Vol. 194, No. 1, 1993 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1993 Pages 560-565

A Secondary Structure Prediction of the Hemorrhagic Metalloprotease Family

Dietlind L. Gerloff, Thomas F. Jenny, Lukas J. Knecht, and Steven A. Benner

Laboratory for Organic Chemistry and Institute for Scientific Computation
E.T.H. Zurich, CH-8092 Switzerland

Received June 7, 1993

Summary: A secondary structure has been predicted for the hemorrhagic metalloproteases using a method developed in Zurich that extracts structural information from patterns of conservation and variation in homologous protein sequences. This prediction tests the limits of the method when applied to a small number of homologous sequences that have undergone only modest evolutionary divergence. Predictions were also obtained using a neural network developed by Sander and coworkers, to date the best fully automated method for predicting secondary structure, and using the classical Chou-Fasman and GOR heuristics. The predictions are different. No crystal structure is known within this protein family, but one is expected shortly. Therefore, this prediction should contribute significantly to the evaluation of the relative merits of these prediction methods.

A method developed at the E.T.H. in Zurich has recently been used to make *bona fide* predictions of various aspects of the conformation of a number of families of proteins [1,2,3,4,5]. This approach has been controversial. To some, the predictions have been "remarkably accurate" [6], finding "core secondary structures much better than ... achieved with standard methods" and providing a "possibility of extending prediction to a tertiary fold [that] is much better" than the per residue score might imply [7], and represent "a spectacular achievement ... that will come to be recognized as a major breakthrough." [8] Others have regarded these comments as "exaggerated" [9], cautioned that "one swallow does not make a summer" [10], questioned the value of *bona fide* predictions in any case, and suggested that because all prediction work is fundamentally statistical, any method that approaches proteins as individual cases violates important rules of procedure in the structure prediction field [10]

Elsewhere, we have made the case that the most productive approach to understanding conformation in proteins is likely to follow procedures that organic chemists have used to address conformational issues in organic molecules generally [1-5, 11]. We recognize that this case will not be universally convincing. Therefore, we present here the first paper where secondary structure predictions by competing methods (the E.T.H. method, a neural network method [12,13], and classical procedures such as Chou-Fasman [14] and GOR [15]) are made for a family of proteins where no crystal structure is available, but where one is imminent,.

For this test, we have chosen the hemorrhagic metalloproteases from snake venom [16], kindly brought to our attention by Prof. Edgar Meyer (Texas A & M University). A crystal structure of a representative of this protein is imminent. Further, the protein family contains

0006-291X/93 $4.00

only 7 sequences in two subfamilies, with a maximum evolutionary distance of only 78 accepted point mutations per 100 amino acids (PAM units)[17]. Therefore, this protein family appears ideal for testing the scope of various methods that build consensus prediction with protein families having only few sequences and relatively little evolutionary divergence. This contrasts with models for families used to predict structures for protein kinase [1], the SH3 domain [2], the MoFe nitrogenase protein [3], and others [4], where at least 10 and often over 50 sequences were used as input, with rather large sequence divergence in several cases (e.g., for the MoFe nitrogenase, the tree was 174 PAM units wide).

MATERIALS AND METHODS

The E.T.H. prediction was made by assigning surface, interior, active site, and parsing residues in the alignment [1-4]. The first three assignments convey tertiary structural information. An unrefined consensus secondary structure is assigned to the alignment based on an analysis of patterns in these assignments, together with special effort to accommodate secondary structures that do not readily reveal themselves by such methods (e.g., internal alpha helices) [1]. Predictions using the neural network developed by Sander and coworkers [12,13] were obtained through a server maintained by these workers in Heidelberg. A consensus prediction for the protein family was kindly provided by B. Rost and C. Sander. Chou-Fasman [14] and GOR [15] predictions were obtained routinely. A consensus prediction was obtained in each following a procedure similar to that introduced by Kirschner *et al.* [18] in their prediction of the tryptophan synthase.

RESULTS AND DISCUSSION

The predictions are shown in Figure 1, with the consensus predictions derived in each method shown in bold. Consensus predictions assume that homologous proteins have similar secondary structures [19]. Given the relatively small overall sequence divergence in this family, this assumption is highly plausible. However a consensus prediction does not correspond exactly to any individual member of the family (although the structure of each member might in principle be predicted from a consensus prediction by homology modelling).

Predictions made by the different methods are different, often markedly so. Indeed, all four methods predict the same secondary structure at only 37% of the residues. Thus, an experimental structure should not only permit evaluation of the relative merits of the prediction methods (at least in this test case), but also indicate whether more reliable predictions might be obtained from a composite prediction method involving all four methods.

Regarding supersecondary and tertiary structure, the small number of sequences in the multiple alignment makes it difficult to assemble a complete model for the protein family from the secondary structure prediction made by the E.T.H. method, primarily because sequence pairs at an evolutionary distance appropriate for performing covariation analysis are not available [1]. Nevertheless, a surface antiparallel beta sheet including strands in positions 50-68 is strongly indicated, flanked by at least one additional strand in position ca. 95-104. Also indicated is an active site flanked by three helices, one containing putative zinc binding ligands (positions 133-145) and two containing putative substrate binding units (positions 33-48, and positions 175-193). Interestingly, the active site helix, assigned using rules developed following the incorrect assignment of an internal helix in protein kinase [1], is suggested by analogy with thermolysin [20] and astacin [21], two well studied metalloproteases dependent on

No.	Sequences	ETH Prediction					Neural Net			GOR			Chou-Fasman		
	abcd efg	1	2	3	4	5	abcd	efg	C	abcd	efg	C	abcd	efg	C
001	QQ-Q -Q-	?								.-.	-.-	-	..-.	-.-	-
002	RRRR NN-	fil spl	A							...	..-	-	....	..-	-
003	FFFF LL-	inside	A							....	..-	-	....	..-	-
004	PPPP PP-	P		A		p				T..T	..-	-	...t	..-	-
005	QQQR QQQ	surf		A						T..T	...	.	tttt	Tt.	T
006	RRRR SRK	surf		A						T..T	.BB		tttt	Tt.	T
007	YYYY YYY	inside		A			BBBB	BBB	B	T..T	BBB		BBBB	.B.	B
008	IIII III	inside	A		b	b	BBBB	BBB	B	T..T	BBB		BBBB	.B.	B
009	KEEK EEE	surf	A		b	b	BBBB	BBB	B	B..B	BBB	B	BBBB	.B.	B
010	LLLL LLL	inside	A		B	B	BBBB	BBB	B	B..B	BBB	B	BBBB	BBB	B
011	GAAA VVV	inside	A		B	B	BBBB	BBB	B	B..B	BBB	B	BBBB	BBB	B
012	IIII VVV	inside	A		B	B	BBBB	BBB	B	B..B	ABA		BBBB	BBB	B
013	FVVV VVV	inside			B	B	BBBB	BBB	B	BBBB	AAA	B	BBBB	BBB	B
014	VVVV AAA	inside			B	B	BBBB	BBB	B	BBBA	AAA	A	BBBB	BBB	B
015	DDDD DDD	D				$	BBBB	BBB	B	BBBA	AAA	A	tttB	A.t	T
016	HHHH HHH	H				$				BBBA	AAA	A	tttB	A.t	T
017	GGGG RRG	surf	A			A		BBB		TTTA	AAA	A	...B	AB.	.
018	MMMI MVM	INSIDE	A			A	BBBB	BBB	B	TTTA	AAA	A	BBbB	ABB	B
019	YYYV FFF	inside	A			A	BBBB	BBB	B	TTTA	AAA	A	BBbB	ABB	B
020	TTKT MMT	amphi	A	s		A	BBBB	BBB	B	TTTA	ATT	T	BBbB	ABB	B
021	KKKK KKK	K	A	s	b	$ A	BBBB	BBB	B	TTTA	TTT	T	BBb.	ABB	B
022	YYYH YYY	inside	A		b	A	BBBB	BBB	B	TTTA	TTT	T	BBb.	.BB	B
023	SSNH NNN	surface	A		b	p A				.TTA	.TT	T	Ttbt	.Bt	T
024	GSQG SSG	Hb var	A		b	p A				.TT.	..T	.	TTbT	TTT	T
025	NNNN DDN	surface	A		b	p A				.AT.	...	.	TT.T	TTT	T
026	SFSL LLL	inside	A		b	p A				.AT.	...	.	taTa	BBB	
027	EKDK NNN	surface	A		b	A	A A	AAA	A	.ATT	...	.	taTa	BBB	
028	RKKK TTT	surface	A			A	AAAA	AAA	A	AATT	..B		Ba.a	BBB	B
029	IIII III	inside	A			A	AAAA	AAA	A	AABT	BBB	B	BaBa	BBB	B
030	TRKR RRR	surface	A			A	AAAA	AAA	A	AABB	BBB	B	BaBa	BBB	B
031	KKVK TTT	surface	A			A	AAAA	AAA	A	AABB	BBB	B	BaBa	BBB	B
032	RRRW RRR	surface	AA				AAAA	AAA	A	AABB	BBB	B	BaBB	BBB	B
033	VVVI VVV	inside	AA			A	BBAA	AAA	A	ABBB	BBB	B	BBBB	BBB	B
034	HHHY HHH	?		A-		A	BBAA	AAA	A	ABBB	BBB	B	BBBB	BBB	B
035	QQQQ EEE	fil spl	A			A	BBAA	AAA	A	.BBB	BBB	B	BBBB	BBB	B
036	MMML III	inside	A			A	AAAA	AAA	A	.BBB	BBB	B	BBBB	BBB	B
037	IVVV VVV	inside	A			A	AAAA	AAA	A	T.BB	BBB	B	BBBB	BBB	B
038	NNNN NNN	N	A			$ A	AAAA	AAA	A	T...	.BB		BBBB	BBB	B
039	NNHT FFT	amphi	A			p A	AAAA	AAA	A	T...	.B.		BBBB	BBB	B
040	IIII IIL	inside	A			A	AAAA	AAA	A	T..B	.B.		BBBB	BB.	B
041	NNNN NNN	N	A			$ A	AAAA	AAA	A	...B	.TT		.BBB	BB.	B
042	MEEN EGG	SURFACE	A			p A	AAAA	AAA	A	.TTB	TTT	T	A..B	Btt	
043	MMMI FFF	inside	A			A	AAAA	AAA	A	BTTB	TTT	T	A..B	Btt	
044	CYYY YYY	inside	AA			A	AA	AA	A	BTTB	TTB		A..B	B..	.
045	RRRR RRR	R	AA			$ A				B..T	TTB		A..B	B..	.
046	APPS SSS	Hb var	AA							B..T	T.B		A...	B..	.
047	LLLL LLL	inside	AA				BBBB	BBB		B..B	TTB		ABBB	BBB	B
048	NNNN NNN	N	AA			$	BBBB	BBB	B	BBBB	TTB	B	.BBB	BBB	B
049	IIII III	inside	A	B		b	BBBB	BBB	B	BBBB	..B	B	BBBB	BBB	B
050	VAAL RHL	surface	A	B		b	BBBB	BBB	B	BBBB	BBB	B	BBBB	BBB	B
051	TIIV VVI	INSIDE	A	B		b	BBBB	BBB	B	BBBB	BBB	B	BBBB	BBB	B
052	TTSA SSS	Hb var	A	B		b	BBBB	BBB	B	BBBB	BBB	B	BBBB	.aa	B
053	LLLL LLL	inside	A	B		b	BBBB	BBB	B	BBBB	BBB	B	BBBB	.aa	B
054	SSNV TTT	surface	A	B		b				BBBB	BBB	B	BBBB	.aa	B
055	VLRY DDD	surface			B	B	BBBB	BBB	B	BBBA	.BB	B	BBBB	.aa	B
056	LLLL LLL	inside			B	B	BBBB	BBB	B	AA.A	...		BBBB	.aa	B
057	EDQE EEE	surface			B	B	BBBB	BBB	B	AATA	T.T		ABBa	.aa	A
058	IVII III	inside			B	B	BBBB	BBB	B	AATA	T.T		ABBa	.aa	A
059	WWWW WWW	inside			b		BBBB	BBB	B	AATA	T.T		ABBa	.aa	A
060	SSSS SSS	S				$	BBBB	BBB	B	AA.A	...		AA.a	TtT	

Figure 1. A multiple alignment for seven hemorrhagic metalloproteases (alignment position numbers and sequences shown vertically, one letter code used, underlining indicates parsing string): Accession numbers (SwissProt): (a) P15503; (b) P14530; (c) P20165; (d) P20164; (e) P20897; (f) P15167; (g) P22796, followed by secondary structure predictions. "A", "a", "B", "b", "T" and "t" indicate strong and weak alpha helix, beta strand, and turn assignments, respectively. A "?" indicates that an assignment is uncertain. **E.T.H. prediction** [1-5]. Column 1 shows surface/interior assignments made using the E.T.H. method (underlining indicates consecutive surface assignments indicative of breaks in secondary structure); "fil spl": hydrophilic split; "amphi": amphiphilic position separating two branches in tree; "Hb var": hydrogen bonding variable. Column 2 shows possible alpha helix assignments made using the

```
061 EEKK DNN SURFACE          p                    AATT T.T T   AAta TtT T
062 KKKQ QEQ SURFACE                               AATT T.T T   AAta TTT T
063 DDDN DDD surface                               AATT TTT T   AABB TTT
064 LLLK FQL amphi        B     B           B      AABT BT.     BABB BBB B
065 IIII III inside       B     B   BBBB BBB B     AABB BT.     BABB BBB B
066 TTTT TNN surface      B     B   BBBB BBB B     AABB B..     BABB BBB B
067 VMVV VIV INSIDE       B     B   BBBB BBB B     AABB B..     BABB BBB B
068 QQKQ QQQ surface      B     B   BBBB BBB B     AABB ...     BA.B BBB B
069 AASS SSS Hb var       B  p  B   BBBB BBB B     .A.B ... .   BA.. ... .
070 SVAA SAA Hb var  a    b  p                     .... ... .   .Att .t.
071 AASS ASA Hb var  a    b  p                     .... ... .   .Att .t.
072 PPNN KSN SURFACE A    b  p  A   A A  A   A     .... ... .   .t.B tT.
073 TTVV NDD surface A       p  A   AAAA AAA A     .BBB ... .   BtaB tT.
074 TTTT TTT T       A       $  A   AAAA AAA A     BBBB ... B   BbaB bbB B
075 LALL LLL inside  A    B     A   AAAA AAA A     BBBB ... B   BbaB bbB B
076 TRED HNK SURFACE A    B     A   AAAA AAA A     BBBB ... B   BbaB bbB B
077 LLSL SAT Hb var  A    B     A   AAAA AAA A     BB.B .A.     BbaB bbB B
078 FFFF FFF inside  A    B     A   AAAA AAA A     .... .A. .   BbaB bbB B
079 GGGG GAG inside  A    B  p  A   AAA  AA  A     .... .A. .   .tTt .Aa
080 ADND EEE surface A    B  p  A   AAAA AAA A     .... AAA .   AtTt aAa A
081 WWWW WWW inside  A    B     A   AAAA AAA A     ...A AAA A   Aaaa aAa A
082 RRRR RRR R       A       $  A   AAAA AAA A     .A.A AAA A   Aaaa aAa A
083 EEEE KEE surface A          A   AAAA AAA A     .A.A AAA A   Aaaa aAa A
084 TTTS STR SURFACE A          A   AAAA AAA A     .A.A AAA A   Aaaa aAa A
085 VVVV VDV surface a i b                AAAA AAA A     .A.A AAA A   Aaaa aAB A
086 LLLL LLL inside  a    b               AAAA AAA A     .A.A AAA A   Aaaa aAB A
087 LLLL LLL inside       b               AAAA AAA A     .ATA AA. A   Aaaa aAB A
088 NKKK NNN surface      b  p              AAA AAA      TATA TAT T   .aaa ..B
089 RQQQ RRR surface          p                         TATT TTT T   taaa ttB
090 TKQR KKI SURFACE          p                         TATT TTT T   ttta ttB T
091 SDNS RSS SURFACE          p                         .ATT T..     ttta ...
092 HHNH HHH surface          p                         .ATT T..     taTt .tt T
093 DDDD DDD D            b  $                         .ATT T..     .aTt TTT T
094 HHCC NNN amphi        b  p              AA   AA A   .ATT .A.     .atB TTT T
095 AAAA AAA inside  n    B     B   AAAA AAA A     .ABT .A.     BaaB BAA A
096 QQHQ QQQ ??      o    B     B   AAAA AAA A     BABB BAB B   BaaB BAA A
097 LLLL LLL inside       B     B   AAAA AAA A     BABB BAB B   BaaB BAA A
098 LLLL LLL inside  3    B     B   BBBB BBB B     BBBB BAB B   BaaB BAA A
099 TTTT TTT T       .    B  $  B   BBBB BBB B     BBBB BAB B   BaaB BAA A
100 ADAT AAA surface 6    B        B  BBBB BBB B     BBBB BAB B   BaaB BAA A
101 TITI III inside       B        B  BBBB BBB B     BB.B BAB B   BaaB BAA A
102 INND VED SURFACE      B        B  BBBB BBB B     B..T BAB     B..B BAA
103 FFLF LLL inside       B        B                  B..T BAB     B..B BAA
104 NTND DDA SURFACE A    B  p     B                  T..T BA.     tttt TAA T
105 GGDG DED SURFACE A       p                        T... BA. .   ttTt TAt T
106 NNNP YEN SURFACE A       p                        .... BA. .   ttTt tAt T
107 VTTT TTT inside  A    B     ?                     .... BA. .   .BBt bbB B
108 IIII LLI inside  A    B     ?   BBBB BBB B     .... BAB .   .BB. bbB B
109 GGGG GGG G       A    B  p  ?   BBBB BBB B     T... BAB     .BB. bbB B
110 RWLK LLI amphi   A    B     ?   BBBB BBB B     T..B BAB     .BB. bbB B
111 AAAA AAA inside  A    B     ?   BBBB BBB B     ..TB B.B     .BB. bbB B
112 PYYY YPY inside  A    B  p  ?   BBBB BBB B     TTTB T.B T   tBBb btB B
113 VMKT LLT SURFACE A                   BBBB BBB B     TTTB TTT T   tBtb btB B
114 GGKA NGG SURFACE a          p                    TTTB TTT T   t.tb b.t
115 GGGS STG Hb var  a          p                    TTTB TBT T   ...b ..t .
116 MMMM MMM inside                      BB  B      TTTB .BT T   tttb ..t T
117 CCCC CCC C               $                       .TTB .BT     tttb ..t T
118 DNND HDY SURFACE          p                      .... .B. .   .... ... .
119 PAPP PPP inside           p                      TTTT TTT T   TtTT TTT T
120 KKKK RKK surface                                 TTTT TTT T   TtTT TTT T
121 RNLR NLN SURFACE      B  p                       T..T T.T T   ttBt T.T T
```

E.T.H. method. Trailing character "s" indicates that amphiphilicity is preserved only if the position is assigned to the surface, while trailing character "i" indicates that amphiphilicity is preserved only if the position is assigned to the inside. Unassigned is a coil. Column 3 shows possible beta strand assignments made using the E.T.H. method. Column 4 shows parses made using the E.T.H. method ("p" indicates a strong parse, "*p*" a weak parse) and active site assignments ($). Column 5 shows consensus secondary structure assignments made using the E.T.H. method. **Neural net prediction.** [12,13]. Provided by Rost and Sander. Consensus prediction under **C. GOR Prediction.** According to reference [15].
Consensus prediction under **C. Chou-Fasman Prediction.** According to reference [14].
Consensus prediction under **C.**

No.	Sequences	ETH Prediction					Neural Net			GOR			Chou-Fasman		
	abcd efg	1	2	3	4	5	abcd	efg	C	abcd	efg	C	abcd	efg	C
122	SSSS SSS	S		B	$	B				B.BT	T.T		.tB.	...	.
123	VVVV VIV	inside		B		B	BBBB	BBB	B	BBBB	TBB	B	BtBB	BBB	B
124	AGGG GGG	inside		B		B	BBBB	BBB	B	BBBB	BBB	B	B.BB	BBB	B
125	IILI LII	INSIDE		B		B	BBBB	BBB	B	BBBB	BBB	B	B.BB	BBB	B
126	VVVV IVV	inside		B		B	BBBB	BBB	B	BBBB	BBB	B	B.BB	BBB	B
127	RKQQ QQQ	surface		B		B	BBBB	BBB	B	BBBB	BBB	B	B.BB	BBB	B
128	DDDD DDD	D			$			B	B	...B	...	.	..tt	ttt	T
129	HHYY HHH	inside								....	...	.	tttt	ttt	T
130	NSSS SSS	surface			p					....	...	.	tTtt	...	T
131	ASPP PPP	Hb var			p					....	..T	.	BTT.	..T	
132	INNI IIK	SURFACE			p			B	B	B...	..T	.	BTTB	BBT	B
133	VVVN NNT	SURFACE	A			A	BBBB	BBB	B	BAAB	..A		BAAB	BBB	B
134	FFFL LLL	inside	A			A	BBBB	BBB	B	BAAB	..A		BAAB	BBB	B
135	VMMV LLL	inside	A			A	BBBB	BBB	B	BAAB	AAA	A	BAAB	BBB	B
136	VVVV MMI	inside	A			A	BBBB	BBB	B	AAAB	AAA	A	BAAB	BBB	B
137	AAAA GGA	inside	A			A	BBBB	BBB	B	AAAA	AAA	A	BAAB	aaB	A
138	VVVV VVV	inside	A			A	BBBB	BBB	B	AAAA	AAA	A	BAAB	aaB	A
139	TTTI TTT	inside	A			A	BBBB	BBB	B	AAAA	AAA	A	BAAB	aaA	A
140	MMMM MMM	inside	A			A	BBBB	BBB	B	AAAA	AAA	A	BAAB	aaA	A
141	TTTT AAA	inside	A			A	BBBA	BBB	B	AAAA	AAA	A	.AA.	aaA	A
142	HHHH HHH	H	A		$	A	AAAA	AAA	A	AAAA	AAA	A	.AA.	aaA	A
143	EEEE EEE	E	A		$	A	AAAA	AAA	A	AAAA	AAA	A	.AA.	aaA	A
144	MILM LLL	inside	A			A	AAAA	AAA	A	AAAA	AAA	A	.AA.	aaA	A
145	GGGG GGG	G	A		p	A				AAAT	AAA	A	....	...	.
146	HHHH HHH	H			$					A..T	AAA	A	....	...	.
147	NNNN NNN	N			$					A...	..A	.	....	...	.
148	LLLL LLL	inside		B		B				A.A.	..A	.	....	aaa	.
149	GGGG GGG	G		B	p	B				AAA.	..A	A	....	aaa	.
150	MMMI MMM	inside		B		B				AAA.	TTA	A	AAA.	aaa	A
151	HEEP EEK	SURFACE		B		B				AAA.	TTA	A	AAAt	aaa	A
152	HHHH HHH	H		B	$	B				AAAT	TTA	A	AAAt	aaa	A
153	DDDD DDD	D			$					AAAT	TTA	A	AAAT	aaa	A
154	EDDG GGE	SURFACE			p					AAAT	TTA	A	AAAT	..a	A
155	-KK- KKN	surface			p					-AA-	TTA	-	-AA-	TT.	-
156	DDDN DDH	SURFACE			p					TAAT	TTT	T	A..T	TT.	
157	KKKS ---	surface			p					TAAT	---	-	TTTT	---	-
158	CCCC CCC	C			$					TAAT	TTT	T	TTTT	..t	T
159	NKKT LLH	SURFACE								TAAT	TTT	T	tttT	..t	T
160	CCCC RRC	surface								TAAT	TTT	T	tttT	..t	T
161	NEEG GGS	SURFACE			p					TAAT	TTT	T	B..T	...	.
162	TAAG AAA	inside			p					TAAT	TTT	T	BBBT	..b	B
163	---F SSS	?	a		p					---.	..T	-	---B	..b	-
164	---P LLF	inside	a		p		BBBB	BBB	B	---T	BBT	-	---B	BBb	-
165	CCCC CCC	C	a	B	$	B	BBBB	BBB	B	BAAT	BBT		BBBB	BBb	B
166	IIII III	inside	a	B		B	BBBB	BBB	B	BAAB	BBT	B	BBBB	BBb	B
167	MMMM MMM	inside	a	B		B	BBBB	BBB	B	BAAB	BB.	B	BBBB	BBb	B
168	SSSS RRP	surface	a	B	p	B	BBBB	BBB	B	BAAB	BB.	B	.BB.	BB.	B
169	KADP PPP	surface	a		p			BB		BAAB	TT.		aBB.	TTT	
170	VVVM GGS	Hb var	a	B	p	B				BAAT	TT.		aBB.	TTT	
171	LIII LLI	INSIDE	a	B		B				BAAT	...		aBB.	...	.
172	SSSS TTS	Hb var	a	B		B				.TT.	...	.	att.	...	.
173	RDDD PKE	SURFACE	a	B		B				.TT.	...	.	att.	TTt	T
174	QKKP GGG	surface			p					....	...	.	a..T	TTt	T
175	PPPP RRP	SURFACE	A		p	A				TTT.	...	.	TTTT	TTT	T
176	SSSS SSS	S	A		$	A				TTT.	..T	T	TTTT	TTT	T
177	KKKE YYY	surface	A			A				TTTT	..T	T	t..t	..t	.
178	YLLL EEE	amphi	A			A	BBBB	BBB	B	TTTT	.TT	T	....	...	.
179	FFFF FFF	inside	A			A	BBBB	BBB	B	TTTT	.TT	T	....	...	.
180	SSSS SSS	S	A		$	A				TTTT	.TT	T	.TTT	.tT	T
181	EDDN DDD	surface	A		p	A				TTTT	.TT	T	tTTT	.TT	T
182	CCCC ADC	surface	A			A				TTTT	.TT	T	tTTt	.TT	T

Figure 1 - Continued

zinc. Despite this similarity, we have been unable to present a convincing case based on either sequence alignment or a secondary structure prediction that hemorrhagic metalloproteases are homologues of these other proteases in other parts of the fold.

```
183 SSSS SSS S       A      $  A                          TTTT ..T T   tttt .bt T
184 KKKK MMK surface A         A   AAAA AAA A           TTTT TAT T   TTTt BbT T
185 DDNA RHD SURFACE A         A   AAAA AAA A           TTTT TAT T   TTTB BbT T
186 TTDY YYY surface A         A   AAAA AAA A           TTTT TAT T   BBtB BbB B
187 YYYY YYY inside  A         A   AAAA AAA A           BBTB TAT     BBBB BbB B
188 QQQQ QEQ surface A         A   AAAA AAA A           BBBB TAT B   BBBB BAB B
189 TTTT KRM surface A         A   BBBA BAA B           BBBB TAT B   BBBB BAB B
190 FFFF FFF inside  A         A   BBBB BBB B           BBBB TAT B   BBBB BAB B
191 LLLL LLL inside  A         A   BBBB BBB B           TTBB TTT T   BBBB BAB B
192 TTTT DKT surface A         A   BBBB BBB B           TTTT TTT T   BtBB .AB B
193 NNKD QQK SURFACE A      p  A                        TTTT TTT T   Bttt TA. T
194 HSYH YYR SURFACE        p                           .TTT TTT T   tttt T.. T
195 NKNK KKK surface        p                           .... ... .   tt.. ... .
196 PPPP PPP P              p                           TTTT TTT T   .... ... .
197 QQQQ QQQ Q       a   B  $  b   BBBB BBB B           TTTT TTT T   BBBB BBB B
198 CCCC CCC inside  a   B     b   BBBB BBB B           BBBB TTT B   BBBB BBB B
199 IIII III inside  a   B     b   BBBB BBB B           BBBB TTT B   BBBB BBB B
200 LILI LLL inside  a   B     b   BBBB BBB B           BBBB TTT B   BBBB BBB B
201 NNNN NNN N       a   B  $  b    BB   BB            BBBB ... B   BBBB BBB B
202 AAAA KKK amphi   a   B     b                       .... ... .   BBBB ... B
203 PPPP PPP P           p                             .... ... .   .... ... .
```

Figure 1 - Continued

REFERENCES

1. Benner, S. A., and Gerloff, D. (1991) Adv. Enzyme Regulat. 31, 121-181.
2. Benner, S. A., Cohen, M. A., and Gerloff, D. L. (1993) J. Mol. Biol. 229, 295-305.
3. Gerloff, D. L., Jenny, T. F., Knecht, L. J., Gonnet, G. H., and Benner, S. A. (1993) FEBS Lett., 318, 118-124.
4. Benner, S. A. (1992) U.S. Patent Application, March 25, 1992.
5. Benner, S. A. (1992) Curr. Opin. Struct. Biol. 2, 402-412.
6. Knighton, D. R., Zheng, J., Ten Eyck, L., Ashford, F. V. A., Xuong, N. H., Taylor, S. S., and Sowadski, J. M. (1991). Science 253, 407-414.
7. Thornton, J. M., Flores, T. P., Jones, D. T., and Swindells, M. B. (1991). Nature 354, 105-106.
8. Lesk, A. M., and Boswell, D. R. (1992) BioEssays 14, 407-410.
9. Rost, B. Schneider, R., and Sander, C.(1993) Trends Biochem. Sci. 18, 120-123.
10. Robson, B., and Garnier, J. (1993) Nature 361, 506 (1993).
11. Benner, S. A. (1989) Adv. Enzyme Regulat. 28, 219-236.
12. Rost, B., and Sander, C. (1992) Nature 360, 540.
13. Rost, B., and Sander, C. (1993) J. Mol. Biol., in press.
14. Chou, P.Y., and Fasman, G.D. (1978) Adv. Enzymol. 47, 45-148.
15. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978). J. Mol. Biol. 120, 97-120.
16. Takeya, H., Arakawa, M. Miyata, T., Iwanaga, S., and Omori-Satoh, T. (1989) J. Biochem. (Tokyo) 106, 151-157.
17. Dayhoff, M. O., Schwartz, R. M., Orcutt, B. C., in "Atlas of Protein Sequence and Structure", M.O. Dayhoff, Ed., (National Biomedical Research Foundation, Washington, D.C., 1978) vol. 5, suppl. 3, p. 345 (1978).
18. Crawford, I.P., Niermann,T., and Kirschner, K. (1987) Proteins 2, 118-129.
19. Chothia, C., and Lesk, A. (1986) EMBO J. 5, 823-826.
20. Colman, P. M., Jansonius, J. N., and Matthews, B. W. (1972) J. Mol. Biol. 70, 701-724 (1972).
21. Bode, W., Gomis-Rueth, F. X., Huber, R., Zwilling, R., and Stoecker, W. (1992) Nature 358, 164-167 (1992).

Reprinted from Biochemical and Biophysical Research Communications 194, 560–565 (1993)