Volume 309, number 3, 316–320 FEBS 11523 September 1992

The disulfide bridge pattern of snake venom disintegrins, flavoridin and echistatin

Juan J. Calvete[a,b], Yuqin Wang[c], Karlheinz Mann[a], Wolfram Schäfer[a], Stefan Niewiarowski[c] and Gwendolyn J. Stewart[c]

[a]*Max-Planck-Institut für Biochemie, W-8033 Martinsried, Germany,* [b]*Instituto de Química-Física 'Rocasolano' C.S.I.C., Madrid, Spain* and [c]*Department of Physiology and Sol Sherry Thrombosis Research Center, Temple University School of Medicine, Philadelphia, PA 19140, USA*

Received 9 July 1992

Flavoridin and echistatin, isolated from the venom of *Trimeresurus flavoviridis* and *Echis carinatus*, respectively, belong to the disintegrin family of integrin β_1 and β_3 inhibitors of low molecular weight RGD-containing, cysteine-rich peptides. Since disulfide bonds are critical for expression of biological activity, we sought to determine their location in these two proteins. In flavoridin, direct evidence for the existence of linkage between Cys^4–Cys^{19} and between Cys^{43} and Cys^{54} was obtained by analysis of proteolytic products, and indirect evidence suggests links between Cys^8–Cys^{14} and Cys^{13}–Cys^{36}. In echistatin, links between Cys^8–Cys^{37} and Cys^{20}–Cys^{39} were identified by direct chemical analysis.

Inhibitor of integrin; Disintegrin; Disulfide bridge; Echistatin; Flavoridin

1. INTRODUCTION

Disintegrins represent a novel family of integrin β_1 and β_3 inhibitory proteins isolated from venoms of various vipers (family Viperidae) [1–4]. They interfere with biological processes such as cell adhesion to extracellular matrix, including adhesion of melanoma cells and fibroblasts to fibronectin [5,6], and are potent inhibitors of platelet aggregation (IC_{50}=20–200 nM) [1–11]. Analysis of the amino acid sequences of disintegrins [1–4, 6–12] shows that the tripeptide, RGD, which represents a cell attachment recognition sequence in a number of adhesive proteins [13], has been conserved in all but one of these molecules. The exception is barbourin, where lysine replacing arginine is thought to contribute to its integrin $\alpha IIb/\beta_3$ antagonist specificity [4]. The appropriate spatial configuration of this recognition motif, determined largely by disulfide bridges, is essential for the inhibitory potency of disintegrins [3,5,7,14,15]. In a previous study, we established the disulfide bond pattern in albolabrin [15]. The purpose of this study was to determine the location of disulfide bridges in flavoridin, as well as in natural and synthetic echistatin, by isolation and sequencing of proteolytic fragments.

2. MATERIALS AND METHODS

Synthetic echistatin was purchased from Bachem Inc. (USA) and was used without further purification. Natural echistatin and flavoridin were purified by reverse-phase HPLC from the lyophilized venom of *Echis carinatus* and *Trimeresurus flavoviridis* (Sigma Chemical Co., St. Louis, MO), respectively, using a Vydac C18 column (Hesperia, CA) (0.44 x 25 cm, 300 Å pore size) and eluting with a mixture of 0.1% (v/v) TFA in water (solution A) and acetonitrile (solution B). The column was first eluted isocratically (100% A) for 10 min, followed by a linear gradient up to 80% B in 50 min. Active peptides were identified by their ability to inhibit ADP-induced platelet aggregation as described [15].

Echistatin, either natural or synthetic, and flavoridin, (2 mg/ml in 100 mM NH_4HCO_3, pH 8.0) were digested with TPCK-trypsin (Sigma) at an enzyme:substrate ratio of 1:50 (w/w) for 4 h at 37°C. Digestion of echistatin was stopped with formic acid (final concentration 30% (v/v)), and the tryptic peptides were separate by reverse-phase HPLC on a Lichrospher 100 RP-18 (5 μM particle size) (Merck, Darmstadt) column equilibrated in a mixture of 0.1% (v/v) TFA in water (solution A) and in acetonitrile (solution B) (95% A/5% B) and eluted at 1 ml/min, at first isocratically for 5 min and then followed by a linear gradient up to 40% B in 70 min.

The enzyme in the tryptic digest of flavoridin was heat-inactivated (100°C for 5 min), endoproteinase Asp-N (Boehringer-Mannheim, Indianapolis, IN) was added up to a final enzyme:substrate ratio of 1:25 (w/w) and incubated 6 h at 37°C. Thereafter, the resulting peptides were separated as above. Elution conditions were: 5 min isocratically (100% A) followed by a linear gradient up to 30% B in 90 min.

The isolated peptides were characterized by amino acid analysis (after acid hydrolysis with 6 N HCl for 24 h at 110°C using a Biotronic (amino acids analyzer) and amino-terminal sequence analysis (using an Applied Biosystems, San Francisco, CA) gas-phase sequencer model 473A. Mass spectra were recorded with a mass spectrometer MAT 900 (Finnigan MAT, Bremen, FRG) equipped with liquid secondary ion ionization system.

3. RESULTS

3.1. *Disulfide bridge pattern of flavoridin*

Proteolytic digestion of flavoridin with trypsin and endoproteinase Asp-N gave nine fractions (Table I).

Correspondence address: J. Calvete, Instituto 'Rocasolano' C.S.I.C., C/Serrano 119, 28006-Madrid, Spain. Fax: (34)(1) 564 2431.

Published by Elsevier Science Publishers B.V.

Table I

Amino acid sequence of flavoridin [1] and of peptides obtained by reverse-phase HPLC after its digestion with trypsin and endoproteinase Asp-N

Flavoridin:

GEECDCGSPSHPCCDAATCKLRPGAQCADGLCCDQCRFKKKRTICRIARGDFPDRCTGLSHDCPRNHL
1 10 20 30 40 50 60 70

↓
↓

Trypsin, Asp-N

F-1:	38	FK	39
	47	IAR	49
	70	DL	70
F-2:	1	GEEC	3
	15	DAATCK	20
F-3:	38	FKKK	41
	68	HN	69
F-4:	21	LRPGAQ	26
F-5:	27	CA	28
	32	CC	33
	58	CTGLSH	63
F-6:	43	TICR	46
	64	DCPR	67
F-7:	58	TGLSH	63
F-8:	5	DCGSPSHPCC	14
	34	DQCR	37
F-9:	21	LRPGAQCADGLCC	33
	58	C	

Table II

Amino acid sequence of echistatin and of peptides obtained by its digestion with TPCK-trypsin (the order of some products has been altered to show a degradation pathway from T-3 through T-5)

Echistatin: ECESGPCCRNCKFLKEGTICKRARGDDMDDYCNGTCDPRNPHKGPAT
1 10 20 30 40

↓
↓

TPCK-Trypsin

T-1:	42	NPHK	45
	46	GPAT	49
T-2:	13	FLK	15
T-3:	1	ECESGPCCRNCKFLKEGTICK	21
	25	GDDMDDYCNGKTCDCPR	41
T-6:	1	ECESGPCCRNCK	12
	16	EGTICK	21
	25	GDDMDDYCNGKTCDCPR	41
T-4:	1	ECESGPCCR	9
	10	NCK	12
	25	GDDMDDYCNGKTCDCPR	41
T-5:	1	ECESGPCCR	9
	10	NCK	12
	16	EGTICK	21
	25	GDDMDDYCNGK	35
	36	TCDCPR	41

Two fractions, F-2 and F-6 (M + H$^+$ = 979.1), each contained two sequences in a 1:1 molar ratio, indicating the existence of disulfide bonds between Cys^4 and Cys^{19} in fraction F-2, and between Cys^{45} and Cys^{65} in fraction F-6. On the other hand, fractions F-5, F-8 and F-9 each contained four cysteine residues involved in the formation of two disulfide bridges. Although at this stage the actual pairing could not be established, the fact that a disulfide bond between adjacent cysteines is sterically hindered limits to four the actual pairing possibilities. Thus, Cys^6 cannot be paired with Cys^{36}, since such a pairing would leave two adjacent cysteines, Cys^{13} and Cys^{14}, unpaired. The pairing of Cys^{27}–Cys^{58} is also forbidden, since this would leave Cys^{32} and Cys^{33} unpaired. Among the remaining possibilities, bonding between Cys^6–Cys^{14}, Cys^{13}–Cys^{36}, Cys^{32}–Cys^{58}, and Cys^{27}–Cys^{33} would be consistent with the recently reported NMR solution structure of kistrin [16], a similar disintegrin [3].

3.2. *Disulfide bridge pattern of echistatin*

The tryptic peptides of synthetic echistatin are shown in Table II. T-1, T-2 an T-5 were final digestion products, while T-3 was a partial digestion product that gave rise to T-6, T-4 and T-5 in that order. This established that the first cleavage sites occur at bonds K^{21}–R^{22}, R^{22}–A^{23}, R^{24}–G^{25}, R^{41}–N^{42} and K^{45}–G^{46}, indicating that the RGD recognition motif and the C-terminal tail are both readily exposed; the other bonds, K^{12}–F^{13}, K^{15}–E^{16}, R^9–N^{10} and K^{35}–T^{36}, were presumably less accessible. The final tryptic degradation product, T-5 (Table II) was composed of five fragments cross-linked by four disulfide bonds giving rise to several possibilities for disulfide bridges.

To establish the actual disulfide bridge arrangement in echistatin, T-5 was further degraded with elastase, and the resulting peptides were isolated by reverse-phase HPLC. This yielded a set of successively degraded products from which a degradation pattern as well as the position of disulfide bridges could be deduced (Table III). On degradation of T-5/E-7 to T-5/E-6, Cys^{20} and Cys^{39} both disappeared indicating that Cys^{20} and Cys^{39} were disulfide linked. This confirms NMR studies [17–21]. Cleavage of T-5/E-5 (M + H$^+$ = 2,239.8) to T-5/E-4 resulted in the loss of Cys^8 and Cys^{37} indicating that they were disulfide bonded. Analysis of fragment T-5/E-1 (M + H$^+$ = 1,311.1) did not show whether Cys^7 was paired to Cys^{11} or Cys^{32}. The second possibility would agree with recently reported ^{1}H NMR studies of synthetic echistatin in solution [16–20]. Furthermore, an S–S bond between Cys^2 and Cys^{11} has been described [18]. The resonance studies could not distinguish between the pairing possibilities Cys^8–Cys^{37}:Cys^7– Cys^{32}, and Cys^8–Cys^{32}:Cys^7–Cys^{37}). Our protein-chemical data established the Cys^8–Cys^{37} linkage, which was previously predicted on the basis of the disulfide bridge pattern reported for albolabrin [14].

Table III

Amino acid sequence of the digestion products of peptide T-5 by porcine pancreatic elastase (a degradation pathway for T-5/E-8 through T-5/E-1 can be followed)

T-5/E-1:	7	CCR	9
		C	20
		C	11 or 32
	36	TCDCPR	41
T-5/E-2:	16	EGTI	19
T-5/E-3:	25	GDDMDDY	31
T-5/E-4:	1	ECESGPC	6
	10	NCK	12
	25	GDDMDDYC	32
T-5/E-5:	1	ECESGPCC	8
	10	NCK	12
	25	GDDMDDYC	32
	37	C	
T-5/E-6:	1	ECESGPCCR	9
	10	NCK	12
	25	GDDMDDYCNGK	35
	36	TCD	38
T-5/E-7:	1	ECESGPCCR	9
	10	NCK	12
	20	CK	21
	25	GDDMDDYCNGK	35
	36	TCDC	38
T-5/E-8:	1	ECESGPCCRC	9
	10	NCK	12
	16	EGTICK	21
	25	GDDMDDYCNGK	35
	36	TCDC	39

When natural echistatin was proteolytically degraded in the same way as the synthetic molecule, a very minor component (estimated to represent less than 5%), which contained the peptides ^{1}ECESG^{5} and ^{25}GDDMDDYC^{32} disulfide bridged (M + H^{+} = 1,454.5) was found. Whether this peptide resulted from limited disulfide bridge rearrangement during isolation and/or degradation of the protein, or wether it reflects the existence of a small echistatin population with an alternative disulfide bridge pattern (which interestingly would be similar to the one found in albolabrin [15]) cannot be concluded from our data. Nevertheless, computer modelling of this alternative pairing showed that it would not introduce significant changes into the conformation of the RGD loop.

4. DISCUSSION

Since disulfide bridges are essential for expression of the biological activity of disintegrins [3,4,11,13,14], we were interested in the disulfide bridge patterns in these peptides. Isolation, amino acid sequencing and analysis of proteolytic fragments of flavoridin (Table I) and echistatin (Tables II and III) enabled us to identify some disulfide bridges (Fig. 1). The disulfide bridge pattern of albolabrin as previously described [15] was added for comparison. We identified two S–S bridges in flavoridin, and the other ones were deduced. Our observations on echistatin confirm the location of two disulfide bridges (Cys^{2}–Cys^{11} and Cys^{20}–Cys^{39}) previously proposed on the basis of NMR studies [17,18]. We were able to add the location of one more (Cys^{8}–Cys^{37}) based on our protein chemistry data. Bond Cys^{7}–Cys^{32} was deduced since these were the only remaining cysteines.

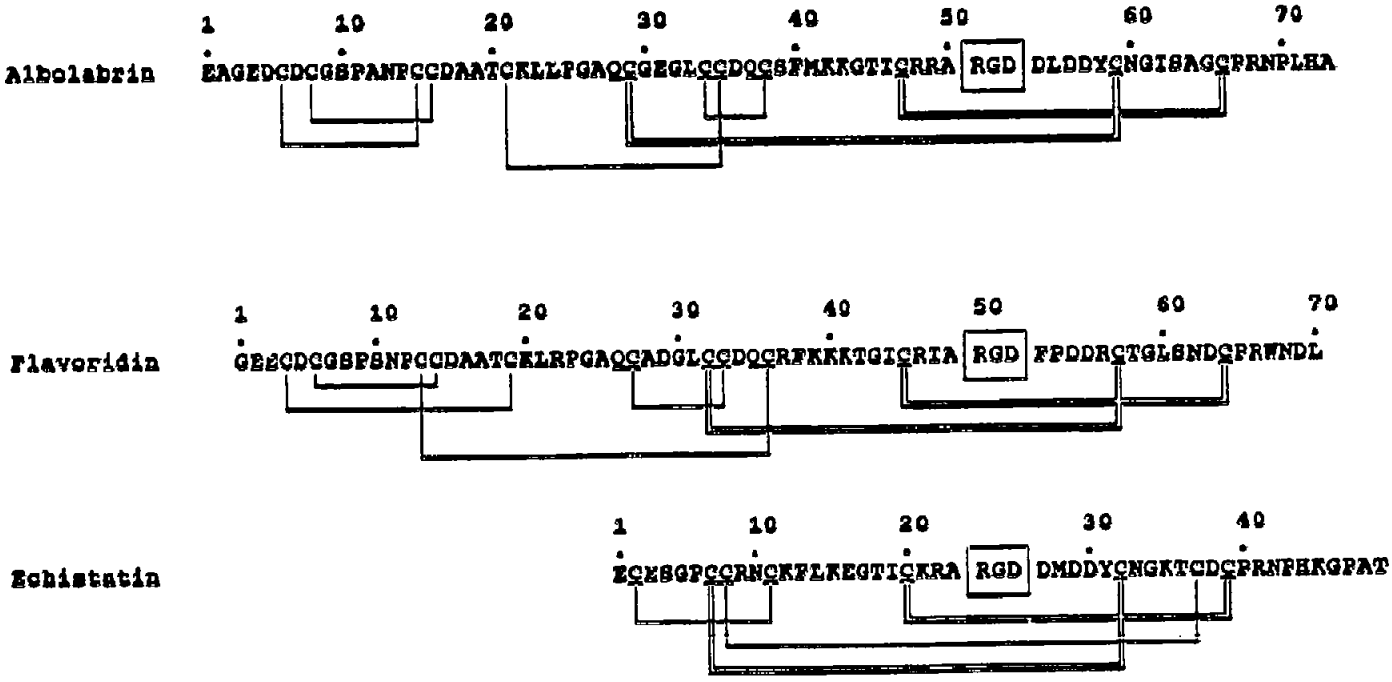

Fig. 1. Schematic representation of the structure of three disintegrins. Disulfide bonds that are conserved or semi-conserved are shown by double lines. The other bonds are shown by single lines.

In flavoridin and echistatin, as well as in albolabrin [15], the disulfide bonds form a complex pattern of overlapping loops and loops within loops (Tables I–III, Fig. 1). This pattern extends from the first to the last cysteine in the peptide, and leaves only a few N-terminal (one to five) and C-terminal (six to ten) amino acids beyond these loops [1–5, 6–10]. In all three disintegrins, two disulfide bridges form an interesting motif with respect to the RGD sequence. A domain containing the RGD sequence is defined by a combination of two disulfide bonds. One is formed between the cysteine closest to the RGD sequence on the N-terminal end of the peptide and the last cysteine in the peptide. This bond is highly conserved. The other is between the cysteine closest to the RGD sequence on the C-terminal end of the molecule and one cysteine of the N-terminal segment.

The formation of these two disulfide bridges has two consequences, (i) delineation of an RGD-containing domain that has 11 amino acids between the cysteines, and (ii) formation of a shorter (6 amino acid) segment of peptide that limits the distance between the S–S bridges defining the RGD-containing domain (a 'retainer' segment). Obviously this arrangement throws the RGD-containing domain into a loop, a finding consistent with observations made by NMR [17–21]. This motif depends on the presence of two or more cysteines, both C-terminal and N-terminal to the RGD sequence, and the formation of at least two trans-RGD disulfide bonds. All disintegrins sequenced to date have two cysteines C-terminal to the RGD sequence, except for echistatin and eristostatin which have three C-terminal cysteines. All sequenced disintegrins have more than two cysteines N-terminal to the RGD sequence, with the number depending on the length of the peptide. If the 'retainer' segment is able to change its length by forming a secondary structure, this arrangement might provide a means for varying the shape of the loop, thereby influencing the configuration of the RGD sequence and possibly the biological activity of the peptide.

Elucidation of the disulfide bridge patterns of flavoridin and echistatin (Fig. 1) provide information for revising our previous prediction that the disulfide bond pattern would be the same in all disintegrins [15]. It is noteworthy that flavoridin and kistrin [16] have identical patterns of S–S bridges [16]. The differences between albolabrin compared to flavoridin and kistrin probably depend on differences in folding which would bring different cysteines within bonding distance of each other. Differences in folding patterns probably depend on differences in amino acid sequences in the peptides.

Three peptides with similar ability to antagonize the platelet fibrinogen receptor have been isolated from other sources and sequenced. They have little resemblence to the disintegrins in either amino acid composition or arrangement of cysteines. Two peptides, decorsin and ornatin, isolated from leeches, have only one cysteine C-terminal to the RGD sequence [22,23]. Thus, only one trans-RGD disulfide bond can be formed, i.e. an RGD-containing loop in these peptides must be defined by one rather than two disulfide bonds. The other peptide, mambin, was isolated from the venom of a snake belonging to the family *Elapidae*. Mambin differs from leech peptides as well as disintegrins [24]. While mambin has three cysteines C-terminal and five N-terminal to the RGD sequence, McDowell et al. [24] suggested that the RGD-containing loop is defined by one disulfide bond, that between Cys^{43}–Cys^{54}. The influence of these different patterns of disulfide bonds on the conformation of the RGD-containing loop and on biological functions of the various peptides will help to establish structure–function relationships.

Acknowledgements: The provision of a Max-Planck-Gesellschaft fellowship to J.J.C. is gratefully acknowledged. This study was also supported by a National Institutes of Health Grant HL 45486 (S.N. and G.J.S.).

REFERENCES

[1] Gould, R.J., Polokoff, M.A., Friedman, P.A., Huang, T.-F., Holt, J.C., Cook, J.J. and Niewiarowski, S. (1990) Proc. Soc. Exp. Biol. Med. 195, 168–171.
[2] Chao, B.H., Jakubowski, J.A., Savage, B., Chow, E.P., Marzec, U.M., Harker, L.A. and Maraganore, J.M. (1989) Proc. Natl. Acad. Sci. USA 86, 8050–8054.
[3] Dennis, M.S., Henzel, W.J., Pitti, R.M., Lipari, M.T., Napier, M.A., Deisher, T.A., Bunting, S. and Lazarus, R.A. (1989) Proc. Natl. Acad. Sci. USA 87, 2471–2475.
[4] Scarborough, R.M., Rose, J.W., Hsu, M.A., Phillips, D.R., Fried, V.A., Campbell, A.M., Nannizzi, L. and Charo, I. (1991) J. Biol. Chem. 266, 9359–9362.
[5] Knudsen, K.A., Tuszynski, G.P., Huang, T.-F. and Niewiarowski, S. (1988) Exp. Cell Res. 179, 42–49.
[6] Gan, Z.-R., Gould, R.J., Jacobs, J.W., Friedman, P.A. and Polokoff, M.A. (1988) J. Biol. Chem. 263, 19827–19832.
[7] Huang, T.-F., Holt, J.C., Kirby, E.P. and Niewiarowski, S. (1989) Biochemistry 28, 661–666.
[8] Williams, J.A., Rucinski, B., Holt, J.C. and Niewiarowski, S. (1990) Biochim. Biophys. Acta 1039, 81–89.
[9] Rucinski, B., Niewiarowski, S., Holt, J.C., Soszka, T. and Knudsen, K.A. (1990) Biochim. Biophys. Acta 1054, 257–262.
[10] Musial, J., Niewiarowski, S., Rucinski, B., Stewart, G.J., Cook, J.J., Williams, J.A. and Edmunds Jr., L.H. (1990) Circulation 82, 261–273.
[11] Huang, T.-F., Sheu, J.-S., Teng, C.-M., Chen, S.-W. and Liu, C.-S. (1991) J. Biochem. 109, 328–334.
[12] Shebuski, R.J., Ramjit, D.R., Bencen, G.J. and Polokoff, M.A. (1989) J. Biol. Chem. 264, 21550–21556.
[13] Ruoslahti, E. and Pierschbacher, M.D. (1987) Science 238, 491–497.
[14] Huang, T.-F., Holt, J.C., Lukasiewicz, H. and Niewiarowski, S. (1987) J. Biol. Chem. 262, 16157–16163.
[15] Calvete, J.J., Schäfer, W., Soszka, T., Lu, W., Cook, J.J., Jameson, B.A. and Niewiarowski, S. (1991) Biochemistry 30, 5225–5229.
[16] Adler, M., Lazarus, R.A., Dennis, M.S. and Wagner, G. (1991) Science 253, 445–448.
[17] Saudek, V., Atkinson, R.A. and Pelton, J.T. (1991) Biochemistry 30, 7369–7372.
[18] Dalvit, C., Widmer, H., Bovermann, G., Breckenridge, R. and Metternich, R. (1991) Eur. J. Biochem. 202, 315–321.
[19] Cooke, R.M., Carter, B.G., Martin, D.M.A., Murray-Rust, P. and Weir, M.P. (1991) Eur. J. Biochem. 202, 323–328.

[20] Saudek, V., Atkinson, R.A., Lepage, P. and Pelton, J.T. (1991) Eur. J. Biochem. 202, 329–338.
[21] Chen, Y., Pitzenberger, S.M., Garsky, V.M., Lumma, P.K., Sanyal, G. and Baum, J. (1991) Biochemistry 30, 11625–11636.
[22] Seymour, J.L., Henzel, W.J., Nevins, B., Stults, J.T. and Lazarus, R.A. (1990) J. Biol. Chem. 265, 10143–10147.
[23] Mazur, P., Henzel, W.J., Seymour, J.L. and Lazarus, R.A. (1991) Eur. J. Biochem. 202, 1073–1082.
[24] McDowell, R.S., Dennis, M.S., Louie, A., Shuster, M., Mulkerrin, M.G. and Lazarus, R.A. (1992) Biochemistry 31, 4766–4772.