Disulphide-bond pattern and molecular modelling of the dimeric disintegrin EMF-10, a potent and selective integrin $\alpha_5\beta_1$ antagonist from *Eristocophis macmahoni* venom

Juan J. CALVETE*[1], Michael JÜRGENS†, Cezary MARCINKIEWICZ‡, Antonio ROMERO§, Michael SCHRADER† and Stefan NIEWIAROWSKI‡

*Instituto de Biomedicina de Valencia, C.S.I.C., Jaime Roig 11, E-46010 Valencia, Spain, †BioVisioN GmbH & Co. KG, D-30625 Hannover, Germany, ‡Department of Physiology, Sol Sherry Thrombosis Research Center, Temple University School of Medicine, Philadelphia, PA 19140, U.S.A., and §Centro de Investigaciones Biológicas, C.S.I.C., E-28006 Madrid, Spain

The disulphide-bond pattern of the heterodimeric disintegrin EMF-10, a potent and selective integrin $\alpha_5\beta_1$ antagonist from *Eristocophis macmahoni* venom, was established by combination of amino-acid analysis, N-terminal sequencing and collision-induced dissociation by nanoelectrospray ionization quadrupole ion-trap MS of fragments isolated by reversed-phase HPLC after degradation of EMF-10 with oxalic acid. Each EMF-10 subunit contains four intrachain disulphide bonds. Two interchain cystine residues join the EMF-10 polypeptides. The intrachain linkages are conserved in monomeric disintegrins. A molecular model of EMF-10 was built using averaged NMR co-ordinates of flavoridin as a template. The active hairpin loops of the EMF-10 subunits occupy opposite locations at the ends of an elongated disulphide-bond ladder. In the EMF-10 model the N-terminal polypeptide of EMF-10B is close to the RGD-loop of the EMF-10A subunit, suggesting that the N-terminal region of the B-subunit could potentially influence the biological activity of the A-subunit.

Key words: fibronectin-receptor inhibitor, mass spectrometry, RGD-containing peptide.

INTRODUCTION

The disintegrins are a family of low-molecular-mass, cysteine-rich peptides present in the venom of *Viperidae* and *Crotalidae* snake species [1,2]. Since the early characterization of the first two disintegrins, trigramin [3] and echistatin [4] from the venoms of *Trimeresurus gramineus* and *Echis carinatus*, respectively, close to 50 disintegrins have been isolated from viper and rattlesnake genera over the last decade, and their primary structures have been elucidated. Disintegrins are specific and potent inhibitors that compete with, and prevent the binding of, adhesive ligands to integrin receptors of the $\beta1$ and $\beta3$ families [1,2]. The activity of disintegrins depends on their tertiary structures, maintained by the appropriate disulphide-bond pairing of 8–14 cysteines [5]. Alanine-scanning mutagenesis [6] and NMR spectroscopic studies [7–10] have shown that the inhibitory mechanism of disintegrins involves a (R/K)GD tripeptide integrin-binding motif presented at the tip of a mobile 13-amino-acid loop joining two short β-strands, which protrude 14–17 Å from the protein core. The amino acids flanking the (R/K)GD sequence are responsible for the shape of the active loop and have a major role in determining the inhibitory potency and integrin-recognition specificity [11–13]. In addition, the C-terminal tail of echistatin induces conformational changes on integrins $\alpha_v\beta_3$ and $\alpha_{IIb}\beta_3$ and potentiates the inhibitory effect of the disintegrin [12].

Disintegrins can be classified in different groups based on the size and number of disulphide bonds [1,2,14]. The short disintegrins (echistatin, eristostatin) contain 49–51 amino acids cross-linked by four disulphide bonds [7,15]. Medium disintegrins (flavoridin, kistrin, albolabrin, etc) have six disulphide bonds within a 68–73-residue polypeptide chain [5,6,9,10]. The long disintegrin bitistatin is a polypeptide with 83 amino acids and seven disulphide bonds [14]. These short, medium and long disintegrins are monomeric molecules. It has been suggested that disintegrins may also occur as multimeric proteins. Gabonin, from *Bitis gabonica* venom, has been reported to be a 21-kDa disulphide-bonded dimer [16], and cerastatin, a 32-kDa protein from *Cerastes cerastes* viper venom, appears to be made up of three or more disulphide-linked subunits [17]. These molecules have not been structurally characterized, however. On the other hand, contortrostatin, a 13.5-kDa disintegrin isolated from the venom of *Agkistrodon contortrix contortrix*, is a homodimer [18]; and EC3 (*Ec. carinatus* fraction 3) and EMF-10 (*Eristocophis macmahoni* fraction 10), purified from the venoms of *Ec. carinatus* [19] and *E. macmahoni* [20], respectively, are heterodimeric disintegrins. Gabonin, cerastatin and contortrostatin are potent antagonists of the fibrinogen receptor (integrin $\alpha_{IIb}\beta_3$) and inhibit platelet aggregation. On the contrary, EC3 is a strong inhibitor of $\alpha_4\beta_1$ and $\alpha_4\beta_7$ integrins and a weak inhibitor of both $\alpha_5\beta_1$ integrin and platelet aggregation, whereas EMF-10 is a strong and very selective inhibitor of $\alpha_5\beta_1$ integrin [20]. In brief, EMF-10 inhibits adhesion of cells expressing $\alpha_5\beta_1$ to fibronectin and cells expressing $\alpha_v\beta_3$ to vitronectin with IC_{50} values of 4 and 5000 nM, respectively [20].

Abbreviations used: TFA, trifluoroacetic acid; MALDI–TOF, matrix-assisted laser-desorption ionization–time-of-flight; QIT, quadrupole ion-trap; CID, collision-induced dissociation; nanoES, nanoelectrospray.

[1] To whom correspondence should be addressed (e-mail jcalvete@ibv.csic.es).

The amino-acid sequences of EC3 [19] and EMF-10 [20] subunits have been determined recently. The amino-acid sequence of contortrostatin subunits has been completed, although not yet published (cited in [21]). Each contortrostatin polypeptide chain contains an RGD motif in the proper location relative to monomeric disintegrins [21]. The EMF-10 A-subunit also displays the RGD motif, but the integrin-binding loop of the EMF-10 B-subunit displays an MGD sequence. RGD is substituted in EC3 A- and B-subunits by VGD and MLD, respectively. The KRAMLDGLNDY sequence of subunit B of EC3 (with the MGD sequence underlined) represents an epitope recognizing α_4 integrins [19], and the WPAMGDWNDDY hairpin-loop sequence appears to be responsible for the selective $\alpha_5\beta_1$ inhibitory activity of EMF-10 [20]. Noteworthy is that the amino-acid sequences of contortrostatin, EC3 and EMF-10 show that each of the six subunits contains 10 cysteines and that, in comparison with monomeric medium disintegrins, there is a 12-residue truncation at the N-terminus. In medium monomeric disintegrins, the two cysteine residues in the missing region are disulphide-bonded to the fourth and fifth cysteine residues, suggesting that an abnormal cysteine pairing might be responsible for dimerization of the 10-cysteine-containing disintegrin polypeptides. The amino-acid sequences and relative spatial orientations of the active loops of both subunits may, in turn, modulate the biological activity of dimeric disintegrins. As a first step to establishing the structure–function correlation of this novel group of disintegrins, we have determined the pattern of intra- and inter-disulphide bonds of EMF-10.

EXPERIMENTAL

Isolation of EMF-10

EMF-10 [30 mg/ml in 0.1 % trifluoroacetic acid (TFA)] was isolated from lyophilized *E. macmahoni* venom (Latoxan, Valence, France) using a reversed-phase Vydac 250 × 10 mm C_{18} HPLC column eluted at 2 ml/min with a linear gradient of 0–80 % solution B for 45 min, where solution A is 0.1 % TFA in water and solution B is 0.1 % TFA in acetonitrile. EMF-10 (eluted at ≈ 40 % solution B) was lyophilized and purified using the same chromatographic conditions but with a gradient from 0–60 % solution B for 45 min. Purity of EMF-10 was checked by N-terminal sequence analysis (see below) and electrospray ionization MS (using a Sciex API-III triple quadrupole LC-MS/MS instrument) [20]. N-terminal sequence analysis yielded two residues in each cycle corresponding to the A- and B-subunits of EMF-10 (EMF-10A, MNSANP…; EMF-10B, ELLQNS…) and MS showed typically two major ions of molecular masses 14947 ± 4 (EMF-10B1) and 14575 ± 2 Da (EMF-10B2) due to Lys/Phe heterogeneity at position 35 and C-terminal trimming of the B-subunit.

Protein cleavage and isolation of fragments

Samples of 0.2 mg of EMF-10 (10 mg/ml in water) were degraded with 250 mM oxalic acid (final concentration) for 4 and 12 h at 100 °C in sealed, evacuated ampoules [22]. Hydrolysates were dried in a Speed-Vac, dissolved in 0.1 % TFA in 5 % aqueous acetonitrile and fragments were isolated by reversed-phase HPLC on a Lichrospher RP100 C_{18} (250 mm × 4 mm, 5 μm pore size) column eluting at 1 ml/min with solutions A and B (as above). Peptides were eluted isocratically (5 % solution B) for 5 min, followed by a gradient of 5–50 % solution B for 90 min, and 50–70 % solution B for 20 min. Elution was monitored by UV absorbance at 220 nm, and chromatographic fractions were collected manually.

Amino-acid analysis and N-terminal sequencing

Oxalic acid-derived fragments were subjected to amino-acid analysis using a Pharmacia AlphaPlus analyser (after sample hydrolysis with 6 M HCl at 110 °C for 18 h in evacuated and sealed ampoules) and N-terminal amino-acid sequence analysis (using an Applied Biosystems Procise 494 sequencer and a Beckman Porton LF3000 instrument) following the manufacturer's protocols.

Matrix-assisted laser-desorption ionization–time-of-flight (MALDI–TOF) MS

Aliquots of 0.5 μl of reversed-phase HPLC fractions were applied automatically using the pipetting robot SymBiot I (PerSeptive Biosystems, Framingham, MA, U.S.A.) to the inner 8 × 8 positions of a 10 × 10-position polished stainless steel sample plate. Each sample was mixed on the plate with the same volume of a matrix composed of 5–10 μg/μl each of α-cyano-4-hydroxycinnamic acid and 6-desoxy-L-galactose [L(−)fucose] (Sigma-Aldrich, Deisenhofen, Germany) in a 1:1 (v/v) mixture of acetonitrile/0.1 % TFA in water. Samples were dried at ambient temperature using a microventilator. Measurements were performed in reflectron mode with a Voyager-DE STR (PerSeptive Biosystems) instrument equipped with a 2-m flight tube (3 m in reflector mode) and a 337-nm nitrogen laser. Positive ions were accelerated at 20 kV and 256 laser shots were accumulated for a single spectrum.

Quadrupole ion-trap (QIT) MS

Lyophilized samples were dissolved in methanol/0.6 % aqueous formic acid (1:1, v/v), and spectra were acquired with a Finnigan/ThermoQuest LCQ (ThermoQuest, San Jose, CA, U.S.A.) mass spectrometer. Ionization was performed using a nanoelectrospray (nanoES) interface and nanoES glass capillaries (Protana, Odense, Denmark). Voltages and trap-filling parameters were optimized for nanoES experiments. For the analysis of each chromatographic fraction 1–5 μl of the analyte solutions were used. Mass spectra were externally calibrated, yielding a mass accuracy better than 300 p.p.m. Because of high background from the samples, only very weak signals could be detected scanning the full mass range. For this reason, monoisotopic MALDI mass data were used to search for the doubly and triply charged ion species. For selecting the precursor ions and thus filling the ion trap only with ions of potential interest, the m/z window was set to ±2.5–4. To enhance resolution and therefore to confirm the charge state of the precursor ions, 'zoom scans' with a m/z window of 10 were applied. Selected ions were subjected to collision-induced dissociation (CID) MS/MS with optimized collision energy. Intense fragment ions were submitted to another CID process (MS^3 experiments) to obtain further structural information. Up to 100 scans were added as centroid peak data and interpreted using the Explore software (ThermoQuest).

Nomenclature of the collision-induced fragmentation pattern

For reasons of simplicity, the standard nomenclature used to identify the fragment ion series produced by CID of protonated monomeric peptides [23] was not adopted here to describe the fragmentation pattern of EMF-10 fragments containing two or more disulphide-bonded peptides. Instead, each peptide of the fragment was identified by a Greek letter (α, β, γ), and numbers

Table 1 Assignment of fragments generated by degradation of EMF-10 with 250 mM oxalic acid at 110 °C for 4 h (A) and 12 h (B; labelled with *) based on amino-acid composition and MS data

Amino-acid sequences that are identical in EMF-10A and EMF-10B are shown in parentheses. Cysteine residues of EMF-10A and EMF-10B are identified with the prefixes A and B, respectively. §, Peptide-sequence information gathered by CID using nanoES QIT MS. Where no definite disulphide bond is shown, bonding patterns are uncertain.

(A)

Fragment	Sequence	Mass (Da)	EMF–10A	EMF–10B	Disulphide bond(s)
1	NPQKSEEED	1133.3	61–69		
7	KSEEED	7.35.3	64–69		
8	KSEEE	620.2	64–68		
13	QNSG	404.1	4–7		
14	ELLQN	615.2			
15	CKKGR	1288.4	39–43		
	DCPRNP		(57–62)	(60–65)	A39–(A58/B61)
16	CKKGR	1456.4	39–43		
	DCPRNPW–OH		57–63		A39–A58
22a	TVCWPAMGD	1299.2	40–48		
	SDC		(56–58)	(59–61)	B42–(A58/B61)
22b	GTVCWPAMGD	1356.2	39–48		
	SDC		(56–58)	(59–61)	B42–*A58/B61)
23a§	ELLQNSGNPCCDPV–OH	2163.4	1–14		
	CDP		8–10		B11–A8
	NCK		(29–31)	(32–34)	B10–(A30/B33)
23b§	QNSGNPCCDPVT	2220.6	1–12		
	CDPIT	(-2 H_2O)		8–12	B11–A8
	DNCK			31–34	B10–B33
24f§	GEHC	2777.5	(18–21)	(21–24)	
	CCDN–OH			29–32	(A21/B24), B29, B30, B54
	WPAMGDWNDDYCTGISS–OH			32–59	
24j§	GEHCVSGPCCDN	2630.3	21–32		
	GDWNDDYCTGISS–OH		47–59		B(24, 29, 30, 54)

(B)

Fragment	Sequence	Mass (Da)	EMF–10A	EMF–10B	Disulphide bond(s)
3a*	DCPRNP	1291.9	(57–62)	(60–65)	
	CKKGR		39–43		A39–(A58/B61)
3b*	DCPRNP	1349.9	(57–62)	(60–65)	
	CKKGRG		39–44		A39–(A58/B61)
3c*	DCPRNP	1464.9	(57–62)	(60–65)	
	CKKGRGD		39–45		A39–(A58/B61)
5a*	DCPRNPV	1175.7		60–66	
	GTVC			39–42	B42–B61
5b*	DCPRNPV	1246.8		60–66	
	AGTVC			38–42	B42–B61
5C*	DCPRNPV-OH	1303.8		60–67	
	GTVC			39–42	B42–B61
6*§	NPGTICKKG-OH	1290.7	34–43		
	SSDC-OH		(55–58)	(58–61)	A39–(A58/B61)
7*	GTICKK	1618.9	36–41		
	SDCPRNPW		56–63		A39–A58
10a*§	PWKSEEE-OH	890.5	62–68		
10b*§	NPWKSEEED-OH	1114.8	61–69		
11*	ELLQNSGNPC-OH	1272.8		1–10	
	NC		(29–30)	(32–33)	(A30/B33)–B10
12*	TICKKGR	1222.4	37–43		
	SDCP		(56–59)	(59–62)	A39–(A58/B61)
13a*§	PWKSEEED-OH	1001.5	62–69		
14a*§	RNPWKSEE	1045.7	60–67		
14b*§	CKKG	1196.8	39–42		
	SSDCPRN-OH		(55–61)	(58–64)	A39–(A58/B61)
14c*§	AGTVCWPAMG	1764.6	38–47		
	SSDCPRN		(55–61)	(58–64)	A39–(A58/B61)
27*	PITCKPKKG-OH	1749.0	10–18		
	PVTCKPR2	13–21	A13–B16		
28*	PITCKPKK-OH	1690.7	10–17		
	PVTCKPR			13–21	A13–B16

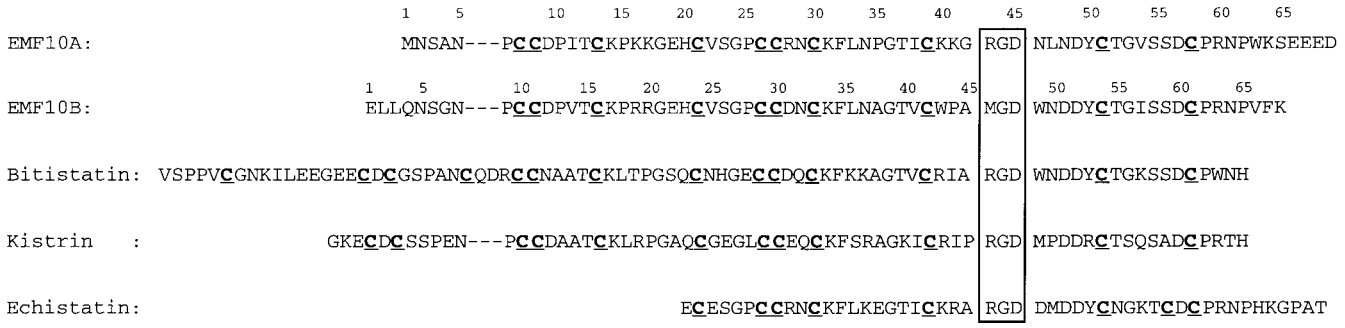

Figure 1 Alignment of the amino-acid sequences of EMF-10 A- and B-subunits with those of long (bitistatin), medium (kistrin) and small (echistatin) monomeric disintegrins

Cysteine residues are in bold and underlined, and the integrin-binding tripeptide sequences are boxed.

following the letters indicate the sequence of cleavage events. The charge status of the generated CID fragments, i.e. $(M+H)^+$, $(M+2H)^{2+}$, $(M+3H)^{3+}$, for the singly, doubly and triply charged quasimolecular ions, respectively, is also given. Thus '$\beta 7$ $(M+2H)^{2+} = 349.6$' means that the seventh fragmentation event involved a residue of chain β of the parent fragment and that this daughter ion is a doubly charged species of $m/z = 349.6$ (i.e. molecular mass of 697.2 Da).

Assignment of disulphide-bond combinations

All theoretical fragments containing two disulphide-bonded peptides were calculated with the program DISULPHIDE [24] using the experimental MALDI MS data as input. The primary outputs were filtered manually, taking into consideration information gathered from amino-acid analysis, N-terminal sequencing and CID MS.

Homology modelling of EMF-10

Three-dimensional models of EMF-10 were built on the basis of kistrin and flavoridin co-ordinates determined by NMR spectroscopy [8,9] (Protein Data Bank accession codes 1KST and 1FVL, respectively) using averaged conformers. In the respective models, residues 1–8 of kistrin or flavoridin were deleted and disulphide bonds between EMF-10A Cys-8 and EMF-10B Cys-11 and between EMF-10A Cys-13 and EMF-10B Cys-16 were created. The models were optimized for stereochemistry and refined by energy minimization using XPLOR [25]. The energy-minimized structures were heated at 1500 K and were further refined with a slow-cooling simulated-annealing molecular-dynamics protocol using XPLOR.

RESULTS AND DISCUSSION

Amino-acid sequence characteristics of oxalic acid fragments

To define the disulphide-bridge pattern of EMF-10, the protein was degraded with oxalic acid. Using this chemical cleavage method, Bauer and colleagues were able to confirm the four disulphide bonds of echistatin [22] and we established the disulphide bridge pattern of eristostatin (four disulphide bonds) [11] and bitistatin (seven disulphide bonds) [14]. Degradation with oxalic acid produces a mixture of fragments, which include incomplete series of disulphide-bonded peptides, such as (abCdef)-S–S-(ghiCjklm), (bCdef)-S–S-(ghiCjklm), (bCdef)-S–S-(iCjklm), (Cde)-S–S-(iCjk), etc. (where C denotes cysteine). Occurrence of dehydration has also been reported [22] and thus the mixture of peptides may include native fragments (M) along with dehydrated ($M-18$ Da) fragments (Table 1). Hence, reversed-phase HPLC separation of such fragment mixtures yields few pure peptide fractions and a broad asymmetric peak formed by the continuous distribution of multiple, partially resolved subpeaks. HPLC fractions were collected and initially characterized by MALDI–TOF MS. The isotope-averaged molecular masses of the largest possible linear peptides without cysteine residues were 1386.4 Da (EMF-10A, ^{59}PRNPWKSEEED^{69}) and 1369.6 (EMF-10B, ^{43}WPAMGDWNDDY^{53}). Thus ions with masses greater than these values must contain disulphide-bonded fragments. Amino-acid analysis of the reversed-phase HPLC fractions helped to identify the presence of particular cysteine residues. Thus single cystine-containing fragments whose amino-acid compositions show methionine and alanine may include Cys-7 of EMF-10A and/or Cys-42 or Cys-54 of EMF-10B. Similarly, tyrosine in a single cystine-containing fragment implies the presence of either Cys-42 (if Met and Ala are also present) or Cys-54 of EMF-10B, and/or Cys-39 or Cys-51 of EMF-10A. Phenylalanine in a single cystine-containing fragment indicates the presence of cysteine residues 30 or 39 of EMF-10A and/or 33, 42 or 61 of EMF-10B. Phenylalanine-containing fragments with Cys-39 (EMF-10A) or Cys-42 (EMF-10B) will also have leucine, threonine and isoleucine (EMF-10A), or leucine, alanine, threonine and valine (EMF-10B). Fragments with a single S–S bond and a histidine residue may contain Cys-13 or Cys-21 of EMF-10A and/or Cys-16 or Cys-24 of EMF-10B. Moreover, fragments including Cys-13 of EMF-10A or Cys-16 of EMF-10B will contain, in addition to histidine, glutamic acid and lysine, or glutamic acid, lysine and arginine, respectively. Except for valine, none of these residues will be present in a fragment including Cys-61 of EMF-10B. The presence of other residues such as isoleucine, which occurs twice in EMF-10A (positions 11 and 38) and only once in EMF-10B (position 57), is also very helpful, although impose less restrictions, for the identification of cysteine residues.

Assignment of disulphide bonds

Intrasubunit disulphide bridges between EMF-10A Cys-39–Cys-58 and EMF-10B Cys-42–Cys-61

Amino-acid and MALDI MS analyses of oxalic acid fragment Ox12* (where an asterisk indicates that the fragment was obtained after 12 h of degradation) indicated that its major ion

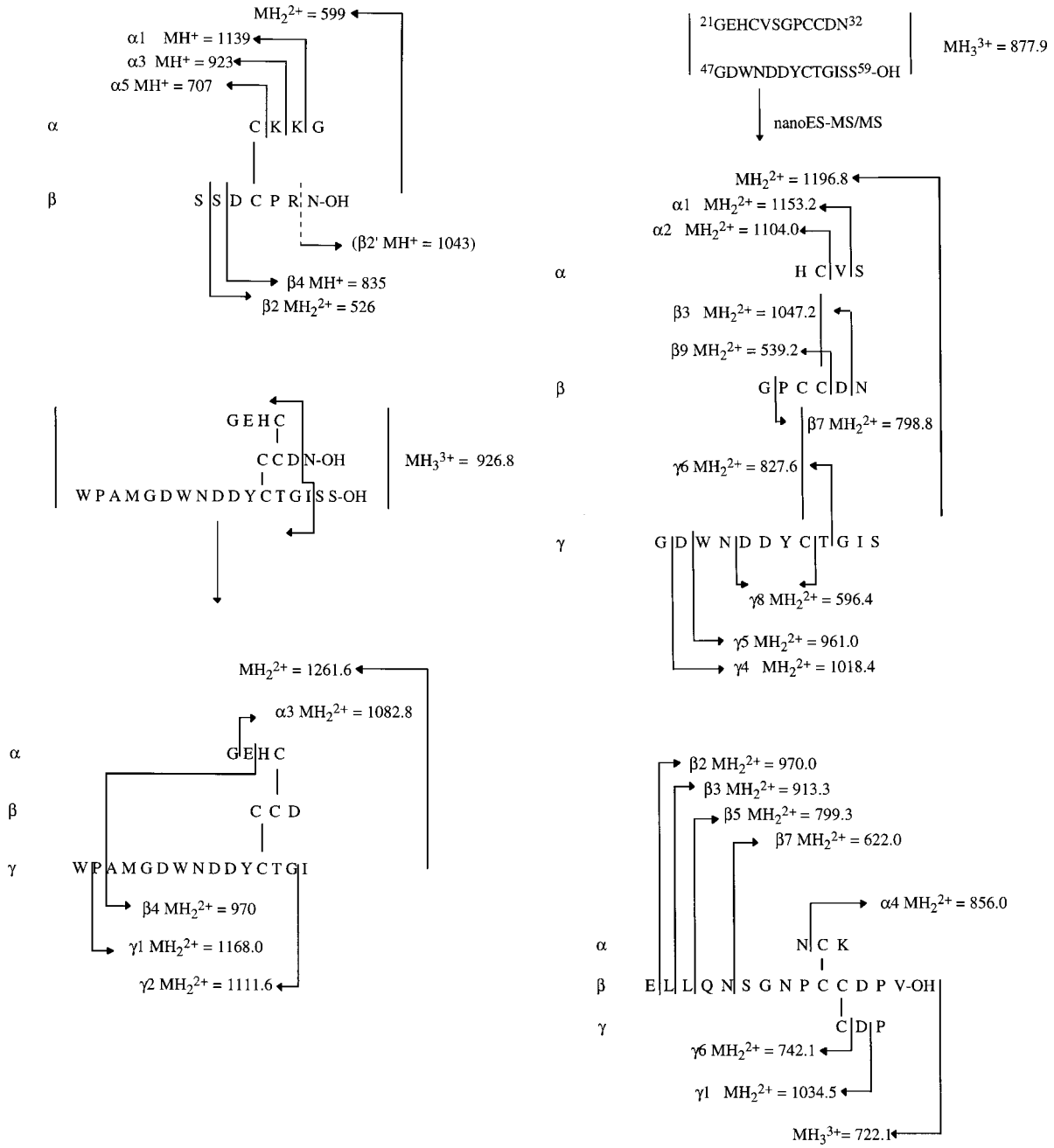

Figure 2 CID fragmentation pattern of fragments Ox14b* (top left), Ox24j (top right), Ox24f (bottom left) and Ox23 (bottom right)

Fragment ions are identified by a letter and a number indicating the parent chain (α, β, γ) from which they were generated and the cleavage order. MH^+, MH_2^{2+} and MH_3^{3+} indicate that the corresponding quasimolecular ions are singly, doubly and triply charged, respectively, and their m/z values are indicated.

of molecular mass (M) of 1222.4 Da (Table 1) corresponded to TICKKGR (EMF-10A 37–43) disulphide-bonded to SDCP. This latter tetrapeptide occurs in both subunits of EMF-10 (Figure 1), making it impossible to ascertain whether the two peptides were joined by an intra- or an intermolecular disulphide. Fragments 15 and 16 had the same amino-acid composition, which was (in nmol): 2.9 D, 2.6 P, 1.2 G, 1.9 C, 2.4 K and 2.1 R. Other amino acids contributed less than 10% to the chromatogram. However, MALDI MS of these fragments showed major ions in each fraction that had different molecular masses (Ox15, 1288.4 Da; Ox16, 1456.4 Da; Table 1). All possible EMF-10A and EMF-10B fragments disulphide-bonded by a single cystine residue were searched initially with the program DISULPHIDE [24] and the results were filtered with the compositional and mass data. Fragment 15 was assigned to peptides CKKGR (EMF-10 39–43) disulphide-bonded to DCPRNP. This latter peptide, which could correspond to EMF-10A 57–62 or to EMF-10B 60–65, was found also disulphide-bonded to EMF-10A ^{39}CKKGR^{43}, ^{39}CKKGRG^{44} and ^{39}CKKGRGD^{45} in ions of fraction 3* of molecular masses of 1291.9, 1349.9 and 1464.9 Da,

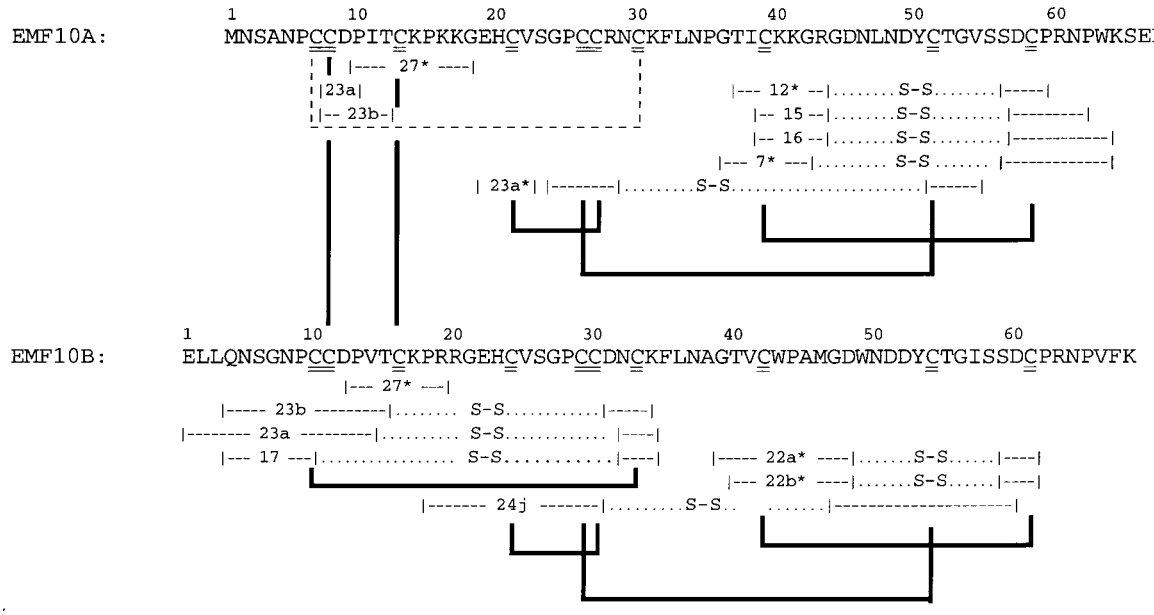

Figure 3 Disulphide-bond pattern of EMF-10

Representative fragments (see Table 1) that yielded structural information for establishing disulphide bonds are shown. Fragments containing the intrachain disulphide bond between EMF-10A cysteine residues Cys-7 and Cys-30 (broken line) were not characterized. All cysteine residues of EMF-10 subunits are double underlined.

respectively (Table 1). However, the CID spectrum of Ox14b* (Figure 2, top left panel) showed clearly that EMF-10A peptides ^{39}CKKG^{42} and dehydrated ^{55}SSDCPRN^{61} were disulphide-bonded (Table 1). In addition, fragment 16 was identified as CKKGR joined to EMF-10A ^{57}DCPRNPW^{63} minus a water molecule (Table 1). This indicated that dehydration took place during oxalic acid degradation (an observation that has been reported previously [22]) and showed clearly that Cys-39 and Cys-58 of EMF-10A were linked by an intramolecular disulphide bond (Figure 3). Fragment 6* (1290.7 Da), which was characterized by CID MS as a double-dehydrated fragment composed of EMF-10A ^{34}NPGTICKKG^{42} disulphide-bonded to SSDC, and fragment 7* (1618.9 Da), which displayed an amino-acid composition corresponding to disulphide-bonded peptides of EMF-10A ^{36}GTICKK41 and ^{56}SDCPRNPW^{63}, confirmed this assignment (Table 1).

Fragment 22 displayed two major MALDI ions of molecular masses 1299.2 and 1356.2 Da. These mass values and the amino-acid composition of fraction 22 were only compatible with fragment SDC (residues 56–58 of EMF-10A or 59–61 of EMF-10B) disulphide-bridged to peptides ^{40}TVCWPAMGD^{48} and ^{39}GTVCWPAMGD^{48} of EMF-10B, respectively. Furthermore, CID of the doubly charged ion with m/z 883.3 found in fragment 14c* identified it as the covalently linked peptides EMF-10B ^{38}AGTVCWPAMG^{47} and SSDCPRN (Table 1). Although it was not possible to discern between the intra- and intermolecular-bond possibilities, the fact that EMF-10A Cys-58 was involved in an intrasubunit disulphide bridge strongly argued for Ox22 containing an intra-EMF-10B disulphide bond (Figure 1). Characterization of fraction Ox5*, which contained three major ions that could be assigned to EMF-10B ^{60}DCPRNPV^{66} disulphide-bonded to ^{39}GTVC^{42} (1175.7 Da), ^{38}AGTVC^{42} (1246.8 Da) and EMF-10B ^{60}DCPRNPVF^{67}-OH disulphide-bridged to ^{39}GTVC^{42} (Table 1), was in line with this hypothesis. A disulphide bond between homologous cysteine residues is conserved in all monomeric disintegrin molecules characterized to date [5,6].

Cysteine residues 21, 26, 27 and 51 of EMF-10A (24, 29, 30 and 54 of EMF-10B) are linked by two intrachain disulphide bonds in each EMF-10 subunit

Amino-acid analysis showed that fraction Ox23* contained (in nmol) 4.6 D, 2.1 T, 2.4 S, 1.4 E, 3.9 P, 4.3 G, 4.4 C, 2.3 V, 1.3 L, 0.8 Y, 0.9 H, 2.9 K and 2.1 R and yielded major ions of masses 1938.4, 1394.8, 1294.8, 1195.7 and 1179.7 Da. A search for fragments containing a single disulphide bridge matching with the amino-acid analysis was unsuccessful. However, using the program DISULPHIDE, a collection of structurally related fragments composed of three peptides of EMF-10A and/or EMF-10B joined by two disulphide bonds were identified that agreed well with both the masses and the amino-acid compositions. These peptides were: (CVS)(GPCC)(EMF-10A ^{44}G-DNLNDYCTGVS^{55}), M_{calc} 1939 Da; (EHCV)(EMF-10A ^{24}GPCCR^{28})(EMF-10A ^{51}CTGV^{54}), M_{calc} 1395 Da; (EMF-10A ^{16}KKGEHC^{21})(EMF-10A ^{25}PCCR^{28})(C), M_{calc} 1295 Da; (EMF-10A ^{16}KKGEHC^{21})(GPCC)(C), M_{calc} 1196 Da; and (CVS)(EMF-10A ^{25}PCCR^{28})(DYC), M_{calc} 1180 Da (where sequences in parentheses represent peptide fragments joined by disulphide bonds, and those segments unambiguously assigned to positions in EMF-10A are shown as such). Sequences EHCV, GPCC and DYC occur in both EMF-10 chains, and it was therefore not possible to discern whether these structures contained only intrachain disulphide bonds or both intra- and interchain cystine residues. Fragment Ox21* showed a major ion at m/z 1236.7. Five cycles of N-terminal sequence analysis yielded (P+Y), V, (S+T), G and I. These data identified Ox21* as peptides CVSG, PCC and EMF-10B ^{53}YCTGI^{57} linked together by two disulphide bonds. The exact pairing of the four cysteine residues could not be determined, however.

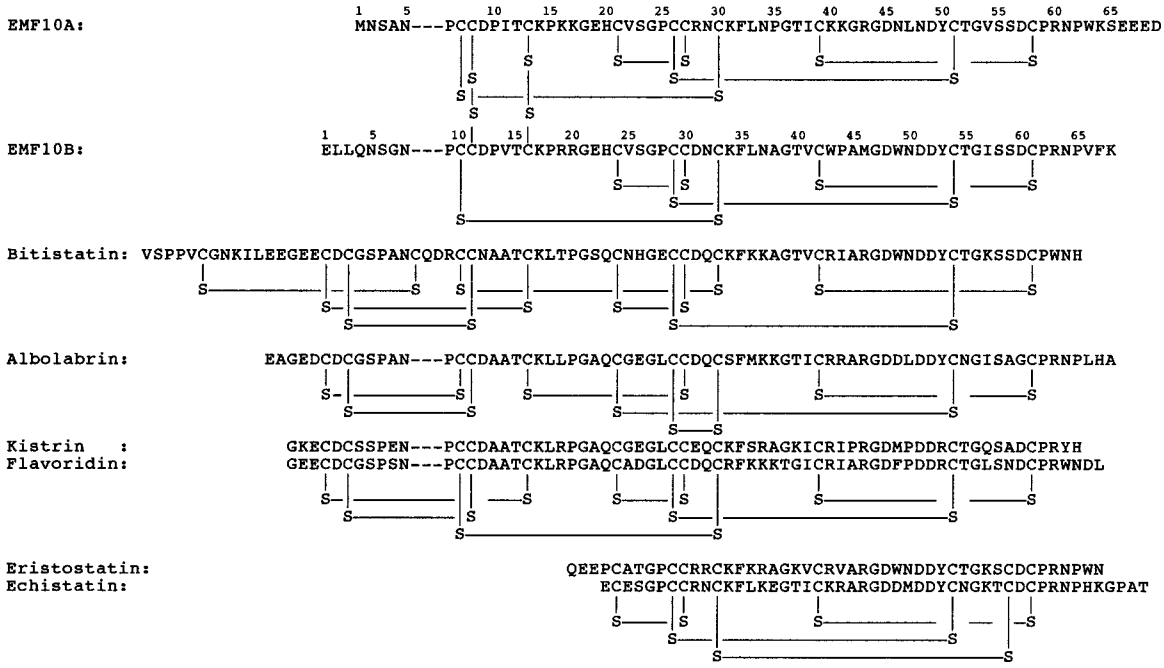

Figure 4 Comparison of the amino-acid sequences and disulphide bonds of dimeric disintegrin EMF-10 and monomeric disintegrins

Structural studies have shown that disintegrins exhibit three different patterns of disulphide bridge: S–S(I), albolabrin [5]; S–S(II), bitistatin, kistrin and flavoridin [14,15]; and S–S(III), eristostatin and echistatin [11,15]. The intrachain disulphide bonds of each EMF-10 subunit correspond to the three disulphide bonds shared by patterns S–S(II) and S–S(III).

MALDI MS analysis of fraction Ox24 showed multiple ions. The sample was analysed by nanoES QIT MS and the major ions were subjected to CID. The CID spectrum of the triply charged ion of Ox24j [$(M+3H)^{3+} = 877.9$ Da] provided the necessary structural information to assign it to GEHCVSGPCCDN (EMF-10B 21–32) disulphide-bonded to dehydrated GDWNDDYC-TGISS (EMF-10B 47–59) and containing an intrachain disulphide (Figure 2, top right panel). This result clearly indicated that cysteine residues 21, 26, 27 and 51 of EMF-10A and the homologous cysteine residues at positions 24, 29, 30 and 54 of EMF-10B were linked by two intramolecular disulphide bonds in each EMF-10 subunit. In addition, the fragmentation pattern of Ox24f yielded three peptides joined by two cystine residues (Figure 2, bottom left panel). From this structure it could be deduced that the linkages between Cys-24–Cys-54 and Cys-29–Cys-30 of EMF-10A are not allowed. Except for albolabrin, the pattern corresponding to EMF-10A Cys-21–Cys-27, Cys-26–Cys-51 (or EMF-10B Cys-24–Cys-30, Cys-29–Cys-54) is conserved in all EMF-10 homologous monomeric disintegrins [5,6]. We therefore propose (Figure 3) that the same arrangement will be, most probably, present in each EMF-10 subunit.

Intrachain disulphide bond EMF-10B Cys-10–Cys-33 and intersubunit linkage between EMF-10A Cys-8 and EMF-10B Cys-11

Oxalic acid fragments Ox17 and Ox18 had amino-acid compositions and MALDI masses (1081.1 and 1322.8 Da) only compatible with peptides EMF-10B ^{4}QNSGNPC^{10} and ^{1}ELLQNSGNPC^{10} disulphide-bonded to the tripeptide NCK and the dipeptide CK, respectively. Similarly, the amino-acid composition and molecular mass (1272.8 Da) of fragment 11* corresponded to disulphide-linked, dehydrated peptides EMF-10B ^{1}ELLQNSGNPC^{10} and NC (Table 1). The sequence NCK occurs in both subunits of EMF-10, restricting the disulphide-bonding possibilities of EMF-10B Cys-10, but not allowing its unambiguous intra- or interchain assignment. However, CID analysis by nanoES QIT MS of the two major ions contained in fraction Ox23 (Ox23a, 2163.4 Da; Ox23b, 2220.6 Da) showed that they corresponded, respectively, to dehydrated EMF-10B peptide ^{1}ELLQNSGNPCCDPVT^{15} disulphide-bonded to NCK and CDP (Figure 2, bottom right panel) and to its related fragment EMF-10B ^{4}QNSGNPCCDPVT^{15} disulphide-bonded to EMF-10B peptides ^{31}DNCK^{34} and EMF-10A ^{8}CDPIT^{13} minus two water molecules (Table 1). These results, along with those obtained for Ox17 and Ox18, clearly showed that EMF-10B cysteine residues 10 and 33 were linked by an intrachain disulphide bond and indicated the existence of an intersubunit cystine residue involving EMF-10A Cys-8 and its homologous residue EMF-10B Cys-11 (Table 1, Figure 3).

EMF-10A Cys-13 and EMF-10B Cys-16 form the second interchain cystine residue

Fraction Ox27* exhibited a major MALDI ion with m/z 1749.0. N-terminal sequence analysis of this fragment yielded P, (V + I), T, X, K, P and (K + R). These data identified fragment Ox27* as EMF-10A ^{10}PITCKPKKG^{18} disulphide-bonded to EMF-10B ^{13}PVTCKPR^{21} (Table 1). The calculated isotope-averaged molecular mass was 1767.5 Da, indicating that fragment Ox27* contained a dehydrated peptide. Since EMF-10B Arg-21 was recovered in the sequence analysis, we suspect that EMF-10A Gly-18 may be the dehydrated residue. Fraction Ox29* contained a single fragment with a molecular mass of 1690.7 Da and the same N-terminal sequences as Ox27*. The mass difference

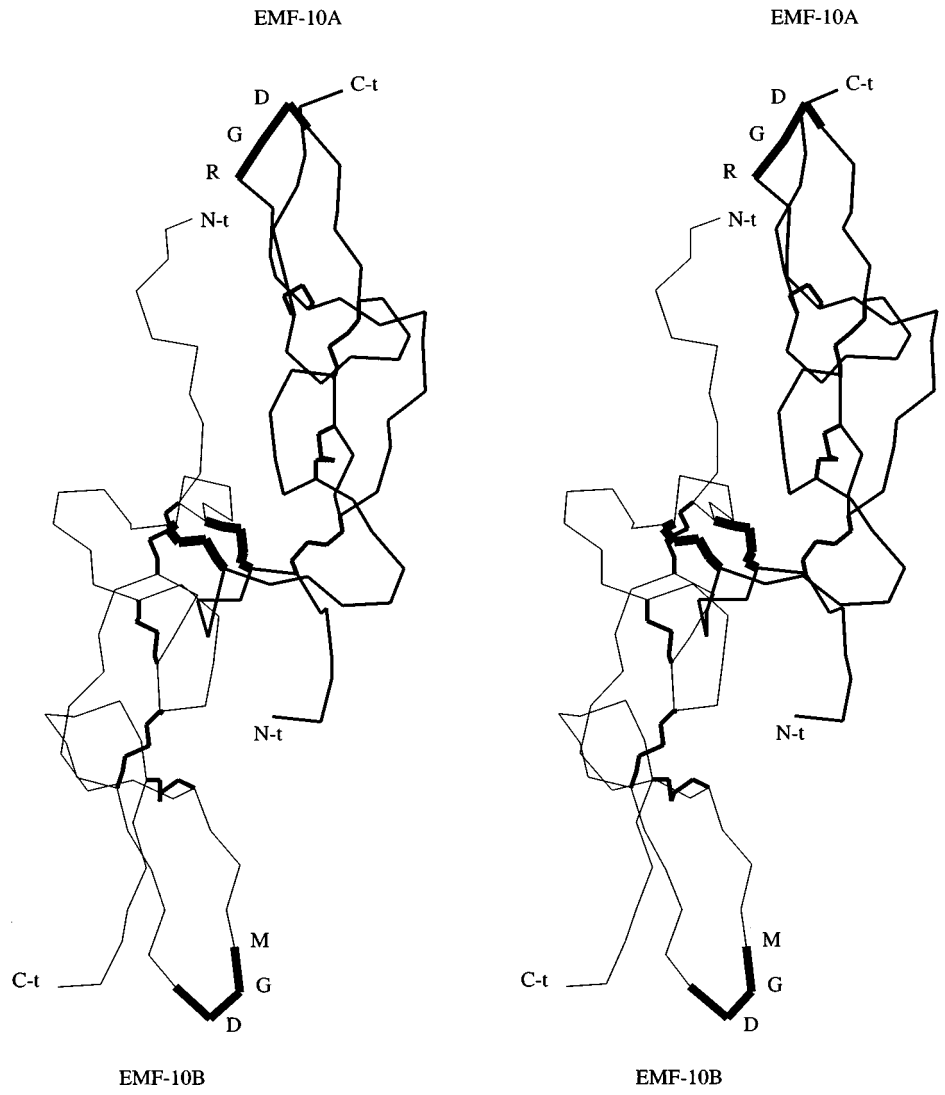

Figure 5 Stereo drawing of the Cα trace of a molecular model of EMF-10 based on flavoridin NMR solution structure

EMF-10 subunits and their integrin-binding tripeptides (RGD in EMF-10A and MGD in EMF-10B) are labelled. The two intersubunit disulphide bonds [(A)Cys-8–(B)Cys-11 and (A)Cys-13–(B)Cys16] are depicted as thick lines. N-t, N-terminus; C-t, C-terminus.

indicated that Ox27* may correspond to EMF-10A 10–17 disulphide-bonded to EMF-10B 13–21 (minus a water molecule). These fragments showed the existence of a second interchain cystine residue composed of the third cysteine residue of each EMF-10 subunit (Figure 3).

The disulphide-bond pattern of EMF-10: a molecular model

The results discussed above led to the identification of nine out of the ten disulphide bonds of EMF-10. The missing cystine residue, corresponding to the intrachain disulphide bond between cysteine residues at amino-acid positions 7 and 30 of EMF-10A, was not found, perhaps because of extensive degradation during the oxalic acid treatment of the dimeric disintegrin. Its existence in the native protein must be inferred, however, because EMF-10 does not contain free thiol groups. A comparison of the disulphide-bond patterns of EMF-10 and representative small (eristostatin [11] and echistatin [5,15]), medium (kistrin [8], flavoridin [9,15] and albolabrin [6,10]), and long (bitistatin [14]) monomeric disintegrins is depicted in Figure 4. Both subunits of EMF-10 display the same intrachain disulphide bonds, which are also conserved in the structures of all other disintegrins, albolabrin being the exception. The second and third cysteine residues of each EMF-10 subunit form two interchain disulphide bridges and are thus responsible for heterodimer formation. The equivalent cysteines of monomeric disintegrins are involved in intrachain disulphide bonds with the two cysteine residues included in the N-terminal polypeptide, which is missing in EMF-10 chains. The NMR solution structures of kistrin [8] and flavoridin [9] have shown that the four disulphide links that are conserved in EMF-10 (Cys-3–Cys-9, Cys-6–Cys-8, Cys-7–Cys-11 and Cys-10–Cys-12; Figure 4) are partially or totally buried in the conformation of these monomeric disintegrins. On the other hand, the two disulphide bridges Cys-1–Cys-5 and Cys-2–Cys-4 (Figure 4) are solvent-exposed on the same side of the molecule and their Sγ atoms point to the same direction. This may be the reason why this particular pair of cysteine residues is involved in

heterodimerization of the N-terminally truncated disintegrin molecules.

Since EMF-10 subunits share cystine linkages with kistrin and flavoridin, the known NMR structures of these disintegrins [8,9] provide templates for building a molecular model of EMF-10. Attempts to model the structure of EMF-10 using an averaged conformer of the kistrin NMR co-ordinates (Protein Data Bank accession code 1KST) were unsuccessful because the structural constraints imposed by the two interchain disulphide bonds forced both subunits to clash. It was possible, however, to model the EMF-10 heterodimer (Figure 5) based on the averaged flavoridin NMR conformer (Protein Data Bank accession code 1FVL). The distances between the Cα atoms of Cys-8 and Cys-13 of EMF-10A (or Cys-11 and Cys-16 of EMF-10B) is 5.4 Å in the kistrin model and 7.4 Å in the flavoridin model. In the first possibility, the unpaired cysteines of each subunit could form an intramolecular bond, whereas in the flavoridin model the two cysteines are too far apart to form a linkage. This can potentially provide a structural reason for why no monomeric chains from EMF-10 [20] or from other dimeric disintegrins, i.e. EC3 [19] and contortrostatin [21], have been reported. According to the flavoridin-based model, the RGD and the MGD active loops of the EMF-10 subunits would occupy opposite locations (180°) at the ends of an elongated disulphide-bond ladder (Figure 5). The distance between Asp-45 of EMF-10A and Asp-48 of EMF-10B is 71.4 Å. Though the N- and C-terminal polypeptides and the loop harbouring the integrin-binding sequence of monomeric RGD-containing disintegrins have been shown to exhibit conformational flexibility, the proposed conformation of EMF-10 does not allow for relative movements of the subunits. Another feature of the EMF-10 model is the location of the N-terminal polypeptide of EMF-10B (^{2}LLQN^{5}) close to the ^{40}KKG^{42} sequence within the RGD-loop of the EMF-10A subunit (Figure 5). This arrangement could potentially influence the RGD-mediated binding activities of the dimeric disintegrin. Noteworthy is the fact that a number of disintegrins containing the RGDN motif are very potent inhibitors of $\alpha_v\beta_3$ integrin [26–28]. On the contrary, although EMF-10A displays the RGDN sequence, the inhibitory activity of EMF-10 on $\alpha_v\beta_3$ integrin is minimal [20]. It has been suggested that the EMF-10A active sequence is sterically hindered by heterodimerization with EMF-10B [20]. However, it must be stressed that the polypeptide backbone extending from residue 1 to 26 of flavoridin is very flexible and unstructured in the solution conformation of the 18 energy-refined conformers [9], and its spatial orientation relative to the RGD loop is not well defined. Thus addressing structure–function correlations based on the homology-modelled EMF-10 molecule awaits experimental backup.

This work was supported by grants PB95-0077 from the Dirección General de Investigación Científica y Técnica, (Madrid, Spain, to J.J.C.), American Diabetes Association (to S.N.), Barra Foundation (S.N.) and American Heart Association Initial Investigatorship (C.M.). We thank Roy R. Lobb (Biogen, Cambridge, MA, U.S.A.) for continuous support.

Received 1 July 1999/30 September 1999; accepted 19 November 1999

REFERENCES

1 McLane, M. A., Marcinkiewicz, C., Vijay-Kumar, S., Wierzbicka-Patynowski, I. and Niewiarowski, S. (1998) Proc. Soc. Exp. Biol. Med. **219**, 109–119
2 Calvete, J. J. (1997) in Integrin-Ligand Interactions (Eble, J. and Kühn, K., eds.), pp. 157–173, Springer, Heidelberg
3 Huang, T.-F., Holt, J. C., Lukasiewicz, H. and Niewiarowski, S. (1987) J. Biol. Chem. **262**, 16157–16163
4 Gan, Z.-R., Gould, R. J., Jacobs, J. W., Friedman, P. A. and Polokoff, M. A. (1988) J. Biol. Chem. **263**, 19827–19832
5 Calvete, J. J., Schäfer, W., Soszka, T., Lu, W. Q., Cook, J. J., Jameson, J. J. and Niewiarowski, S. (1991) Biochemistry **30**, 5225–5229
6 Dennis, M. S., Carter, P. and Lazarus, R. A. (1993) Proteins Struct. Funct. Genet. **15**, 312–321
7 Saudek, V., Atkinson, R. A. and Pelton, J. T. (1991) Biochemistry **30**, 7369–7372
8 Adler, M., Lazarus, R. A., Dennis, M. S. and Wagner, G. (1991) Science **253**, 445–448
9 Senn, H. and Klaus, W. (1993) J. Mol. Biol. **232**, 907–925
10 Smith, K. J., Jaseja, M., Lu, X., Williams, J. A., Hyde, E. I. and Trayer, I. P. (1996) Int. J. Pept. Protein Res. **48**, 220–228
11 McLane, M. A., Vijay-Kumar, S., Marcinkiewicz, C., Calvete, J. J. and Niewiarowski, S. (1996) FEBS Lett. **391**, 139–143
12 Marcinkiewicz, C., Vijay-Kumar, S., McLane, M. A. and Niewiarowski, S. (1997) Blood **90**, 1565–1575
13 Rahman, S., Aitken, A., Flynn, G., Formstone, C. and Savidge, G. F. (1998) Biochem. J. **335**, 247–257
14 Calvete, J. J., Schrader, M., Raida, M., McLane, M. A., Romero, A. and Niewiarowski, S. (1997) FEBS Lett. **416**, 197–202
15 Calvete, J. J., Wang, Y., Mann, K., Schäfer, W., Niewiarowski, S. and Stewart, G. J. (1992) FEBS Lett. **309**, 316–320
16 Huang, T.-F., Peng, H. C., Peng, I. S., Teng, C. M. and Ouyang, C. (1992) Arch. Biochem. Biophys. **298**, 13–20
17 Marrakchi, N., Barbouche, R., Bon, C. and el-Ayeb, M. (1997) Toxicon **35**, 125–135
18 Trikha, M., De Clerk, Y. A. and Markland, F. S. (1994) Cancer Res. **54**, 4993–4998
19 Marcinkiewicz, C., Calvete, J. J., Marcinkiewicz, M. M., Raida, M., Vijay-Kumar, S., Huang, Z., Lobb, R. R. and Niewiarowski, S. (1999) J. Biol. Chem. **274**, 12468–12473
20 Marcinkiewicz, C., Calvete, J. J., Vijay-Kumar, S., Marcinkiewicz, M. M., Raida, M., Schick, P., Lobb, R. R. and Niewiarowski, S. (1999) Biochemistry **38**, 13302–13309
21 Markland, F. S. (1998) Toxicon **36**, 1749–1800
22 Bauer, M., Sun, Y., Degenhardt, C. and Kozikowski, B. (1993) J. Protein Chem. **12**, 759–764
23 Biemann, K. (1988) Biomed. Environ. Mass Spectrom. **16**, 99–111
24 Caporale, C., Sepe, C., Caruso, C., Pucci, P. and Buonocore, V. (1996) FEBS Lett. **393**, 241–247
25 Brünger, A. T. (1992) X-PLOR version 3.1, a System for X-ray Crystallography and NMR, Yale University Press, New Haven
26 Juliano, D., Wang, Y., Marcinkiewicz, C., Rosenthal, L. A., Stewart, G. J. and Niewiarowski, S. (1996) Exp. Cell Res. **225**, 2482–2487
27 Pfaff, M., McLane, M. A., Beviglia, L., Niewiarowski, S. and Timpl, R. (1994) Cell Adhes. Commun. **2**, 491–500
28 Scarborough, R. M., Rose, J. W., Naughton, M. A., Phillips, D. R., Nannizzi, L., Arfsten, A., Campbell, A. M. and Charo, I. F. (1993) J. Biol. Chem. **268**, 1058–1065