Accelerating cryptic pocket discovery with deep learning

Gregory R. Bowman
@drGregBowman
Director of Folding@home
Louis Heyman University Professor
Depts. of Biochemistry & Biophysics, and Bioengineering
University of Pennsylvania

Disclaimer: Co-founder of Decrypt Bio
Structural snapshots are just the tip of the iceberg
Cryptic pockets present new opportunities

Horn and Shoichet. JMB 2004.
We're getting good at finding cryptic pockets with simulations.
But where should we look for cryptic pockets?
What data do we have to work with?

PDB ~200K structures

Only ~100 cryptic pockets
A second of simulation is a lot of data

SARS-2 Nsp16

SARS-2 Spike
Zimmerman et al. Nature Chemistry 2021

Ebola VP35
Cruz et al. Nature Communications 2022

ApoE
Stuchell-Brereton et al. PNAS 2022
PocketMiner algorithm

Node Features → Input Graph → Message Passing Layers → Output Graph

Cryptic pocket likelihood

Meller et al. bioRxiv 2022.
PocketMiner performs well on simulation data
PocketMiner performs well on crystal structures
Cryptic pockets dramatically expand the potentially draggable proteome

- **Cryptic pocket**
  - N=3172 (29.4%)

- **No pocket**
  - N=2001 (18.5%)

- **Ground state pocket**
  - N=5633 (52.1%)

PocketMiner is predictive of simulations

Kinase PIM 2 (crystal structure)

Orthosteric ligand
PocketMiner is predictive of simulations

Kinase PIM 2 (crystal structure)

Orthosteric ligand

Predicted cryptic pocket

Cryptic pocket likelihood
PocketMiner is predictive of simulations

Kinase PIM 2 (crystal structure)

Predicted cryptic pocket

Simulated structure Apo/template structure

Orthosteric ligand

Cryptic pocket likelihood

Meller et al. bioRxiv 2022.
Can AlphaFold do it again?
AlphaFold sometimes helps

Figure 2. Stochastic clustering of its input multiple sequence alignment allows AlphaFold to generate structures with open or partially open cryptic pockets across multiple systems.

A) In 6 out of 10 examples, AlphaFold samples the open state of a known cryptic pocket. The box-and-whisker plots show cryptic pocket root mean square deviation (RMSD) to a holo crystal structure (defined by heavy atoms within 5 Å of the ligand that binds at the cryptic pocket). For the top 5 examples, the holo structure was part of the training dataset for AlphaFold but the bottom 5 examples had their holo crystal structures deposited after AlphaFold was trained. The red line indicates 1.2 Å RMSD, a proposed cutoff for sampling the open state.

B) Structural overlay of an AlphaFold-generated structure with the holo structure of Neimann-Pick C2 Protein (NPC2) shows that AlphaFold samples the open state. The ligand which binds in the cryptic pocket is shown in magenta; the apo structure is shown in gray; the holo structure is shown in blue; and the AF structure is shown in orange. Residues that change rotamer state between apo and holo experimental structures are shown in sticks.

C) Structural overlay of an AlphaFold-generated structure of plasmepsin II with a holo structure containing a cryptic pocket shows that AlphaFold partially samples cryptic pocket opening. Select residues that change rotamer state between apo and holo experimental structures show that AlphaFold samples holo-like tryptophan orientations in the plasmepsin II cryptic pocket. As in B, the ligand which binds in the cryptic pocket is shown in magenta; the apo structure is shown in gray; the holo structure is shown in blue; and the AF structure is shown in orange.
Example of a success

Figure 2. Stochastic clustering of its input multiple sequence alignment allows AlphaFold to generate structures with open or partially open cryptic pockets across multiple systems.

A) In 6 out of 10 examples, AlphaFold samples the open state of a known cryptic pocket. The box-and-whisker plots show cryptic pocket root mean square deviation (RMSD) to a holo crystal structure (defined by heavy atoms within 5 Å of the ligand that binds at the cryptic pocket). For the top 5 examples, the holo structure was part of the training dataset for AlphaFold but the bottom 5 examples had their holo crystal structures deposited after AlphaFold was trained. The red line indicates 1.2 Å RMSD, a proposed cutoff for sampling the open state.

B) Structural overlay of an AlphaFold-generated structure with the holo structure of Neimann-Pick C2 Protein (NPC2) shows that AlphaFold samples the open state. The ligand which binds in the cryptic pocket is shown in magenta; the apo structure is shown in gray; the holo structure is shown in blue; and the AF structure is shown in orange. Residues that change rotamer state between apo and holo experimental structures are shown in sticks.

C) Structural overlay of an AlphaFold-generated structure of plasmepsin II with a holo structure containing a cryptic pocket shows that AlphaFold partially samples cryptic pocket opening. Select residues that change rotamer state between apo and holo experimental structures show that AlphaFold samples holo-like tryptophan orientations in the plasmepsin II cryptic pocket. As in B, the ligand which binds in the cryptic pocket is shown in magenta; the apo structure is shown in gray; the holo structure is shown in blue; and the AF structure is shown in orange.
Example of a partial success

PM 2

Apo

Holo

AlphaFold Conformers

AF Structs.

Holo Xtals

W41-Y77 [nm]

W41-β6 [nm]

Example of a partial success

PM 2

Apo

Holo

AlphaFold Conformers

AF Structs.

Holo Xtals

W41-Y77 [nm]

W41-β6 [nm]
AlphaFold still helps jumpstart MD

Figure 3. Launching simulations from AlphaFold-generated structures improves sampling of cryptic pocket opening in Plasmepsin II. A) Structure of PM II's flap domain showing key residues involved in PM II's cryptic pocket. Trp41 and Tyr77, part of the flap domain, are shown in sticks. We use the distances indicated in dotted lines to capture pocket opening. Specifically, the cryptic pocket is open when the minimum distance between Y77 and W41 is large (indicated with blue line) and the distance between the W41 sidechain (either atom CZ3 or CH2 depending on which is closer) and a reference residue in the 6th beta-sheet (K72) is small (indicated with red line).

B) Pocket distances for a set of 32 AlphaFold-generated conformers (grey dots) and holo-crystal structures (black triangles) show that the AlphaFold ensemble includes partially open states for PM II. Trp41 is in its holo-orientation in one of the AlphaFold structures, but the distance between Trp41 and Tyr77 is smaller than it is in holo-crystal structures.

C) A free energy surface from a Markov State Model from apo-seeded simulations shows that these simulations do not sample cryptic pocket opening. Though the flap dissociates as indicated by large Trp41-Tyr77 distances, Trp41 does not adopt the holo-orientation, despite 32 microseconds of sampling.

D) A free energy surface from a Markov State Model generated from AlphaFold-seeded simulations shows robust sampling of the open state. Both requirements for cryptic pocket opening are fulfilled as indicated by the overlay of holo-crystal structures (black triangles) on the free energy surface.
AlphaFold still helps jumpstart MD

Figure 3. Launching simulations from AlphaFold-generated structures improves sampling of cryptic pocket opening in Plasmepsin II. A) Structure of PM II's flap domain showing key residues involved in PM II's cryptic pocket. Trp41 and Tyr77, part of the flap domain, are shown in sticks. We use the distances indicated in dotted lines to capture pocket opening. Specifically, the cryptic pocket is open when the minimum distance between Y77 and W41 is large (indicated with blue line) and the distance between the W41 sidechain (either atom CZ3 or CH2 depending on which is closer) and a reference residue in the 6th beta-sheet (K72) is small (indicated with red line). B) Pocket distances for a set of 32 AlphaFold-generated conformers (grey dots) and holo-crystal structures (black triangles) show that the AlphaFold ensemble includes partially open states for PM II. Trp41 is in its holo-orientation in one of the AlphaFold structures, but the distance between Trp41 and Tyr77 is smaller than it is in holo-crystal structures. C) A free energy surface from a Markov State Model from apo-seeded simulations shows that these simulations do not sample cryptic pocket opening. Though the flap dissociates as indicated by large Trp41-Tyr77 distances, Trp41 does not adopt the holo-orientation, despite 32 microseconds of sampling. D) A free energy surface from a Markov State Model generated from AlphaFold-seeded simulations shows robust sampling of the open state. Both requirements for cryptic pocket opening are fulfilled as indicated by the overlay of holo-crystal structures (black triangles) on the free energy surface.
How do we find the key degrees of freedom when we don’t know the answer?
Making comparisons is hard given the assumptions made by existing methods.

- Variant 1
- Variant 2

Decision boundary

DiffNets automate the discovery of biochemically-relevant traits

DiffNets automate the discovery of biochemically-relevant traits

DiffNets finds subtle structural differences that explain biochemical variation

- Variant 1
- Variant 2
- Decision boundary
Relaxation of the labels
VP35's cryptic pocket is coupled to the blunt end-binding interface.
Thanks!

foldingathome.org