Applications of molecular graphics for the study of recognition

A. C. T. North

Astbury Department of Biophysics, University of Leeds, Leeds, UK

Molecular recognition depends upon a precise structural complementarity between the pairs of molecules concerned. For example, the strong affinity between an antigen and its specific antibody can be negated by replacement of a single amino-acid residue. The protein β-lactoglobulin from cows' milk is a member of a family of ligand-binding proteins with dual molecular recognition properties—for ligand and for a target receptor cell. This paper describes the use of molecular mechanics and energy calculations in attempting to understand the basis of such molecular recognition phenomena.

Keywords: interactive computer graphics, molecular modeling, molecular mechanics, molecular recognition

INTRODUCTION

There is general agreement that specific affinities between pairs of molecules—the process now known as "molecular recognition"—occurs as a result of a pattern of favorable interactions between the molecules concerned. Such favorable interactions arise from a complementarity between the shapes of the molecular surfaces that maximize van der Waals interactions; from the juxtaposition between hydrogen bond donor and acceptor groups or between groups carrying opposite charges; and from the apposition of hydrophobic patches on the molecular surfaces. A proper understanding of the interactions concerned requires a detailed knowledge of the three-dimensional (3D) structures of the interacting molecules, both separately and as a complex.

An exquisite example of molecular recognition occurs in the binding of an antibody to its antigen. The best understood example of an antibody/antigen complex is that between hen egg white lysozyme (HEWL) and the monoclonal antibody D1.3, for which it has been possible to obtain crystals of the complex between HEWL and the antigen-binding fragment of the antibody.[1] Color Plate 1 shows that the two molecules in the complex are in contact over an area of about 20 Å diameter. This contact area involves discontinuous segments of both the antigen and the antibody polypeptide chains (i.e., the epitope is discontinuous). A comparison of the structure of HEWL in the complex with that of native HEWL in the uncomplexed form shows that the protein structure has been distorted little, if at all, on binding to the antibody and that it involves a relatively rigid part of the HEWL molecule (i.e., one that has comparatively low crystallographic thermal parameters). While it would not be appropriate to generalize from a single example, it clearly shows that flexibility of the epitope is not required for antibody recognition and binding.

Detailed examination of the contact surface (Color Plate 2) shows that a particularly favorable interaction involves the fitting of the side chain of HEWL Gln 121 into an invagination in the antibody. The contact comprises several side chains of the antibody and, specifically, a hydrogen bond to the Gln side chain. While the antibody D1.3 has a high affinity for HEWL, it has a negligible affinity for some other avian lysozymes in which Gln 121 is substituted by a different residue, most notably California Quail lysozyme, in which the only change in the epitope is the substitution of Gln 121 by His. Color Plate 3 shows a computer graphics model in which a His side chain has been substituted for Gln, showing that there is apparently no difficulty in accommodating the His side chain within the antibody pocket still with a stereochemically plausible hydrogen bond. At first site, it therefore seems very surprising that this substitution should cause such a dramatic change in the affinity between antigen and antibody. In both cases, a hydrogen bond can be formed and the fit appears to be snug (i.e., van der Waals interactions are apparently favorable).

Phillips and Haneef[2] have carried out an energy minimization procedure in order to compare the energies of interaction between the molecules in the two complexes. A major difficulty in molecular mechanics energy minimization methods is that it is very easy for the molecular conformations to reach false local minima; in two

Address reprint requests to Dr. North, Astbury Dept. of Biophysics, University of Leeds, Leeds LS2 9JT, UK.

This paper was presented at the Symposium on Molecular Recognition, held in Sopron, Hungary, 24–27 August 1988.

Table 1. Calculation of interaction energies between HEW lysozyme and the D1.3 antilysozyme antibody by molecular mechanics, using a modified Kollman force field and the SIMMIN method of Haneef. The upper part of the table is for native HEWL with Gln at position 121 and the lower part for a model in which Gln has been replaced by His. In each case, the first column shows the energy of the D1.3 fragment, the second that of the HEWL, the third that of the complex, and the final column the difference between the energy of the complex and the sum of the energies of the separate molecules. No allowance has been made for solvent. All values are in kcal/mol

	D1.3	HEWL(Gln)	Complex	Change
Electrostatic term	−5721	−3234	−9116	−161
Van der Waals term	−2091	−1230	−3387	−66
Total enthalpy	−7812	−4464	−12503	−227
	D1.3	HEWL(His)	Complex	Change
Electrostatic term	−5721	−3235	−9098	−142
Van der Waals term	−2091	−1229	−3384	−64
Total enthalpy	−7812	−4464	−12482	−206

Difference in enthalpies of complex formation: 21 kcal/mol

Observed differences in affinity of Hen and California Quail egg lysozymes for D1.3 require an energy difference of at least 8 kcal/mol

different simulations, it would be possible for the structures to reach two unrelated minima, so giving an incorrect estimate for the energy difference between them. Haneef has, however, developed a novel minimization technique[3] in which the structures to be compared are required to follow similar minimization trajectories, to ensure they reach corresponding minima. Table 1 shows the results of such a study in which the enthalpies of the antibody, the lysozyme antigen and the complex between the two are minimized simultaneously. The computation was carried out both for the native hen egg lysozyme and for a model in which Gln 121 had been replaced by His. The results show that the lysozyme with Gln 121 binds significantly more strongly than that with His 121, the energy difference being large enough to account for the observed relative affinities of the two antigens, arising primarily from the Gln side chain making more favorable electrostatic interactions.

Repeated minimizations, from different starting coordinates, yielded estimates of binding enthalpies that, while different, were consistent in showing that the diminished binding energy of the His side chain was sufficient to explain the negligible affinity of the His enzyme. Thus, while a simple examination of the system by computer graphics would appear to show that either of the lysozymes should form a productive complex with the antibody, careful evaluation of the interaction energies reveals a significant difference in affinity. A conclusion to be drawn from this work is that, while examination of a model by computer graphics could undoubtedly show that certain amino-acid substitutions would prevent formation of a viable complex, the ability to model an apparently favorable complex on the graphics does not necessarily prove that such a complex would be sufficiently favorable to form *in vivo*.

STRUCTURAL STUDIES OF β-LACTOGLOBULIN

As will be shown later, β-lactoglobulin is a member of a family of proteins that show a dual molecular recognition property, namely for a ligand being transported by the protein and for a cell-surface receptor site. β-lactoglobulin itself is a protein that is found in high concentration in the soluble whey fraction of the milk of ruminants. It occurs also in some other mammalian species, but its presence in human milk is unproved. It is normally present as a dimer of identical molecules, each comprising 162 amino-acid residues. In Leeds, we have determined the structure of a triclinic crystal form of β-lactoglobulin, the present resolution limit being 2.5 Å.[4] Color Plate 4 shows the β-lactoglobulin dimer; although they are crystallographically independent, there are no significant conformational differences between the two monomers. A rotated view of the dimer, Color Plate 5, reveals two characteristic aspects of the molecule. First, from certain views the chain fold appears as two β sheets packed on top of each other, with the strands running in perpendicular directions so as to give a criss-cross pattern. The second feature is that the chain fold forms a deep pocket between the two β sheets, which take the shape of a calyx.

The same features had been seen in the structure of a second crystalline form of the protein and the striking criss-cross pattern had also been apparent in the structure of serum retinol-binding protein (RBP).[5] The crystal structure of RBP had shown retinol to be bound in the pocket. This structural similarity provided a rational explanation for observations that β-lactoglobulin too could bind retinol, in addition to other hydrophobic molecules; in fact, BLG shows a higher affinity for retinol than does RBP.

The computer graphics system was used to align the BLG and RBP molecules and it was found that 129 out of the 162 amino-acid residues of BLG could be aligned with corresponding residues of RBP with a root-mean-square error of 2.76 Å. These residues constitute the framework of the polypeptide chain fold, with much greater differences being shown between the loop regions joining the β strands. The position occupied by retinol in RBP fell within the pocket of BLG, but replacements in the amino acids facing the interior of the pocket result in the BLG pocket being considerably deeper than the RBP one. We therefore used the computer graphics system to model a position for retinol more deeply within the pocket, followed by energy minimization procedures to obtain an optimized position. In this, Color Plate 6 shows the retinol molecule to be fitting snugly within the pocket and forming a hydrogen bond between the retinol hydroxyl group and one of the protein side chains. The side chain of a tryptophan residue, Trp 28, is in contact with the isoprene tail of the retinol molecule.

It has been known for many years that β-lactoglobulin shows a number of conformational transitions, probably

the most important being the Tanford transition, which occurs at about pH 7.5 and involves a change in O.R.D. spectrum, in the accessibility of a sulphydryl group and in the titration of a carboxyl group. The two crystal forms, X and Y, have been grown respectively on the low and high pH side of this transition, and it has been possible to compare their structures. While there is very little change in the overall chain fold, there is a significant alteration in the relative positions of the free sulphydryl group and two aromatic side chains. We believe that these differences are sufficient to account for the change in sulphydryl accessibility and also that in the O.R.D. spectrum arising from the altered environment of the aromatic rings. There are also differences in the orientation of some of the side chains that face the internal pocket, though it is not clear whether these are a direct consequence of the structural transition or whether they merely represent arbitrary conformations taken up by these side chains, whose positions do not seem to be constrained by neighboring parts of the molecule.

Despite its abundance, the primary function of β-lactoglobulin has never been established. It is a very stable protein, resistant to the low pH conditions found in the intestine and presumably not present in milk just as a source of amino acids for the young animal. The similarity with RBP, and the fact that BLG is known to bind retinol and other hydrophobic molecules, suggests that it is present primarily as a transport protein, with the function of carrying retinol or other metabolites from the adult to the young animal. Preliminary experiments have indeed shown that a radioactively labeled BLG/retinol complex binds specifically in the lower segments of one-week-old calf intestine, but that such specific binding sites are no longer present in six-month-old calves.[5] These results demonstrate that BLG can indeed function as a transport protein targeted to specific receptors, although it remains to be established that retinol is the principal ligand *in vivo*.

A FAMILY OF LIGAND-BINDING PROTEINS

Recently, two more crystal structures have been obtained of proteins that have closely analogous features to those of BLG and RBP. These are the bilin-binding protein from the insect Pieris brassicae[6,7] and the insecticyanin from the tobacco hornworm moth.[8] Both these proteins bind bilin, the breakdown product of haem (protoporphyrin IX). They are almost identical in conformation with each other, very close in amino-acid sequence, and show a very similar central framework of β sheets to that of BLG and RBP, though again with substantial differences in the loop regions connecting the β strands. In both of them, the bilin molecule is found in the central pocket, though in slightly different orientations in the two cases. The coloration exhibited by the protein/bilin complex is used as camouflage by the insects, although it is not clear whether it has another function, such as an involvement in photoreception.

Despite substantial differences in amino-acid sequence, certain regions of the molecules show a high degree of homology. Analysis of sequences in the protein sequence databases has revealed a number of other proteins that share these regions of high homology with BLG, RBP and the two bilin-binding proteins.[9-12] Table 2 lists the members of this family that have so far been identified. They are of comparable size (150 to 200 amino acid residues) and exhibit two regions of high homology, respectively a sequence -h-a-x-x-u-h-x-Gly-x-Trp-y-x-h-h- near residue 20 (where a is usually acidic, u often basic, y aromatic and h hydrophobic), and a second sequence -h-h-x-Thr-Asp-Tyr-x-x-y-h- near residue 100. In addition, a disulphide bridge occurs in most of the proteins between Cys 66 and another Cysteine near to the carboxyl end of the chain. The presence of these highly homologous features suggests strongly that all these proteins have a common chain fold in three dimensions. The Gly-x-Trp feature is almost universal, and the Trp is in fact the only amino acid that is completely invariant throughout all of the proteins of the family that have been identified so far. The Thr-Asp-Tyr- triplet is very common, though not universal.

What is remarkable about the highly conserved residues is that, in the positions in which they occur in BLG, they form a cluster in space; the side chain of Trp 19 faces into the pocket where, as has already been said, it is in contact with retinol in our model. The majority of the remaining near-invariant or homologous residues are on

Table 2. Properties and functions of members of the ligand-binding protein family

Protein	Size	Ligands	Properties
Beta-lactoglobulin	2 × 162	Retinol and other hydrophobic molecules	Vitamin transport to gut receptor
Serum retinol-binding protein	182	Retinol in complex with transthyretin	Retinol transport to the eye
Purpurin	196	Retinol, heparin	Cell survival
Alpha-1-acid-glycoprotein	187	Progesterone	Mediates inflammatory response
Androgen-dependent secretory protein	165	?	Binds to sperm membrane
Endometrial alpha-2-globulin	~21Da	?	Synthesized in early pregnancy
Alpha-1-microglobulin (Protein HC)	183	IgA and a yellow-brown retinoid	Mediation of neutrophil chemotaxis
Apo-lipo-protein D	169	Lecithin, cholesterol	Lipid transport
Major urinary protein (alpha-2μ-globulin)	162	Pheromones?	Signaling?
Bowman's gland odorant-binding protein	160	Odorants	Presentation of odorants to receptor?
Insecticyanin (Bilin-biding prot.)	189	Biliverdin IX	Camouflage
Crustacyanin	170 + 177	Astaxanthin	Coloration, photo-reception

the exterior surface of the protein (Color Plate 7).

As can be seen from Table 2, the function of many of these proteins is to transport small hydrophobic molecules through an aqueous or otherwise hostile environment. Particularly interesting examples are the proteins that are secreted in urine and that transport odorant molecules and those that are found in nasal mucosa, which also bind odorant molecules. Presumably, these proteins are involved in the transmission and reception of odorants that are used in signaling between animals.

For some members of the family, it has been shown that these proteins will bind specifically to cell-surface receptors, and it may be presumed that in general, such transport proteins must have affinities for specific cell surfaces from which they would pick up or to which they would deliver their bound ligands. Thus, these proteins must have a dual molecular recognition property—for the ligand that they are transporting and for the cells that secrete or absorb it.

It seems unlikely that the conserved surface residues can have any role in ligand recognition or an important role in stabilizing the chain fold of these proteins. The alternative possibility is that their function is to promote binding to the surfaces of specific cells and presumably the *common* features of the transport protein surfaces engage with *common* features of the different receptor cells. Color Plate 8 shows the area of the BLG molecule that includes the near-invariant and highly homologous side chains and shows that adjacent to these are other side chains that vary between the different members of the family; presumably, these variable side chains are concerned with the specific recognition of the correct cell to which the ligand is to be delivered or from which it is to be picked up.

For efficient transport of a ligand to a receptor cell, loaded protein molecules should be able to displace molecules that have already released their ligand. This has been shown experimentally to be the case with RBP.[13] A conformational change, such as the pH-induced Tanford transition, is likely to be an essential feature of such a process, and we suggest that the invariant Trp residue that faces the ligand-binding pocket in BLG might act as a trigger in inducing the conformational change.

ACKNOWLEDGEMENTS

I am grateful to Drs. S.E.V. Phillips and I. Haneef for permission to report unpublished work on the HEWL/D1.3 complex; to Drs. R. Cooper, E.E. Eliopoulos and S.J. Yewdall for contributions to the BLG X form structure; to Dr. I. Haneef and Miss H. Gilman for their role in modeling the BLG/retinol complex; to Dr. L. Sawyer and his group in Edinburgh for their collaboration with the BLG project as a whole; to Dr. J.B.C. Findlay for helpful discussion and for making available his unpublished observations on crustacyanin; and to the UK Science and Engineering Research Council for financial support.

REFERENCES

1 Amit, A.G., Marriuzza, R.A., Phillips, S.E.V., and Poljak, R.J. *Science* 1986, **233**, 747–753
2 Phillips, S.E.V., and Haneef, I. (unpublished work)
3 Haneef, I. (in course of publication)
4 North, A.C.T., and Yewdall, S.J. (unpublished work)
5 Papiz, M.Z., Sawyer, L., Eliopoulos, E.E., North, A.C.T., Findlay, J.B.C., Sivaprasadarao, A., Jones, T.A., Newcomer, M.E., and Kraulis, P.J. *Nature (London)* 1986, **324**, 373–385
6 Huber, R., Schneider, M., Epp, O., Mayr, I., Messerschmitt, A., Pflugrath, J., and Kayser, H. *J. Molec. Biol.* 1987, **195**, 423–434
7 Huber, R., Schneider, M., Mayr, I., Muller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H., and Kayser, H. *J. Molec. Biol.* 1987, **198**, 499–513
8 Holden, H.M., Rypniewski, W.R., Law, J.H., and Rayment, I. *EMBO J.* 1987, **6**, 1565–1570
9 Sawyer, L. *Nature (London)* 1987, **327**, 659
10 Godovac-Zimmermann, J. *Trends Biochem. Sci.* 1988, **13**, 64–66
11 Pevsner, J., Reed, R.R., Feinstein, P.G., and Snyder, S.H. *Science* 1988, **241**, 336–339
12 Pervaiz, S., and Brew, K. *Science* 1985, **228**, 335–337
13 Sivaprasadarao, A., and Findlay, J.B.C. *Biochem. J.* 1988, **255**, 561–569
14 Jones, T.A. *J. Appl. Crystallogr.* 1978, **11**, 268–272