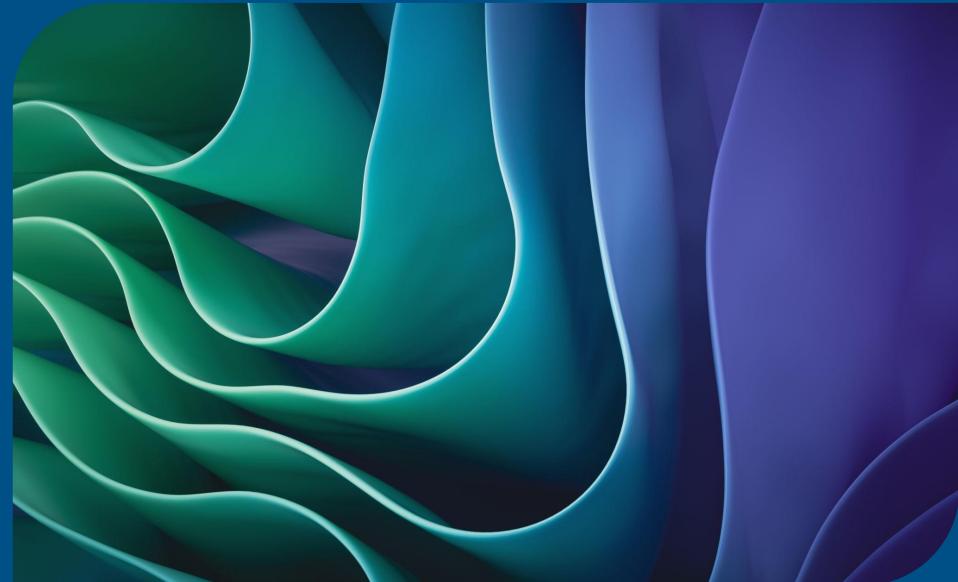


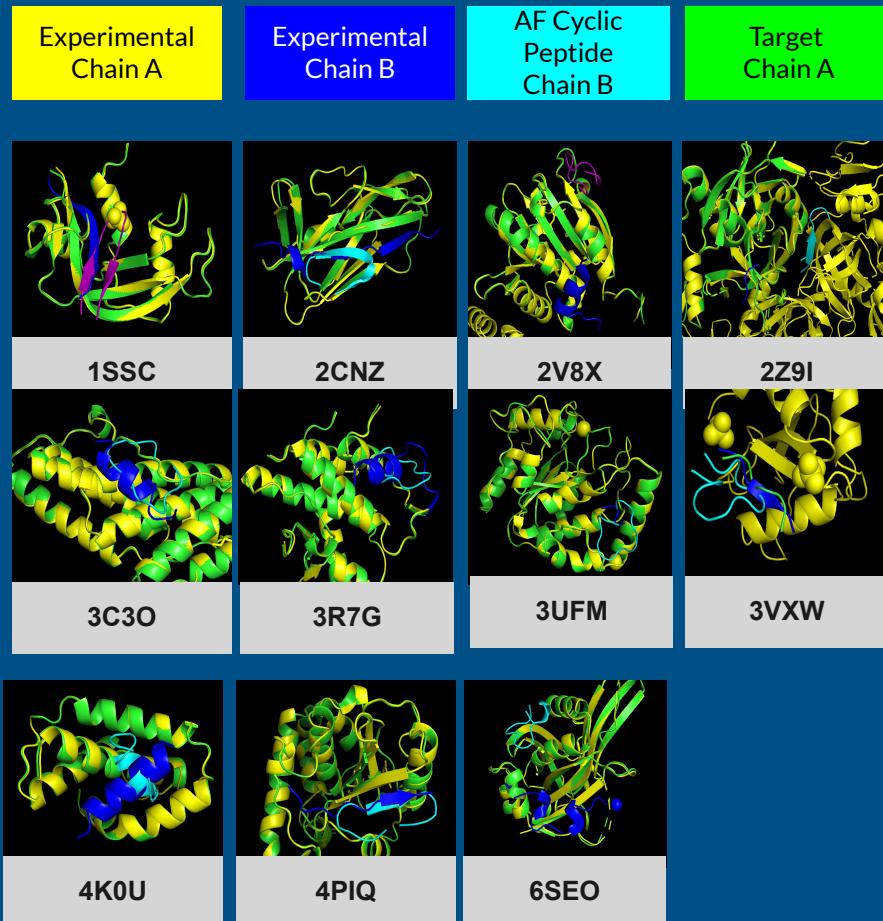
AlphaFold for Cyclic Peptide Design

Fall Independent Research
Jethro Cheuk Sau Au



1. Overview of Methodology & Approaches
2. Exploring peptide control with hotspots
3. Enhancing precision of peptide generations
4. Experimental Ligand Recreation
5. Discussion & Further Work

A quick recap from last term



A quick recap from last term

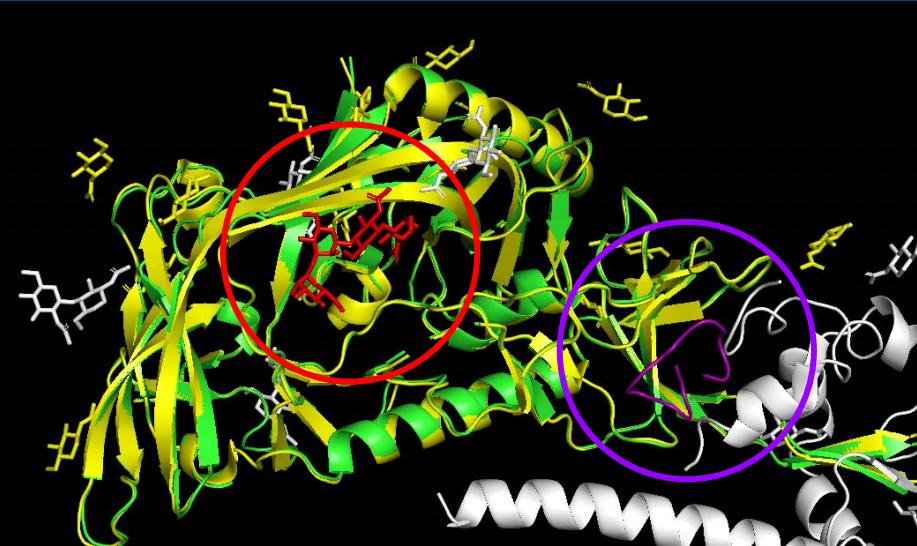
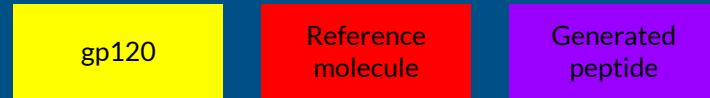
Challenges:

Distance too far from experimental molecule

No control over binding regions

Low confidence score in conformation

Unoptimized generation



How can we improve cyclic peptides generation for HIV gp120 trimer using AlphaFold ?

Evaluation - BMS-818251 molecule

6MU7

Crystal Structure of HIV-1 BG505 SOSIP.664 Prefusion Env Trimer Bound to Small Molecule HIV-1 Entry Inhibitor BMS-818251 in Complex with Human Antibodies 3H109L and 35O22 at 3.1 Angstrom

PDB DOI: <https://doi.org/10.22110/pdb6MU7/pdb>

Classification: IMMUNE SYSTEM/INHIBITOR

Organism(s): Human immunodeficiency virus 1, Homo sapiens

Expression System: Homo sapiens

Mutation(s): Yes

Deposited: 2018-10-22 Released: 2019-01-16

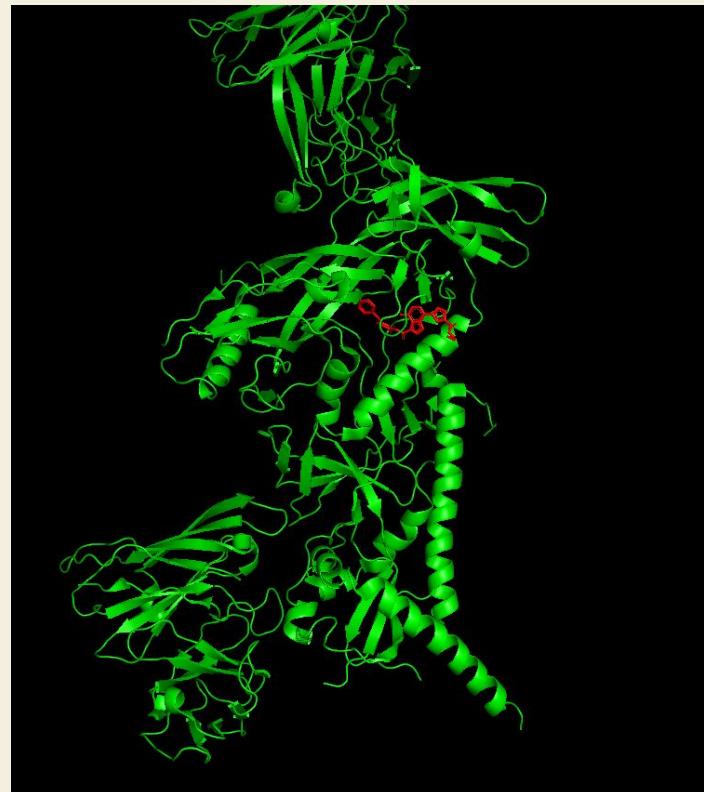
Deposition Author(s): Lai, Y.-T., Kwong, P.D.

Lattice engineering enables definition of molecular features allowing for potent small-molecule inhibition of HIV-1 entry

Yen-Ting Lai ¹, Tao Wang ², Sijy O'Dell ¹, Mark K Louder ¹, Arne Schön ³, Crystal S F Cheung ¹, Gwo-Yu Chuang ¹, Aliaksandr Druz ¹, Bob Lin ¹, Krisha McKee ¹, Dongjun Peng ¹, Yongping Yang ¹, Baoshan Zhang ¹, Alon Herschhorn ⁴ ⁵, Joseph Sodroski ⁴, Robert T Bailer ¹, Nicole A Doria-Rose ¹, John R Mascola ¹, David R Langley ⁶, Peter D Kwong ⁷

Affiliations + expand

PMID: 30604750 PMCID: PMC6318274 DOI: 10.1038/s41467-018-07851-1



1. Overview of Methodology & Approaches
2. Exploring peptide control with hotspots
3. Enhancing precision of peptide generations
4. Experimental Ligand Recreation
5. Discussion & Further Work

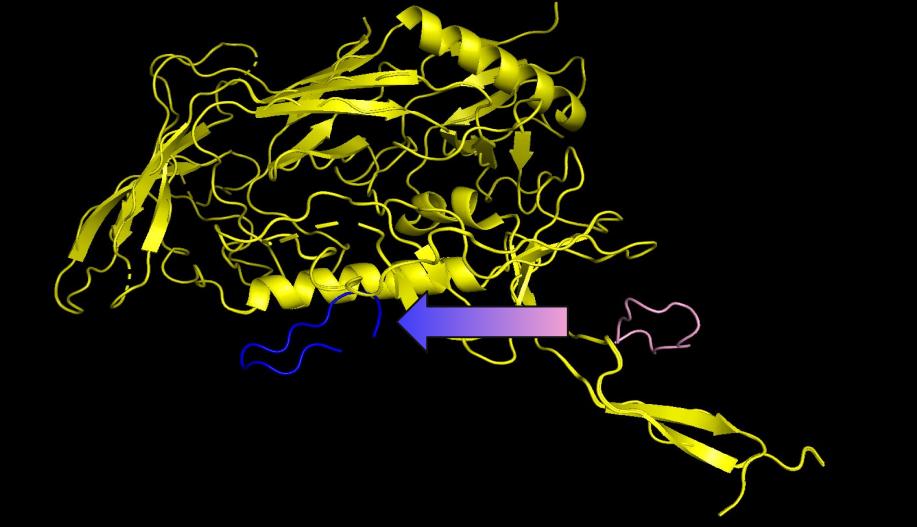
Hotspot Configuration

Where the cyclic peptides bind can be configured based on a variable set during the generative process.

This allows more fine-grained control of cyclic peptide generations.

- Default (no hotspot)
- Hotspots = “100-105”

Close-up view of shifting to peptide using hotspots



Experiment 1:

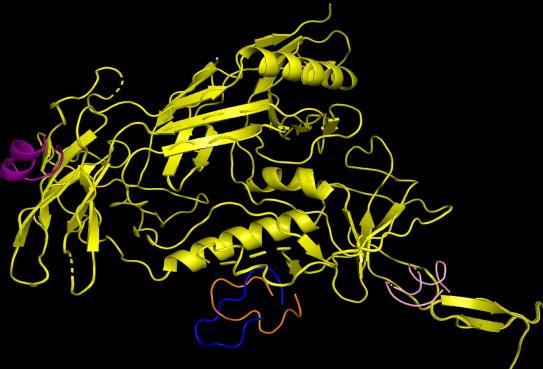
Evaluating how hotspots impact final peptide location

Observation 1: When the hotspot is located on the outer regions of the gp120 trimer, variations of cyclic peptides are consistent

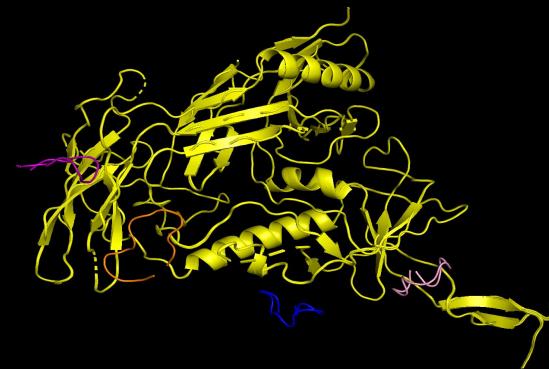
Observation 2: When the hotspot is located on the inner regions, the cyclic peptide binding region becomes unstable after generation

- Default (no hotspot)
- Hotspots = "100-105"
- Hotspots = "150-155"
- Hotspots = "200-205"

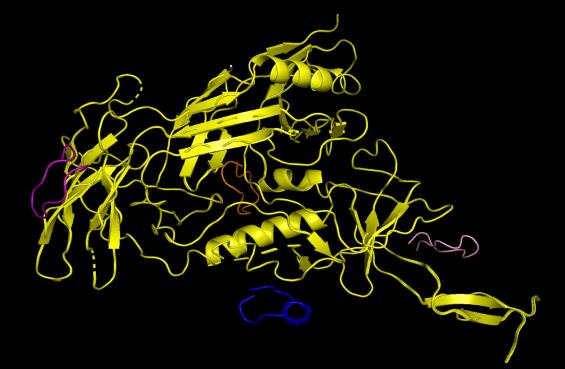
Seed=1



Seed=2

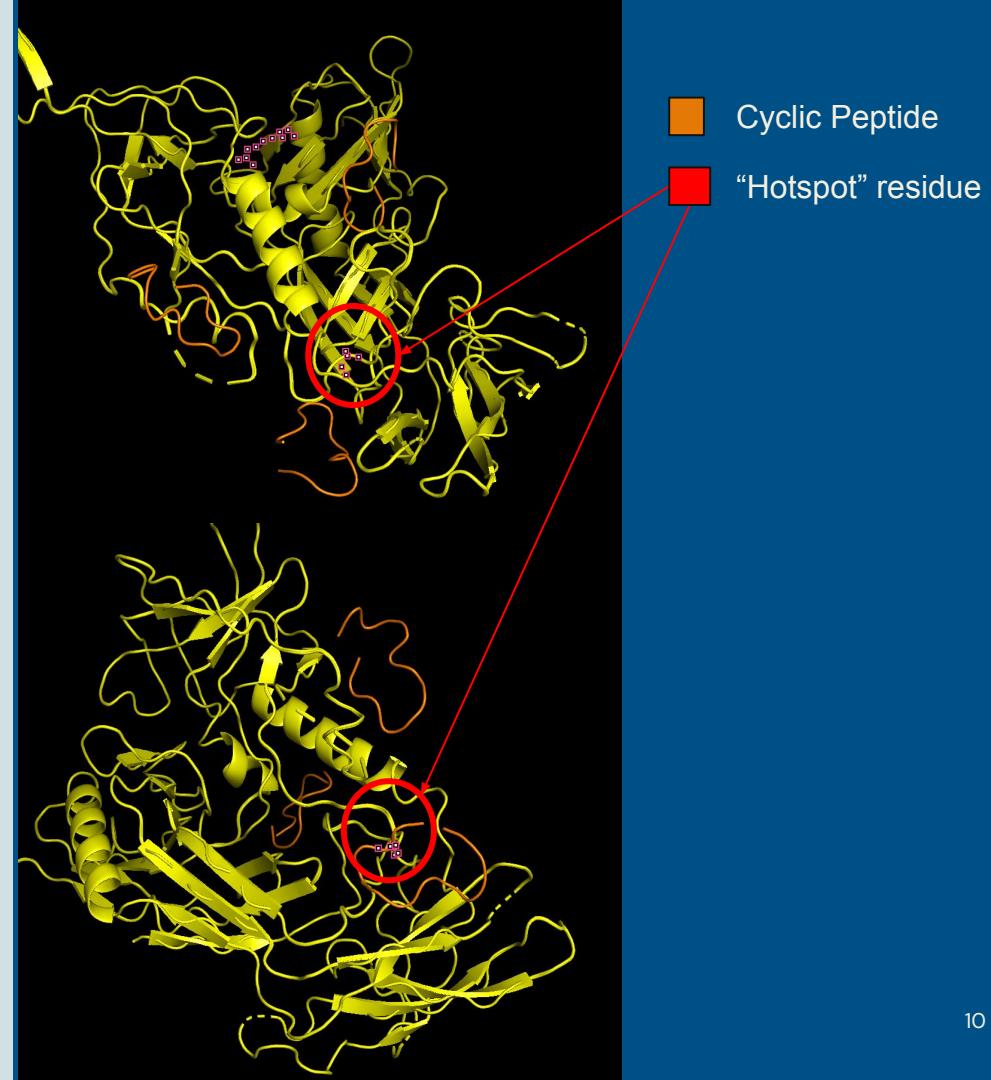


Seed=3



Closer look into hotspots in the inner regions

During generation, the current model is unable to identify conformation within the gp120 structure, and settles for an “easier” but ineffective location



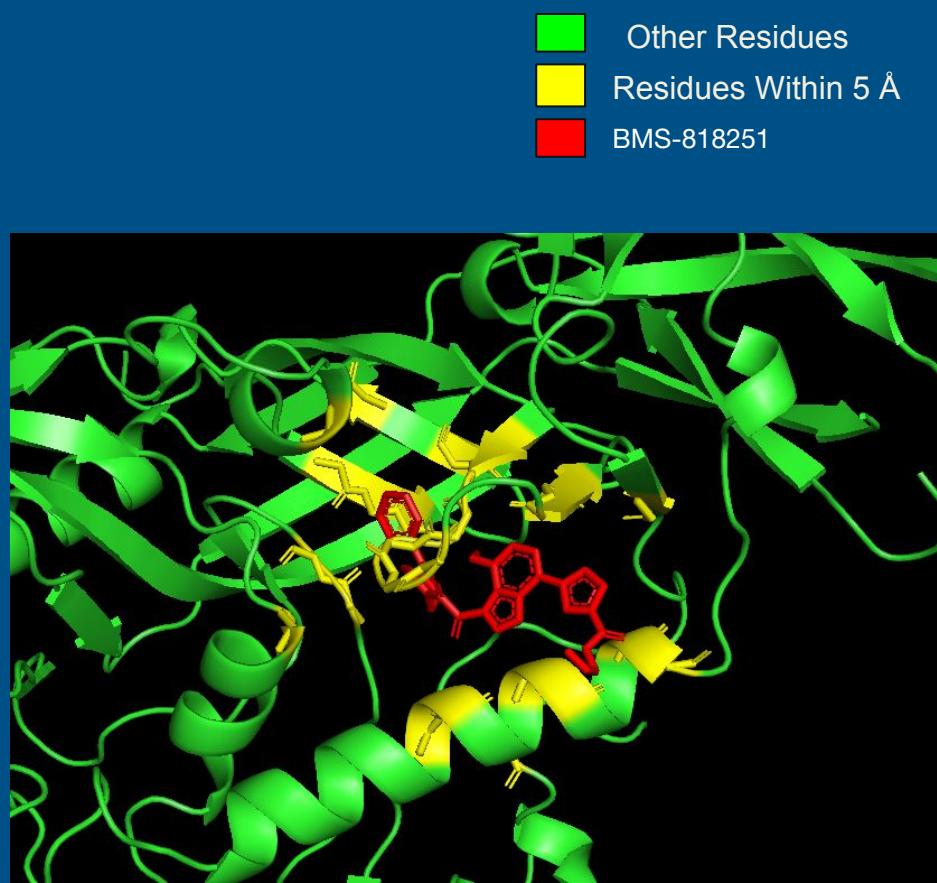
What if we generate cyclic peptides targeting the gp120 CD4 binding site?

Proximity Mapping of hotspot regions of CD4bs

Taking reference from experimental structure
from PDB 6MU7

Hotspots:

"69,108,109,112,113,116,117,121,202,255,256,25
7,370,375,376,377,382,384,424,425,426,427,4
28,429,432,433,434,475"

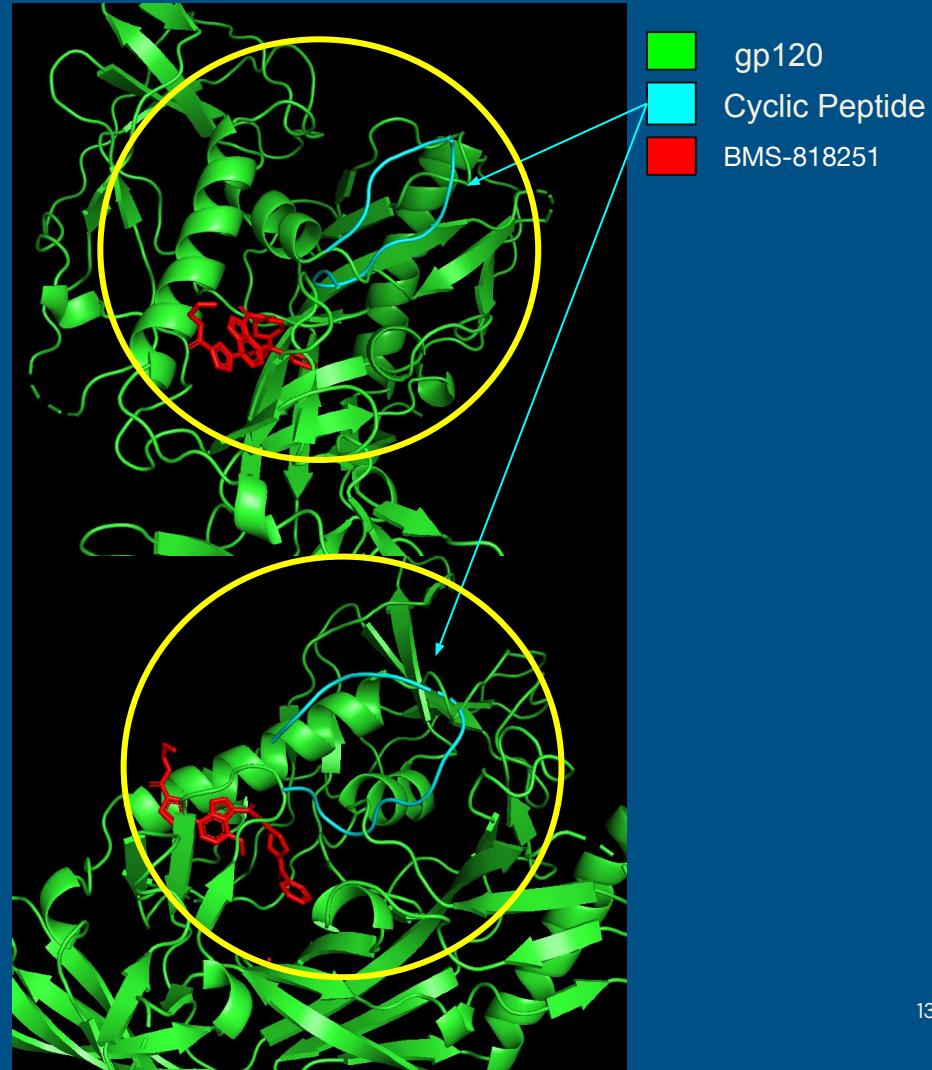


Generating cyclic peptides using hotspots

Generation of cyclic peptides was unstable due to the CD4 binding site being in an inner region of gp120

After visual inspection of results one was close but not within the CD4 binding site

- Cyclic peptide does not go into the binding pocket
- Similar to before, conforming to an “easy” option by remaining outside
- Default generation methods need to be changed

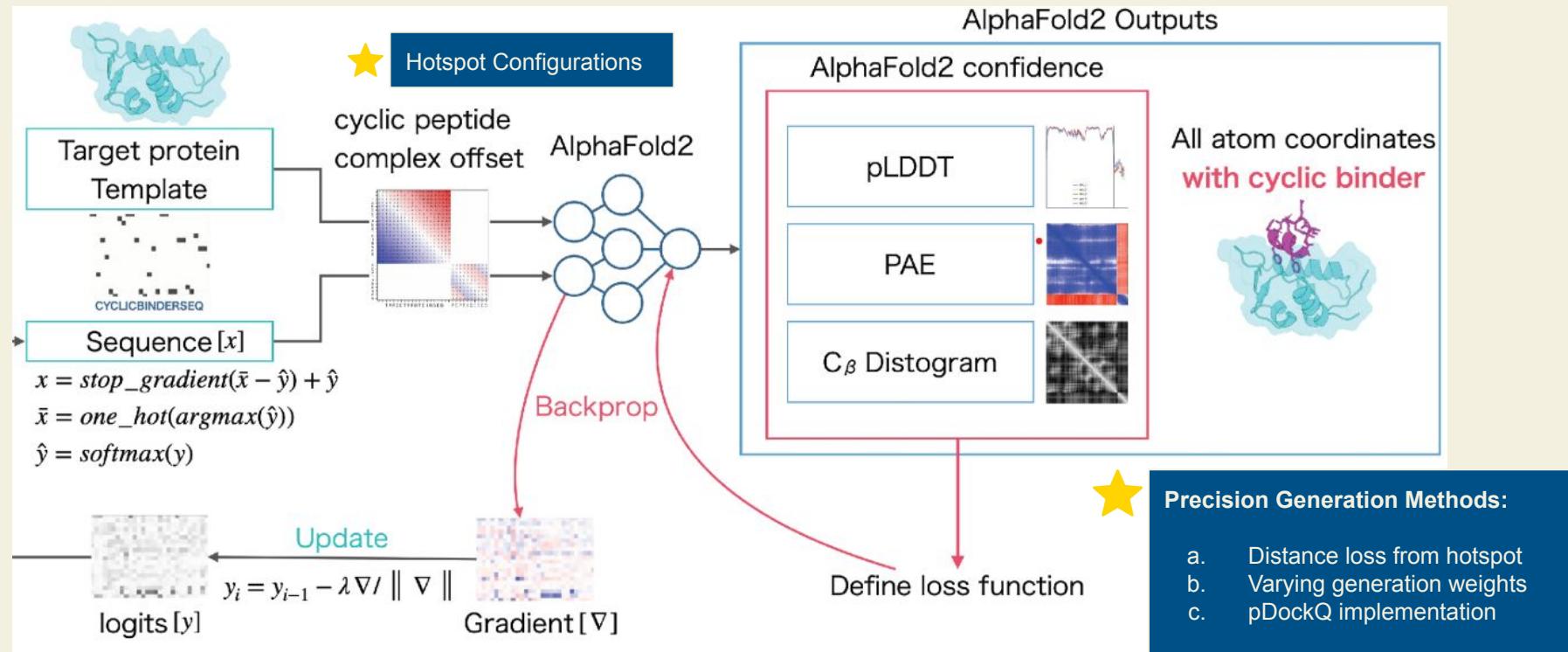


1. Overview of Methodology & Approaches
2. Exploring peptide control with hotspots
3. Enhancing precision of peptide generations
4. Experimental Ligand Recreation
5. Discussion & Further Work

How can I better generate cyclic peptides to be within the CD4 binding-site pocket?

Improve the loss function

Recall...

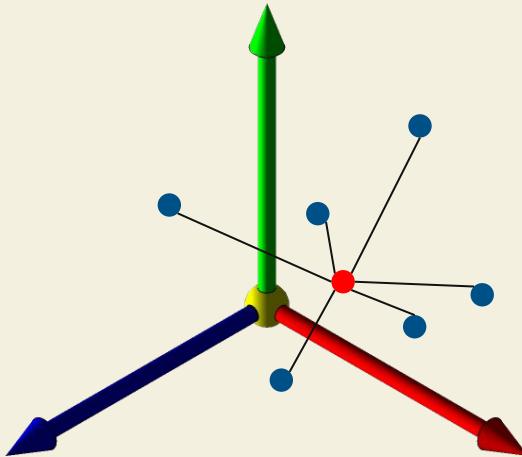


Adapted from: Kosugi T, Ohue M. Design of Cyclic Peptides Targeting Protein–Protein Interactions Using AlphaFold. *International Journal of Molecular Sciences*. 2023; 24(17):13257.

Experiment 2:

Developing better methods to enhance precision for cyclic peptides to CD4bs binding site

Centroid loss calculation



MSE Loss

1. Centroid calculation of cyclic peptide points
2. Calculation of a MSE loss from peptide centroid to all hotspot regions during each generation
3. Assign custom weight to MSE loss to enhance precision of generated peptide

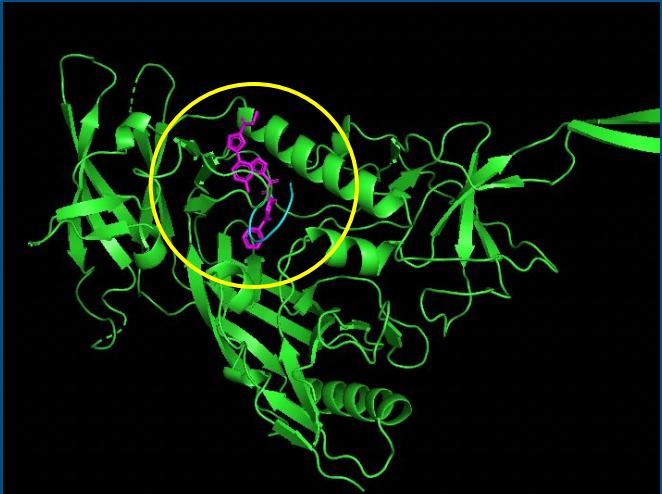
$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Experiment 2

By lowering binder length and add MSE loss, cyclic peptides were able to be generated in CD4bs

But the resulting generations had low confidence scores

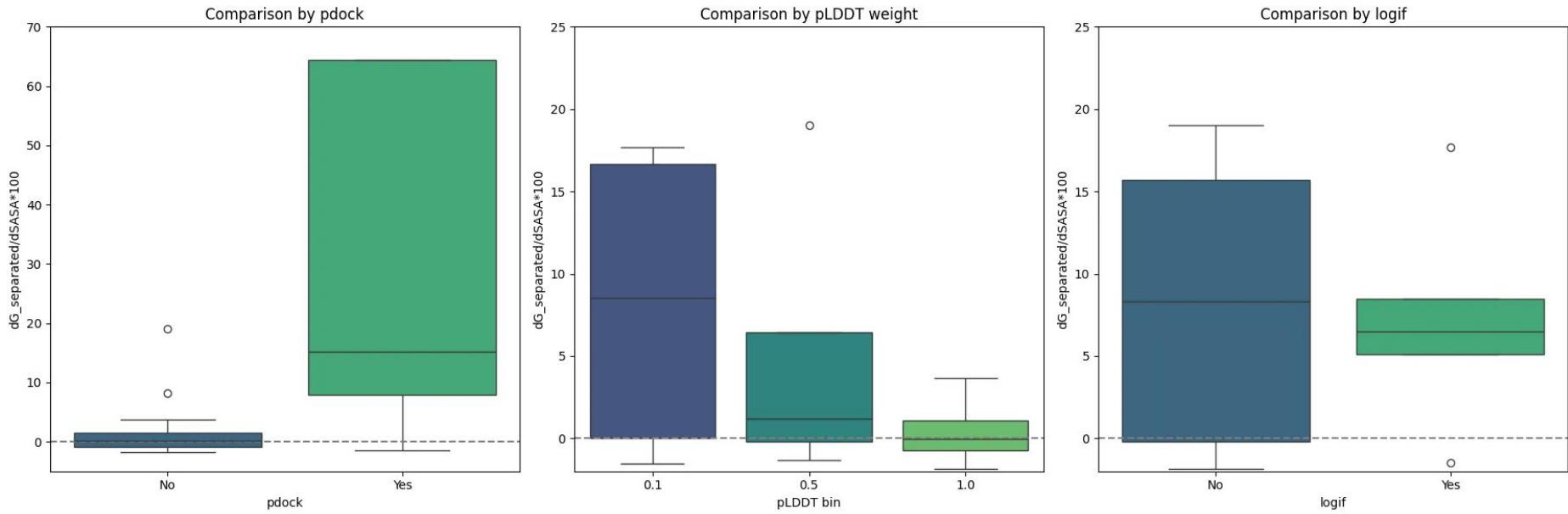
- Generated cyclic peptide
- BMS-818251



Experiment 3: How can we increase model confidence?

Variations in AlphaFold Model	Assigned Name	Model Loss Weights	Custom Loss Function
Baseline - No Change	Baseline	(i_con = 1.0,plddt=0.1)	-
Increasing weighting of per residue confidence - Variation #1	w1	(i_con = 1.0,plddt=0.5)	-
Increasing weighting of per residue confidence - Variation #2	w2	(i_con = 1.0,plddt=1.0)	-
Enhancing model with a predicted DockQ score - Variation #3	PDQw0	(i_con = 1.0,plddt=0.1, pdockq=0.1)	pDockQ
Enhancing model with a predicted DockQ score - Variation #4	PDQw1	(i_con = 1.0,plddt=0.1, pdockq=0.5)	pDockQ
Enhancing model with a predicted DockQ score - Variation #5	PDQw2	(i_con = 1.0,plddt=0.1, pdockq=1.0)	pDockQ
Enhancing model with a predicted DockQ score - Variation #6	logif	(i_con = 1.0,plddt=1.0, logif=1.0)	LogIF

Experiment 3: Increasing per-residue confidence works best



1. Overview of Methodology & Approaches
2. Exploring peptide control with hotspots
3. Enhancing precision of peptide generations
4. Experimental Ligand Recreation
5. Discussion & Further Work

Combining it all together

Generated a total of 168 cyclic peptides using the following model:

Changes to the default AlphaFold Model:

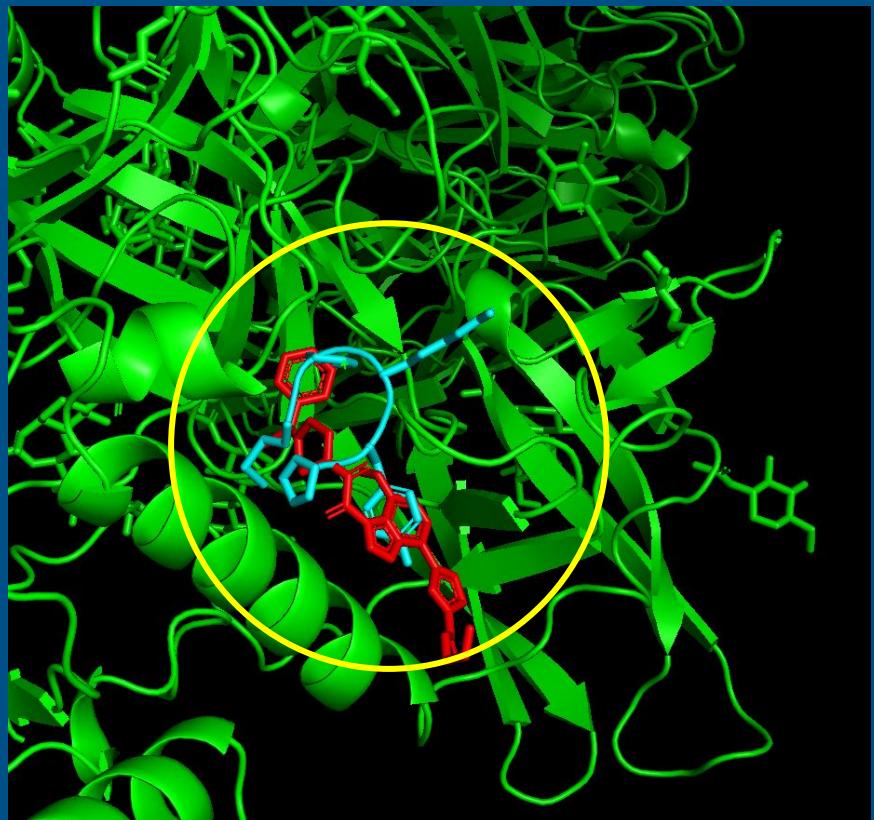
- CD4 binding-site hotspot proximity mapping
- A confidence-weighted centroid distance loss
- A boosted confident weighting from generation

Binder Length = 5

Sequence: PYRPM

Generated cyclic peptide

BMS-818251



Combining it all together

Generated a total of 168 cyclic peptides using the following model:

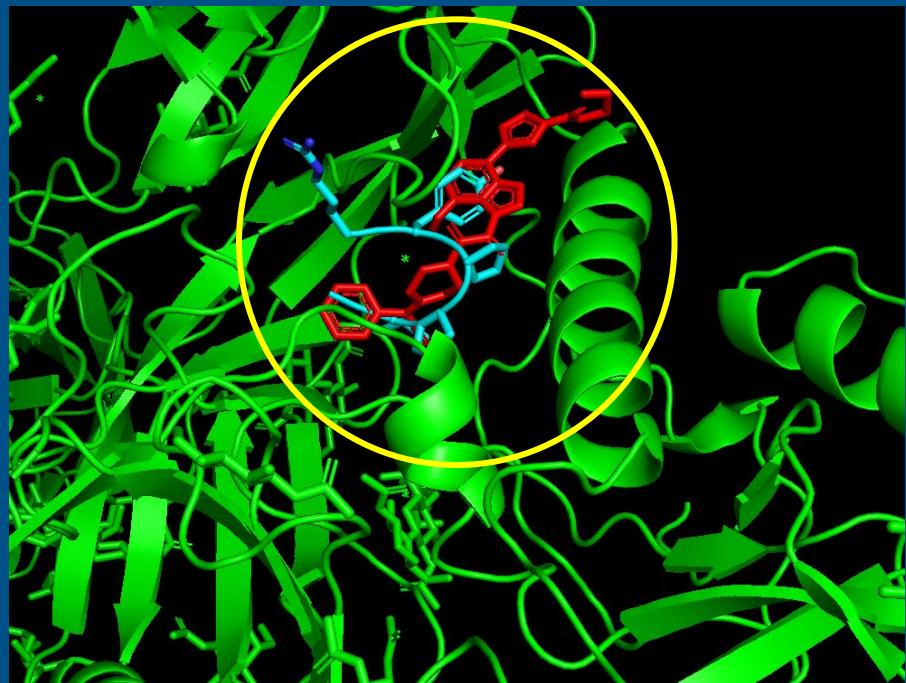
Changes to the default AlphaFold Model:

- CD4 binding-site hotspot proximity mapping
- A confidence-weighted centroid distance loss
- A boosted confident weighting from generation

Binder Length = 6

Sequence: GIPPYR

Generated cyclic peptide
BMS-818251



Combining it all together

Generated a total of 168 cyclic peptides using the following model:

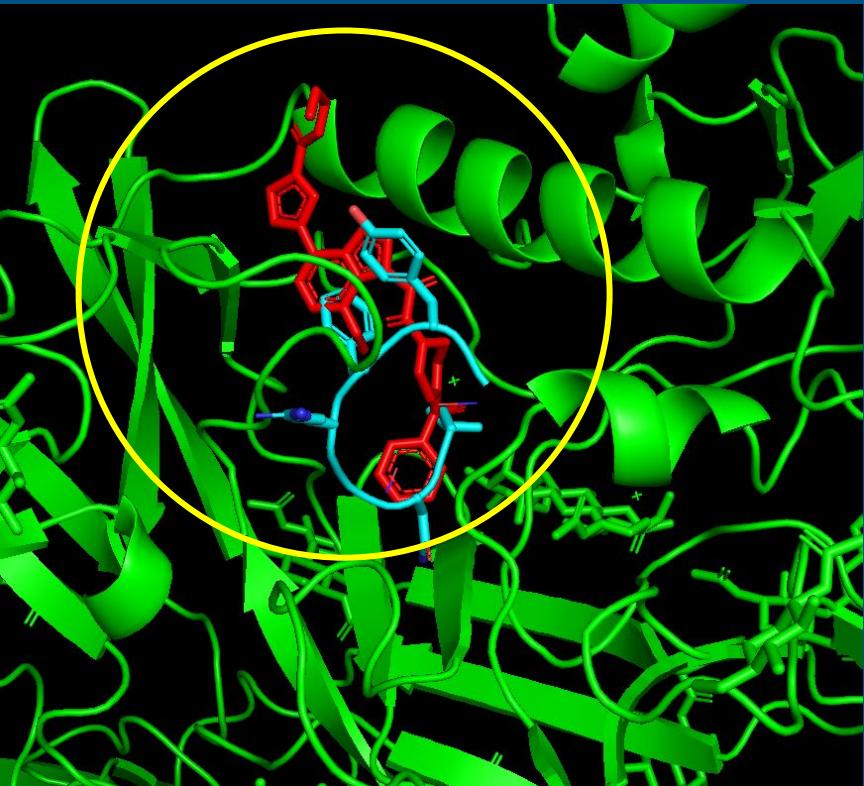
Changes to the default AlphaFold Model:

- CD4 binding-site hotspot proximity mapping
- A confidence-weighted centroid distance loss
- A boosted confident weighting from generation

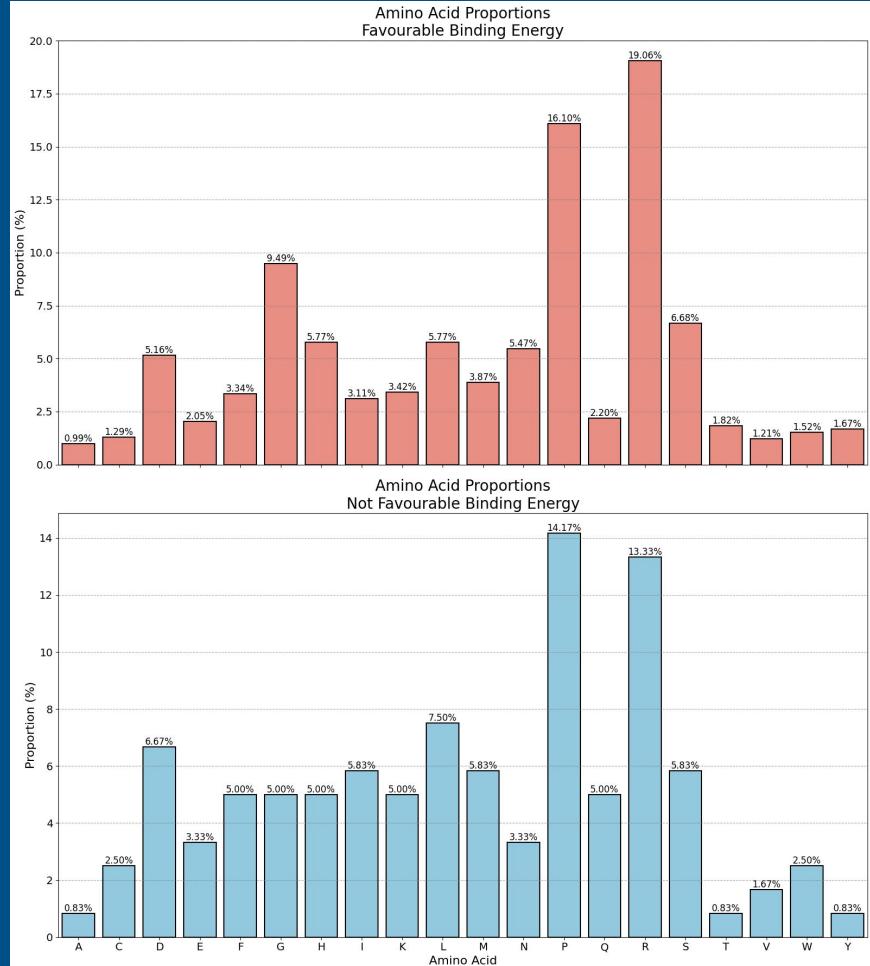
Binder Length = 7

Sequence: GYFRGNL

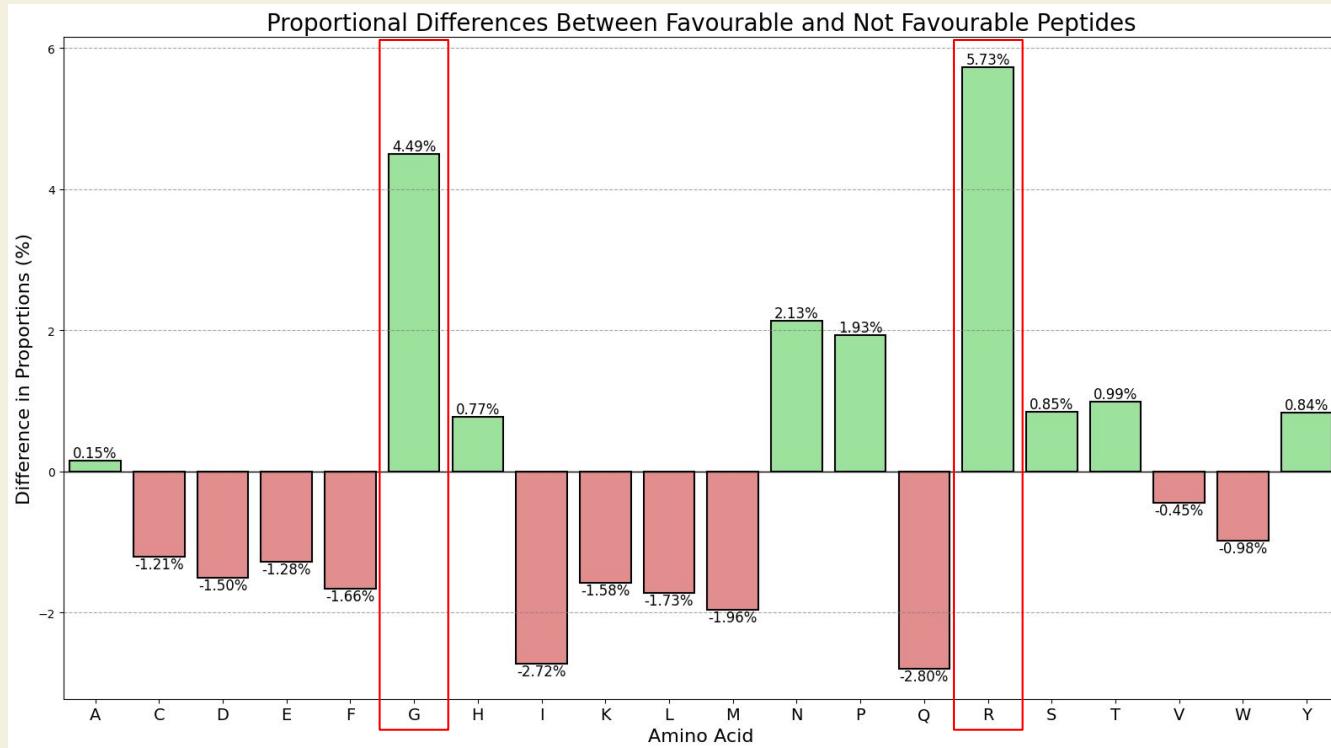
Generated cyclic peptide
BMS-818251



Distribution binning of favourable vs not favourable bindings



Distribution binning of favourable vs not favourable bindings



- 1. Overview of Methodology & Approaches**
- 2. Exploring peptide control with hotspots**
- 3. Enhancing precision of peptide generations**
- 4. Experimental Ligand Recreation**
- 5. Discussion & Further Work**

Conclusion

- Able to generate cyclic peptides with binding close to CD4-binding site for gp120
- Demonstrated increased control and precision of generated peptides, through the following:
 - Proximity hotspot mapping
 - Centroid distance calculation
 - pLDDT weighted MSE loss
 - Boosted plDDT loss
 - Binder length variation

Limitations & Further Areas

- Further improvement of the generation process by developing a search-process
- Generation and control of cyclic peptides linking the disulfide bridge in embedding
- Exploration of the non-canonical amino acids beyond the naturally occurring ones
- Optimize model generation through quantization

Thank you

Appendix 1

Experiment 4:

Does hotspot impact generation time of peptide?

results_df

i_ptm	loss	models	pae	plddt	...	seed	GD_method	optimizer	learning_rate	norm_seq_grad	dropout	num_recycles	num_models	hotspot	seconds
0.588371	3.358747	0	0.522446	0.390946	...	1	sgd	pssm_semgreedy	0.1	True	True	0	1	None	2722.599177
0.682958	3.130018	0	0.426358	0.586948	...	1	sgd	pssm_semgreedy	0.1	True	True	0	1	100	2726.515175
0.477004	3.720087	0	0.610313	0.341876	...	1	sgd	pssm_semgreedy	0.1	True	True	0	1	100-105	2733.477499
0.663349	3.155167	0	0.437303	0.582951	...	1	sgd	pssm_semgreedy	0.1	True	True	0	1	100-110	2734.201046
0.610113	3.367026	0	0.506754	0.428611	...	1	sgd	pssm_semgreedy	0.1	True	True	0	1	100-115	2735.500011
0.683624	3.090057	0	0.438786	0.560630	...	1	sgd	pssm_semgreedy	0.1	True	True	0	1	100-120	2739.833050

Appendix 2

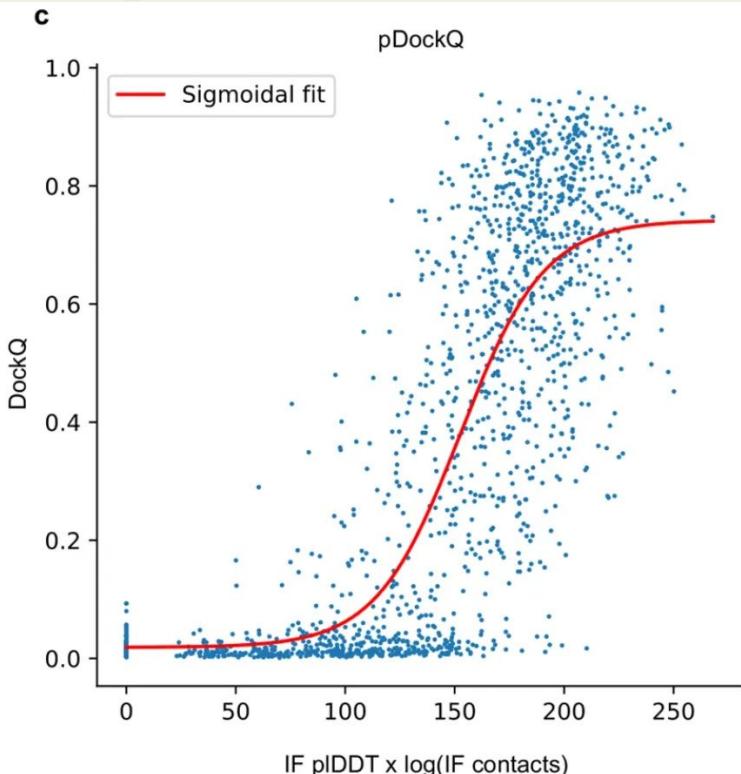
How does hotspot configuration work?

```
def _loss_binder(self, inputs, outputs, aux):
    '''get losses'''
    opt = inputs["opt"]
    mask = inputs["seq_mask"]
    zeros = jnp.zeros_like(mask)
    tL,bL = self._target_len, self._binder_len
    binder_id = zeros.at[-bL: ].set(mask[-bL: ])
    if "hotspot" in opt:
        target_id = zeros.at[opt["hotspot"]].set(mask[opt["hotspot"]])
        i_con_loss = get_con_loss(inputs, outputs, opt["i_con"], mask_1d=target_id, mask_1b=binder_id)
    else:
        target_id = zeros.at[:tL].set(mask[:tL])
        i_con_loss = get_con_loss(inputs, outputs, opt["i_con"], mask_1d=binder_id, mask_1b=target_id)

    # unsupervised losses
    aux["losses"].update({
        "plddt": get_plddt_loss(outputs, mask_1d=binder_id), # plddt over binder
    })
```

Appendix 3

The pDockQ score



pDockQ

As it is not only desirable to know when a model is accurate but also how accurate this model is, we developed a predicted DockQ score, pDockQ. This score is created by fitting a sigmoidal curve (Fig. 2c) using “curve_fit” from SciPy v.1.4.1⁵⁶, to the DockQ scores using the average interface plDDT multiplied with the logarithm of the number of interface contacts, with the following sigmoidal equation:

$$p\text{DockQ} = \frac{L}{1 + e^{-k(x-x_0)}} + b \quad (7)$$

where

$$x = \text{average interface plDDT} \cdot \log(\text{number of interface contacts}) \quad (8)$$

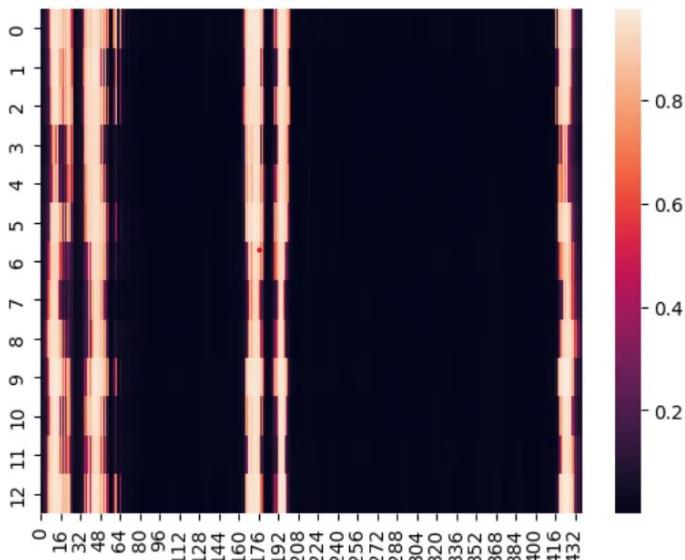
and we obtain $L = 0.724$, $x_0 = 152.611$, $k = 0.052$ and $b = 0.018$.

Appendix 4

How do you calculate the proximity for pDockQ?

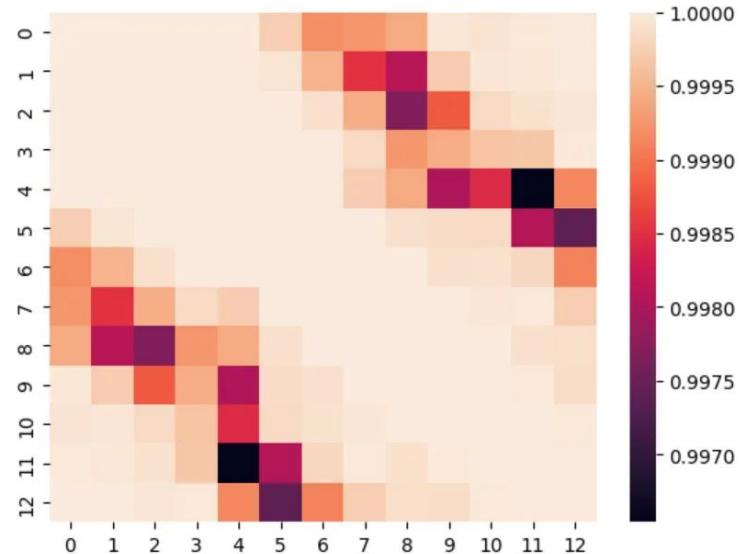
```
bL = 13
bin_plddt = w0_seed20['aux']['plddt'][:-bL:]
bin_cmap = w0_seed20['aux']['i_cmap'][:-bL:,:-bL]
sns.heatmap(bin_cmap)
```

<Axes: >



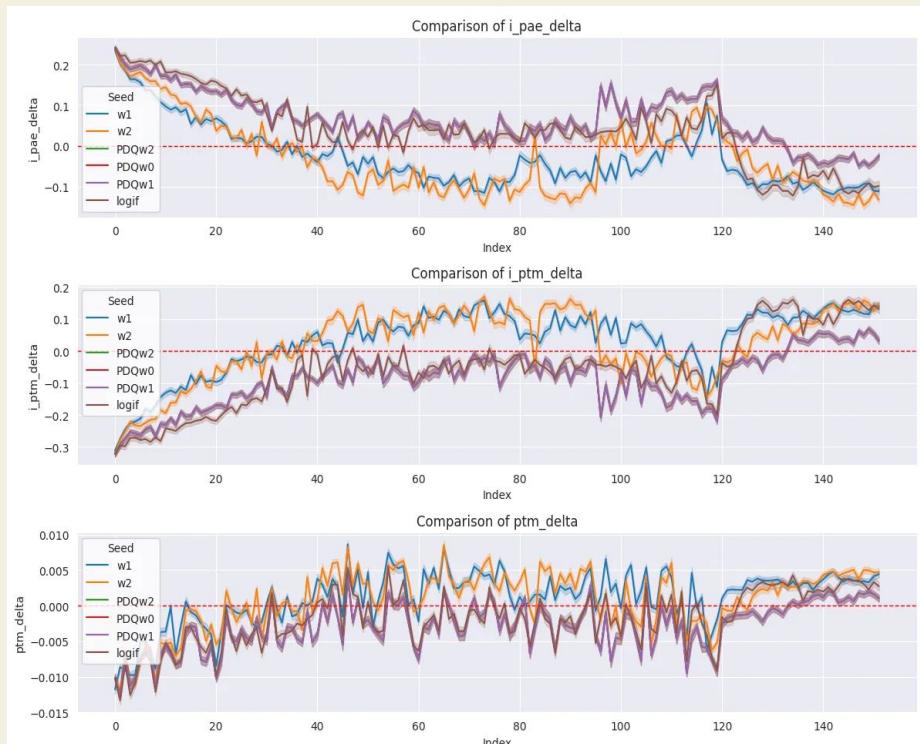
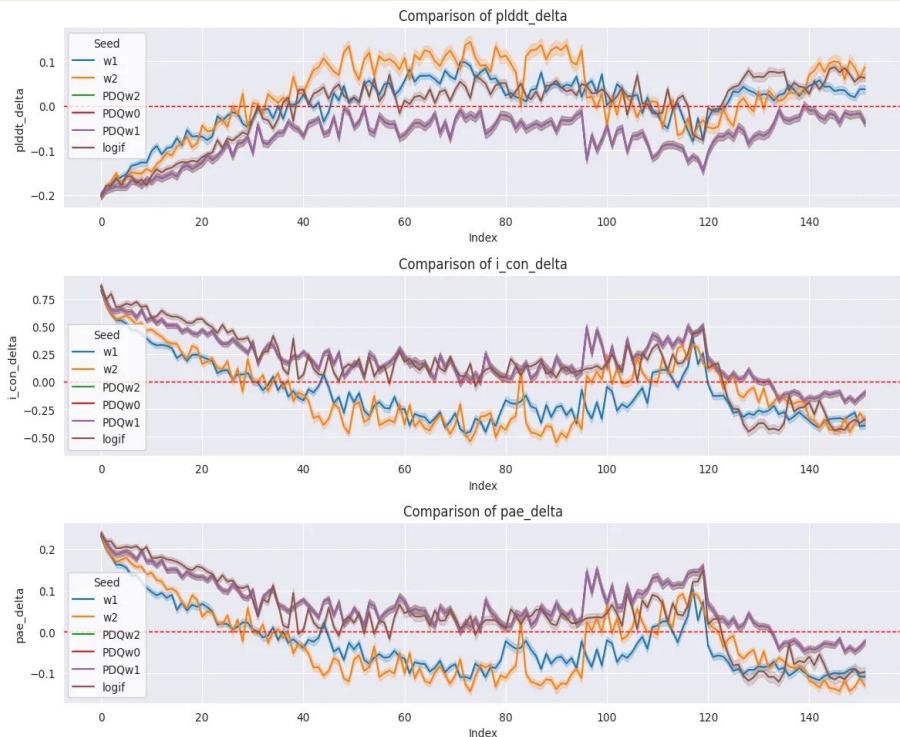
```
 sns.heatmap(aux_w2['i_cmap'][-bL:,-bL:])
```

<Axes: >



Appendix 5

Experiment 3: Training loss comparison



Appendix 6

Experiment 3: More charts & comparison

