nature reviews molecular cell biology

https://doi.org/10.1038/s41580-024-00718-y

Review article

Check for updates

Opportunities and challenges in design and optimization of protein function

In the format provided by the authors and unedited

Supplementary Box 1. **Methods for statistical comparison of computed and natural proteins**

We extracted *de novo* designed proteins from the Protein Data Bank, including 84 monomeric proteins and 17 designed binders complexed with their targets. We randomly sampled 1,000 natural proteins from the CATH database (version 4.3)[1]. The natural protein binders were taken from the molecular surface interaction fingerprinting (MaSIF) testset, which contains 936 structures[2].

Relative Contact Order (RCO) is determined by measuring the sequence distance between secondary structures for all residue pairs within 8 Å (defined as contacts). If the contacts are separated by more than four residues in sequence, the average distance of these contacts is calculated. DSSP was used to determine the secondary structure element (SSE) content of the protein[3].

The script used for the generation of these figures is available at https://github.com/casperg92/opportunities_in_protein_design_review.

References

1. Sillitoe, I. *et al.* CATH: increased structural coverage of functional space. *Nucleic Acids Res.* **49**, D266–D273 (2021).

2. Gainza, P. *et al.* De novo design of protein interactions with learned surface fingerprints. *Nature* **617**, 176–184 (2023).

3. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).