

**Table 5.  $P$  values corresponding to the average pairwise sequences identities listed in Table 4**

| <b>Fold</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  | 3.7e-44        | 1.1e-28        | 6.5e-22         | 8.6e-12         | 8.6e-8          | 1.7e-4          |
| Cold-shock  | 1.2e-19        | 2.6e-17        | 1.7e-14         | 2.4e-10         | 3.7e-3          | 2.1e-3          |
| HTH         | 5.7e-6         | 5.1e-12        | 3.0e-4          | 2.1e-5          | 1.2e-1          | 1.7e-2          |
| Protein G   | 1.2e-47        | 4.8e-14        | 7.5e-20         | 2.2e-4          | 3.0e-9          | 2.1e-3          |
| SH3         | 2.4e-8         | 4.8e-11        | 1.9e-7          | 2.2e-4          | 1.2e-1          | 1.7e-2          |
| Homeobox    | 2.6e-17        | 4.8e-11        | 1.7e-14         | 1.8e-3          | 2.6e-2          | 1.7e-2          |
| CI2         | 1.8e-18        | 3.4e-16        | 2.0e-11         | 2.2e-4          | 3.9e-4          | 2.1e-3          |

To compute these  $P$  values, we used a uniform random sequence model and a binomial distribution of the sequence identity values. The fold name is given in the first column. The number of sequences for each fold and all other details are as in Table 2. For all other information, the reader is referred to the text of the main paper.