Volume 319, number 1,2, 5–11 FEBS 12181 March 1993

Modelling the carbohydrate recognition domain of human E-selectin

Alan Mills

Laboratory of Molecular Biology, Crystallography Department, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

Received 4 November 1992; revised version received 18 January 1993

The three-dimensional structure of the carbohydrate recognition domain of the human E-selectin endothelial-leucocyte adhesion molecule (ELAM-1) was modelled on the basis of the recently determined X-ray crystallographic structure of the homologous domain found in the rat mannose-binding protein. The EGF-like module, which is contiguous to the E-selectin lectin domain, and which is also involved in binding the tetrasaccharide ligand, was modelled on the recently determined NMR structure of the EGF-like module of human factor IX. The rule-based homology modelling procedures developed at Birkbeck and encoded in the program COMPOSER were used. The model of the two domains combined is discussed in terms of cation and ligand binding.

E-selectin; ELAM-1; Adhesion; Modelling; Lectin; EGF

1. INTRODUCTION

E-selectin, or endothelial leucocyte adhesion molecule-1 (ELAM-1), is a mosaic glycoprotein of 610 amino acids (116 kDa) found on the external membrane surface of vascular endothelial cells. It is transiently expressed in response to cytokines such as Il-1 and TNF [1]. Its function appears to be to recruit neutrophils into sites of inflammation. E-selectin is one of a small class of 'selectin' molecules, all of which have similar primary structures, and which include PADGEM (CE62), GMP-140, and Leu-8 (TQ1 or LAM-1). These latter are implicated in the 'homing' of leucocytes to the appropriate lymph nodes.

From consideration of amino acid sequence the structure of E-selectin is believed to consist of a stem of six cystine rich repeats (as found in complement regulatory proteins) surmounted by an N-terminal lectin domain of 119 residues, which has been shown to bind to a specific oligosaccharide. There is a single EGF-like module connecting this domain to the stem, and the bottom of the stem passes via a single transmembrane α helix to a small C-terminal cytoplasmic domain. This mosaic structure is mirrored at the nucleotide level by each separate structural module being coded for by its own exon [1].

The lectin domain is homologous to a relatively large family of C-type (i.e. calcium-dependent) animal lectins [2]; versions of this module are found in a variety of extracellular mosaic proteins, some soluble, some structural, and some membrane bound.

Correspondence address: A. Mills, Laboratory of Molecular Biology, Crystallography Department, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK. Fax: (44) (71) 436 8918.

The specific oligosaccharide counter ligand to E-selectin is a glycolipid tetrasaccharide known as sialyl-Lewis X, the sialylated version of the Lewis X blood antigen, with the structure *Neu5*→*Ac*α2→3*Gal*β1→4-(*Fuc*α1→3)*GlcNac* [3]. It is constitutively expressed on both neutrophils and monocytes.

Thus far, no X-ray crystallographic or NMR structure determinations have been reported for the E-selectin molecule. However, the recent determination [4] of the X-ray structure of a homologous C-type animal lectin domain from the rat mannose-binding protein offered the opportunity to build a structural model of the E-selectin lectin domain, using well established modelling techniques [5–8]. The EGF-like domain, which joins the lectin module to the stem, is modelled on the basis of a set of NMR coordinates for the epidermal growth factor-like module from human factor IX [9].

The rat mannose-binding protein carbohydrate recognition domains (115 residues) show a mixed $\alpha + \beta$ structure with no clear relationship to any previously known fold, and with a fair proportion of irregular loop structure. There are also two distinct calcium binding sites, although in the X-ray structure determination, holmium ions were bound.

This paper describes the process of building the two-domain model of E-selectin, and discusses those features of the model which may contribute to ligand and cation binding.

2. METHODS

2.1. *Coordinates*

The crystallographic coordinates of the rat mannose-binding domain (code 1MSB) were obtained from the Brookhaven Protein Data-

base, January 1992 Release [10,11]. In the crystals, the domain exists as a dimer, and the detailed conformation of some of the loops, and especially the amino- and carboxy-terminal tails, is slightly different between the two copies.

A set of preliminary coordinates for an averaged solution NMR structure of the factor IX EGF-like module were distributed (prior to release by Brookhaven Protein Data Bank) on computer diskette along with the quoted reference [9].

2.2. *Sequences*

A set of 39 examples of homologous sequences of C-type animal lectin domains were extracted from a composite sequence database [12], and these were multiply aligned (see Fig. 1) using a Needleman and Wunsch-based programme [13].

As a prerequisite to the structural modelling, an alignment was prepared between the sequence of the human E-selectin lectin domain and that of the rat mannose-binding structure, exhibiting 27% identity.

Structural information was brought to bear in confirming the correctness of the resultant alignment; the JOY typographical representation [14] (see Fig. 2a) was used to reveal this data. An adjustment was introduced in the alignment comprising a two-residue shift in the section following the second α helix, and including β strands three and four. The sequence identity in this region is very weak, although it includes two residues (Asp-161 and Glu-165) in the known structure which are involved in metal ion coordination. The alignment from sequence alone not only fails to align at least one such coordinating residue, but also would require a deletion of two residues at this point in the model. The modified alignment (see Fig. 2a) maintains hydrophobic residues at those points in β strands three and four which are buried, and avoids the deletion at the coordinating loop. In consequence, however, a coordinating aspartate in the mannose-binding protein is replaced by an isoleucine in the model by this alignment. This is discussed further below.

The modified alignment shows that four insertions were necessary. It was clear that these could be accommodated in exposed surface loops between secondary structural elements, the two disulphide bridges were conserved, and a proline aligned with the only *cis*-proline in the X-ray structure. Apart from the instance already mentioned, all existing metal-coordinating residues were either unchanged or conservatively mutated.

The alignment of the EGF-like module with the sequence of the EGF-like domain from human Factor IX preserves their three disulphides (see Fig. 2b), and requires no insertion or deletion.

2.3. *Structural modelling*

The COMPOSER suite of modelling software [5–8] was used to model E-selectin. The alignment was supplied, along with nominations for the boundaries between the structurally conserved regions (SCRs) and the loops or structurally variable regions (SVRs), as indicated in Fig. 2a. The SVRs were selected to be just wide enough to permit the necessary conformational adjustments to accompany the insertions at the four loops. COMPOSER extracted the conserved main chain framework from a structural alignment of the two crystallographically independent copies of the known structure, and then built on conformers of the side chains for the E-selectin sequence selected from its internal libraries.

COMPOSER then searched a database of high resolution protein structures to find loop fragments which met the appropriate geometric criteria for each loop. The retrieved candidate fragments were then automatically ranked on criteria of both goodness of fit and on sequence similarity, and filtered for steric clashes with the framework. The best selections were accepted, and the side chains for the loops were again built in their most probable conformations.

The crude model was then superimposed on the known structures and examined on a computer graphics system using FRODO [15]. A few steric clashes were removed by manual adjustment of side chain torsion angles; no adjustment of the polypeptide backbone was necessary. The conserved disulphides were checked and found to have good stereochemistry. The metal-binding sites were then examined, and calcium ions inserted into the model at positions almost identical to the equivalent holmium ions in the X-ray structure. Manual adjustment of the coordinating side chains was carried out to optimise the coordination at the binding sites. The whole model was then subjected to several rounds of energy minimisation within SYBYL, using its default parameters to relieve any steric strain resulting from the modelling process.

A similar procedure was followed in the case of the EGF-like model using the alignment shown in Fig. 2b, concluding with a few rounds of minimisation; it was apparent that the modelled structure was relatively strain-free.

The two sub-models were brought together on a graphics system, and the C-terminal strand from the lectin domain joined to the N-terminus of the EGF-like module in such a way as to largely preserve the extended conformation of the linking polypeptide. The interface region between the two domains was examined for the presence of hydrophobic patches which might indicate a common contact area but none were apparent.

No obvious relative orientation for the two domains suggested itself. The C-terminus of the EGF-like domain ended up at a point furthest from the calcium binding loop of the lectin domain (see Fig. 3).

3. RESULTS AND DISCUSSION

The backbone structure of the resultant model is shown (Fig. 3), with the four loop insertions in the lectin domain indicated. In this modelling exercise no difficulties were experienced with the COMPOSER method, except that it was necessary to nominate manually the SCRs and then to adjust the alignment of sequence to structure over them.

An analysis of the main chain ϕ–ψ dihedral angles

→

Fig. 1. Sequence alignment of 39 sequences of C-type animal lectin domains. The identifying codes (with the exception of ELAM1 and RATMABP) are as obtained from the OWL composite database [18], except that they have been truncated by two leters; the sequences are: 1, human E-selectin; 2, rat mannose-binding protein; 3, human pulmonary surfactant protein A; 4, rabbit pulmonary surfactant protein A; 5, dog pulmonary surfactant protein A; 6, rat RATSPD locus 1265bp ssRNA; 7, bovine conglutinin; 8, human mannose-binding protein-C; 9, rat prepulmonary surfactant-associated protein A; 10, human tetranectin; 11, reef shark tetranectin analogue; 12, human cartilage specific proteoglycan; 13, bovine cartilage specific proteoglycan; 14, chicken cartilage specific proteoglycan; 15, acorn barnacle lectin BRA-3; 16, rat hepatic asialoglycoprotein receptorR2/3; 17, mouse hepatic asialoglycoprotein receptor 2; 18, human hepatic asialoglycoprotein receptor H2; 19, rat hepatic asialoglycoprotein receptor 1; 20, human hepatic asialoglycoprotein receptor H1; 21, rat Gal/GalNac-specific lectin; 22, rat Kupfer cell receptor; 23, human low affinity Ig Fc_ε receptor; 24, mouse low affinity Ig Fc_ε receptor; 25, chicken hepatic lectin; 26, bovine BOVLE-CAM1 locus 2650bp ssRNA; 27, human LECAM1 (Leu-8, GP90-Mel, TQ1); 28, human granule membrane protein (GMP140); 29, mouse MUSGMP140A locus 2558bp ssRNA; 30, rat pulmonary surfactant-associated protein A; 31, human mannose receptor precursor; 32, rattlesnake galactose-specific lectin; 33, polyandrocarpa misakiensis lectin; 34, human natural killer protein 2-A; 35, human natural killer protein 2-C; 36, mouse NKR-P1 gene-40 protein; 37, mouse NKR-P1 gene-34 protein; 38, mouse NKR-P1 gene-2 protein; 39, rat NK-cell receptor P1. The degree of local conservation, as assessed within a 7-residue wide sliding window, is plotted below the alignment.

Cation site residues ----->
1 ELAM1
2 RATMABP
3 PSPA_HU
4 PSPA_RA
5 PSPA_CA
6 RATSPD
7 CONG_BO
8 MABC_HU
9 RATPSPA
10 TETN_HU
11 A37287
12 PGCA_HU
13 PGCA_BO
14 PGCA_CH
15 LEC3_ME
16 LECI_RA
17 LECI_MO
18 LECI_HU
19 LECH_RA
20 LECH_HU
21 RATGGSL
22 KUCR_RA
23 FCE2_HU
24 FCE2_MO
25 LECH_CH
26 BOVLECA
27 LEM1_HU
28 LEM3_HU
29 MUSGMP1
30 PSPA_RA
31 MANR_HU
32 LECG_CR
33 LECC_PO
34 PT0372
35 PT0374
36 MUSNKRP
37 MUSNKRP
38 MUSNKRP
39 A35917
C 100-91%
O P 90-81%
N O 80-71%
S S 70-61%
E I 60-51%
R T 50-41%
V I 40-31%
E O 30-21%
D N 20-11%
S 10- 0%
0010
0020
0030
0040
0050
0060
0070
0080
0090
0100
0110
0120
0130
0140

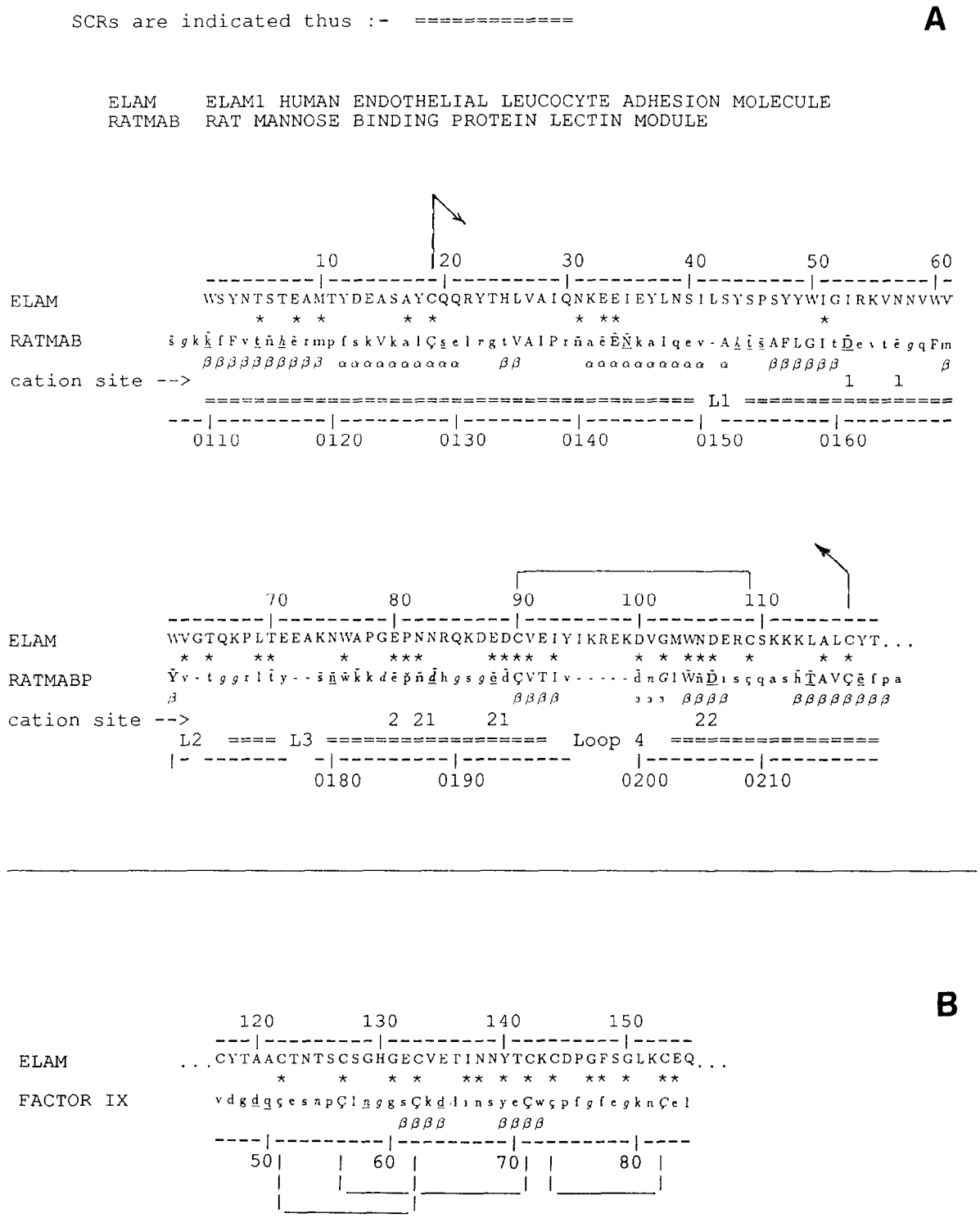

Fig. 2. Sequence alignments of the E-selectin sequence, together with nominations of structurally conserved regions as used in the modelling in COMPOSER, over the stretches of homology with (a) rat mannose-binding protein (RATMABP), and (b) human epidermal growth factor-like domain from human Factor IX. The RATMABP structural features are illustrated using the JOY typographic convention [20] thus: *italic* for positive ϕ angle; UPPER CASE for solvent inaccessible residues; lower case for solvent accessible residues; bold type for hydrogen bonds to main chain amide; underline for hydrogen bonds to main chain carbonyl; tilde (~) for side chain to side chain hydrogen bonds; çedilla indicates a disulphide bonded residue. The secondary structure is indicated beneath the alignment, and the cation coordinating residues below that. Conserved residues are indicated by asterisks. The residue numbering for RATMABP in Fig. 2(a) concurs with that in [10].

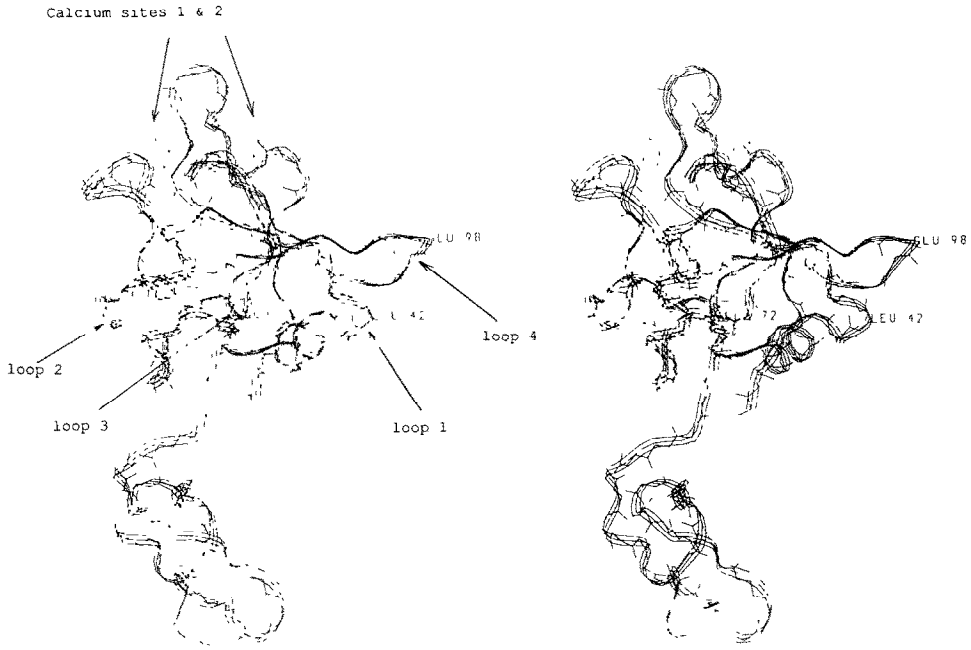

Fig. 3. Stereoscopic view of the backbone of the completed model with a superimposed ribbon trace for clarity. The dotted spheres indicate the calcium ions. The four loops which are the sites of insertions in the lectin domain are indicated. The relative disposition of the lectin and the EGF-like domains should be regarded as speculative.

indicated acceptable stereochemistry angles, but then the model's main chain is very similar to the rat protein experimental structure.

The areas which are likely to be different in fine detail are:
the exact conformations of the modelled loops;
the fine structure of the calcium coordination sites;
the relative disposition of the lectin and EGF-like domains.

3.1. *Calcium coordination sites*

The author is not aware of any available information concerning the affinity for calcium of the binding site(s) in human E-selectin.

It can be seen from Fig. 4a that whilst most of the coordinating residues at cation binding site 1 are conserved (or conservatively mutated) between the known and modelled sequences, Asp-161 in the rat is mutated to an isoleucine in E-selectin. This side chain obviously cannot partake in coordinating a metal ion, and the implication of this for E-selectin is unclear. It is likely that the binding affinity of this site for calcium is considerably reduced. Another possibility is that the alignment of Fig. 2a is incorrect at this point, but this is thought to be unlikely, since it cannot be sensibly adjusted to improve this particular detail and conserve an oxygen-bearing side chain.

Whilst the coordination is lost from one residue, the possibility of a compensating gain is introduced by the mutation of Gly-190 to a glutamine in E-selectin; in the model this side chain is positioned 'above' site 1. Thus, a third possibility is that the calcium ion at this site may be shifted, and the side chains re-arranged somewhat to reform the coordination site at a slightly different location. Further data are needed to clarify this point.

The coordination at cation site 2 is shown in Fig. 4b; It is seen that all coordinating side chains are absolutely conserved in the model, except it seems that an additional stabilising residue may be gained through the mutation of Ile-207 to a glutamate. This may then hold a coordinating water molecule between its carboxylate and the calcium ion.

As noted by Weis et al. [4], the two cation sites are effectively linked: "Glu-193 binds to site 2 with its carboxylate and to site 2 with its main chain carbonyl, and Asp-206 binds to site 2 with one of its carboxylate oxygens and its main chain carbonyl, while its other carboxylate oxygen atom forms a hydrogen bond with a water molecule that serves as a ligand in site 1." Both these residues are conserved in the E-selectin model, and the water molecule also is likely to be.

3.2. *Sugar binding considerations*

An analysis of the sequence alignment for the C-type lectins in the light of the known structure of the rat mannose-binding protein suggests that certain surface residues which are moderately conserved yet not involved in metal coordination, may instead be sugar-binding residues. The difficulty is that the various lectin sequences in Fig. 1 are expected to exhibit a variety of

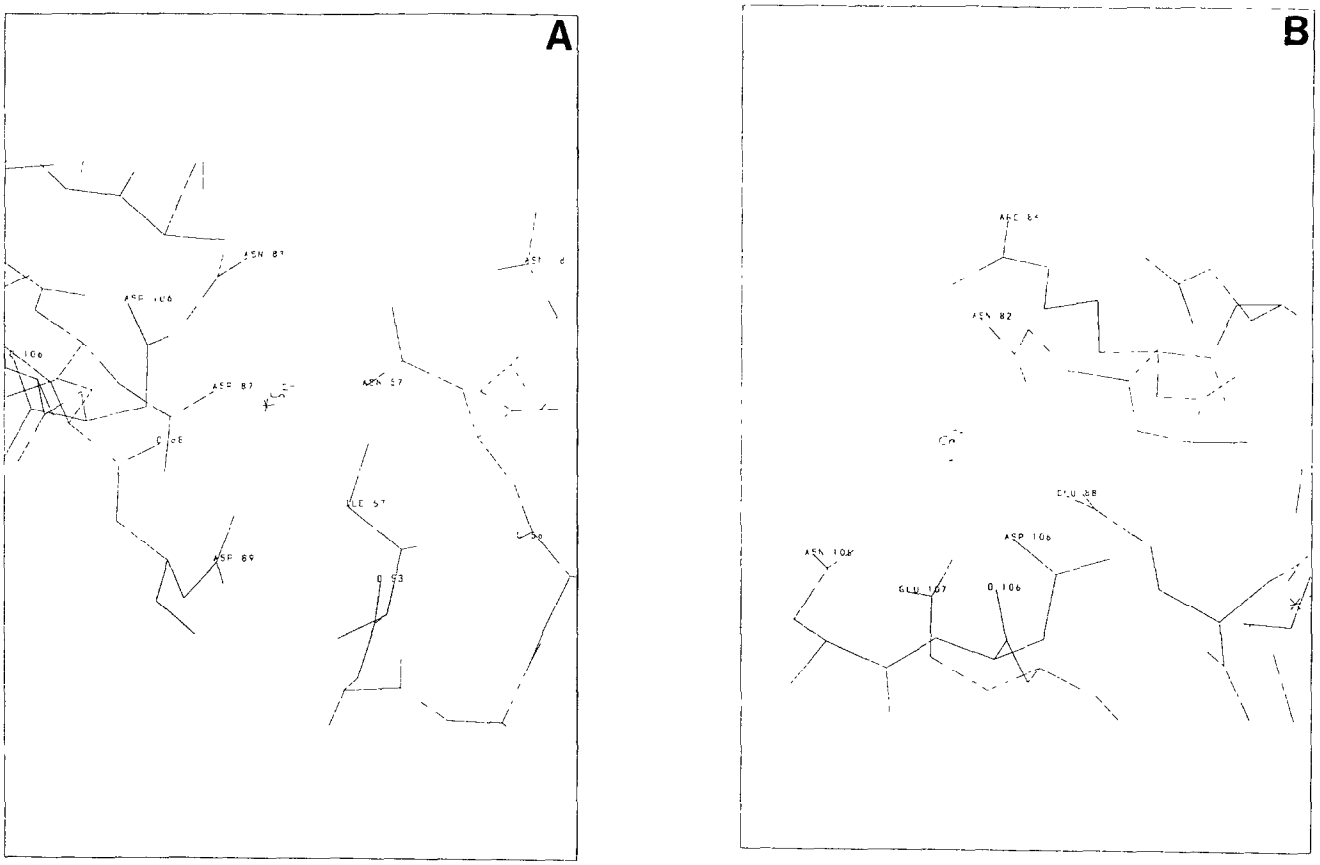

Fig. 4. Calcium binding sites in E-selectin model (a) site 1, and (b) site 2. The residue numbering corresponds to that of E-selectin as in Fig. 2a.

specificities for different sugar structures, and accordingly, we expect a degree of variation (rather than conservation) of liganding side chains. Nevertheless, Weis et al. [4] identified Glu-143, Gly-173, Pro-186, and Gly-202 in the rat sequence as likely such candidates.

In the alignment of Fig. 1 we see that the position corresponding to Glu-143 is an aspartate or glutamic acid with only one exception among the 39 sequences. The *cis*-Proline-186 is conserved in all sequences except the natural killer group of proteins. Its carbonyl is virtually at the 'tip' of the molecule, and could be involved in ligand binding.

Indeed, most of the main chain carbonyls and amides around the large calcium-stabilised loop are solvent-exposed. Taken together with the exposed atoms of the side chains around the calcium ions, the result is a region excessively rich in hydrogen-bonding potential, a factor likely to contribute to the sugar binding potential of this domain.

3.3. *Conclusions*

The three-dimensional structure of the carbohydrate binding domain of human E-selectin, comprised of its lectin domain and its EGF-like module, have been modelled on the basis of the X-ray structure of the C-type lectin domain of rat mannose-binding protein and of the solution NMR structure of the EGF-like domain of human Factor IX. One of the two potential calcium-binding sites is modified from the rat mannose-binding protein, and probably has reduced affinity for calcium ions.

At this stage, little could be deduced about the nature of the binding of E-selectin to its oligosaccharide ligand. However, this model should provide a useful starting point for interpreting any new data that may become available from NMR and other experiments.

Acknowledgements: Thanks are due to Drs. M. Johnson and N. Srinivasan of Birkbeck College, and Dr. P. Finn of Pfizer, for use of alignment programmes and assistance with some of the computational details, Dr. J.P. Overington of Pfizer for use of his JOY programme, and to Dr. R.L. Laskowski of University College, London for use of the PROCHECK programme. This work was supported by Pfizer Central Research.

REFERENCES

[1] Collins, T., Williams, A., Johnston, G.I., Kim, J., Eddy, R., Shows, Gimbrone Jr., M.A. and Bevilacqua, M.P. (1991) J. Biol. Chem. 266, 2466–2473.
[2] Drickamer, K. (1988) J. Biol. Chem. 263, 9557–9560.
[3] Lowe, J.B., Stoolman, L.M., Nair, R.P., Larsen, R.D., Berhend, T.L. and Marks, R.M. (1990) Cell 63, 475–484.
[4] Weis, W.I., Kahn, R., Fourme, R., Drickamer, K. and Hendrickson, W.A. (1991) Science 254, 1608–1615.

[5] Sutcliffe, M.J., Haneef, I., Carney, D. and Blundell, T.L. (1987a) Protein Eng. 1, 377–384.
[6] Sutcliffe, M.J., Hayes, F.H.F. and Blundell, T.L. (1987b) Protein Eng. 1, 385–392.
[7] Blundell, T.L., Carney, D., Gardener, S., Hayes, F., Howlin, B., Hubbard, T., Overington, J.P., Singh, D.A., Sibanda, B.L. and Sutcliffe, M. (1988) Eur. J. Biochem. 172, 513–520.
[8] Sali, A., Overington, J.P., Johnson, M.S. and Blundell, T.L. (1990) Trends Biochem. Sci. 15, 235–240.
[9] Baron, M., Norman, D.G., Harvey, P.A., Handford, P.A., Mayhew, M., Tse, A.G.D., Brownlee, G.G. and Campbell, I.D. (1992) Protein Sci. 1, 81–90.
[10] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[11] Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T. and Weng, J., in: Crystallographic Databases – Information Content, Software Systems, Scientific Applications, (F.H. Allen, G. Bergerhoff and R. Sievers, Eds.), Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, 1987, pp. 107–132.
[12] Bleasby, A.J. and Wootton, J.C. (1990) Protein Engineering 3, 153–159.
[13] Feng, D.-F. and Doolittle, R.F. (1987) J. Mol. Evol. 25, 351–360.
[14] Overington, J.P., Johnson, M.S., Sali, A. and Blundell, T.L. (1990) Proc. R. Soc. Lond. B. 241, 132–145.
[15] Jones, T.A. (1978) J. Appl. Cryst. 11, 268–272.
[16] Weis, W.I., Drickamer, K. and Hendrickson, W.A. (1992) Nature 360, 127–134.
[17] Drickamer, K. (1992) Nature 360, 183–186.

NOTE ADDED IN PROOF

Since initial submission of this paper a full report of a complex of the rat mannose-binding protein with an oligosaccharide ligand has been published [16]. It reveals that the 3- and 4-hydroxyl groups of a terminal mannose are directly ligated to calcium ion 2 to complete the formation of an 8-coordinate calcium ion complex. The simultaneous publication [17] of mutagenesis studies confirms the interpretation of the X-ray results by mutating key residues to successfully engineer specificity for galactose rather than mannose. In the selectins all coordinating side chains at site 2 are conserved with respect to the mannose-binding proteins. This suggests that the 2- and 3-hydroxyl of L-fucose binding to this calcium site may be partly responsible for the specificity of E-selectin for the s-Le^x ligand. If this is the case then an additional site may be expected to bind the terminal sialic acid.