Recombinant decorsin: Dynamics of the RGD recognition site

ANDRZEJ M. KREZEL,¹ JANA S. ULMER,² GERHARD WAGNER,³ AND ROBERT A. LAZARUS²

¹Department of Molecular Biology, Vanderbilt University, Box 1820, Station B, Nashville, Tennessee 37235

²Department of Protein Engineering, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080

³Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115

(RECEIVED January 31, 2000; FINAL REVISION May 13, 2000; ACCEPTED May 26, 2000)

Abstract

Decorsin is an antagonist of integrin $\alpha_{IIb}\beta_3$ and a potent platelet aggregation inhibitor. A synthetic gene encoding decorsin, originally isolated from the leech *Macrobdella decora*, was designed, constructed, and expressed in *Escherichia coli*. The synthetic gene was fused to the stII signal sequence and expressed under the transcriptional control of the *E. coli* alkaline phosphatase promoter. The protein was purified by size-exclusion filtration of the periplasmic contents followed by reversed-phase high-performance liquid chromatography. Purified recombinant decorsin was found to be indistinguishable from leech-derived decorsin based on amino acid composition, mass spectral analysis, and biological activity assays. Complete sequential assignments of ¹H and proton bound ¹³C resonances were established. Stereospecific assignments of 21 of 25 nondegenerate β -methylene groups were determined. The RGD adhesion site recognized by integrin receptors was found at the apex of a most exposed hairpin loop. The dynamic behavior of decorsin was analyzed using several independent NMR parameters. Although the loop containing the RGD sequence is the most flexible one in decorsin, the conformation of the RGD site itself is more restricted than in other proteins with similar activities.

Keywords: dynamics; leech protein; NMR; platelet aggregation inhibitor; resonance assignments, RGD

Leeches are known to contain a variety of proteins in their salivary secretions that can affect the hemostatic process (Sawyer, 1991; Arocha-Pinango et al., 1999). Many of these have been characterized as relatively small disulfide-rich proteins that inhibit platelet aggregation or serine proteases such as thrombin or Factor Xa, which play a key role in the coagulation cascade. Decorsin, a 39-residue protein originally isolated from the leech *Macrobdella decora*, is a potent integrin $\alpha_{IIb}\beta_3$ antagonist and inhibitor of platelet aggregation (Seymour et al., 1990). It contains six cysteine residues as well as an RGD adhesion site-recognition sequence (Ruoslahti & Pierschbacher, 1987), found in adhesive proteins

such as fibrinogen (Fg), vitronectin, and von Willebrand factor. These proteins bind to $\alpha_{IIb}\beta_3$ (Phillips et al., 1988, 1991), a calcium-dependent heterodimer belonging to the integrin family of receptors that mediate cell–cell and cell–matrix interactions (Giancotti & Ruoslahti, 1999). Because platelet aggregation, which is mediated by Fg binding to $\alpha_{IIb}\beta_3$, is a key element in thrombosis, antagonists of the $\alpha_{IIb}\beta_3$ receptor are excellent therapeutic candidates for the treatment of thrombotic diseases (Coller, 1992).

Amino acid sequence analysis of decorsin indicated that it is related to the ornatin, similar $\alpha_{IIb}\beta_3$ antagonists found in the leech *Placobdella ornata* (Mazur et al., 1991). However, structural analysis of decorsin revealed that it shared the same tertiary fold as hirudin, a potent thrombin inhibitor from medicinal leech *Hirudo medicinalis* (Krezel et al., 1994). Further sequence analysis suggested that antistasin, a Factor Xa inhibitor from the Mexican leech *Haementeria officinalis* (Nutt et al., 1988), might also be structurally related. Because these leech proteins all disrupt processes leading to blood clot formation, albeit by different mechanisms, and share structural elements, they may have evolved from a common ancestral gene.

The sequence-specific assignments and NMR structures of the disintegrins kistrin (Adler et al., 1991; Adler & Wagner, 1992), flavoridin (Senn & Klaus, 1993), echistatin (Chen et al., 1991; Cooke et al., 1991; Dalvit et al., 1991; Saudek et al., 1991), and albolabrin (Jaseja et al., 1993; Smith et al., 1996) have been re-

Reprint requests to: Andrzej M. Krezel, Vanderbilt University, Department of Molecular Biology, Box 1820, Station B, Nashville, Tennessee 37235; e-mail: Andrzej.M.Krezel@vanderbilt.edu.

Abbreviations: 2D, two-dimensional; 3D, three-dimensional; $\alpha_{IIb}\beta_3$, integrin also known as glycoprotein IIb-IIIa; δ , chemical shift in ppm; CSI, chemical shift index; COSY, correlated spectroscopy; DQF, double quantum filtered; EDTA, ethylenediaminetetraacetic acid; ELISA, enzyme-linked immunosorbent assay; FAB-MS, fast-atom bombardment mass spectrometry; Fg, fibrinogen; HPLC, high-performance liquid chromatography; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy; PMSF, phenylmethyl sulfonyl fluoride; ppb, parts per million; RMSD, root-mean-square deviation; SDS, sodium dodecyl sulfate; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy; [d₄]TSP, sodium 3-(trimethylsilyl) propionate-2,2,3,3,-²H.

ported. These proteins show little regular secondary structure. They display the RGD binding epitope at the apex of an extended loop that is the least well-defined and most flexible part of the structure. Reduction of the disulfide bonds of these proteins greatly diminishes their activity, and mutational analysis of kistrin has demonstrated that only the RGD epitope is required for high affinity binding to $\alpha_{IIb}\beta_3$; no evidence for any secondary binding sites has been found (Dennis et al., 1993). Mambin, also known as dendroaspin, is also a potent $\alpha_{IIb}\beta_3$ antagonist that contains an RGD sequence at the apex of an extended loop (McDowell et al., 1992). However, mambin is from an elapid venom and has little sequence similarity to the disintegrins, but is structurally similar to the short chain neurotoxins (Jaseja et al., 1994). It therefore appears that many different scaffolds can present the RGD epitope, which is critical for binding to integrin receptors. The small amount of decorsin obtained by purification from the leech *M. decora* (Seymour et al., 1990) previously limited the extent of structural, functional, and biological characterization that could be addressed. Recently, decorsin was also synthesized using solid-state peptide synthetic methods (Polverino de Lauro et al., 1998). Here, we describe the expression, purification, and characterization of recombinant decorsin from *E. coli*, as well as the ^1H and ^{13}C resonance assignments and the dynamics of decorsin tertiary structure.

Results

Gene synthesis and phagemid construction

A decorsin gene was synthesized from six oligonucleotides that were phosphorylated and ligated into the 4.1 kb *NarI*/*StyI* frag-

ment of pZKis (Dennis et al., 1993). This construct encoded decorsin as a fusion protein with the Z domain of protein A and with the sequence IEGR inserted as a Factor Xa cleavable linker (Fig. 1A). A second construct was made in which the Z domain and linker regions were removed by deletion mutagenesis. The resultant direct expression phagemid, pDecorsin (Fig. 1B), placed transcription of the decorsin gene under the control of the alkaline phosphatase promoter (*phoA*), which induces high expression levels under phosphate limiting culture conditions (Chang et al., 1987). The presence of the *stII* leader sequence directed secretion of decorsin into the periplasmic space.

Protein expression and purification

Recombinant decorsin was expressed in *Escherichia coli* at ~700 mg/L in a 10 L fermentor. Based on $\text{Fg}/\alpha_{IIb}\beta_3$ ELISA titers, about half of the decorsin was found in the periplasmic space and half was present in the media. No refractile bodies were noted upon microscopic inspection of the cells.

Recombinant decorsin was purified by molecular weight exclusion filtration of the periplasmic contents and C_{18} reversed-phase chromatography. Following the penultimate HPLC purification step, two peaks of inhibitory activity with slightly different HPLC retention times were noted; these peaks are referred to as rDecorsin (recombinant), which eluted at ~22.5% acetonitrile, and mDecorsin (misfolded), which eluted at ~25% acetonitrile. Both of these peaks were rechromatographed and characterized. From 402 g of cell paste, 358 mg of rDecorsin was isolated. Based on HPLC peak height, an approximately equal amount of mDecorsin was recovered.

Fig. 1. A: Construction of the synthetic decorsin gene. The junctions between synthetic oligonucleotides are indicated by brackets (]). These oligos share a unique four-base overlap with neighboring oligonucleotides and the 5' and 3' termini were designed to accommodate *NarI* and *StyI* sticky ends, respectively. The Factor Xa linker region is underlined; the stop codon is marked with an asterisk. Selected restriction sites are noted. The RGD recognition sequence is boxed. **B:** pDecorsin expression phagemid. The annealed oligonucleotides were ligated into the 4.1 kb *NarI*/*StyI* fragment of pZKis, which encodes the alkaline phosphatase promoter, *stII* signal sequence, one synthetic Z domain, the ampicillin resistance gene, and pMB1 and f1 origins of replication. Upon ligation, the *StyI* site was destroyed due to the A/T base pair immediately upstream from the *StyI* sticky end of the synthetic insert. Deletion mutagenesis of pZDecorsin as described in Materials and methods removed the Z domain and the Factor Xa linker region, and resulted in pDecorsin.

Amino acid analysis of each peak resulted in compositions equivalent to authentic decorsin (Table S2 in Supplementary material in the Electronic Appendix). FAB-MS analysis of each peak revealed molecular masses for rDecorsin and mDecorsin to be within 1 amu of the mass calculated from the decorsin sequence (4,377.9) assuming three disulfide bonds are present (Table 1). Mass spectra obtained following reduction of both peaks with thioglycerol revealed molecular masses (Table 1) consistent with that calculated for reduced decorsin (4,383.9); no evidence of any proteolytic fragments was found. Furthermore, when aliquots of both peaks were reduced and subjected to C₁₈ reversed-phase analytical HPLC, they coeluted. Fg/ $\alpha_{1\text{Ib}}\beta_3$ ELISA activity measurements on both peaks showed that rDecorsin and mDecorsin exhibited IC₅₀ values of ~3.5 and 30 nM, respectively. Finally, platelet aggregation assays revealed that rDecorsin inhibited platelet aggregation with an IC₅₀ of 270 nM, while mDecorsin inhibited with an IC₅₀ of 12 μ M (Table 1). The specific activities for rDecorsin are consistent with those obtained with material purified from leeches (Seymour et al., 1990).

Resonance assignments

Sequential assignments of proton resonances were derived from 2D ¹H NMR experiments using established methodology (Wagner & Wüthrich, 1982; Wüthrich, 1986). Sufficient dispersion of chemical shifts and signal-to-noise ratio allowed identification of virtually all expected TOCSY cross peaks in 2D spectra. Spin systems corresponding to different amino acid residues of decorsin were identified in a TOCSY spectrum recorded in water. DQF-COSY experiments were used to verify spin system assignments and to assign side-chain methylene protons. Sequential assignments were completed using the NOESY spectra, which were recorded with mixing times ranging from 50 to 200 ms. Ambiguous and overlapped peaks were resolved with the help of NOESY and TOCSY spectra recorded at 18, 20, 23, 27, and 30 °C.

Assignment of proton-bearing ¹³C atoms was based on natural abundance ¹H-¹³C HSQC spectra recorded at 23 and 30 °C. Carbon resonances were correlated to the previously assigned ¹H resonances. Figure 2 summarizes the observed sequential connectivities as well as the chemical-shift index (CSI) values for ¹H α and ¹³C α resonances (Wishart & Sykes, 1994a). The largest observed CSI values of 0.9 ppm for ¹H α and -5.9 ppm for ¹³C α were found for residue C22.

The estimates of H α -H β coupling constants for almost all observed nondegenerate β -methylene groups are listed in Figure 2. These H α -H β and H^N-H α vicinal coupling constants combined with the relative intensities of H^N-H β and H α -H β intraresidue NOE cross peaks were used for the stereospecific assignments of

diastereotopic β -methylene protons. This procedure was conducted manually for the majority of the assignments.

Amide proton exchange and chemical shift temperature dependence

Seventeen of 32 amide resonances present in decorsin could be seen in TOCSY experiments recorded immediately after dissolving lyophilized protein in 100% D₂O at pH 5.5, 27 °C. For the most protected amide protons, an exponential fit of the exchange rates (k_{ex}) produced values on the order of 10⁻³ min⁻¹ (L16, C22, Q26, R28, Y37, and E39) (Table 2). The amide protons of D11, E13, C15, C17, D20, C27, F29, and C38 had k_{ex} on the order of 10⁻² min⁻¹. The amide protons of Q6, C7, Q8, G9, K14, G25, A34, and D35 produced k_{ex} on the order of 10⁻¹ min⁻¹. The protection of amide protons is represented graphically in Figure 4 by the size of the sphere surrounding them. The larger radius corresponds to slower exchange rate.

Temperature coefficients of chemical shifts for most of the aliphatic resonances were between -1 and 0 ppb/°C. Exchangeable protons showed considerably larger and more variable temperature coefficients. Amide protons with temperature coefficients between -4.5 and 0 ppb/°C are also indicated as spheres around amide nitrogen atom in Figure 4.

Disulfide bonds

Disulfide bonds as shown in Figures 2 and 4 were identified from H_i β -H_j β and H_i β -H_j α NOE cross peaks observed in the 50 ms mixing time experiment. These cross peaks were clearly visible between residues 22 and 38 and between residues 7 and 15. Residue 7 also showed crosspeaks to residue 27. The unique disulfide-bonding pattern was confirmed by structure calculations. Calculations without explicit disulfides positioned the side chains of potentially bonded residues. Addition of explicit disulfide bonds (C7-C15, C17-C27, and C22-C38) did not increase violations of other distance constraints. On the other hand, addition of other combinations of disulfides led to the significant violations of other distance constraints.

Secondary and tertiary structure

Complete resonance assignments allowed identification of all diagnostic cross peaks in NOESY spectra. Figure 2 shows the summary of those NOEs. Inspection of backbone to backbone NOEs on a residue versus residue 2D map indicated two antiparallel double-stranded β -sheets. The patterns of amide proton protection, CSIs, and temperature dependence of chemical shifts as well as H^N-H α vicinal coupling constants confirmed the presence of two β -sheets. Four reverse turns of the polypeptide chain were also identified, which involved residues 8-11, 17-20, 23-26, and 31-34.

Analysis of 2D NOE experiments provided 448 distance constraints. Of those, 129 were intraresidue, 137 sequential, 43 medium range, and 139 long range. Additionally, 26 vicinal coupling constants were used as torsion angle constraints and 11 hydrogen bonds were used as distance constraints. The highlights of the 3D structure were described previously (Krezel et al., 1994). The structure of decorsin is very well defined, except for the first three N-terminal residues and some surface side chains. The dominant features of the decorsin structure are two β -sheets, which are linked by three disulfide bonds. The three disulfides form the core of the

Table 1. Characterization of rDecorsin and mDecorsin

	Activity (IC ₅₀ , nM)		Mass (amu)	
	Fg/ $\alpha_{1\text{Ib}}\beta_3$ ELISA	Platelet aggregation	Reduced	Non-reduced
rDecorsin	3.5	270	4,384.3	4,378.8
mDecorsin	30	12,000	4,385.0	4,378.6

Fig. 2. Amino acid sequence of decorsin and summary of observed NMR measurements. The RGD recognition sequence is boxed. The strength of amide protection (k_{ex}) is indicated by the height of the black bar above the sequence. Below the sequence, estimates of coupling constants are given in Hz (see text for details). Intensity of observed sequential NOEs is indicated by the line thickness. Gray lines represent $d_{\alpha\beta}(i, i+1)$ and $d_{\beta\beta}(i, i+1)$ interactions with proline residues. Differences between observed and random coil chemical shifts are indicated for $^1\text{H}^\alpha$ and $^{13}\text{C}^\alpha$, positive values are shifts to lower field. Disulfide bonds connecting cysteine residues (solid lines) as well as β -strands (zigzag lines) and turns deduced from the 3D structure are indicated.

protein, which contains no bulky hydrophobic side chains. The torsion angles of cysteine residue side chains are well defined.

Backbone dynamics

The dynamics of the decorsin backbone was measured by hydrogen exchange (see above) and $^{13}\text{C}^\alpha$ linewidth analysis (Fig. 3). Linewidths were measured for all $^{13}\text{C}^\alpha$. The narrowest lines were found for N-terminal residues 1–3. Somewhat broader although still narrower than the average for all residues (9.1 Hz) were the lines for residues 32–34. These include G32 and D33 of the RGD site.

Discussion

Expression and purification

A synthetic gene for decorsin has been constructed and put under the control of the alkaline phosphatase promoter and stII signal sequence. The original decorsin expression phagemid, pZdecorsin, was constructed with six synthetic oligonucleotides that encoded the Factor Xa linker region and the decorsin gene. The purpose of this construct was to express decorsin as a fusion protein with the Z domain of protein A, thus facilitating rapid purification by IgG-Sepharose affinity chromatography. The fusion protein could then be proteolytically cleaved by Factor Xa, liberating free decorsin. Although this protocol worked well on a small scale, the cleavage did not scale up efficiently (data not shown). For this reason, the Z-decorsin phagemid was mutagenized using oligonucleotide site-directed mutagenesis to delete the Z domain and Factor Xa linker regions, resulting in the final direct secretion phagemid, pDecorsin.

Decorsin is expressed at high levels and is found both in the periplasm and the media following a 10 L fermentation run. The protein was purified from the periplasm because it contained fewer contaminants than did the media, as assessed by SDS polyacrylamide gel electrophoresis (data not shown). Following purification by size-exclusion filtration and C_{18} reversed-phase chromatography, two forms of decorsin were found. Although the two forms were indistinguishable in amino acid composition, mass spectra, and HPLC retention time of reduced samples, they differed in HPLC retention time of non-reduced samples and in specific activity. rDecorsin was found to be equivalent to the authentic, leech-derived decorsin in both ELISA and platelet aggregation assays, whereas mDecorsin was considerably less potent in these assays (Table 1). The supposition that mDecorsin is a clipped form of rDecorsin is inconsistent with both the mass spectral data and the fact that the two reduced forms co-elute. Therefore, we conclude that rDecorsin is correctly folded decorsin, while mDecorsin consists of one or more misfolded form(s) due to either different disulfide pairings or different proline conformers resulting from the *cis-trans* isomerization. The misfolded form, while still retaining some ability to inhibit both $\text{Fg}/\alpha_{\text{IIb}}\beta_3$ binding and platelet aggregation, apparently lacks the requisite conformation present in the native decorsin molecule necessary for tight binding to $\alpha_{\text{IIb}}\beta_3$. Somewhat similar results were observed for recombinant ornatin E expressed in *E. coli*, where both correctly folded and misfolded disulfide variants were found (Mazur et al., 1991). Ornatin E did not express as well as decorsin in *E. coli* and was purified from the extracellular media.

Chemical synthesis of decorsin (Polverino de Laureto et al., 1998) produced a protein with similar activity ($\text{IC}_{50} \sim 100 \text{ nM}$ vs. ours 270 nM). The overall yield of the native, fully oxidized form was about 10%, including peptide chain assembly, resin cleavage,

Table 2. Amide proton exchange rates and relative solvent accessibilities for backbone NH groups in decorsin at pH 5.5 and 27°C

Residue	k_{ex} (min ⁻¹)	Intrinsic k_{ex} (min ⁻¹)	Solvent accessibility (%)
Ala1	>1		100
Pro2			16
Arg3	>1	146.35	74
Leu4	>1	2.27	66
Pro5			20
Gln6	0.235	3.41	77
Cys7	0.224	28.63	1
Gln8	0.217	16.88	0
Gly9	0.250	15.05	19
Asp10	>1	9.16	17
Asp11	0.066	8.20	0
Gln12	>1	7.78	48
Glu13	0.038	4.91	44
Lys14	0.226	5.36	61
Cys15	0.013	23.82	0
Leu16	0.016	3.91	0
Cys17	0.015	11.17	0
Asn18	>1	45.34	2
Lys19	>1	9.74	3
Asp20	0.070	8.16	0
Glu21	>1	4.11	75
Cys22	0.008	20.70	0
Pro23			0
Pro24			0
Gly25	0.250	5.50	9
Gln26	0.007	8.68	9
Cys27	0.015	28.63	7
Arg28	0.004	17.67	0
Phe29	0.018	4.91	71
Pro30			27
Arg31	>1	3.57	69
Gly32	>1	15.76	88
Asp33	>1	9.16	77
Ala34	0.250	6.78	45
Asp35	0.225	6.21	59
Pro36			0
Tyr37	0.002	1.62	0
Cys38	0.049	20.28	68
Glu39	0.009	0.26	0

and oxidative refolding. In particular, the oxidative refolding step for the synthetic material proved to be relatively difficult. This remains the method of choice for incorporation of unnatural amino acids into this protein. The recombinant method was able to produce relatively large amounts of protein; however, it did result in two separable isoforms of decorsin. Clearly, depending on the specific application either method of decorsin production may be appropriate. For a large-scale production as well as uniform labeling with stable isotopes the recombinant approach is probably faster and more economical.

Resonance assignments

Sequential assignments of the protons in decorsin were completed using ¹H 2D NMR techniques. There were 14 AMX spin systems

Fig. 3. Backbone dynamics of decorsin reflected by different parameters. **A:** Number of nonredundant NOE constraints per residue. **B:** RMSD in Å for the backbone atoms among 25 calculated structures of decorsin. **C:** ¹³Cα linewidths in Hz. **D:** $\Delta\delta_{sc} = 20 \times (1.5 - |\delta_{obs} - \delta_{rc}|)$ as previously defined (Wishart & Sykes, 1994b).

in the polypeptide sequence and only 13 types of residues. The high proline content prevented long sequential walks with amide protons. Minor peaks corresponding to small amounts of impurities, most likely due to acid hydrolysis of the D35–P36 peptide bond, were observed in the spectra. The relatively high content of reverse turns provided a number of strong sequential H^N–H^N cross peaks in NOESY spectra, facilitating the alignment of spin systems. Experiments at temperatures a few degrees above and below 27 °C aided in the resolution of overlapping peaks and in the identification of resonances near the water signal. The dispersion of chemical shifts of decorsin protons is similar to other proteins of this size. The only outstanding signal is the downfield-shifted amide proton of Q8. The dipolar interactions between this proton and the aromatic ring protons of F29 indicate a large ring current shift. An additional contributing factor is a probable hydrogen bond to the side-chain carboxyl of D11, confirmed by results of pH titration experiments (data not shown). In most cases, CSI analysis (Fig. 2) showed good agreement of the chemical shift deviations from the random coil values and the position of the secondary structure elements as they were gleaned from the tertiary structure. Used in the predictive mode, CSI analysis indicated four out of five β -strands. Only residues 20–23 were not classified into potential β -strand. Interestingly, it is in this very fragment where C22 is positioned. This residue has the highest CSIs for ¹H α and ¹³C α resonances, both consistent with the β -strand conformation. The preceding residue E21 does not show the CSI characteristic for the β -strand. This situation is mirrored, although with somewhat lower

intensity on the opposite strand with the CSI of C15 and L16. In an isolated double-stranded antiparallel β -sheet, every other residue forms two hydrogen bonds with the opposite strand; remaining residues do not form backbone hydrogen bonds, and consequently, their CSI are not likely to be characteristic for β -strands.

Hydrogen bonds and temperature dependence of chemical shifts

The amide protons of decorsin exchange with rates similar to those observed for hirudin ((Szyperski et al., 1992), pH 4.5, 22 °C), a close structural relative of decorsin. We were able to observe half of all possible amide protons in a TOCSY experiment, which ended 30 min after the lyophilized decorsin sample was dissolved in 100% D₂O at 27 °C, pH 5.5. Most of the slowly exchanging amide protons were involved in hydrogen bonds stabilizing the antiparallel β -sheets. Amide protons of Q6, C7, Q8, and E13 were not involved in the secondary structure elements. The carbonyl oxygen of L16 and the side-chain carboxyl group of D11 serve as acceptors for C7 and Q8, respectively (Krezel et al., 1994). Table 2 shows the averaged solvent accessibilities for amide groups in 25 NMR structures of decorsin along with observed exchange rates and the expected random coil intrinsic exchange rates (Bai et al., 1993). The only residue showing amide protection without correlation to either proposed hydrogen bond or deep burial was F29. The side chains of Q8 and/or D10 can possibly interact with this group. The structurally homologous residue L30 of hirudin also shows the protected amide interacting with the side chain of S9. The overall pattern of protection from amide exchange for hirudin and decorsin are very similar (Szyperski et al., 1992). The lack of correlation of the predicted intrinsic exchange rates with observed ones suggests that despite the small-size, decorsin amide proton exchange rates are dominated by the hydrogen bonding and steric effects. The observed thermal stability of decorsin, $T_m \sim 75$ °C (Polverino de Laureto et al., 1998), supports this conclusion.

The above arguments suggest that chemical shift temperature dependencies could also show trends observed for stably folded proteins (Baxter & Williamson, 1997). The temperature coefficients of chemical shifts for amide protons varied from -10.9 to -0.3 ppb/°C. The larger (smaller magnitude) values correlated very well with observed and/or potential hydrogen bonding. Residues $i + 3$ of the tight reverse turns (D11, D20, Q26, A34) showed uniformly low temperature coefficients, although other residues involved in turns did not show any discernible trend. A striking example here was A34 following RGD loop with its amide proton positioned to form a hydrogen bond with the carbonyl oxygen of R31. Despite small temperature coefficients, the amide protons of A34, G9, D11, and G25 exchanged rapidly with the solvent. On the other hand, the relatively slowly exchanging amide proton of L16 had a temperature coefficient of -9.5 ppb/°C. Other well-protected amides showing large temperature coefficients included Q6, C7, K14, and F29. The range of temperatures used to measure these coefficients was rather narrow (18 to 30 °C) and far from the T_m of 75 °C for decorsin. Nevertheless, small rearrangements of structure could contribute to some larger than expected temperature coefficients.

Secondary and tertiary structure

Secondary structure elements found in decorsin involve residues 14–17 and 20–23 in one double-stranded antiparallel β -sheet, and

residues 26–28 and 37–39 in another. These elements were clearly defined by the number of dipolar interactions. The key NOEs in aligning the strands were L16 H $^{\alpha}$ to E21 H $^{\alpha}$ and C27 H $^{\alpha}$ to C38 H $^{\alpha}$. The predominantly large H N –H $^{\alpha}$ coupling constant values support a β -type structure. Also, the CSI temperature dependence of chemical shifts and hydrogen protection patterns supported the presence of the secondary structure elements shown in Figure 4.

Despite its small size, decorsin exhibits a considerably better defined secondary structure than do other small, potent, RGD-containing $\alpha_{IIb}\beta_3$ antagonists. Kistrin, echistatin, and flavoridin contain very little regular secondary structure (Adler et al., 1991; Chen et al., 1991; Cooke et al., 1991; Dalvit et al., 1991; Saudek et al., 1991; Senn & Klaus, 1993). The dominant structural feature of these proteins is the disulfide bond network.

The homology of the tertiary structure of decorsin to hirudin was an unanticipated result of the structure determination (Krezel et al., 1994). Based on the sequences alone, we proposed that antistasin, another leech protein, may share the same fold. Another highly related protein to antistasin, termed hirustasin from the leech *H. medicinalis* that inhibits tissue kallikrein has been characterized (Sollner et al., 1994). Recently, structures have been determined for both of these proteins (Lapatto et al., 1997; Mittl et al., 1997). Although our sequence alignment for antistasin was incorrect, a large portion of the fold and the disulfide connectivities for both antistasin and hirustasin are similar to residues 16–39 of decorsin. It is worth noting that the small size and unusual amino acid composition defeat most attempts at finding homologs of decorsin by the popular sequence searching algorithms. One exception is a group of isoforms of ornatins (Mazur et al., 1991), another potent $\alpha_{IIb}\beta_3$ antagonist isolated from leech *P. ornata*. They have the same cysteine residue spacing and position of the RGD adhesion recognition site as decorsin. In addition, decorsin and ornatins have glycine residues in positions 9 and 25 (numbering of decorsin residues, alignment as shown in Fig. 4 in Mazur et al., 1991). These glycine residues are crucial for the formation of β -type reverse turns. Decorsin has a lysine in position 19, which is the $i + 2$ position of another β -turn. All ornatins have a glycine in the equivalent position making the formation of a β -turn even more favorable. These observations suggest that both secondary and tertiary structure of ornatins and decorsin will be very similar. Because ornatins do not have proline residues in the RGD containing loop, their RGD sequence could occur in a very regular β -turn.

Stereospecific assignments could be determined for 21 of 25 nondegenerate β -methylene groups. Overlap of H $^{\alpha}$ –H $^{\beta}$ cross peaks in the DQF-COSY prevented the estimation of coupling constants in those four cases left unassigned. The qualitative nature of data used in this process made the manual method of assigning each individual prochiral group more reliable than automated approaches (Güntert et al., 1989).

Decorsin contains six prolines and six cysteines among its 39 residues. The determination of the tertiary structure of decorsin was critically dependent upon the correct assignment of the disulfides (Gray, 1993). We were able to identify a number of H $^{\beta}$ –H $^{\beta}$ dipolar interactions among cysteine residues. The two disulfide bridges between C22 and C38 and between C7 and C15 were clearly defined. Therefore, the remaining disulfide was deduced to be between C17 and C27. The chemical shift of β -methylene protons for C17 and C27 are so close that cross peaks between them could not be observed. We also found a number of NOEs between C7 and C17 and between C7 and C27. From these observations we

Fig. 4. Stereo diagram of 25 superimposed model structures of decorsin (Protein Data Bank code 1DEC). Residue numbers and disulfide bonds (green) are shown for the first model. Only backbone atoms are shown for the remaining models. Residues with C^α linewidths below 6 Hz are in yellow, 6–8 Hz in orange, 8–10 Hz in red, and above 10 Hz in black. Additionally, for the first model the following representation is used. The spheres are shown for slowly exchangeable amide protons, largest spheres correspond to slowest exchange. Amide nitrogens of the amide protons with chemical shift temperature dependence greater than -4.5 ppb/ $^\circ\text{C}$ are also indicated with small spheres. Backbone carbonyl groups are shown as solid sticks.

expected that the interior of the molecule could largely be formed by the disulfide bonded cysteine residues. The results of structure calculations served as a final confirmation of the disulfide bonding pattern. Only the pattern indicated in Figure 2 produced structures without constraints violations.

The RGD adhesion site recognition sequence of decorsin is found in one of the reverse turns. The placement of this sequence at the apex of flexible loops has been found in all disintegrin $\alpha_{\text{IIb}}\beta_3$ antagonists containing the RGD sequence thus far (Adler et al., 1991; Chen et al., 1991; Cooke et al., 1991; Dalvit et al., 1991; Saudek et al., 1991; Adler & Wagner, 1992; Lazarus & McDowell, 1993; Senn & Klaus, 1993). In addition, the RGD sequence has also been found at the apex of a mobile loop in the tenth type III module of fibronectin, one of the natural protein ligands that binds to $\alpha_{\text{IIb}}\beta_3$ (Main et al., 1992). The RGD sequence (residues 31–33) in decorsin is flanked by two proline residues (P30 and P36). This precludes the formation of a canonical antiparallel β -sheet conformation. On the other hand, the coupling constants and a number of observed NOEs indicate that this region has a well-defined conformation. In this sense, decorsin is exceptional among active RGD-containing protein $\alpha_{\text{IIb}}\beta_3$ antagonists that have been studied structurally.

Dynamics of decorsin

Rather unexpected conformational rigidity of the RGD recognition site in decorsin prompted us to examine this feature more closely.

The large number of nonredundant distance and angle constraints is the primary source of the conformational definition in NMR structures. Both constraints and average backbone RMSD are shown in Figure 3. The RGD loop shows densities of NOEs similar to the residues 15–22 hairpin loop; consequently, the RMSD are also similar. More sensitive measures of molecular dynamics are relaxation times of nuclei and exchange of amide protons for deuterons. The relaxation of backbone $^{13}\text{C}^\alpha$ nuclei was estimated by the linewidth measurements in ^1H – ^{13}C HSQC spectra of natural abundance protein. These measurements are not readily translated into the relaxation parameters of individual nuclei (Dayie et al., 1996). Nevertheless, they can be used as qualitative estimate of the relative flexibility of various parts of the macromolecule. The first three N-terminal residues of decorsin showed the narrowest lines. Residues 32–34 also had very narrow lines. Similarly, the amide exchange rates for residues 31–35 were very high. None of the residues in this loop showed any measurable protection in contrast to the residues in a 15–22 loop. Chemical shifts indexes show very strong statistical correlation with protein conformation (Wishart & Sykes, 1994a). The analysis of deviation of chemical shifts from the random coil values was also proposed as another estimate of protein mobility (Wishart & Sykes, 1994b). The results of such analysis are included in Figure 3. Again, in this case, the N-terminal residues and the tip of the RGD loop appear as the most dynamic parts of decorsin. As it was pointed out in the discussion of CSI analysis of the secondary structure, chemical shifts data have to be

interpreted carefully because the number of chemical shift values can be strongly affected by the proximity of the side-chain and backbone groups. Taken together, all available evidence indicates that the RGD loop of decorsin, despite covalent constraints imposed by proline residues flanking the five middle residues, remains rather dynamic. Figure 4 summarizes these observations. The “extremities” of the protein, N-terminus, and RGD loop contain no protected amide protons and the narrowest linewidths. The disulfide-bond interconnected loops form the globular core with many protected amide protons. The RGDAD loop begins and ends with conformationally restrictive proline residues anchored in the very stable antiparallel β -sheet. Additionally, P30 and P36 interact with aromatic rings of Y37 and F29, respectively. The stacking interactions between proline and aromatic side chains have been observed to stabilize conformation in short peptides in aqueous solutions (Yao et al., 1994). This suggests that conformational changes within the RGD loop of decorsin are slow enough to allow the observation of a number of NOE interactions. On the other hand, during biological assays on a much longer timescale, this loop most likely can undergo significant conformational change. This could explain the fact that the levels of the biochemical and biological activities of decorsin are similar to the ones observed for other proteins containing extremely flexible RGD sites.

Materials and methods

Materials

Restriction enzymes were purchased from New England Biolabs (Beverly, Massachusetts) or Bethesda Research Laboratories. T4 DNA ligase and T4 DNA kinase were purchased from Bethesda Research Laboratories, and T7 DNA polymerase, T7 DNA polymerase and dNTPs were purchased from Pharmacia (Uppsala, Sweden). All enzymes were used according to the manufacturers' recommendations. ^{35}S ATP was purchased from Amersham. The pZKis phagemid was provided by Mark Dennis (Dennis et al., 1993). The *E. coli* strains used were 16C9, a derivative of W3110 tonA (ATCC 27325), JM101 (Δ lac-pro supE thi[F' traD36 proAB+ lacI^Q Z Δ M5]) and BMH 71-18 *mutL*. Reagents used in the Fg/ $\alpha_{\text{IIB}}\beta_{\text{III}}$ ELISA have been previously described (Seymour et al., 1990); oligonucleotides were synthesized using hydrogen phosphonate chemistry (Froehler et al., 1986) and purified by polyacrylamide gel electrophoresis prior to use. Nuge C₁₈ resin was purchased from Separation Industries (Metuchen, New Jersey). Bradford dye reagent and bovine serum albumin were obtained from Bio-Rad (Hercules, California).

Gene synthesis and phagemid construction

All DNA manipulations were performed according to standard procedures (Sambrook et al., 1989) unless otherwise indicated. Initially, six synthetic oligonucleotides, ranging in length from 52 to 66 bases, were designed to encode a Factor Xa cleavage site and the decorsin gene. The synthetic oligonucleotides were phosphorylated individually, and annealed together in pairs by gradual cooling from 85 to 20°C. The double-stranded segments were then ligated together with the purified 4.1 kb *NarI*/*StyI* fragment of pZKis (Dennis et al., 1993). The ligation mix was digested with *StuI* to eliminate any uncut pZKis phagemid and transformed into JM101; the cells were then plated on minimal glucose plates (Carter,

1987) containing 50 $\mu\text{g}/\text{mL}$ carbenicillin. Twelve carbenicillin-resistant colonies (out of ~ 350) were picked for restriction analysis and sequencing (Sanger et al., 1977). One isolate of the original 12 contained the correct sequence and is referred to as pZDecorsin.

pZDecorsin was then mutagenized to delete the Z domain and Factor Xa linker regions as follows. A deletion primer designed as a perfect match between the final 15 bases of the *stII* leader sequence and the initial 15 bases of the decorsin gene (5'ACAAATGCCTATGCTGCGCCGCGGCTGCCG3') was phosphorylated and annealed to a single-stranded pZDecorsin template by gradual cooling from 100 to 20°C. The annealed template/primer was then extended and ligated by the addition of ATP (1 mM), dNTPs (1 mM), T4 DNA ligase (3U), T4 polynucleotide kinase (5 U), and T7 DNA polymerase (0.6 U), with appropriate buffers. The extension/ligation mix was transformed into BMH 71-18 *mutL* cells, the cells were grown overnight in 2YT media containing 50 $\mu\text{g}/\text{mL}$ carbenicillin, and DNA was isolated directly from this pool of transformed cells. Restriction selection with *MluI* and *BglII* eliminated nonmutagenized phagemids. *E. coli* JM101 was then transformed and plated onto minimal glucose carbenicillin plates. Twelve carbenicillin resistant colonies (out of ~ 500) were picked for restriction analysis and sequencing. Eight of these clones were correct and one was chosen as pDecorsin.

Activity assays

The Fg/ $\alpha_{\text{IIB}}\beta_3$ ELISA was used to follow activity throughout the purification. This assay as well as *in vitro* human platelet aggregation assays were performed as previously described (Seymour et al., 1990).

Protein expression and purification

pDecorsin was transformed into *E. coli* strain 16C9 (a derivative of *E. coli* strain W3110 tonA). Cultures of 16C9 pDecorsin were grown for 32 h at 37°C in an aerated 10 L fermentor in medium that contained 50 mg/L ampicillin, 9 g/L digested casein, 1.2 g/L yeast extract, 26 mM (NH₄)₂SO₄, 6 mM NaH₂PO₄, 10 mM K₂HPO₄, 2.6 mM sodium citrate, 14 mM KCl, 7.5 mM MgSO₄, 0.9 mM FeCl₃, and 10 μM each of ZnSO₄, MnSO₄, CuSO₄, CoCl₂, H₃BO₃, and NaMoO₄. Glucose was provided to maintain glucose excess or avoid anaerobiosis, depending on cell density, and the pH was maintained at 7.0 with the addition of NH₄OH and H₂SO₄. The cell density at harvest was about 80 OD₅₅₀. Following centrifugation to recover the cells, the cell paste was stored at -20°C prior to use.

Recombinant decorsin was purified from the cell paste as follows. The periplasmic contents from 402 g of cell paste were obtained by resuspending the paste in 10 mM Tris, 1 mM EDTA, pH 7.6 containing 1 mM PMSF (2.5 L). The resuspension was stirred for 25 min at room temperature to thaw the paste and then for 75 min at 4°C to complete the osmotic shock. The resuspension was then spun at 10,800 $\times g$ for 10 min in a RSCB refrigerated centrifuge. The supernatant was still cloudy and was respun at 30,600 $\times g$ for 20 min. The supernatant from the second spin (2.1 L) was then passed through a YM30 membrane in a 400 mL Amicon concentrator to remove high molecular weight contaminants. The flow through from the YM 30 membrane (2.0 L) was then loaded on a 5 $\times$ 15 cm² Nuge C₁₈ preparative reversed-phase column that had been previously equilibrated in 0.1% TFA. Following the load, the column was washed with one column volume

of 0.1% TFA followed by four column volumes of 10% acetonitrile, 0.1% TFA. Recombinant decorsin was then eluted with 50% acetonitrile, 0.1% TFA. The eluted decorsin (985 mL) was partially lyophilized to remove the acetonitrile, thawed, and loaded onto a 2.5×50 cm², 300 Å, Vydac C₁₈ column equilibrated in 0.1% TFA. The column was eluted with a linear 12–26% acetonitrile gradient, 0.175%/min, at a flow rate of 18 mL/min; 1 min fractions were collected. The fractions were assayed for the ability to inhibit Fg binding to $\alpha_{\text{Ib}}\beta_3$; aliquots of each fraction were also subjected to analytical HPLC. The fractions were pooled according to inhibitory activity and retention time and repurified on the same HPLC column. Acetonitrile and TFA were removed from the purified pools by rotary evaporation and were characterized as described below.

Protein quantitation

Protein concentrations of impure samples were determined by the method of Bradford (1976) with bovine serum albumin as a standard, and of pure samples by amino acid analysis. Amino acid analysis of decorsin was determined after hydrolysis (24 h with 6 N constant boiling HCl under vacuum) using a Millipore Picotag workstation. The hydrolysates were dried on a Savant Speed-Vac concentrator and analyzed on a Beckman model 6300 amino acid analyzer equipped with a ninhydrin detector. Protein sequence analysis by automated Edman degradation was performed on an Applied Biosystems 473A protein sequencer equipped with on-line 120A PTH analyzer; PTH-amino acids were integrated with a PE-Nelson model 3000 data acquisition system as previously described (Henzel et al., 1987).

Other analytical methods

Analytical HPLC was performed using a 250×4.6 mm Vydac C₁₈ reversed-phase column (5 μ m) as previously described (Seymour et al., 1990). FAB-MS spectra were obtained with a JEOL HX110HF/HX110HF tandem double-focusing mass spectrometer as reported earlier (Seymour et al., 1990). Electrospray mass spectra were obtained using a Sciex API-III triple quadrupole mass spectrometer equipped with the Ionspray interface operating at 5 kV and calibrated with a solution of polypropylene glycols. Thioglycerol at alkaline pH was used to reduce decorsin samples on the probe tip.

NMR sample preparation

Lyophilized decorsin was dissolved in 0.5 mL of 90% (v/v) H₂O/D₂O or 100% D₂O at ~4 mM concentration. In the following text we refer to the decorsin dissolved in 90% (v/v) H₂O/D₂O as decorsin in water. The pH of the sample was adjusted to 4.7 without correction for deuterium isotope effect.

NMR data acquisition and analysis

A series of homonuclear 2D NMR experiments were recorded with a decorsin sample at 30 °C. The sample contained 0.4 mM [d₄]TSP as an internal standard. Referencing of ¹H resonances was done by setting δ [d₄]TSP equal to -0.0063 at pH 4.7. Referencing of ¹³C resonances was done indirectly assuming ¹³C/¹H frequency ratio of 0.25144953 (Wishart et al., 1995). Experiments were done on AMX500 (11.7 T) and AMX600 (14.2 T) instruments (Bruker

GmbH, Karlsruhe, Germany), operating at 500.13 and 600.13 MHz, respectively. NOESY spectrum was recorded on Varian Unity operating at ¹H 750 MHz. Sweep widths were adjusted to 7,042 Hz at 500.13 MHz and 8,474 Hz at 600.13 MHz. The sizes of acquired free induction decays were 2,048 or 4,096 complex points. The carrier frequency was set on the H₂O or the residual HDO signal. Time proportional phase incrementation was used for 1,024 data points collected in the indirect (*t*₁) dimension (Marion & Wüthrich, 1983). In the majority of spectra, water signal was suppressed by low-level presaturation during the relaxation delay (1.3 to 1.5 s). In NOESY experiments recorded at 750 MHz, gradient water suppression was used (Piotto et al., 1992). For the sample in water, at least 96 transients were acquired for each *t*₁ value and at least 32 transients for the sample in D₂O. TOCSY spectra (Braunschweiler & Ernst, 1983) were recorded with a DIPSI-II mixing pulse sequence (Shaka & Freeman, 1983). Three different mixing times (32, 55, and 85 ms) were used in high resolution TOCSY experiments.

The amide proton exchange was measured in a sequence of short TOCSY experiments carried out at pH 5.5, 27 °C. A lyophilized decorsin sample was dissolved in 100% D₂O and placed in the pre-shimmed magnet. A series of 20 min TOCSY experiments was then recorded. The first experiment was started about 10 min after dissolving the protein and further experiments were started after 30, 50, 70, 100, 190, and 490 min. Intrinsic exchange rates for amide protons were calculated as previously described (Bai et al., 1993).

The percent of solvent accessibility for amide groups was calculated in the following way. Solvent accessible surfaces were calculated and stored for all 25 NMR structures of decorsin using standard algorithm (Connolly, 1983) available in InsightII package (MSI, San Diego, California). The areas for amide groups (N and HN atoms of peptide bonds) were averaged and compared to the area of the amide group in a fully extended polyglycine.

Time domain data were processed using Felix (MSI) software on SGI computers. The final digital resolution in both dimensions of 2D spectra was 2 Hz per point or higher. Baseline distortions in spectra were corrected with XWINNMR software (Bruker GmbH). The H^N–H ^{α} coupling constants were derived from the DQF-COSY spectrum (Piantini et al., 1982; Rance et al., 1983; Shaka & Freeman, 1983) as described by (Kim & Prestegard, 1989). The error of these estimates was ± 2 Hz, mainly because of the low signal-to-noise ratio. The estimates of H ^{α} –H ^{β} coupling constants were made by comparison with calculated cross-peak patterns (Hyberts et al., 1987). These estimates were qualitative and were done for the sole purpose of stereospecific assignments. The temperature coefficients of chemical shifts for all signals were derived from NOESY experiments (Jeener et al., 1979; Kumar et al., 1980) recorded at 18 and 30 °C.

We measured linewidths of ¹³C ^{α} resonances in the natural abundance sample of decorsin. High-resolution ¹H–¹³C HSQC experiments (512 *t*₂ and 680 *t*₁ complex points) were processed without window multiplication of the data points. For each C ^{α} cross peak 8 to 20 columns from HSQC spectrum were added, inverse Fourier transformed, zero filled to 16,384 complex points, and Fourier transformed again. Line fitting routine in Felix was used to estimate the linewidths. The estimated error of measurements was ± 1 Hz.

Structure determination

Secondary structure elements were identified based on observed NOE patterns and values of H^N–H ^{α} vicinal coupling constants

(Wüthrich et al., 1991). Disulfide bonds between cysteine residues were first tentatively assigned based on observed $H_i^\beta-H_j^\beta$ NOEs and confirmed during structure calculations. Three-dimensional structures were calculated using distance geometry-simulated annealing protocol (Nilges et al., 1988). Embedded structures (Crippen & Havel, 1978) were optimized with a simulated annealing and energy minimized with distance constraints (Krezel et al., 1994).

Supplementary material in the Electronic Appendix

The 1H and ^{13}C sequential resonance assignments, $^{13}C^\alpha$ line-widths, and chemical shift temperature coefficients for amide protons are provided (Table S1). The amino acid analysis data for decorsin are also included (Table S2).

Acknowledgments

This research was supported by NIH grant GM38608. We are grateful to Mark Vasser, Peter Ng, and Parkash Jhurani for DNA synthesis, Mark Velligan for *E. coli* fermentations, Allan Padua for amino acid composition data, John Stults and Karl Clauser for mass spectral data, and Sherron Bullens for the platelet aggregation assay data.

References

- Adler M, Lazarus RA, Dennis MS, Wagner G. 1991. Solution structure of kistrin, a potent platelet aggregation inhibitor and GP IIb-IIIa antagonist. *Science* 253:445–448.
- Adler M, Wagner G. 1992. Sequential 1H NMR assignments of kistrin, a potent platelet aggregation inhibitor and glycoprotein IIb-IIIa antagonist. *Biochemistry* 31:1031–1039.
- Arocha-Pinango CL, Marchi R, Carvajal Z, Guerrero B. 1999. Invertebrate compounds acting on the hemostatic mechanism. *Blood Coagul Fibrinolysis* 10:43–68.
- Bai Y, Milne JS, Mayne L, Englander SW. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins* 17:75–86.
- Baxter NJ, Williamson MP. 1997. Temperature dependence of 1H chemical shifts in proteins. *J Biomol NMR* 9:359–369.
- Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–254.
- Braunschweiler L, Ernst RR. 1983. Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J Magn Reson* 53:521–528.
- Carter P. 1987. Improved oligonucleotide-directed mutagenesis using M13 vectors. *Methods Enzymol* 154:382–403.
- Chang CN, Rey M, Bochner B, Heyneker H, Gray G. 1987. High-level secretion of human growth hormone by *Escherichia coli*. *Gene* 55:189–196.
- Chen Y, Pitzenger SM, Garsky VM, Lumma PK, Sanyal G, Baum J. 1991. Proton NMR assignments and secondary structure of the snake venom protein echistatin. *Biochemistry* 30:11625–11636.
- Coller BS. 1992. Antiplatelet agents in the prevention and therapy of thrombosis. *Annu Rev Med* 43:171–180.
- Connolly ML. 1983. Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221:709–713.
- Cooke RM, Carter BG, Martin DM, Murray-Rust P, Weir MP. 1991. Nuclear magnetic resonance studies of the snake toxin echistatin. 1H resonance assignments and secondary structure. *Eur J Biochem* 202:323–328.
- Crippen GM, Havel TF. 1978. Stable calculation of coordinates from distance information. *Acta Crystallogr A* 34:282–284.
- Dalvit C, Widmer H, Bovermann G, Breckenridge R, Metternich R. 1991. 1H NMR studies of echistatin in solution. Sequential resonance assignments and secondary structure. *Eur J Biochem* 202:315–321.
- Dayie KT, Wagner G, Lefevre JF. 1996. Theory and practice of nuclear spin relaxation in proteins. *Annu Rev Phys Chem* 47:243–282.
- Dennis MS, Carter P, Lazarus RA. 1993. Binding interactions of kistrin with platelet glycoprotein IIb-IIIa: Analysis by site-directed mutagenesis. *Proteins* 15:312–321.
- Froehler BC, Ng PG, Matteucci MD. 1986. Synthesis of DNA via deoxynucleoside H-phosphonate intermediates. *Nucleic Acids Res* 14:5399–5407.
- Giancotti FG, Ruoslahti E. 1999. Integrin signaling. *Science* 285:1028–1032.
- Gray WR. 1993. Disulfide structures of highly bridged peptides: A new strategy for analysis. *Protein Sci* 2:1732–1748.
- Güntert P, Braun W, Billeter M, Wüthrich K. 1989. Automated stereospecific 1H NMR assignments and their impact on the precision of protein structure determinations in solution. *J Am Chem Soc* 111:3997–4004.
- Henzel WJ, Rodriguez H, Watanabe C. 1987. Computer analysis of automated Edman degradation and amino acid analysis data. *J Chromatogr* 404:41–52.
- Hyberts SG, Marki W, Wagner G. 1987. Stereospecific assignments of side-chain protons and characterization of torsion angles in Eglin c. *Eur J Biochem* 164:625–635.
- Jaseja M, Lu X, Williams JA, Sutcliffe MJ, Kakkar VV, Parslow RA, Hyde EI. 1994. 1H -NMR assignments and secondary structure of dendroaspin, an RGD-containing glycoprotein IIb-IIIa (α IIb- β 3) antagonist with a neurotoxin fold. *Eur J Biochem* 226:861–868.
- Jaseja M, Smith KJ, Lu X, Williams JA, Trayer H, Trayer IP, Hyde EI. 1993. 1H -NMR studies and secondary structure of the RGD-containing snake toxin, albolabrin. *Eur J Biochem* 218:853–860.
- Jeener J, Meier BH, Bachmann P, Ernst RR. 1979. Investigation of exchange processes by two-dimensional NMR spectroscopy. *J Chem Phys* 71:4546–4553.
- Kim Y, Prestegard JH. 1989. Measurement of vicinal couplings from cross peaks in COSY spectra. *J Magn Reson* 84:9–13.
- Krezel AM, Wagner G, Seymour-Ulmer J, Lazarus RA. 1994. Structure of the RGD protein decorsin: Conserved motif and distinct function in leech proteins that affect blood clotting. *Science* 264:1944–1947.
- Kumar A, Ernst RR, Wüthrich K. 1980. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem Biophys Res Commun* 95:1–6.
- Lapatto R, Kregel U, Schreuder HA, Arkema A, de Boer B, Kalk KH, Hol WG, Grootenhuis PD, Mulders JW, Dijkema R, Theunissen HJ, Dijkstra BW. 1997. X-ray structure of antistasin at 1.9 Å resolution and its modelled complex with blood coagulation factor Xa. *EMBO J* 16:5151–5161.
- Lazarus RA, McDowell RS. 1993. Structural and functional aspects of RGD-containing protein antagonists of glycoprotein IIb-IIIa. *Curr Opin Biotechnol* 4:438–445.
- Main AL, Harvey TS, Baron M, Boyd J, Campbell ID. 1992. The three-dimensional structure of the tenth type III module of fibronectin: An insight into RGD-mediated interactions. *Cell* 71:671–678.
- Marion D, Wüthrich K. 1983. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of 1H - 1H spin-spin coupling constants in proteins. *Biochem Biophys Res Commun* 113:967–974.
- Mazur P, Henzel WJ, Seymour JL, Lazarus RA. 1991. Ornatin: Potent glycoprotein IIb-IIIa antagonists and platelet aggregation inhibitors from the leech *Placobdella ornata*. *Eur J Biochem* 202:1073–1082.
- McDowell RS, Dennis MS, Louie A, Shuster M, Mulkerrin MG, Lazarus RA. 1992. Mambin, a potent glycoprotein IIb-IIIa antagonist and platelet aggregation inhibitor structurally related to the short neurotoxins. *Biochemistry* 31:4766–4772.
- Mittl PR, Di Marco S, Fendrich G, Pohlig G, Heim J, Sommerhoff C, Fritz H, Priestle JP, Grutter MG. 1997. A new structural class of serine protease inhibitors revealed by the structure of the hirustasin-kallikrein complex. *Structure* 5:253–264. [Published erratum appears in *Structure* 5:585.]
- Nilges M, Clore GM, Gronenborn AM. 1988. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett* 229:317–324.
- Nutt E, Gasic T, Rodkey J, Gasic GJ, Jacobs JW, Friedman PA, Simpson E. 1988. The amino acid sequence of antistasin. A potent inhibitor of factor Xa reveals a repeated internal structure. *J Biol Chem* 263:10162–10167.
- Phillips DR, Charo IF, Parise LV, Fitzgerald LA. 1988. The platelet membrane glycoprotein IIb-IIIa complex. *Blood* 71:831–843.
- Phillips DR, Charo IF, Scarborough RM. 1991. GPIIb-IIIa: The responsive integrin. *Cell* 65:359–362.
- Piantini U, Sørensen OW, Ernst RR. 1982. Multiple quantum filters for elucidating NMR coupling networks. *J Am Chem Soc* 104:6800–6801.
- Piotto M, Saudek V, Sklenar V. 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2:661–665.
- Polverino de Laureto P, Scaramella E, De Filippis V, Marin O, Doni MG, Fontana A. 1998. Chemical synthesis and structural characterization of the RGD-protein decorsin: A potent inhibitor of platelet aggregation. *Protein Sci* 7:433–444.
- Rance M, Sørensen OW, Bodenhausen G, Wagner G, Ernst RR, Wüthrich K. 1983. Improved spectral resolution in cosy 1H NMR spectra of proteins via double quantum filtering. *Biochem Biophys Res Commun* 117:479–485.
- Ruoslahti E, Pierschbacher MD. 1987. New perspectives in cell adhesion: RGD and integrins. *Science* 238:491–497.

- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467.
- Saudek V, Atkinson RA, Pelton JT. 1991. Three-dimensional structure of echistatin, the smallest active RGD protein. *Biochemistry* 30:7369–7372.
- Sawyer RT. 1991. Thrombolytics and anti-coagulants from leeches. *Biotechnology* 9:513–515, 518.
- Senn H, Klaus W. 1993. The nuclear magnetic resonance solution structure of flavoridin, an antagonist of the platelet GP IIb-IIIa receptor. *J Mol Biol* 232:907–925.
- Seymour JL, Henzel WJ, Nevins B, Stults JT, Lazarus RA. 1990. Decorsin. A potent glycoprotein IIb-IIIa antagonist and platelet aggregation inhibitor from the leech *Macrobdella decora*. *J Biol Chem* 265:10143–10147.
- Shaka AJ, Freeman R. 1983. Simplification of NMR spectra by filtration through multiple-quantum coherence. *J Magn Reson* 51:169–173.
- Smith KJ, Jaseja M, Lu X, Williams JA, Hyde EI, Trayer IP. 1996. Three-dimensional structure of the RGD-containing snake toxin albolabrin in solution, based on ^1H NMR spectroscopy and simulated annealing calculations. *Int J Pept Protein Res* 48:220–228.
- Sollner C, Mentele R, Eckerskorn C, Fritz H, Sommerhoff CP. 1994. Isolation and characterization of hirustasin, an antistatin-type serine-proteinase inhibitor from the medical leech *Hirudo medicinalis*. *Eur J Biochem* 219: 937–943.
- Szyperski T, Guntert P, Stone SR, Wüthrich K. 1992. Nuclear magnetic resonance solution structure of hirudin(1–51) and comparison with corresponding three-dimensional structures determined using the complete 65-residue hirudin polypeptide chain. *J Mol Biol* 228:1193–1205.
- Wagner G, Wüthrich K. 1982. Sequential resonance assignments in protein ^1H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. *J Mol Biol* 155:347–366.
- Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD. 1995. ^1H , ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6:135–140.
- Wishart DS, Sykes BD. 1994a. The ^{13}C chemical-shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J Biomol NMR* 4:171–180.
- Wishart DS, Sykes BD. 1994b. Chemical shifts as a tool for structure determination. *Methods Enzymol* 239:363–392.
- Wüthrich K. 1986. *NMR of proteins and nucleic acids*. New York: Wiley.
- Wüthrich K, Spitzfaden C, Memmert K, Widmer H, Wider G. 1991. Protein secondary structure determination by NMR. Application with recombinant human cyclophilin. *FEBS Lett* 285:237–247.
- Yao J, Dyson HJ, Wright PE. 1994. Three-dimensional structure of a type VI turn in a linear peptide in water solution. Evidence for stacking of aromatic rings as a major stabilizing factor. *J Mol Biol* 243:754–766.