Method of extending rotamer libraries and its usage in protein active site studies

Algirdas Grybauskas[1], Antanas Vaitkus[1], Andrius Merkys[1], Saulius Gražulis[1]

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

Abstract

Rotamer libraries are extensively used during the protein modelling process and have been shown to greatly improve the prediction of protein-ligand and protein-protein interactions as well as suggest amino acid mutations. Most of the strategies predicting the side-chain conformations involve calculating the dihedral angle averages for each side-chain from a subset of high-quality protein structures. However, these methods, while fast to apply, tend to average out rarely observed dihedral angles and ignore the surrounding atoms. Also, these methods exclude heteroatoms in their statistics. In order to create a well-rounded rotamer library, we have developed a method [1] to generate rotamer libraries from a single structure including heteroatoms without the need to process large subsets of side-chain occurrences.

Methods

The method scans for dihedral angles and utilises a force-field to incorporate interactions with surrounding atoms and can detect rare dihedral angle occurrences. Modified *ff14SB* parameters from *Amber18* are used in the force-field, such as vdW radii, partial charge. vdW radii for ions were selected using the work of Šidlauskaitė et al., 2023 [2]. However, weights for each force-field term had to be optimised.

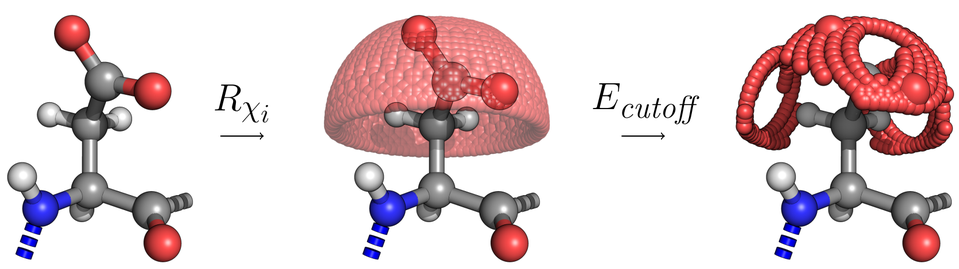

Figure 1: simplified scheme for generating rotamer libraries

Due to the method's flexibility, we decided to incorporate heteroatoms, thereby opening up a possibility to store and generate heteroatom positions in the same format as the rotamer libraries. Initially, metal ions and water molecules were analysed.

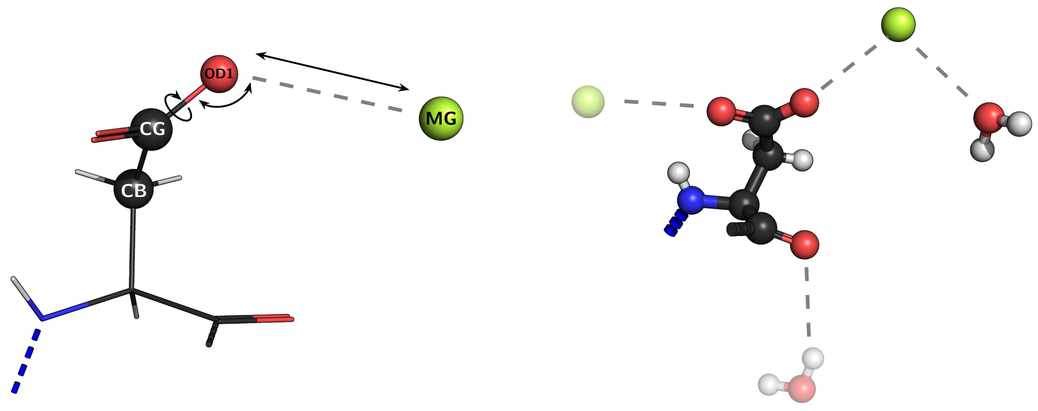

Figure 2: possible bond parameter changes for heteroatoms (left) and placements of heteroatoms with respect to the interacting side-chain atoms (right)

Currently, the research focuses on Type II restriction endonuclease active sites to demonstrate the utility of the extended rotamer libraries by determining the quantity of metal ions in the active site – a currently unsolved problem [3].

Conformational model

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

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;
- $\boldsymbol{R}_{A_i}$ – rotational matrix that changes the bond angle;
- $\boldsymbol{T}_{B_i}$ – translational matrix that changes the bond length.

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

CIF usage

CIF is an excellent format for using chemical information data, such as atom positions, storing rotamer library information and passing it to other programs in UNIX-like manner.

```
loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.auth_asym_id
_atom_site.auth_seq_id
_atom_site.pdbx_PDB_model_num
ATOM   2  N  N   . ASP A 1 27 -31.459 -9.382  -0.630 1.00 . A 54  1
ATOM   3  C  CA  . ASP A 1 27 -32.444 -9.886  0.303  1.00 . A 54  1
ATOM   4  C  C   . ASP A 1 27 -33.790 -10.232 -0.318 1.00 . A 54  1
ATOM   5  O  O   . ASP A 1 27 -34.560 -10.956 0.317  1.00 . A 54  1
ATOM   6  C  CB  . ASP A 1 27 -32.650 -8.892  1.445  1.00 . A 54  1
ATOM   7  C  CG  . ASP A 1 27 -33.510 -7.732  1.055  1.00 . A 54  1
ATOM   8  O  OD1 . ASP A 1 27 -33.182 -7.066  0.057  1.00 . A 54  1
ATOM   9  O  OD2 . ASP A 1 27 -34.527 -7.497  1.737  1.00 . A 54  1
ATOM   10 H  H   . ASP A 1 27 -31.436 -8.527  -0.724 1.00 . A 54  1
ATOM   11 H  HA  . ASP A 1 27 -32.095 -10.702 0.695  1.00 . A 54  1
ATOM   12 H  HB2 . ASP A 1 27 -33.079 -9.347  2.187  1.00 . A 54  1
ATOM   13 H  HB3 . ASP A 1 27 -31.787 -8.546  1.723  1.00 . A 54  1
HETATM 15 MG MG  . MG  B 1 .  -37.397 -8.848  1.441  1.00 . A 101 1
#                 *    *   *                              * *   *
loop_
_struct_conn.id
_struct_conn.conn_type_id
_struct_conn.ptnr1_label_seq_id
_struct_conn.ptnr1_label_asym_id
_struct_conn.pdbx_ptnr1_label_alt_id
_struct_conn.ptnr1_auth_seq_id
_struct_conn.ptnr1_auth_asym_id
_struct_conn.[local]_ptnr1_pdbx_PDB_model_num
_struct_conn.ptnr1_label_atom_id
_struct_conn.ptnr2_label_seq_id
_struct_conn.ptnr2_label_asym_id
_struct_conn.pdbx_ptnr2_label_alt_id
_struct_conn.ptnr2_auth_seq_id
_struct_conn.ptnr2_auth_asym_id
_struct_conn.[local]_ptnr2_pdbx_PDB_model_num
_struct_conn.ptnr2_label_atom_id
metalc1 metalc 27 A . 54 A 1 OD1 . B 101 . A 1 MG
#               *  * * *  * *    * * *   * * *
```

```
loop_
_rotag_rotamer_bond_parameter.id
_rotag_rotamer_bond_parameter.rotamer_id
_rotag_rotamer_bond_parameter.label_seq_id
_rotag_rotamer_bond_parameter.label_comp_id
_rotag_rotamer_bond_parameter.label_asym_id
_rotag_rotamer_bond_parameter.auth_seq_id
_rotag_rotamer_bond_parameter.auth_asym_id
_rotag_rotamer_bond_parameter.pdbx_PDB_model_num
_rotag_rotamer_bond_parameter.label_alt_id
_rotag_rotamer_bond_parameter.frequency
_rotag_rotamer_bond_parameter.type
_rotag_rotamer_bond_parameter.value
_rotag_rotamer_bond_parameter.units
1 1 .  MG  B 101 A 1 . 0.001792 CB-CG-OD1.MG -180.000 degrees
2 1 .  MG  B 101 A 1 . 0.001792 CG-OD1.MG    -180.000 degrees
3 1 .  MG  B 101 A 1 . 0.001792 OD1.MG          1.500 angstroms
4 1 27 ASP A 27  A 1 . 0.001792 chi1         -108.000 degrees
5 1 27 ASP A 27  A 1 . 0.001792 chi2         -180.000 degrees
#
loop_
_rotag_rotamer_energy.id
_rotag_rotamer_energy.rotamer_id
_rotag_rotamer_energy.calculation_method
_rotag_rotamer_energy.value
1 1 composite -47.307025
```

Figure 3: `ATOM_SITE` data items that are essential to generate rotamer libraries with heteroatoms (left). `STRUCT_CONN` data items are essential to track covalent and non-covalent bond connections/interactions. `ROTAG_ROTAMER_BOND_PARAMETER` and `ROTAG_ROTAMER_ENERGY` data items were added to store bond parameter information that could be used for generating new atom positions (right).

Results

Ion and water positions suggested by rotamer libraries in the active site were compared to the PDB structures.

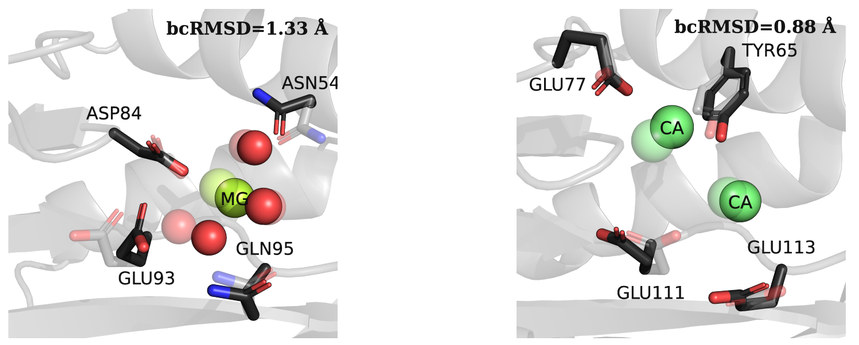

Figure 4: Type II restriction endonuclease active sites (PDB ID: 1D2I and PDB ID: 2BAM)

Conclusions

- CIF is a good format for storing and passing rotamer library information to other programs;
- our method is suitable for suggesting potential ion and water positions in protein active sites.

References

[1] Grybauskas, A. & Gražulis, S. (2023). Bioinformatics, 39(7), btad429.

[2] `http://crystallography.net/archives/2023/posters/OpenReadings/poster-P1-45.pdf`

[3] Pingoud, A., Wilson, G. G. & Wende, W. (2014). Nucleic acids res., 42(12), 7489-7527.

This research has received funding from the Research Council of Lithuania under Grant agreement No. MIP-23-87.