Journal of Molecular Structure, 194 (1989) 203–214
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

COMPUTER SIMULATION OF PROTEIN–CARBOHYDRATE COMPLEXES: APPLICATION TO ARABINOSE-BINDING PROTEIN AND PEA LECTIN*

V.S.R. RAO, MARGARET BISWAS, CHAITALI MUKHOPADHYAY and P.V. BALAJI
Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012 (India)
(Received 22 March 1988)

ABSTRACT

The CCEM method (Contact Criteria and Energy Minimisation) has been developed and applied to study protein-carbohydrate interactions. The method uses available X-ray data even on the native protein at low resolution (above 2.4 Å) to generate realistic models of a variety of proteins with various ligands.

The two examples discussed in this paper are arabinose-binding protein (ABP) and pea lectin. The X-ray crystal structure data reported on ABP-β-L-arabinose complex at 2.8, 2.4 and 1.7 Å resolution differ drastically in predicting the nature of the interactions between the protein and ligand. It is shown that, using the data at 2.4 Å resolution, the CCEM method generates complexes which are as good as the higher (1.7 Å) resolution data. The CCEM method predicts some of the important hydrogen bonds between the ligand and the protein which are missing in the interpretation of the X-ray data at 2.4 Å resolution. The theoretically predicted hydrogen bonds are in good agreement with those reported at 1.7 Å resolution. Pea lectin has been solved only in the native form at 3 Å resolution. Application of the CCEM method also enables us to generate complexes of pea lectin with methyl-α-D-glucopyranoside and methyl-2,3-dimethyl-α-D-glucopyranoside which explain well the available experimental data in solution.

INTRODUCTION

A knowledge of the three-dimensional structure of proteins and their complexes is extremely important in understanding various biological processes. The X-ray diffraction technique is, perhaps, the only method that gives information about the precise position of atoms and hence about the three-dimensional structure of proteins. Recently, however, it has been noted that unless the structure is solved at high resolution (below 2 Å) the information about protein–ligand interactions derived from this technique is not always reliable [1]. Most of the protein data that is available in the literature has been solved at above 2 Å resolution. Further, these studies are often limited to the native proteins and not many protein–ligand complexes have been solved.

*Dedicated to Professor K.S. Pitzer on the occasion of his 75th birthday.

0022-2860/89/$03.50

Because of these limitations, the wealth of data on protein structure that is available in the literature has not been fully exploited in understanding biomolecular interactions. However, computer modelling techniques have made it possible not only to generate new protein–ligand complexes using the available data but also provide information which is as good as that derived from X-ray studies at high resolution. It is worth mentioning that the concepts about torsional barriers and some of the data on model compounds which were reported by Prof. Pitzer as early as 1951, are still widely used in the computer-modelling of biological macromolecules.

We have been involved, for quite some time, in developing and applying computer modelling techniques to various protein–ligand complexes. The CCEM method (Contact Criteria and Energy Minimisation) developed in our laboratory enables the generation of realistic models of a variety of proteins with various ligands. The method involves two steps: (i) fitting of the ligand in the protein binding site using contact criteria [2–5] and (ii) discrimination of the allowed conformations by minimising the energy of the protein–ligand complex. To demonstrate the power of the CCEM method, two examples where it has been applied are discussed here. The arabinose-binding protein, is a carrier protein which has been solved using the X-ray diffraction technique as a complex with *β-L*-arabinose at 2.8 Å [6], 2.4 Å [7] and 1.7 Å [8] resolution. It was only at 1.7 Å resolution that the authors were able to determine the binding mode of *β-L*-arabinose unequivocally. Pea lectin is a protein with similar specificity for *α*-methyl-*D*-mannose and *α*-methyl-*D*-glucose [9]. It has been solved by X-ray crystallography to 3 Å resolution only in the native form [10]. The molecular structure of the pea lectin is very similar to that of concanavalin A (Con A) and there is some crystallographic evidence that the carbohydrate in the pea lectin binds in a similar location to that in Con A [10]. In the former case using 2.4 Å data we show that not only can the mode of binding of *β-L*-arabinose be unambiguously determined, but also the hydrogen bonding scheme predicted theoretically agrees well with the X-ray data reported at high resolution (1.7 Å) [8]. In the latter case, starting with the X-ray data on the native protein the possible modes of binding of methyl-*α-D*-glucose and a derivative have been determined. These are shown to be consistent with solution data.

METHOD OF CALCULATION

The coordinates of the proteins were taken from available X-ray crystallographic data. The ligand is generated in the protein binding site using standard geometry [11]. The rigid-body rotation method is used to move the sugar in the binding site using rotational and translational parameters.

In the first step, contact criteria are used to identify the possible orientations of the ligand in the protein binding site. Those orientations for which there

are no severe steric conflicts between the ligand and the protein are taken as the allowed orientations. Three-dimensional steric maps which give the allowed regions are constructed for each complex. At this stage, possible hydrogen bonds between the protein and the ligand can also be estimated. In order to identify the most likely orientations, in the second step, an energy minimisation procedure is followed. Prior identification of the stereochemically allowed orientations avoids an unproductive search of the whole conformational space during minimisation. The energy of the complex is calculated using empirical potential energy functions, which sum the nonbonded, electrostatic, torsional and hydrogen bond energies. A term which estimates the contribution to the energy of the exo-anomeric effect is also included for ligands which contain carbohydrate. Energy terms for bond length stretching and bond angle bending were not used since a fixed geometry was assumed for the protein and the ligand.

During energy minimisation, the side chains of the protein in the binding site are treated as flexible while backbone and other side chains are held rigid. The ligand is also treated as flexible and the energy of the ligand is included in the calculations. Also allowed to vary are the three rotational parameters (the Eulerian angles ϕ, θ and ψ) and the three translational parameters (x, y, z) that define the orientation of the ligand in the binding site [4]. Energy minimization is then carried out using either the search method of Rosenbrock [12] or using the gradient-based method of Dennis and Mei [13].

Potential energy functions and constants used

(1) Nonbonded energy is calculated using the modified Lennard–Jones 6-12 potential,

$$V_{nb} = FE\ (r_o/r_{ij})^{12} - 2.0E\ (r_o/r_{ij})^6$$

where, E is the depth of the minimum at $r_{ij}=r_o$ and F is the scaling factor and has a value of 0.5 for 1–4 interactions and 1.0 for 1–5 interactions and higher. The constants used are those reported by Scheraga and coworkers [14,15].

(2) Electrostatic energy is calculated using the standard expression. The dielectric constant was taken to be four. In the case of pea lectin a distance-dependent dielectric constant was used. The use of different dielectric constants does not significantly affect the preferred modes of binding. The sigma charges on various atoms were taken from Del Re and the pi charges were calculated using Hückel's method [16]. The partial charges on the sugar atoms were taken from Yathindra and Rao [17].

(3) The torsional potential energy was calculated according to Momany et al. [14].

(4) The 10-12 potential function of Scheraga and coworkers [18] was used to calculate the hydrogen bond energy. This is of the form,

$$V_{hb} = A_k/r^{12} - B_k/r^{10}$$

where, A_k and B_k are constants which differ for the different types of hydrogen bonds and r is the distance between the acceptor and the hydrogen atom. The constants used are those reported by Nemethy et al. [19].

(5) The expression used to estimate the energy due to exo-anomeric effect has been discussed previously [20,21].

RESULTS AND DISCUSSION

Arabinose-binding protein (ABP)

Only the data solved for the protein at 2.4 Å resolution is available to us and this lower resolution data was used as the starting data for the protein. Quiocho et al. have solved the ABP-β-L arabinose complex but have not reported the sugar coordinates. From the information about the approximate location of the sugar [7] we were able to fit the sugar in the ABP binding site.

Using the CCEM method, steric maps were constructed to show the allowed orientations of the β-L-arabinose in the ABP binding site (Fig. 1). These maps show that four regions are allowed for the sugar in the binding site of the protein suggesting that the sugar can reach the binding site in any of these allowed orientations. To determine the most favourable orientation for the sugar, energy minimisation was carried out using starting points from all the four al-

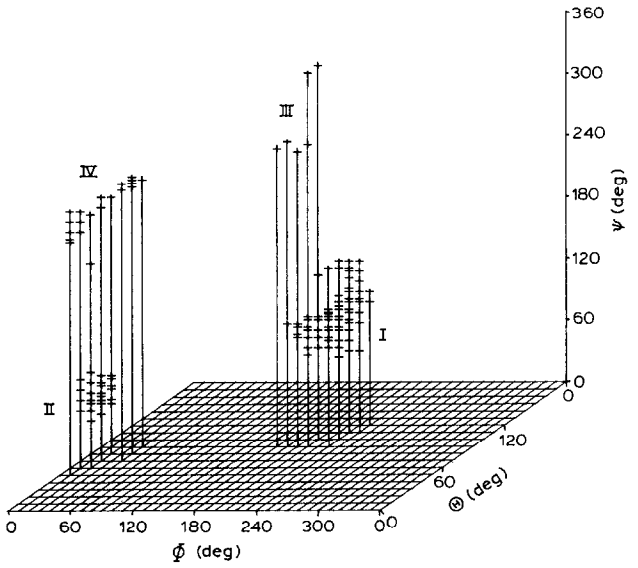

Fig. 1. Steric map representing the allowed orientations for β-L-arabinose in the ABP binding site. The Eulerian angles ϕ, θ and ψ can be read as: the length of the line joining the mark "+", and the point on the grid gives the value of the angle ψ; the grid point gives the values of ϕ and θ.

lowed regions. The energetically most favoured orientation was obtained in region 1 where the stabilising interactions between the protein and the ligand came from hydrogen bonds and van der Waals interactions (Table 1). The next most favourable region is region 4 where the energetically most favoured orientation is 9 kcal mol^{-1} more than the best orientation in region 1. Since L-arabinose has inversion symmetry, from contact criteria alone, it was difficult to discriminate one orientation over the other. Energy minimisation clearly shows that an orientation in region 1 is energetically more favoured than those in region 4. X-ray crystallographic data at 2.4 Å resolution also could not distinguish between these two possible orientations of β-L-arabinose. The other two regions of the steric map, regions 2 and 3, are energetically much less favourable than regions 1 and 4 (Table 1).

The hydrogen bonding pattern for interaction between ABP and β-L-arabinose in the best orientation from region 1 is shown in Table 2 and a projection of the complex is given in Fig. 2. It is gratifying to note that although the starting coordinates for ABP were taken from the 2.4 Å resolution X-ray data [7], the CCEM method has predicted the preferred orientation of β-L-arabinose in the binding site and a hydrogen bond scheme which is in good agreement with the scheme proposed by Quiocho et al. from the 1.7 Å resolution data [8] (Table 2). Arg (151) and Asn (232) are predicted to form complex

TABLE 1

Energy minimisation of ABP-β-L-Ara complex

Region	Orientation of sugar						Binding energy (kcal mol^{-1})	Hydrogen bonding scheme
	ϕ(°)	θ(°)	ψ(°)	X(Å)	Y(Å)	Z(Å)		
(1)	196	90	95	14.2	56.6	53.6	−59.2	OH(1)-OD2(89) OH(2)-NZ(10) OH(3)-ND2(205) OH(3)-ND2(232) OH(4)-NH2(151) OH(4)-OD1(232) O5-NH1(151)
(2)	14	74	41	14.4	56.8	53.9	−28.8	OH(1)-NH1(151) OH(3)-OD2(89)
(3)	170	115	270	14.4	56.9	53.9	−30.3	OH(1)-OD2(89) OH(2)-OD2(89) OH(4)-NH1(151)
(4)	4	92	254	14.4	56.8	53.9	−50.2	OH(1)-NH1(151) OH(2)-OD1(232) OH(2)-ND2(232) OH(4)-NZ(10) OH(4)-O(89) OH(4)-OD2(89)

TABLE 2

Hydrogen bond interactions in ABP-β-L-Ara complex as proposed from various studies

Nature of study	Possible hydrogen bonds with the sugar				
	OH(1)	OH(2)	OH(3)	OH(4)	O5
X-ray study at 2.4 Å resolution	LYS(10) ASP(90)	LYS(10)	GLU(14) (possible through water) ASN(205)	ASN(232)	
X-ray study at 1.7 Å resolution	ASP(90)OD2	LYS(10)NZ	GLU(14)OE2 ASN(205)ND2 ASN(232)ND2	ASN(232)OD1 ARG(151)NH2	ARG(151) NH1
Present theoretical studies using 2.4 Å resolution data	ASP(89)OD2	LYS(10)NZ	GLU(14)OE2 (possible through water) ASN(205)ND2 ASN(232)ND2	ASN(232)OD1 ARG(151)NH2	ARG(151) NH1

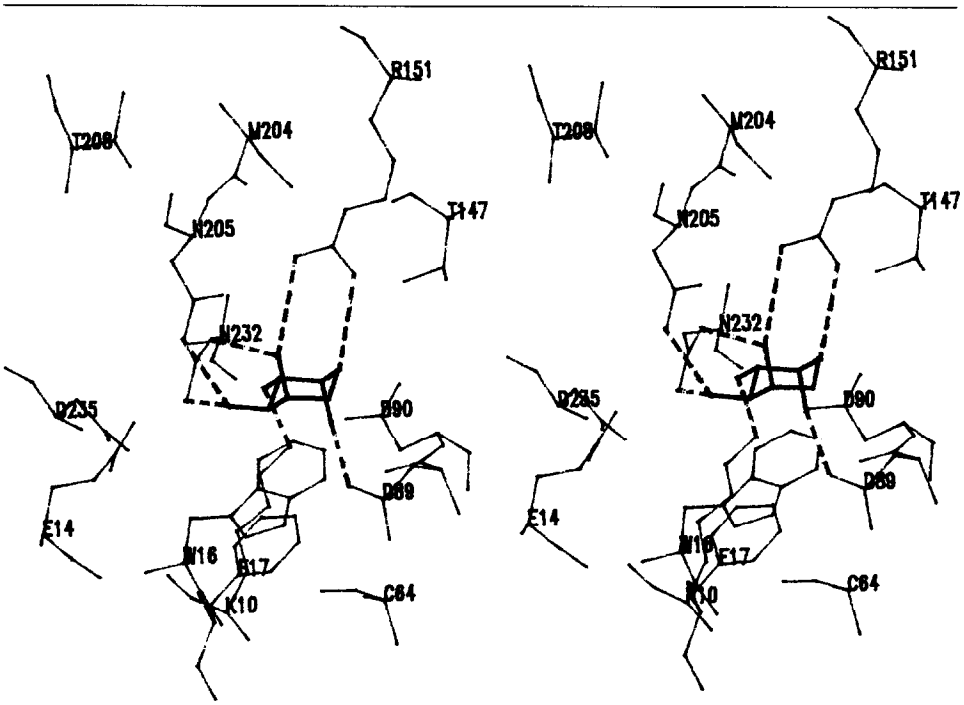

Fig. 2. Stereoscopic projection of β-L-arabinose in the ABP binding site in the minimum energy conformation obtained theoretically. This represents the best orientation from region 1 of Fig. 1. Details of the possible hydrogen bonds are given in Table 2. All protein hydrogens and some of the ligand hydrogen atoms have been removed for greater clarity.

hydrogen bonds, the former with OH(4) and O5 and the latter with OH(3) and OH(4) [21], in agreement with the high resolution X-ray data [8]. Interestingly, these hydrogen bonds, except the one between Asn (232) and OH(4) were not assigned from X-ray studies at 2.4 Å resolution.

Thus, starting from low resolution X-ray data, the CCEM method was able not only to fix the sugar in the correct orientation in the binding site but also to predict the hydrogen bonding scheme which agrees with the X-ray crystallographic data solved at 1.7 Å resolution.

Pea lectin

The steric map for pea lectin with liganded methyl-α-*D*-glucopyranoside is shown in Fig. 3. Comparison of this steric map with the steric map for methyl-α-*D*-glucopyranoside in the sugar binding site of Con A [4] shows that the former is much less restricted, indicating a lower degree of specificity for the pea lectin. Energy minimisation of pea lectin with methyl-α-*D*-glucopyranoside shows that there are three favoured orientations of which the orientation $\phi=250°$, $\theta=110°$, $\psi=120°$ is the most favoured (Table 3). In this orientation and in the third most favoured orientation, $\phi=355°$, $\theta=115°$, $\psi=33°$, the sugar hydroxyls, O3, O4 and O6 are involved in the formation of strong hydrogen bonds with the protein. In the second most favoured orientation, $\phi=20°$,

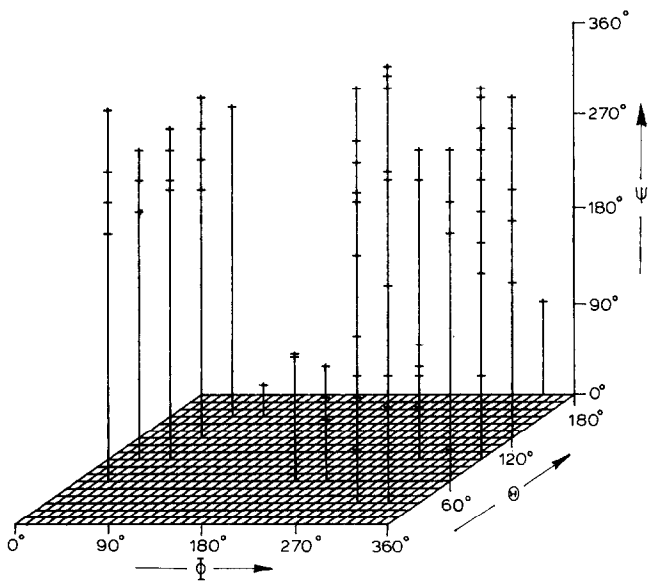

Fig. 3. Steric map representing the allowed orientations for methyl-α-D-glucopyranose in the pea lectin binding site.

TABLE 3

Energy minimisation of the complex of pea lectin with methyl-α-D-glucopyranoside

Orientation of sugar						Binding energy (kcal mol^{-1})	Hydrogen bonding scheme
ϕ(°)	θ(°)	ψ(°)	X(Å)	Y(Å)	Z(Å)		
1. 250	110	120	24.4	23.5	25.9	−72.8	O3H···OE2(218) O4H···OD1(81) O6H···OD2(81)
2. 20	75	225	25.0	23.5	25.4	−69.5	O2H···OD2(81) O3H···OE2(218) O4H···OE2(218)
3. 355	115	330	25.2	23.4	24.0	−62.9	O3···NH(99) O4···NH2(125) O4H···OD1(81) O6H···OD2(81)

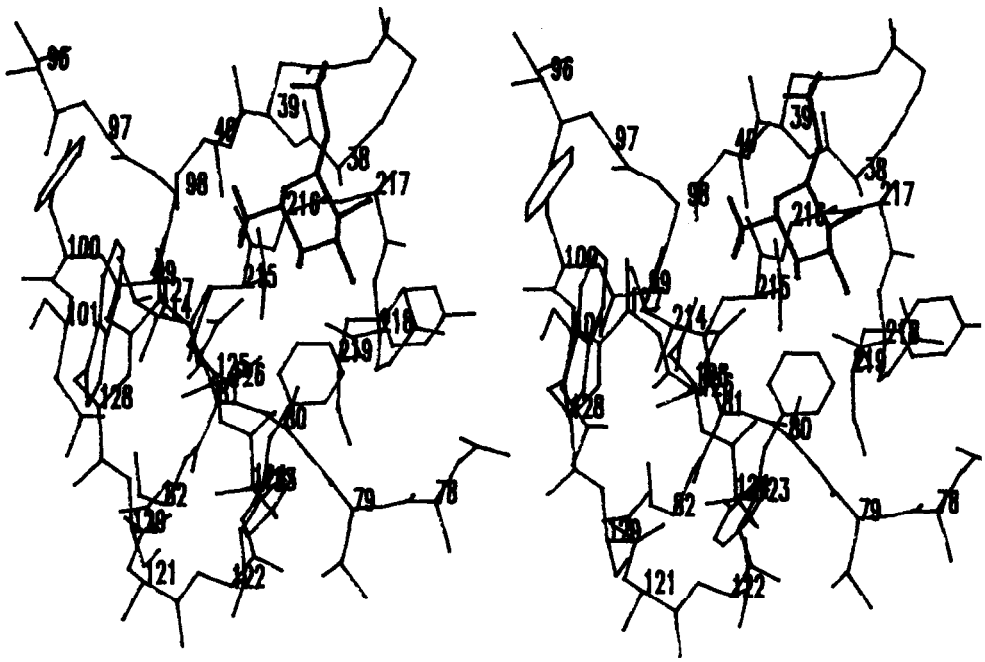

Fig. 4. Stereoscopic projection of methyl-α-D-glucopyranose in the pea lectin binding site in the best orientation (ϕ=250°, θ=110° and ψ=120°) as given in Table 3. The O3, O4 and O6 hydroxyl groups are involved in hydrogen bonds with the protein.

θ=75°, ψ=225°, the O6 hydroxyl group does not seem to be involved in the formation of hydrogen bonds (Table 3).

The first and third orientations preferred by methyl-α-*D*-glucopyranoside in binding to the pea lectin are very similar to the allowed orientations of

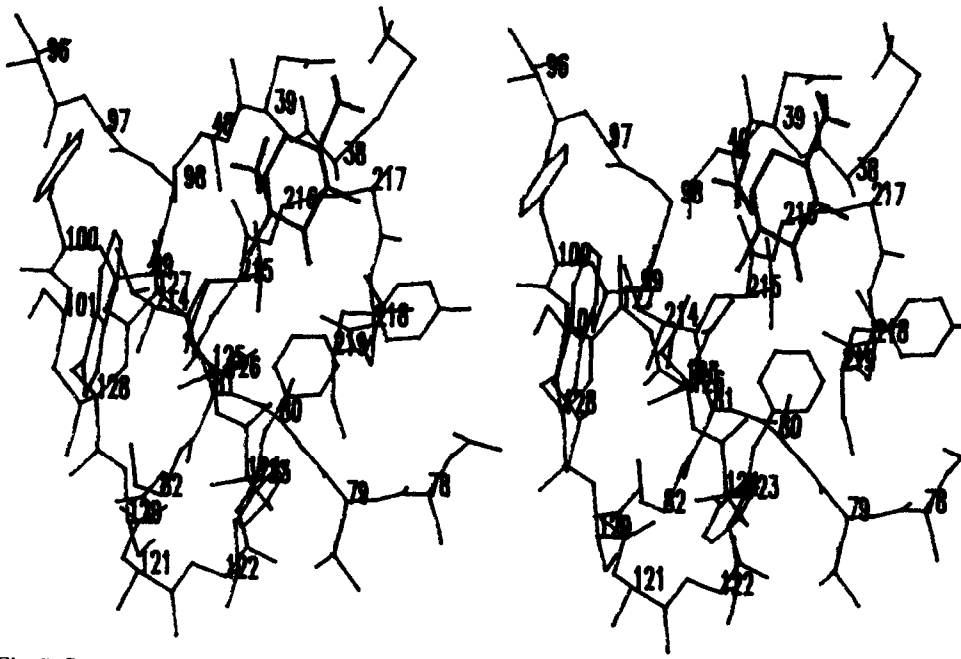

Fig. 5. Stereoscopic projection of methyl-α-D-glucopyranose in the pea lectin binding site in the second best orientation ($\phi=20°$, $\theta=75°$ and $\psi=225°$) as given in Table 3. In this orientation the O2, O3 and O4 hydroxyl groups are involved in hydrogen bonds with the protein.

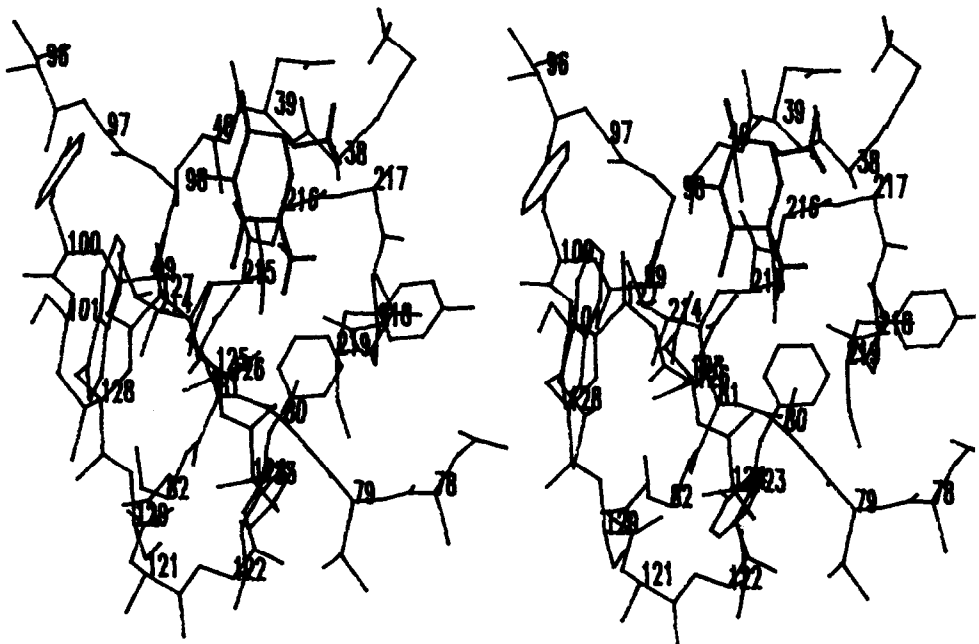

Fig. 6. Stereoscopic projection of methyl-α-D-glucopyranose in the pea lectin binding site in the third best orientation ($\phi=355°$, $\theta=115°$ and $\psi=330°$) as given in Table 3. The O3, O4 and O6 hydroxyl groups are involved in hydrogen bonds with the protein.

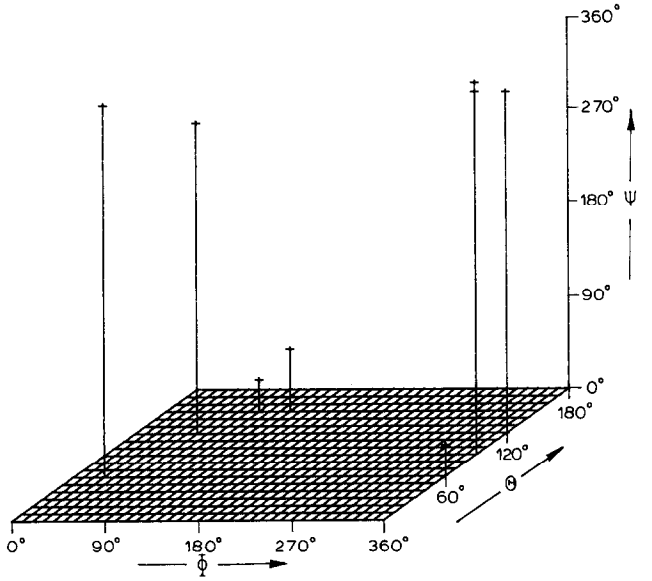

Fig. 7. Steric map of the allowed orientations for methyl-2,3-dimethyl-α-D-glucopyranoside in the pea lectin binding site. The number of the allowed orientations are very few compared with the large number possible for methyl-α-D-glucopyranoside (cf. Fig. 3).

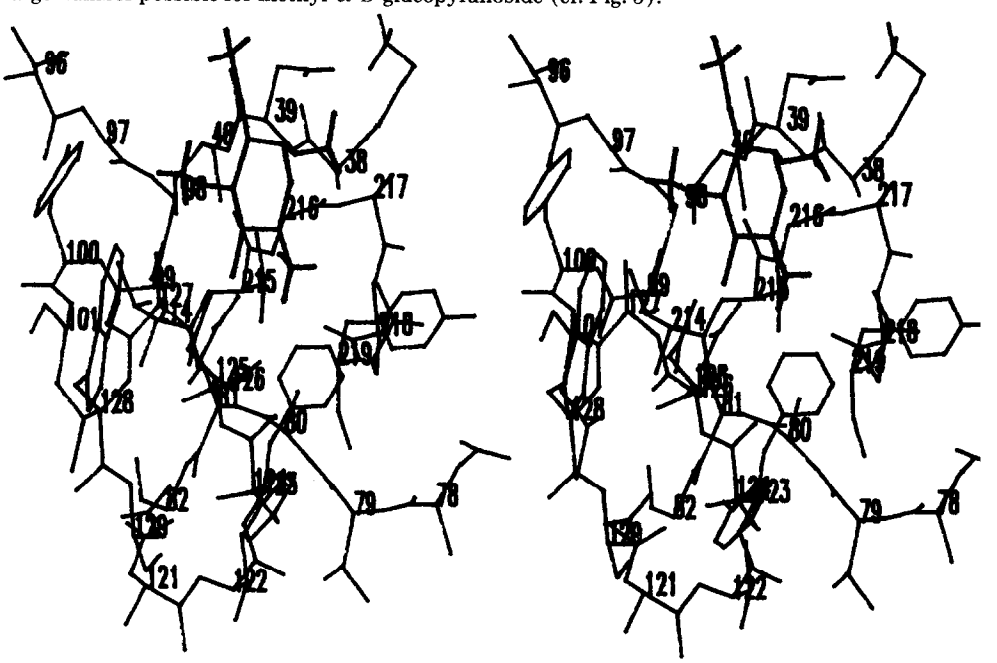

Fig. 8. Stereoscopic projection of methyl-2,3-dimethyl-α-D-glucopyranose in the pea lectin binding site in the best orientation (ϕ=355°, θ=115° and ψ=330°) arrived at from preliminary energy calculations. Possible hydrogen bonds are between O3···NH(99), O4···NH2(125), OD2(81) and O6··OD1(81). The methyl group at C3 is placed close to a hydrophobic region of the protein constituted by Gly(97), (98), (99) and Tyr(100) and the hydroxymethyl hydrogens lie close to Phe(123).

methyl-α-*D*-glucopyranoside in the Con A binding site [4]. However, the second most preferred orientation in pea lectin is not favoured in Con A. Projections of all three orientations for methyl-α-*D*-glucopyranose in the pea lectin binding site are shown in Fig. 4, 5 and 6.

Comparison of the steric map of methyl-2,3-dimethyl-α-*D*-glucopyranoside in the pea lectin binding site (Fig. 7) with that of the map for methyl-α-*D*-glucopyranoside (Fig. 1) shows that the addition of the two methyl groups at O2 and O3, highly restricts the possible favoured orientations of the ligand in the pea lectin binding site. The first two orientations which are preferred for methyl-α-*D*-glucopyranoside (Table 3) are completely eliminated for the 2,3 substituted derivative. Preliminary energy calculations have, in fact, indicated that the most favoured orientation of methyl-2,3-dimethyl-α-*D*-glucopyranoside is close to the third most favoured orientation for methyl-α-*D*-glucopyranoside. In this orientation, $\phi = 355°$, $\theta = 115°$, $\psi = 330°$, besides the hydrogen bonds which are probable with the O3, O4 and O6 hydroxyl groups, the methyl group at C3 is placed close to a hydrophobic region of the protein constituted by Gly(97), Gly(98), Gly(99) and Tyr(100) and the hydrogens of the hydroxymethyl group (C6) are close to Phe (123) (Fig. 8). This may lead to additional stabilising hydrophobic interactions. In this orientation the methyl group at C2 is not involved in interaction with the protein.

Our results, obtained theoretically using the CCEM method, agree well with the hapten inhibition measurements on the precipitin reaction between the pea lectin and a phosphomannan which show that methyl-2,3-dimethyl-α-*D*-glucopyranoside is a very potent inhibitor [9]. It is interesting to note that a similar methyl substitution on the O3 position of the sugar ligand drastically reduces its ability to bind to Con A.

In conclusion, the CCEM method allows the interactions of protein–carbohydrate complexes to be studied using the data obtained at low resolution (2.4 Å and 3.0 Å) and also enables the favoured binding orientations of the ligand in the protein binding site to be determined using data of the native protein, as in the case of pea lectin.

ACKNOWLEDGEMENTS

We express our thanks to Prof F.L. Suddath for making available the coordinates of pea lectin at 3 Å resolution.

The work reported here was supported by the Department of Science and Technology, New Delhi, India.

REFERENCES

1 F.A. Quiocho, Ann. Rev. Biochem., 55 (1986) 287.

2 M. Biswas, Y.C. Sekharudu and V.S.R. Rao, Carbohydr. Res., 160 (1987) 151.
3 I. Ghosh and V.S.R. Rao, Int. J. Biol. Macromol., 4 (1982) 130.
4 Y.C. Sekharudu and V.S.R. Rao, J. Biomol. Struct. Dyn., 2 (1984) 41.
5 C.R. Vinayaka and V.S.R. Rao, J. Biomol. Struct. Dyn., 2 (1984) 663.
6 F.A. Quiocho, G.L. Gilliland and G.N. Phillips, Jr., J. Biol. Chem., 252 (1977) 5142.
7 G.L. Gilliland and F.A. Quiocho, J. Mol. Biol., 146 (1981) 341.
8 F.A. Quiocho and N.K. Vyas, Nature, 310 (1984) 381.
9 J.P. van Wauwe, F.G. Loontiens and C.K. de Bruyne, Biochim. Biophys. Acta, 379 (1975) 456.
10 H. Einspahr, E.H. Parks, K. Suguna, E. Subramanian and F.L. Suddath, J. Biol. Chem., 261 (1986) 16518.
11 S. Arnott and W.E. Scott, J. Chem. Soc., Perkins Trans. 2, (1972) 324.
12 H.H. Rosenbrock, Comput. J., 3 (1960) 175.
13 J.E. Dennis, Jr. and H.H.W. Mei, J. Optimization Theory and Applications, 28 (1979) 453.
14 F.A. Momany, R.F. McGuire, A.W. Burgess and H.A. Scheraga, J. Phys. Chem., 79 (1975) 2361.
15 L.G. Dunfield, A.W. Burgess and H.A. Scheraga, J. Phys. Chem., 82 (1978) 2609.
16 D. Polland and H.A. Scheraga, Biochemistry, 6 (1967) 3791.
17 N. Yathindra and V.S.R. Rao, Carbohydr. Res., 25 (1972) 256.
18 R.F. McGuire, F.A. Momany and H.A. Scheraga, J. Phys. Chem., 76 (1972) 375.
19 G. Nemethy, M.S. Pottle and H.A. Scheraga, J. Phys. Chem., 87 (1983) 1883.
20 S. Prakash, Ph.D. Thesis, Indian Institute of Science, Bangalore, India, 1980.
21 C. Mukhopadhyay and V.S.R. Rao, Int. J. Biol. Macromol., 10 (1988) 217.