The Effect of Sequence Evolution on Protein Structural Divergence

Simon G. Williams and Simon C. Lovell
Faculty of Life Sciences, University of Manchester, Manchester, UK

The complex constraints imposed by protein structure and function result in varied rates of sequence and structural divergence in proteins. Analysis of sequence differences between homologous proteins can advance our understanding of structural divergence and some of the constraints that govern the evolution of these molecules. Here, we assess the relationship between amino acid sequence and structural divergence. Firstly, we demonstrate that the relationship between protein sequence and structural divergence is governed by a variety of evolutionary constraints, including solvent exposure and secondary structure. Secondly, although compensatory substitutions are widespread, we find many radical size-changing mutations that are not compensated by neighboring complementary changes. Instead, these noncompensated substitutions are mitigated by alteration of protein structure. These results suggest a combined mechanism of accommodating substitutions in proteins, involving both coevolution and structural accommodation. Such a mechanism can explain previously observed correlated substitutions of residues that are distant both in sequence and structure, allowing an integrated view of sequence and structural divergence of proteins.

Introduction

Subsequent to duplication or speciation events, homologous proteins accumulate sequence substitutions, resulting in diverged families of related proteins. In any homologous protein family, sequence divergence can proceed to the extent that shared evolutionary history can barely be detected (Orengo and Thornton 2005). This includes not only residues on the protein’s surface that make few interactions, but also those that are in the core of the protein, and so under substantial evolutionary constraint (Overington et al. 1990, 1992), and those directly in functional and binding sites (Todd et al. 2001). Despite the possibility of radical substitutions, protein structure, stability, and function must be maintained (Blundell and Wood 1975; DePristo et al. 2005). However, the mechanisms by which the apparently radical substitutions in highly evolutionarily constrained regions of proteins can be accommodated remain unclear.

It has long been known that structural divergence in the core of proteins correlates with sequence divergence (Lesk and Chothia 1980) and are related by an exponential function (Chothia and Lesk 1986). This exponential relationship is due to the varying mutation rates of buried and accessible residues. In closely related proteins, the majority of sequence differences are in their peripheral, solvent exposed residues, whereas more distantly related proteins will additionally have mutations in their buried residues resulting in more structural change. Such an explanation is in accordance with the observation that residues buried in the protein core are more conserved both in terms of sequence and structure than those that are solvent accessible (Hubbard and Blundell 1987). These studies have been repeated with larger data sets (Flores et al. 1993; Russell et al. 1997) with broadly similar conclusions.

Insertions and deletions also have a dramatic effect on structural divergence, although major conformational characteristics of proteins can be well conserved even at low sequence identities (<20%) (Flores et al. 1993). This would indicate that using sequence identity within the twilight zone (20–30% sequence identity) could be fairly reliable in terms of structural identification during homology modeling if the effect of substitutions on the protein structure is well understood.

Regions of proteins with strong evolutionary constraint can potentially have those constraints altered by mutations elsewhere in the structure. Changing the amino acid at one position can act to modify the “landscape” of evolutionary constraint at neighboring positions. This may result in the sites evolving in a codependent manner, that is, site-specific coevolution (Fitch and Markowitz 1970). Many cases of identified coevolving sites in proteins are found to have side chains that interact (Altschuh et al. 1987, 1988; Gobel et al. 1994). Typical changes that could alter the constraints at neighboring site are radical size changes. For example, a substitution from a large residue to a small one can introduce destabilizing cavities (Xu et al. 1998; Cuff and Martin 2004), and so increase the chances of a compensating “small-to-large” substitution at a neighboring site becoming fixed in the population. It has also been proposed that substitutions increase the thermodynamic stability of the molecule may allow greater tolerance to destabilizing substitutions elsewhere (Bloom et al. 2005).

Correlated changes in multiple sequence alignments have been identified by a variety of methods (Altschuh et al. 1987, 1988; Korber et al. 1993; Gobel et al. 1994; Clarke 1995; Pollock et al. 1999; Atchley et al. 2000; Pritchard and Dufton 2000; Pritchard et al. 2001; Suel et al. 2003; Choi et al. 2005, 2006; Gloor et al. 2005; Wang and Pollock 2005; Fares and Travers 2006; Wang and Pollock 2007). Many identified coevolving amino acids are coupled to one or two other positions in the molecule with interacting side chains (Altschuh et al. 1987; Clarke 1995; Atchley et al. 2000; Gloor et al. 2005; Yeang and Haussler 2007). In some protein families, chains of coevolving residues have been identified (Atchley et al. 2000; Socolich et al. 2005; Travers et al. 2007; Wang and Pollock 2007). However, other studies have identified coevolving sites in spatially distant locations (Gobel et al. 1994; Clarke 1995; Gloor et al. 2005; Fares and Travers 2006). Such substitutions at one site can seemingly alter evolutionary constraints at distant sites with no direct interaction and no intervening changes. It has been suggested that these substitutions may act to increasing overall stability thus facilitating sequence change at another site (Bloom et al. 2005).
Here, we analyze both the effect of sequence substitution on protein structural evolution and the occurrence of compensating mutations in proteins. We find the relationship between structural change and sequence evolution is both fold specific and dependent on a protein's secondary structure. Surprisingly, we find that many substitutions between residues of differing size are not mitigated by compensating size-change mutations. Instead, we find small structural shifts in neighboring main-chain conformations. These shifts are site specific, and both their magnitude and direction can be rationalized. These results suggest that in order to understand evolution and selection in proteins, knowledge of coevolution and structural change must be integrated, and we suggest how this may be done.

Methods
Data Sets

Protein families were downloaded from the HOMSTRAD database (Mizuguchi et al. 1998). HOMSTRAD contains annotated sequences of homologous protein families, aligned on the basis of structure using MNYFIT (Sutcliffe et al. 1987) and COMPARER (Sali and Blundell 1990). The HOMSTRAD families are classified mainly on their secondary structure content and those used in this investigation come from the "all alpha," "all beta," or "alpha beta" groups. These categories were chosen as they cover the majority of globular folds; they contain different types of secondary structure and an appreciable buried core, allowing the distinction of solvent accessible (surface) or solvent inaccessible (buried) residues. Homologous proteins from 17 families were selected from these categories and can be seen in supplementary table 1, Supplementary Material online, were selected as being representative of the proteins in the HOMSTRAD database. The local structural environment of each amino acid residue (main-chain conformation and solvent accessibility) was calculated using the program JOY (Mizuguchi et al. 1998). The accessibility of each residue was categorized as solvent accessible, if 7% or more of its side-chain area was accessible to a water-sized (1.4 Å) probe, or buried if <7% of the surface area was accessible. This measure of solvent accessibility was calculated by Hubbard and Blundell (1987) as demonstrating the difference in sequence and structural conservation within these regions of proteins.

Local Structural Environment Differences in Structural Conservation

Amino acids were superimposed and their structural divergence compared with their sequence identities. Residues in β-sheet conformations were found in 47 pairs of proteins, α-helical conformations in 49 pairs of proteins, and the coil conformations in all 96 protein pairs. These protein families, highlighted in supplementary table 1, Supplementary Material online, were selected as being representative of the proteins in the HOMSTRAD database. The local structural environment of each amino acid residue (main-chain conformation and solvent accessibility) was calculated using the program JOY (Mizuguchi et al. 1998). The accessibility of each residue was categorized as solvent accessible, if 7% or more of its side-chain area was accessible to a water-sized (1.4 Å) probe, or buried if <7% of the surface area was accessible. This measure of solvent accessibility was calculated by Hubbard and Blundell (1987) as demonstrating the difference in sequence and structural conservation within these regions of proteins.

The Effect of Volume Changing Substitutions on Local Structure

Those α-helices and β-strands common between all 1,168 pairs of homologous proteins from all families (supplementary table 1, Supplementary Material online) were used to assess the effect on structure of amino acid substitutions where the side-chain volume of the residues are altered. A greater number of homologous families were used for this analysis to maximize the number of volume-changing substitutions with corresponding changes in the contacting residues so that relationships could be established. α-Helices were recognized if at least eight residues in succession were identified as helical and β-strands regions, tending to have fewer residues than helices, were recognized if at least six consecutive residues were identified as sheet. Although α-helices and β-strands can be found containing fewer amino acids, these numbers were chosen to ensure that we only included the larger secondary structure regions that form the core of the protein folds. These common regions within homologous pairs of proteins were then used to
assess the effect of volume change on the neighboring amino acids. Amino acid side-chain volumes were obtained from Harpaz et al. (1994) and were calculated from atoms buried in the interior of proteins using a Voronoi polyhedra method.

PROBE (Word, Lovell, LaBean, et al. 1999) was used to identify neighboring amino acids with atomic interactions to those residues of interest. PROBE uses the rolling probe algorithm (Connolly 1983) to identify regions of neighboring amino acids that can be considered to be interacting through van der Waals or hydrogen bonding. At each site, the neighboring residues considered to have these types of interactions between their amino acids were recorded. A number of measurements were made both at the site in question and in the contacting residues. Amino acid volume change, chi1 angle deviations (calculated using DANG; Word 2000), and distances to neighboring residues were assessed.

Angle Definition

When volume-changing substitutions occur, the orientation of the amino acid side chain in relation to the surrounding residues may affect where local structural divergence and potential compensatory substitutions arise. In order to determine whether this is the case, we calculated the angle of the substituted amino acids to the contacting residues when analyzing the effects of volume change. To calculate the angle of amino acid $i$ to the contacting residues, each side chain was represented as a vector beginning at the C$\beta$ atom and using the terminal nonhydrogen atom as the end point. Where there is more than one terminal nonhydrogen atom (e.g., Asp, Asn, Thr, Glu, Val, Gln, Ile, Leu, and Arg), the mean position of the two terminal atoms was used. Proline amino acid terminals were calculated as the mean between the C$\delta$ and C$\gamma$ atoms and for histidine the mean between the N$\varepsilon$2 and C$\varepsilon$2. No residues have three or more terminal nonhydrogen atoms. Glycine and alanine amino acids were not considered as it is not possible to determine a vector in this manner. A second vector was defined from the C$\beta$ of residue $i$ to the C$\alpha$ of contacting residues (as defined by PROBE). The angle between these vectors was determined (fig. 1) and used to assess where the side-chain volume alterations had most effect.

The Effect of Amino Acid Contact Density on Compensatory Substitutions

In order to identify the relationship between contact density and compensatory mutations, hydrogen atoms were added using the REDUCE software (Word, Lovell, Richardson SC, and Richardson DC 1999) and all-atom contacts calculated with PROBE (Word, Lovell, LaBean, et al. 1999). The contact density of a residue is defined as the number of points on that residue in contact with other noncovalently bonded atoms. Because amino acids vary considerably in surface area, this value is normalized by dividing by the total number of points that may potentially be in contact for a given residue type. The total number of potential points in contact was calculated by running PROBE on a set of amino acid side chains built with standard geometry (Engh and Huber 1991). All amino acid pairs within the 0–90° contact range were assessed.

Results

Sequence and Structural Divergence within Protein Families

Within each family, all members were superimposed in turn onto a representative member. Superposition started with the five pairs of equivalent C$\alpha$ atoms from residues that were most conserved in sequence throughout the family. Iteratively, the C$\alpha$ atom with the shortest Cartesian distance was added to the set of equivalences and the structures re-superimposed.

Figure 2 shows the structural divergence between proteins (measured in RMSD) as equivalent pairs are added to a superposition. Because each superposition begins with the five most conserved residue pairs from the sequence alignment (those with highest CRESCEndo score (Chelliah et al. 2004), these may not be the most structurally conserved positions. As a consequence the RMSD initially decreases as the most conserved positions are subsequently added. As more structurally diverged positions are added to the superposition, the overall RMSD increases until all equivalent amino acid pairs have been included and the final RMSD is calculated.

Figure 2A shows the members of the cytochrome C5 family as they are superimposed onto a structural homologue (pdb code 1CCH; Cai et al. 1992). The members have a range of sequence identities with respect to 1CCH. Superimposing in this manner indicates that percent sequence
identity does not necessarily reflect the overall structural divergence between the proteins. The protein 1CC5 (Carter et al. 1985) has the lowest sequence identity with 1CCH (25%), and the overall structural difference between the two proteins is greater than the other members. Two proteins (pdb codes 1AYG [Hasegawa et al. 1998] and 1A8C [Timkovich et al. 1998]) have ~56% sequence identity to the target protein but are seen to have different patterns of structural deviation as equivalent residues are added to the superposition resulting in large differences in their final RMSDs. 1A8C has a final RMSD of over 2.5 Å, whereas 1AYG has a final RMSD of ~1.5 Å, comparable with proteins sharing a much higher sequence identity.

Figure 2B indicates the superposition and structural divergence of proteins in the alpha beta hydrolase family. In this case, percent sequence identity is representative of the overall structural differences between the proteins. The curve of the slopes indicates that some proteins may show more structural similarity in some regions than others. This is demonstrated by 1CLE (Ghosh et al. 1995), 1TRH (Grochulski et al. 1994), and 1THG (Schrag and Czygan 1993) having greater structural conservation as the first 250 residues are superimposed than 2BCE. Members of the glycosyl hydrolase 18 family (fig. 2C) show a similar trend, with 1CNV (Hennig, Jansonius, et al. 1995) having greater structural conservation to the target protein 1EDT (Rao et al. 1995) than 1NAR (Hennig, Pfeffer-Henning, et al. 1995) through most of the superposition, but with these two proteins having similar overall structural divergence of 4–4.5 Å RMSD.

The shapes of the lines in figure 2 differ between proteins, demonstrating that an assessment of divergence based on a single figure like RMSD does not completely capture the relationship between sequence and structural divergence. It is also evident from these assessments that sequence identity, indicated by the position on the y-axis when all residues are included, cannot always be an accurate indicator of the level of overall structural divergence between two proteins. This is in accordance with the work of Wood and Pearson (1999) who demonstrated that the “mutational sensitivity” (i.e., the structural change resulting from sequence substitutions) can vary greatly between families. The differences in sequence between two proteins in terms of the amino acid types and the local environment of each substitution that occurs is likely to have a greater influence on structural divergence as a whole than simple sequence identity. Different regions of proteins are likely to maintain structural integrity, whereas others may be more amenable to structural divergence depending on their evolutionary constraints. Figure 3 demonstrates these differences by comparing the structural conservation of solvent accessible and buried regions as well as sheet, helix, and coil regions with sequence divergence.

Compensatory Mutations

When amino acid substitutions within a protein structure increase the volume of the side chain, the positions of the neighboring amino acids are likely to be affected due to extra atoms that need to be accommodated. The effect of these volume changes has been assessed in buried regions of the proteins, where residues are tightly packed and variations in volume would have the greatest impact (Liu et al. 2000). It has been suggested that the effect of amino acid volume change on neighboring residues may result in
neighboring amino acid pairs demonstrate compensation for volume change, whereas 70% are noncomplementary, either by both amino acids substituting to add extra strain on the local structure (25%) or by one amino acid in the pair remaining conserved (45%).

Contact residues in the range of 31–60° to the volume-changing side chain show the best correlation for compensatory substitutions although even this correlation is poor ($r = 0.15$, compared with $r = 0.09$ for 0–30° and $r = 9 \times 10^{-5}$ for 61–90°). This correlation is significantly different from the correlation in the 61–90° category ($P = 0.011$, Fisher’s $Z$-transformation). The limited data in the 0–30° category ($n = 140$ compared with $n = 473$ and $n = 879$ for 31–60 and 61–90° categories, respectively) result in no significant difference with this correlation and the others ($P = 0.601$ and $P = 301$ for comparisons with 31–60 and 61–90° groups, respectively). As the angle between substituted position contact residue increases to 61–90° no correlation is observed and hence no indication that compensatory substitutions would occur meaningfully at this range.

Analysis of amino acid contact density between contacting amino acid pairs in the 0–90° range and their neighboring residues reveal that there is a small difference between compensating pairs and pairs in which one residue remains conserved, although this difference is not significant at the 95% confidence level ($P = 0.077$, with an average of 10.5% and 10.7% of potential contacts made, respectively). It has been suggested that proteins with high contact density are more accepting of amino acid mutations (Bloom et al. 2006), and you might then expect there to be less requirement for compensatory substitutions in these regions. We do not observe any significant correlation on our data set; however, it is likely that this method of calculating contact density will be sensitive to the quality of the X-ray structures.

### Structural Adjustments in Response to Residue Substitutions

The sites in buried $\alpha$-helices and $\beta$-strands where increased volume change is compensated by complementary substitutions are as common as those where no compensation occurs. This implies that there may be other mechanisms in place that allow these substitutions to occur while causing minimal structural disruption. Adjustment of side-chain conformation is one possibility. However, we found that there was no significant difference in the chi-1 side-chain dihedral angles in compensated contacting pairs as compared with noncompensated contacting pairs. This is the case in both $\alpha$-helices and $\beta$-sheets ($\alpha$-helix $P < 0.81$ and $\beta$-sheet $P < 0.91$, $t$-test).

We find, however, in sites where there is volume-changing substitution, the local main-chain conformation adjusts to accommodate the additional atoms. Figure 5 shows how the angles between the side chains of the substituted amino acid and the Cz atom of the residues in van der Waals contact dictates the effect of volume change. When the angle is small, the effect that volume changes have on the distance that the contact residue shifts in relation to its homologue is greater. As the angle increases, the effect that volume change has on the movement directly contacting the initial site (Altschuh et al. 1987; Clarke 1995; Atchley et al. 2000; Gloor et al. 2005; Fares and Travers 2006; Yeang and Haussler 2007). The angle of the amino acid side chain to the contacting residue is likely to be important in determining where the increased volume causes the most structural disruption and therefore where compensatory substitutions are likely to occur.

Figure 4 demonstrates the volume change associated with the substitutions in buried $\alpha$-helix and $\beta$-strand residues and the corresponding volume changes in the contacting residues at a range of angles to the initial site side chain. The overall correlation of volume change and compensatory volume change in contact residues is poor. Where a substitution increases the volume of the side chain, there are many substitutions in the contact residues that also increase in volume. The angle of the side chain to the contacting residue Cz is shown to have a small influence on the trend of compensatory substitutions. Overall, 30% of

**Structural Divergence in Relation to Sequence Identity**

**FIG. 3.—Structural divergence in relation to sequence identity in pairs of superimposed homologous proteins.** (*A*) The relationship in different secondary structure regions. Equivalent helix, sheet, and coil regions between homologues were superimposed and their overall RMSD calculated and compared with the percent sequence identity between the respective regions in the two proteins. (*B*) The relationship in regions of differing solvent accessibility. Equivalent buried and solvent accessible amino acid regions were superimposed and their overall RMSD compared with percent sequence identity.
of contacting residues decreases, and at 75–90°, there is very little correlation.

The volume change associated with amino acid substitutions in α-helices and β-strands can cause the surrounding residues to shift in order to accommodate them. The degree of movement between the volume-changing positions and contacting residues is shown in figure 6. At each site in the homologues where the volume changes between equivalent positions, the distance to equivalent contacting residues was measured. The mean of the distance changes are taken at each volume change in buried and solvent accessible positions to make the relationship clearer. In the α-helices (fig. 6A), volume change in the buried sites has a greater effect on the movement of the contacting residues than in solvent accessible sites. The local structural movement caused by volume change in β-strands (fig. 6B) shows that buried and solvent accessible sites have a similar impact on the movement of the contacting residues.
Discussion

Analysis of the relationship between protein sequence and structure divergence demonstrates that structural divergence is influenced by both the type of secondary structure and degree to which residues are buried from solvent through the varying evolutionary constraints imposed at certain positions (Overington et al. 1992).

On average, substitutions in $\beta$-sheets result in less structural divergence of that position than substitutions in $\alpha$-helices (fig. 3A). This is likely to be due to differences in hydrogen bonding between the two main-chain conformations. The hydrogen bonding in an $\alpha$-helix is mainly confined to residues that are local within the sequence. This bonding links the $>\text{N}^{-}\text{H}$ group of an amino acid to the $>\text{C} = \text{O}$ group of the fourth amino acid away from it, maintaining the helical structure. In contrast, the hydrogen bonding in $\beta$-sheets can form a network of interactions with the surrounding parallel strands that are nonlocal in sequence. Mutations within a $\beta$-sheet that could lead to structural

![Fig. 5.—The change in volume of substituted amino acid side chains and the corresponding change in distance to equivalent neighboring amino acids in pairs of homologous proteins. A small-to-large substitution would be represented as positive on the $x$-axis and is shown to correspond to an increase in distance to the equivalent contacting residue (positive on the $y$-axis). The contacting pairs are separated, based upon the angle of the contacting residue to the original substituted side chain, into 0–15°, 16–30°, 31–45°, 46–60°, 61–75°, and 76–90° categories demonstrating the importance of side-chain orientation on the surrounding structural divergence.](https://academic.oup.com/mbe/article-abstract/26/5/1055/1036784)
divergence are more likely to be compensated by the surrounding hydrogen bonds that maintain the rigid conformation whereas mutations in α-helices could potentially result in more structural divergence as the whole secondary structural unit can shift relative to the rest of the protein. Coil regions generally make up the periphery of proteins and have fewer constraints placed upon them allowing greater flexibility over the type of substitutions that can occur and the extent to which structural deviation will be tolerated. Unsurprisingly, substitutions at sites with amino acids buried in the protein structure are also likely to be more resistant to structural divergence than those in solvent accessible regions (fig. 3B). Presumably, this is due to the requirement to maintain intricate packing interactions in the protein core (Word, Lovell, LaBean, et al. 1999).

The negative structural effects of a size-changing mutation in the tightly packed core of a protein may be mitigated by a compensating mutation at a neighboring site. Such an evolutionary coupling will result in intramolecular coevolution. Coevolution of this type has been observed between amino acids with directly interacting side chains (Altschuh et al. 1987; Clarke 1995; Atchley et al. 2000; Gloor et al. 2005; Yeang and Haussler 2007). This would suggest that these coupled sites are acting to maintain the local structural stability of the proteins.

We may therefore expect that substitutions altering amino acid volume would be surrounded by complementary volume changes. However, we find that volume compensation is found in only a subset of volume-change mutations. Moreover, there is no correlation between volume change at a site and the size of a complementary volume change at a neighboring site. This is the case regardless of the spatial orientation of the residue pairs investigated. In early work on homologous globin structures, Lim and Ptitsyn (1970) observed the hydrophobic core has a largely constant volume, despite amino acid substitutions. This has been confirmed with larger data sets (Gerstein et al. 1994). However, the randomly generated sequences, which by definition do not have volume-compensating mutations, also display largely constant volumes when viewed as a set. Our results, in accordance with those previously published (Gerstein et al. 1994) suggest that size-compensating mutations may not be important in many positions in a range of
protein folds. By contrast, we find that structural rearrange-
ments are common, and correlate with occurrence of vol-
ume-change mutations. These structural rearrangements
are larger when more directly in line with the side chain
of the substituted residue.

We do not suggest that intramolecular coevolution
does not occur; several previous studies have identified cor-
related changes in the sequence evolution of proteins
(Altschuh et al. 1987, 1988; Korber et al. 1993; Gobel
et al. 1994; Clarke 1995; Pollock et al. 1999; Atchley
et al. 2000; Pritchard and Dufton 2000; Suel et al.
2003; Choi et al. 2005, 2006; Gloor et al. 2005; Wang
and Pollock 2005, 2007; Fares and Travers 2006), and this
signal is robust to a wide range of methods for detecting it.
We do find, however, that for a large proportion of buried
residues, volume-change mutations are accommodated by
compensating substitutions, by structural rearrange-
ment, or by a combination of the two. The structural plasticity
associated with specific positions could explain coevolu-
tion of spatially distant sites. Previous studies have iden-
tified amino acid pairs located in spatially distant sites as
coevolving (Gobel et al. 1994; Clarke 1995; Gloor et al.
2005; Fares and Travers 2006). If compensating mutation
were the only available mechanism, this would be some-
what puzzling.

We propose a combined model of coevolution and
structural divergence in response to substitution. A substitu-
tion at a given site may result in compensating mutations at
neighboring sites, allowing the maintenance of structural in-
tegrity. Alternatively, depending on the local structural en-
vironment, structural rearrangement may be possible. This
rearrangement will be larger in line with the direction of
the side chain, and proportional to the volume change asso-
ciated with the substitution (fig. 6). We may also imagine that
such changes can lead to knock-on effects such as further
structural change. Alternatively, the result may be compen-
satory change in a distant site where structural divergence
cannot be tolerated due to the local structural environment.
The relationship between sequence and structure divergence
over spatially distant positions in a protein has also been rec-
ognized by Sinha and Nussinov (2001) who found that dis-
tributed point mutations often result in structural divergence
of common regions in homologues.

Although substitutions that alter amino acid volume
do not invoke compensatory substitutions at every site in
these homologous proteins, there are other properties of
amino acids that may be compensated for in protein evolu-
tion. Side-chain charge (Fukami-Kobayashi et al. 2002;
Choi et al. 2005) and shape complementarity are important
for maintaining protein structures and, as such, offer adi-
tional constraints over the sequence change at certain sites.
It is also possible that any compensatory response could be
distributed among multiple surrounding sites and not
limited to individual contacting residues.

Unless there is redundancy arising from gene duplica-
tion, protein structure and function must be maintained
within a protein as it accommodates substitutions. Our com-
bined proposed mechanism of compensating mutations and
structural rearrangements not only explains coevolution of
distant amino acids, but also suggests an additional mecha-
nism that allows proteins to be robust toward substitutions.

Supplementary Material

Supplementary table 1 is available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/)

Acknowledgments

We thank Mario Fares and David Robertson for sci-
entific discussion and the EPSRC for funding in the form of
a studentship for S.G.W.

Literature Cited


Arndt von Haeseler, Associate Editor

Accepted January 26, 2009