The rotag library: generating protein structure-specific side-chain rotamer libraries

Algirdas Grybauskas[1], Saulius Gražulis[1]
[1]Vilnius University Life Sciences Center, Lithuania
Email: algirdas.grybauskas@gmc.vu.lt

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. We have already presented an alternative approach for generating rotamer libraries from the single target protein in our previous work (https://bit.ly/2YJ0X7h) and, now, we present more refined results supporting the application of the method.

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* [Maier et al., 2015] 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 [Berman et al., 2003] structures ($Resolution \leq 1.9$ and $R_{free} \leq 0.20$). The results then were compared against widely used rotamer libraries, such as Dunbrack [Shapovalov et al., 2011], Ultimate [Hintze et al., 2016] and Dynameomics [Towse et al., 2016] (both backbone-dependent and backbone-independent). The sample size of 1000 bcRMSD comparisons per residue for each rotamer library was chosen.

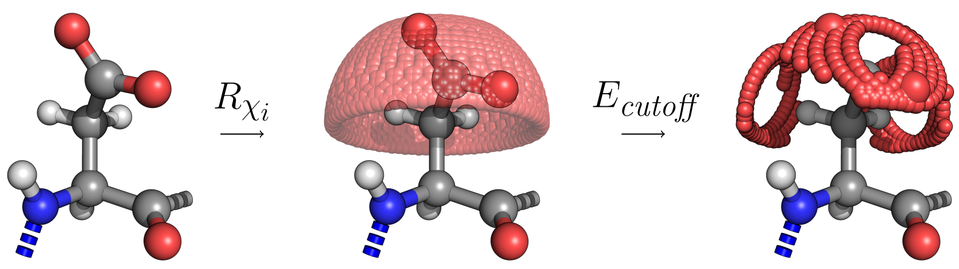

Figure 1: simplified scheme for generating rotamer libraries

Conformational model

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

where:

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

Energy function

$$E_{\text{Total}} = \sum_{i} \sum_{j \neq i} (w_1 E^{LJ}_{ij} + w_2 E^{C}_{ij} + w_3 E^{H}_{ij}) + \sum_{\text{d}} w_4 E^{T}_{d} \qquad (2)$$

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 and the atom positions with RMSD calculations (Fig. 2), the *rotag* program, on average, is able to capture both the variety and the accuracy of the side-chain positions (Fig. 3).

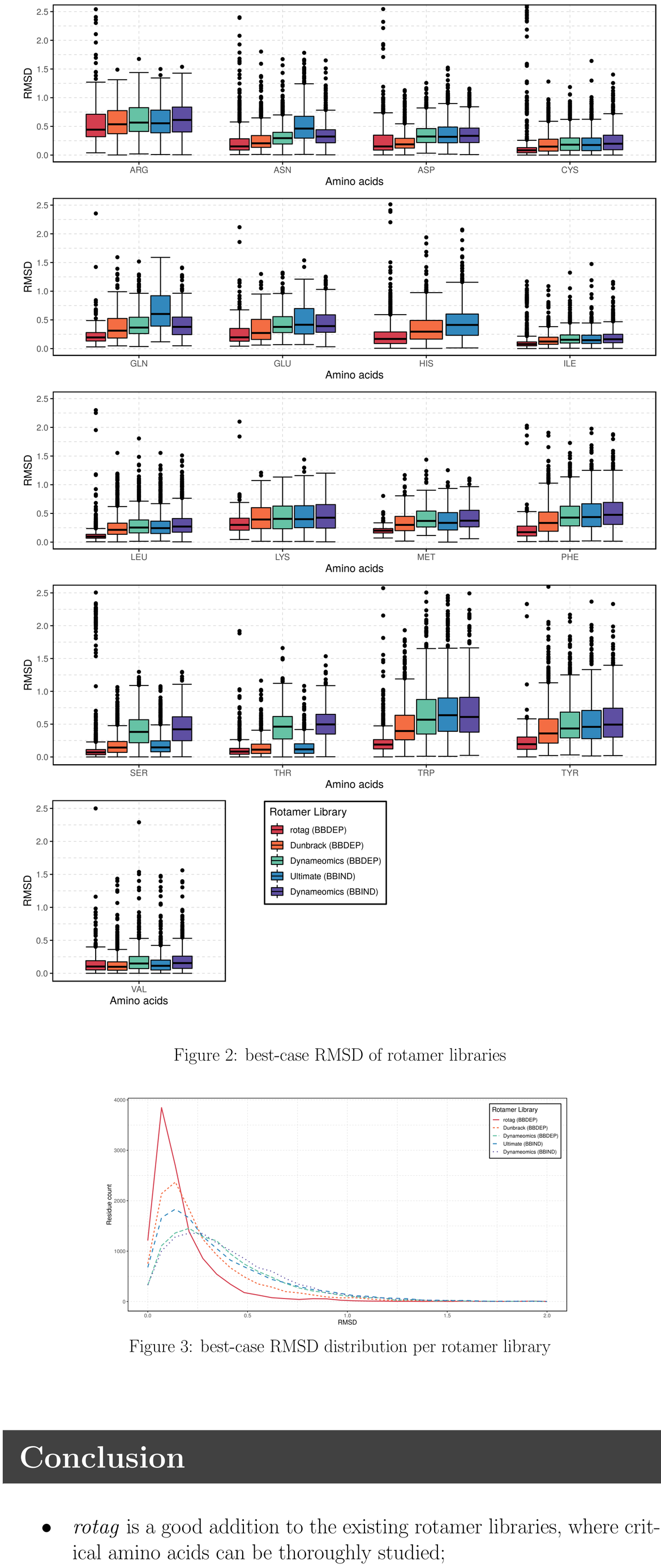

Figure 2: best-case RMSD of rotamer libraries

Figure 3: best-case RMSD distribution per rotamer library

Conclusion

- *rotag* is a good addition to the existing rotamer libraries, where critical amino acids can be thoroughly studied;
- *rotag* could be applied to the analysis of non-standard amino-acids.