



STRUCTURAL
BIOLOGY

Volume 77 (2021)

Supporting information for article:

The β -link motif in protein architecture

David P. Leader and E. James Milner-White

Supplementary Table S1 **β -links in different SCOPe classes represented in the *Protein Motif 2* database**

Class b folds have been restricted to those with at least four entries, and class c or d folds to those with at least 10 entries.

	Database entries	β -Links			
		Raw	E	H Bond	
b.1	Immunoglobulin-like β -sandwich	89	44	15	15
b.2	Common fold of diphtheria toxin/TFs/cyt. f	17	8	7	4
b.3	Prealbumin-like	11	0	0	0
b.6	Cupredoxin-like	28	36	36	23
b.7	C2 domain-like	11	2	2	2
b.11	Gamma-Crystallin-like	8	0	0	0
b.18	Galactose-binding domain-like	28	7	6	6
b.21	Virus attachment protein globular domain	7	2	0	0
b.22	TNF-like	4	5	4	4
b.29	Concanavalin A-like lectins/glucanase	40	15	14	13
b.33	ISP domain	7	5	5	5
b.34	SH3-like barrel	37	33	31	31
b.35	GroES-like	7	7	7	7
b.36	PDZ domain-like	30	30	26	2
b.40	OB-fold	34	21	21	21
b.42	β -Trefoil	18	14	7	4
b.43	Reductase/isomerase/EF common domain	8	7	7	7
b.45	Split barrel-like	17	2	1	1
b.47	Trypsin-like serine proteases	9	15	15	11
b.49	Domain of α and β subunits of F1 ATP synthase-like	4	6	5	5
b.50	Acid proteases	9	0	0	0
b.52	Double psi β -barrel	4	4	4	4
b.55	PH domain-like barrel	6	0	0	0
b.60	Lipocalins	23	3	3	2
b.61	Streptavidin-like	6	0	0	0
b.62	Cyclophilin-like	9	11	2	2
b.68	6-bladed β -propeller	12	1	1	0
b.69	7-bladed β -propeller	4	1	0	0
b.71	Glycosyl hydrolase domain	7	4	4	3
b.74	Carbonic anhydrase	6	6	5	5
b.80	Single-stranded right-handed β -helix	9	2	2	2
b.81	Single-stranded left-handed β -helix	11	0	0	0
b.82	Double-stranded beta-helix	33	47	38	34
b.84	Barrel-sandwich hybrid	10	15	15	9
b.85	β -clip	14	23	17	6
b.115	Calcium-mediated lectin	4	0	0	0
b.121	Nucleoplasmin-like/VP	6	2	2	2
b.122	PUA domain-like	5	3	3	3
		592	381	305	233

Supplementary Table S1 (cont.)

	Database entries	β-Links			
		Raw	E	H Bond	
c.1	TIM beta/α-barrel	186	5	4	2
c.2	NAD(P)-binding Rossmann-fold domains	93	2	1	1
c.3	FAD/NAD(P)-binding domain	20	0	0	0
c.6	7-stranded β/α barrel	13	5	0	0
c.14	ClpP/crotonase	21	0	0	0
c.23	Flavodoxin-like	61	0	0	0
c.26	Adenine nucleotide alpha hydrolase-like	21	0	0	0
c.37	P-loop containing NTP hydrolases	81	3	0	0
c.41	Subtilisin-like	12	0	0	0
c.45	Phosphotyrosine phosphatases II	13	2	2	2
c.47	Thioredoxin fold	75	15	1	1
c.55	Ribonuclease H-like motif	43	6	4	3
c.56	Phosphorylase/hydrolase-like	19	9	9	10
c.61	PRTase-like	14	0	0	0
c.66	SAM-dependent methyltransferases	33	17	16	12
c.67	PLP-dependent transferase-like	42	3	0	0
c.68	Nucleotide-diphospho-sugar transferases	13	0	0	0
c.69	α/β-Hydrolases	42	0	0	0
c.72	Ribokinase-like	14	0	0	0
c.80	SIS domain	10	1	0	0
c.82	ALDH-like	11	0	0	0
c.92	Chelatase-like	12	1	1	1
c.93	Periplasmic binding protein-like I	13	3	3	3
c.94	Periplasmic binding protein-like II	34	1	1	1
c.95	Thiolase-like	37	5	5	4
c.97	Cytidine deaminase-like	10	0	0	0
c.108	HAD-like	42	0	0	0
		985	78	47	40

Supplementary Table S1 (cont.)

	Database entries	β -Links		
		Raw	E	H Bond
d.3 Cysteine proteinases	28	6	4	4
d.15 β -Grasp (ubiquitin-like)	25	0	0	0
d.17 Cystatin-like	41	0	0	0
d.26 FKBP-like $\beta(2)$ - α - $\beta(2)$;antiparallel β -sheet	12	12	11	12
d.32 Glyoxalase/Bleomycin resistance protein	21	3	3	0
d.37 CBS-domain pair	10	0	0	0
d.38 Thioesterase/thiol ester dehydrase-isomerase	18	7	10	10
d.54 Enolase N-terminal domain-like	13	12	0	0
d.58 Ferredoxin-like	89	4	3	3
d.79 Bacillus chorismate mutase-like	19	0	0	0
d.90 FMN-dependent nitroreductase-like	17	0	0	0
d.108 Acyl-CoA N-acyltransferases	23	1	0	0
d.110 Profilin-like	19	7	1	1
d.129 TBP-like	14	3	3	3
d.144 Protein kinase-like	35	0	0	0
d.145 FAD-binding/transporter-assoc. domain-like	11	6	6	6
d.157 Metallo-hydrolase/oxidoreductase	14	9	8	7
d.166 ADP-ribosylation	9	8	3	3
d.169 C-type lectin-like	12	0	0	0
	430	78	52	49

Key

Raw: β -links defined by hydrogen bonds and dihedral angles as in Figure 1 of main text.

E: β -links with four flanking residues having the β R conformation.

H-Bond: β -links with four flanking residues making two hydrogen bonds.

Supplementary Table S2
Conservation of β-links in some different SCOPe Families

Key						
+, β-link						
- †, β-turn only						
- ‡, β-bulge only						
-, no β-link, turn or bulge						

(a) Example of β-sandwich family

b.1.8.1 (Superoxide dismutase)

		-1	0	1	2		
<i>Homo sapiens</i>	1mfm*	I 112	I 113	G 114	R 115	I 149	+
<i>Bos taurus</i>	1cb4	I 110	I 111	G 112	R 113	I 147	+
<i>Xenopus laevis</i>	1xs0	I 110	I 111	G 112	R 113	I 147	+
<i>Bombyx mori</i>	3l9e	I 112	I 113	G 114	R 115	I 149	+
<i>Schistosoma mansoni</i>	1to4	I 111	I 112	G 113	R 114	V 148	+
<i>Saccharomyces cerevisiae</i>	1b4l	V 112	V 113	G 114	R 115	I 149	+
<i>Spinacia oleracea</i>	1srd	V 112	V 113	G 114	R 115	V 149	+
<i>Photobacterium leiognathi</i>	1bzo	L 110	K 111	G 112	H 113	I 147	+
<i>Escherichia coli</i>	1eso	I 110	K 111	D 112	K 113	I 147	+
<i>Mycobacterium tuberculosis</i>	1pzs	G 118	A 119	K 120	T 121	I 167	- †
<i>Salmonella typhimurium</i>	1eqw	L 110	K 111	G 112	H 113	I 147	+
<i>Haemophilus ducreyi</i>	1z9n	V 143	K 144	G 145	H 146	I 176	- ‡
<i>Actinobacillus pleuropneumoniae</i>	2aps	V 133	K 134	G 135	H 136	I 166	- ‡
<i>Brucella abortus</i>	2aqm	I 120	K 121	Q 122	R 123	I 153	+
<i>Neisseria meningitidis</i>	2aqn	V 152	R 153	G 154	H 155	I 185	+
						79%	

*Prototype (excluded from percentage calculation)

(b) Example of small β-barrel family

b.43.4.2 (Ferredoxin Reductase)

		K	P	G	A	C	+
<i>Spinacia oleracea</i>	1fnc	140	141	142	143	42	
<i>Pisum sativum</i>	1qfz	K	P	G	S	C	+
		134	135	136	137	36	
<i>Zea mays</i>	1gaq	Q	P	G	D	C	+
		140	141	142	143	42	
<i>Azotobacter vinelandii</i>	1a8p	K	E	G	D	V	+
		84	85	86	87	9	
<i>Pseudomonas aeruginosa</i>	3crz	K	E	G	D	V	+
		84	85	86	87	9	
<i>Escherichia coli</i>	1fdr*	K	P	G	D	V	+
		83	84	85	86	9	
							100%

(c) Serine proteases

b.47.1.2 (Trypsin)

		D	S	G	G	V	+
<i>Rattus norvegicus</i>	1amh	194	195	196	197	213	
<i>Homo sapiens</i>	1trn	D	S	G	G	V	+
		194	195	196	197	213	
<i>Sus scrofa</i>	1aks	D	S	G	G	V	+
		194	195	196	197	213	
<i>Bos taurus</i>	1bty*	D	S	G	G	V	+
		194	195	196	197	213	
<i>Salmo salar</i>	1a0j	D	S	G	G	V	+
		194	195	196	197	213	
<i>Fusarium oxysporum</i>	1fn8	D	S	G	G	V	+
		194	195	196	197	213	
<i>Pontastacus leptodactylus</i>	2f91	D	S	G	G	V	+
		194	195	196	197	213	
							100%

b.47.1.2 (Serine proteases — eukaryotic)

		D	S	G	G	I	+
Plasminogen activator	1bda	194	195	196	197	213	
<i>Homo sapiens</i>							
Urokinase	1fv9	D	S	G	G	V	+
<i>Homo sapiens</i>		196	197	198	199	215	
Coagulation factor XI	1xx9	D	S	G	G	T	+
<i>Homo sapiens</i>		194	195	196	197	213	
Kallikrein 6	1gvl	D	S	G	G	V	+
<i>Homo sapiens</i>		194	195	196	197	213	
Trypsin	1trn*	D	S	G	G	V	+
<i>Homo sapiens</i>		194	195	196	197	213	
Elastase	1b0f	D	S	G	S	A	+
<i>Homo sapiens</i>		194	195	196	197	213	
Granzyme A	1op8	D	S	G	S	T	+
<i>Homo sapiens</i>		194	195	196	197	213	
							100%

b.47.1.1 (Serine proteases — prokaryotic)

Stress sensor protease DegS <i>Escherichia coli</i>	1sot	N	S	G	G	N	+
		200	201	202	203	216	
Glutamyl endopeptidase <i>Bacillus intermedius</i>	1p3c	N	S	G	S	H	+
		170	171	172	173	186	
Protease I <i>Achromobacter lyticus</i>	1arb	S	S	G	S	L	+
		193	194	195	196	209	
α -Lytic Protein <i>Lysobacter enzymogenes</i>	2alp	D	S	G	G	M	+
		194	195	196	197	213	
Epidermal Toxin A <i>Staphylococcus aureus</i>	1agj	N	S	G	S	H	+
		194	195	196	197	210	
V8 protease <i>Staphylococcus aureus</i>	1qy6	N	S	G	S	H	+
		168	169	170	171	184	
							100%

b.47.1.3 (Serine proteases — viral)

NS3 protease <i>West Nile virus</i>	2fom	T	S	G	S	Y	+
		134	135	136	137	150	
NS4 protease <i>Equine arteritis virus</i>	1mbm	D	S	G	S	H	+
		119	120	121	122	134	
Capsid protein <i>Semliki forest virus</i>	1vcp	D	S	G	R	V	+
		218	219	220	221	234	
							100%

(d) Cysteine proteases (trypsin-like)

b.47.1.4 (Cysteine proteases — viral)

C3 protease <i>Hepatitis A Virus</i>	1hav	M	C	G	G	H	+
		171	172	173	174	191	
Protease (cys) <i>Tobacco Etch Virus (Classified as 47.1.3)</i>	1vlm	N	C	G	S	H	+
		150	151	152	153	167	
Main protease <i>Human coronavirus 229E (SARS-CoV)</i>	1p9s	A	C	G	S	H	+
		143	144	145	146	162	
Mpro <i>Human SARS-CoV-2</i>	7bqy	S	C	G	S	H	+
		144	145	146	147	163	
							100%

Supplementary Table S3
Amino Acid Composition of β -links in β -sandwiches and small β -barrels

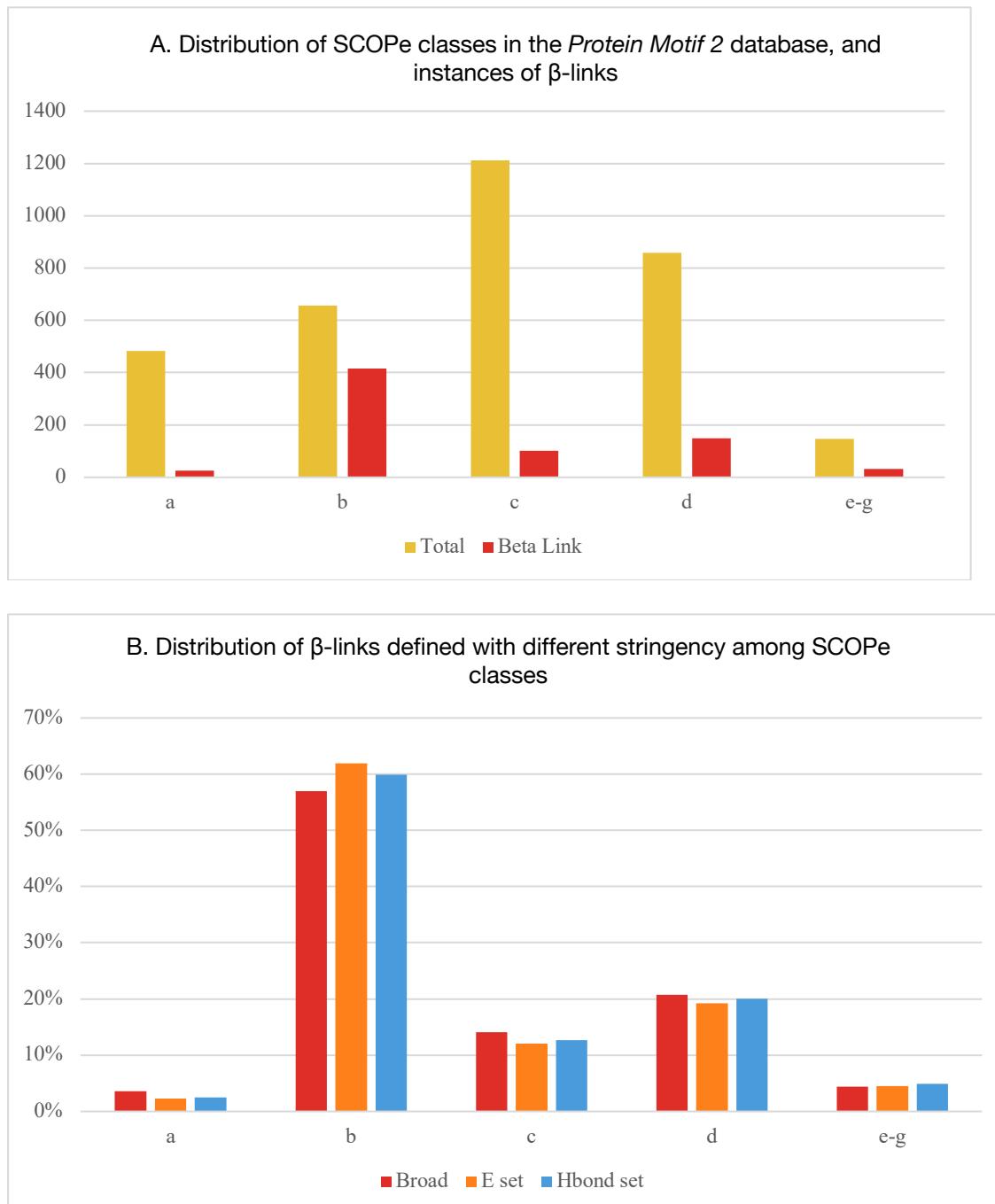
Position -1	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
Sandwich	8	5	7	4	0	7	9	5	4	3	3	16	0	1	7	11	3	1	2	4
Barrel	8	10	2	4	2	8	7	3	5	1	6	14	2	3	6	5	2	0	6	4
Position 0	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
Sandwich	12	3	1	4	0	7	10	2	2	3	1	10	0	2	22	5	2	1	1	12
Barrel	11	4	6	5	1	2	10	1	1	10	1	11	0	0	17	7	4	0	1	8
Position 1	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
Sandwich	2	1	2	3	0	1	1	87	0	0	0	1	1	0	0	1	0	0	1	0
Barrel	0	1	5	1	0	1	3	84	0	0	1	1	0	1	0	1	0	1	0	0
Position 2	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
Sandwich	5	3	2	27	1	8	20	4	2	0	3	8	2	1	0	5	6	1	1	1
Barrel	4	4	1	41	2	4	9	4	1	0	1	4	1	1	0	13	9	0	1	2
Position X	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
Sandwich	16	3	14	2	3	1	0	1	3	10	11	0	2	8	0	2	1	1	1	21
Barrel	11	0	1	1	2	0	0	2	1	34	4	1	0	0	0	1	1	0	4	36

The five positions (-1, 0, 1, 2, X) are as in Figure 1. The numbers are the percentages of the total for β -links in a 182 β -sandwiches or 166 small β -barrels contributed by each individual amino acid.

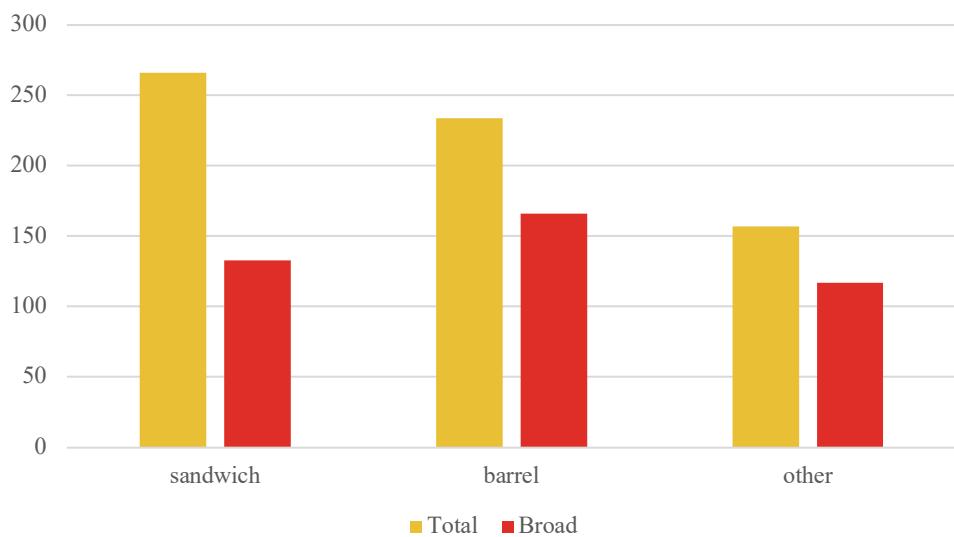
Supplementary Figure S1

Distribution of β -links among SCOPe classes and sub-classes

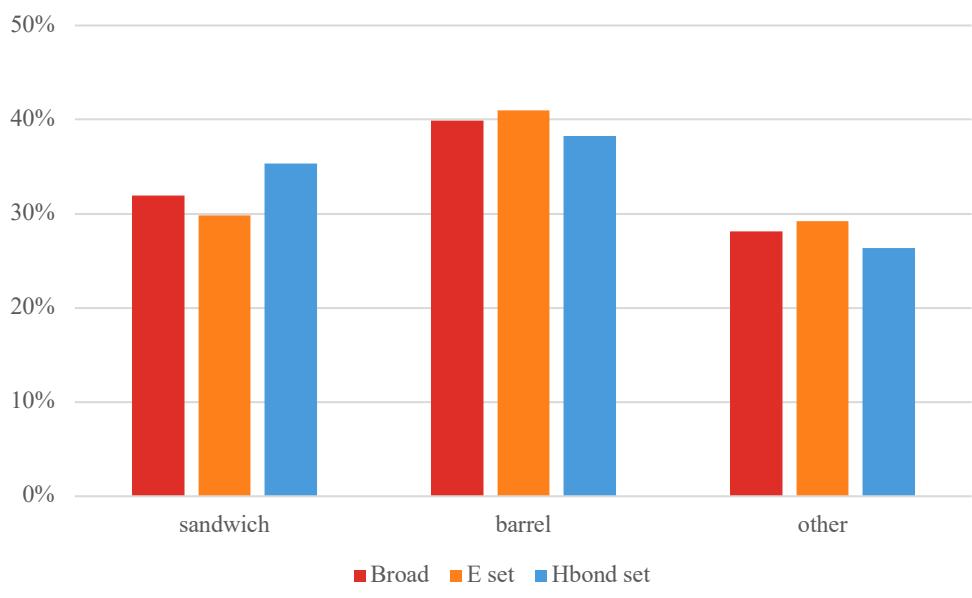
- Broad: β -links defined by hydrogen bonds and dihedral angles as in Figure 1 of main text.
E: β -links with four flanking residues having the βR conformation.
H-Bond: β -links with four flanking residues making two hydrogen bonds.



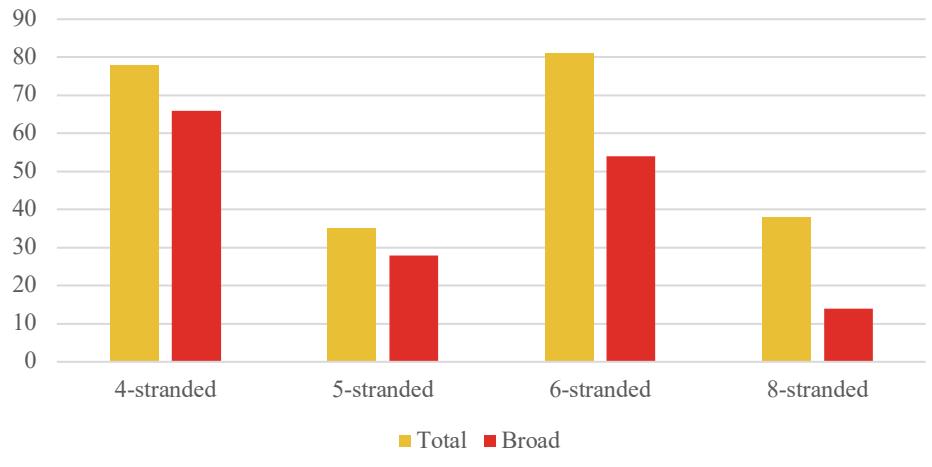
C. Distribution of three general SCOPe class b fold types in the *Protein Motif 2* database, and instances of β -links



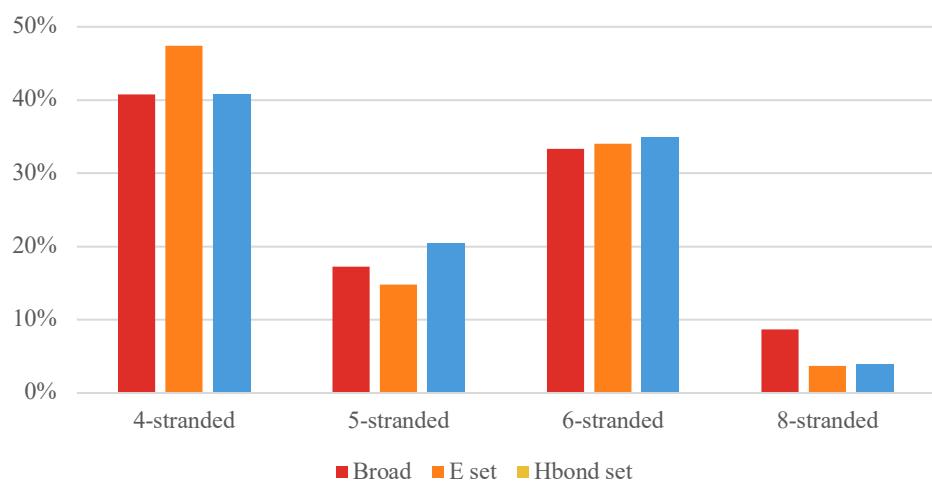
D. Distribution of β -links defined with different stringency among three general SCOPe class b fold types



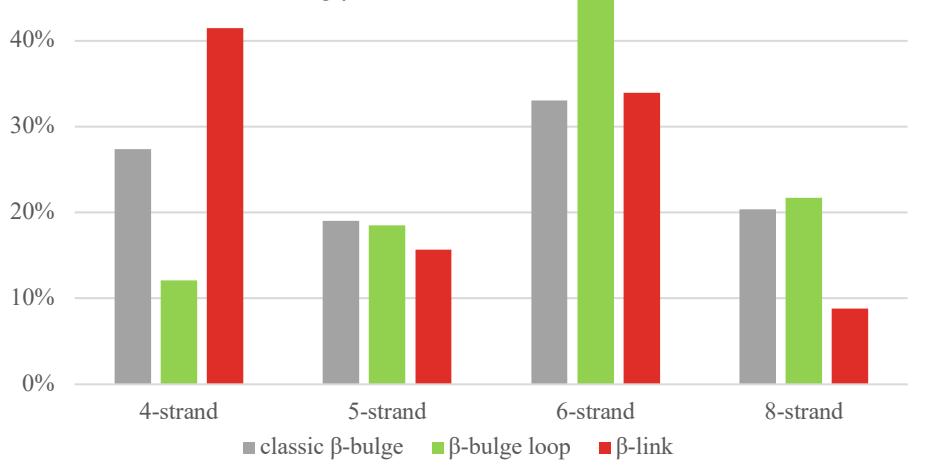
E. Distribution of different β -barrels in the *Protein Motif 2* database, and instances of β -links



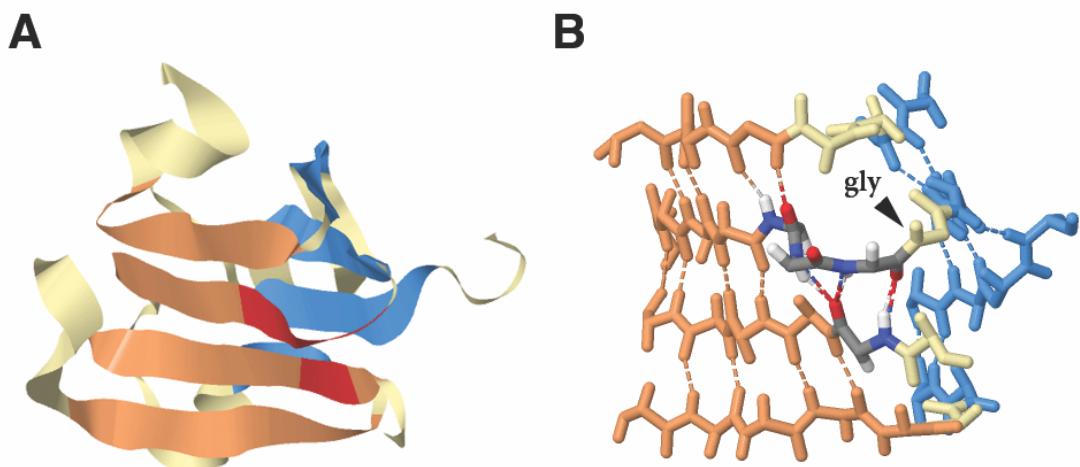
F. Distribution of β -links defined with different stringency among different β -barrels



G. Distribution of different β -bulge structures among β -barrels



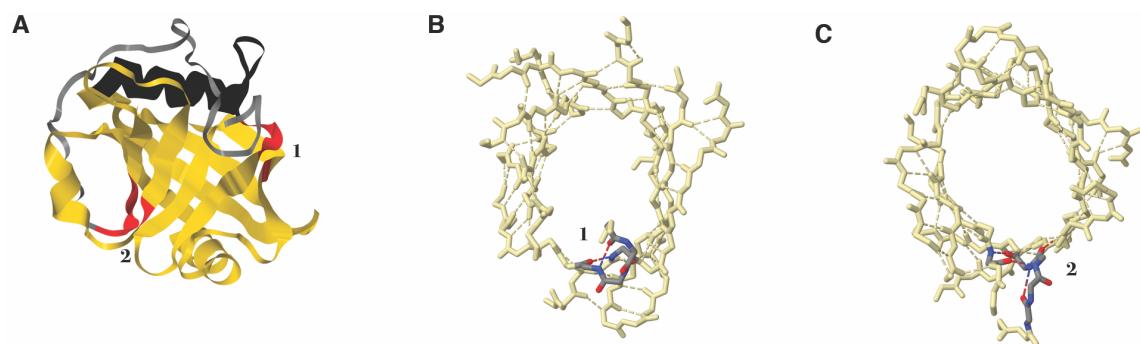
Supplementary Figure S2
Classic β -bulges in Sandwich Proteins of SCOPe Fold b.11



Human β -crystallin A4, SCOPe domain d3lwka1, SCOPe family b.11.1.0 (3lwk). (A) One domain of the pseudo-symmetrical two-domain protein is shown as a ribbon with the β -bulges in red, and other colouring as in Figure 4. (B) Wireframe diagram of section of the domain near the connection between the two sheets of the β -sandwich.

The second domain of the protein is similar, as are both domains of other proteins of fold b.11 in the ProtMotif 2 database: 3qk3(A), 1h4a(X), 1oki(B) and 1ytq(A). Gly occupies position -1 of the classic β -bulge in all these domains, and the residue at position X is Lys or Arg.

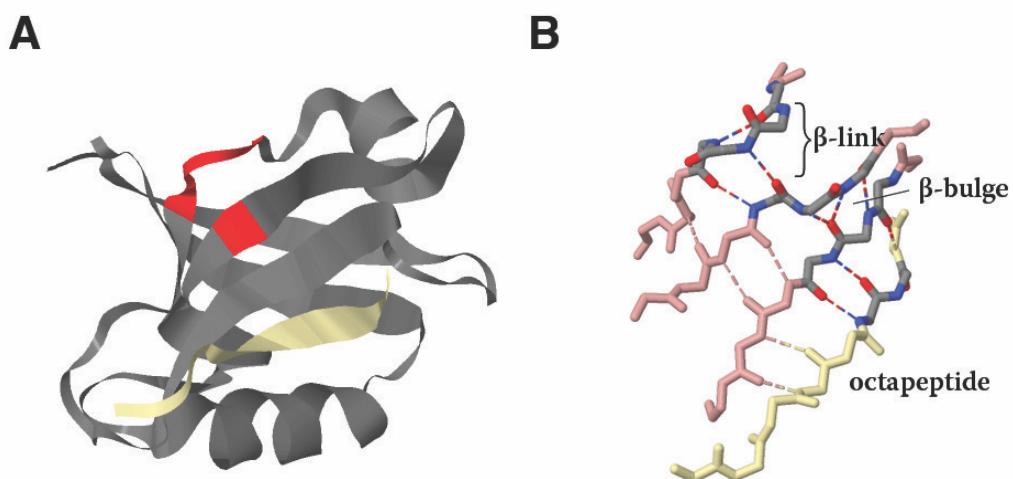
Supplementary Figure S3
Two types of β -link in cyclophorins



8-strand, closed β -barrel: cyclophilin (2cpl) — SCOPe domain d2cpla_, SCOPe family b.62.1.1. (A) Ribbon diagram. The basic colour is gold, with β -links in red. (B) Wireframe diagram viewed from the top of A, but showing only the β -strands (cream), β -link 1 (cpk colouring), and the residue *N*-terminal to its β -turn (cream). (C) As B, but with the barrel turned through 180° (i.e. inverted) and β -link 2 coloured cpk.

Supplementary Figure S4

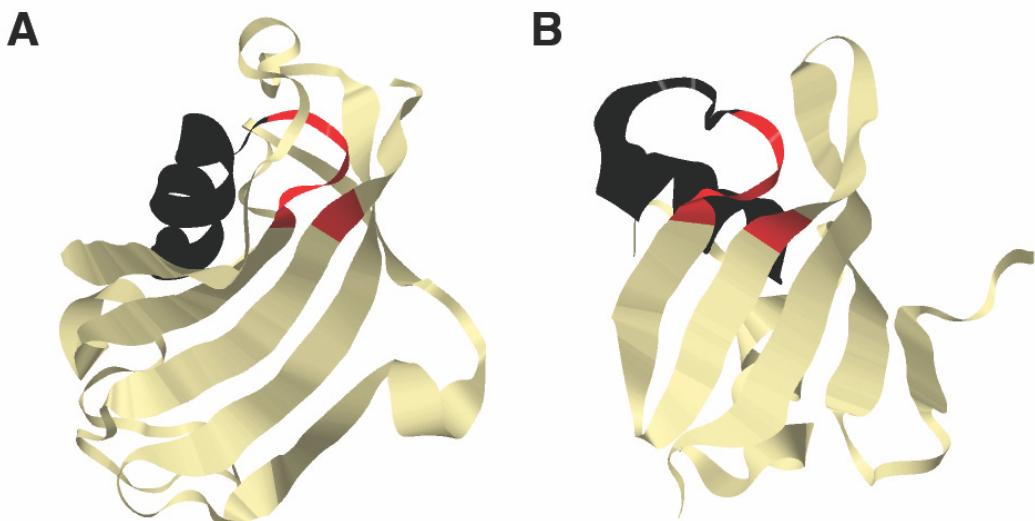
The extended hydrogen-bonded network from β -links in ligand-binding PDZ domains



Glutamate receptor-interacting protein 1 (1nfz) — SCOPe domain d1n7fa_ SCOPe family

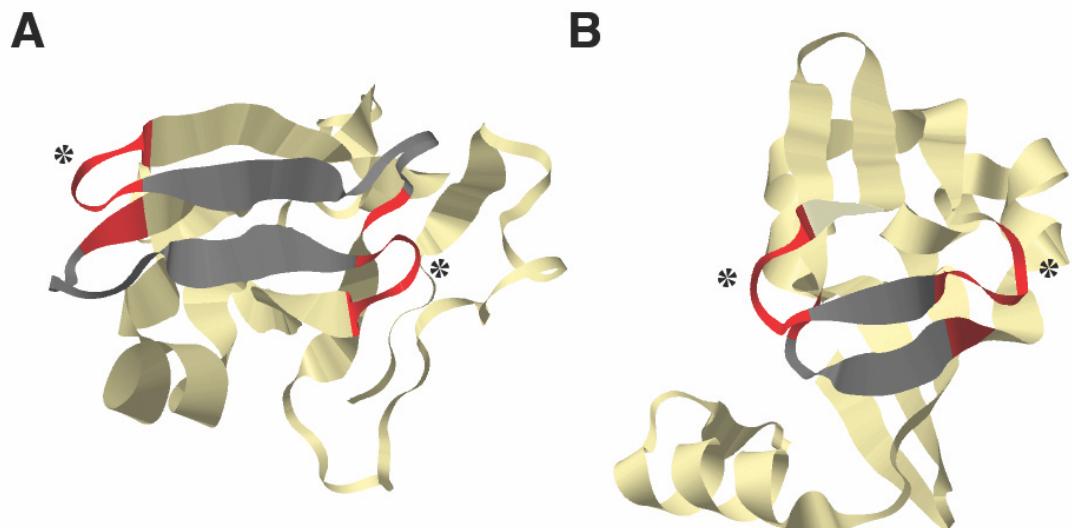
b.36.1.1 — bound to liprin C-terminal octapeptide. (A) Ribbon diagram. The β -link is shown in red and the octapeptide in cream. (B) Wireframe diagram of the four-stranded β -sheet with the β -link and extended hydrogen-bonding residues coloured cpk, and adjacent regions in salmon pink or cream for protein or peptide, respectively.

Supplementary Figure S5
 β -links connecting a single β -sheet to an α -helix



(A) Ascomycin (1fkj) — SCOPe family d.26.1.1. The β -link is coloured red and the α -helix black. (B) Haemolysin domain CorC_HlyC (2oai) — SCOPe domain d2oai1, SCOPe family d.145.1.4. Colour scheme as in A.

Supplementary Figure S6
 β -links flanking short two-stranded segments of protein



In each case the two β -links are shown in red, with the region between in grey. The strand of the β -link bearing the β -turn and β -bulge doublet is indicated by an asterisk. (A) β -clip:
deoxyuridine 5'-triphosphatase (1euw) — SCOPe domain d1euwa1, SCOPe family b.85.4.1.
(B) Barrel-sandwich hybrid: Glycine cleavage system protein H (3mxua) — SCOPe domain d3mxua1, SCOPe family b.84.1.0.