HAL
open science

AnAnaS: Software for Analytical Analysis of Symmetries in Protein Structures

Guillaume Pagès, Sergei Grudinin

To cite this version:

Guillaume Pagès, Sergei Grudinin. AnAnaS: Software for Analytical Analysis of Symmetries in Protein Structures. Protein Structure Prediction. Methods in Molecular Biology, pp.245-257, 2020, 10.1007/978-1-0716-0708-4_14 . hal-02931690

HAL Id: hal-02931690

https://hal.science/hal-02931690v1

Submitted on 7 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AnAnaS : software for analytical analysis of symmetries in protein structures

Chapter template for

Methods in Molecular Biology, Protein Structure Prediction 4th Edition

Guillaume Pagès, Sergei Grudinin

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP*, LJK, 38000 Grenoble, France

* Institute of Engineering Univ. Grenoble Alpes

Summary

Symmetry is very common among proteins found in structural databases such as the PDB. We present novel software, called AnAnaS, that finds positions and orientations of the symmetry axes in all types of symmetrical protein assemblies. It deals with five symmetry groups, cyclic, dihedral, tetrahedral, octahedral, and icosahedral. The software also assesses the quality of symmetry and can detect symmetries in incomplete cyclic assemblies. Internally, AnAnaS comprises discrete and continuous optimization steps and is applicable to assemblies with multiple chains in the asymmetric subunits or to those with pseudo-symmetry. The method is very fast, as most of the steps are performed analytically.

Keywords : Protein Symmetry, Protein Assembly, Continuous optimization, Symmetry axes detection, Quaternion arithmetic

1. Introduction

Symmetry in protein, and, more generally, in macromolecular assemblies is a key point to understand their structure, stability and function. Many symmetrical assemblies are currently

present in the Protein Data Bank (PDB) and some of them are among the largest solved structures. Thus an efficient computational method is required for the exhaustive analysis of these. The cyclic symmetry groups represent the most common assemblies in the PDB. However, complex protein structures very often possess higher order symmetries. These are dihedral and cubic, i.e., tetrahedral, octahedral, and icosahedral, groups. Detection and analysis of these symmetries has been a challenging problem and no efficient algorithms have been developed until recently (1; 2).

We present a novel formulation with the corresponding software, called AnAnaS, to find the positions and the orientations of the symmetry axes in all types of symmetrical protein assemblies. These are determined with a machine precision. The software also assesses the quality of symmetry and can detect symmetries in incomplete cyclic assemblies. The method comprises discrete and continuous optimization steps and is applicable to assemblies with multiple chains in the asymmetric subunits or to those with pseudo-symmetry (3; 4).

We implemented AnAnaS in C++ and exhaustively tested on all 51,358 symmetrical assemblies from the Protein Data Bank (PDB, as for May 2018). Its speed (most of the operations are analytical) and accuracy allows studying structural organization of symmetrical assemblies solved by X-ray crystallography, and also to routinely assess the symmetry annotation in the PDB and other structural databases. For example, we demonstrated that 1.6% of the symmetrical structures in the PDB have a higher symmetry compared to the PDB annotation. We also detected multiple cases with incorrect annotation. The method is available at http://team.inria.fr/nano-d/software/ananas as a standalone binary. The graphical user interface of the method built for the SAMSON platform is available at http://samson-connect.net.

2. Materials

2.1. Brief Theory

All amino acids, except glycine, are chiral. Hence, symmetry groups that can be present in protein assemblies cannot contain any reflection, inversion, or improper rotation. The only remaining finite point groups are the cyclic *Cn* for the cyclic group of order *n*, dihedral *Dn* for the dihedral group of order *n*, tetrahedral *T*, octahedral *O*, and icosahedral *I* groups. The last three (*T*, *O*, and *I*) constitute the three cubic groups. Figure 1 shows five examples of protein assemblies in different point groups.

2.2. Loss function

Generally, point groups can have multiple rotation (or symmetry) axes of different order. Therefore, it is useful to formally define the loss function that we aim to minimize. First, let us associate each symmetry axis $\vec{n}_k$ of order N_k with $N_k - 1$ non-trivial rotations $\hat{R}_k^i(2i\pi/N_k, \vec{n}_k)$ about this axis. Then, we define the loss function for a structure *A* as

$$\mathrm{Loss}^2 = \sum_k \sum_{i=1}^{N_k-1} \mathrm{RMSD}^2(A, \hat{R}_k^i(2i\pi/N_k, \vec{n}_k)A).$$

This loss function can be seen as a sum of root mean square deviations (RMSDs) between the original assembly and the rotated assemblies for every rotation in a certain symmetry group. In our method, we assign the same weights to each atom. We should mention that this loss function is very natural, since it is only based on Euclidean 3D distances, no adjustable parameters are required and all the rotations have equal importance. Without loss of generality, we can assume that each subunit has the same number of reference points. Technically, we achieve it by performing a multiple sequence alignment of the subunits and keeping only the aligned parts for the subsequent analysis. Our reference points are located at

the positions of the aligned Cα atoms. We should also mention that the loss function is minimized analytically using quaternion arithmetic, as it was shown earlier (3-5).

When a rotation axis is determined, the associated loss function is additionally decomposed in three components corresponding to the three dimensions in the cylindrical coordinate system, whose *z* axis is aligned with the symmetry axis. These are RMSD_R – the radial RMSD, RMSD_T – the tangential RMSD, and RMSD_Z – the axial RMSD.

2.3 Input data and output format

The method expects a file in the PDB or MMCIF format as input. A set of flags can control the behavior of the program. For example, if the user knows the exact correspondence between the chains in the input protein assembly, this can be additionally given to the program. The program returns a text output to the standard output (terminal). Additionally, the program can write the output in the JSON format to an external file. It can also output small scripts to help the user with the visualization of the results in one of the popular molecular viewers, e.g. Pymol. Finally, we also distribute an AnAnaS module for the SAMSON software platform with a graphical user interface for the input and the output of the method.

3. Methods

3.1 Basic usage of AnAnaS in the text mode

We will start by computing the symmetry of a tetrahedral assembly with pdb code 5x47. To do the analysis of the 5x47 structure, please type in the terminal “AnAnaS 5x47.pdb”. This produces the following output,

```
=======================Parsing the Command Line=======================
No symmetry groups specified, will be detected automatically.......... :
==========================Reading PDB file==========================
Read PDB file..........................................................: 5x47.pdb1
Number of chains read..................................................: 12
Number of atoms read...................................................: 13440
===================Auto-detecting Compatible Groups===================
Detected groups........................................................: t d6 d3 d2 c12 c6 c4 c3 c2
==========================Detecting Symmetry==========================
Cutoff for symmetry measure............................................: 7 A
Symmetry group : t
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
  0.769  0.356  0.515  0.447 29.397     3   0.536  -0.000   0.844   -26.543    -0.053    34.814
  0.967  0.451  0.660  0.545 29.427     2   0.098  -0.777   0.621   -26.543    -0.053    34.814
  0.873  0.403  0.556  0.539 29.482     3   0.423   0.897   0.127   -26.543    -0.053    34.814
  0.945  0.587  0.537  0.511 29.467     3  -0.600  -0.198   0.775   -26.543    -0.053    34.814
  0.833  0.427  0.457  0.551 29.469     2   0.153  -0.605  -0.781   -26.543    -0.053    34.814
  0.875  0.382  0.601  0.509 29.483     2   0.983   0.172   0.060   -26.543    -0.053    34.814
  0.928  0.534  0.542  0.531 29.492     3  -0.713   0.699   0.058   -26.543    -0.053    34.814
Average RMSD : 0.884844
Symmetry group : d2
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
  0.967  0.451  0.660  0.544 29.427     2  -0.098   0.777  -0.621   -26.543    -0.053    34.814
  0.833  0.427  0.457  0.551 29.469     2  -0.153   0.605   0.781   -26.543    -0.053    34.814
  0.875  0.382  0.601  0.509 29.483     2   0.983   0.172   0.060   -26.543    -0.053    34.814
Average RMSD : 0.893615
Symmetry group : c3
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
  0.769  0.355  0.515  0.447 29.397     3   0.536  -0.000   0.844   -26.543    -0.053    34.814
Average RMSD : 0.769347
```

```
Symmetry group : c3
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
  0.873  0.403  0.556  0.539 29.482     3  -0.423  -0.897  -0.127   -26.543    -0.053    34.814
Average RMSD : 0.872645
Symmetry group : c3
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
  0.945  0.587  0.537  0.511 29.467     3  -0.600  -0.198   0.775   -26.543    -0.053    34.814
Average RMSD : 0.945243
Symmetry group : c3
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
  0.928  0.534  0.542  0.531 29.492     3  -0.713   0.699   0.058   -26.543    -0.053    34.814
Average RMSD : 0.928172
Symmetry group : c2
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
  0.967  0.451  0.660  0.544 29.427     2   0.098  -0.777   0.621   -26.543    -0.053    34.814
Average RMSD : 0.967010
Symmetry group : c2
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
  0.833  0.427  0.457  0.551 29.469     2   0.153  -0.605  -0.781   -26.543    -0.053    34.814
Average RMSD : 0.833259
Symmetry group : c2
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
  0.875  0.383  0.601  0.509 29.483     2  -0.983  -0.172  -0.060   -26.543    -0.053    34.814
Average RMSD : 0.875335
```

The first section lists the group of symmetry provided by the user. Here, we did not provide any, so, since the assembly has 12 chains, the program will try the tetrahedral T, the dihedral D6, D3 and D2 and the cyclic C12, C6, C4, C3 and C2 symmetry groups. The second section lists the structural parameters, the number of atoms and chains in the structure. The last section shows the results for each tested symmetry group. For each symmetry operator, it lists the associated RMSD, symmetry order, axis direction and axis position. The results are ordered starting with

the higher order group first, as they are more significant, if the corresponding RMSD falls within a predefined threshold value (by default it is 7Å). Please note that the results for the D3 and D6 groups are not presented, as they fall outside the default threshold. The method suggests that the example structure has the tetrahedral T symmetry with the corresponding RMSD measure of 0.884844 Å. Indeed, this structure is annotated as a tetrahedral assembly in the PDB.

3.2 Changing the RMSD threshold

To get more exhaustive list of results (for more symmetry groups), the method allows modifying the threshold RMSD measure. In the following example, we set the RMSD threshold to 40 Å. This gives results for the missing the D3 and D6 groups from the previous example. Please type in the terminal “AnAnaS 5x47.pdb -C 40”. This produces the following additional output,

```
Symmetry group : d6
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
 30.711 14.092 13.490 23.705 29.427     6  -0.097   0.778  -0.620   -26.543    -0.053    34.814
 38.185 19.485 25.611 20.555 29.446     2  -0.360  -0.608  -0.707   -26.543    -0.053    34.814
 38.104 19.515 25.512 20.500 29.498     2  -0.624   0.438   0.647   -26.543    -0.053    34.814
  0.838  0.418  0.450  0.570 29.469     2   0.152  -0.604  -0.782   -26.543    -0.053    34.814
  0.881  0.384  0.601  0.516 29.483     2  -0.984  -0.170  -0.060   -26.543    -0.053    34.814
 38.107 12.966 33.464 12.814 29.453     2  -0.776  -0.449  -0.443   -26.543    -0.053    34.814
 38.184 12.926 33.586 12.764 29.506     2  -0.928   0.155   0.339   -26.543    -0.053    34.814
Average RMSD : 32.533732
Symmetry group : d3
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
 38.184 19.484 25.611 20.555 29.446     2  -0.360  -0.606  -0.710   -26.543    -0.053    34.814
 38.104 19.515 25.512 20.501 29.498     2   0.623  -0.436  -0.649   -26.543    -0.053    34.814
 38.145 17.498 16.695 29.497 29.427     3   0.097  -0.781   0.617   -26.543    -0.053    34.814
  0.881  0.384  0.601  0.517 29.483     2   0.984   0.170   0.060   -26.543    -0.053    34.814
Average RMSD : 34.119843
Symmetry group : d3
```

RMSD	RMSD_R	RMSD_T	RMSD_Z	RADGYR	ORDER	AXIS X	AXIS Y	AXIS Z	CENTER X	CENTER Y	CENTER Z
0.883	0.420	0.453	0.631	29.468	2	-0.152	0.602	0.784	-26.543	-0.053	34.814
38.145	17.499	16.695	29.497	29.427	3	0.096	-0.781	0.617	-26.543	-0.053	34.814
38.107	12.966	33.463	12.815	29.453	2	-0.776	-0.447	-0.445	-26.543	-0.053	34.814
38.184	12.926	33.586	12.765	29.506	2	-0.928	0.154	0.340	-26.543	-0.053	34.814

Average RMSD : 34.120379

We can see that the two additional tested groups, D3 and D6, have very high symmetry measures. Thus, the corresponding structure is very unlikely to posses one of these symmetries.

3.3 Testing input structure with only one specific symmetry group

The user may get the results for only one or several specific symmetry groups. The user can provide the group codes as program arguments without any flag. Supported symmetry group codes are “t” for the tetrahedral, “o” for the octahedral, “i” for the icosahedral, “dn” for the dihedral of order n>1 and “cn” for the cyclic of order n>1 groups. To test the previous example for only the tetrahedral symmetry, please type in the terminal “AnAnaS 5x47.pdb t”.

3.4 Symmetry-based reconstruction of missing subunits

The AnAnaS software is a powerful method for finding symmetry axis in cyclic assemblies with one or more missing subunits. We should specifically note that AnAnaS supports missing subunits only for cyclic symmetries. If there are any missing subunits, the result will give, for each permutation of the chain in the input structure, the best rotation with the expected symmetry-constrained angle. However, the different transformations will not form a group. That is why no average RMSD is provided, because each RMSD is potentially obtained with a different axis. The treatment of assemblies with missing subunits is not as automated as the

analysis of the complete assemblies. The user has to explicitly provide the symmetry group to be tested. Figure 2 presents an example of incomplete assembly.

3.5 Visualization of the results with PyMOL

We provide a PyMol script to visualize the predicted axes. Please use the "-y" option to output the PyMol commands, then simply copy-paste it into your PyMol console. One can paste the full output, as the non-relevant part will be ignored by the Python interpreter. Please note that the "cgo_arrow.py" script has to be loaded in advance by 'run cgo_arrow.py' in the PyMol console. The script can be found in the examples folder of the AnAnaS distribution or at https://raw.githubusercontent.com/Pymol-Scripts/Pymol-script-repo/master/cgo_arrow.py.

Please type in the terminal "AnAnaS 5x47.pdb1 t -y". This produces the following output:

```
Symmetry group : t
   RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
  0.769  0.356  0.515  0.447 29.397     3   0.536  -0.000   0.844   -26.543    -0.053    34.814
cgo_arrow [0.753636,-0.0641983,77.8158], [-53.8396,-0.0414718,-8.18717]
  0.967  0.451  0.660  0.545 29.427     2   0.098  -0.777   0.621   -26.543    -0.053    34.814
cgo_arrow [-20.6024,-47.2003,72.5106], [-32.4836,47.0946,-2.88193]
  0.873  0.403  0.556  0.539 29.482     3   0.423   0.897   0.127   -26.543    -0.053    34.814
cgo_arrow [-5.99587,43.5509,40.9687], [-47.0901,-43.6566,28.66]
  0.945  0.587  0.537  0.511 29.467     3  -0.600  -0.198   0.775   -26.543    -0.053    34.814
cgo_arrow [-56.1212,-9.8402,73.0655], [3.03524,9.73453,-3.43679]
  0.833  0.427  0.457  0.551 29.469     2   0.153  -0.605  -0.781   -26.543    -0.053    34.814
cgo_arrow [-17.2952,-36.6277,-12.3879], [-35.7908,36.522,82.0165]
  0.875  0.382  0.601  0.509 29.483     2   0.983   0.172   0.060   -26.543    -0.053    34.814
cgo_arrow [32.8768,10.3173,38.4204], [-85.9628,-10.423,31.2082]
  0.928  0.534  0.542  0.531 29.492     3  -0.713   0.699   0.058   -26.543    -0.053    34.814
cgo_arrow [-62.7094,35.4285,37.7446], [9.62341,-35.5341,31.8841]
Average RMSD : 0.884844
```

In the output above, after each axis, a command to display it in PyMol is provided. In PyMol, type the following commands to visualize 5x47 with all its symmetry axes:

```
fetch 5x47
run https://raw.githubusercontent.com/Pymol-Scripts/Pymol-script-repo/master/cgo_arrow.py
cgo_arrow [0.753636,-0.0641983,77.8158], [-53.8396,-0.0414718,-8.18717]
cgo_arrow [-20.6024,-47.2003,72.5106], [-32.4836,47.0946,-2.88193]
cgo_arrow [-5.99587,43.5509,40.9687], [-47.0901,-43.6566,28.66]
cgo_arrow [-56.1212,-9.8402,73.0655], [3.03524,9.73453,-3.43679]
cgo_arrow [-17.2952,-36.6277,-12.3879], [-35.7908,36.522,82.0165]
cgo_arrow [32.8768,10.3173,38.4204], [-85.9628,-10.423,31.2082]
cgo_arrow [-62.7094,35.4285,37.7446], [9.62341,-35.5341,31.8841]
```

3.6 Saving results in the JSON format

To facilitate the usage of AnAnaS in automated pipelines, we provide an option to output the result in the JSON format. To obtain this output, use the option "--json <jsonFilename>". It will output a JSON array with the specification described in the json-schema.json file.

Please type in the terminal "AnAnaS 5x47.pdb t --json out.json". This example creates an output file "out.json" containing the detailed result. This file is easy to read and understand for humans and also easy to parse. Libraries to parse JSON files are available for most programming languages.

3.7 A GUI interface

A SAMSON module is available for the AnAnaS method at samson-connect.net. It provides a convenient interactive graphical user interface, as shown in Figure 3. To use this module, select the structure on which you want to run the symmetry analysis. Then choose the symmetry group

you want to test or keep it "Automatic" otherwise, and click on "Compute Symmetry". The list of the results appears for each of tested symmetry groups. To visualize the axes, just click on an element in the list. One can also highlight specific axes by selecting them in the list, or move the camera along the chosen axis direction by double-clicking on the axis information in the list.

3.8 Symmetrize a protein assembly

A common task in structural bioinformatics is to modify a structure that is approximately symmetric to make it perfectly symmetric. AnAnaS provides this feature with the option "--symmetrize <symmetric pdb output>". When used with this option, AnAnas will replicate and rotate the first subunit to create a perfectly symmetric assembly. One should explicitly provide a symmetry group when using this option. This option may be particularly useful to generate full assembly from ones with missing subunits. Please type in the terminal "AnAnaS 2gza.pdb c6 –symmetrize 2gza_c6.pdb" then "AnAnaS 2gza.pdb c7 –symmetrize 2gza_c7.pdb" to obtain two full assemblies, with different symmetry order from the partial assembly 2gza.

4. Case Studies

4.1 How good are PDB symmetry annotations?

We use AnAnaS to assess the quality of symmetry annotations in the PDB (4). In 98.1% of the cases, the annotation from PDB and the AnAnaS results were identical. AnAnaS was also able to find a higher order symmetry group in 1.6% of the cases. Figure 4 shows the number of structures with different symmetry groups, according to AnAnaS or PDB.

*Table 1 Summary of the symmetry groups annotated in the PDB (rows) against the ones discovered by AnAnaS (columns). For example, the first cell shows that there are 54 structures annotated as C2 in the PDB for which AnAnaS did not find any symmetry. * compatible groups, ** incompatible groups.*

		AnAnaS Detection																		
		C1	C2	C3	C4	C5	C6	C7	C8	D2	D3	D4	D5	D6	D7	D8	T	O	I	Total
PDB Annotations	C2	54*	33091	8**	23*		6*			470*	15*	7*	1*	1*	205*	2*				33883
	C3	2*	3**	4188			16*				60*									4269
	C4	1*	2*		1046				7*			4*								1060
	C5	6*				561							1*							568
	C6		2*	2*			411							1*						416
	C7							104							6*					110
	C8								34							3*				37
	D2	3*	26*							6571		6*		1*			2*			6609
	D3		8*	5*							1939									1952
	D4	1*	1*									654				5*				661
	D5		1*										236							237
	D6													106						106
	D7	1*						1*							99					101
	D8															34				34
	T					2**											359	3*		364
	O																	329		329
	I	6*															2*		617	625

4.2 Are the big assemblies more symmetrical than the small ones?

A simple geometric intuition would suggest that as the angular uncertainty in the packing of subunits should stay constant with the size of the assembly, the imperfection of its symmetry becomes more pronounceable as it grows larger. Therefore, one could expect a linear correlation between the RMSD symmetry measure and the radius of gyration of the assemblies. We used AnAnaS to conduct an experiment to compare these values for all the assemblies in the PDB (4). We surprisingly found the opposite. More precisely, the bigger assemblies seems to be better organized and have smaller imperfections compared to the small ones. Figure 4 provides more details of this experiments.

5. FAQ

Why do I obtain different RMSD for the same axis in different symmetry groups?

To perform computation, AnAnaS uses sequentially aligned alpha-carbons as the reference points. Different symmetry groups will require different sets of subunits to be sequentially aligned and thus change the reference points used.

When should I change the default symmetry threshold value?

The default threshold of 7 Å is usually a good choice to consider an assembly as symmetric, when the radius of gyration of this assembly is between 15 and 100 Å. For very small assemblies however, this cutoff might be too large, and one may reduce it to keep fewer results. In the opposite case, if one wants to work with large assemblies with large deformation amplitudes, one may increase the threshold.

The symmetry of my assembly is not detected, what can I do?

The first step is to check the number of chains of your assembly. AnAnaS will try to find a symmetry involving every single chain of the assembly. If some chains are missing or additional chains are present, the recognition of the symmetry group will fail. Make sure you removed chains not involved in the symmetry. Provide the symmetry group in input if your assembly has missing subunits.

Second, you may try to increase the RMSD threshold. If your assembly presents large amplitude deformation, it may fall within the predefined threshold.

Last, it may happen that AnAnaS is unable to properly determine correspondences between the different chains of the assembly. In this case, you may provide manually the correspondences.

How can I provide manually the correspondence between the chains?

The correspondence may be provided using the -P <correspondence> option. For cyclic assemblies, the correspondence between the chains must be provided as shown in Figure 5. The chains are labeled from zero in the order of appearance in the input file. For the dihedral and cubic assemblies, two correspondences should be given, one corresponding to a rotation of order n (or 3 for cubic), and one corresponding to a rotation of order 2. This option is rather difficult to use, thus we recommend the user to consult with the detected permutations typing the "-p" flag.

Acknowledgment

The authors thank Nikolay Mayorov for his 2D trust-region algorithm developed during Google Summer of Code 2015, and Sergei Khashin from Ivanovo State University for his forth order polynomial solver available at http://math.ivanovo.ac.ru/dalgebra/Khashin/poly/index.html. The authors also thank Elvira Kinzina and Andrei Kazennov from MIPT Moscow for their support at the initial stage of the project. This work has been supported by L'Agence Nationale de la Recherche (grant number ANR-15-CE11-0029-03).

References

1. *On the root mean square quantitative chirality and quantitative symmetry measures.* **Petitjean, M.** 1999, J Math Phys, pp. 4587-4595.

2. *Analytical methods for caluclating continuous symmetry measures and the chirality measure.* **Pinsky, M., et al.** 2008, J Comput Chem, pp. 2712-2721.

3. *Analytical symmetry detection in protein assemblies. I. Cyclic symmetries.* **Pagès, G., Kinzina, E. et Grudinin, S.** 2018, Journal of Structural Biology, pp. 142-148.

4. *Analytical symmetry detection in protein assemblies. II.Dihedral and Cubic symmetries.* **Pagès, G. et Grudinin, S.** 2018, Journal of Structural Biology, pp. 185-194.

5. *Rapid determination of RMSDs corresponding to macromolecular rigid body motions.* **Popov, P. et Grudinin, S.** 2014, Journal of Computational Chemistry, 35 (12), pp.950-956.

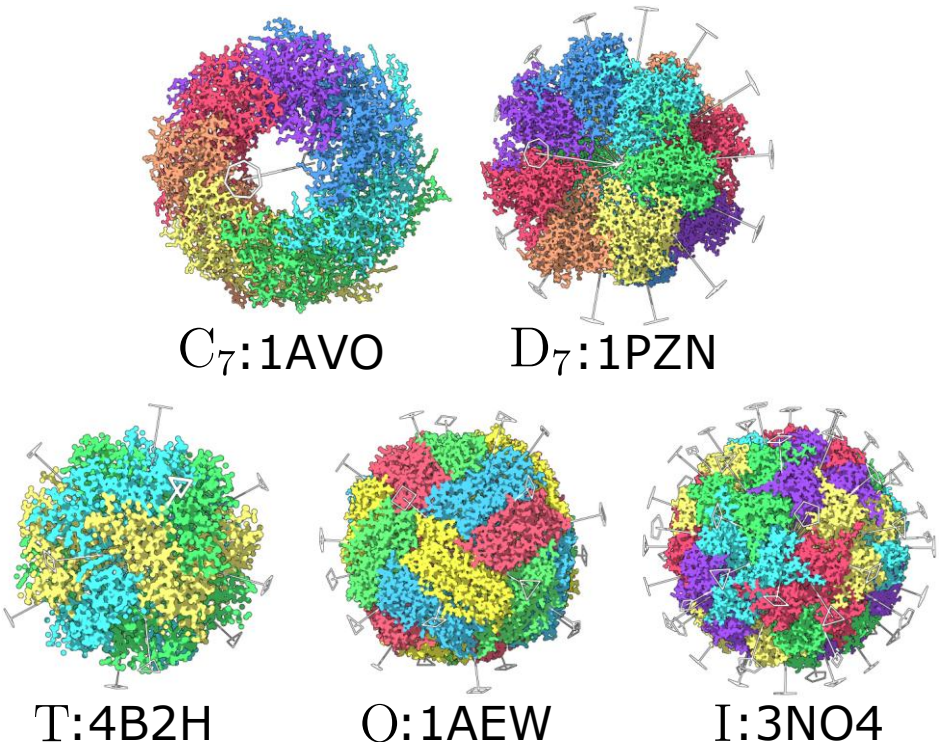

Figure 1 Five examples of symmetrical assemblies with symmetry axes. These are a C7 cyclic assembly, a D7 dihedral assembly, a T tetrahedral assembly, an O octahedral assembly, and an I icosahedral assembly. The order n of each axis is represented with a regular n-gone, except of order 2 represented with a rhombus.

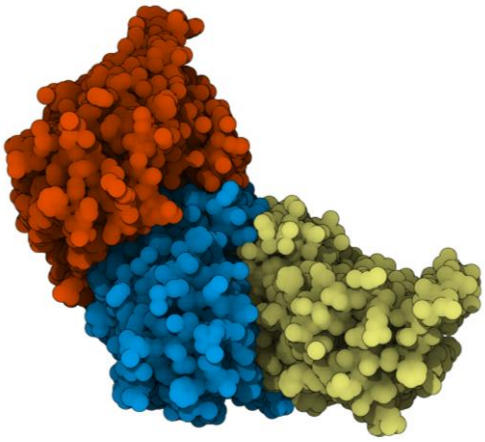

Figure 2 2gza, a C6 assembly with 3 missing chains

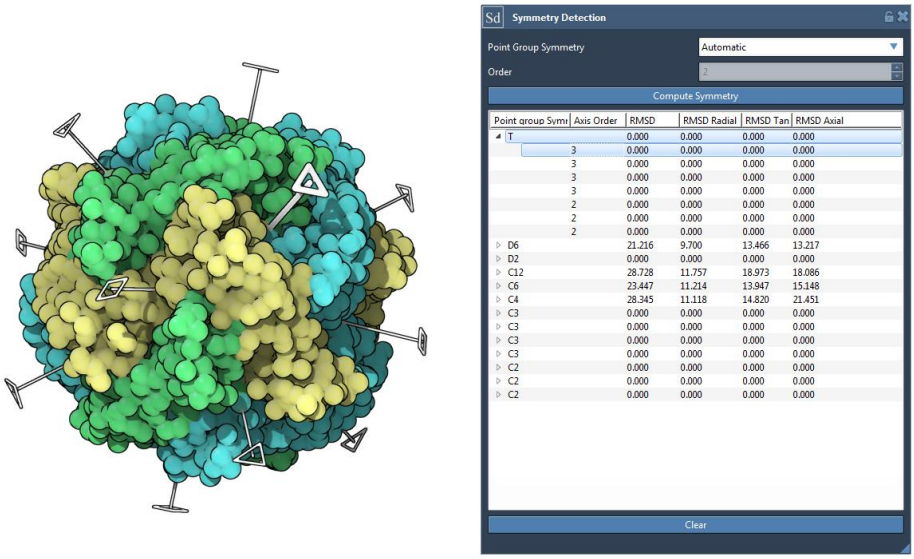

Figure 3 A SAMSON module that provides an interactive graphical user interface for the AnAnaS method. The structure in the example is 4B2H. All the point group symmetries have been automatically detected and are listed with their RMSD measures. When a point group is selected, the axes are displayed in the viewport and the detail of axis order, and RMSD by axis is displayed. Clicking on an axis in the list highlights it on the viewport.

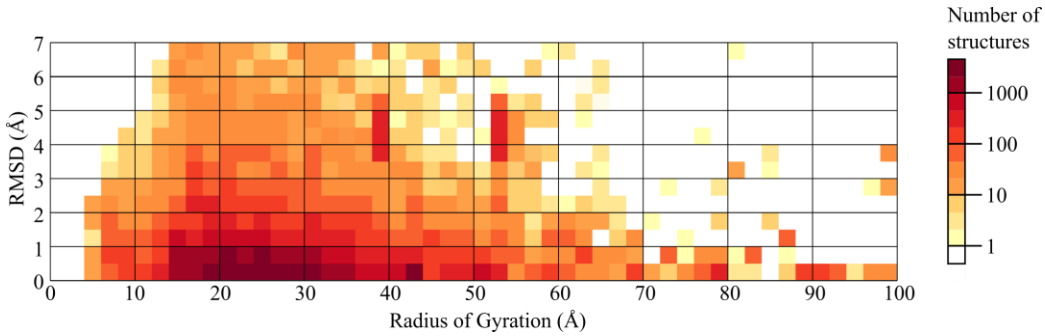

Figure 4 Relation between the RMSD symmetry measure and the radius of gyration of all structures annotated as symmetric in the PDB. Surprisingly, there is no positive correlation between the two.

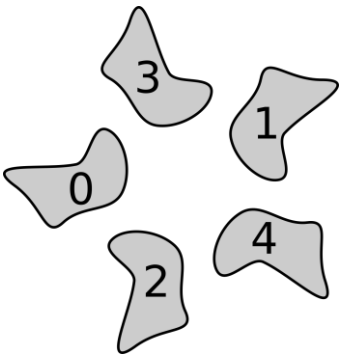

*Figure 5 Example of how to provide the correspondence between the chains. Here the provided correspondence should be (**3,4,0,1,2**) as after a fifth of a turn, the subunit 0 goes to the place of the subunit 3 (0->**3**), 1->**4**, 2->**0**, 3->**1**, 4->**2**.*