3. D. Camus and T. J. Hadley, *Science* **230**, 553 (1985).
4. P. A. Orlandi, F. W. Klotz, J. D. Haynes, *J. Cell Biol.* **116**, 901 (1992).
5. T. J. Hadley *et al.*, *J. Clin. Invest.* **80**, 1190 (1987).
6. B. K. L. Sim *et al.*, *J. Cell Biol.* 111, 1877 (1990).
7. F. W. Klotz *et al.*, *Mol. Biochem. Parasitol.* **51**, 49 (1992).
8. J. H. Adams *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7085 (1992).
9. The various gene fragments of EBA-175 were cloned into the Pvu II and Apa I sites of HSV gD in plasmid pRE4 (*20*). Cloning into these restriction sites resulted in constructs of chimeric EBA-175–HSV gD signal sequence and a COOH-terminal hydrophobic cytoplasmic tail, which allowed targeting of the malarial proteins to the surface of transfected COS-7 cells (*12*). The gene fragments encoding the various regions of EBA-175 were amplified with the polymerase chain reaction (PCR). Oligonucleotide primers used were 5′-ATCGATCAGCTGAAAGCTAGGAATGAATAT-3′ and 5′-ATCGATGGGCCCATTACTTAAAACT-TCGTT-3′ for amplification of region I; 5′-ATCGATCAGCTGGGAAGAAATACTTCATCT-3′ and 5′-ATCGATGGGCCCCGAAGTTTGTTCAT-TATT-3′ for amplification of region II; 5′-ATC-GATCAGCTGCCAGAAGTAAAGGATGTA-3′ and 5′-ATCGATGGGCCCTTTATTTATTTCTGCT-AA-3′ for amplification of regions III to V; and 5′-ATCGATCAGCTGAACAGAAATGATAGTA-CA-3′ and 5′-ATCGATGGGCCCTGAAAAAGCC-TCCTTTCT-3′ for amplification of region VI. The reverse primer for region F_1 was 5′-ATC-GATGGGCCCCTTAATTGTATTGTCGTT-3′ and the forward primer for region F_2 was 5′-ATCGATCAGCTGGAAAAGCGTGAACATATT-3′. The PCR products were size-fractionated on low-melt agarose, gel-purified with GENECLEAN II (BIO 101), restricted with Pvu II and Apa I, and ligated into pRE4. The 5′ and 3′ ends of each insert were sequenced, and an open reading frame read-through established. Fresh monolayers of COS-7 cells (American Type Culture Collection CRL 1651) cultured in Dulbecco's modified Eagle's medium and 10% fetal bovine serum (FBS) in 3.5-cm-diameter wells were transfected with 3 to 5 μg of Qiagen-purified, phenol chloroform–extracted plasmids with the use of calcium phosphate precipitation [F. L. Graham and A. J. van der Eb, *Virology* **52**, 456 (1973)].
10. To follow expression, we grew COS cells on cover slips in 3.5-cm-diameter wells, then transfected them and used them in immunofluorescence assays 40 to 60 hours after transfection. Immunofluorescence on 2% formaldehyde-fixed COS cells with HSV gD monoclonal antibodies ID3 and DL6 (*20*) (specific for amino acids 11 to 19 and amino acids 272 to 279 of HSV gD, respectively, contained in the resultant chimeric HSV gD–EBA-175 constructs) was performed by standard methods (*6*).
11. Rosetting assays were performed on COS cells 40 to 60 hours after transfection in 3.5-cm-diameter wells. Two hundred microliters of 10% hematocrit erythrocytes in media were added to the 2 ml of media in transfected wells, mixed, and incubated for 2 hours at 37°C. The COS cells were washed three times with phosphate-buffered saline to remove nonadherent erythrocytes.
12. C. E. Chitnis and L. H. Miller, *J. Exp. Med.*, in press.
13. D. J. Anstee, *Semin. Hematol.* **18**, 13 (1981); W. Dahr, K. Beyreuther, H. Steinbach, W. Gielen, J. Kruger, *Hoppe-Seyler's Z. Physiol. Chem.* **361**, 895 (1980).
14. M. Tomita, H. Furthmayr, V. T. Marchesi, *Biochemistry* **17**, 4756 (1978).
15. O. O. Blumenfeld and A. M. Adamany, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2727 (1978).
16. Metabolic labeling of EBA-175 was performed as described [J. D. Haynes *et al.*, *J. Exp. Med.* **167**, 1873 (1988)], with the exception that the label used was Trans-^{35}S (ICN Radiochemicals, Irvine, CA). Forty microliters of labeled EBA-175 (3.75×10^5 schizont equivalents/ml) was pre-incubated with glycophorin A or B or glycopeptides in a final volume of 100 μl with RPMI-1640 with 10% FBS for 1 hour at room temperature. This mixture was then added to 5×10^8 packed erythrocytes and rocked for 20 min at room temperature for binding to occur. SDS–polyacrylamide gel electrophoresis (SDS-PAGE) analysis was performed on eluates of erythrocytes eluted with 9 μl of RPMI, 1.5 M NaCl, 10% FBS, and 2 mM phenylmethylsulfonyl fluoride. Unenhanced gels were dried and quantitated with a Fuji phosphorescent image plate analyzer.
17. H. Furthmayr, *Nature* **271**, 519 (1978).
18. D. Sako *et al.*, *Cell* **75**, 1179 (1993).
19. O. O. Blumenfeld, A. M. Adamany, K. V. Puglia, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 747 (1981).
20. G. H. Cohen *et al.*, *J. Virol.* **62**, 1932 (1988).
21. L. Warren, *J. Biol. Chem.* **234**, 1971 (1959).
22. The NLIN procedure is in *SAS/STAT User's Guide Version 6* (SAS Institute, Cary, NC, ed. 2, 1989), vol. 2, pp. 1135–1193.
23. Methods for isolation, digestion, and purification of glycophorins and glycopeptides were as described [E. Lisowska, L. Messeter, M. Duk, M. Czerhinski, A. Lundblad, *Mol. Immunol.* **24**, 605 (1987); K. Wasniowska *et al.*, *Glycoconjugate J.* **2**, 163 (1985)]. Crude glycophorins were isolated by phenol extraction of membranes of human erythrocytes. Glycophorin A was purified by gel filtration in the presence of SDS, and column fractions containing bands corresponding to monomeric and dimeric glycophorin A were pooled, dialyzed against 50% ethanol to remove SDS, dialyzed against distilled water, and lyophilized. Column fractions enriched in glycophorin B were pooled for further purification of glycophorin B with the use of reverse-phase, high-pressure liquid chromatography (LKB, Bromme, Sweden). Tryptic and chymotryptic fragments of glycophorin A were obtained by digestion with TPCK-trypsin (Sigma) or TLCK-chymotrypsin (Sigma) and purified. The purity of products was analyzed by SDS-PAGE followed by PAS (periodate acid Schiff) staining and immunoblotting with monoclonal antibody to N (mAb 453) and showed clean single bands, with appropriate mobilities. Glycophorin A and the glycopeptides of glycophorin A contained no detectable contaminants. The tryptic peptide (MT1) contained two peptides (1–31 and 1–39), with 1–39 predominating. Glycophorin B, when probed with antibodies to N, in an immunoblot showed a band below the glycophorin A dimer that was probably a heterodimer of glycophorins A and B. The heterodimer was not detected on PAS-stained gels, which indicates a minimal contamination of glycophorin A in the glycophorin B preparation. Amounts of glycophorins and glycopeptides obtained were also analyzed by quantitative amino acid analysis (Table 2), and sialic acid contents were determined by the thiobarbituric acid method (*21*). Molalities of the glycopeptides and sialic acid were the expected concentration for each glycopeptide, which indicates that the glycopeptides of glycophorin A were unaltered during enzyme treatment and purification.
24. Erythrocytes for binding assays were washed three times with 10 volumes of RPMI-1640 before use. Enzymatic treatment of erythrocyte was performed as described with neuraminidase and trypsin (*3*) and peptide *N*-glycosidase (Oxford Glycosystems) (12 U) for 16 hours at 37°C.
25. This assay was identical to that described (*16*), except that the step for pre-incubation with inhibitor was omitted.
26. We thank M. Duk and E. Lisowska from the Institute of Immunology and Experimental Therapy in Poland for providing NCH3 and NT3 glycopeptides; G. Cohen and R. Eisenberg for plasmid pRE4 and antibodies ID3 and DL6; and D. Alling, National Institute of Allergy and Infectious Diseases, NIH, for statistical analysis of the inhibition curves.

23 February 1994; accepted 27 May 1994

Structure of the RGD Protein Decorsin: Conserved Motif and Distinct Function in Leech Proteins That Affect Blood Clotting

Andrzej M. Krezel, Gerhard Wagner,* Jana Seymour-Ulmer, Robert A. Lazarus*

The structure of the leech protein decorsin, a potent 39-residue antagonist of glycoprotein IIb-IIIa and inhibitor of platelet aggregation, was determined by nuclear magnetic resonance. In contrast to other disintegrins, the Arg-Gly-Asp (RGD)–containing region of decorsin is well defined. The three-dimensional structure of decorsin is similar to that of hirudin, an anticoagulant leech protein that potently inhibits thrombin. Amino acid sequence comparisons suggest that ornatin, another glycoprotein IIb-IIIa antagonist, and antistasin, a potent Factor Xa inhibitor and anticoagulant found in leeches, share the same structural motif. Although decorsin, hirudin, and antistasin all affect the blood clotting process and appear similar in structure, their mechanisms of action and epitopes important for binding to their respective targets are distinct.

Blood-sucking leeches are anatomically and physiologically adapted to their ecto-parasitic life-style (*1*). Leech salivary glands secrete several proteins that affect blood clotting, a complex process involving platelet adhesion and aggregation, the coagulation and fibrinolytic systems, and the endothelium (*2*). Hirudin, from the leech *Hirudo medicinalis*, potently inhibits thrombin, an important serine protease in the coagulation cascade (*3*). Other leech proteins that affect the hemostatic process include the bdellins, which inhibit plasmin (*4*);

A. M. Krezel and G. Wagner, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.
J. Seymour-Ulmer and R. A. Lazarus, Department of Protein Engineering, Genentech, South San Francisco, CA 94080, USA.

*To whom correspondence should be addressed.

antistasin, which inhibits Factor Xa (5); and hementin, which degrades fibrinogen and fibrin (1). Decorsin and ornatin, also from the leech, block fibrinogen binding to the platelet receptor glycoprotein IIb-IIIa (GP IIb-IIIa) and potently inhibit platelet aggregation (6, 7).

We present the solution structure of decorsin, a 39-residue protein from the leech *Macrobdella decora* (6, 8). It contains the RGD adhesion site recognition sequence, which is found in fibrinogen, fibronectin, vitronectin, and other GP IIb-IIIa ligands and is thought to be essential for binding to many integrins (9). Decorsin is related to the disintegrins, a family of RGD-containing GP IIb-IIIa antagonists from snake venom (10). Previous nuclear magnetic resonance (NMR) structural studies on the disintegrins kistrin, echistatin, and flavoridin have shown that the RGD sequence, which is the only epitope required for binding (11), lies at the apex of a conformationally ill-defined extended loop (12); similar results have been reported for the solution and x-ray structures of the RGD-containing tenth type III module of fibronectin (13). The RGD sequence in the x-ray structure of the fibronectin type III domain from tenascin is well defined; however, it is uncertain whether this constitutes a recognition site (14).

The structure of decorsin is well defined, except for the first three NH_2-terminal residues and some surface side chains (Fig. 1) (15). The average pairwise root-mean-square (rms) difference for the backbone atoms (N, C^{α}, C′) of residues 4 through 39 was 0.35 Å; the average rms difference from the mean coordinates was 0.24 Å (16). The dominant features of the decorsin structure are two β sheets, which are linked by three disulfide bonds. The three disulfides form the protein core, which does not contain any bulky hydrophobic side chains. The disulfide connectivities (Cys^7 to Cys^{15}, Cys^{17} to Cys^{27}, and Cys^{22} to Cys^{38}) were identified from nuclear Overhauser enhancements between methylene groups of the cysteines and were confirmed by calculations without explicit disulfide bond constraints. The final set of structures was calculated with disulfide bonds included as covalent bonds; their geometry is well-defined (17). The first β sheet is short and consists of three strands. Strands S2 (Cys^{15} to Cys^{17}) and S3 (Asp^{20} to Cys^{22}) form a hairpin; strand S1 (Pro^5 to Cys^7) forms one edge of the sheet and is parallel to S2. The second sheet is a two-stranded antiparallel hairpin consisting of strands S4 (Gln^{26} to Phe^{29}) and S5 (Pro^{36} to Glu^{39}). All reverse turns are well-defined. These include an atypical turn (residues 8 to 11), a type VIII turn (11 to 14) (18), a type II′ turn (17 to 20), a type II turn (23 to 26), and a

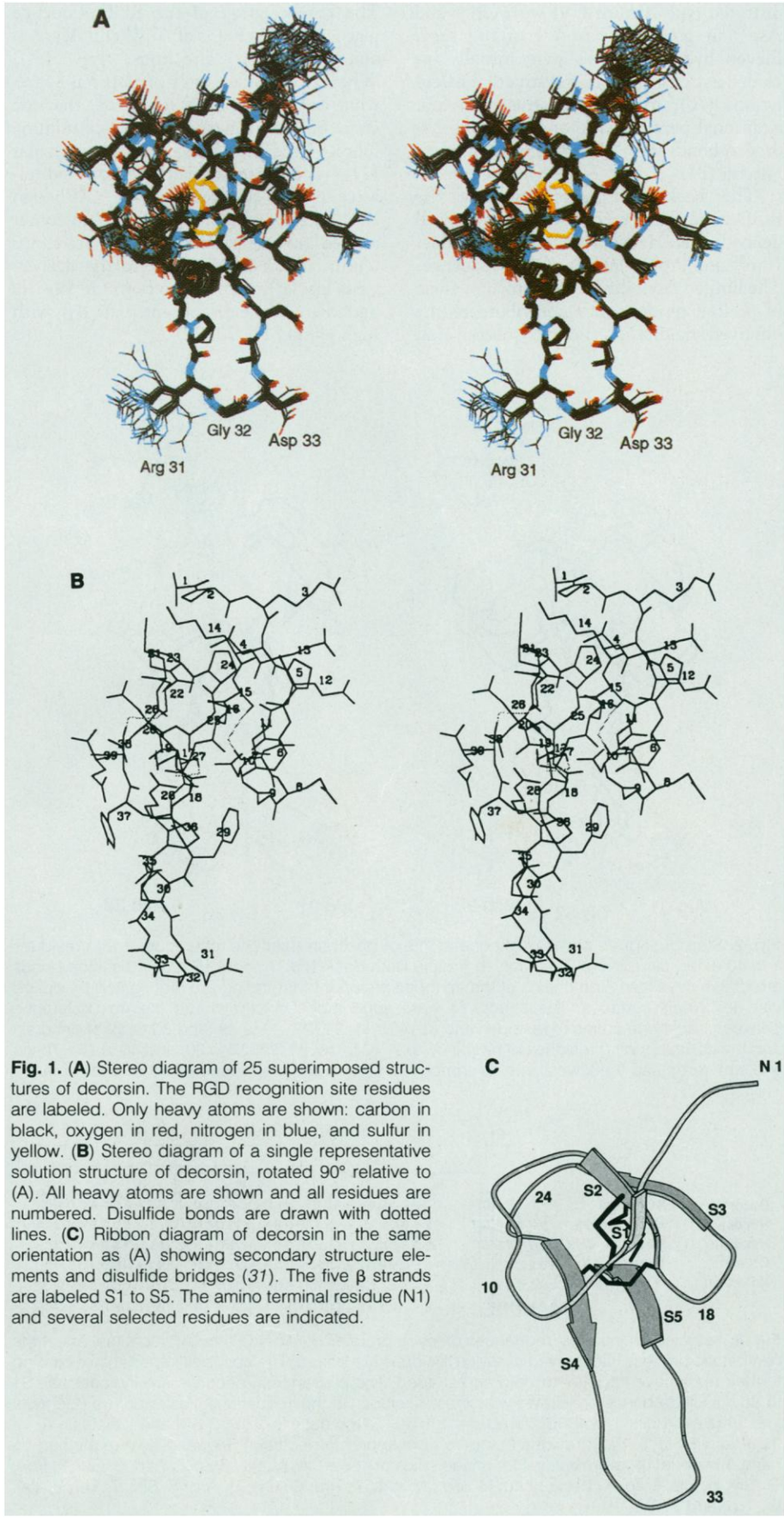

Fig. 1. (**A**) Stereo diagram of 25 superimposed structures of decorsin. The RGD recognition site residues are labeled. Only heavy atoms are shown: carbon in black, oxygen in red, nitrogen in blue, and sulfur in yellow. (**B**) Stereo diagram of a single representative solution structure of decorsin, rotated 90° relative to (A). All heavy atoms are shown and all residues are numbered. Disulfide bonds are drawn with dotted lines. (**C**) Ribbon diagram of decorsin in the same orientation as (A) showing secondary structure elements and disulfide bridges (31). The five β strands are labeled S1 to S5. The amino terminal residue (N1) and several selected residues are indicated.

distorted type II′ turn (*19*) with Gly^{32} and Asp^{33} in positions 2 and 3 of the turn. Eleven hydrogen bonds were initially included as constraints in the structure calculations (*20*); in the final structures, four additional peptide backbone-backbone hydrogen bonds were identified from distance criteria (*21*).

The backbone conformation of the RGD-containing recognition loop is well defined (Fig. 1A). Two proline residues (Pro^{30} and Pro^{36}) flank the RGD sequence. The limited flexibility of the φ torsion angle of proline residues may contribute to the observed rigidity of the recognition site. The conformation of the RGD sequence places the side chains of Arg^{31} and Asp^{33} in almost opposite directions (Fig. 1A). Whereas the side chain of Asp^{33} is somewhat conformationally restricted, the side chain of Arg^{31} is not. The RGD-containing loop of decorsin—and Asp^{33}, in particular (*11*)—is conformationally well defined relative to the disintegrins (*12*). Although this difference might be expected to confer altered binding affinity or specificity to integrins, decorsin is similar to the disintegrins insofar as it binds to both GP IIb-IIIa and the vitronectin receptor ($\alpha_v\beta_3$) with high affinity (*6*, *10*, *22*).

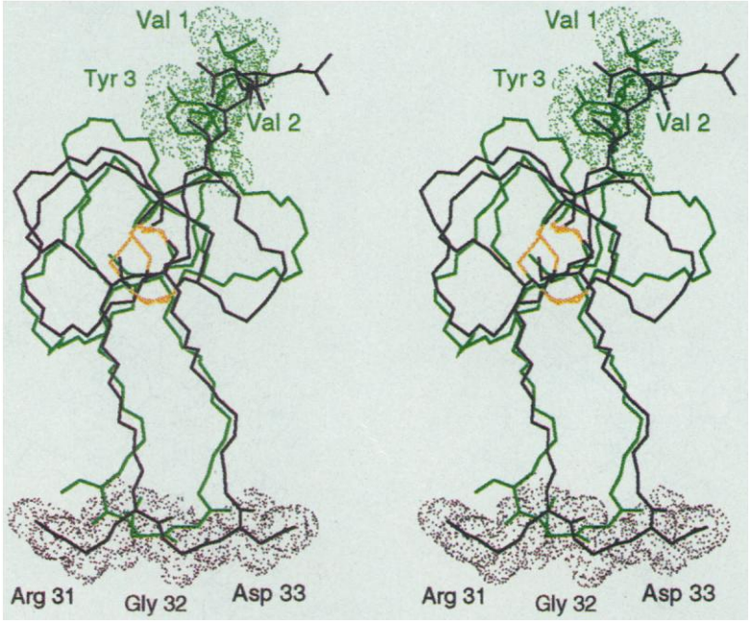

Fig. 2. Stereo diagram of the backbone atoms of decorsin (black) superimposed with residues 1 to 40 of hirudin (green) [Brookhaven Protein Data Bank 1HIC, model 15 (*23*)]. Disulfide bonds are drawn in yellow. Side chains of the epitopes important for binding are highlighted by dotted van der Waals surfaces. The structures were superimposed to minimize the rms difference between backbone atoms of residues 4 to 8, 14 to 17, 21, 22, 26 to 29, and 35 to 39 of decorsin and the corresponding residues of hirudin (3 to 7, 13 to 16, 21, 22, 27 to 30, and 36 to 40). These two structures had the lowest rms difference from the average structure of either ensemble.

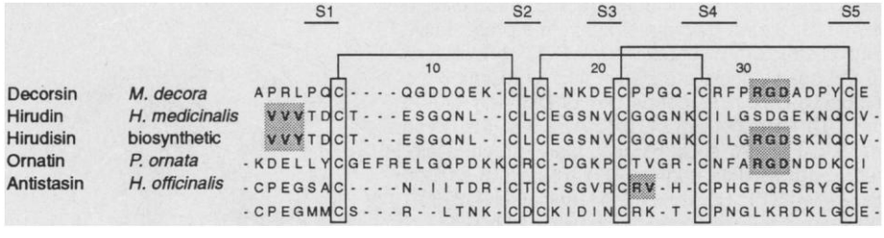

Fig. 3. Selected amino acid sequences of leech proteins containing the LAP motif that affect the hemostatic process. Conserved cysteine residues are boxed. The epitopes important for binding to their respective targets are bold and shaded. The β strands in decorsin are indicated as S1 to S5. Disulfide bonds are drawn with lines. Residues in the sequences illustrated are decorsin, 1 to 39 (*6*); hirudin, 1 to 40 (*3*); hirudisin, 1 to 40 (*29*); ornatin E, 7 to 50 (*7*); and antistasin A, 13 to 49 and 67 to 104 (*5*); residue numbers correspond to decorsin. Sequences were aligned by hand. Single-letter abbreviations for amino acids are used: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

After the structure was solved, we discovered that the fold of decorsin closely resembles the fold of the first 40 residues of hirudin (Fig. 2) (*23–25*). The sequence identity between decorsin and any of the hirudin sequences does not exceed 24% (*26*); cysteines alone account for 15% of the identity (Fig. 3). The amino acid composition also differs; decorsin has more aspartic and arginine residues, as well as six proline residues, that are completely absent among the first 40 residues of hirudin. The key to structural similarity is the spacing of the cysteine residues forming the disulfide bonds. In addition, the presence of residues facilitating the formation of the reverse turns (Gly and Asn) and some of the bulky hydrophobic side chains are conserved. Residues 8 to 12 of hirudin form a short loop, whereas corresponding residues 9 to 13 in decorsin form two consecutive reverse turns, accommodating the partially buried side chain of Asp^{11}. The decorsin and hirudin structures also differ in the location of the type II′ and type II reverse turns formed by residues 17 to 20 and residues 23 to 26, respectively. The two β sheets and the overall conformation of the hairpin loop bearing the RGD sequence in decorsin are similar to those in hirudin (*23*, *24*). However, the loop region in decorsin (residues 30 to 36) is relatively well defined in contrast to that of hirudin (*23*). The average rms difference between the backbone atoms (N, C^{α}, and C′) of the core residues of decorsin and hirudin was 0.93 Å (Fig. 2).

Because the decorsin and hirudin structural similarity involves the spacing of the six cysteine residues, we searched the Dayhoff protein database for the following motif

$$\text{Cys-X}_{6-12}\text{-Cys-X-Cys-X}_{3-6}\text{-}$$
$$\text{Cys-X}_{3-6}\text{-Cys-X}_{8-14}\text{-Cys}$$

where X_n is a string of any n residues except cysteine. A total of 181 sequences out of the 140,192 sequences searched fit this criterion. Of these, 3 were decorsin, 150 were hirudin, 9 were ornatin, and 9 were antistasin (*27*). Ornatin is a potent GP IIb-IIIa antagonist isolated from leech *Placobdella ornata* (*7*). Several isoforms exist, which share from 34 to 42% sequence identity with decorsin. The position of the RGD sequence in ornatin is the same as in decorsin (Fig. 3). The sequence and functional similarity of ornatin and decorsin suggest that their structures are also similar.

Of particular interest is antistasin, a 119-residue, cysteine-rich protein from the leech *Haementaria officinalis*. Antistasin is a potent anticoagulant that stoichiometrically and selectively inhibits Factor

Xa, another serine protease in the coagulation cascade (*5*). Antistasin contains a twofold internal repeat that represents two closely related disulfide-bonded structures (*5*). On the basis of the conservation of cysteines and some of the key turn residues (Fig. 3), we speculate that each domain of antistasin shares a similar fold to that of decorsin and hirudin (*28*). We propose the term "LAP motif" (leech antihemostatic protein) to refer to the general tertiary fold of this protein family.

Although decorsin, hirudin, and antistasin are all from leeches and all affect the hemostatic process, their sequences, activities, and binding epitopes differ. Hirudin forms a tight complex with thrombin, having numerous significant binding contacts: residues 1 to 3 bind at the active site and 9 out of the final 11 COOH-terminal residues have interactions with the exosite (*24*). Hirudin does not contain an RGD sequence and has no binding affinity for GP IIb-IIIa; only inhibition of thrombin-induced platelet aggregation is observed (*29*). The epitopes in hirudin that are necessary for activity either have been replaced by different side chains or are missing altogether in decorsin, which does not inhibit thrombin (*22*). Antistasin is a slow tight-binding inhibitor of Factor Xa and inhibits by the standard mechanism common to many serine protease inhibitors (*30*). The P_1 active site residue in antistasin is Arg^{34} (*5*), which corresponds to Pro^{23} in decorsin and Gly^{23} in hirudin (Fig. 3). Antistasin does not inhibit thrombin (*5*); no reports of any GP IIb-IIIa antagonist activity exist. Furthermore, neither hirudin nor decorsin inhibit Factor Xa (*3*, *22*).

The RGD sequence was recently introduced into hirudin, producing chimeric proteins termed hirudisins (*29*). However, their activity in adenosine diphosphate–induced platelet aggregation assays was about one-hundreth the activity of decorsin. Possible explanations for poor activity include suboptimal conformation of the recognition site loop and the presence of a lysine residue in position 36 of the hirudisin sequence (Fig. 3), a position that generally is Asp in decorsin and ornatin isoforms or flanked by Asp in most disintegrins (*7*, *10*, *29*).

The LAP motif apparently provides a framework for entirely different, yet highly specific activities of leech salivary proteins that potently affect hemostasis. Proteins that share the LAP motif likely arose as a result of divergent evolution. An interesting class of leeches to be studied in this respect would be the genus *Acanthobdella*, which are considered living fossils. The isoforms of hirudin, antistasin, and ornatin may result from multiple alleles present in a single leech or from multiple genes in the leech population (*3*, *5*, *7*, *26*).

It is not unexpected that hematophagous leeches might possess multiple proteins directed toward different targets to prevent clot formation. However, that various species of leeches would use the same protein scaffold but different binding epitopes and diverse mechanisms to affect hemostasis is notable.

REFERENCES AND NOTES

1. R. T. Sawyer, *Leech Biology and Behaviour* (Clarendon, Oxford, 1986); *Biotechnology* **9**, 513 (1991); N. U. Bang and M. D. Clayman, *Trends Cardiovasc. Med.* **2**, 183 (1992).
2. R. W. Colman, V. J. Marder, E. W. Salzman, J. Hirsh, in *Hemostasis and Thrombosis*, R. W. Colman, J. Hirsh, V. J. Marder, E. W. Salzman, Eds. (Lippincott, Philadelphia, PA, ed. 3, 1994), pp. 3–18.
3. F. Markwardt, *Methods Enzymol.* **19**, 924 (1970); S. R. Stone and J. Hofsteenge, *Biochemistry* **25**, 4622 (1986); J. W. Fenton II, *Ann. N.Y. Acad. Sci.* **370**, 468 (1981); *ibid.* **485**, 5 (1986); F. Markwardt, *Thromb. Res.* **74**, 1 (1994).
4. U. Seemüller, J. Dodt, E. Fink, H. Fritz, in *Proteinase Inhibitors*, A. J. Barrett and G. Salvesen, Eds. (Elsevier, Amsterdam, 1986), pp. 337–359.
5. G. P. Tuszynski, T. B. Gasic, G. J. Gasic, *J. Biol. Chem.* **262**, 9718 (1987); E. Nutt *et al.*, *ibid.* **263**, 10162 (1988); C. Dunwiddie *et al.*, *ibid.* **264**, 16694 (1989).
6. J. L. Seymour *et al.*, *ibid.* **265**, 10143 (1990).
7. P. Mazur, W. J. Henzel, J. L. Seymour, R. A. Lazarus, *Eur. J. Biochem.* **202**, 1073 (1991).
8. Decorsin was expressed in *Escherichia coli* with a synthetic gene under control of the alkaline phosphatase promoter and was purified from the periplasm by size-exclusion filtration and rpHPLC (A. M. Krezel, J. Seymour-Ulmer, R. A. Lazarus, G. Wagner, in preparation).
9. E. Ruoslahti and M. D. Pierschbacher, *Science* **238**, 491 (1987); R. O. Hynes, *Cell* **69**, 11 (1992).
10. R. A. Lazarus and R. S. McDowell, *Curr. Opin. Biotechnol.* **4**, 438 (1993); M. S. Dennis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2471 (1990); R. J. Gould *et al.*, *Proc. Soc. Exp. Biol. Med.* **195**, 168 (1990).
11. M. S. Dennis, P. Carter, R. A. Lazarus, *Proteins Struct. Funct. Genet.* **15**, 312 (1993).
12. M. Adler *et al.*, *Science* **253**, 445 (1991); V. Saudek *et al.*, *Biochemistry* **30**, 7369 (1991); H. Senn and W. Klaus, *J. Mol. Biol.* **232**, 907 (1993).
13. A. L. Main *et al.*, *Cell* **71**, 671 (1992); C. D. Dickinson *et al.*, *J. Mol. Biol.* **236**, 1079 (1994).
14. D. J. Leahy *et al.*, *Science* **258**, 987 (1992).
15. Distance constraints were derived from nuclear Overhauser effect spectroscopy (NOESY) experiments with mixing times up to 100 ms. The NMR experiments were performed at 11.7 and 14.2 T, 25°C, and pH 4.7 with 5 mM decorsin in a 9:1 mix of H_2O and D_2O or 100% D_2O. Structures were calculated with the NMR_refine module [NMRchitect, version 2.0; Biosym Technologies, San Diego (1992)] with 129 intraresidue, 137 sequential, 43 medium-range, and 139 long-range distance constraints. Embedded structures were optimized with a simulated annealing protocol [M. Nilges, M. G. Clore, A. M. Gronenborn, *FEBS Lett.* **229**, 317 (1988)] and energy minimized with distance constraints. The force constants used for experimental restraints were set to 10 kcal/mol. A detailed description of the resonance assignments and structure calculations will be published elsewhere.
16. The average rms deviation for residues 4 to 39 in 25 structures superimposed pairwise, incorporating all heavy atoms, was 1.10 Å with a standard deviation of 0.16 Å. The average rms deviation from the mean structure was 0.77 Å with a standard deviation of 0.10 Å. The 25 structures had on average 14 violations of distance constraints above 0.1 Å; no violation was greater than 0.18 Å. The final energies of all the structures were between 363 and 380 kcal/mol, and the total forcing potential contributions were between 10 and 14 kcal/mol.
17. The average disulfide bond torsion angles χ^2, χ^3, and $\chi^{2'}$, respectively, were −44°, 96°, and 81° for 21 structures (4 structures showed values 101°, −87°, and 177°) for Cys^7 and Cys^{15}; 73°, 82°, and 57° for Cys^{17} and Cys^{27}; and 76°, 77°, and 75° for Cys^{22} and Cys^{38}. The variation of disulfide bond torsion angles was smaller than ±8°.
18. P. Y. Chou and G. D. Fasman, *J. Mol. Biol.* **115**, 135 (1977); J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981); C. M. Wilmot and J. M. Thornton, *J. Mol. Biol.* **203**, 221 (1988).
19. Ideal torsion angle values for a type II′ reverse turn are $\varphi_{i+1} = 60°$, $\Psi_{i+1} = -120°$, $\varphi_{i+2} = -80°$, $\Psi_{i+2} = 0°$. Values for the RGD sequence in decorsin are $\varphi_{31} = -80 \pm 7°$, $\Psi_{31} = 171 \pm 16°$, $\varphi_{32} = 78 \pm 4°$, $\Psi_{32} = -116 \pm 10°$, $\varphi_{33} = -120 \pm 16°$, and $\Psi_{33} = 70 \pm 6°$. Values for the RGD residues in tenascin åre $\varphi_{877} = -158°$, $\Psi_{877} = 100°$, $\varphi_{878} = 63°$, $\Psi_{878} = -119°$, $\varphi_{879} = -75°$, and $\Psi_{879} = -25°$ (*14*).
20. The following hydrogen bonds, used as constraints, were identified from the amide exchange, pH titration, and NOESY data: Cys^{15} H^N–Cys^{22} O′, Cys^{17} H^N–Asp^{20} O′, Asp^{20} H^N–Cys^{17} O′, Cys^{22} H^N–Cys^{15} O′, Arg^{28} H^N–Tyr^{37} O′, Ala^{34} H^N–Arg^{31} O′, Tyr^{37} H^N–Arg^{28} O′, Glu^{39} H^N–Gln^{26} O′, Gln^8 H^N–Asp^{11} O^δ, Gly^9 H^N–Asp^{11} O^δ, and Arg^{28} H^ϵ–Glu^{39} $O^{\epsilon 1}$.
21. The criteria used for postulating hydrogen bonds in calculated structures were (i) the average distance between donor hydrogen and acceptor oxygen was less than 3.3 Å, (ii) the average dihedral angle for the four atoms involved was between 120° and 240°, and (iii) the hydrogen was exchanging slowly in D_2O solution. The following backbone hydrogen bonds met those criteria: Cys^7 H^N–Leu^{16} O′, Leu^{16} H^N–Pro^5 O′, Gln^{26} H^N–Pro^{23} O′, and Cys^{27} H^N–Gly^{25} O′.
22. J. Seymour-Ulmer and R. A. Lazarus, unpublished results.
23. G. M. Clore *et al.*, *EMBO J.* **6**, 529 (1987); P. J. M. Folkers *et al.*, *Biochemistry* **28**, 2601 (1989); H. Haruyama and K. Wüthrich, *ibid.*, p. 4301; T. Szyperski *et al.*, *J. Mol. Biol.* **228**, 1193 (1992); T. Szyperski *et al.*, *ibid.*, p. 1206.
24. M. G. Grütter *et al.*, *EMBO J.* **9**, 2361 (1990); T. J. Rydel *et al.*, *Science* **249**, 277 (1990); T. J. Rydel *et al.*, *J. Mol. Biol.* **221**, 583 (1991); J. Vitali *et al.*, *J. Biol. Chem.* **267**, 17670 (1992).
25. The database of families of structurally similar proteins contains no proteins with folds homologous to hirudin [L. Holm, C. Ouzounis, C. Sander, G. Tuparev, G. Vriend, *Protein Sci.* **1**, 1691 (1992)].
26. J. Dodt *et al.*, *FEBS Lett.* **165**, 180 (1984); V. Steiner *et al.*, *Biochemistry* **31**, 2294 (1992).
27. These include natural isoforms (*26*) as well as synthetic variants in the case of hirudin.
28. A protein homologous to antistasin termed hirustasin, isolated from *H. medicinalis*, was reported during the preparation of this manuscript. Hirustasin is not an anticoagulant; however, it inhibits cathepsin G and tissue kallikrein and possesses the LAP motif [C. Söllner, R. Mentele, C. Eckerskorn, H. Fritz, C. P. Sommerhoff, *Eur. J. Biochem.* **219**, 937 (1994)].
29. A. Knapp, T. Degenhardt, J. Dodt, *J. Biol. Chem.* **267**, 24230 (1992).
30. M. Laskowski Jr. and I. Kato, *Annu. Rev. Biochem.* **49**, 593 (1980).
31. Figure 1C was produced with the program Molscript [P. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991)].

32. This work was supported in part by the National Institutes of Health and the Keck Foundation.

7 March 1994; accepted 16 May 1994