Supporting Information

Kister and Gelfand 10.1073/pnas.0909714106

Table S1. Sequence alignment of proteins with the SSS shown in Fig. 1*A*

PDB code	Chain	Start	End	Strand 1	Loop	Strand 2	Loop	Strand 3	Loop
							1		
				12	345	6 7 8	9 0	12345	67 8
1kcr	H	117	218	...SVYP...	...AAA...	...LGCLVK...	...PEP...	...VTLTWH...	...GSLS...
1m7d	B	114	213	...SVYP...	...GSS...	...LGCVVK...	...PEP...	...VTVKWN...	...GALS...
2fbj	H	119	220	...TIYP...	...SSD...	...IGCLIH...	...PSG...	...MNVTWG...	...SGKD...
1ow0	A	242	342	...SLHR...	...GSE...	...LTCTLT...	...GLR...	...VTFTWT...	...SSGK...
1fp5	A	336	438	...SAYL...	...KSP...	...ITCLVV...	...PSK...	...VNKTWS...	...SG-K...
1hxm	A	121	206	...SVFV...	...NGT...	...VACLVK...	...PKD...	...IRINLV...	...SK-K...
1c16	A	181	276	...KAHV...	...EGD...	...LRCWAL...	...PAD...	...ITLTWQ...	...GE-E...
1svb	A	303	395	...TWKR...	...SGH...	...VVMEVT...	...GTK...	...IPVRAV...	...GSPD...
1oke	A	298	398	...KFKV...	...HGT...	...IVIRVQ...	...GDG...	...IPFEIM...	...EKRH...
1f42	A	88	211	...TFL-...	...SGR...	...FTCWWL...	...STD...	...LTFSVK...	...GS-S...

	Strand 4	Loop	Strand 5	Loop	Strand 6	Loop	Strand 7
		2					3
	9	0	1	2 3	4 56	7 8	9 O
1kcr	...-FHTFP...	...QSQ...	...SSSVT...	...TS-SP...	...VTCNV...	...PAS...	...KVDKK...
1m7d	...-VRTVS...	...QSG...	...SSLVT...	...PS-ST...	...VICNV...	...PAS...	...DLIKE...
2fbj	...TVNFPP...	...SGG...	...SNQLT...	...P--AV...	...VKCSV...	...DSN...	...ELDVN...
1ow0	...AVQGPP...	...CG-...	...VSSVL...	...GC-AE...	...FTCTA...	...PES...	...PLTAT...
1fp5	...-VNHST...	...NG-...	...TSTLP...	...G--TR...	...YQCRV...	...P-H...	...ALMRS...
1hxm	...EFDPAI...	...PSG...	...GVKLG...	...ED-SN...	...VTCSV...	...D-N...	...TVHST...
1c16	...-MELVE...	...AGD...	...WAAVV...	...PL-GK...	...YTCHV...	...E-G...	...DPILR...
1svb	...-VAMLI...	...NPT...	...FIEMQ...	...PL-P-...	...IYVG-...	...ELS...	...QWFQK...
1oke	...RLITV-...	...NPI...	...-VNIE...	...EP-PF...	...IIIGV...	...EPG...	...NWFKK...
1f42	...AATLSA...	...-GD...	...EYSVE...	...QEDSA...	...VMVDA...	...KLK...	...NYTSS...

Protein Data Bank (PDB) code of proteins is provided in the first column, chain name in the second column (″chain″), and the beginning and end positions of each sequence in a domain in the third and fourth (″start″ and ″end″) columns, respectively. The sequences of the strands and loops are not shown in their entirety, and the ″skipped″ segments of sequences are indicated by ellipses (. . .). The spaces within the sequences separate between secondary structure units (strands and loops). The conserved hydrophobic and hydrophilic positions are numbered. Per the SCOP database, structures 1C5C, 1NCW, 1KCR, 1M7D, 1IK3, 1FBJ, 1OW0, 1FP5, 1HXM, and 1C16 belong to Superfamily Ig (Family: C1-set domains); structures 1SVB and 1OKE to Superfamily E-set domains (Family: Class II viral fusion proteins C-terminal domain); and structure 1F42 to Superfamily Fibronectin (Family: Fibronectin type III).

Table S2. Sequence alignment of proteins with the SSS shown on Fig. 1*B*

PDB				Strand 1	Loop	Strand 2	Loop	Strand 3	Loop
code	Chain	Start	End				1		
				1 2	34 5	6 7	8 90 1	2 3	45 6
1gxe	A			…VVAG…	…NK-LR…	…LDVPI…	…D—-PA-P…	…VIWQ-…	…KAITQ-G…
1bih	A	5	98	…VLFR…	…EN-NP…	…LECI-…	…N—-DQ-G…	…YSWK-…	…KD----G…
2cs6	A			…LFPE…	…EADIG…	…WSCVA…	…GK-PR-P…	…VRWLR…	…DGQ---P…
1o6v	A	417	496	…WTNA…	…NYKAN…	…VSIP-…	…N—-KA-D…	…ATIS-…	…DG----G…
1nci	A			…INLP…	…NSRGP…	…VRIRS…	…D--KNLS…	…YSVTG…	…DQ----P…
1l3w	A	1	100	…IKVS…	…NERGP…	…VQIKS…	…NSMKD-R…	…YSITG…	…QGAADNP…

	Strand 4	Loop	Strand 5	Loop	Strand 6	Loop	Strand 7
			2				3
	78	9	0 1	23 4	5 6 7	8 9	01
1gxe	…RVET…	…K…	…IFTV-…	…GAEKE…	…VYTVTVKN…	…PVGE…	…VNL…
1bih	…ALRK…	…G…	…SLVFL…	…QASDE…	…HYQCFAET…	…PAG-…	…VAS…
2cs6	…EVS-…	…G…	…ELRFS…	…KLVLE…	…MYQCVAEN…	…KHGT…	…LTV…
1o6v	…IYT-…	…P…	…-WNLP…	…SYTGE…	…-VTYKISD…	…TIGK…	…ISE…
1nci	…IINP…	…G…	…QLSVT…	…PLDRE…	…RFHLRAHA…	…N-GN…	…VEN…
1l3w	…RIEW…	…G…	…WMLVT…	…PLDRE…	…KYVLSSHA…	…ENGS…	…VEE…

This table follows the format of Table S1. According to the SCOP database, 1GXE, 1IRA, 1BIH, and 1CS6 protein structures are classified into Superfamily Ig (Family: I-set domains); 1F1S into Superfamily E-set domains (Family: E-set domains of sugar-utilizing enzymes); 1O6V and 1H6T into Superfamily: E-set domains (Family: Internalin); and 1NCI, 1EDH, 1O6S, and 1L3W into Superfamily Cadherin (Family: Cadherin). PDB, Protein Data Bank.