Crystal Structure of the Disintegrin Heterodimer from Saw-Scaled Viper (*Echis carinatus*) at 1.9 Å Resolution

Sameeta Bilgrami,[‡] Savita Yadav,[‡] Punit Kaur,[‡] Sujata Sharma,[‡] Markus Perbandt,[§] Christian Betzel,[§] and Tej P. Singh*[,‡]

Department of Biophysics, All India Institute of Medical Sciences, New Delhi 110 029, India, and Institute of Medical Biochemistry and Molecular Biology, Hamburg 22603, Germany

Received May 9, 2005; Revised Manuscript Received June 27, 2005

ABSTRACT: Disintegrins constitute a family of potent polypeptide inhibitors of integrins. Integrins are transmembrane heterodimeric molecules involved in cell−cell and cell−extracellular matrix interactions. They are involved in many diseases such as cancer and thrombosis. Thus, disintegrins have a great potential as anticancer and antithrombotic agents. A novel heterodimeric disintegrin was isolated from the venom of saw-scaled viper (*Echis carinatus*) and was crystallized. The crystals diffracted to 1.9 Å resolution and belonged to space group $P4_32_12$. The data indicated the presence of a pseudosymmetry. The structure was solved by applying origin shifts to the disintegrin homodimer schistatin solved in space group $I4_122$ with similar cell dimensions. The structure refined to the final R_{cryst}/R_{free} factors of 0.213/0.253. The notable differences are observed between the loops, (Gln39−Asp48) containing the important Arg42-Gly43-Asp44, of the present heterodimer and schistatin. These differences are presumably due to the presence of two glycines at positions 43 and 46 that allow the molecule to adopt variable conformations. A comparative analysis of the surface-charge distributions of various disintegrins showed that the charge distribution on monomeric disintegrins occurred uniformly over the whole surface of the molecule, while in the dimeric disintegrins, the charge is distributed only on one face. Such a feature may be important in the binding of two integrins to a single dimeric disintegrin. The phylogenetic analysis developed on the basis of amino acid sequence and three-dimensional structures indicates that the protein diversification and evolution presumably took place from the medium disintegrins and both the dimeric and short disintegrins evolved from them.

Disintegrins are a family of small cysteine-rich proteins, found in *Viperidae* and *Crotalidae* snake venoms that bind to integrins of the β1 and β3 classes (*1*, *2*). They can be classified into small, medium, long, and dimeric disintegrins depending on the size and the number of cysteines. Short disintegrins are 49−50 amino acids long and have 8 cysteines (*3*). The medium disintegrins contain approximately 70 amino acids with 12 cysteines. Most of the polypeptides that have been characterized so far belong to the class of medium disintegrins (*4*−*10*). Long disintegrins have ∼80 amino acid residues including 14 cysteines (*11, 12*). The dimeric disintegrins have been identified recently, and their number is increasing gradually. Each of the monomers of the dimeric disintegrins has approximately 67 amino acid residues including 10 cysteines. Each monomer contains two unpaired cysteine residues at the N terminus that form two disulfide bridges with the other monomer to generate the dimer (*13*−*21*). An exception is bilitoxin-I (*22*), a homodimeric disintegrin, each monomer of which has 15 cysteine residues. The binding of disintegrins to integrins is typically characterized by the presence of the Arg-Gly-Asp motif, although mutations of Arg and Gly have been reported (*15*−*19*). The Asp residue is always conserved (*23*). The Arg-Gly-Asp is located at a long surface-exposed loop that extends out of the molecule and is highly flexible. Integrins are present on the cell surfaces and are known to be involved in cell−cell interactions and cell−extracellular matrix interactions (*24*, *25*). The integrins have been reported to participate in a number of important biological processes such as angiogenesis, tumor invasion, inflammatory responses, platelet aggregation, and tissue repair (*26*−*31*). It has been reported that the disintegrins interfere with the functions of integrins as antagonists (*1*) and hence act as anticancer and antithrombotic agents. In view of such important functions of disintegrins, the elucidation of their detailed three-dimensional structures is necessary so that the mechanism of their actions could be understood at the structural level. It will help in the design of synthetic polypeptides as potent antagonists of integrins. Unfortunately, the structural information about disintegrins is still scanty because only one crystal (*32*) and a few NMR structures (*33*−*37*) of monomeric medium disintegrins and only one structure of a dimeric disintegrin, schistatin, are known (*23*). The most important feature of dimeric disintegrins is having two Arg-Gly-Asp-binding sites that diverge away from each other in the dimer. It has been shown that, in comparison to monomers, the dimeric disintegrins show increased affinity for integrins (*38*). We have isolated a novel heterodimer of

* To whom correspondence should be addressed. Telephone: 91-11-2658-8931. Fax: 91-11-2658-8663. E-mail: tps@aiims.aiims.ac.in.
[‡] All India Institute of Medical Sciences.
[§] Institute of Medical Biochemistry and Molecular Biology.

10.1021/bi050849y CCC: $30.25
Published on Web 07/28/2005

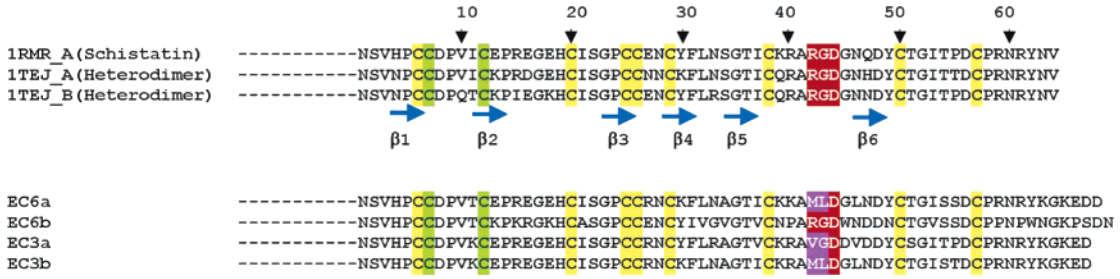

FIGURE 1: Amino acid sequence multiple alignment of various disintegrins. The Arg-Gly-Asp sequence is shaded in red. Cys residues involved in intramolecular disulfide bonds are shaded in yellow, while those involved in intermolecular disulfide linkages are indicated in green. The blue arrows represent the regions of the β strands.

disintegrin from the venom of *Echis carinatus*. The crystal structure has been solved and refined to 1.9 Å resolution. The heterodimer sheds a new light on the organization and the mode of binding of the dimeric disintegrins to integrins.

EXPERIMENTAL PROCEDURES

Isolation and Purification of Disintegrin. The crude venom of *E. carinatus* was obtained from the Irula Cooperative snake farm, Mahabalipuram, India. A total of 250 mg of lyophilized venom was dissolved in 20 mM ammonium acetate buffer at pH 5.0 to make the final concentration of venom at 10 mg/mL. This solution was centrifuged at 25 000 rotations per minute for a total period of 15 min at 293 K. The insoluble material was removed, and the supernatant was diluted 3 times with the same buffer. This was loaded onto a Blue-Sepharose CL-6b column (15 × 2.5 cm). The column was pre-equilibrated with 20 mM ammonium acetate buffer at pH 5.0 and was washed repeatedly with the same buffer to remove other unbound proteins. The bound proteins were eluted using a continuous NaCl gradient (0.0−0.6 M). The fractions corresponding to 14 kDa as indicated by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE)[1] were collected. These were pooled and dialyzed against 20 mM ammonium acetate at pH 5.0 to remove NaCl. The dialyzed samples were loaded on an SP-Sephadex column C-25 (8.0 × 2.5 cm). The column was washed thoroughly for the unbound proteins. The bound proteins were eluted using a linear gradient of NaCl (0.0−0.5 M). The peak corresponding to 14 kDa was used for further analysis.

Molecular Weight Determination by Matrix-Assisted Laser Desorption Ionization−Time-of-Flight (MALDI−TOF). The purified and lyophilized samples of disintegrin were dissolved in distilled water to attain an approximate concentration of 10 mg/mL. The mass spectrometric analysis was carried out with MALDI−TOF (KRATOS, Shimadzu, Japan). Prior to the acquisition of spectra, 100 μL of the protein solution was mixed with 100 μL of 0.2% aqueous trifluoroacetic acid. A total of 1 μL of the acidified solution was spotted onto a stainless steel sample slide by 1 μL of 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinnic acid) matrix solution (10 mg/ml) in 50:50 ethanol containing 0.1% trifluoroacetic acid.

Sequence Determination. The complete amino acid sequence of the present heterodimeric disintegrin was determined using the principle of Edman degradation (*39*) on the PPSQ-21A (Shimadzu) protein sequencer. The purified samples of disintegrin were lyophilized. A 10 mg/mL solution of disintegrin was prepared in deionized water. At least 10 μL of the protein solution was loaded on the machine, and sequencing was run up to 70 cycles. The sequences of the two chains were derived from the output of the sequencing procedure. Density modifications done with the help of DM (*40*) and RESOLVE (*41*) and omit-map calculations clearly indicated the differences in the sequence of A and B chains. The $|2F_o - F_c|$ and $|F_o - F_c|$ maps calculated by the warpNtrace protocol of ARP/wARP (42) use the hybrid model. Such maps are unbiased from initial models and noncrystallographic symmetry (NCS) restraints and thus could reveal structural differences. The maps calculated from ARP/wARP also showed the similar differences in the A and B chains. The sequences of the two polypeptide chains of the heterodimer and their comparisons with the sequence of the polypeptide chain of schistatin homodimer are given in Figure 1.

Crystallization. The purified samples of the dimeric disintegrins were used for crystallization with the sitting drop vapor-diffusion method. A 3 mg/mL protein solution was prepared in 10 mM sodium cacodylate buffer at pH 6.0. The reservoir solution containing 1.6 M ammonium sulfate was prepared as a precipitating agent. The final protein solution was made by mixing equal volumes of the protein solution and the reservoir solution. A total of 20 μL drops of the new protein solution were placed in microbridges. These were allowed to equilibrate with the reservoir solution in the sealed dishes. Diamond-shaped crystals were obtained after 10 days, which grew to the maximum dimensions of $0.3 \times 0.2 \times 0.2$ mm^3.

X-ray Data Collection. X-ray intensity data were collected using synchrotron beamline X13 at DESY−Hamburg from one single flash-frozen crystal. The diffraction data set was processed and scaled with DENZO and SCALEPACK (*43*). The details of X-ray intensity data collection and processing are given in Table 1.

Structure Solution. The X-ray intensity data of disintegrin to 1.9 Å resolution were collected in the tetragonal P lattice. However, it could be indexed, using DENZO (*43*), in the I tetragonal or P tetragonal lattices depending on the number of spots taken for indexing. This suggested that the I-specific molecule at 0.5, 0.5, 0.5, had shifted slightly from its position during the freezing procedures applied at the synchrotron, giving rise to the P tetragonal lattice and thus pseudosymmetry. Such a lattice will be "approximately" I tetragonal but will actually be P tetragonal, and all of the reflections satisfying the condition $h + k + l = 2n + 1$ will be weak.

[1] Abbreviations: SDS−PAGE, sodium dodecyl sulphate−polyacrylamide gel electrophoresis; NCS, noncrystallographic symmetry.

Table 1: Overall Statistics of Data Processing

space group	$P4_32_12$
cell dimensions (Å)	
a	90.7
b	90.7
c	55.5
V_m (Å^3/Da)	4.1
solvent content (%)	69.8
Z (number of molecules in the unit cell)	8
resolution range (Å)	20.0−1.9
number of unique reflections	17 734
completeness (1.97−1.90 Å) (%)	99.9
completeness in the highest resolution shell (1.97−1.90 Å) (%)	99.8
R_{sym} overall (%)	6.2
R_{sym} in the highest resolution shell (1.97−1.90 Å) (%)	65.9
mean $I/\sigma(I)$ overall	10.8
mean $I/\sigma(I)$ in the highest resolution shell (1.97−1.90 Å)	2.5

To confirm the pseudosymmetry, a native Patterson map was calculated up to 3.0 Å using the data indexed in P tetragonal lattice. An off-origin peak at 0.5, 0.5, 0.5 confirmed the presence of pseudosymmetry. This peak was approximately 50% of the origin peak, indicating a high level of pseudosymmetry.

Attempts were made to determine the structure with the molecular replacement method using the structure of schistatin as a model (PDB 1RMR) (*23*). AMoRe (*44*) and MOLREP (*45*) were used for structure determination. However, no clear solution was obtained in the P tetragonal lattice, most likely because of weak P specific reflections.

It may be mentioned here that the structure of schistatin was solved in the spacegroup $I4_122$ with only one chain of the dimer in the asymmetric unit. The X-ray crystallographic data of the present structure, with the cell parameters similar to that of schistatin but having a P tetragonal lattice, suggested that the I-specific molecule had shifted from its 0.5, 0.5, 0.5 position, and thus, the whole dimer was present in the asymmetric unit. The overall organization of the present crystal however is expected to be the same as schistatin. Hence, the most probable space group of the present structure could be one of the subgroups of $I4_122$, i.e., $P4_32_12$, $P4_12_12$, $P4_322$, or $P4_122$. Origin shifts were applied to the schistatin dimer to get each of the subgroups of $I4_122$. The structure refined only in $P4_32_12$. The solution was improved further with FITFUN of AMoRe (*44*). An accurate solution was finally derived with a good correlation coefficient of 79.5% and R_{factor} of 38.2%. LSQKAB from CCP4 package (*46*) was used to translate and rotate the model. The structure was subjected to the initial round of rigid-body refinement with CNS (*47*) by defining the two chains of the dimer as two rigid bodies. This improved the positions of the two chains of the dimer further and helped in orienting them accurately.

This was followed by additional cycles of restrained refinement with REFMAC5 (*48*). The initial maps indicated differences in the electron density for some of the residues of the two chains of dimer. The density modifications were carried out with the help of DM (*40*) and RESOLVE (*41*). The omit-map calculations clearly indicated the differences in the sequences of A and B chains (Figure 2). The $|2F_o - F_c|$ and $|F_o - F_c|$ Fourier maps calculated by the warpNtrace protocol of ARP/wARP (42) could reveal precise structural differences. The observed differences in the electron density

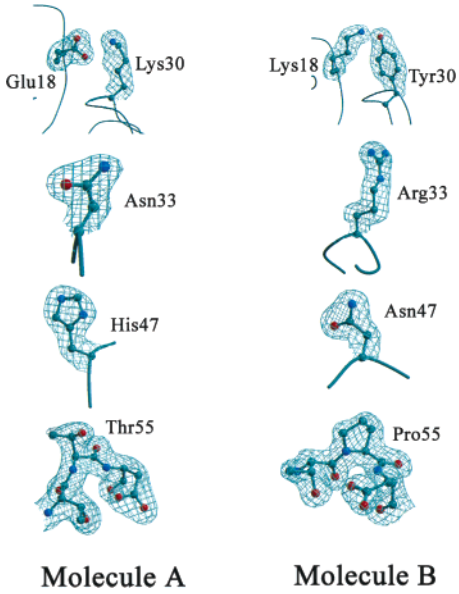

FIGURE 2: Electron densities for some of the differences observed in the sequences of molecules A and B of the heterodimer. The figure was prepared using Bobscript (*56*).

Table 2: Summary of Crystallographic Refinement

PDB code	1TEJ
refinement	
resolutions limits (Å)	20.0−1.9
number of reflections	17 734
working set (95%)	16 779
test set (5%)	955
R_{cryst} (%)	21.3
R_{free} (%)	25.3
model	
protein atoms	940
water molecules	161
rms deviations from ideal values	
bond lengths (Å)	0.009
bond angles (deg)	1.4
overall *G* factor	−0.06
average *B* factors (Å^2)	
main-chain atoms	38.8
side-chain atoms	48.8
for all atoms	43.8
Ramachandran plot (non-Gly, non-Pro)	
residues in the most favored regions (%)	84.7
residues in the additionally allowed regions (%)	15.3

for various residues of the two chains matched well with the sequences determined chemically. The model was further improved manually with the help of the graphics program O (*49*). The refinement finally converged to R_{cryst}/R_{free} of 21.3/25.3%. The refinement parameters are listed in Table 2.

RESULTS AND DISCUSSION

Sequence Analysis. The precise molecular weight of the dimeric disintegrin determined using MALDI−TOF was 12 791.9 Da. The molecular weight calculated from the sequence that was determined approximately using automatic protein sequencer was 12 792.4 Da. These values indicated that the sequences of molecules A and B of the disintegrin were determined correctly.

The sequence identity between the A and B chains is 84%. The identities of A and B chains with the monomer schistatin are 85 and 87%, respectively. The sequence identities of

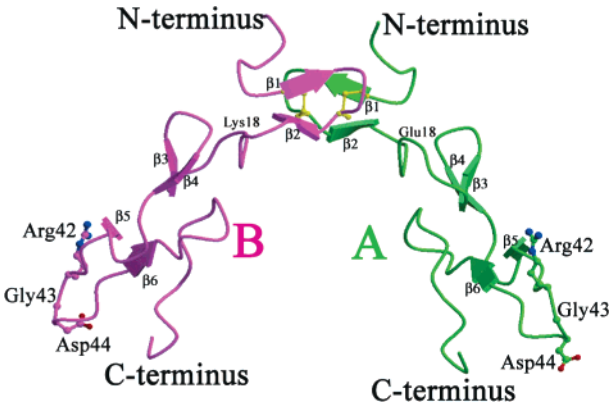

FIGURE 3: Overall fold of the dimer. The side chains of the Arg-Gly-Asp motif are also indicated. The two molecules of the dimer are designated as A and B. The β strands β1, β2, β3, β4, β5, and β6 in each chain are also indicated. The figure was prepared using Moslcript and Raster3D (*57*, *58*).

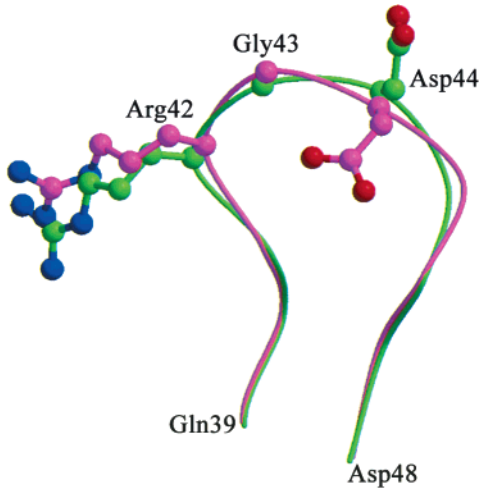

FIGURE 4: Showing the superposition of the Arg-Gly-Asp loops of the A and B chains of the heterodimer.

molecules A and B of the present heterodimer with the monomers of other heterodimers from *E. carinatus* EC3A, EC3B, EC6A, and EC6B vary from 69 to 78%. Although the Arg-Gly-Asp sequence is considered as an important requirement for binding to integrin, the variations from this have been observed in other heterodimers (Figure 1). However, the Asp residue of the Arg-Gly-Asp sequence is invariably conserved. The Arg-Gly-Asp loops, Cys38−Cys50 of various disintegrins, show higher sequence variations presumably to allow it to acquire a higher degree of freedom to adopt suitable conformation for binding to specific integrins. The presence of four intramolecular and two intermolecular disulfide bridges and several conserved glycine residues ensures the correct relative orientations of the disintegrin polypeptides.

Quality of the Model. The quality of electron density for the main chains and the side chains of both chains A and B was good and allowed an accurate interpretation of the structure. It clearly showed the differences between the nonidentical residues of the two chains of the heterodimer (Figure 2). The final crystallographic model has 940 protein atoms from 128 amino acid residues from the heterodimer and 161 water molecules. The root-mean-square (rms) deviations in the bond lengths and angles were 0.009 Å and 1.4°, respectively. The geometry of the model as analyzed with PROCHECK (*50*) showed 84.7% of the residues in the core region of the Ramachandran plot (*51*), while the remaining 15.3% were present in the additionally allowed regions.

Overall Structure. The overall fold of the monomers as well as the general organization of the present heterodimer (Figure 3) is essentially similar to that of the schistatin homodimer (*23*). Each of the monomers in the present structure can be divided into two distinct domains, an N-terminal domain (1−18) and a C-terminal domain (19−63). The N-terminal domains of the two monomers are responsible for the dimer formation by swapping into each other to generate intermolecular interactions including two covalently formed disulfide bridges. The C-terminal domain contains the functionally important Arg42-Gly43-Asp44 loop (to be referred hereafter as the Arg-Gly-Asp loop) and the C-terminal segment. The C-terminal segment lends crucial support to the Arg-Gly-Asp loop through several stabilizing interactions resulting in a well-formed domain. The C-terminal domains in the heterodimeric arrangement are widely separated from each other.

The structure is dimerized through two N-terminal disulfide bonds formed between Cys7 (A chain) and Cys12 (B chain) and Cys12 (A chain) and Cys7 (B chain). This confirms the pattern of intermolecular disulfide bonds as observed in the homodimer of schistatin (*23*) and negates the scheme of intermolecular disulfide bond pattern deduced by chemical methods on the heterodimeric disintegrin EMF10 (*52*). In addition to the two disulfide bridges, the dimer is held by two hydrogen bonds involving the main-chain nitrogen atoms of the Asn4(A) and Asn4(B) residues and the carbonyl atoms of the Pro9(A) and Pro9(B) residues. The two protomers of the dimer are linked with each other only at their N termini and as seen from Figure 3, diverge away at the C termini, allowing them the flexibility of reaching the distant binding sites in integrins (*38*). The distance between the Cα atoms of Gly43(A) and Gly43(B) of the Arg-Gly-Asp loops of molecules A and B is nearly 61 Å.

The rms shift between the Cα atoms of the A and B chains is 0.39 Å. Both the A and B chains of the dimer consists of three pairs of six antiparallel β strands comprising residues 6−8 (β1), 13−16 (β2), 25−27 (β3), 30−32 (β4), 36−39 (β5), and 48−50 (β6). The β strands are connected by flexible loops of different lengths consisting of 4−10 residues. The most prominent segment is formed with Gln39−Asp48 and is called as the Arg-Gly-Asp loop. The backbone conformations of this loop are essentially similar in both chains, except for the orientations of Asp44(A) and Asp44(B), which are disposed in the opposite directions (Figure 4).

The C termini of both of the monomers lie adjacent to the Arg-Gly-Asp loop, forming several interactions with it. The ability to bind to integrins depends on the structure and potentials of the combined surfaces of the C termini and the Arg-Gly-Asp loops (*32*, *53*). Thus, the interactions between the C termini and the Arg-Gly-Asp loops are important for keeping the two segments together. In the present structure, the close alignment of the C termini with the Arg-Gly-Asp loop brings together many residues such as Asn46 and Asp48 (from the Arg-Gly-Asp loops) and Asn63, Asn60, Arg59,

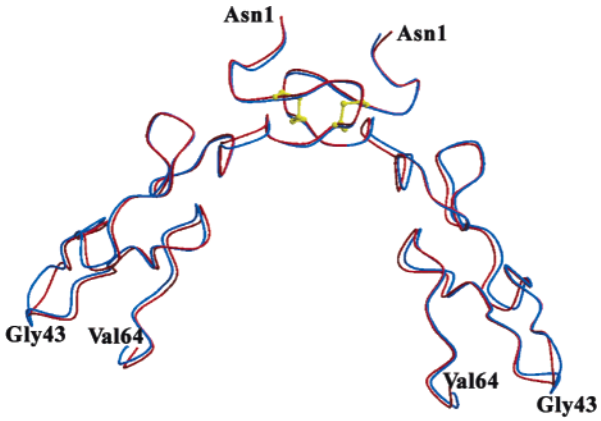

FIGURE 5: Superposition of the heterodimer with Schistatin homodimer. The heterodimer is shown in red, whereas the schistatin homodimer is in blue.

and Tyr62 (from the C-terminal segments), which form the continuous patches of polar residues that promote the interactions with the integrins.

Comparative Analysis with Other Disintegrins: Comparison with the Schistatin Structure. The overall fold of the monomers as well as the organization of the heterodimer is very similar to that of schistatin (Figure 5). The rms shifts between the Cα atoms of the two dimers is 0.63 Å. The most notable differences are observed between their Arg-Gly-Asp loops. The Arg-Gly-Asp loop in schistatin acquires a relatively broader arrangement in which Gly43 adopts an extended conformation ($\phi = 155$, and $\varphi = 100$). In contrast, the corresponding loops in the present heterodimer are significantly narrower with Gly43 adopting a cis conformation ($\phi = -108$ and $\varphi = -2$ for chain A, and $\phi = -85$ and $\varphi = -10$ for chain B) (parts a and b of Figure 6). Furthermore, the kink formed by the residues 44−46 in the Arg-Gly-Asp loop is much less pronounced in the two monomers of the heterodimer as compared to those of schistatin, though it is not completely absent.

Thus, even though the Arg-Gly-Asp-Gly sequence is conserved in the two monomers of the heterodimer as well as in the schistatin homodimer, the shapes and widths of the loops are different. These observations further confirm our earlier analysis (*23*) that glycines at these positions in the Arg-Gly-Asp loop will make the disintegrin amenable to bind to both types of integrins αIIβ3 (which requires a broader and stouter Arg-Gly-Asp loop in its ligands) and α5β3, αvβ3 integrins (that prefer a thinner Arg-Gly-Asp loop) (*54*).

Comparison with the Monomeric Disintegrins. The overall folds of the two polypeptide chains of the present structure are similar to the structures of the other known medium disintegrins but showed significant differences in several parts. The rms shifts for the Cα atoms were found to be 1.4 Å for flavosatin (PDB 1FVL), 1.3 Å for trimestatin (PDB 1J2L), and 1.5 Å for rhodostomin (PDB 1JYP), when the A chain was superimposed on the above disintegrins, and were 1.5, 1.3, and 1.4 Å for flavostatin (PDB 1FVL), trimestatin (PDB 1J2L), and rhodostomin (PDB 1JYP), respectively, when the B chain was superimposed. The N-terminal region, the loop containing the Arg-Gly-Asp motif, and the C terminus show rms displacements greater than 2.0 Å when the corresponding Cα positions of the two polypeptides were

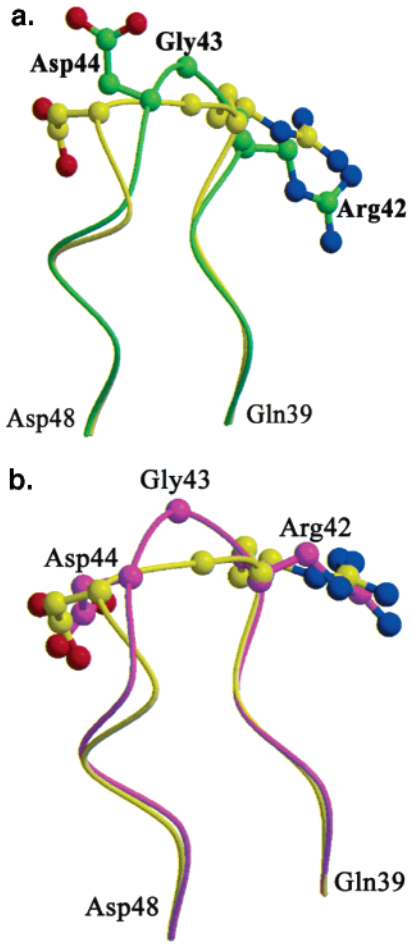

FIGURE 6: (a) Showing the superposition of the Arg-Gly-Asp loop of the A chain of the heterodimer and that of the monomer of the schisatin homodimer. The loop of the A chain is shown in green, and that of the schistatin homodimer is shown in yellow. (b) Showing the superposition of the Arg-Gly-Asp loop of the B chain of the heterodimer and that of the monomer of the schisatin homodimer. The loop of the B chain is shown in pink, and that of the schistatin homodimer is shown in yellow.

compared. The only region that shows a significant overlap corresponds to the central region consisting of residues 21−37 and 49−57 (numbering scheme of the present structure). The rms shifts for echistatin are 1.6 and 1.5 Å when the Cα atoms of the A and B chains of the present structure are compared. Echistatin is a short disintegrin and consists of only the region corresponding to the C-terminal domain of the present structure. As observed for the medium disintegrins, large deviations were observed in the Arg-Gly-Asp loop and the C terminus region. The structures of two more short disintegrins, obtustatin and salmosin, are known, but they seem to indicate large variations (*36*, *37*).

Comparative Analysis of the Electrostatic Surface Potentials of Disintegrins. Disintegrins bind integrins primarily through the polar and charged residues found in the Arg-Gly-Asp loop and the C terminus (*32*, *53*). A comparative analysis of the surface-charge distribution showed considerable differences in the monomeric and dimeric disintegrins (Figure 7). It was found that all monomeric disintegrins (medium as well as short) had a surface-charge distribution all over the C-terminal domain and not restricted to one particular side. The surface could be roughly divided into a negatively charged region (on the side where the Asp of the

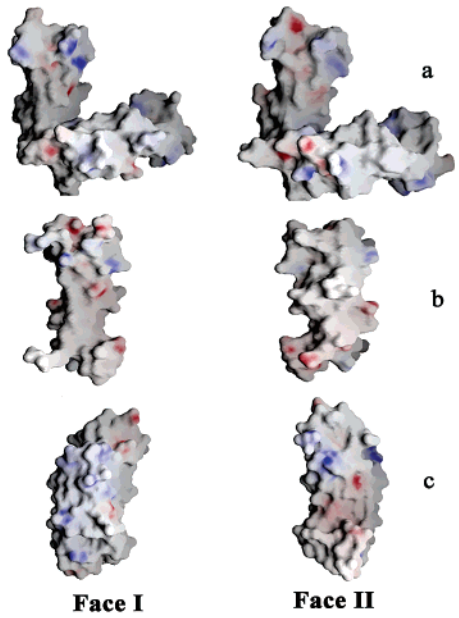

FIGURE 7: Showing the electrostatic potential distribution of the various integrins. Positive and negative charges are colored blue and red, respectively. (a) 1TEJ, Face-I has the A molecule vertical (showing its charged surface), whereas the B molecule is horizontal (showing its neutral face). The same can be seen for Face-II with the B molecule vertical (showing its charged surface) and the A molecule now horizontal (showing its neutral face). This clearly shows that one face of each of the molecules of the dimer is charged and the other is essentially neutral and that the two charged faces lie opposite to each other that allow two integrin molecules to bind easily. (b) 1RMR is a homodimer; therefore, both of the molecules are the same. Face-I shows the charged face of the molecule, and Face-II shows the essentially neutral face. The negatively charged area seen here lies in the N terminus, which is not used for integrin binding. (c) 1J2L, both the faces of the molecule are charged. The figure was drawn using GRASP (*59*).

Arg-Gly-Asp loop lies) and a positively charged region (on the side of the Arg). Exceptions were the short disintegrins, obtustatin and salmosin, where obtustatin is nearly neutral and salmosin is mostly negatively charged. Even in these disintegrins, the charges are distributed all over the surface and are not restricted to a particular region. In contrast, in the dimeric disintegrins, schistatin, and the present structure, the charges are distributed on one face of the C-terminal domain and the other face is essentially neutral. The face bearing the charge lies opposite to the charged face of the other monomer. Because the disintegrins bind integrins primarily through their charged surfaces and polar residues, such an arrangement will facilitate the binding of two integrins to the dimeric disintegrins. Thus, not only are the two integrin-binding sites in the disintegrins widely separated from each other (*23*), but they also lie on opposite sides to each other, which will further facilitate the binding of two integrins to a single dimeric disintegrin. Upon examining the reasons for the differences in the surface-charge distribution of the monomeric and dimeric disintegrins, it was found that the differences were due to the placement of the C termini with respect to the Arg-Gly-Asp loops. In the dimeric disintegrins, the C terminus is placed parallel to the Arg-Gly-Asp loop and has many interactions with it (*23*), whereas in the medium disintegrins, the C terminus lies across the Arg-Gly-Asp loop. In the short disintegrin echistatin, a portion of the C terminus lies parallel to the Arg-Gly-Asp loop in a manner similar to schistatin and the present heterodimer. However, the C terminus is longer in echistatin as compared to the medium and dimeric disintegrins. This extended portion of the C terminus is placed in such a manner that the surface charge is distributed uniformly and is not confined to a particular region. The other two short disintegrins are structurally different, and thus such comparisons cannot be made.

Both of the dimeric disintegrins, schistatin and the present structure, have the same general pattern of charge distribution as described above, although there are subtle differences. In schistatin, the charge distribution is more extensive than in the present dimer. This may be partly due to the presence of Lys39 in schistatin, which is replaced by a neutral Gln39 in the A and B chains of the present dimer. When the two chains of the present dimer are compared, the A chain has more extensive charge distribution than the B chain. This is because the A chain has Thr55, which accentuates the negatively charged environment at this point. On the other hand, the B chain has a proline at this position, thus not contributing to the negative charge. It is observed that not only the differences in the type of residues but also the differences in their placement lead to differences in the surface-charge distributions. Thus, there is a strongly negatively charged area when the Asp44 is pointing inside (as in schistatin and the B chain of the present dimer) and not when it is pointing outside. This is because the negative charge is accentuated by a number of carbonyl oxygens in the same area when the Asp is pointing inside. Thus, small changes in the conformation of the loops and segments, changes in the conformation of the side chains of residues, as well as mutations all have the effects on the surface-charge distribution and hence on the binding to integrins.

Phylogenetic Analysis. The medium and dimeric disintegrins contain distinct N-terminal and C-terminal domains, with the C-terminal domain being responsible for the binding to integrins and the N-terminal domain being used as an anchor in the dimeric arrangement. The short disintegrins, which are shorter than the medium disintegrins by about 20−25 amino acid residues from the N terminus have only the portion homologous to the C-terminal portions of the medium and dimeric disintegrins. A phylogenetic analysis carried out based on the available cDNA sequences (*55*) indicated that all of the disintegrins have evolved from a common Adam-like ancestor by successive deletions and mutations. It appears that the dimeric disintegrins arose from medium disintegrins by the deletion of the two N-terminal cysteines. It is however not clear whether the short disintegrins arose from medium or dimeric disintegrins because of the unavailability of cDNA sequences of short disintegrins. The present structural analysis allows us to infer the parentage of the short disintegrins. It is likely that both the dimeric and short disintegrins arose from the medium disintegrins. It is highly unlikely that the short disintegrins arose from the dimeric disintegrins because in these proteins the N-terminal domain is being used for the dimerization in contrast to the medium disintegrins, where this domain does not serve a known purpose. Thus, it appears that the diversification and evolution took place from the medium disintegrins by playing with the N-terminal domain, which was not useful in integrin binding in these disintegrins. It was shortened and used for dimerization in the dimeric disintegrins, whereas it was

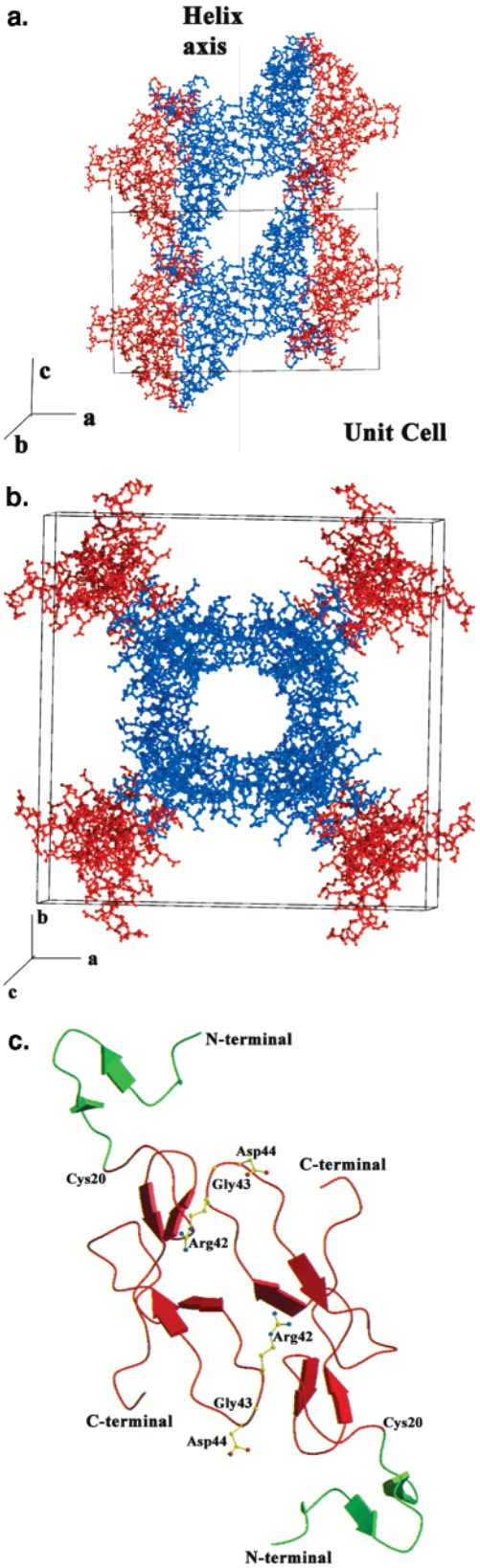

FIGURE 8: (a) Helical arrangement created by the packing of the monomers is observed parallel to the *z* axis. The helices in red are parallel to the ones in blue (not visible completely). (b) Top view of the helix formed by the monomers. The blue is the central helix surrounded by four parallel helices (not visible completely). The helix is formed by the monomers of the dimer parallel to the *z* axis. The red regions show part of the adjacent helices. The figure was made by Pymol (*60*). (c) Packing of monomers A and B in the unit cell. The green portions are outside the helical arrangement and are involved in dimerization.

completely deleted to give rise to short disintegrins. Evidence of the common parentage of dimeric and short disintegrins can also be seen from the phylogenetic tree made by Calvete et al. The short and dimeric disintegrins cluster together and seem to have a common ancestor.

Crystal Packing. The heterodimeric units are packed in the crystal in such a way that the monomers of the dimer are placed adjacent to each other to form a helix parallel to the *z* axis (parts a and b of Figure 8). The other partner monomer of the dimer lies in another helix that is parallel to the helix in which the first monomer lies. There are 8 monomers/turn of the helix. The pitch of the helix is equal to the *c* axis, i.e., 55.5 Å. The diameter of the helix is ~25 Å. Only residues 19−64 of the monomers form a part of the helix (Figure 8c). Residue numbers 1−18 lie in the space between the two parallel helices. The crystal packing also shows clear demarcation between the N-terminal (residues 1−18) and C-terminal (residues 19−64) regions, each of which are independent of each other and can have different orientations with respect to each other.

CONCLUSIONS

This structure furthers our understanding of the binding of disintegrins to integrins and also confirms some of the hypotheses proposed from the schistatin structures. The Arg-Gly-Asp loops of the monomers of the present heterodimer as well as that of the schistatin homodimer have different shapes and widths even though the Arg-Gly-Asp-Gly sequence is conserved, thus confirming our hypotheses that glycines at these positions in the Arg-Gly-Asp loop will make the disintegrin amenable to bind to both types of integrins αIIbβ3 (which requires a broader and stouter Arg-Gly-Asp loop in its ligands) and α5β3, αvb3 integrins (that prefer a thinner Arg-Gly-Asp loop) (*54*). An analysis of the surface charge distribution of all of the available structures of disintegrins revealed that in both short and medium monomeric disintegrins the surface-charge distribution was found to be all over the C-terminal domain and not restricted to one particular side. However, in the dimeric disintegrins, the present structure, and schistatin, the surface charges were distributed on one face of the C-terminal domain only and the charged faces lie opposite to each other. As the disintegrins bind integrins through their charged surfaces and polar residues, the arrangement facilitates the binding of two integrins to the dimeric disintegrins. Thus, not only are the two integrin-binding sites in the disintegrins widely separated from each other (as observed in schistatin and the present heterodimer), but they also lie on opposite sides to each other, which further helps the binding of two integrins to a single dimeric disintegrin. The two structures of dimeric disintegrins reveal the mechanism of binding of two integrin molecules to a single dimeric disintegrin molecule. An analysis of the surface-charge distribution also showed that small changes in the conformation of the loops and segments, changes in the conformation of the side chains of residues, and mutations have an effect on the surface-charge distribution and on the binding of disintegrins to integrins.

ACKNOWLEDGMENT

The authors acknowledge the financial assistance from the Department of Science and Technology, New Delhi, under

the FIST program. They thank Dr. Z. Dauter for many helpful discussions. S. B. thanks the Council of Scientific and Industrial Research, New Delhi, for the award of fellowship.

REFERENCES

1. Gould, R. J., Polokoff, M. A., Friedman, P. A., Huang, T. F., Holt, J. C., Cook, J. J., and Niewiarowski, S. (1990) Disintegrins: A family of integrin inhibitory proteins from viper venoms, *Proc. Soc. Exp. Biol. Med. 195*, 168−171.
2. McLane, M. A., Marcinkiewicz, C., Vijay-Kumar, S., Wierzbicka-Patynowski, I., and Niewiarowski, S. (1998) Viper venom disintegrins and related molecules, *Proc. Soc. Exp. Biol. Med. 219*, 109−119.
3. Gan, Z. R., Gould, R. J., Jacobs, J. W., Freidman, P. A., and Polokof, M. A. (1988) Echistatin: A potent platelet aggregation inhibitor from the venom of the viper *Echis carinatus*, *J. Biol. Chem. 263*, 19827−19832.
4. Au, L. C., Chou, J. S., Chang, K. J., Teh, G. W., and Lin, S. B. (1993) Nucleotide sequence of a full-length cDNA encoding a common precursor of platelet aggregation inhibitor and hemorrhagic protein from *Calloselasma rhodostoma* venom, *Biochim. Biophys. Acta 1173*, 243−245.
5. Kishimoto, M., and Takahashi, T. (2002) Molecular cloning and sequence analysis of cDNA encoding flavoridin, a disintegrin from the venom of *Trimeresurus flavoviridis*, *Toxicon 40*, 1033−1040.
6. Huanf, T. F., Holt, J. C., Lukasiewicz, H., and Niewiarowski, S. (1987) Trigramin: A low molecular weight peptide inhibiting fibrinogen interaction with platelet receptors expressed glycoprotein IIb-IIIa complex, *J. Biol. Chem. 262*, 16157−16163.
7. Dennis, M. S., Henzel, W. J., Pitti, R. M., Lipari, M. T., Napier, M. A., Deisher, T. A., Bunting, S., and Lazarus, R. A. (1990) Platelet glycoprotein IIb-IIIa protein antagonists from snake venoms: Evidence for a family of platelet-aggregation inhibitors, *Proc. Natl. Acad. Sci. U.S.A. 87*, 2471−2475.
8. Huang, T. F., Sheu, J. R., Teng, C. M., Chen, S. W., and Liu, C. S. (1991) Triflavin, an antiplatelet Arg-Gly-Asp-containing peptide, is a specific antagonist of platelet membrane glycoprotein IIb-IIIa complex, *J. Biochem. 109*, 328−334.
9. Scarborough, R. M., Rose, J. W., Hsu, M. A., Phillips, D. R., Fried, V. A., Campbell, A. M., Nannizzi, L., and Charo, I. F. (1991) Barbourin. A GPIIb-IIIa-specific integrin antagonist from the venom of *Sistrurus m. barbouri*, *J. Biol. Chem. 266*, 9359−9362.
10. Maruyama, K., Kawasaki, T., Sakai, Y., Taniuchi, Y., Shimizu, M., Kawashima, H., and Takenaka, T. (1997) Isolation and amino acid sequence of flavostatin, a novel disintegrin from the venom of *Trimeresurus flavoviridis*, *Peptides 18*, 73−78.
11. Calvete, J. J., Schrader, M., Raida, M., McLane, M. A., Romero, A., and Niewiarowski, S. (1997) The disulphide bond pattern of bitistatin, a disintegrin isolated from the venom of the viper *Bitis arietans*, *FEBS Lett. 416*, 197−202.
12. Park, D., Kang, I., Kim, H., Chung, K., Kim, D. S., and Yun, Y. (1998) Cloning and characterization of novel disintegrins from *Agkistrodon halys* venom, *Mol. Cell 8*, 578−584.
13. Calvete, J. J., Fow, J. W., Agelan, A., Niewiarowski, S., and Marcinkiewicz, C. (2002) The presence of the WGD motif in CC8 heterodimeric disintegrin increases its inhibitory effect on αII-(b)β3, αVβ3, and α5β1 integrins, *Biochemistry 41*, 2014−2021.
14. Gasmi, A., Sriari, N., Guermazi, S., Dkhil, H., Karoui, H., and Ayeb, M. E. (2001) Amino acid structure and characterization of a heterodimeric disintegrin from *Vipera lebetina* venom, *Biochem. Biophys. Acta 1547*, 51−56.
15. Marcinkiewicz, C., Calvete, J. J., Marcinkiewicz, M. M., Raida, M., Vijay-Kumar, S., Huang, Z., Lobb, R. R., and Niewiarowski, S. (1999) EC3 a novel MLD dependent disintegrin from *Echis carinatus*, is a potent antagonist for α4 integrins, *J. Biol. Chem. 274*, 12468−12473.
16. Marcinkiewicz, C., Taooka, Y., Yokosaki, Y., Calvete, J. J., Marcinkiewicz, M. M., Lobb, R. R., Niewiarowski, S., and Sheppard, D. (2000) Inhibitory effects of MLDG-containing heterodimeric disintegrins reveal distinct structural requirements for interaction of the integrin α9β1 with VCAM-1, tenascin-C, and osteopontin, *J. Biol. Chem. 275*, 31930−31937.
17. Marcinkiewicz, C., Calvete, J. J., Vijay-Kumar, S., Marcinkiewicz, M. M., Raida, M., Schick, P., Lobb, R. R., and Niewiarowski, S. (1999) Structural and functional characterization of EMF10, a heterodimeric disintegrin from *Eristocophis macmahoni* venom that selectively inhibits α5β1 integrin, *Biochemistry 38*, 13302−13309.
18. Okuda, D., and Morita, T. (2001) Purification and characterization of a new RGD/KGD-containing dimeric disintegrin, piscivostatin, from the venom of *Agkistrodon piscivorus* piscivorous: The unique effect of piscivostatin on platelet aggregation, *J. Biochem. 130*, 407−415.
19. Okuda, D., Koike, H., and Morita, T. (2002) A new gene structure of the disintegrin family: A subunit of dimeric disintegrin has a short coding region, *Biochemistry 41*, 14248−14254.
20. Trikha, M., De Clarke, Y. A., and Markland, F. S. (1994) Contortrostatin a snake venom disintegrin, inhibits β1 integrin mediated human metastatic melanoma cell adhesion and blocks experimental metastasis, *Cancer Res. 54*, 4993−4998.
21. Zhou, Q., Hu, P., Ritter, M. R., Swenson, S. D., Argounova, S., Epstein, A. L., and Markland, F. S. (2000) Molecular cloning and functional expression of contortrostatin, a homodimeric disintegrin from southern copperhead snake venom, *Arch. Biochem. Biophys. 375*, 278−288.
22. Nikai, T., Taniguchi, K., Komori, Y., Masuda, K., Fox, J. W., and Sugihara, H. (2000) Primary structure and functional characterization of bilitoxin-1 a novel dimeric p-II snake venom metalloproteinase from *Agkistrodon bilineatus* venom, *Arch. Biochem. Biophys. 378*, 6−15.
23. Bilgrami, S., Tomar, S., Yadav, S., Kaur, P., Kumar, J., Jabeen, T., Sharma, S., and Singh, T. P. (2004) Crystal structure of Schistatin a disintegrin homodimer from saw-scaled viper (*Echis carinatus*) at 2.5 Å resolution, *J. Mol. Biol. 341*, 829−837.
24. Hynes, R. (1992) Integrins: Versatility, modulation, and signaling in cell adhesion, *Cell 69*, 11−25.
25. Hynes, R. O. (1987) Integrins: A family of cell surface receptors, *Cell 48*, 549−554.
26. Sheppard, D. (2000) *In vivo* functions of integrins: Lessons from null mutations in mice, *Matrix Biol.* 9, 203−209.
27. Byzova, T. V., Rabbani, R., D'Souza, S. E., and Plow, E. F. (1998) Role of integrin α5β3 in vascular biology, *Thromb. Haemostasis 80*, 726−734.
28. Tucker, G. C. (2003) αV integrin inhibitors and cancer therapy, *Curr. Opin. Invest. Drugs 4*, 722−731.
29. Schuppan, D., and Ocker, M. (2003) Integrin-mediated control of cell growth, *Hepatology 38*, 289−291.
30. Brooks, P. C., Clark, R. A. F., and Cheresh, D. A. (1994) Requirement of vascular integrin α5β3 for angiogenesis, *Science 264*, 569−571.
31. Hapke, S., Kessler, H., Luber, B., Benge, A., Hutzler, P., Hofler, H., Schmitt, M., and Reuning U. (2003) Ovarian cancer cell proliferation and motility is induced by engagement of integrin αVβ3/Vitronectin interaction, *Biol. Chem. 384*, 1073−1083.
32. Fuji, Y., Okuda, D., Fujimoto, Z., Horii, K., Morita, T., and Mizuno, H. (2003) Crystal structure of Trimestatin, a disintegrin containing cell adhesion recognition motif RGD, *J. Mol. Biol. 332*, 1115−1122.
33. Adler, M., Lazarus, R. A., Dennis M. S., and Wagner, G. (1991) Solution structure of Kistrin, potent platelet aggregation inhibitor and GP IIb-IIIa antagonist, *Science 253*, 445−448.
34. Saudek, V., Atkinson, R. A., Lepage, P., and Pelton, J. T. (1991) Three-dimensional structure of echistatin, the smallest active RGD protein, *Biochemistry 30*, 7369−7372.
35. Senn, H., and Klaus, W. (1993) The nuclear magnetic resonance solution structure of flavoridin, an antagonist of the platelet GP IIa-IIIa receptor, *J. Mol. Biol. 232*, 907−925.
36. Paz Moreno-Murciano, M., Monleon, D., Marcinkiewicz, C., Calvete, J. J., and Celda, B. (2003) NMR solution structure of the non-RGD disintegrin obtustatin, *J. Mol. Biol. 329*, 135−145.
37. Shin, J., Hong, S. Y., Chung, K., Kang, I., Jang, Y., Kim, D. S., and Lee, W. (2003) Solution structure of a novel disintegrin, salmosin, from *Agkistrondon halys* venom, *Biochemistry 42*, 14408−14415.
38. Ritter, M. R., and Markland, F. S., Jr. (2001) Differential regulation of tyrosine phosphorylation in tumor cells by contortrostatin, a homodimeric disintegrin, and monomeric disintegrins echistatin and flavoridin, *Toxicon 39*, 283−289.
39. Edman, P. (1970) Sequence determination, *Mol. Biol. Biochem. Biophys. 8*, 211−255.
40. Cowtan, K. (1994) Joint CCP4 and ESF-EACBM, *Newsl. Protein Crystallogr. 31*, 34−38.
41. Terwilliger, T. C. (2000) Maximum-likelihood density modification, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 56*, 965−972.

42. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol. 6*, 458−463.
43. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collection in oscillation mode, *Methods Enzymol. 276*, 307−326.
44. Navaza, J. (2001) AMoRe: An automated package for molecular replacement, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 57*, 1367−1372.
45. Vagin, A., and Teplyakov, A. (1997) MOLREP: An automated program for molecular replacement, *J. Appl. Crystallogr. 30*, 1022−1025.
46. Collaborative Computational Project, Number 4. (1994) The CCP4 suite: Programs for protein crystallography, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 50*, 760−763.
47. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905−921.
48. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 53*, 240−255.
49. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr., Sect. A: Found. Crystallogr. 47*, 110−119.
50. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
51. Ramachandran, G. N., and Sasisekharan, V. (1968) Conformation of polypeptides and proteins, *Adv. Protein Chem. 23*, 383−438.
52. Calvete, J. J., Jurgens, M., Marcinkiewicz, C., Romero, A., Schrader, M., and Niewiarowski, S. (2000) Disulfide-bond pattern and molecular modelling of the dimeric disintegrin EMF-10, a potent and selective integrin α5β1 antagonist from *Eristocophis macmahoni* venom, *Biochem. J. 345*, 573−581.
53. Marcinkiewicz, C., Vijay-Kumar, S., McLane, M. A., and Niewiarowski, S. (1997) Significance of RGD loop and C-terminal domain of echistatin for recognition of αIIbβ3 and αVβ3 integrins and expression of ligand-induced binding site, *Blood 90*, 1565−1575.
54. Pfaff, M., Tangemann, K., Muller, B., Gurrath, M., Muller, G., Kessler, H., Timpl, R., and Enge, J. (1994) Selective recognition of cyclic RGD peptides of NMR defined conformation by αIIbβ3, αVβ3, and α5β1 integrins, *J. Biol. Chem. 269*, 20233−20238.
55. Calvete, J. J., Moreno-Murciano, M. P., Theakston, R. D., Kisiel, D. G., and Marcinkiewicz, C. (2003) Snake venom disintegrins: Novel dimeric disintegrins and structural diversification by disulphide bond engineering, *Biochem. J. 372*, 725−734.
56. Esnouf, R. M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities, *J. Mol. Graphics Modell. 15*, 132−134.
57. Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structure, *J. Appl. Crystallogr. 24*, 946−950.
58. Merritt, E. A., and Murphy, M. E. P. (1994) Raster3D version 2.0: A program for photorealistic molecular graphics, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 50*, 869−873.
59. Nicholls, A., Sharp, K., and Honig, B. (1991) Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins: Struct., Funct., Bioinf. 11*, 281−296.
60. DeLano, W. L. (2002) *The PyMOL User's Manual*, DeLano Scientific, San Carlos, CA.

BI050849Y