

PockMan

Examples explanation

Wael Badr, Narine Fischer, Radostina Kislava

Master in Bioinformatics for Health Sciences

Universitat Pompeu Fabra

Structural Bioinformatics and Introduction to Python joint project

April 2025

3HTB - T4 Lysozyme

Bacteriophages are distributed across the Earth, being found in both terrestrial and aquatic environments, and being able to survive extreme conditions [1,2]. Among the existing bacteriophages, T4, a double-stranded DNA phage [2], has been used to infect strains of *Escherichia coli* to understand the underlying mechanism of molecular biology [1].

The lysozyme analyzed in this example has a role in the lytic cycle of the T4 phage. This protein is an endolysin responsible for degrading the peptidoglycan layer of the host's cell wall, causing the lysis of the host bacteria, and the release of mature T4 phages into the environment [3].

From Protein Data bank, the crystallized structure can be obtained under the code 3HTB [4]. This protein was co-crystallized alongside three different ligands:

- PO4: a phosphate ion
- JZ4: 2-propylphenol
- BME: β -mercaptoethanol

To assess whether PockMan successfully detects the ligand binding sites of this protein, its results were compared against the specific residues annotated in the PDB file to be in contact with the ligands, signaled by the header "SITE". In the PDB, it can be found a list of interacting residues for either PO4 or JZ4, but not for BME, so the results were compared for the two known binding sites.

With PockMan, a total of eight predicted ligand-binding sites were identified (Fig.1A). Of these, two showed to partially overlap either the PO4 binding site or JZ4 binding site described in the PDB file. Specifically:

- Predicted ligand-binding site nº 3 partially overlaps with the residues annotated interacting with JZ4 (Fig.1B).
- Predicted ligand-binding site nº 7 partially overlaps with the residues annotated interacting with PO4 (Fig.1C).

Regarding the rest of pockets signaled by PockMan, they might represent alternative or potential ligand-binding sites in the protein, each one characterized by different degrees of depth and surface accessibility.

In conclusion, these results show that PockMan is able to detect known ligand-binding sites with a certain level of accuracy. Not only determining the interaction with small organic compounds (2-propylphenol, JZ4), but also of ions that interact with the protein (phosphate ion, PO4).

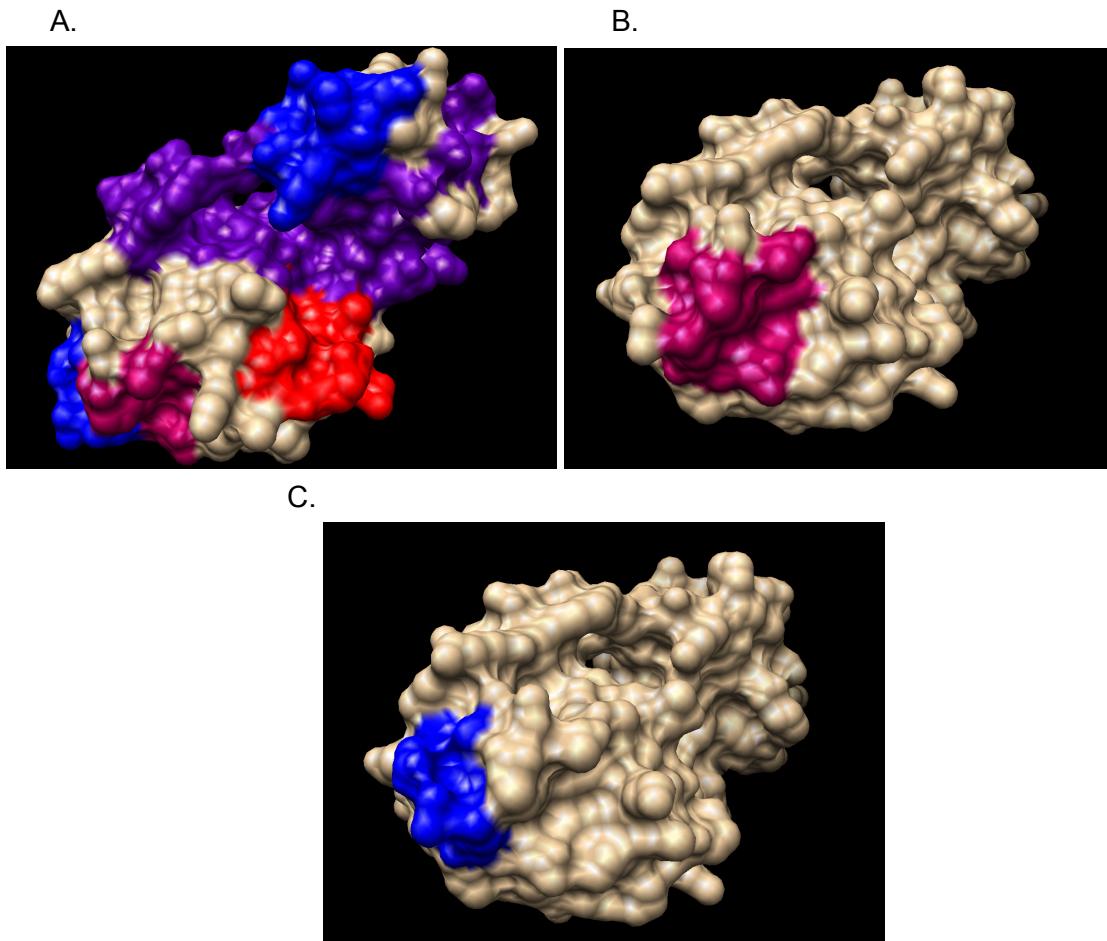


Figure 1. A) Visualization of the lysozyme of the T4 phage with all predicted ligand-binding sites predicted by PockMan marked. The color is related to the score obtained, the sites with a higher score (more deep) have a color resembling more to red, while those with lowest (less deep), their colour resembles more to blue. At the bottom left, the two predicted ligand-binding sites overlapping the ones specified in the PDB can be observed. B) Predicted ligand-binding site nº 3, which partially overlaps the stated residues that interact with JZ4 in the PDB. C) Predicted ligand-binding site nº 7, which partially overlaps the stated residues that interact with PO4 in the PDB.

3PTB - Serine protease 1

Serine proteases play an important role in the regulation of several cellular processes. One of its functions is to activate some protein precursor, such as pro-enzymes, prohormones or protease-activated receptors [5]. They have also been observed participating in the degradation of proteins to facilitate bacterial infection [6].

The serine protease analyzed is from *Bos taurus*. It is encoded by the gene PRSS1 [7].

From Protein Data bank, the crystallized structure can be obtained under the code 3PTB [8]. This protein was co-crystallized alongside two different ligands:

- CA: a calcium ion
- BEN: benzamidine

As with the previous protein, to assess whether PockMan successfully detects the ligand binding sites of this protein, its results were compared against the specific residues annotated in the PDB file to be in contact with the ligands, signaled by the header “SITE”. In the PDB, it can be found a list of interacting residues for either CA or BEN. Both these lists have been obtained with a software.

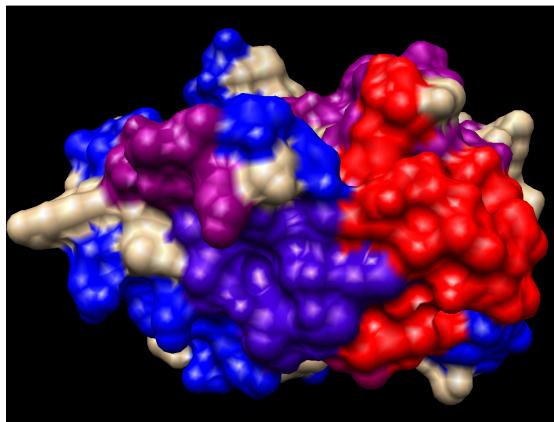
With PockMan, a total of fourteen predicted ligand-binding sites were identified (Fig.2A). Of these, two showed complete overlap with either the CA binding site or BEN binding site described in the PDB file. Specifically:

- Predicted ligand-binding site nº 1 completely overlaps with the residues annotated interacting with CA (Fig.2B). Despite this, extra residues are present in our prediction, which may be close in space with the residues listed in the PDB. The overlapping residues are: Glu70, Asn72, Val 75, Glu80.
- Predicted ligand-binding site nº 4 completely overlaps with the residues annotated interacting with BEN (Fig.2C). As before, extra residues can be found in our prediction, in addition to the residues listed in the PDB. The overlapping residues are: Asp189, Ser190, Ser195, Val213, Gly216, Gly219, Gly226.

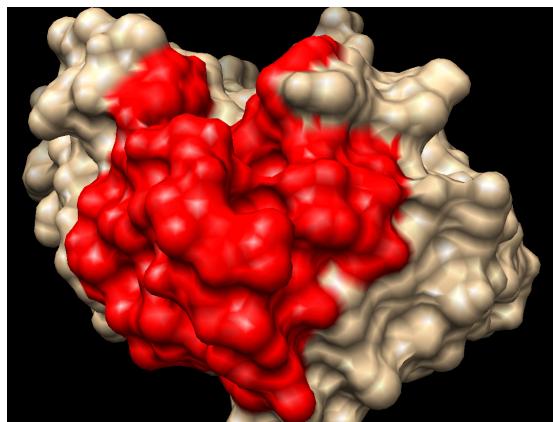
The rest of the pockets might be other possible ligand-binding sites. On the other hand, PockMan is just a geometrical-approach, it is possible that some of the predictions are false positives. Therefore, the addition of other ways to detect ligand-binding sites might improve the results of PockMan, removing possible false positives, for example using an energy-approach with the geometrical.

Nevertheless, these results from Pockman show that the program is able to detect the ligand-binding sites with accuracy, as it was able to detect the ligand binding site of not only the calcium ion, but also

A.



B.



C.

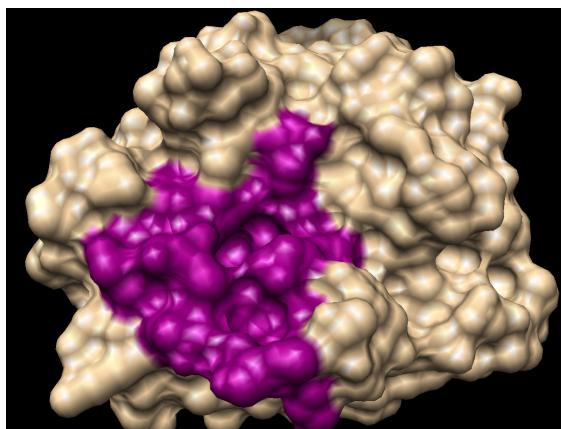


Figure 2. A) Visualization of the serine protease from *Bos taurus* with all predicted ligand-binding sites predicted by PockMan marked. The color is related to the score obtained, the sites with a higher score (more deep) have a color resembling more to red, while those with lowest (less deep), their colour resembles more to blue. B) Predicted ligand-binding site n° 1, which completely overlaps the stated residues that interact with CA in the PDB. C) Predicted ligand-binding site n° 4, which completely overlaps the stated residues that interact with BEN in the PDB.

1HVR - HIV-1 Protease

For this analysis, PockMan was run using the default parameters: a grid size of 1.0 Å, a border size of 5.0 Å, a voxel score cutoff of 4, and an atom distance threshold of 4 Å. Additionally, planar diagonal PSP detection was enabled by answering “yes” when prompted.

HIV-1 protease is a homodimeric aspartyl protease essential for the replication cycle of HIV-1. It catalyzes the cleavage of the viral polyprotein into functional subunits required for the assembly of infectious virions [10]. Given its indispensable role, the enzyme has long been a major target for antiretroviral therapy. The structure analyzed in this example, deposited under the PDB code 1HVR, represents the protease in complex with a nonpeptidic cyclic urea inhibitor, a class of molecules rationally designed for enhanced potency, bioavailability, and resistance to proteolytic degradation. These inhibitors mimic the transition state of the natural substrate and bind tightly within the active site, thereby blocking protease function and viral maturation [10]. This example is a s

To determine the ligand-binding site of HIV-1 protease, we used UCSF Chimera to analyze the structure deposited under the PDB code 1HVR, which is co-crystallized with a cyclic urea inhibitor labeled as XK2 [9]. The residues within 5 Å and less of XK2 were assumed to be the binding site of the protease.

To evaluate the accuracy of our predicted binding sites, we compared the top-ranking pocket detected by PockMan (shown in Figure 3, left) with the actual ligand-binding site of the HIV-1 protease structure (PDB: 1HVR), which contains a co-crystallized cyclic urea inhibitor (XK2, shown in pink in Figure 3, right).

The predicted binding site, visualized in red, closely overlaps with the region where the ligand is bound in the crystal structure. This spatial correspondence strongly suggests that

binding site #1, as identified by PockMan, accurately reflects the true ligand-binding pocket of the protein.

This validation supports the capability of PockMan to detect biologically relevant binding sites, even in complex homodimeric structures such as HIV-1 protease.

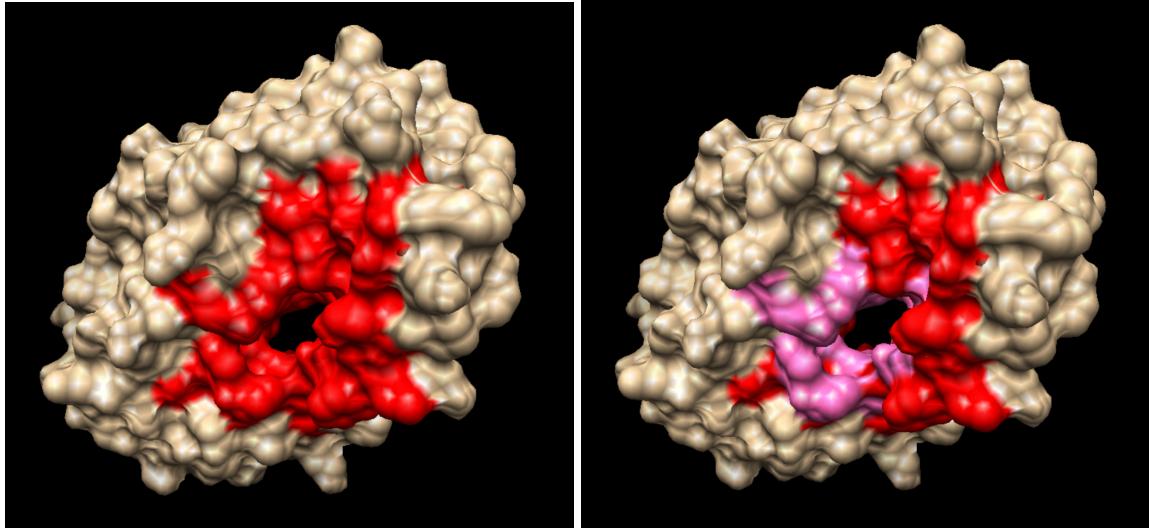


Figure 3. PockMan-predicted binding site #1 (left) vs. predicated & known inhibitor site (right).

On the other hand, the second predicted binding site (Figure 4) does not appear to correspond to any known or plausible ligand-binding region in the protein as it is basically highlighting the whole protein. Given that the first site already shows a strong match with the actual inhibitor's location, it is possible that the second site represents a lower-confidence or non-functional surface feature.

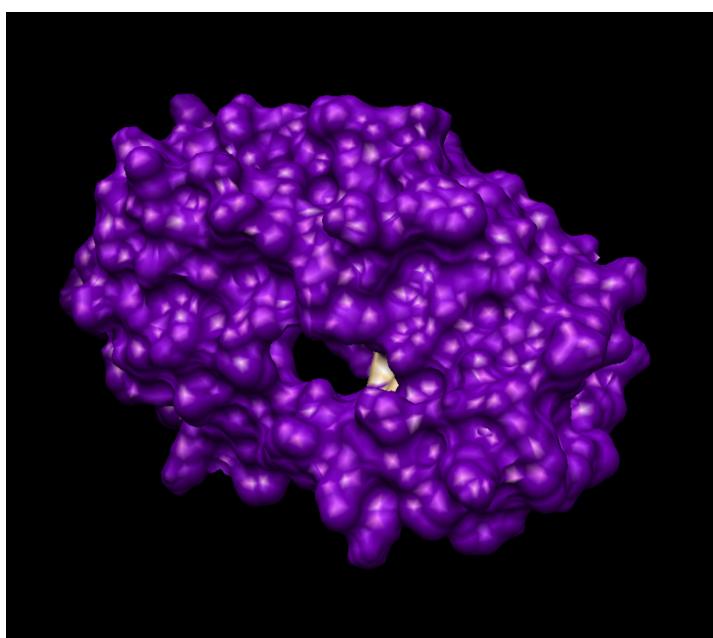


Figure 4. PockMan-predicted binding site #2.

This example also serves to highlight that PockMan performs reliably even in the context of dimeric proteins. HIV-1 protease functions as a homodimer, with its active site formed at the interface between the two monomers. Despite the structural complexity introduced by dimerization, PockMan was still able to correctly identify the main ligand-binding pocket. This demonstrates that the tool is capable of handling multimeric assemblies without compromising accuracy.

REFERENCES

1. Rao, V. B., Fokine, A., Fang, Q., & Shao, Q. (2023). Bacteriophage T4 Head: Structure, Assembly, and Genome Packaging. *Viruses*, 15(2), 527. <https://doi.org/10.3390/v15020527>
2. Rao, V. B., & Zhu, J. (2022). Bacteriophage T4 as a nanovehicle for delivery of genes and therapeutics into human cells. *Current opinion in virology*, 55, 101255. <https://doi.org/10.1016/j.coviro.2022.101255>
3. UniProt Consortium. (2023). *ENLYS_BPT4 (Endolysin)*. UniProtKB. <https://www.uniprot.org/uniprotkb/P00720>
4. Boyce, S. E., Mobley, D. L., Rocklin, G. J., Graves, A. P., Dill, K. A., & Shoichet, B. K. (2009). *2-propylphenol in complex with T4 lysozyme L99A/M102Q [Protein structure]*. RCSB Protein Data Bank: 3HTB. <https://www.rcsb.org/structure/3HTB>
5. Chen, X., Leahy, D., Van Haeften, J., Hartfield, P., Prentis, P. J., van der Burg, C. A., Surm, J. M., Pavasovic, A., Madio, B., Hamilton, B. R., King, G. F., Undheim, E. A. B., Brattsand, M., & Harris, J. M. (2019). A Versatile and Robust Serine Protease Inhibitor Scaffold from *Actinia tenebrosa*. *Marine drugs*, 17(12), 701. <https://doi.org/10.3390/md17120701>
6. Rosales-Islas, V., Ramírez-Paz-Y-Puente, G. A., Montes-García, F., Vázquez-Cruz, C., Sánchez-Alonso, P., Zenteno, E., & Negrete-Abascal, E. (2024). Isolation and characterization of a Mannheimiahaemolytica secreted serine protease that degrades sheep and bovine fibrinogen. *Microbial pathogenesis*, 192, 106706. <https://doi.org/10.1016/j.micpath.2024.106706>
7. UniProt Consortium. (2023). *TRY1_BOVIN (Serine Protease 1)*. UniProtKB. <https://www.uniprot.org/uniprotkb/P00760/entry>
8. Bode, W., Schwager, P., Walter, J. (1983). *THE GEOMETRY OF THE REACTIVE SITE AND OF THE PEPTIDE GROUPS IN TRYPSIN, TRYPSINOGEN AND ITS COMPLEXES WITH INHIBITORS [Protein structure]*. RCSB Protein Data Bank: 3PTB. <https://www.rcsb.org/structure/3PTB>
9. Koyano, K., & Nakano, T. (2008). Interaction of HIV-1 aspartic protease with its inhibitor, by molecular dynamics and *ab initio* fragment molecular orbital method. *Journal of Synchrotron Radiation*, 15(3), 239–242. <https://doi.org/10.1107/s0909049507068586>
10. Lam, P. Y., Jadhav, P. K., Eyermann, C. J., Hodge, C. N., Ru, Y., Bacheler, L. T., Meek, J. L., Otto, M. J., Rayner, M. M., Wong, Y. N., Chang, C.-H., Weber, P. C., Jackson, D. A., Sharpe, T. R., & Erickson-Viitanen, S. (1994). Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science*, 263(5145), 380–384. <https://doi.org/10.1126/science.8278812>

1HVR - HIV-1 Protease

For this analysis, PockMan was run using the default parameters: a grid size of 1.0 Å, a border size of 5.0 Å, a voxel score cutoff of 4, and an atom distance threshold of 4 Å. Additionally, planar diagonal PSP detection was enabled by answering “yes” when prompted.

HIV-1 protease is a homodimeric aspartyl protease essential for the replication cycle of HIV-1. It catalyzes the cleavage of the viral polyprotein into functional subunits required for the assembly of infectious virions [2]. Given its indispensable role, the enzyme has long been a major target for antiretroviral therapy. The structure analyzed in this example,

deposited under the PDB code 1HVR, represents the protease in complex with a nonpeptidic cyclic urea inhibitor, a class of molecules rationally designed for enhanced potency, bioavailability, and resistance to proteolytic degradation. These inhibitors mimic the transition state of the natural substrate and bind tightly within the active site, thereby blocking protease function and viral maturation [2]. This example is a s

To determine the ligand-binding site of HIV-1 protease, we used UCSF Chimera to analyze the structure deposited under the PDB code 1HVR, which is co-crystallized with a cyclic urea inhibitor labeled as XK2 [1]. The residues within 5 Å and less of XK2 were assumed to be the binding site of the protease.

To evaluate the accuracy of our predicted binding sites, we compared the top-ranking pocket detected by PockMan (shown in Figure 1, left) with the actual ligand-binding site of the HIV-1 protease structure (PDB: 1HVR), which contains a co-crystallized cyclic urea inhibitor (XK2, shown in pink in Figure 1, right).

The predicted binding site, visualized in red, closely overlaps with the region where the ligand is bound in the crystal structure. This spatial correspondence strongly suggests that binding site #1, as identified by PockMan, accurately reflects the true ligand-binding pocket of the protein.

This validation supports the capability of PockMan to detect biologically relevant binding sites, even in complex homodimeric structures such as HIV-1 protease.

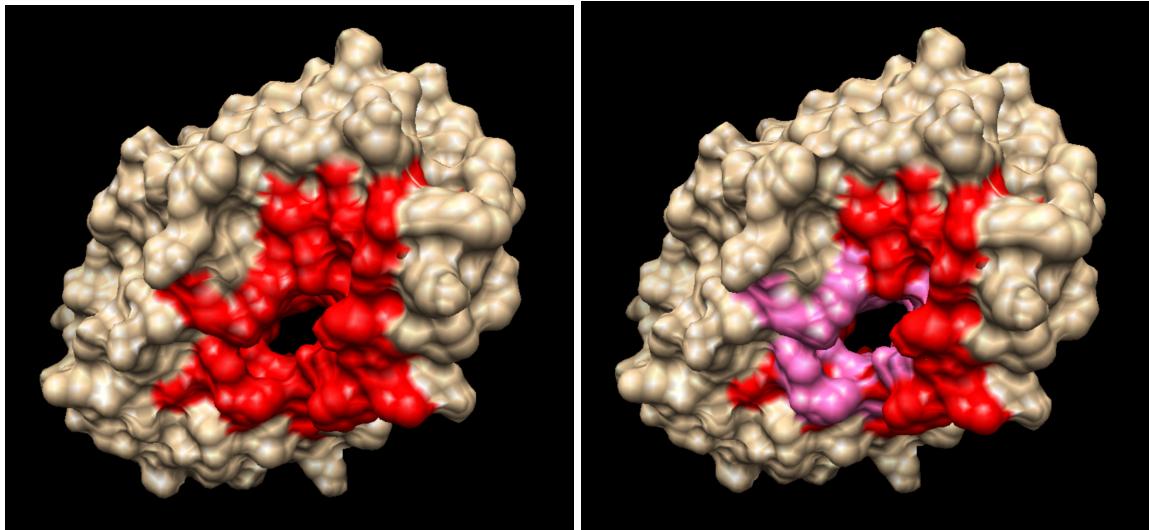


Figure 1. PockMan-predicted binding site #1 (left) vs. predicted & known inhibitor site (right).

On the other hand, the second predicted binding site (Figure 2) does not appear to correspond to any known or plausible ligand-binding region in the protein as it is basically highlighting the whole protein. Given that the first site already shows a strong match with the actual inhibitor's location, it is possible that the second site represents a lower-confidence or non-functional surface feature.

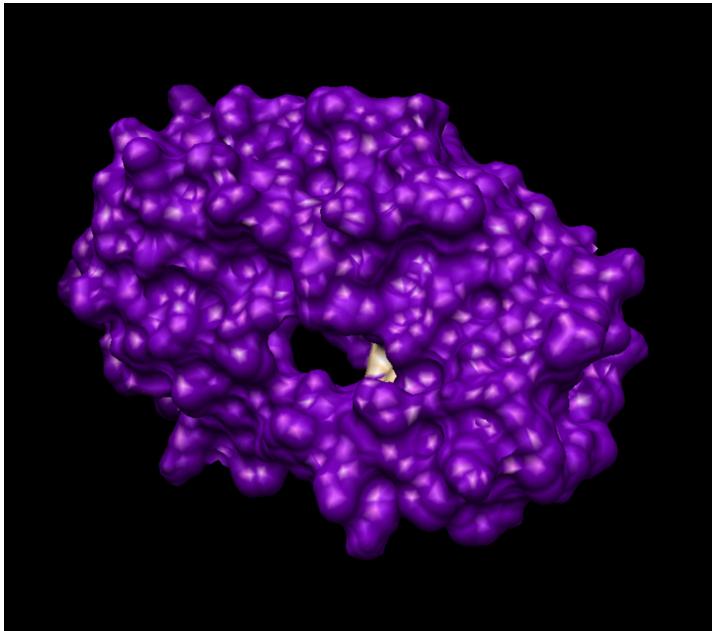


Figure 2. PockMan-predicted binding site #2.

This example also serves to highlight that PockMan performs reliably even in the context of dimeric proteins. HIV-1 protease functions as a homodimer, with its active site formed at the interface between the two monomers. Despite the structural complexity introduced by dimerization, PockMan was still able to correctly identify the main ligand-binding pocket. This demonstrates that the tool is capable of handling multimeric assemblies without compromising accuracy.

REFERENCES

1. Koyano, K., & Nakano, T. (2008). Interaction of HIV-1 aspartic protease with its inhibitor, by molecular dynamics and *a b initio* fragment molecular orbital method. *Journal of Synchrotron Radiation*, 15(3), 239–242.
<https://doi.org/10.1107/s0909049507068586>
2. Lam, P. Y., Jadhav, P. K., Eyermann, C. J., Hodge, C. N., Ru, Y., Bacheler, L. T., Meek, J. L., Otto, M. J., Rayner, M. M., Wong, Y. N., Chang, C.-H., Weber, P. C., Jackson, D. A., Sharpe, T. R., & Erickson-Viitanen, S. (1994). Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science*, 263(5145), 380–384. <https://doi.org/10.1126/science.8278812>
- 3.