Building the protein structure-specific side-chain rotamer libraries

Algirdas Grybauskas[1], Saulius Gražulis[1]
[1]Vilnius University Institute of Biotechnology, Lithuania

Abstract

Identifying the probable positions of the protein side-chains is one of the protein modelling steps that can improve the prediction of protein-ligand and protein-protein interactions. Most of the strategies predicting the side-chain conformations use predetermined angles, also called rotamer libraries, that are usually generated from the subset of high-quality protein structures. Then the grouping of amino acids by their type and the calculation of the frequencies of the most occurring dihedral angles are performed. Depending on the rotamer library, groups can be further differentiated by the structure of the main-chain atoms, solvent accessibility or other criteria.
Although these methods are fast, they tend to be too generalised and ignore basic assumptions about the geometry and surroundings of the target protein that can result in inaccuracies when predicting the possible side-chain atom positions. Hence we propose an approach that takes into account both of these terms by scanning through the conformationally restricted side-chain positions and generating dihedral angle libraries specific to the target proteins. The method avoids the drawbacks of lacking conformations due to unusual or rare protein structures. The core principle of the approach is a limited movement of side-chains imposed by a fixed number of degrees of freedom. Combinatorial explosion is avoided by using dead-end elimination and cell-list neighbour search while calculating potential energy.
Building such dynamic libraries could increase the likelihood of detecting rare rotamers.

Methods

The dihedral angle was chosen as a variable for constructing a simple rotamer model, because rotation around the bond produces the most extreme changes on the positions of the atoms. The possible atom position of the specific side-chain is determined by first changing the dihedral angles of all rotatable bonds (Fig. 1). It can be achieved by using matrix multiplication of the rotation matrices (Eq. 1).
Then the potential energy of each side-chain atom is calculated (Eq. 2) and high-energy rotamers are excluded by specified threshold. Modified *ff14SB* parameters from *Amber18* (`http://ambermd.org`) are used in the force field, such as VdW radii, partial charge. However, weights for each force field term had to be optimised.
The generation of rotamers was validated by comparing predicted rotamers against the side-chains from high-quality PDB (`https://www.wwpdb.org`) structures (< 1.9Å). First, the side-chains were randomly selected and χ angles were reset to 0°. Then the conformations were found by scanning dihedral angles and choosing the most energetically favourable rotamers.

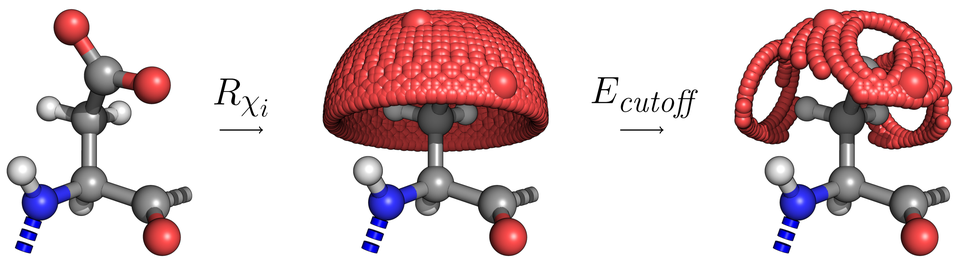


Figure 1: simplified scheme for generating rotamer libraries

Conformational model

$$\boldsymbol{p}^{0\prime} = \boldsymbol{T}^{0}_{n} \cdot (\prod_{i=1}^{n} \boldsymbol{R}_{\chi i} \cdot \boldsymbol{T}^{i}_{i-1}) \cdot \boldsymbol{p}^{0} \tag{1}$$

where:

$\boldsymbol{p}^{0}$	– initial atom coordinates;
$\boldsymbol{p}^{0\prime}$	– transformed atom coordinates;
$\boldsymbol{T}^{i}_{i-1}$	– transformation matrix that changes one frame of reference to another;
$\boldsymbol{R}_{\chi i}$	– rotational matrix that changes the dihedral angle.

Energy function

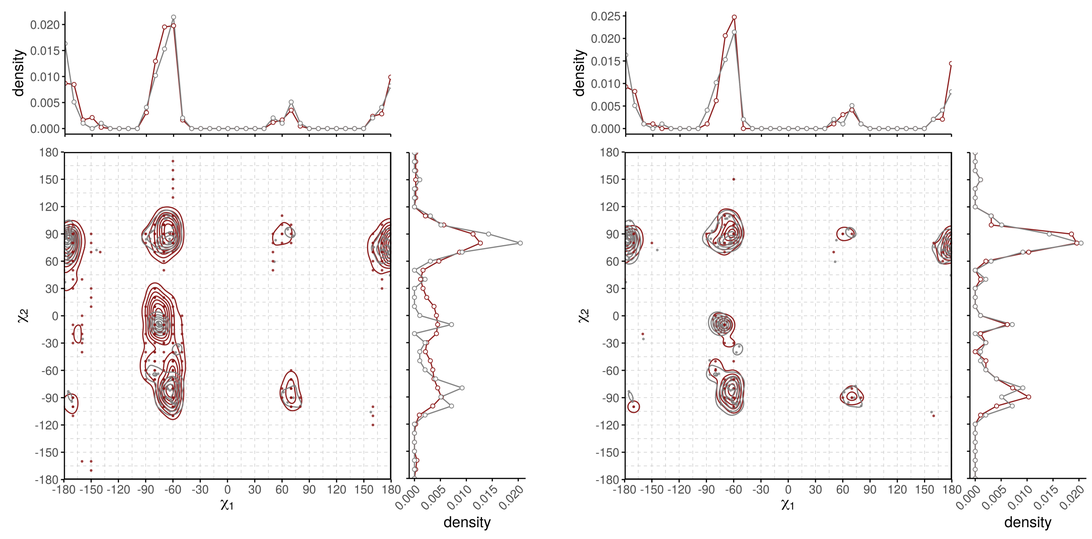
$$E_{\text{Total}} = \sum_{i} \sum_{j \neq i} (w_1 E_{\text{LJ}}(i,j) + w_2 E_{\text{C}}(i,j) + w_3 E_{\text{H}}(i,j)) + \sum_{\text{d}} w_4 E_{\text{T}}(d) \tag{2}$$

where:

E	– energy value;	LJ	– Lennard-Jones;
w	– weight;	C	– Coulomb;
i, j	– atom indexes;	H	– hydrogen bond;
d	– dihedral angle index;	T	– torsional.

Results

By comparing dihedral angles (Fig. 2) and the atom positions with RMSD calculations (Fig. 3), it seems that the program is able to capture both the variety (Fig. 2a) and the accuracy (Fig. 2b) of the side-chain positions. However, not all amino acids are easy to predict – more parameter estimations have to be performed.



(a) (b)

Figure 2: comparison of experimentally observed (grey) and predicted (red) dihedral angles of phenylalanine ($N_{residues} = 100$). The graphs depict 25% (a) and 1% (b) of the most energetically favourable side-chain angles that fall under the specified energy cutoff.

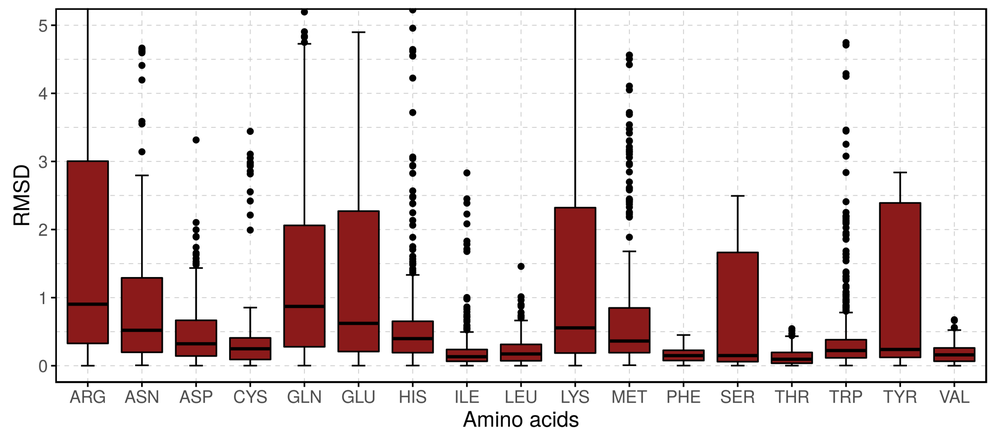


Figure 3: RMSD calculations for predicted side-chain positions ($N_{residues} = 100$ per amino acid)

Applications

- analysis of the point mutation impact to the protein structure and function;
- prediction of synthetic side-chain rotamers;

Contacts

E-mail: `algirdas.grybauskas@yahoo.com`