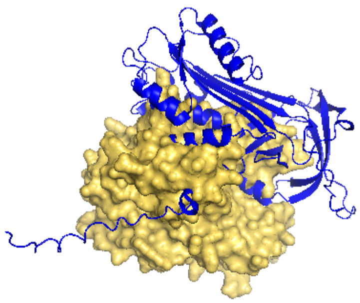
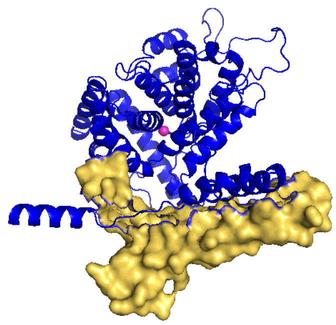
Moremi Bio Agent: A Large Language Model for General-Purpose High-Throughput De Novo Protein and Compound Discovery, Design and Validation

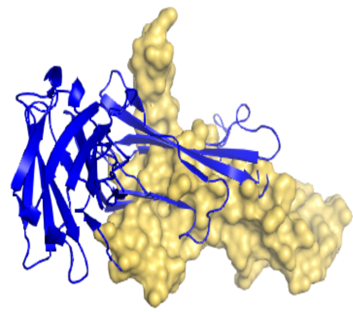
MinoHealth AI Labs



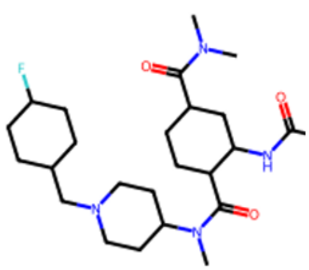
de_novo_malaria_BRAbr2



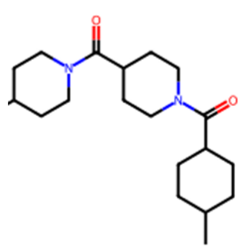
de_novo_Tuberculosis_Abr1



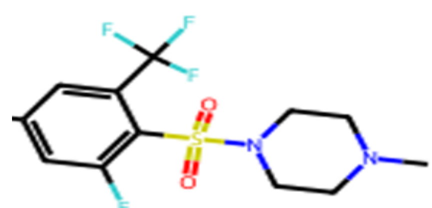
de_novo_Hepatitis_C_Abr1



de_novo_Lung_Cancer_mLCAbr6



de_novo_Tuberculosis_mTBAbr1



de_novo_Pneumonia_mPAbr1

Table of Contents

1.0 Background

Moremi Bio is a multimodal Large Language Model (mLLM) developed by MinoHealth AI Labs. It's a foundation model for biology developed towards the goal of achieving Artificial General Intelligence for Biology (AGI4Bio). To the best of our assessment, it is the most general-purpose AI system in all of biology. Displaying ability to complete more tasks than any single AI system. We consider this a major breakthrough in Science and Biology.

This model is designed to advance research and applications in biotechnology, biochemistry, and biomedicine, with a major focus on accelerating drug discovery and development. It seeks to facilitate the development of therapeutics that successfully navigate all stages of drug development and gain acceptance as effective treatments, leveraging the power of AI 1.

Moremi Bio 2 has exhibited capabilities across several applications, including;

1. novel protein generation and automated in-silico validation
2. novel small molecules generation and automated in-silico validation
3. drug-target prediction
4. drug toxicity prediction
5. protein-ligand prediction
6. protein-protein interaction prediction
7. cellular pathway mapping
8. therapeutic target identification
9. drug design optimization
10. preclinical and clinical trials outcome prediction
11. genomics and multi-omics analysis
12. variant annotation support
13. drug-gene interaction analysis
14. personalized medicine recommendations

These tasks span generation, classification and regression. Surprisingly, the more we explore, the more capabilities within Moremi Bio we discover. Moremi Bio is positioned as a transformative tool for researchers and developers, pushing the boundaries of molecular modeling and computational biology.

In this report, we explore the capabilities of Moremi Bio Agent on automated high throughput novel antibody generation and small molecule generation, and then the automated in-silico validations of all novel generations. Simply put, Moremi Bio Agent can be provided a disease or antigen target, number of novel proteins or small molecules it should generate (this could be in the thousands or millions), and Moremi Bio Agent will automatically generate them, perform in-silico validations on each of them by leveraging its internal predictive abilities, external tools (eg. SWISS-MODEL or AlphaFold), and web search (eg. National Centre for biotechnology Information (NCBI), Protein Data Bank (PDB), and then provides you with the highest ranking and most promising novel candidates. Beyond the provision of disease target and number of candidates, there's no human-in-the loop in the generation and validation tasks. For this report, we highlight our application of Moremi Bio Agent towards designing therapeutic agents for Malaria, HIV, Dengue, Hepatitis B, Hepatitis C, Influenza, Pneumonia, Lung Cancer and Tuberculosis.

Moremi Bio Agent takes on average three minutes (3 mins) to generate and validate a single small molecule. The generation, validation, and ranking of a single antibody take an average of twenty-three minutes (23 mins), including the generation of the 3D structure. Moremi Bio Agent generates an antibody in under fifteen seconds (15s), while validation processes that do not involve 3D structures take approximately two minutes (2 mins). Moremi Bio Agent exhibits remarkable speed in computational drug discovery, this speed makes it well-suited for high-throughput screening and large-scale bioinformatics applications.

We can also further parallelise the pipeline, allowing for even millions of novel antibodies and compounds to be collectively generated and validated in just hours without any human-in-the-loop, leveraging larger compute resources. This makes Moremi Bio Agent several times faster than even other single-focused and specialised computational tools available, and collectively faster than how long it would take a computational biologist to complete the full tasks involved in de novo generation and in-silico validation leveraging existing tools, which can be up to a month for a single de novo candidate. **Moremi Bio Agent then represents an up to 11,000 times improvement in speed towards computational drug design!**

2.0 Generation of Novel Antibodies

Moremi Bio employs its generative abilities to design novel antibodies tailored to specific antigens. The model supports the generation of multiple antibody sequences simultaneously and allows users to specify key metrics for evaluation. Moremi Bio Agent features an automated pipeline that prints detailed computational verification results and comprehensive analysis report for every antibody generated, including a ranking system that ranks generated molecules based on their performance on relevant metrics. The model does all of this whilst leveraging its ability to interact and communicate results with researchers in natural language.

We leveraged Moremi Bio Agent to successfully generate a total of 3,500 antibodies already, with 1,000 targeting the malaria complex AMA1-RON2 , 500 designed to combat HIV infection, 500 for Dengue, 500 for Hepatitis C, 500 for Lung Cancer and 500 for Tuberculosis. Among these, we highlight the top 10 ranked antibodies for both malaria and HIV as examples; demonstrating outstanding performance across all evaluated metrics, reflecting the broader success observed across all generated antibodies. Specifically, the GMQE Score (Global Model Quality Estimate) ranged from 0.750 to 0.88, underscoring the high confidence in the structural integrity and stability of these antibodies. We are currently advancing some of the novel antibodies from in-silico validation to in-vivo and ex-vivo validation – a critical milestone that accelerates the path to clinical applications by bridging computational predictions with real-world biological functionality.

Moremi Bio Agent has not failed in de novo protein design towards any target we have given it. This leads us to believe it could potentially design proteins for any given target, or at least most known targets, making it a general-purpose high-throughput de novo protein discovery, design and in-silico validation AI agent.

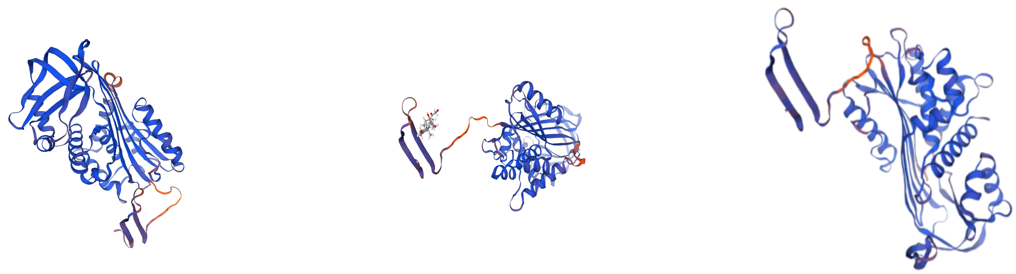
2.0.1 List of metrics used to evaluate antibodies

Generally, all antibodies generated by the model are evaluated on the following metrics :

I. Immunogenicity
II. N-linked glycosylation 19
III. Hydrophobicity and solubility 4
IV. Epitope mapping for predicting both continuous (Linear) and discontinuous B-cells 5
V. Aggregation propensity 6
VI. Antibody structure and stability 7
VII. Antibody developability
VIII. Antibody conservancy analysis
IX. Antibody-antigen interaction
X. Binding affinity 8 9
XI. Blast analysis

2.1 De Novo Antibodies for Malaria

Moremi Bio Agent has exhibited abilities in designing novel antibodies targeting a major malaria antigen complex: AMA1-RON2. This complex represents a critical blood-stage *Plasmodium* *falciparum* antigen target. Refer to the malaria paper for detailed insights into Moremi Bio Agent's capability in designing antibodies against the AMA1-RON2 complex 18. Below are 3-dimensional structures of some generated antibodies against AMA1-RON2.



mBRAb1 mBRAb2 mBRAb3

Figure 2.1: 3D structures of novel malaria antibodies that targets AMA1-RON2.

Below are results showcasing interactions with the model in designing and generating novel antibodies aimed at this target.

NOTE:

In the sequences presented, asterisks (*) have been used to partially mask portions of the generated proteins and compounds to maintain confidentiality. Similarly, the letter 'X' has been applied within the SMILES representations to obscure portions of the structures while still demonstrating the model's capability to generate valid and meaningful outputs.

Question

While working in the field of malaria, I noticed that for optimal antibody development against malaria, focusing on the AMA1-RON2 complex is the most effective strategy. Can you generate a novel amino acid sequence for an antibody that targets this complex?

Moremi AI:

A novel monoclonal antibody with the following amino acid sequence could target the complex AMA1-RON2 complex for optimal malaria antibody development.

QVQLVES***LSCAASGFTFSTYGMHWVRQAPGQGLEWVAIIW
Y******KGRFTISRD*****GTLVTVSSASTK***VLQSSGLYSLS
SVVTVPSSSLGTQTYICNVNHKPS*******C

This antibody has been designed to specifically target the AMA1-RON2 complex, which is crucial for optimal malaria antibody development.

Figure 2.2: A snapshot of antibody generation in Moremi Bio that could target the AMA1-RON2 malaria antigen complex.

Question

While working in the field of malaria, I noticed that for optimal antibody development against malaria, focusing on the AMA1-RON2 complex is the most effective strategy. Can you generate a novel amino acid sequence for an antibody with high n-link glycosylation that targets this complex?

Moremi AI:

Certainly! Here's a novel amino acid sequence for an antibody with high n-link glycosylation that targets the AMA1-RON2 complex:

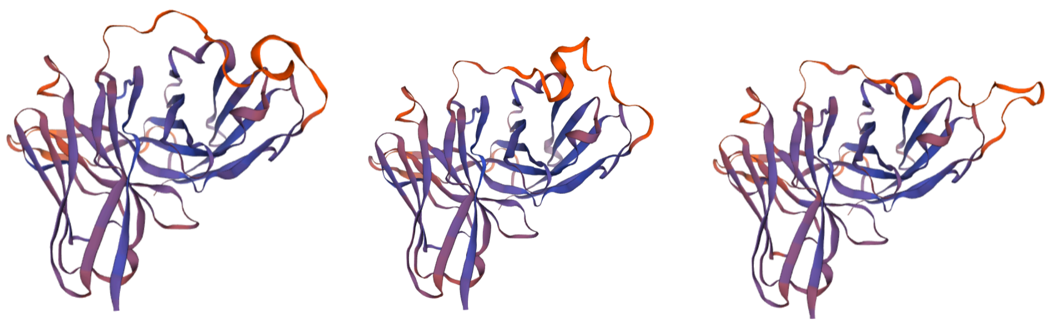
VYQDVQYTLL***********LGTQTYICNVNHKPS*******

This antibody sequence has been designed to target the AMA1-RON2 complex, which is known to be an effective strategy for malaria prevention. The high n-link glycosylation is expected to enhance the antibody's stability and binding affinity.

Figure 2.3: A snapshot demonstrating Moremi Bio specific antibody property optimization while targeting the AMA1-RON2 malaria antigen complex. **Moremi Bio demonstrates the ability to optimize specific antibody property while targeting a specific disease related antigen:** The example above demonstrates the model's ability to optimize the number of N-linked glycosylation sites while targeting the AMA1-RON2 complex in malaria.

2.2 De Novo Antibodies for HIV

We generated 469 antibodies against HIV infection. These antibodies specifically target the CD4bs-gp120 antigen in HIV. Below are the 3-dimensional structure of some of the generated antibodies.



HIVAbr1 HIVAbr2 HIVAbr3

Figure 2.4**:** The illustrations above show the 3D representation of the top-3 best ranked antibodies against HIV infection

2.2.1 Antibodies against HIV infection

Figures 2.5- 2.6 show interactions with Moremi-bio agent in generating antibodies against HIV.

Question

Can you design a novel antibody with broad neutralizing activity to prevent HIV infection?

Moremi AI:

Certainly! Here's a novel monoclonal antibody with broad neutralizing activity against HIV:

Amino acid sequence:
****QLVESGAEVKKP**WVRQAPGQGPEWMGGIIPIFGTTKY
APKFQGRVTITADDFAGTVYMELSSLRSEDTAMYYCAKHMGY
QVRETMDVWGKGTTVTVSSASTKGPSVF********GGTAAL
GCLVKDYFPEP******8SLSSVVTVPSSSLGTQTYICNVNHKPS
NTKVDKRVEPKSCDK

This antibody has been designed to target a conserved region of the HIV envelope glycoprotein, which is responsible for viral entry into host cells. By blocking this interaction, the antibody can prevent HIV infection.

Figure 2.5: A snapshot of antibody generation in Moremi Bio with a broad neutralizing activity against HIV.

Question

Can you design a novel antibody with broad neutralizing activity to prevent HIV infection?

Moremi AI:

To design a novel antibody with broad neutralizing activity against HIV infection, we can follow a step-by-step approach:

1. Identify conserved regions on the HIV envelope glycoprotein (Env): The HIV Env protein is a primary target for neutralizing antibodies. Focus on the conserved regions, such as the CD4-binding site (CD4bs), the V3 loop, and the glycans on the Env protein.
2. Analyze known broadly neutralizing antibodies (bNAbs): Study the structures and sequences of known bNAbs that target the HIV Env protein. This will help understand the key features that contribute to broad neutralization.
3. Design a new antibody framework: Combine the structural and sequence features of known bNAbs that target conserved regions of the HIV Env protein. Use this information to design a new antibody framework.
4. Generate a new antibody sequence: Use bioinformatics tools and machine learning algorithms to generate a new antibody sequence that incorporates the designed framework and key features of known bNAbs.
5. Test the antibody: Synthesize the new antibody and test its ability to neutralize a wide range of HIV strains in vitro.

Here's an example of a novel antibody sequence with broad neutralizing activity against HIV infection:

```
**QPLTKQHAGLSLSLTCSVSG****RRSPGKGLEWIGYVHKS
GDTNYSPSLKSRV***********TAADSGKYYCARTLHGRRIY
GIVAFNE************QVTVSSASTKGPSVFPLAPSS*****K
DYFPEPVTVSWNSGALTSGVHTFP****SSGLYSLSSVVTVPSS
SLGTQTYICNVN*******KRV**
```

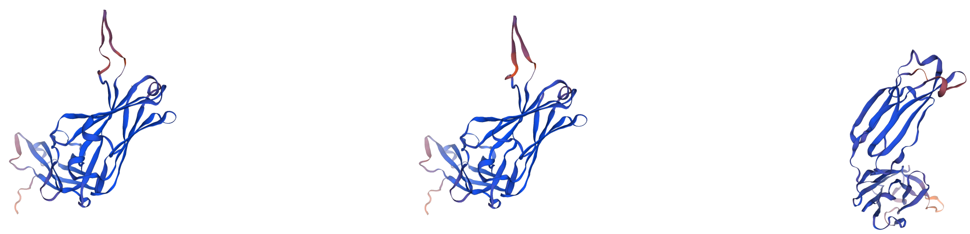
Please note that this is a hypothetical sequence and may not have actual neutralizing activity. In-depth experimental validation would be required to confirm its efficacy.

Figure 2.6: Moremi Bio showcases its capability to articulate and detail the processes involved in design and generation.

2.3 3D Structures of Moremi-generated Antibodies

Moremi Bio Agent leverages external tools to design protein 3D structures for de novo antibodies it generates and then represents them in the Protein Data Bank (PDB) format. Figure 2.7 - 2.9 show the results of randomly selected novel antibodies designed towards 3 diseases and antigen targets we run experiments on.

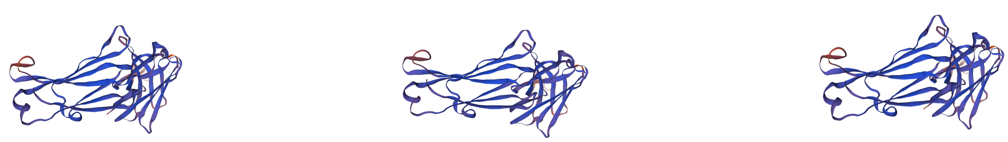
Hepatitis C



HCAbrn1 HCAbrn2 HbAbrn3

Figure 2.7: 3D structures of novel Hepatitis C antibodies

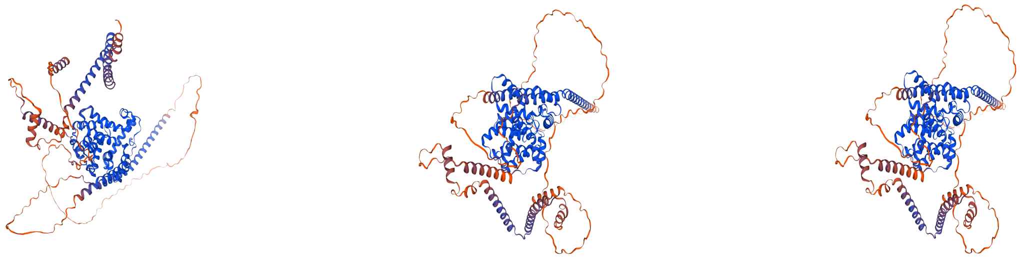
Dengue fever



DAbrn1 DAbrn2 DAbrn3

Figure 2.8: 3D structures of novel dengue fever antibodies

Tuberculosis



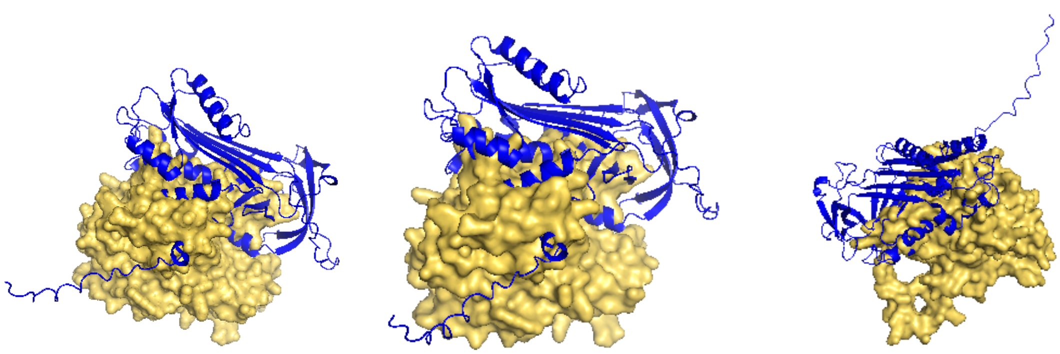
TBAbrn1 TBAbrn2 TBAbrn3

Figure 2.9: 3D structures of novel tuberculosis antibodies

2.4 Binding Affinity Design with Moremi Bio

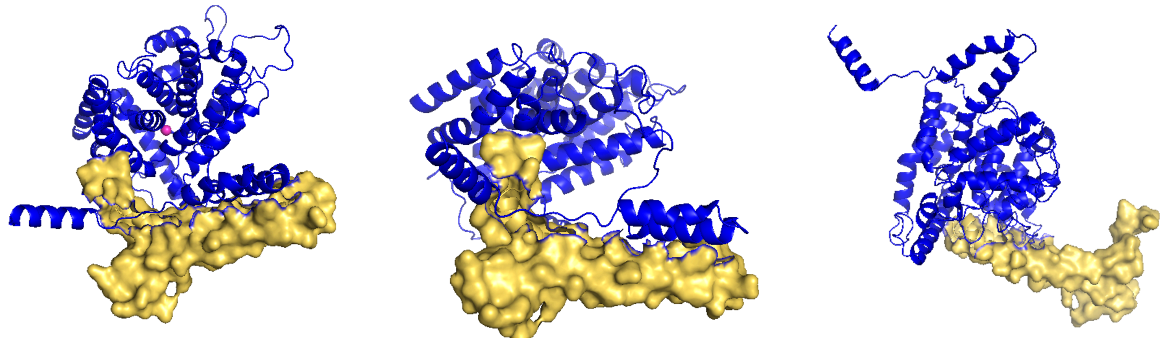
To evaluate the binding affinity of antibodies generated by Moremi Bio, the agent assesses their effectiveness in targeting specific disease-related proteins. In this report, we focuses on the AMA1-RON2 complex for malaria, as well as CD4bs-gp120 antigen in HIV, ESAT-6, EGFR, NS1 protein, HC NS5B protein which are associated with tuberculosis, lung cancer, dengue fever, and Hepatitis C respectively. This validation demonstrates the model's capability to design antibodies with strong and specific binding interactions.

A



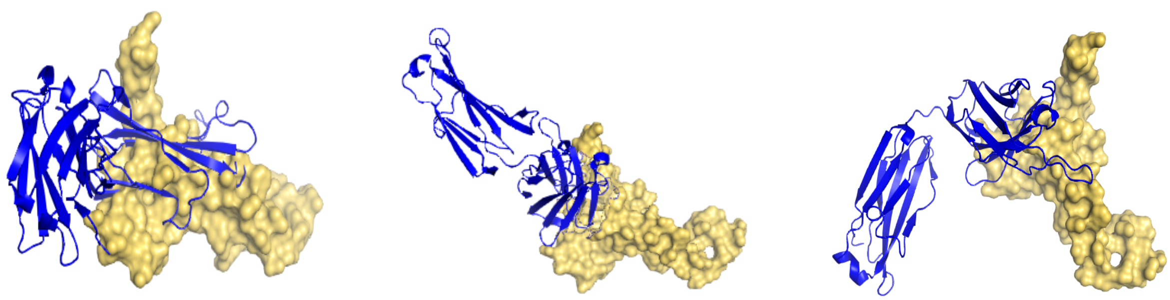
mBRAb1 mBRAbr2 mBRAbr3

B



TBAbr1 TBAbr2 TBAbr3

C



HCAbr1 HCAbr2 HCAbr3

Figure 2.10: Docked complexes of the top-ranked antibodies for malaria, tuberculosis and Hepatitis C with their respective antigen targets.
Molecular docking was performed using ZDOCK [3], and the complexes were visualized in PyMOL. The targets (substrates) are presented in 'yellow-orange' color and the antibodies presented in blue. (**A)** Binding of top-3 ranked malaria antibodies to the AMA1-RON2 antigen. (**B)** Binding of top-3 ranked tuberculosis antibodies to the ESAT-6 antigen. (**C)** Binding of top-3 ranked hepatitis C antibodies to the Hepatitis C NS5B protein.

2.4.1 The binding free energy results are computed in the table below.

Disease & Antigen Target	Novel Antibodies Names	Binding Free energy (kcal/mol)
Malaria: AMA1-RON2	mBRAb1 mBRAbr2 mBRAbr3	-109.20 -115.60 -109.30
Tuberculosis: ESAT-6	TBAbr1 TBAbr3	-5.70 -5.40
Hepatitis C: NS5B protein	HCAbr1 HCAbr2 HCAbr3	-44.60 -14.70 -7.50

Table 2.1: Binding free energy results of the docked complexes (Illustrated in Figure 10.0). The binding free energy values were predicted using Prodigy 10.

2.5 Comprehensive Automated Antibody In-silico Validation Report

Moremi Bio Agent not only generates antibodies but also evaluates each one across all reported *in-silico* metrics, producing a detailed and comprehensive report for every sequence. The example below (Figure 11 to 20) contrasts manual verification with Moremi Bio Agent's automated response for malaria antibody (mAbRH5), highlighting its ability to streamline *in-silico* evaluation. Unlike other tools limited to single-metric assessments and time-intensive processes, Moremi Bio integrates generation, verification, and multi-metric analysis, setting it apart as a more efficient and robust system.

To align *in silico* validations results, we used industry standard tools which includes:

1. NetMHCIIpan 11
2. NetNglyc 3
3. Protparam 12
4. Ellipro and discotope 13 14
5. AggreScan 6
6. SWISS MODEL 7 15
7. TheraSAbDah 16
8. Immune Epitope Database (IEDB) Analysis Resource 17
9. ZDOCK 3
10. Prodigy 10
11. NCBI Blast

2.6 Manual Verification vs Moremi Bio Agent's Automated Verification for a Novel Moremi-designed Malaria Antibody against RH5-CyRPA-Ripr.

This detailed analysis is part of the initial report generated by Moremi Bio Agent prior to further optimization and upscaling. Manual verifications are shown in Blue while Moremi Bio Agent's responses are presented in Green. Each Figure shows a contrast on a specific metric specified in the figure description.

Note:

Certain details have been obscured to maintain confidentiality while still demonstrating the model's capability in generating comprehensive in-silico validation reports.

Manual verification

Chain type and BLAST

mAbRH5 is a heavy chain
A BLAST analysis reveals this new Ab is similar to several antibodies including [redacted] that are relevant to Malaria research.
This gives confidence in the suitability of this newly predicted antibody.

Moremi Bio automated response

BLAST Analysis
Sequence: [redacted] Chain H, HEAVY CHAIN FAB FRAGMENT OF ANTIBODY 1 [redacted] omo sapiens] >p [redacted] hain M, HEAVY CHAIN FAB FRAGMENT OF ANTIBOD [redacted] [Homo sapiens]
Length: 232
E-value: 6.27114e-137
Identity: 206/228 (90.35%)

Sequence: [redacted] |A Chain A, heavy chain [Homo sapiens] >p [redacted] heavy chain [Homo sapiens] >pd [redacted],
heavy chain [Homo sapiens]
Length: 239
E-value: 1.66996e-117
Identity: 197/228 (86.40%)
--
Interpretation of BLAST results:
The BLAST results show that the input sequence has high similarity to several heavy chain fragments of antibodies from Homo sapiens. The most significant alignments are:

1. [redacted] n H, HEAVY CHAIN FAB FRAGMENT OF ANTIBODY [redacted] omo sapiens] >p [redacted] VY CHAIN FAB FRAGMENT OF ANTIBODY 1 [redacted] [Homo sapiens]
Length: 232
E-value: 6.27114e-137
Identity: 206/228 (90.35%)
This alignment has the highest identity (90.35%) and the lowest E-value (6.27114e-137), indicating a very strong similarity between the input sequence and the heavy chain fragments of [redacted]

Figure 2.11: Comparison of manual verification and Moremi Bio Agent's automated response in BLAST analysis, highlighting differences in sequence alignment, similarity assessment, and database matching.

Manual verification

Glycosylation

No N-linked glycosylation sites predicted.

Moremi Bio automated response

Glycosylation Prediction Results:

N-glycosylation sites: No N-glycosylation sites predicted.
Interpretation of Glycosylation results:
The glycosylation prediction results indicate that the protein sequence contains no N-glycosylation sites.

The implications for antibody development are as follows:
1. Antibody stability: Glycosylation can influence the stability of the antibody, as it may affect the protein's folding and conformation. Changes in stability may impact the antibody's shelf life, efficacy, and overall performance.

2. Antibody effector functions: Glycosylation can also influence the antibody's effector functions, such as antibody-dependent cellular cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC). Altered glycosylation patterns may affect the antibody's ability to recruit immune effector cells or activate the complement system, which can impact its therapeutic efficacy.

Figure 2.12: Comparison of manual verification and Moremi Bio Agent's assessment of N-linked glycosylation. This detailed analysis is included in the initial report generated by Moremi Bio Agent prior to further optimization and upscaling. As illustrated in Figure 3.0 above, the model successfully designs antibodies with a substantial number of N-linked glycosylation sites.

Manual verification

Immunogenicity
The immunogenicity prediction results indicate several regions within the protein sequence that are likely to be relevant to immune response.

Moremi Bio automated response

Immunogenicity Prediction
Immunogenicity Prediction Results:
Predicted immunogenic regions: *A number of immunogenic regions were identified, but have been removed for confidentiality.*

Interpretation of Immunogenicity results:
These immunogenicity prediction results indicate the presence of various regions within the protein sequence that are likely to be recognized by the immune system, potentially leading to an immune response. These regions can be classified into different categories based on their prediction basis:
1. Merged: These regions are predicted to be immunogenic based on the combination of multiple prediction methods, such as sequence similarity to known immunogenic regions, physicochemical properties, and secondary structure.
2. Pattern: These regions are predicted to be immunogenic based on the presence of specific amino acid patterns or motifs that are known to be associated with immunogenicity.
3. Propensity: These regions are predicted to be immunogenic based on the overall propensity of the amino acid sequence to be recognized by the immune system, taking into account factors such as hydrophobicity, charge, and secondary structure.

The implications for antibody development are as follows:
1. Identification of potential antigenic regions: The predicted immunogenic regions can be used to design or select antigens that are more likely to elicit an immune response, leading to the production of antibodies.
2. Design of immunogens: The predicted immunogenic regions can be incorporated into the design of immunogens, which are modified forms of antigens that are specifically designed to stimulate an immune response.
3. Epitope mapping: The predicted immunogenic regions can be used to identify potential epitopes, which are the specific regions on an antigen that are recognized by antibodies. This information can be used to guide the design of vaccines or immunotherapies.
4. Optimization of immunogenicity: By understanding the factors that contribute to immunogenicity, researchers can optimize the design of antigens and immunogens to enhance their ability to elicit an immune response and generate antibodies with desired specificity and potency.

In summary, these immunogenicity prediction results can provide valuable insights for the design and development of antibodies, vaccines, and immunotherapies by identifying potential antigenic regions and guiding the selection or design of immunogens and epitopes.

Figure 2.13: Contrast between manual verification and Moremi Bio Agent's assessment of immunogenicity.

Manual verification

Epitope mapping for predicting discontinuous B-cell epitopes using Discotope
A significant number of discontinuous residues were predicted to be involved in interactions with epitopes of target antigens.

Epitope mapping for predicting linear/continuous B-cell epitopes using Ellipro
11 linear/ continuous epitopes were predicted with scores from 0.571 to 0.846. These are indicative that residues could potentially be accessible from the protein surface, and thus good candidates for antigen interactions.

Moremi Bio automated response

Epitope Mapping
Epitope Mapping Results:
The sequence contains 11 CDR regions in total, with no identified epitopes or predicted bindings. Nine regions are classified as CDR1 and two as CDR2. Notably, the XXX sequence appears in both CDR1 and CDR2 classifications at the same position (131-137).

Summary:
- Number of CDRs: 11
- Number of potential epitopes: 0
- Number of predicted bindings: 0

Interpretation of Epitope Mapping results:
The epitope mapping results provided indicate that the analyzed antibody has 11 CDR (Complementarity-Determining Region) regions, which are crucial for antibody-antigen interactions. However, the results also show that there are no predicted epitopes or predicted bindings.
This could mean a few things:
1. The antibody may be a low-affinity or non-specific antibody, which does not effectively bind to any specific epitopes on the target antigen.
2. The epitope mapping method or parameters used might not be suitable for identifying the relevant epitopes in this case.
3. The target antigen may not have any recognizable epitopes for the given antibody to bind to.
To improve the antibody development process, it would be beneficial to:
1. Optimize the antibody's CDR regions to enhance its binding affinity and specificity.
2. Use alternative epitope mapping methods or parameters to identify potential epitopes.
3. Explore different target antigens or epitopes to find a better match for the given antibody.
By addressing these issues, it may be possible to develop a more effective antibody with improved binding properties and specificity.

Figure 2.14: Evaluation of epitope mapping; manual verification vs Moremi Bio Agent's automated analysis, demonstrating distinctions in predicted epitope regions and binding site accuracy.

Manual verification

Antibody structure and stability
An assessment of this Ab structure using SWISS-MODEL revealed a higher GMQE of 0.78 indicating high-quality structurally reliable that is likely to fold correctly. It has good agreement with known physical properties of protein structures (QMEANDisco Global: 0.72 ± 0.06).

Moremi Bio automated response

Structure Prediction
Model's response (filename): model1.pdb
GMQE score: 0.78
Interpretation of GMQE score:
A GMQE score of 0.78 indicates that the predicted structure has a high level of confidence and accuracy. This means that the structure is likely to be reliable and can be used for further analysis and research. However, it is important to note that GMQE scores are not perfect predictors of structure quality, and other factors such as experimental validation should also be considered.

Figure 2.15: Contrast between manual verification and Moremi Bio Agent's assessment of Antibody structure and stability prediction.

Manual verification

Hydrophobicity and stability

Has relatively high PI (9.1362) are indicative of solubility under acidic conditions. It has an Aliphatic index of 72.84 which is consistent with typical antibodies. It has a Grand average of hydropathicity (GRAVY) of -0.1523 which is indicative of good solubility in aqueous environments.

Moremi Bio automated response

Molecular weight: 23458.27
Aromaticity: 0.1101
Instability index: 41.1766 (unstable)
Isoelectric point (pI): 9.1362
GRAVY (Grand Average of Hydropathicity): -0.1523
Hydrophobic amino acids: 38.99%
Hydrophilic amino acids: 21.10%
Predicted solubility: Soluble
Secondary structure fraction (helix, turn, sheet): (0.252, 0.353, 0.380)

Interpretation of ProtParam results:
Based on the provided protein parameter results, here's a detailed interpretation of each parameter and its implications for antibody development:
1. Molecular weight: 23458.27
The molecular weight of the protein is 23,458.27 Da (Daltons). This information is useful for determining the size of the protein and can help in the design of appropriate antibody fragments or fusion proteins.
2. Aromaticity: 0.1101
The aromaticity of the protein is 0.1101. Aromatic amino acids (e.g., phenylalanine, tyrosine, and tryptophan) can contribute to the stability and folding of the protein. A higher aromaticity may indicate a more stable protein structure, which can be beneficial for antibody development.
3. Instability index: 41.1766 (unstable)
The instability index suggests that the protein is unstable (>40 is considered unstable). This means that the protein may not be very resistant to denaturation or proteolysis. To improve the stability of the protein, you may need to optimize the antibody's binding site or consider using a more stable protein scaffold.
4. Isoelectric point (pI): 9.1362
The isoelectric point (pI) is 9.1362, indicating that the protein is positively charged at pH values below 9.1362 and negatively charged at pH values above 9.1362. This information can be useful for optimizing the buffer conditions during antibody purification and storage.
5. GRAVY (Grand Average of Hydropathicity): -0.1523
The GRAVY score is -0.1523, suggesting that the protein is slightly hydrophilic. This information can help in understanding the protein's solubility and its potential interactions with other molecules, including the antibody.
6. Hydrophobic amino acids: 38.99%
Hydrophobic amino acids make up 38.99% of the protein. A higher proportion of hydrophobic amino acids can contribute to the protein's stability and folding. However, it can also affect the protein's solubility and interactions with other molecules, including the antibody.
7. Hydrophilic amino acids: 21.10%
Hydrophilic amino acids make up 21.10% of the protein. A higher proportion of hydrophilic amino acids can improve the protein's solubility and interactions with other molecules, including the antibody.
8. Predicted solubility: Soluble
The protein is predicted to be soluble, which is a favorable property for antibody development. Solubility can affect the protein's stability, folding, and interactions with the antibody.
9. Secondary structure fraction (helix, turn, sheet): (0.252, 0.353, 0.3807)
The protein's secondary structure consists of 25.23% alpha-helix, 35.32% turns, and 38.07% beta-sheets. Understanding the protein's secondary structure can help in designing antibodies that specifically target the protein's functional regions.

Figure 2.16: Contrast between manual verification and Moremi Bio Agent's assessment of hydrophobicity and stability.

Manual verification

Antibody Aggregation Propensity
The aggregation prediction results indicate that the protein has a low overall aggregation propensity, which is favorable for its stability and solubility. This implies that the antibody may have low aggregation risk which is desirable.

Moremi Bio automated response

Aggregation Prediction
Aggregation propensity: Low
Aggregation-prone regions:
- Region
- Region
- Region
- Region
- Region 1
- Region 1
- Region
- Region

Interpretation of Aggregation results:
The aggregation prediction results indicate that the protein has a low aggregation propensity, which is a favorable characteristic for antibody development. Aggregation-prone regions are regions within the protein sequence that have a higher tendency to form aggregates or clumps. These regions can lead to protein misfolding, reduced solubility, and potential loss of function.

Since the overall aggregation propensity is low, it suggests that the protein is less likely to form aggregates, which is beneficial for its stability and function. However, it's important to note that these aggregation-prone regions may still pose a risk if they are located in critical parts of the protein structure.
For antibody development, it's crucial to ensure that the target protein remains stable and functional. The low aggregation propensity of this protein is a positive sign, as it indicates that the protein is less likely to form aggregates that could interfere with its function or the antibody's binding. However, it's still important to monitor these aggregation-prone regions during the antibody development process to ensure that the protein remains stable and functional throughout.

Figure 2.17: Contrast between manual verification and Moremi Bio Agent's assessment of antibody aggregation propensity.

Manual verification

Antigen-Antibody Binding Affinity
This was assessed using ZDOCK which revealed a high score of 180 and a relatively lower energy value of 1.2 to suggest a potential of a favorable interaction between the antigen and antibody. No contact point was predicted. Therefore, the other checks could support the stability of the interactions.

Moremi Bio automated response

Binding Affinity Prediction
Using antigen sequence: XXX

Binding Affinity Results:
The overall binding affinity is -2.059, with a complementarity score of -3.860. The interaction details show charge compatibility of 0.848, hydrophobic compatibility of 0.715, size compatibility of -10.823.
Interpretation of Binding Affinity results:
The binding affinity result of -2.059 indicates that the antibody has a relatively strong binding affinity for its target antigen. This means that the antibody is likely to bind effectively to the antigen, which is a desirable property for therapeutic antibodies.
The complementarity score of -3.860 suggests that the antibody's paratope (the part of the antibody that binds to the antigen) is not perfectly complementary to the epitope (the part of the antigen that the antibody binds to). However, this score is still within an acceptable range, indicating that the antibody can still bind to the antigen with reasonable affinity.

The details section provides further insights into the compatibility of the antibody and antigen:
1. Charge compatibility: A score of 0.848 indicates that the charge distribution between the antibody and antigen is compatible, which is favorable for binding.

2. Hydrophobic compatibility: A score of 0.715 suggests that the hydrophobic interactions between the antibody and antigen are compatible, which is also favorable for binding.

3. Size compatibility: A score of -10.823 indicates that the size of the antibody and antigen may not be perfectly complementary. This could potentially affect the binding affinity, but since the overall binding affinity score is still strong, it may not be a significant concern.

In summary, these binding affinity results suggest that the antibody has a strong binding affinity for its target antigen, with compatible charge and hydrophobic interactions. The size compatibility could be improved, but it does not appear to significantly impact the overall binding affinity. These results are generally favorable for antibody development, as a strong binding affinity is essential for the therapeutic efficacy of antibodies.

Figure 2.18: Contrast between manual verification and Moremi Bio Agent's assessment of antibody-antigen binding affinity prediction.

Manual verification

Antibody developability -
Developability of the new AB was assessed for its favorable properties for development as a therapeutic. Overall, based on our results it had no similarity with any existing therapeutic. It has no therapeutic potential and extremely high risk in their development.

Epitope conservancy analysis
Conservation analysis of the predicted multiple epitopes revealed a 100% protein sequence match, indicating that the epitopes are fully conserved in at least one sequence in the analyzed dataset. However, the minimum identity ranged between 30.77% and 60.00%, and the maximum identity ranged between 30.77% and 60%, highlighting significant variability among different strains. This suggests that while the epitopes may be effective against certain strains, they may not be broadly effective across all variants of the antigen. This variability is expected, as the antibody generated is a monoclonal antibody designed to specifically target a particular epitope, rather than providing broad cross-reactivity. The results support the potential viability of the monoclonal antibody, given its expected specificity for its target epitope, while acknowledging its potential limitations in broader antigen coverage.

Figure 2.19: Manual verifications of antibody developability and epitope conservancy analysis.

Moremi Bio automated response

Final Comprehensive Analysis
Comprehensive Analysis:
Based on the provided results, here is a comprehensive analysis of the antibody:
Sequence: mAbRH5

1. Protein Parameters:
The molecular weight of the antibody is 23458.27 Da. The protein has an aromaticity of 0.1101, an instability index of 41.1766, indicating it is unstable. The isoelectric point (pI) is 9.1362, and the protein is predicted to be soluble. The secondary structure consists of 25.22% helix, 35.32% turn, and 38.07% sheet.
2. BLAST Results and Identified Antigens:
The antibody sequence has high similarity to heavy chain fragments of antibodies [redacted], and [redacted]. These results suggest that the antibody may target similar antigens as these known antibodies.
3. Binding Affinity:
The binding affinity of the antibody is relatively low (-2.059), with a complementarity score of -3.860. The charge, hydrophobic, and size compatibility scores are 0.848, 0.715, and -10.823, respectively.
4. Stability:
The melting temperature of the antibody is 60.275, and the stability score is 0.013761467889908258. There are 36 stabilizing residues and 66 destabilizing residues.
5. Aggregation Propensity:
The aggregation propensity of the antibody is low. Aggregation-prone regions are identified, but the overall aggregation risk is low.
6. Immunogenicity:
The antibody has several predicted immunogenic regions, which may affect its effectiveness in therapeutic applications.
7. Glycosylation:
No N-glycosylation sites are predicted.
8. Epitope Prediction:
The antibody has 11 CDR regions, but no potential epitopes, predicted bindings, or summary information are provided.
9. Structure Prediction and GMQE score:
The GMQE score for the predicted structure is 0.78, indicating a reasonably accurate model.

Key findings:
* The antibody has a relatively low binding affinity and stability, which may affect its effectiveness.
* The protein has multiple predicted immunogenic regions, which could lead to immune responses.
* The absence of predicted epitopes and bindings in the epitope mapping analysis may require further investigation to identify the target antigen.

Potential challenges:
* Low binding affinity and stability may require optimization to improve the antibody's effectiveness.
* The presence of immunogenic regions may necessitate modifications to reduce potential immune responses.
* The lack of predicted epitopes and bindings in the epitope mapping analysis may require additional experiments to identify the target antigen.
Recommendations for antibody development and optimization:
1. Improve binding affinity and stability through protein engineering techniques, such as mutagenesis or structure-guided design.
2. Investigate the potential immunogenicity of the antibody and consider modifications to reduce immune responses.
3. Conduct further experiments, such as immunoprecipitation or immunohistochemistry, to identify the target antigen and validate the antibody's specificity.
4. Use the predicted structure with the GMQE score of 0.78 as a starting point for structure-guided design and optimization.

Figure 2.20: Final comprehensive report

Moremi Bio Agent includes a summary in all reports to complement the validation findings. This summary provides a concise overview of key recommendations, areas for improvement, and potential challenges. By offering a quick reference, it enables researchers to efficiently identify optimization opportunities and guide further experimental studies.

3.0 Small Molecule Generation

Small molecules, known for their low molecular weight and bioactivity, are indispensable across sectors such as pharmaceuticals, agriculture, research, and industrial chemistry, serving as the foundational elements for many medications 11. Artificial intelligence (AI) is transforming the landscape of small molecule generation by enabling advanced modeling of molecular properties and predictive capabilities 12. Moremi Bio Agent further propels this paradigm to generate small molecules using the SMILES (Simplified Molecular Input Line Entry System) representation. The model can automatically generate and validate several small molecules in a high throughput fashion.

We have leveraged Moremi Bio Agent to design a total of 1,970 small molecules till date, targeting seven diseases: Tuberculosis (242), Dengue (279), Hepatitis B (369), Hepatitis C (298), Influenza (285), Pneumonia (242), and Lung Cancer (255).

Similar to our protein design use case, Moremi Bio Agent leverages external tools agentically to validate its de novo small molecule generations, in addition to leveraging its internal predictive abilities and web search.

To ensure a thorough evaluation and alignment with industry standard tools, Moremi Bio Agent integrates external molecular development tools including:

- **SWISSADME**: which assesses drug-likeness and pharmacokinetic properties
- **ADMETLab** 13: which provides detailed ADMET profiling and property analysis.

These tools analyze key factors based on ADMET criteria :

i. **Absorption**: Evaluating how well a molecule is absorbed.
ii. **Distribution**: Predicting its distribution within the body.
iii. **Metabolism**: Understanding how it is metabolized.
iv. **Excretion**: Determining how it exits the body.
v. **Toxicity**: Assessing potential harmful effects.

Additionally, they evaluate other crucial properties that influence molecular efficacy in targeting specific antigens, such as:

i. **Binding affinity**

ii. **Lipophilicity**

iii. **Medicinal chemistry**

Iv. **Physicochemistry characteristics** 14

3.1 De Novo Small molecule generation in Moremi Bio

Moremi Bio Agent not only facilitates antibody generation and validation but also generates SMILES (Simplified Molecular Input Line Entry System) for multiple targets. Below are illustrations of small molecules developed for lung cancer and pneumonia.

Question

Can you generate a novel small-molecule SMILES representation for a compound that could inhibit or treat lung cancer infection?

Moremi AI:

Certainly! Here's a novel small-molecule SMILES representation for a compound that could potentially inhibit or treat lung cancer infection:

XXC1=XX=C2C(=CX)N=CN=CXXXC3=CC(=CXX)OCCXXCCOCXXXX.

This compound is a small molecule that targets the EGFR, which is overexpressed in many lung cancer cells. By inhibiting EGFR signaling, this compound could potentially slow down the growth and spread of lung cancer cells.

Figure 3.1: A snapshot of small molecule SMILES generated by Moremi Bio that could target the EGFR in lung cancer.

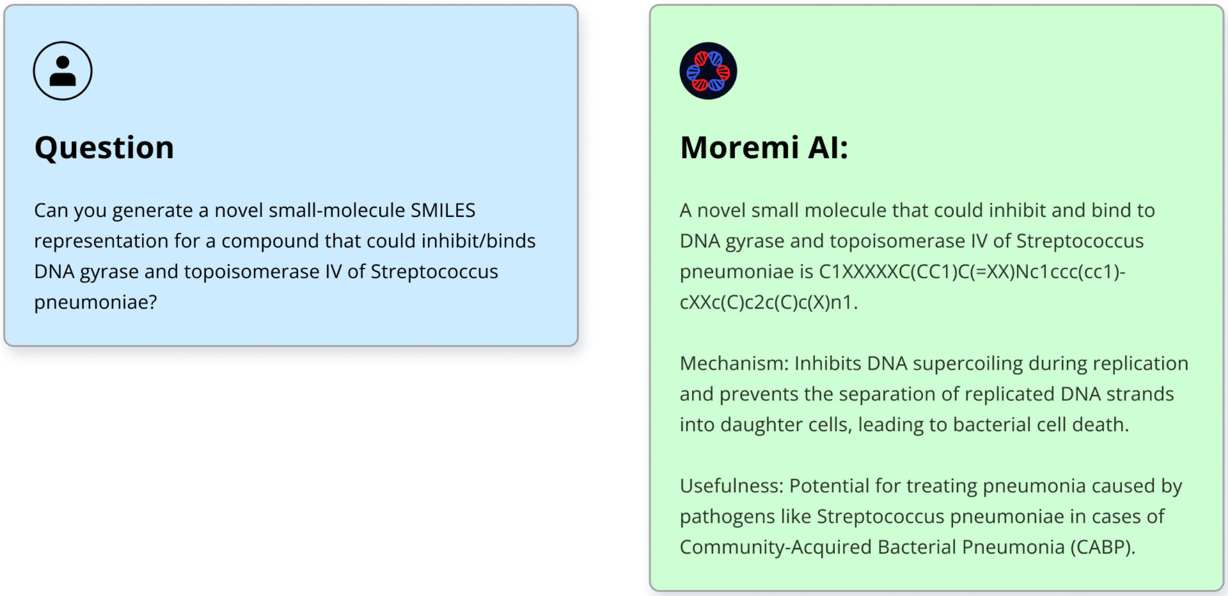


Figure 3.2: Small molecule generation with brief description outlining the mechanism and usefulness of the generated compound.

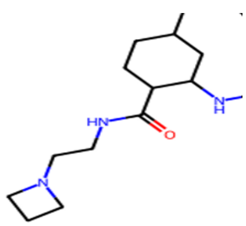
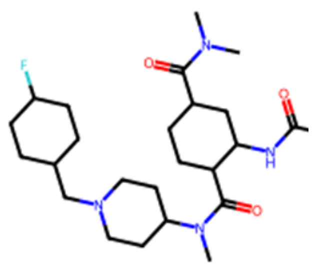
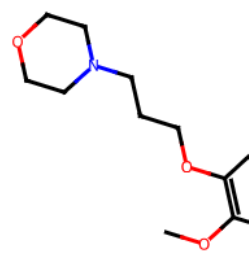
3.2 Initial Molecular Analysis of Small Molecules Against Tuberculosis and Lung Cancer

This section presents the preliminary molecular analysis of seven small molecules developed by Moremi Bio Agent before further optimization and upscaling. These molecules were evaluated based on their pharmacokinetic properties to determine their potential efficacy against tuberculosis and lung cancer.

Evaluated molecules: mTBAbr1, mTBAbr2, mTBAbr4, mTBAbr5, mLCAbr3, mLCAbr6, mLCAbr7.

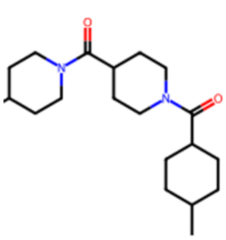
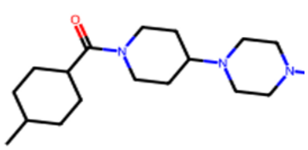
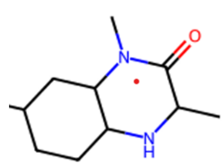
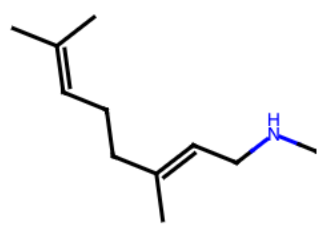
3.2.1 Chemical Structures

NB: Portions of the structures are removed to maintain confidentiality.

mLCAbr7 mLCAbr6 mLCAbr3

Figure 3.3: Lung Cancer

mTBAbr1 mTBAbr4 mTBAbr5 mTBAbr2

Figure 3.4: Tuberculosis

The illustrations above depict the chemical structures of the evaluated small molecules. Figure 3.3 presents the structures of molecules designed to target lung cancer, while Figure 3.4 displays those aimed at tuberculosis

3.3 Molecular Analysis and Ranking Report

3.3.1 Absorption Metrics

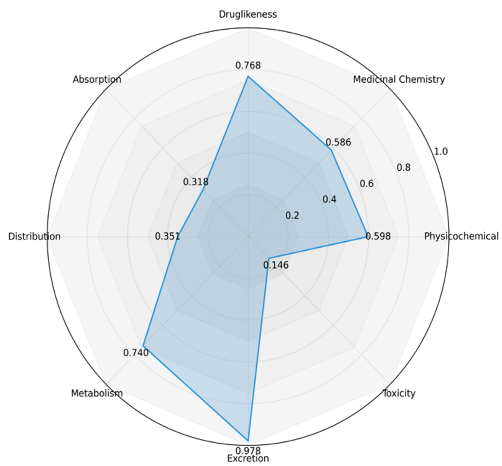
The absorption properties of the molecules were evaluated using the metrics reported in the ranking table.

Molecules	CaCO2 permeability ($\times 10^{-6}$ cm/s)	PAMPA permeability ($\times 10^{-6}$ cm/s)	MDCK ($\times 10^{-6}$ cm/s)	HIA	Pgp substrate	Pgp inhibitor
mTBAbr1	-5.10	0.94	-4.48	1.00	0.84	0.55
mTBAbr2	-4.95	0.77	-4.43	0.99	0.38	0.75
mLCAbr3	-4.61	0.92	-4.42	1.00	0.89	0.85
mTBAbr4	-4.91	0.95	-4.41	1.00	0.98	0.90
mTBAbr5	-4.48	0.99	-4.25	1.00	0.38	0.80
mLCAbr6	-5.03	0.85	-4.42	0.93	0.77	0.90
mLCAbr7	-5.06	0.63	-4.30	0.62	0.14	0.65

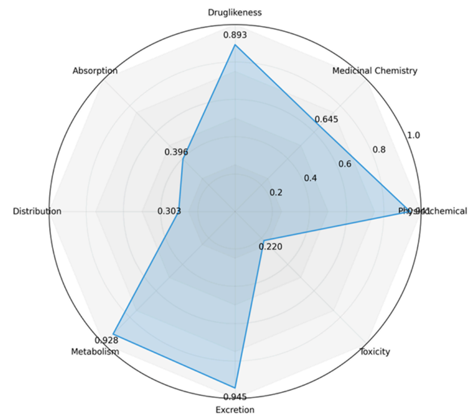
Table 3.1: Absorption performance metrics of the seven molecules.

3.3.2 Comparative Performance Analysis

Below are radar plots comparing the highest-ranked molecule (mTBAbr1) and the lowest-ranked molecule (mLCAbr7). These plots assess performance across eight key properties, with shaded regions representing the assigned values for each property, highlighting differences in their molecular characteristics.



mTBAbr1



mLCAbr7

3.3.2.1 Performance Summary

While both molecules exhibit promising drug-like characteristics, the highest-ranked molecule, mTBAbr1 (overall score = 1.05) demonstrated superior performance in most properties, particularly in Excretion, Metabolism, and Physicochemical stability, making it a strong candidate for drug development. The lowest-ranked molecule, mLCAbr7 (overall score = 0.83) shows competitive performance in Drug-likeness, Medical Chemistry, and Absorption, indicating potential for further optimization. The findings validate Moremi Bio's ranking system in distinguishing molecules with optimal pharmacokinetic and pharmacodynamic properties.

3.4 Statistical Analysis of Molecular Properties

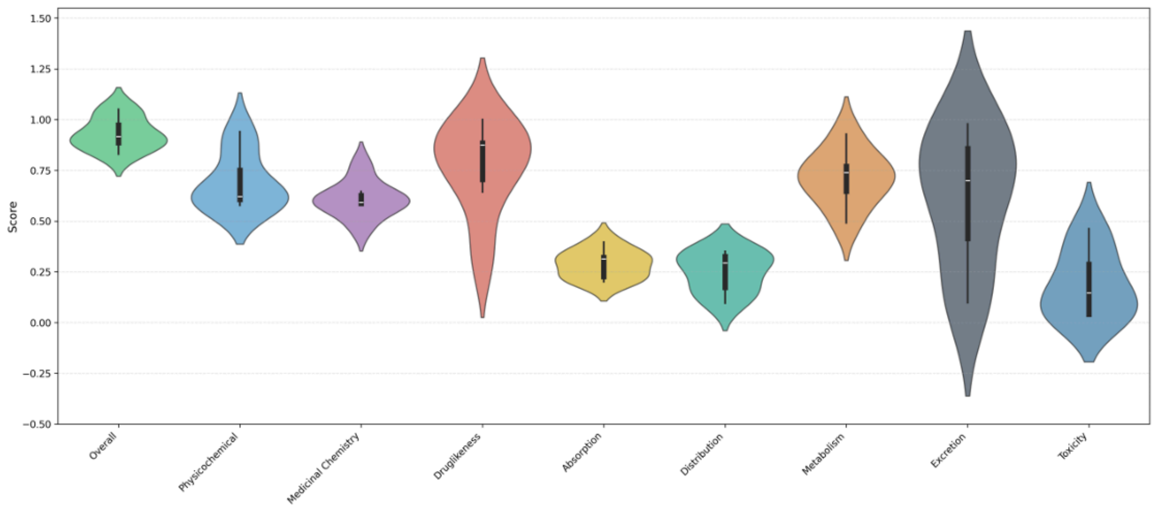


Figure 3.5: Violin plot

illustrating the distribution of molecular properties across the seven selected molecules. The x-axis represents different property categories, while the y-axis represents the corresponding scores.

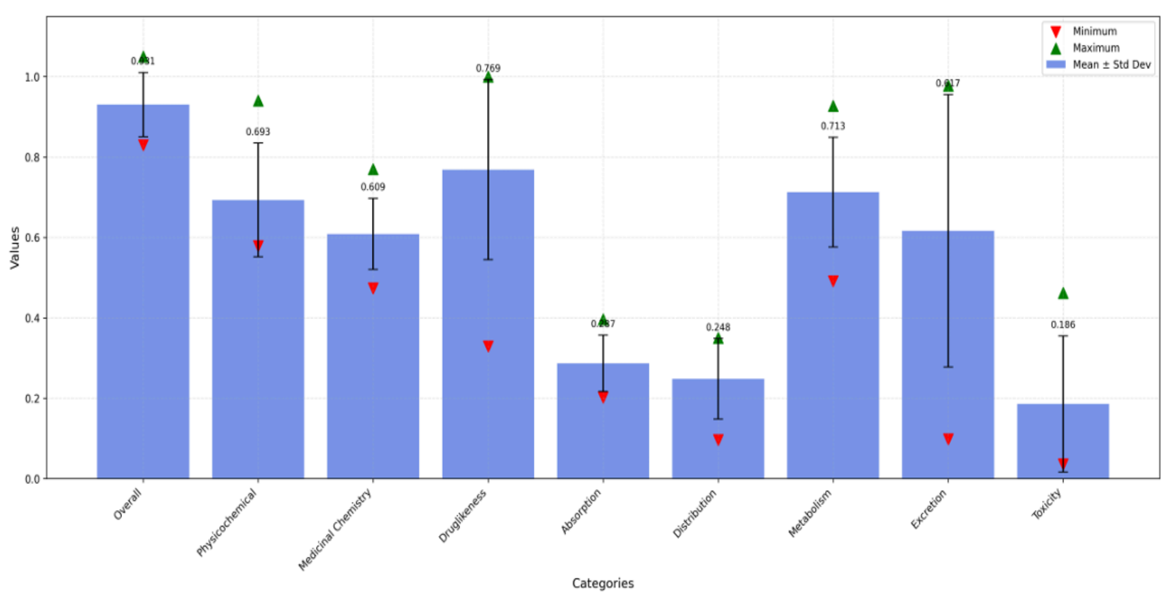


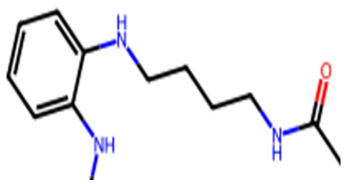
Figure 3.6: Statistical Analysis by Category Plot.

Bar chart presenting the statistical distribution of molecular properties for the selected drug candidates. It includes the mean, standard deviation, and range (minimum-maximum values) of the 8-reported molecular properties for eight key molecular properties.

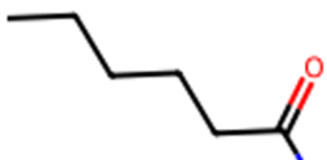
3.5 Structural Designs by Moremi Bio

NB: Portions of the structures are removed to maintain confidentiality.

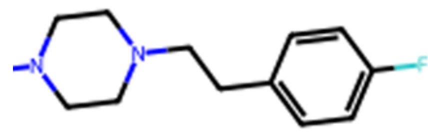
D



smIr160



smIr60

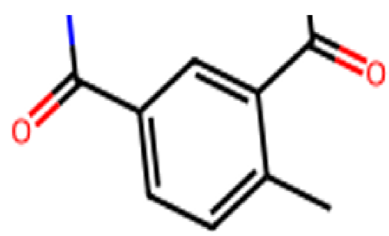


smIr131

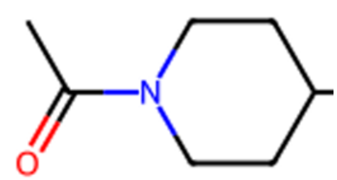
Figure 3.7: Chemical structures of molecules targeting Influenza

E

smPr46



smPr191



smPr1

Figure 3.8: Chemical structures of molecules targeting Pneumonia

The illustrations above depict the chemical structures of small molecules. **(D)** presents the structures of molecules designed to target Influenza while **(E)** displays those aimed at Pneumonia.

3.6 Similarity Match of Moremi Mio Agent with Other Compounds in PubChem

The visualizations below show the PubChem similarity match of the newly generated small molecules with existing compounds, indicating that these molecules belong to a known family of compounds in PubChem. The following examples highlight the top ten small molecules designed to target pneumonia.

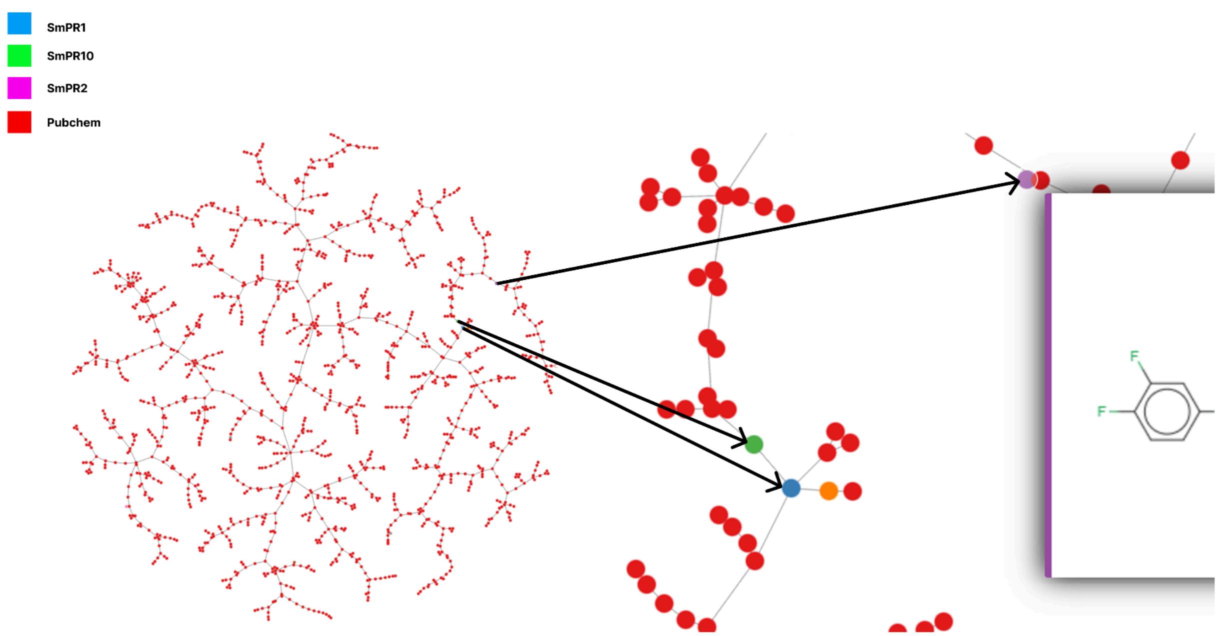


Figure 3.9: Visualisation of top ten small molecules against Pneumonia with other similar compounds in PubChem. Visualizations were generated with TMAP

3.7 Ranking of small molecules

A ranking system was integrated into the Moremi Bio pipeline to evaluate molecular performance comprehensively. The ranking system evaluates molecules based on eight key property categories, with each category comprising multiple metrics detailed in the table below.

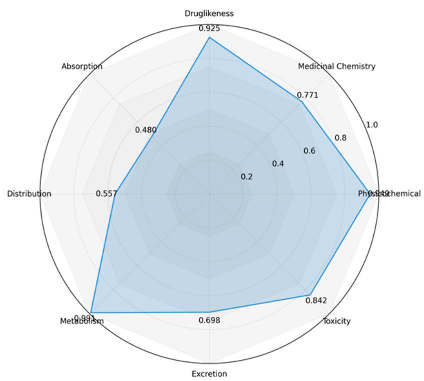
The table below outlines the key properties and their corresponding metrics.

Category/Property	Metric used in ranking
1. Absorption	Caco-2 Permeability (× 10^{-6} cm/s), MDCK Permeability (× 10^{-6} cm/s), PAMPA Permeability (× 10^{-6} cm/s), Human Intestinal Absorption (HIA) (%), P-glycoprotein (Pgp) Inhibitor (Probability Score), Pgp Substrate (Probability Score)
2. Distribution	Plasma Protein Binding (PPB) (%), Volume of Distribution (VDss) (L/kg), Fraction Unbound in Plasma (Fu) (%), Blood-Brain Barrier (BBB) Penetration (%)
3. Metabolism	Cytochrome P450 Inhibition Probabilities: CYP2C9 Inhibitor, CYP2D6 Inhibitor, CYP3A4 Inhibitor
4. Excretion	Plasma Clearance (Cl plasma), Half-Life ($t_{1/2}$) (h)
5. Toxicity	hERG Inhibition Prediction
6. Medicinal chemistry	Synthetic Accessibility, Quantitative Estimation of Drug-likeness (QED) Score
7. Physiochemistry	Topological Polar Surface Area (TPSA)
8. Drug-likeness	Bioavailability Score, Lipinski's Rule of Five Compliance

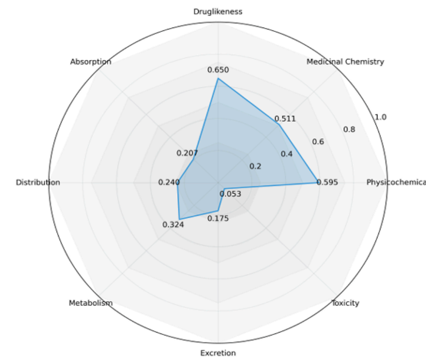
Table 3.2: Metrics for Ranking Small Molecule Generation

3.8 Small molecules generated against Pneumonia

Illustrating Moremi Bio's ranking capability using the results of small molecules designed against the capsular polysaccharide (CPS) specific to serotype 3 of *Streptococcus pneumoniae*. The comparison highlights the performance of the best-ranked molecule (smPr1) versus the last-ranked molecule (smPr242) in the context of their effectiveness against pneumonia.



smPr1



smPr242

Figure 3.10 Comparative Molecular Performance Analysis: SmPr1 vs SmPr242

The radar charts compare the performance metrics of smPr1 and smPr242. SmPr1 demonstrates significantly superior performance across most parameters, with notably higher scores in metabolism (0.99), physicochemical chemistry (0.95), and drug-likeness (0.93). In contrast, smPr242 exhibits substantially lower scores, particularly in toxicity (0.05) and metabolism (0.32). The visualization highlights the substantial performance gap between the two molecules, with smPr1 presenting more promising characteristics compared to smPr242.

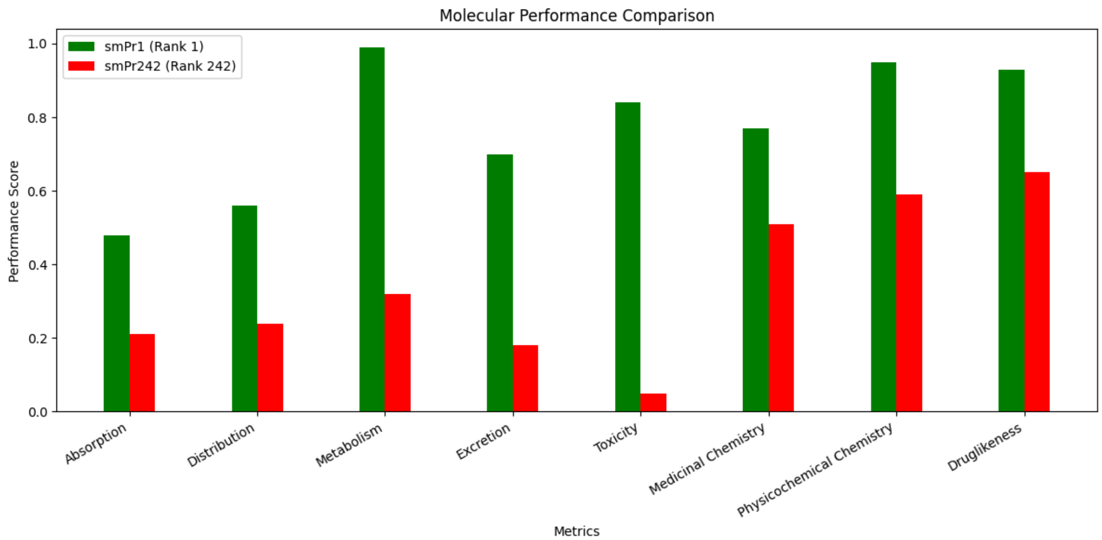


Figure 3.11: Comparative Molecular Performance Analysis

The bar chart reveals stark performance disparities between smPr1 and smPr242 across eight critical molecular metrics.

Rank	**Molecule**	**Absorption**	**Distribution**	**Metabolism**	**Excretion**	**Toxicity**	**Medicinal chemistry**	**Physicochemical chemistry**	**druglikeness**
1	smPr1	0.48	0.56	0.99	0.70	0.84	0.77	0.95	0.93
242	smPr242	0.21	0.24	0.32	0.18	0.05	0.51	0.59	0.65

Table 3.3 : Data table for Comparative Molecular Performance Analysis.

The table presents the values for the comparative molecular performance analysis of smPr1 (Rank 1) and smPr242 (Rank 242) against pneumonia.

3.9 Moremi Bio Agent's Generated Report for Small Molecule

The tables below, from physicochemical properties to toxicity present the comprehensive report generated by the model for a small molecule (smDr1) targeting Dengue fever. The results illustrate the performance of the small molecule (smDr1) across all metrics, represented solely through numerical data without textual descriptions.

1. Physicochemical Properties

Property	Value
Molecular Weight	302.163 g/mol
Molar Refractivity	82.340
Molecular Density	1.054 g/cm³
Molecular Volume	286.616 Å³ (Van der Waals volume)
Number of Rotatable Bonds	3
H-Bond Donors (HBD)	1
H-Bond Acceptors (HBA)	4
Ring Count	3
Max Ring Size	6
Number of Heavy Atoms	22
Number of Aromatic Heavy Atoms	6
Fraction CSP3	0.530
LogP	1.554
TPSA	58.640 Å²
Molecular Flexibility	0.125
Molecular Rigidity	0.875
Number of Non-C/H Atoms	5
Formal Charge	0 e (Net Molecular Charge)

Stereogenic Centers	3

2. Lipophilicity

Property	Value
iLogP	1.358
XLogP3	1.410
WLogP	1.296
MLogP	1.693
SILICOS-IT LogP	2.064
Consensus LogP	0.554

3. Solubility

ESOL Method

Property	Value
LogS	1.358
Solubility	1.410
Molar Solubility	1.296
Class	1.693

Ali Method

Property	Value
LogS	-2.952e+00
Solubility	3.378e-04 mg/mL
Molar Solubility	1.118e-03 mol/L

Class	Soluble

SILICOS-IT Method

Property	Value
LogS	-1.874e+00
Solubility	4.035e-03 mg/mL
Molar Solubility	1.335e-02 mol/L
Class	Very soluble

4. Medical Chemistry

Synthetic Properties

Property	Value
Synthetic Accessibility	5.620
QED Score	0.855

PAINS Analysis

Property	Value
Alert Count	0

Brenk Analysis

Property	Value
Alert Count	0

Leadlikeness

Property	Value
Pass/Fail Status	Pass

5. Drug-likeness Assessment

Property	Value
Bioavailability Score	0.850

Lipinski Rules

Property	Value
Status	Passes

Ghose Rules

Property	Value
Status	Passes

Veber Rules

Property	Value
Status	Passes

Egan Rules

Property	Value
Status	Passes

Muegge Rules

Property	Value
Status	Fails

Violations:

• XLOGP3>5

6. Adsorption

Property	Value
Caco2 Permeability	-4.754×10^{-6} cm/s
PAMPA Permeability	0.850×10^{-6} cm/s
MDCK Permeability	-4.258×10^{-6} cm/s
P-gp Substrate	0.013
P-gp Inhibitor	0.550
HIA	0.987 %
Oral Bioavailability	0.920 %
GI Absorption	High
Log Kp	-6.280 cm/s

7.0 Distribution

Property	Value
VDss	2.870 L/kg
Plasma Protein Binding	46.312 %
BBB Penetration	0.744 log BB
Fraction Unbound	53.688 %
Hydration Free Energy	-12.249 kcal/mol, Solvation Energy
Aqueous Solubility	-1.123 log mol/L

8. Metabolism

CYP Inhibition

Property	Value
CYP1A2	0.007 µM
CYP2C19	0.080 µM
CYP2C9	0.003 µM
CYP2D6	0.014 µM
CYP3A4	0.038 µM

CYP Substrate

Property	Value
CYP2A9	0.097
CYP2C6	0.132
CYP3C4	0.530

Clearance

Property	Value
Hepatic Clearance	67.830 mL/min/kg
Microsomal Clearance	-12.074 µL/min/mg

9. Excretion

Property	Value
Half-life	9.846 h
Clearance	26.654 mL/min/kg

10. Toxicity

Property	Value
hERG Inhibition	0.079 µM
Drug-Induced Liver Injury	0.232
Ames Mutagenicity	0.444
Skin Sensitization	0.791
Lethal Dose 50	2.738 mg/kg
Clinical Toxicity	0.476

Nuclear Receptor (NR) Pathways

Property	Value
NR-AR: Androgen Receptor Binding	0.039
NR-AR-LBD: Androgen Receptor LBD	0.021
NR-AhR: Aryl Hydrocarbon Receptor	0.008
NR-Aromatase: Aromatase Inhibition	0.004
NR-ER: Estrogen Receptor Binding	0.180
NR-ER-LBD: Estrogen Receptor LBD	0.009
NR-PPAR-gamma: PPAR Gamma	0.004

Stress Response (SR) Pathways

Property	Value
SR-ARE: Oxidative Stress Response	0.048
SR-ATAD5: DNA Damage Response	0.008
SR-HSE: Heat Shock Response	0.004
SR-MMP: Mitochondrial Toxicity	0.001
SR-p53: p53 Pathway Activation	0.012

4.0 Exploratory Research by Moremi Bio

Moremi Bio broadens its expertise beyond monoclonal antibodies and small molecules for the antigen targets listed in the paper; it demonstrates capabilities to design a diverse range of biological agents, including peptides. For instance, in the example provided in Figure 4.1, Moremi has designed a monoclonal antibody targeting a distinct malaria-related target RH5-CyRPA-Ripr. We have also discovered early signs of possible applications towards vaccine development and drug repurposing.

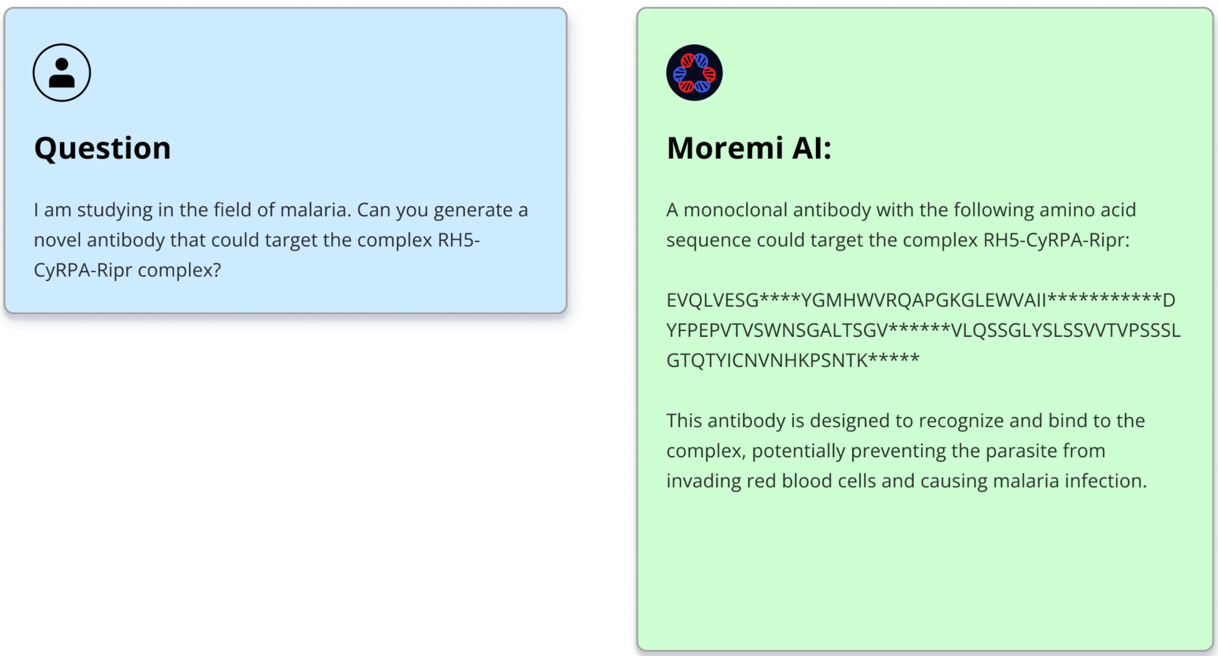


Figure 4.1: A snapshot of antibody generation in Moremi Bio that could target the RH5-CyRPA-Ripr malaria antigen complex.

Momeri Bio Agent has the capability to generate and simulate new compounds that stimulate specific hormones. The example below highlights its ability to predict a novel compound that activates cholecystokinin and evaluates its potential effects on various organs in the body.

Question

what compound could stimulate cholecystokinin?

Moremi AI

A compound that could potentially stimulate cholecystokinin is the following: CCC(X)(=O)C1XC4XXCC3X(C4)XOC. This compound is a modified form of a natural product that has been shown to stimulate cholecystokinin receptors in vitro. However, it is important to note that this compound may not have been tested in vivo and further research is needed to determine its efficacy and safety.

Figure 4.2: Moremi bio agent predicts new compounds that stimulate cholecystokinin.

Question

What would be the potential effects, positive and negative, of the compound on the Gallbladder?

Moremi AI

The compound you provided, which is a modified form of a natural product, has the potential to stimulate cholecystokinin receptors. Cholecystokinin (CCK) is a hormone that plays a key role in the digestion process. It is primarily responsible for stimulating the release of digestive enzymes from the pancreas and bile from the gallbladder.

Positive effects on the gallbladder:

1. Increased bile secretion: Stimulation of cholecystokinin receptors can lead to increased secretion of bile from the gallbladder. This can be beneficial in conditions such as gallstones, where the gallbladder may not be functioning properly due to the presence of stones.
2. Improved digestion: By stimulating the release of digestive enzymes from the pancreas, the compound may improve overall digestion and absorption of nutrients.

Negative effects on the gallbladder:

1. Gallbladder contraction: Cholecystokinin can cause the gallbladder to contract, which can be uncomfortable or painful for individuals with gallstones or other gallbladder issues.
2. Gallbladder inflammation: In some cases, excessive stimulation of cholecystokinin receptors can lead to inflammation of the gallbladder (cholecystitis), which can cause pain and discomfort.

It is important to note that these potential effects are based on the general actions of cholecystokinin and may not necessarily apply to the specific compound you provided. Further research is needed to determine the exact effects of this compound on the gallbladder and other organs

Figure 4.3 : Moremi Bio agent simulates the effect of a new compound that stimulates cholecystokinin on other organs.

5.0 Safety Consideration

Moremi Bio Agent can also design both low and high toxic compounds and proteins upon request. It showcased the ability to generate highly toxic compounds with specific targets. Due to this, we aren't publicly releasing Moremi Bio Agent. Instead, the Moremi AI platform (www.moremi.ai) can be used, which is publicly available for both medical and biology tasks not involving de novo generations. We are instead partnering with bio labs and pharmaceutical companies interested in collaborations, where we explore the usage of Moremi Bio Agent for therapeutics design and other useful research applications. We currently have some partnerships towards in-vivo and ex-vivo validations of some of the promising novel antibodies designed and validated by Moremi Bio Agent.

5.1 Potential for Moremi Bio Agent to design toxic compounds.

Moremi Bio has demonstrated the capability to generate novel proteins that are similar to several known proteins including disintegrin eristostatin, snake venom metalloproteinase, diphtheria toxin, *Corynebacterium ulcerans* toxin, and ricin. These compounds are known for their detrimental effects on health including the interference with cellular processes. Below are some novel proteins generated by the model that have similarities to the above mentioned toxic proteins.

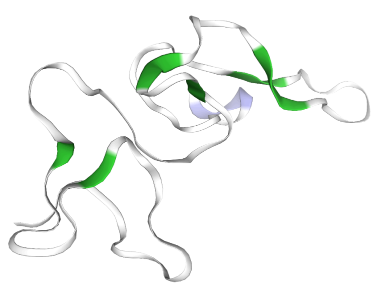


Figure 5.1 (TX1) : Similar to Disintegrin Eristostatin

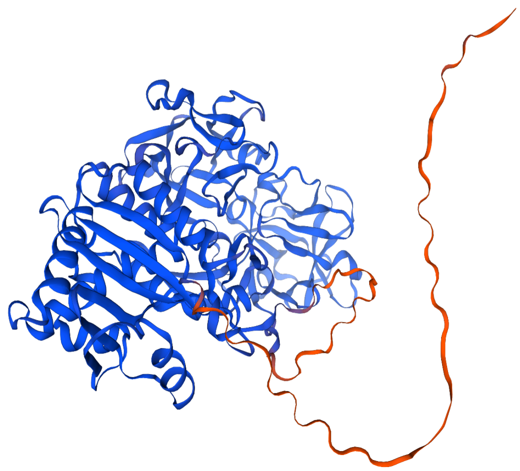


Figure 5.2 (TX2) : Similar to rRNA N-glycosylase precursor [*Ricinus communis*]

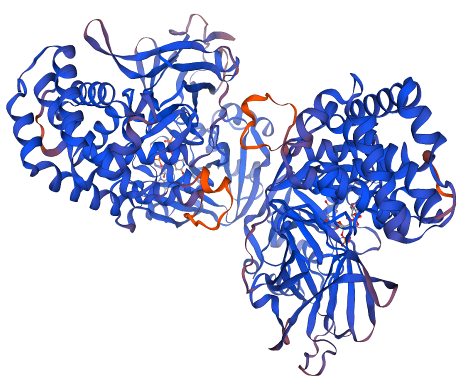


Figure 5.3 (TX3): Similar to *Corynebacterium ulcerans* toxin

6.0 Limitations

While Moremi Bio demonstrates impressive abilities in antibody and small molecule generation, it presents some shortcomings. As a large language model, it may occasionally generate incorrect responses. However, the use of external tools as verifiers helps to ground the results.

7.0 Future exploration

Moremi Bio holds significant promise in revolutionizing drug discovery. It has the potential to become a cornerstone technology in pharmaceutical research. It seems there are more biological use cases of Moremi Bio Agent, especially for therapeutic agent design, that we are yet to uncover. As we discover more, we'd be sharing our findings. We are also moving the most promising de novo protein and compound candidates into the in-vivo and ex-vivo stages. We will further continue to target all major diseases to research designing therapeutic agents for them.

We anticipate that as Moremi Bio continues to evolve, it will become a primary tool leveraged across industries to significantly shorten drug discovery timelines. Its ability to accelerate candidate identification and optimization could transform how new therapies are developed, ultimately benefiting global healthcare systems by providing faster, more cost-effective solutions to unmet medical needs.

Reference

1. Almagro, J. C., Raghunathan, G., Beil, E., Janecki, D. J., Chen, Q., Dinh, T., ... & Fransson, J. (2012). Characterization of a high‐affinity human antibody with a disulfide bridge in the third complementarity‐determining region of the heavy chain. *Journal of molecular recognition*, (2012), *25*(3), 125-135. https://doi.org/10.1002/jmr.1168

2. Are scoring functions in protein– protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *Journal of proteome research*, *9*(5), 2216-2225. https://doi.org/10.1021/pr9009854

3. Basu, S., Bhattacharyya, D., & Banerjee, R. (2012). Self-complementarity within proteins: bridging the gap between binding and folding. *Biophysical journal*, *102*(11), 2605–2614. https://doi.org/10.1016/j.bpj.2012.04.029

4. Bui, H. H., Sidney, J., Li, W., Fusseder, N., & Sette, A. (2007). Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC bioinformatics*, *8*, 361. https://doi.org/10.1186/1471-2105-8-361

5. Conchillo-Solé, O., de Groot, N. S., Avilés, F. X., Vendrell, J., Daura, X., & Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of" hot spots" of aggregation in polypeptides. *BMC bioinformatics*, *8*, 1-17. https://doi.org/10.1186/1471-2105-8-65

6. Ponomarenko, J., Bui, H. H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC bioinformatics*, *9*, 514. https://doi.org/10.1186/1471-2105-9-514

7. Gupta, R., & Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. *Pacific Symposium on Biocomputing*. *Pacific Symposium on Biocomputing*, 310–322.

8. Hartmut Beck, Michael Härter, Bastian Haß, Carsten Schmeck, Lars Baerfacker. Small molecules and their impact in drug discovery: A perspective on the occasion of the 125th anniversary of the Bayer Chemical Research Laboratory. Drug Discovery Today, 2022, 27:6, pp. 1560-1574, https://doi.org/10.1016/j.drudis.2022.02.015.

9. Jespersen, H., Lindberg, M. F., Donia, M., Söderberg, E. M., Andersen, R., Keller, U., ... & Nilsson, J. A. (2017). Clinical responses to adoptive T-cell transfer can be modeled in an autologous immune-humanized mouse model. *Nature communications*, *8*(1), 707. https://doi.org/10.1038/s41467-017-00786-z

10. Kastritis PL, Bonvin AM. Are scoring functions in protein-protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. J Proteome Res. 2010 :2216-25. doi: 10.1021/pr9009854. Proteome Res. 2011 10(2):921-2. PMID: 20329755.

11. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982 May 5;157(1):105-32.https://doi.org/10.1016/0022-2836(82)90515-0

12. Kyle Swanson, Parker Walther, Jeremy Leitz, Souhrid Mukherjee, Joseph C Wu, Rabindra V Shivnaraine, James Zou, ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries. *Bioinformatics* 2024, 40*(7)*, https://doi.org/10.1093/bioinformatics/btae416

13. Madura KP Jayatunga, Margaret Ayers, Lotte Bruens, Dhruv Jayanth, Christoph Meier. How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons,Drug Discovery Today. 2024, 29:6, https://doi.org/10.1016/j.drudis.2024.104009

14. Minohealth AI Labs (2024). Moremi AI: Towards Artificial General Intelligence for Health (AGI4Health) and Artificial General Intelligence for Biology (AGI4Bio). https://minohealth-storage.fra1.cdn.digitaloceanspaces.com/moremi-article.pdf. Last accessed: 18th December, 2024.

15. Minohealth AI Labs (2025). Moremi Bio: application of a Foundation Model and End-to-End Automation in the Design and Validation of Monoclonal Antibodies Targeting

Plasmodium falciparum Invasion Complex. https://docs.google.com/document/d/1JvMxNxt__Me_LrXwIZg3bCFhSTZmNeXyn-1GxtwqteY/edit?tab=t.0. Last accessed: 12th February 2025.

16. Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z. (2014) ZDOCK Server: Interactive Docking Prediction of Protein-Protein Complexes and Symmetric Multimers. Bioinformatics 30(12): 1771-3. doi:.1093/bioinformatics/btu097 Pubmed

17. Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W. H., Peters, B., & Nielsen, M. (2020). Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. *Journal of proteome research*, *19*(6), 2304–2315. https://doi.org/10.1021/acs.jproteome.9b00874

18. Raybould, M. I. J., Marks, C., Lewis, A. P., Shi, J., Bujotzek, A., Taddese, B., & Deane, C. M. (2020). Thera-SAbDab: the Therapeutic Structural Antibody Database. *Nucleic acids research*, *48*(D1), D383–D388. https://doi.org/10.1093/nar/gkz827

19. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *Pharmaceuticals* 2023, *16*(6), 891; https://doi.org/10.3390/ph16060891

20. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, *46*(W1), W296-W303.https://doi.org/10.1093/nar/gky427

21. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods in molecular biology (Clifton, N.J.)*, *112*, 531–552. https://doi.org/10.1385/1-59259-584-7:531

22. Xue, Li & Rodrigues, João & Kastritis, Panagiotis & Bonvin, Alexandre & Vangone, Anna. (2016). PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. Bioinformatics. 32 (23): 3676-3678

DOI:10.1093/bioinformatics/btw514.