Structural Characterization of EMS16, an Antagonist of Collagen Receptor (GPIa/IIa) from the Venom of *Echis multisquamatus*[†,‡]

Katsunori Horii,[§] Daiju Okuda,[||] Takashi Morita,[||] and Hiroshi Mizuno*[,§]

Department of Biochemistry, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan, and Department of Biochemistry, Meiji Pharmaceutical University, 2-522-1, Noshio, Kiyose, Tokyo, 204-8588, Japan

Received May 29, 2003; Revised Manuscript Received September 9, 2003

ABSTRACT: Snake venoms contain a number of hemostatically active C-type lectin-like proteins (CLPs), which affect the blood coagulation system, endothelial cells, and platelets. CLPs have broad similarities in structure and possess distinct biological functions. EMS16, a CLP from *Echis multisquamatus* venom, which is a potent and selective inhibitor of the collagen receptor, glycoprotein Ia/IIa (integrin $\alpha 2\beta 1$), has been used in the present study to examine structure−function relationships in venom CLPs by X-ray crystallography. The structure of EMS16, determined at a resolution of 1.9 Å, revealed a heterodimer involved with domain swapping of the central loop as observed in the structures of other CLPs. A part of the glycan was observed and identified as *N*-acetyl-D-glucosamine (GlcNAc) in the electron density map at Asn21 of subunit B, an expected glycosylation site. EMS16 had a unique, positively charged electrostatic potential patch on the concave surface that may qualify as a site for interaction with the I-domain of the glycoprotein Ia/IIa.

Snake venoms contain a wide variety of protein components affecting haemostasis. One such group consists of the C-type lectin-like proteins (CLPs).[1] The name of this family is derived from the primary structure and disulfide bond arrangement of the coagulation factor IX/X-binding protein (IX/X-bp) from habu snake venom (*1, 2*). Up to the present, various CLPs that affect either coagulation factors or platelets have been sequenced and characterized. CLP is an heterodimeric protein consisting of homologous subunits A (α) and B (β) linked by a disulfide bond. The amino acid sequence of each subunit shows homologies with carbohydrate recognition domains of the classic C-type lectins, although lectin activity has not been demonstrated. We first determined the crystal structure of a CLP using IX/X-bp (*3*). The structure showed that the two subunits were intimately associated by domain swapping, and this dimerization resulted in the disruption of the lectin active site and the creation of a concave surface serving as an expected coagulation factor-binding site. Examination of the crystal structure of the complex between factor X-binding protein (X-bp) and the Gla domain of factor X revealed that the binding site is, in fact, the concave surface (*4*). Structures of CLPs solved so far by X-ray crystallography (*3*−*11*) show that they are very similar to each other but differ in their functions. This observation suggests that electrostatic, as well as size and shape matching, interactions are important for binding with ligands. The concave surface, which is the binding site of the Gla domain for X-bp, as mentioned previously, has been considered the common binding site even for different ligands. In this connection, we previously showed that the electrostatic potentials on the concave surface are considerably different among IX/X-bp, bitiscetin, and botrocetin (*8*). Thus, further crystallographic studies of CLPs, particularly those characterized as having different functions, are deemed necessary for better understanding of structure−function relationships in CLPs.

Recently, EMS16 from *Echis multisquamatus* venom has been identified as a novel CLP antagonist that acts as a potent and selective inhibitor of a collagen receptor, GPIa/IIa (also commonly referred to as integrin $\alpha 2\beta 1$) (*12*). The latter is a major receptor expressed in many different cell types for native collagen types I and IV and laminin (*13*). GPIa/IIa, expressed on platelet membranes, plays a role in the initial event in vascular response to injury with another collagen receptor, GPVI. However, it is still not clear what the mechanism of the interaction is between GPIa/IIa and GPVI during platelet aggregation because there are numerous complexes produced during signal transduction. Hence, an antagonist to GPIa/IIa would be a useful tool in elucidating the mechanism of collagen-induced platelet aggregation. Currently, the GPVI-agonist, convulxin (*14*, *15*), and the GPIa/IIa agonist, aggretin (rhodocytin) (*16*, *17*), are used in such studies.

The crystal structure of the I-domain of GPIa/IIa, containing the metal-ion dependent adhesion site (MIDAS) (also known as the critical domain in the α-subunit for collagen binding), has been determined in the absence of ligand (*18*)

[†] This work was supported in part by Scientific Research Grants-in-Aid (T.M.) and by Special Coordination Funds for Promoting Science and Technology (H.M.) from the Ministry of Education, Science, Sports and Culture of Japan.

[‡] The atomic coordinates of EMS16 are available from the Protein Data Bank as entry 1UKM.

* To whom correspondence should be addressed. Telephone: +81-29-838-7877. Fax: +81-29-838-7408. E-mail: mizuno@affrc.go.jp.

[§] National Institute of Agrobiological Sciences.

[||] Meiji Pharmaceutical University.

[1] Abbreviations: CLP, C-type lectin-like protein; GPIa/IIa, glycoprotein Ia/IIa; GlcNAc, *N*-acetyl-D-glucosamine; IX/X-bp, coagulation factor IX/X-binding protein; X-bp, coagulation factor X-binding protein; GPVI, glycoprotein VI; MIDAS, metal-ion dependent adhesion site; RMS, root-mean-square; VWF, von Willebrand factor.

10.1021/bi034890h CCC: $25.00
Published on Web 10/10/2003

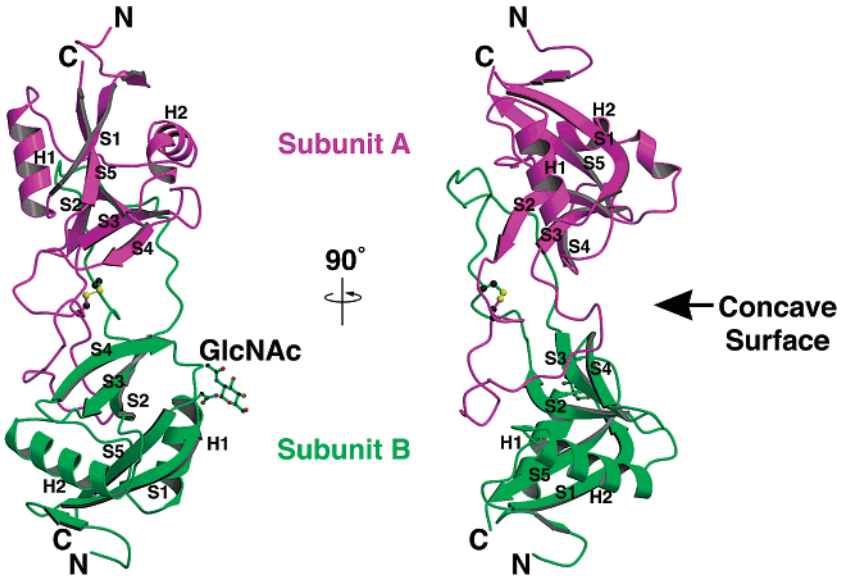

FIGURE 1: Overall structure of EMS16. The left ribbon diagram is viewed from the concave surface, and the one on the right is rotated 90° from the one on the left around the vertical axis. The subunits A and B are shown in magenta and green, respectively. The N and C termini and each secondary structural element are labeled for both subunits. The side chains of cysteine residues forming an interchain disulfide bond (Cys81 in subunit A and Cys74 in subunit B), the glycosylated Asn21 in subunit B, and the attached *N*-acetyl glucosamine are shown as ball-and-stick models. The figure was created with the programs MOLSCRIPT (*30*) and Raster3D (*31*).

and in the presence of a triple helical collagen peptide (*19*). The I-domain, containing about 200 amino acid residues, adopts a dinucleotide-binding fold that is very homologous to the A-domain of VWF and the I-domains of other integrins α1, αL, αM, and αX. In particular, GPIa/IIa is closely related to integrin α1β1 that binds to collagens I and IV with a different affinity. It is of interest, however, that EMS16 does not interact with the I-domain of integrin α1β1 (*12*).

The complete amino acid sequence of EMS16 has been recently determined (*20*) and found to be approximately 30−50% identical to those of other CLPs (Figure 2). In this paper, we report on the three-dimensional structure of EMS16 at a resolution of 1.9 Å.

EXPERIMENTAL PROCEDURES

Crystallization and Data Collection. EMS16 was isolated from the crude venom of *E. multisquamatus* and crystallized as previously described (*20*). Briefly, the purified EMS16 was crystallized by the hanging-drop vapor-diffusion method at 293 K. Some diamond-shaped crystals were obtained in a week by mixing 1 μL of protein solution (5 mg of protein/mL in 50 mM Tris-HCl buffer, pH 7.5) and 3 μL of precipitant solution (0.04 M potassium dihydrogen phosphate, 16% PEG8000, 20% glycerol) against 0.5 mL of reservoir solution. X-ray diffraction data on a crystal were collected at 120 K using an R-AXIS IV^{++} imaging-plate detector mounted on a copper rotating-anode generator equipped with an Osmic mirror system (Rigaku). The diffraction was measured at 1.9 Å resolution and then integrated and scaled using the program CRYSTAL CLEAR (Rigaku). The data processing statistics are tabulated in Table 1.

Structure Determination. The structure of EMS16 was determined by the molecular replacement method with the program AMoRe (*21*) from the CCP4 suites (*22*). A search model was constructed by homology modeling based on the coordinates of coagulation factor IX-bp (*5*), using Insight II 2000 (Accelrys). After rotation and translation calculations over the resolution range of 10−4.0 Å, a clear solution was found with a correlation coefficient of 62.4% and *R* factor of 39.0%.

Table 1: Data Processing and Refinement Statistics

Crystallographic Data	
space group	$P2_12_12_1$
cell dimensions (Å)	$a = 46.57$, $b = 59.93$, $c = 115.74$
resolution range (Å)	30.0−1.90 (1.97−1.90)[a]
measured reflections	91401
unique reflections	26199
average $I/\sigma(I)$	16.1 (5.6)
completeness (%)	99.7 (99.7)
mosaicity (deg)	0.6
R_{merge} (%)[b]	3.1 (11.6)
Refinement Statistics	
resolution range used in refinement (Å)	20.0−1.90
R factor (%)[c]	19.6
R_{free} (%)[d]	23.4
model	
no. of non-hydrogen protein atoms	2128
no. of glycan atoms	14 (GlcNAc)
no. of chloride ions	1
no. of glycerol molecules	4
no. of water molecules	250
average B factor (Å^2)	
subunit A	24.7
subunit B	28.0
glycosylated sugar	56.3
solvent	37.5
Ramachandran plot	
residues in most favored	205
residues in additional allowed	24
residues in generously allowed	1
residues in disallowed	0
RMS deviations	
bond lengths (Å)	0.008
bond angles (deg)	1.50

[a] Values in parentheses refer to the highest resolution shell. [b] $R_{merge} = \sum_h\sum_i|I(h, i) - \langle I(h)\rangle|/\sum_h\sum_i I(h, i)$, where $I(h, i)$ is the intensity value of the *i*th measurement of *h*, and $\langle I(h)\rangle$ is the corresponding mean value of $I(h)$ for all *i* measurements. [c] *R* factor $= \sum||F_o| - |F_c||/|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively. [d] R_{free} is the same as the *R* factor but for a 5% subset of all reflections.

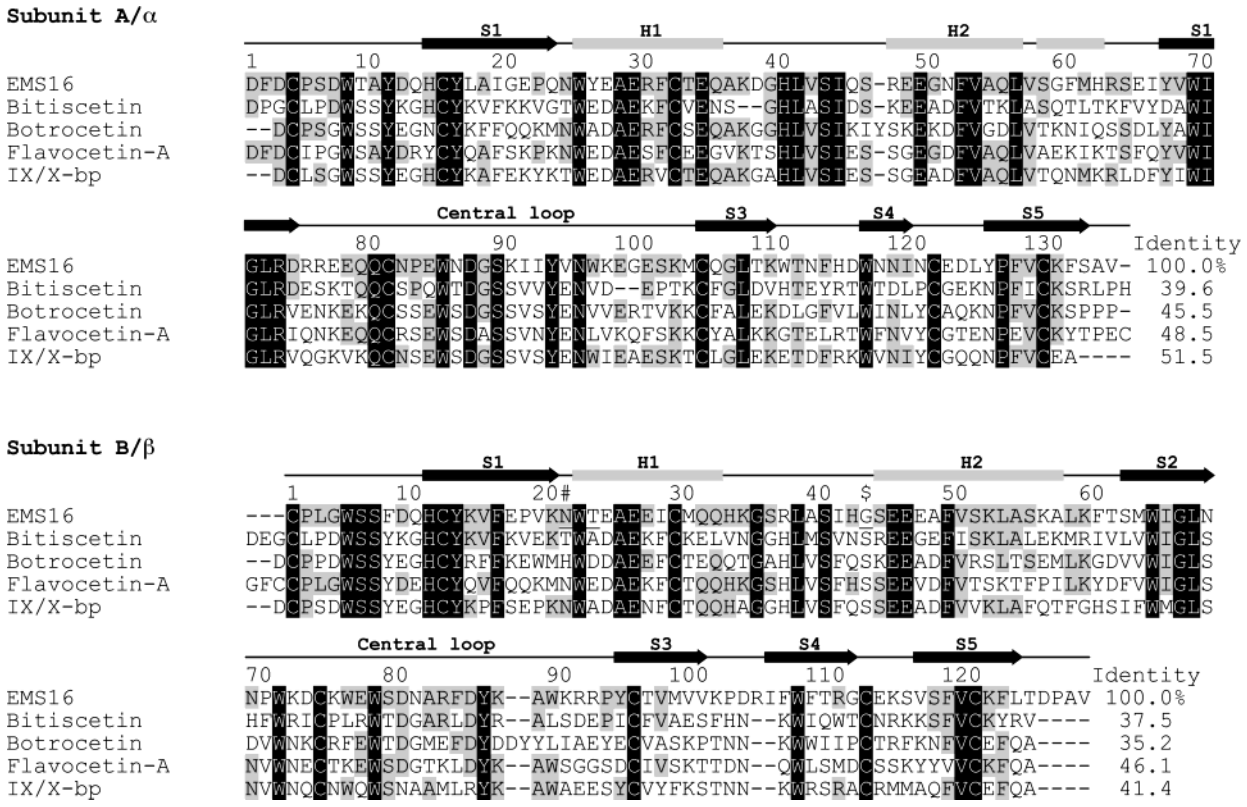

FIGURE 2: Amino acid sequence alignments of EMS16 and other venom CLPs. Sequences are aligned based on structural superposition. Identity indicates the differences in amino acid residues of CLPs in comparison with that of EMS16. The secondary structural elements are shown as arrows for β-strands and rectangles for α-helices. The symbols # and $ denote the glycosylation site (amino acid sequence NXT) and the mutated site (G43S in subunit B) of EMS16, respectively (see text for details).

Structure Refinement. All crystallographic refinements were carried out with the program CNS (*23*) using the maximum-likelihood target function. Throughout the refinement process, the reflection data in the 20−1.9 Å resolution range were used, except for 5% of the randomly selected reflections for cross-validation. The $2F_o - F_c$ and $F_o - F_c$ electron density and composite omit maps were generated by CNS and used to manually rebuild the model with XtalView (*24*). Refinements were performed with several cycles of torsion-angle simulated annealing, energy minimization, individual *B* factor refinement, and manual model rebuilding. When the value of the *R* factor dropped to 25%, the solvent molecules were gradually included in the model. Coordinate, topology, and parameter files for glycerol were obtained from the HIC-UP database (*25*). In the final model, the *R* factor and R_{free} converged to 19.6 and 23.4%, respectively. The structure of EMS16 was determined for most of the protein with some exceptions, namely, Asp1, Phe2, and Val134 of subunit A, and Pro126, Ala127, and Vla128 of subunit B. These undetermined residues are located in the N or C terminal regions of the subunits and may be unduly flexible. The stereochemistry of the model was checked with the program PROCHECK (*26*). Refinement statistics are listed in Table 1.

RESULTS AND DISCUSSION

Structural Overview. EMS16 consisted of two subunits, A and B, which were connected by an interchain disulfide bond between Cys81 of subunit A and Cys74 of subunit B (Figure 1). The overall features and secondary elements of EMS16 were very similar to those of the snake venom CLPs, such as bitiscetin, botrocetin, flavocetin-A, and IX/X-bp, reflecting the high degree of identity in the amino acid sequence (Figure 2). Each subunit was divided into a globular domain and a central loop region. The globular domain was comprised of residues 1−74 and 104−134 in subunit A and residues 1−68 and 95−125 in subunit B. It was further composed of two major helices (H1 and H2), five major β-strands (S1-S5), and two short β-strands. Subunits A and B were found dimerized by domain swapping (*3*, *5*, *27*) in the central loops of residues 75−103 of subunit A and residues 69−94 of subunit B. They made contacts with globular domains of the neighboring subunits mainly by hydrophobic interactions (Figure 1). Dimerization by domain swapping resulted in a concave surface in the middle of the two subunits, as has been observed in the structures of other CLPs (*3*−*11*).

In a recent study, it has been found that Asn21 of subunit B in EMS16 is glycosylated and that the molecular mass of the attached glycan is about 2.4 kDa (*20*). From the electron density, a part of the glycan was identified as *N*-acetyl-D-glucosamine (GlcNAc) covalently linked to Asn21. Because of poor electron density, it was not possible to find the continuous sugar residues connecting to GlcNAc. We attribute this to the high mobility of the attached sugars, which may be expected from the relatively high *B* factors of most of the atoms in the defined GlcNAc residue (Table 1).

The complete amino acid sequence of EMS16 is given in Figure 2 (*20*), with residue 43 of subunit B shown as glycine. However, the extra electron density and its shape suggested that this residue is not glycine but serine (Figure 3). Assuming that residue 43 is serine, the hydroxyl group would

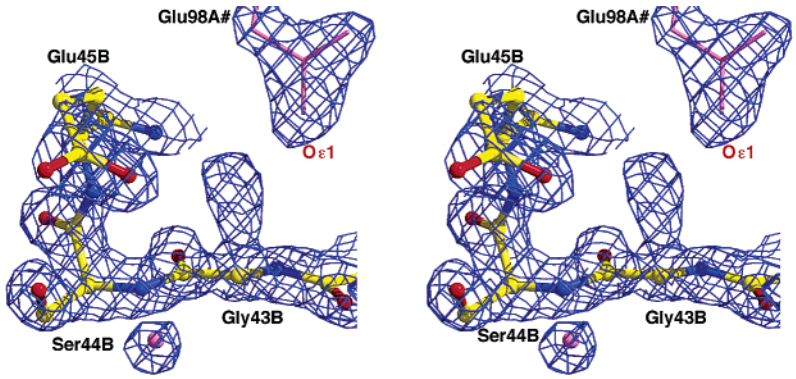

FIGURE 3: Stereoview of the region around residue 43 in subunit B with the $2F_o - F_c$ electron density map contoured at 2 σ. The side chain of residue 43 is omitted from the refinement and map calculations. The symmetry related molecule is shown in magenta. The figure was drawn with the programs XtalView (*24*) and Raster3D (*31*).

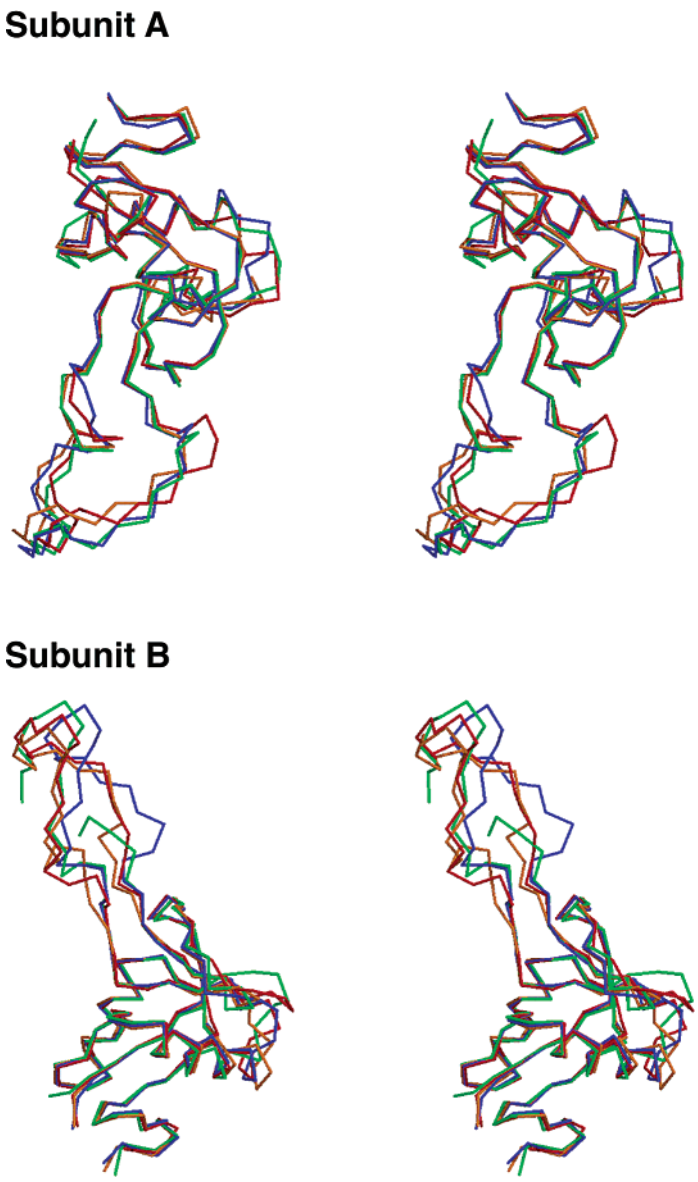

FIGURE 4: Superposition of subunits A and B of EMS16 over those of the other CLPs. Stereoview of the C_α traces of EMS16 (red), bitiscetin (orange), botrocetin (green), and IX/X-bp (blue) is shown after superposition using C_α atom pairs of major secondary elements (β-strands: S1-S5 and α-helices: H1-H3) in each of the subunits. This orientation corresponds to the molecule on the right in Figure 1.

make a hydrogen bond with a distance of 2.54 Å to the carboxyl group of Glu98 of subunit A in the symmetry-related molecule. This interpretation is supported by the fact that serine is highly conserved in this position among snake venom CLPs, as shown in Figure 2. Therefore, residue 43 of subunit B is considered likely to be serine. Such an amino acid difference in this position is probably due to single nucleotide polymorphism.

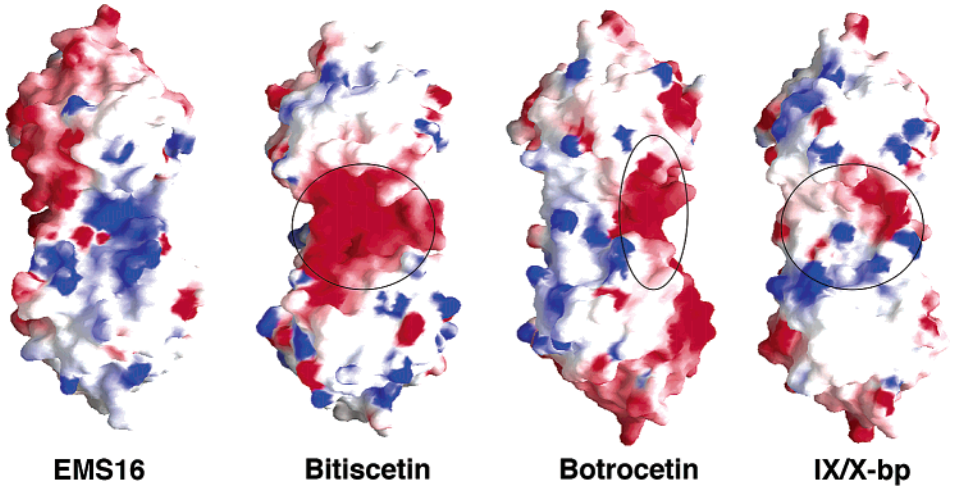

FIGURE 5: Comparison of the surface electrostatic potentials of EMS16 relative to the other CLPs. The drawn circles for bitiscetin, botrocetin, and IX/X-bp indicate the binding site of each target molecule. The charge was assigned to each normally ionizable residue (i.e. aspartate, glutamate, lysine, and arginine). Negatively charged regions are represented in red and positively charged regions in blue. This orientation corresponds to the molecule on the left in Figure 1. The figure was generated with GRASP (*28*).

Table 2: Comparison of Structures between EMS16 and the Other CLPs (Å)[a]

	bitiscetin	botrocetin	IX/X-bp
subunit A	0.54 (0.43)	0.72 (0.52)	0.53 (0.45)
subunit B	1.01 (0.71)	1.12 (0.73)	0.70 (0.54)
all	1.11 (0.92)	1.97 (1.64)	1.72 (1.53)

[a] The RMS deviations and average differences (in parentheses) are given for bitiscetin, botrocetin, and IX/X-bp after least squares superposition using C_α atoms of major secondary elements (α-helices: H1-H3 and β-strands: S1-S5) in each or both subunits.

Comparison with Other C-Type Lectin-Like Proteins. EMS16 is the first CLP found specifically to inhibit collagen binding to GPIa/IIa. With the objective of understanding the structural details that may account for the biological activity of EMS16, we compared the structure of EMS16 with those of other snake venom CLPs (bitiscetin, botrocetin, and IX/X-bp) that have the ability to recognize distinct target molecules. The structural superposition of each of the CLP subunits is shown in Figure 4. The main-chain conformation of each subunit is well-conserved among the CLPs, except for the central loop and the regions between H2 and S2, and between S3 and S4. The root-mean-square (RMS) deviations are summarized in Table 2. The relative orientation of the two subunits of EMS16 is similar to that of bitiscetin, but different from that of botrocetin and IX/X-bp, by having a more closed conformation (about ≤20° rotation) as reported by Hirotsu et al. (*8*). Interestingly, these differences are concentrated mainly in the concave surface, which may play an important role in having determinants expressing distinct biological activities. In actuality, the crystal structures of the complex with the target molecule for IX/X-bp (*4*), X-bp (*10*), botrocetin (*9*), and bitiscetin (*11*) show that the binding site to the target molecule is located on the concave surface.

The calculated molecular electrostatic potentials of IX/X-bp, botrocetin, bitiscetin, and EMS16, using the program GRASP (*28*), are shown in Figure 5. The surface potentials on the concave surface may be viewed in a contrasting outline shown for the different CLPs. The IX/X-bp has a positively charged patch on the concave surface where the negatively charged residues of the Gla domain of factor IX/X-bp are involved in binding (*4*). On the other hand,

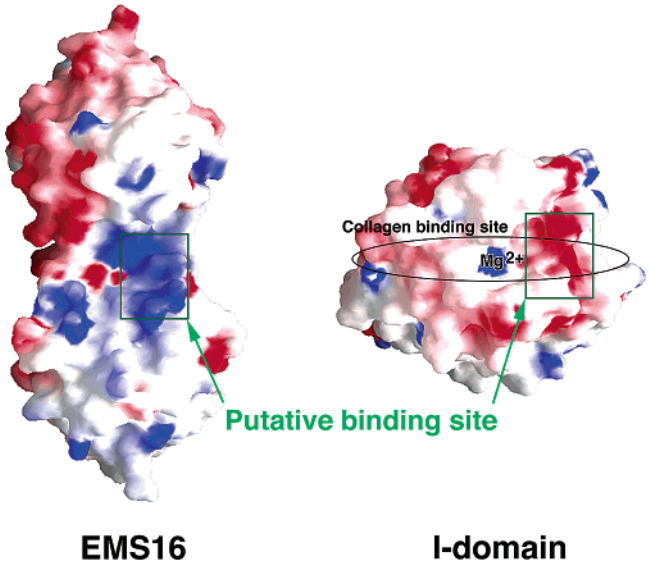

FIGURE 6: Surface electrostatic potentials of EMS16 and the I-domain of GPIa/IIa (*18*). The possible binding site is shown. The charge of the magnesium ion of the I-domain was assigned a value of 2.0 to calculate the surface electrostatic potentials using the program GRASP (*28*).

botrocetin and bitiscetin have negatively charged patches where the VWF A1 domain binds as a target molecule. For EMS16, the positively charged and hydrophobic patches exist on the concave surface, even though the total surface potential is negatively charged. The positively charged patch is made up of residues Lys97 and Lys102 in subunit A and Arg92, Lys101, Arg110, and Lys114 in subunit B. It should be noted that the positively charged patch exists only on the concave surface in the case of EMS16, which is expected to be a binding site of the I-domain of GPIa/IIa, as described next.

Potential Binding Site to Collagen Receptor. On the basis of the crystal structure of the complex between the I-domain of GPIa/IIa and a triple helical collagen peptide (*19*), the collagen-binding site of the I-domain is presumed to be negatively charged (Figure 6). Therefore, the positively charged patch on the concave surface of EMS16 may be considered a candidate binding site to the I-domain. Further,

it has been reported that the cyclic peptide, CTRKKHDNAQC, whose design is based on the metalloproteinase domain of jararhagin, binds to the I-domain of GPIa/IIa and prevents the binding of collagen types I and IV (*29*), wherein only three positively charged amino acid residues, RKK, are essential in inhibiting the collagen binding. This result supports the present assumption that the positively charged patch on the concave surface of EMS16 is the site for binding to the I-domain.

Nonetheless, it is inexplicable why EMS16 does not inhibit the binding of collagen to the I-domain of integrin α1 (*12*), whose conformation and electrostatic potential are essentially identical to those of GPIa/IIa (*18*). To resolve this problem, crystal structure analysis of the complex is now in progress. From a biomedical application point of view, such studies should help in developing more effective antagonists against GPIa/IIa.

REFERENCES

1. Atoda, H., Hyuga, M., and Morita, T. (1991) *J. Biol. Chem. 266*, 14903−11.
2. Atoda, H., and Morita, T. (1993) *J. Biochem. 113*, 159−63.
3. Mizuno, H., Fujimoto, Z., Koizumi, M., Kano, H., Atoda, H., and Morita, T. (1997) *Nat. Struct. Biol. 4*, 438−41.
4. Mizuno, H., Fujimoto, Z., Atoda, H., and Morita, T. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 7230−4.
5. Mizuno, H., Fujimoto, Z., Koizumi, M., Kano, H., Atoda, H., and Morita, T. (1999) *J. Mol. Biol. 289*, 103−12.
6. Fukuda, K., Mizuno, H., Atoda, H., and Morita, T. (2000) *Biochemistry 39*, 1915−23.
7. Sen, U., Vasudevan, S., Subbarao, G., McClintock, R. A., Celikel, R., Ruggeri, Z. M., and Varughese, K. I. (2001) *Biochemistry 40*, 345−52.
8. Hirotsu, S., Mizuno, H., Fukuda, K., Qi, M. C., Matsui, T., Hamako, J., Morita, T., and Titani, K. (2001) *Biochemistry 40*, 13592−7.
9. Fukuda, K., Doggett, T. A., Bankston, L. A., Cruz, M. A., Diacovo, T. G., and Liddington, R. C. (2002) *Structure 10*, 943−50.
10. Shikamoto, Y., Morita, T., Fujimoto, Z., and Mizuno, H. (2003) *J. Biol. Chem. 278*, 24090−4.
11. Maita, N., Nishio, K., Nishimoto, E., Matsui, T., Shikamoto, Y., Morita, T., Sadler, J. E., and Mizuno, H. (2003) *J. Biol. Chem. 278*, 37777−81.
12. Marcinkiewicz, C., Lobb, R. R., Marcinkiewicz, M. M., Daniel, J. L., Smith, J. B., Dangelmaier, C., Weinreb, P. H., Beacham, D. A., and Niewiarowski, S. (2000) *Biochemistry 39*, 9859−67.
13. Hynes, R. O. (1992) *Cell 69*, 11−25.
14. Berlanga, O., Tulasne, D., Bori, T., Snell, D. C., Miura, Y., Jung, S., Moroi, M., Frampton, J., and Watson, S. P. (2002) *Eur. J. Biochem. 269*, 2951−60.
15. Suzuki-Inoue, K., Tulasne, D., Shen, Y., Bori-Sanz, T., Inoue, O., Jung, S. M., Moroi, M., Andrews, R. K., Berndt, M. C., and Watson, S. P. (2002) *J. Biol. Chem. 277*, 21561−6.
16. Suzuki-Inoue, K., Ozaki, Y., Kainoh, M., Shin, Y., Wu, Y., Yatomi, Y., Ohmori, T., Tanaka, T., Satoh, K., and Morita, T. (2001) *J. Biol. Chem. 276*, 1643−52.
17. Bergmeier, W., Bouvard, D., Eble, J. A., Mokhtari-Nejad, R., Schulte, V., Zirngibl, H., Brakebusch, C., Fassler, R., and Nieswandt, B. (2001) *J. Biol. Chem. 276*, 25121−6.
18. Emsley, J., King, S. L., Bergelson, J. M., and Liddington, R. C. (1997) *J. Biol. Chem. 272*, 28512−7.
19. Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J., and Liddington, R. C. (2000) *Cell 101*, 47−56.
20. Okuda, D., Horii, K., Mizuno, H., and Morita, T. (2003) *J. Biochem. 134*, 19−23.
21. Navaza, J. (1994) *Acta Crystallogr. D50*, 157−63.
22. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. D50*, 760−7.
23. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−21.
24. McRee, D. E. (1999) *J. Struct. Biol. 125*, 156−65.
25. Kleywegt, G. J., and Jones, T. A. (1998) *Acta Crystallogr. D54*, 1119−31.
26. Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993) *J. Mol. Biol. 231*, 1049−67.
27. Bennett, M. J., Schlunegger, M. P., and Eisenberg, D. (1995) *Protein Sci. 4*, 2455−68.
28. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins 11*, 281−96.
29. Ivaska, J., Kapyla, J., Pentikainen, O., Hoffren, A. M., Hermonen, J., Huttunen, P., Johnson, M. S., and Heino, J. (1999) *J. Biol. Chem. 274*, 3513−21.
30. Kraulis, P. (1991) *J. Appl. Crystallogr. 26*, 946−50.
31. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol. 277*, 505−24.

BI034890H