Journal of Computer-Aided Molecular Design, 1 (1987) 17–22
ESCOM
J-CAMD 007

Modelling of α-lactalbumin from the known structure of hen egg white lysozyme using molecular dynamics

Barry Robson and Eric Platt

Epsitron Peptide and Protein Engineering Unit, Department of Biochemistry and Molecular Biology, Stopford Building, University of Manchester, Oxford Road, Manchester M13 9PT, U.K.

Received 17 February 1987
Revised 27 February 1987
Accepted 2 March 1987

Key words: Protein homology; Tertiary prediction; Energy Minimisation; Calcium-binding Loop

SUMMARY

The prediction of protein conformation by homology is being widely pursued using interactive computer graphics. However, there have been a limited number of energy minimization and/or molecular dynamics studies for such predictions. This paper reports one such study on α-lactalbumin, a system that can be tested as the X-ray crystal structure has recently been determined. The differences in structure of the Ca^{2+} binding loop reported for the holo-protein (Stuart et al., Nature, 324 [1986] 84–87) and that predicted for the apo-protein could be attributed to the presence or absence of the Ca^{2+} ion.

INTRODUCTION

X'ray crystallography is the recognized standard tool for determining the three-dimensional structure of a globular protein in atomic detail, but it cannot be described as routine. Following attempts over several years [1,2] the X-ray crystal structure of an α-lactalbumin has recently been determined. Details of the Ca^{2+} binding loop of baboon α-lactalbumin have been reported [3]. Without prior knowledge of this structure, human α-lactalbumin has been computer modelled by analogy with hen white lysozyme [4,5] the three-dimensional structure of which has long been determined and, in fact, represents the pioneering application of X-ray crystallography to an enzyme [6].

Because of these difficulties, calculation of protein conformation would be a valuable substitute for X-ray crystallography, or a powerful aid where the interpretation of the raw diffraction data presents difficulties. Despite the notorious problem of calculating protein conformation from the amino acid residue sequence alone, i.e. from the chemical formula which is fairly easily obtained, this approach is generally considered more viable when the three-dimensional structure of a related protein is used as a starting point. Hence previous attempts have been made to calculate α-lactalbumin conformation from that of hen egg white lysozyme [7,8].

Such a study is still far from trival since the proteins differ in some two-thirds of their amino acid residues, after allowing for several insertion/deletions between the two chains. Modelling in atomic detail a protein against a relative of this *weak* degree of relatedness remains an unproven approach, where the answer is not known in advance. Thus this report is timely as only a few specific details of the recent structural studies on holo α-lactalbumin have been reported [3]. Previous modelling of α-lactalbumin may have a subjective element in that although knowledge of interatomic interactions was used to varying extents, a simulation based on fundamental physicochemical principles applied simultaneously to the entire molecule was not employed. This is not surprising, since the most complete technique of molecular dynamics [9,10] which applies Newton's laws of motion to all atoms of the molecule has only recently been applied to proteins and remains computationally expensive.

METHODS

As a first step, the backbone of the computer model for residues 1–115 of human α-lactalbumin was fitted to the homologous sections of hen egg white lysozyme [4] shown in Table 1. The relevant backbone coordinates were taken from the X-ray structure designated RS5D by Diamond [11]. This RMS-fitting procedure has been described previously [5] for fitting a computer model of avian pancreatic polypeptide to the X-ray coordinates. The X-ray determined torsion angles of the homologous sections were used as starting values and each section was fitted individually before the overall fit was performed. Both procedures save computer time. An RMSD (root mean square deviation between interatomic distances in the model and in lysozyme) deviation less than 0.5 Å. RMS was achieved. Due to the general lack of homology in the region of residues 116–123, this part of the backbone was modelled automatically by closing the disulphide bridge between Cys-6 and Cys-120, using the X-ray torsion angles as starting values.

At this stage, the side chains of the amino acids were added. For computational economy, the starting values of the torsion angles were chosen as follows. If the sidechains in both proteins are identical, then the X-ray torsion angles of lysozyme were used. If they are not, then the angles were chosen so that the atoms of the model overlay as much of the lysozyme sidechains as possible, e.g. when Arg-45 is replaced by Glu-43 in α-lactalbumin the $\chi 1$ and $\chi 2$ angles of Arg-45 were used for these angles of Glu-43. Any sidechains torsion angles not assigned above, were set in one of the possible sidechain conformations observed in X-ray structures [12], e.g. when Gly-102 in lysozyme was replaced by Ile-98 in α-lactalbumin, the $\chi 1$ torsion angle of Ile-98 was assigned the g^+ conformation, whilst $\chi 2$ was assigned the t conformation. This combination of $\chi 1$ and $\chi 2$ torsion angles of an Ile sidechain accounts for 46% of the observed Ile side chain conformations [12]. This computer model was then energy-minimised using the LUCIFER (Logical Use of Conformational Information and Fast Energy Routines) program [5] using the SIMPLEX and GLOBEX procedures [13] with the backbone conformation fixed.

Finally the whole computer model was subjected to a momentum-directed minimisation (MDM) technique, which is similar to the stochastic and constraint molecular dynamics approaches [14, 15] which rescale velocities and eliminate fast mode vibrations respectively. In this approach, the velocities of each atom is rescaled at every iteration of the solution of the recursive Newtonian equation [16] so that the total kinetic energy is consistent with a constant temperature. A final temperature of 4K is chosen to obtain a more or less unique conformation. Higher temper-

TABLE I

HOMOLOGY ALIGNMENT OF THE HUMAN α -LACTALBUMIN (SEQUENCE B) WITH HEN EGG WHITE LYSOZYME (SEQUENCE A) USED IN THE COMPUTER MODELLING. AMINO ACIDS WHICH ARE IDENTICAL IN BOTH SEQUENCES ARE PRINTED IN CAPITAL LETTERS

```
                              6                                       15
A: LYS-Val-PHE-Gly-Arg-CYS-GLU-LEU-Ala-Ala-Ala-Met-LYS Arg-His-
B: LYS-Gln-PHE-Thr-Lys-CYS-GLU-LEU-Ser-Gln-Leu-Leu-LYS-
                              6                    13

                                                                      30
A: Gly-Leu-ASP-Asn-TYR-Arg-GLY-Tyr-Ser-LEU-Gly-Asn-Trp-Val-CYS-
B: Asp-Ile-ASP-Gly-TYR-Gly-GLY-Ile-Ala-LEU-Pro-Glu-Leu-Ile-CYS-
                                                                      28

                                                                      45
A: Ala-Ala-Lys-Phe-Glu-SER-Asn-Phe-Asn-THR-GLN-ALA-Thr-Asn-Arg-
B: Thr-Met-Phe-His-Thr-SER-Gly-Tyr-Asp-THR-GLN-ALA-Ile-Val-Glu-
                                                                      43

         47                                                           60
A: ASN-Thr-ASP-Gly-SER-THR-Asp-TYR-GLY-Ile-Leu-GLN-ILE-Asn-Ser-
B: ASN      ASP-Gln-SER-THR-Glu-TYR-GLY-Leu-Phe-GLN-ILE-Ser-Asn-
             45                                                       57

                    64                                                75
A: Arg-Trp-TRP-CYS-Asn-Asp-Gly-Arg-Thr-PRO-Gly-SER-ARG-ASN-Leu-
B: Lys-Leu-TRP-CYS-Lys-Ser-Ser-Gln-Val-PRO-Gln-SER-ARG-ASN-Ile-
                    61                                                72

    76                  80                                            90
A: CYS-Asn-ILE-Pro-CYS-Ser-Ala-Leu-LEU-Ser-Ser-Asp-ILE-THR-Ala-
B: CYS-Asp-ILE-Ser-CYS-Asp-Lys-Phe-LEU-Asn-Asp-Asn-ILE-THR-Asn-
    73                  77                                            87

                    94                          100                   105
A: Ser-Val-Asn-CYS-ALA-LYS-LYS-ILE-Val-Ser-ASP-Gly-Asp-GLY-Met-
B: Asn-Ile-Met-CYS-ALA-LYS-LYS-ILE-Leu      ASP-Ile-Lys-GLY-Ile-
                    91                           97              101

                                                                     120
A: ASN-Ala-TRP-Val-ALA-Trp-Arg-Asn-Arg-CYS-Lys-Gly-Thr-Asp-Val-
B: ASN-Tyr-TRP-Leu-ALA-His-Lys-Ala-Leu-CYS-Thr-Glu-Lys-Leu
                                                              115

                        125 126 127          129
A: GLU-Ala-TRP-Ile-Arg-Gly-CYS      Arg-LEU
B: GLU-Gln-TRP-Leu           CYS-Glu-Lys-LEU
                             120 121     123
```

atures of 3000K are chosen to increase the computational efficiency in crossing energy barriers, which would be negotiable by the real molecule. More details of the technique will be described elsewhere.

RESULTS

The conformation of the calcium binding loop of α-lactalbumin, i.e. residues 79–88 of the structure obtained from the LUCIFER rigid geometry conformation by 6ps of MDM at 4K, then 1ps at 310K, 0.5ps at 3000K, followed by 5ps of annealing at 4K is shown in Fig. 1. the RMSD deviation for the Cα atoms is 0.44Å with respect to those of residues 82–91 of hen egg white lysozyme. This is 0.68Å in r.m.s. deviation of atomic positions as used by crystallographers, which compares with the value of 0.6Å reported by Stuart *et al.* [3]. On inspection of the calcium binding loop, all 5 oxygen ligands, which are reported to bind Ca^{2+} are pointing into a cavity, easily accessible to the solvent. If a Ca^{2+} ion is placed at the centroid of these 5 ligands, then 4 of the Ca-O distances are in the range 2.0–2.3Å whilst the fifth Ca-O distance, which is between the Ca^{2+} and a oxygen of Asx-88 is 3.25Å. At this stage it should be emphasised that no Ca^{2+} ion was included in the procedure given above, and that the structure obtained is probably better treated as a prediction of the apo-protein. In the holo-protein structure reported recently [3] the Ca^{2+} ion is coordinated by 5 oxygens from the protein molecule and 2 oxygens from water molecules forming a slightly distorted pentagonal bipyramid, the Ca^{2+} -O distances are in the range 2.3 – 2.5Å. A stereo plot of the Cα atoms is shown in Fig. 2a.

In practice, two model structures in deep energy wells separated by substantial energy barriers have been located. (See Fig. 2b for the second conformation.) Both have an RMSD deviation of less than 1.2Å r.m.s. with respect to lysozyme.

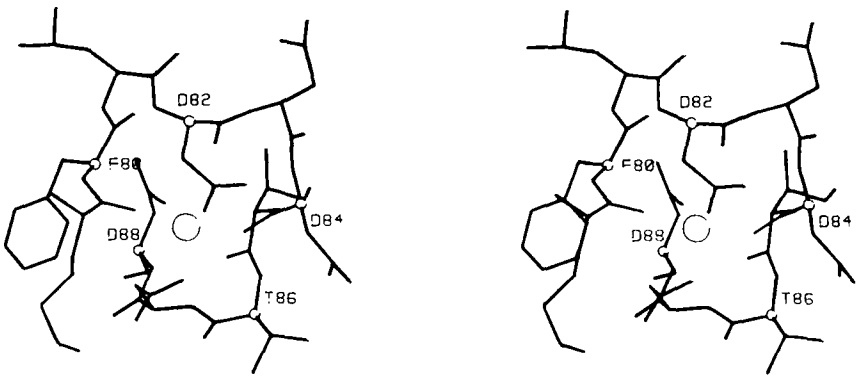

Fig. 1. Stereo plot of the Ca^{2+} binding loop of computer modelled α-lactalbumin. The Ca^{2+} ion is placed at the centroid of the five oxygen atoms of the protein, that bind to it (Stuart *et al.*, 1986). The IUPAC 1-letter code for the amino acids is used.

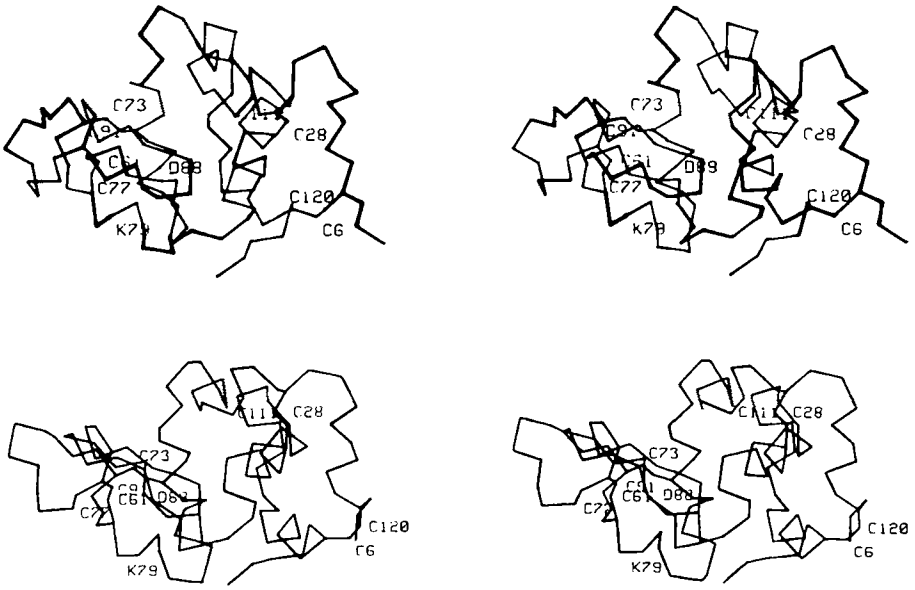

Fig. 2. Stereo plot of the Cα atoms of the computer model of α-lactalbumin. The Cys, Lys-79 and Asp-88 residues are labelled using the IUPAC 1-letter amino acid code a) model 1 and b) model 2.

DISCUSSION

The two models agree in detail in many features, with a RMSD deviation of 0.9Å between them for residues 1–111, but differ in relationship to the conformation of the C-terminal fragment (residues 112–123), well known to be problematic due to the lack of significant homology with lysozyme [7,8]. Model 1, shown in Fig. 2a, has a more extended structure in this region. Model 2, shown in Fig. 2b, has more helical structure, following the proposal of the T3 conformation of Warme *et al.* [8]. We support their proposal for this region in that the energy of the model 2 structure is 27 kcal/mol lower than that of model 1, in the above simulation. Almost all of this favourable energy, however comes from hydrogen bonding and it is likely that the energy difference would be substantially reduced in a simulation explicitly representing the solvent. Because the simulation treats the whole molecule at the same time, there are necessarily other differences of detail between the two conformations due to long-range direct and indirect effects in the molecule. Model 1 has two right-handed (90°) and two left-handed (−90°) C-S-S-C disulphide connections, model 2 has three right-handed (although one is 140°) and one left-handed disulphide connections, whilst the proposal by Warme *et al.* [8] for the whole molecule has three right-handed and one nearly *trans* (180°) disulphide connection. The optical activity data [17] is more consistent with model 2, and is thus further evidence that model 2 is, in fact, the lower energy structure.

As the coordinates of the holo-protein become available, a more detailed comparison of the two structures will be conducted.

ACKNOWLEDGEMENTS

E.P. was supported by the S.E.R.C. The computer calculations were carried out on the CYBER 205 at UMRCC.

REFERENCES

1 Aschaffenberg, R., Fenna, R.E. and Phillips, D.C., J. Mol. Biol., 67 (1972) 525–528.
2 Fenna, R.E., J. Mol. Biol., 161 (1982) 203–210.
3 Stuart, D.I., Acharya, K.R., Walker, N.P.C., Smith, S.G., Lewis, M. and Phillips, D.C., Nature, 324 (1986) 84–87.
4 Brew, K., Vanaman, T.C. and Hill, R.L., J. Biol. Chem., 242 (1967) 3747–3749.
5 Robson, B. and Platt, E., J. Mol. Biol., 188 (1986) 259–281.
6 Blake, C.C.F, Mair, G.A., North, A.C.T., Philips, D.C. and Sarma, V.R., Proc. Roy. Soc. London (Ser B) 167 (1967) 365–377.
7 Browne, W.J., North, A.C.T., Phillips, D.C., Brew, K., Vanaman, T.C. and Hill, R.L., J. Mol. Biol., 42 (1969) 65–86.
8 Warme, P.K., Momany, F.A., Rumball, S.V., Tuttle, R.W. and Scheraga, H.A., Biochemistry, 13 (1974) 768–782.
9 McCammon, J.A., Gelin, B.R. and Karplus, M., Nature, 267 (1977) 585–589.
10 Levitt, M., J. Mol. Biol., 170 (1983) 723–764.
11 Diamond, R., J. Mol. Biol., 82 (1974) 371–391.
12 Janin, J., Wodak, S., Levitt, M. And Maigret, B., J. Mol. Biol., 125 (1978) 357–386.
13 Robson, B., Platt, E., Fishleigh, R.F., Marsden, A. and Millard, P., J. Mol. Graph., 5 (1987) 8–17.
14 Van Gunsteren, W.F. and Berendsen, H.J.C., Molec. Phys., 45 (1982) 637–647.
15 Van Gunsteren, W.F. and Berendsen, H.J.C., Molec. Phys., 34 (1977) 1311–1327.
16 Hockney, R.W. and Eastwood, J.W., Computer simulation using Particles, McGraw Hill, New York, 1981.
17 Cowburn, D.A., Brew, K. and Gratzer, W.B., Biochemistry, 11 (1972) 1228–1234.