

**Table 4. Sequence identity levels for the families of designed sequences computed by DESIGNER, for the seven considered protein folds, and for the corresponding families of natural sequences**

<b>Fold</b>	<b>No. nat</b>	<b>No. des</b>	<b>Nat-all</b>	<b>Des-all</b>	<b>Nat-core</b>	<b>Des-core</b>	<b>Nat-surf</b>	<b>Des-surf</b>
Antifreeze	23	717	73.5 (14.2)	57.8 (8.0)	91.4 (5.8)	67.5 (11.8)	60.9 (23.6)	47.7 (13.3)
Cold-shock	113	211	45.4 (18.1)	42.3 (11.9)	70.5 (15.7)	59.4 (16.8)	30.4 (22.1)	34.7 (19.6)
HTH	184	880	22.3 (7.0)	33.7 (13.6)	34.5 (13.4)	38.2 (21.4)	17.2 (9.2)	31.7 (16.8)
Protein G	20	198	76.7 (18.1)	36.9 (13.1)	87.5 (14.4)	36.2 (19.8)	66.5 (22.2)	36.7 (17.5)
SH3	534	2665	28.3 (10.0)	33.0 (13.1)	47.7 (15.9)	36.5 (16.6)	20.4 (11.7)	25.3 (20.0)
Homeobox	1225	1380	41.9 (13.8)	33.3 (10.2)	70.6 (13.8)	27.8 (16.9)	21.7 (14.9)	27.3 (14.0)
CI2	35	882	44.8 (16.2)	40.7 (12.2)	63.9 (15.9)	34.5 (20.7)	37.0 (18.5)	35.4 (18.0)

The listed sequence identity represents average pairwise sequences identity computed over all the designed (or natural) sequences available for each fold. For each fold, several backbones were used in the protein design calculations and the low energy sequences computed for all the backbones were combined (see *Materials and Methods* of main manuscript). The first column lists the fold name, and the second and third columns list the number of sequences considered in each family. Nat-all, Nat-core, and Nat-surf refer to the natural sequences of the full protein, the core, and surface residues, respectively. Des-all, Des-core, and Des-surf refer to the designed sequences of the corresponding residue subsets. The listed values represent average identities, taken over all the sequences of the considered family, and the parenthesis represent the standard deviations. See text of the main paper for discussion of the results.