol. Biol.*, **128**, 74–79.

Orengo, C. A., Pearl, F. M. G., Bray, J. E., Todd, A. E., Martin, A. C., Lo Conte, L. and Thornton, J. M. (1999) The CATH database provides insights into protein structure/function relationships, *Nucleic Acids Research*, **27(1)**, 275–279.

Pace, C. N., Horn, G., Hebert, E. J., Bechert, J., Shaw, K., Urbanikova, L., Scholtz, J. M. and Sevcik, J. (2001) Tyrosine hydrogen bonds make a large contribution to protein stability, *J. Mol. Biol.*, **312**, 393–404.

Pauling, L. and Corey, R. B. (1951) Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets, *Proc. Natl. Acad. Sci. USA*, **37**, 729–740.

Pauling, L., Corey, R. B. and Branson, H. (1951) The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain, *Proc. Natl. Acad. Sci. USA*, **37**, 205–211.

Ryckaert, J.-P., Ciccotti, G. and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes, *J. Comput. Phys.*, **23**, 327–341.

Sibanda, B. L., Blundell, T. L. and Thornton, J. M. (1989) Conformation of beta-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering, *J Mol Biol*, **206**, 759–777.

Sippl, M. J., Ortner, M., Jaritz, M., Lackner, P. and Flockner, H. (1996) Helmholtz free energies of atom pair interactions in proteins, *Fold Des*, **1**, 289–298.

Stickle, D. F., Presta, L. G., Dill, K. A. and Rose, G, D. (1992) Hydrogen bonding in globular proteins, *J. Mol. Biol.*, **226**, 1143–1159.

Takano, K., Scholtz, J. M., Sacchettini, J. C. and Pace, C. N. (2003) The contribution of polar group burial to protein stability is strongly context-dependent, *J. Biol. Chem.*, **278**, 31790–31795.

Takano, K., Yamagata, Y. and Yutani, K. (2001) Contribution of polar groups in the interior of a protein to the conformational stability, *Biochemistry*, **40**, 4853–4858.

Vogt, G. and Argos, P. (1997) Protein thermal stability: hydrogen bonds or internal packing, *Folding and Design*, **2**, S40–S46.

Yang, A. and Honig, B. (1995a) Free energy determinants of secondary structure formation: I. alpha-helices, *J. Mol. Biol.*, **252**, 351–365.

Yang, A. and Honig, B. (1995b) Free energy determinants of secondary structure formation: II. antiparallel beta-sheets, *J. Mol. Biol.*, **252**, 366–376.