Pure & Appl. Chem., Vol. 66, No. 1, pp. 43–50, 1994.
Printed in Great Britain.

Comparative modelling of human renin: A retrospective evaluation of the model with respect to the X-ray crystal structure

Carlos Frazao, Chris Topham, Venugopal Dhanaraj and Tom L. Blundell.

Laboratory of Molecular Biology and ICRF Unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, London WC1E 7HX, U.K.

Abstract

A three-dimensional model of human renin has been constructed using COMPOSER, a rule-based approach to comparative modelling. The model was constructed from the three-dimensional structures of two homologous aspartic proteinases, pepsin and chymosin, which have been defined by X-ray analysis. The variable regions were obtained from aspartic proteinase structures for most of the loops, but also from unrelated structures where appropriate conformers were not available within the family. Using a 3.5Å cut-off for pairs of equivalent atoms, 280 C-α atom pairs of the model and the experimentally defined structure superpose with an rmsd of 0.84Å. Larger deviations occur in some of the loop regions, especially around residues 290 where there is a proline-rich sequence with CIS-peptides. The analysis shows that comparative modelling can give a reasonably accurate estimate of homologous protein structures especially in the active-site region and that these provide a useful basis for structure-based design in the absence of experimental data.

INTRODUCTION

Divergent evolution has given rise to families of homologous proteins that differ in sequence but adopt the same general fold. This provides an opportunity to learn about the three-dimensional structures of proteins if their sequences have been defined and at least one other member of the family has a structure defined by X-ray analysis or nuclear magnetic resonance.

The first application recorded in the literature of this procedure was the construction of a model for alpha-lactalbumin on the basis of lysozyme (ref. 1) using physical models. The advent of computerised techniques, particularly of the computer graphics program FRODO (ref. 2), made the task of replacing sidechains and making insertions and deletions more straightforward. However, modelling was rarely performed applying rigorous rules, although some systematic procedures were suggested, for example, for the use of loops from homologous proteins (ref. 3).

More recently there has been interest in developing methods for constructing a model on the basis of rules derived not only from comparison of related structures but also from the analyses of protein structures in general. The use of these rules depends very importantly on the alignment of the sequence of the protein to be modelled with the template for the family fold. Most methods depend on the assembly of rigid fragments (ref. 4-7). This modelling procedure is most successful where several known structures cluster around that to be predicted and where the percentage sequence identity to the unknown is high (greater than 30%) (ref. 8,9). Similarly, 80% of sidechain conformations are correctly predicted for closely homologous structures (ref. 10).

Renin (ref. 11) is the key enzyme of the renin-angiotensin cascade which regulates the formation of angiotensin II and the control of vascular tone, fluid volume and sodium excretion in the circulation. Because of the unique specificity of renin, its inhibition is widely expected to provide selective therapy for hypertension, congestive heart failure and associated degenerative disorders linked to angiotensin II. The 3-D structures of renin-inhibitor complexes have long been sought as an aid to the discovery of clinically effective antihypertensives (ref. 12).

Sequences of renins (ref. 13,14) show that they belong to the family of aspartic proteinases, several of whose structures are already known (ref. 15 and references therein). These structures have provided a basis for modelling mouse (ref. 16) and human renins (ref. 17-19) using interactive computer graphics. Later when COMPOSER, a rule-based computer modelling approach, became available (ref. 6), we remodelled human renin using the three-dimensional structures of pepsin (ref. 20) and chymosin (ref. 15) defined at that time in our laboratory.

X-ray analysis of an uncomplexed, partially deglycosylated form of recombinant human renin was reported by Sielecki et al. (ref. 21). However, the coordinates of the three-dimensional structure were not generally available until very recently and so detailed evaluation of the models could not be made. X-ray structural studies of renin inhibitors complexed with other aspartic proteinases (for example ref. 22) have defined the conformation of the extended main-chain of the inhibitor, the location and nature of the specificity subsites and the interactions of various transition-state isosteres with the catalytic aspartates of these aspartic proteinases. However, these analyses left unanswered many questions concerning the specificity of renin for angiotensinogen and the differences in specificities of the human, mouse and other renins. More recently we (ref. 23) reported the X-ray analyses and refinement of inhibitor-complexed structures of mouse submandibular and recombinant human renin defined at 1.9 and 2.8A resolution respectively. Similar analyses of human renins have been reported by others (ref. 24). These structures give a basis for understanding the specificity of renin in terms of the three-dimensional structures and indicate an important role for the ordered loop regions, which exist uniquely at the periphery of the binding cleft and are disordered in the uncomplexed enzyme.

We can now retrospectively evaluate the success of the comparative modelling approach. In this paper we describe the comparative modelling of human renin using the coordinates derived from the crystal structures of pepsin, chymosin and several fungal homologues. We then compare the model with the structure defined at a later stage by X-ray analysis.

THE COMPOSER APPROACH TO COMPARATIVE MODELLING

Our approach to comparative modelling uses knowledge from both homologues and from unrelated proteins (ref. 6,25,26). The first stage involves the alignment of three-dimensional structures of homologues using COMPARER (ref. 27,28) and the calculation of percentage sequence identities based on the resulting alignment. This is used to cluster the structures (ref. 29,30) and to select the homologous structures that are most useful for construction of the model. The three sets of fragments selected are :

(i) fragments from the framework defined by multiple least-squares superposition of the structures (ref. 8).

(ii) fragments for regions outside the framework selected from the database of loop sub-structures using a distance filter in a similar way to Jones and Thirup (ref. 4). Sequence templates for each of the selected fragments are derived using the environment-dependent substitution tables (ref. 31,32) and are compared to the sequence to be modelled (ref. 26). The top ranking fragment is annealed to the core and checked for overlap with other parts of the model. If it is rejected on these grounds, the next ranking fragment is processed in the same way.

(iii) fragments of sidechains selected by using a set of rules derived from the analysis of sidechain dihedral angles at topologically equivalent positions in homologous structures (ref. 10). The 1200 rules derived from this analysis include one for each of the 20 by 20 amino acid replacements in each of the three secondary structure types (α-helix, ß-strand or irregular). Where there is no applicable rule, the most probable conformation is chosen from a rotamer library, and where there is more than one prediction, the one closest to the median of all predictions is chosen.

Finally, the model is energy minimized to remove steric clashes and to obtain an acceptable geometry and conformation.

All computations were performed using the Birkbeck Crystallographic Computer Network involving a VAX 11/750, a microVAX11 or a VAX station. Amino acid sequences and three-dimensional coordinates were obtained from the Brookhaven Data Bank (ref. 33), with the exception of those from pepsin and chymosin which were available in Birkbeck (now available from Protein Data Bank). Manual inspection of models was performed using FRODO (ref. 2) on an Evans and Sutherland PS300 or PS390.

CONSTRUCTION OF THE HUMAN RENIN MODEL

At the time of modelling in 1989 we had available the three-dimensional structures of three fungal aspartic proteinase structures (rhizopuspepsin, endothiapepsin and penicillopepsin) and two mammalian aspartic proteinases, porcine pepsin and calf chymosin, which had recently been defined at Birkbeck (ref. 15,20,34). These were superposed as rigid bodies first using MNYFIT (ref. 8) and then using a more flexible approach, encoded in the computer program COMPARER, which equivalences local protein structural features and relationships (ref. 27). The percentage sequence identities and the root mean square differences (rmsd) for each pair of structures are shown in Table 1. It is clear that the sequences of porcine pepsin and calf chymosin cluster much closer to human and mouse renins. These were used to select the structurally conserved regions (SCRs).

TABLE 1. A matrix giving the percentage sequence identities (upper right) and rmsds (in Å) for pairwise comparisons of X-ray structures described in this paper.

Protein code	4ape	3app	2apr	pep	4cms	hren	2ren
Endothiapepsin	**	54.7	39.2	28.8	27.4	25.2	25.0
Penicillopepsin	0.71	**	41.1	32.9	30.7	24.6	26.9
Rhizopuspepsin	0.87	0.86	**	37.3	32.7	25.9	28.0
Porcine pepsin	1.08	0.98	0.95	**	59.6	39.8	43.0
Calf chymosin	1.04	0.94	0.81	0.69	**	38.3	38.8
Human renin	1.19	1.12	1.04	0.93	0.89	**	70.9
Mouse renin	1.25	1.20	1.13	1.09	1.02	0.51	**

A distance of 2Å between equivalent amino acid residues was used to construct a framework comprising the weighted average of Cα positions for the seventeen structurally conserved regions. Weights were dependent on percentage sequence identities. These are defined in Fig. 1, which shows the alignments of sequences of porcine pepsin, calf chymosin and human renin. To obtain a real mainchain, each of the SCRs was then modelled by fitting to the framework the fragment of pepsin or chymosin with the closest percentage sequence identity.

The structurally conserved regions were modelled from fragments of the structures of chymosin and pepsin, indicated by c and p respectively; the structurally variable regions (SVRs) are indicated by v.

The structurally variable regions (SVRs) were selected from a data base of fragments derived from the Brookhaven Data Bank on geometric criteria using a three residue overlap with each part of the framework linked by the fragment selected for the SVRs (ref. 3,6,26). The seventeen selected SVRs are listed in Table 2; thirteen derive from other aspartic proteinases, including the fungal enzymes, whereas the remaining four are from proteins that are not homologous. In certain regions we were able to extend the framework by a "collar region" obtained from either pepsin or chymosin, where the sequence of the SVR was similar but the length slightly different. The shorter insertion region was selected in the same way as other SVRs.

Fragments selected for each variable region were then least squares fitted to their respective SCRs using a three residue overlap at each connection. The SVR is melded into the existing framework using a sliding weighting scheme to position mainchain atoms in the overlapping regions. Fig. 2 shows a stereo view of the energy-minimised human renin model.

Fig 1. Alignments of sequences of pepsin (5pep), chymosin (4cms) and human renin (hren) based on comparisons of three-dimensional structures of aspartic proteinases. The amino acid code is the standard one-letter code formatted using the following convention:

italic	positive Phi;
UPPER CASE	solvent inaccessible residues;
lower case	solvent accessible residues;
bold type	hydrogen bonds to mainchain amide;
underline	hydrogen bonds to mainchain carbonyl;
tilde	sidechain-sidechain hydrogen bonds;
ç (cedilla)	disulphide bridge partner.

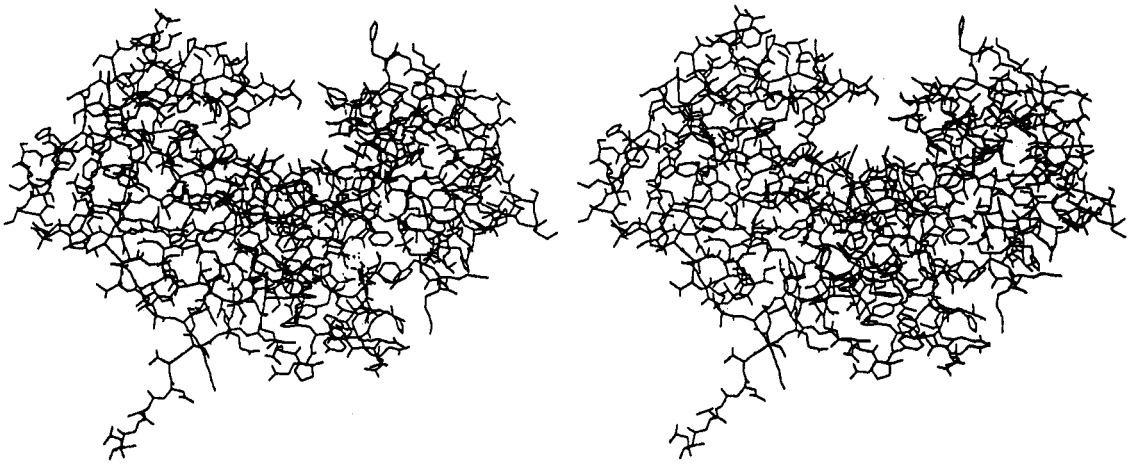

Fig 2. A stereo view of the energy-minimised human renin model.

TABLE 2. The structurally variable regions (SVRs) of human renin sequence and the structures from which these fragments were selected. Pepsin numbering scheme is used for the residues and insertions relative to pepsin in the renin sequence are underlined.

SVR No	Sequence number	Fragment chosen from the structure of	Sequence of fragment	Sequence in renin
1	-6 - 1	Pepsinogen	LVKVPLV	LTLGNTT
2	8 - 14	Pepsin	NYLDTEY	NYMDTQY
3	22 - 24	Chymosin	TPP	TPP
4	46 - 49	Glycolate oxidase	YKDRNV	SRLYTA
5	66 - 68	Pepsin	ATS	HNG
6	72 - 80	Pepsin	SITYGTGSM	TLRYSTGTV
7	97 - 100	γ-II crystallin	HSL	TQM
8	130 - 133	Chymosin	SEYS	IGRV
9	158 - 161	Trypsin *S. griseus*	GYNGTGKD	DSENSQSL
10	174 - 176	Pepsin	YYTGS	HYEGN
11	185	Chymosin	T	I
12	200 - 201	Pepsin	DG	GS
13	224 - 227	Pepsin	PTSA	STSS
14	240 - 243	Carbonic anhydrase B	NVGH	KRLF
15	256 - 257	Pepsin	PD	PD
16	275 - 283	Endothiapepsin	DFGPISTGSSSCF	VFQESYSSKKLCT
17	289 - 299	Pepsin	MDVPTSSGELW	MDIPPPTGPTW

COMPARISON OF THE RENIN MODEL WITH THE X-RAY STRUCTURE

Fig. 3 illustrates the structures of human recombinant and mouse submandibular renins defined by X-ray analysis (ref. 23,35). They have very similar three-dimensional structures, reflecting the high degree of sequence identity. As seen from the stereo view of the superimposed Cα plots of the renin structure (solid line) and the model in Fig. 4, the model shows a significant degree of overlap with

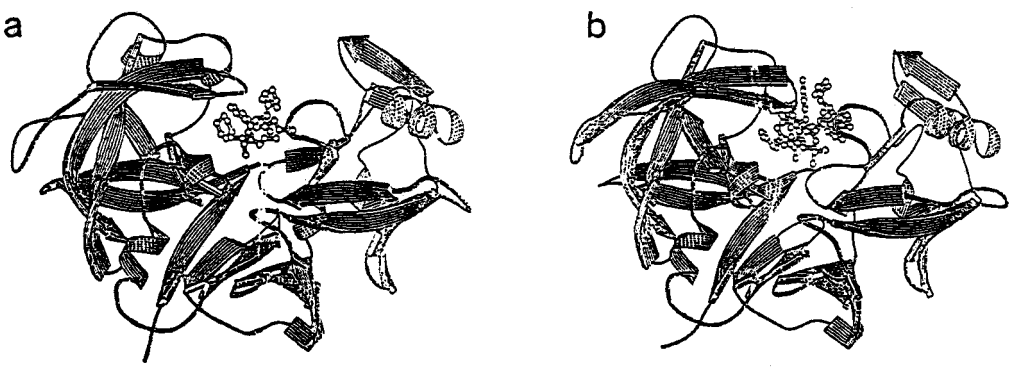

Fig 3. The crystal structures of (a) human recombinant and (b) mouse submandibular renins defined by X-ray analysis.

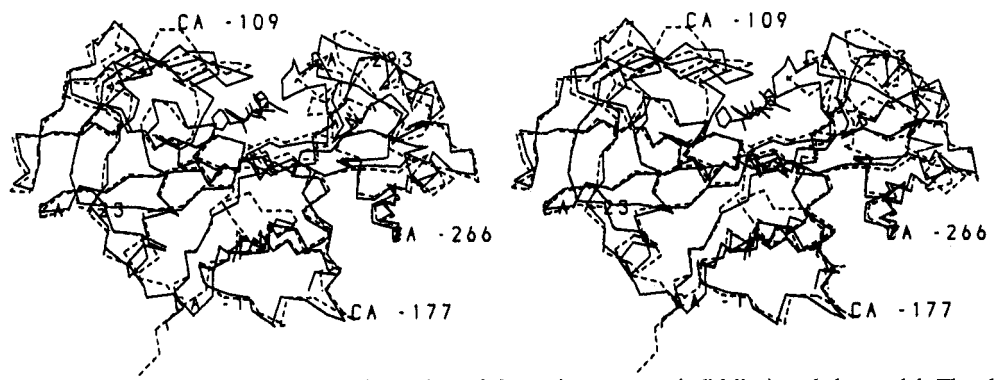

Fig 4. A stereo view of the superimposed Cα plots of the renin structure (solid line) and the model. The bound inhibitor is shown in full.

the structure. Using a 3.5Å cut-off for pairs of equivalent atoms, 280 Cα atom pairs superpose with a rmsd of 0.84Å. This value is lower in comparison with the rmsd's from the least-squares superposition of the known aspartic proteinase structures, listed in Table 1, with the renins. Thus, the model is closer to the structure of human renin than pepsin and chymosin, or indeed any other aspartic proteinase with the exception of mouse renin.

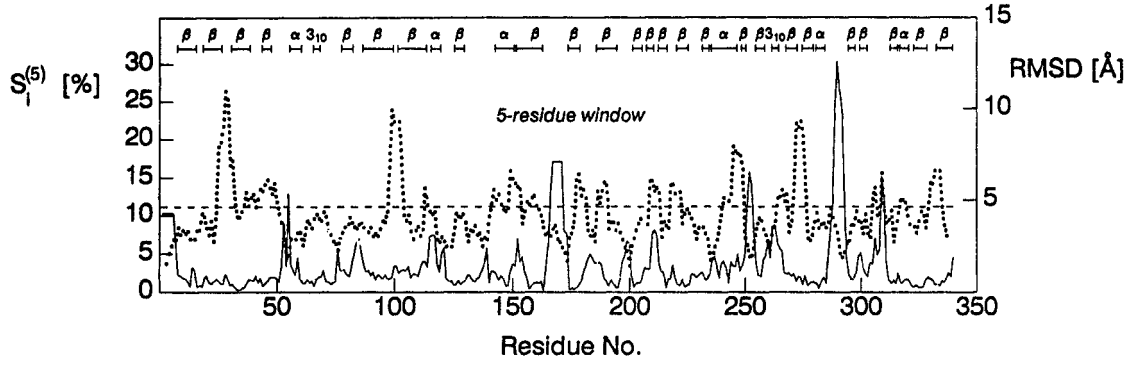

Fig 5. The 3-D profile showing the average percentage score [$S_i^{(n)}$] according to the appropriate physical environment, for residue i at the centre of a window of length n plotted as a function of residue position, for the comparison of the model with the crystal structure of human renin.

The framework regions of the model, which are derived mainly from the core portions of the parent structures, show very small deviations from the experimental structure. From the 3-D profile for the comparison of the model with the structure illustrated in Fig. 5, a good agreement can be observed in most regions of well-defined secondary structure. The major deviations occur mainly at the surface loops, particularly in the regions where there are considerable differences between the sequences (i.e. insertions and deletions) between the sequences of pepsin, chymosin and renin. There are also some minor deviations ($\sim$0.6Å) in regions at the periphery of the active site. These deviations become meaningful when one considers the fact that the model was based on the native structures of pepsin and chymosin whilst the renin structure has a bound inhibitor.

Although the structural similarity of the conserved central core containing the catalytic regions is very high amongst all aspartic proteinases, the active site cleft has a less open arrangement in renins than in the other aspartic proteinases. Many loops as well as the helix h_C (residues 224 - 236) belonging to the C-domain (residues 190-302, as defined by Sali et al., ref. 36) are significantly closer to the active site in the renin structures compared to those of endothiapepsin-inhibitor complexes. This is in accordance with previous structural studies (ref. 20,36-39) which have shown that in different crystalline environments the aspartic proteinases can undergo rigid body movements of a domain comprising residues 190-302 (pepsin numbering) relative to the rest of the molecule. Thus, there is a rotation of $\sim 5^{\circ}$ of this domain in the renin complexes with respect to the structures of pepsin and chymosin. This would account for many of the deviations of the model from the renin structure in the C-domain. The rotation is even higher in the cases of the fungal proteinases. Such differences are extremely difficult to model accurately since there is considerable variation in the magnitude and direction of the rigid body movements in the known structures of aspartic proteinases.

Apart from differences in the relative orientations of the rigid domains, the specificity pockets in renins differ from other aspartic proteinases as a consequence of changes in the position and composition of several well-defined loops and secondary structure elements. Unique to the renins is a *cis*-proline, Pro 111, which caps the helix h_{N2} and lines the subsites S_5 and S_3. This helix is nearer to the active site in renins than in other aspartic proteinases. Further, there is a proline-rich segment (residues 292-297) in the C-lobe, also unique to the renins among the aspartic proteinases, with Pro 294 and Pro 297 in a *cis* configuration. This loop is much closer to the active site in the

renin-inhibitor complexes and forms part of the subsites S1' and S3'. Such proline-rich structures probably provide an effective means of constructing well defined pockets from loops which would otherwise be more flexible. However, they pose a considerable problem for those who try to construct models based on them since there are not many representative structures available in the data base of known structures.

CONCLUSION

This analysis shows that an automated and rule-based approach to comparative modelling using COMPOSER gives a useful representation of the human renin structure. The renin model is closer to the real structure than the two mammalian aspartic proteinases from which it was constructed. This justifies retrospectively the use of the model for studying renin-angiotensin interactions and in designing inhibitors as antihypertensive agents.

The difficulties with modelling unusual loops, such as the proline-rich regions, underline the obvious problem with knowledge-based procedures; they depend on the extent of our present knowledge. It is interesting that similar proline-rich loops occur in other aspartic proteinases such as yeast proteinase A and cathepsin D; we have since been able to use the renin crystal structures in constructing models of these aspartic proteinases (ref. 40). Thus, as our knowledge increases, the next challenge becomes easier. Methods for the systematic generation of loop conformations are certainly important, but in this case, where cis-prolines are involved, are less likely to have selected the correct model.

Apart from the conformation of loop regions, the major challenge in the construction of precise models is the prediction of rigid body shifts of individual elements of secondary structure, occasionally leading to shifts of domains as in the aspartic proteinases. Part of the shift may be dependent on the solution or crystal environment, and cannot be predicted easily. However, much of the difference between homologues is independent of crystal packing and reflects the difference in amino acids at the interfaces. These should be predictable and we have derived useful relationships between sequence and packing of secondary structures, which should be helpful in modelling (B. Reddy and T. L. Blundell, unpublished results). We have yet to apply these to the model of renin.

REFERENCES

1. W.J. Browne, A.C.T. North, D.C. Phillips, K. Brew, T.C. Vanaman and R.L. Hill, *J. Mol. Biol.* **42**: 65-86 (1969).
2. A. Jones, *J. Appl. Crystallogr.* **11**: 268-272 (1978).
3. J. Greer, *J. Mol. Biol.* **153**: 1027-1042 (1981).
4. T.A. Jones and S. Thirup, *EMBO J.* **5**: 819-822 (1986).
5. T.L. Blundell, B.L. Sibanda, M.J.E. Sternberg and J.M. Thornton, *Nature* **326**: 347-352 (1987).
6. T.L. Blundell, D. Carney, S. Gardner, F. Hayes, B. Howlin, T. Hubbard, J. Overington, D.A. Singh, B.L. Sibanda and M.J. Sutcliffe, *Eur. J. Biochem.* **172**: 513-520 (1988).
7. M. Claessens, E.V. Cutsem, I. Lasters and S. Wodak, *Prot. Engg.* **2**: 335-345, (1989).
8. M.J. Sutcliffe, I. Haneef, D. Carney and T.L. Blundell, *Prot. Engg.* **1**: 377-384 (1987).
9. N. Srinivasan and T.L. Blundell, *Prot. Engg.* Submitted.
10. M.J. Sutcliffe, F.R.F. Hayes and T.L. Blundell, *Prot. Engg.* **1**, 385-392 (1987).
11. R. Tigerstedt and P.G. Bergman, *Scand. Arch. Physiol.* **8**: 223-271 (1898).
12. W.J. Greenlee, *Med. Res. Rev.* **10**: 173-236 (1990).
13. P.M. Hobart, M. Fogliano, B.A. O'Connor, I.M. Schaeffer and J.M. Chirgwin, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5026-5030 (1984).
14. T. Imai, H. Miyazaki, S. Hirose, H. Hori, T. Hayashi, R. Kageyama, H. Ohkubo, S. Nakanishi and K. Murakami, *Proc. Natl. Acad. Sci. U.S.A.* **80**: 7405-7409 (1983).
15. M. Newman, M. Safro, C. Frazao, G. Khan, A. Zdanov, I.J. Tickle, T.L. Blundell and N.S. Andreeva, *J. Mol. Biol.* **221**: 1295-1309 (1991).
16. T.L. Blundell, B.L. Sibanda and L.H. Pearl, *Nature* **304**: 273-275 (1983).
17. B.L. Sibanda, T.L. Blundell, P.M. Hobart, M. Fogliano, J.S. Bindra, B.W. Dominy and J.M. Chirgwin, *FEBS Lett.* **174**: 102-111 (1984).
18. K. Akahane, H. Umeyama, S. Nakagawa, I. Moriguchi, S. Hirose, K. Iizuka and K. Murakami, *Hypertension* **7**: 3-12 (1985).

19. W. Carlson, M. Karplus and E. Haber, *Hypertension* **7**: 13-20 (1985).
20. J.B. Cooper, G. Khan, G. Taylor, I.J. Tickle and T.L. Blundell, *J. Mol. Biol.* **214**: 199-222 (1990).
21. A.R. Sielecki, K. Hayakawa, M. Fujinaga, M.E.P. Murphy, M. Fraser, A.K. Muir, C.T. Carilli, J.A. Lewicki, J.D. Baxter and M.N.G. James, *Science* **241**: 1346-1351 (1989).
22. B. Veerapandian, J.B. Cooper, A. Sali and T.L. Blundell, *J. Mol. Biol.* **216**: 1017-1029 (1990).
23. V. Dhanaraj, C.G. Dealwis, C. Frazao, M. Badasso, B.L. Sibanda, I.J. Tickle, J.B. Cooper, H.P.C. Driessen, M. Newman, C. Aguilar, S.P. Wood, T.L. Blundell, P.M. Hobart, K.F. Geoghegan, M.J. Ammirati, D.E. Danley, B.A. O'Connor and D.J. Hoover, *Nature* **357**: 466-472 (1992).
24. J. Rahuel, J.P. Priestle and M.G. Grtter, *J. Struc. Biol.* **107**: 227-236 (1991).
25. A. Sali, J.P. Overington, M.S. Johnson and T.L. Blundell, *Trends Biochem. Sci.* **15**: 235-240 (1990).
26. C. Topham, A. McLeod, F. Eisenmenger, J.P. Overington, M.S. Johnson and T.L. Blundell, *J. Mol. Biol.* **229**, 194-220 (1993).
27. A. Sali and T.L. Blundell, *J. Mol. Biol.* **212**: 403-428 (1990).
28. Z.-Y. Zhu, A. Sali and T.L. Blundell, *Prot. Engg.* **5**: 43-51 (1992).
29. M.S. Johnson, M.J. Sutcliffe and T.L. Blundell, *J. Mol. Evol.* **30**: 43-59 (1990).
30. M.S. Johnson, A. Sali and T.L. Blundell, *Meth. Enzymol.* **183**: 670-690 (1990).
31. J.P. Overington, M.S. Johnson, A. Sali and T.L. Blundell, *Proc. Roy. Soc. Lond.* **B241**: 132-145 (1990).
32. J.P. Overington, D. Donnelly, M.S. Johnson, A. Sali and T.L. Blundell, *Prot. Sci.* **1**: 216-226 (1992).
33. F.C. Bernstein, T.F. Koetzle, J.B. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi and M.J. Tasumi, *J. Mol. Biol.* **112**: 535-542 (1977).
34. P. Strop, J. Sedlacek, J. Stys, J. Kaderabkova, I. Blaha, L. Pavlickova, J. Pohl, M. Fabry, V. Kostka, M. Newman, C. Frazao, A. Shearer, I.J. Tickle and T.L. Blundell, *Biochemistry* **29**: 9863-9871 (1990).
35. C.G. Dealwis, C. Frazao, M. Badasso, J.B. Cooper, I.J. Tickle, H. Driessen, T.L. Blundell, K. Murakami, H. Miyazaki, J. Sueiras-Diaz, D.M. Jones and M. Szelke, *J. Mol. Biol.* Submitted.
36. A. Sali, B. Veerapandian, J.B. Cooper, S.I. Foundling, D.J. Hoover and T.L. Blundell, *EMBO J.* **8**: 2179-2188 (1989).
37. A.R. Sielecki, A.A. Fedorov, A. Boodhoo, N.S. Andreeva and M.N.G. James, *J. Mol. Biol.* **214**: 143-170 (1990).
38. C. Abad-Zapatero, T.J. Rydel and J. Erickson, *Proteins* **8**: 62-81 (1990).
39. A. Sali, B. Veerapandian, J.B. Cooper, D.S. Moss, T. Hofmann and T.L. Blundell, *Proteins* **12**: 158-170 (1992).
40. P.E. Scarborough, K. Guruprasad, C. Topham, G.R. Richo, G.E. Conner, T.L. Blundell and B.M. Dunn, *Protein Science* **2**: 264-276 (1993).