FEBS LETTERS

AN INTERNATIONAL JOURNAL FOR THE RAPID PUBLICATION OF SHORT REPORTS IN BIOCHEMISTRY, BIOPHYSICS AND MOLECULAR CELL BIOLOGY

(a) P-selectin lectin domain

(b) L-selectin lectin domain

(c) E-selectin lectin domain

ublished by Elsevier on behalf of the Federation of European Biochemical Societies

olume 363, Number 1,2, 17 April 1995
EBLAL 363 (1,2) 1–210

ISSN 0 014 5793

Cover figure taken from: Kuo-Chen Chou, The convergence-divergence duality in lectin domains of selectin family and its implications, ***FEBS Letters***, 1995, **363**, 123-126 (Figure 2).

Elsevier Science B.V., PO Box 211, 1000 AE
Amsterdam, The Netherlands

The convergence–divergence duality in lectin domains of selectin family and its implications

Kuo-Chen Chou*

Computer-Aided Drug Discovery, Upjohn Laboratories, Kalamazoo, MI 49007-4940, USA

Received 17 February 1995

Abstract A comparison of the three-dimensional structures of P-, L-, and E-selectin lectin domains reveals that there is a convergence–divergence duality for the 77–107 polypeptide in the three domains; i.e. part of the peptide is folded into a closely similar conformation, and part of it into a highly different one. Since the 77–107 residues are associated with the putative binding sites of the selectin family for ligands, this kind of duality might well reflect the common character of ligands to the selectin family as well as the specificity to each of their respective receptors. The finding may be of use for rationally designing selectin inhibitors with a given specificity and possible antiadhesion drugs.

Key words: Selectin P-, L-, E-; Cell adhesion receptor; Sialyl Lewis^x binding site; Fucose binding site

The selectin family of carbohydrate-binding proteins consists of P-selectin [1], L-selectin [2–4], and E-selectin [5–7]. P-selectin is found in alpha granules of platelets and Weible–Palade bodies of endothelial cells [8]. L-selectin is found on leukocytes and is involved with the trafficking of lymphocytes to peripheral lymphoid tissue [9] and with acute neutrophil-mediated inflammatory responses [10]. E-selectin is an endothelial adhesion molecule that is induced by various inflammatory stimuli [11] and that recognizes the neutrophil and monocyte cell surface carbohydrate, sialyl Lewis^x (sLe^x). The members of selectin family share a common structural organization; i.e. they all contain an amino-terminal lectin domain (lec), followed by an epidermal growth factor-like element (EGF), a variable number of complement regulatory repeat units (cr), a single membrane-spanning region, and a carboxyl-terminal cytoplasmic domain. The selectin family has a singular role in initiating the 'adhesion cascade' by which leukocytes move from the blood stream into tissue [12–15]. Adhesion proteins called selectins make the endothelial cells sticky to sLe^x-containing white blood cells, which adhere. After binding to the endothelial cells, the leukocytes are able to squeeze past gaps between them and enter the adjoining tissue, where they can help repair injury, but may also sometimes do damage. Therefore, the selectin family has become an important focus of antiadhesion drug design.

The lectin domain is analogous to other C-type Ca^{2+}-dependent animal lectins, suggesting where the binding site for sLe^x is located. In order to understand the biochemical mechanism of the cell-adhesion process involving binding sLe^x to the selectins it is necessary to find at least the 3D (dimensional) structures of their lectin domains. Recently, the structure of E-selectin lectin domain has been modeled [5,6] and determined by X-ray technology [7]. As expected, the X-ray structure of E-selectin lectin domain is very similar to the lectin domain of the mannose binding protein (MBP) determined by Weis et al. [16] since both E-selectin and MBP belong to the C-type lectin family [17]. Meanwhile, as an independent approach, the lectin domain of MBP [16] has been used as a structural template for modelling the P-, L-, and E-selectin lectin domains. The procedure can be briefed as follows. First, the sequence alignment of the human P-selectin [1], L-selectin [18] and E-selectin [19] with MBP [20] was performed with the pattern-induced multi-sequence alignment (PIMA) algorithm [21], in which the secondary structure-dependent gap penalties were employed for use in comparative protein modelling. The result of alignment is given in Fig. 1, where the second structural elements of the MBP are indicated on the top of the amino acid sequences by the symbols: h, α-helix; e, β-strand; t, β-turn; and c, coil. Thus, the starting conformations of P-, L-, and E-selectin lectin domains for energy refinement can be generated based on the atomic coordinates of MBP [16] and the sequence alignment of Fig. 1. However, the existence of insertion or deletion in their sequence alignment has made the problem by no means a straightforward or trivial one. As shown in Fig. 1, there are three insertions for each of the three selectin sequences that occur at residues 44–46, residue 63, and residues 96–100 (Fig. 1). Therefore, a smooth connection treatment is needed for these insertion sites. Besides, since φ in proline is always fixed at $-75°$, to deal with such a rigidity, for any proline in P-, L-, or E-selectins whose counterpart in the MBP is not a proline, a smooth connection operation is also needed by adjusting its neighboring residues. The detailed procedure and the technique for performing the desired smooth connection were described in [22–24]. The side-chain dihedral angles of the lectin chain were assigned according to the data on energy-minimized amino acid residues [25], thus completing the generation of the starting conformations for the P-, L-, and E-selectin lectin domains, respectively. Subsequently, each of these starting conformations derived from the MBP templet was subject to energy minimization. The energy was computed with the updated version (ECEPP/2) of the ECEPP algorithm (Empirical Conformational Energy Program for Peptides [26–27]). The total energy is the sum of the electrostatic, non-bonded, hydrogen bond, and torsional energies. A general unconstrained optimizing algorithm was used for minimization [28]. The computations were carried out on an Silicon Graphics IRIS Indigo work station (Elan 4000) at Upjohn Laboratories. Energy minimization was carried out according to the procedure as described in [24].

The final energies thus obtained for P-, L-, and E-selectin lectin domains are -556.5, -667.1, and -533.5 kcal/mol, respectively. The calculated rms (root-mean-square) deviations

*Corresponding author. Fax: (1) (616) 385 7373.

SSDI 0014-5793(95)00240-5

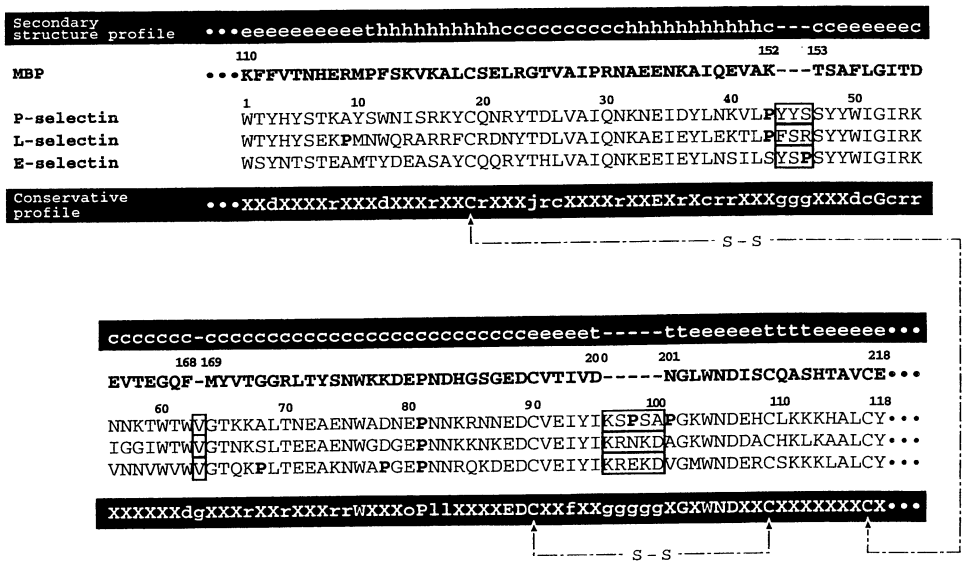

Fig. 1. Sequence alignment of human P-selectin [1], L-selectin [18], E-selectin [19], and MBP, i.e. rat mannose-binding protein [20]. The sequences were aligned with the pattern-induced multi-sequence alignment (PIMA) algorithm [21], in which the secondary structure-dependent gap penalties were employed for use in comparative protein modelling. The second structural elements of the MBP are indicated on the top of the amino acid sequences and are labeled as follows: h, α-helix; e, β-strand; t, β-turn; and c, coil. In the conservative profile: upper-case characters, one-letter amino acid codes; lower-case characters, designated amino acid classes as defined by Fig. 2 of ref. [21]; X, wild-card character (one amino acid of any type); and g, gap character.

for the best backbone structure fit between P- and L-, between P- and E-, and between L- and E-selectin lectin domains are 3.10 A, 2.48 A, and 3.07 A, respectively. The resulting overlay of these backbone structure is shown in Fig. 2, where for distinction P-selectin lectin domain is drawn in red, L-selectin lectin domain in green, and E-selectin lectin domain in blue. It can be seen from Fig. 2 that the three lectin domains of the selectin family possess all the typical features observed in the lectin domain of MBP [16]; i.e. they all contain two α-helices, and two antiallel β-sheets of which one is formed by two strands (strands 1 and 5) and the other by three (strands 2, 3, and 4). The sequence distributions of these secondary structures in the three domains are listed in Table 1, where for facilitating comparison the corresponding distribution in MBP is also given. It can be seen from Table 1 that compared with the lectin domain of MBP, the lengths of α-helices in P-, L-, and E-selectin lectin domains remain the same, but the lengths of β-strands become shorter. This is due, in part, to the various insertions and deletions. The first major insertion in the P-, L-, and E-selectin sequences occurs at residues 44–46 (Fig. 1), which increases the loop size between the second α-helix and the second β-strand. The other major insertion occurs at residues 96–100, which increases the loop size between the third and fourth β-strands.

The other common feature is that there are two disulfide bonds, formed by the cystine pair (residues 19 and 117), and the cystine pair (residues 90 and 109), in each of the three structures. This is reflected by the following fact: The S–S distances in the P-, L-, and E-selectin lectin domains between Cys^{19} and Cys^{117} are 2.00, 2.02, and 2.07 A, respectively, while those between Cys^{90} and Cys^{109} are 2.04, 1.93, and 2.04 A. All these distances are very close to 2.04 A, the standard length for the S–S bond as described in ECEPP25. The disulfide bond which covalently links Cys^{19} in helix 1 to Cys^{117} in strand 5 provides a tether to hold the helix bent to the strand, and the disulfide between the Cys^{90} and Cys^{109} provides another tether to restrict the conformation of the loop between stands 3 and 4.

On the other hand, it can be seen from Fig. 2 that the backbone conformations of the three structures follow a similar track from residues 77 to 93 (with a rms deviation less than 0.8 A), but diverge remarkably from residues 94 to 103. The part following a similar track is located at the top of Fig. 2 and extended down to the β-strand 3, while the divergent part located in the loop between β-strands 3 and 4. As a result of such a divergence in the loop region, the distance between the C^{α} atom of residue 98 in the P-selectin lectin domain and that in the L-selectin lectin domain is as large as 10.3 A, and the

Table 1
Sequence distribution of secondary structures in the lectin domains of P-selectin, L-selectin, E-selectin, and MBP

Lectin domain	Helix-1	Helix-2	Strand-1	Strand-2	Strand-3	Strand-4	Strand-5
P-selectin	Trp^{12}–Asn^{21}	Lys^{32}–Leu^{42}	Thr^{2}–Tyr^{5}	Tyr^{48}–Ile^{51}	Cys^{90}–Glu^{92}	Trp^{104}–His^{108}	Ala^{115}–Tyr^{118}
L-selectin	Trp^{12}–Asp^{21}	Lys^{32}–Leu^{42}	Trp^{1}–Tyr^{5}	Tyr^{48}–Ile^{51}	Cys^{90}–Ile^{95}	Trp^{104}–Ala^{108}	Ala^{115}–Tyr^{118}
E-selectin	Tyr^{12}–Gln^{21}	Lys^{32}–Leu^{42}	Trp^{1}–Thr^{5}	Tyr^{48}–Ile^{51}	Cys^{90}–Ile^{93}	Trp^{104}–Arg^{108}	Ala^{115}–Tyr^{118}
MBP[a]	Phe^{121}–Glu^{130}	Ala^{141}–Ala^{151}	Lys^{110}–Met^{119}	Ala^{155}–Thr^{160}	Cys^{195}–Val^{199}	Leu^{203}–Ser^{208}	His^{213}–Glu^{218}

[a]From Weis et al. [16].

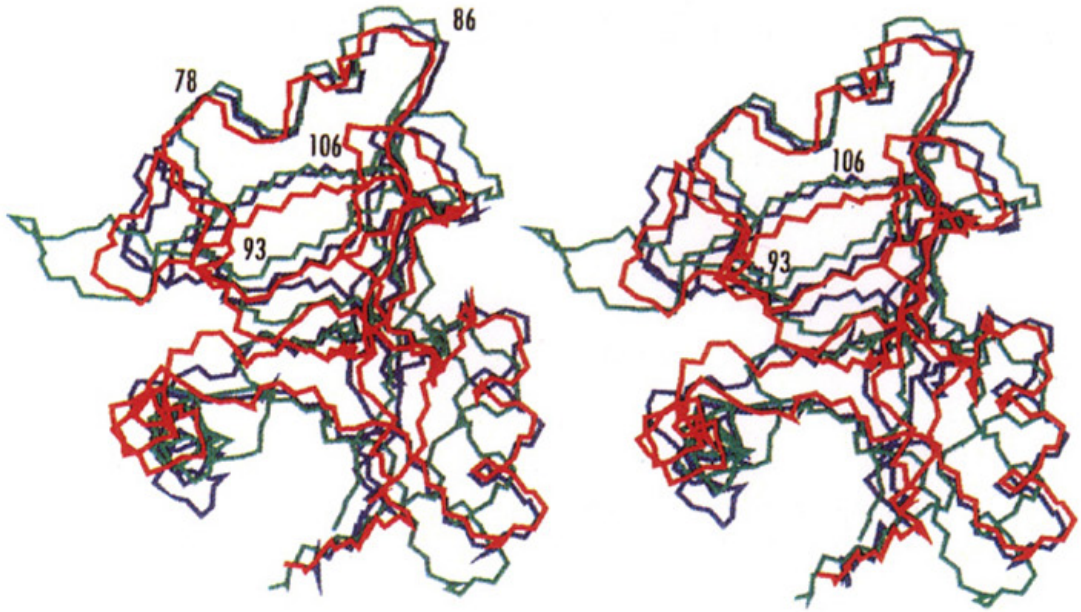

Fig. 2. Stereoscopic drawing to show the overlay of lectin domains of P-selectin (red), L-selectin (green), and E-selectin (blue). The overlay was achieved by means of a best molecular fit for all the backbone atoms N, C^{α}, and C′. It can be seen that the backbone conformations of the three structures follow almost an identical track from residues 77 to 93 (at the top), but suddenly diverge dramatically from residues 94 to 103 (in the loop between β-strands 3 and 4). This presents a dual feature for the folding of the 77–107 sequence, around which the putative binding sites for ligands are located. See the text for further explanation.

corresponding distance between P- and E-selectin lectin domains is 5.4 A, while that between L- and E-selectin lectin domains 9.3 A. Since residue 98 is located at the middle of the loop, these deviation quantities in its C^{α} atom may reflect the divergent extent of the loop in the three domains. This finding is quite consistent with the recent observation by Ma et al. [29]. According to their report that the binding site for ligand is very likely formed by residues 77–107 of the selectin molecules. In order to bind their ligands, all these three selectin molecules need two binding sites: one is for fucose and the other for sLe^{x}. This requires a common character in conformation around the binding region. On the other hand, the ligand of each different selectin molecule usually has its own specificity, and this will require a special character for its binding region as well. To realize such a relationship of the unity of opposites in the selectin family, it requires a structural *convergence–divergence duality*, i.e. an existence of both the closely similar and highly different conformations around the ligand-binding region of the three lectin domains. The convergent part is responsible for binding the chemical group common to all ligands of the selectin family, while the divergent part is responsible for binding the chemical group special to the ligand of a given selectin. Such a deduction is also consistent with the recent finding by Erbe et al. [6] and Graves et al. [7] that the loop between β-strands 3 and 4 of E-selectin lectin domain is a sensitive region for binding ligands. Since the conformation around the binding region for ligands can provide useful insights for revealing the binding mechanism (see e.g. [30]), the predicted structures of the three lectin domains, particularly the unique feature of their convergence–divergence duality around the binding site, can be of immediate use for the design of potential inhibitors against adhesion proteins, leading to find antiadhesion drugs.

Acknowledgements: Illuminative discussions with J.G. Geng, D. Asa, R.L. Heinrikson, and Donald Anderson are gratefully acknowledged. I would like to thank Joan Baker for helping to draw Fig. 1. I would particularly like to express special gratitude to Bi-Kun Luo. Without her encouragement I would not have been able to overcome the difficulties that occurred during this study.

References

[1] Johnston, G.I., Cook, R.G. and McEver, R.P. (1989) Cell 56, 1033–1044.
[2] Siegelman, M.H., van de Rijn, M. and Weissman, I.L. (1989) Science, 243, 1165–1172.
[3] Siegelman, M.H. and Weissman, I.L. (1989) Proc. Natl. Acad. Sci. USA 86, 5562–5566.
[4] Lasky, L.A., Singer, M.S., Yednock, T.A., Dowbenko, D., Fennie, C., Rodriguez, H., Nguyen, T., Stachel, S. and Rosent, S.D. (1989) Cell 56, 1045–1055.
[5] Mills, A. (1993) FEBS Lett. 319, 5–11.
[6] Erbe, D.V., Wolitzky, B.A., Presta, L.G., Norton, C.R., Ramos, R.J., Burns, D.K., Rumberger, J.M., Rao, B.N.N., Foxall, C., Brandly, B.K. and Lasky, L.A. (1992) J. Cell Biol. 119, 215–227.
[7] Graves, B.J., Crowther, R.L., Chandran, C., Rumberger, J.M., Li, S., Huang, K.S., Presky, D.H., Familletti, P.C., Wolitzky, B.A. and Burns, D.K. (1994) Nature 367, 532–538.
[8] Bonfanti, R., Furie, B. and Wagner, D. (1989) Blood, 73, 1109–1112.
[9] Gallatin, M., St. John, T.P., Siegelman, M., Reichert, R., Butcher, E.C. and Weissman, I.L. (1986) Cell 44, 673–680.
[10] Watson, S.R., Fennie, C. and Lasky, L.A. (1991) Nature 349, 164–167.
[11] Bevilacqua, M.P.A., Pober, J.S., Mendrick, D.L., Cotran, R.S. and Gimbrone Jr., M.A. (1987) Proc. Natl. Acad. Sci. USA, 84, 9238–9242.
[12] Butcher, E.C. (1991) Cell 67, 1033–1036.
[13] Lasky, L.A. (1992) Science 258, 964–969.
[14] Lasky, L.A., Singer, M.S., Dowbenko, D., Imai, Y. Henzel, W.J., Grimley, C., Fennie, C., Gillett, N., Watson, S.R. and Rosent, S.D. (1992) Cell 69, 927–938.
[15] Bevilacqua, M.P.A. (1993) Annu. Rev. Immunol., 11, 767–804.
[16] Weis, W.I., Kahn, R., Fourme, R., Drickamer, K. and Hendrickson, W.A. (1991) Science 254, 1608–1615.
[17] Drickamer, K.J. (1988) Biol. Chem. 263, 9557–9560.

[18] Camerini, D., James, S.P., Stamenkovic, I. and Seed, B. (1989) Nature 342, 78–82.
[19] Bevilacqua, M.P., Stengelin, S., Gimbrone Jr., M.A., and Seed, B. (1989) Science 243, 1160–1165.
[20] Drickamer, K., Dordal, M.S. and Reynolds, L. (1986) J. Biol. Chem. 261, 6878–6887.
[21] Smith, R.F. and Smith, T.F. (1992) Protein Eng. 5, 35–41.
[22] Chou, K.C., N'emethy, G., Pottle, M. and Scheraga, H.A. (1989) J. Mol. Biol. 205, 241–249.
[23] Carlacci, L., Chou, K.C. and Maggiora G.M. (1991) Biochemistry 30, 4389–4398.
[24] Chou, K.C. and Carlacci, L. (1991) Proteins: Structure Function and Genetics 9, 280–295.
[25] V'asquez, M., N'emethy, G. and Scheraga, H.A. (1983) Macromolecules 16, 1043–1049.
[26] Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A. (1975) J. Phys. Chem. 79, 2361–2381.
[27] N'emethy, G., Pottle, M.S. and Scheraga, H.A. (1983) J. Phys. Chem. 87, 1883–1887.
[28] Gay, D.M. (1983) Assoc. Comput. Mach. Trans. Math. Software 9, 503–524.
[29] Ma, L., Chou, K.C., Kaytes, P.S., Leone, B.A., Toombs, C.F., Staples, D.J., Babcock, M.S., Fischer, H.D., Smith, C.W. and Geng, J.G. (1995) Biochim. Biophys. Acta, in press.
[30] Chou, K.C. (1992) J. Mol. Biol. 223, 509–517.