PROTEINS: Structure, Function, and Genetics 50:272–282 (2003)

Protein Side-Chain Rearrangement in Regions of Point Mutations

Eran Eyal,* Rafael Najmanovich, Marvin Edelman, and Vladimir Sobolev
Department of Plant Sciences, Weizmann Institute of Science, Rehovot, Israel

ABSTRACT **A major problem in predicting amino acid side-chain rearrangements following point mutations is the potentially large search space. We analyzed a nonredundant data set of 393 Protein Data Bank protein pairs, each consisting of structures differing in one amino acid, to determine the number of residues changing conformation in the region of mutation. In 91–95% of cases, two or fewer residues underwent side-chain conformational change. If mutation sites with backbone displacements were excluded, the number increased to 97%. The majority of rearrangements (over 60%) were due to the inherent flexibility of side-chains, as derived from analysis of a control set of protein subunits whose crystal structures were determined more than once. Different amino acids demonstrated different degrees of flexibility near mutation sites. Large polar or charged residues, and serine, are more flexible, while the aromatic amino acids, and cysteine, are less so. This pattern is common to the inherent side-chain flexibility, as well as the increased flexibility at ligand binding sites and mutation sites. The probability for conformational change was correlated with B-factor, frequency of the side-chain conformation in proteins and solvent accessibility. The last trend was stronger for aromatic and hydrophilic residues than for hydrophobic ones. We conclude that the search space for predicting side-chain conformations in the region of mutation can be effectively restricted. However, the overall ability to predict a particular side-chain conformation, or to check predictions according to individual existing structures, is limited. These findings may be useful in deriving empirical rules for modeling side-chain conformations. Proteins 2003;50:272–282.** © 2002 Wiley-Liss, Inc.

Key words: side-chain flexibility; side-chain prediction; Protein Data Bank; B-factor

INTRODUCTION

Changes in protein structure, stability, and chemical properties resulting from point mutations are of major importance in the evolution of proteins. They also play a major role in assisting the biologist to probe the function of a protein. The influence of a single mutation on protein function can vary from no affect, to partial loss of function due to local changes in inter- and intramolecular interactions, to complete loss of function due to protein misfolding upon mutation of core residues (for reviews, see Refs. 1 and 2 and references therein). The increasing size of single nucleotide polymorphism data sets makes the task of predicting relations between point mutations and changes in structure and function timely and relevant.

A first step in understanding the consequences of point mutations is identification of possible changes in protein structure. Analysis of individual crystals has shown that, in most cases, the effects are limited to the side-chain conformations of the amino acids having immediate contact with the mutation.[1–3] Predictive rules about the conformation of the residue in the mutated position exist[4]; however, comprehensive analysis of structural changes in the residues surrounding the mutation has yet to be done.

Modeling side-chain conformations is an important step in current methods for protein structure prediction. A serious limitation is the enormous theoretical number of combinations available for a set of residues within the search space of dihedral angles. From analysis of known protein structures, it was shown that for every amino acid side-chain there is a set of preferable conformations, or rotamers.[5–7] The population of each rotameric state depends mainly on torsion energy (similar to that in small organic molecules) and on the backbone conformation.[8] Interactions within secondary structure elements, such as hydrogen bonds in formation of β-turns,[9] repulsion preventing side-chain steric clashes in β-sheets,[10] or hydrogen bonding of Ser/Thr and Asx side-chains in an α-helix terminal,[5,11,12] also contribute to stabilizing some side-chain rotamers. Based on such analyses, several rotamer libraries were created.[13–16]

Testing the full set of protein side-chain conformations using a rotamer library is, however, an intractable computational task. To further restrict the search space, various methods have been applied, such as simulated annealing[17–21] and application of empirical procedures both with rotamer libraries[22–28] and without.[29–34] The dead-end elimination theorem is also commonly used,[35–38] as well as self-consistent rearrangement of several residues at a time.[15,39,40] Comparison between some of these methods

Grant sponsor: Israeli Ministry of Science; Grant sponsor: Ministry of Absorption, the Center for Absorption of Scientists.

*Correspondence to: Eran Eyal, Department of Plant Sciences, Weizmann Institute of Science, 76100 Rehovot, Israel. E-mail: eran.eyal@weizmann.ac.il

Received 12 March 2002; Accepted 29 July 2002

was previously preformed.[41] Nonetheless, it is still possible that all these approaches can miss the global minimum during the search. Modification of the dead-end elimination approach[42,43] resolves this problem, but at the cost of a large increase in the size of the search space. Although considerable advances were achieved in this field lately,[16,44] any information that allows additional restrictions of the search space is valuable.

Conformational change around a point mutation can be estimated by analyzing the side-chain flexibility of amino acid residues in contact with the mutation site and comparing it to that of the nonmutated protein. To this end, we previously built a structural database of protein pairs differing in sequence by one or two amino acids (http://bioinfo.weizmann.ac.il/MutaProt[45]). In the current work we use a nonredundant subset of this data to statistically analyze structural rearrangements around point mutations.

MATERIALS AND METHODS

Definition of Mutation Region

CSU software[46] was used to determine the residues in contact with a particular residue site. The approach takes into account the location of proximal atoms and interatomic distances. A "mutation region" in an entry of our data set of file pairs is composed of the set of residues in contact with the point mutation site in at least one file of the pair.

Gly and Ala are, necessarily, excluded from analyses of side-chain dihedral angles. Changes in Pro side-chain conformation cause backbone rearrangement; therefore, Pro was not considered either. In addition, a residue was usually omitted from analysis if it had at least one heavy atom with B-factor > 50 Å^2 (see Discussion below), had atoms having more than one possible location assigned by the Protein Data Bank (PDB), or lacked the coordinate assignment for one or more atoms.

Construction of Data sets

X-ray structures with resolution 2.0 Å or better were selected from the PDB[47,48] (November 2000 release) and divided into groups such that all entries in an individual group contained the same number of amino acids and the same set of ligands. A redundant list of PDB file pairs, differing by a single amino acid position, was created by screening inside each of these groups for pairs satisfying this criterion. Similarly, a redundant data set of PDB file pairs differing by two positions was created.

A nonredundant data set of point mutation pairs (M-Set) was achieved in a series of steps. We define a cell as a collection of entries from the same protein differing in a single amino acid at the same specific position. For each cell, the two PDB entries with the best resolution were taken. In the double-mutation list, we considered each amino acid change independently to handle redundancy. Every pair from this list was listed twice, once for each amino acid change. A nonredundant list for the double mutation was then derived as above. A unified nonredundant list was obtained by addition of pairs from the double-mutation list to the single-mutation list if the mutation position was not already present. Thus, any mutated position in the M-Set for any specific protein is present only once. The procedure is schematically shown in Figure 1(A).

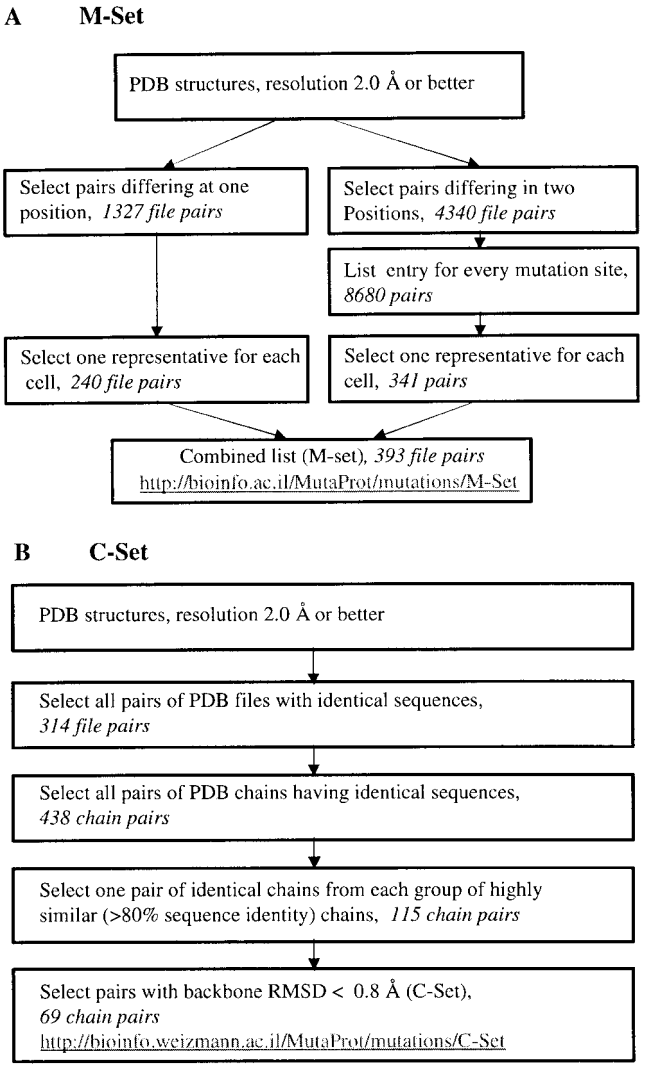

Fig. 1. Schematic procedure for creating the data sets: (**A**) Nonredundant data set for analysis of structural changes in mutation regions (M-Set). (**B**) Nonredundant data set for analysis of structural differences between PDB files with the same amino acid sequence and set of ligands (C-Set).

The procedure used to create a nonredundant data set of PDB pairs with identical sequences (C-Set) is schematically represented in Figure 1(B). An initial list was obtained by including all the combinations of PDB pairs with exactly the same amino acid sequence. These pairs were assembled into groups, each pair in a group sharing $> 80\%$ identity with all other pairs in the same group. This relatively high value (80% identity) was chosen to retain data for a single protein family within a group. A lower threshold (e.g., 30%) is not necessary because we are dealing with side-chain positions, while a higher threshold (e.g., 95%) could miss similarity between proteins with a

few mutations. From each group, the pair with the best resolution was taken. The two structures of the pair were then superimposed by minimizing the C_α root mean square deviation (RMSD). Only pairs with RMSD < 0.8 Å were included in the final list.

A data set of ligand binding sites (L-Set), in which each different ligand appears once, was constructed as described before.[49] However, more stringent criteria were introduced for resolution ($\leq$2.0 Å vs. $\leq$2.5 Å) and selection of files (those with the highest resolution were chosen).

Calculation of Dihedral Angle Changes

Structural differences between analogous protein side-chains can be quantified by estimating the change in dihedral angles or by measuring atomic displacement. Both methods are used in this study.

Conformations with deviations of up to $\pm$ 40° from the reference rotamer values are in general considered to belong to the given rotamer.[50] Therefore, absolute differences of $\approx$ 80° can be found for a particular angle with the side-chains still belonging to the same rotamer.[49] We define a conformational change as having occurred if an absolute difference between two angles larger than 60° exists for at least one dihedral angle within the residue. The chosen cutoff value of 60° provides a high probability for correct identification and a low probability for misidentification of actual rotamer alterations.[49]

Several amino acids are symmetrical in their terminal dihedral angle. As a result, there are two possible values for differences between the dihedral angles. In these cases, the smaller value represents the actual difference. The following residues have such symmetry: Glu (at χ_3), Asp (at χ_2), Tyr (at χ_2), and Phe (at χ_2). Some other residues do not exhibit real symmetry but the position of heavy atoms might be wrongly assigned, causing incorrect calculation of the associate dihedral angles. This group includes: Gln (at χ_2), Asn (at χ_2), and His (at χ_2). For these residues, we chose not to consider this ambiguity and to rely on the PDB notation. Thus, the flexibility of these residues might be slightly overestimated.

To allow comparison of data concerning χ_1 and χ_2 with other works (e.g., Refs. 29, 51, and 52), we calculated the probability for changes of χ_1 or χ_2. We also calculated the probability for changes of any side-chain dihedral angle because modeling side-chains for practical use should strive to consider all the angles. The probability for dihedral angle change is:

$$P_i = \frac{N_c^i}{N_t^i} \pm \sqrt{\frac{F_c \times (1 - F_c)}{N_t^i}} \quad (1)$$

$$F_c = \frac{\sum_{i=aa} N_c^i}{\sum_{i=aa} N_t^i}, \quad (2)$$

where N_c^i is number of amino acids of type i that undergo conformational change and N_t^i is the total number of amino acids of type i.

TABLE I. Amino Acid Properties

Amino acid	Number of rotatable dihedral angles	Maximum solvent accessibility of side-chain (Å^2)	Terminal atom of side-chain
Arg	4[a]	190	NH1, NH2
Asn	2	113	OD1, ND2
Asp	2	101	OD1, OD2
Cys	1	81	SG
Gln	3	132	OE1, NE2
Glu	3	136	OE1, OE2
His	2	147	CE1, NE2
Ile	2	129	CD1
Leu	2	141	CD1, CD2
Lys	4	153	NZ
Met	3	150	CE
Phe	2	179	CZ
Ser	1	70	OG
Thr	1	88	OG1, CG2
Trp	2	228	CH2
Tyr	2	195	OH
Val	1	103	CG1, CG2

[a]Rotation of the NE-CZ bond in Arg was not considered because the CD, NE, CZ, NH1, and NH2 atoms form a structure close to planar that does not change in shape (atom names follow PDB nomenclature).

The probability of conformational change was normalized to compare residues with different numbers of dihedral angles (Table I). For an amino acid of type i with n_d independent flexible dihedral angles, the probability of change of a single dihedral angle (P_d^i) is[49]

$$P_d^i = 1 - \sqrt[n_d]{1 - P_i}, \quad (3)$$

where P_i is probability of side-chain conformational change as described above.

Backbone Changes

We measure backbone changes in the vicinity of residue X by calculating the maximum displacement between pair-wise distances of any two C_α atoms of residue-X-neighbors in the two files. A neighbour is defined as any residue within 8.0 Å from the C_α atom of residue X. For side-chain flexibility analysis, we used only residues with backbone displacement smaller than a threshold value of 0.5 Å.

Atom Displacement

The spatial difference for a residue between two structures was estimated by displacement of its terminal atom (Table I). Displacement was measured by superimposing backbone atoms N, C_α, and C, and noting the distance between the positions for the side-chain terminal atom of each. For residues with more than one terminal atom, the maximum displacement was chosen.

Solvent Accessibility

Solvent accessibility[53] of side-chains was calculated using CSU software.[46] Relative accessibility of side-chain

atoms for any residue was defined as the ratio between absolute solvent accessibility in the experimental structure and maximum solvent accessibility of the side-chain. This maximum for residue X was calculated as the accessibility of X in an extended form of the peptide GGXGG (Table I). Residues with a relative accessibility of < 0.15 were considered buried, while those with relative accessibility ≥ 0.4 were considered exposed. These definitions are comparable to those accepted in the literature (e.g., ref. 51). Residues with intermediate values were not included in any of the accessibility classes.

B-Factor Analysis

The average observed B-factor values vary widely between different structures. Therefore, for statistical analysis, B-factor values were normalized[54]:

$$B' = \frac{\sum B - < B >}{\sigma(B)}, \quad (4)$$

where, for a given file, $\langle B \rangle$ and $\sigma(B)$ are the average and standard deviation of the B-factor values, respectively.

Rotamer Library

We modified the backbone-independent rotamer library of Dunbrack and Cohen[55] to describe χ_1 and χ_2 only. For residues with more than two flexible bonds, χ_1 and χ_2 values were derived as weighted averages of all the rotamers that share this same angle space region. The weights are the probabilities of the rotamers in the original library. A residue of interest was assigned a rotamer from this reduced library by taking a rotamer having both χ_1 and χ_2 values within 60°.

RESULTS

Data Sets

A nonredundant data set of high-resolution PDB structures (≤2.0 Å) was created to analyze structural changes upon point mutations. In this data set (called the M-Set), a mutated position in a given protein appears only once. The M-Set is composed of 393 pairs of protein structures. A nonredundant control set (called the C-Set) of 69 pairs of high-resolution PDB structures was constructed as well. The two members of a pair in the C-Set have identical sequence, identical ligands, and a small backbone displacement (RMSD < 0.8 Å) following superimposition. The M- and C-Sets are available on the Web (http://bioinfo.weizmann.ac.il/MutaProt/mutations/).

Influence of Specific Types of Mutations

The residue type at the mutated position was not considered in creating the M-Set. Ala showed the maximal number of entries for any amino acid type, being found at the mutation position 145 times. This constitutes 37% of all the entries in the M-Set. Ala often appears against Val (23 cases), Gly (19 cases), or Ser (19 cases). Thus, a sizable fraction of the changes in the M-Set are from one small amino acid to another. Other abundant combinations are Asp–Asn (21 cases), Tyr–Phe (15 cases), and Glu–Gln (12 cases). A desirable goal would be to analyze the structural changes caused by different types of substitutions separately. However, even for the relatively abundant combinations, the data is not currently sufficient for reliable statistical analysis.

We analyzed changes in flexibility of the substituted amino acids at the level of various chemical properties. A redundant database of pairs of PDB files was scanned for the desired specific type of substitution. A nonredundant data set for each specific property was then created in a manner similar to that described for the M-Set, except that the representative pair from each "cell" of the "unified nonredundant list" (see Materials and Methods) is the one with the desired specific type of substitution.

We found a marginally larger flexibility upon hydrophobic/hydrophilic substitutions compared to hydrophobic/hydrophobic ones ($P < 0.05$). In addition, there was a noticeable, but statistically nonsignificant, increase in flexibility upon changes in residue size (residues with a maximum of two side-chain heavy atoms replaced by residues with at least five heavy atoms), compared to replacement of side-chains with the same number of heavy atoms. Surprisingly, no differences were observed upon substitution with residues of opposite charge compared to replacements with the same charge. These analyses and their data sets can be found at http://bioinfo.weizmann.ac.il/MutaProt/mutations/.

Region Affected by the Mutation

We examined the structural changes in different shells around the mutations. We found significant difference between the influence of the mutation on the buried residues of the first shell (probability for conformational change of $4.2 \pm 0.4\%$) compared to buried residues of the second shell ($2.7 \pm 0.3\%$) and beyond ($2.6 \pm 0.2\%$). For exposed residues the trend was less clear. The flexibility level throughout the protein was about the same (19.2 ± 1.2, 18.0 ± 1.0, and $16.6 \pm 0.5\%$ for first shell, second shell, and beyond, respectively). Therefore, the contribution of a point mutation to the flexibility of exposed residues is relatively small compared to their natural flexibility. Because the contribution of the mutation to side-chain flexibility is mainly at the first shell, we restricted our analysis to this area. It should be noted, however, that changes beyond the first shell, although less frequent, do happen.

B-Factor and Flexibility

The B-factor is a prevalent measure of mobility or motion of individual atoms or a group of bonded atoms in X-ray resolved structures. A large B-factor indicates that the probability for an atom to be located close to its assigned position is small. It may also indicate errors in the refinement or disorder in exact location of molecules in different unit cells. In light of this, Betts and Sternberg[52] ignored residues having atoms with high B-factor (> 50 Å^2) in their statistical analysis of protein structures. However, B-factor also indicates mobility of individual atoms and side-chains.[56,57] Therefore, ignoring residues having atoms with high B-factors might bias the results toward lower flexibility.

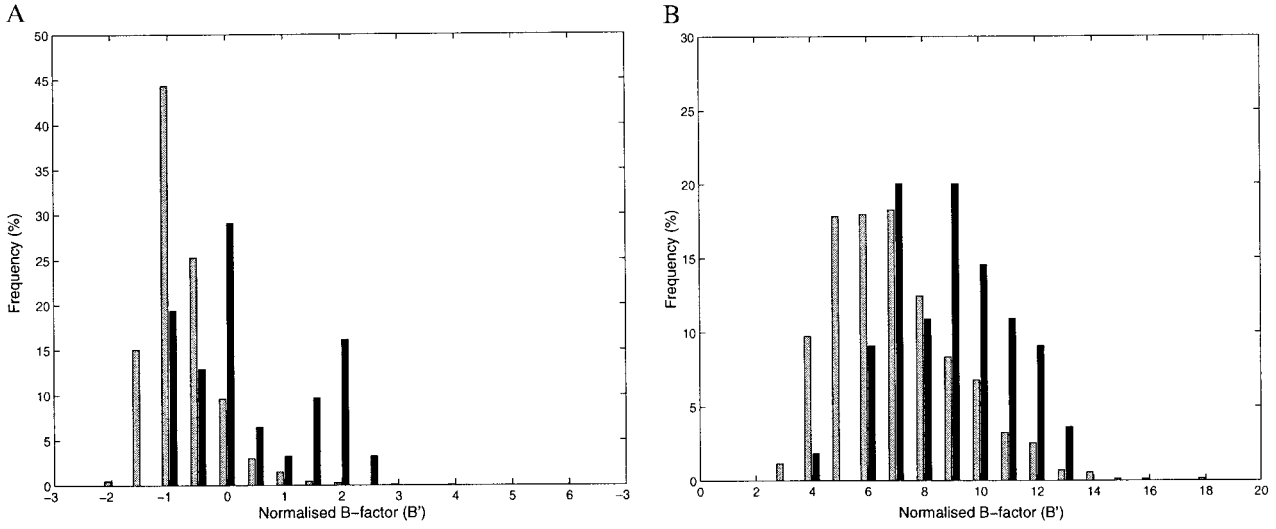

Fig. 2. Normalized B-factor (B') of Cγ/Oγ atoms of equivalent residues in pair of PDB files from the M-Set. Black bars, B' distribution for residues undergoing conformational change. Grey bars, B' distribution for residues that do not change conformation. (**A**) Buried residues. (B) Exposed residues.

We calculated the probability of conformational change in the mutation region considering all residues or only those with atoms having a B-factor < 50 Å^2. The probability calculated to 10.7 and 8.5%, respectively, exposed residues showing probabilities of 23.6 and 19.2% and buried residues 4.5 and 4.2%. We also analyzed the relationship between B-factor and probability for conformational change. The normalized B-factor (B') distributions for the Cγ/Oγ atoms of residues that change conformation and those that do not change are shown separately in Figure 2 for buried and exposed residues. In each case the distributions are statistically different (t-test; $P \ll 0.001$). Thus, B-factor values of Cγ atoms are higher in side-chains that adopt alternative χ_1 conformations. This is in agreement with analyses of specific pairs of homologous proteins,[29,58] and supports the use of probability of conformational change as a valid measure for flexibility. In our study, we analyzed flexibility levels with and without B-factor restriction. Relative residue flexibilities proved to be practically the same. In these cases, we present only the data with B-factor restriction.

Structural Changes at Mutation Regions

In the M-Set, the average number of residues in the region of a mutation is 8.3 ± 2.9 (ignoring Ala, Gly, and Pro, which are excluded from the study). We determined the number of residues undergoing conformational change for each entry. With B-factor restrictions, 60% of cases had no side-chain conformational changes in the mutation region, while two or fewer changes accounted for 95% of cases [Fig. 3(A)]. Without B-factor restrictions, the numbers were 49 and 91%, respectively [Fig. 3(B)].

The above analysis includes both side-chain and backbone influences. To distinguish between them we repeated this experiment excluding backbone displacement. In 17% of cases, backbone displacement of more than the threshold value of 0.5 Å occur; their distribution is shown in Figure 3. Cases with three or more side-chain conformational changes have a higher percentage of backbone shifts than those with fewer side-chain changes. When we excluded backbone shifts from the analyses, two or fewer side-chain changes accounted for 97% of cases irrespective of B-factor restriction. Unless noted otherwise, analysis of side-chain rearrangements is restricted to those residues with only minor backbone displacement (< 0.5 Å).

The probability of different amino acid types to undergo side-chain conformational change in the vicinity of a mutation was examined. The data collected from 2172 residues in mutation regions having backbone displacements of less than 0.5 Å are summarized in Table II. The probability for conformational change for each amino acid type was derived. In Figure 4(A), values are presented for χ_1 (black bars), χ_2 (dark grey), and χ_3 and χ_4 (light grey). A change at a specific χ angle is scored only when all lower-denomination angles were unchanged. Probability for conformational change is the sum of the individual χ angle probabilities (the composite height of the left bars). As can be seen in Figure 4(A), for any χ angle change the large polar residues (Glu, Gln, Arg, Lys) are the most flexible, while the aromatic ones (Tyr, Trp, Phe, His) and Cys are the least. Hydrophobic and medium-sized polar amino acids have intermediate levels of side-chain flexibility.

It can be argued that summation of the changes in side-chain dihedral angles does not necessarily reflect comparable spatial rearrangement, especially for the large residues. For these, changes in consecutive angles might compensate each other to yield a reduced, net atomic displacement. We examined the probability for terminal atom displacement of more than 1.7 Å for each side-chain type [Fig. 4(A), white bars]. That threshold was chosen to score the great majority of rotamer switches for residues with one flexible bond. A comparison of the tops of bars for each amino acid type shows that the two different ap-

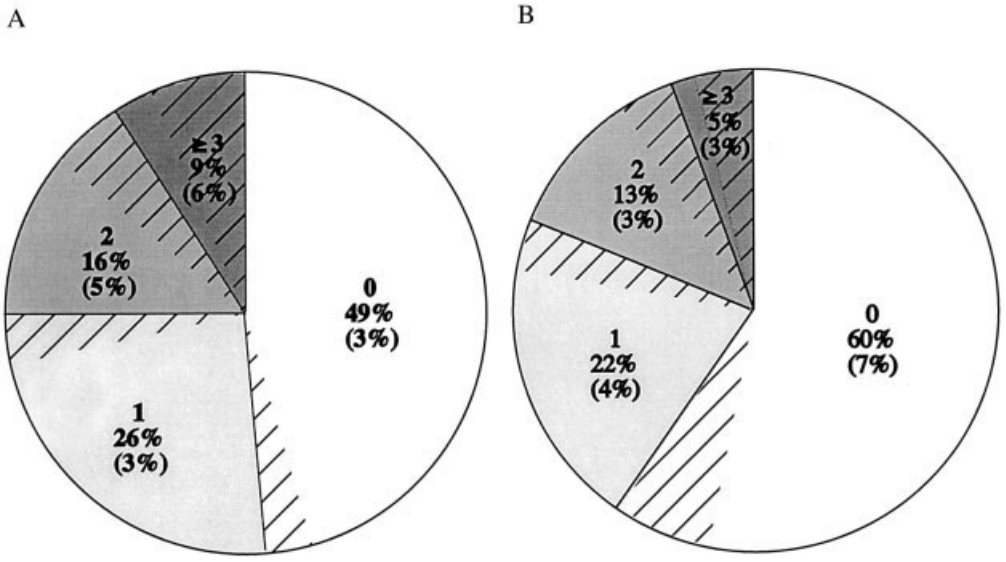

Fig. 3. Distribution of mutation regions as a function of the number of side-chains undergoing conformational change. The lined areas indicate cases with above-threshold backbone displacement and the numbers in parentheses are their percentages of the total. (A) Without B-factor restriction (all residues). (B) With B-factor restriction (< 50 Å^2).

TABLE II. Side-Chain Conformational Change of Amino Acids in Contact with the Mutation Site

	All					Buried					Exposed				
Amino acid	Total no.	χ_1 change	χ_2[a] change	χ_3/χ_4[b] change	Any[c] change	Total no.	χ_1 change	χ_2[a] change	χ_3/χ_4[b] change	Any[c] change	Total no.	χ_1 change	χ_2[a] change	χ_3/χ_4[b] change	Any[c] change
Arg	155	5	9	19	33	43	0	2	4	6	76	5	5	13	23
Asn	145	2	14	—	16	44	1	3	—	4	45	1	6	—	7
Asp	114	1	3	—	4	35	0	0	—	0	45	1	2	—	3
Cys	93	0	—	—	0	82	0	—	—	0	0	0	—	—	0
Gln	89	3	2	7	12	30	0	0	0	0	38	1	1	4	6
Glu	92	4	6	2	12	29	2	2	0	4	29	2	3	0	5
His	79	1	0	—	1	48	1	0	—	1	11	0	0	—	0
Ile	165	2	10	—	12	142	1	7	—	8	11	0	2	—	2
Leu	256	2	10	—	12	195	2	3	—	5	31	0	3	—	3
Lys	91	4	6	14	24	10	1	0	1	2	55	2	5	12	19
Met	70	1	1	2	4	53	0	0	1	1	1	1	0	0	1
Phe	139	1	0	—	1	115	0	0	—	0	5	0	0	—	0
Ser	121	7	—	—	7	78	5	—	—	5	28	0	—	—	0
Thr	154	3	—	—	3	77	0	—	—	0	46	2	—	—	2
Trp	84	0	1	—	1	65	0	1	—	1	4	0	0	—	0
Tyr	135	1	0	—	1	92	0	0	—	0	10	1	0	—	1
Val	190	12	—	—	12	131	7	—	—	7	31	1	—	—	1
Total	2172	49	62	44	155	1269	20	18	6	44	466	16	27	30	73

[a]Number of cases in which χ_2 undergoes change and χ_1 does not.
[b]Number of cases in which χ_3 or χ_4 (or both) undergo change when χ_1 and χ_2 do not.
[c]Sum of all cases in which at least one angle change has occurred. Division of the numbers from this column by the number of residues gives the measured probability of conformation change.

proaches—changes of dihedral angles (left bars) and spatial displacement (right bar)—yield remarkably similar patterns of side-chain flexibility. The only exception is Lys (and to a much lesser extent Arg), in which the probabilities for terminal atom displacement were reduced compared to the probability for conformational change. These results show that measurement of terminal atom displacement is a faithful representation of probability for rotamer switch for all residue types with one to three side-chain dihedral angles.

The number of flexible bonds in a given residue is an obvious major factor in determining flexibility. Residues with more bonds are more flexible because they simply contain more degrees of freedom for side-chain movements. Still, after normalizing this factor for side-chain changes in mutation regions [Fig. 4(B)], significant differ-

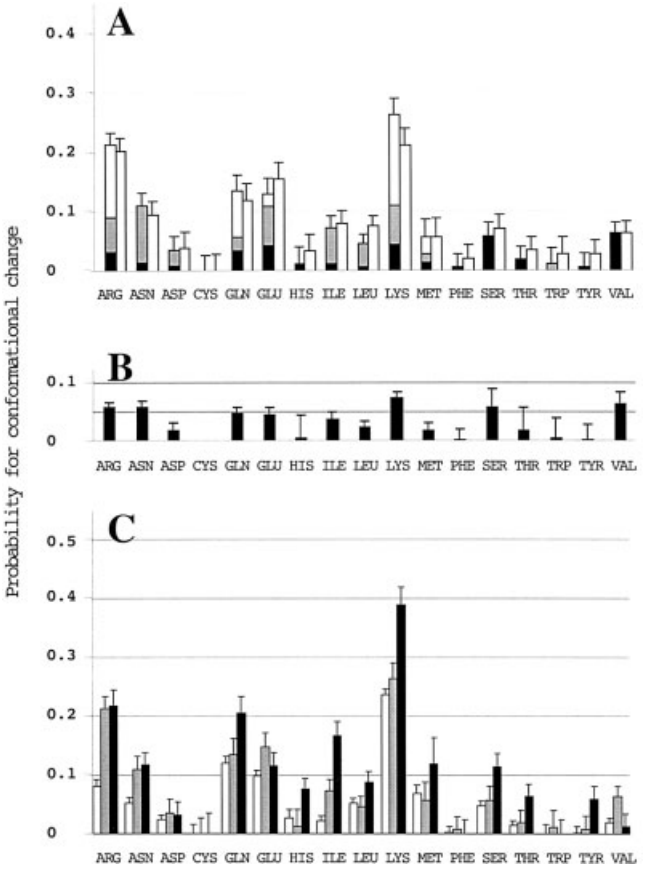

Fig. 4. Amino acid side-chain flexibility at mutation regions. The error bars represent the standard deviation of the binomial distribution and reflect the sample size for each amino acid. (**A**) Changes in dihedral angles versus spatial displacement. The left bar for each amino acid shows the probability for conformational change. Starting from the backbone, only the first angle undergoing change is scored. Black, χ_1; dark grey, χ_2; light grey, χ_3/χ_4. The right bars (white) for each amino acid show the probability of the terminal atom of that residue to be displaced by more than 1.7 Å. (**B**) Probability of a single dihedral angle to change. Values obtained from the left bar tops of (A) were normalized by the number of dihedral angles for each amino acid using eq. 3. (**C**) Comparison of flexibilities: Inherent, mutation-derived, and ligand binding. The probabilities for conformational change of side-chains are derived from different data sets of structures at a resolution 2.0 Å or better. White bars (C-Set), derived from a data set of pairs of structures having identical sequence. Grey (M-Set), derived from a data set of residues surrounding mutation sites (A, left bars). Black (L-Set), derived from a new version of the data set of ligand binding sites described by Najmanovitch et al.[49]

ences are observed among the various amino acids. A similar phenomenon was recorded for side-chain changes induced by ligand binding.[49] The normalized pattern is similar to that found for χ_1 angles of paired, uncomplexed protein structures.[59] The χ_1 angle changes for mutation regions are shown as black bars in Figure 4(A). Considering χ_1 and the normalized data of Figure 4(B), Ser, Val, and the large hydrophilic amino acids register as the most flexible residue types.

Comparisons with Other Data Sets

Table III summarizes statistical data on side-chain conformational changes in 9064 residues from a nonredundant control set (C-Set) of structures crystallized more than once. The overall probability of a residue to have different conformations in two crystals of the same protein is 4.7% in helical regions, 4.9% for sheets, and 5.9% in other regions. Intuitively, one expects to find a smaller degree of flexibility in the relatively well-packed sheet and helix structures. Indeed, flexibility is somewhat lower (by 20–25%) in these regions versus others ($P < 0.01$). Thus, the rigidity of secondary structure resulting from backbone hydrogen bonds is reflected also in side-chain flexibility, but to a relatively small extent. Overall, 2.0% of the buried residues and 10.1% of the exposed ones in the C-Set change conformation (Table III) as compared to 3.5 and 15.6%, respectively, in mutation regions in the M-Set (Table II). The differences between sets are statistically highly significant ($P \ll 0.001$) and represent changes of about 75% for the buried residues and about 55% for the exposed ones. Thus, the inherent flexibility of amino acid side-chains (independent of backbone influences) contributes substantially to the apparent flexibility in the vicinity of a point mutation.

The probability for conformational change of individual amino acid types in the region of a mutation (M-Set) was compared to that for analogous changes in the C-Set of identical protein pairs and to that for changes in a high-resolution set (L-Set) of ligand binding sites (pairs of holo and apo-proteins). For the L-Set, we observed few differences from the original work of Najmanovich et al.[49] Val, Thr, and Met are found here to be less flexible, while Ser is more flexible at the higher resolution. The overall patterns for the three data sets are similar [Fig. 4(C)]. The correlation of flexibility patterns for amino acid types among these three sets supports the idea that side-chain flexibility is an intrinsic property of the amino acids.

However, there are several notable differences. Arg demonstrates a high propensity for conformational change, both in mutation regions and in binding sites, compared to the control set. For Gln (surprisingly different from Asn, Asp, and Glu), there is a large increase in flexibility upon ligand binding. Ser, as well as the aromatic polar residues Tyr and His, also show increased propensities for change at ligand binding sites. Likewise, β-branched amino acids, Thr and Ile, show enhanced probabilities for conformational change upon ligand binding, while Val and Ile show it for mutation sites. Interestingly, Val is the only amino acid that demonstrates higher flexibility at mutation regions compared to ligand binding sites. The long hydrophobic residues show increased propensity for change upon ligand binding compared to mutations. This can be explained by a necessity for their hydrophobic surfaces to be hidden from the solvent in the apo condition.

The solvent accessibility of side-chains affects flexibility in the M- and C-Sets in a similar way (Tables II and III). The statistics for individual residues is not sufficient; therefore, analysis was done for groups of residues. In both data sets, exposed residues on average are about fivefold more flexible than buried ones (Fig. 5). These results are in accord with several other studies.[51,52,59] As can be seen in Figure 5, the difference in flexibility between exposed and

TABLE III. Side-Chain Conformational Change of Amino Acids in Different Structures of the Same Protein (C-Set)

	All					Buried					Exposed				
Amino acid	Total no.	χ_1 change	χ_2[a] change	χ_3/χ_4[b] change	Any[c] change	Total no.	χ_1 change	χ_2[a] change	χ_3/χ_4[b] change	Any[c] change	Total no.	χ_1 change	χ_2[a] change	χ_3/χ_4[b] change	Any[c] change
Arg	416	9	7	18	34	115	1	1	3	5	158	8	2	11	21
Asn	520	2	25	—	27	139	0	0	—	3	231	2	17	—	19
Asp	692	8	8	—	16	137	0	0	—	0	394	8	7	—	15
Cys	208	0	—	—	0	171	0	—	—	0	13	0	—	—	0
Gln	398	10	13	25	48	100	1	0	4	5	197	7	12	16	35
Glu	566	29	17	10	56	103	4	0	0	4	324	19	14	8	41
His	289	1	7	—	8	143	0	1	—	1	69	0	3	—	3
Ile	648	3	11	—	14	485	2	4	—	6	62	1	3	—	4
Leu	966	2	49	—	51	717	1	31	—	32	85	0	7	—	7
Lys	532	15	28	83	126	41	0	1	3	4	362	13	22	64	99
Met	219	5	3	7	15	153	2	3	4	9	26	1	0	1	2
Phe	551	0	2	—	2	424	0	1	—	1	28	0	0	—	0
Ser	787	38	—	—	38	324	5	—	—	5	358	29	—	—	29
Thr	723	10	—	—	10	240	0	—	—	0	343	9	—	—	9
Trp	190	0	0	—	0	149	0	0	—	0	8	0	0	—	0
Tyr	463	0	1	—	1	294	0	0	—	0	44	0	0	—	0
Val	896	17	—	—	17	637	14	—	—	14	129	3	—	—	3
Total	9064	149	171	143	463	4372	30	45	14	89	2831	100	87	100	287

[a]Number of cases in which χ_2 undergoes change and χ_1 does not.
[b]Number of cases in which χ_3 or χ_4 (or both) undergo change when χ_1 and χ_2 do not.
[c]Sum of all cases in which at least one angle change has occurred. Division of the numbers from this column by the number of residues gives the measured probability of conformation change.

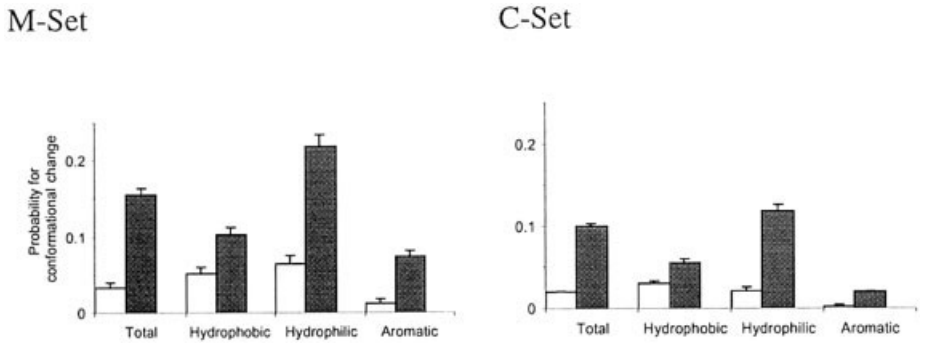

Fig. 5. Side-chain conformational changes as a function of solvent accessibility. White bars, buried residues; black bars, exposed residues; total, all residues; hydrophobic residues, Ile, Leu, Met, and Val; hydrophilic residues (polar and charged), Arg, Asn, Asp, Gln, Glu, Ser, Thr, and Lys; aromatic residues, His, Phe, Tyr, and Trp.

buried residues is considerably reduced for hydrophobic residues (twofold), while for aromatic residues it is more than sixfold. The large restriction in side-chain rearrangement for buried aromatic residues is likely connected with the reduced probability for maneuverable space within the protein core for the large, rigid aromatic rings. The average flexibility for all residues in the M-Set is 1.6 times that of the C-Set. While this is also the approximate ratio for hydrophobic and exposed polar residues, buried polar residues are three times more flexible near mutations.

Conformations of exposed residues depend on interactions between molecules in the crystal, among other factors. These conformations could differ because of dissimilar relative molecular orientations in the crystal lattices. To analyze this effect, we subdivided the C-Set into those pairs crystallized with the same space group (62 pairs) or different space groups (7 pairs). We find that, indeed, exposed residues are more affected in the case of different space groups (18% flexibility as compared to 9% in the same space group). Because 90% of the C-Set is composed of pairs with the same crystal space group, the overall results derived for the C-Set would be only slightly affected and are not adjusted.

Correlation Between Flexibility and the Original Side-Chain Conformation

The effect of the original side-chain conformation on the probability for conformation change was examined for equivalent residues from M- and C-Set file pairs. Two rotamers were assigned for each residue under consideration using a reduced backbone-independent rotamer library. The data, summed for all residues, are available as

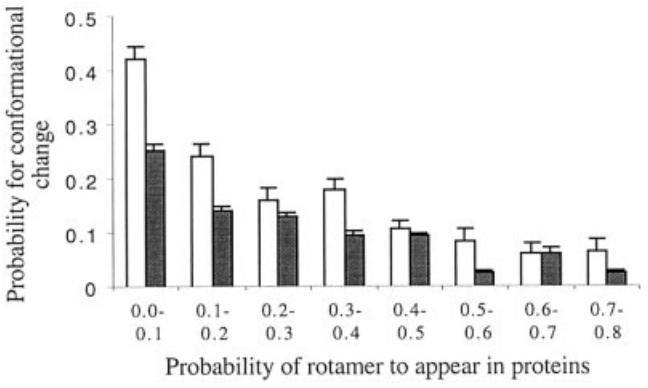

Fig. 6. Probability for side-chain conformational change as a function of probability for side-chain conformations to be found in native structures. This probability for side-chain conformations is taken from a backbone-independent rotamer library.[55] White bars, residues in mutation regions (M-Set); black bars, residues in the control set (C-Set).

an all-against-all rotamer matrix for each amino acid type (see, http://bioinfo.weizmann.ac.il/MutaProt/mutations/). Using this data, we determined the correlation between the probability of a side-chain to change conformation and the probability of its conformation to appear in the structural database (Fig. 6). Clearly, residues having less frequent conformation (reflecting unfavorable internal interaction) have a greater tendency to undergo conformational change. This trend is stronger for structural changes induced by a mutation.

DISCUSSION

Increased Flexibility at Mutation and Ligand Binding Sites

On average, point mutation causes an increase of about 60% in changed rotamers among contacting residues versus the inherent flexibility (Tables II and III). Arg, Ile, and Val show large increases (at least 35% after considering statistical error) in flexibility at mutation regions. Asn and Glu show smaller increases ($\approx$ 15%), while for all the rest the increase in flexibility is not statistically significant [Fig. 4(C)]. Ligand binding, which can evoke larger steric disturbances and solvent effects, results in greater structural change, amounting to an average overall increase of about 120% above the inherent flexibility [cf., Fig. 4(C)]. Some residues (Arg, Ile, and Asn) demonstrate increased side-chain flexibility independent of backbone movement, both in ligand binding sites and in mutation regions. This suggests that they are relatively sensitive to any structural change in their surrounding.

Modeling of Side-Chain Conformations

The results of this study suggest some empirical rules for modeling regions of point mutations, given the structure of the original protein. Knowledge-based rules can significantly reduce the size of search space that, otherwise, might be too large for evaluation, even using rotamer libraries. Our data (Fig. 3) indicate that in 91–95% of cases two or fewer residues change conformation in the vicinity of a point mutation. This means that if the analysis is restricted to the first shell only a few residues (plus the mutated one) need be simultaneously considered as flexible during the modeling, resulting in significant reduction of the dihedral angle combinations. Of course, other factors as well correlate with the probability of a given side-chain to change its conformation. Among them are the residue type, the solvent accessibility of the residue, and the original conformation of the residue. We found that side-chains having conformations that are uncommon in protein structures have a higher probability to undergo conformational change than commonly found ones (Fig. 6). This is likely due to their relatively higher potential energy. As seen in Figure 6, the trend is stronger for mutation regions than for the protein overall. An explanation for this might be that an induced structural change affords an opportunity to escape from an unfavorable conformation enforced by a previous environment. Our results agree with the demonstrated inverse correlation between B-factor values and the probability for a particular side-chain conformation to appear in a protein.[55]

All the above can be combined in an algorithm that chooses small sets of simultaneously flexible residues. The B-factor also correlates with flexibility (Fig. 2). Therefore, it is worthwhile to incorporate B-factor values in algorithms for modeling side-chains.

Background Flexibility of Side-Chains and Prediction Limits

Several studies included analysis of side-chain dihedral angles for equivalent residues in similar or identical proteins.[5,29,49,51,52,56] Direct comparison with our results is complicated due to differences in quality of the crystallographic data included and criteria for analyzing conformational changes. Our results demonstrate high conservation (95%) for the side-chain conformations compared with some of these works, such as Janin et al.[5] (70%), Flores et al.[51] (71%), and Betts and Sternberg[52] (81%). In all these cases, analyses were for crystal structures with resolutions 2.5–2.7 Å or better, while we considered only structures with resolutions 2.0 Å or better. This is probably the main source of differences between our results and the works listed above.

The importance of resolution is well demonstrated by a subset of 6 pairs of independently resolved structures with resolution of 2.0 Å or better that we extracted from the 12 pairs studied by Betts and Sternberg.[52] For the six pairs of high-resolution structures, the conservation of side-chain conformations increases from 81 to 89% (for χ_{1+2}). We considered the possibility that much of the remaining difference with our data is due to a larger fraction of pairs of structures in our control set that were resolved by molecular replacement. Assuming that a pair of structures resolved by the same author has a greater chance to involve molecular replacement, we extracted a subset of 14 pairs of structures from our control data in which pair members were submitted to the PDB by different authors. The conservation in this case was 91% for χ_{1+2} (86% for exposed residues and 94% for buried ones).

The results obtained for these two subsets suggest that about 10% of side-chains accept different conformations in independently resolved crystals. These numbers must be kept in mind when one is validating modeling results for individual PDB structures. Xiang and Honig[16] almost reached the upper limit of correct predictions for side-chain conformations of buried residues (94 and 89% for χ_1 and χ_{1+2}, respectively, predicted to within 20° of crystal structure dihedral angles values). Because there is little reason to suspect that the increase in inherent flexibility for independently determined structures changes the amino acid type pattern of flexibility, we argue that, in such cases, the set of incorrectly predicted residues will overlap significantly with the flexibility pattern for the different amino acids shown in Figure 4.

Flexibility of Specific Amino Acids

The flexibility of specific amino acids can be analyzed based on the results presented in Tables II and III and Figure 4. Lys is the most flexible residue in the majority of measurements. While Lys and Arg have the same number of dihedral angles and similar charges, they are statistically different in their flexibility levels ($P < 0.001$ in the C-Set; $P < 0.05$ in the M-Set). This is only partly explained by the different distribution of these residues in globular protein structures (Lys is found on the surface more frequently). Possibly, the larger number of stabilizing interactions that can be formed by the guanidium group of Arg as compared to the amide group of Lys contributes to the relative rigidity of the former. In addition, the explanation may lie in steric difficulties to find suitable position for the large, rigid, nearly planar terminal group of Arg. Such speculations remain to be tested.

Glu and Asp appear more flexible than Gln and Asn, respectively, until the terminal dihedral angle, where the trend is reversed [Fig. 4(A)]. This is likely due to incorrect atomic assignment of Gln OE1-NE2 and Asn OD1-ND2. From our results in Tables II and III, one can estimate what is the fraction of such incorrect assignment. Assuming the probability of terminal angle changes for Glu and Gln (or for Asp and Asn) to be similar, we estimate 5–10% incorrect assignments of terminal atoms. Based on optimization of hydrogen bonds networks, Hooft et al.[60] previously concluded that some portion of Gln and Asn residues (14–18%) had incorrect terminal atoms assignment.

There are several arguments to explain why Ser is found to be so flexible [Fig 4(B) and ref. 56]. Energy barriers for rotation around the C_α—C_β bond are relatively low; thus, switching between rotamers is frequent. Further, the exposed Ser side-chains are often poorly defined in electron density maps,[5] and it was suggested they are often resolved as a weighted average of two discrete conformations.[7] Thus, a relatively large fraction of Ser residues might be wrongly positioned. In line with the above, Ser has an almost equal probability to be found in each of its canonical rotamers.[5,6,14] The large differences in flexibility between Ser and Thr (Fig. 4) are explained by restrictions in mobility caused by steric interactions of the extra γ carbon in Thr. On the other hand, this carbon atom also facilitates more accurate determination of the side-chain position and eliminates errors usually scored as conformational change.

Our results are consistent with the well-known rigidity of the aromatic side-chains. For these residues, displacement analysis of the side-chain terminal atoms is a more sensitive measure than dihedral angle analysis due to the large rigid component in their side-chains. Cys is the only nonaromatic amino acid that demonstrates a comparably low level of flexibility. The reason here is mainly the disulfide bonds and interactions with metals, which stabilize its conformation.[5]

ACKNOWLEDGMENTS

The authors thank M. Babor for programming assistance and M. Eisenstein and G. Vriend for their valuable comments on the article. This work was supported in part by a Strategic Infrastructure Grant from the Israeli Ministry of Science and by the Ministry of Absorption, the Center for Absorption of Scientists. E.E. is supported by a grant from the Israel Ministry of Science.

REFERENCES

1. Shortle D. Mutational studies of protein structures and their stabilities. Rev Biophys 1992;25:205–250.
2. Eigenbrot C, Kossiakoff AA. Structural consequences of mutation. Curr Opin Biotech 1992; 333–337.
3. Rejto PA, Freer ST. Protein conformational substates from X-ray crystallography. Progr Biophys Mol Biol 1996;66:167–196.
4. De Filippis V, Sander C, Vriend G. Predicting local structural changes that result from point mutations. Protein Eng 1994;7: 1203–1208.
5. Janin J, Wodak S, Levitt M, Maigret B. Conformation of amino acid side-chains in proteins. J Mol Biol 1978;125:357–386.
6. Schrauber H, Eisenhaber F, Argos P. Rotamers: To be or not to be? An analysis of amino acid side-chain conformations in globular proteins. J Mol Biol 1993;230:592–612.
7. MArthur MW, Thornton JM. Protein side-chain conformations: a systematic variations of $\chi 1$ mean values with resolution—a consequence of multiple rotameric states? Acta Crystallogr 1999;D55: 994–1004.
8. Dunbrack RL, Karplus M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. Struct Biol 1994;1:334–340.
9. Hutchinson EG, Thornton JM. A revised set of potentials for β-turn formation in proteins. Protein Sci 1994;3:2207–2216.
10. Street AG, Mayo SL. Intrinsic β-sheet propensities result from van der Waals interactions between side chains and the local backbone. Proc Natl Acad Sci USA 1999;96:9074–9076.
11. Eswar N, Ramakrishnan C. Secondary structures without backbone: an analysis of backbone mimicry by polar side chains in protein structures. Protein Eng 1999;12:447–455.
12. Eswar N, Ramakrishnan C. Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures. Protein Eng 2000;13:227–238.
13. Tuffery P, Etchebest C, Hazout S, Lavery R. A new approach to the rapid determination of protein side chain conformations. J Biomol Struct Dyn 1991;8:1267–1289.
14. Dunbrack RL, Karplus M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. J Mol Biol 1993;230:543–574.
15. Koehl P, Delarue M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. J Mol Biol 1994;239:249–275.
16. Xiang Z, Honig B. Extending the accuracy limits of prediction for side-chain conformations. J Mol Biol 2001;311:421–430.
17. Lee C, Subbiah S. Prediction of protein side-chain conformation by packing optimization. J Mol Biol 1991;217:373–388.
18. Holm L, Sander C. Fast and simple Monte Carlo algorithm for side

chain optimization in proteins: Application to model building by homology. Proteins 1992;14:213–223.
19. Hwang, JK, Liao, WF. Side-chain prediction by neural networks and simulated annealing optimisation. Protein Eng 1995;8:363–370.
20. Shenkin PS, Farid H, Fetrow JS. Prediction and evaluation of side-chain conformations for protein backbone structures. Proteins 1996;26:323–352.
21. Kussell E, Shimada J, Shakhnovich EI. Excluded volume in protein side-chain packing. J Mol Biol 2001;311:183–193.
22. Sutcliffe MJ, Hayes FRF, Blundel TL. Knowledge based modeling of homologous proteins, part II: rules for the conformations of substituted side chains. Protein Eng 1987;1:385–392.
23. Laughton CA. Prediction of protein side-chain conformations from local three-dimensional homology relationships. J Mol Biol 1994; 235:1088–1097.
24. Chinea G, Padron G, Hooft RWW, Sander C, Vriend G. The use of position-specific rotamers in model-building by homology. Proteins 1995;23:415–421.
25. Bower MJ, Cohen FE, Dunbrack RL. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. J Mol Biol 1997;267:1268–1282.
26. Mendes J, Bapitista AM, Carrondo MA, Soares CM. Improved modeling of side chains in proteins with rotamer-based method: A flexible rotamer model. Proteins 1999;37:530–543.
27. Feig M, Rotkiewicz P, Kolinski A, Skolnick J, Brooks CL. Accurate reconstruction off all-atom protein representations from side-chain-based low-resolution models. Proteins 2000;41:86–97.
28. Samudrala R, Moult J. Determinants of side-chain conformational preferences in protein structures. Protein Eng 1998;11:991–997.
29. Summers NL, Carlson WD, Karplus M. Analysis of side-chain orientations in homologous proteins. J Mol Biol 1987;196:175–198.
30. Reid LS, Thornton JM. Rebuilding flavodoxin from C_α coordinates: A test study. Proteins 1989;5:170–182.
31. Levitt M. Accurate modeling of protein conformation by automatic segment matching. J Mol Biol 1992;226:507–533.
32. Eisenmenger F, Argos P, Abagyan R. A method to configure protein side-chains from the main-chain trace in homology modelling. J Mol Biol 1993;231:849–860.
33. Wilson C, Gregoret LM, Agard DA. Modeling side-chain conformation for homologous proteins using an energy-based rotamer search. J Mol Biol 1993;229:996–1006.
34. Cheng B, Nayeem A, Scheraga HA. From secondary structure to three-dimensional structure: Improved dihedral angle probability distribution function for use with energy searches for native structures of polypeptides and proteins. J Comput Chem 1996;17: 1453–1480.
35. Desmet J, De Maeyer M, Hazes B, Lasters I. The dead-end elimination theorem and its use in protein side-chain positioning. Nature 1992;356:539–542.
36. Goldstein RF. Efficient rotamer elimination applied to protein side-chains and related spin-glasses. Biophys J 1994;66:1335–1340.
37. Lasters I, De Maeyer M, Desmet J. Enhanced dead-end elimination in the search for the global minimum energy conformation of a collection of protein side-chains. Protein Eng 1995;8:815–822.
38. De Maeyer M, Lasters I. All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of side-chains by dead-end elimination. Fold Des 1997;2:53–66.
39. Lee C. Predicting protein mutant energetics by self-consistent ensemble optimization. J Mol Biol 1994;236:918–939.
40. Vasquez M. An evaluation of discrete and continuum search techniques for conformational analysis of side-chains in proteins. Biopolymers 1995;36:53–70.
41. Voigt CA, Gordon DB, Mayo SL. Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. J Mol Biol 2000;299:789–803.
42. Lasters I, Desmet J. The fuzzy-end elimination theorem—correctly implementing the side-chain placement algorithm based on the dead-end elimination theorem. Protein Eng 1993;6:717–722.
43. Keller DA, Shibata M, Marcus E, Ornstein RL, Rein R. Finding the global minimum: A fuzzy end elimination implementation. Protein Eng 1995;8:893–904.
44. Liang SD, Grishin NV. Side-chain modeling with an optimized scoring function. Protein Sci 2002;11:322–331.
45. Eyal E, Najmanovich R, Sobolev V, Edelman M. MutaProt: a Web interface for structural analysis of point mutations. Bioinformatics 2001;17:381–382.
46. Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M. Automated analysis of interatomic contacts in proteins. Bioinformatics 1999;15:327–332.
47. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: A computer based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
48. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
49. Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side-chain flexibility in proteins upon ligand binding. Proteins 2000;39:261–268.
50. Petrella R, Lazaridis T, Kurplus M. Protein side chain conformer prediction: a test of the energy function. Fold Des 1998;3:353–377.
51. Flores TP, Orengo CA, Moss DS, Thornton JM. Comparison of conformational characteristics in structurally similar protein pairs. Protein Sci 1993;2:1811–1826.
52. Betts MJ, Sternberg MJE. An analysis of conformational changes on protein–protein association: implications for predictive docking. Protein Eng 1999;12:271–283.
53. Lee B, Richards FM. The interpretation of protein structure: Estimation of static accessibility. J Mol Biol 1971;55:379–400.
54. Parthasarathy S, Murthy MRN. Analysis of temperature factor distribution in high-resolution protein structures. Protein Sci 1997;6:2561–2567.
55. Dunbrack RL, Cohen FE. Bayesian statistical analysis of protein side chain rotamer preferences. Protein Sci 1997;6:1661–1681.
56. Petsko GA, Ringe D. Fluctuations in protein structure from x-ray diffraction. Annu Rev Biophys Bioeng 1984;13:331–371.
57. Carugo O, Argos P. Correlation between side-chain mobility and conformation in protein structures. Protein Eng 1997;10:777–787.
58. Kamphuis IG, Drenth J, Baker EN. Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. J Mol Biol 1985;182:317–329.
59. Zhao SR, Goodsell DS, Olson AJ. Analysis of a data set of paired uncomplexed protein structures: New metrics for side-chain flexibility and model evaluation. Proteins 2001;43:271–279.
60. Hooft RWW, Sander C, Vriend G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. Proteins 1996;26:363–376.