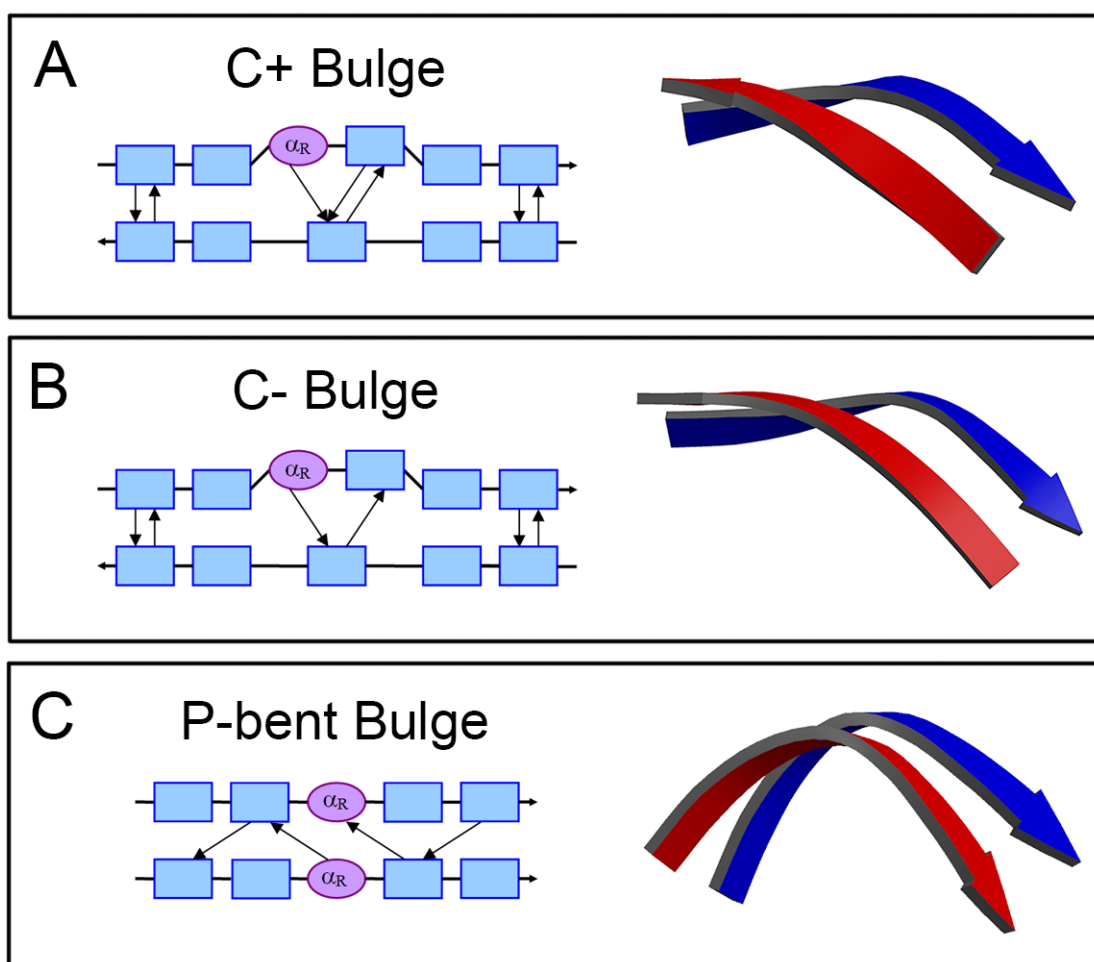


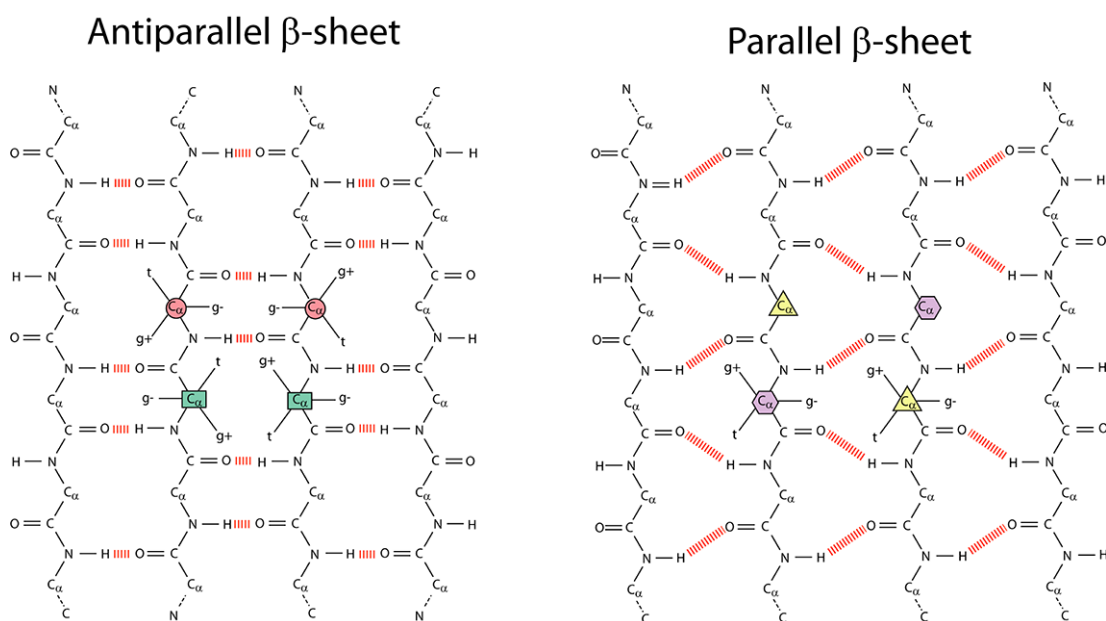
**Table S1.** Total number of structures for each BSD type in our database distributed according  $\beta$ -strand alignment and location in the hydrogen bonding system. The number of non-redundant structures for each type (number of clusters) is given in parentheses.

BSD Type	Antiparallel		Parallel
	NHB	HB	
CSD	4861 (252)	114 (20)	212 (26)
ETD	69 (13)	0	0
BDD	272 (21)	929 (36)	712 (49)
BFD	233 (5)	0	20 (3)

**Fig. S1.** Canonical  $\beta$ -bulge structures related to BFD and ETD BSD motifs. **A** C+ bulge in antiparallel  $\beta$ -sheet; PDB: 2sga, A 86-93, 103-108. **B** C- bulge in antiparallel  $\beta$ -sheet; PDB: 9rub, A 24-31, 120-128. **C** P-bent bulge in parallel  $\beta$ -sheet; PDB: 3cox, A 10-17, 280-287. In each case the hydrogen bonding pattern which defines the bulge is shown schematically. Residues which adopt a  $\beta$  backbone conformation are shown as blue rectangles, and those which have an  $\alpha$  backbone conformation as purple ovals. Hydrogen bonds are shown as arrows pointing from the NH group to the carbonyl. For each  $\beta$ -bulge type a ribbon diagram for the  $\beta$ -ladder containing the bulge is also shown (prepared using Molscript<sup>69</sup> and Raster3D<sup>70</sup>). These ribbon diagrams show the deformation of the structure due to the presence of the  $\beta$ -bulge.



**Fig. S2.** Schematic representations of the extended hydrogen bonding networks in antiparallel and parallel  $\beta$ -sheet. The three low-energy sidechain conformations ( $g^+$ ,  $g^-$  and  $t$ ) are shown for residues in different sites within the network. Red circles indicate the  $C_\alpha$  atoms of a pair of residues forming an HB site in the central  $\beta$ -ladder of the antiparallel  $\beta$ -sheet. Green rectangles indicate  $C_\alpha$  atoms of residues forming an NHB site in the central  $\beta$ -ladder of the antiparallel  $\beta$ -sheet. Purple hexagons and yellow triangles show the  $C_\alpha$  atoms of residues which form the narrow and wide sites of a parallel sheet site respectively.



**Fig. S3.** A portion of the sequence alignment of CD4 proteins showing full conservation of the C-terminal Cys of the redox-active aCSDn amongst all euteleosts, while the N-terminal Cys is found only in primates and muroidea.

155	HLSIKLFNINNDIVTTEIL----	QEEAPQKYILKQKQKKAIDSGTWM	CHVYSN-SPSINQ	209	Q6Q4G2	Q6Q4G2_CHICK	
155	HLGIELFNISISKIVTTEIL----	QNEAPQKYILKQKQKKAIDSGTWM	CHVYSN-SPLINQ	209	A7WPR9	A7WPR9_MELGA	BIRDS
151	RFSITLFNHNSRVTPPEVL----	QNETPQKYALKVKQLQPTDSGTWIC	NMHSD-SPSINE	205	B8YEL2	B8YEL2_ANAPL	
156	-----LTECKHKHKGV-----	VSGSKVLSMSNLRVQDSDFWNC	TVT---LDQKKN	197	P06332	CD4_MOUSE	MUROIDEA
155	-----PIECKHKSSNI-----	VKDSKAFSTHSLRIQDSGIWNC	TVT---LNQKKH	196	P05540	CD4_RAT	
34	-----SIECKGPGNSI-----	VKGSKTLSMPNLRIQDSGIWTC	TVT---QSQHKN	75	Q0MWX7	Q0MWX7_PERMA	
152	-----SVQCRSPRGKN-----	IQGGKTLVSQLELQDSGTWTC	TVL---QNQKKV	193	P01730	CD4_HUMAN	
152	-----SVQCRSPRGKN-----	IQGGKTLVSQLELQDSGTWTC	TVL---QNQKKV	193	P16004	CD4_PANTR	PRIMATES
152	-----SVKCRSPGGKN-----	IQGGRTISVPQLERQDSGTWTC	TVS---QDQKTV	193	P16003	CD4_MACMU	
152	-----SVKCRSPRGKN-----	IQGGRTISVPQLERQDSGTWTC	TVS---QDQNTV	193	Q08338	CD4_CHLAE	
110	-----SVKFKQP-GNK-----	ISTDVEVSVPNVGFQDSGTWTC	FIS---QDQKSL	150	A7U476	A7U476_MARMO	WOODCHUCK (non-muroidean rodent)
152	-----SVQWKPGDNN-----	KRDVKLSLSLAQVGLQDSGTWTC	TIS---QSQOTL	193	A7YY52	A7YY52_BOVIN	
152	-----SVQWKPGNNR-----	KEELKLSLSLAQVGLQDSGTWTC	TIS---QSQOTL	193	Q19VJ6	Q19VJ6_SHEEP	
152	-----SVQWKPGNKR-----	KNEAKLSLSPQVGLQDSGTWTC	TVS---QAQOTL	193	Q9XS78	CD4_DELLE	
152	-----SVQWKPGNKR-----	KNEAKLSLSPQVGLQDSGTWTC	TVS---QAQOTL	193	Q71QE2	Q71QE2_TURTR	
154	-----TVQWKPGNKS-----	KNDVKSLLLPQVGLQDSGLWTC	TVS---QDQKTL	195	Q6R3N4	Q6R3N4_PIG	OTHER MAMMALS
160	-----SVQWKPGNKS-----	KHGGQNLSSLWPQLQDGGTWT	CIIS---QSQKTV	201	P33705	CD4_CANFA	
171	-----SVQWKPGNKS-----	KSGVHLSLSLQLELQESGTCT	CTVS---QSQKTL	212	P79355	P79355_FELCA	
170	-----SIVWESPGSK-----	KYEDKLSLTLQGRQESGTWEC	IVS---YNKCTL	210	Q3LG90	Q3LG90_ROUAE	
156	-----SVQWKS PENKI-----	IETGPTCSMPKLRQLQDSGTW	SCHLSF---QDQNKL	198	P46630	CD4_RABIT	
156	-----KVEWRGPGDKSK-----	QILNQDKKTLNLVKMGPNETGL	WDCIVS---VSEKTL	201	A6MUU7	A6MUU7_MACEU	MARSUPIALS
149	-----KIEWHGPGNTSK-----	RILSSNKKTLNLLQVDSEEEGE	WSSCTVS---INGKSL	194	F7CXR5	F7CXR5_MONDO	
143	KP-----EYWLNPRGER-----	IKNNQGTIVTRVTSQDDGMWIC	VVA---EEKQ--	184	B9UYP5	B9UYP5_HIPHI	
151	KP-----GIHWLNPOGEK-----	IT--QATHSVQVSSRHSGRWT	CVVT---LDRKEA	192	Q68AX5	Q68AX5_TAKRU	FISH
150	-----QRRWLSPOKQDLNEDKRA	QIRNDGSLTVMSVTDQDHGEWT	CVVT---YQGREA	199	Q32QB0	Q32QB0_ONCMY	
147	-----SVTIWISPKNEG-----	CQEHKNAKDTVLVPSVSTCHNG	VWTCQLK---YENKKT	192	B5KH12	B5KH12_DANRE	
131	-----TKPTIQK-----	VISDPLVTQPLGIEDSGPWKC	IVSVPSRSTEPL	170	F6WFZ6	F6WFZ6_ORNAN	MONOTREME
147	-----LRITWETPRRGK-----	IEDKRGITVANVQINDGGTYNC	HLWID--GENKA	190	E9LP47	E9LP47_XENLA	FROG

..\*: