Article No. jmbi.1998.2401 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 285, 1735–1747

JMB

Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-chain Amide Orientation

J. Michael Word, Simon C. Lovell, Jane S. Richardson and David C. Richardson*

Biochemistry Department
Duke University, Durham
NC 27710-3711, USA

Small-probe contact dot surface analysis, with all explicit hydrogen atoms added and their van der Waals contacts included, was used to choose between the two possible orientations for each of 1554 asparagine (Asn) and glutamine (Gln) side-chain amide groups in a dataset of 100 unrelated, high-quality protein crystal structures at 0.9 to 1.7 Å resolution. For the movable-H groups, each connected, closed set of local H-bonds was optimized for both H-bonds and van der Waals overlaps. In addition to the Asn/Gln ''flips'', this process included rotation of OH, SH, NH_3^+, and methionine methyl H atoms, flip and protonation state of histidine rings, interaction with bound ligands, and a simple model of water interactions. However, except for switching N and O identity for amide flips (or N and C identity for His flips), no non-H atoms were shifted. Even in these very high-quality structures, about 20% of the Asn/Gln side-chains required a 180° flip to optimize H-bonding and/or to avoid NH_2 clashes with neighboring atoms (incorporating a conservative score penalty which, for marginal cases, favors the assignment in the original coordinate file). The programs Reduce, Probe, and Mage provide not only a suggested amide orientation, but also a numerical score comparison, a categorization of the marginal cases, and a direct visualization of all relevant interactions in both orientations. Visual examination allowed confirmation of the raw score assignment for about 40% of those Asn/Gln flips placed within the ''marginal'' penalty range by the automated algorithm, while uncovering only a small number of cases whose automated assignment was incorrect because of special circumstances not yet handled by the algorithm. It seems that the H-bond and the atomic-clash criteria independently look at the same structural realities: when both criteria gave a clear answer they agreed every time. But consideration of van der Waals clashes settled many additional cases for which H-bonding was either absent or approximately equivalent for the two main alternatives. With this extra information, 86% of all side-chain amide groups could be oriented quite unambiguously. In the absence of further experimental data, it would probably be inappropriate to assign many more than this. Some of the remaining 14% are ambiguous because of coordinate error or inadequacy of the theoretical model, but the great majority of ambiguous cases probably occur as a dynamic mix of both flip states in the actual protein molecule. The software and the 100 coordinate files with all H atoms added and optimized and with amide flips corrected are publicly available.

© 1999 Academic Press

Keywords: side-chain amide orientation; hydrogen atom placement; Asn/Gln flips; hydrogen bond network; small-probe contact dots

*Corresponding author

Abbreviation used: PDB, Protein Data Bank.

E-mail address of the corresponding author: dcr@kinemage.biochem.duke.edu

0022-2836/99/041735–13 $30.00/0

© 1999 Academic Press

Introduction

Correct assignments of the NH_2 *versus* the O branches of Asn and Gln side-chain amide groups are a relatively minor part of a protein structure determination, but they can be quite important if the residue is involved in H-bonding at the active site, or if one wants to analyze bound water molecules, H-bond networks, or detailed electrostatics. However, such assignments have been considered difficult, and sometimes are not even attempted. The Protein Data Bank (PDB; Bernstein *et al*., 1977) format actually provides a special ambiguous atom designator of A, meaning ''either N or O'', for use by crystallographers who want to keep the uncertainty explicit. Finding the NH_2 hydrogen atoms or telling apart the N and O atoms by direct observation in the electron-density map is not possible except at extremely high resolution. The distinction, therefore, is almost always made on the indirect evidence of H-bonding possibilities, usually done by inspection as part of the model-fitting process.

The most secure assignments are when the environment includes H-bonding groups that are either obligate donors such as a peptide NH group (which can interact only with the O branch of the amide) or obligate acceptors such as a carboxyl group (which can interact only with the NH_2 branch). In the frequent cases where the surrounding groups are ambiguous donors or acceptors (OH, His, other amide groups, or water molecules), assignment involves analyzing the entire local network of H-bonds. The energy terms in refinement protocols are capable, in principle, of favoring the amide orientation with the best H-bonding, but in practice the 180° flip between the two orientations is a large enough step that minimization will always, and molecular dynamics sometimes, be trapped in one of the local minima. Even after full analysis, many H-bond networks have two equally favorable solutions that involve concerted exchange of all donors and acceptors, many Asn or Gln amide groups are undetermined because they are exposed at the surface, and some make only non-polar interactions. Histidine residues also have a similar assignment problem, since a 180° flip of the imidazole ring exchanges N and C atoms in the δ and ε positions. In place of the amide ambiguities of H-bond donor *versus* acceptor, for His the choice is a more drastic one between a polar, or even charged, NH and a CH with only very weak H-bonding potential; however, imidazole orientation can also be ambiguous.

Several automated methods have been developed to help deal with this problem. HBPLUS (McDonald & Thornton, 1994) tries the flip states of each Asn, Gln, and His, and chooses the alternative that minimizes unsatisfied buried H-bonding groups, dividing the prior Asn/Gln/His orientations into highly favored, slightly favored, indifferent, slightly suspect, and highly suspect; however, it does not deal with pairs or larger interacting groups. NETWORK (Bass *et al*., 1992) analyzes H-bond networks to optimize polar H placement, but does not allow for amide or imidazole flips. WhatIf (Hooft *et al*., 1996) deals with both aspects of the problem, including even the assignment of H positions for all crystallographically located water molecules with occupancy >0.5; it builds in crystal symmetry, and has a penalty bias against flips in marginal cases. Inclusion of the water hydrogen atoms makes the combinatorial problem so huge that it cannot possibly be treated exhaustively, so it is done by a variant of simulated annealing. WhatIf does a thorough and careful job of analyzing the H-bond networks, coming out with a decision for all the ambiguous polar groups; using it would improve assignments for the majority of structures. This feature is just one part of an overall package with many other valuable functionalities. Its disadvantages for amide assignment are that it relies strongly on placement of water molecules, which are the least reliable feature in macromolecular structures; its output is not convenient, and its answers cannot be critically evaluated because there are no estimates of confidence and the reasons for its choices are well hidden inside a complex, stochastic process.

We are now revisiting this problem because our small-probe contact dot methodology (Word *et al*., 1999) uncovers a source of independent new information from the analysis of van der Waals clashes for explicit H atoms, so that these decisions become less complex and much less subtle. The reasons for a given choice can easily be expressed both as numerical scores and in a visual display that explicitly shows all relevant positive and negative interactions (e.g. Figure 1), so that the user can easily evaluate confidence levels for a given choice. The method is applied here to optimizing the H-bond networks and assigning Asn, Gln, and His flips in a set of very high-resolution crystal structures.

Results

Individual Asn/Gln examples

The basic point of this new approach is that the van der Waals interactions of polar H atoms are crucial to ruling out incorrect amide orientations, even if they are not necessary for evaluating the energy of correct hydrogen bonding (see Procedures). The NH hydrogen extends 0.6 Å farther than the bare oxygen, which alters the van der Waals interactions so drastically that these amide flip choices generally become blatantly evident rather than subtle, once hydrogen atoms are added by the program Reduce and contacts are scored and examined with small-probe dots generated by the program Probe (see Procedures). Figure 2 shows one of the many really obvious cases, which has excellent H-bonding in the correct orientation and extreme van der Waals clashes in the incorrect

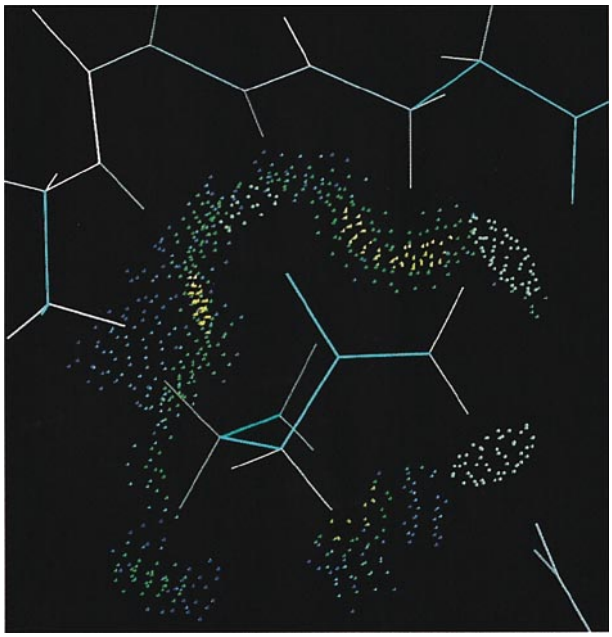

Figure 1. Small-probe (radius 0.25 Å) contact dots around Gln71 of cutinase (1CUS), colored by contact gap and including favorable van der Waals contacts (green and blue dots) as well as H-bonds (pale green dots), slight overlaps (short yellow spikes), and clashes (orange or red spikes; none present).

orientation; the un-normalized score comparison (see Procedures) is +0.3 *versus* −7.0. This Gln could be assigned correctly by any person or computer program explicitly considering it, since the H-bonding is completely unambiguous. However, some cases this obvious were found to be misassigned in our reference dataset, including some in proteins refined using molecular dynamics and even a few in structures at atomic resolution. This particular example, Gln90 from the immunoglobulin V_L dimer 1REI at 2.0 Å resolution, used "A" atom designations to leave the amide assignment explicitly ambiguous.

Many other cases cannot be determined by the H-bonds alone, but are unambiguous if the van der Waals interactions are included. Figure 3 shows an example (from ribonuclease F1) which has no H-bonds, but where the non-polar interactions accommodate one amide orientation very nicely but not the other. In one flip state the NH_2 group of Gln57 nestles neatly against the backbone, while in the other flip state it collides with a proline side-chain. The score comparison is +1.0 *versus* −1.4.

Figure 4 illustrates a pair of doubly H-bonded side-chain amide groups from the fungal peroxidase 1ARU, for which two of the four possible orientations are equally good if only H-bonding is considered. Such situations are rather common, either for pairs or for larger H-bond networks, in which switching all donors and acceptors in unison can produce equally good H-bonding. However, as in Figure 4, van der Waals clashes usually rule out one of the two best H-bond possibilities: in this case, the original assignment in the coordinate file has good H-bonds but bad clashes of both amide

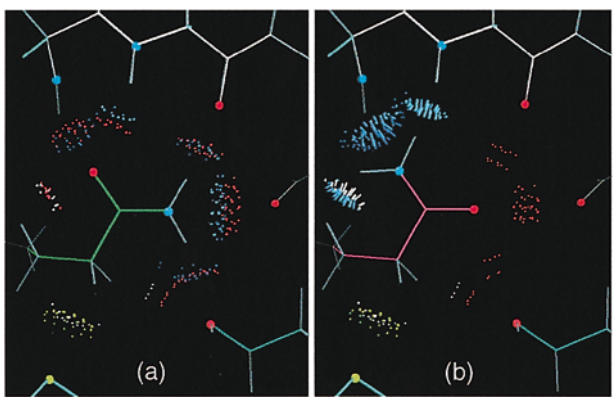

Figure 2. (a) and (b) Amide flip comparison for Gln90 from the immunoglobulin V_L dimer of 1REI (Epp *et al.*, 1975), colored by atom type (O, red; N, blue; C, white) and with clashes emphasized by spikes. There are three H-bonds and no clashes in the correct flip position (a) *versus* no H-bonds and three serious clashes of the NH_2 hydrogen atoms in the incorrect flip position (b). The probe radius is 0.25 Å.

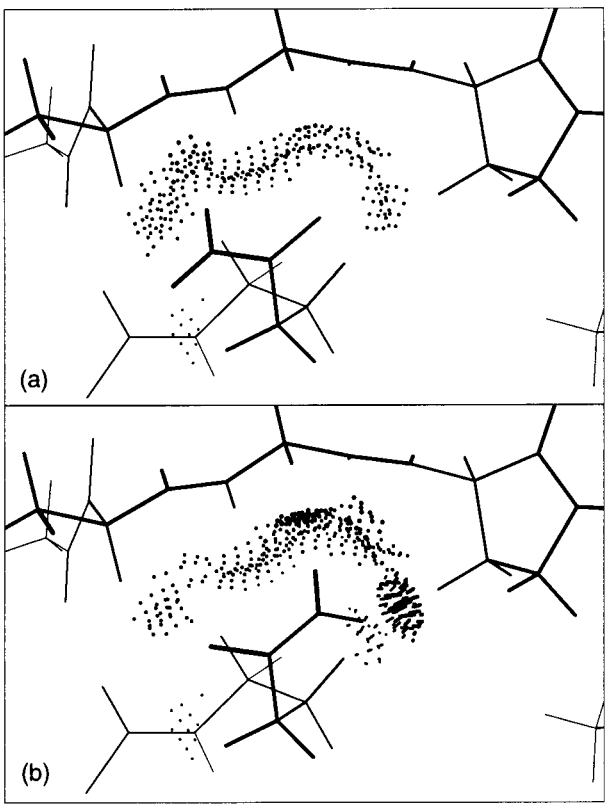

Figure 3. (a) and (b) Amide flip comparison for Gln57 in the 1FUS ribonuclease F1 (Vassylyev *et al.*, 1993), which has no H-bonding but whose orientation is unambiguously determined by the NH_2 clashes in the position of (b), mainly with the H^{β} of a Pro side-chain. In contrast, it fits well against the backbone in (a). Probe radius is 0.25 Å.

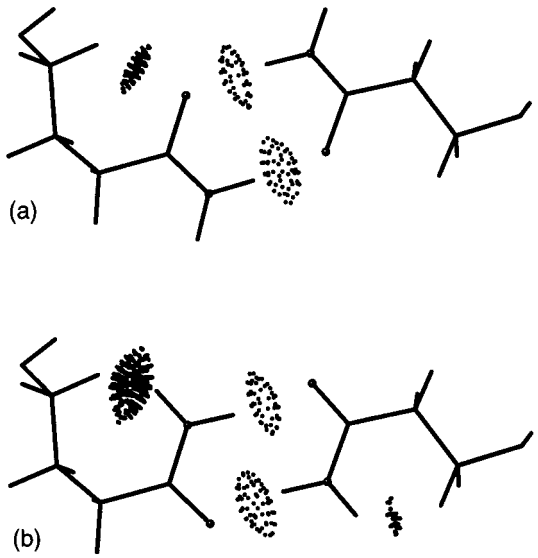

Figure 4. (a) and (b) A double amide flip of the Asn128-Gln34 pair in the fungal peroxidase 1ARU (Flory, 1969) that cannot be resolved just by H-bonding. In the incorrect double flip (b), there is a very bad clash of the Gln NH_2 with H^{α} and a smaller clash of the Asn NH_2, whereas the amide flip state shown in (a) is accommodated well. There is also a shear offset between the amide groups that puts the two NH groups at a further, and more favorable, distance in (a). Here and in Figures 6 and 8, the contact dots are simplified by showing only the H-bonds and the overlaps, not the attractive van der Waals contacts.

NH_2 groups with their own H^{α} atoms, while flipping both amide groups gives very much better van der Waals contacts and equally good H-bonding. The score comparison is -0.4 *versus* $+3.6$ for the original and double-flip states, respectively (compared with -5.0 and -6.7 for the two single-flip states). A slight shear offset between the two side-chains is visible in Figure 4, which puts the two polar H atoms further apart (2.5 Å) in the better orientation. That does not necessarily mean that the amide H radius needs to be larger than 1.0 Å, however, because the two polar H atoms are presumably shifted apart by electrostatic as well as van der Waals repulsion. Unfortunately, our dataset cannot calibrate the consistency of such a shear offset, because it happens that each of the three other doubly H-bonded Asn/Gln pairs has one of the amide groups incorrectly oriented by our criteria, which seems to have caused refinement to slightly distort the interactions.

Systematic surveys

The reference datasets for 100 proteins contain 1554 unique Asn and Gln residues, 1539 of which have no missing atoms, metal ligation, or covalent modifications. To provide a cross-check on this new methodology (and to identify unusual situations that could be handled by improving the algorithm), the Asn and Gln residues were systematically surveyed three times, each time by a different combination of contact score comparison (see Procedures) and inspection of their small-probe contact dots in the Mage display program. The first time through, each Asn or Gln with $B < 40$ was examined if its individual flipped score was not clearly worse than its original score; H-bond interactions between multiple residues were assessed visually. The above process resulted in flipping 252 Asn/Gln residues (17 % of the total) in 71 of the files. For several files the Asn/Gln flip rate was approximately random (near 50 %), implying that the amide groups had not been examined (note: 451C (Matsuura *et al.*, 1982) uses the ambiguous ''A'' atom designations).

For the second-round survey, an algorithm was developed to analyze full H-bond networks automatically, as part of the Reduce program. The orientations of 47 His, 17 Cys, ten OH, nine Asn, and three Gln residues are fixed by metal ligation, and three Asn and several Cys, Ser, and Tyr residues by various other covalent modifications; the Asn/Gln modifications are listed in Table 1. This leaves 6548 unique movable-H groups, including Asn, Gln and His residues, OH, SH, NH_3^+ and Met methyl groups, and similar functionalities in the small-molecule ''heterogens'' (see the accompanying paper to explain why methyl groups are rotated only in Met side-chains). Those groups were then partitioned into closed sets of local interacting networks or ''cliques'' (see Procedures). Of the movable groups, 5050 were found to be isolated from any other; there are 557 interacting pairs, 94 triples, 14 cliques of four, eight cliques of five, one clique of six, and no larger groupings.

The clique score (using both favorable H-bonds and unfavorable overlaps, including a simple model for interaction with the crystallographic water molecules) is evaluated for all combinations of possible H atom positions, in order to choose the optimal arrangement. Most movable groups have either two, four, or six potential H atom positions, while we restrict an OH or SH to something between two and as many as 18 in the most crowded environments (see Procedures). Since the largest clique found had only six members, an exhaustive search is computationally tractable: it takes three hours to do all 100 proteins on an R10000 SGI Indigo II.

For each residue in a clique, the best total clique score and the best conformation are reported plus, for Asn/Gln/His, the best total clique score found with that residue in the opposite flip state. This score comparison directly shows how sensitive, or well-determined, the flip state is for that specific residue. The H atom positions for the best-scoring arrangement are added to the output PDB format coordinate file, and N *versus* O or N *versus* C identities are switched where indicated for Asn/Gln/His flips. Agreement with decisions made in the first-round survey was used to optimize the value of a penalty against changing the depositor-

Table 1. Side-chain amide flips of Asn and Gln (round 3)

PDB	N + Q[a]	Fixed[b]	Keep	Clash[c]	Unk.[d]	Flip
1aac	3		3			
1ads	27		15		5	7
1aky	20		14		5	1
1amm	16		9	1		6
1arb	26		16		1	9
1aru	26	1CHO	15		4	6
1benAB	6		4		1	1
1bkf	6		2		2	2
1bpi	4		4			
1cem	37		24		9	4
1cka	4		2		1	1
1cnr	3		3			
1cnv	25		15		4	6
1cpcB	12	1 CH3	3	1	2	5
1cse	33	2 Ca	21		3	7
1ctj	10		6		1	3
1cus	15		9		3	3
1dad	18		9		4	5
1dif	9		5		3	1
1edmB	5	1 Ca	4			
1etm	1		1			
1ezm	27		9	1	4	13
1fnc	17		12		3	2
1fus	14		11		3	
1fxd	3		3			
1hfc	14		8	1	1	4
1ifc	12		8			4
1igd	4		3			1
1iro	4		4			
1isuA	7		6		1	
1jbc	19	1 Ca	12	1	1	4
1kap	53	1 Ca	44		2	6
1knb	20		10		4	6
1lam	36		24		1	11
1lit	12		6		4	2
1lkk	10		5	1	3	1
1lucB	34		19		3	12
1mctI	0					
1mla	25		12		3	10
1mrj	29		19		4	6
1nfp	30		18	1	7	4
1nif	21		17		1	3
1not	1		1			
1osa	11	2 Ca	4		1	4
1phb	34		12	1	8	13
1php	15		8		3	4
1plc	7		7			
1poa	14		9		2	3
1ptf	6		3			3
1ptx	5		4			1
1ra9	9		4		1	4
1rcf	17		13		2	2
1rgeA	7		5	1		1
1rie	6		5		1	
1rro	11	1 Ca	4		2	4
1sgpI	5		3		1	1
1smd	52	1 Ca	42		4	5
1snc	10		6		3	1
1sriA	10		6		1	3
1tca	32	1 CHO	28			3
1ttaA	3		2		1	
1whi	6		4			2
1xic	21		14	1	2	4
1xsoA	12		9	1		2
1xyzA	45		27	1	7	10
256bA	12		8			4
2ayh	23		15	1	2	5
2bopA	11	1 Yb	5	2	2	1
2cba	21		16		3	2
2ccyA	8		6		2	
2cpl	12		12			
2ctc	25		16	2	1	6
2end	9		8			1
2er7	15		4		2	9
2erl	3		1		2	
2hft	21		16		3	2
2ihl	16		12		2	2
2mcm	7		6		1	
2mhr	6		5		1	
2msbA	10	2 Ca	6			2
2olb	54		46	1	4	3
2phy	11		8			3
2rhe	8		7		1	
2rn2	15		8	1		6
2trxA	7		7			
3b5c	5		5			
3chy	10		8			2
3ebx	7		4		1	2
3grs	25		15	1	3	6
3lzm	17		7		7	3
3pte	34		23		3	8
3sdhA	16		11			5
451c	8		2		1	5
4fgf	9		8			1
4ptp	26		13		6	7
5p21	15		7		8	
7rsa	17		14		3	
8abp	20		17		2	1
bio1rpo	5		3		2	
bio2wrp	10		3		1	6
Totals	1554	15	1006	20	195	318

[a] Total number of Asn + Gln.
[b] Number of amide orientations fixed by covalent modifications: by metals (Ca or Yb), carbohydrates (CHO), or methylation (CH_3).
[c] Number with severe clashes (overlap $\geqslant$0.4 Å) in both orientations.
[d] Number classified as ''Unknown'' (score difference <0.5), even after individual examination.

assigned flip state; the penalty was set at 0.5, which means that the score difference must favor the flipped state by at least 0.5 or Reduce would not assign a flip. Any flip state for which an atom in the movable group has a serious clash (overlap $\geqslant$ 0.4 Å) is flagged with a !. If both the best state and the best flip state have clashes, then non-H atoms must be badly placed and *B*-factors are usually high on at least one side of the clash. If it is not practical to evaluate these cases individually, then they should be omitted from any further analysis. For the automated algorithm, therefore, we have adopted the conservative policy of not assigning any flips for these double-clash cases.

The result of the automated analysis of round 2 is 100 coordinate files with all H atoms added and optimized, and all changes and assignments documented in the file headers, plus contact dot kinemages which animate between the two flip states, one set for Asn/Gln and another set for His side-chains. Out of the 1554 Asn + Gln side-chains, Reduce flipped 290, or 19 %, of them (see Figure 5). The rest were all left in their original orientation, including 49 with bad clashes both ways (3 %) and 314 with small score differences between −0.5 and +0.5. Out of 379 His side-chains, 30 were flipped (8 %), 27 had bad clashes both ways (7 %), and 49 had small score differences. Only 17 Asn, 29 Gln, and 12 His (3 % of the total) residues are so com-

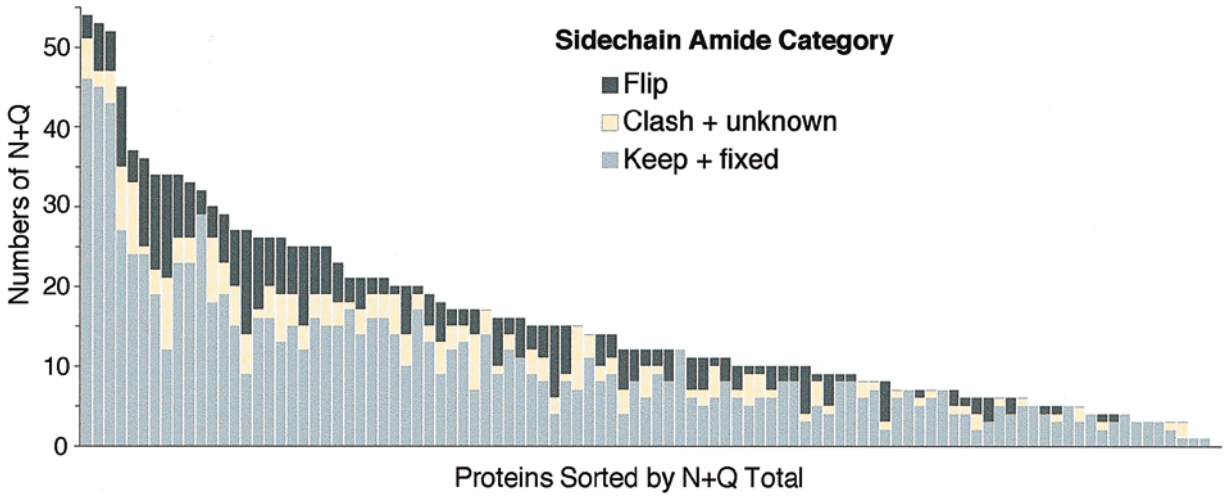

Figure 5. Categories of side-chain amide assignments for each of the 100 proteins, sorted in decreasing order of total Asn + Gln residues. Here, the ''Keep + fixed'' category includes the few fixed by covalent modifications, and the ''Clash + unknown'' category includes both the ''C'' (double-clash) and ''X'' (low-score difference) groups. N stands for Asn, Q for Gln.

pletely exposed that they had scores of exactly zero in both orientations.

As an example of Reduce's automated analysis of an H-bond network, Figure 6 shows the two principal alternatives for the linear Asn138-His123-His131 clique in cellulase (correctly assigned in the PDB file 1CEM); there are two states for the Asn and six for each His, giving a total of 72 possibilities. There is a water molecule at each end of the network, and internally all H-bond acceptors and donors can be switched, giving four equivalent H-bonds in the two best states. When H atom clashes are considered, however, they occur in (and unambiguously disfavor) only one of these two overall flip states: the score comparison is +5.4 *versus* −0.7.

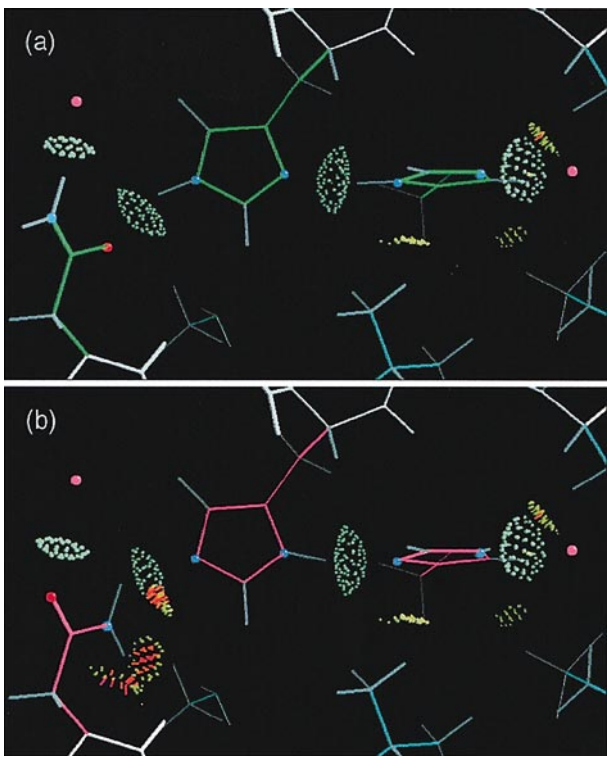

Figure 6. (a) Correct *versus* (b) flipped arrangements of the Asn138-His123-His131 H-bonding network from the 1CEM cellulase (Alzari *et al.*, 1996). All four H-bonds are equivalent in the two forms, which are distinguished by clashes of the Asn NH_2 group in (b). van der Waals contact dots are not shown.

Because this is a new method, and because we want these modified coordinate files to be a reliable basis for future analyses, we undertook a third-round survey in which both flip alternatives and their contact dots were visually examined in Mage for all of the Asn/Gln/His residues that the automated algorithm had flipped, for all of the small number of cases where Reduce (round 2) disagreed with the round 1 assignments, and for all cases in a subset of 20 files. Out of 290 amide flips recommended by Reduce, only 15 were rejected in round 3 (5% of the flips, or 1% of all Asn/Gln amides), of which 12 were declared ambiguous and only three as clear ''Keep'' residues.

During this process, it became obvious that many of the double-clash cases and the marginal cases with low score differences could actually also be assigned an orientation with confidence. Therefore, round 3 was expanded to include visual examination of all Asn/Gln/His in those two categories. Most of the resulting reassignments involve confirming the orientation indicated by the flip scores: for example, Gln61 of 1MLA, malonyl CoA carrier protein (Serre *et al.*, 1995), with a score difference of 0.48 was promoted from marginal to flipped, because it can make weak H-bonds both to backbone and to a Glu O^{ε} in the preferable flipped orientation. However, there are a small but significant number of cases where the visual assignment contradicts the direction of the score

difference. Sometimes these involve factors that are not yet included in the algorithm but which could be (such as the influence of a charged group slightly too far away to score as an H-bond), while sometimes the analysis involves judging the relative probability of different types of errors in ways it is hard to imagine automating. As an example of the latter sort, Gln260 of 1ARU (fungal peroxidase) has a score difference of -1.0 *versus* $+0.1$ and a bad clash of the NH_2 with C^{δ} of Glu325 in the original orientation, yet our judgment agrees unambiguously with the original crystallographic assignment: the low *B*-factor Gln260 in its flipped orientation has only a water H-bond and O^{ε} is very close to two other oxygen atoms, while the original orientation adds a backbone CO H-bond and improves contacts, and a minor rotation of the high *B*-factor Glu325 could convert the clash into an H-bond with the Glu O^{ε}. Such cases should emphasize the point that although the automatic algorithm does a very good job, the most reliable assignments combine the automated analysis with visual inspection.

Summary of final assignments

The third-round survey results in five categories of Asn/Gln residues summarized in Table 1. First is the small subset whose orientation is fixed by metal ligation or covalent modification. The largest category by far (65% of the total) are the 1006 Asn/Gln, with a clear, unambiguous flip assignment that agrees with their identification in the deposited PDB file. Then there are the side-chain amide groups which unambiguously require flipping: 318 Asn/Gln residues (20.5%) in 73 of the 100 files. For 20 of the Asn/Gln side-chains (1%) whose movable group has a severe clash in both orientations, either they or a neighboring group is positioned incorrectly in such an arrangement that it is unclear which would be the correct amide orientation once the problem was fixed. The fifth, final category are 195 still-ambiguous Asn/Gln cases (12.5%), mostly with small score differences ($-0.5 \leqslant D < 0.5$). All examples in the two ambiguous categories are left in their original orientation.

For each of the 100 proteins, sorted by total Asn + Gln residues, Figure 5 plots the number of ''Keep + fixed'' Asn/Gln (in light gray), the number of ''Clash + unknown'' (in cream), and the number of ''Flip'' Asn/Gln (in dark gray). Twenty-seven files had no flips and four files had 50% or more flips, but overall the distribution is relatively uniform. Somewhat surprisingly, the percentage of amide flips shows no significant relation with resolution. However, the percentage of flips does show a small but significant ($p = 0.017$) positive relation to the size of the protein, perhaps reflecting time limitations for evaluating correct amide orientation as protein size increases.

For histidine side-chains, there are 47 fixed by metal ligation, 250.5 which unambiguously should be kept in their original orientation (66%), 37.5 which unambiguously require flipping (only 10%), 13 with unresolvable clashes both ways (3.4%), and 31 ambiguous cases with small score differences (8.2%). The non-integral values occur at a dimer interface, as discussed below. Histidine residues show fewer flips than Asn/Gln amide groups, but more often have unresolvable clashes. Those unresolved clashes are almost all with O atoms and, therefore, may be cases of CH···O H-bonding (Derewenda *et al*., 1995) in His rings.

The marginal Asn/Gln/His cases with a small score difference include: (1) completely exposed side-chains with no neighboring atoms or self clashes; (2) H-bond networks with no external clashes to any alternative H positions and nearly equal scores when donors and acceptors are all switched; (3) H-bond networks across symmetrical dimer interfaces; and (4) conflicts in which each flip state has a favorable interaction incompatible with the other state. Some of these examples make it clear that a flip state can genuinely alternate between two possibilities. For the ROP protein dimer-interface clique of SerA42a-HisA46-HisB46-SerB42b in 1RPO (Vlassi *et al*., 1994), both the Ser alternate conformation and the His flip state must differ across the two subunits; 1RPO His46 shows up as a half-integer value in the His assignments above, because the two interacting copies must be positioned asymmetrically, with one flipped and the other not. Figure 7 shows the two flip states of His54 in 1PTX (scorpion toxin), each of which has one good H-bond to a water molecule which clashes in the other flip state. Allowance for the positive effect of CH···O H-bonding would decrease the severity of the clashes and allow a water molecule to stay (a bit farther out) when the ring flipped. However, this histidine residue almost certainly occurs in a mixture of the two orientations.

The cases originally assigned as double-clash by Reduce predominantly consist of side-chains that are themselves well defined but are bumped by another slightly mispositioned group, often with high *B*-factors: for example, the double-clash Asn366 in 2OLB (oligopeptide-binding protein; Tame *et al*., 1995) overlaps H^{δ} of Arg413, but both from the scores of $+0.2$! *versus* -8.8!, and from our visual examination, it is clear that four good H-bonds in the original orientation are clearly preferable to just one in the flipped orientation. Such cases were reassigned in round 3.

Three of the Asn double-clash examples are consistent with the possibility of chemical de-amidation of the side-chain: 1CSE Asn158E (subtilisin; Bode *et al*., 1987), 1HFC Asn206 (collagenase; Spurlino *et al*., 1994), and 1LUC Asn12 (luciferase; Fisher *et al*., 1996), which is shown in Figure 8. The tight interactions around Asn12, including two H-bonds to backbone NH groups, definitely require an O^{δ} for the left-hand branch of the amide. There is some room around the arginine guanidinium group, and in the unmodified protein it would need to move farther to the right, away

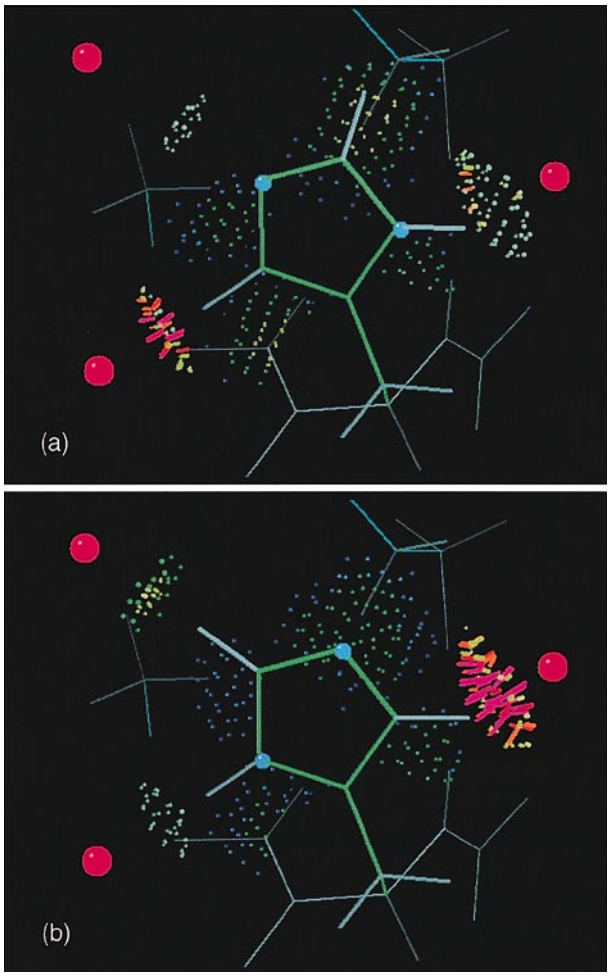

Figure 7. (a) and (b) Evidence for dynamic equilibrium in the flip state of His54 in the scorpion toxin 1PTX (Housset *et al.*, 1994). Although this His ring has good *B*-factors and makes good contact with the structure behind it, each possible ring flip position makes an N^{δ} H-bond to a well-ordered water molecule that clashes with the other flip state. Although those clashes can be partly mitigated by considering CH···O H-bonds, presumably His54 occupies both conformations. The probe radius is 0.25 Å.

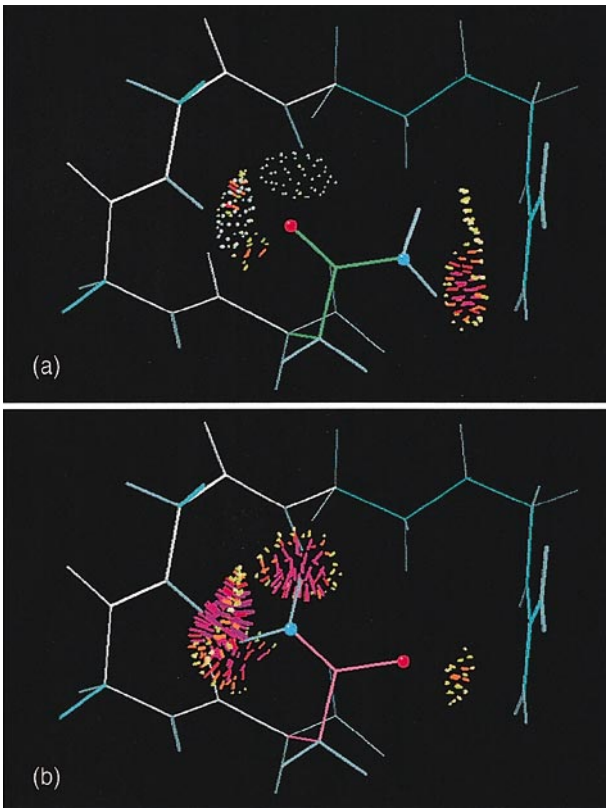

Figure 8. (a) and (b) Asn12 of 1LUCb luciferase (Fisher *et al.*, 1996), an example whose interactions are consistent with possible side-chain deamidation. The left-hand side is clearly compatible only with an O^{δ} that can H-bond tightly with two backbone NH groups as in (a), rather than an N^{δ} that would clash impossibly as in (b). The fact that the Arg guanidinium group was observed in this close position also implies an O^{δ} on the right-hand branch of the Asn, both for steric and for electrostatic compatibility.

from what would then be the NH_2 group of Asn. However, the fact that the Arg was observed this close to the side-chain of residue 12 in the crystal structure provides circumstantial evidence that Asn12 may have become de-amidated to Asp.

Overall, less than 14% of all the Asn/Gln amide orientations in the 100 proteins remain undetermined here, once H atom van der Waals interactions are considered. Some of those ambiguous cases represent our inadequate level of knowledge, but most of them probably indeed occur in the protein molecule as a mixture of both orientations.

Discussion

The present analysis departs from common practice in two main ways: the use of small-probe contact dots for explicit visualization and quantification of molecular interactions, and the placement of all hydrogen atoms, both polar and non-polar, and inclusion of their van der Waals as well as H-bonding contributions. That additional information made the process of orienting side-chain amide groups much more straightforward and more often definitive than the H-bonding analyses used in previous work. H-optimized coordinate files for the 100 high-resolution proteins of our dataset are now available for further structural analysis. The programs Reduce, Probe, and Mage are available for adding and optimizing H atoms, for analyzing the contacts in known macromolecular structures, and to help in the determination of new structures.

It appears that most side-chain amide groups do indeed have surroundings in the equilibrium protein structure that enforce a unique, and readily identifiable, amide orientation. Assigning those orientations correctly will help in the details of refinement and the identification of water molecules in crystal structure determinations, and will

aid in analyses of hydrogen bonding, water structure, side-chain conformations, and ligand binding. In particular, the discovery of serious internal clashes in Asn/Gln side-chains as they are fit even in the highly accurate structures of our dataset (e.g. the severe Gln H^{ε}-H^{α} clash in Figure 4(b)) implies that there are problems with existing side-chain rotamer libraries. We plan to use the methods and datasets described here to address that issue.

Even using H-bonds and van der Waals interactions, there still remain between 10 and 15% of the side-chain amide groups (and also of the His rings) whose orientation is ambiguous. A few of these cases are due to unresolved problems in the coordinates, and there would be somewhat more such cases in lower-resolution structures. However, at least 10% of these side-chains are probably in dynamic equilibrium in the actual protein molecules, such that both flip states would be significantly populated. For some of the unassigned cases (e.g. Figure 7) the amide or ring plane is well defined but the flip is not, while for some of the fully exposed, high *B*-factor cases (more common for Gln than Asn) even the plane orientation is probably dynamic. Thus we feel that the 14% of side-chain amide groups left unassigned here mainly represent not a failure of the method, but a successful identification of the cases that should not be assigned. It also follows that such unassignable cases should be omitted or treated separately in any statistical analyses of side-chain conformations or interactions. To this end, we flag these cases in the headers of the modified PDB files.

Here, we have tried to start out with a simple and straightforward model and add complications only when we are convinced of their necessity and are also sure that they will contribute in the right direction even for the pathological cases caused by occasional coordinate errors. Our initial aim was simply adding hydrogen atoms in order to use contact dots to quantitatively analyze interior packing in proteins. It was immediately obvious this could not be done without first correcting side-chain amide flips, which led to the study described here. In addition, both the ring flip and protonation state of histidine residues had to be considered, although that treatment was developed only far enough to avoid incorrect His influence on Asn/Gln orientations, since a complete analysis of His protonation equilibria would require detailed electrostatics and knowledge of the pH.

The major, completely necessary, complication in the present method is of course the combinatorial analysis of the H-bond network cliques. The tractability of the clique analysis depends, in turn, upon keeping several other aspects of the model simple: Met methyl groups are the only ones rotated; the probe radius is set to zero so that only H-bond and overlap terms enter the combinatorial search; and, most importantly, interactions with crystallographically located water molecules are included but water-water interactions are not. Inclusion of water molecules is crucial for success of the algorithm, but a simplified model that treats their possible H-bonds as completely independent has worked quite well. We preferred not to attempt explicit placement of hydrogen atoms on the water molecules, because the errors in position or occupancy for a significant fraction of water molecules (for example, those that are impossibly close to side-chain atoms) could often produce problems that would propagate through the network of water molecules. Interactions across crystal contacts were also deliberately omitted, because we are more interested in what can be learned about the molecular structure than in the crystal structure for its own sake. There are, of course, many efficient search algorithms that could be applied to optimizing the clique scores. However, since the simplest and most guaranteed method of complete enumeration is indeed fast enough for the actual cases found, it is the preferred choice.

The contact-dot analysis, in general, is based on geometry and atom types, rather than on energies. In particular, its treatment of electrostatics is indirect and, so far, only short range: hydrogen-bonding interactions based on degree of atomic overlap, with charged H-bonds stronger only because they can overlap further. Certainly those short-range H-bonds are the most dominant single factor affecting amide orientation. As a future modification, we could incorporate a weighting scheme for the non-overlap contact dots based on atom types, in order to add mid-range electrostatic preferences for distances between grazing contact and a probe diameter further out. We do not intend to add long-range electrostatics, however. It would not combine easily with the geometrical terms, and any simple, dielectric-based treatment is likely to overestimate the contribution.

Although adding and optimizing hydrogen atoms and determining side-chain amide orientations is an apparently simple set of tasks, the number and detail of considerations involved and the variety of unexpected special cases that occur in 100 protein structures are very large. Therefore, the automated algorithm in Reduce is gradually developing into an expert system. Fortunately, for the most part those developments make it more robust and easier to use.

Procedures

In exchanging the two possible flip orientations for an Asn or Gln amide, we exchange the identities of the N and the O atoms rather than doing a bond rotation of 180°. These two methods give slightly different results, because the bond lengths and angles are not identical for the N and O branches. In an electron density map, the positions of the amide N/O atoms are more precisely known than that of the central C atom, so that this method seems more conservative. Similarly, to flip a His ring, we exchange identities of the δ and ε C and N atoms; then the assigned ring nitrogen atoms are considered in three protonation states (see details below).

Definition of contact dots

Contact dot surfaces are loosely related to the Connolly algorithm for calculating solvent-accessible surfaces on the outside of a molecule (Connolly, 1983), in that a spherical probe is rolled around the van der Waals surface of each atom, visiting each of a set of predefined points, and a dot is drawn if certain tests are satisfied in that position. The differences are that the contact dot algorithm, as implemented by the program Probe and for some calculations in Reduce, uses a very small probe (typically 0.25 Å in radius or smaller, rather than the 1.4 Å radius used for Connolly surfaces) and leaves a dot when the probe does touch another ''not-covalently-bonded'' atom, rather than when it does not touch another atom. Small-probe dots form discontinuous surfaces, the patches of which directly show the location, extent, and shape of close atomic contacts (e.g. Figure 1).

One color scheme used here for these contact dots (e.g. Figure 2) reflects atom type: C, white; N, blue; O red; S, yellow; and H in the color of its bonded heavy atom. The NH···O hydrogen bonds show as interpenetrating lens shapes in red and blue. Overlapped van der Waals shells of non-polar atoms are emphasized by showing spikes instead of dots: a spike is a line drawn from the dot position to the contact midplane, along the atom radius. An alternative color scheme (as in Figure 1 and Figures 6 to 8) reflects the gap distance between atoms at each dot position: green or yellow for good contact (greens for narrow gaps, yellows for slight overlaps), pale green dots for H-bonds, blues for wider gaps (>0.25 Å), orange or red spikes for unfavorable interpenetrations, or clashes, and hot pink spikes for severe clashes $\geqslant$0.4 Å overlap.

Contact dots or spikes are output by Probe as a simple text file of dot lists or vector lists in kinemage format (Mime standard: chemical/x-kinemage), with color, source atom, and contact type specified. They are shown in the Mage display program (Richardson & Richardson, 1992, 1994), which supports alternate color schemes, atom or dot identification by picking, turning on or off groups by atom type or by contacts *versus* clashes *versus* H-bonds, saving many local views within a large structure, and animating between different forms. Probe can also format output for display as graphics objects in O (Jones *et al*., 1991) or in XtalView (McRee, 1993), so that contact dots can be used to help rebuild models in electron density maps.

Adding H atoms

Small-probe contact dots require the use of all explicit H atoms. The program Reduce adds them to PDB-format coordinate files, using local geometry. Methyl hydrogen atoms are added in staggered positions, and only the Met methyl groups are rotationally optimized. OH, SH, and NH_3^+ hydrogen atoms are rotationally optimized and His protonation assigned as part of the H-bond network analysis described below. Water molecules are treated by presuming that they can always orient so as to present whatever is needed for each interaction. If water molecules are extremely close, we restrict their H-bonding score to a reasonable value. The details of these procedures, the choice of parameters for bond lengths and van der Waals radii, and further details of the contact-dot methodology are explained in the accompanying paper (Word *et al*., 1999).

Especially in the context of evaluating amide orientations, we must justify using any van der Waals terms at all for polar H atoms, since they are set to zero in many energy calculations. For instance, Hagler *et al*. (1974) optimized force-field parameters to agree with crystal dimensions, vaporization energies, etc. for a variety of small molecules with H-bonded amide groups, reaching the conclusion that van der Waals terms from the polar H atoms do not make a significant contribution beyond what can be fit with only electrostatic terms and van der Waals interactions for the heavy atoms. Similarly, Berendsen *et al*. (1981) found polar H van der Waals terms unnecessary for obtaining a satisfactory model of water interactions. Those conclusions are undoubtedly justified for the cases analyzed, where all amide or water interactions are through H-bonds. However, such studies do not address the issue of what parameters are needed in the less typical but still fairly common cases where non-polar groups contact amide H atoms. In our current analysis, such parameters are absolutely essential for evaluating the counterfactual non-polar-to-amide clashes that are characteristic of wrong flip choices for side-chain amide groups. Our approach returns to application of very simple physical principles: as stated for instance in the classic crystallographic text by Stout & Jensen (1968), ''Any postulated arrangement of atoms…must fulfil the simple steric requirement that no two atoms should approach closer than the sum of their van der Waals radii unless they are bonded together.''

In order to decide the best polar H radius for use in calculating contact dots, we have analyzed the spacings actually seen for non-polar-to-amide contacts and then compared the proposed radii to the shape of electron-density distributions calculated by quantum mechanics. Figure 9(a) shows calculated electron density contours for an NH group (Bader *et al*., 1967), the nearly round shape of which has been used to argue for a negligible effect of the polar H van der Waals terms (Hagler *et al*., 1974). However, equivalent contours lie about 0.5 Å farther out in the H direction than elsewhere, a difference that can be quite crucial in tightly packed regions. Shown overlaid in Figure 9(a) are the simple radii of 1.55 Å for N and 1.0 Å for the polar H, which fit the shape of the electron-density contours almost perfectly. For comparison, Figure 9(b) shows the equivalent overlay for the non-polar CH group; standard radii fit the calculated electron density equally well in both cases. The difference in contour level matched is within the uncertainty in our present parameters and has the advantage of keeping the NH radii conservatively on the small side.

Scoring

Quantitative measures for goodness-of-fit are then defined in ways that seek to capture the insights gained from the contact-dot visual representation of packing interactions. We have found three levels of scoring to be useful in analyzing protein structures. As in the standard definition of van der Waals energies, our general scoring system is a sum of competing terms, but the contact scores are evaluated per dot, not per atom pair, and are then summed. Hydrogen bonds and other overlaps are quantified by the volume of overlap. Each non-overlapped contact dot is counted with an error-function weight of:

$$w(\mathrm{gap}) = \mathrm{e}^{-\left(\frac{gap}{err}\right)^2}$$

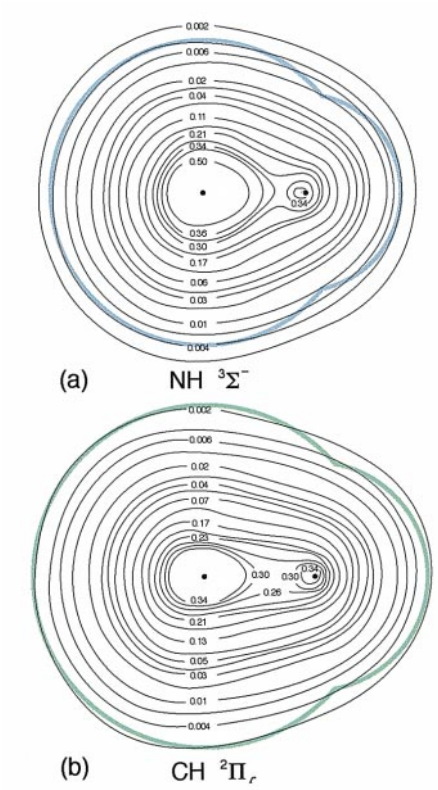

Figure 9. Total molecular charge density contours in atomic units (Figure modified from Bader *et al.*, 1967), overlaid with, in (a) a nitrogen van der Waals radius of 1.55 Å and a hydrogen radius of 1.0 Å at a distance of 1.0 Å; (b) a carbon radius of 1.75 Å and a hydrogen radius of 1.17 Å at a distance of 1.1 Å.

where *gap* is the distance from the dot to the other atom's surface, and *err* is taken as the probe radius, typically 0.25 Å for general evaluation of goodness-of-fit, as used in the accompanying paper. If the probe radius is zero that term is not produced. Multiplying overlap volume by 10 and H-bond volume by 4 before adding the terms, gives an overall scoring profile similar in shape to the van der Waals function for an isolated pairwise interaction, thus:

$$\text{score} = \sum_{\text{dots}} w(\text{gap}) + 4\,Vol(\text{Hbond}) - 10\,Vol(\text{Overlap})$$

Scores can then be normalized by possible surface area. For all of the scores used here (such as output from the clique analysis in Reduce), only the overlap and H-bond terms are included; these scores are not normalized by area. This is equivalent to, and achieved by, setting the probe radius to zero. (The accompanying paper uses normalized scores with the contact dot term and also a ''clashscore'', which is the number of severe overlaps $\geqslant$0.4 Å per 1000 atoms.)

Asn, Gln survey in reference datasets

The set of 100 protein structures used here is listed in Table 1. The accompanying paper (Word *et al.*, 1999) describes the criteria for their choice, including resolution (1.7 Å or better), *R*-value, non-homology, and absence of any unusual problems, starting from the PDB index of January 13, 1997. Only one identical subunit was used from each file, except for three cases with unusually tight interactions: 1DIF, 1RPO, and 2WRP, where interactions to the second subunit are scored but only one set of side-chains is tabulated.

These 100 proteins contain 1554 unique Asn and Gln residues, 15 of which are fixed by metal ligation or covalent modifications. For an initial screen (round 1), contact dots and scores were calculated for each one of them (only using 'a' if there are alternate conformations), both in the originally assigned position and also with the side-chain N and O atoms interchanged (flipped), including interactions to water molecules, but omitting any clashes of atoms with *B*-factor $\geqslant$40 or with alternate conformations, producing comparisons like those represented in Figures 2 and 3. Cases for which the flipped score was not obviously inferior were examined in Mage in order to learn what range of circumstances to expect. However, since side-chain amide groups often interact either with one another or with other hydrogen atoms (such as OH groups) that require positional optimization, definitive assignments of amide flips must be done within interacting networks rather than individually.

These interacting closed sets of side-chains with flippable groups or rotatable H atoms (most often local H-bond networks) were analyzed in a series of steps. First, Reduce identified all metal-liganding or covalently modified groups (listed for Asn/Gln in Table 1) and fixed their orientations. Then all potentially-interacting pairs and larger closed sets among the remaining side chains or heterogen groups are identified by considering their full range of possible hydrogen positions (in both orientations for flips and each 10° for rotations) along with the positions of flippable heavy atoms. At each such position, a sphere is placed with the van der Waals radius of the corresponding atom. If a sphere from one adjustable group overlaps a sphere of another group, then those two groups can interact. Such pairwise interactions are then gathered into disjoint sets, which we call cliques, in the sense that their members all interact internally, but not with any adjustable group outside the clique. The cliques do not propagate through water molecules, because a water molecule is assumed here to be able to act as either an H-bond donor or acceptor independently, even if it makes more than one polar interaction. Only water molecules with $B < 40$ and occupancy $\geqslant$0.66 are used in this analysis. If a side-chain has alternate conformations, the 'a' is used but not the 'b'.

An Asn or Gln amide has only two possible states, and all of its interactions must switch between donor and acceptor in synchrony. A His has two flip states for the ring as a whole; within each of those we consider three possible protonation states (H only on N^{δ}, H only on N^{ε}, or doubly protonated with a small penalty of 0.05), so that its two H-bonds usually but not always change in correlation. Double deprotonation is allowed only if the His ligands two metal ions, such as for His61 in the 1XSO superoxide dismutase (Carugo *et al.*, 1996).

For fully rotatable H atoms, before undertaking the combinatorial step, we use the following process to reduce the number of states that must be considered. For each OH or SH, orientations for the rotatable H atom are selected in the direction of each one of the potential H-bond acceptors surrounding it. If the acceptor is too close for an acceptable straight-line H-bond, then potential H orientations are also defined 15° (and,

if necessary, also 30°) on either side of it. Finally, an additional orientation is located which avoids these acceptors and has minimal interaction with all surrounding atoms. For an OH surrounded by three H-bond acceptors, for instance, we would thus define four potential positions to be tried in the combinatorial search; but if the acceptors were all close, there would be ten potential H positions (three near each acceptor and one that avoids them all). Each Met methyl and each Lys or N-terminal NH_3^+ is considered in four possible orientations, 30° apart. Each side-chain in an interacting clique has a fairly small number of possible H arrangement states that must be considered. Since the largest clique found in our 100 reference structures contains six members (for a later set of 240 proteins tested, one clique had eight members), an exhaustive search is computationally tractable in practice.

The isolated OH, SH, NH_3^+, and Met methyl groups are rotationally optimized in Reduce by sampling every preassigned orientation and then testing 1° increments around the best one; the final rotation angle and score for the best position are reported, and the optimized H atom is added to the output PDB file. Each clique of two or more is then analyzed by an exhaustive search through all combinations of the preassigned potential H positions for all residues in the clique, plus a final rotational optimization around the best preassigned position. For each residue, the program accumulates its best score, the best total score for its clique, and (for Asn, Gln, or His) the best total clique score found with this residue in its opposite flip state. Those scores and the best assignment (e.g. ''FLIP - + both NH'' for a His) are written both to the screen and onto the header of the output PDB file, and H atoms for the chosen clique conformation are added to the PDB output file. Each Asn/Gln/His (unless fixed in advance by a covalent modification) is assigned to one of four categories: Keep (K), Flip (F), double-Clash (C), or unknown (X). An adjustable penalty can be applied to the difference between the best score and the best flipped score, in order to automatically leave the marginal cases in the state originally assigned by the depositors. Alternatively, they can be examined and the flip state decided by the user.

In order to obtain a visual comparison of the alternative states, Reduce is rerun, using the file header information from its previous output to specify that it now should optimize cliques (OH rotations, etc.) with each flippable group fixed in the non-preferred orientation. Both output PDB files are used by a Unix script that produces a kinemage with precalculated views scaled and centered on each Asn/Gln/His, so that they can easily be examined in Mage using animation to switch between the two alternative arrangements. The user can then decide whether to reject any of the automated assignments.

Program and data availability

The annotated list of 100 high-resolution structures, the coordinate files with H atoms added and Asn/Gln flips corrected, and the contact dot kinemage files with animated Asn/Gln comparisons, plus the programs Reduce, Probe, Prekin, Mage, and supporting scripts and utilities are available from the anonymous FTP site (ftp://kinemage.biochem.duke.edu) or the WorldWide-Web site (http://kinemage.biochem.duke.edu). Probe is a generic Unix C program; the current version (v2) of Reduce that includes H-bond clique optimization is in C^{++}. Mage and Prekin are in C, available for Mac, PC, Linux, and SGI Unix, and can be re-compiled to run on most Unix platforms where Motif is available. A stripped-down version of Mage written in Java is used on our Web site to provide real-time interactive display of small kinemages with contact dots.

Acknowledgments

This work was supported by NIH research grant GM-15000, by use of the Duke Comprehensive Cancer Center Shared Resource for Macromolecular Graphics, and by an educational leave for J.M.W. from the Glaxo Wellcome Inc.

References

Alzari, P. M., Souchon, H. & Dominguez, R. (1996). The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. *Structure,* **4**, 265-275.

Bader, R. F. W., Keaveny, I. & Cade, P. E. (1967). Molecular charge distributions and chemical bonding. II. First-row diatomic hydrides, AH. *J. Chem. Phys.* **47**, 3381-3402.

Bass, M. B., Hopkins, D. F., Jaquysh, W. A. N. & Ornstein, R. L. (1992). A method for determining the positions of polar hydrogens added to a protein structure that maximizes protein hydrogen bonding. *Proteins: Struct. Funct. Genet.* **12**, 266-277.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. (1981). Interaction models for water in relation to protein hydration. In *Intermolecular Forces* (Pullman, B., ed.), pp. 331-342, D. Reidel Publishing Company, Boston.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Bode, W., Papamokos, E. & Musil, D. (1987). The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech, *Hirudo medicinalis*. *Eur. J. Biochem.* **166**, 673-692.

Carugo, K. D., Battistoni, A., Carri, M. T., Polticelli, F., Desideri, A., Rotilio, G., Coda, A., Wilson, K. S. & Bolognesi, M. (1996). Three-dimensional structure of *Xenopus laevis* Cu,Zn superoxide dismutase *b* determined by X-ray crystallography at 1.5 Å resolution. *Acta Crystallog. sect. D,* **52**, 176-188.

Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science,* **221**, 709-713.

Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). The occurrence of C-H···O hydrogen bonds in proteins. *J. Mol. Biol.* **252**, 248-262.

Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. (1975). The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0-Å resolution. *Biochemistry,* **14**, 4943-4952.

Fisher, A. J., Thompson, T. B., Thoden, J. B., Baldwin, T. O. & Rayment, I. (1996). The 1.5-Å resolution crystal structure of bacterial luciferase in low salt conditions. *J. Biol. Chem.* **271**, 21956-21968.

Flory, P. J. (1969). *Statistical Mechanics of Chain Molecules* (Jackson, J. G. & Wood, C. J., eds), 1st edit., vol. 3, Interscience Publishers, New York.

Hagler, A. T., Huler, E. & Lifson, S. (1974). Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J. Am. Chem. Soc.* **96**, 5319-5327.

Hooft, R. W. W., Sander, C. & Vriend, G. (1996). Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins: Struct. Funct. Genet.* **26**, 363-376.

Housset, D., Habersetzer-Rochat, C., Astier, J.-P. & Fontecilla-Camps, J. C. (1994). Crystal structure of toxin II from the scorpion *Androctonus australis* Hector refined at 1.3 Å resolution. *J. Mol. Biol.* **239**, 88-103.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A,* **47**, 110-119.

Matsuura, Y., Takano, T. & Dickerson, R. E. (1982). Structure of cytochrome c_{551} from *Pseudomonas aeruginosa* refined at 1.6 Å resolution and comparison of the two redox forms. *J. Mol. Biol.* **156**, 389-409.

McDonald, I. K. & Thornton, J. M. (1994). The application of hydrogen bonding analysis in X-ray crystallography to help orientate asparagine, glutamine and histidine side chains. *Protein Eng.* **8**, 217-224.

McRee, D. E. (1993). *Practical Protein Crystallography*, 1st edit., Academic Press, San Diego.

Richardson, D. C. & Richardson, J. S. (1992). The kinemage: a tool for scientific illustration. *Protein Sci.* **1**, 3-9.

Richardson, D. C. & Richardson, J. S. (1994). Kinemages: simple macromolecular graphics for interactive teaching and publication. *Trends Biochem. Sci.* **19**, 135-138.

Serre, L., Verbree, E. C., Dauter, Z., Stuitje, A. R. & Derewenda, Z. S. (1995). The *Escherichia coli* malonyl-CoA:acyl carrier protein transacylase at 1.5-Å resolution. *J. Biol. Chem.* **270**, 12961-12964.

Spurlino, J. C., Smallwood, A. M., Carlton, D. D., Banks, T. M., Vavra, K. J., Johnson, J. S., Cook, E. R., Falvo, J., Wahl, R. C., Pulvino, T. A., Wendoloski, J. J. & Smith, D. L. (1994). 1.56 Å structure of mature truncated human fibroblast collagenase. *Proteins: Struct. Funct. Genet.* **19**, 98-109.

Stout, G. H. & Jensen, L. H. (1968). *X-ray Structure Determination: A Practical Guide*, vol. 12, The MacMillan Company Collier-MacMillan Ltd, London.

Tame, J. R. H., Dodson, E. J., Murshudov, G., Higgins, C. F. & Wilkinson, A. J. (1995). The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure,* **3**, 1395-1406.

Vassylyev, D. G., Katayanagi, K., Ishikawa, K., Tsujimoto-Hirano, M., Danno, M., Pahler, A., Matsumoto, O., Matsushima, M., Yoshida, H. & Morikawa, K. (1993). Crystal structures of ribonuclease F1 of *Fusarium moniliforme* in its free form and in complex with 2′ GMP. *J. Mol. Biol.* **230**, 979-996.

Vlassi, M., Steif, C., Weber, P., Tsernoglou, D., Wilson, K. S., Hinz, H.-J. & Kokkinidis, M. (1994). Restored heptad pattern continuity does not alter the folding of a four-α-helix bundle. *Nature Struct. Biol.* **1**, 706-716.

Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S. & Richardson, D. C. (1999). Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* **285**, 1711-1733.

Edited by J. Thornton

(Received 28 May 1998; received in revised form 2 November 1998; accepted 3 November 1998)