Homology Modeling of Protein Structures

1. Major Facilitator Superfamily (MFS)

Major facilitator superfamily MFS transporters play a crucial role in moving various small compounds across biological membranes. In bacteria, they facilitate nutrient uptake and expulsion of harmful substances. Human MFS transporters, while performing similar functions such as nutrient absorption and clearance, have evolved to transport a wider range of substances between cells and within cells. Implicated in diseases like cancer and metabolic disorders, MFS transporters are potential drug targets. Additionally, they are key players in pharmacokinetics, transporting toxins and drugs. These transporters can be uniporters, symporters, or antiporters, depending on whether they transport a single substrate, a substrate coupled with an ion, or two different molecules in opposite directions.

2. Determining the MFS Protein

The MFS profile domain-containing protein **SACS2** has been determined experimentally as an MFS protein without a solved structure. The UniProt ID of the relevant protein is Q97XW7 and according to the information obtained, it has been associated with the organism Sulfolobus solfataricus.

Figure 1. Displayed the Q97XW7 in UniProt Database.

3. Function of Q97XW7 Protein

Enables the transfer of a substance, usually a specific substance or a group of related substances, from one side of a membrane to the other. The protein related to the function of substrate-specific transmembrane transporter activity, uptake transmembrane transporter activity, uptake permease activity,and substrate-specific transporter activity.

4. Created Homology Modelling with SWISS-MODEL

The sequence of the corresponding protein obtained from UniProt will be used for homology modeling in SWISS-MODAL (Figure 2).

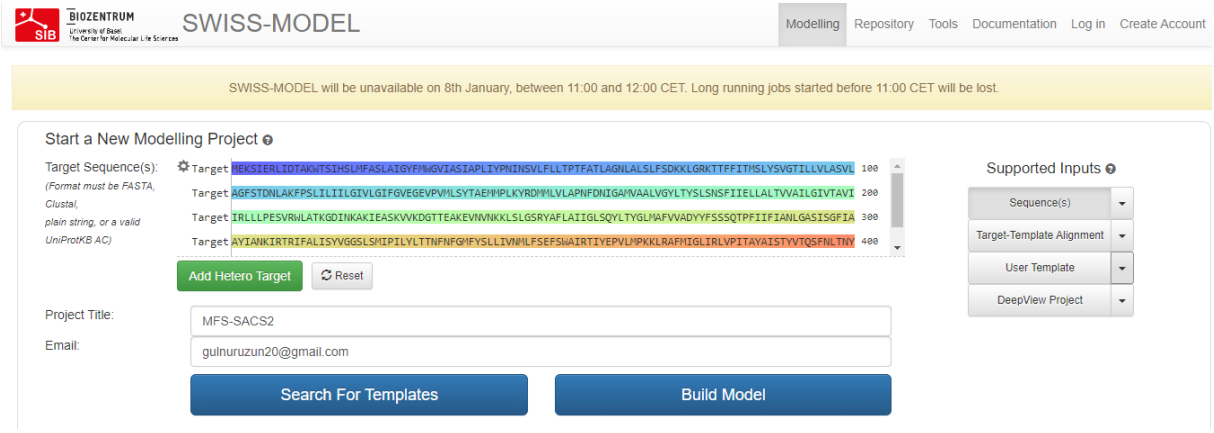

Figure 2. First step of Homology Modelling on SWISS-MODAL.

The template with the closest and highest identity as the reference sequence model is displayed in Figure 3.

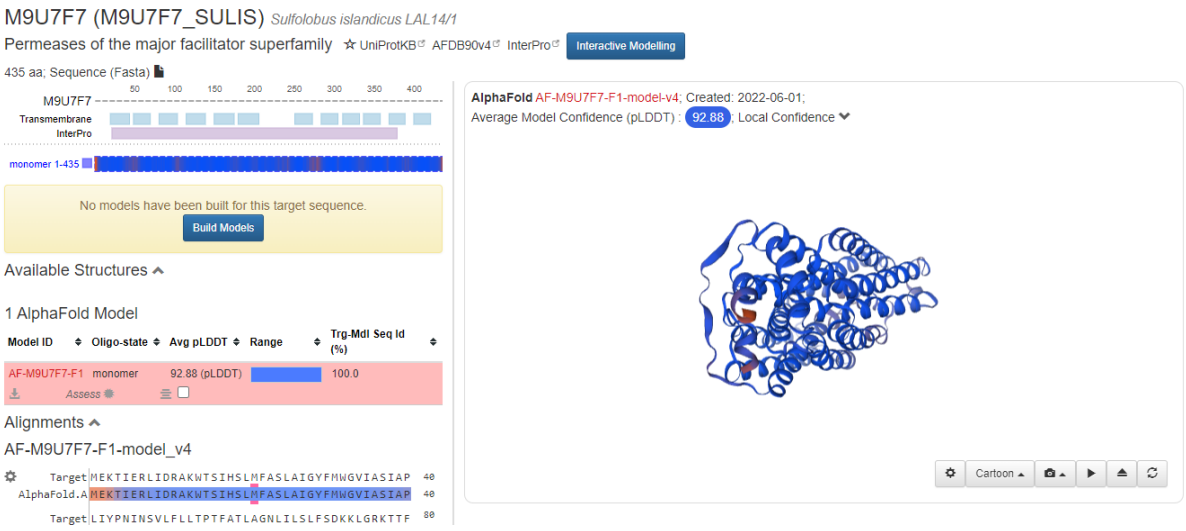

Figure 3. A Template with a High Identity: M9U7F7

5. Interpretation and Reporting using Structure Assessment

We compared this model with other templates by determining the seq identity that is closest to it as the highest template i target. Displayed as the target modal M9U7F7.1.A Permeases of the major facilitator superfamily.

GMQE 0.93, Seq-identity 98.62, Seq-similarity 0.59

GMQE (Global Model Quality Estimate): With a value of 0.93, it indicates that the model is generally reliable and of high quality.

Sequence Identity: A high value of 98.62% suggests a significant similarity between the model and the target sequence, indicating a reliable alignment.

Sequence Similarity: The value of 0.59 suggests a moderate level of similarity between the model and the target sequence.

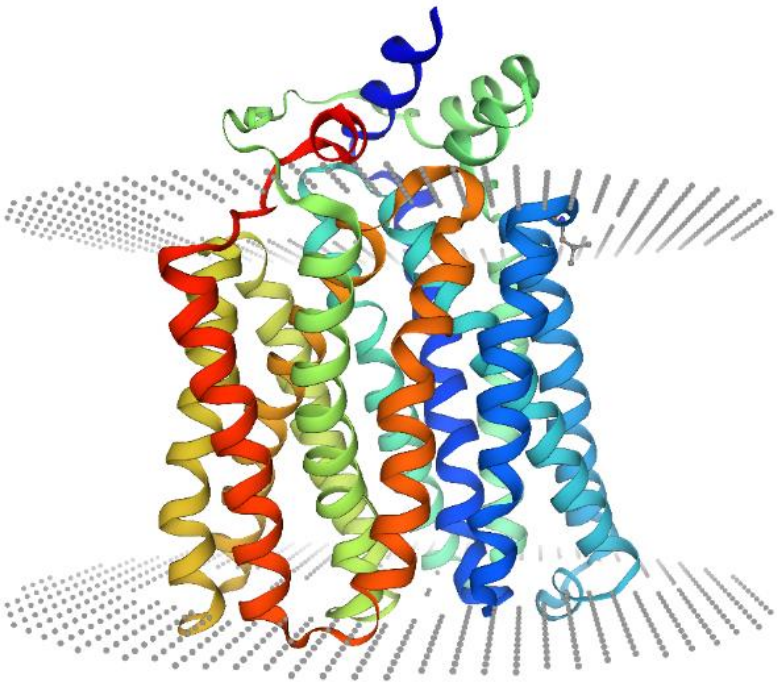

Figure 4. Target Model: M9U7F7.1.A Permeases of the Major Facilitator Superfamily.

Ramachandran Favoured **98.85%**

Ramachandran Outliers **0.00%**

Ramachandran Favoured: The high percentage of 98.85% indicates that the majority of the modeled protein structures have a physically acceptable conformation.

Ramachandran Outliers: The absence of outliers (0.00%) in the Ramachandran plot suggests that the model produces protein structures with minimal undesirable conformations.

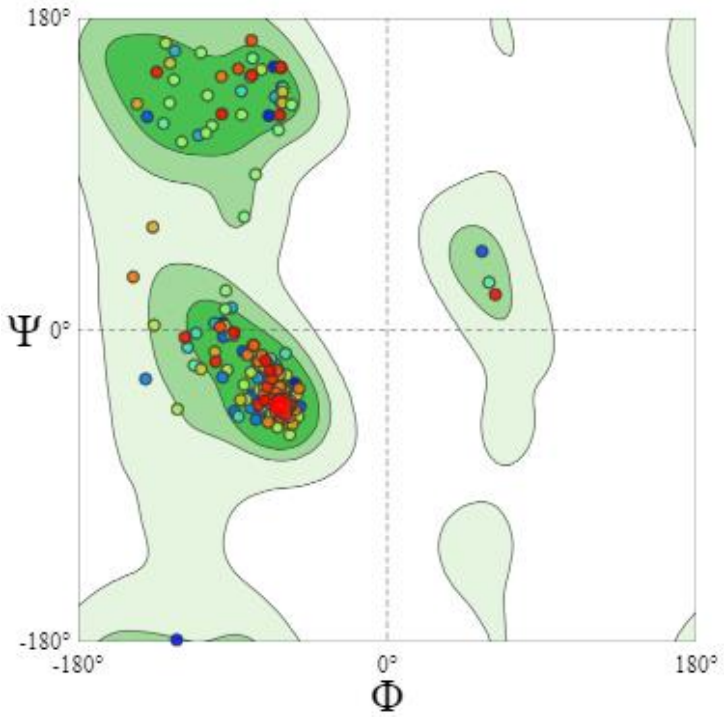

Figure 5. Ramachandran Plot of Target Modal M9U7F7.1.A

In conclusion, based on these values, it can be stated that the template used in homology modeling is generally reliable and of high quality. The high sequence identity, low Ramachandran outliers, and high GMQE values suggest that the model successfully mimics the target sequence and is likely to produce biologically meaningful results.

7aar.1.A Sugar transport protein 10 (Figure 6)

GMQE 0.53, Seq-identity 18.07, Seq-similarity 0.29

Based on the provided values, the following comments can be made regarding the results of homology modeling:

GMQE: A value of 0.53 indicates that the overall quality of the model is lower, suggesting that significant improvements are needed. The presence of semi-stable states in MFS proteins may imply that the model struggles to effectively capture dynamic conformational changes.

Sequence Identity: The value of 18.07% indicates a low similarity between the model and the target sequence. This suggests that the model struggles to accurately mimic the structure of the target protein.

S**equence Similarity:** With a value of 0.29, it signifies a low similarity between the model and the target sequence, indicating a lack of compatibility.

For seq-identity and seq-similarity The semi-stable states in the energy landscape of MFS proteins may contribute to the model's challenge in capturing various conformational states and result in a lack of compatibility with the target sequence.

Based on these values, it can be concluded that the model in the homology modeling process is not reliable and requires improvement. The low GMQE, low sequence identity, and low sequence similarity suggest that the model does not adequately match the target, and the resulting structures are less likely to be biologically meaningful. The existence of semi-stable states in the energy landscape of MFS proteins suggests that corrective measures, such as model optimization or choosing a different template, are necessary. This is essential for the model to better understand dynamic conformational changes and simulate the energy landscape more effectively.

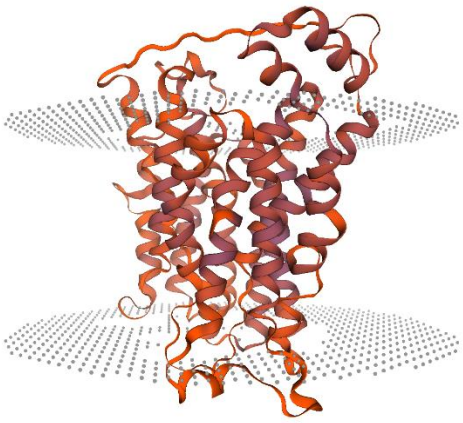

Figure 6. Template 1: 7aar.1.A Sugar transport protein 10

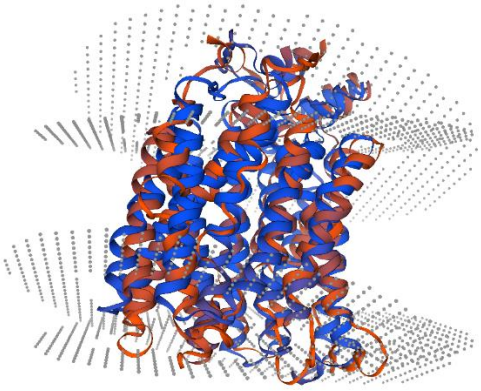

Figure 7. Target Modal-Template 1

Ramachandran Favoured 91.29%

Ramachandran Outliers 3.29% (A190 VAL, A91 THR, A412 GLY, A40 PRO, A413 ILE, A10 ASP, A278 SER, A90 GLY, A189 VAL, A243 LYS, A188 THR, A127 GLY, A12 ALA, A112 PRO)

Regarding the provided Ramachandran plot values and outliers, the following comments can be made as the Figure 8:

Ramachandran Favoured: With a percentage of 91.29%, it indicates that the majority of the protein structures generated by the model have a physically acceptable conformation. This suggests that the model generally produces well-shaped protein structures.

Ramachandran Outliers: At 3.29%, it shows that some of the structures obtained by the model have undesired conformations in the Ramachandran plot. Although this percentage is relatively low, careful attention may be needed to address outliers at the specified amino acid positions. The outliers are specified for amino acid positions A190 VAL, A91 THR, A412 GLY, A40 PRO, A413 ILE, A10 ASP, A278 SER, A90 GLY, A189 VAL, A243 LYS, A188 THR, A127 GLY, A12 ALA, A112 PRO. The presence of outliers at these positions suggests that specific regions of the model may require careful adjustments or optimization.

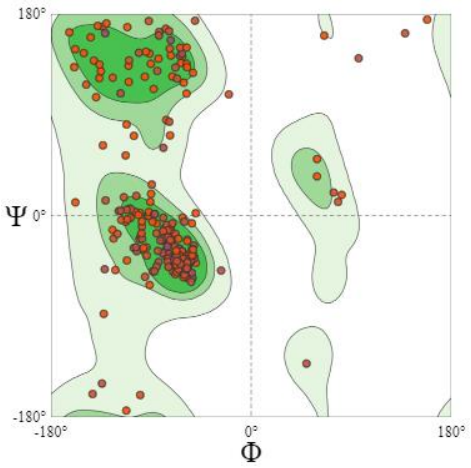

Figure 8. Ramachandran Plot of Template 1.

7aaq.1.A Sugar transport protein 10 (Figure 9)

GMQE 0.50, Seq-identity 18.02, Seq-similarity 0.29

Based on the provided values, the following comments can be made regarding the results of homology modeling:

GMQE: A value of 0.50 indicates that the overall quality of the model is low. This suggests that the reliability of the model is low, and improvements are needed. The semi-stable states of MFS proteins may suggest that the model struggles to effectively capture dynamic conformational changes and transitions between semi-stable states.

Sequence Identity: The value of 18.02% indicates a low similarity between the model and the target sequence. This suggests that the model struggles to accurately mimic the structure of the target protein.

Sequence Similarity: With a value of 0.29, it signifies a low similarity between the model and the target sequence. This indicates that the model is incompatible with the target and requires improvement.

Fort seq-similarity and seq-identity the semi-stable states in the energy landscape of MFS proteins may contribute to the model's difficulty in capturing various conformational states and lead to incongruence with the target sequence.

Based on these values, it can be concluded that the model in the homology modeling process is not reliable and is in need of significant improvement. The low GMQE, low sequence identity, and low sequence similarity suggest that the model does not adequately match the target, and the resulting structures are less likely to be biologically meaningful. Consideration should be given to optimizing the model, selecting a different template, or taking other corrective measures.

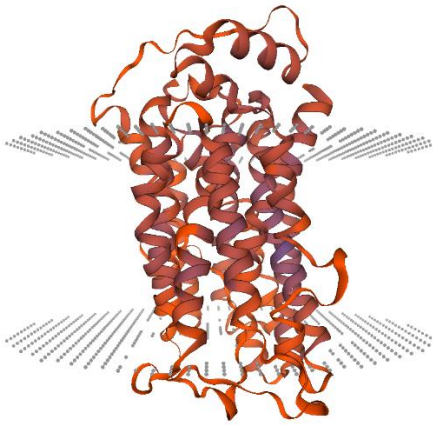

Figure 9. Template 2: 7aaq.1.A Sugar transport protein 10

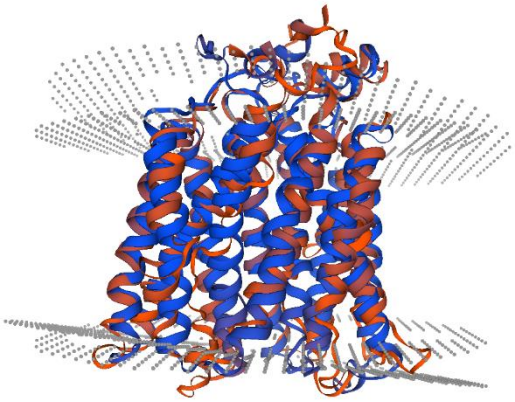

Figure 10. Target Modal-Template 2.

Ramachandran Favoured 93.10%

Ramachandran Outliers 0.71% (A197 THR, A230 LYS, A308 ARG)

Ramachandran Favored: A percentage of 93.10% indicates that the majority of the protein structures generated by the model will have a physically acceptable conformation. This suggests that the model will generally produce well-shaped protein structures.

Ramachandran Outliers: At 0.71%, it shows that a small percentage of the structures obtained by the model will have undesired conformations in the Ramachandran plot. The specific amino acid positions with outliers are A197 THR, A230 LYS, A308 ARG. Although the percentage is low, it will be important to consider these outliers and pay attention to improving the model in these specific regions.

In summary, while the Ramachandran plot values are generally positive, indicating a high percentage of favored conformations, the presence of outliers at specific positions suggests that adjustments or optimizations in those areas will be beneficial for further improving the model **(Figure 11).**

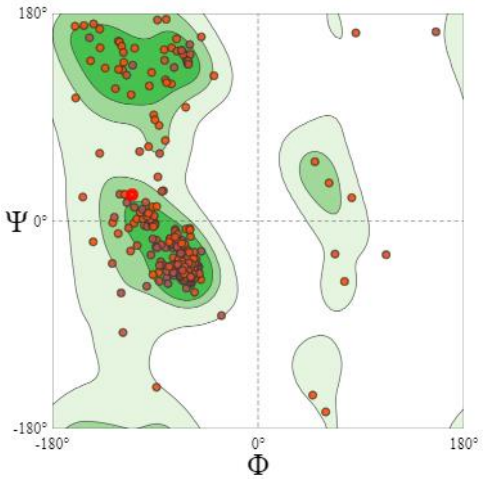

Figure 11. Ramachandran Plot of Template 2.

7sps.1.A Solute carrier family 2, facilitated glucose transporter member 3 (Figure 12)

GMQE 0.50, Seq-identity 16.58, Seq-similarity 0.28

Based on the provided values, the following academic interpretation can be made regarding the results of homology modeling:

GMQE: The value of 0.50 indicates a low overall quality of the model. This suggests a reduced reliability in the model, necessitating improvements.

Sequence Identity: A value of 16.58% signifies a diminished similarity between the model and the target sequence. This implies challenges in accurately emulating the structure of the target protein.

Sequence Similarity: With a value of 0.28, the model demonstrates a low similarity to the target sequence. This indicates a lack of congruence with the target and highlights the need for enhancement.

The low GMQE value of 0.50 suggests a reduced reliability of the model. In the context of MFS proteins and their energy landscape, this could indicate that the model might not effectively capture the dynamic conformational changes or transitions between different semi-stable states. The inherent flexibility and multiple energy minima in MFS proteins may pose a challenge for accurately predicting the structure, and the current model might not adequately represent these complexities.

The low sequence identity of 16.58% and sequence similarity of 0.28 further emphasize the challenges in accurately emulating the structure of the target protein. In the context of MFS proteins with multiple energy minima, it suggests that the model may struggle to capture the diverse conformational states, leading to a lack of congruence with the target sequence. Considering the energy landscape hint, it is crucial to improve the model to better account for the dynamic nature of MFS proteins. This might involve refining the modeling approach to incorporate information about the different semi-stable conformational states and optimizing the model parameters to better simulate the energy landscape.

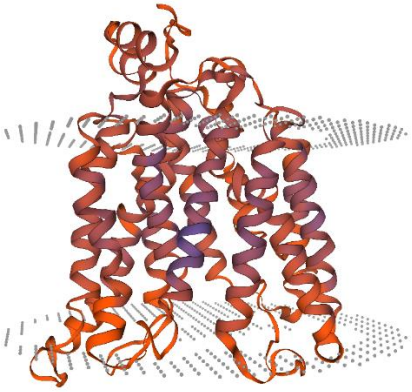

Figure 12. Template 3: 7sps.1.A Solute carrier family 2, facilitated glucose transporter member 3

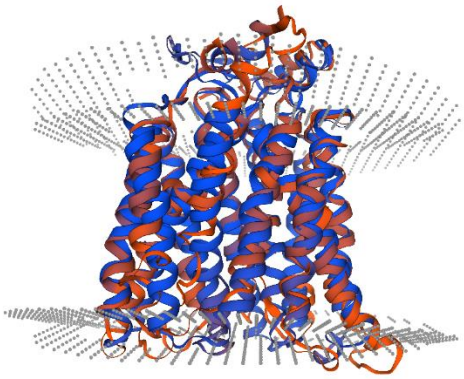

Figure 13. Target Modal-Template 3

Ramachandran Favoured 90.52%

Ramachandran Outliers 1.18% (A13 LYS, A402 LEU, A268 MET, A90 GLY, A49 VAL)

For the given Ramachandran plot values, an academic interpretation can be provided as follows (Figure 14):

Ramachandran Favoured: With a percentage of 90.52%, the model suggests that the majority of generated protein structures exhibit a physically acceptable conformation. This implies a generally satisfactory performance in terms of structural conformation.

Ramachandran Outliers: At 1.18%, a small percentage of structures deviate from the preferred regions in the Ramachandran plot. Specific amino acid positions with outliers are A13 LYS, A402 LEU, A268 MET, A90 GLY, A49 VAL. Although the overall percentage of outliers is relatively low, attention should be directed towards these specific positions to improve the model's performance in these regions.

In summary, while the Ramachandran plot values indicate a predominantly favorable conformation, the presence of outliers at specific positions suggests that targeted improvements or adjustments may enhance the model, particularly in the specified regions.

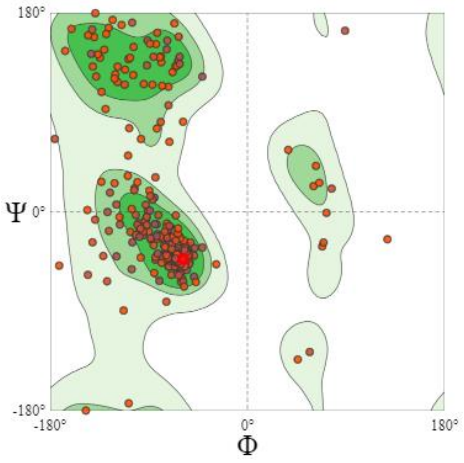

Figure 14. Ramachandran Plot of Template 3.

References (the sources used to make a statement in the first question):

1. https://www.nature.com/articles/nrm.2015.25#Sec2
2. https://doi.org/10.1021/acs.chemrev.0c00983
3. https://doi.org/10.1002/pro.2759