PROTEINS: Structure, Function, and Genetics 22:299–310 (1995)

PREDICTION REPORT

A Predicted Consensus Structure for the Protein Kinase C2 Homology (C2H) Domain, the Repeating Unit of Synaptotagmin

Dietlind L. Gerloff, Gareth Chelvanayagam, and Steven A. Benner
Department of Chemistry, Swiss Federal Institute of Technology, CH-8092 Zurich, Switzerland

ABSTRACT **A secondary structure has been predicted for the protein kinase C2 regulatory domain found in homologous form in synaptotagmin, some phospholipases, and some GTP activated proteins. The proposed structure is built from seven consecutive beta strands followed by a terminal alpha helix. Considerations of overall surface exposure of individual secondary structural elements suggest that these are packed into a 2-sheet beta sandwich structure, with one of only three of the many possible folds being preferred.**
© 1995 Wiley-Liss, Inc.

Key words: prediction contest, beta sandwich, protein sequence alignment

INTRODUCTION

One of the defining problems in modern protein chemistry challenges the biological chemist to deduce the conformation (secondary and tertiary structure) of a protein from sequence information (primary structure). Both at the Swiss Federal Institute of Technology in Zurich[1] and elsewhere,[2–6] progress towards solution of this problem has come through an analysis of patterns of conservation and variation in the sequences of homologous proteins. Such an analysis is especially powerful when it is aided by detailed models of divergent evolution.[7] Predictions made using this approach are "consensus" models for conformation of a protein family; they assume that proteins related by common ancestry have similar conformations.[8]

The value of these methods has been demonstrated by their application to bona fide predictions, those published before an experimental structure becomes available. To date, over a dozen bona fide predictions have been made using these methods (reviewed in ref. 9). For about half of them, a subsequently determined crystal structure has emerged to allow these predictions to be evaluated. In each case, the predictions have proved to be remarkably accurate, especially in comparison with those obtained by classical methods.[9] Further, the misassignments that are made by these tools come in only a very few types; for example, mistakes are often made in predicting secondary structure near an active site. Recognition of this fact allows researchers to focus on a very small number of problematic aspects of the prediction strategy to develop improved heuristics. As a result, "perfect" predictions are possible, defined as secondary structural models that miss no core secondary structural elements, missassign no alpha helices as beta strands (or vice versa), and do not overpredict any significant secondary structural element.[10] Predictions that meet this criterion are satisfactory as starting points for assembly of a tertiary structural model of a protein family. Predicted secondary structures for the pleckstrin homology domain,[11,12] the hemorrhagic metalloproteinases,[13] and the Src homology 2 domains[2,3] come close to perfection by this definition.

Continuing bona fide prediction efforts are necessary to define the scope of this or any other prediction method. Gradually, a large set of examples will emerge that, in time, will become statistically representative of proteins as a whole. As part of the structure prediction contest being organized by Dr. John Moult (Center of Advanced Research in Biotechnology, Bethesda MD), we now add to this growing collection by predicting the secondary and tertiary structure of the repeating domain of synaptotagmin (also designated in the literature as p65). This protein is presumed to be involved in the release of neurotransmitters through calcium-triggered fusion of synaptic vesicles; it contains two segments of an internal repeat, each having ~ 120 amino acids.[14] This polypeptide segment is found in homologous form in the regulatory C2-region of protein kinase C (PKC), which is believed to confer

Received October 25, 1994; revision accepted February 13, 1995.
Address reprint requests to Steven A. Benner, Department of Chemistry, ETH, CH-8092 Zürich, Switzerland.

© 1995 WILEY-LISS, INC.

```
    Pos q w zyAx tsurv DCBEFGH  p nomlkji h fedcba g | K IJ SIAPred       Parse
      1 C C CCCC CCCCC CCCCCCC|                      |
      2 G G GGGG GGGGG GGGGGGG| S GKEEEEE H QKKKKK K | P F F  s   (*)
      3 A C VVVV MMMM TTTTTTT| E GEEEEEE Q SEEEEE E | T N N  s
      4 D D DDDD DDDDD DDDDDDD| K QAPPQQQ K EEPDEE E | S S S  s
      5 I H HHHH HHHHH HHHHHHH| A EEEEEEE V QEEEEE V | P R R  s
      6 S T TTTT TTTTT TTTTTTT| D KKKKKKK N KKNKKK K | E A A  s
      7 E E EEEE EEEEE EEEEEEE| L LSLLLLL C LLLLLL L | R L L
      8 V R RRRR KKKKK RRRRRRR| _ _L_____ _ ______ _ | I T A      (**)
      9 R R RRRR RRRRR RRRRRRR| _ _______ _ ______ _ | K Q H      (***)
     10 G G GGGG GGGGG GGGGGGG  G GGGGGGG G GGGGGG G | S G G      *
     11 K R RRRR RRRRR RRRRRRR  E DDDDDDD R RKKKKK R | S P P  S                B1
     12 L I LLLL IIIII IIIIIII  I IIIIIII I LILLLL I | _ W W  I                B1
     13 L Y QQQQ YYYYY YYYYYYY  N CCCCCCC N NQQQQQ Q | _ W W  s                B1
     14 L L LLLL LLLLL IIIIIII  F FFTTFFF F FFFFYY Y | _ _ _  i                B1
     15 Y E EEEE KKKKK QQQQQQQ  S SSSSSSS M KSSSSS K | _ _ _  S                B1
     16 V I IIII AAAAA AAAAAAA  L LLLLLLL L LLLLLL L | _ _ _  I                B1
     17 E N RRRR EEEEE HHHHHHH  C RRRRRRR R EDDDDD D | _ _ _  S                B1
     18 L V AAAA VVVVV IIIIIII  Y YYYYYYY Y YYYYYY Y | _ _ _  I                B1
     19 K K PPPP ATTTT EDDDDDD  L VVVVVVV T DDDDDD D | _ _ _  i      **
     20 _ _ ____ _____ _______  P PPPPPPP Y FFFFFF F | _ _ _         **
     21 _ _ TTTT _____ _______  T TTTTTTT T NQQQQQ Q | K R T  S
     22 G E SSAS DDDDD RRRRRRR  A AAAAAAA T SAANNN Q | Y P P  S
     23 N N DDDD EEEEE EEEEEDD  G GGGGGGG E NNNNNN G   S E K  S
     24 N L EEEE KKKKK VVVVVVV  R KKKKKKK Q SQQQQQ Q   R R R  s                B2
     25 L L IIII LLLLL LLLLLLL  L LLLLLLL L LLLLLL L   L L L  I                B2
     26 K T HHHH HHHHH IIIIIII  T TTTTTTT V ATTILL T   I R N  s                B2
     27 V V VVVI VVVVV VVVVVVV  I VVVVVVV V VVVVVV V   V V V  I                B2

     28 D Q TTTT TTTTT VVVVVLL  T VVCCVVV K TGGGGG T   N R R  s                B2
     29 I I VVVV VVVVV VVVVVVV  I IIIIIII I VIVIII V   V I V  I                B2
     30 K K GGGG RRRRR RRRRRRR  I LLLLLLL L IILIII I   I I I                   B2
     31 E E EEEE DDDDD DDDDDDD  K EEEEEEE K QQQQQQ Q   S S S  S                b2
     32 A G AAAA AAAAA AAAAAAA  A AAAAAAA A AAAAAA A   A G G  i                b2
     33 A R RRRR KKKKK KKKKKKK  T KKKKKKK L EAAAAA E   R Q Q  S                b2
     34 N N NNNN NNNNN NNNNNNN  N NNNNNNN D EEEEEE D   Q Q Q  s
     35 L L LLLL LLLLL LLLLLLL  L LLLLLLL L LLLLLL L   L L L  I
     36 I I IIII IIIII VVVVVVV  K KKKKKKK P PPPPPP P   P P P         *?
     37 P P PPPP PPPPP PPPPPPP  A KKKKKKK A AAAAAA G   K K K         *?
     38 M M MMMM MMMMM MMMMMMM  M MMMMMMM K LLLLLL M   Y V V  I
   a 39 D D DDDD DDDDD DDDDDDD  D DDDDDDD D DDDDDD D   T N N  S
     40 T P PPPP PPPPP PPPPPPP  L VVVVVVV A MMMVMM M   K K K         *
     41 N N NNNN NNNNN NNNNNNN  T GGGGGGG N GGGGGG S   S N N  s      ***
     42 _ _ ____ _____ _______  _ _______ _ ______ _   T _ _         ***
     43 _ _ ____ _____ _______  _ _______ _ ______ _   K K K         ***
     44 _ _ ____ _____ _______  _ _______ _ ______ _   G N N         ***
     45 G G GGGG GGGGG GGGGGGG  G GGGGGGG G GGGGGG G   E S S         ***
     46 F L LLLL LLLLL LLLLLLL  F LLLLLLL F TTTTTT T   V I I  I      *
   a 47 S S SSSS SSSSS SSSSSSS  S SSSSSSS S SSSSSS S   I V V  i      *
     48 D D DDDD DDDDD DDDDDDD  D DDDDDDD D DDDDDD D   D D D  A      *
     49 P P PPPP PPPPP PPPPPPP  P PPPPPPP P PPPPPP P   P P P  .      *
     50 Y Y YYYY YYYYY YYYYYYY  Y YYYYYYY Y YYYYYY Y   Y K K                   B3
     51 I V VVVV VVVVV VVVVVVV  V VVVVVVV V VVVVVV V   V V V  I                B3
     52 A K KKKK KKKKK KKKKKKK  K KKKKKKK K KKKKKK K   T I T  s                B3
     53 V V LLLL LLLLL LLLLLLL  A IIIIIII I VVVVVV L   L V V  I                B3
     54 Q K KKKK KKKKK KKKKKKK  S AVHHHHH Y YFFFFF Y   S E E  S                B3
     55 M L LLLL LLLLL LLLLLLL  L ILLLLLL L LLLVLL L   I I I  I                B3
     56 H I IIII IIIII IIIIIII  I MMLMMM L LLLLLL L   V H H  i                B3
     57 P P PPPP PPPPP PPPPPPP  C QQQQQQQ P PPPPPP P   G G G  .      *
```

Fig. 1.

	58	D	D	DDDD	DDDDD	DDDDDDD	D	NGNNNNN	D	DDDDDE	E	T	V	V	S	*	
	59	R	D	PPPP	PPPPP	PPPPPPP	E	GGGGGGG	R	KKKKKK	K	H	G	T	S	*	
	60	S	K	RRRR	KKKKK	KKKKKKK	R	KKKKKKK	_	______	_	F	R	G	S	*	
	61	G	D	NNNN	NNNNN	SSSSSSS	R	RRRRRRR	_	______	_	D	D	D	s	*	
	62	R	Q	LLLL	EEEEE	EEEEEEE	L	LLLLLLL	K	KKKKKK	K	Q	T	V			
	63	T	S	TTTT	SSSSS	SSSSSSS	K	KKKKKKK	K	KKKKKK	K	K	G	A			
	64	K	K	KKKK	KKKKK	KKKKKKK	K	KKKKKKK	K	KKKKKK	K	V	S	S			
	65	K	K	QQQQ	QQQQQ	QQQQQQQ	R	KKKKKKK	F	FYYYFF	V	E	R	R	s		
	66	K	K	KKKK	KKKKK	KKKKKKK	K	KKKKKKK	Q	EEEEEE	E	K	Q	Q	s		
	67	T	T	TTTT	TTTTT	TTTTTTT	T	TTTTTTT	T	TTTTTT	T	T	T	T	A		b4
	68	K	R	RRRK	KKKKK	KKKKKKK	S	SSTTTTT	K	KKKKKK	K	K	A	A	s		b4
	69	T	T	TTTT	TTTTT	TTTTTTT	I	VIVVIII	V	VVVVVV	V	V	V	V	I		b4
	70	I	I	VVVV	IIIII	IIIIIII	K	KKKKKKK	H	HQHHHH	H	I	I	V			b4
	71	Q	K	KKKK	RRRRR	KKKKKKK	K	KKKKKKK	R	RKRRRR	R	D	T	T	S		
	72	K	A	AAAA	SSSSS	CCCCCCC	N	CCNKNNN	K	KKKKKK	K	N	N	N	S		
	73	N	C	TTTT	TTNTT	SSSSSSS	T	TTTTTTT	T	TTTTTT	T	N	N	N		*	
	74	L	L	LLLL	LLLLL	LLLLLLL	L	LLLLLLL	L	LLLLLL	L	G	G	G	I	*	
	75											F	F	F		***	
	76	N	N	NNNN	NNNNN	NNNNNNN	N	NNNNNNN	N	SNNNNN	N	N	N	N	s	*	
	77	P	P	PPPP	PPPPP	PPPPPPP	P	PPPPPPP	P	PPPPPP	P	P	P	P	.	*	
	78	V	V	VVVV	QQQQR	EEEEEEE	V	YYYYYYY	I	VTAVVV	V	H	R	W	S		
	79	F	W	WWWW	WWWWW	WWWWWWW	Y	YYYFYYY	F	FFFFFF	F	W	W	W	I		
	80	N	N	NNNN	NNNND	NNNNNNN	N	NNNNNNN	N	NNNNNN	N	G	D	D	S		
a	81	E	E	EEEE	EEEEE	EEEEEEE	E	EEEEEEE	E	EEEEEE	E	E	M	T	A		
	82	T	T	TTTT	SSSSS	TTTTTTT	A	SSSSSSS	T	TSTSQQ	T	E	E	E	S		B5
	83	F	L	FFFF	FFFFF	FFFFFFF	L	FFFFFFF	F	FFFFFF	F	F	F	L	I		B5
	84	T	T	VVVV	TTTTT	RRRRRRR	V	SSSSSSS	Q	TVTITT	I	E	E	E	s		B5
	85	F	Y	FFFF	FFFFF	FFFFFFF	F	FFFFFFF	F	FFFFFF	F	F	F	F	I		B5
	86	E	D	NNNN	KKKKK	QQQQQQQ	D	EEEEEEE	N	KKKKKK	K	P	E	E	s		B5
	87	L	L	LLLL	LLLLL	LLLLLLL	I	VVIIVVV	V	SVVIVV	V	L	V	V	I		B5
	88	Q	K	KKKK	KKKKK	KKKKKKK	P	PPPPPPP	P	LPPPPP	A	Y	T	A	s	*	
	89									P						**?	
	90	P	P	PPPP	PPPPP	EEEEEEE	N	FFFFFFF	F	YYYYYY	F	N	V	V	s	*	
	91	Q	E	GGGG	SSSSS	SSSSSSS	E	EEEEEEE	N	AQQSSS	N	S	P	P	s	*	
	92	D	D	DDDD	DDDDD	DDDDDDD	N	QQQQQQQ	E	DEEEEE	E	Q	D	E	S		
	93	R	K	VVVV	KKKKK	KKKKKKK	M	MIIIIII	L	ALLLLL	I	L	L	L			
	94	D	D	EEEE	DDDDD	DDDDDDD	E	QQQQQQQ	Q	MGGGGG	T	_	_	_	s	*	
	95						H	KKKKKKK	N	NGGGGG	A				s	**	
	96	K	R	RRRR	RRRRR	RRRRRRR	V	IVVVVVV	R	KKKKKK	K	S	A	A	s		
	97	R	R	RRRR	RRRRR	RRRRRRR	N	CSQQQQQ	K	TTTTTT	T	M	L	L	s		B6
	98	L	I	LLLL	LLLLL	LLLLLLL	V	LLVVVVV	L	LLLLLL	L	L	V	V	I		B6
	99	L	L	SSSS	SSSSS	SSSSSSS	I	VMCVVVV	H	VMVVVV	V	L	R	R			B6
	100	I	I	VVVV	VVVVE	VVVVVVV	I	VIVVVVV	F	FMMMMM	F	I	F	F			B6
	101	E	E	EEEE	EEEEE	EEEEEEE	A	TTTTTTT	S	AAAAAA	A	R	M	V	s		B6
	102	V	V	VVVV	IIIII	IIIIIII	V	VVVVVVV	V	IVIVVV	I	V	V	V	I		B6
	103	W	W	WWWW	WWWWW	WWWWWWW	M	VMLLLLL	Y	FYYYYY	Y	D	E	E			
	104	D	D	DDDD	DDDDD	DDDDDDD	D	DDDDDDD	D	DDDDDD	D	D	D	D	A		
	105	W	W	WWWW	WWWWW	WWWWWWW	Y	YYYYYYY	F	FFFFFF	F	K	Y	Y			
	106	D	D	DDDD	DDDDD	DDDDDDD	D	DDDDDDD	D	DDDDDD	D	D	D	D	A		
	107	R	R	RRRR	RRRRR	LLLLLLL	C	RKKKKKK	R	RRRRRR	R	K	S	A	s		
	108	T	T	TTTT	TTTTT	TTTTTTT	I	ILILIII	F	FFFFFF	F	V	S	S	i		
	109	S	S	SSSS	TTTTT	SSSSSSS	G	GGGGGGG	S	SSSSSS	S	G	S	S		*?	
	110	R	R	RRRR	RRRRR	RRRRRRR	H	TSKKKKK	R	KKKKKK	K	H	K	K	s		
	111	N	N	NNNN	NNNNN	NNNNNNN	N	SNNNNNN	H	HHHHHH	H	N	N	N	s		
	112	D	D	DDDD	DDDDD	DDDDDDD	E	EDDEDDD	D	DDDDDD	D	R	D	D	s		
	113	F	F	FFFF	FFFFF	FFFFFFF	V	PAAAAAA	L	QCIVII	Q	I	F	F	i		b7
	114	M	M	MMMM	MMMMM	MMMMMMM	I	IIIIIII	I	IIIIII	I	G	I	I	I		b7
	115	G	G	GGGG	GGGGG	GGGGGGG	G	GGGGGGD	G	GGGGGG	G	H	G	G		*	b7
	116	S	A	AAAA	SSSSS	SSSSSSS	M	RRKKKKK	Q	EQEEEE	Q	H	Q	Q	s		B7
	117	F	L	MMMM	LLLLL	LLLLLLL	C	CCIIVVV	V	VVVAFF	V	C	S	S	i		B7

Fig. 1 continued.

```
118 S S SSSS SSSSS SSSSSSS R ILFFFFF V KTKKKK L   I T T s       B7
119 F F FFFF FFFFF FFFFFFF V LLVVVVV L VVVVVV I   R I I         B7
120 S G GGGG GGGGG GGGGGGG G GGGGGGG D PLPPPP P   _ P P   *     B7
121 _ _ ____ _____ _______ _ _______ N ______ _   _ _ _   *
122 _ _ ____ _____ _______ _ _______ L ______ _   _ _ _   *
123 _ _ ____ _____ _______ _ CCSSYYY L LMMMM  L | V W L i *
124 L I VVVV VVVVV IIIIIII N MNNNNNN E CTNNNN G | E N K
125 E S SSSS SSSSS SSSSSSS A GGAASSS F TKTTTT K | N S S s
126 E E EEEE EEEEE EEEEEEE T TTSTTTT S IVVVVV I | I L L
127 L I LLLL LLLLL LLLLLLL D GGGGAGG D DDDDDD D | R K K         A
128 Q I LLLL MMMM  QQQQQQQ G TATTAAA F LLLFFF L | P Q Q         A
129 K K KKKK KKKKK KKKKKKK P EEEEEEE S AGGGGG G | G G G s       A
130 E N AAAA MMMM  AAAAAAA _ _______ E QQQHHH A | _ _ _ s       A
131 P P PPPP PPPPP GGGGGGS G _______ D TQPVVV V | _ _ _   *     A
132 V T VVVV AAAAA VVVVVVV R LLLLLLL T ILITTT I | Y Y Y i       A
133 D N DDDD SSSSS DDDDDDD E RRRRRRR T EEEEEE E | R R R s       A
134 G G GGGG GGGGG GGGGGGG H HHHHHHH I EEEEEE E | I H H         A
135 W W WWWW WWWWW WWWWWWW W WWWWWWW W WWWWWW W | L V I i       A
136 Y F YYYY YYYYY FFFFFFF N SMSSSSS R RRRRRR K | K H H s       A
137 K K KKKK KKKKK KKKKKKK E DDDDDDD D DDDDDD D | L L L         A
138 F L LLLL LLALL LLLLLLL M MMMMMM  I LLLLLL I | K L L i       A
139 L L LLLL LLHLL LLLLLLL L LLLLLLL L VEQQQQ A | N S S         A
140 S T NNNN NNNNN SSSSSSS A AAAAAAA E SSGGSS P | N K K   *?    A
141 Q Q QQQQ QQQQQ QQQQQQQ N SSNNNNN A VAGAAA P | F N N s       A
142 V D EEEE EEEEE EEEEEEE P PPPPPPP T EEEEEE P | N G G
```

Fig. 1. Residue-by residue secondary structure prediction for the C2H domain. The column "SIAPred" lists positions as they are predicted to the surface (S, s), to the interior (I, i), or to lie near the "active site." Asterisks denote parse positions. Regions where the alignment is uncertain are marked with vertical lines between the subfamilies. Sequences are designated by single letters, from the SwissProt Version 29 database.

a (P21707) SYT1_RAT Pos 140 to 266 of SYNAPTOTAGMIN I (P65). RATTUS NORVEGICUS (RAT).
b (P21579) SYT1_HUMAN Pos 141 to 267 of SYNAPTOTAGMIN I (P65). HOMO SAPIENS (HUMAN).
c (P24505) SY61_DISOM Pos 146 to 272 of SYNAPTOTAGMIN A (SYNAPTIC VESICLE PROTEIN O-P65-A). DISCOPYGE OMMATA (ELECTRIC RAY).
d (P29101) SYT2_RAT Pos 141 to 270 of SYNAPTOTAGMIN II. RATTUS NORVEGICUS (RAT).
e (P24506) SY62_DISOM Pos 158 to 284 of SYNAPTOTAGMIN B (SYNAPTIC VESICLE PROTEIN O-P65-B). DISCOPYGE OMMATA (ELECTRIC RAY).
f (P21521) SY65_DROME Pos 191 to 318 of SYNAPTOTAGMIN (P65). DROSOPHILA MELANOGASTER (FRUIT FLY).
g (P34693) SYT1_CAEEL Pos 155 to 285 of SYNAPTOTAGMIN I. CAENORHABDITIS ELEGANS.
h (P24507) SY63_DISOM Pos 235 to 365 of SYNAPTOTAGMIN C (SYNAPTIC VESICLE PROTEIN O-P65-C). DISCOPYGE OMMATA (ELECTRIC RAY).
i (P21707) SYT1_RAT Pos 266 to 381 of SYNAPTOTAGMIN I (P65). RATTUS NORVEGICUS (RAT).
j (P21579) SYT1_HUMAN Pos 267 to 382 of SYNAPTOTAGMIN I (P65). HOMO SAPIENS (HUMAN).
k (P24505) SY61_DISOM Pos 272 to 387 of SYNAPTOTOGAMIN A (SYNAPTIC VESICLE PROTEIN O-P65-A). DISCOPYGE OMMATA (ELECTRIC RAY).
l (P29101) SYT2_RAT Pos 267 to 382 of SYNAPTOTAGMIN II. RATTUS NORVEGICUS (RAT).
m (P24506) SY62_DISOM Pos 284 to 399 of SYNAPTOTAGMIN B (SYNAPTIC VESICLE PROTEIN O-P65-B). DISCOPYGE OMMATA (ELECTRIC RAY).
n (P21521) SY65_DROME Pos 315 to 435 of SYNAPTOTAGMIN (P65). DROSOPHILA MELANOGASTER (FRUIT FLY).
o (P34693) SYT1_CAEEL Pos 285 to 405 of SYNAPTOTAGMIN I. CAENORHABDITIS ELEGANS.
p (P24507) SY63_DISOM Pos 360 to 470 of SYNAPTOTAGMIN C (SYNAPTIC VESICLE PROTEIN O-P65-C). DISCOPYGE OMMATA (ELECTRIC RAY).
q (P13677) KPC2_DROME Pos 185 to 300 of PROTEIN KINASE C (EC 2.7.1.-) (PKC) (DPKC53E(EY)) (PROTEIN INAC). DROSOPHILA MELANOGASTER (FRUIT FLY).
r (P10102) KPCA_RABIT Pos 150 to 265 of PROTEIN KINASE C, ALPHA TYPE (EC 2.7.1.-) (PKC-ALPHA). ORYCTOLAGUS CUNICULUS (RABBIT).
s (P05696) KPCA_RAT Pos 150 to 265 of PROTEIN KINASE C, ALPHA TYPE (EC 2.7.1.-) (PKC-ALPHA). RATTUS NORVEGICUS (RAT).
t (P17252) KPCA_HUMAN Pos 150 to 265 of PROTEIN KINASE C, ALPHA TYPE (EC 2.7.1.-) (PKC-ALPHA). HOMO SAPIENS (HUMAN).
u (P20444) KPCA_MOUSE Pos 150 to 265 of PROTEIN KINASE C, ALPHA TYPE (EC 2.7.1.-) (PKC-ALPHA). MUS MUSCULUS (MOUSE).
v (P04409) KPCA_BOVIN Pos 150 to 265 of PROTEIN KINASE C, ALPHA TYPE (EC 2.7.1.-) (PKC-ALPHA). BOS TAURUS (BOVINE).
w (P05130) KPC1_DROME Pos 155 to 275 of PROTEIN KINASE C (EC 2.7.1.-) (PKC) (DPKC53E(BR)). DROSOPHILA MELANOGASTER (FRUIT FLY).
x (P05697) KPCG_RAT Pos 150 to 265 of PROTEIN KINASE C, GAMMA TYPE (EC 2.7.1.-) (PKC-GAMMA). RATTUS NORVEGICUS (RAT), MUS MUSCULUS (MOUSE).
y (P05128) KPCG_BOVIN Pos 135 to 250 of PROTEIN KINASE C, GAMMA TYPE (EC 2.7.1.-) (PKC-GAMMA) (FRAGMENT). BOS TAURUS (BOVINE).
z (P10829) KPCG_RABIT Pos 150 to 265 of PROTEIN KINASE C, GAMMA TYPE (EC 2.7.1.-) (PKC-GAMMA) (DELTA). ORYCTOLAGUS CUNICULUS (RABBIT).
A (P05129) KPCG_HUMAN Pos 150 to 265 of PROTEIN KINASE C, GAMMA TYPE (EC 2.7.1.-) (PKC-GAMMA). HOMO SAPIENS (HUMAN).
B (P05772) KPC1_RABIT Pos 150 to 265 of PROTEIN KINASE C, BETA-I TYPE (EC 2.7.1.-) (PKC-BETA-1). ORYCTOLAGUS CUNICULUS (RABBIT).
C (P04410) KPC1_RAT Pos 150 to 265 of PROTEIN KINASE C, BETA-I TYPE (EC 2.7.1.-) (PKC-BETA-1). RATTUS NORVEGICUS (RAT).
D (P05126) KPC2_BOVIN Pos 150 to 265 of PROTEIN KINASE C, BETA-II TYPE (EC 2.7.1.-) (PKC-BETA-2). BOS TAURUS (BOVINE).
E (P05773) KPC2_RABIT Pos 150 to 265 of PROTEIN KINASE C, BETA-II TYPE (EC 2.7.1.-) (PKC-BETA-2). ORYCTOLAGUS CUNICULUS (RABBIT).
F (P04411) KPC2_RAT Pos 150 to 265 of PROTEIN KINASE C, BETA-II TYPE (EC 2.7.1.-) (PKC-BETA-2). RATTUS NORVEGICUS (RAT), MUS MUSCULUS (MOUSE).
G (P05771) KPC1_HUMAN Pos 150 to 265 of PROTEIN KINASE C, BETA-I TYPE (EC 2.7.1.-) (PKC-BETA-1). HOMO SAPIENS (HUMAN).
H (P05127) KPC2_HUMAN Pos 150 to 265 of PROTEIN KINASE C, BETA-II TYPE (EC 2.7.1.-) (PKC-BETA-2). HOMO SAPIENS (HUMAN).
I (P10688) PIP6_RAT Pos 607 to 725 of 1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE DELTA 1 (EC 3.1.4.11) (PLC-DELTA-1) (PHOSPHOLIPASE C-DELTA-1) (PLC-III). RATTUS NORVEGICUS (RAT).
J (P10895) PIP6_BOVIN Pos 545 to 665 of 1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE DELTA 1 (EC 3.1.4.11) (PLC-DELTA-1) (PHOSPHOLIPASE C-DELTA-1) (PLC-III) (FRAGMENT). BOS TAURUS (BOVINE).
K (Q02158) PIPA_DICDI 1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE (EC 3.1.4.11) (PLC) (PHOSPHOINOSITIDE-SPECIFIC PHOSPHOLIPASE C). DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).

Ca^{++} sensitivity upon protein kinase C activity.[15,16] A homologous domain is also found in the phosphatidylinositol-specific phospholipase C[17] and in cytosolic phospholipase A_2 (which displays Ca^{++}-dependent translocation), and in the GTPase activating protein.[18] Collectively, this family of proteins will be called the C2 homology (C2H) domain.

A multiple alignment for the protein family was built from sequences extracted from SwissProt 29[19] using the DARWIN system.[20,21] Surface and interior residues were assigned by automated procedures similar to those described elsewhere,[22] the multiple alignment was parsed into units forming independent secondary structures, and elements of secondary structure were predicted within the parsed segments from patterns of conservation and variation, as described elsewhere.[9,11,13,23] [Many of the automated routines used in this prediction are available to the public on a server accessible via electron mail at the address cbrg@inf.ethz.ch, or using the World Wide Web (WWW) with URL http://cbrg.inf.ethz.ch/.]

SECONDARY STRUCTURE PREDICTION

The secondary structure prediction is presented residue by reside in Figure 1 and is summarized in Table I. The protein is predicted to be nearly entirely built from beta strands. A beta strand is shared by the p65 and PKC subfamilies at positions 11–18; the poor alignment with the phospholipase domain suggests that it is not a core strand in the folded structure (see below). A beta strand is predicted in all three subfamilies for positions 24–30. A short helix was also considered for positions 29–37 based purely on surface/interior assignments. Its short length and the absence of a parse at the beginning suggested that this segment is best modeled as a coil when assembling a tertiary structure, although this coil may resemble a helix in some family members. Positions 37–49 form a coil. Positions 51 (possibly 50) to 56 form a beta strand, and positions 57–61 form a parse. The following region contributes to the "active site" of the protein. Secondary structures in these regions are difficult to assign reliably. Our preference is coil (62–66), beta (67–70), and coil (71–77). Positions 82–87 form another beta strand, as do positions 97–103. Positions 104 and 106 contain two absolutely conserved Asp residues, proposed to be at a regulatory site, presumably to bind Ca^{++}. Positions 116–120 form a beta strand, possibly starting at position 113. Provided that a parse is moved at positions 130–131, an alpha helix can be assigned following position 132 (132–141), with an additional turn at the beginning possible for some subfamilies. Although this helix lies in a region where the three subfamilies do not have statistically significant sequence identity, putative helices can be found in all three subfamilies.

We then asked whether the evolutionary analysis might suggest that the C2H domain form as a unit a single folding unit. Relevant to this was the suggestion that the GTPase activating protein (GAP) contains a domain that is similar to the C2H domain, but over only 45 residues.[18] This highly conserved unit begins with the parse separating strand 2 from strand 3, and ends with the parse separating strand 5 from strand 6. The correspondence of the sequence with the two parses increases the likelihood that the sequence similarity is not fortuitous. However, if the GAP domain is homologous *only* over strands 3, 4, and 5, and if the regions before and after are not to be homologous between the GAP and C2 proteins, this would suggest that strands 3, 4, and 5 form a folding unit by themselves. Examination of the GAP protein family in more detail using DARWIN showed, however, that a full C2H domain could plausibly be constructed within the GAP protein, not only for strands 3, 4, and 5, but also for beta strands 2, 6, and 7, and the following helix. Therefore, we conclude that at least six of the seven strands in the C2H domain form a single folding unit; again, strand 1 could not unambiguously be established as part of this domain.

We then asked whether information contained within patterns of divergence and conservation in the C2H domain family could suggest an overall fold for this unit. To answer this question, we began with tools outlined elsewhere, where accurate guesses (but only guesses) of supersecondary and domain structure have been possible by careful analysis of patterns of variation and conservation.[9,10,13,23] For the C2H domain, approximately 39% of the residues are assigned to the interior of the structure; a full half of these are strongly assigned to the interior. This is consistent with a globular structure rather than an extended structure, as this is the only way to bury so many residues in a protein of this size. The seven beta strands can form one, two, or possibly three beta sheets. However, because there is only a single alpha helix, most of the interior residues must be buried by contact with other strands. This constrains possible folds to beta barrels and sandwiches.

The assignments of interior and surface residues to the seven predicted strands allows us to rank them in order of decreasing surface exposure (most exposed strand $4 > 1 > 5 \approx 2 > 7 \approx 3 > 6$ least exposed). This suggests that strands 3 and 6 are either central in a sheet, or are in contact with the helix. Further, strand 4 is weakly predicted. We know from past experience that surface edge strands are frequently confused with coils, especially when near an "active site."[9,10,23] Further, as noted above, strand 1 is not obviously homologous in the lipase subfamilies. This suggests substantial divergence, also consistent with an edge position and its high overall surface exposure. The presence of one, and possibly two edge strands is inconsistent with a beta

TABLE I. Secondary Structure Assignments in the C2H Domain

Unit	Preferred length	Comments
Loop 0-1	1–10	Loop surface not at active site
Beta 1	11–18	Clear alternation; surface positions 15 and 17, one side exposed
Loop 1-2		Short (three residues minimum), surface, not at active site
Beta 2	24–30	Clear alternation
Loop 2-3		Asp 48, calcium binding (?), conserved SDP, long loop (16–19 residues)
Beta 3	51–56	Largely inside
Loop 3-4 AS		Lys-rich region, Thr 67, active site loop (10–12 residues)
Beta 4	67–70	Weak assignment, edge strand
Loop 4-5 Ca		Glu 81, calcium binding (?), long loop (10–11 residues)
Beta 5	82–87	Partly exposed, surface residue at position 86
Loop 5-6		Short (5–7 residues), surface, not at active site
Beta 6	97–103	Fully buried
Loop 6-7		Asp 104, Asp 106, calcium binding (?), long loop (10 residues)
Beta 7	116–120	Largely buried
Loop 7-a		Medium length (5–9 residues), not at active site
Alpha 1	132–141	Single helix

barrel structure, but is consistent with beta sandwich structures, which have been intensively investigated in recent years.[24]

Secondary structural elements are often oriented relative to assignments of active site regions. As the active site function of the C2H domain is not known, it is difficult to apply this approach in this case. Further, active site regions are assigned both at the beginning (Asp-48) and end (Lys-rich region, and conserved Thr-67) of strand 3, and at the beginning (Lys-rich region, and conserved Thr-67) and end (conserved Glu 81) of strand 4. Assuming that the ends of a strand cannot come together in space, this suggests two "active sites," which makes the problem still more difficult. Interesting, antiparallel orientation of strands 3 and 4 would permit Asp 48 and Glu 81 to form a ligand site, possibly a regulatory site binding Ca^{++}, distinct from the active site formed by the Lys-rich region and the conserved Thr-67. If these two active sites are presumed, and if strands 5 and 6 are again assembled antiparallel, conserved Asp-104 and Asp-106 also can participate in the putative "calcium site," which in turn allows the assembly of a fully antiparallel beta sheet structure, found in many beta sandwich structures.[24]

The question then arose as to the order of the seven beta strands in the two sheets. A systematic approach (see Fig. 3)[25] was taken to assemble all possible "connectivities." (We use the word "connectivities" to designate what is elsewhere referred to as "topologies.") The seven strands were represented as an opened beta barrel (Fig. 2). The sequence was then threaded on to these strands and the fold designated by the order in which the strands in Figure 2 are traversed by the polypeptide chain. Thus, the connectivity designated ABCDGFE implies that the polypeptide chain traces the strands in Figure 2 in that order, and that beta 1 corresponds to strand A in the fold, beta 2 corresponds to strand B, beta 3 corresponds to strand C, beta 4 corresponds to strand D, beta 5 corresponds to strand G, beta 6 corresponds to strand F, and beta 7 corresponds to strand E. In the decision tree in Figure 3, beta 1 in the polypeptide chain is always assumed to correspond with the edge strand A in one of the sheets and is arbitrarily designated as pointing "down." The following assumptions were then used to thread the opened barrel in Figure 2:

1. No strands adjacent in a sheet have a parallel relative orientation.

2. No strands consecutive in the polypeptide chain have a parallel relative orientation.

3. Connections between loops are of the "+1," "−1," "+3," "−3," "+2," and "−2" types.[26] This nomenclature arranges beta strands clockwise in order in a barrel (Fig. 2). A "+1" connection joins strain i with strand $i + 1$ (for example, strand A with strand B). A "+3" connection joins strain i with strand i + 3 (for example, strand A with strand D).

According to the constraints above, beta 2 can be associated with only three other strands in Figure 2: via a "+1" connection to strand B, via a "+3" connection to strand D, and via a "−2" connection to strand F. The connections, and the strands in Figure 2 that correspond to beta 1 and beta 2, are designated (for example) "+1 AB" to indicate that beta 1 corresponds to strand A in Figure 2, beta 2 corresponds to strand B in Figure 2, and the two strands

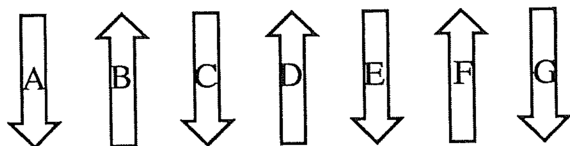

Fig. 2. Template for an opened seven strand beta barrel that forms the core of the C2 homology (C2H) domain.

are joined by a "+1" connection. Figure 3 shows 36 possible connectivities.

Two criteria were then hypothesized to exclude some of these connectivities:

Criterion 1. Two sheets are assumed to be present, one with three strands, one with four strands. Further, beta 4 is an edge strand (see above). Thus, any connectivity where beta 4 is not strand C, D, or G creates a two-strand sheet, which is not allowed under the assumptions above. If consecutive betas in the polypeptide chain are antiparallel (and that beta 1 points "down"), beta 4 must point "up," implying that strand 4 can be only strand D.

Criterion 2. Two connecting polypeptide chains that join by loops two beta strands cannot "cross over."[24] This implies that a loop joining (for example) strand A with strand D excludes a loop joining strand C with strand F (Fig. 2).

Of the 36 connectivities, 16 are excluded by both criterion 1 and 2, 8 are excluded by criterion 1 alone, and 6 are excluded by criterion 2 alone. This leaves just six connectivities. Connectivity ABCDEFG is related in a general toplogical sense to AFEDCBG, ABEDCFG is related in a general topological sense to ABGDEFC, and ABCDGFE is related in a general topological sense of AFCDEBG. Examples of these folds in the literature are retinol binding protein (connectivity ABCDEFG), the pleckstrin homology domain (connectivity ABCDGFE), and pseudoazurin (connectivity ABEDCFG).[22]

As this discussion makes clear, an analysis of this sort requires that the number, nature, and order (but not so significantly the lengths) of secondary structure elements be perfectly assigned in a secondary structure prediction. A single error creates pro-

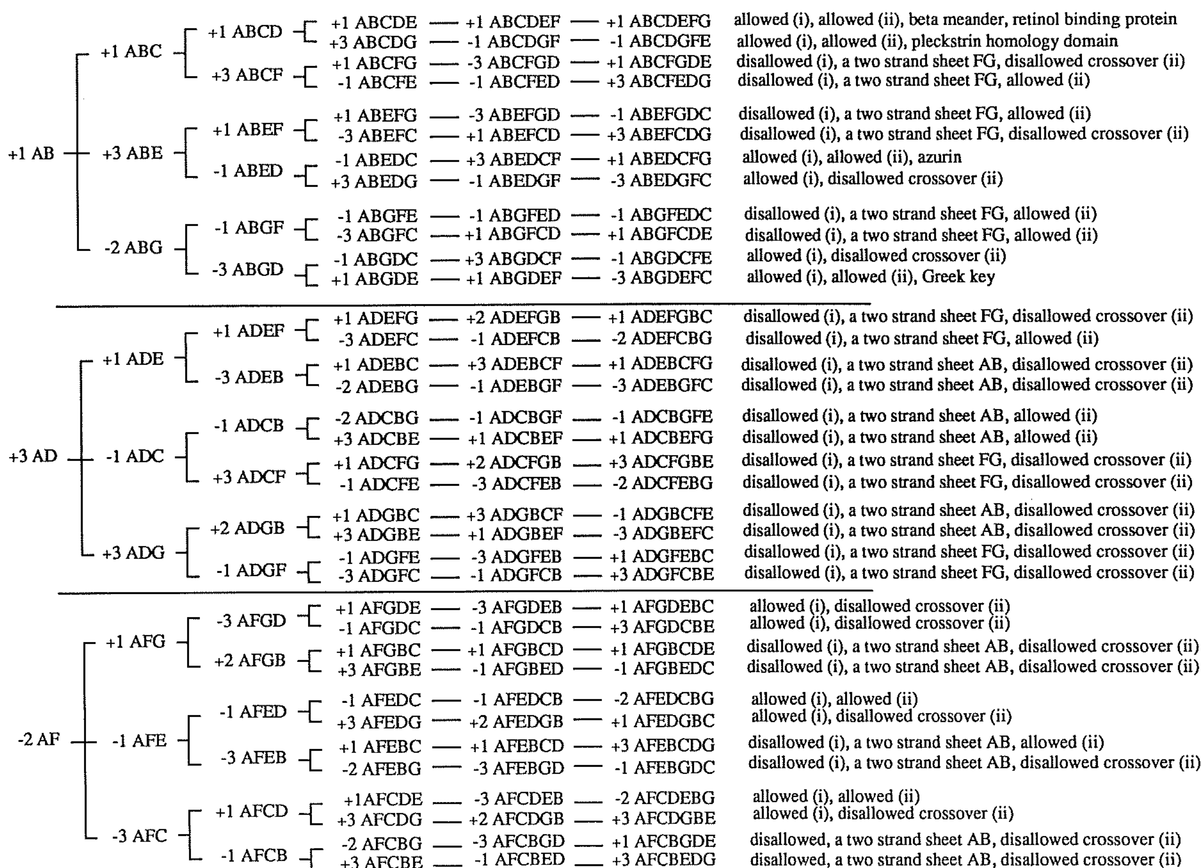

Fig. 3. Decision tree assigning possible beta connectivities. The sequence was then threaded onto the strands in the open beta barrel in Figure 2. The fold is designated by the order in which the strands in Figure 2 are traversed by the polypeptide chain. Thus, the connectivity designated ABCDGFE implies that beta 1 corresponds to strand A in the fold, beta 2 corresponds to strand B, beta 3 corresponds to strand C, beta 4 corresponds to strand D, beta 5 corresponds to strand G, beta 6 corresponds to strand F, and beta 7 corresponds to strand E. Beta 1 in the polypeptide chain is always assumed to correspond with the edge strand A in one of the sheets.

Fig. 4. The alpha carbon tracing of a possible model for the C2H domain based on the crystal structure of retinol binding protein. Putative "active site" residues are highlighted. Sequence numbering is taken from the multiple alignment in Figure 1.

found problems. In this particular example, the reliability of the assignment of a beta structure to strand 4 is critical. If (for example) strand 4 is omitted, then the analysis based on active site residues becomes implausible, and an entirely different packing analysis ensues. Thus, a secondary structure prediction must meet a high standard for it to serve as the starting point for modelling a tertiary structure. Further, the most commonly used methods for scoring secondary structure predictions[27–29] do not adequately measure (and to a large extent are irrelevant to measuring) the value of a secondary structure prediction as a starting point for tertiary structure modeling.[10]

We then asked whether we could infer a preference for one of the six folding connectivities from the available data. Several alternative approaches are conceivable. For example, representative sequences from the C2H domain family might be threaded onto known experimental structures from each family. Because the coordinates for the pleckstrin homology domain were not available, we were able to try this only for the retinol binding protein (ABCDEFG)[30] and the pseudoazurin (ABEDCFG)[31] structures.

The sequences were aligned so as to best overlap the assigned and predicted secondary structures while preserving the phase in the surface and interior assignments. The alignment was then used to generate coordinates for the model structure with a distance geometry routine (G. Chelvanayagam, L. Knecht, S. A. Benner, G. H. Gonnet, unpublished data). In the resulting model (Fig. 4), Asp 48, located at the beginning of strand 3, lies on the wrong side of the sheet face and is too distant to be part of a "calcium binding site." Likewise, Asp-104 and Asp-106 are distant from other binding sites. Thus, this threading yields a model that does not support the "two active site" speculation noted above.

The large number of insertions in the C2H domain made the alignment with pseudoazurin difficult. Further, pseudoazurin has no strand that is an an-

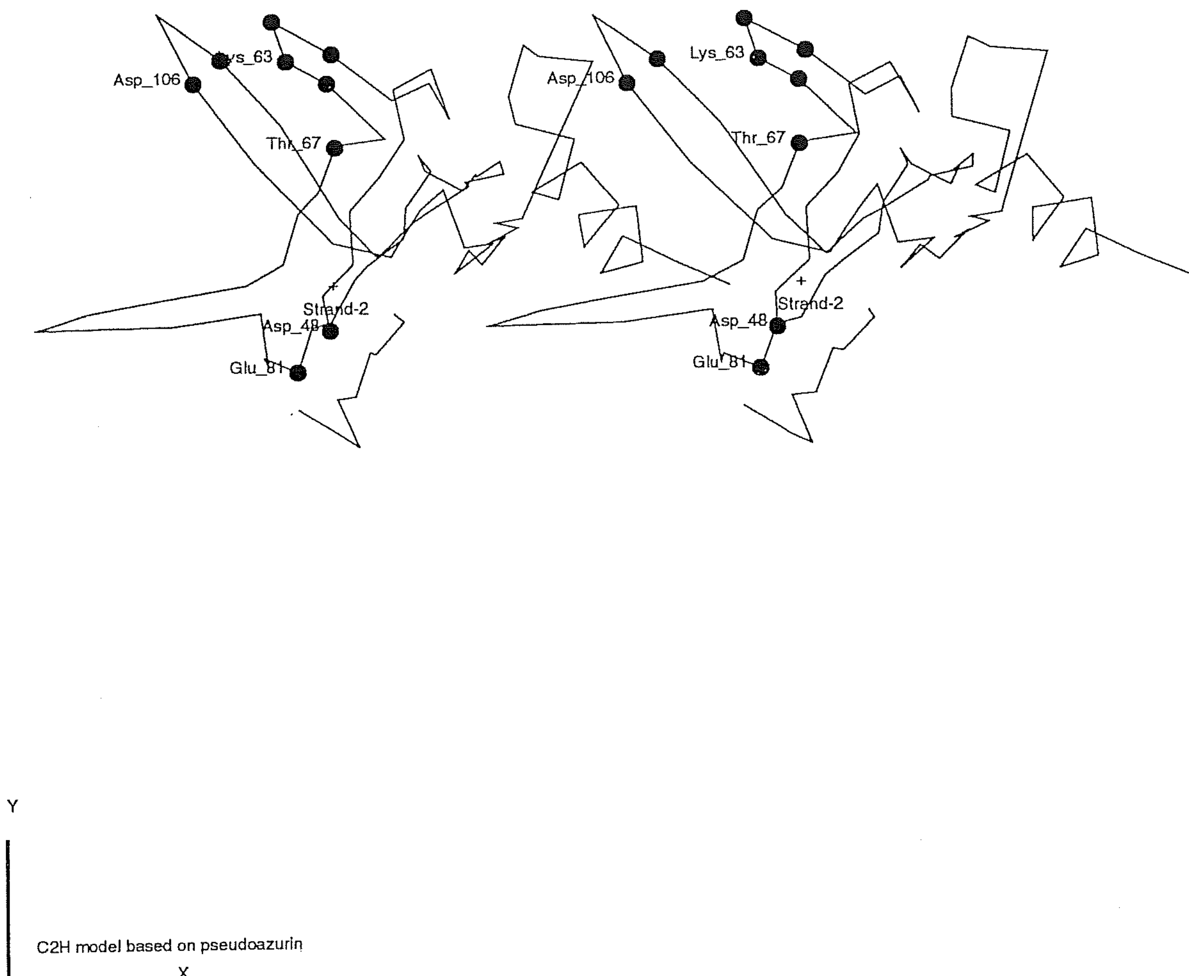

Fig. 5. The alpha carbon tracing of a possible model for the C2H domain based on the crystal structure of pseudoazurin. The first beta strand and loop (positions 35–49) were poorly built and therefore omitted. Strand 1 should bond antiparallel to strand 2. Putative "active site" residues are highlighted. Sequence numbering is from Figure 1.

alog of strand 1 in the C2H domain. In the resulting structure (Fig. 5), Glu 81 cannot participate as an "active site" residue. However, reconfiguration of the loop containing Asp-104 and Asp-106 would allow the formation of a pocket containing these as Asp-48, a potential metal binding site. The Lys-rich region and Thr-67 segregate together at the other end of the model, consistent with the "two active site" speculation.

Because coordinates were not available for the pleckstrin homology domain, a model was interpreted from the hydrogen bonding pattern in the published work. Again, two distinct active sites were formed (Fig. 6). However, although Asp-48 is somewhat distant from Asp-104, Asp 106, and Glu 81, these residues might be brought together by introducing more twists in the strands.

Such model building is, of course, not conclusive. We prefer the ABCDGFE packing, but simply because it comes closest to fitting the "two active site" model. This model is, of course, far from secure. Better would be a full packing of the possible structures using force fields or pairwise contact potentials, dynamics, and, perhaps, a compensatory covariation analysis.[23] From these, a more informed judgment might be possible about which structure is preferred. Because the experimental structure was set to be announced on October 1, 1994, there was not enough time to approach the problem in this direction. Nevertheless, we can use this opportunity to encourage again those who work in this area to consider the application of their tools to assemble tertiary structural models from well-defined secondary structure predictions such as those described here.[32] If it is possible to model tertiary structure from sequence data alone, it should certainly be possible to model tertiary structure from sequence data together with a reliable secondary structure prediction, and still easier to choose the best of six alternative tertiary structural models.

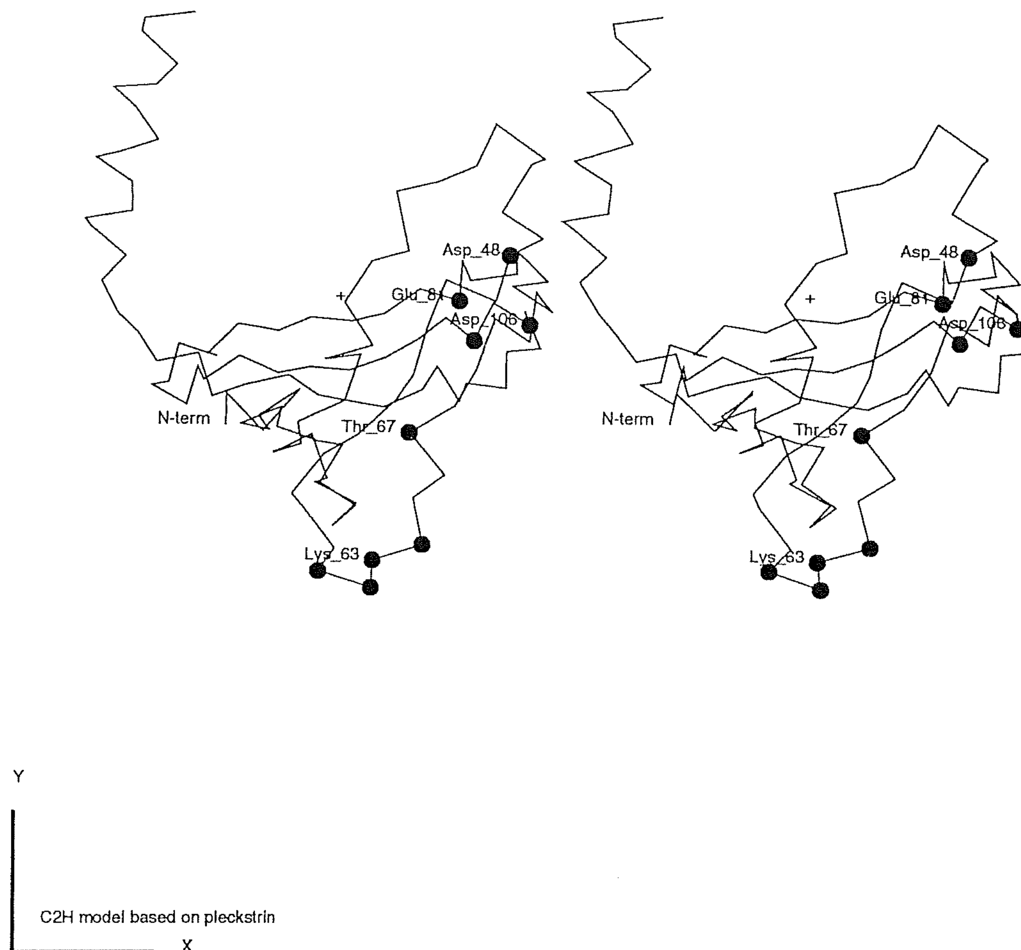

Fig. 6. The alpha carbon tracing of a possible model for the C2H domain based on the hydrogen bonding network from the NMR structure of the pleckstrin homology domain.[34] Putative "active site" residues are highlighted. Sequence numbering is from Figure 1.

ALTERNATIVE APPROACHES

Alternative approaches for analyzing the six connectivities were also explored. Each of these are hypothesis bordering on speculation, and we do not wish to imply by this analysis anything other than that we are using a bona fide prediction opportunity to test some unorthodox ideas. This is, of course, one of the great values of the bona fide prediction setting.

Strand Exposure

The different beta strands have decreasing exposure to solvent in the following order: (most exposed) $4 > 1 > 5 \approx 2 > 7 \approx 3 > 6$ (least exposed). With other factors being equal, the most exposed strands are expected to be edge strands in a beta sheet, while the least exposed should be in the middle of a beta sheet. Complicating the analysis in the C2H domain is the helix, which will bury some of the external side chains of beta strands in the barrel.

Unfortunately, the six preferred connectivities are quite similar in terms of the disposition of the various beta strands. Thus, strand exposure can be only a weak criterion for distinguishing between the various connectivities. All connectivities place strand 6 at an internal position in one of the sheets (Table II). None place strand 2 on the edge. Instead, all place either strand 7 or strand 3 on the edge. We may give the ABGDEFC connectivity a "demerit" because it places strand 3 on the edge and strand 7 inside, when there is a slight preference for the other arrangement. We may give the ABEDCFG connectivity a "demerit" because it places all three internal strands (3, 6, and 7) in one sheet, leaving no clearly internal strands for the other sheet. However, it is equally plausible that placing all three largely internal strands on one sheet is acceptable, as the helix approaches the domain from this side.

TABLE II. Overall Surface Exposure of Beta Strands as a Criterion for Assessing Sheet Connectivity*

	Beta strand						
Connectivity	1	2	3	4	5	6	7
ABCDEFG	Edge	In	In	Edge	Edge	In	Edge
ABCDGFE	Edge	In	In	Edge	Edge	In	Edge
ABEDCFG	Edge	In	In	Edge	Edge	In	Edge
ABGDEFC	Edge	In	Edge	Edge	Edge	In	In
AFEDCBG	Edge	In	In	Edge	Edge	In	Edge
AFCDEBG	Edge	In	In	Edge	Edge	In	Edge
Prediction	Edge	Edge	In	Edge	Edge	In	In

*Edge and interior positions for a strand are judged by the number of surface and interior residues assigned in the strand.

TABLE III. Lengths of Loops Between Beta Strands as a Criterion for Assessing Sheet Connectivity*

	Loop*					
Connectivity	12	23	34	45	56	67
ABCDEFG†	Short	Short	Short	Short	Short	Short
ABCDGFE†	Short	Short	Short	Long	Short	Short
ABEDCFG	Short	Long	Short	Short	Long	Short
ABGDEFC†	Short	Medium	Long	Short	Short	Long
AFEDCBG	Medium	Short	Short	Short	Short	Medium
AFCDEBG	Medium	Long	Short	Short	Long	Medium
Fact	Short	Long	Medium	Medium	Short	Long

*Loops are designated by the number of the strand that precedes them and follow them (e.g., loop 23 is the loop connecting strand 2 and strand 3). Loop lengths designated "short" have five or fewer residues. Loop lengths designated "medium" have six to ten residues. Loop lengths designated "long" have more than ten residues. Connectivities marked with a dagger are preferred by this criterion.

Loop Lengths

The remarkable similarities between the connectivities being considered raises the question: What could the protein chains themselves be using to prefer one fold over another? One possibility is that the fold is determined by the loop length, where shorter loops between strands identify strand pairs that are adjacent in the sheet, and longer loops indicate crossovers. Studies with known structures show that such a rule cannot possibly be absolute. However, it might be useful to create a preference for one fold over another, where we consider a short loop as a strong indicator that the strands joined by the loop are adjacent in a sheet. The opposite inference, that a long loop indicates that the strands joined by the loop are not adjacent in a sheet, cannot be made (for empirical reasons).

The loop lengths are ranked (shortest to longest) (12) < (56) < (34) < (45) < (67) < (23) (Table III). The ranking suggests that that beta 1 and beta 2 are adjacent (antiparallel) in the sheet. This is the case for connectivities ABCDEFG, ABCDGFE, ABEDCFG, and ABGDEFC. However, loop 56 is also short, implying that strands 5 and 6 are adjacent in a sheet. This is the case only in the connectivities ABCDEFG, ABCDGFE, ABGDEFC, and AFEDCBG. The intersection of the two sets identifies two preferred connectivities, the ABCDEFG connectivity and the ABCDGFE connectivity.

CONCLUSIONS

In this discussion, it is important to recognize that several elements of the analysis are new, potentially controversial and, more importantly, untested. In particular, the proposal of two distinct "active sites," the collection of functionality separately into these two sites, and the construction of an antiparallel beta sheet structure based on this separation, is pure inference. Further, there is little experience to suggest that the order of strands within a sheet can be assigned from the predicted overall surface exposure of the individual strands. Finally, it should be kept in mind that the sheets in a beta sandwich are normally twisted with respect to each other. Thus, the analysis based on loop lengths is tenuous and is applicable in the best case only if the strands and sheets are relatively flat. This prediction is interesting as a test of these conjectures.

The criteria discussed above suggest that we prefer the ABCDEFG and ABCDGFE connectivities over all other antiparallel connectivities. Interestingly, the ABCDGFE connectivity is the same as that found in the pleckstrin homology domain, recently determined by nuclear magnetic resonance (NMR) spectroscopy in two laboratories.[33,34] The secondary structure of the pleckstrin homology domain was also accurately predicted by two groups before the experimental structure became available.[11,12] The pleckstrin homology domain binds to

phosphatidylinositol-4,5-bisphosphate.[35] Interestingly, recombinant p65 binds to phosphatidylserine. Phosphatidylserine is also an activator of PKC. The binding site is not known, but is believed not to be the same as the binding site for diacylglycerol and phorbol esters in the C1 domain,[36] implying that the C2 domain in PKC might interact with phosphatidylserine. It is tempting from these facts to speculate that the C2H domain is a distant homolog of the pleckstrin homology domain. The comparison is, however, not entirely straightforward. For example, no evidence was found that Ca^{++} binds to the p65 protein.[14] Further, although the C2 domain of PKC has been suspected as having a site for binding calcium, this has not been proved to be the case. We must await an experimental structure.

NOTE ADDED IN PROOF

A crystal structure of synaptotagmin, just determined[37] shows that the protein does in fact adopt one of the three preferred folding connectivities predicted here, the ABEDCFG topology shown in Figure 5.

REFERENCES

1. Benner, S.A. Patterns of divergence in homologous proteins as indicators of tertiary and quaternary structure. Adv. Enzyme Regul. 31:121–181, 1989.
2. Pananoyotou, G., Bax, B., Gout, I., Federwisch, M., Wroblowski, B., Dhand, R., Fry, M.J., Blundell, T.L., Wollmer, A., Waterfield, M.D. Interaction of the p85 subunit of PI 3-kinase and its N-terminal SH2 domain with a PDGF receptor phosphorylation site. EMBO J. 11:4261–4272, 1992.
3. Russell, 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.
4. 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.
5. Bazan, J.F. Structural design and molecular evolution of a cytokine receptor superfamily. Proc. Natl. Acad. Sci. USA 87:6934–6937, 1990.
6. Moe, G.R., Koshland, D.E. Jr. Transmembrane signalling through the aspartate receptor. In: "Microbial Energy Transduction, Genetics, Structure and Function of Membrane Proteins." Youvan, D.C., Daldal, F., eds. Cold Spring Harbor, New York: The Cold Spring Harbor Laboratory Press, 1986; pp. 163–168.
7. Benner, S.A., Ellington, A.D. Evolution and structural theory: The frontier between chemistry and biochemistry. Bioorg. Chem. Frontiers 1:1–70, 1990.
8. Chothia, C., Lesk, A.M. The relation between the divergence of sequence and structure in proteins. EMBO J. 5:823–826, 1986.
9. Benner, S.A., Gerloff, D.L., Jenny, T.F. Predicting protein crystal structures. Science 265:1642–1644, 1994.
10. Jenny, T.F., Benner, S.A. Evaluating predictions of secondary structure in proteins. Biochem. Biophys. Res. Commun. 200:149–155, 1994.
11. Jenny, T.F., Benner, S.A. A prediction of the secondary structure of the pleckstrin homology domain. Proteins 20: 1–4, 1994.
12. Musacchio, A., Gibson, T., Rice, P., Thompson, J., Saraste, M. The pH domain—a common piece in the structural patchwork of signaling proteins. Trends Biochem. Sci. 18: 343–348, 1993.
13. 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.
14. Perin, M.S., Fried, V.A., Mignery, G.A., Jahn, R., Suedhof, T.C. Phospholipid binding by a synaptic vesicle protein homologous to the regulatory region of protein kinase C. Nature 345:260–263, 1990.
15. Ono, Y., Fujii, T., Ogita, K., Kikkawa, U., Igarashi, K., Nishizuka, Y. The structure, expression and properties of additional members of the protein kinase C family. J. Biol. Chem. 263:6927–6932, 1988.
16. Ohno, S., Akita, Y., Konno, Y., Imajoh, S., Suzuki, K. A novel phorbol ester receptor/kinase, nPKC, distantly related to the protein kinase C family. Cell 53:731–741, 1988.
17. Stahl, M.L., Ferenz, C.A., Kelleher, K.L., Kriz, R.W., Knopf, J.L. Sequence similarity of phospholipase C with the non-catalytic region of src. Nature 332:269–272, 1988.
18. Clark, J.D., Lin, L.-L., Kriz, R.W., Ramesha, C.S., Sultzman, L.A., Lin, A.Y., Milona, N., Knopf, J.L. A novel arachadonic acid-selective cytosolic PLA2 contains a Ca(2+)-dependent translocation domain with homology to PKC and GAP. Cell 65:1043–1051, 1991.
19. Bairoch, A., Boeckmann, B. The SWISS-PROT protein sequence data bank. Nucleic Acids Res. 20:2019–2022, 1992.
20. Gonnet, G.H., Benner, S.A. Computational Biochemistry Research at ETH. Technical Report 154, Departement Informatik, March, 1991.
21. Gonnet, G.H., Cohen, M.A., Benner, S.A. Exhaustive matching of the entire protein sequence database. Science 256:1443–1445, 1992.
22. 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.
23. 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.
24. Woolfson, D.N., Evans, P.A., Hutchinson, E.G., Thornton, J.M. Topological and stereochemical restrictions in beta-sandwich protein structures. Prot. Eng. 6:461–470, 1993.
25. Cohen, F.E., Kuntz, I.D. Tertiary structure prediction. In: "Prediction of Protein Structure and the Principles of Protein Conformation." Fasman, G.D., ed. New York: Freeman, 1989:647–705.
26. Richardson, J.S., Richardson, D.C., Thomas, K.A., Silverton, E.W., Davies, D.R. Similarity of three dimensional structure between the immunoglobulin domain and the copper-zinc superoxide dismutase subunit. J. Mol. Biol. 102:221–235, 1976.
27. 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.
28. Robson, B., Garnier, J. Protein structure prediction. Nature 361:506, 1993.
29. Rost, B., Sander, C., Schneider, R. Redefining the goals of protein secondary structure prediction. J. Mol. Biol. 235: 13–26, 1994.
30. Zanotti, G., Berni, R., Monaco, H.L. Crystal structure of liganded and unliganded forms of bovine plasma retinol binding protein. J. Biol. Chem. 268:10727–10738, 1993.
31. Petratos, K., Banner, D.W., Beppu, T., Wilson, K.S., Tsernoglou, D. The crystal structure of pseudoazurin from *Alkaligenes faecalis* SS-6 determined at 2.9 Å resolution. FEBS Lett. 218:209–214, 1987.
32. Monge, A., Friesner, R.A., Honig, B. An algorithm to generate low, resolution protein tertiary structures from knowledge of secondary structure. Proc. Natl. Acad. Sci. USA 91:5027–5029, 1994.
33. Yoon, H.S., Hajduk, P.J., Petros, A.M., Olejniczak, E.T., Meadows, R.P., Fesik, S.W. Solution structure of a pleckstrin-homology domain. Nature 369:672–675, 1994.
34. Macias, M.J., Musacchio, A., Ponstingl, H., Nilges, M., Saraste, M., Oschkinat, H. Structure of the pleckstrin homology domain from beta-spectrin. Nature 369:675–677, 1994.
35. Harlan, J.E., Hajduk, P.J., Yoon, H.S., Fesik, S.W. Pleckstrin homology domains bind to phosphatidylinositol-4,5,-bisphosphate. Nature 371:168–170, 1994.
36. Huang, F.L., Huang, K.P. Interaction of protein kinase C isozymes with phosphatidylinositol 4,5-bisphosphate. J. Biol. Chem. 266:8727–8733, 1991.
37. Sutton, R.B., Davtetov, B.A., Berghuis, A.M., Sudhof, T.C., Sprang, S.R. Structure of the first C-2 domain of synaptotagmin. A novel Ca^{2+}/phospholipid binding mode. Cell 80:929–938, 1995.