Tolerance to the Substitution of Buried Apolar Residues by Charged Residues in the Homologous Protein Structures

S. Balaji,[1] S. Aruna,[2] and N. Srinivasan[1]*
[1]*Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India*
[2]*Centre for Biotechnology, Anna University, Chennai, India*

ABSTRACT Occurrence and accommodation of charged amino acid residues in proteins that are structurally equivalent to buried non-polar residues in homologues have been investigated. Using a dataset of 1,852 homologous pairs of crystal structures of proteins available at 2Å or better resolution, 14,024 examples of apolar residues in the structurally conserved regions replaced by charged residues in homologues have been identified. Out of 2,530 cases of buried apolar residues, 1,677 of the equivalent charged residues in homologues are exposed and the rest of the charged residues are buried. These drastic substitutions are most often observed in homologous protein pairs with low sequence identity (<30%) and in large protein domains (>300 residues). Such buried charged residues in the large proteins are often located in the interface of sub-domains or in the interface of structural repeats, Beyond 7Å of residue depth of buried apolar residues, or less than 4% of solvent accessibility, almost all the substituting charged residues are buried. It is also observed that acidic sidechains have higher preference to get buried than the positively charged residues. There is a preference for buried charged residues to get accommodated in the interior by forming hydrogen bonds with another sidechain than the main chain. The sidechains interacting with a buried charged residue are most often located in the structurally conserved regions of the alignment. About 50% of the observations involving hydrogen bond between buried charged sidechain and another sidechain correspond to salt bridges. Among the buried charged residues interacting with the main chain, positively charged sidechains form hydrogen bonds commonly with main chain carbonyls while the negatively charged residues are accommodated by hydrogen bonding with the main chain amides. These carbonyls and amides are usually located in the loops that are structurally variable among homologous proteins. Proteins 2003;53:783–791. © 2003 Wiley-Liss, Inc.

Key words: buried residues; charged residues; comparative modeling; hydrophobic effect; protein structures; site-directed mutagenesis

INTRODUCTION

It is well known that three-dimensional structures of homologous proteins are conserved better than their amino acid sequences. The extent of conservation of the three dimensional (3-D) structures, as measured by parameters such as root mean square deviation, varies as a function of extent of sequence similarity.[1–6] The overall structural difference is the cumulative effects of local structural variations due to sequence divergence in the homologous proteins. As changes in the solvent exposed residues, which are often in the loops, are more extensive than the changes in the core of the structure, the divergence in the structure of protein surfaces is usually higher than in the interior of the homologous proteins.[4] However, it was also observed that the sequence variations could take place even in the cores of homologous proteins, hence causing the divergence in the structures of the cores.[1,4,7] Naturally occurring substitution mutations of hydrophobic to polar residues, especially the charged residues, that are in the core of the homologous proteins are rare but they are nevertheless observed in the diverse family of proteins.[1,7] Such changes may be expected more commonly in the distantly related proteins, with low sequence similarity between them, rather than in the closely related proteins. Drastic mutations in the core, generated using site-directed mutagenesis, could cause substantial variations in the core structure resulting from movements in the backbone to accommodate such a substitution.[8,9] But such a mutation affects the stability of the protein in some cases[10–15] and, in turn, may influence its function.[16]

Comparative studies of drastic mutations in the core of homologous proteins would have implications on protein engineering and design and also in the comparative modeling of protein structures. Currently, the most accurate structure prediction method is comparative modeling in which the three-dimensional structure of the target sequence is built on the basis of the homologous protein

structures.[17–22] In a comparative modeling exercise, given the alignment of target sequence to the template structure(s), if there is a buried non-polar residue in the template aligned with a charged residue in the target sequence, it is not straightforward to decide if the charged residue should be exposed or buried. Even in the cases where the decision is made to bury the charged residue, the question of how to accommodate the charge in the interior remains to be answered. Hence the analysis of the tolerance to substitution of the buried non-polar to charged residues in the structurally conserved regions of the homologous proteins would give clues to decision making in these circumstances.

The present analysis uses a large number of pairs of homologous protein structures and investigates the instances of substitution of a buried apolar residue in a protein by a charged residue in the structurally equivalent position of a homologue. Our analysis suggests that such drastic substitutions do occur in the homologous proteins. We also present strategies adopted by the proteins for the accommodation of the charged group in the interior.

MATERIALS AND METHODS

The data set analyzed comprises of 1,852 pairs of structurally aligned homologous protein domains extracted from the PALI (release 1.3) database[23,24] (http://pauling.mbu.iisc.ernet.in/~pali) such that both the domains in a pair have 2Å or better crystallographic resolution. The version of PALI database used in the analysis has 614 protein families (families with two or more members) and more than 17,000 pair-wise alignments of homologous protein structures involving 3,050 protein domains aligned by rigid body superposition using the program STAMP.[25]

Analysis of the data set involved the following steps:

1. The structurally conserved regions (SCR) in every aligned pair of protein domains have been identified. SCRs in a homologous protein domain pair are the equivalent regions of high structural similarity. In the present analysis, SCRs have been assigned for the homologous protein pairs as indicated by the STAMP, which designates structurally conserved regions by taking in to consideration the deviation of equivalent Cᵅs and also the conformational similarity of the neighboring residues.
2. Instances of substitution of non-polar residues (Leu, Ile, Ala, Val, Trp, Phe, Met) by polar residues (Lys, Arg, His, Glu, Asp, Gln, Asn, Ser, Thr, Cys) in the SCRs have been identified in the 1,852 pairs of homologous proteins in the data set. Tyrosine is not considered for the present analysis because of its dual nature to behave as a polar and a non-polar residue.
3. The subset of buried non-polar residues that are substituted by polar residues have been identified. A residue is defined as buried if its side-chain solvent accessibility[26] is less than or equal to 7%.
4. Hydrogen bonding partners of polar chemical groups in the buried charged residues have been identified and analyzed further. Identification of potential hydrogen bonding is based on the method used by Overington et al.[27] and encoded in the computer program HBOND. The criteria for identification of potential hydrogen bond involve distance between putative donors and acceptors that is set at 3.5Å as the upper limit. The salt bridges are recognized if there is a potential hydrogen bond between the oppositely charged chemical groups coming from the sidechains of acidic and basic residues.

Depths of buried non-polar residues that are substituted by charged residues and the substituting charged residues have been calculated using the DEPTH[28] program. The algorithm employed by this program involves placing the protein structure of interest in a pre-equilibrated box of water molecules obtained from a Monte-Carlo simulation. As in solvent accessible surface area calculations,[26] a water molecule is approximated to a sphere of diameter 2.8Å. A number of water molecules are removed from consideration based on the short contact among water molecules as well as between water molecules and protein atoms as detailed by Chakravarty and Varadarajan.[28] The final picture that emerges is the gathering of water molecules all around the protein surface. The surface of the protein that is in contact with water molecules is very similar to the water contact surface of the protein involved in solvent accessible surface area calculations.[26] The depth of an atom in a protein is defined as the distance of the atom from the centre of the nearest surface water molecule. A residue depth is calculated as the mean of depths of the atoms in the residue. The parameter depth is capable of distinguishing the buried residues that are close to the surface and those at the interior of a protein molecule.

RESULTS
Frequency of Occurrence of Drastic Residue Substitutions in the Homologous Proteins

In the 1,852 pairs of structurally aligned homologous protein domains in the data set, the total number of residue-residue alignment positions in the structurally conserved regions (SCRs) is 189,125. The number of non-polar to polar residue substitutions in SCRS is about 16% (29,601) of the total number of alignment positions (189,125). This clearly indicates that non-polar to polar residue substitution is not uncommon in the SCRs of homologous proteins. Of these 29,601 non-polar to polar residue substitutions, the number of positions that have non-polar to charged residue substitution is 14,024 (about 47% of non-polar to polar residue substitutions). Out of these, the number of buried non-polar to charge residue substitutions is 2,530 (which is about 18% of non-polar to charged residue substitutions). Among 2,530 cases of such substitutions, there are 1,677 (66.3%) cases where the substituting charged residue is exposed and 853 cases the charged residue is buried. These 853 cases correspond to 379 independent examples of buried charged residues that are spread over 194 protein domains and these are listed in the *Supplementary information* with the electronic release of this paper.
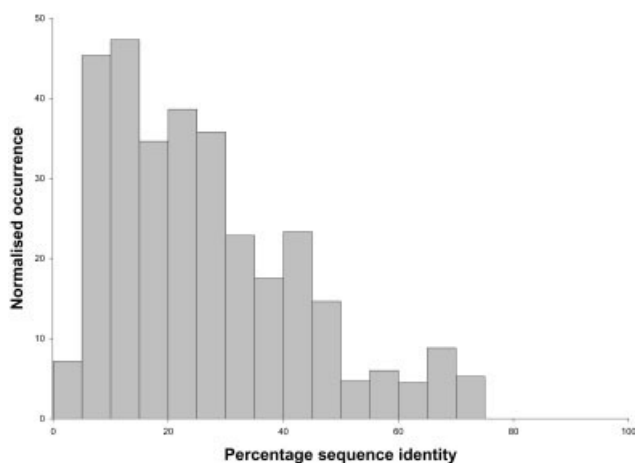
Fig. 1. Distribution of normalized occurrence of protein pairs that have buried apolar to charged residue substitutions in SCRs. The distribution is averaged for every 5% of percentage sequence identity

Circumstances of Buried Apolar to Charged Residue Substitutions

To learn about the circumstances under which buried apolar to charged residue substitutions occur, investigations have been made to understand the dependency of such substitutions on the percentage sequence identity of the aligned proteins and also on the protein sequence length.

The total number of unique domains that have at least one buried charged residue substituting a buried non-polar residue in a homologue is 194. Out of these, 75 are from different families in PALI. The normalized occurrence of the pairs of protein domains that have buried non-polar to charged residue substitutions in SCRs are plotted against sequence identity of the proteins in the pair (Fig. 1). The normalization corresponds to the number of pairs involving buried non-polar to charged residue substitutions in a range of sequence identity divided by the number of pairs in that range in the entire data set. Figure 1 clearly shows pronounced peaks for the pairs that have less than 30% sequence identity. The distribution is skewed towards the low sequence identity regions (less than about 20%). The distribution up to 30% sequence identity corresponds to 907 pairs in the data set, which is nearly half the number of pairs in the data set (1,852 pairs). This clearly indicates that these substitutions are more common in the pairs with low sequence identity.

An earlier work[29] reported that buried charged residues occur more commonly in larger proteins than in smaller proteins. In the distribution (not shown) corresponding to the present dataset, we too find that the number of buried charged residues generally increases with an increase in protein size. However, the number of domains with lengths longer than 400 residues (12 domains) is very small and hence such examples are not considered in deriving the conclusions. It is interesting to note that these domains come from either α class or α/β class of folds and the highest representation from proteins adopting α-α toroid folds in the α class. These buried charged residues are often in regular secondary structures and, in most cases, these residues are involved in hydrogen bond interactions with a residue from another compact unit of the fold that tends to form a separate structural sub-domain/module. The secondary structures with which these buried charged residues interact are often solvent exposed in the face opposite to the face that interacts with the buried charged residue. If one imagines the loss of such interactions enabled by subtle gross movement of the sub-domain away from the rest of the fold, then this might result in the formation of a separate structural domain with the charged residue now in the inter-domain interface and possibly better exposed to the solvent. In the cases of a few examples from the α-α toroid fold, the charged residues come from one of the repeats of antiparallel α-helices and it interacts with the adjacent repeat. Figure 2 shows a representative example from a large protein cholesterol esterase from Bos Taurus corresponding to the protein data bank[30] entry of 2bce.[31]

Depths of the Buried Non-Polar Residues and Substitutions

From the above, it is clear that the buried non-polar to charged residue substitutions are not uncommon in the pairs of homologous proteins with low sequence identity. The nature of residues getting substituted under these circumstances and how deep the substituted non-polar residues and the substituting charged residues are from the surface of the structure are discussed here. Figure 3(a) is the plot of the depths of buried apolar residues in a set of proteins vs. the depths of buried and exposed charged residues present in the topologically equivalent positions in the homologues. Average depths of the exposed [Fig. 3(a) open circles] and buried [Fig. 3(a) filled circles] charged residues for every 1Å range of depth of substituted buried non-polar residue are represented. The plot shows a linear increase in the depth of buried charged residue as the depth of buried non-polar residue increases. Such a trend is not seen for the depth of exposed charges [see also scatter plot: Fig. 3(b)]. From the plot, it is also evident that the depths of buried charged residues are slightly less than those of the substituted buried non-polar residues. Figure 3(b) is the scatter plot of depth of exposed charged residues against the depth of substituting buried apolar residues and Figure 3(c) is the scatter plot of depth of buried charged residues vs. the depth of substituting buried apolar residues. It appears from Figure 3(c) that there is a linear increase of depth of charged residue with an increase in depth of substituting apolar residue.

The normalized number of occurrences of buried non-polar residues substituted by exposed charged residues in the data set calculated for every 1Å range of depth of non-polar residues is given in Table Ia. The corresponding distribution of normalized occurrence of buried non-polar residues substituted by buried charged residues is shown in Table Ib. Figure 4 is a plot of the depths of buried non-polar residues in the above-mentioned cases and the normalized occurrence. From Figure 4, it can be seen that the peak corresponding to the data in Table Ia is broad, in
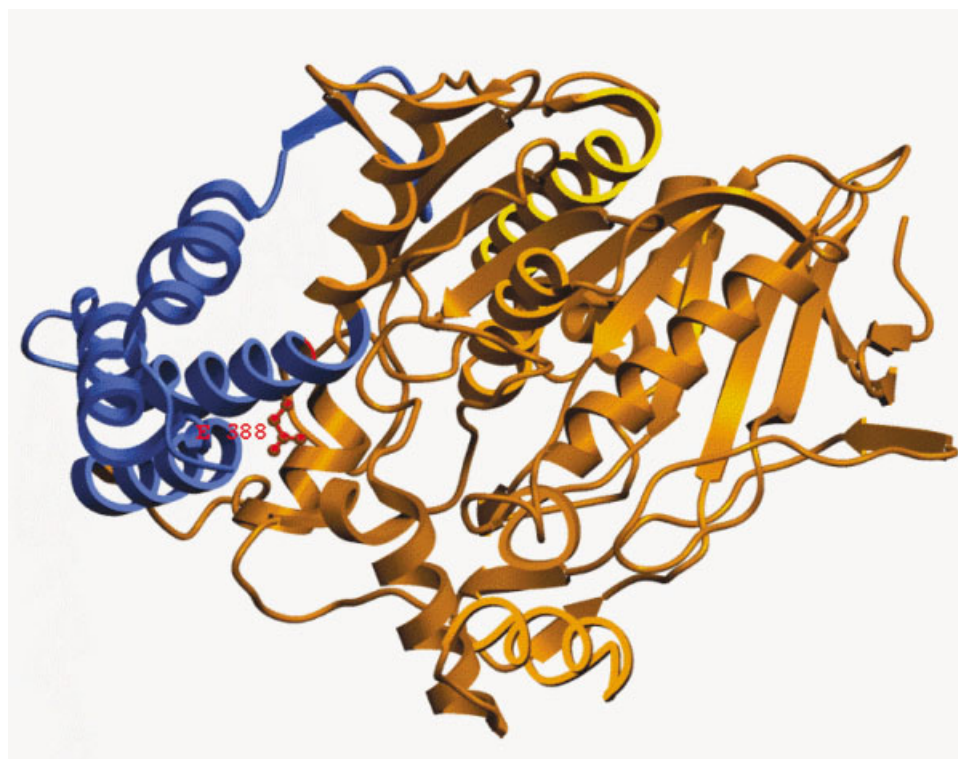
Fig. 2.   Illustration of buried charged residue (Glu 388) in a sub-domain interacting with the adjacent sub-domain in the structure of Cholesterol Esterase from Bos Taurus.[31] The two subdomains are shown in different colors. The figure has been prepared using Setor.[32]

the range 4–7Å, while the peak corresponding to the data in Table Ib is sharper and steeper and falls in the range of 5–6Å of depth. Due to a significant overlap in the ranges of depths in the two distributions shown in Figure 4, it may be difficult to predict if a charged residue that replaces a buried apolar residue will be exposed or buried. It is evident from tables IA and IB and Figure 4 that beyond 7Å of depth of the buried apolar residue, the normalized frequency of occurrence of buried charged is slightly higher than that of the exposed charged residues. Hence, it is more likely that the substituting charged residues are also buried when the buried non-polar residues are deeper than 7Å. Beyond 7Å depth, the normalized occurrence of charged residues, excluding histidine, are higher for buried apolar to buried charged residue substitutions (Table Ib: 45% for Lys, 28% for Arg, 41% for Glu, and 34% for Asp) as compared to the corresponding ones for buried apolar to exposed charged residue substitutions (Table Ia: 21% for Lys, 24% for Arg, 27% for Glu, and 14% for Asp). This means that all the charged residues except His have a higher tendency to be buried in the structure when they replace an apolar residue buried beyond 7Å. It is deduced that the average depth of the buried apolar residues that are substituted by buried charged residues is 6.5Å and the average depth of substituting charged residues is 6.3Å. It is also interesting to note that the average depth of other buried apolar residues in the data set used is 6.1Å. Hence, it could be inferred that the buried apolar residues that are being substituted by buried charged residues are deeper

on an average than other buried apolar residues in the protein.

We have compared the depths of charged groups, such as carboxylate and amino groups, in the charged sidechains to the depth of charged residue taken as a whole. In general, the depth of the charged group is only slightly smaller than that of the entire sidechain if the residue is buried. For example, the average depth of buried charged groups, shown in Figure 4, is 6.0Å and the average depth of buried charged sidechains is 6.3Å. The average depths of charged groups and charged sidechains for the exposed charged residues are 3.4 and 4.2Å, respectively.

Preferred Substitution Patterns of Apolar to Charged Residues Dependent on Extent of Burial and Sizes

In the majority of the cases where the buried non-polar residues is substituted by charged residues, the charged residues usually get exposed (1,677 out of 2,530 instances) to the solvent. For the situations of buried charged residues (853 out of 2,530 instances), Figure 5 shows the plot of side-chain solvent accessibility of buried non-polar residues vs. the normalized number of charged residue substitutions. Figure 5 shows that the distribution falls as the side-chain solvent accessibility of the buried non-polar residue increases up to about 4% of the accessibility. Then the distribution raises slightly. This means that buried charged residues more often replace the non-polar residues that are buried to a greater extent (less than 4%) than
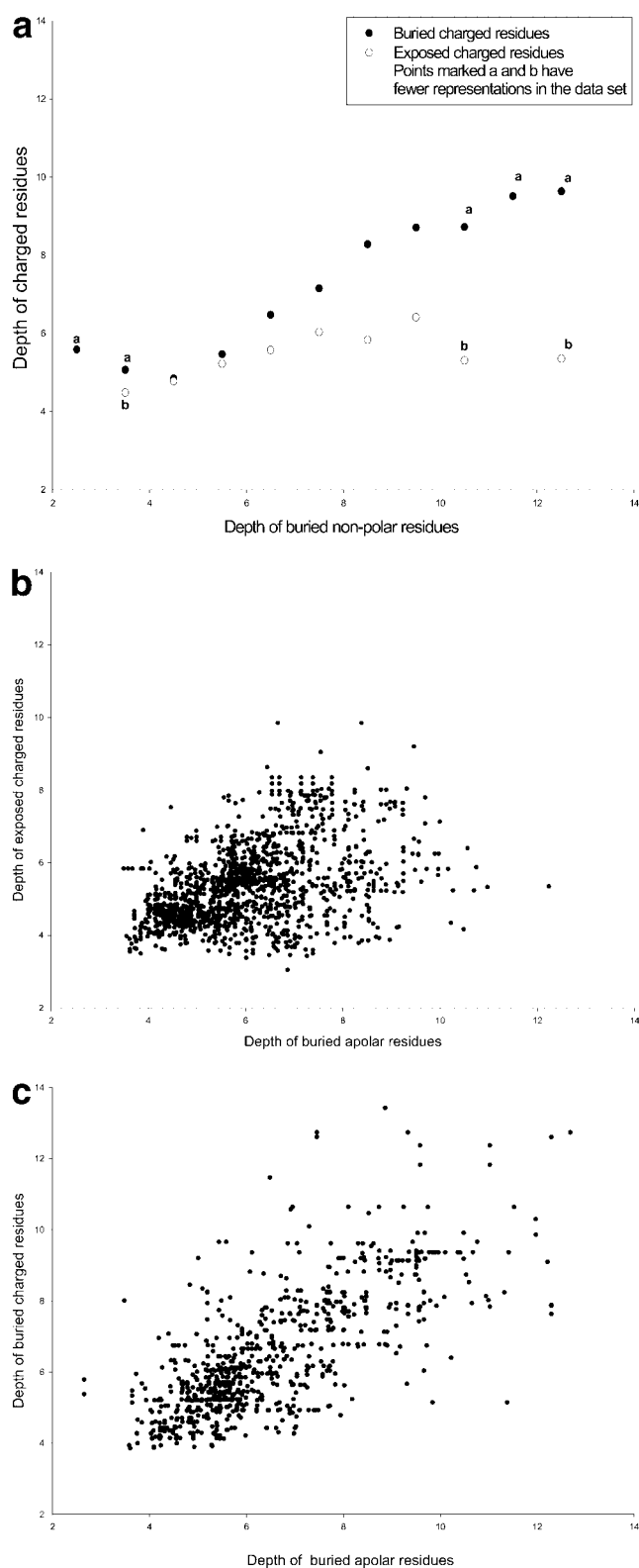
Fig. 3. **a:** Plot of the depth of buried apolar residues (in Å) and the average depth of substituting charged residues (in Å) for every 1Å interval for the cases when the charged is exposed and buried. **b:** Plot of the depth of buried apolar residues (in Å) and the depth of substituting exposed charged residues (in Å). (*a* and *b* indicate those intervals of depths with small numbers of observations) **c:** Plot of the depth of buried apolar residues (in Å) and the depth of substituting buried charged residues (in Å).

TABLE IA. Distribution of the Depth of Buried Apolar Residues Replaced by Exposed Charged Residues in a Homologue

Range of depth (in Å)		Normalized occurrence of substitutions (in %)					
		K	R	H	E	D	Total
0	1	0	0	0	0	0	0
1	2	0	0	0	0	0	0
2	3	0	0	0	0	0	0
3	4	1.3	2.3	5.7	1.5	6	2.8
4	5	26.4	22.1	26.5	15.0	32.5	24.1
5	6	25.3	27.4	26.5	27.8	33	27.5
6	7	25.1	24.2	22.3	28.6	14.5	23.7
7	8	12.6	15.3	10.4	12.8	7	12.4
8	9	6.6	7.1	4.7	7.5	5.5	6.5
9	10	2.0	1.6	3.8	4.9	1	2.5
10	11	0.7	0	0	1.5	0.5	0.5
11	12	0	0	0	0	0	0
12	13	0	0	0	0.4	0	0.1
13	14	0	0	0	0	0	0

TABLE IB. Distribution of the Depth of Buried Apolar Residues Replaced by Buried Charged Residues in a Homologue

Range of depth (in Å)		Normalized occurrence of substitution					
		K	R	H	E	D	Total
0	1	0	0	0	0	0	0
1	2	0	0	0	0	0	0
2	3	0	0	0.7	0	0	0.3
3	4	6.4	0	1.4	1.4	3.6	1.9
4	5	12.8	12.6	25.1	10.6	14.5	17.4
5	6	10.6	27.4	38.6	30.5	30.9	31.9
6	7	25.5	22.2	11.9	15.6	17.0	16.2
7	8	14.9	11.9	7.1	23.4	13.9	12.8
8	9	19.1	8.1	5.8	5.7	10.9	8.0
9	10	4.3	9.6	5.8	10.6	8.5	7.8
10	11	6.4	1.5	2.7	0.7	0	1.8
11	12	0	3.0	1.0	0.7	0.6	1.1
12	13	0	3.7	0	0.7	0	0.8
13	14	0	0	0	0	0	0

those less buried (4–7%). Thus, it appears that when the accessibility of the apolar residue, which is replaced by a charged residue, is 4% or lower the replacing charged residue is more likely to be buried than when the accessibility of the apolar residue is greater than 4%.

This could be extended to understand if there are preferences for the substitution by charged residues depending on the nature of non-polar residues that is being substituted. The frequency of substitutions (Table IIa) for each type of charged residue topologically equivalent to a buried non-polar residue in a homologue is more for negatively charged residues than the positively charged residues (with the exception of histidine). This is also reflected in the plots shown in Figure 6(a,b). Figure 6(a) shows the distribution of normalized occurrence of buried and exposed positively charged residues (Lys and Arg) substituting buried apolar residues. Figure 6(b) shows a
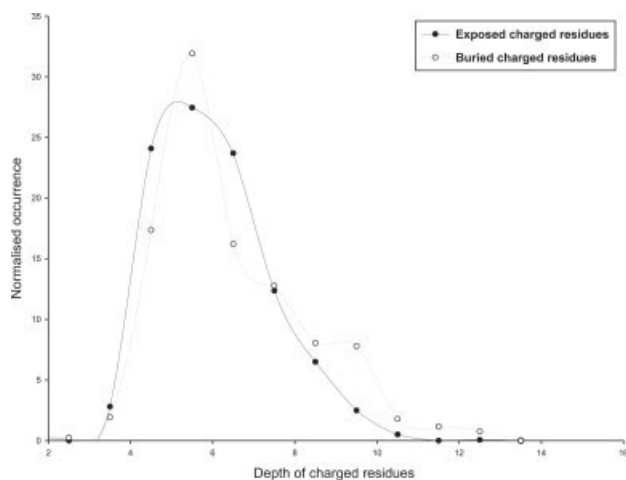
Fig. 4. Plot of average depth of buried apolar residues, averaged for every 1Å interval of depth, and the normalized occurrence for the cases when they are substituted by buried charged residues and by exposed charged residues.
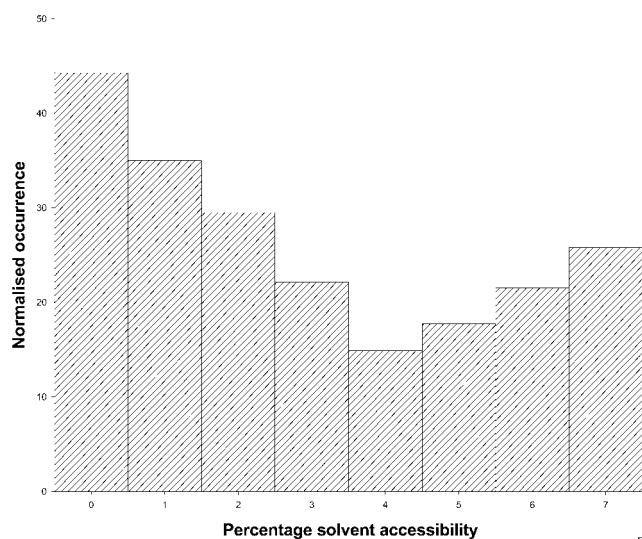

Fig. 5. Distribution of accessibility of buried apolar residues and the normalized occurrence of substituting buried charged residues for every 1Å range.

TABLE IIA. Normalised Occurrence of Buried Apolar to Buried Charged Residue Substitutions

	L	I	A	V	W	F	M	Average
K	9.9	5.2	6.6	6.4	8.7	13.8	22.3	10.4
R	28.3	27.9	16.3	25.2	13.0	16.3	27.3	22.0
D	50.1	67.8	33.8	63.8	80.1	25.1	34.7	50.8
E	22.3	24.4	49.3	31.6	23.1	35.5	64.6	35.8
H	68.1	64.2	45.3	49.4	61.6	82.2	63.4	62.0

similar distribution for the negatively charged residues (Asp and Glu) and histidine.

It is also seen from Table II(a) and Figure 6(b) that the buried Trp to Asp and buried Phe to His substitutions are most frequent and hence indicates a preference for the burial of the charged residues. Substitution by buried
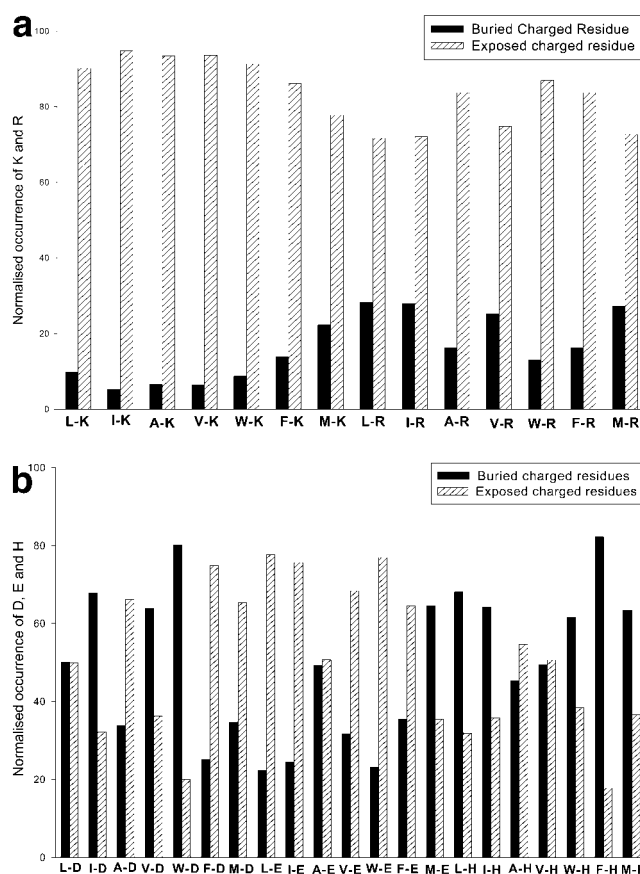

Fig. 6. **a:** Distribution of normalized occurrence of positively charged residues substituting buried apolar residues for all possible substitutions. **b:** Distribution of normalized occurrence of negatively charged residues and histidine substituting buried apolar residues for all possible substitutions.

TABLE IIB. Normalised Occurrence of Buried Apolar to Exposed Charged Residue Substitutions

	L	I	A	V	W	F	M	Average
K	90.1	94.8	93.4	93.6	91.3	86.2	77.8	89.6
R	71.7	72.1	83.8	74.8	87.0	83.7	72.7	78.0
D	49.9	32.2	66.2	36.2	20	74.9	65.3	49.2
E	77.7	75.6	50.7	68.4	76.9	64.5	35.4	64.2
H	31.9	35.8	54.7	50.6	38.4	17.8	36.6	38.0

positively charged residues, Lys and Arg, are in general least frequent. Substitution of any buried apolar residue by buried Lys is less frequent than substitution by buried Arg although both are positively charged. The opposite trend is seen in Table IIb, which corresponds to the frequencies of occurrence of exposed charged residues substituting for buried apolar residues. The substitutions of buried Ala to Glu, Leu to Asp and Val to His have equal preference, as indicated by almost equal peaks in Figure 6(b) and also evident from Table IIa and IIb, for the burial and exposure of the charged residues.

In order to get an idea of the nature of replacement by buried charged residues as a function of the size of the substituted non-polar residues, we have categorized the

TABLE III. Distribution of Small, Medium, and Bulky Apolar to Charged Residue Substitutions

Buried apolar residue type	Substitution (to buried charges) frequency	Nature of substitution
Small-sized residue	31% (785 out of 2,530 instances)	55% of the instances substituted by acidic residues
Medium-sized residues	51% (1,287 out of 2,530 instances)	74% of the instances substituted by basic residues
Bulky residues	18% (458 out of 2,530 instances)	70% of the instances substituted by basic residues

non-polar residues on the basis of their size as small, medium, and bulky. Ala belongs to the first category. Ile, Leu, and Val belong to the category of medium-sized residues, and Phe, Trp, and Met have been classified as the bulky group. Table III gives information about the substitution of these categories of apolar residues by buried charged residues. Medium sized buried non-polar residues are more often substituted (51%) than small-sized (31%) or bulky non-polar residues (18%) and in these instances they are more frequently substituted by charged basic residues (Arg, Lys, His), perhaps due to their large sizes (Table IIa). Basic residues prefers to replace bulky apolar residues while the small apolar residues are replaced more often by acidic sidechains.

Mode of Stabilization of the Buried Charged Residues

In the situations where the buried non-polar residue is substituted by a buried charged residue in SCRs, the buried charged residue could hydrogen bond either to other polar side chains (SC) or to main chain (MC) polar atoms or to both (MSC) or neither. The cases with no hydrogen bond with the protein atoms are usually hydrogen bonded to water molecules in the crystal structures. The MC interactions are observed in 411 substitutions, which involve 271 positively charged residues and 140 negatively charged residues and SC (which includes MSC) cases, are observed in 646 instances of which 371 involve positively charged residues and 275 involve negatively charged residues. It is interesting to note that almost all of these interactions (more than 95%) are seen at the charged group (such as carboxylate and amino groups) depths greater than 8Å. About 80% of the number of interactions correspond to the buried charged residue depths greater than 8Å. In general, the minimum and maximum depths of charged chemical groups in the buried charged residues is 3.3 and 11.7 Å, respectively. The average and standard deviation of the depths of the charged groups are 6.0 and 1.5Å, respectively.

In the MC cases, both from the pure main chain interaction or bifurcated interaction involving main chain (MSC), following are the observations:

1. The carbonyl oxygen atom of the MC is involved in the hydrogen bond to the positively charged side-chain in the majority of the instances (63%, 260 out of 411 instances). In general, 271 of the 411 instances involve positively charged residues. All the remaining 11 instances correspond to buried histidine residue interacting with MC amide.
2. In the majority of cases (65%, 269 out of 411 instances), there is no regular secondary structure at these main-

chain positions. Such substitutions being in loop probably allows some accommodative variations in the main chain conformation and such changes are not easily possible in the helical and β-strand regions. The rest of the residues (142 examples) occur in α-helical and β-sheet regions.
3. Many of the residues (55%, 226 out of 411 instances) whose mainchain forms hydrogen bond with the charged sidechains that substitute buried apolar residues are also in SCR. This means that in these examples, there is not much deviation in the backbone conformation despite the formation of a hydrogen bond with the charged residue in one of the homologues. The remaining 45% (185 out 411 instances) are in SVRs (structurally variable regions) with 47 of these instances corresponding to an insertion or deletion, in alignment with the homologue, suggesting entirely different conformations in these regions.
4. Interestingly, all the buried negatively charged residues, corresponding to 140 instances, are hydrogen bonded to the mainchain amide group.

In the cases of SC and MSC, which form the majority, corresponding to the buried charges interacting with the other sidechain polar groups, following are the observations.

1. 365 out of 646 instances (more than 50% cases) involve salt bridges covering about 70% of the interactions. The other interactions correspond to hydrogen bonds between buried charged sidechains and uncharged polar sidechains. Hence, many of the interactions involve salt-bridges.
2. Many of the residues whose sidechain form hydrogen bond with the buried charged sidechain are (547 out of 646 instances, 85%) also structurally conserved in the homologue in the pair. The rest are in structurally variable regions. Thus, in an overwhelming majority of the homologous pairs the interaction of a polar sidechain with the buried charged sidechain in one of the homologues did not result in a significantly altered conformation at the backbone of the polar residue.
3. The compensatory substitution, involving replacement of hydrophobic interaction between two buried nonpolar residues in one of the homologues by an interaction between buried charged sidechain and another polar sidechain in a homologue, is seen in 50% (375 out of 646) of the instances. In the rest of the examples of no compensatory substitution, no conservation of interaction could be found between the two homologues in a pair. The polar residue in the homologue, without a buried charged residue, that is equivalent to the polar

residue that interacts with the buried charged residue in the other homologue interacts with other sidechains or is exposed to interact with the water. Out of 375 instances, 142 cases correspond to salt bridges. In a vast majority (85%, 319 out of 375) of these cases, both the residues are in SCRs. This suggests that, most often, replacement of interaction between two apolar sidechains by interaction between polar sidechains in a homologue is not accompanied by major changes in the backbone conformation.

DISCUSSION

The observed number of instances of replacement of a buried non-polar residue by a charged residue in a homologue is substantial. Such substitutions could be considered "drastic." In many cases, the substituting charged residues have a higher tendency to get exposed and solvated rather than buried. The 853 instances of buried apolar residues in a protein replaced by a buried charged residue could, in principle, be considered highly undesirable but nevertheless are observed in homologous protein structures. Protein engineers interested in maintaining protein stability in a site-directed mutant do not usually substitute a buried apolar residue by a charged residue.

As the occurrence of buried charged residues is, in general, not a frequent observation in protein structures it could be expected that the larger a protein is the higher the chances of observing buried charged residues. Thus, as expected, the normalized distribution of length (not shown) of the domains in the pairs that have buried charged residue substituting a buried non-polar residue in the other domain in a given pair generally increases with the increase in the domain length. This indicates that the substitutions in SCRs are in general related to the length of the domain that contains the buried charged residue and increases with the increase in the length of the domain that contains the buried charged residue. It has also been observed that the number of such substitutions is more common in pairs of proteins with less than 30% sequence identity. This observation is particularly relevant while modeling a protein structure on the basis of a homologue of known structure with low sequence similarity.

If the depth of the buried apolar residue, which is substituted by a charged residue, is greater than 7Å then the substituting charged residue is also usually buried with the exception of histidine. It appears that a depth of 7Å or higher seems too long for most charged residues to get exposed. Also, if the solvent accessibility of the buried apolar residue is less than 4%, the substituting charged residue in a homologue is also buried. These observations would influence the placing of the sidechains concerned in comparative modeling of protein structures.

Table IIa,b and the distributions in Figure 6(a,b) indicate clearly that negatively charged residues tend to be buried in general more than positively charged residues while substituting the buried non-polar residues. This may be due to the larger size of the positively charged residues (Lys and Arg) compared to negatively charged residues (Asp and Glu) and hence more difficult to accommodate in place of a buried apolar residue. But in the situations where the buried apolar residues are substituted by exposed charged residues [Table IIb and Fig. 6(a),] the frequency of the substitution is higher for positively charged residues than their negatively charged counterparts, especially lysine. Hence, the most positively charged residues (Lys and Arg) usually get exposed when they substitute buried non-polar residues.

The observations on the accommodation of buried charged residues by hydrogen bond formation with the main chain raise the question whether or not there were changes in overall hydrogen bonding/electrostatic interactions between the homologues. In other words, if a buried charged side chain is interacting with a mainchain, is it disrupting an interaction present in the other homologous structure? Although 225 instances correspond to the structurally conserved regions (of the homologues) at the interaction site, only 165 of these instances preserve the secondary structural states of interacting main-chain in the homologues and, hence, only these 165 instances could be strictly considered as a similar backbone structure between two homologues. For every pair of proteins contributing to these 165 instances, we have examined the interaction in the equivalent main chain region of the homologue, which has no buried charged residue to interact with. In 60 out of 165 instances, the homologue without buried charged residue corresponds to mainchain-mainchain interactions without any bifurcation. These instances correspond to conservation of mainchain-mainchain interactions between the homologues and the addition of a bifurcated interaction between the main chain and buried charged sidechain in one of the homologues. Seventeen of the 165 instances correspond to bifurcated (both mainchain-mainchain and mainchain-sidechain) interactions in the homologue without a buried charged residue. These cases correspond to conservation of the mainchain-mainchain interaction, but the bifurcated interaction with the sidechains involves different residues in the homologues. Twenty-three of the instances have only mainchain to sidechain interactions and these sidechains are not in topologically equivalent positions to the buried charged residue in the homologue. We did not detect hydrogen bond in the equivalent mainchain group of the homologue, without a buried charged residue, in 65 of the 165 instances. However, it is likely that most of these mainchain groups are hydrogen bonded to the water molecules. Our dataset did not permit a detailed analysis of this feature as the structures concerned are not available at ultra-high resolution.

On the contrary, if the buried charged residue is accommodated by the formation of a salt bridge or a hydrogen bond formation with a polar sidechain, the backbone conformation at these residues is conserved with the homologue without a buried charged residue. This retention of the backbone structure is seen even if a salt bridge interaction is replaced by interaction between two apolar sidechains. Retention of these tertiary interactions in most instances suggests that these interactions, either salt

bridge or between two apolar groups, are potentially important in the formation or stability of the protein fold.

CONCLUSIONS

From the present analysis, it is evident that the buried apolar to charged residue substitutions in structurally conserved regions of homologous proteins are not uncommon in low sequence identity situations. Also from the analysis, it is found that substituting charged residues is usually solvent exposed. But in some situations, it is buried depending on the nature of substitution, its size, the depth of the substituted apolar residue, and also the presence of an appropriate hydrogen-bonding partner in the vicinity of the substituting position. These results could have implications in comparative modeling of distantly related proteins and also in the protein engineering to design mutants, as in both practices buried non-polar to charged residue substitution is usually considered undesirable. The results could be influential especially in the comparative modeling exercises in the decision making to expose or to bury the charged residue that replaces a non-polar residue.

ACKNOWLEDGMENTS

REFERENCES

1. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in protein. EMBO J 1986;5:823–826.
2. Chothia C, Lesk AM. The evolution of protein structures. Cold Spring Harb Symp Quant Biol 1987;52:399–405.
3. Hubbard TJ, Blundell TL. Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. Protein Eng 1987;1:59–71.
4. Flores TP, Orengo CA, Moss DS, Thornton, JM. Comparison of conformational characteristics in structurally similar protein pairs. Protein Sci 1993;2:1811–1826.
5. Russell RB, Barton GJ. Structural features can be unconserved in proteins with similar folds. An analysis of side-chain to side-chain contacts secondary structure and accessibility. J Mol Biol 1994;244:332–350.
6. Balaji S, Srinivasan N. Use of a database of structural alignments and phylogenetic trees in investigating the relationship between sequence and structural variability among homologous proteins. Protein Eng 2001;14:219–226.
7. Lesk AM, Chothia C. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. J Mol.Biol 1980;136:225–270.
8. Lim WA, Farruggio DC, Sauer RT. Structural and energetic consequences of disruptive mutations in a protein core. Biochemistry 1992;31:4324–4333.
9. Xu J, Baase WA, Baldwin E, Matthews BW. The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. Protein Sci 1998;7:158–177.
10. Wilcock D, Pisabarro MT, Lopez-Hernandez E, Serrano L, Coll M.

11. Blaber M, Lindstrom JD, Gassner N, Xu J, Heinz DW, Matthews BW. 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 1993;32:11363–11373.
12. Dao-pin S, Anderson DE, Baase WA, Dahlquist FW, Matthews BW. Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. Biochemistry 1991;30:11521–11529.
13. Waldburger CD, Schildbach JF, Sauer RT. Are buried salt bridges important for protein stability and conformational specificity? Nat Struct Biol 1995;2:122–128.
14. Stites WE, Gittis AG, Lattman EE, Shortle D. In a staphylococcal nuclease mutant the side-chain of a lysine replacing valine 66 is fully buried in the hydrophobic core. J Mol Biol 1991;221:7–14.
15. Hendsch ZS, Tidor B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci 1994;3:211–26.
16. Shoichet BK, Baase WA, Kuroki R, Matthews BW. A relationship between protein stability and protein function. Proc Natl Acad Sci USA 1995;2:452–456.
17. Blundell T, Carney D, Gardner S, Hayes F, Howlin B, Hubbard T, Overington J, Singh DA, Sibanda BL, Sutcliffe M. 18th Sir Hans Krebs lecture. Knowledge-based protein modelling and design. Eur J Biochem 1988;172:513–520.
18. Sali A, Overington JP, Johnson MS, Blundell, TL. From comparisons of protein sequences and structures to protein modelling and design. Trends Biochem Sci 1990;15:235–240.
19. Srinivasan N, Blundell, TL. An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. Protein Eng 1993;6:501–512.
20. Johnson MS. Srinivasan N, Sowdhamini R, Blundell TL. Knowledge-based protein modelling. CRC Crit Rev Biochem Mol Biol 1994;29:1–68.
21. Sanchez R, Sali A. Advances in comparative protein-structure modelling. Curr Opin Struct Biol 1997;7:206–214.
22. Bates PA, Sternberg MJ. Model building by comparison at CASP3: using expert knowledge and computer automation. Proteins 1999; 37(S3):47–54.
23. Balaji S, Sujatha S, Kumar SSC, Srinivasan N. PALI- a database of Phylogeny and ALIgnment of homologous protein structures. Nucleic Acids Res 2001;29:61–65.
24. Gowri VS, Pandit SB, Karthik PS, Srinivasan N, Balaji S. Integration of related sequences with protein three dimensional structural families in an updated version of PALI database. Nucleic Acids Res 2003;31:486–488.
25. Russell RB, Barton GJ. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. Proteins 1992;14:309–323.
26. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
27. Overington J, Johnson MS, Sali A, Blundell TL. Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction. Proc R Soc London 1990;B241:132–145.
28. Chakravarty S, Varadarajan R. Residue depth: a novel parameter for the analysis of protein structure and stability. Structure 1999;7:723–732.
29. Kajander T, Kahn PC, Passila SH, Cohen DC, Lehtio L, Adolfsen W, Warwicker J, Schell U, Goldman A. Buried charged surface in proteins. Structure Fold Des 2000;8:1203–1214.
30. 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.
31. Chen JC, Miercke LJ, Krucinski J, Starr JR, Saenz G, Wang X, Spilburg CA, Lange LG, Ellsworth JL, Stroud RM. Biochemistry 1998;37:5107–5117
32. Evans SV. SETOR: hardware lighted three-dimensional solid model representatitons of macromolecules. J Mol Graph 1993;11:134–138.