Modelling Somatic Hypermutations with Reinforcement Learning for PD1 and Pembrolizumab

Anmol Singh, Subhashini Srinivasan, Nithya Ramakrishnan
Institute of Bioinformatics and Applied Biotechnology - Biotech Park, Electronics City Phase 1, Electronic City, Bengaluru, Karnataka 560100

IBAB Institute of Bioinformatics and Applied Biotechnology

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 C-alpha 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

- Affinity maturation (AM) in B-cells occurs via Somatic Hypermutations (SHM).
- 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

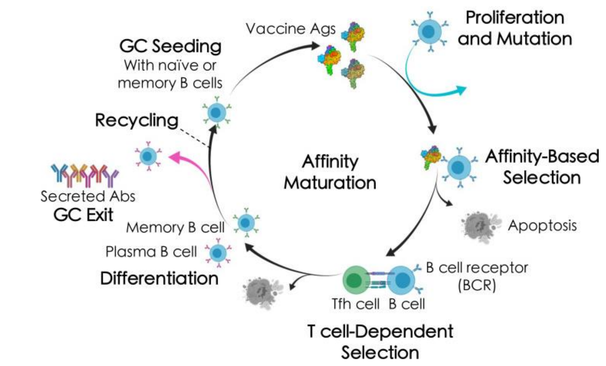


Figure 1. Broad overview of the affinity maturation (AM) process by which antibodies (Abs) evolve against vaccine-candidate antigens (Ags) in a germinal center (GC) reaction (Faris J et al., 2022)

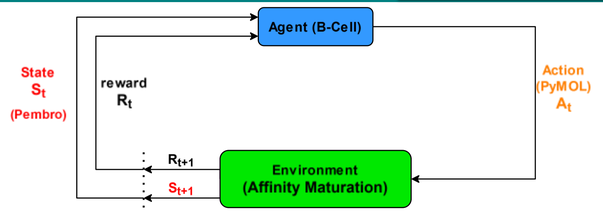
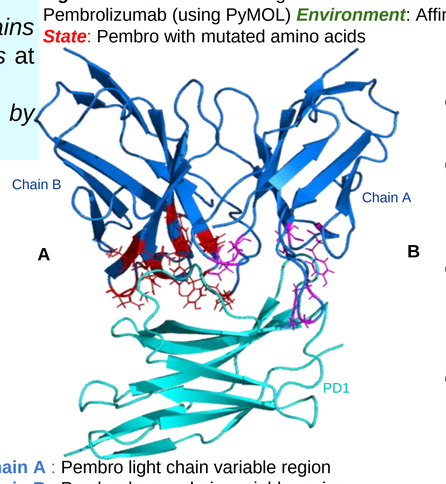
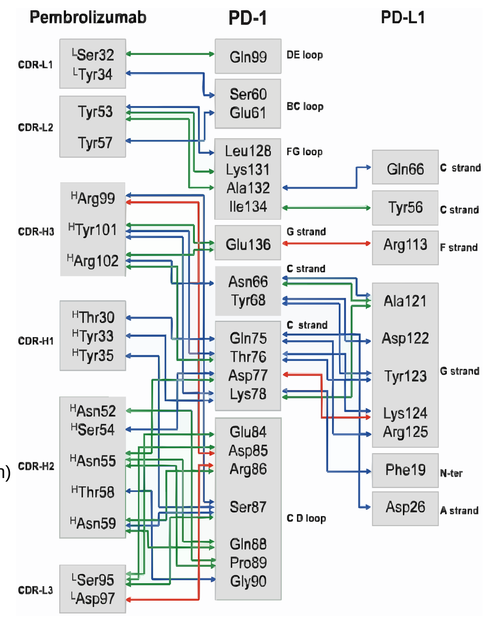


Figure 2. The interaction of agent and environment in a MDP ***Agent***: B-cells ***Action***: Point mutations on Pembrolizumab (using PyMOL) ***Environment***: Affinity Maturation of Pembrolizumab-PD1 ***State***: Pembro with mutated amino acids



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

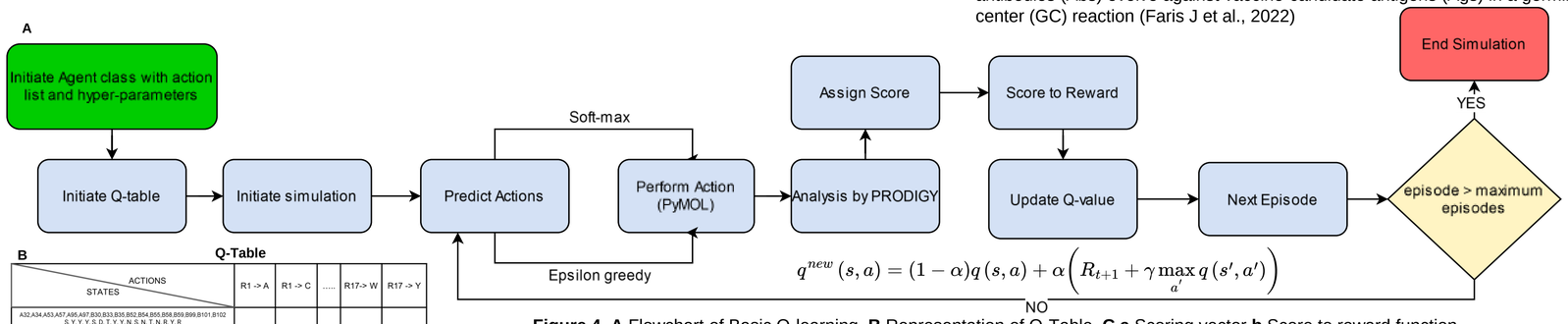
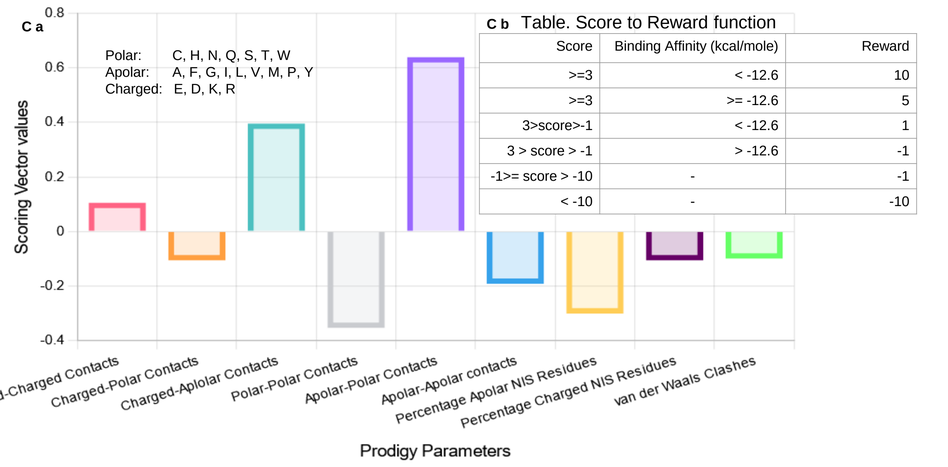


Figure 4. A Flowchart of Basic Q-learning, **B** Representation of Q-Table, **C a** Scoring vector **b** Score to reward function



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

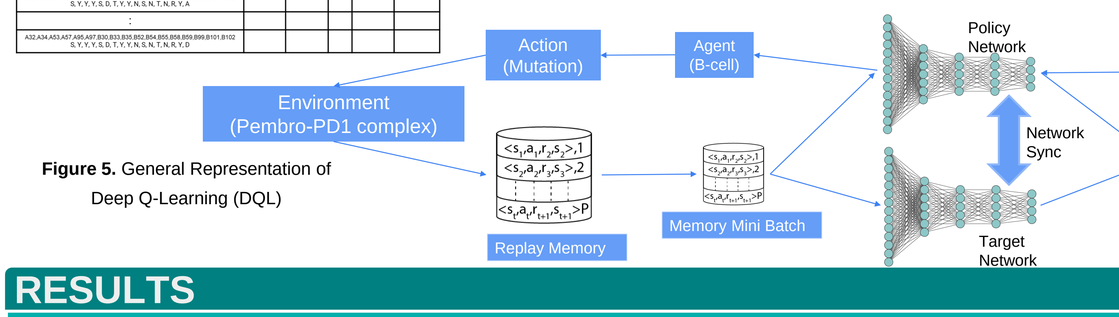


Figure 5. General Representation of Deep Q-Learning (DQL)

RESULTS

A Parameters of Simulations

Q-Learning

Epsilon Decay

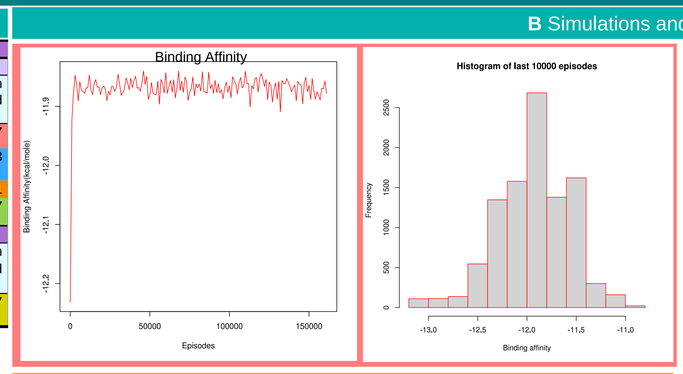
Chain	Residues position	No. of Mutation a.a.	Mutations Residues	Exploration rate	Exploration decay	Exploration decay interval	Minimum Exploration	Learning rate	Discount factor	Random seed
Both	All	20	All essential a.a.	1	0.01	1000	0.05	0.01	0.9	17
Light	Ser32, Tyr34, Tyr53, Tyr57, Ser95, Asp97	6	S,Y,N,R,T,D	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,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,R,S,T,V,W,Y	1	0.005	1000	0.05	0.01	0.9	17

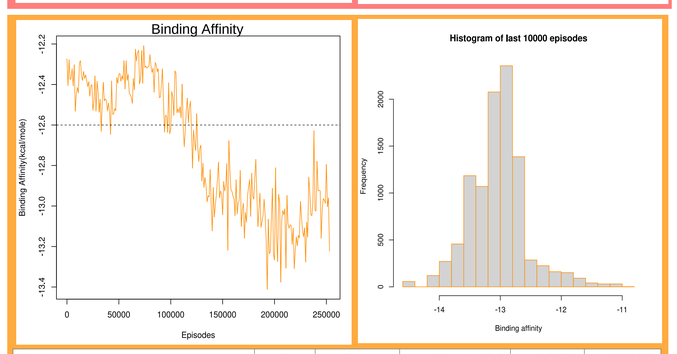
Deep Q Learning

Chain	Residues position	No. of Mutation a.a.	Mutations Residues	Exploration rate (epsilon)	Epsilon decay	Epsilon decay interval	Minimum Exploration	Learning rate	Discount factor	Network Sync Rate	Replay Memory Size	Mini Batch Size	Random seed
Heavy	Ser54, Arg99, Tyr101, Arg102	16	A,D,E,F,G,H,I,K,L,N,R,S,T,V,W,Y	1	0.05	1000	0.05	0.01	0.9	2000	5000	200	17

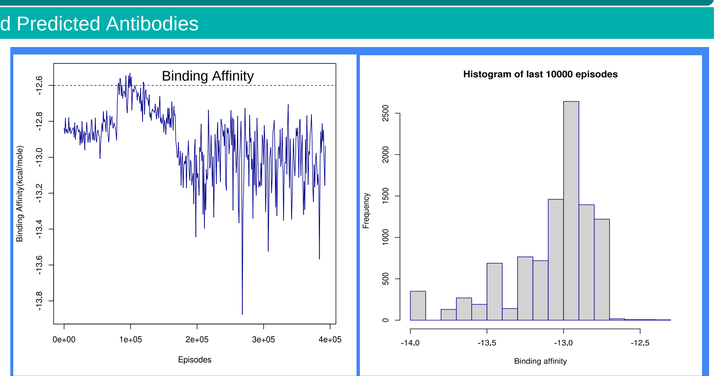
Figure 6 A Parameter Table, **B** Simulations and AlphaFold2 (AF) Predicted structures with Binding Affinity(BA) and RMSD. The parameters for the simulation. The simulations and the higher affinity states from them are shown in the Simulation section corresponding to the colour in the Parameter table as Red, Blue, Orange and Green(*Apoptosis loop). The DQL simulation (Yellow) is mentioned in the Discussion and Future work section.

B Simulations and Predicted Antibodies

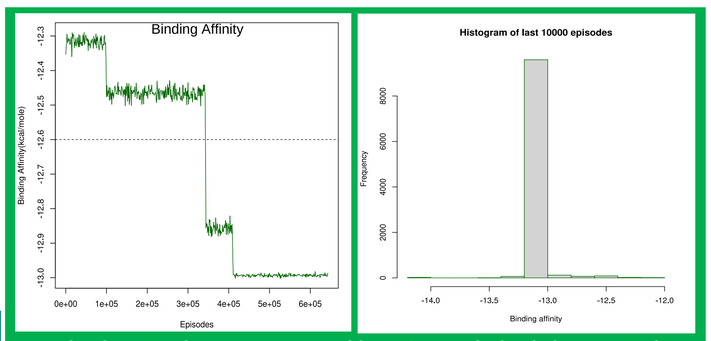




State	Freq	Reward	Crystal BA	AF BA	RMSD
SYYYSDTYYNYNTNEWE	34	1	-13.7	-13.5	2.195
SYYYSDTYYNYNTNNWE	34	1	-12.9	-13.5	2.172
SYYYSDTYYNLNTNAWW	77	1	-12.7	-13.4	2.101
SYYYSDTYYNFNTNAWY	45	10	-14.0	-13.3	2.217
SYYYSDTYYNFNTNIYY	39	10	-14.1	-13.0	2.185



State	Freq	Reward	Crystal BA	AF BA	RMSD
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



*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

Structure Validation using AlphaFold2 and C-alpha distance plots

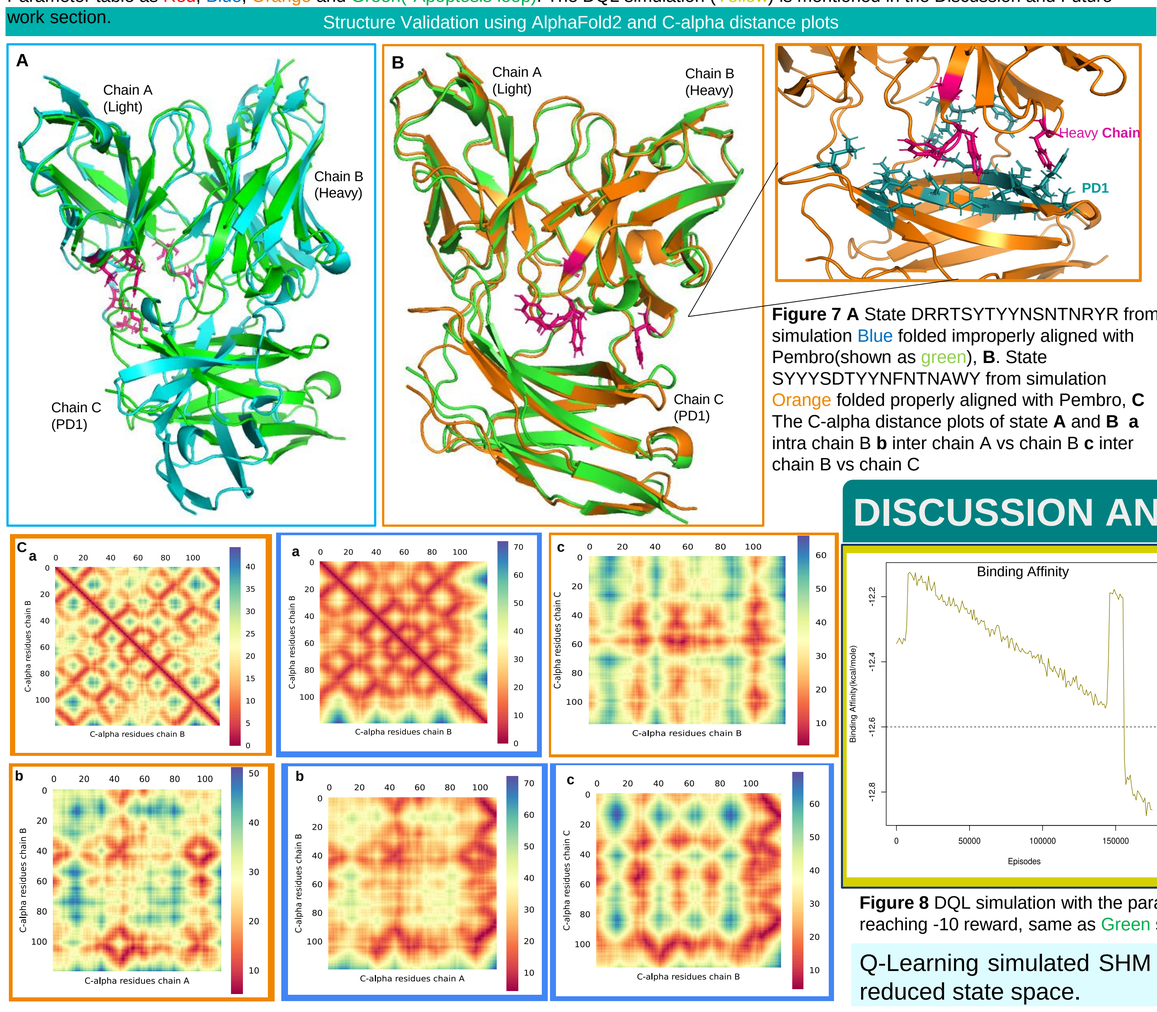


Figure 7 A State DRRTSYTYYNSNTNRYR from simulation Blue folded improperly aligned with Pembro(shown as green), **B**. State SYYYSDTYYNFNTNAWY from simulation Orange folded properly aligned with Pembro, **C** The C-alpha distance plots of state **A** and **B** **a** intra chain B **b** inter chain A vs chain B **c** inter chain B vs chain C

DISCUSSION AND FUTURE WORK

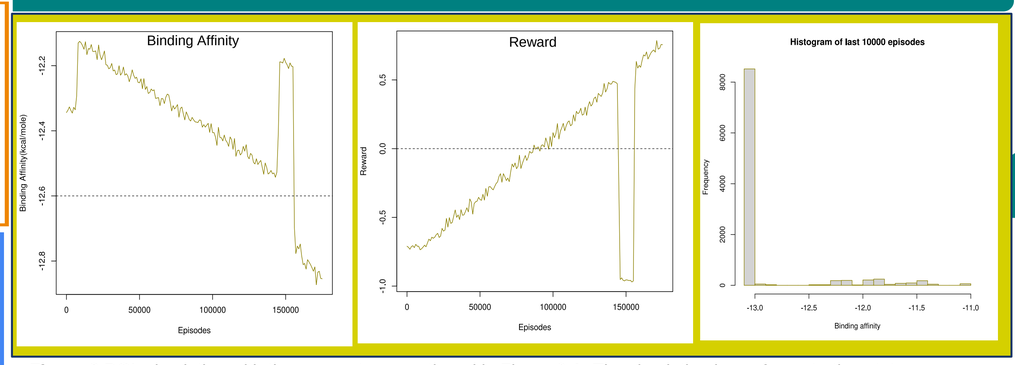


Figure 8 DQL simulation with the parameters mentioned in Figure 6 A. The simulation loops from starting state upon reaching -10 reward, same as Green simulation mentioned in Figure 6 B

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