Protein Science (2000), *9*:2251–2259. Cambridge University Press. Printed in the USA.

Exploring steric constraints on protein mutations using MAGE/PROBE

J. MICHAEL WORD,[1,3] ROBERT C. BATEMAN, JR.,[2] BRENT K. PRESLEY,[1,4] SIMON C. LOVELL,[1] AND DAVID C. RICHARDSON[1]

[1]Department of Biochemistry, Duke University, Durham, North Carolina 27710-3711
[2]Department of Chemistry and Biochemistry, University of Southern Mississippi, Hattiesburg, Mississippi 39406

(RECEIVED May 17, 2000; FINAL REVISION September 1, 2000; ACCEPTED September 1, 2000)

Abstract

When planning a mutation to test some hypothesis, one crucial question is whether the new side chain is compatible with the existing structure; only if it is compatible can the interpretation of mutational results be straightforward. This paper presents a simple way of using the sensitive geometry of all-atom contacts (including hydrogens) to answer that question. The interactive MAGE/PROBE system lets the biologist explore conformational space for the mutant side chain, with an interactively updated kinemage display of its all-atom contacts to the original structure. The Autobondrot function in PROBE systematically explores that same conformational space, outputting contact scores at each point, which are then contoured and displayed. These procedures are applied here in two types of test cases, with known mutant structures. In ricin A chain, the ability of a neighboring glutamate to rescue activity of an active-site mutant is modeled successfully. In T4 lysozyme, six mutations to Leu are analyzed within the wild-type background structure, and their Autobondrot score maps correctly predict whether or not their surroundings must shift significantly in the actual mutant structures; interactive examination of contacts for the conformations involved explains which clashes are relieved by the motions. These programs are easy to use, are available free for UNIX or Microsoft Windows operating systems, and should be of significant help in choosing good mutation experiments or in understanding puzzling results.

Keywords: all-atom contact analysis; explicit hydrogens; interactive molecular graphics; kinemages; side-chain conformational maps; site-directed mutagenesis

Substitutions—replacing one amino acid side chain with another—are an important class of protein mutations. In nature, a substitution can arise due to an error in replication, transcription, or translation or even through post-translational reactions. In molecular biology, numerous techniques are routinely employed to deliberately alter the amino acid type at a specific sequence location, usually to test some hypothesis about the function of that residue. A single substitution may give rise to either subtle or profound changes in molecular properties: in function, in stability, or in structure.

When planning a substitution mutation or when analyzing experimental results obtained from such a mutant, one of the most important issues is whether or not the structure remains essentially unaltered. If it does, then the measured change in properties can be interpreted straightforwardly, whereas if the structure changes significantly, then such results will always be equivocal.

Even in the absence of high-resolution three-dimensional (3D) structures for both the background and mutant proteins, many conclusions can be drawn by extrapolating from a single accurate structure. The techniques described here, interactive MAGE/PROBE and Autobondrot, can take a static set of atomic coordinates and permit some parts to move by rotations around one or more axes. Typically, these are side-chain atoms and the axes define the side-chain torsion angles: χ_1, χ_2, etc., where χ_1 is defined as the dihedral angle around the central bond for the N–Cα–Cβ–Cγ atoms, and the other χ angles are measured progressively further out the side chain. At each conformation sampled, PROBE calculates the small-probe, all-atom contact surface for interactions between the side chain and other nearby atoms. This contact surface can be represented visually as a set of contact dots and clash spikes on the van der Waals surface of atoms at points less than 0.5 Å from the surface of a nonbonded nearby atom. Alternatively, the contact surface can be summarized by a numerical contact score with terms for H-bonds, unfavorable overlaps, and favorable van der Waals contacts. All-atom contact analysis evaluates detailed surface complementarity; the contact score is not an

Reprint requests to: David C. Richardson, Department of Biochemistry, 211 Nanaline Duke Building, Duke University, Durham, North Carolina 27710-3711; e-mail: dcr@kinemage.biochem.duke.edu.
[3]Present address: Glaxo Wellcome, Inc., Research Triangle Park, North Carolina 27709.
[4]Present address: University of Texas Southwestern Medical School, Dallas, Texas 75219.

energy, and as currently formulated, the better the packing, the more positive the score. For a detailed description of small-probe contact displays, contact scores, and the PROBE program, see Word et al. (1999a). An early example of using all-atom contacts to evaluate a mutation is described in Ghaemmaghami et al. (1998), where a buried Trp mutant chosen for its good PROBE score was found experimentally to have essentially unchanged (if anything, slightly improved) stability and also maintained equally rapid folding kinetics. Use in nucleic acids, to model a phosphate attached to O5′ of an HDV ribozyme cleavage product, is described in Wickham and Word (1999).

The contact surface generated by PROBE can be thought of as the inverse of solvent-exposed surface in the sense that it displays that which is thrown out by the traditional surface algorithms (such as the Langridge–Connolly dot surfaces; Langridge et al., 1981). Dot representations are useful in both these techniques because of the advantages in interactive molecular graphics of both seeing the surface and seeing beyond the surface. Furthermore, dots and spikes are more distinct and easier to pick or label than are continuous surfaces, allowing positive identification of the atoms responsible for each part. Other common illustration techniques that use continuous surfaces—such as CPK spheres (Porter, 1978), semi-transparent continuous surfaces (Nicholls et al., 1993), and "egg-shell" solid contact surfaces (MacKenzie et al., 1997)—produce a more familiar appearance in static illustrations, but contact dots are much easier to interpret for serious scientific detail in crowded molecular environments.

Simple mutation analysis with all-atom contacts can be done either interactively in the MAGE display (Richardson & Richardson, 1992, 2000; i.e., the MAGE/PROBE system) or by generating plots of contact score vs. conformation with the Autobondrot function of PROBE. From either such exploration of how the contact surface varies with changes in side-chain conformation, one can readily determine whether or not nearby groups, in their given positions, can accommodate the substituted side chain and over what range of angles. This simple process provides most of the payoff obtainable by a predictive analysis. However, if desired, an additional step can be done to test the benefit of moving another side chain. Movement of main chain, however, is not only more difficult to model, but is presumed to have unpredictable and perhaps widespread consequences and is not attempted here.

For this sort of analysis to be reliable, the reference structure needs to be accurate enough that the positions of groups near the site of mutation can be trusted to better than half an Ångstrom; in our experience, crystal structures better than 2.0 Å resolution are desirable. NMR structures with more than about 20 restraints per residue in the region of interest should support this method, but we have not tested such use. Even given a high-resolution structure, reliable analysis cannot be performed in regions of disorder as indicated by high crystallographic temperature factors (*B*-factors) or multiple conformations.

Another requirement is that hydrogen atoms must be explicitly included in the molecular models, because van der Waals surface complementarity is very sensitive to the local geometry. United-atom or implicit representations of hydrogen atoms (in which the van der Waals radius of the nonhydrogen atom is increased), although commonly employed, ignore the individual hydrogens' significant influence on local packing and often produce ambiguous or incorrect results in this context. Therefore, the program REDUCE (Word et al., 1999b) is first used to add hydrogens to the structure and to optimize the positions of movable ones.

Many other approaches to modeling side-chain substitutions have been described, most of which attempt to determine the actual mutant conformation and its energy. One popular family of techniques involves building a model of the mutant and performing energy minimization or molecular dynamics runs. Some or all of the side chains and prosthetic groups are free to move. Depending on the force field and strategy, the backbone is either fixed or free to move. Comparison of both energy and structure with the nonmutated background is used to evaluate the mutation. This family of techniques is powerful, but energy-based methods almost always give a plausible result—they rarely say no. Ironically, this means that yes may not always mean yes, and it is left up to the user to decide how much change is unacceptable. Explicit estimation of the change in stability ($\Delta\Delta G_u$) for a substitution involves a thermodynamic cycle in which atoms and even charges are added or removed (e.g., Zeng et al., 1999); this can be quite accurate for relatively small conformational changes, but is not easy for the nonexpert to perform and is computationally demanding. If changes are larger, especially if the backbone moves, then it is unlikely that current techniques will produce the right answer.

The method described here provides an easy way of answering a more limited question: is or is not the proposed mutation compatible with an essentially unchanged surrounding structure? This question is probably the most important one that can be asked in advance, because if the surroundings must move, then even if the mutant is stable (which it may well be) there is no straightforward way to analyze the meaning of whatever changed properties are observed. Similarly, if a mutation is found to produce unexpected consequences, this simple analysis can show in hindsight whether conformational changes are likely to be the confounding factor.

Results

Ricin

As an initial example of MAGE/PROBE and Autobondrot use, a previously successful mutant predicton can be recapitulated and further understood.

Castor bean seeds contain the toxic protein ricin whose A chain (RTA) is an enzyme that inactivates eukaryotic ribosomes by removing a specific conserved adenine base from 28S rRNA (Endo et al., 1987). Yeast genetic studies (Frankel et al., 1989) and analogy with the homologous Shiga-like toxin I (Hovde et al., 1988) suggested that glutamic acid 177 may be required for catalysis. In an attempt to determine the precise role of Glu177, Schlossman et al. (1989) cloned and tested several active site mutants. The mutation E177D lowered activity by a factor of 80 but, disturbingly, E177A lowered activity only by a factor of 20. At the time, we observed that in the wild-type ricin crystal structure Glu208 appeared to be close enough to stand in and rescue activity (Frankel et al., 1990). The double mutant testing this hypothesis, E177A–E208D, was found to be completely inactive. Then, definitively, a crystal structure of the E177A RTA mutant (Kim et al., 1992) confirmed that Glu208 does indeed adopt the proposed alternative conformation.

The modeling then used to predict the role of Glu208 was quite primitive: it consisted of noting that, in the 2.3 Å ricin structure 1RTC (Mlsna et al., 1993), the side chain could be oriented to reach the active site, and thus speculating that it must be able to support catalysis, based on the unexpected experimental result. In advance of that result, one could not predict whether this carbox-

ylate position would produce activity. However, it would now be possible in advance to determine whether or not there is enough room to fit the side chain into the new position without deforming chemical bonds, putting atoms on top of one another, or significantly altering the surrounding structure. Such steric constraints are particularly limiting when all the hydrogen atoms are explicitly represented in the model. Figure 1 shows the interactive MAGE/PROBE display of an E177A model (based on the more recent 1.8 Å RTA structure 1IFT, Weston et al., 1994) showing the all-atom contacts to Glu208. In this picture, the Glu is positioned to hydrogen bond with Tyr123's phenolic oxygen (as Glu177 did in the wild-type), as represented by the lens-shaped set of dots to the left. Contact dots are on the van der Waals surface of an atom; clicking on a dot displays the atom name. The kinemage is interactive; as the χ angles are varied using the sliders, the contact display is automatically updated. (The reader can try this out, using Kin.3 of supplementary file WordRicn.kin.) Exploring different conformations, the great majority of them are found to clash with the atoms of neighboring groups. The conformation shown in Figure 1, however, places the carboxylic acid in approximately the same spot as Glu177 previously occupied, with good positive interactions and no steric conflicts.

Simple interactive exploration suggested that there are at least two conformations that accomplish the goal of avoiding clashes and putting the carboxylate near the right spot. Are there any others? And how large are the acceptable regions? An automated conformational survey was performed over all three side-chain torsion angles for Glu208 using the coordinates from 1IFT. The Autobondrot procedure was used to sample each angle at 5° increments (10° for χ_3) and calculate a PROBE score for the side chain assuming fixed conformations for the main chain and all other side chains (waters were not considered). The PROBE score was combined with a torsional factor for χ_1 and χ_2, but not χ_3, as described in Methods. The scan required 81 min on our Silicon Graphics Indigo2 workstation.

Figure 2 shows the results of this scan, a map of the three-dimensional space that describes Glu208 side-chain conformational freedom. The gray mesh encloses conformations in which Glu208 has acceptable contact scores (> -1; see Methods). Conformations outside of this mesh are clashing with neighboring atoms. Also, there are indeed only two conformational regions, outlined in black, which put the carboxylate oxygens in approximately the same position as those in Glu177 (the two are centered at $\chi_1 = -105°$, $\chi_2 = -51°$, $\chi_3 = 90°$ and $\chi_1 = -167°$, $\chi_2 = 53°$, $\chi_3 = -5°$). Each of these regions encloses a narrow range of conformations that also have acceptable contacts—candidates for how Glu208 might be rescuing enzymatic activity in the mutant. The PROBE scores do not distinguish between these two alternatives, because the nearly eclipsed side-chain dihedral of the first alternative is compensated by better contacts. In the end, the first region (at top in Fig. 2) is further favored by a slight shift of Arg180 to ion pair with Glu208; this conformation is the one observed in the E177A mutant crystal structure (Kim et al., 1992).

T4 lysozyme

Bacteriophage T4 lysozyme is in many ways a molecular "lab rat"—a system that is well behaved and well characterized and that can be readily manipulated for study. Over the past three decades, Matthews and co-workers have extensively studied T4 lysozyme, developing a large and unique database of mutants for which there are activity and stability measurements and in most cases high-

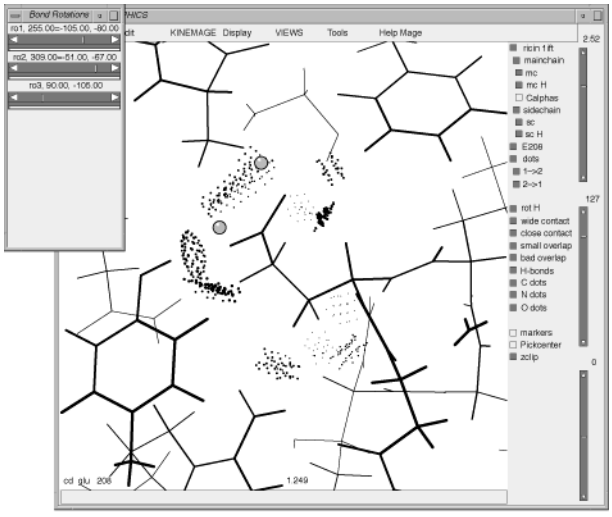

Fig. 1. MAGE/PROBE display of ricin E177A active site mutant showing the alternative catalytic "rescue" conformation of Glu208 (centered), with its interactive contact surface dots. The surrounding structure is from the wild-type ricin A chain structure 1IFT (Weston et al., 1994). REDUCE was used to add hydrogens and a kinemage was created in PREKIN, with an Ala substituted for Glu at position 177 and a rotatable Glu208 side chain. Ala177 is visible above the two gray balls that mark the oxygen positions of wild-type E177. Adjusting the sliders at the upper left results in rotation of side-chain torsion angles χ_1, χ_2, and χ_3 of Glu208 and the recalculation of the all-atom contacts. In the interactive display, contact surfaces are color coded to indicate van der Waals surface complementarity. This interactive kinemage is in file WordRicn.kin of the electronic supplement.

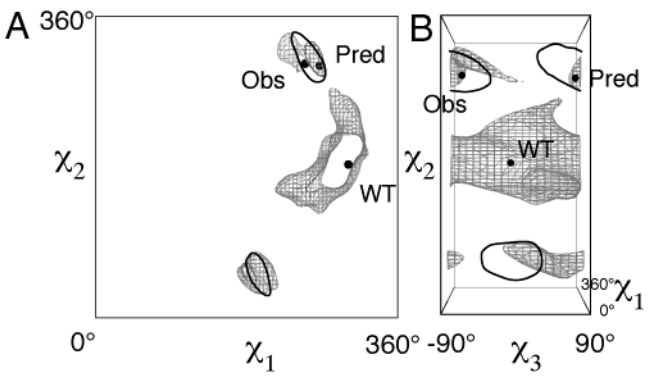

Fig. 2. (**A**) Face and (**B**) side views of a 3D contour map summarizing an Autobondrot scan of Glu208 side-chain conformations in a model of the ricin E177A mutant. The model is constructed from the high-resolution wild-type structure 1IFT, in which Glu177 was trimmed back to Ala, waters were omitted, and hydrogens were added with REDUCE. The gray contour mesh encloses favorable conformations where the contact score is greater than -1. The heavy black lines enclose the two regions (near the top and bottom) where the carboxylate oxygens for Glu208 have a summed distance of <2.5 Å from those in the original Glu177. The full range is shown for each torsion angle, so opposite edges wrap and a line that extends beyond the right edge continues on the left. Solid points mark: (WT) the conformation Glu208 had in the wild-type structure, (Pred) our original prediction, and (Obs) the conformation observed in the crystal structure of the E177A mutant (Kim et al., 1992). The full 3D contour maps are in supplementary file WordRicn.kin.

resolution X-ray crystal structures (Matthews, 1995). This database has been gleaned, by that laboratory and by others, to yield many significant insights about the sensitivity of protein stability and structures to mutations and the determinants of those effects (e.g., Karpusas et al., 1989; Faber & Matthews, 1990; Baldwin et al., 1993, 1996; Blaber et al., 1993a; Gassner et al., 1996; Vetter et al., 1996). In the process, the early simple description of "temperature-sensitive" mutants has evolved into a sophisticated understanding of the thermodynamics of protein stability.

Here we make use of this resource, building models of several mutants based on the "pseudo-wild-type" structure WT* (C54T/C97A with PDB code 1L63; Matsumura & Matthews, 1989) and comparing those with the actual observed structure for each mutant protein. We have chosen six single-site mutants, of resolution 1.7–2.05 Å, which substitute a leucine for either alanine, serine, phenylalanine, or methionine. Analysis of just $x \rightarrow L$ mutations facilitates case-to-case comparisons, and the set covers a range of local environments as well as a range of responses to the mutation. In each case, structural changes (or lack thereof) in response to the mutation were described in detail in the primary reference for each structure, which we will quote from; the intention here is to focus on the limited but very useful inferences that can be made when using conformational modeling of steric interactions with fixed backbones and neighboring side chains. It should be noted that each of the selected mutations was successful in yielding a stable, well-ordered protein.

The six sites vary in the extent to which the wild-type structure could accommodate the substitution without moving, as revealed in contour maps of the total score (PROBE score + torsional penalty; see Methods) for all combinations of side-chain dihedral angles, χ_1 and χ_2 (Fig. 3). The iso-volume substitution M120L (Lipscomb et al., 1998) is at a site with "partial solvent exposure" near the center of a short helix. The maximum total score for the model (Fig. 3A) is +8 in a conformation with considerable positive van der Waals contacts. In the mutant structure (Fig. 3B), the conformations of the main chain and neighboring side chains do not substantially change from that in WT*, so the maps are very similar. The regions predicted from the model to accept the substitution (score > -1; shaded areas in Fig. 3) do indeed contain the observed mutant conformation.

Site S44L (Blaber et al., 1993b, 1994) is in many ways even simpler (Fig. 3C,D). Being "fully solvent exposed," the acceptance regions for both the model and the observed mutant outline the major leucine rotamers and primarily reflect contacts with local main-chain atoms. A complication, however, is that the leucine conformation observed in the mutant structure (marked with an X) lies outside even the generous acceptance criteria for predictive models, implying considerable strain (score −3); also, that position is slightly less, rather than more, favorable in the mutant than in WT*. There seems no reason for Leu44 to avoid one of the more favorable available conformations, because it does not appear to be constrained by its surroundings. When crystal contacts are considered, the score for the reported conformation is not improved, although χ_1 trans is now excluded (Fig. 3X). We note that this side chain has been built into a conformation (which we will call mp* using nomenclature described in Lovell et al., 2000) that is pseudo-symmetric with the conformation of the unstrained principal leucine rotamer mt (marked in Figs. 3C,D,X with a circled-plus; see Lee & Subbiah, 1991; Petrella et al., 1998; Lovell et al., 2000). Both conformations position the terminal methyls in approximately the same location and can often appear to fit less-than-optimal electron density equally well (unfortunately, we could not check the density directly, because structure factors were not deposited). However, Cγ is offset in opposite directions in the two conformations—the difference is about an Ångstrom—and in this case, Cδ1 differs by more than Cδ2. A comparison of the crystallographic *B*-factors of the side-chain carbons can be diagnostic; they describe not only atomic motion but also the local quality of the electron density and how well the atom's position matches that density. For Leu44 (*B*-factors: Cα 24 Å^2, Cβ 24 Å^2, Cγ 35 Å^2, Cδ1 35 Å^2, Cδ2 25 Å^2), the higher *B*-value for Cγ than for Cδ2 suggests that Cγ may not be centered in the electron density and confirms that the rotameric conformation mt is probably a better choice for this side chain (score +4). We have previously suggested (Lovell et al., 2000) that the leucine mp* conformation and an analogous tt* conformation are, in fact, essentially always mistakes, resulting from ambiguous electron density, the confusing similarity to major rotamers, and the recycling of these mistaken conformations from earlier structures into entries in standard rotamer tables. If we accept the suggestion that the actual conformation is mt, we see from the similar contour maps that the unstrained local environment for the model of S44L (Fig. 3C) is a good representation of the actual mutant.

In M106L (Lipscomb et al., 1998), a mutation of "the most solvent exposed methionine of T4 lysozyme," the score map made using the original WT* Cβ (Fig. 3Y) has a maximum score of +4.6, but the mutant shows a dramatically different conformational map (Fig. 3F) due primarily to a change in main-chain angles, which tilts the Cα–Cβ bond vector and shifts the space sampled by the 106 side chain. However, Met106 in the background structure has significantly nonideal bond angles, especially around Cα. Therefore, we calculated a score map (Fig. 3E) with an idealized Cβ position (see Methods), which differs by 0.9 Å; Figures 3E and 3F now match quite well, and both show good maxima at the observed mutant conformation. This result persuaded us to adopt the strategy of always idealizing the Cβ in predictive model calculations. According to Lipscomb et al. (1998), "the introduction of the leucine side chain does not result in steric interference with neighboring protein atoms or in the formation of cavities." If the mutant is not forcing the backbone movement, the change may come from the release of strain in WT*. Alternatively, the WT* structure may be misleading in this region: Met106 is more disordered than Leu106 in the mutant, and perhaps it adopts the mutant backbone structure (or even the mutant side-chain χ_1) part of the time. In any case, if the modeling is done with idealized Cβ, it correctly predicts that the M106L mutation is consistent with very minimal perturbation of the surrounding structure.

For mutant F153L (Figs. 3G, 3H; Eriksson et al., 1993), the model's acceptance region is extremely small, with a maximum score of +1. This represents a borderline case, where there are "moderately large shifts of several side chains toward the mutation site" (to close down around the deleted atoms) and slight shifts in the main chain: twisting of the alpha helix containing the mutation by ~2° and slight shift of an adjacent helix (providing a bit more room for the Leu methyls). The model's score map (Fig. 3G), with a very small acceptable region, should be read as indicating that some modest reorganization would probably be required to accommodate the mutation comfortably. Therefore, the conservative interpretation of such results should consider size as well as height of the most favorable region.

Finally, mutants M102L (Hurley et al., 1992) and A129L (Baldwin et al., 1996) can be seen to require significant structural reor-

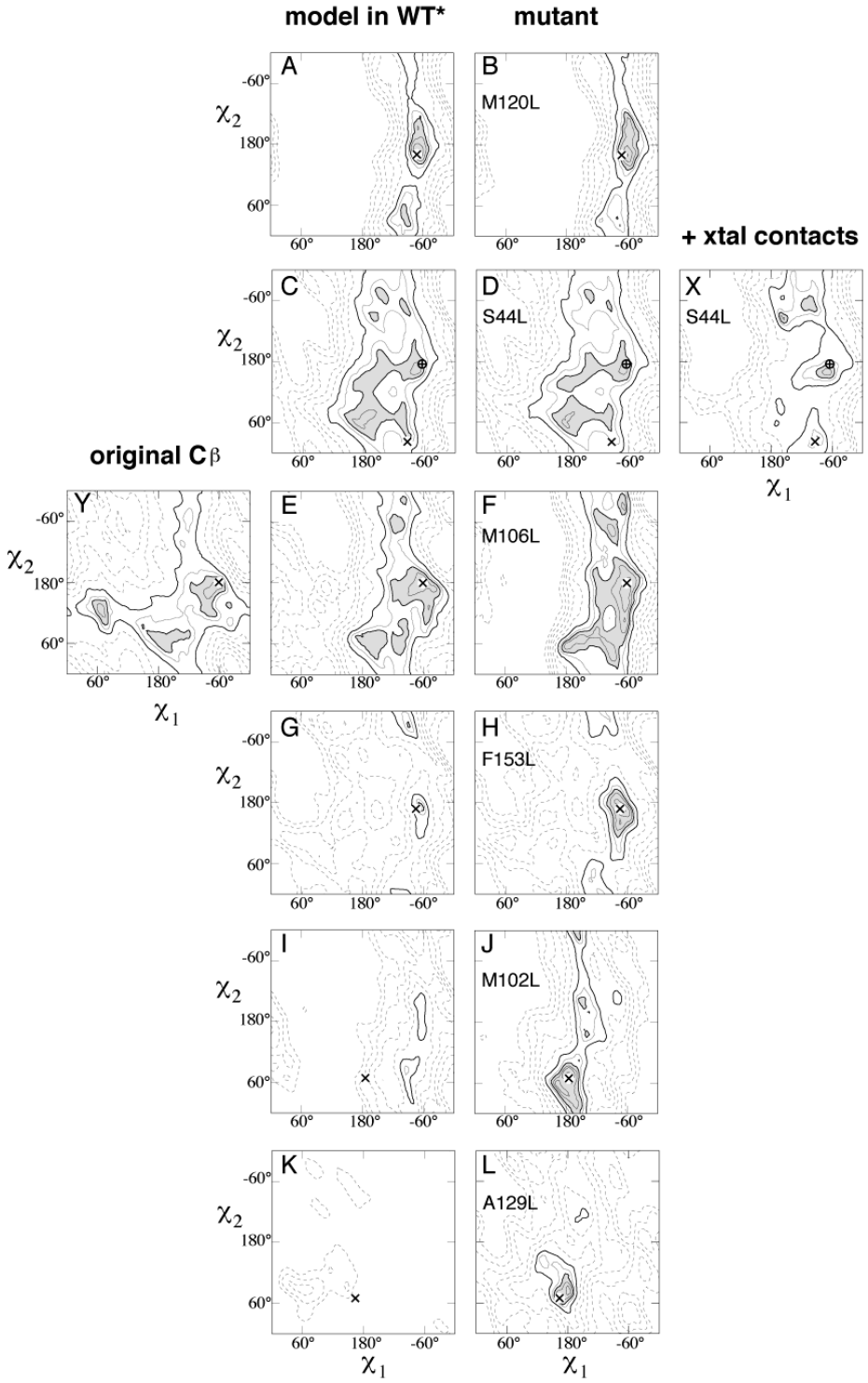

Fig. 3. Autobondrot side-chain conformational maps for six T4 lysozyme mutants, calculated in the context of the static protein structure. For each mutant, total score (PROBE score + torsional potential) contours are shown for a leucine with idealized geometry, both in the "pseudo-wild-type" structure WT* (left column) and in the observed crystal structure of the mutant (right column). Shaded regions have scores greater than −1, indicating acceptable side-chain packing. Solid black contours are at scores of −10 and −1. Dashed lines are contoured by 10s from −60 to −20, and solid gray contours are by 3s between −4 and the highest level of +8. Conformations observed in the mutant crystal structure are marked with an X. In (**C**), (**D**), and (**X**), the circled-plus sign marks the conformation of the highest-populated leucine rotamer ($\chi_1 = -65°$ and $\chi_2 = 175°$), which occupies a position in space close to the conformation reported for the mutant crystal structure. S44 is on the exterior of the protein and (**X**) shows the map for the S44L mutant when crystal contacts are considered; the map changes but the conclusion remains. The background structure used in the models (**A, C, E, G, I, K,** and **Y**) has PDB identifier 1L63. The mutant structures used were: (**B**) 233L, (**D**, **X**) 110L, (**F**) 234L, (**H**) 1L87, (**J**) 1L77, and (**L**) 195L. Hydrogens were added to each structure with REDUCE. The position of the Cβ was idealized to standard bond lengths and angles when modeling each of the substitutions in WT*: panel (**Y**) was done with the original Cβ for M106, the only site for which this made a significant difference. Examples of these structures and their all-atom contacts are animated in supplementary file WordT4Lz.kin.

ganization (Figs. 3I, 3J, 3K, 3L). The WT*-based models have maximum scores of −2 and −37, respectively. Unlike the other cases considered above, the change in free energy of unfolding found for the mutant protein relative to WT* is negative (−0.7 kcal/mol for M102L and −1.3 kcal/mol for A129L). That is, these two mutations are destabilizing, and the induced strain from A129L is calculated to be −2.6 kcal/mol (Baldwin et al., 1996). The Cβs shift by 0.5–0.7 Å and several nearby side-chain atoms move by 1 Å or more; in M102L the Cϵ of nearby Met106 moves 2.8 Å. For an Ala to Leu mutation in a protein interior, this is hardly surprising, but even the iso-volume Met to Leu mutation changes the side-chain shape and reduces the degrees of freedom available to the side chain. In each of these two cases, the model score map clearly signals that the mutation may not be stable but that even if it is, the background structure is not a good stand-in for the mutant structure. Therefore, any simple interpretation of measured mutant properties would be rendered suspect.

For each mutation, an important aid to the interpretation of the contour map involves going back to the interactive MAGE/PROBE kinemage and studying the all-atom contacts and clashes for conformations in each of the candidate regions of the map (see, e.g., file WordT4Lz.kin in Supplementary material in the Electronic Appendix). If polar groups are involved, H-bond geometry can be evaluated directly, as in the ricin case. Even when the map shows good scores, looking at the structure can reveal complicating factors (e.g., buried charges or cavity formation) that are not considered in this method but may destabilize the mutant protein or force the mutant to adopt a different structure. To verify, either interactively or quantitatively, whether a charged or polar group can reach a solvent-exposed position, PROBE can be run in a mode that displays or evaluates solvent-accessible area. On the other hand, neighboring groups that constrain the mutant side chain may be shown to be relatively unconstrained themselves. Autobondrot provides an overview of conformation space, revealing the trends, while critical inspection of the atomic structure and its contacts provides a complementary detailed view of the elements—their identity and the weight they should be given—of which these trends are composed.

Discussion

The MAGE/PROBE and Autobondrot methodology has a restricted but important purpose, and in keeping with its circumscribed ambitions it is simple and easy to use. It determines whether or not the surrounding structure must move to accommodate a mutation. If this method shows an acceptable mutant conformation within the background context (a substantial region in χ space with contact score = −1) then, even in the absence of determining the new structure, one can with reasonable confidence attribute any observed changes in function or stability to direct effects of the side-chain substitution. Steric considerations are quite generally an issue in such mutational work, because very few useful types of amino acid substitutions are sterically harmless. For instance, in testing the importance of a hydrogen bond, Thr → Val is not safe in the general case, because the new methyl group is quite constrained and much larger than the OH; even Glu → Gln introduces two additional H atoms. MAGE/PROBE, however, could quickly show whether such a change actually presents any problem within a specific context. These tools also illustrate some of the very substantial enhancements to kinemages and the MAGE display program that have been made since it was first developed for *Protein Science* (Richardson & Richardson, 1992).

The procedure described here is designed to encourage the bench scientist to consider the molecular geometry when planning or analyzing a mutation. It is a natural extension to simple visual inspection of a molecular model from X-ray crystallography or NMR, and it promotes exploration. Interactive graphics encourages thinking about spatial arrangements and consideration of the relative impact of nearby groups, complementing the analysis of summary statistics such as contact scores or energies. MAGE, PREKIN, and PROBE cooperate to make exploration of side-chain contacts straightforward. Autobondrot surveys conformational space systematically, showing the number and size of favorable regions, usually leading back to the interactive display to model some point on the score map.

This methodology does have a number of limitations that should be kept in mind, although none of them invalidates the proposed uses. Electrostatics are only used to recognize hydrogen bonds; long-distance charge–charge attraction or repulsion is ignored. Similarly, because the all-atom contacts are strictly local (within 0.5 Å), cavity size is not measured. The energetic consequences of a buried charge or of cavity formation would have to be calculated independently. In fact, no energetic terms are evaluated directly in this method: its essence consists of isolating the purely geometrical issue of structural compatibility and presenting it in a form that is especially sensitive to details and that shows the individual positive and negative interactions explicitly. Fixed, idealized bond lengths and angles are used for the mutant side chain. Small changes in bond angles, and any plausible change in bond lengths, are generally within the tolerances of the acceptable score limits adopted here. In our experience, larger distortions in bond angles do not occur in accurate protein structures except where there are other reasons to suspect a fitting error or perhaps occasionally as part of a strained active site conformation; therefore, we believe that allowing such variability would probably degrade rather than improve these mutation analyses.

The most obvious limitation is reliance on fixed nearby side chains and fixed main-chain position. It is feasible to model some changes to nearby side chains, and, in fact, the ricin case is an example of that process, because Glu208 is a neighbor of the original E177A mutation. As shown convincingly by Matthews and co-workers for T4 lysozyme, it is not at all unusual for the main chain to change conformation (Baldwin et al., 1993). They argue that "protein backbones are more flexible than generally assumed," and that the modeling of side-chain steric conformations using fixed main chain and neighboring groups is "overly restrictive." That is certainly true, and it poses a real difficulty for modeling done with the purpose of reliably predicting a protein's response to mutation or for homology modeling, because the current state of the art is not capable of accurately calculating shifts in backbone conformation. Such shifts are also especially likely to produce nonlocal effects. Our technique pursues a more modest goal: predicting whether or not the structure must change but not how it will change. In searching molecular conformations, these procedures act as conservative filters. As seen for the A129L mutation above, they may very well "reject" a mutation that later proves to be stable; however, they are not rejecting its possible viability, they are rejecting it as an experiment that can produce easily-interpretable results.

The major alternative method for analyzing the suitability of proposed mutations would be energy minimization or molecular

dynamics calculations on both wild-type and mutant, which is capable of yielding a wider range of information. However, if done by simple protocols, it has some of the same limitations discussed above, while doing more sophisticated and reliable calculations requires considerable expertise. In either case, interpreting the significance of the resulting structures and energies is not at all straightforward. MAGE/PROBE and Autobondrot, on the other hand, are easily accessible to the nonexpert, are scripted for this particular use, and provide more limited but readily interpreted results. These MAGE/PROBE/Autobondrot methods act as screens for whether or not a proposed mutation can be accommodated without other shifts (so that its results can be interpreted straightforwardly). For the cases examined so far, the method indeed works as intended. Its practical significance is to provide an easy way for improving the experimental design of substitution mutations

Methods

The computer program MAGE displays interactive three-dimensional graphics from a kinemage, a structured text file containing object descriptors, coordinates, identifiers, and display parameters (Richardson & Richardson, 1992, 1994, 2000). MAGE versions 5.4 and above for Unix and Microsoft Windows (95 or later) have been modified to interact with separate external programs that generate new geometric objects that are then incorporated into the current graphics display. The hydrogens necessary for all-atom contact analysis are first added to the Protein Data Bank (PDB) (Bernstein et al., 1977; Berman et al., 2000) file with the program REDUCE (Word et al., 1999b), using the command "reduce -build 1xyz.pdb >1xyzH.pdb." Then a kinemage display file is created with PREKIN, using the command "prekin - lots 1xyzH.pdb 1xyzH.kin."

MAGE's program-to-program communication is used in the work described here in two ways. First, a rotatable side chain (either mutated or not) can be constructed with the assistance of PREKIN. PREKIN is the primary feeder program for MAGE, designed to read PDB format atomic coordinates and create a kinemage to illustrate them with a choice of styles or subset selections. The version of PREKIN used for this work is 5.71. Although PREKIN can independently construct a kinemage with a mutation or a rotatable side chain, the MAGE/PREKIN link can even more conveniently be used to modify a structure while it is being viewed and analyzed in MAGE. Clicking on a residue, and then choosing the "Remote Update . . ." tool sets up a command instructing PREKIN to either make the selected residue rotatable or to replace it with a rotatable version of a different amino acid. PREKIN uses standard Engh and Huber (1991) geometry to build mutant side chains. In this work, models using the actual mutant crystal structures were built with the reported Cβ coordinates. However, because the bond angles out to a Cβ are sometimes significantly distorted, in our predictive models the Cβ position for the mutant was standardized to better represent an unstrained side chain. This was accomplished by constructing two locally ideal-geometry Cβs: first using the backbone N (i.e., N–Cα–Cβ angle and C–N–Cα–Cβ dihedral), and second, using the backbone C (i.e., C–Cα–Cβ angle and N–C–Cα–Cβ dihedral), averaging those two constructed positions, and then idealizing the Cα–Cβ bond length. That procedure is built into PREKIN 5.71 and later, for mutations but not for rotation of an existing side chain.

The second way MAGE is used to interact with external programs is by calling PROBE, to generate an all-atom contact display of the sort described in Word et al. (1999a) and shown in Figure 1. Once the link is set up (using the "Remote Update . . ." dialog), the generated contact information is coupled with the rotation of torsion angles in MAGE: each time an angle is adjusted, MAGE calls PROBE, passes PROBE the modified coordinates, reads the results, and displays the updated contact dots and clash spikes. (To try this, use Kin.3 of file WordRicn.kin.) Despite the fact that PROBE is being restarted each time the conformational angles are adjusted, the process is very quick, allowing interactive exploration. Usually, PROBE is instructed to combine the rotated coordinates from MAGE with the static atomic coordinates from the original PDB file. To do this, MAGE generates a command line that uses PROBE's flexible method of selecting sets of atoms to make the rotated coordinates supercede the originals.

In all the above interactions, MAGE communicates with the external programs using pipes, a computing facility widely supported under Unix and some other operating systems. The only requirement for an external program to work with this technique is the ability to read ATOM records from the "standard input" stream and write geometry to "standard output" in kinemage format. The MAGE/PREKIN and MAGE/PROBE links operate locally. Although pipes are a simple, robust technique for communication between programs running on the same machine, the way we are using them is not well suited for communication between programs running on separate machines because the overhead of repeatedly restarting a program remotely can be much greater than doing the same locally. For MAGE to efficiently call programs remotely, we would need to modify the remote program to act more like a server by maintaining a persistent connection. Currently, on the Macintosh a relatively high process-creation overhead also seems to recommend a persistent connection. Versions of PROBE that persistently communicate with MAGE may be developed in the future, but it is clear that one of the advantages of our current implementation is flexibility in the choice of external programs, because they now need no modification to communicate with MAGE and can, therefore, continue to evolve independently without losing the ability to be linked.

The MAGE/PROBE link described above permits interactive exploration of conformational space. However, if this space is large because several torsion angles must be explored together, then sampling by hand is inadequate, and a more automated method of examining conformations is needed. PROBE version 2.0 has a new feature called "Autobondrot" to iterate over a range of dihedral angles, determining the contact surface at each step and calculating a numerical score. This score, along with each of the angular coordinates, is written out in tabular format. The Autobondrot procedure is controlled by a rotation script (a .rotscr file), which specifies the dihedral axes, angle ranges, and sampling frequency, and the static and rotated atoms. A conversion tool (mkrotscr) will translate any rotations in a kinemage file (usually sidechain rotations set up by PREKIN) into the dihedral scan pattern in a .rotscr file, which the user can customize further if needed (e.g., to change the sampling interval).

Relying on exhaustive enumeration at a sampling rate of 10 to 20 conformations per second on a 250 MHz R10000 Silicon Graphics workstation, Autobondrot is most appropriate for scans involving one to four torsion angles. For most uses, side chains with up to three torsion angles are sufficiently well sampled every 5° in χ_1, 5° or 10° in χ_2, and 10° or 15° in χ_3. The longer side chains of Lys and Arg with four torsion angles can be rapidly surveyed by making three runs, each with a fixed χ_1 at one of the three preferred

staggered angles −65°, −177°, and +62° (Lovell et al., 2000), and then plotting each block of data separately in 3D. Sampling only near "rotameric" conformations where side chains have previously been observed in a number of proteins can speed the scanning somewhat, but the benefits of a comprehensive survey map usually outweigh the marginal increase in speed. The Autobondrot scores for an isolated side chain recapitulate the boundaries of allowed regions within which the rotamers occur, so that the information is included implicitly.

PROBE scores are determined by dividing the contact dots into categories of favorable contacts and H-bonds or unfavorable clashes, scaling each appropriately and summing (see Word et al., 1999a, for a detailed description of how scores are computed). In the procedure described here, scores were calculated only for movable atoms in the rotated side chain. An important step for many types of conformational scans is the addition of a penalty function representing a torsional barrier to bond rotation, as is done in most force fields. This is necessary even with explicit hydrogen van der Waals contacts included, because the primary torsional effect is an intrinsic property of the bond, related to the hybridization of the bonded atoms (Momany et al., 1975; Streitwieser & Heathcock, 1976). The torsional penalty varies with the cosine of the dihedral angle (e.g., from 0 at staggered angles to s at eclipsed angles) and is added to the PROBE score as follows:

$$\text{Total score } (\chi^*) = \text{Probe score } (\chi^*) + \frac{1}{2}\sum_i s_i(1 - \cos[n_i(\chi_i - \delta_i)])$$

where χ^* refers to the conformation of the group as a whole, s_i is the scale factor between torsional strain and PROBE score for the i^{th} χ dihedral angle, n_i is the number of barriers in a full rotation, and δ_i is the phase angle where the penalty vanishes. For χ_1 of all residues except Gly, Ala, or Pro, and for other Csp^3–Csp^3 dihedrals, $n = 3$, $\delta = 60°$, and $s = -3$. For χ_3 of disulfides (the S–S bond), $n = 2$, $\delta = 90°$, and $s = -6$. For dihedrals adjacent to flat groups (such as χ_2 of Phe or Asp), and for dihedrals that rotate an OH or SH group, s is taken to be 0. Torsional scale factors for S–C bonds and several other special cases have not been determined, but they should almost certainly be smaller than −3; by default, we set them to 0. The torsional scale factors, and the conformational acceptance criterion of TotalScore > -1.0, were selected to approximately match PROBE scores to observed side-chain distributions from a database of 240 high-resolution structures with high *B*-factor residues omitted (Lovell et al., 2000). Side chains with total scores above the acceptance cutoff, even if they clash slightly, are assumed capable of relieving that strain by deviating modestly from standard bond angles or making very small shifts of surrounding atoms, without large effects on stability or conformation.

The output PROBE scores for each combination of dihedral angles are most readily analyzed graphically. Our programs, KIN2DCONT and KIN3DCONT (Word, 2000), are used to generate 2D or 3D contour maps in kinemage format (e.g., see file WordRicn.kin) from tabular data. The contour maps summarize the trends in the data, while individual values can be identified by clicking on the contours in MAGE. The contouring routines can also write PostScript output directly, as done to produce Figures 2 and 3, with some editing in Adobe Illustrator.

Although the Autobondrot procedure is described here in terms of sampling torsional angles, it can also rotate around or translate along arbitrary directions in space. This could be used to perform simple docking calculations (a capability we have not yet exploited). Autobondrot is also suitable for calculations on nucleic acid conformations, as was done in Wickham and Word (1999). The carboxyl-distance calculations shown for ricin E208 in Figure 2 were done outside of PROBE in a separate scanning procedure (bondrotscan) that can evaluate an arbitrary function at each conformation. This procedure is implemented as an AWK script and is, therefore, much slower than the Autobondrot function in PROBE, which is written in C.

All of the programs used in this work (including MAGE and PREKIN by D.C.R. and PROBE, REDUCE, KIN2DCONT, and KIN3DCONT by J.M.W.) are freely available from the Web/FTP site at kinemage.biochem.duke.edu, in versions for UNIX or Microsoft Windows operating systems.

Supplementary material in the Electronic Appendix

Two kinemage files and a small PDB file are included. WordRicn.kin introduces all-atom contacts, gives the 3D version of Figure 2, and (along with the PROBE program and the 1iftE2Aw.pdb file) allows interactive MAGE/PROBE exploration of the ricin mutant. See above for availability of the latest MAGE and PROBE programs needed to run the interactive demo on UNIX or PC. WordT4Lz.kin shows an overview of the six T4 lysozyme mutations discussed, and for both structure-compatible and incompatible examples animates between WT*, the modeled mutation, and the actual mutant structure, with all-atom contacts.

Acknowledgments

Jane Richardson contributed extensive useful discussions. This work was supported by research grant GM-15000 to D.C.R. from the National Institutes of Health, by sabbatical leave to R.C.B. from the University of Southern Mississippi, and by educational leave to J.M.W. from Glaxo Wellcome, Inc.

References

Baldwin EP, Hajiseyedjavadi O, Baase WA, Matthews BW. 1993. The role of backbone flexibility in the accommodation of variants that repack the core of T4 lysozyme. *Science 262*:1715–1718.

Baldwin EP, Xu J, Hajiseyedjavadi O, Baase WA, Matthews BW. 1996. Thermodynamic and structural compensation in "size-switch" core repacking variants of bacteriophage T4 lysozyme. *J Mol Biol 259*:542–559.

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*:235–242.

Bernstein FC, Koetzle TF, Williams GJ, Meyer EE, Brice MD, Rodgers JR, 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.

Blaber M, Lindstrom JD, Gassner N, Xu J, Heinz DW, Matthews BW. 1993a. Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala → Ser and Val → Thr substitutions in T4 lysozyme. *Biochemistry 32*:11363–11373.

Blaber M, Zhang X-j, Lindstrom JD, Pepiot SD, Baase WA, Matthews BW. 1994. Determination of alpha-helix propensity within the context of a folded protein. *J Mol Biol 235*:600–624.

Blaber M, Zhang X-j, Matthews BW. 1993b. Structural basis of amino acid alpha helix propensity. *Science 260*:1637–1640.

Endo Y, Mitsui K, Motizuki M, Tsurugi K. 1987. The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. *J Biol Chem 262*:5908–5912.

Engh RA, Huber R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr A 47*:392–400.

Eriksson AE, Baase WA, Matthews BW. 1993. Similar hydrophobic replace-

ments of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *J Mol Biol 229*:747–769.
Faber HR, Matthews BW. 1990. A mutant T4 lysozyme displays 5 different crystal conformations. *Nature 348*:263–266.
Frankel A, Schlossman D, Welsh P, Hertler A, Withers D, Johnston S. 1989. Selection and characterization of ricin toxin A-chain mutations in *Saccharomyces cerevisiae*. *Mol Cell Biol 9*:415–420.
Frankel A, Welsh P, Richardson JS, Robertus JD. 1990. Role of arginine 180 and glutamic acid 177 of ricin toxin A chain in enzymatic inactivation of ribosomes. *Mol Cell Biol 10*:6257–6263.
Gassner NC, Baase WA, Matthews BW. 1996. A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc Natl Acad Sci USA 93*:12155–12158.
Ghaemmaghami S, Word JM, Burton RE, Richardson JS, Oas TG. 1998. Folding kinetics of a fluorescent variant of monomeric λ repressor. *Biochemistry 37*:9179–9185.
Hovde CJ, Calterwood SB, Mekalanos JJ, Collier RJ. 1988. Evidence that glutamic acid 167 is an active-site residue of Shiga-like toxin I. *Proc Natl Acad Sci USA 85*:2568–2572.
Hurley JH, Baase WA, Matthews BW. 1992. Design and structural analysis of alternative hydrophobic core packing arrangements in bacteriophage T4 lysozyme. *J Mol Biol 224*:1143–1159.
Karpusas M, Baase WA, Matsumura M, Matthews BW. 1989. Hydrophobic packing in T4 lysozyme probed by cavity-filling mutants. *Proc Natl Acad Sci USA 86*:8237–8241.
Kim Y, Mlsna D, Monzingo AF, Ready MP, Frankel A, Robertus JD. 1992. Structure of a ricin mutant showing rescue of activity by a noncatalytic residue. *Biochemistry 31*:3294–3296.
Langridge R, Ferrin TE, Kuntz ID, Connolly ML. 1981. Real-time color graphics in studies of molecular interactions. *Science 211*:661–666.
Lee C, Subbiah S. 1991. Prediction of protein side-chain conformation by packing optimization. *J Mol Biol 217*:373–388.
Lipscomb LA, Gassner NC, Snow SD, Eldridge AM, Baase WA, Drew DL, Matthews BW. 1998. Context-dependent protein stabilization by methionine-to-leucine substitution shown in T4 lysozyme. *Protein Sci 7*:765–773.
Lovell SC, Word JM, Richardson JS, Richardson DC. 2000. The penultimate rotamer library. *Proteins Struct Funct Genet 40*:389–408.
MacKenzie KR, Prestegard JH, Engelman DM. 1997. A transmembrane helix dimer: Structure and implications. *Science 276*:131–133.
Matsumura M, Matthews BW. 1989. Control of enzyme activity by an engineered disulfide bond. *Science 243*:792–794.
Matthews BW. 1995. Studies on protein stability with T4 lysozyme. *Adv Protein Chem 46*:249–278.
Mlsna D, Monzingo AF, Katzin BJ, Ernst S, Robertus JD. 1993. Structure of recombinant ricin A chain at 2.3 Å. *Protein Sci 2*:429–435.
Momany FA, McGuire RF, Burgess AW, Scheraga HA. 1975. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic changes, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J Phys Chem 79*:2361.
Nicholls A, Bharadwaj R, Honig B. 1993. GRASP–Graphical representation and analysis of surface-properties. *Biophysical Society Journal, 37th Annual Meeting Abstracts 64*:A166.
Petrella RJ, Lazaridis T, Karplus M. 1998. Protein sidechain conformer prediction: A test of the energy function. *Fold Des 3*:353–377.
Porter TK. 1978. Spherical shading. *Comput Graphics 12*:282–285.
Richardson DC, Richardson JS. 1992. The kinemage: A tool for scientific illustration. *Protein Sci 1*:3–9.
Richardson DC, Richardson JS. 1994. Kinemages—Simple macromolecular graphics for interactive teaching and publication. *Trends Biochem Sci 19*:135–138.
Richardson JS, Richardson DC. 2000. MAGE, PROBE, and Kinemages. In: Rossmann MG, Arnold EA, eds. *International tables for crystallography*, vol. F, chap. 25.2.8. Dortrecht: Kluwer Academic Publishers.
Schlossman D, Withers D, Welsh P, Alexander A, Robertus J, Frankel A. 1989. Role of glutamic acid 177 of the ricin toxin A chain in enzymatic inactivation of ribosomes. *Mol Cell Biol 9*:5012–5021.
Streitwieser A, Heathcock CC. 1976. *Introduction to organic chemistry*. New York: Macmillan Publishing Co., Inc.
Vetter IR, Baase WA, Heinz DW, Xiong J-P, Snow S, Matthews BW. 1996. Protein structural plasticity exemplified by insertion and deletion mutants in T4 lysozyme. *Protein Sci 5*:2399–2415.
Weston SA, Tucker AD, Thatcher DR, Derbyshire DJ, Pauptit RA. 1994. X-ray structure of recombinant ricin A-chain at 1.8Å resolution. *J Mol Biol 244*:410–422.
Wickham GS, Word JM. 1999. Evaluation of the resultant van der Waals interactions after modeling the cleavage site phosphate into the crystal structure of the HDV genomic ribozyme. *Nucleic Acids Symp Series 41*:182–183.
Word JM. 2000. All-atom small-probe contact surface analysis: An information-rich description of molecular goodness-of-fit [PhD dissertation]. Durham, NC: Duke University.
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*:1709–1731.
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*:1735–1747.
Zeng J, Fridman M, Maruta H, Treutlein HR, Simonson T. 1999. Protein–protein recognition: An experimental and computational study of the R89K mutation in Raf and its effect on Ras binding. *Protein Sci 8*:50–64.