PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Response of Protein Structures to Amino-Acid Sequence Changes

A. M. Lesk and C. H. Chothia

Phil. Trans. R. Soc. Lond. A 1986 **317**, 345-356
doi: 10.1098/rsta.1986.0044

Phil. Trans. R. Soc. Lond. A **317**, 345–356 (1986)
Printed in Great Britain

The response of protein structures to amino-acid sequence changes

By A. M. Lesk[1]† and C. H. Chothia[1,2]

[1] *M.R.C. Laboratory of Molecular Biology, Cambridge CB2 2QH, U.K.*
[2] *Christopher Ingold Laboratory, University College London, 20 Gordon Street, London WC1H 0AJ, U.K.*

The general response of protein structures to mutations, insertions and deletions is conformational change.

Comparisons of related proteins show that a large part of the polypeptide chain retains its basic folding pattern. This 'core' of the structures comprises major elements of secondary structure and residues flanking them, including active-site peptides. The core may amount to as little as 40% of the structure for distantly related proteins, but is 90% or more for proteins with amino-acid sequence homologies of 50% or more.

There is a direct relation between the root mean square deviation of the main-chain atoms of the core residues of a pair of proteins and the overall amino-acid sequence homology. The deviation primarily reflects the shifts and rotations of packed secondary structures with respect to one another. For distantly related proteins, shifts of 3–5 Å (1 Å = 10^{-1} nm = 10^{-10} m) are typical and shifts of up to 7 Å have been observed. For proteins with sequence homologies of 50% or more the shifts are much smaller, lying in the range 0.3–1.5 Å. Such closely related proteins are also characterized by a conservation of over 85% of the conformational angles of the backbone and of the side chains of unmutated residues.

These observations suggest that successful model building of an unknown protein structure depends on knowing the structure of a reasonably close relative. As an application of these results we propose a model for the V_L and V_H domains of the antilysozyme antibody D1.3, the crystal structure determination of which is in progress.

Introduction

In a series of investigations over the past several years we have compared the structures of related proteins. The results present a consistent picture of the nature of the conformational changes produced by mutations, insertions and deletions in the amino-acid sequence.

In this paper we begin with a description of residue–residue interactions in proteins. It is because most residues make multiple contacts with other residues that changes in the amino-acid sequence of a protein produce changes in the structure. Next we describe the nature of the structural changes observed in comparing pairs of related proteins. By extracting the 'core' structure which retains the basic fold, we can derive a quantitative relation between the divergence of the three-dimensional structure and amino-acid sequence homology.

We describe in outline the structural changes that can occur within the core, showing that individual α-helices and β-sheets retain their structure, and that mutations at interfaces between them change their relative positions and orientations in their spatial assembly. We next show how the magnitude of these structural changes increases as the amino-acid sequence homology

† Permanent address: Fairleigh Dickinson University, Teaneck-Hackensack Campus, Teaneck, New Jersey 07666, U.S.A.

decreases. Pairs of proteins having 50% homology or higher are characterized by internal shifts of no more than 1.5 Å,† and a retention of 85% of the conformational angles of the backbone and unmutated sidechains to within 30°.

We discuss the implications of these observations for model building of related structures, by using the concept of the core to suggest how accurate models can be over different ranges of sequence homology. We present a model for the antilysozyme antibody IgG D1.3.

Residue–residue interactions in proteins

The stability and uniqueness of a native protein structure arises from the set of interactions among its residues. There are several kinds of chemical forces involved – including van der Waals interactions, hydrogen bonds and hydrophobicity – but the net result is a constellation of very intimate interactions, with the native fold determined by the compatibility within a single structure of all the individual residue–residue interactions. It is the intimacy and specificity of the interactions that requires the conformation to change when the amino-acid sequence changes.

It is well known that protein interiors are close-packed: the mean volumes of residues in the interiors of proteins are as high as in amino-acid crystals, despite the stereochemical constraints of the backbone (Klapper 1971; Richards 1974; Chothia 1975). Cavities as large as water molecules occur very rarely. Close packing contributes to protein stability by maximizing van der Waals attraction. In addition, nearly all polar groups buried within a protein form hydrogen bonds (Baker & Hubbard 1984).

A mutation changing a side chain will alter the interaction of a residue with its neighbours. To get an idea of how many residues would be directly affected by a mutation, we calculated the distribution of side chain–side chain contacts in six proteins.

We selected three small proteins (sperm whale myoglobin (Phillips 1980), poplar leaf plastocyanin (Guss & Freeman 1983), and human lysozyme (Artymiuk & Blake 1981), 99–153 residues) and three larger proteins (carboxypeptidase *a* (Rees *et al.* 1983), thermolysin (Holmes & Matthews 1982) and penicillopepsin (James & Sielecki 1983), 306–323 residues), all structures determined at high resolution (1.4–1.7 Å). For each residue we counted the number of other side chains with which its own side chain (including the C-α) is in van der Waals contact. The results (figure 1) show that very few side chains are free of contacts, even among those on the protein surface.

In the small proteins 83% of side chains are in contact with at least one other residue; for large proteins, 88% make at least one contact, reflecting the reduced surface to volume ratio. Approximately half of the side chains making no contacts are glycines or alanines. For some of these, certainly, mutation to a larger side chain would create new contacts that the structure would have to accommodate.

Thus, for almost all residues, a mutation will alter the interactions between the side chain and residues with which it is in contact, and produce *some* structural change.

† 1 Å = 10^{-1} nm = 10^{-10} m.

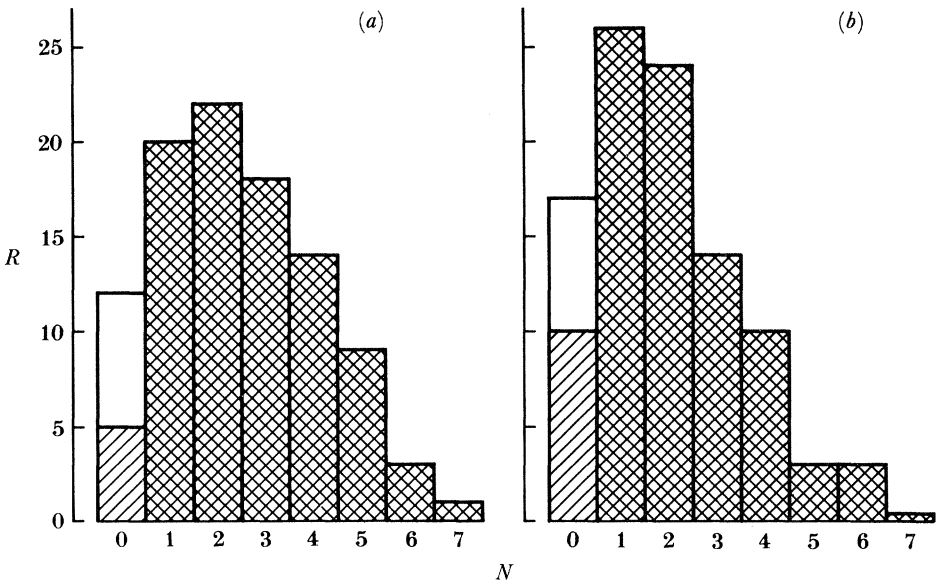

FIGURE 1. Percentage of residues, R, whose side chains are in contact with N other side chains (see text). For residues making zero contacts, Gly and Ala are indicated by ▨ and larger side chains by □. (*a*) Large proteins. (*b*) Small proteins.

THE RELATION BETWEEN DIVERGENCE OF AMINO-ACID SEQUENCE AND DIVERGENCE OF STRUCTURE

Homologous proteins share a common folding pattern over much of the polypeptide chain. Certain regions, however, change conformation entirely; these regions usually consist of peripheral elements of secondary structure, the loops between major elements of secondary structure, the ends of helices, or strands on the edges of sheets. Insertions and deletions are almost exclusively confined to such parts of the structure, and, together with mutations, contribute to the conformational changes.

A quantitative comparison of homologous structures shows that the major elements of secondary structure individually retain their conformation. We measure differences in conformation in terms of Δ, the root mean square (r.m.s.) deviation of atomic positions after an optimal superposition. The main-chain atoms of individual homologous helices in distantly related proteins 'fit' to $\Delta \approx 0.4$ Å or less. As we shall see, an important component of the structural divergence arises from changes in the way these homologous regions are assembled in space. By contrast, individual loops can so radically change their conformation that they cannot sensibly be superposed; indeed, if extensive insertion or deletion has occurred it may even be impossible to align the sequence, in such regions.

Thus, to describe the effect of amino-acid sequence changes quantitatively, it is essential to distinguish and treat separately those parts of a set of homologous proteins in which the fold is similar and those in which it is different. In comparing each pair of structures, we extracted a 'core' of individually well fitting segments by the following procedure (Chothia & Lesk 1986).

(1) First, the main-chain atoms (N, C-α, C, O) of homologous major elements of secondary structure – helices, or two adjacent strands of β-sheet – were individually superposed.

(2) To *extend* such a superposition is to recompute the superposition with the inclusion of

additional atoms along the chain at either end. The fits of step (1) were extended as long as the deviations in position of atoms in the last residue included were no greater than 3 Å.

(3) All these well fitting segments were taken together to form what we call the 'core' structure of the homologous pair.

The core comprises the main secondary structural elements and some residues flanking them, a set which generally includes all the active-site peptides. For pairs of proteins of high amino-acid sequence homology (more than 50% residue identity) the core includes over 90% of the residues. For proteins of low homology (less than 20% residue identity) the core can drop in size to less than half the residues.

The superpositions of core structures give useful information about the relation between divergence of amino-acid sequence and divergence of structure. For 32 pairs of homologous protein structures, we determined the core and calculated the r.m.s. deviation of the main-chain atoms. These values fall into the range 0.62–2.31 Å. The 32 pairs of structures represent eight protein families (identified in the caption to figure 2). All structures used were determined at a resolution of 1.5–2.0 Å.

How much of the observed structural differences is genuinely produced by changes in the amino-acid sequences, and how much arises from the effects of differences in environment (for example, solvent conditions or crystal packing) and experimental error? We can estimate the

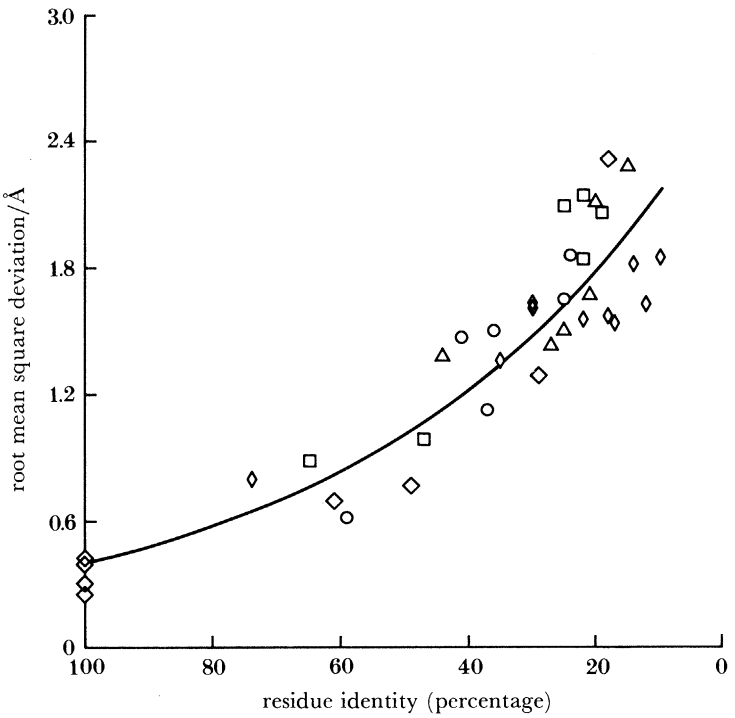

FIGURE 2. The structural divergence of the cores of homologous proteins plotted against sequence divergence. The structural divergence is measured by the root mean square deviation of main-chain atomic positions after the cores have been optimally superposed. The sequence homology is measured by the percentage of residues that are identical in the cores. Eight families of proteins are represented in this figure; four families by more than one comparison: globins (△), cytochromes *c* (○), serine proteases (□) and immunoglobulin domains (◊); and four by single comparisons: lysozymes, papain-actinidin, dihydrofolate reductases and plastocyanin-azurin (◇). The four points at 100% residue identity are for proteins whose structure has been determined in more than one crystal environment (see text). The curve is a least-squares fit to the data (see text).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

magnitude of these effects from proteins for which structures have been determined in different crystal forms, or for which there are two crystallographically independent copies in the asymmetric unit of one crystal form. For five such examples structures have been determined and refined to high resolution: bovine pancreatic trypsin inhibitor (two crystal forms) (Wlodawer *et al.* 1984), and *Alcaligenes denitrificans* azurin (Norris *et al.* 1983), tuna cytochrome *c* (Takano & Dickerson 1981), rat protease (Anderson *et al.* 1978), and deoxyhaemoglobin (Fermi *et al.* 1984) (two molecules per asymmetric unit of one crystal form). Superpositions of the main-chain atoms of the cores of these pairs of structures gave values in the range 0.25–0.40 Å; the average of the values is 0.33 Å. The deviations of non-identical pairs of proteins are significantly higher than this.

Figure 2 shows the variation in the r.m.s. deviation of the main-chain atoms of the core, Δ, with the overall amino-acid sequence homology, expressed in terms of H, the fraction of unconserved residues in the core. The data are fitted by the following function, which is plotted in figure 2:

$$\Delta = 0.40\ e^{1.87H}.$$

Structural differences in protein families

The nature of the structural changes that occur during evolution have been described in detail for five protein families: the globins (Lesk & Chothia 1980), the cytochromes *c* (Chothia & Lesk 1984), the immunoglobulin domains (Lesk & Chothia 1982), plastocyanin and azurin (Chothia & Lesk 1982), and the serine proteases (Lesk *et al.* 1986). These families contain members that have very low sequence homology (less than 20% amino acid identity). The studies of these families, and other unpublished analyses, create a consistent picture of the effects of amino-acid sequence changes on the structure of the helices and sheets, and the changes in the way they are assembled.

Our previous investigations treated families containing distantly related structures, as these allow the structural changes arising from mutations, insertions and deletions to be clearly determined. To complete the analysis of structural change over the entire range of homology covered in figure 2, we have also analysed five cases of more closely related proteins (amino-acid sequence homology more than 49%). We may anticipate the conclusions of the analysis of these systems by suggesting that whereas the comparison of distantly related structures shows how large the changes can be, remaining consistent with a similar fold and function, the comparison of closely related structures shows how well the structures are conserved when sequence divergence is low.

(*a*) *Distantly related proteins*

The core structures of pairs of distantly related proteins (amino-acid sequence homology less than 20%) consist, almost exclusively, of certain major elements of secondary structure and of peptides associated with the active site. The residues buried within these structures are at the interfaces between packed helices and sheets. These residues retain their hydrophobic character, but are otherwise fairly free to mutate. For distantly related proteins the homology of the buried core residues is similar to that of the residues on the protein surface.

The structural differences between the cores of homologous proteins arise primarily from

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

changes in the side chain volume of the hydrophobic residues buried in the core. The mean variation in size of such residues in distantly related proteins is *ca.* 60 Å^3. This is exemplified by a change of an alanine to a valine or leucine.

Mutations that change the size of residues at helix interfaces produce changes in the packing at the interface, and cause shifts and rotations of the helices with respect to each other. The main-chain atoms of helices tend to move as rigid bodies during conformational changes. This retention of structure of individual helices is measured by the r.m.s. differences (Δ) in atomic position of the main-chain atoms of homologous helices; typical values are 0.2–0.4 Å. Local conformational changes can occur in one to three residues at the ends of helices.

Thus protein structures accommodate mutations that change the size of residues at helix interfaces by moving the helices relative to each other. The maximum difference in relative geometry of packed helices is a relative shift of 7 Å and a rotation of 30°. Typical values, for proteins of 20% homology, are shifts of 3–5 Å.

Mutations of residues at interfaces between β-sheets also produce shifts and rotations. In this case, shifts and rotations of up to 3.5 Å and 30° have been observed. Because sheets are intrinsically more flexible than helices, they can also respond to mutation by local conformational change. Homologous β-sheets in core structures have r.m.s. differences in the positions of main-chain atoms of *ca.* 1.5 Å (3–4 times the value for helices).

The β-sheets show other mechanisms for accommodating mutations. Non-polar side chains from the loops joining strands of sheet can infiltrate laterally into the region between two layers of sheet, to fill space created by mutations that reduce the sizes of residues at the interface. The formation of β-bulges can accommodate insertions in the strands at the edges of sheets.

The very large structural changes that occur during protein evolution have taken place while maintaining function. We have observed two kinds of mechanisms used by families of proteins to retain a functional active site. In the globins, the changes in relative geometry of different pairs of packed helices are *coupled* to conserve the relative geometry of the residues that form the haem pocket (Lesk & Chothia 1980). In the cytochromes *c*, there has been a reconstruction of part of the active site to accommodate mutations, insertions and deletions (Almassy & Dickerson 1978; Chothia & Lesk 1984). In all proteins studied, there has been an integration of the response to mutation over all or a large portion of the molecule.

(*b*) *Proteins of high homology*

There are five pairs of published structures of closely related proteins, refined at high resolution (1.4–1.7 Å; see table 1). The errors in coordinates are *ca.* 0.15 Å. Each of these pairs of structures shows the following points of similarity.

(1) *The core structure of each homologous pair is very large*: *it contains between* 95 *and* 100% *of the residues*, *including much of the loop regions as well as the secondary structural elements*. This means that the number of loop regions which have changed their conformation is small. Partly responsible for this observation is the fact that the amino-acid sequences of homologous pairs differ very little in length. There have been few insertions and deletions in the loops between secondary structural elements. Rice embryo cytochrome *c* has an N-terminal extension of eight residues. These five pairs of sequences are consistent with the insertion or deletion of a maximum of 14 residues in positions not at the end of the chains. The immunoglobulin domains V_LRHE and V_LKOL show no insertions or deletions in the first 107 residues.

The few residues outside the core are in loops and most of them are adjacent to insertions or deletions.

Table 1. Proteins that have high sequence homologies and whose structures have been determined at high resolution

proteins	sequence identity (percentage)	reference for structure analysis
V_L:RHE KOL	74	Furey *et al.* (1983) Marquart *et al.* (1980)
Streptomyces griseus protease {A B}	65	Sielecki *et al.* (1979) Read *et al.* (1983)
lysozymes {human hen egg}	61	Artymiuk & Blake (1981) Grace (1979)
cytochromes *c* {rice tuna}	59	Ochi *et al.* (1983) Takano & Dickerson (1981)
papain actinidin	49	Kamphuis *et al.* (1985) Baker (1980)

(2) *The homologous elements of secondary structure have very similar conformations.* The main-chain regions of fifteen of the seventeen pairs of homologous helices fit with r.m.s. deviations in atomic position of 0.15–0.27 Å. In some cases, one or two residues at the ends of the helices were 'trimmed' because of small local conformational changes. Two of the helices fit with r.m.s. deviations of 0.33 and 0.49 Å; the latter is the small, irregular and peripheral helix in papain and actinidin.

The r.m.s. difference of atomic position of the main-chain atoms of the β-sheet in papain and actinidin was $\Delta = 0.34$ Å. The value for the smaller β-sheet in the V_L domains of RHE and KOL was $\Delta = 0.25$ Å. For the larger β-sheet in V_LRHE and V_LKOL, a somewhat poorer fit, $\Delta = 0.46$ Å, is produced by small, local, conformational changes in the edge strands and at the ends of strands.

The bacterial serine proteases SGPA and SGPB contain four β-sheets. The individual β-sheets of domain 1 fit with $\Delta = 0.40$ and 0.48 Å. Those in domain 2 fit with $\Delta = 0.22$ and 0.24 Å.

(3) *Homologous pairs of secondary structures that pack against each other show significant differences in their relative positions and orientations, although these effects are smaller than those observed in distantly related pairs of structures.* Helix–helix packings show relative shifts and rotations in the range 0.3 Å–2° to 1.1 Å–13°; for helix-sheet packings the values are between 0.3 Å–1° and 1.3 Å–12°; for the β-sheet–β-sheet packing in the immunoglobulin V_L domains they are 0.4 Å–1°, and for the β-sheet–β-sheet packing in domain 1 of the bacterial serine proteases they are 0.5 Å–2°.

(4) *The main-chain atoms of the core structures fit with r.m.s. deviations in the range from* 0.62 *to* 0.89 Å (see figure 2). This is approximately twice as large as the deviation of the main-chain atoms of the cores of pairs of proteins of 100% homology, but one-third of the typical values for pairs of proteins of 20% homology. This combination of large core (over 90% of the residues; see point (1) above) and relatively well fitting core, imply that the conformations of the backbones of these closely related structures are highly conserved.

We also analysed the conformations of the closely related proteins in terms of their conformational angles. We tabulate the changes in backbone conformational angles ψ and ϕ in table 2. Table 3 contains the differences in conformational angles χ_1 and χ_2 for residues that are conserved between the homologous structures, and the differences in χ_1 for residues that are mutated between the two structures.

TABLE 2. DIFFERENCES IN THE ϕ AND ψ VALUES OF HOMOLOGOUS RESIDUES

range of differences in torsion angles (degrees)	proportion of residues in each range[a] ϕ	ψ
0–20	0.86	0.86
20–40	0.08	0.08
40–60	0.02	0.02
60–100	0.00	0.02
100–180	0.02	0.02

[a] These proportions are for residues in the five pairs of homologous structures listed in table 1. The results for the individual pairs are all close to the average values given here.

TABLE 3. DIFFERENCES IN THE χ VALUES OF HOMOLOGOUS RESIDUES

range of torsion angle differences (degrees)	proportion of residues in each range: conserved residues[a] χ_1	χ_2	mutated residues[a] χ_1	residue in proteins with identical sequences[b] χ_1
0–20	0.82	0.73	0.52	0.86
20–40	0.05	0.14	0.07	0.08
40–60	0.05	0.08	0.04	0.04
60–80			0.04	
80–100			0.05	
100–120	0.09	0.05	0.11	0.03
120–140			0.10	
140–160			0.04	
160–180			0.02	

[a] Results for residues in the five pairs of homologous structures listed in table 2.
[b] Results for residues in five proteins whose structures have been determined in different crystals or different crystal environments (see text and figure 2).

Consistent with the large, well fitting core, over 90% of the backbone conformational angles ψ and ϕ change by less than 30°. Unmutated residues tend to conserve their side chain conformations as well: approximately 85% of the values of χ_1 and χ_2 of these residues change by less than 30°, and nearly 90% change by less than 60°. Kamphuis *et al.* (1985) report similar results from a comparison of papain and actinidin. (For the pairs of independent structure determinations of the same proteins, 95% of the residues had values of ψ and ϕ that changed by less than 30°, and 90% of the χ_1 values of side chains changed by less than 30°.) For mutated residues, 60% of the values of χ_1 change by less than 30°.

The significance of the distribution of conformational angle changes can be understood with reference to the overall distribution of conformational angles in proteins. Allowed regions of backbone angles are summarized by the familiar Ramachandran plot. Conservation of secondary structure constrains the backbone angles of these regions of the chain. (In some cases, coupled changes of ψ_{i-1} and ϕ_i rotate a whole peptide group without substantially affecting the course of the chain in its vicinity.) Side-chain angles in proteins tend to cluster in discrete sets of preferred conformations (Janin *et al.* 1978). These are similar to the preferred conformations of isolated units; interactions with the rest of the protein tend to select one of the minima and to sharpen it (Gelin & Karplus 1979). Thus the angles that vary in a range of $\pm 30°$ have undergone internal rotation *within* a local minimum. The rotation about the bond in question does not flip over into another local minimum.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The conservation of conformational angles of mutated side chains is not significantly greater than might be expected from the general statistics of side chain conformations in proteins (Janin *et al.* 1978). The conservation of side chain conformation in unmutated side chains is much higher. This retention of side chain conformation of unmutated residues can be understood in terms of the following model. A side chain occupies a cage defined by the backbone and side chains of its nearest neighbours. Nearly all side chains make contacts with one or more neighbours (figure 1). Because mutations usually occur one at a time during evolution, only a part of the cage is altered at each step. When the structure of the cage is largely conserved, the side chain will tend to occupy it in the same way.

Contacts with mutated residues, and shifts in the backbone, will tend to deform the structure of the cage. The highest conservation of structure will occur when the amount of mutation is very small – especially if there is only a single mutation in the molecule, as produced in protein engineering experiments. In this case, the neighbours of the mutated residues might so highly retain their conformation, and give such an accurate picture of the environment of the *mutated* side chain, that it is possible to predict its conformation (Shih *et al.* 1985).

Applications to model building

The development of rapid techniques for DNA sequencing, and the growth in the number of known protein structures, have generated considerable interest in the problem of predicting the structure of a protein of known sequence from a related protein of known structure. The results described here suggest that the degree of success expected in the prediction of related structures depends on the degree of sequence homology. As a rough guideline, we suggest that related structures can provide useful models of closely related proteins (homology greater than about 50%) but not of more distantly related ones.

Given the amino-acid sequence and structure of one protein and only the sequence of a related one, the basic steps in generating a model are as follows.

(1) Determine an alignment of the sequences. For closely related proteins the standard Needleman–Wunsch–Sellers methods give nearly correct results. For distantly related proteins, however, the correct determination of the alignment cannot always be achieved without structural information; in fact, not all regions of the sequence are necessarily homologous and alignable (Chothia & Lesk 1982; Dickerson 1980).

(2) For residues in the common portions of the backbone, i.e. those positions not affected by insertions and deletions, substitute for the mutated residues, and, as far as possible, set the conformations of mutated residues to the same conformation as the homologous residue in the known structure. Information is lacking in many situations in which the mutations substitute a larger residue for a smaller, and there is ambiguity when a branched side chain is replaced by an unbranched one.

Suppose we use this procedure to build a model of an unknown protein structure from a distantly related protein, with a homology of *ca.* 20%. The model will contain the following errors:

(1) between one-tenth and one-half of the protein – the portion outside the core – have radically changed conformation;

(2) mutations in the core will have caused shifts and rotations of elements of secondary structure, typically of 3–5 Å in magnitude;

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

(3) the effect of insertions and deletions is not accounted for.

No method exists at present which can predict these effects of the amino-acid sequence changes, and therefore there is no way of achieving a more accurate model of the target structure. However, we note that in some families the active site itself tends to conserve its geometry much more highly than the rest of the core, and in such cases it would be possible to interpret the effect on function of mutations in the active site.

By contrast, consider building a model of unknown protein structure from a closely related one, with a homology of 50% or greater. After applying the same procedure, we should expect that:

(1) over 90% of the structure will have a fold similar to that of the parent structure;

(2) the major secondary structural elements will individually have the correct structure and will be in the proper relative geometries to within 0.5–1.5 Å, and 80–90% of the conserved residues will have the proper side-chain conformations, with χ_1 and χ_2 correct to within $\pm 30°$; approximately 60% of the mutated side chains will have the correct conformational angle χ_1, to within 30°;

(3) there will be relatively few insertions and deletions to account for, and these will have local effects.

Even for closely related proteins, we know of no general method that we could apply with confidence to improve all these features. A skilled investigator could improve individual regions, on an *ad hoc* basis, by manually editing the structure.

By basing our model-building procedure on the analysis of closely related structures, we have been able to specify, *quantitatively and in advance of experimental test*, the quality of the model and the features that are likely to be correct.

A model for the V_L and V_H domains of the antilysozyme antibody IgG D1.3

We have applied the procedure described in the preceding section to build a model for the V_L–V_H dimer of the immunoglobulin IgG D1.3. The amino-acid sequences of these domains was determined by Martine Verhoeyen of the M.R.C. Laboratory of Molecular Biology. The crystal structure analysis of this molecule, in complex with lysozyme, is currently being done in the laboratories of Roberto Poljak at the Institute Pasteur and Simon E. V. Phillips at Leeds (Amit *et al.* 1985).

It may be remarked that the variable domains of immunoglobulins are inauspicious candidates for model building. A model of a globin (for example) in which the helices were accurately placed but the loops predicted less accurately would be deemed by most people a satisfactory result, because the loops are peripheral to function as well as structure. But the antigen-binding site of immunoglobulins is formed from six loops connecting the strands of β-sheet in the V_L and V_H domains.

When we received the sequences of these domains, we were fortunate to find close relatives in the immunoglobulin domains whose atomic structures have been determined by Epp *et al.* (1974), Marquart *et al.* (1980) and by Segal *et al.* (1974). The V_L domain of IgG D1.3 has a 67% sequence homology to V_LREI. The V_H domain of IgG D1.3 has 50% homology to the V_H domain of IgG KOL and 48% homology to the V_H domain of MCPC603. By using the procedure described in the previous section we prepared models of the two variable domains

of IgG D1.3. It is known that part of the V_L–V_H domain interface is conserved in sequence and structure. The conservation in sequence extends to IgG D1.3, and this provided a way to replace V_L domain of KOL with the model of the V_L domain of D1.3, to give a model for the V_L–V_H complex.

The antigen-binding loops of antibody-variable domains are so important that they require a separate investigation of the determinants of their structure. From an analysis of solved antibody structures, we determined how the conformations of these loops depend on:

(1) mutations within individual loops;
(2) insertions and deletions within individual loops;
(3) mutations in the framework regions, to which the loop is attached;
(4) interactions between different loops;

(C. Chothia & A. M. Lesk, unpublished results).

The details of this model will be tested against the experimental structure when it becomes available.

We thank John Creswell for the figure drawings, Sir David Phillips, F.R.S., for the atomic coordinates of hen egg lysozyme, and the Royal Society, the National Science Foundation (PCM83-20171), the National Institute of General Medical Sciences (GM25435), and the European Molecular Biology Organization (ASTF 4475) for grants.

References

Almassy, R. J. & Dickerson, R. E. 1978 *Proc. natn. Acad. Sci. U.S.A.* **75**, 2674–2678.
Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. 1985 *Nature, Lond.* **313**, 156–158.
Anderson, W. F., Matthews, B. W. & Woodbury, R. G. 1978 *Biochemistry* **17**, 819.
Artymiuk, P. J. & Blake, C. C. F. 1981 *J. molec. Biol.* **152**, 737–762.
Baker, E. N. 1980 *J. molec. Biol.* **141**, 441–484.
Baker, E. N. & Hubbard, R. E. 1984 *Prog. Biophys. molec. Biol.* **44**, 97–179.
Chothia, C. 1975 *Nature, Lond.* **254**, 304–308.
Chothia, C. & Lesk, A. M. 1982 *J. molec. Biol.* **160**, 309–323.
Chothia, C. & Lesk, A. M. 1984 *J. molec. Biol.* **182**, 151–158.
Chothia, C. & Lesk, A. M. 1986 (Submitted.)
Dickerson, R. E. 1980 In *Evolution of protein structure and function, U.C.L.A. forum in medical science*, vol. 22, pp. 173–202. New York: Academic Press.
Epp, O., Colman, P., Fehlhammer, H., Bode, W., Schiffer, M. & Huber, R. 1974 *Eur. J. Biochem.* **45**, 513–524.
Fermi, G., Perutz, M. F., Shaanan, B. & Fourme, R. 1984 *J. molec. Biol.* **175**, 159–174.
Furey, W., Wang, B. C., Yoo, C. S. & Sax, M. 1983 *J. molec. Biol.* **167**, 661–692.
Gelin, B. R. & Karplus, M. 1979 *Biochemistry* **18**, 1256–1268.
Grace, D. E. P. 1979 D. Phil. thesis, University of Oxford.
Guss, J. M. & Freeman, H. C. 1983 *J. molec. Biol.* **169**, 521–562.
Holmes, M. A. & Matthews, B. W. 1982 *J. molec. Biol.* **160**, 623–639.
James, M. N. G. & Sielecki, A. R. 1983 *J. molec. Biol.* **163**, 299–361.
Janin, J., Wodak, S., Levitt, M. & Maigret, B. 1978 *J. molec. Biol.* **125**, 357–386.
Kamphuis, I. G., Drenth, J. & Baker, E. N. 1985 *J. molec. Biol.* **182**, 317–329.
Klapper, M. H. 1971 *Biochem. biophys. Acta* **229**, 557–566.
Lesk, A. M. & Chothia, C. 1980 *J. molec. Biol.* **136**, 225–270.
Lesk, A. M. & Chothia, C. 1982 *J. molec. Biol.* **160**, 325–342.
Lesk, A. M., Chothia, C., James, M. N. G. & Sawyer, L. 1986 (In preparation.)
Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. 1980 *J. molec. Biol.* **141**, 369–391.
Norris, G. E., Anderson, B. F. & Baker, E. N. 1983 *J. molec. Biol.* **165**, 501–521.
Ochi, H., Hata, Y., Tanaka, N., Kakudo, M., Sakurai, T., Aihara, S. & Morita, Y. 1983 *J. molec. Biol.* **166**, 407–418.
Phillips, S. E. V. 1980 *J. molec. Biol.* **142**, 531–554.
Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. G. 1983 *Biochemistry* **22**, 4420–4433.
Rees, D. C., Lewis, M. & Lipscomb, W. N. 1983 *J. molec. Biol.* **168**, 367–387.
Richards, F. M. 1974 *J. molec. Biol.* **82**, 1–14.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Segal, D., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. 1974 *Proc. natn. Acad. Sci. U.S.A.* **71**, 4298–4302.
Shih, H. H.-L., Brady, J. & Karplus, M. 1985 *Proc. natn. Acad. Sci. U.S.A.* **82**, 1697–1700.
Sielecki, A. R., Hendrickson, W. A., Broughton, C. G., Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. 1979 *J. molec. Biol.* **134**, 781–804.
Takano, T. & Dickerson, R. A. 1981 *J. molec. Biol.* **153**, 95–115.
Wlodawer, A., Walter, J., Huber, R. & Sjölin, L. 1984 *J. molec. Biol.* **180**, 301–329.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES