Supplementary Material

Implementation of all tools

- StructureAlign

StructureAlign superimposing query protein on reference protein to evaluate their structural similarity. The StructureAlign page provides two algorithms for users to choose from:

In Algorithm 1, the program first extracts sequences from the PDB files of the query protein and the reference protein, and performs Needleman-Wunsch alignment. Subsequently, the protein structures are translated to the center of mass. Following this step, an optimal overlay rotation matrix and translation vector are calculated for aligning the two sets of spatial points, resulting in an optimal superimposed conformation and outputting the RMSD value. A notable feature of this algorithm is that users can specify amino acid positions or domains of interest on multiple chains for protein structure alignment.

In Algorithm 2, the program utilizes the heuristic iterative searching algorithm US-align[1] to find the optimal superimposed conformation of two structures, and outputs TM-score and RMSD values. This represents a global-scale alignment where users cannot specify specific regions for comparison.

- BreakCheck

The program will traverse the input PDB file to check if the distance between the nitrogen atom of one residue and the third carbon atom of the previous residue is less than 1.6 Å, check whether the distance between the carbon atom of the same residue and the nitrogen atom of the latter residue satisfies less than1.6 Å and, if not simultaneously, treat this site as a break point.

- PDB2FAS

The program parses ATOM coordinates from PDB file and converts them into a FASTA format sequence file. For modified residues using the HETATM group name, such as Selenomethionine (MSE), the program converts it to standard residue Methionine (MET), but if it is other unrecognized residues, the program will ignore this site.

- AA-Mutate & PE-Minimize

The functionalities of protein energy minimization and amino acid mutation display are both implemented based on our laboratory's previously developed side-chain modeling method CIS-RR[3]. This method conducts clash-detection guided iterative search (CIS) of side-chain rotamers while continuously optimizing side-chain conformations using a conjugate gradients method.

In AA-Mutate, the program will call the -S parameter of CIS-RR to re-model the mutated residue and search for all possible rotamers of the residue to find the optimal conformation.

In PE-Minimize, the program will call -X parameter of CIS-RR for rotamer relaxation to minimize energy. Rotamer relaxation is implemented by using a conjugate gradients method to optimize the side-chain dihedral angles near the rotamer value. The conjugate gradients method follows Fletcher–Reeves algorithm.

- PDB-Repair

The function of PDB-Repair is to repair atomic missing in the main or side chains of input protein file:

When fixing the main-chain, each essential atom (N, Cα, C, O) required for every backbone residue is checked for presence. If only one atom is missing, the program calculates the position of the newly inserted atom based on atomic distances, bond angles, and dihedral angles, ensuring the relative spatial positioning and geometric configuration conform to the structural norms of the protein backbone. If multiple atoms are missing, our CCD (Cyclic coordinate descent) iterative search algorithm[2] is employed to insert the missing atoms.

For side-chain fixing, the program will call the -m parameter of a well-established side-chain modeling method, CIS-RR[3] for repair.

In cases of extensive atom loss or even missing amino acids, users need provide an aligned reference template sequence; the program will use this reference sequence to repair backbone or side-chain atoms. Users can also choose to provide a template PDB file, this allows the program to match the missing regions with the corresponding template structure, making the repaired structure more reasonable.

- PDB-Split

In PDB-Split, the program will read each chain of the input protein PDB file and analyze the different chains by sequence alignment. Then the original PDB file will be divided into several different single-chain PDB files.

- PDB-Assemble

PDB-Assemble contains two child tools: Assemble and ComplexAssemble.

In Assemble, user needs to provide a PDB file containing assembly parameters. The program will extract the atomic and positional information of the bio-units that need to be assembled from the assembly parameters in the PDB file, and then assemble these units to form a complete complex PDB file.

In ComplexAssemble, user needs to provide each individual unit PDB file and upload the assembly reference template file. The program assembles the complex by superimposing each unit file onto the template file.

- PDB-ReNumber

The program renumbers the atoms and residues of each chain of the input PDB file based on user-customized begining number (starting from 1 by default).

- PDB-AddH

The program will add hydrogen atoms to the input PDB file based on hydrogen atoms defined by CHARMM22.

References

1. Zhang C, Shine M, Pyle A M, et al. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. Nature Methods, 2022, 19(9): 1109-1115.
2. Canutescu A A, Dunbrack Jr R L. Cyclic coordinate descent: A robotics algorithm for protein loop closure. Protein Science, 2003,12(5): 963-972.
3. Cao Y, Song L, Miao Z, et al. Improved side-chain modeling by coupling clash-detection guided iterative search with rotamer relaxation. Bioinformatics, 2011, 27(6): 785-790.