PROTEINS: Structure, Function, and Genetics 20:1–3 (1994)

PREDICTION REPORT

A Prediction of the Secondary Structure of the Pleckstrin Homology Domain

Thomas F. Jenny and Steven A. Benner
Department of Chemistry, Swiss Federal Institute of Technology, CH-8092 Zurich, Switzerland

ABSTRACT **A consensus prediction for the secondary structure of the pleckstrin homology (PH) domain is presented. The prediction is based on an analysis of patterns of conservation and variation of homologous protein sequences. The structure is predicted to be formed largely from beta strands with a single alpha helix.** © 1994 Wiley-Liss, Inc.

Key words: protein secondary structure prediction, pleckstrin homology domain

Efforts to predict secondary structure in proteins have recently come to include methods that use as input a multiple alignment of homologous protein sequences.[1,2] Particularly effective have been tools that extract conformational information from patterns of conservation and variation within these alignments.[3–5] Five bona fide predictions made using these tools can now be evaluated using one or more subsequently determined experimental structures.[6]

Many predictions made with these tools focused on proteins and domains involved in signal transduction, including the src homology 1 (SH1) domain, a protein kinase,[7] the src homology 2 (SH2) domain,[8] a unit that binds peptides containing phosphotyrosine, and the src homology 3 (SH3) domain,[9,10] a unit that presumably binds to proline-rich peptide sequences. The recently identified pleckstrin homology (PH) domain may also be involved in similar type interactions in signal transduction.[11,12] As an experimental structure may be imminent, we present here a predicted consensus model for the secondary structure of this protein family.

Sequences of pleckstrin homology domains were extracted from entries in SwissProt 27 and a multiple alignment built by DARWIN.[13] The multiple alignment was then adjusted by hand (Fig. 1) in light of a multiple alignment of Musacchio et al.[14] Additional sequences present in the multiple alignment from Musacchio et al. but not listed in SwissProt 27 were not incorporated. Surface and interior residues were assigned by an automated procedure similar to that described elsewhere.[15] The multiple alignment was then parsed using procedures described elsewhere,[7,16] and elements of secondary structure were predicted within the parsed segments from patterns of conservation and variation, as described elsewhere.[7] Many of the automated routines are available on a server via electronic mail at the address cbrg@inf.ethz.ch.

The prediction is shown in Figure 1. In addition to serving as a documentation of a prediction in advance of the appearance of an experimental structure, this prediction contributes to the discussion of methodology in three ways. First, testing of the heuristics that parse the alignment based on strings of consecutive Pro, Gly, Asp, Asn, and Ser heuristic is more advanced in this prediction than in any previous prediction.[16] Second, the prediction was made independently of that of Musacchio et al.[14] Nevertheless, it corresponds well to their prediction. As Musacchio et al.[14] use an approach that resembles the part of the ETH method that focuses on periodicity in patterns of variation and conservation of amino acids,[3] this correspondence is not surprising. It does, however, illustrate the transferrability of the method and the reproducibility of its predictions, topics that have been the source of some discussion in the recent literature.[17]

Finally, one feature of the pleckstrin homology domain family that separates it from other families of proteins for which predictions have been recently been published (for example, the hemorrhagic metalloprotease family)[16] is its enormous sequence divergence. The PH domain family has an overall divergence greater than 250 PAM units, while the hemorrhagic metalloprotease family had diverged by only ca. 75 PAM units. The large sequence diver-

Received April 14, 1994; revision accepted May 24, 1994
Address reprint requests to Dr. Steven A. Benner, Laboratorium für Organische Chemie, E.T.H. Zentrum, CH-8092 Zurich, Switzerland.

Alignment Position Number	Surface Interior Parse Prediction	Pleckstrin Homology Domains Sequences lk m	hqgd j	abc pf	eo i n	Predicted Secondary Structure
1	0.41 S	YY E	G L E	ED RN	LI G L	
2	0.84 S	AA N	AMML R	EA QQ	DG Q A	
3	1.79 S P	ML D	GNSA V	MFF LD	TH N N	
4	2.04 S P	GG I	GEDQ G	ERY GE	SK C Y	
5	0.80 S P	KK S	SVVR V	PGK NI	QK N G	
6	1.30 S P	DD N	GSAD V	KVN QL	TG E R	
7	0.17 S p	CC S	SVIG V	RII VV	FA F P	
8	0.27 I	II I	AIVT S	IIV II	VT I K	
9	0.91 S	MM K	RKKR K	RKK RR	RK R I	
10	0.94 S	HH N	EEEK K	EQK KK	QM E D	
11	0.50 I	GG G	GGGG G	GGG GG	GK D G	strand
12	0.35 I	YY I	WWWY Y	YCY HW	SD S E	strand
13	5.47 I	MM L	LLLL M	LLL ML	LL L L	strand
14	0.60 S	LS Y	FHHS N	VLL VT	IA S K	strand
15	1.00 S	KK L	KKKK F	KKK II	QR K I	
16	0.58 S	LM E	WRRR L	KQK QN	VF L T	
17	0.93 S P	GG D	TGGS E	GGG NN	PK G S	
18	0.61 S P	NN P	NEES E	SHK LI	M_ S V	
19	0.51 S P	PP V	YYYD K	VRG GG	S_ G E	
20	- P	__ _	____ _	___ __	G_ _ _	
21	$ P	__ _	____ _	___ __	S_ _ _	
22	0.11 I P	__ _	____ _	___ II	L_ _ _	
23	0.08 S P	FF _	____ _	___ MM	S_ _ _	
24	0.79 S P	LL N	IIIN T	FR_ KK	L_ K R	
25	0.82 S P	TT H	KKKP Q	NKK GG	K_ R R	
26	0.85 S P	QQ E	GTTK G	TNR GG	K_ I S	
27	0.55 I	WW W	YWWW W	WWW SS	EP W K	
28	0.69 S	QQ Y	QRRQ T	KKK RK	GM S T	
29	0.96 S	RR P	RPPT R	PVN PE	EQ E D	
30	0.51 S	RR H	RRRK R	MRL YY	RR R R	strand
31	0.61 I	YY Y	WYYW W	WKY WW	QH K Y	strand
32	4.01 I	FF F	FFFF V	VFF FF	CL V A	strand
33	5.47 I	YY V	VLLA I	VII VV	FF F F	strand
34	5.47 I	LL L	LLLL V	LLL LL	LL L L	strand
35	0.25 S	FF T	SKKL R	LRE TT	FY F L	strand
36	1.78 S	PP S	NSNQ R	EEG SA	SE D D	
37	0.24 S P	__ _	_DD_ _	_DS __	__ _ _	
38	0.10 S P	__ _	_GG_ _	_PD __	__ _ _	
39	1.30 S P	NN S	GSTN P	DAA EE	KK _ K	
40	0.27 S	RR K	LFFL Y	GYQ SN	HA G A	strand
41	5.47 I	LL I	LIIL I	ILL IL	LI L L	strand
42	0.27 S	EE Y	SGGF L	EHI SS	IV M L	strand
43	5.47 I	WW Y	YYYY L	FYY WW	IF V I	strand
44	0.05 S	RR S	YKKF F	YYF YY	CC L C	strand
45	0.94 S P	GG E	REEE R	KDE KK	TK C K	
46	0.79 S P	EE E	SRRS D	KPS DD	RR K R	
47	0.85 S P	GG T	KPPD D	KAE ED	GR A R	
48	P	__ _	____ _	___ __	__ N _	
49	P	__ _	____ _	___ __	__ T _	
50	P	__ _	____ _	___ __	__ K _	
51	0.26 S P	__ _	____ _	___ __	_V K _	
52	0.26 S P	__ _	____ _	___ __	_E Q _	
53	0.24 S P	__ _	____ _	___ __	_S T _	
54	- P	__ _	____ _	___ __	_G P _	
55	0.26 S P	__ _	____ _	___ __	_E S _	
56	- P	__ _	____ _	___ __	_G A _	
57	0.09 S P	__ _	____ _	___ __	SS G _	
58	0.43 S P	__ N	____ _	___ __	GD A _	
59	0.47 S P	__ S	A___ _	___ __	SR T _	
60	0.48 S P	__ P	E___ _	__K __	KY A G	
61	0.56 S P	__ L	M__S R	SGR DE	LP Y D	
62	2.18 S P	__ G	REQS D	DAA EE	HS D S	
63	1.78 S P	EE D	HADS L	NET KK	LY Y Y	
64	0.85 S P	SA L	TPVR V	SDK EE	TS R D	
65	1.25 S P	RP L	CDDP I	PPP KK	KF L L	
66	0.92 S P	QQ R	RQQS R	KLK KK	NK K K	
67	1.98 S P	NS G	GTRG G	GGG FY	GH E A	
68	1.78 S	LL V	TLEL I	MAL MM	VC K S	strand (weak)
69	5.01 I P	LL L	IPAY I	III LL	IW Y V	strand (weak)
70	0.76 S P	TT D	NPPL N	PHD PS	SK F N	strand (weak)
71	5.47 I	MM V	LLLL L	LLL LV	LM M L	strand (weak)
72	0.95 S	EE P	ANNE A	KRS DD	ID R H	
73	1.36 S	QE A	TNNG N	GGV GN	DE R S	strand
74	3.52 I	II C	AFFS A	SCC LL	CV V F	strand
75	0.94 S	MQ Q	NSSI R	TVS KK	TG D Q	strand
76	3.75 I	SS I	IVVC I	LVV LL	LI I V	strand
77	0.48 S	VV A	TAAK E	TTY RR	LT N R	strand
78	2.21 S P	EE I	VEQR H	SSV DD	DE D D	
79	0.41 S P	EE R	ECCM S	PVV IV	DY R D	
80	0.66 S P	TT P	DQQP E	CEH EE	PV P S	
81	0.91 S P	QQ E	SLLS D	QSD QK	EK D S	
82	0.56 S P	II G	_MMP Q	DNS GG	NG S G	
83	0.37 S P	KK K	_KKK Q	FSL FF	MD D E	
84	1.11 S P	DE N	_TTR A	GNF MM	DN D R	
85	0.98 S P	RR N	_EEG M	KGG SS	D_ L D	
86	0.72 S P	__ R	_RRT V	RRR MS	D_ K N	
87	P	__ _	___S _	___ __	___ _ _	
88	0.22 S P	__ _	___S _	___ __	G_ _ _	
89	$ P	__ _	___K _	___ __	K_ _ _	
90	0.24 S P	__ _	___E _	___ __	G_ _ _	
91	0.24 S P	__ _	___S _	_K_ __	Q_ _ _	
92	0.25 S P	__ _	___D _	_S_ __	E_ _ K	
93	0.61 S P	__ _	_PPK K	_E_ S_	V_ _ K	
94	0.75 S P	__ _	_RRQ V	_E_ R_	D_ _ W	
95	0.64 S P	__ L	_PPH P	MEP RK	H_ _ S	
96	0.87 S P	KK F	CNNH N	FNN VH	LR N H	
97	0.44 S	CC V	NTTY T	VLC TI	DK S M	strand
98	5.47 I	IL F	FFFF F	FFF FF	FF F F	strand
99	0.24 S	LL S	IVIT S	KEQ AA	KE E L	strand
100	5.47 I	LL I	IIIV V	III LL	II L L	strand
101	0.82 S	RK S	SRRN C	TIV FF	WW A I	strand
102	0.44 S P	VI M	NCCF T	TTV SN	VY P E	
103	1.05 S P	KR A	GLLS N	TAQ PT	EG R D	
104	0.84 S P	GG S	GQQN Q	KDH DE	PE M Q	
105	0.54 S P	GG V	AWWD R	QEF GQ	KK Q G	
106	0.79 S	KK A	QTTS G	QVS RR	DE P A	
107	$ P	__ _	_____ _	____ NN	__ _ _	
108	0.11 I P	__ _	_____ _	___ VV	__ _ _	
109	0.11 I P	__ _	_____ _	___ YY	__ _ _	
110	0.79 S.P	__ _	_____ _	__E RK	S_ _ _	
111	0.12 S P	__ _	_TT_ _	__E DD	P_ _ _	
112	0.06 S P	__ H	_VVQ _	__H YY	P_ _ _	
113	0.68 S P	__ W	_IIK _	__Y KR	FE _ Q	
114	0.73 S P	QQ S	TEES _	DHI QQ	TV P G	
115	0.04 S	FF L	YRRL F	HYF LL	VY I Y	strand
116	0.72 S	VI D	HTTE L	FFY EE	II V E	strand
117	5.47 I	LL V	LFFL M	FLF LL	LV L L	strand
118	0.69 S	QQ A	KHHR Q	QQA SA	VQ T F	strand
119	3.30 I	CC A	AVVT M	AAG CC	AA A F	strand
120	2.54 S P	ED D	SDED M	AAE EE	SS K K	
121	1.13 S P	SS S	SSTD P	FTT TT	SN N T	
122	0.78 S P	DD Q	EPPS G	LPP VQ	RV A R	
123	2.18 S P	PP E	VDEK D	EKE EE	QD Q E	
124	0.87 S	EE E	EEED E	EEQ DE	EV H L	helix
125	0.29 S	FL L	RRRC M	RRA VV	KK K K	helix
126	0.81 S	AV Q	QEED Y	DTE ED	AM H K	helix
127	0.86 S	QQ D	REEE D	AED SS	AT D K	helix
128	5.47 I	WW W	WWWW W	WWW WW	WW W W	helix
129	1.38 S	LK V	VMTV L	VIM KK	TL M M	helix
130	2.18 S	KK K	TRTA Y	RKK AA	SK A E	helix
131	0.31 S	EE K	AAAA A	DAG SS	DE D Q	helix
132	5.47 I	LL I	LIII I	IIL FF	II L F	helix
133	1.20 S P	TR R	EQQA N	NQQ LL	IR L E	helix
134	1.56 S P	CD E	LMTR P	KMA RR	QN M M	helix
135	3.14 I	TA V	AVVA L	AAF AA	CI V A	helix
136	0.02 I	FY A	KAAS M	ISC GG	VL I I	helix
137	2.53 S	NR Q	ANDY A	KRN VV	DL T S	helix
138	1.13 S	EE T	KSGK G	CTL YY	NK K N	helix
139	0.14 S	AA A	ALLI Q	IGR PP	IQ S I	
140	0.93 S	QQ D	VKKL M	EKK EE	RQ M Y	
141	2.65 S	RQ A	KQKA K	G S KR	CE L P	

Fig. 1. Consensus secondary structure prediction for the pleckstrin homology (PH) domain surface (S) and interior (I) predictions are stronger with increasing index value; "P" indicates a parse. SwissProt 27 accession numbers for sequences of the pleckstrin homology domains: a, P08567; b, P08567; c, P20936; d, P28818; e, P28818; f, P21575; g, P31749; h, P22059; i, P26675; j, P08567; k, P25098; l, P08567; m, P19174; n, P27870; o, P08567; p, P08567; q, P08567.

gence makes it impossible to align reliably the PH domain family by a fully automated process, and suggests that substantial divergence of secondary structure has occurred within the family. For example, the strand at alignment positions 68–71 would be strongly assigned if the PP dipeptide did not appear in sequence q. This could represent divergence of conformation within the family. Thus, just as the prediction for the hemorrhagic metalloprotease family[16] tested the scope of the ETH prediction method for narrowly divergent protein families, this prediction tests the scope of our method for very divergent families.

Evaluation of the prediction will also undoubtedly be complicated by this sequence divergence, as it was with the SH3 domain.[18] Approaches for evaluating consensus secondary structure predictions have been discussed elsewhere,[19–21] and this discus-

sion should be consulted before evaluating this or any other consensus prediction.

REFERENCES

1. Nishikawa, K., Ooi, T. Amino acid sequence homology applied to the prediction of protein secondary structure, and joint prediction with existing methods. Biochem. Biophys. Acta 871:45–54, 1986.
2. Crawford, I.P., Niermann, T., Kirschner, K. Prediction of secondary structure by evolutionary comparison: application to the a subunit of tryptophan synthase. Proteins 2:118–129, 1987.
3. Benner, S.A. Patterns of divergence in homologous proteins as indicators of tertiary and quaternary structure. Adv. Enzym. Regul. 28:219–236, 1989.
4. Benner, S.A. Predicting de novo the folded structure of proteins. Curr. Opin. Struct. Biol. 2:402–412, 1992.
5. Bazan, J.F. Structural design and molecular evolution of a cytokine receptor superfamily Proc. Natl. Acad. Sci. U.S.A. 87:6934–6938, 1990.
6. S.A. Benner. Predicting the conformation of proteins from sequences. Progress and future progress. J. Mol. Recog., in press.
7. Benner, S.A., Gerloff, D. Patterns of divergence in homologous proteins as indicators of secondary and tertiary structure: The catalytic domain of protein kinases. Adv. Enzyme Regul. 31:121–181, 1991.
8. Russel, R.B., Breed, J., Barton, G.J. Conservation analysis and structure prediction of the SH2 family of phosphotyrosine binding domains. FEBS Lett. 304:15–20, 1992.
9. Benner, S.A., Cohen, M.A., Gerloff, D. A predicted secondary structure for the Src homology domain 3. J. Mol. Biol. 229:295–305, 1993.
10. Musacchio, A., Gibson, T., Lehto, V.-P., Saraste, M. SH3-An abundant protein domain in search of a function. FEBS Lett. 307:55–61, 1992.
11. Haslam, R.J., Kolde, H.B., Hemmings, B.A. Pleckstrin domain homology. Nature (London) 363:309–310, 1993.
12. Mayer, J.B., Ren, R., Clark, K.L., Baltimore, D. A putative modular domain present in diverse signalling proteins. Cell 73:629–630, 1993.
13. Gonnet, G.H., Benner, S.A. Computational biochemistry research at ETH. Technical Report 154, Departement Informatik, March 1991.
14. Musacchio, A., Gibson, T., Rice, P., Thompson, J., Saraste, M. The PH domain. a common piece in the structural patchwork of signalling proteins. Trends. Biochem. Sci., 18:343–348, 1993.
15. Benner, S.A., Badcoe, I., Cohen, M.A., Gerloff, D.L. Bona fide prediction of aspects of protein conformation: Assigning interior and surface residues from patterns of variation and conservation in homologous protein sequences. J. Mol. Biol. 235:926–958, 1994.
16. Gerloff, D.L., Jenny, T.F., Knecht, L.J., Benner, S.A. A secondary structure prediction of the hemorrhagic metalloprotease family. Biochem. Biophys. Res. Commun. 194: 560–565, 1993.
17. Robson, B., Garnier, J. Protein structure prediction. Nature (London) 361:506, 1993.
18. Gerloff, D.L., Benner, S.A. Predicting the conformation of proteins: Man versus machine. FEBS Lett. 325:29–33, 1993.
19. Gerloff, D.L., Jenny, T.F., L.J. Knecht, G.H. Gonnet, G.H., Benner, S.A. The Nitrogenase MoFe Protein: A secondary structure prediction. FEBS Lett. 318:118–124, 1993.
20. Russell, R.B., Barton, G.J. The limits of protein secondary structure prediction accuracy from multiple sequence alignment. J. Mol. Biol. 234:951–957, 1993.
21. Jenny, T.F., Benner, S.A. Evaluating predictions of secondary structure in proteins. Biochem. Biophys. Res. Commun. 200:149–155, 1994.