
ALL-ATOM CONTACTS: A NEW APPROACH TO STRUCTURE VALIDATION

Jane S Richardson

The enormous wealth of macromolecular structure data already available and the even greater wealth soon to come—from structural genomics, from the push for atomic-resolution structures, and from the push to solve much larger biological complexes—provide a treasure trove of functional, interactional, and evolutionary data that will change how one can do biology. But, in order to make effective use of this great resource, it is important, among other things, to take into account the very large spread of accuracy in those data. Relatively low-resolution structures can be among the most valuable if they are of critical molecules or of large and complex cellular machinery. These structures show overall fold and relative positioning of their parts and they often illuminate function in surprising ways, but one should not expect to learn from them fine details in an active site or critical differences that determine substrate or inhibitor specificity. At the other extreme, increasing numbers of structures are being solved at better than 1 Å resolution, where one can reliably detect minute changes and disentangle multiple conformations of side chains and waters. Within a given structure there can be even wider variability in quality. Regardless of resolution, most structures have some parts disordered enough that they are not visible in a crystallographic electron-density map (or have no observable NMR constraints). In some cases their coordinates will actually be missing, but more often disordered areas are indicated by a high crystallographic B-factor or highly divergent conformations in an NMR ensemble. If a particular part of a structure is important to the question being asked, such telltale signs should always be heeded.

Structural Bioinformatics

Edited by Philip E. Bourne and Helge Weissig

ISBN 0-471-20199-5 Copyright © 2003 by Wiley-Liss, Inc.

Many of the basic quality indicators such as resolution, *B*-factor, R and free R residuals (measures of how well the model accounts for the observed data), and model root-mean-square deviation (rmsd) are directly reported in the Protein Data Bank (PDB) coordinate file (Chapter 8; Berman et al., 2000). Beyond those indicators, the subject known as structure validation (Chapter 14) provides further tools for assessing both overall and local accuracy of structures. Standard validation programs such as ProCheck and WhatIf provide an excellent set of widely-used tools, centering especially on ideality of molecular geometry and on whether backbone ϕ , ψ angles occur outside the preferred *core* regions. Of special importance in validation are independent criteria not explicitly part of the target function optimized by the structure-refinement process, because their deviations are much more sensitive indicators of problems. The two classic such indicators are the ϕ , ψ or Ramachandran plot (Ramachandran et al., 1963; Laskowski et al., 1993; Lovell et al., 2002), since ϕ , ψ values are not in the target function, and the free R factor (Brunger 1992), the agreement with a designated 5–10% of the data that are deliberately kept out of refinement in order to provide an unbiased indicator of progress in model quality.

Recently we have discovered, in a surprisingly simple place, a new source of information for an unbiased and sensitive validation criterion: the hydrogen atoms. They constitute about half of the atoms, but they usually are ignored for technical or expediency reasons. H atoms are, of course, important and present in NMR structures (see Chapter 5), although often not treated with full radius. In macromolecular crystallography, polar H atoms are typically added to better define H bonds but with no van der Waals terms, while nonpolar H atoms are added and refined against the data only at ultra-high (near 1 Å) resolution. The main reason for this omission is that hydrogens diffract X rays very poorly, so that they can be directly detected only under the best of conditions.¹ Another reason is that including hydrogens doubles the number of parameters if they are treated as fully independent, which is acceptable only when there is a large enough number of experimental observations. Finally, only recently has computer speed allowed the extra cost in time, either for structure refinements or for theoretical calculations. H atom volume is standardly accounted for by using larger united-atom radii for the other atoms, but the directionality and specificity of H interactions are not represented. The net result of all this is that the crystallographers have obligingly ignored half their atoms in refinement, managing to do quite well without them but now giving us the opportunity to use the correctness of the hydrogens' tight and specific packing interactions as both a global, and especially a local, validation criterion. This new method (Word et al., 1999a) is called all-atom contact analysis.

As applied to the structural database, all-atom contact analysis has two different goals. The first, long-term goal is to actually improve the accuracy of the data, by having structural biologists apply the criteria themselves and fix many errors before coordinates are deposited (a similar process occurred several years ago with routine application of free-R and Ramachandran-plot criteria). The second goal is to give users of the database an easy and effective way to assess local structural accuracy. The first goal would produce higher-grade ore for data mining, while the second improves the extraction process.

¹The invisibility of hydrogens is actually very fortunate, because it produces the beautifully clear separation between aliphatic side chains in protein interiors at moderate resolution.

THE METHOD OF ALL-ATOM CONTACT ANALYSIS

The all-atom method must start off with a reliable way to add H atoms and optimize their positions, which is done by the program Reduce (Word et al., 1999b). A great many of the hydrogen positions are completely determined by the heavier atoms: methylene, methine, backbone NH, aromatic H, and so forth. OH rotations and His protonation, however, must clearly be optimized relative to the surrounding structure. Less obviously, the 180° flip orientation of Asn and Gln side-chain amides (and also flips of His rings) need to be optimized; they are fairly often incorrect as reported, because the N and O atoms of the amide are not easily distinguished by the experimental X-ray data. However, the choice can reliably be made if both H bonding and potential clashes of the NH₂ are considered (Word et al., 1999b). We have found, surprisingly, that most methyl rotations do not actually need to be optimized because they are remarkably relaxed in protein structures, with departures from staggered orientation seldom much above 10°. NH₃ groups and Met side-chain methyls do however need rotational optimization. The Reduce program handles nucleic acids and small-molecule ligands as well as proteins, and interactions with individual bound waters are treated by a simplified model. The reason hydrogen addition is a complex process is that the movable H atoms often occur in interacting H-bond networks and must be optimized as a group rather than individually. In practice, such H-bonding cliques are small enough, given our simplified model for water molecules, that exhaustive evaluation of all possible hydrogen positions is computationally tractable. A single simple command runs Reduce rapidly and produces a commented, properly formatted output PDB file with all H atoms present.

All-atom contacts are calculated by the program Probe (Word et al., 1999a) from a Reduce-modified PDB file that now includes hydrogens. The usual output is contact surfaces as color-coded dots in the “kinemage” format for display in the Mage graphics program (Richardson and Richardson 1992; Richardson and Richardson 2001) as shown in the color figures for this chapter, but other display formats, numerical scores, or lists of serious clashes can also be produced (see Current Facilities and Their Use below). Typically, Probe is run on an entire PDB file, but it can also calculate the internal contacts for a small region or just the contacts between two pieces (i.e., a ligand and a protein), using a flexible syntax of atom selection.

Figure 15.1 illustrates a simple example of all-atom contact surfaces for a small region, to show the appearance of favorable van der Waals contacts, favorable H bond overlaps, and unfavorable atomic overlap, color-coded by the local gap distance between the two contacting atoms. The all-atom contact algorithm rolls a small spherical probe (shown as a gray ball) on the surface of each atom, drawing a colored dot only when the probe intersects another noncovalently-bonded atom. This method is a bit like the inverse of solvent-exposed surface (Connolly 1983; Lee and Richards 1971), where here only occluded surface is shown; however, our much smaller probe means that only atom pairs within 0.5 Å of touching will count as contacts. These contacts are extremely sensitive to fine details of how well the structure fits together. If a local conformation is in the right energy well but not quite correct, it will usually produce just yellow and orange overlap dots. However, it is very difficult to fit anything in a completely wrong conformation without producing red clash overlaps, even after refinement has done its best at adjustments. Therefore, the primary way of interpreting the all-atom contact results is simply that lots of soothing green (such as seen in Fig. 15.2a and 15.2b) means the structure is correct, while an area of red spikes has some

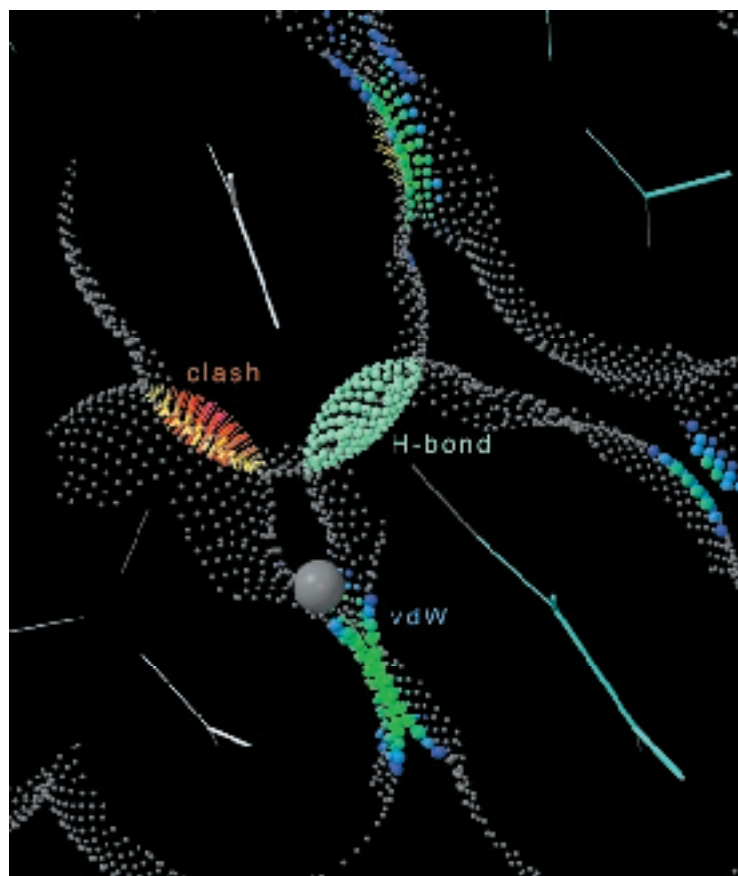


Figure 15.1. Slice through a small section of protein structure (stick figure, backbone in white and side chains in cyan) showing the relation of all-atom contact surfaces (colored dots) to the atomic van der Waals (vdW) surfaces (gray dots) and to the 0.25 Å-radius probe sphere (gray ball) used in the calculation. The small probe sphere is rolled over the surface of each atom, leaving a contact dot only when the probe touches another noncovalently-bonded atom. The dots are colored by the local gap width between the two atoms: blue when nearly maximum 0.5 Å separation, shading to bright green near perfect van der Waals contact (0 Å) gap. When suitable H-bond donor and acceptor atoms overlap, the dots are shown in pale green, forming lens or pillow shapes. When incompatible atoms interpenetrate, their overlap is emphasized with spikes instead of dots, and with colors ranging from yellow for negligible overlaps to bright reds and hot pinks for serious clash overlaps ≥ 0.4 Å. Kinemage-format contact dots also carry color information about their source atom (e.g., O's are red, S's are yellow, etc.); in Mage, one can toggle between the two color schemes.

sort of problem. In fact, for an all-atom kinemage interactively displayed in Mage one can turn off everything but the bad clashes and quickly spot all problem areas even in a large structure, as shown for the 324-residue dimer in Figure. 15.2c.

In addition to graphic display, several scoring schemes suitable for different purposes produce numerical evaluations of the contact, H-bond, and clash terms (Word et al., 1999a).

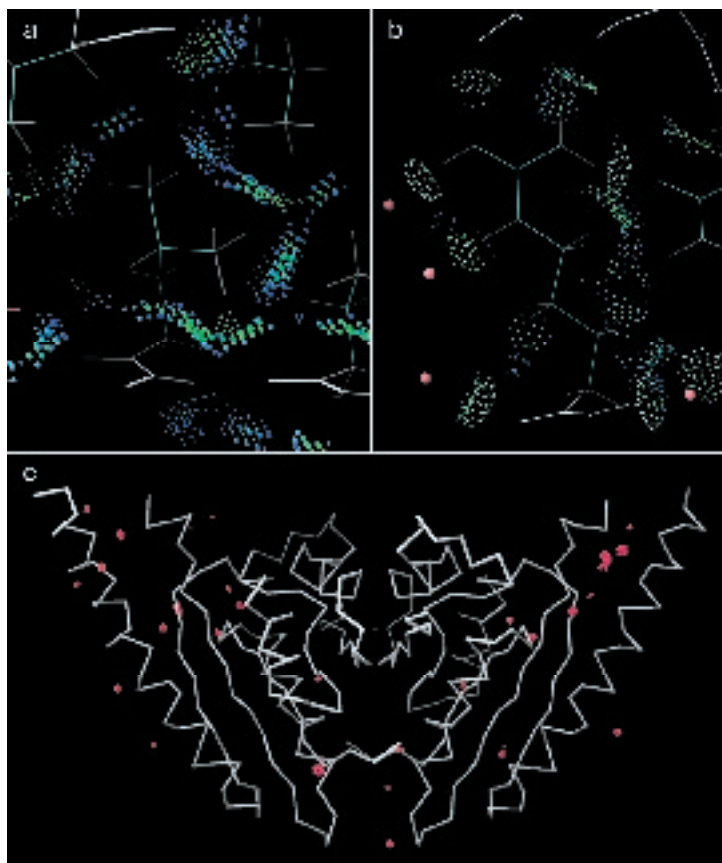


Figure 15.2. All-atom contact examples from the dimer of 1MJH (Zarembinski et al. 1998), a well-determined structural-genomics protein at 1.7 Å resolution. (a) All contacts for one of the typically well-packed and well-fit regions of aliphatic side chains, with the green of close van der Waals contacts predominant. (b) All contacts for an Arg side chain, with all 5 planar H-bonds (lens-shaped groups of pale green dots) of its guanidinium NH's formed either to protein O atoms or to waters (pink balls). (c) An overview of the dimer, with only the C α backbone and the serious clashes ≥ 0.4 Å shown. When interactively displayed in Mage, it is easy to locate and fix the small number of isolated problems, including two flipped-over His rings at the putative active site and a high-*B* Lys squeezed into insufficient space between two hydrophobic side chains.

These scores are not energies, however, because the serious clash overlaps represent model errors, not real strains in the structure. When used to understand features of molecular architecture, such as side-chain packing, overlaps are treated simply as tight contacts, but for structure-validation and error-correction purposes, the clash overlaps are very much the dominant issue. We consider a serious clash (one that usually indicates some sort of misfitting) to occur where two incompatible atoms overlap by 0.4 Å or more. The overall clash score of a structure is the number of serious clashes per 1000 atoms.

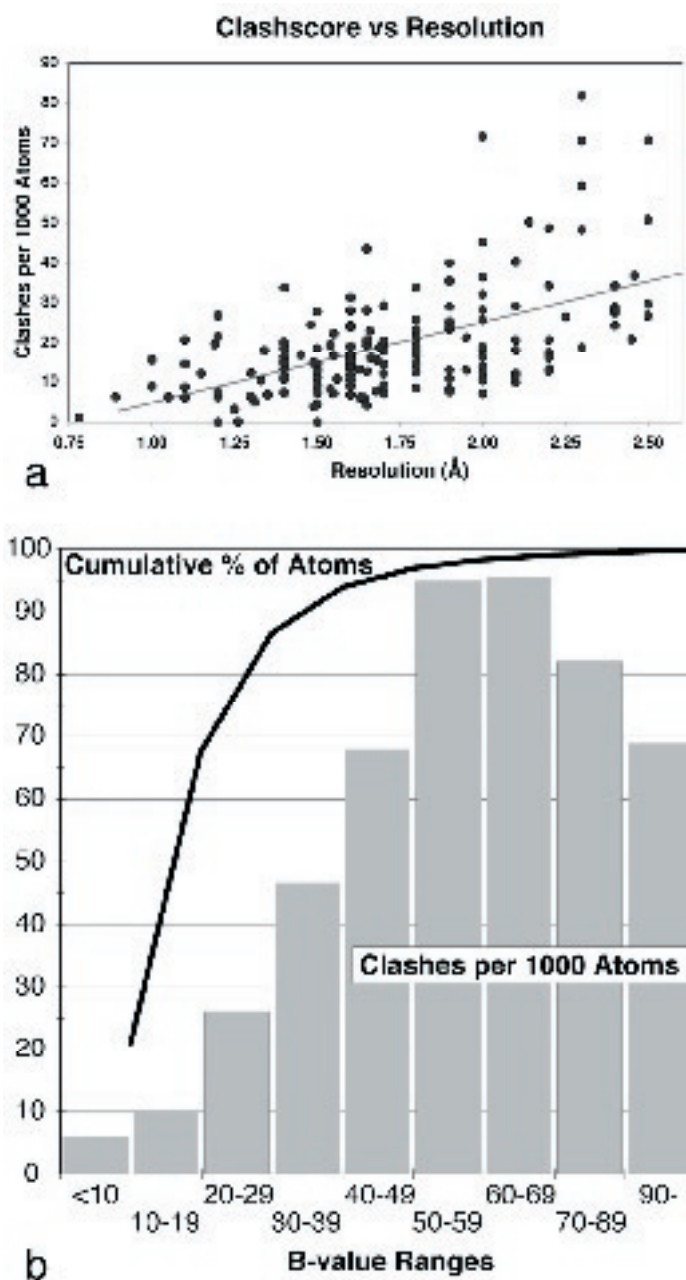
RELATIONSHIP TO MORE TRADITIONAL CRITERIA

The well-ordered parts of the very best X-ray and NMR structures fit the all-atom contact criteria nearly perfectly, with extensive contacts throughout the interior, an absence of even modest clashes, and most atoms showing the green dot patches of ideal van der Waals contact as in Figure. 15.2a and 15.2b (and, at even higher resolution, in Fig. 15.5a below). Such agreement is strong confirmation that our algorithms and parameters have been chosen correctly. Clash score is strongly correlated with other indicators of structure quality: overall parameters such as resolution or number of NMR restraints correlate with overall score (Fig. 15.3a), and local crystallographic *B*-factor correlates especially strongly with locally measured clash score (Fig. 15.3b).

If structure factors are available, enabling examination of the electron-density map in the area of a serious clash, it usually turns out that the density is either weak or its shape is somewhat ambiguous, making a misfitting more likely than in clearer areas. For example, electron density for a side chain that branches at the C β such as Thr or Val fairly often has a straight bar shape rather than a tetrahedral junction, making it possible to misfit the χ 1 angle by 180°. When that happens (as for the Val in Fig. 15.4a), there are always clashes with the H β or H γ atoms, the side-chain rotamer will be poor, and the bond-angle geometry around the C α will almost always be badly distorted through forcing the C γ atoms to fit into the bar-shaped electron density although connected to a C β that has been fit on the wrong side of the bar. Figure. 15.4b shows both the original and the refit side chains, emphasizing the great difference in their geometry and conformations though occupying nearly the same space; Figure. 15.4c shows the excellent fit obtainable in a good rotamer with ideal geometry and no backbone movement. Traditionally, electron-density difference maps are used as indicators of this kind of problem (for instance, often but not always showing a pair of positive and negative peaks at the real and the misfit C β), but they are difficult for noncrystallographers to calculate and interpret. The Uppsala Electron Density Server (<http://portray.bmc.uu.se/eds/>) is a valuable source of viewable electron density maps and related quality criteria, for those PDB files with available structure factors that could successfully be processed automatically. In contrast to the technicalities of electron-density maps, user-friendly validation tools are available for assessing all-atom contacts, rotamers, geometric ideality, and whether backbone ϕ, ψ angles are unfavorable. These tools are best used in concert with one another, because a given problem usually shows up only in a subset of them. For instance, if the refinement terms for ideal geometry were heavily weighted

Figure 15.3. Correlation of all-atom clash scores with other indicators of structure quality. (a) Overall clash score (number of serious overlaps ≥ 0.4 Å per 1000 atoms, after correction of amide flips) as a function of resolution, for 328 protein structures between 0.8 Å and 2.5 Å resolution. The relationship is highly significant and is still improving down near 1 Å. (b) Serious clashes per 1000 atoms, grouped into ranges of crystallographic *B*-factor values, for 100 proteins at 1.7 Å resolution or better. Note that an atom with *B* > 50 is 10 times as likely to clash as one with *B* between 10 and 20. Clashes fall off again at the very highest *B* range, because those atoms are exposed at the surface with few neighbors. Note that for high-resolution structures only about 5% of the atoms have *B* > 40, so that the ones most prone to error can be omitted from empirical studies with essentially no loss in sample size. Part b is reproduced from Word et al., (JMB, 1999a) by permission of Academic Press.

relative to agreement with the experimental data, then bond angles will not be distorted but clashes will show; however, if clashes are between non-H atoms, then refinement may remove them at the expense of geometry. In our experience, the two most sensitive and reliable indicators of local problems in a structure are bond-angle distortions (Lovell et al. 2002) and all-atom clashes. Nonideal torsion angles or bond distances often tell more about



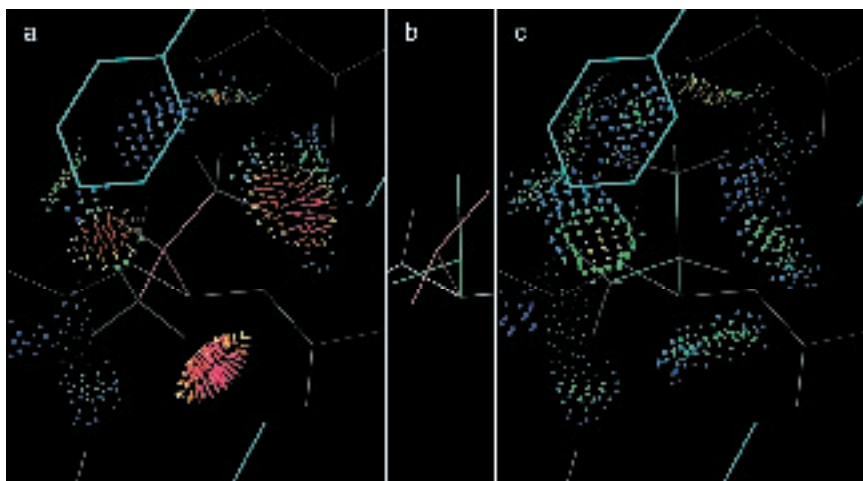


Figure 15.4. Diagnosis and correction of a backward-fit valine side chain. (a) All-atom contacts for the original side chain, with substantial clashes (hot pink) and an eclipsed χ_1 angle. (b) Original and refit side chains, showing how both occupy the same space but in opposite orientations. Bond-angle distortions in the original put its $C\beta$ 0.48Å from the idealized position. (c) Good all-atom contacts for the refit Val, which has ideal geometry and staggered χ_1 without backbone movement. Even without deposited structure-factor data, one can be fairly confident that the electron density must have been ambiguous and that the conformation shown in (c), not (a), is in the correct local energy well. From the 2SIM neuraminidase at 1.6 Å resolution (Crennell et al., 1996).

how the refinement was set up than about the structural accuracy, whereas a bond angle off by 6–8° from ideal or a clash of 0.5 Å (or both!) almost always means that something is seriously misfit.

In drawing conclusions from a structure or comparison, it seldom matters if one or two parameters are slightly off (e.g., a torsion angle by 15°), but it is often critical if the backbone or side chain are actually in the wrong conformation (e.g., a torsion off by 90–180°): that will change which atoms are in position to interact, say, with a ligand. Local problems in polypeptide chain-tracing, such as a sequence out of register by two within a β strand, are often flanked at each end by clusters of all-atom clashes and bad bond angles. Neither all-atom contacts nor geometrical ideality are suitable in general, however, for identifying incorrect chain folds—both are too sensitive and too local. That task is probably best done by the sort of threading methods used in fold recognition (Chapter 26) and homology modeling (Chapter 25).

As an extra bonus, we have found that filtering experimental datasets by all-atom clashes and B -factors as well as by resolution can greatly improve the quality of Ramachandran-plot criteria (Lovell et al. 2002) and of side-chain rotamer libraries (Lovell et al., 2000), thus indirectly improving these more traditional validation tools. The new rotamers have no internal clashes and all occupy valid local energy minima. The Ramachandran plots are much cleaner, allowing defensible separation of disfavored but allowed

regions from forbidden regions and the definition of core regions for Pro and Gly. Separate criteria for glycines are important to validation, because the lack of a C β makes Gly ϕ, ψ the most error-prone for either X-ray or NMR structures. These new rotamers and Ramachandran plots are available from the Kinemage Web site (RichardsonLabWebSite 2001).

CURRENT FACILITIES AND THEIR USE

Probe and Reduce for calculating all-atom contacts and Mage and Prekin for interactive display of molecules and contacts are available as free, open-source software from our Web site (RichardsonLabWebSite 2001). Probe and Reduce run on Unix, Linux, Mac OSX, or (less conveniently) PC; the Mage/Prekin display runs on Unix, Linux, Mac, PC, or Java.

The most basic and general function of all-atom contact analysis for structure validation is to generate a clash report on a particular PDB file, either in graphic form or in list form. To set up for analysis of a PDB file called `lxyz`, first add and optimize H atoms by running the command:

```
reduce -build lxyz > lxyzH
```

and make a kinemage file of the structure (including backbone, side chains or bases, H atoms, small-molecule ligands, and waters) with:

```
prekin -lots lxyzH > lxyzH.kin
```

Now calculate all-atom contacts and append them to the kinemage file with:

```
probe lxyzH >> lxyzH.kin
```

This graphic contact report is viewed in Mage by typing:

```
mage lxyzH.kin
```

For example, on a small, very high-resolution structure such as the 1BRF thermophilic rubredoxin in Figure 15.5 the contacts are excellent throughout (green, with some yellow and blue). In this case if everything is turned off except the bad overlaps, it is immediately obvious that there is a single serious clash between two surface side chains. If the crystallographer is looking at this clash report, he or she should investigate that region to see if it can be corrected; if a bioinformaticist is doing the evaluation, he or she now knows all of this structure is of extremely high quality except for the two clashing side chains, whose detailed conformation cannot be trusted. This example is very small for clarity of presentation in static two dimension (2D), but in the interactive display it is easy to locate the problem regions even on a large, lower-resolution structure and to zoom in and examine them.

The two areas in which such clash reports have had the greatest impact are both for crystal structures: detecting and fixing protein side chains fit in the wrong rotamer, and finding places where nucleic-acid backbone conformations are incorrect. The most common side-chain misfittings are for Asn/Gln flips (case below), Thr/Val/Leu tetrahedral branches

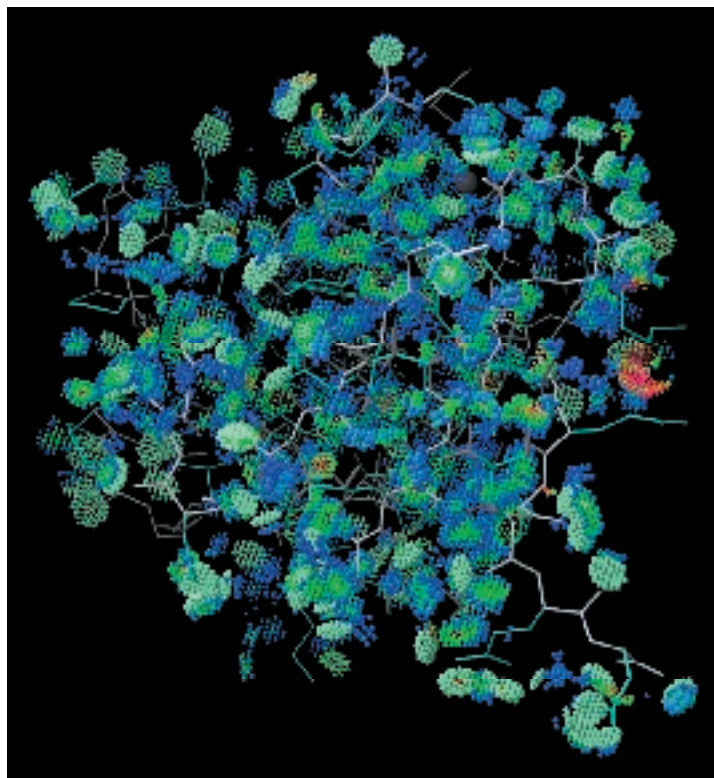


Figure 15.5. All-atom contacts for the entire structure of 1BRF rubredoxin (Bau et al., 1998), a highly accurate small protein structure at 0.95 Å resolution. The dense green dot patches signifying well-packed contacts in the molecule and a well-fit model are seen consistently throughout the structure, except for a single red clash between two surface side chains. 1BRF thus illustrates both how precisely the all-atom contact criteria are satisfied in atomic-resolution protein structures and also how occasional local errors can be found even in such extremely high-quality structures.

(as in Fig. 15.4), and Met conformations. The reasons for problems with Thr/Val and with Leu are discussed in detail in Lovell et al. (2000); for validation purposes it suffices to know that these problems occur fairly often and that they almost always produce bad clashes and usually distort C α –C β geometry (Lovell et al., 2002). Met can be difficult because the heavy S δ atom produces diffraction ripples in the electron density that weaken the information for the nearby C γ and C ϵ ; all-atom clash and rotamer information can usually make the correct choice clear.

For nucleic-acid crystal structures, the bases are large, rigid, and well determined, and the P atom density is generally unambiguous, even at the moderate resolution (often around 2.5 Å) typical of the most biologically interesting structures. In those same structures, however, the rest of the sugar-phosphate backbone has too many free parameters per observable atom (see Chapter 3 for description of the six backbone dihedrals per residue) and is quite prone to errors when in conformations less well understood than standard B-DNA or A-RNA. H atom clashes, however, mark the incorrect conformations extremely clearly.

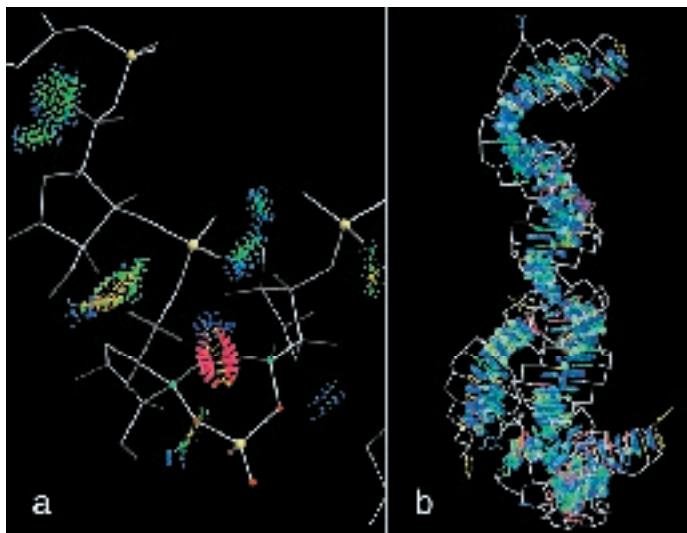


Figure 15.6. Base and backbone all-atom contacts in the 5S RNA from the 2.4 Å ribosome structure of 1FFK (Ban et al., 2000). (a) A section of the backbone-backbone contacts, mostly very nicely packed but with one impossible overlap of C 3' and C 5' hydrogens (red spikes). (b) Base-base contacts, showing the long columns of well-fit base stacking.

Figure 15.6a shows all-atom contacts for just the backbone of part of a 5S RNA; most areas show excellent contacts, but one residue is in a physically impossible conformation. To calculate such a backbone display, use:

```
probe -mc "mc" 1rnaH >> 1rnaH.kin
```

When analyzing nucleic acid structures, all-atom contacts also provide a quick and pleasing way to visualize base stacking (see Fig. 15.6b), done with:

```
probe "base" 1rnaH >> 1rnaH.kin
```

and the NAContacts script (RichardsonLabWebSite 2001) gives a sensitive numerical measure of stacking quality.

As a clash report for those who prefer working with lists and scores rather than visual displays, the Clashlistcluster script produces a text file that lists all clashes ≥ 0.4 Å, clustered into local groupings in space (often a single problem is responsible for several nearby clashes). After running Reduce, the command is:

```
clashlistcluster 1xyzH > 1xyzHclcl.txt
```

That file gives the total clash score, for an overall evaluation, or one can look to see whether the residues of interest have any bad clashes. Since all of these programs can be run and controlled from the command line, they can be combined into shell scripts that perform a desired sequence of operations on an entire list of structures. For more information on options, type reduce -help or probe -help.

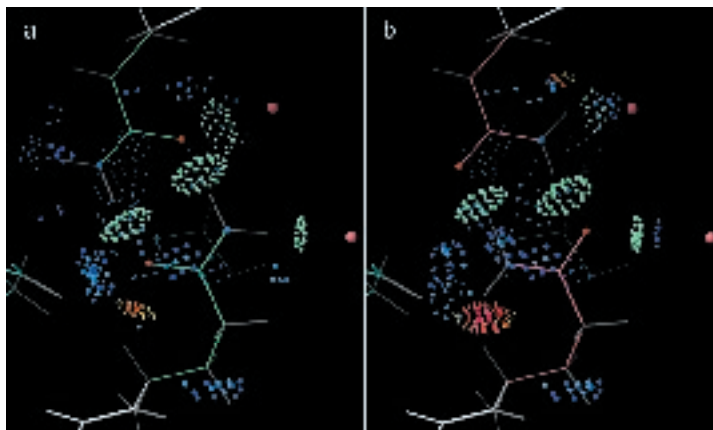


Figure 15.7. Resolving the ambiguity in a pair of doubly H-bonded side-chain amides, which have equivalent H bonds both to each other and to waters in the two possible flip states. (a) The correct flip orientation, with only a minor overlap. (b) The next-best, but incorrect, flip state with a large, physically impossible clash of the Gln Nε' H with Hε\ (red spikes). From the 1.6 °A peroxidase of 1ARU (Fukuyama et al., 1995)

As explained above, automatic correction of 180° flip alternatives for Asn, Gln, and His is done as part of the H-bond network optimization in Reduce (invoked by the “-build” flag). In order to see why Reduce made each change and to evaluate its level of certainty, a Perl script called Flipkin is used:

```
flipkin 1xyzH > 1xyzNQflip.kin
```

This produces an output kinemage file with a preset view for each Asn/Gln in the structure (or His, if run with the “-h” flag added), with the ones Reduce chose to flip marked with “*” on the Views menu. An animation is set up between the two possible flip states with display of the contacts, H bonds, and clashes in each state. Figure 15.7a and 15.7b show the two displays for a doubly interacting Asn-Gln pair whose H bonds are equally good in either flip state, but where the original choice has an impossibly bad clash with the Gln CαH, whereas the flipped state fits well. The flip of a side-chain amide is a small change but can be crucial if it affects an H bond at an active, allosteric, or binding site.

In addition to assessing the reliability of particular pieces of three-dimensional structure, all-atom contact analysis can be used in the interactive Mage/Probe system (Word et al., 2000) to try out the plausibility of a proposed alternative conformation, or to see whether a modified sequence would be compatible with a known structure. This methodology was developed to tell whether or not a single-site mutation can be accommodated without movement of the original structure around it (essential information in determining whether observed mutant properties are actually due to the altered side chain itself); however, it would also be useful in structural bioinformatics for assessing the degree of structural change between closely related sequences, or deciding whether or not one protein could assume a different conformation seen for other members of the related family.

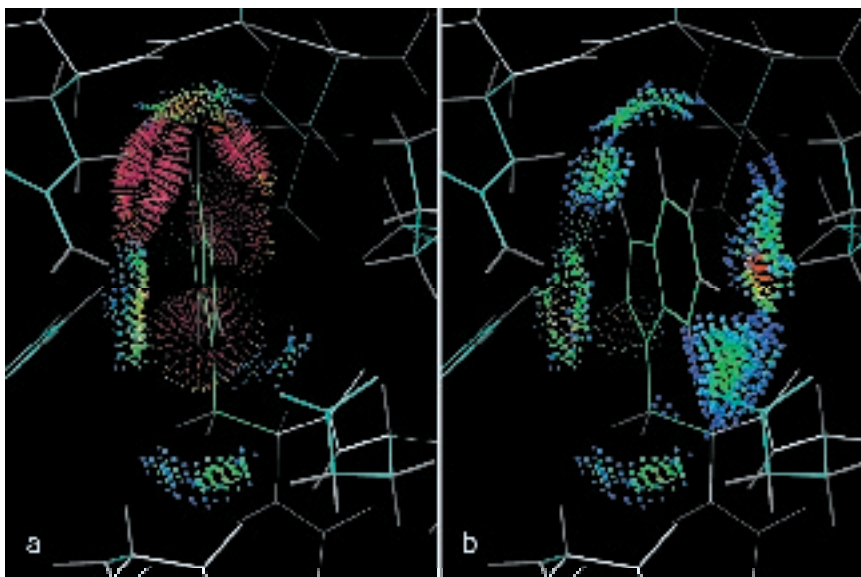


Figure 15.8. A test of alternative sequence possibilities substituting Trp for Tyr at a buried position in the N-terminal domain of λ repressor, using the “remote update” function in the interactive Mage/Probe system. (a) One of the initial rotamer trials, with impossibly bad clashes on both sides of the Trp ring. (b) The best of the exact rotamers, with only two minor overlaps in orange, indicating that the Trp side chain can indeed fit without perturbing the structure significantly. Starting coordinates from 1LMB (Beamer and Pabo, 1992).

In the interactive Mage/Probe system, the programs are set up to communicate directly with each other, so that an all-atom contact display changes as you rotate bonds in the structure. When looking at a stick-figure kinemage of your structure (in Unix, Linux or OSX), center on the residue of interest and choose “remote update” on the Tools menu. Ask for Prekin to mutate the residue (edit the three-letter code to what you want, in the command line proposed by Mage), and the new sidechain will appear in green, with idealized geometry, along with sliders to change its χ angles. (Note that the PDB file 1xyzH must be in the directory, for Prekin and Probe to use.) Go to “remote update” again, and ask for Probe to calculate contacts around the rotatable side chain. First you see contacts around the present position, and they will be updated automatically as you change the conformation.

Figure. 15.8a shows unsatisfactory Probe dots around a buried Tyr to Trp mutation that has been made rotatable, but not yet properly optimized. In the text window is a list of the rotamers for that amino acid (as defined in Lovell et al., 2000); if you click on one of the Q6 listed rotamers, the side chain will be put in that conformation. Usually most of the rotamers will have terrible clashes like the one in Figure. 15.8a, which is actually the second-best rotamer (χ_1 trans, χ_2 - 105°). Identify the rotamers that are most nearly acceptable, such as the good rotamer shown in Figure 15.8b (χ_1 trans, χ_2 90°). Then move the χ angles by modest amounts (up to 20–30°) to look for a position with green, blue, and yellow dots, and perhaps the pale green pillows of H-bond dots, without any appreciable amount of red spikes; in this case, there is a very well-fitting conformation only 3° away from the best rotamer. This Trp mutant was produced and found to have a stability and folding rate at least as good as, and

an NMR spectrum very close to, that of the parent λ repressor domain (Ghaemmaghami et al., 1998). In general, if a satisfactory conformation can be found in Mage/Probe for the mutated side chain, that means that the new amino acid can be accommodated without changing anything else. If it looks as though moving another side chain would help, you can make it rotatable as well. If no acceptable conformation can be found, the mutation might still be stable and functional, but it could not be so without the structure rearranging. Predicting such rearrangements and their functional consequences is currently beyond the capabilities even of the most sophisticated modeling tools, and so this simple method has given you an answer nearly as good as can be done.

In a more systematic or formal context, the conformational search described above can be done by a function called *Autobondrot* built into the Probe program. It surveys all dihedral-angle values on a specified grid (e.g., for the χ angles of a mutated side chain) and outputs a contoured map of contact score (Word et al., 2000). If there is a sizable area in the map with score > -1 , then it is considered that the mutant can be accommodated without significant structure change. *Autobondrot* is somewhat more complex to set up than the interactive Mage/Probe exploration, but it can then be run automatically.

For crystallographers solving new structures and wanting to improve database quality directly, our Web site has the tools and instructions for generating all-atom clash and H-bond displays interactively while rebuilding models in the commonly used fitting programs O and XtalView (Jones et al., 1991; McRee, 1999; Richardson and Richardson, 2001). The improved rotamer library (Lovell et al., 2000) is available as a drop-in replacement file for either program.

The all-atom contact tools are also valuable for NMR structures, but that use is less powerful and less straightforward. H atoms are explicitly included already in NMR refinement, and NMR structures are solved in terms of local distances not absolute Cartesian coordinates; injudicious application of contact criteria could just expand the structure undesirably. So far, the most general conclusion from all-atom contact analysis of NMR structures is that for the best-determined cases the interiors excellently fit all-atom criteria and the surface regions would then benefit from a final refinement step with all atoms at 100% radius (rather than the maximum of 75% radii currently standard). All-atom contact analysis can also be used for validation of theoretical model structures, but again its interpretation is much less robust than when applied to crystallographic structures. A serious clash still means that something must be wrong, but a lack of clashes does not mean the model is necessarily correct. One must also beware, when using software that may not output completely valid PDB format, of occasional apparent clashes produced by incorrectly resolved atom-name ambiguities such as `_hg_` for an H γ hydrogen confused with `hg__` for an Hg⁺⁺ ion.

FUTURE DIRECTIONS

For the purposes of bioinformatics and structure validation, our most important plans are to develop service-provision on the kinemage Web site that will run all the functions described above on client-submitted files, now available in a preliminary form on the Mol-Probity subsite. It is always helpful to bypass the necessity for downloading, installing, and learning new programs, since even for very user-friendly software those steps are always more of a barrier than one feels they ever should be. It is now possible, from either specified PDB codes or uploaded files, to run clash reports with either visual or numerical output

directly on the web, to receive a modified file with H atoms added and optimized, and to view directly on line in Java Mage an animated kinemage showing the Asn/Gln/His flips. Along with the all-atom contact functions, we will also provide visual and numerical evaluations, both in 3D on the structure and in one dimension on the sequence, of the ideality or favorability of sidechain rotamers, of ϵ_p, ϵ_r values, and of bond-angle geometry, using our updated criteria.

Another important area for development will be further automation of the evaluation and correction functions for other side-chain types in the style now provided for Asn/Gln/His flips, and eventually for some kinds of backbone corrections. More complete automation is especially vital for use in the structural genomics effort, but it will help other users as well. We would appreciate feedback, both about patterns of use and effectiveness and also suggestions about needs and improvements.

RELEVANT WEB SITES

<http://kinemage.biochem.duke.edu/> The Richardson laboratory web site and kinemage home page is the primary source for up-to-date software, information, and other resources relevant to all-atom contact analysis; it includes documentation, datasets, and validation examples interactively illustrated in JavaMage, and the MolProbity service that runs our software on a selected or uploaded file.

<http://www.sdsc.edu/CCMS/Packages/XTALVIEW/xtalview>. The source for XtalView version 4.0 and later, which supports real-time display of all-atom contacts during crystallographic model rebuilding.

<http://origo.imsb.au.dk/~mok/o>. The source for the O crystallographic rebuilding software, with links to the kinemage site for drop-in rotamer files and macros for updating an all-atom contact display.

REFERENCES

- Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000): The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289:905–20. [1FFK]
- Bau R, Rees DC, Kurtz DM, Scott RA, Huang HS, Adams MWW, Eidsness MK (1998): Crystal structure of rubredoxin from *Pyrococcus furiosus* at 0.95 angstrom resolution, and the structures of N-terminal methionine and formylmethionine variants of Pf Rd. Contributions of N-terminal interactions to thermostability. *J Biol Inorg Chem* 3: 484–93. [1BRF]
- Beamer LJ, Pabo CO (1992): Refined 1.8 Å crystal structure of the lambda repressor-operator complex. *J Mol Biol* 227:177–96. [1LMB]
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000): The Protein Data Bank. *Nucleic Acids Res* 28:235–42.

- Brünger AT (1992): Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355:472–5.
- Connolly ML (1983): Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221:709–13.
- Crennell SJ, Garman EF, Philippon C, Vasella A, Laver WG, Vimr ER, Taylor GL (1996): The structures of *Salmonella typhimurium* LT2 neuraminidase and its complexes with three inhibitors at high resolution. *J Mol Biol* 259:264–80 [2SIM].
- Fukuyama K, Kunishima N, Amada F, Kubota T, Matsubara H (1995): Crystal structures of cyanide- and triiodide-bound forms of *Arthromyces ramosus* peroxidase at different pH values. Perturbations of active site residues and their implication in enzyme catalysis. *J Biol Chem* 270:21884–92 [1ARU].
- Ghaemmaghami S, Word JM, Burton RE, Richardson JS, Oas TE (1998): Folding Kinetics of a Fluorescent Variant of Monomeric λ Repressor. *Biochemistry* 37: 9179–85.
- Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M (1991): Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in the Models. *Acta Cryst, Sec. A* 47: 110–9.
- Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993): PROCHECK—a program to check the stereochemical quality of protein structures. *J Applied Crystallogr* 26:283–91.
- Lee BK, Richards FM (1971): The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 55:379–400.
- Lovell SC, Word JM, Richardson JS, Richardson DC (2000): The penultimate rotamer library. *Proteins Struct Func Gen* 40:389–408. Development of an improved side-chain rotamer library using a large, high-resolution dataset filtered by all-atom clashes and B-factors; analyzes systematic side-chain misfittings prevalent in the structural database.
- Lovell SC, Davis IW, Arendall WB III, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2002): *Proteins Struct Func Gen* (in press).
- McRee DE (1999): XtalView/Xfit: a versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* 125:156–65.
- Richardson DC, Richardson JS (1992): The Kinemage: a tool for scientific illustration. *Protein Sci* 1:3–9. Richardson JS, Richardson DC (2001): MAGE, PROBE, and Kinemages. In: Rossmann MG, Arnold E, editors. *International Tables for Crystallography, Volume F*. Dordrecht: Kluwer Academic Publishers, pp727–30. A brief survey of the crystallographic uses of all-atom contacts calculated by Reduce and Probe, and of the graphics system of Mage and kinemages.
- Ramachandran GN, Ramakrishnan C, Sasisekharan V (1983): Stereochemistry of Polypeptide Chain Configurations. *J Mol Biol* 7: 95–9.
- Word JM, Lovell SC, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC (1999a): Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogens. *J Mol Biol* 285:1711–33. Original description, test cases, and discussion of the all-atom contact method; explains parameter and algorithm choices in the Probe program and analyzes the contacts in a database of 100 high-resolution structures.

- Word JM, Lovell SC, Richardson JS, Richardson DC (1999b): Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* 285:1735–47. Primary reference for the program Reduce; explains H addition and optimization for protein, nucleic acid, and small-molecule “heterogens,” including the analysis of H-bond networks and the correction of Asn/Gln/His flips.
- Word JM, Bateman RC Jr, Presley BK, Lovell SC, Richardson DC (2000): Exploring steric constraints on protein mutations using Mage/Probe. *Protein Sci* 9:2251–9. Describes both interactive and batch-calculation methods for evaluating whether a specific sequence change is compatible with a given protein conformation.
- Zarembinski TI, Hung LW, Mueller-Dieckmann HJ, Kim KK, Yokota H, Kim R, Kim, S-H (1998): Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc Natl Acad Sci USA* 95:15189–93 [1MJH].