Journal of Computer-Aided Molecular Design, 4 (1990) 369–379
ESCOM

J-CAMD 098

Comparison of the X-ray structure of baboon α-lactalbumin and the tertiary predicted computer models of human α-lactalbumin

Barry Robson and Eric Platt

Proteus Molecular Design Limited, 48 Stockport Road, Marple, Cheshire SK6 6AB, U.K.

Received 24 January 1990
Accepted 26 June 1990

Key words: α-Lactalbumin; Molecular modelling; X-ray

SUMMARY

The previously predicted structures of human α-lactalbumin by homology with hen egg white lysozyme by an automatic method, after the alignment stage, are compared to the X-ray determined structure of baboon α-lactalbumin. The root mean square by rotation method (RMSR) deviations for 122 C-α atoms between the two models and the X-ray structure are 2.0 Å and 2.3 Å. The RMSR deviations for all atoms, except for differences in human and baboon sequences, are 2.8 Å and 3.1 Å. If the flexible C-terminus (residues 112-122) are removed then these RMSR deviations are reduced to 2.4 Å and 2.3 Å respectively. These results are consistent with the fact that the RMSR deviation between the human and baboon X-ray structures increases from residue 112 onwards and is conformationally flexible.

INTRODUCTION

We previously reported in this journal the automatic modelling of human α-lactalbumin [1]. Some of the coordinates generated in modelling were filed at the same time [1]. The paper reported 2 distinct (if by no means grossly dissimilar) models. Although it was tempting simply to conclude that the merits of the two structures could not be distinguished within the state of the art of modelling, we carefully inspected the two structures and their dynamics, and made the proposal that the models were likely to represent 2 alternative conformations of equivalent energy in the presence of solvent. Subsequently, Acharya et al. [2] obtained the experimental X-ray crystallographic structure of baboon α-lactalbumin at 1.7 Å. The experimental coordinates have recently become available for analysis. In considering by comparison with the crystal structure the extent to which the results obtained are good or useful, several conceptual difficulties come to light which are well known in many aspects of protein modelling but usually not well highlighted or discussed. In particular, the experimental approach may imply much higher uncertainty than normally stated, if

comparison could be made with the free solution form. The opportunity is taken to analyze briefly these conceptual difficulties, using the lysozyme family as a particular example.

Although this study [1] was carried out some time ago and does not by any means represent the current state of our art, essentially the same picture has been obtained (see below) by our more recent methods. The earlier data publically described are used here to demonstrate objectivity. The particular interest is that it is still rare to have fairly detailed predictive calculations made not only before the publication of the experimental results, but also before completion of the experimental studies, as in this case. We also note that the study was by no means exhaustive by traditional standards: it was in essence a control study to assess whether a simple 'brute force' might do just as well as a complex approach involving human experts. It may as such represent the first simple non-interactive, non-graphic, attempt to carry out such modelling. Automatic fitting to the homologous protein hen egg white lysozyme was followed by molecular dynamics alternately at very high and low temperatures. The rationale was that this would help anneal the structure despite the presence of many local potential energy minima. Because of the admitted simplicity and hence deficiency in this approach compared with more recent modelling methods, it is perhaps of even greater interest that the comparison is between a simple, cheap, fast and automatic computational procedure, and an experimental study which represents the culmination of considerable effort and of a series of crystallographic investigations spanning many years.

In comparing model and experiment, it has become customary to employ two major methods. The first (RMSR – root mean square by rotation method) measures the root mean square deviation of atomic positions and is the classic method for comparing structures used by crystallographers. It requires superimposing the two structures by rotation and translation and minimizing the deviation. The second (RMSD – root mean square deviation of interatomic distances) is much used by modellers and does not require superimposition. It is better suited to molecules or parts of molecules that differ more drastically in conformation.

RESULTS AND DISCUSSION

Many authors choose either one or the other of the above root mean square measures but in the interest of completeness both measures are reported here, in the format (RMSD/RMSR). Since the modelling led to the proposal of 2 presumed co-existing conformational states of α-lactalbumin, we also speak below of such results for State 1 and State 2. The measures RMSD/RMSR were 1.6/2.0 Å for State 1 and 1.7/2.2 Å for State 2 with respect to Brookhaven entry 1ALC [3], counting C-α backbone atoms only (see Fig. 1). Coordinates were filed only for the slightly lower energy State 2. It is emphasised that the two states are by no means separate models in the normal sense: it was noted [1] that there is rather negligible difference in the conformational energies of the two model structures when any reasonable water model is included. This has been reconfirmed in subsequent studies, though the precise difference in energy depends on the water model and presence of counterions – it is most generally of the order of one kcal/mol. The filed coordinates thus represent the conformer then expected to be slightly more prevalent in solution, though the relative balance of the two could of course be significantly perturbed in the crystal, and indeed could be sensitively dependent on variations of ionic composition between crystals. Be this as it may, the two proposed states are in general rather similar.

The most significant difference between States 1 and 2 is in the region near the C-terminus. Be-

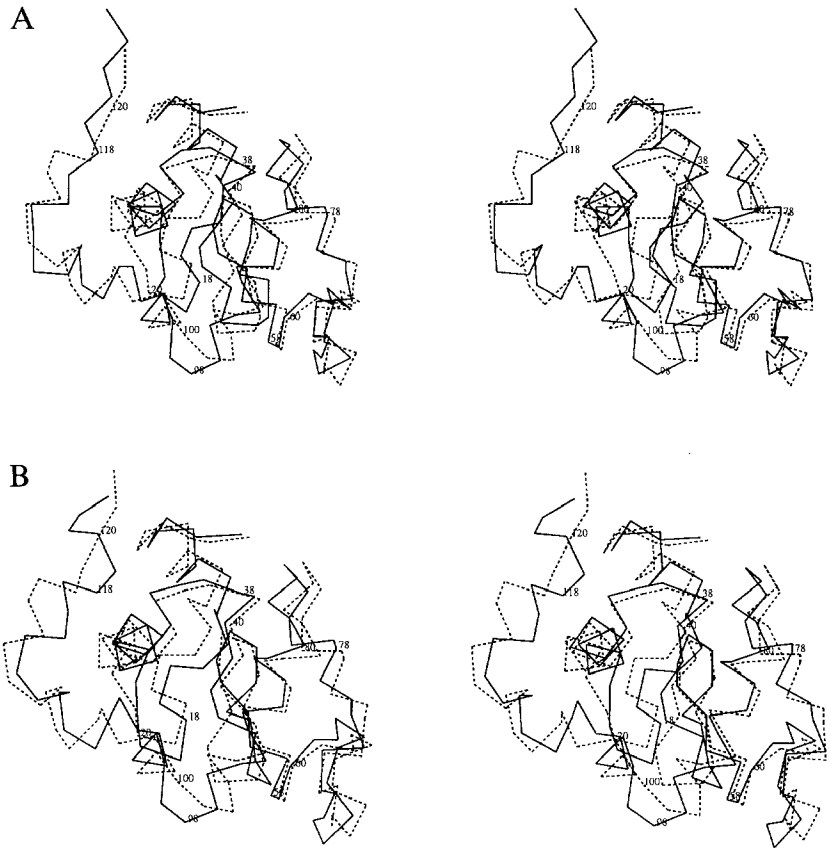

Fig. 1. Optimal superimposition of the C-α coordinates of each of the 2 models with those of the X-ray structure [2]. Solid lines for a model and dotted for the X-ray structure, (A) State 1 with 1 ALC, (B) State 2 with 1 ALC.

cause of our particular interest in calculating the complex repertoire of conformational behaviour of globular proteins which we believe must be considered in evaluating free energies, it is of particular interest to us that this conformational variation predicted in the C-terminal region was indeed experimentally observed. The C-terminal fragment (residues 112–123) is experimentally rather less well defined in the raw electron density map than that of the rest of the structure, with variation more-or-less increasing down this segment to the C-terminal residue (Walker, personal communication). This is consistent with the major distinction between States 1 and 2. In fairness, it must be said that this region has provided difficulties not only for crystallographers [2,4] but also for modellers [1,5,6], though to our knowledge our model is the only one to express clearly the view that the problem arises because alternative conformational details are of equivalent stability and co-exist. Warme et al. [6] predicted 3 different conformations of nearly equivalent energy for the C-terminus region, but these were much more similar to one another than are States 1 and 2. Confusion may have arisen in the various modelling attempts because of the ambiguity in identifying equivalent residues in the alignment (see Table 1), which in turn disagrees with that deduced from the X-ray structure (see Ref. 2, Table 1). However, States 1 and 2 are the only states

TABLE 1
ALIGNMENTS OF THE C-TERMINUS, (A) USED IN MODELLING [1], (B) USED BY WARME ET AL. [6], (C) SUGGESTED BY X-RAY STRUCTURE [2] AND (D) SUGGESTED BY COMPARISON OF 1ALC AND 2LYZ φ/ψ ANGLES (SEQUENCE A IS HEN EGG WHITE LYSOZYME, SEQUENCE B IS HUMAN α-LACTALBUMIN)

```
a)
   115                        121                        127     128
A: CYS-Lys-Gly-Thr-Asp-Val-GLU-Ala-TRP-Ile-Arg-Gly-CYS     Arg-LEU
B: CYS-Thr-Glu-Lys-Leu     GLU-Gln-TRP-Leu          CYS-Glu-Lys-LEU
   111                        116                        120     122

b)
   115          118                                     127     128
A: CYS-Lys-Gly-Thr-Asp-Val-GLU-Ala-TRP-Ile-Arg-Gly-CYS     Arg-LEU
B: CYS-Thr-     Glu-Lys-Leu-GLU-Gln-TRP-Leu          CYS-Glu-Lys-LEU
   111          113                                     120     122

c)
   115                                 123              127     128
A: CYS-Lys-Gly-Thr-Asp-Val-GLU-Ala-TRP-Ile-Arg-Gly-CYS     Arg-LEU
B: CYS-Thr-Glu-Lys-Leu-Glu-Gln-     TRP-Leu          CYS-Glu-Lys-LEU
   111                              118              120     122

d)
   115                           121                        127     128
A: CYS  -Lys-  Gly-Thr-Asp-Val-GLU-Ala-TRP-Ile-Arg-Gly-CYS     Arg-LEU
B: CYS-Thr-Glu-Lys-         Leu-GLU-Gln-TRP-Leu          CYS-Glu-Lys-LEU
   111                        116                        120     122
```

of low energy so far identified by minimization and dynamics, and on that basis predict a conformational dynamism irrespective of alignment considerations. Although Acharya et al. [2] state in the synopsis of their paper that 'The C-terminus appears to be rather flexible in α-lactalbumin compared to lysozyme', their detailed experimental analysis of the nature of the deviation has not yet been published (see, however, Addendum). The authors elsewhere state that 'The largest shifts occurred in the C-terminal region, which is flexible, and in some of the loop regions that are exposed. Estimating the accuracy of the model coordinates is difficult. In general, we believe the $R = 0.220$ model to be mostly rather accurate'. No coordinates at all could be assigned to Glu^{123} (Leu^{123} in human α-lactalbumin). With these difficulties in mind, the prediction of a variation in the C-terminal region was both unusual as a prediction and with hindsight appears qualitatively correct. If comparisons exclude residues 112–122, the RMSD/RMSR values improve markedly (see Table 2) and as a model State 2 becomes that in closest agreement with the reported X-ray structure.

Since modelling used a protein of known homology (hen egg white lysozyme) it is not too surprising that the prediction of backbone conformation is sound, though the assumption that conformational similarity follows sequence homology is an aspect of the prediction and was not assumed from the outset. Viral and bird lysozymes are distantly related, but show important differences in conformation. There are also potential difficulties even in the case of two closely re-

TABLE 2
COMPARISON OF MODEL RMSD AND RMSR VALUES WITH RESPECT TO THE CORRESPONDING ATOMS OF THE X-RAY STRUCTURE

State/X-ray	Residue numbers	C-α RMSD	C-α RMSR	All-atom RMSD	All-atom RMSR
1/1 ALC	1–122	1.6	2.0	2.3	2.8
2/1 ALC	1–122	1.7	2.2	2.2	3.1
2 LYZ/1 ALC	1–122	1.2	1.5	–	–
1/2 LYZ	1–122	1.2	1.6	–	–
2/2 LYZ	1–122	1.3	1.9	–	–
1/1 ALC	1–111	1.5	1.8	1.8	2.4
2/1 ALC	1–111	1.4	1.7	1.8	2.3
2 LYZ/1 ALC	1–111	1.0	1.3	–	–
1/2 LYZ	1–111	1.1	1.4	–	–
2/2 LYZ	1–111	0.9	1.2	–	–
1/1 ALC	79–88	0.6	0.6	0.9	1.3
2/1 ALC	79–88	0.6	0.6	0.9	1.2
2 LYZ/1 ALC	79–88	0.6	0.6	–	–
1/2 LYZ	79–88	0.4	0.4	–	–
2/2 LYZ	79–88	0.5	0.5	–	–

lated sequences. For example, one of the main reasons for the current interest in applying 'spin-glass' theory to proteins is the fact that small differences in chemistry can have, in at least occasional circumstances, major ramifying effects on conformation. Be that as it may, in general the secondary structures represented by the two lowest States 1 and 2 agree fairly well with those of the X-ray structure. The greatest differences occur at the insertion/deletion sites of the lysozyme/α-

TABLE 3
COMPARISON OF PREDICTED AND X-RAY SECONDARY STRUCTURES

Secondary structure	X-ray	State 1	State 2	2 LYZ[a]
α-Helix	5–11	5–12	5–12	5–13
α-Helix	23–34	23–32	23–34	23–34
α-Helix	86–99	86–96	86–96	86–97
α-Helix	105–109	105–109	105–109	105–110
3_{10}-Helix	12–16	–	–	
3_{10}-Helix	17–21	–	–	
3_{10}-Helix	76–82	77–82	77–82	77–80
3_{10}-Helix	101–104	101–104	101–104	100–103
3_{10}-Helix	115–119	–	115–119	115–118
β-Strand	40–43	40–44	40–43	40–44
β-Strand	47–50	47–50	47–50	47–50

[a] α-Lactalbumin numbering system used.

lactalbumin alignment used in the modelling (see Table 3). State 2 reflects the features assigned by the crystallographers, better than does State 1. Of course, a great deal of modelling from first principles must take place in the insertion/deletion regions, whether interactive graphics or pure simulation is involved. It should be recalled from the above, however, that loop regions seem less well defined in the experimental result, and as it happens (though by no means unexpectedly) these include the insertion/deletion sites. α-Helices and β-strands are predicted better than 3_{10} helices.

Agreement on an all-atom basis is expected to be less satisfactory not simply because of even greater demands on modelling from basic physicochemical principles, but because there are significant influences on conformation of side chains due to crystal formation and associated solvent effects. The corresponding RMSD/RMSR values of all the non-hydrogen atoms, excepting gaps due to the 8 differences in sequence between baboon and human α-lactalbumin as determined by the X-ray structure [2], are for State 1 2.3/2.8 Å and for State 2 2.2/3.1 Å. These results are more than adequate to support the view that when there is close homology between a protein with unknown structure and one of known X-ray structure, automatic computer modelling (rather than computer graphics modelling by a human expert) can be useful. However, as regards our reported α-lactalbumin modelling [1] there is a recognised need for further molecular dynamics time to anneal the side chains. Also, a better strategy assigning starting conformations of the side chains would probably diminish the need for a great deal of computer time. Subsequent longer studies have been carried out. They confirm the above criticisms by leading to more convincing side-chain interactions and in general there is a higher degree of classical intramolecular hydrogen bonding involving side chains, consistent with observation. Although less apparent in the reported structure, it is of scientific interest to note that with powerful optimization techniques and ample computer time the customary difficulty in forming hydrogen-bonding networks switches to one of excessive intramolecular hydrogen bonding in protein modelling in vacuo, and solvent has to be added to produce more realistic conformations of major flexible loops in particular. A detailed comparison of these aspects for several proteins has been carried out and in our view deserves deeper analysis which will be made and reported elsewhere. We note here briefly the interesting preliminary observation that regions of the surface containing charged groups appear to be responsible for some two-thirds of the deviation from experimental observation, and that addition not only of water but counterions to the simulations can indeed be shown to be important.

The complex role of the crystal lattice and of solvent and counterions thus provides protein modellers with several valid excuses for deficiencies in agreement with experiment! The authors even suspected that, while there is considerable room for improvement in the model reported here, any predictions of any protein structure giving dramatically superior RMS agreement with the crystal structure, yet which do not model the crystal lattice and the precise aqueous ionic medium, should be viewed with some suspicion. It therefore came as a pleasant surprise to find that functionally important regions of the earlier in vacuo model [1] showed reasonable agreement with experiment even at the level of atomic detail and even with charged side chains associated with them. The obvious example both from the view of functional significance and ion binding is the case of the calcium binding loop [1]. The RMSD/RMSR values of the loop are given in Table 2. Even though the calcium ion was not present until the very end of the simulations, the predicted structure of this loop was in very good agreement with the X-ray structure, with the RMSR/RMSD values being less than the values of the overall structure.

Another feature worth commenting on in detail at this stage is the conformation of the disulph-

ide bridges. This is largely because subsequent analysis has suggested to us that a major feature in the transition between States 1 and 2 is a transition of CYS6-CYS120 from a left-hand to a right-hand conformation. Subject to the reservations below and assuming that the conformational transition of the C-terminus in free solution is first validated, this hypothesis would be of interest to test by judiciously chosen experimental methods in different solvents. The prediction is by no means a trivial one since details of disulphide bridging conformations have varied between models [1,5,6], and could be a sensitive means of testing the relative merits of both modelling and potential functions. State 2 was consistent with the spectroscopic optical activity [7] indicating that for the dominant conformer there are 3 right-handed and 1 left-handed C-S-S-C disulphide connections. This is in turn consistent with the X-ray structure. Table 4 compares the torsion angles of the disulphide bridge of the two states with the X-ray structures. The agreement is quite good for all disulphide bridges; unfortunately, it is for the very case of interest, CYS6-CYS120, that the results are worst. Note that in particular x_2' and x_1' of CYS6...CYS120 would be expected to be heavily influenced by (and in turn influence) transitions in the flexible C-terminus.

In overall summary it may be said that if State 1 and State 2 are regarded as distinct models, then State 2 has greater merit as more similar to the X-ray structure in terms of qualitative features – even though the overall RMSD and RMSR values are greater for State 2 than for State 1. State 2 is the lowest energy state in vacuo (and of similar energy in water). On the positive side and irrespective of whether States 1 and 2 are considered separately or as a single dynamic structure, they represent a tolerable prediction attempt for a relatively early automatic procedure. Indeed, for many purposes encountered in applied protein technology, we would make the controversial point that even these preliminary predicted structures would serve just as well as the experimental

TABLE 4
COMPARISON OF THE DISULPHIDE BRIDGE TORSION ANGLES

Disulphide bridge		x_1	x_2	x_3	x_2'	x_1'
6–120	X-ray	−96	−60	103	123	164
	State 1	−56	29	−93	−105	−172
	State 2	−79	−69	144	−35	−88
6–127	2 LYZ	−90	−21	−99	−52	−40
28–111	X-ray	171	−82	−74	−86	−38
	State 1	−168	−98	−100	−84	−46
	State 2	158	−106	−109	−46	−75
30–115	2 LYZ	−180	−85	−103	−74	−67
61–77	X-ray	53	92	97	−65	−56
	State 1	72	93	72	−24	−88
	State 2	80	112	53	−19	−81
64–80	2 LYZ	67	79	94	−45	−70
73–91	X-ray	−64	173	71	64	−179
	State 1	−72	−169	78	40	−169
	State 2	−78	−166	85	31	−172
76–94	2 LYZ	−85	146	79	43	−177

X-ray crystallographic study and represent a relatively negligible effort and cost. While of course we continue to support the merit of *combined* computational and experimental techniques, the importance of such considerations lies in the fact that relatively few proteins can yet be subjected to X-ray crystallographic analysis.

On the other hand, the model is typical of most detailed protein models in that it borrows heavily from the known structure of hen egg white lysozyme, so putting great emphasis on the value of obtaining a crystallographic structure. One could argue as follows: if the model α-lactalbumin structure could serve for some purposes as well as the experimental result, then the original unmodified structure of lysozyme might have served just as well also (see Tables 2–4). In such a case a prediction by modelling is of purely academic interest, as a step in the development of modelling technologies. Prediction of some unexpected feature (as in the case of a degree of pronounced conformational dynamism) is then of even greater importance. The counter argument exists that one should not be penalised for finding a correct solution just because it lies close at hand. Formally, this is a complex problem in completeness, bound up with the statistical mechanical problem of ergodicity and in practice the need to study divergence and convergence of simulations over long simulation times.

In assessing results and in considering how improvements might be obtained in modelling one runs into a difficulty which is, in our opinion, inadequately stressed even if inherently obvious. The structure reported by a protein crystallographer is not a universal constant: it is not perfectly refined, assumption-free, model-free, nor error-free, nor is it invariant with environment. For some proteins, 2.5–3.0 Å might be the absolute experimental error, if comparison with free-solution forms could be made. In particular the choice of entry 2LYZ from the Brookhaven data base was a somewhat arbitrary standard choice as there are a large number of lysozyme entries, all of which differ in detail. Whereas the agreement between 2LYZ lysozyme and experimental and modelled α-lactalbumins is about 1.6 Å (C-α atoms, crystallographer's measure), the model and experimental α-lactalbumins differ by some 2.0 Å. However, the latter, our model proposes, is a consequence of a genuine flexibility in the C-terminal portion. The agreement for the rest of the backbone is about 1.7 Å. Beyond this specific consideration, there is the expectation that inherent flexibility of protein and effect of environment and chemical differences upon the relative stabilities of different forms can even further complicate the issue. This is reflected in part by the differences between the three-dimensional structures of lysozymes on the Brookhaven data base, which are particularly numerous and provide a rare opportunity to study the problem. The variation in crystallographer's fit measures based on the C-α carbons of all interpretations of the electron-density data recorded on the Brookhaven data base, between different mutant and related species of lysozymes, and between different crystal forms, is about 0.4 Å (chicken and turkey lysozyme, which one might expect to be rather similar, differ by 0.8 Å on the same C-α basis). This is the same scale as the deviations in the measures of fit. One could in consequence pick a specific experimental lysozyme structure to serve directly as a better model of α-lactalbumin. Similarly, on removing a lysozyme from its crystal form, in simulation and almost certainly in reality too, changes in RMS agreement (on the same basis as above) occur which are again of the same scale as deviations in the above measures of fit (sometimes higher), e.g. 0.7–1.0 Å, depending on authors and potential functions. Workers who repeatedly claim smaller variations in the technique of 'energy refinement' when using previous results for the same protein as a starting point are most often using only local gradient methods or dynamics which trap the structure in kinetically facile local

minima: such apparently pleasing results are particularly obviously artifacts when modelling is carried out in vacuo, since the real protein in vacuo probably adopts a quite different local structure even in the extremely unlikely event that a comparable globular structure was stable in a vacuum at all. Finally, there is ample indication from NMR spectroscopy of protein solutions, from studies on diffusing in substrates and substrate analogues into protein crystals, and from spectroscopic studies of protein multimer formation, that conformational changes potentially much larger than those discussed above can take place due to environmental changes and new intermolecular contacts.

CONCLUSION

Can modelling be improved to a point where the above considerations pose a theoretical limit, and is modelling of close homologues then useful at all? One important justification for modelling at all is the 'imposition of reasonableness' argument. Irrespective of the ability to assess correctness one has to impose the quality of *reasonableness* on the model. In energy refinement of an X-ray structure, for example, it is sometimes considered desirable that there are no van der Waals overlaps or other stereochemically unlikely features for reasons of convenience quite distinct from improvement in R factor. For most purposes there would be no point in simply taking the unmodified lysozyme structure model for α-lactalbumin even if it were equally good from an RMS point of view. One still has to do the modelling to allow for insertions/deletions and side-chain substitutions, and ideally one should do that objectively and reproducibly, by an automatic modelling method as done for the present study. For many purposes in pharmacology and biotechnology, we would also prefer that any changes introduced in modelling tend towards what is happening in the solution, rather than crystalline, case. One must also impose realistic dynamics as well as statics. For example, neither direct use of the lysozyme data without modification, nor the imposition of the simplest stereochemical considerations in homology modelling, could themselves predict a dynamic feature of the α-lactalbumin structure. At this level, the distinction between simple reasonableness and genuine agreement with experiment is blurred and shown to be something of an illusion deriving simply from imperfect knowledge and technologies. Current studies on α-lactalbumin are being performed to identify the theoretical limits rather more precisely. These studies are largely represented by better refined potential functions and more detailed water models, as well as by faster dynamics covering longer time scales. Though the study is not yet complete, improvement in agreement as opposed to simple reasonableness does seem possible. One aspect of this study that was surprising is that it turns out to be better (at least in the case of α-lactalbumin) to follow many protein modellers and to transfer directly coordinates from the homologous protein of known conformation, rather than perform an RMS fitting procedure. It should be noted, however, that our findings are the reverse of this when the sequence and conformation homology is weaker, and RMS fitting has been found to be better even for some proteins which are 40% homologous in sequence. Analysis of the α-lactalbumin conformations so far suggests that side chains have conformations more comparable with those observed: 65% of side chains have χ_1 and χ_2 angles in the same rotameric state (e.g. $+/-$ 60 degrees) as observed. It is of some concern that this is actually a drop of about 2% from a somewhat earlier point in the simulation. However, this phenomenon of losing ground in order to make more ground is by no means unique in our modelling experience, and may simply be analogous to the case of the Rubik Cube in which a par-

tial solution must be disassembled in order to reach a fuller solution. It seemed less surprising that in a simulation of myoglobin starting from the known helices but flexible side chains, the arbitrary starting structure (defined by assigning extended regions between helices) blew open to a structure of worse than 10 Å RMSD backbone agreement with experimental, before falling to a more satisfactory 3.0 Å RMSD.

We thus believe that such studies as these are important for three reasons: they demonstrate that convenient automatic modelling is feasible, they highlight the theoretical limits and the kind of further studies which are required simply to *evaluate* objectively the degree of success, and they signpost the way to best approach even more challenging predictive studies in which homology is, at best, confined to the recognition of certain subdomains.

ADDENDUM

On October 5th we were kindly accepted to Dr David Stuart's laboratory to discuss the latest experimental results regarding the α-lactalbumins. In particular we thank Mr Ren Jin Chen for data not yet publically available, regarding the recently completed human α-lactalbumin structure. As an extension to our policy of carrying out predictions of protein structure prior to the availability of the X-ray crystallographic results, it is interesting to compare separately the above conclusions, based on a single set of experimental coordinates, with those further data subsequently available. The further data represent both a deeper analysis of the dynamics in the crystal and exploration of conformational variation between α-lactalbumins. In brief, the human and baboon forms are as expected very similar, both in terms of conformation and of conformational flexibility. Principal differences in the xyz coordinates of C-α carbon atoms are at residues 45-47 and 111-123, both tolerably consistent with regions of flexibility identified in our model although we directed attention only to the C-terminal fragment. The flexibility predicted by our model at residues 45–47 is also the largest deviation excepting the C-terminal segment, and has an RMS deviation of 3-4 Å compared with 2 Å for the experimental variation between human and baboon.

Consideration of the C-terminal segment is best done by considering that segment in two parts. The experimental deviation between human and baboon is about 1 Å for residues 111-119 and between about 2.5 and 4 Å for residues 121-123. This is significantly less than the variation predicted in our model, which although showing a roughly similar form of variation along the sequence predicts 5 Å for 111-119 and 9-11 Å for the last three residues. If comparison between baboon and human α-lactalbumins is a meaningful measure of variability, then our model is qualitatively correct but grossly overestimates the variability at the very end of the chain. Since using such comparison is subject to criticism, it is important to note that the B-factor is in excess of 25 for residues from 106 on, and generally 2-3 times the B-factor for the rest of the chain, both for human and baboon α-lactalbumins, with peaks around 109–116 and 120–123. With hindsight, we might have considered it desirable to distinguish the C-terminal sequences 111-119 and 120–123 since 120 is anchored to the rest of the protein by a disulphide bridge. This remains a valid criticism but in fact analysis of variation between α-lactalbumins and inspection of B-factors give a slightly different impression, which suggests that there is no single, convenient residue at which to divide the behaviour of the C-terminal segment. From the viewpoint of B-factors the α-carbon atom of residue 120 appears to display almost as much motion as 121-123, and 120-123 provide the largest B-factors in excess of about 50 both for human and baboon.

Somewhat unexpectedly, B-factors are large not only around residues 45-47 and 111-123, but also around 13-21 and 57-58 where there are no significant differences between the spatial positions of residues in the human and baboon forms. We note that there is a relation but certainly not equivalence between conformational variation in time and between species variants and crystal forms. The new data for the lactalbumins taken into account both with modelling and data for all the lysozymes provide a rich ground for study of this relation, which will be reported elsewhere.

REFERENCES

1 Robson, B. and Platt, E., J. Comput.-Aided Mol. Design, 1 (1987) 17. By arrangement with this Journal, coordinates were filed at the Center for Molecular Design at Washington University. For details see e.g. Vol 3 No 3, advertising information section.
2 Acharya, K.R., Stuart, D.I., Walker, N.P.C., Lewis, M. and Phillips, D.C., J. Mol. Biol., 208 (1989) 99.
3 Berstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shzmanouchz, T. and Tasumz, M., J. Mol. Biol., 112 (1977) 535.
4 Smith, S.G., Lewis, M., Aschaffenburg, R., Fenna, R.E., Wilson, I.A., Sundaralingam, M., Stuart, D.I. and Phillips, D.C., Biochem. J., 242 (1987) 353.
5 Browne, W.J., North, A.C.T., Phillips, D.C., Brew, K., Vanaman, T.C. and Hill, R.L., J. Mol. Biol., 42 (1969) 65.
6 Warme, P.K., Momany, F.A., Rumball, S.V., Tuttle, R.W. and Scheraga, H.A., Biochemistry, 13 (1974) 768.
7 Cowburn, D.A., Brew, K. and Gratzer, W.B., Biochemistry, 11 (1972) 1228.