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

Three-Dimensional Structural Aspects of the Design of New Protein Molecules [and Discussion]

T. L. Blundell, D. Barlow, B. L. Sibanda, J. M. Thornton, W. Taylor, I. J. Tickle, M. J. E. Sternberg, J. E. Pitts, I. Haneef, A. M. Hemmings and B. Robson

Phil. Trans. R. Soc. Lond. A 1986 **317**, 333-344
doi: 10.1098/rsta.1986.0043

Phil. Trans. R. Soc. Lond. A **317**, 333–344 (1986)
Printed in Great Britain

Three-dimensional structural aspects of the design of new protein molecules

By T. L. Blundell, F.R.S., D. Barlow, B. L. Sibanda, J. M. Thornton, W. Taylor, I. J. Tickle, M. J. E. Sternberg, J. E. Pitts, I. Haneef and A. M. Hemmings

Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, London University, Malet Street, London LWC1 E7HX, U.K.

A knowledge of the three-dimensional structure of proteins is an essential prerequisite for the design of new molecules. When the tertiary structure is not available from high-resolution X-ray or n.m.r. analysis, the success of prediction is improved by using a relational database of known protein structures. This can be searched to provide information on secondary structure motifs and domains which are recognized by characteristic sequence patterns and which are assembled as 'spare parts' by using computer graphics. Similar techniques can be used to give approximate structures for amino-acid replacements, deletions and insertions introduced by mutagenesis. The resulting structures are optimized by using interactive graphics, energy minimization and molecular dynamics.

Every engineer – and the protein engineer is no exception – requires appropriate tools and a suitable blueprint if a design is to be effective. In protein engineering, site-directed mutagenesis provides an efficient and flexible tool for modifying a natural or synthetic DNA coding sequence, which can then be expressed in bacterial, yeast or mammalian cells, and the gene product purified in an active form by using the well established techniques of preparative biochemistry. However, a rational approach to the design of the new protein is more problematic as the three-dimensional structure may not be known, and in any case the prediction of the effects of site-specific replacements, insertions and deletions is still a formidable challenge to the theoretician.

At the present time the *ab initio* prediction of the folding pathway of a protein is beyond our technical capability, not only because there are many conformers (*ca.* 10^{100}) which must be explored even for a small protein, but also because the free energies of the unfolded, intermediate and folded states involve large enthalpic and entropic terms with only small differences between them. Even the calculation of free energy differences between wild-type and single-point mutated proteins is not yet straightforward. We conclude that predictions of protein structure and effects of site-specific mutations are best achieved by using the wealth of data already available from the high-resolution X-ray analyses of proteins rather than *ab initio* methods. In this paper we describe our approach to this problem. For site-specific replacements, insertions or deletions we use information derived from analyses of structural differences in homologous sequences and conformations derived from supersecondary motifs such as β-hairpins, αβ-loops and β-bulges, which have been defined by X-ray analysis. When the three-dimensional structure has not been defined by X-ray analysis, the tertiary structure of a homologous protein may be used as a basis for modelling with the use of similar techniques. Finally, when there

is no significant sequence homology, the tertiary structure may still be modelled approximately if a pattern of residue types or a 'fingerprint' can be identified which is characteristic of a known structure. These models may then be sufficiently accurate to allow the global energy minimum to be reached by using molecular dynamics or energy minimization. Such computer-aided modelling techniques will be most powerful when they are linked to a data base of protein sequences and structural motifs. These knowledge-based techniques provide a rational approach to the design of new molecules into which new features – binding sites, catalytic groups, stabilizing crosslinks, etc. – can be engineered.

Engineering amino-acid replacements

Our first assumption, especially if the amino acid occupies a position in the core of the protein, is that the amino-acid main chain and side chain of the mutant occupies a volume similar to that of the wild type. For many amino acids this is equivalent to placing the amino acid with the torsion angles defined by X-ray analysis for the wild type. The modelling is usually achieved by using an interactive computer graphics system with one of the subroutines of the program FRODO (Jones 1980). The protein can be displayed with a van der Waals surface; alternatively interatomic distances can be monitored on the graphics screen; and the conformation can be adjusted to avoid disallowed contact distances. A more quantitative approach, used widely in drug design, involves the continuous recalculation of the internal energy as torsion angles are varied. Alternatively a systematic search of conformational space can be performed (see, for example, Shih *et al.* 1985). The conformation or conformations selected can then be evaluated by allowing them to find local minima with the use of energy minimization or molecular dynamics procedures.

The energy minimization procedure must be carefully monitored and possibly restrained, as most sets of potentials give rise to compaction of the structure (Wodak *et al.* 1984; Weiner *et al.* 1984; Hemmings *et al.* 1985). This must arise partly from the contraction of volume at 0 K, but it may also be contributed to by van der Waals radii which are too small, lack of solvent in the simulation, surface tension effects which are equivalent to a large pressures in aperiodic structures and the neglect of induced charges and dipoles. In the absence of a cut-off distance, the compaction is large (more than 8% in volume), and the effect generally increases with the size of the protein. Even with a computationally advantageous cut-off of 9Å† and the potential set of Weiner *et al.* (1984) there is a compaction of more than 5% in volume for endothiapepsin of *ca.* 330 amino acids (Hemmings *et al.* 1985). For a single-point mutation or local changes, restrained energy minimization can be used. In the study of Shih *et al.* (1985), atoms within 5Å of the mutant amino acid were unrestrained, and atoms at successively larger distances were subjected to increasing harmonic restraints, with all atoms further than 15Å from the mutation kept fixed.

Many authors omit solvent from the calculations and use an effective dielectric constant for the electrostatic terms that is proportional to the interatomic distance. We have attempted to introduce water molecules. If the wild-type structure is defined at high-resolution by X-ray analysis, we begin by using the water molecule positions which are not excluded by the mutation. Other waters are generated by modelling or from Monte Carlo generated bulk water. As high-resolution X-ray analysis does not define the hydrogen atoms, possible positions for

† 1 Å = 10^{-1} nm = 10^{-10} m.

the water hydrogens need to be identified from an analytical procedure which finds a self-consistent hydrogen-bonding network starting from the protein surface (A. M. Hemmings 1985, unpublished results). This can then be energy-minimized or subjected to molecular dynamics techniques. Calculations on endothiapepsin show that the introduction of water molecules prevents large decreases of solvent-accessible surface area which result from the folding of polar side chains onto the surface of the molecule.

Energy calculations ignore the entropic contributions that arise from different vibrations, conformations, and configurations of the system. Many of these will be included in calculations based on molecular dynamics of the protein solvent system. Absolute thermodynamic quantities for the unfolded polypeptide–globular-protein equilibrium depend on a knowledge of the conformation(s) of the unfolded state. At first sight, modelling site-specified mutations appears to be a less difficult problem as only free energies of the wild type and mutant need be compared to assess their relative stabilities. In principle, this can be achieved by using perturbation techniques in which the molecule slowly 'grows' from wild type to mutant as molecular dynamics simulations are done (Tembe & McCammon 1984). Although this procedure has proved useful in simulating ion–water systems, calculations for methanol in water involving 'hydrophobic' effects appear to be less successful and clearly more work needs to be done on developing these theoretical approaches (W. van Gunsteren 1985, unpublished results). In the meantime we must hope that the modelling procedures based on analogy with known polypeptide conformations provide a useful prediction of three-dimensional structure, if not of the free energies.

Apart from modelling changes of active-site residues or of residues in specificity or binding pockets, it is likely that a major requirement of the protein engineer will be to increase the resistance to denaturation in non-aqueous solvents, at extremes of pH, or at high temperatures. In this respect the introduction of bridges to cross-link the structure is attractive. It is likely that ion pairs have a major effect on thermal stability and hydration. A recent analysis of ion pairs (Barlow & Thornton 1983) provides a data base by which tentative positions, chosen in the protein to be subjected to mutation, can be tested for analogous ion pairs in known structures. In a similar way, the analysis of disulphide bridges (Thornton 1981) can be used to predict suitable positions for their insertion.

Engineering insertions and deletions

Much can be learnt about engineering insertions and deletions from the natural evolution of proteins. Insertions and deletions generally occur in the loops and turns between secondary structure elements close to the surface of the protein, although insertions in β-strands can be achieved by introducing a β-bulge (Richardson 1975). Greer (1981) has suggested that loops of a certain length in a family of proteins are likely to have the same conformation, and that this may be useful in modelling where the structure of several homologous proteins have been defined by X-ray analysis. Although this may not be strictly true, it is clear that the conformations of loops are much less variable than has been previously assumed.

At Birkbeck we have undertaken an analysis of turns and loops between the standard secondary structures. Sibanda & Thornton (1985) have shown that loops between two antiparallel β-strands (β-hairpins) have well defined conformations that frequently recur in protein structures. They analysed 107 β-hairpins in 39 proteins and defined loop residues

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

(L_1, L_2, etc.) as those residues not participating in the antiparallel β-hairpin hydrogen-bonded ladder.

Two-residue loops comprised 29 of the 107 hairpins. Although Lewis *et al.* (1973) found that, in general, 60% of turns were type I and type III and 15% are type II, Sibanda & Thornton (1985) show that for β-hairpins, 25 were the mirror images (type I′ or type II′) (figures 1, 2) and only four were type I. It appears that although type I′ and II′ turns may

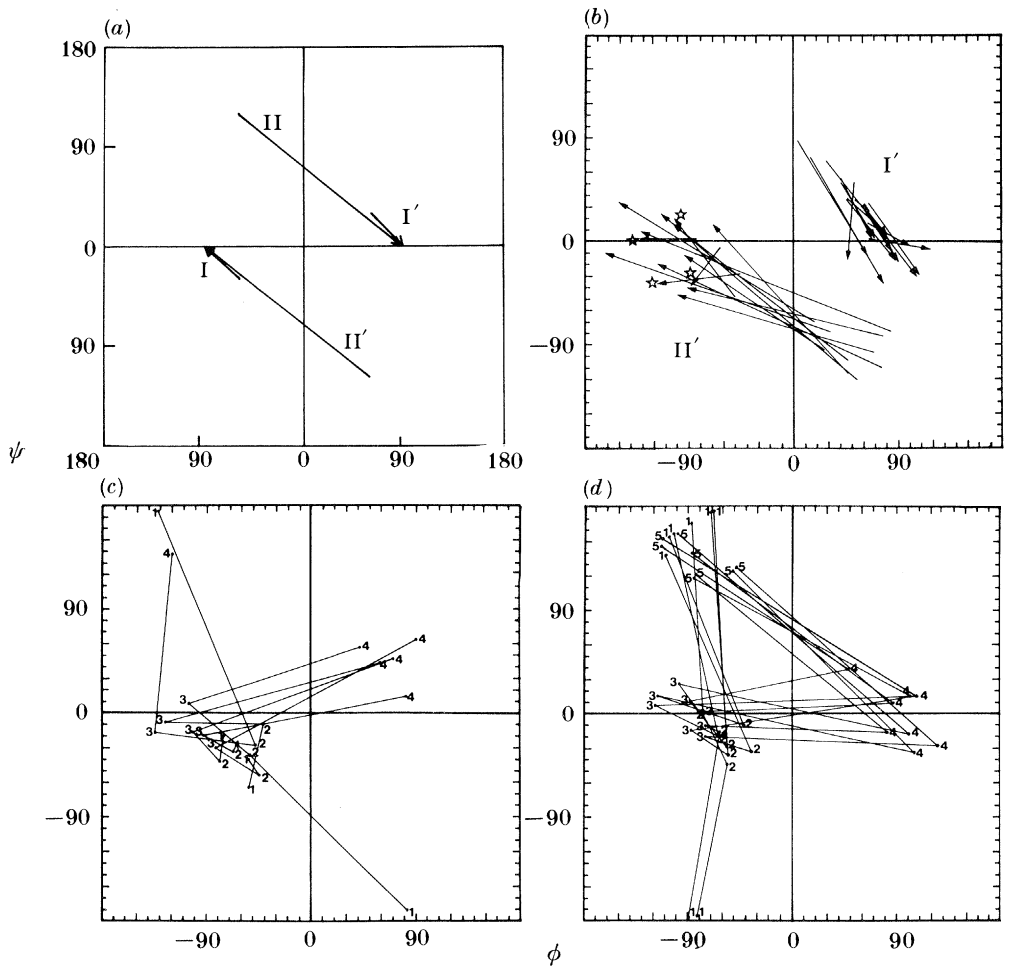

FIGURE 1. Conformation of loop regions shown by plotting ϕ, ψ values for the loop residues. (*a*) Classical β-turn conformations types I and II and their mirror images I′ and II′. These are defined by the ϕ, ψ values of the two central residues of the turn (usually referred to as residues 2 and 3) which in the context of, say, a two-residue hairpin, will be described as residues L1 and L2. The ϕL1, ψL1, ϕL2, ψL2 values plotted are type I = −60°, −30°, −90°, 0°; type I′ = +60, +30°, +90°, 0°; type II = −60°, 120°, 90°, 0°; type II′ = 60°, −120°, −90°, 0°. Because type III turns do not form a group distinct from the type I turns (Richardson 1981), they have been omitted. (*b*) The ϕ, ψ plot for the 29 two-residue hairpin loops. For each loop the ϕ, ψ values of residue L1 are connected by a line to the ϕ, ψ values of residue L2 with the arrow-head denoting the L2 position. Types I′ and II′ turns predominate with only four type I turns, and no type II turns, observed. (*c*) The ϕ, ψ plot for the six four-residue hairpin loops with type I turns. For each loop the ϕ, ψ values of residues L1, L2, L3 and L4 are plotted and joined together sequentially. Five of these hairpins also have residue L4 in the all-positive quadrant of the ϕ, ψ plot. The ϕ, ψ angles calculated by averaging over the four similar loops for residues L1 to L4 are −62°, −38°; −58°, −35°; −101°, −19°; 75°, 42°. (*d*) The ϕ, ψ plot for the eight five-residue hairpin loops which form a family. For each loop the ϕ, ψ values of residues L1, L2, L3, L4, and L5 are plotted and joined together sequentially. Note the type I turn for residues L2 and L3 and the positive ϕ value for residue L4. The standard ϕ, ψ angles for residues L1 to L5, calculated by averaging over the eight hairpins in the family, are −90°, 161°; −56°, −24°; −92°, 2°; 95°, 2°; −85°, 36°.

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

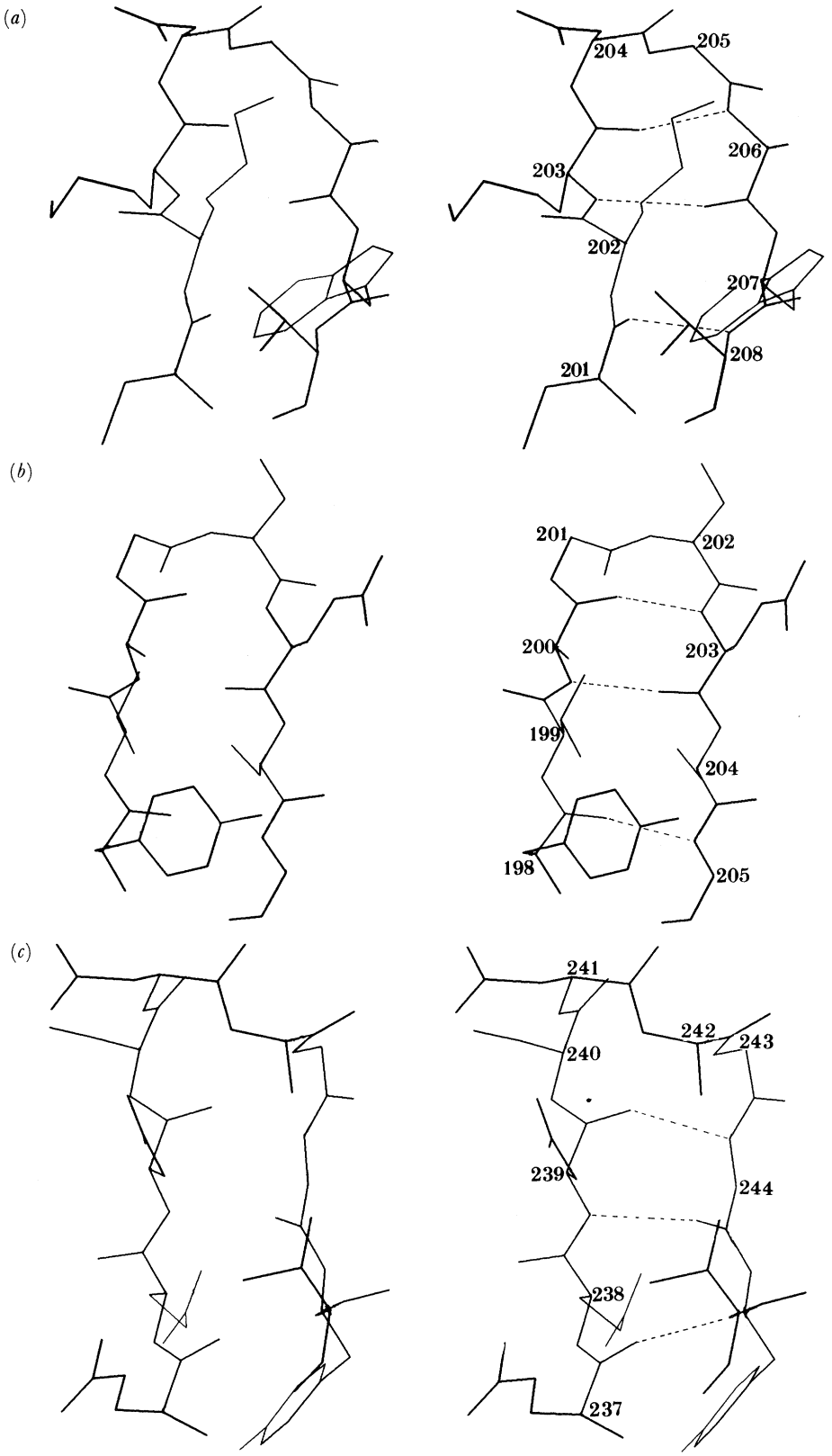

FIGURE 2 (*a–c*). For description see next page.

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

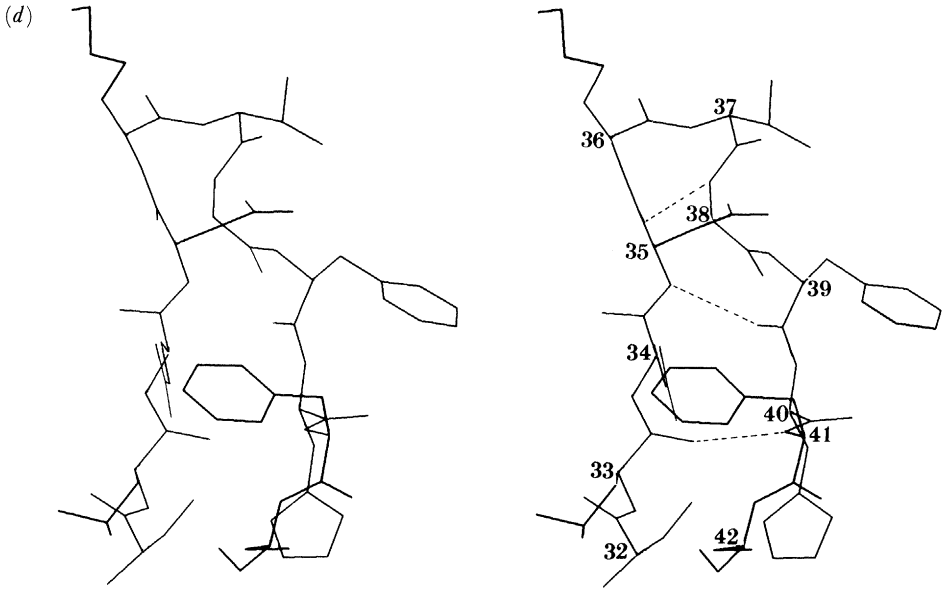

FIGURE 2. Stereodiagrams of representative structures for the two; four-, and five-residue hairpin-loop families. (*a*) Two-residue hairpin loop with a type I′ β-turn; γ-chymotrypsin A, residues 201–208. (*b*) Two-residue hairpin loop with a type II′ β-turn; penicillopepsin, residues 198–205. (*c*) Four-residue hairpin loop with a type I β-turn; penicillopepsin, residues 237–246. Notice how the amide NH groups of residues L2, L3 and L4 point inwards towards the centre of the turn and to the carbonyl oxygen of the residue immediately before the loop (−B1). The side chain of this residue may help to neutralize the cluster of NH groups. (*d*) Five-residue hairpin loops with a type I β-turn and a G1 β-bulge; γ-chymotrypsin, A, residues 32–42. This structure is reminiscent of the four-residue family in that the nitrogens of residues L3, L4 and 5 all point inwards towards the carbonyl of residue L1, and an aspartyl oxygen is often observed nearby.

be less favourable due to steric hindrance, they give a twist which is compatible with that of the β-hairpin ladder. Thus we learn that modelling two-residue turns in β-hairpins, we should first consider types I′ or II′, a conclusion that will be a surprise to most protein modellers.

Three-residue turns are less common than those with four or five residues. Figures 1 and 2 show conformations and ϕ–ψ plots for some characteristic four- and five-residue loops. For example, six of the thirteen four-residue loops contain a type I turn, with residues L1, L2 and L3 in the α-helical region of the ϕ–ψ plot and residue L4 (glycine or asparagine) in the all-positive quadrant. In a similar way, eight of the nineteen five-residue loops contain a type I turn, with a hydrogen bond between the carbonyl of residue L1 and the amide of residue L4, and a G1-type β-bulge. This structure has a glycine, asparagine or aspartate at L4.

Each of the types of loops have associated restrictions on the sequence both in the loop and on each side of the loop. Type I′ turns prefer L1 as Asn, Asp or Gly and L2 as Gly; type II′ prefer L1 as Gly and L2 as Ser or Thr; and four- and five-residue loops often have the preferences noted above for L4 as glycine. These observations often allow the loop conformations to be identified from a combination of length and sequence.

We are performing similar analysis for other loop structures including β-arches (i.e. between different β-sheets), αβ-loops and αα-loops, which are revealing patterns of conformations. For example, the residue pattern

—hydrophobic—hydrophobic—X—residue favouring β-conformation—

PRO—X—hydrophobic—hydrophobic—

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

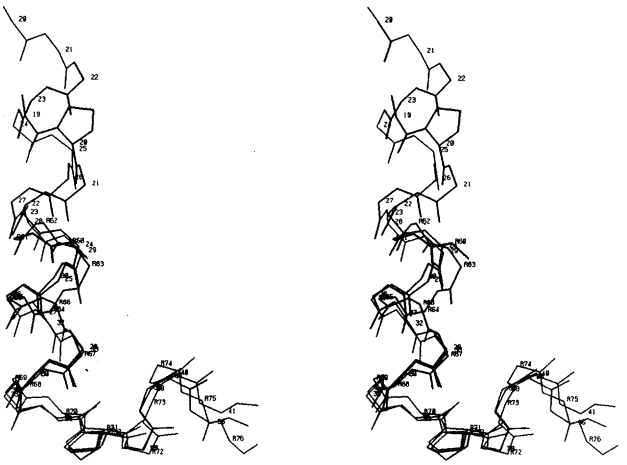

FIGURE 3. A family of loops joining consecutive α-helices, in which the key residue is a proline. The three structures illustrated (involving residues 20–41 in myoglobin, residues 19–36 in erythrocruorin, and residues 60–75 in cytochrome c) are superimposed with an r.m.s. error (calculated for the C-α atoms of the proline, the three residues following it and the four residues preceding it) of *ca.* 0.7 Å.

gives rise to helices related by 50–90° (figure 3; D. J. Barlow & J. M. Thornton 1985, unpublished results). Specific sequence and structural patterns are particularly evident for short loops of all classes – αα and αβ as well as ββ – in which glyine is usually involved, and these conformations will be most straightforward to use in modelling. For longer loops the conformations are more variable.

How can these be used in modelling insertions and deletions? We are adopting the following procedure. If an insertion or deletion is to be made in a loop region, we begin by characterizing the residues before and after the loop to be modified in terms of their main chain torsion angles and the interatomic distances between C-α atoms. Similar information will be stored in a relational data base for all precisely defined loops of proteins whose three-dimensional structures are defined by X-ray analysis. We then search for all loops of the correct length, matching the torsion angles and distance matrices of the secondary structure strands and the sequence patterns defined by the analysis of protein structure. The distance matrices are particularly important when α-helices are involved. We select the best matches and weight these by their frequency of occurrence to obtain a rank order.

We must then test the loops against the rest of the known wild-type structure. We do this by docking the loop so that it occupies approximately the volume of that in the wild type and does not overlap with the rest of the structure, using logical operations on volumes with the program BILBO of Honegger & Blundell (1984). We join the chains by using the Hermans & McQueen (1974) algorithm to regularize the geometry. The conformations of the side chains are guided by those in the data base of structures and the whole system is energy minimized, constraining movements of the protein that are more than 10 Å distant from the loop residues. Other local minima can be explored by using restrained molecular dynamics. The optimal conformations are chosen not only from energy criteria but also in terms of the accessibilities of hydrophobic and hydrophilic groups, ion pairs, etc.

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

Modelling homologous proteins

So far we have assumed that the structure of the wild-type protein has been defined by X-ray analysis at high resolution. In many cases this will not be so, and we need to model the protein before we can start to consider the conformational effects that result from mutagenesis. If the structure of a homologous protein is available this may provide a basis for the generation of such a model.

The first model of this kind reported was for α-lactalbumin on the basis of lysozyme (Browne *et al.* 1979). We have used the method over the past fifteen years to model many proteins, including members of the insulin family (Wood & Blundell 1975; Bedarkar *et al.* 1977; Blundell *et al.* 1978, 1983), the aspartic proteinases (Sibanda 1980; Blundell *et al.* 1983; Sibanda *et al.* 1984), lens β/γ-crystallins (Wistow *et al.* 1981; Inana *et al.* 1983), and histocompatability antigens (Travers *et al.* 1984). Although the general approach is the same, the quality of the model will be very much dependent on the degree of divergence of the homologous proteins.

The three-dimensional structures of several insulins of differing sequences have been defined by X-ray analyses. Although crystal packing, ionic strength and the presence of specific ions can cause large conformational changes, it has been shown that a closely similar conformer can exist for primitive vertebrate insulins and those of mammals (Dodson *et al.* 1979). This is because the hydrophobic core is conserved and so the secondary structural elements can retain an identical scaffold. For the insulin-like growth factors the residues of the hydrophobic core are also identical and the pattern of glycines in the main chains at A1, B8, B20 and B23 is also retained. We are thus confident that our models of the insulin-like growth factors are reliable indications of their true structures (Blundell *et al.* 1978; Blundell *et al.* 1983). The main ambiguities are in the C- and D-peptide regions, which represent small loops on the surface and can be predicted by the methods described in the previous section.

What happens when there are differences in the residues of the hydrophobic core between two homologous proteins? Chothia & Lesk (1980) have shown for α-helical proteins (globins) that the secondary structural elements tend to have different relative positions although the nature of the interactions (ridge to groove) is largely retained. There are, of course, complementary changes which can lead to a rough retention of the volume. For example, in modelling relaxin on the homologous insulin structure we noted that, although most residues of the core were varied, they were retained as hydrophobic with complementary changes such as A2 Ile, A16 Leu in porcine insulin compared with A2 Leu, A16 Ile in porcine relaxin, and B6 Leu, B14 Ala in porcine insulin and B6 Ala, B14 Leu in porcine relaxin (Bedarkar *et al.* 1977). Similar changes also occur on the surface; thus in porcine insulin we have A5 Gln, A15 Gln in close proximity, while in relaxin there is A5 Asp, A15 Arg, and in IGF there is A5 Arg, A15 Asp. Such complementarity allows a rough model to be constructed by assuming a rigid scaffold of secondary structural elements. We have then allowed the molecule to relax into a minimum by using energy minimization; the minimum will only be the global minimum if the initial model was sufficiently correct. For the model of protein S based on γ-crystallin (Wistow *et al.* 1985) this has led to a contraction of volume (S. Najmudin, A. M. Hemmings & T. L. Blundell 1985, unpublished results), which is greater than that for γ-crystallin as there is a decrease in core volume. However, it is unlikely that energy minimization techniques will effectively model the reorientation of secondary structural elements noted by Chothia & Lesk (1980). It will be necessary to attempt systematic searches of packing of rigid secondary structure elements to

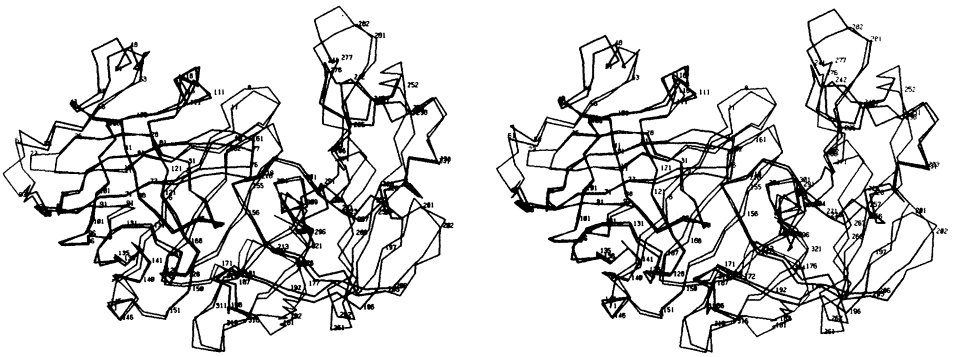

FIGURE 4. A least-squares fit of endothiapepsin and penicillopepsin illustrates the similar main-chain arrangements in the central core regions and in the active-site cleft.

TABLE 1. VARIATION OF THE TERTIARY STRUCTURE BETWEEN THE EQUIVALENT GREEK KEY MOTIFS (I–IV) OF γII-CRYSTALLIN DEFINED AT 1.9 Å RESOLUTION

r.m.s. separation (ångstroms) of equivalent C-α atoms

	I	II	III	IV	
I	—	12	13	7	
II	1.26	—	8	19	number of chemically
III	1.05	1.48	—	8	identical equivalent residues
IV	1.4	0.69	1.22	—	

R.m.s. values and chemical identities are for 38 pairs of equivalent amino acids except for the comparison between motifs II and IV, which used 43 pairs; 82 pairs of C-α atoms in the N-terminal (motifs I and II) and C-terminal domains were separated by 0.99 Å at best fit.

increase the radius of convergence, although molecular dynamics calculations or the searching of energy minima along low-frequency normal modes appear to offer attractive alternative approaches.

For structures that have an evolving hydrophobic core we have made systematic searches of the packing relation of secondary structures in members of the family already defined by X-ray analysis. Table 1 shows the r.m.s. differences between topologically equivalent C-α positions for the Greek key motifs of γII-crystallin defined at 1.9 Å resolution (Blundell *et al.* 1981; Wistow *et al.* 1983; Summers *et al.* 1984). These demonstrate a strong relation between the sequence divergence and the three-dimensional structural relations. In modelling protein S we were careful to use those motifs that showed closest sequence homology; thus we need motifs I and II of γII to model both S_2, S_1 and S_4, S_3 (Wistow *et al.* 1985).

For the aspartic proteinases we have compared the known three-dimensional structures of porcine pepsin, penicillopepsin, endothiapepsin, and rhizopuspepsin (Sibanda *et al.* 1984). For penicillopepsin and endothiapepsin, sequences are defined by chemical methods and the structures are refined to $R < 17\%$ by using restrained least squares at 1.8 and 2.1 Å resolution respectively. These have 55% homology and have 207 pairs of C-α atoms with an r.m.s. difference of 0.48 Å. Closer examination of the structures with the use of computer graphics (see figure 4) shows that the active-site regions are very closely similar in conformation as they are in sequence. Most differences occur in loops at the periphery and in the hydrophobic cores towards the edges of the bilobal enzymes. The fit with the rhizopuspepsin is less good; this is probably in part because the sequence defined by X-ray analysis was only 50% correct and

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

the structure was incompletely refined; and in part because the enzyme is less similar. For pepsin the differences are much greater; here the structure is at lower resolution and even less well refined.

How then should we model a mammalian enzyme such as human renin which has 40% homology with pepsin but only 22% homology with endothiapepsin and penicillopepsin? We decided that the pepsin structure was not sufficiently well refined to provide a good model. As the sequence of rhizopuspepsin was not available from chemical sequencing and the structure was not completely refined, this did not appear to be the best candidate for a starting model; however, Carlson *et al.* (1985) and Akahane *et al.* (1985) have since modelled renin with the use of this structure. Instead, we modelled on the basis of endothiapepsin but considered the structures of the other enzymes as the modelling proceeded, especially for loops and other divergent regions in the enzyme structures (Blundell *et al.* 1983; Sibanda *et al.* 1984). This approach was also justified by the high affinity of endothiapepsin for analogues and inhibitors with human angiotensinogen sequences (Hallet *et al.* 1985). In general we opt for a structure that is well defined both chemically and by X-ray analysis. However, we remember that those regions (active site and cleft) which are conserved in evolution between other aspartic proteinases are most likely to be correctly modelled in human renin.

Perhaps the most critical stage in modelling by homology is the alignment of the sequences. The alignment is best obtained by noting the relation between the sequences of the known three-dimensional structures and listing residues at topologically equivalent positions. This will often reveal the equivalence of residues which do not give the highest sequence homology when based on a simple residue identity count. It will also identify residues that are invariant and positions that are conservatively varied for structural reasons within certain categories of residues. Alignment of the new sequence then proceeds by first identifying the critical residues that are invariant or nearly invariant. These often turn out to be glycines (either with conformations with ϕ positive or where no side chain is possible); *cis*-prolines; buried disulphides such as those in the immunoglobulins; or buried polar residues such as Thr-33 and Thr-216 in aspartic proteinases. We then check patterns of hydrophobics required by the tertiary structure. A program has been devised to perform this automatically (W. R. Taylor 1985, unpublished results). If these criteria are met it is likely that those of Novotny *et al.* (1984), concerning accessibility of hydrophobics and hydrophilics, will also be satisfied. Although our criteria are more subjective at first sight, they will probably distinguish between local modelling differences to which the overall criteria will be relatively insensitive. In summary, before any model building is started we must understand the structural data already available for the homologous proteins.

The elaboration of the homologous structure once the main relations between secondary structural elements is achieved can proceed in the manner already described for point mutations, insertions, and deletions.

Designing novel proteins

The ultimate objective of protein engineering must be to design novel proteins with new binding functions, catalytic activities, electrical properties, etc. There are two different approaches. We can begin with a known structure as a scaffold, for example an immunoglobin fold or an α-helical fold of the kind found in tobacco mosaic virus, etc. (Richardson 1981),

and elaborate this locally to give a binding site. This we may term 'local design'. Alternatively, we may design one of the classic folds from scratch. This may be termed 'global design'. In the near future, binding proteins with a range of specificities could have implications for purification processes from affinity chromatography to the extraction of rare metals from minerals. In the long term, the design of biological microchips for electronic circuitry is a real possibility. In all these processes the design aspects of protein engineering will become increasingly central to successful synthesis of novel proteins. We believe that the rational approach will be to use our detailed knowledge of protein structure – available in an easily accessible and adaptable relational data base – to aid in the modelling process. Protein engineering in the future will be dependent on knowledge-based and computer-aided design.

References

Akahane, K., Umeyama, H., Nakagawa, S., Moriguchi, I., Hirose, S., Iizuka, K. & Murakami, K. 1985 *Hypertension* **7**, 3–12.

Barlow, D. J. & Thornton, J. M. 1983 *J. molec. Biol.* **168**, 867–885.

Bedarker, S., Turnell, W. G., Blundell, T. L. & Schwabe, C. 1977 *Nature, Lond.* **270**, 449–451.

Bedarkar, S., Blundell, T. L., Gowan, L. K., McDonald, J. K. & Schwabe, C. 1982 *Ann. N.Y. Acad. Sci.*, 22–23.

Blundell, T. L., Bedarkar, S. & Humbel, R. E. 1983*a* *Fedn. Proc. Fedn Am. Socs exp. Biol.* **42**, 2592–2597.

Blundell, T. L., Bedarkar, S., Rinderknecht, E. & Humbel, R. E. 1978 *Proc. natn. Acad. Sci. U.S.A.* **75**, 180–184.

Blundell, T. L., Lindley, P., Miller, L., Moss, D., Slingsby, C., Tickle, I. J., Turnell, W. G. & Wistow, G. 1981 *Nature, Lond.* **289**, 771–777.

Blundell, T. L., Sibanda, B. L. & Pearl, L. H. 1983*b* *Nature, Lond.* **304**, 273–275.

Blundell, T. L. & Wood, S. P. 1975 *Nature, Lond.* **257**, 197–203.

Browne, W. J., North, A. C. T., Phillips, D. C., Brew, K., Vanamann, T. C. & Hill, R. L. 1969 *J. molec. Biol.* **42**, 65–86.

Carlson, W., Karplus, M. & Haber, E. 1985 *Hypertension* **7**, 13–26.

Dodson, E. J., Dodson, C. G., Hodgkin, D. C. & Reynolds, C. D. 1969 *Can. Journ. Biochem.* **57**, 469–479.

Greer, J. 1981 *J. molec. Biol.* **153**, 1027–1042.

Hallet, A., Jones, D. M., Atrash, B., Szelke, M., Leckie, B. J., Beattie, S., Dunn, B. M., Valler, M. J., Rolph, C. E., Kay, J., Foundling, S. F., Wood, S. P., Pearl, L. H., Watson, F. E. & Blundell, T. L. 1985 In *Aspartic proteinases* (ed. V. Kostka). Berlin: Walter de Gruyter. (In the press.)

Hemmings, A. M., Foundling, S. F., Sibanda, B. L., Woods, S. P., Pearl, L. H. & Blundell, T. L. 1985 *Trans. Biochem. Soc.* (In the press.)

Hermans, J. & McQueen, J. E. 1974 *Acta crystallogr.* A30, 730–738.

Honegger, A. M. & Blundell, T. L. 1984 In *Insulin-like growth factors/somatomedius: basic chemistry, biology and clinical importance* (ed. E. M. Spencer), pp. 9–113. Berlin: Walter de Gruyter.

Inana, G., Piatigorsky, J., Norman, B., Slingsby, C. & Blundell, T. L. 1983 *Nature, Lond.* **302**, 310–315.

Jones, T. A. 1978 *J. appl. Crystallogr.* **11**, 286–272.

Lesk, A. M. & Chothia, C. 1980 *J. molec. Biol.* **136**, **136**, 225–270.

Novotny, J., Bruccoleri, R. & Karplus, M. 1984 *J. molec. Biol.* **117**, 787–818.

Richardson, J. S. 1981 *Adv. Protein Chem.* **34**, 167–339.

Richardson, J. S., Getzoff, E. D. & Richardson, D. C. 1978 *Proc. natn. Acad. Sci. U.S.A.* **75**, 2574–2578.

Shih, H.-L., Brady, J. and Karplus, M. 1985 *Proc. natn. Acad. Sci U.S.A.* **82**, 1697–1700.

Sibanda, B. L. 1980 *Molecular modelling of chymosin.* M.Sc. thesis, University of London.

Sibanda, B. L., Blundell, T. L., Hobart, P. M., Fogliano, M., Bindra, J. S., Dominy, B. W. & Chirgwin, J. M. 1984 *FEBS Lett.* **174**, 102–111.

Sibanda, B. L. & Thornton, J. M. 1985 *Nature, Lond.* **316**, 170–174.

Summers, L., Wistow, G., Narebor, M., Moss, D. S., Lindley, P. F., Slingsby, C., Blundell, T. L., Bartunik, H. & Bartels, K. 1984 *Pept. Protein Rev.* **3**, 147–168.

Tembe, B. L. & McCammon, J. A. 1984 *Comp. Chem.* **8**, 204–210.

Thornton, J. M. 1981 *J. molec. Biol.* **151**, 261–287.

Travers, P., Blundell, T. L., Sternberg, M. J. E. & Bodmer, W. F. 1984 *Nature, Lond.* **310**, 235–238.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, V. C., Ghio, C., Alagone, G., Proteta, S. & Weiner, P. 1984 *J. Am. chem. Soc.* **106**, 765–784.

Wistow, G., Slingsby, C., Blundell, T. L., Driessen, H., de Jong, W. & Bloemendal, H. 1981 *FEBS Lett.* **133**, 9–16.

Wistow, G., Summers, L. & Blundell, T. L. 1985 *Nature, Lond.* **315**, 771–773.

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

Wistow, G., Turnell, W. G., Summers, L., Slingsby, C., Moss, D. S., Miller, L., Lindley, P. F. & Blundell, T. L. 1983 *J. molec. Biol.* **170**, 175–202.
Wodak, S. J., Alard, P., Delhaise, P. & Renneboog-Squilbin, C. 1984 *J. molec. Biol.* **181**, 317–322.

Discussion

B. Robson (*Theoretical Biochemistry Laboratory, The Medical School, Manchester*). Professor Blundell's controversial statement, that energy minimization is of little value as it leads only to the nearest minimum in the potential surface, implies a very limited definition of energy minimization. As a general definition it would only be applicable to very old minimization technology. Global minimization, which explores many minima in a directed maner, is what is now being developed, and is demonstrably the crucial approach to opening up the further potential of protein engineering. By using his own example of β-hairpins, he will recall the collaboration between ourselves and Drs Sternberg and Thornton of his laboratory, using our programs to evaluate the intrinsic relative stabilities of different turn types to compare with their relative abundance in proteins. Leaving aside the question of the good quality of the results, many minima were certainly automatically and rapidly negotiated in a directed manner, implying an extensive search of the potential surface. The definition may or may not apply to Birkbeck technology in this area but this is not what we should be looking for.

T. L. Blundell. Dr Robson is correct that it is possible to search through conformational space to find the global minimum for small peptides where the number of degrees of freedom is small; and this technique may be used to investigate loop conformations in proteins, as I mentioned in my talk. However, this approach becomes prohibitively expensive with large proteins. I thought it important to make the point forcefully that the normal energy-minimization techniques find a local minimum; it is not always realized that very little conformational change is required to find a local minimum even with an incorrect structure. At Birkbeck we use our knowledge of protein structure along with interactive computer graphics to approach the global minimum before the expensive calculations are performed. Until our knowledge of proteins is properly included in a data base – a project under way at Birkbeck – it is certainly better to think carefully before entering into an automated procedure.

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