Fabian Gobet
Bioinformatics #5

Disclaimer: The code developed for this assignment may be found in my repository:
https://github.com/FabianGobet/USI-projects/tree/main/Bioinformatics/assignment5

1. Recovery of the primary structure of Human Adenosine Receptor A3 (hAA3R)

By Searching UniProt for adenosine receptor A3:

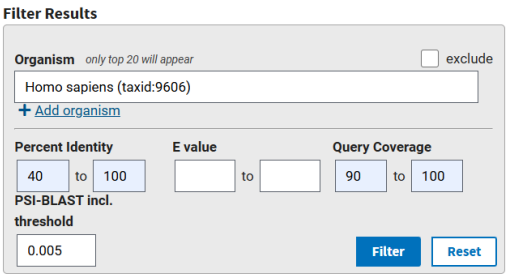
- PDB entry ID: P0DMS8
- Gene Name: ADORA3
- Organism: Homo Sapiens
- Length: 318
- Link: https://www.uniprot.org/uniprotkb/P0DMS8/entry
- Sequence:
  PNNSTALSLANVTYITMEIFIGLCAIVGNVLVICVVKLNPSLQTTTFYFIVSLALADIAVGVLVMPLAIVVSLGITIHFYSCLFMTCLLLIFTH
  ASIMSLLAIAVDRYLRVKLTVRYKRVTTHRRIWLALGLCWLVSFLVGLTPMFGWNMKLTSEYHRNVTFLSCQFVSVMRMDYMVYFSF
  LTWIFIPLVVMCAIYLDIFYIIRNKLSLNLSNSKETGAFYGREFKTAKSLFLVLFLFALSWLPLSIINCIIYFNGEVPQLVLYMGILLSHANS
  MMNPIVYAYKIKKFKETYLLILKACVVCHPSDSLDTSIEKNSE

2. Identification of possible templates for hAA3R

The goal is to model A3R3 in holo-active conformation.

Using the NCBI PSI-BLAST tool with the sequence corresponding to P0DMS8 (PDB ID), and the PDB database as the search target, an analysis was conducted to identify homologous sequences. During the second iteration of PSI-BLAST, the number of sequences retrieved was increased to 1000 to avoid constraints on the results. After sorting the results by percentage identity, it was observed that new entries had a percentage identity of 37.50% or lower. Based on this, further iterations were deemed unnecessary, and attention was focused on refining the results.

Filters were applied to the results of the second iteration, and it was found that there were still valid new entries in the dataset. This discrepancy occurred because the visible list of sequences was limited to 500 entries, despite the total number of sequences being 661. To ensure comprehensive analysis, the filters were reset, and all sequences were selected to run a third iteration of PSI-BLAST. However, after applying filters to the results of the third iteration, only seven sequences were retained, none of which were new. Additionally, some valid sequences from the previous iteration appeared to be missing.



Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
Chain R, Adenosine receptor A3 [Homo sapiens]	Homo sapiens	321	321	100%	6e-109	100.00%	318	8X16_R	✓	✓	
Chain R, Chimera protein of Muscarinic acetylcholine receptor M4 and Adenosine receptor A1 [H...	Homo sapiens	253	253	98%	2e-81	47.37%	387	7LD3_R	✓	✓	
Chain R, Chimera protein of Muscarinic acetylcholine receptor M4 and Adenosine receptor A1 [H...	Homo sapiens	253	253	98%	2e-81	47.37%	389	6D9H_R	✓	✓	
Chain A, Adenosine receptor A2a,Soluble cytochrome b562,Adenosine receptor A2a [Homo sapi...	Homo sapiens	213	312	95%	2e-65	41.89%	447	5JTB_A	✓		✓
Chain A, Adenosine receptor A2a,Soluble cytochrome b562 [Homo sapiens]	Homo sapiens	210	304	92%	3e-64	41.63%	423	8A2O_A	✓	✓	
Chain A, ADENOSINE RECEPTOR A2A [Homo sapiens]	Homo sapiens	270	270	92%	4e-89	41.04%	314	5G53_A	✓	✓	
Chain A, Adenosine receptor A2a,Soluble cytochrome b562 [Homo sapiens]	Homo sapiens	212	312	95%	5e-65	40.99%	431	8RLN_A	✓		✓
Chain A, Adenosine receptor A2a, soluble cytochrome b562 chimera,Soluble cytochrome b562,A...	Homo sapiens	211	306	94%	1e-64	40.85%	447	8DU3_A	✓		✓
Chain A, Adenosine receptor A2a [Homo sapiens]	Homo sapiens	275	275	93%	6e-91	40.65%	326	3VG9_A	✓	✓	
Chain A, Adenosine receptor A2a,Soluble cytochrome b562 [Homo sapiens]	Homo sapiens	214	310	95%	6e-66	40.54%	433	8CIC_A	✓	✓	
Chain A, Adenosine receptor A2a,Soluble cytochrome b562,Adenosine receptor A2a [Homo sapi...	Homo sapiens	214	310	95%	6e-66	40.54%	433	5NLX_A	✓		✓
Chain A, ADENOSINE RECEPTOR A2A [Homo sapiens]	Homo sapiens	271	271	93%	4e-89	40.00%	325	2YDO_A	✓	✓	

To address this, the analysis returned to the results from the second iteration. Filters were reapplied to study the 12 valid entries identified at this stage. Among these, two sequences were found to be chimeric and were subsequently excluded from further consideration. This approach allowed for a focused evaluation of the relevant homologous sequences while addressing anomalies in the dataset.

7 sequences selected | sequences newly added this iteration | GenPept | Graphics | Distance tree of results | Multiple alignment | MSA Viewer

No new sequences were found above the 0.005 threshold

Sequences with E-value BETTER than threshold

select all 7 sequences selected — **PSI-BLAST iteration 3**

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
Chain R, Adenosine receptor A3 [Homo sapiens]	Homo sapiens	225	225	100%	2e-71	100.00%	318	8X16_R	✓	✓	
Chain R, Chimera protein of Muscarinic acetylcholine receptor M4 and Adenosine receptor A1 [H...	Homo sapiens	203	203	99%	5e-62	47.22%	387	7LD3_R	✓	✓	
Chain R, Chimera protein of Muscarinic acetylcholine receptor M4 and Adenosine receptor A1 [H...	Homo sapiens	203	203	99%	5e-62	47.22%	389	6D9H_R	✓	✓	
Chain A, ADENOSINE RECEPTOR A2A [Homo sapiens]	Homo sapiens	218	218	91%	1e-68	41.72%	314	5G53_A	✓	✓	
Chain A, Adenosine receptor A2a [Homo sapiens]	Homo sapiens	224	224	93%	5e-71	40.91%	326	3VG9_A	✓	✓	
Chain A, ADENOSINE RECEPTOR A2A [Homo sapiens]	Homo sapiens	221	221	93%	5e-70	40.26%	325	2YDO_A	✓	✓	
Chain A, Adenosine receptor A2a,Soluble cytochrome b562 [Homo sapiens]	Homo sapiens	175	258	93%	5e-51	40.09%	423	8A2O_A	✓	✓	

From the results of the second iteration of PSI-BLAST, 12 valid entries were identified. After a detailed analysis, two of these sequences were ruled out due to being chimeric. The remaining ten entries were further investigated through a search of the PDB database, revealing that all belonged to the same family and class (adenosine receptors, class A), albeit with varying subclasses. This structural conservation within the transmembrane (TM) region, combined with functional diversity across subclasses, was considered advantageous for further analysis.

Upon closer examination of the structures:

- 8X16: This represents the A3AR receptor bound to CF101 (IB-MECA), a synthetic agonist. It captures the receptor in a holo-active state, directly representing A3AR in this conformation. Solved using Cryo-EM, this structure is highly suitable for further analysis.
- 3VG9: This structure represents A2AR bound to an allosteric inverse-agonist antibody (Fab2838) in an inactive state. Solved by X-ray crystallography, it is deemed not suitable.
- 5G53: A2AR is bound to a synthetic agonist and G-protein, representing a holo-active state. Solved by X-ray crystallography, it is suitable for the analysis.
- 2YDO: This structure features a thermostabilized A2AR bound to adenosine, an endogenous agonist. It represents an intermediate holo-active state. Although the thermostabilizing mutations may alter the natural conformation, the structure is moderately suitable for the study.
- 8CIC: This represents A1AR bound to a natural ligand (endogenous agonist adenosine) and Gi protein, with high biological relevance. It is part of the same receptor family (class A) but a different subclass (Type 1), allowing structural conservation while incorporating functional diversity. The holo-active state and conserved TM region make this structure highly suitable.
- 5NLX: A2AR is bound to a synthetic agonist (UK-432097) and a mini-Gs protein, representing a holo-active state. The G-protein mimic helps model the cytoplasmic active-state conformation. Although the synthetic agonist might induce subtle differences in the ligand-binding pocket, this structure is highly suitable.
- 5JTB: This structure shows A2AR bound to a synthetic agonist (NECA) in an intermediate active state without a G-protein. As it does not fully represent the holo-active state, it is moderately suitable.
- 8RLN: This structure captures A2AR bound to caffeine, a natural antagonist, in an inactive state. It is not suitable for the analysis.
- 8DU3: An engineered structure of A2AR (Star-bRil) bound to theophylline, an antagonist, representing an inactive conformation. It is not suitable.
- 8A2O: An engineered A2AR bound to theophylline and stabilized with StaR mutation and BRIL fusion, representing an inactive conformation. It is not suitable.

Based on this analysis, the selected structures for further study are 8X16_R, 2YDO_A, 5G53_A, 8CIC_A, 5NLX_A, and 5JTB_A. These structures were chosen for their representation of the holo-active state and their relevance in capturing functional and structural diversity within the adenosine receptor family.

After selecting six possible templates from PSI-BLAST, HHPred was used to further validate and refine the choices. The sequence corresponding to PDB ID P0DMS8 was used as input in HHPred, and the results were cross-referenced with the PSI-BLAST findings:

- 8X16: While 8X16 was not directly identified in HHPred, the tool provided a match to 8X17_R. Upon investigation, it was confirmed that 8X16 and 8X17 originate from the same study and authors, with HHPred grouping them together. Although HHPred labels 8X17_R as a structure from Rattus norvegicus, further examination in the PDB database revealed that the A3AR structural component comes from Homo sapiens. The structure is bound to the synthetic agonist CF102 (Namodenoson) and coupled with a heterotrimeric Gi protein, representing a holo-active state. Given these attributes, this structure remains highly suitable for further analysis.

- 2YDO: This structure was not identified in HHPred. As no supporting evidence was available for its inclusion, it was deemed not suitable.

- 5G53: Similarly, this structure was not found in HHPred and is therefore considered not suitable.

- 8CIC: This structure did not appear in HHPred, and without additional validation, it is also deemed not suitable.

- 5NLX: This structure was absent from the HHPred results and is thus not suitable.

- 5JTB: Like the others, this structure was not identified in HHPred and is not suitable.

From this analysis, 8X17_R was confirmed as a highly suitable template, validated both by HHPred and further investigation in the PDB. The remaining structures identified through PSI-BLAST were not supported by HHPred and are excluded from further consideration.

In our initial round of analysis, 8X16_R was identified as a suitable template through HHPred. To expand the search and identify at least three additional templates, I conducted a deeper analysis using the HHPred output data.

First, the .hhr file from the HHPred results was downloaded, and the data from all entries were programmatically processed. To refine the dataset, the following steps were taken:

- Entries containing the keyword 'chimer' in their names were filtered out to exclude chimeric structures, ensuring the biological relevance and integrity of the selected templates.

- The query coverage percentage was calculated for each entry using the formula:

- $Query\ Coverage\ (\%) = (Length\ of\ Aligned\ Query\ Region / Length\ of\ Query) \times 100$

- The results were ordered based on identity, similarity, and query coverage percentages for systematic evaluation.

- A filter was applied to exclude entries that did not meet the following criteria:

    - Identity ≥ 36%
    - Similarity ≥ 55%
    - Query coverage ≥ 86%

These thresholds were chosen to ensure the structural and functional relevance of the templates, maintaining a balance between sequence identity and alignment quality. This methodical filtering process significantly reduced the pool of potential candidates while retaining entries most likely to be suitable for further analysis.

[PAGE BREAK]

PDB ID	State	Ligand Type	Synthetic Ligand	Method	Resolution	Suitability	Effector	Protein Name
8X17	Active	Agonist	CF102 (Namodenoson)	Cryo-EM	3.19 Å	Highly Suitable	Gi protein	A3 adenosine receptor (A3AR)
7LD3	Active	Agonist	Adenosine (endogenous)	Cryo-EM	3.20 Å	Suitable	Gi2 protein	A1 adenosine receptor (A1AR)
5UEN	Inactive	Antagonist	DU172 (synthetic)	X-ray	3.20 Å	Not Suitable	None	A1 adenosine receptor (A1AR)
5N2S	Inactive	Antagonist	PSB36 (synthetic)	X-ray	3.30 Å	Not Suitable	None	A1 adenosine receptor (A1AR)
7T32	Inactive	Antagonist	ZM241385 (synthetic)	Cryo-EM	3.40 Å	Not Suitable	None	A2A adenosine receptor (A2AR)
8JWY	Active	Agonist	2-118 (synthetic)	X-ray	2.33 Å	Suitable	None	A2A adenosine receptor (A2AR)
5NM4	Active	Agonist	NECA (synthetic)	X-ray	1.70 Å	Suitable	None	A2A adenosine receptor (A2AR)
6GDG	Active	Agonist	NECA (synthetic)	Cryo-EM	4.1 Å	Suitable	mini-Gs protein	A2A adenosine receptor (A2AR)
8HDO	Active	Agonist	BAY 60-6583 (synthetic)	Cryo-EM	2.87 Å	Suitable	mini-Gs protein	A2B adenosine receptor (A2BR)
8GNG	Inactive	Antagonist	Istradefylline (synthetic)	X-ray	3.20 Å	Not Suitable	None	A2A adenosine receptor (A2AR)
7XY6	Active	Agonist	NECA (synthetic)	Cryo-EM	2.99 Å	Suitable	Gs protein	A2B adenosine receptor (A2BR)
4UG2	Active	Agonist	CGS21680 (synthetic)	X-ray	2.60 Å	Suitable	None	A2A adenosine receptor (A2AR)

The analysis of the 12 possible entries from the later procedure, excluding the first entry already considered (8X17), was conducted systematically. The gathered data was summarized in the table above, and the initial observations were further validated using the PDB database.

At first glance, it was observed that 8GNG, 5UEN, 5N2S, and 7T32 are in the inactive state and were therefore excluded as they do not meet the criterion of representing the holo-active state. Further inspection revealed that 7LD3 is a chimera of the muscarinic acetylcholine receptor M4 and A1AR, and 4UG2 represents a thermostabilized human A2A receptor with engineered mutations that could alter its natural conformation. Both were also excluded.

The remaining entries were evaluated based on extracted metrics from HHPred, including similarity, identity, query coverage, and the number of columns covered. The selection process prioritized these metrics in the following order: similarity, identity, suitability, and inverse resolution (favouring lower resolution for higher structural detail). This filtering process, summarized in the next table, ensured that the final entries selected for further study aligned with the criteria for structural and functional relevance, as well as alignment quality. This approach strikes a balance between biological accuracy and computational rigor, optimizing the selection of templates for subsequent modelling efforts.

The refined list of templates will now be analysed to select at least three additional candidates that meet the structural and functional requirements for our study.

To enhance our selection of templates beyond simply picking the top entries, I considered the critical factor of overall query coverage provided by the combination of templates. Ensuring robust and comprehensive coverage of the query sequence requires evaluating how well different templates collectively align across the entire sequence.

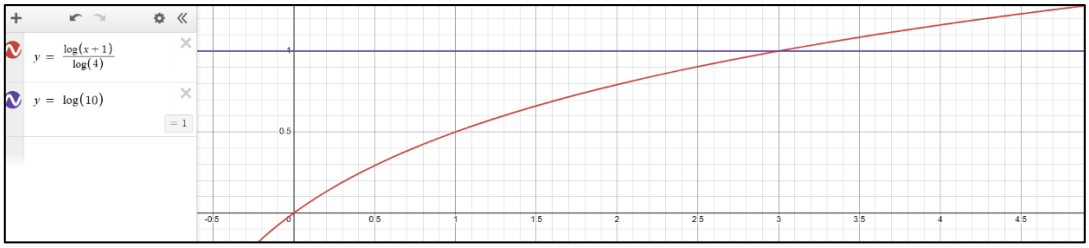
To achieve this:

- Aligned Column Frequency Analysis: The frequency of alignment for each query column across all potential templates was calculated. This provided insight into which regions of the query sequence were consistently aligned and which were underrepresented.

- Combination Analysis: All combinations of three to five potential templates (including 8X17, which was already selected) were generated to evaluate their collective coverage of the query sequence. This step ensured we explored combinations that might provide complementary coverage.

- Frequency Plotting: The alignment column frequencies for each combination were plotted to visually assess the distribution and identify gaps or inconsistencies in coverage.

- Log-Weighted Frequency Score: An average log-weighted frequency score was computed for each combination. This score penalized sparse or inconsistent coverage and rewarded combinations that provided comprehensive and evenly distributed coverage of the query. The log-weighted scoring approach emphasized both the presence and density of alignment across the query sequence, allowing for more nuanced evaluation of the templates.



By combining these analyses, I identified the combinations that provided the most effective coverage while ensuring alignment quality. The final selection of templates prioritized these combinations to achieve optimal representation of the query sequence's structural and functional regions.

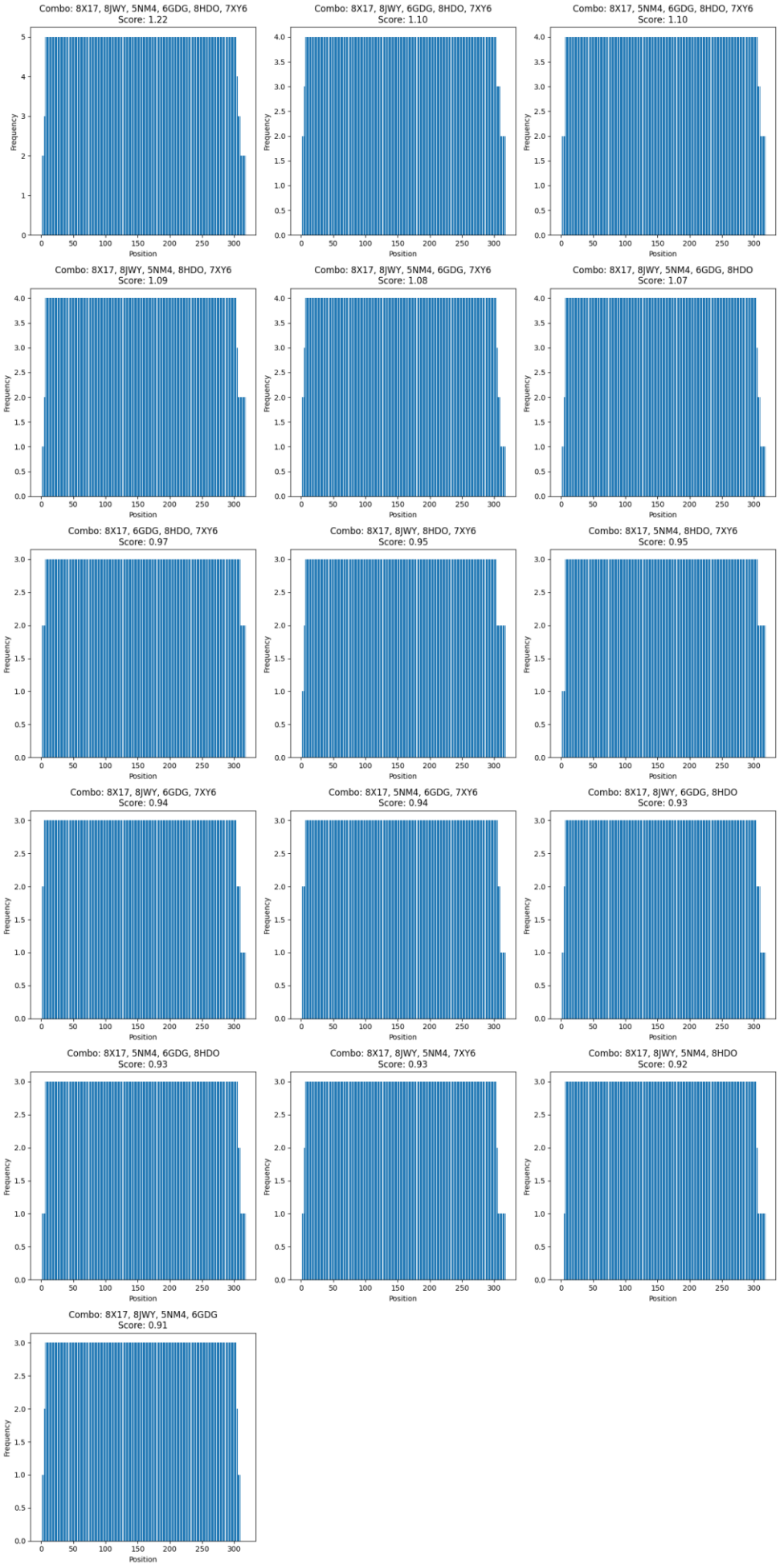
PDB ID	Resolution	Suitability	Effector	Similarity	Identities	Query Coverage	Query HMM
8X17	3.19 Å	Highly Suitable	Gi protein	1.551	100	96.226	[1, 306]
8JWY	2.33 Å	Suitable	None	0.782	42	94.025	[6, 304]
5NM4	1.70 Å	Suitable	None	0.762	41	94.025	[8, 306]
6GDG	4.1 Å	Suitable	mini-Gs protein	0.752	42	96.855	[3, 310]
8HDO	2.87 Å	Suitable	mini-Gs protein	0.734	39	97.484	[8, 318]
7XY6	2.99 Å	Suitable	Gs protein	0.716	38	99.057	[3, 318]

I could directly pick the first three or more entries, but another important aspect to consider is the overall coverage provided by the combination of these templates. To investigate this, I determined the frequency of the aligned columns of the query across all the templates.

To analyse this, I computed all possible combinations of three to five potential templates, including 8X17, which I had already selected. I plotted the alignment column frequencies for these combinations to visualize the distribution of coverage and identify regions that might be underrepresented. Additionally, I calculated an average log-weighted frequency score for each combination. This score penalized sparse or inconsistent coverage while delineating constraints in the alignment coverage, ensuring that the selected combination provided robust and evenly distributed alignment across the query sequence.

This approach allowed me to evaluate the suitability of template combinations not just on individual merits but on how well they collectively covered the query sequence, ensuring comprehensive and high-quality structural modelling.

[PAGE BREAK]


Combo: 8X17, 8JWY, 5NM4, 6GDG, 8HDO, 7XY6
Score: 1.22
Combo: 8X17, 8JWY, 6GDG, 8HDO, 7XY6
Score: 1.10
Combo: 8X17, 5NM4, 6GDG, 8HDO, 7XY6
Score: 1.10
Combo: 8X17, 8JWY, 5NM4, 8HDO, 7XY6
Score: 1.09
Combo: 8X17, 8JWY, 5NM4, 6GDG, 7XY6
Score: 1.08
Combo: 8X17, 8JWY, 5NM4, 6GDG, 8HDO
Score: 1.07
Combo: 8X17, 6GDG, 8HDO, 7XY6
Score: 0.97
Combo: 8X17, 8JWY, 8HDO, 7XY6
Score: 0.95
Combo: 8X17, 5NM4, 8HDO, 7XY6
Score: 0.95
Combo: 8X17, 8JWY, 6GDG, 7XY6
Score: 0.94
Combo: 8X17, 5NM4, 6GDG, 7XY6
Score: 0.94
Combo: 8X17, 8JWY, 6GDG, 8HDO
Score: 0.93
Combo: 8X17, 5NM4, 6GDG, 8HDO
Score: 0.93
Combo: 8X17, 8JWY, 5NM4, 7XY6
Score: 0.93
Combo: 8X17, 8JWY, 5NM4, 8HDO
Score: 0.92
Combo: 8X17, 8JWY, 5NM4, 6GDG
Score: 0.91
Frequency
Position

After cross-referencing the results with the data available in the latter image, an evidence-based decision was made regarding the selection of templates. Among the various combinations analysed, the set comprising ['8X17', '6GDG', '8HDO', '7XY6'] emerged as a strong candidate, ranked in position 7.

This combination stood out for several reasons:

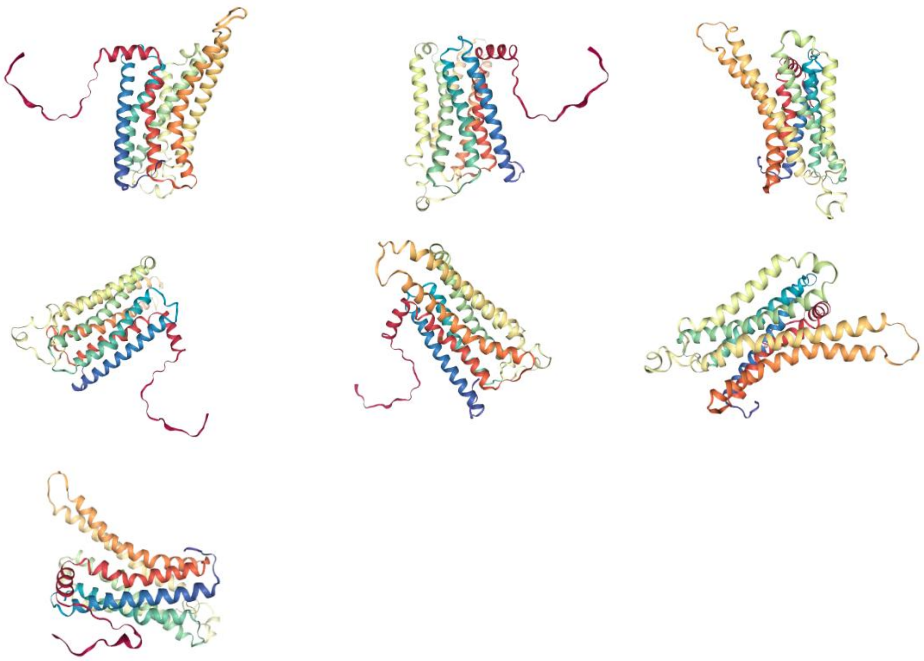
- Holo-Active State Representation: All proteins in this combination are in the holo-active state and bound to a G-complex, which aligns well with the structural and functional requirements of the study.
- Coverage: The overall coverage provided by this combination is robust, adequately covering the query sequence except for the terminal regions, a common limitation in structural modelling due to inherent flexibility and lack of structural constraints in these regions.
- Evidence-Based Parameters: The combination balances critical parameters, including query coverage, alignment quality, and biological relevance, making it an optimal choice for template selection.

To support the final decision, the following table provides a comprehensive overview of all relevant parameters, enabling transparency and justification for the chosen combination. This data-driven approach ensures that the selected templates offer the best compromise between structural accuracy, functional relevance, and query coverage, laying a strong foundation for subsequent modelling and analysis.

Protein name	PBD ID	Identity	Similarity	Resolution	Query cover	Ligand name	Ligand type	Effector
A3 adenosine receptor (A3AR)	8X17	100	155.1	3.19	96.23	CF102 (Namodenoson)	Agonist	Gi protein
A2A adenosine receptor (A2AR)	6GDG	42	75.2	4.1	96.86	NECA (synthetic)	Agonist	mini-Gs protein
A2B adenosine receptor (A2BR)	8HDO	39	73.4	2.87	97.48	BAY 60-6583 (synthetic)	Agonist	mini-Gs protein
A2B adenosine receptor (A2BR)	7XY6	38	71.6	2.99	99.06	NECA (synthetic)	Agonist	Gs protein

3. *Assembly of the homology model of hAA3R and comprehensive validation*

Having selected these templates, I proceeded to use them with MODELLER to generate a structural model of the target protein.

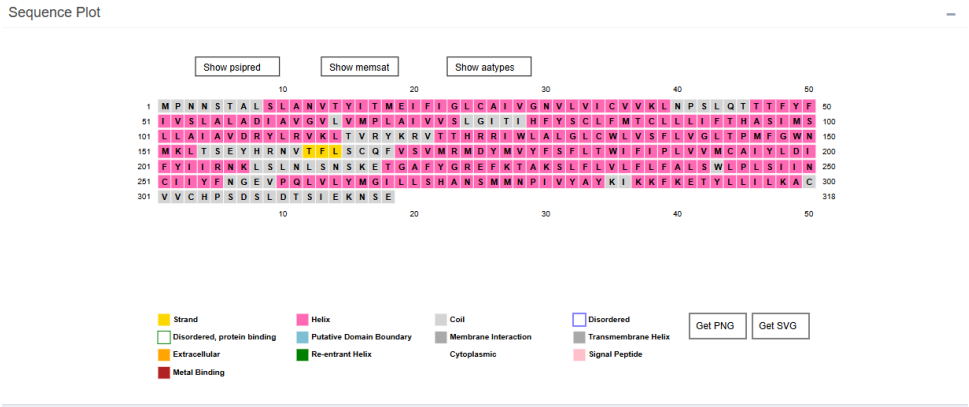


The resulting models revealed several key observations. The transmembrane helices (TM) were well-defined, with no significant distortions or acute bending observed, indicating robust structural consistency across the modelled regions. However, TM7

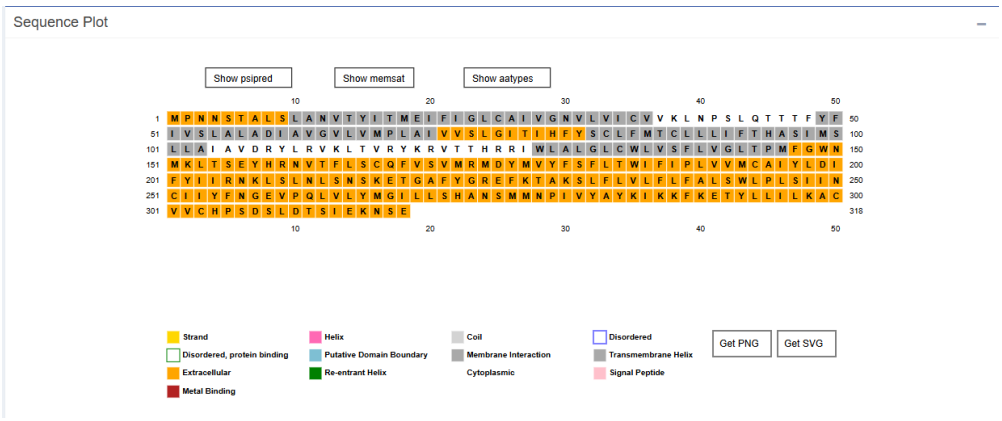
displayed an accentuated skew, causing the C-terminal to extend into the lipid bilayer. This feature is characteristic of GPCRs, including A3AR, and was consistently observed across different template combinations, reinforcing its expected biological relevance.

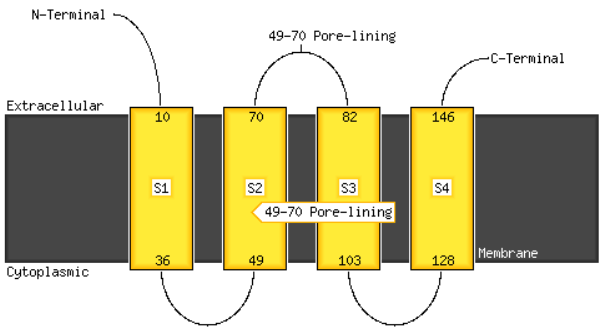
Additionally, the extracellular loop (ECL) 2 showed some distortions. This could be attributed to the nature of the ligand binding present in all the templates used for modelling. Despite this, the overall structure was cohesive, with no gaps or holes, and the termini were not excessively extended. These results suggest a reliable and biologically relevant model that captures the essential structural features of the target protein while reflecting the inherent variability of the templates used.

To perform an a-priori analysis of the secondary structure, I utilized the PsiPred tool, selecting PSIPRED 4.0 and MEMSAT-SVM as the algorithms. The results from PSIPRED were cross-referenced with the atom positions in the 3D model generated by MODELLER. The analysis showed that the predicted positions did not deviate significantly from the modelled positions, with an average difference of 1.64. Some regions predicted as coils by PSIPRED were identified as helices in the 3D model, likely representing small, transient helical segments. Taking this into account, a total of seven helices were predicted, though their alignment indexes showed slight skewing. These findings were further validated using subcellular feature annotations from UniProt, which confirmed the consistency of the predicted helices.

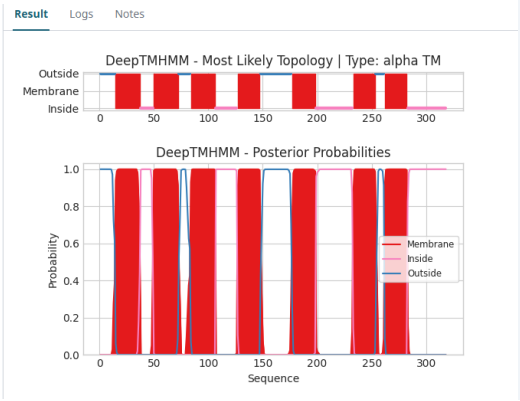


In contrast, MEMSAT, being a transmembrane-specific tool, failed to accurately predict the secondary structure of the model. It identified only four helices in total, significantly underestimating the structural complexity. This demonstrates that MEMSAT is not a reliable a-priori validation tool for this specific case.





To enhance the prediction, I used DeepTMHMM, which offers greater accuracy for secondary structure prediction. However, the results from DeepTMHMM showed an average deviation of 5.07 from the modelled protein, a larger discrepancy compared to PSIPRED. This suggests that while DeepTMHMM is effective for general predictions, its results may require careful integration with structural data for precise modelling.




```
##gff-version 3
# Unnamed Length: 318
# Unnamed Number of predicted TMRs: 7
Unnamed outside 1   14
Unnamed TMhelix 15  37
Unnamed inside  38  49
Unnamed TMhelix 50  72
Unnamed outside 73  83
Unnamed TMhelix 84  106
Unnamed inside  107 126
Unnamed TMhelix 127 147
Unnamed outside 148 176
Unnamed TMhelix 177 198
Unnamed inside  199 232
Unnamed TMhelix 233 253
Unnamed outside 254 261
Unnamed TMhelix 262 282
Unnamed inside  283 318
```

The summarized data is presented in the following table, where absolute differences (diff) are mapped to a color-coded scale for clarity:

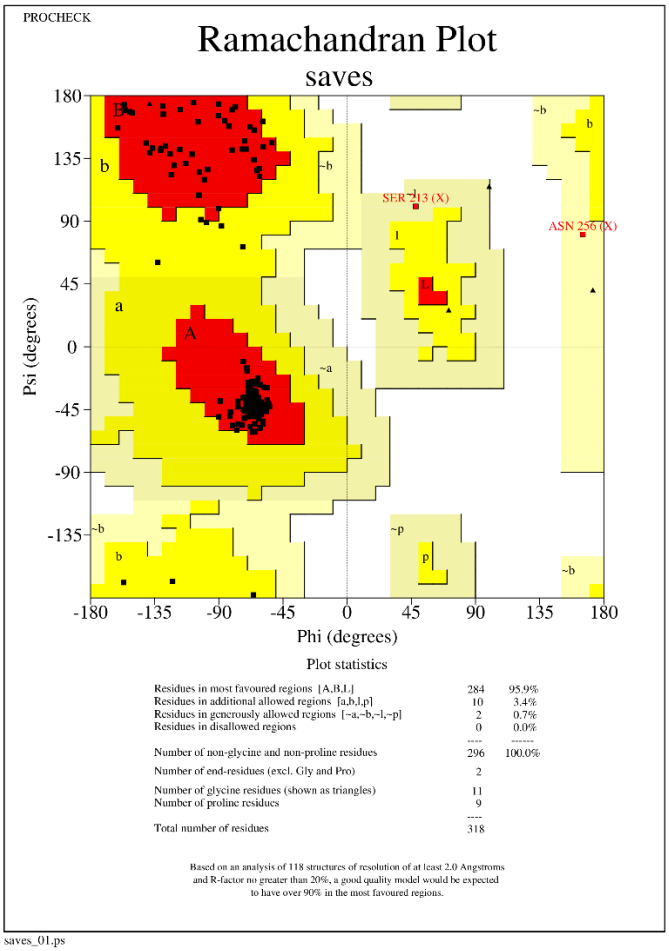
- Green: diff ≤ 2
- Yellow: 3 ≤ diff ≤ 4
- Red: diff > 4

Helix	My model	DeepTMHMM	PSIPRED	MEMSATSVM	UNIPROT
H1	9	15	9	10	15
	39	37	39	36	37
H2	47	50	46	49	49
	73	72	74	70	72
H3	81	84	80	82	85
	112	106	115	103	106
H4	125	127	123	128	127
	149	147	153	146	148
H5	174	177	170		178
	210	198	207		198
H6	221	233	219		232
	255	253	255		255
H7	261	262	261		262
	298	282	300		284
Average absolute difference		**5.07**	**1.64**	**3.13**	**4.71**

legend:
diff<2
2<diff<5
diff>4

[PAGE BREAK]

To make an a-posteriori analysis we make use of ProCheck Saves v6.1 by uploading the PDB file we got from our model during MODELLER.

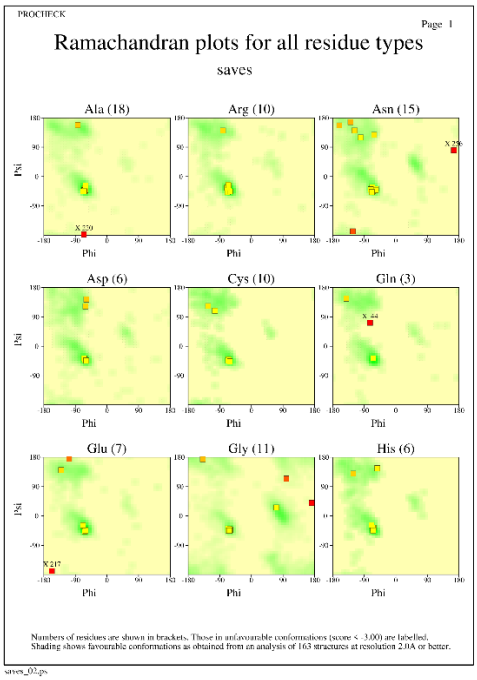


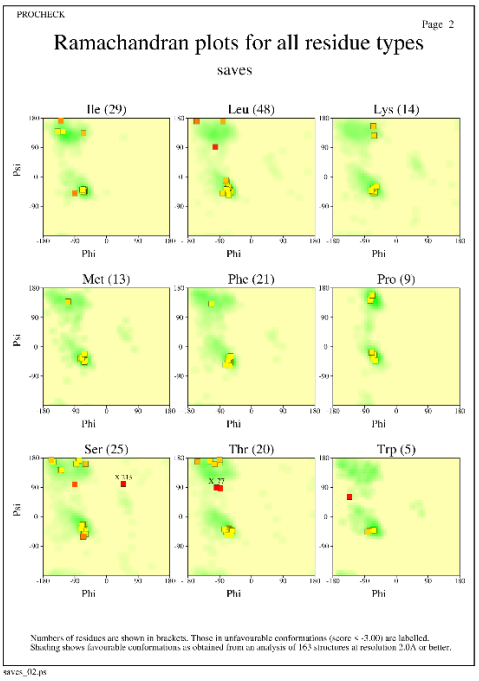
saves_01.ps

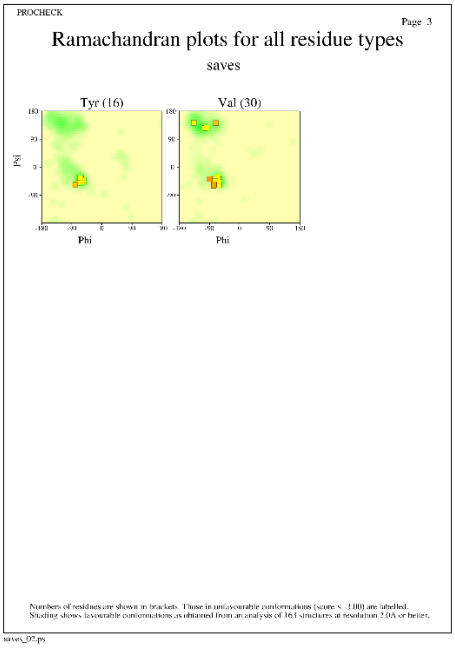
The Ramachandran plot provides a summary of the backbone torsion angles (ϕ and ψ) of residues in the protein structure and is a key indicator of structural quality. In this model, 95.9% of residues lie in the most favored regions (A, B, L), which means most residues adopt energetically favorable backbone conformations. This percentage exceeds the 90% threshold typically used to define a high-quality model. Additionally, 3.4% of residues are in the additional allowed regions (a, b, l, p), and 0.7% are in generously allowed regions (~a, ~b, ~l, ~p).

These regions correspond to less common but still permissible backbone conformations, often associated with flexibility or local structural constraints. Importantly, no residues are found in disallowed regions, which indicates that the structure is free from steric clashes or highly unfavorable conformations. The presence of 11 glycine residues, which are marked as triangles, is consistent with their well-known flexibility and ability to adopt unusual conformations. Similarly, the 9 proline residues, marked as squares, cluster in specific regions due to the restricted torsion angles imposed by the cyclic side chain. With a total of 318 residues analyzed, the structure achieves an excellent overall quality, with deviations likely reflecting functional requirements or localized flexibility.

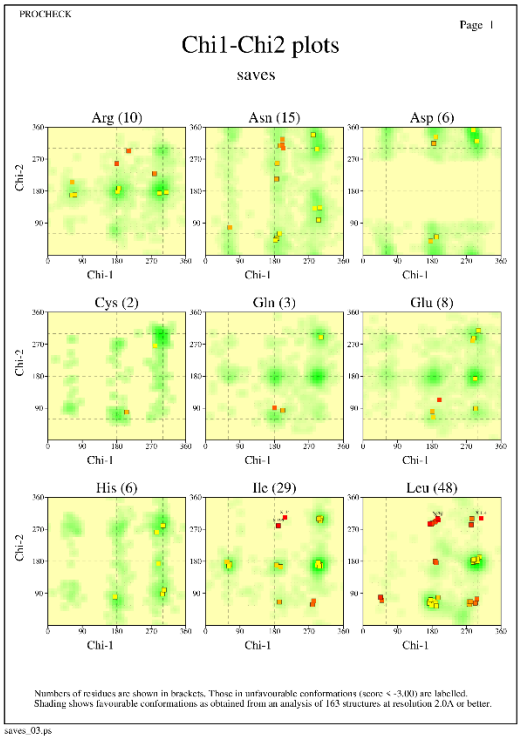
Residues 213 (SER) and 256 (ASN) are highlighted as red squares on the plot, indicating that their torsion angles place them in additional allowed or generously allowed regions. For residue 213 (SER), the conformation is influenced by the structural context, as serine is a small and flexible residue often found in loops or turns where non-ideal torsion angles are more common. For residue 256 (ASN), the amide side chain may be forming hydrogen bonds or interacting with neighboring residues, pulling the backbone into a less favored conformation. These deviations from ideal geometry reflect localized flexibility or specific functional roles, such as stabilizing loops or binding pockets. Moreover, both residues are in allowed regions and are not disallowed, so their conformations are permissible and likely justified by the structural environment.

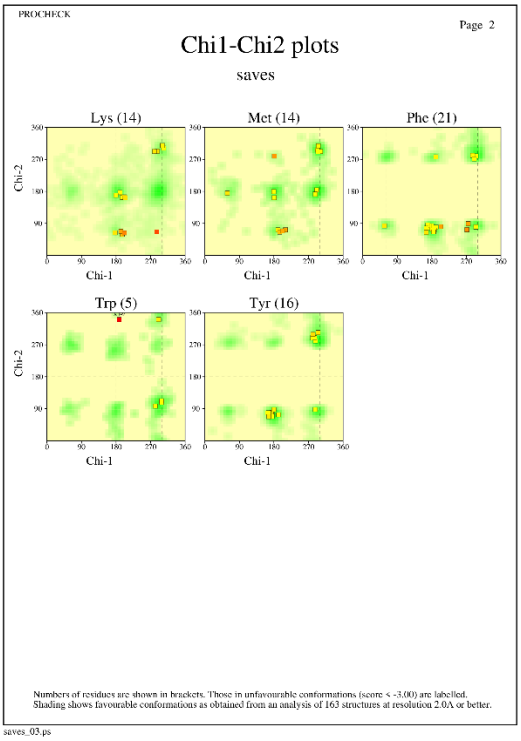






These plots analyze the backbone torsion angles (ϕ and ψ) for each amino acid type and highlight residues with conformations that deviate from expected regions. The plots include shading for favored conformations (green/yellow) and highlight outliers (red squares), which represent residues in unfavorable conformations. In the shown images, specific residues are marked with red squares. These include Ser 213, which appears in an unfavorable conformation located in the additional or generously allowed region. Serine's small and flexible nature means it is often found in loops or turns, which may explain this deviation. Asn 256 is also located in an unfavorable region, likely due to its amide side chain causing steric or conformational strain. Thr 77 is highlighted as an outlier, and threonine's side chain hydroxyl group may cause steric hindrance or hydrogen bonding, leading to deviations. Gln 44 is another outlier in an unfavorable region, where glutamine's bulky side chain may be influencing its backbone torsion angles. Ala 220, despite alanine being highly flexible, appears as an outlier, possibly due to constraints from its structural environment. Finally, Gly 217 is marked as an outlier, which is unusual because glycine, due to the absence of a side chain, often adopts unusual backbone angles. However, this residue may have extreme flexibility or local structural strain causing this deviation. These outliers (red squares) indicate that the residues adopt unusual torsion angles, but they may still be biologically or structurally relevant if located in loops, active sites, or other regions requiring flexibility.





Chi1-Chi2 plots analyse the side chain torsion angles (χ1 and χ2) of amino acids in the protein structure. These torsion angles describe the rotational conformations of the side chains around their bonds. The plots show favoured conformations as green and yellow regions based on statistical data from high-resolution protein structures, while residues in unfavourable conformations are marked with red squares.

In the shown plots, specific residues fall into unfavourable regions. For example, in the Asn (15) plot, several residues have side chain torsion angles that deviate significantly from favoured regions, as indicated by the red squares. This is likely due to steric clashes, hydrogen bonding, or constraints imposed by the local structural environment. Similarly, in the Ile (29) plot, a red square highlights an outlier with torsion angles in a highly strained position, which might result from packing constraints or interactions with nearby residues. In the Leu (48) plot, multiple outliers can be seen, likely due to leucine's branched side chain experiencing steric interference in crowded environments. The same applies to Cys (2) and Glu (8), where red squares indicate strained side chain conformations, possibly due to disulfide bond formation (for cysteine) or charge interactions (for glutamate).

4. Final summary

The homology modeling process for the Human Adenosine Receptor A3 (hAA3R) was executed methodically, leveraging sequence and structural data to produce a high-quality model with robust biological relevance. The process began with the identification of suitable templates through PSI-BLAST and HHPred, focusing on structures in the holo-active conformation. Key templates were selected based on alignment metrics, functional relevance, and structural consistency.

The final model, generated using MODELLER, demonstrates well-defined transmembrane helices (TMs) without significant distortions. The characteristic skew of TM7, leading to the C-terminal's interaction with the lipid bilayer, aligns with known GPCR features, reinforcing the model's biological accuracy. Minor distortions in extracellular loop 2 are attributed to ligand interactions within the templates but do not detract from the overall structural cohesion.

Validation of the model was rigorous. A-priori analyses using PSIPRED and DeepTMHMM confirmed the alignment of predicted and modeled secondary structures, with an average deviation of 1.64 Å for PSIPRED. A-posteriori evaluation with ProCheck revealed 95.9% of residues in favored regions of the Ramachandran plot, exceeding the threshold for high-quality models. Importantly, no residues were found in disallowed regions, underscoring the model's structural reliability.

While a few outliers were identified in backbone torsion and side chain angles, their conformations were justifiable by local structural constraints or biological roles, such as loop flexibility and active site stabilization. These deviations, although minor, reflect the inherent dynamics and functional specificity of the receptor.

Overall, the modeling process yielded a reliable homology model that faithfully captures the structural and functional nuances of hAA3R. This model provides a strong foundation for downstream studies, including ligand-binding analyses and mutational research.