STRUCTURAL MODELLING

RES-Xre Toxin-Antitoxin System

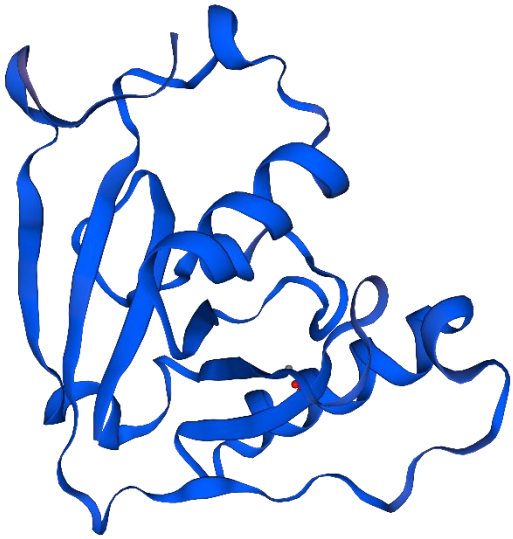
Elvan Wiyarta
Faculty of Medicine Universitas Indonesia

Overview

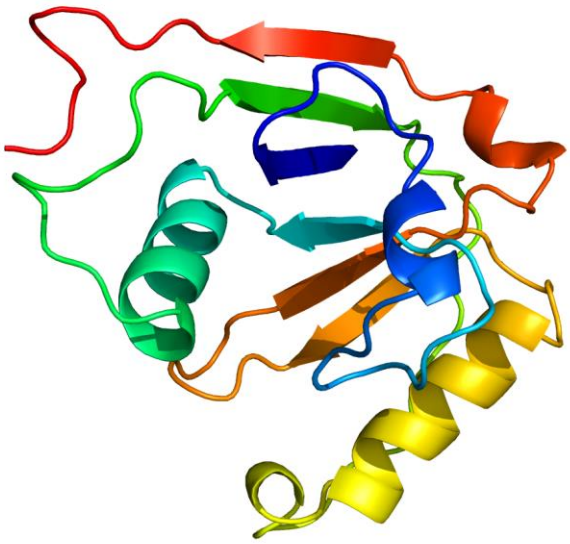
In collaboration with **National University of Singapore** (NUS), I modelled a 3D protein structure of **RES Toxin** and **Xre Antitoxin.** As known, structural protein modelling is a predictive modelling base on biological database that could assemble the real biological structure of the protein itself. On this project, I conducted a protein modelling through **homology modelling.** Then, I checked the result of the modelling and did some tests to make sure that the modelled protein has high reliability. As requested from NUS Drylab Team, the modelled structure will then be test for its **stability** which will be measured by **Gibbs Free Energy (ΔG°).**

RES TOXIN

To model the toxin, I used **SwissModel** and **PHYRE2** to predict the best homology model for the toxin. Both models are shown below.



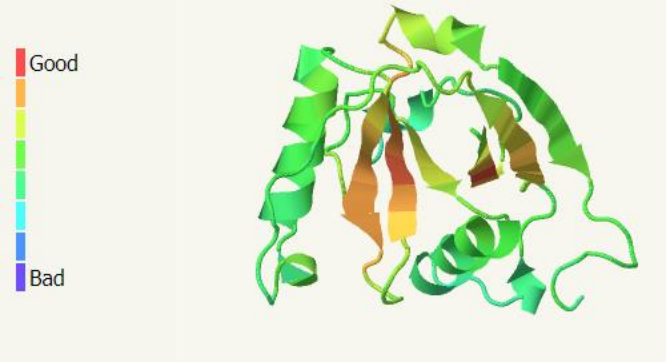
SwissModel



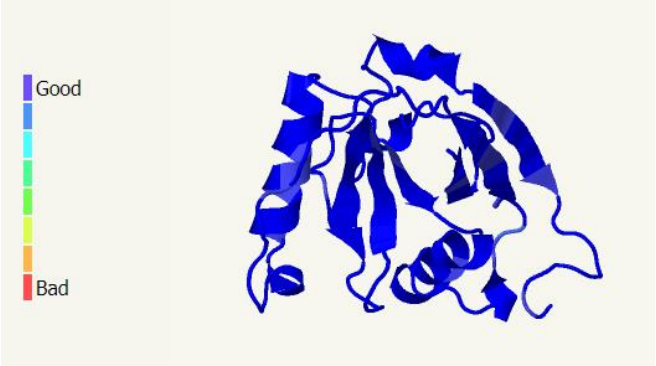
PHYRE^2

Both model have 100% sequence identity with template (6GW6.1.A). In order to match both of the model, I aligned them in **YASARA** using **MUSTANG.** The result for the alignment is that these model has an RMSD of **0.167 A** over 144 aligned residues with 100.00 sequence identity. Based on the RMSD, these model are **similar.**

After model the structural modelling, I wanted to make the model more reliable so that it can predict the real biological structure. I ran on some investigator tools, then come out with the following results.



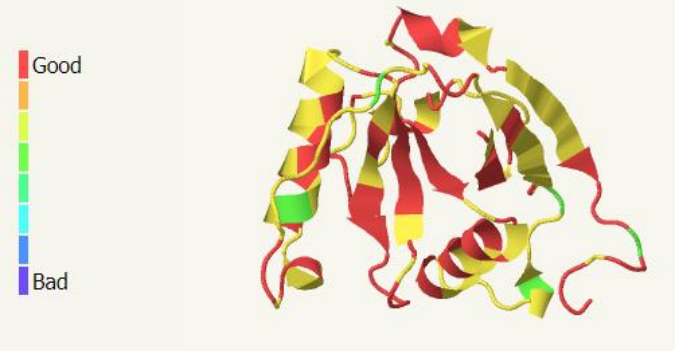
ProQ2 Quality Assessment



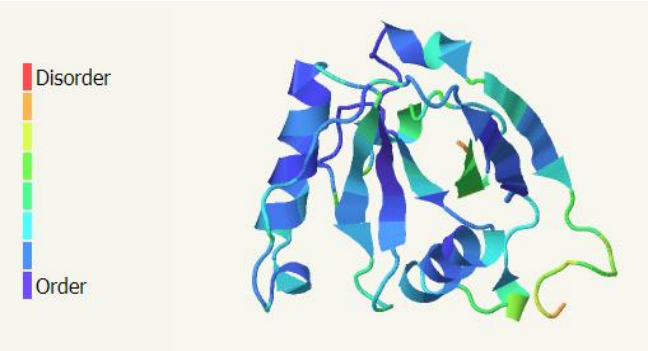
Clashes

Rotamers

Ramachandran Analysis



Alignment Confidence



Disorders

Analysis of each tools are mention bellow.

1. **ProQ2 quality assessment**
   ProQ2 is a model quality assessment algorithm that uses support vector machines to predict local as well as global quality of protein models. The RES toxin model has an overall **good quality.**
2. **Clash Analysis**
   Some atoms in some residues may lie too close to one another in the model. Residues are coloured by how many clashes are observed. A large number of clashes could mean bad sidechain placement, or possibly an incorrect backbone in this region. The clash criteria are particularly strict. The RES toxin model **has no clashes.**
3. **Rotamer analysis**
   Some sidechains in the model may not have been modelled ideally. These are coloured in red and may indicate a problem with the backbone or underlying alignment in this region. The RES toxin has **a very good** model with allowed sidechain.
4. **Ramachandran analysis**
   Some residues in the model may lie in favourable (blue), allowed (green) or disallowed (red) regions of the Ramachandran plot. This colouring indicates residues that may have problems with the backbone phi/psi angles. The RES toxin model has **no disallowed** residue.
5. **Alignment Confidence**
   This is the reliability of the pairwise query-template alignment as reported by HHsearch. The confidence values are obtained from the posterior probabilities calculated in the Forward-Backward algorithm. This model has **high** confidence level.
6. **Disorder Prediction**
   Disordered regions are important in the function of many proteins. Disordered regions are dynamically flexible and are distinct from irregular loop secondary structures, which are static in solution. Disorder prediction is also likely to be a valuable tool for identifying flexible regions that may hinder successful protein crystallization. This prediction has been made by the knowledge-based **Disopred** method. Based on that, this model has **low disorder** residue which could be an insight for the protein stability.

To make sure about the protein model, I conducted a secondary structure analysis on **PROCHECK.**

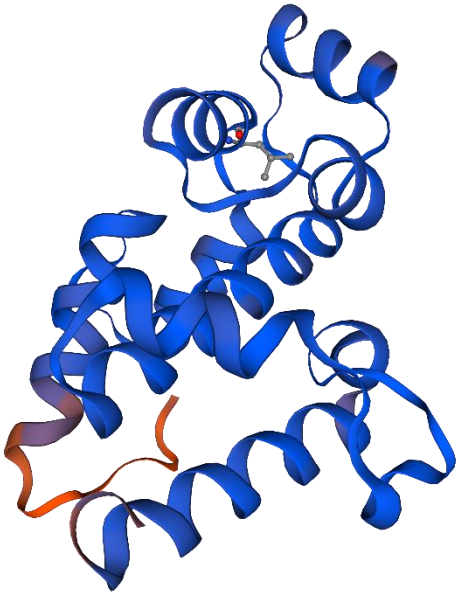
```
+----------<<<  P  R  O  C  H  E  C  K     S  U  M  M  A  R  Y  >>>----------+
|                                                                             |
| /var/www/PROCHECK/Jobs/3976131/3976131.pdb   1.5              145 residues |
|                                                                             |
| Ramachandran plot:   92.5% core    7.5% allow    0.0% gener    0.0% disall |
|                                                                             |
| All Ramachandrans:    0 labelled residues (out of 143)                      |
+| Chi1-chi2 plots:      1 labelled residues (out of  78)                      |
| Side-chain params:    5 better     0 inside      0 worse                    |
|                                                                             |
*| Residue properties: Max.deviation:     3.9             Bad contacts:    0 |
*|                     Bond len/angle:    6.5    Morris et al class:  1  1  2 |
|                                                                             |
| G-factors           Dihedrals:  -0.19  Covalent:   0.01    Overall:  -0.10 |
|                                                                             |
+| Planar groups:    91.7% within limits   8.3% highlighted                    |
|                                                                             |
+-----------------------------------------------------------------------------+
  + May be worth investigating further.  * Worth investigating further.
```

Based on the PROCHECK analysis, the RES toxin model has a **high reliability.**

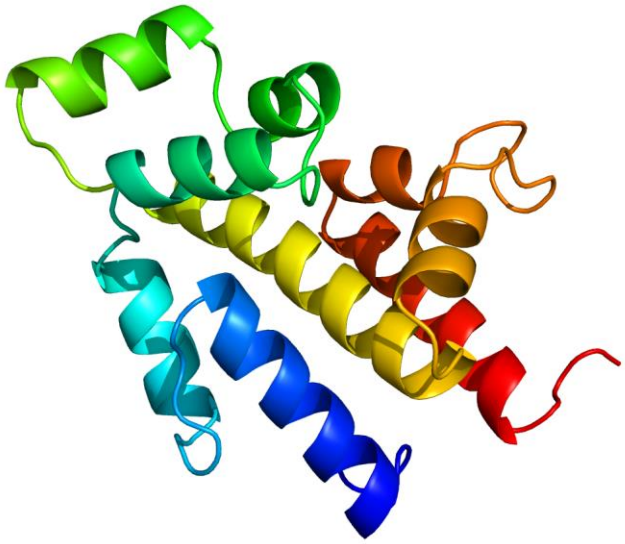
Regarding all analyses that have been conducted, **the RES toxin model has a high reliability** and could be used for further analysis on the protein stability.

Xre ANTITOXIN

To model the antitoxin, I used **SwissModel** and **PHYRE2** to predict the best homology model for the antitoxin. Both models are shown below.



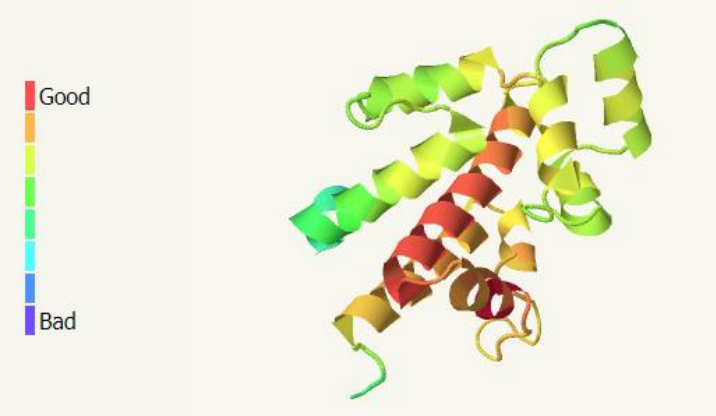
SwissModel



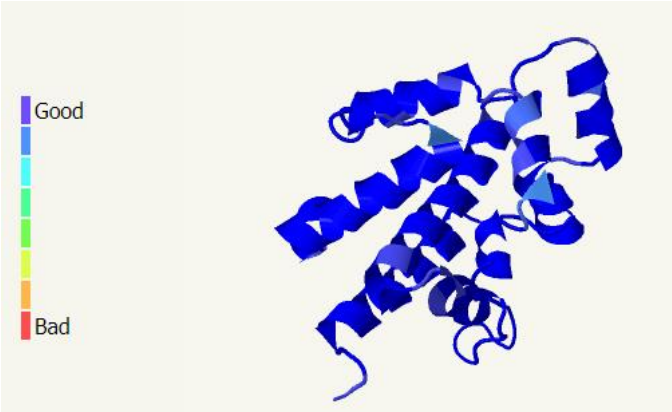
PHYRE2

Both model have 100% sequence identity with template (6GW6.1.(B: PHYRE; C: SwissModel). In order to match both of the model, I aligned them in **YASARA** using **MUSTANG**. The result for the alignment is that these model has an RMSD of **0.795 A** over 144 aligned residues with 100.00 sequence identity. Based on the RMSD, these model are **similar.**

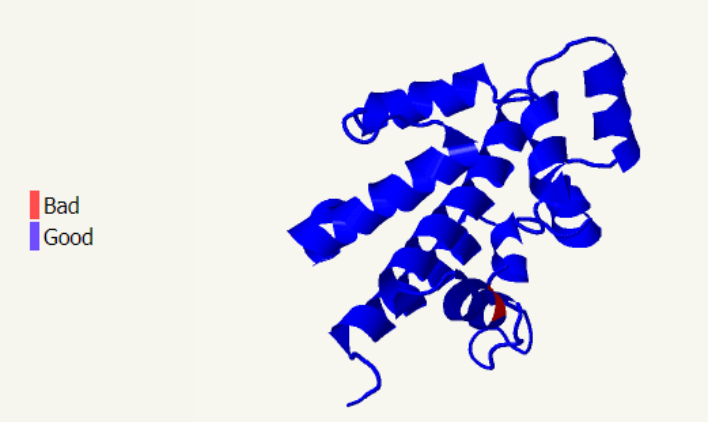
After model the structural modelling, I wanted to make the model more reliable so that it can predict the real biological structure. I ran on some investigator tools, then come out with the following results.



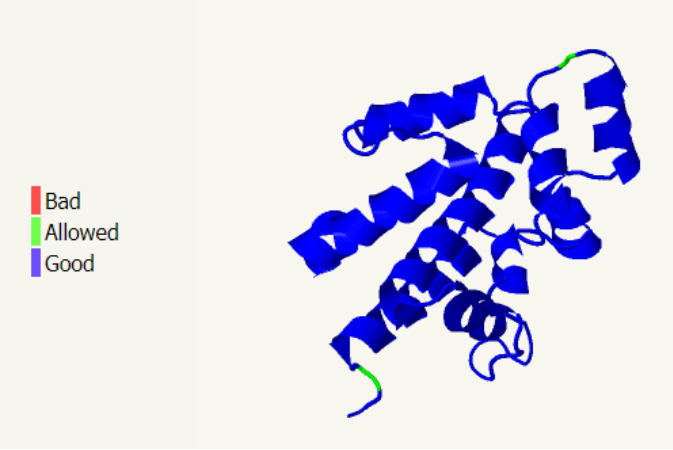
ProQ2 Quality Assessment



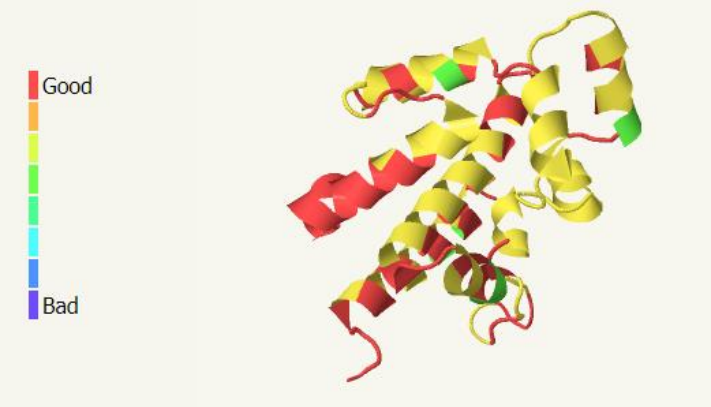
Clashes



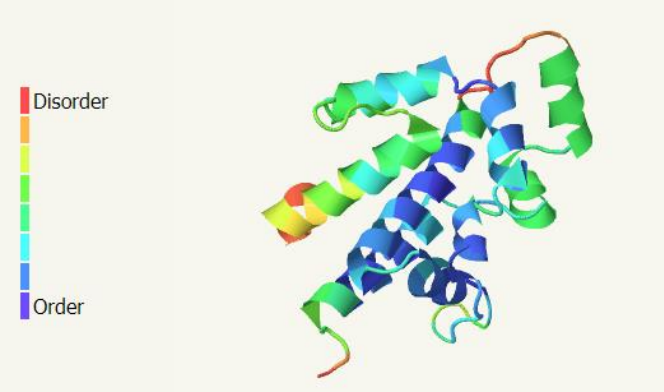
Rotamers



Ramachandran Analysis



Alignment Confidence



Disorders

Analysis of each tools are mention bellow.

1. **ProQ2 quality assessment**
   ProQ2 is a model quality assessment algorithm that uses support vector machines to predict local as well as global quality of protein models. The Xre Antitoxin model has an overall **good quality.**
2. **Clash Analysis**
   Some atoms in some residues may lie too close to one another in the model. Residues are coloured by how many clashes are observed. A large number of clashes could mean bad sidechain placement, or possibly an incorrect backbone in this region. The clash criteria are particularly strict. The Xre antitoxin model **has no clashes.**
3. **Rotamer analysis**
   Some sidechains in the model may not have been modelled ideally. These are coloured in red and may indicate a problem with the backbone or underlying alignment in this region. The Xre antitoxin has a **two** sidechain that highlight in red (**Arg7** and **Lys108**). This result causes no problem on model since both of the residue indeed has many variations on rotamers.
4. **Ramachandran analysis**
   Some residues in the model may lie in favourable (blue), allowed (green) or disallowed (red) regions of the Ramachandran plot. This colouring indicates residues that may have problems with the backbone phi/psi angles. The Xre antitoxin model has **no disallowed** residue.
5. **Alignment Confidence**
   This is the reliability of the pairwise query-template alignment as reported by HHsearch. The confidence values are obtained from the posterior probabilities calculated in the Forward-Backward algorithm. This model has **high** confidence level.
6. **Disorder Prediction**
   Disordered regions are important in the function of many proteins. Disordered regions are dynamically flexible and are distinct from irregular loop secondary structures, which are static in solution. Disorder prediction is also likely to be a valuable tool for identifying flexible regions that may hinder successful protein crystallization. This prediction has been made by the knowledge-based Disopred method. Based on that, this model has **moderate disorder** residue which could be an insight for the protein stability.

To make sure about the protein model, I conducted a secondary structure analysis on **PROCHECK.**

```
+----------<<<  P  R  O  C  H  E  C  K     S  U  M  M  A  R  Y  >>>----------+
 |                                                                            |
 | /var/www/PROCHECK/Jobs/1206642/1206642.pdb   1.5              149 residues |
 |                                                                            |
*| Ramachandran plot:   93.9% core    4.5% allow    0.8% gener    0.8% disall |
 |                                                                            |
+| All Ramachandrans:    2 labelled residues (out of 147)                     |
 | Chi1-chi2 plots:      0 labelled residues (out of  95)                     |
 | Side-chain params:    5 better     0 inside      0 worse                   |
 |                                                                            |
+| Residue properties: Max.deviation:     5.1             Bad contacts:    0  |
+|                     Bond len/angle:    3.8    Morris et al class:  1  1  2  |
 |                                                                            |
 | G-factors           Dihedrals:  -0.05  Covalent:   0.02    Overall:  -0.01 |
 |                                                                            |
+| Planar groups:    93.0% within limits   7.0% highlighted       1 off graph |
 |                                                                            |
 +----------------------------------------------------------------------------+
  + May be worth investigating further.  * Worth investigating further.
```

Based on the PROCHECK analysis, the Xre antitoxin model has a **high reliability**.

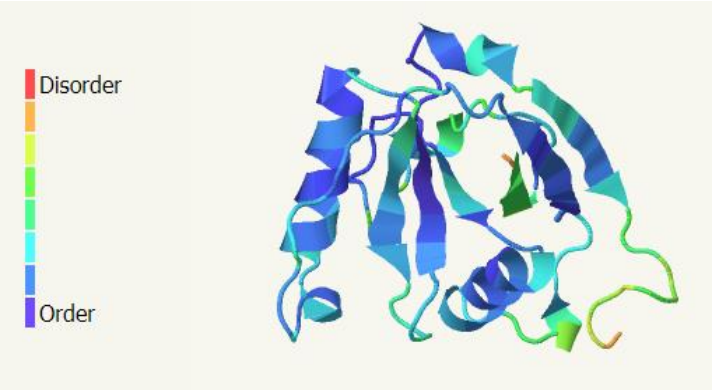
Regarding all analyses that have been conducted, **the Xre antitoxin model has a high reliability** and could be used for further analysis on the protein stability.

PROTEIN STABILITY

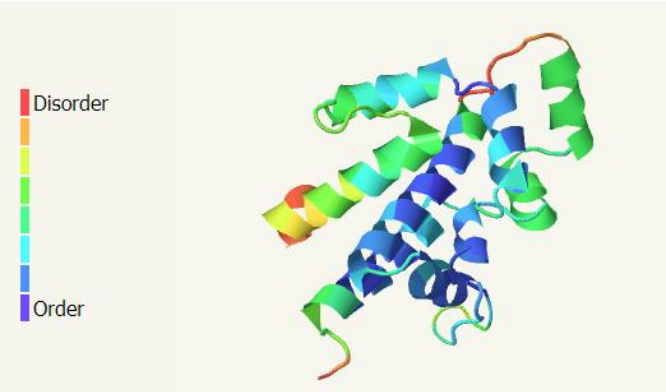
After make sure that the models have a high reliability, then I could use the model to proceed to the core parameter for this report of structural modelling. I conducted the protein stability using **FoldX** on **YASARA** then generate the stability using **Gibbs Free Energy (ΔG°)** as the parameter. The results are shown below.

Protein	ΔG°
RES toxin	32.32 kcal/mol
Xre antitoxin	54.33 kcal/mol

Based on the data, we could conclude that the Xre antitoxin need **more energy** in order to maintain the certain folding formation of the model rather than the RES toxin. This could be linked to the **disorder test** that I conducted on the previous section. The comparison of both model in the disorder test are shown below



RES Toxin



Xre Antitoxin

From the picture above we could see that the RES toxin **has more blue-highlighted residues** rather than the Xre antitoxin which means that the RES toxin has more **residues-in-order** rather than Xre antitoxin. These disorder residues could affect the protein folding energy thus make the protein more unstable.

Based on the structural analysis and comparison on both model, I could conclude that **RES toxin is more stable than the Xre antitoxin.**

n.b. this analysis needs to be continued to produce more accurate predictions.

Reference

- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296-W303 (2018).

- Benkert, P., Biasini, M., Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27, 343-350 (2011).

- Berman, H., Henrick, K., Nakamura, H. and Markley, J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res, 35, D301-303.

- The UniProt Consortium. (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res, 45, D158-D169.

- Kelley LA et al. Nature Protocols 10, 845-858 (2015)

- Pontius J, Richelle J, Wodak SJ. Deviations from standard atomic volumes as a quality measure for protein crystal structures. J Mol Biol. 1996 Nov 22;264(1):121-36.

- Arjun Ray, Erik Lindahl and Bjorn Wallner. BMC Bioinformatics 2012, 13:224.