Modelling Somatic Hypermutations with Reinforcement Learning for PD1 and Pembrolizumab

Anmol Singh, *Prof. Subhashini Srinivasan, Dr. Nithya Ramakrishnan*

Institute of Bioinformatics and Applied Biotechnology - Biotech Park, Electronics City Phase 1, Electronic City, Bengaluru, Karnataka 560100

ABSTRACT

We present a reinforcement learning (RL) model for Somatic Hypermutations (SHM), which mimics the natural selection process in a very short time scale. In this model, the agent can learn to preferentially mutate amino acids in the antibody, leading to affinity maturation. The model thus predicts a higher binding affinity antibody than the initial antibody-antigen complex. We have used the Pembrolizumab-PD1 (5b8c) complex to create the model, as Pembrolizumab (Pembro) is widely used in immunotherapy.

We use Q-Learning in RL to model SHM on a reduced state space to provide better binding affinity antibodies. We validated the structure of the antibodies predicted by the RL model using AlphaFold2 and inter-residue distance plots to check for proper folding of chains and protein-protein interactions. This study provides a proof of concept that RL can be used for modeling the biological process of SHM and can further be employed for creating novel antibodies.

OBJECTIVES

Major Objective: To create a reinforcement learning model for somatic hypermutations using Pembro and PD1.

Minor Objective: To find states/antibodies which have better binding affinity than Pembro, and validating the good states using AlphaFold2 and C-alpha distance plots of the predicted structures.

DESIGN AND WORKFLOW

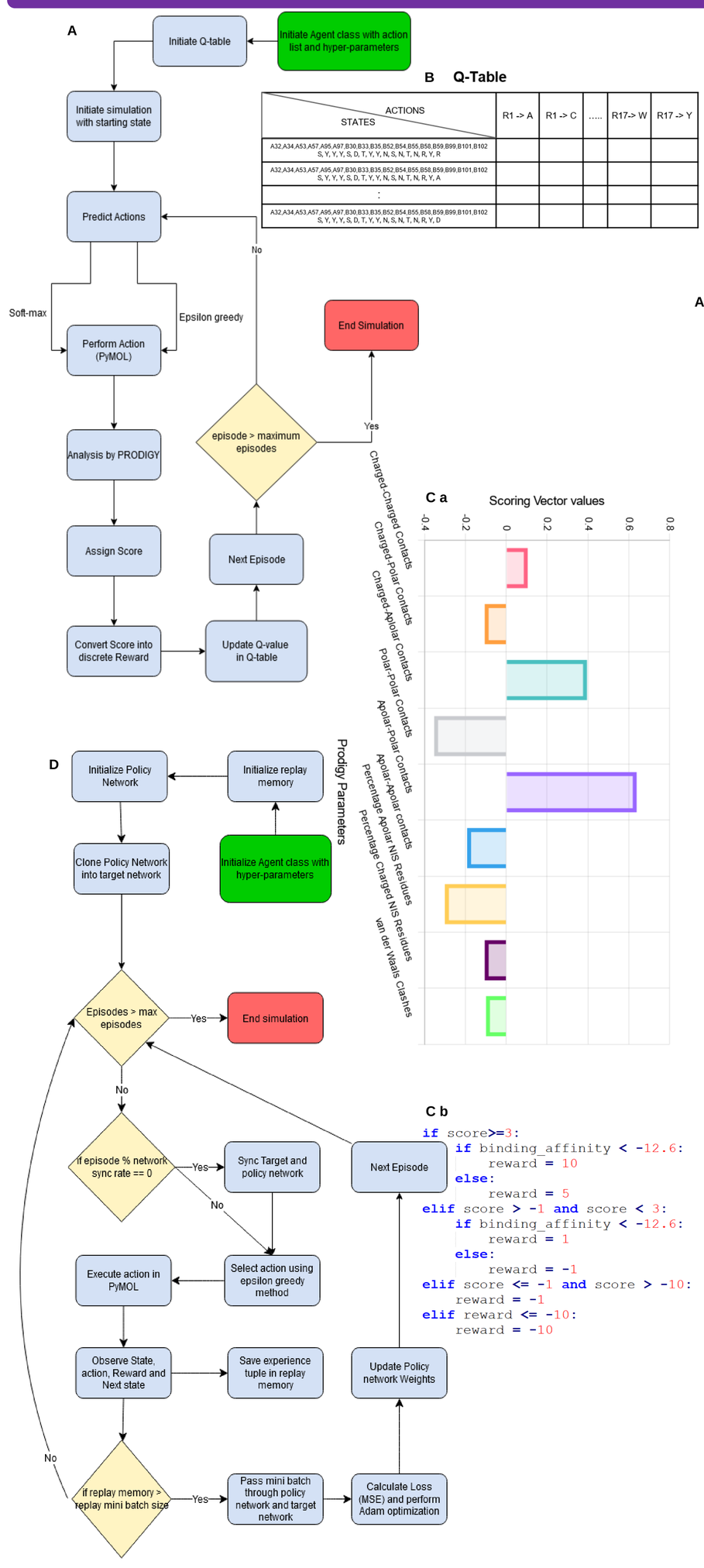


Figure 4. A Flowchart of Basic Q-learning, **B** Representation of Q-Table, **C a** Scoring vector, **b** Score to reward function, **D** Flowchart of Deep Q Learning,

INTRODUCTION

Somatic Hypermutations (SHM) are an important part of the affinity maturation (AM) in B-cells.

We express the biological process of AM as a Markov decision process (MDP), creating a RL model of SHM and AM.

Model Assumptions:

- The relative configuration of PD1-Pembro complex remains the same irrespective of the type of amino acids mutations at the 17 positions
- The 3D folds of light and heavy chains are not disrupted by the mutations in loops on Pembro

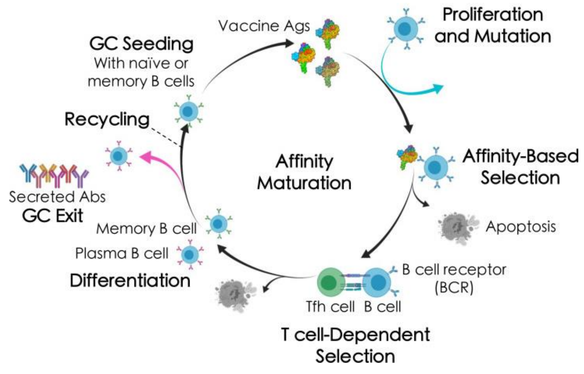


Figure 1. Broad overview of the affinity maturation (AM) process by which antibodies (Abs) evolve against vaccine-candidate antigens (Ags) in a germinal center (GC) reaction (Faris J et al., 2022)

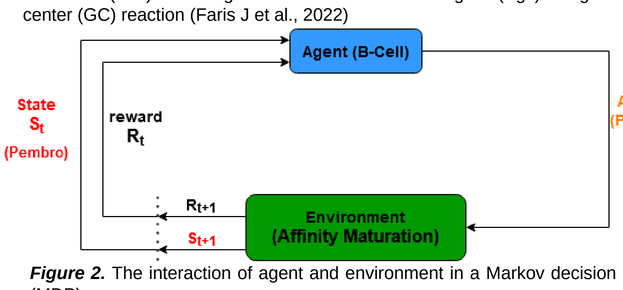
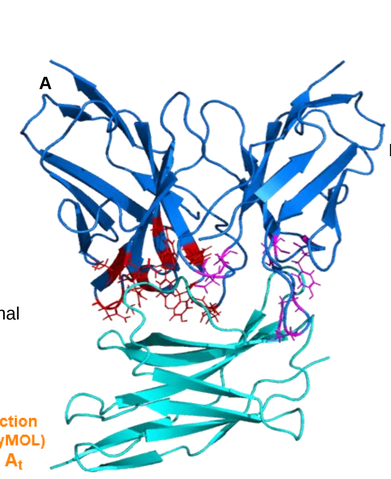


Figure 2. The interaction of agent and environment in a Markov decision process (MDP),
Agent: B-cells
Action: Point mutations on Pembrolizumab (using PyMOL)
Environment: Affinity Maturation of Pembrolizumab-PD1
State: Pembrolizumab with mutated amino acids



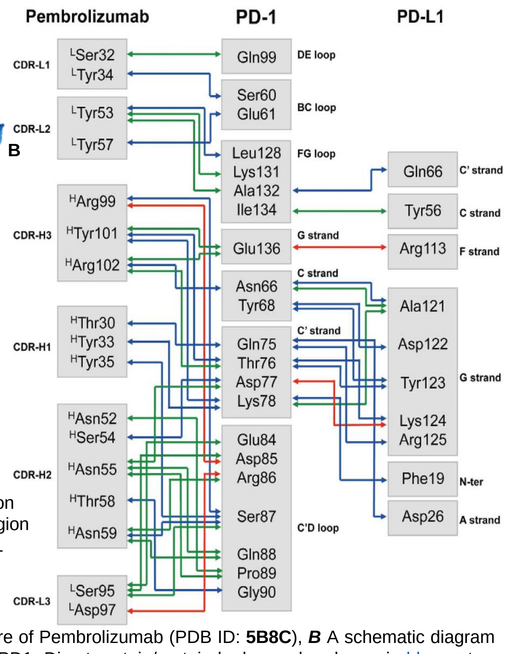


Figure 3. A Reference crystal structure of Pembrolizumab (PDB ID: **5B8C**), ***B*** A schematic diagram of interactions between PD1 and antiPD1. Direct protein/protein hydrogen bonds are in blue;water-mediated hydrogen bonds are in green; and salt bridges are in red (Horita S et al., 2016)

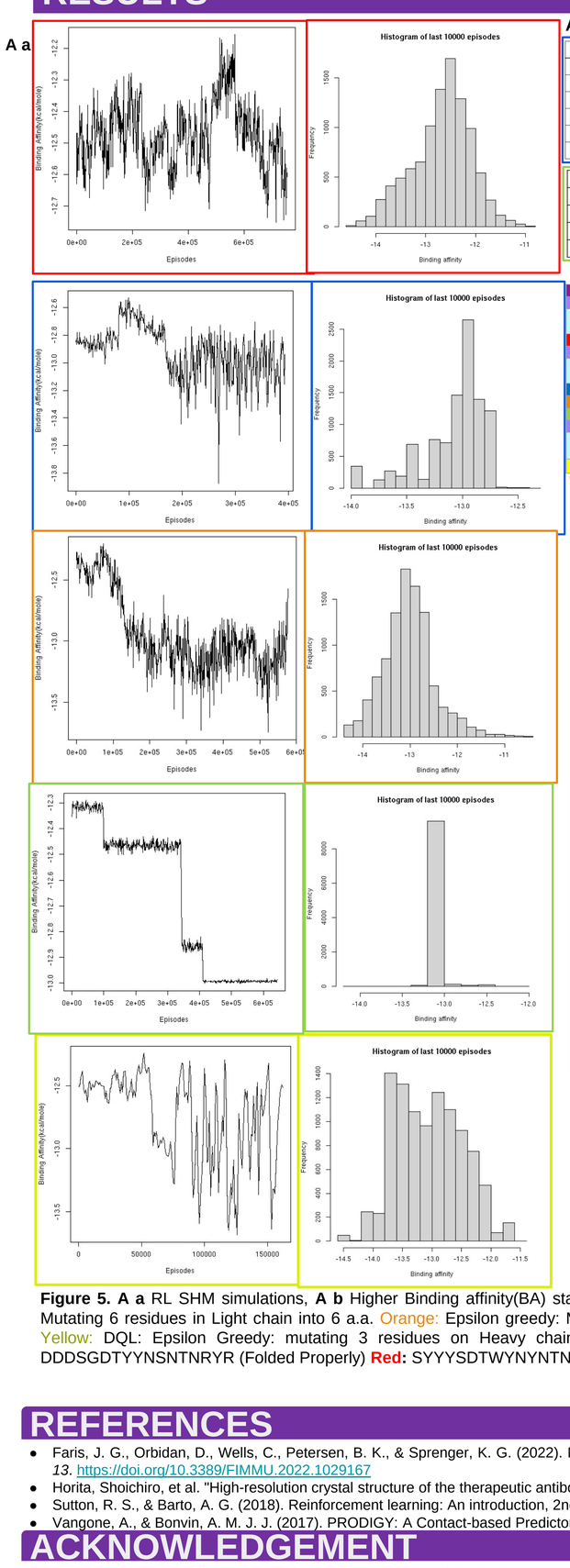
DISCUSSION AND FUTURE WORK

Q-Learning simulated SHM and provided a better binding Pembro-PD1 complex using a reduced state space.

Deep Q-Learning can be employed to predict alternate complexes to Pembro-PD1 – Currently underway.

Protein Language Models (PLM) can be employed in conjunction with our SHM-RL model for drug discovery.

RESULTS



A b Higher Affinity States/Abs

State	Freq	Reward	Crystal BA	AF BA	RMSD
DRRTSYTYYNSNTNRYR	65	10	-13.4	-23.3	19.48
STRNYYTYYNSNTNRYR	45	10	-13.6	-13.5	2.061
YRNNSRTYYNSNTNRYR	141	10	-13.4	-13.4	2.058
DSRNYYTYYNSNTNRYR	41	10	-13.7	-13.3	2.072
DTRNYYTYYNSNTNRYR	89	10	-13.7	-13.3	1.987
YRNNNRTYYNSNTNRYR	101	10	-13.3	-13.3	2.072

State	Freq	Reward	Crystal BA	AF BA	RMSD
SYYYSDTYYNFNTNRYR	5	1	-13.4	-13.1	2.179
SYYYSDTYYNLNTNRYR	6	1	-13.5	-13	2.153
SYYYSDTYYNWNTNRYR	6	1	-13.4	-13	2.286
SYYYSDTYYNYNTNRYR	9	1	-13.6	-12.8	2.193

State	Freq	Reward	Crystal BA	AF BA	RMSD
SYYYSDTYYNWNTNTHL	60	1	-12.9	-13.5	2.195
SYYYSDTYYNYNTNIFN	53	1	-13.0	-13.5	2.172
SYYYSDTYYNYNTNWAK	143	1	-13.0	-13.4	2.101
SYYYSDTYYNFNTNAWY	45	10	-14.0	-13.3	2.217
SYYYSDTYYNFNTNIYY	39	10	-14.1	-13.0	2.185

State	Freq	Reward	Crystal BA	AF BA	RMSD
SYYYSDTGYNLNTNDYR	165	1	-13	-13.9	2.169
SYYYSDTSYNLNTNDYR	117	1	-12.8	-13.6	2.177
SYYYSDTSYNLNTNLYR	15	1	-13.9	-13.2	2.161
SYYYSDTVYNLNTNDYR	94	1	-13.5	-13.1	2.147
SYYYSDTYYNVNTNEYR	52	1	-13.2	-13.1	2.141

Table: Simulation parameters

Q-Learning										
Soft-max										
Chain	Residues position	No. of Mutation a.a.	Mutations Residues	Temperature	Temperature decay	Temperature decay interval	Minimum Temperature	Learning rate	Discount factor	Random Seed
Heavy	Tyr33, Ser54, Arg99	17	A,D,E,F,G,H,I,K,L,N,Q,R,S,T,V,W,Y	100	0.5	1000	1	0.01	0.9	717
Epsilon Decay										
Chain	Residues position	No. of Mutation a.a.	Mutations Residues	Exploration rate	Exploration decay	Exploration decay interval	Minimum Exploration	Learning rate	Discount factor	Random Seed
Light	Ser32, Tyr34, Tyr53, Tyr57, Ser95, Asp97	6	S,Y,N,R,T,D	1	0.005	1000	0.01	0.01	0.9	643
Heavy	Ser54, Arg99,Tyr101, Arg102	16	A,D,E,F,G,H,I,K,L,N,Q,R,S,T,V,W,Y	1	0.005	1000	0.05	0.01	0.9	71
Heavy*	Ser54, Arg99,Tyr101, Arg102	16	A,D,E,F,G,H,I,K,L,N,Q,R,S,T,V,W,Y	1	0.005	1000	0.05	0.01	0.9	17
Deep Q Learning										
Chain	Residues position	No. of Mutation a.a.	Mutations Residues	Exploration rate	Exploration decay	Exploration decay interval	Minimum Exploration	Learning rate	Discount factor	Network Sync Rate	Replay Memory Size	Mini Batch Size	Random Seed
Heavy	Tyr33, Ser54, Arg99	17	A,D,E,F,G,H,I,K,L,N,Q,R,S,T,V,W,Y	1	0.01	1000	0.01	0.01	0.9	5000	5000	200	7

*Looping from starting state upon reaching -10 reward

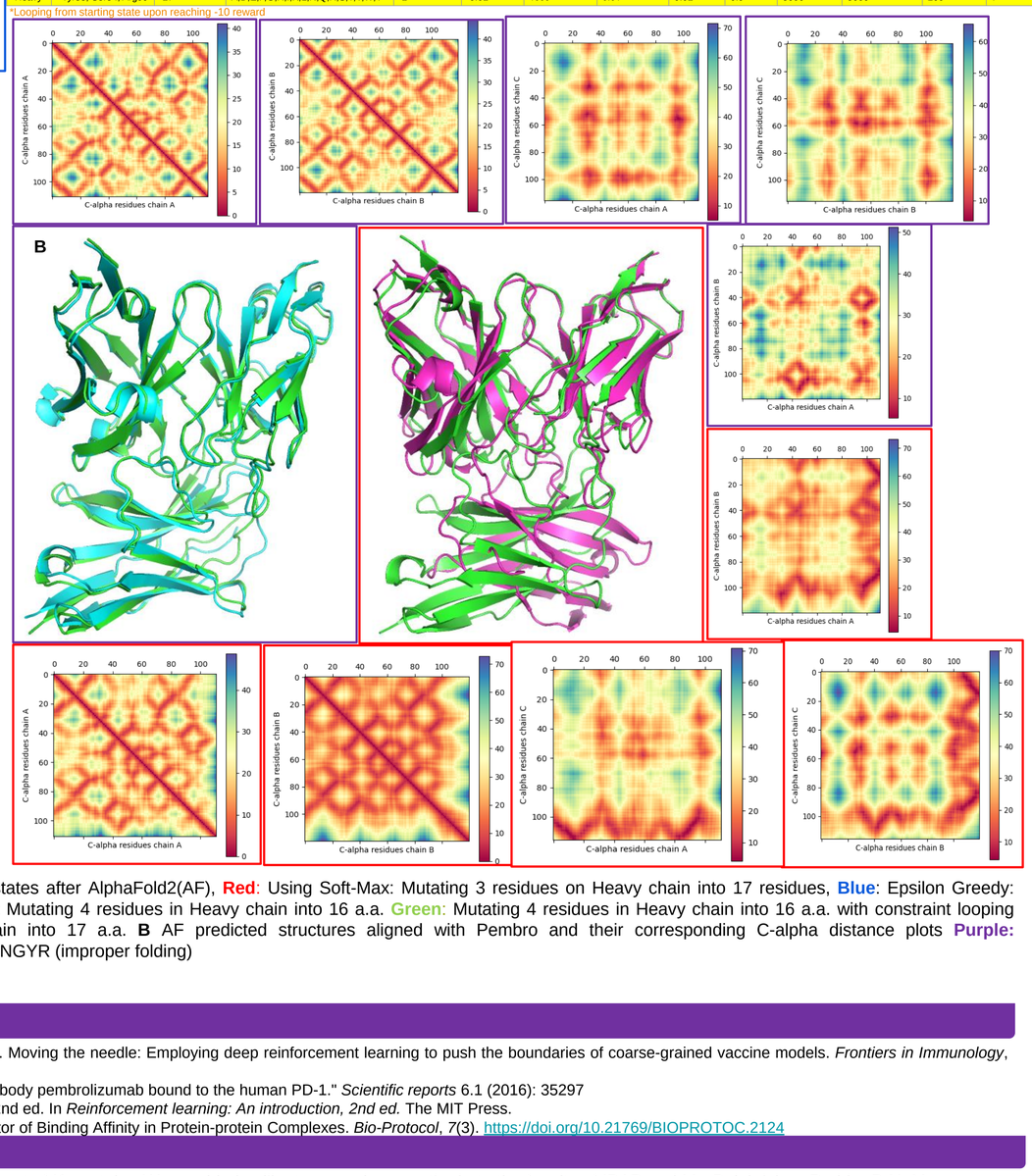


Figure 5. A a RL SHM simulations, **A b** Higher Binding affinity(BA) states after AlphaFold2(AF), **Red**: Using Soft-Max: Mutating 3 residues on Heavy chain into 17 residues, **Blue**: Epsilon Greedy: Mutating 6 residues in Light chain into 6 a.a. Orange: Epsilon greedy: Mutating 4 residues in Heavy chain into 16 a.a. **Green**: Mutating 4 residues in Heavy chain into 16 a.a. with constraint looping Yellow: DQL: Epsilon Greedy: mutating 3 residues on Heavy chain into 17 a.a. **B** AF predicted structures aligned with Pembro and their corresponding C-alpha distance plots **Purple:** DDDSGDTYYNSNTNRYR (Folded Properly) **Red:** SYYYSDTWYNYNTNGYR (improper folding)

REFERENCES

- Faris, J. G., Orbidan, D., Wells, C., Petersen, B. K., & Sprenger, K. G. (2022). Moving the needle: Employing deep reinforcement learning to push the boundaries of coarse-grained vaccine models. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/FIMMU.2022.1029167
- Horita, Shoichiro, et al. "High-resolution crystal structure of the therapeutic antibody pembrolizumab bound to the human PD-1." *Scientific reports* 6.1 (2016): 35297
- Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction, 2nd ed. In *Reinforcement learning: An introduction, 2nd ed.* The MIT Press.
- Vangone, A., & Bonvin, A. M. J. J. (2017). PRODIGY: A Contact-based Predictor of Binding Affinity in Protein-protein Complexes. *Bio-Protocol*, *7*(3). https://doi.org/10.21769/BIOPROTOC.2124

ACKNOWLEDGEMENT

This study is supported by Dept. of electronics, IT, BT and S&T, Government of Karnataka. We would also like to extend our gratitude to the faculty of IBAB and Mr. Balakrishna Prabhu B N, Ms. Namita Menon, Mr. Yash Chindarkar and Ms. Apoorva Ganesh for their contribution and guidance.

Modelling Somatic Hypermutations with Reinforcement Learning for PD1 and Pembrolizumab

Anmol Singh, *Prof. Subhashini Srinivasan, Dr. Nithya Ramakrishnan*

Institute of Bioinformatics and Applied Biotechnology - Biotech Park, Electronics City Phase 1, Electronic City, Bengaluru, Karnataka 560100

ABSTRACT

We present a reinforcement learning (RL) model for Somatic Hypermutations (SHM), which mimics the natural selection process in a very short time scale. In this model, the agent can learn to preferentially mutate amino acids in the antibody, leading to affinity maturation. The model thus predicts a higher binding affinity antibody than the initial antibody-antigen complex. We have used the Pembrolizumab-PD1 (5b8c) complex to create the model, as Pembrolizumab (Pembro) is widely used in immunotherapy.

We use Q-Learning in RL to model SHM on a reduced state space to provide better binding affinity antibodies. We validated the structure of the antibodies predicted by the RL model using AlphaFold2 and inter-residue distance plots to check for proper folding of chains and protein-protein interactions. This study provides a proof of concept that RL can be used for modeling the biological process of SHM and can further be employed for creating novel antibodies.

OBJECTIVES

Major Objective: To create a reinforcement learning model for somatic hypermutations using Pembro and PD1.

Minor Objective: To find states/antibodies which have better binding affinity than Pembro, and validating the good states using AlphaFold2 and C-alpha distance plots of the predicted structures.

INTRODUCTION

- Somatic Hypermutations (SHM) are an important part of the affinity maturation (AM) in B-cells.
- We express the biological process of AM as a Markov decision process (MDP), creating a RL model of SHM.

Model Assumptions:

- The relative configuration of PD1-Pembro complex remains the same irrespective of the type of amino acids mutations at the 17 positions
- The 3D folds of light and heavy chains are not disrupted by the mutations in loops on Pembro

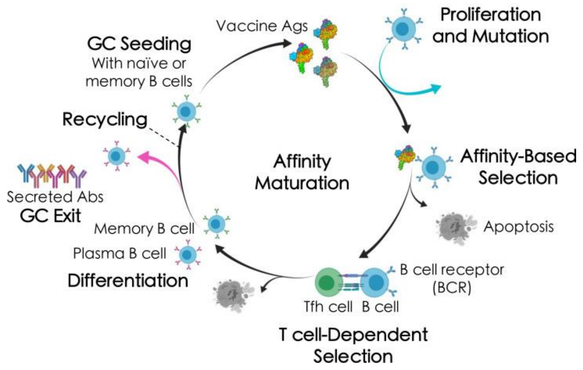


Figure 1. Broad overview of the affinity maturation (AM) process by which antibodies (Abs) evolve against vaccine-candidate antigens (Ags) in a germinal center (GC) reaction (Faris J et al., 2022)

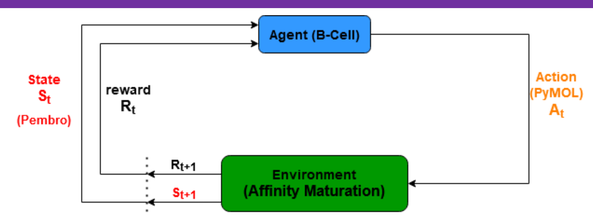
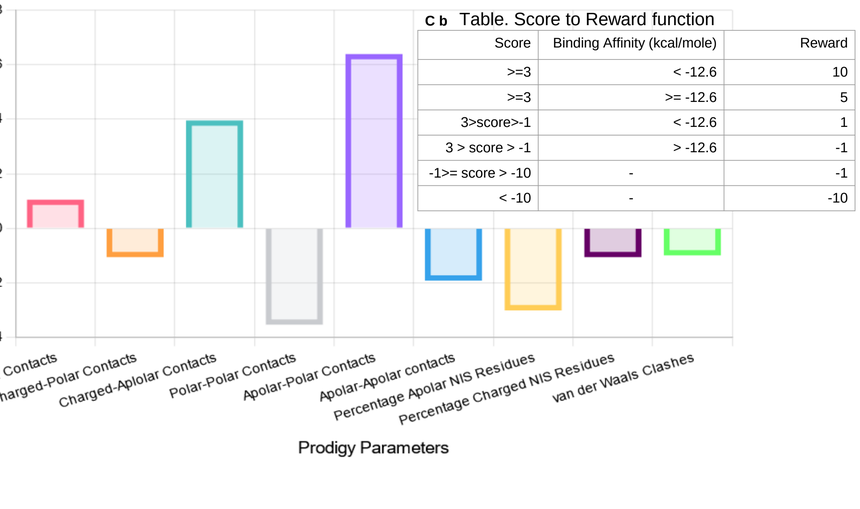


Figure 2. The interaction of agent and environment in a MDP ***Agent***: B-cells ***Action***: Point mutations on Pembrolizumab (using PyMOL) ***Environment***: Affinity Maturation of Pembrolizumab-PD1 ***State***: Pembro with mutated amino acids



Chain A : Pembro light chain variable region
Chain B : Pembro heavy chain variable region
Chain C : Programmed cell death protein 1 (PD1-Antigen)

Figure 3. A Reference crystal structure of Pembrolizumab (PDB ID: **5B8C**), ***B*** A schematic diagram of interactions between PD1 and antiPD1. Direct protein/protein hydrogen bonds are in blue, water-mediated hydrogen bonds are in green and salt bridges are in red (Horita S et al., 2016)

DESIGN AND WORKFLOW

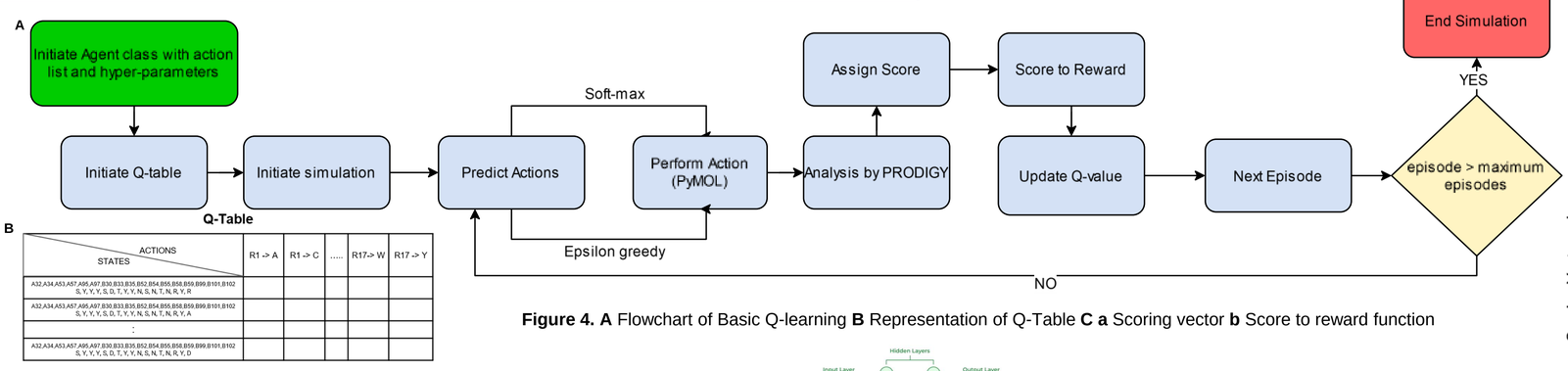


Figure 4. A Flowchart of Basic Q-learning **B** Representation of Q-Table **C a** Scoring vector **b** Score to reward function

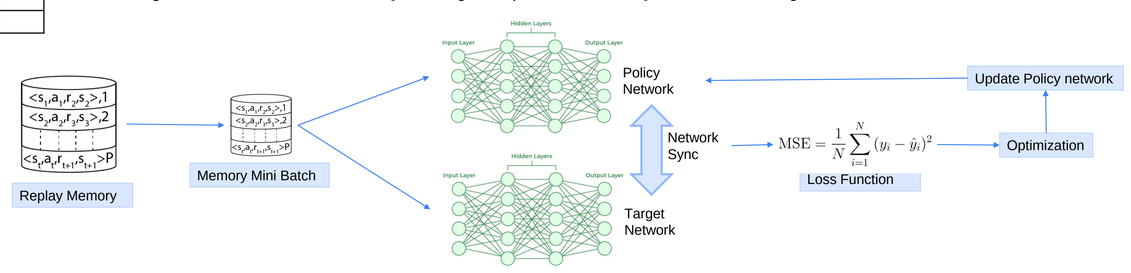


Figure 5. General Representation of Deep Q-Learning

C b Table. Score to Reward function

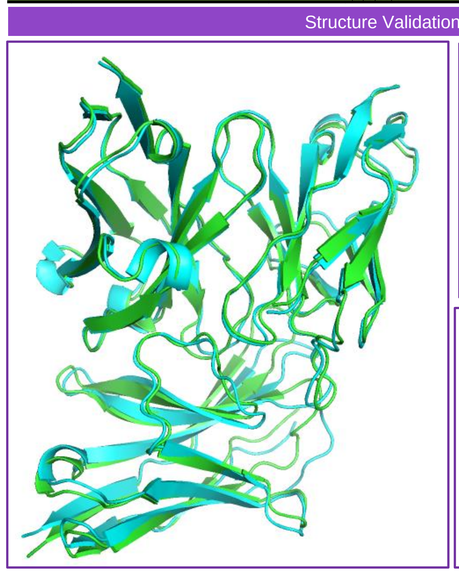
Score	Binding Affinity (kcal/mole)	Reward
>=3	< -12.6	10
>=3	>= -12.6	5
3>score>-1	< -12.6	1
3 > score > -1	> -12.6	-1
-1>= score > -10	-	-1
< -10	-	-10

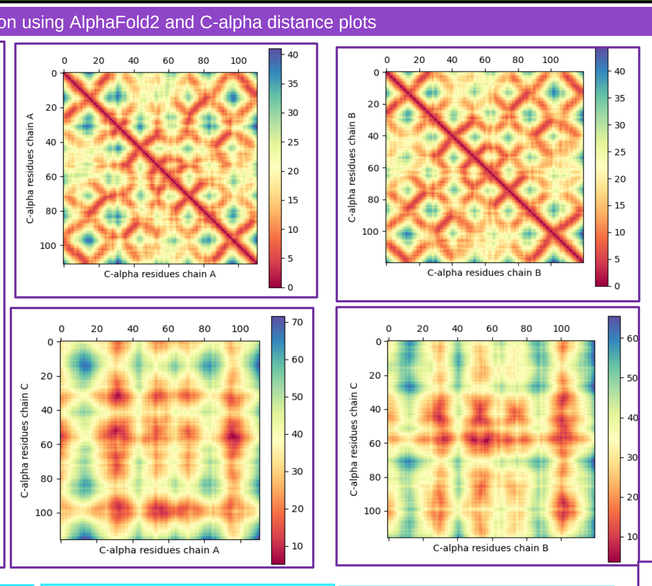
RESULTS

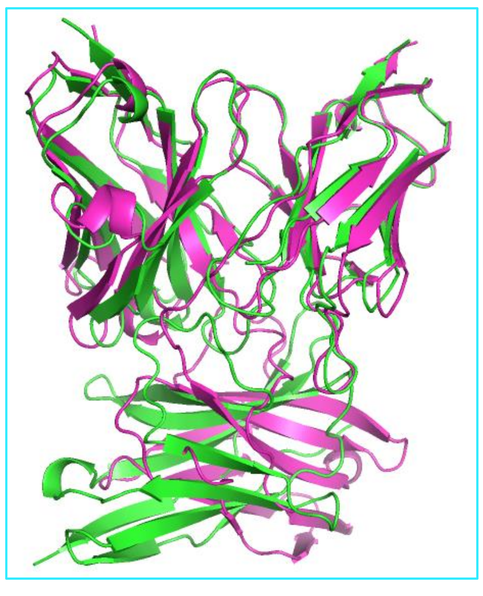
Table. Parameters of Simulations

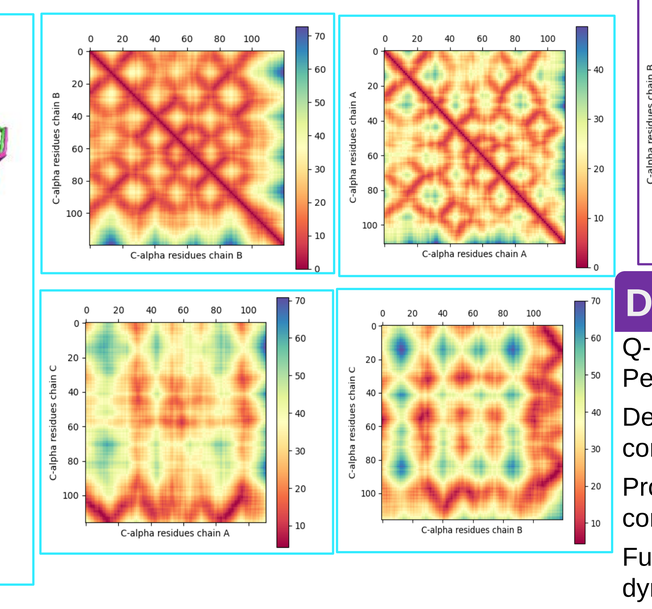
Chain	Residues position	No. of Mutation a.a.	Mutations Residues	Temperature	Temperature decay	Temperature decay interval	Minimum Temperature	Learning rate	Discount factor	Random Seed
Q-Learning										
Soft-max										
Heavy	Tyr33, Ser54, Arg99	17	A,D,E,F,G,H,I,K,L,N,Q,R,S,T,V,W,Y	100	0.5	1000	1	0.01	0.9	717
Epsilon Decay										
Chain	Residues position	No. of Mutation a.a.	Mutations Residues	Exploration rate	Exploration decay	Exploration decay interval	Minimum Exploration	Learning rate	Discount factor	Random Seed
Light	Ser32, Tyr34, Tyr53, Tyr57, Ser95, Asp97	6	S,Y,N,R,T,D	1	0.005	1000	0.01	0.01	0.9	643
Heavy	Ser54, Arg99,Tyr101, Arg102	16	A,D,E,F,G,H,I,K,L,N,Q,R,S,T,V,W,Y	1	0.005	1000	0.05	0.01	0.9	71
Heavy*	Ser54, Arg99,Tyr101, Arg102	16	A,D,E,F,G,H,I,K,L,N,Q,R,S,T,V,W,Y	1	0.005	1000	0.05	0.01	0.9	17

Structure Validation using AlphaFold2 and C-alpha distance plots

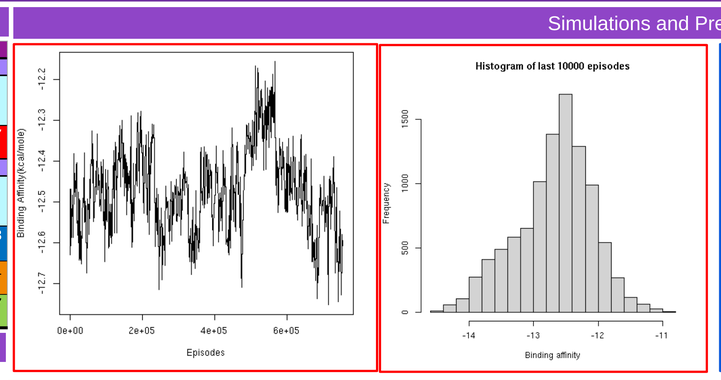




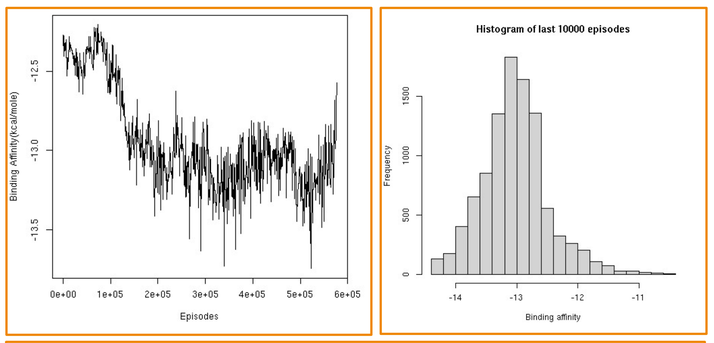




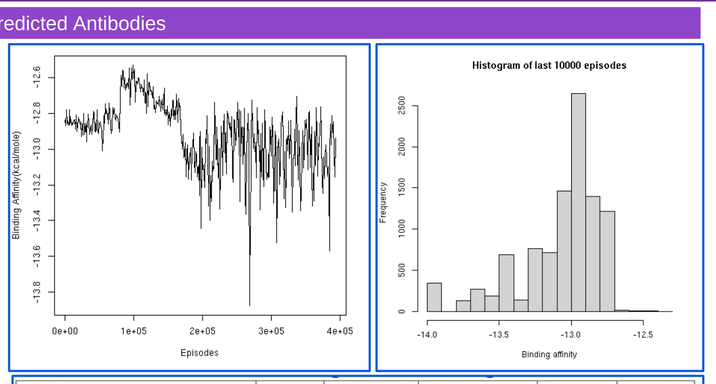
Simulations and Predicted Antibodies



State	Freq	Reward	Crystal BA	AF BA	RMSD
DRRTSYTYYNSNTNRYR	65	10	-13.4	-23.3	19.48
STRNYYTYYNSNTNRYR	45	10	-13.6	-13.5	2.061
YRNNSRTYYNSNTNRYR	141	10	-13.4	-13.4	2.058
DSRNYYTYYNSNTNRYR	41	10	-13.7	-13.3	2.072
DTRNYYTYYNSNTNRYR	89	10	-13.7	-13.3	1.987
YRNNNRTYYNSNTNRYR	101	10	-13.3	-13.3	2.072

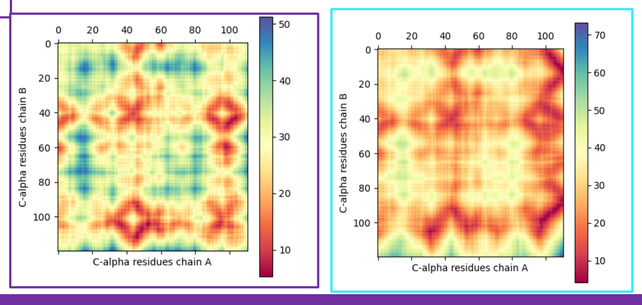


State	Freq	Reward	Crystal BA	AF BA	RMSD
SYYYSDTYYNWNTNTHL	60	1	-12.9	-13.5	2.195
SYYYSDTYYNYNTNIFN	53	1	-13.0	-13.5	2.172
SYYYSDTYYNYNTNWAK	143	1	-13.0	-13.4	2.101
SYYYSDTYYNFNTNAWY	45	10	-14.0	-13.3	2.217
SYYYSDTYYNFNTNIYY	39	10	-14.1	-13.0	2.185



*Looping from starting state upon reaching -10 reward, simulating apoptosis

State	Freq	Reward	Crystal BA	AF BA	RMSD
SYYYSDTYYNFNTNRYR	5	1	-13.4	-13.1	2.179
SYYYSDTYYNLNTNRYR	6	1	-13.5	-13	2.153
SYYYSDTYYNWNTNRYR	6	1	-13.4	-13	2.286
SYYYSDTYYNYNTNRYR	9	1	-13.6	-12.8	2.193



DISCUSSION AND FUTURE WORK

Q-Learning simulated SHM and provided a better binding Pembro-PD1 complex using a reduced state space.

Deep Q-Learning can be employed to predict alternate complexes to Pembro-PD1 – Currently underway.

Protein Language Models (PLM) can be employed in conjunction with our SHM-RL model for drug discovery.

Further validation of predicted antibodies using molecular dynamics and different binding affinity tools is required.

ACKNOWLEDGEMENT

This study is supported by Dept. of electronics, IT, BT and S&T, Government of Karnataka. We would also like to extend our gratitude to the faculty of IBAB and Mr. Balakrishna Prabhu B N, Ms. Namita Menon, Mr. Yash Chindarkar and Ms. Apoorva Ganesh for their contribution and guidance.

REFERENCES

- Faris, J. G., Orbidan, D., Wells, C., Petersen, B. K., & Sprenger, K. G. (2022). Moving the needle: Employing deep reinforcement learning to push the boundaries of coarse-grained vaccine models. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/FIMMU.2022.1029167
- Horita, Shoichiro, et al. "High-resolution crystal structure of the therapeutic antibody pembrolizumab bound to the human PD-1." *Scientific reports* 6.1 (2016): 35297
- Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction, 2nd ed. In *Reinforcement learning: An introduction, 2nd ed.* The MIT Press.
- Vangone, A., & Bonvin, A. M. J. J. (2017). PRODIGY: A Contact-based Predictor of Binding Affinity in Protein-protein Complexes. *Bio-Protocol*, *7*(3). https://doi.org/10.21769/BIOPROTOC.2124