Dear editors,

We wish to submit an original paper entiteled Python based pipeline for detection and characterization of allosteric pockets in GPCR molecular dynamics simulations by Bioinformatics journal.

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

G protein-coupled receptors (GPCRs) are a large family of seven-transmembrane-domain proteins that are activated by a diverse range of ligands and plays an essential role in initiating intracellular functional responses. Since GPCRs are fundamental to transduce cellular responses, these receptors become a drug target of major research focus in the pharmaceutical industry.

However, most of these efforts on finding new drugs have focused on targeting the orthosteric binding site which is highly conserved among GPCR subtypes. As a consequence, important difficulties appear when designing an orthosteric drug due to the poor selectivity and off-target side effects. In order to achieve greater ligand selectivity on different GPCR subtypes, allosteric regulation has emerged as an innovation strategy towards GPCR drug design.

In this project it is developed a python-based automated pipeline for pocket detection in molecular dynamics simulations for allosteric modulation of GPCRs using the program MDpocket which is available at https://github.com/OriolCanal/Master_thesis. This pipeline is implemented into GPCRmd server for the detection and characterization of druggable binding sites. In addition, a case study of a co-crystallized structure of CCR5 with its allosteric modulator maraviroc is used to show the potential of the pipeline to detect druggable binding sites.

Our findings could be of great interest as it will allow one to have a specific view of the differences and similarities of allosteric pockets between different GPCRs subtypes and enhancing allosteric drug development on GPCRs..

Thank you for your consideration of this manuscript,

Yours sincerely,

Signature:

Oriol Canal Pujol

Jana Selent

Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Structural Bioinformatics

Structural Bioinformatics

Python based pipeline for detection and characterization of allosteric pockets in GPCR molecular dynamics simulations.

Canal Oriol[1,*], Aranda David[1] and Selent Jana[1]

[1]GPCR drug discovery group. Research Programme on Biomedical Informatics(GRIB), Hospital del Mar Medical Research Institute (IMIM), 08003 Barcelona, Spain

Abstract

Motivation: G-protein coupled receptors (GPCR) are the largest family of human membrane proteins. The implication of GPCRs in a plethora of physiological processes makes them a drug target of high relevance. However, the orthosteric binding site of GPCRs is highly homologous between different subtypes. For this reason, allosteric drug design has emerged with the great potential to improve GPCR selectivity. Here, a working pipeline to automatically detect, identify and characterize allosteric binding sites in GPCRs using MDpocket as a cavity detection algorithm is described. This pipeline has been applied to GPCRmd database with more than 1000 GPCR molecular dynamics simulations from more than 60 receptor subtypes.
Results: A case study based on detection of Maraviroc binding pocket proved the pipeline to be useful for detecting allosteric pockets.
Availability: Freely available on https://github.com/OriolCanal/Master_thesis
Contact: oriol.canal01@estudiant.upf.edu
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

G protein-coupled receptors (GPCRs) are a large family of seven-transmembrane-domain cell membrane proteins. By binding to a wide variety of extracellular ligands (e.g., hormones, neurotransmitters, ions, or other stimuli) they are able to be activated and, as consequence, transmit a specific signal into the cell. The implication of GPCRs in plethora of physiological processes makes them a drug target of high relevance for a broad range of different disorders such as Alzheimer, cancer, obesity, among others (Hauser et al., 2017) . It is estimated that drugs that target for GPCRs accounts for 35-45% of the global market of therapeutic drugs (Serrano-Marín et al., 2020) proving the strong interest of pharmaceutical industries over these receptors.

Most of the approved drugs interact with the orthosteric site, which is the binding site of the endogenous ligand. However, there are some GPCRs characteristics that makes orthosteric drugs laborious to design. The main issue is that the orthosteric binding site is highly homologous between different subtypes causing difficulties to develop drugs that bind to a unique receptor subtype. As consequence, synthetic ligands for only a small fraction of GPCRs are available, despite the strong effort to obtain highly selective ligands (Conn et al., 2009).

As an alternative strategy, allosteric drug design has emerged with the great potential to improve GPCR selectivity (Wu, Want et al., 2014, Miao et al., 2018) . Allosteric ligands bind at distinct sites from those of endogenous ligands causing structural rearrangements that are transmitted across the receptor causing either enhancement or inhibition of the specific signaling response. The allosteric and the orthosteric ligand bind together to the GPCR forming a complex where the allosteric modulator affects the binding affinity and/or efficacy of the orthosteric ligand. Generally, orthosteric sites have faced high evolutionary pressure to keep an efficient binding to their orthosteric ligand. In contrast, allosteric pockets have less conserved sequences and more variable sites among GPCRs subtypes. Hence, targeting allosteric sites can provide improved receptor subtype selectivity.

However, targeting allosteric sites comes with some challenges. Allosteric binding sites are not always evident from the static structure as pockets can be transiently formed due to inherent protein dynamics. In this scenario, molecular dynamics (MD) simulations have been shown to be useful for detecting potential allosteric binding sites (Matosin et al., 2014).

So far, no standardized procedure for detecting and analyzing allosteric pockets in GPCRs has been developed. Therefore, this project aims to design and develop python-based pipeline to identify and characterize

allosteric binding sites in GPCRs using MD simulations as input. These MD simulations will be obtained from GPCRmd database (Rodriguez, Torrens et al., 2020) where more than 3000 simulations from over 60 GPCR subtypes can be found. In order to validate this tool, we have performed a case study of the GPCR CCR5 with its allosteric modulator Maraviroc to show the potential of the pipeline to correctly detect allosteric pockets and to differentiate the promising ones.

With this implementation, we aim to provide information of great interest as it will show if cavities are maintained or not in different GPCR subtypes, being a great tool for predicting the selectivity of a modulator over different GPCR subtypes. This implementation will provide different descriptors of the pockets (e.g., volume, residues involved, hydrophobicity density) to easily detect the most interesting ones.

2 Methods

GPCR MD simulations were obtained from the GPCRmd database. More than 1500 molecular dynamics simulations, with an accumulated time of 750 μs, have been analyzed through a python-based pipeline aiming to detect and characterize potential allosteric pockets.

2.1 MDpocket Concepts

This pipeline is mainly based on MDpocket pocket detection program (Schmidtke et al., 2011). MDpocket is based on Fpocket (Le Guilloux, Schmidtke et al., 2009) pocket detection algorithm. Fpocket aims to identify and characterize binding sites from a static structure (PDB) using a geometry-based cavity detection algorithm. In contrast, MDpocket runs Fpocket algorithm in every frame of the simulation allowing the qualitative and quantitative analysis of pockets.

Both programs are based on alpha spheres theory, an approach that relies on Voronoi tessellation (Poupon et al., 2004). Alpha spheres are spheres that contact four atoms on their border and do not contain any atom inside the sphere. In order to detect the pocket, both programs identify alpha spheres in the protein structure. The radii of these alpha spheres are used to identify binding pockets: whereas alpha spheres of large and very small radii are located at the exterior and within the protein respectively, intermediate radii correspond to spheres that are located in cavities. Thus, to differentiate between candidate pockets from the ones of poor interest, clusters of alpha spheres of proper radius have to be identified.

In the case of MDpocket, alpha spheres of proper radius are detected on every frame of the trajectory. Then, a 1 Å spaced grid is placed over the first frame of the trajectory and each filtered alpha-sphere of the whole trajectory is assigned to the closest grid point to its center (figure S1). To visualize the results, a density map and a frequency map are generated. The density map allows capturing very rare openings of cavities as it is generated by summing the number of alpha spheres assigned to each grid point and then normalizing the resulting value by the number of snapshots.

2.2 Python-Based Pipeline For Automatization Of The MDpocket Workflow

An automated pipeline has been developed for the detection and tracking of pockets over the GPCRmd database. The pipeline consists of 4 major steps: input processing, MDpocket detection, DBSCAN clustering, and MDpocket characterization shown in Figure 1:

Dataset: The dataset consisted of GPCR MD simulations obtained from GPCRmd server. GPCRmd is a database of GPCR MD simulations containing more than three thousand GPCR simulations from over 60 different GPCR subtypes. From most of the static structure, GPCRmd provides 3 replicates from both apo and complex state. As a result, it has been analyzed more than 1500 simulations providing by more than 500 different static structures.

Input processing: As input, the trajectory file with extensions .xtc or .dcd and its corresponding PDB model file are required. If the files are downloaded and consequently named in GPCRmd database format, the algorithm automatically matches the trajectory file with its corresponding model file. In case the trajectory and model file are not obtained from GPCRmd database, the user can run the program over them using -t and -p flags respectively. As MDpocket needs as input a prealigned trajectory, an optional argument (-a, --structural_alignment) is available to perform a structural alignment using the GMXapi python package from GROMACS (Lindahl et al., 2001). In addition, using the same package, the trajectory is transformed into a set of PDB files, one for each frame, that only contain the protein structure. Aiming to save computational cost in further analyses, not all the snapshots are transformed into PDB files, as the user can select a step value to reduce the trajectory (e.g. step value of 5 will transform 1 in 5 frames into a PDB file). Lastly, a text file is generated listing the path to every snapshot which is required as input to run MDpocket.

MDpocket: The program is run for the first time in order to obtain the frequency and density maps being the last one used in further steps as it better suits the detection of transient pockets. From the density map, the isovalue parameter is crucial for the identification of conserved pockets. The isovalue can be expressed as the number of alpha sphere centers in an 8 Å cube around each grid point per snapshot, so, the more conserved a cavity is the higher the isovalue of the corresponding grid points. Hence, in the density map, we obtain a grid map that contains the coordinates that have been detected as pockets at a given isovalue. In the MDpocket manual, is recommended to use an isovalue between 2 and 3 to detect transient pockets. Consequently, the pipeline extracts the grid points that have an isovalue higher than 3 to continue with further steps.

Clustering of grid points into pockets: Once we have the coordinates that are considered potential pockets, we need to cluster these coordinates into different pockets. Thereby, the algorithm uses Density-based spatial clustering of applications with noise (DBSCAN) using the python package scikit-learn (Pedregosa et al., 2011).

This algorithm identifies distinctive clusters of points (pockets) in the data (density map), by looking at the local density of the data points. There are some characteristics that make DBSCAN algorithm fit our requirements as it does not require specifying the number of clusters beforehand, performs well with arbitrary cluster shapes, and most importantly, it is robust to outliers being able to detect small pockets of poor interest.

Two essential parameters need to be set in the DBSCAN algorithm: epsilon, which is the maximum distance between two points to be considered as the same cluster and minPoints which is the minimum number of data points required in a cluster to not be considered as an outlier. These parameters were adjusted (epsilon = 1.5 and minPoints = 8) to GPCR after a calibration considering different GPCR-MD simulations.

Finally, one PDB file for each cluster containing dummy atoms at the positions of grid points that we consider as a pocket is created. These resulting PDB files are required for MDpocket as input for a 2nd run of the program in order to track the properties of the pockets.

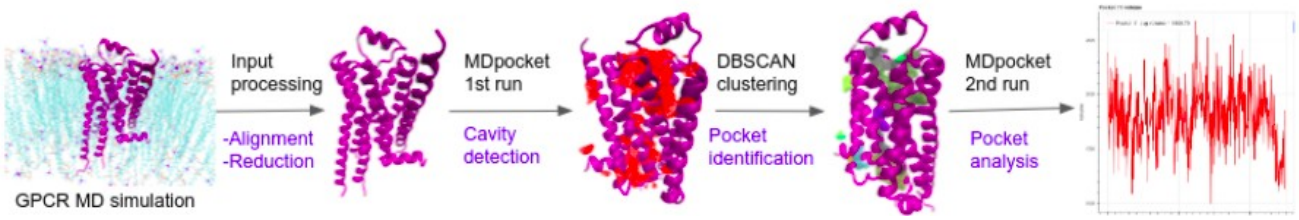

Fig. 1. Workflow of the automatization of pocket detection. Workflow of the pipeline used for the automatization of pocket detection .

Pocket characterization: A second run of MDpocket is needed to analyze the selected pockets to determine all pocket descriptors corresponding to the selected area for the whole ensemble. The calculation of descriptors for each pocket is the time-limiting step, to speed up this analysis, it has been implemented a parallelization step that allows using multiple cores of a CPU to analyze multiple pockets in parallel.

To run a subsequent MDpocket analysis, the PDB file created in the previous step is used to delimit the area of interest to be analyzed for each timeframe of the trajectory.

Once the pockets are analyzed, a text file is created containing the most important descriptors (e.g., average volume, receptor atoms defining the binding pocket, hydrophobicity density, polarity score, solvent accessible surface area of the pocket) of each pocket found over simulation. This file aims to easily detect the most interesting pockets with a quick view over the descriptors. Two PDB files are generated for the visualization of the pocket: a PDB file that contains all the alpha spheres of the pocket for each snapshot which allow you to visualize the movement of the pocket along with the simulation. This PDB file can be visualized using PYMOL. The second PDB file contains all receptor atoms defining the binding pocket along the simulation.

Analysis of detected pockets: Druggability predictions are important to differentiate pockets with the potential to accommodate drug-like ligands that modulate the protein biological function with high affinity. MDpocket provides a flag to obtain the mean druggability in the grid maps. However, much less pockets are detected using this flag. Consequently, transient pockets are rarely detected and for this reason, the druggability score flag have not been used in the pipeline.

The druggability score used by MDpocket uses 3 descriptors obtained from the pocket characterization: local hydrophobic density, hydrophobicity score and polarity score. However, when running the MDpocket 2nd run, we are delimiting the pocket using a pocket coordinate file as input, as consequence calculating the druggability score from the obtained descriptors is not relevant as the score function has not been trained with such input. For this reason, in order to differentiate between promising pockets to the ones that are poorly druggable, we will use the descriptor mean local hydrophobic density as an approximation to the druggability score. This descriptor reflects local densities of hydrophobic alpha spheres clusters in a binding site. Hydrophobicity density has been shown to be the descriptor that has a higher correlation with druggability (Schmidtke and Barril, 2010) as interactions between hydrophobic surface regions and the ligand are important, especially seen the rather hydrophobic characteristics of drug-like molecules (Vieth et al., 2003).

3. Results

Class A GPCR allosteric modulators have been discovered during the last years and some of them have been approved for clinical use (Table S1). Co-crystallized structures of the GPCR with its allosteric modulator provide crucial information on allosteric binding sites. However, crystallization of GPCRs is still a challenging task due to conformational flexibility and instability of the proteins removed from the membrane. One example of a co-crystalized structure of a GPCR with its allosteric marketed drug is the PDB 4MBS which contains the static structure of class A GPCR CCR5 with its allosteric modulator Maraviroc determined at 2,7 Å resolution (Tan et al., 2013). Its trajectory in both the apo and the complex state can be found in GPCRmd database. These simulations are used for discussion purposes in order to validate the described pipeline for allosteric pocket detection.

3.1 Case Study: Allosteric Pocket Detection On CCR5 With Allosteric Modulator Maraviroc.

CC chemokine receptor (CCR5) is a GPCR that is responsible for immune and inflammatory responses by regulating the trafficking and effector functions of leukocytes. In addition, CCR5 and CXCR4 are required as co-receptors for human immunodeficiency virus type 1 (HIV-1) infection.

Intense research in the development of antivirals to block the infection of HIV-1 through inhibition of CCR5 led to the approval of maraviroc by the US Food and Drug Administration (FDA) in 2007. Maraviroc is the first CCR5-receptor allosteric antagonist used in the treatment of HIV-1 infection. It acts as inverse agonist of CCR5, stabilizing CCR5 in an inactive conformation and, consequently, preventing the binding of HIV-1 glycoprotein gp120 to CCR5 which is necessary for blocking the entrance of HIV into human cells (Sharon et al., 2008).

In the PDB structure, Maraviroc occupies the bottom of a pocked defined by residues from helices I, II, III, V, VI and VII. As reported by Q. Tan et al. (Tan et al., 2013) there are some residues that are crucial for the anti-HIV infection activity of Maraviroc, being residues Tyr251, Tyr37, Thr195 and Thr259 interacting via hydrogen bond, Glu283 forming a salt bridge and Tyr108, Phe109, Phe112 and Trp248 creating hydrophobic interactions with the allosteric ligand. The location of the mentioned residues can be observed in figure 2A.

We have analyzed 3 apoform and 3 complex trajectories from GPCRmd which use the PDB 4MBS structure as starting coordinates with the described pipeline. To assess if the binding pocket was detected by the pipeline or not we will use the Mutual Overlap criteria which has been used in previous studies (Le Guilloux, Schmidtke et al., 2009). This criterion considers a pocket successfully identified if at least 50% of the ligand atoms lies within 3 Å of at least one alpha sphere, and if at least 20% of the pocket alpha spheres lies within 3 Å of the ligand. As expected, the results show that in all the 6 replicas, we were able to detect a pocket that overlaps with the known Maraviroc binding site and fulfill the conditions required for the MO criteria. However, in 2 complex simulations, the binding pocket of Maraviroc has been detected as 2 different pockets.

Table 1. Results obtained from the analyses of predicted pockets over the 6 GPCR trajectories. We present the total number of pockets detected over each trajectory (pockets detected), the rank of the Maraviroc pocket according to the hydrophobicity density parameter, the percentage of frames where the key residues for the binding of Maraviroc are detected as residues that are being part of the pocket and finally if the trajectory analyzed contains the ligand (complex) or not (apoform).

Trajectory	Pockets detected	Rank	Tyr251	Tyr37	Thr195	Thr259	Glu283	Tyr108	Phe109	Trp248	State
1	13	2	100	100	100	0	100	100	100	0	apoform
2	12	2	0	100	0	0	100	100	100	0	apoform
3	14	2	100	100	100	0	100	100	100	0	apoform
4	15	1	100	100	100	0	100	100	100	0	complex
5	16	1, 2	100	100	0	0	100	100	100	100	complex
6	20	1, 3	0	100	100	100	100	100	100	0	complex

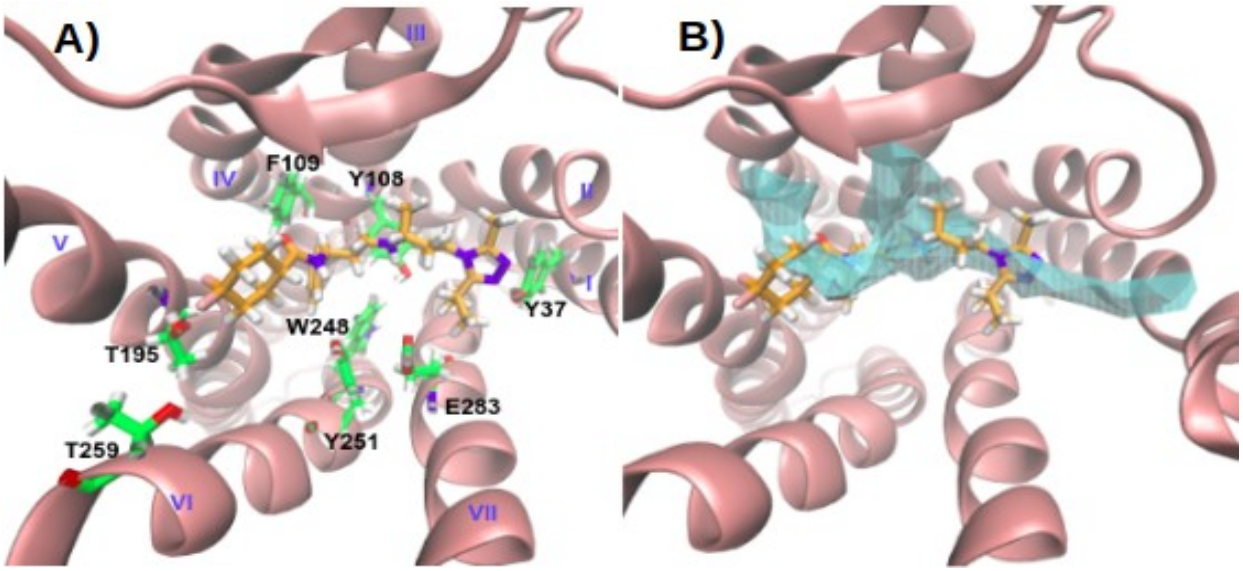

Fig. 2. Maraviroc binding pocket. A) Maraviroc (orange carbons) and crucial receptor residues involved in ligand binding (green carbons) are shown in stick representation. Other elements are colored as follows: oxygen: red, nitrogen: blue, hydrogen: white and CCR5 helices are indicated in blue. **B)** Pocket detected by the automatization of MDpocket pipeline over the trajectory 1 shown in light blue. A structural alignment of CCR5 has been performed to place Maraviroc in the apoform structure.

Table 1 provides details about geometric and physicochemical properties, rank of the predicted binding site considering the hydrophobicity density descriptor and the percentage of frame that the important residues of the binding site of Maraviroc have been detected. These descriptors show the potential of the algorithm to predict binding pockets and to differentiate between the ones that are more druggable as all the predicted pockets are in the top three ranked considering the average of the hydrophobicity density as the ordering measure. In addition, almost all the key residues for Maraviroc interaction, excluding Thr259 and Trp248 which are positioned at the most upper and lower part of the pocket respectively, have been identified in almost all the frames of the trajectories implying a precise positioning of the pocket in all the frames of the trajectory. This is also an indicator that the pocket is stable as it has been detected during the whole trajectory in all the 6 replicas.

4. Discussion

We have introduced a new open-source algorithm for the prediction of allosteric sites based on the MDpocket program. It has proved useful for the qualitative and quantitative analysis of pockets from MD GPCR simulations. The accuracy of pocket prediction is limited by two factors: the grid-based nature of MDpocket, and the choice of parameters for filtering and clustering of alpha spheres.

An optional argument to perform a structural superimposition has been implemented to face the grid-based nature of MDpocket. However, molecular dynamics with high protein motions, where for example two domains exhibit large relative motion with respect to each other, pocket detection and characterization will be altered for all pockets on the other subunit. In GPCRmd just the trajectory of the GPCR protein is obtained, for this reason a structural superimposition of the whole protein is enough to obtain a good performance of the algorithm. However, if the algorithm wants to be run in trajectories that exhibit large relative motion with respect to other subunits, it should be considered to perform a structural alignment on the cavity of interest and take into consideration

that pockets found on other places of the protein might not be representative.

On the other hand, the parameter sets (epsilon, minPoints and isovalue) have a major influence on the results. These parameters have been assessed for detection pockets on GPCR trajectories. However, these parameters might have to be adjusted to perform well with no GPCR simulations. For this reason, optional arguments have been implemented in order to decide which are the parameter sets that fits better with user's requirements. The choice of these parameters allows adjusting the distance between two points to be considered the same pocket (epsilon), the minimum number of points that a cluster should have in order to be considered a pocket and not an outlier (minPoints), and the frequency of the pocket over the simulation (isovalue). The influence of these parameters is shown in figures S2, S3 and S4.

The final limitation is linked to the length of the trajectories. To be able to detect and measure the frequency of opening of transient channels in the density map, a considerable simulation time might be required. To face this problem, GPCRmd provides three MD replicas of 0,5 μs originated from each static structure. As a result, an accumulated time of 1,5 μs is analyzed from each static structure.

A web-based interactive tool for visualizing and analyzing the results obtained from running the algorithm over GPCRmd database is being developed. This tool will be part of the GPCRmd web server and will provide information about the pockets detected over the different simulations. It will also provide the most important descriptors of each pocket to see which are the most promising ones. We surely envisage that this implementation is aspiring to be a great tool for the researchers to improve the discovery of new potential allosteric modulators on GPCRs.

The computational cost of the pipeline is dependent on some parameters: protein size, trajectory length and number of pockets detected. For the characterization of each pocket, a MDpocket 2nd round must be run over each pocket detected being the most time limiting step. To reduce the computational cost a parallelization processing using multiprocessing module have been implemented allowing to characterize multiple pockets on the same computer at the same time. The reduction of time obtained with this implementation is substantial: the pipeline using 10 CPUs takes 34 minutes to run over a Beta-2 adrenergic receptor trajectory (PDB: 5JQH) whereas it takes 3 hours and 9 minutes to run without the parallelization step.

We aim that our finding could be of great interest as it aims to improve the knowledge about allosteric regulation on GPCRs as it will allow one to have a specific view of the differences and similarities of allosteric pockets between different GPCRs subtypes enhancing allosteric drug development on GPCRs.

Acknowledgements

Molecular graphics performed with VMD, developed by Theoretical and Computational Biophysics Group, NIH Center for Macromolecular Modeling and Bioinformatics, at the Beckman Institute, University of Illinois at Urbana-Champaigs.

Conflict of Interest: none declared.

References

Conn, P., Christopoulos, A., Lindsley, C. (2009) Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. Nat Rev Drug Discov 8, 41–54.

Hauser, A., Attwood, M., Rask-Andersen, M. et al. (2017) Trends in GPCR drug discovery: new agents, targets and indications. Nat Rev Drug Discov 16, 829–842.

Le Guilloux, V., Schmidtke, P. & Tuffery, P. (2009) Fpocket: An open source platform for ligand pocket detection. BMC Bioinformatics 10, 168.

Lindahl, E., Hess, B. & van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 7, 306–317.

Matosin. N, et al. (2014) Metabotropic glutamate receptor mGluR2/3 and mGluR5 binding in the anterior cingulate cortex in psychotic and nonpsychotic depression, bipolar disorder and schizophrenia: implications for novel mGluR-based therapeutics. Journal of Psychiatry and Neuroscience, 39, 407-416.

Miao Y., Bhattarai A., Nguyen A., Christopoulos A. (2018) May Structural basis for binding of allosteric drug leads in the adenosine A 1 receptor. Science. Rep., 8.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011) Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Rodríguez-Espigares, I., Torrens-Fontanals, M., Tiemann, J.K.S. et al. (2020) GPCRmd uncovers the dynamics of the 3D-GPCRome. Nat Methods 17, 777–787.

Schmidtke, P., Bidon-Chanal, A., Luque FJ, Barril X. (2011) MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*. 27(23), 3276-3285.

Schmidtke P, Barril X. Understanding and predicting druggability. (2010) A high-throughput method for detection of drug binding sites. J Med Chem. ;53(15):5858-5867.

Serrano-Marín J., Reyes-Resina I., Martínez-Pinilla E., Navarro G., Franco R. (2020) Natural Compounds as Guides for the Discovery of Drugs Targeting G-Protein-Coupled Receptors. Molecules. 25(21):5060.

Sharon S., Horatio B., Juan C. (2008). Maraviroc: A CCR5-receptor antagonist for the treatment of HIV-1 infection. Clinical Therapeutics 30(7), 1228-1250.

Poupon A. Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. (2004). Curr Opin Struct Biol. 14(2):233-241.

Tan Q., Zhu Y., Li J., Chen Z., Han G., Kufareva I., Li T., Ma L., Fenalti G., Li J., Zhang W., Xie X., Yang H., Jiang H., Cherezov V., Liu H., Stevens R., Zhao Q., Wu B. (2013). Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. Science (New York, N.Y.), 341(6152), 1387–1390.

Vieth. M., Siegel. G., Higgs, E., Watson, A., Robertson, H., Savin, A., Durst, L., Hipskind, A. (2003) Characteristic physical properties and structural fragments of marketed oral drugs. J. Med. Chem 47(1), 224-232.

Wu H., Wang C., Gregory K., Han G., Cho H., Xia Y., Niswender C., Katritch V., Meiler J., Cherezov V, Conn P., Stevens R. (2014) Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. Science 344, 58–64.

Supplementary Materials

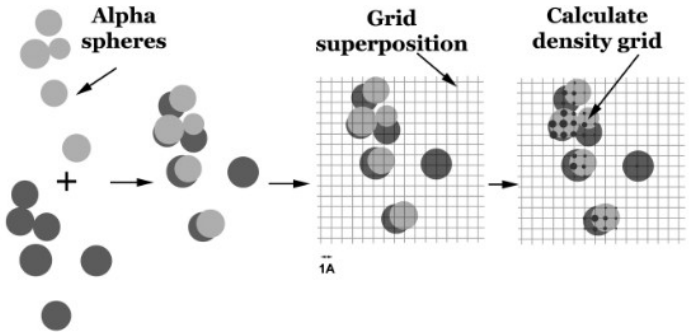

Figure S1: Schematic representation of the obtention of the density grid. Alpha spheres are detected on different snapshots (dark grey and light grey indicating different snapshots). A 1 A spaced grid is superimposed to the alpha spheres and on each grid point the density of surrounding alpha spheres is calculated. Image extracted from MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*. 27(23), 3276-3285

Table S1: Allosteric modulators of class A GPCRs approved for clinical use.

Family	Target	Name	Mechanism of action	Indication
P2Y receptor	P2Y	ticagrelor	Allosteric antagonist	anti-thrombosis
Chemokine receptors	CCR5	maraviroc	NAM	HIV infection
Chemokine receptor	CXCR4	plerixafor	NAM	Bone marrow transplantation

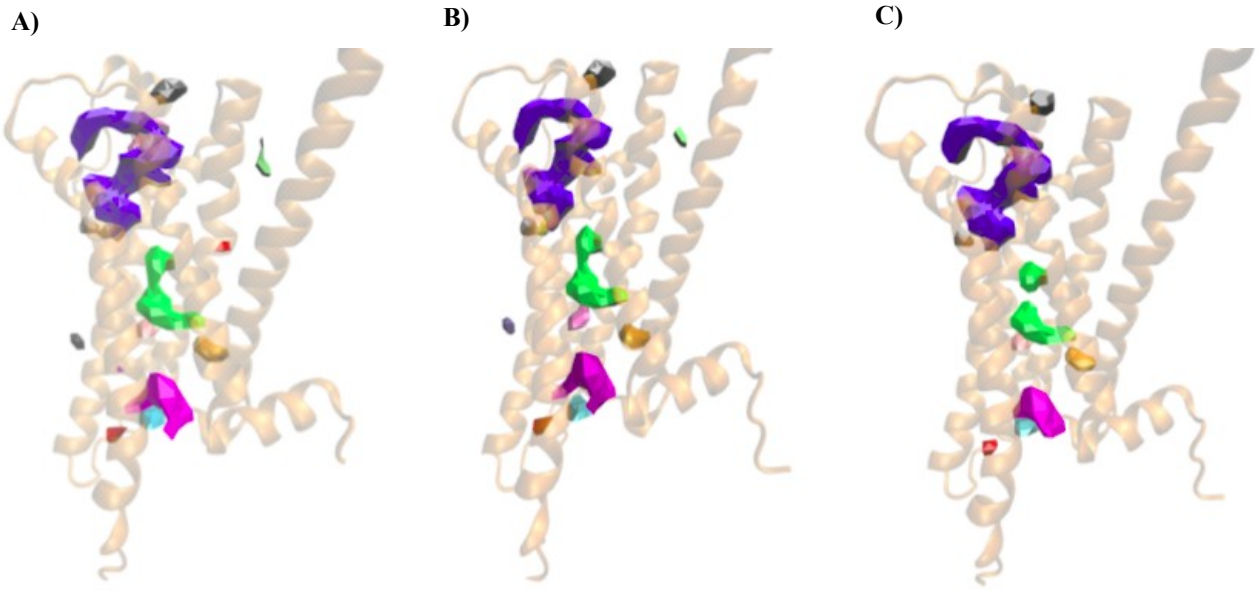

Figure S2: Representation of how minPoints parameter effects the characterization of pockets. Three different minPoints values (0, 8, 15) have been used to run the pipeline over the beta-2 adrenergic receptor (PDB: 6PS5 and GPCRmd trajectory file ID: 15443) and the resulting pockets considered for the characterization step are shown in figures A, B and C respectively. Each coordinate point detected is colored according to the pocket it belongs. The pipeline detects a total of 18 different pockets over the trajectory. In this situation, MinPoints parameter allow to filter the pockets that have to be characterized in posterior steps depending on its size: in the figure 3A using a minPoints value of 0, all the pockets detected are characterized as even the pockets that only contains one coordinate point are considered for the analysis. In figure C the pockets that have less than 15 coordinate points are considered outliers, and consequently, are not characterized using a second round of MDpocket.

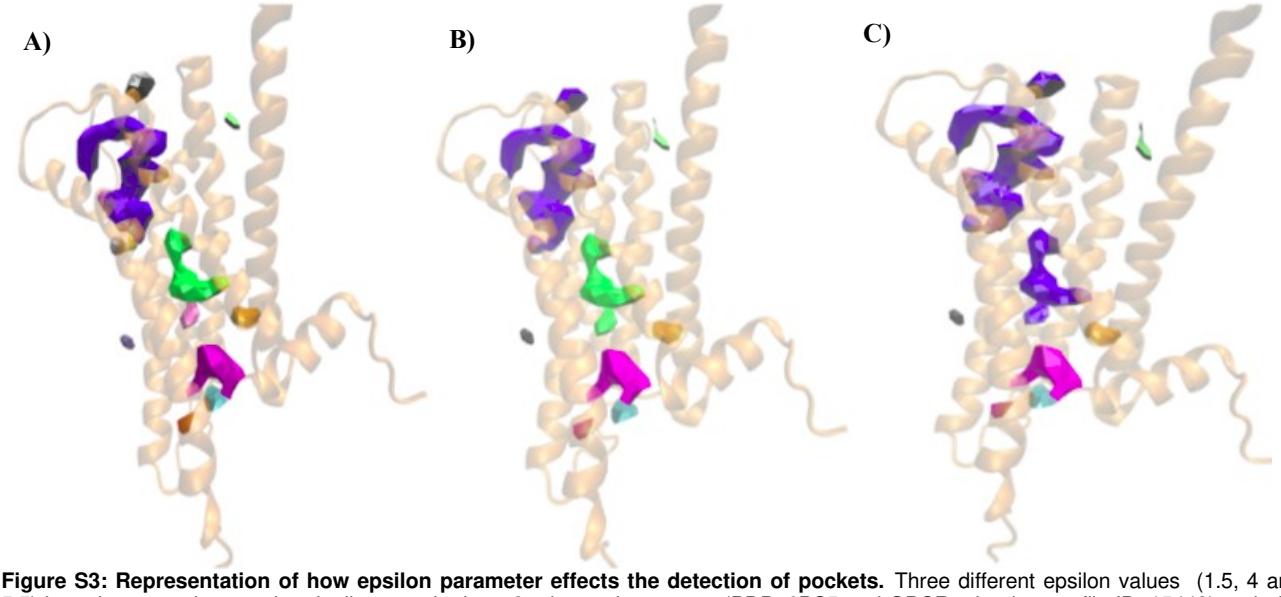

Figure S3: Representation of how epsilon parameter effects the detection of pockets. Three different epsilon values (1.5, 4 and 5.5) have been used to run the pipeline over he beta-2 adrenergic receptor (PDB: 6PS5 and GPCRmd trajectory file ID: 15443) and the resulting pockets considered for the characterization step are shown in figures A, B and C respectively. Each coordinate point detected is colored according to the pocket it belongs. Epsilon allows adjusting the distance between two points to be considered the same pocket. In figure A a low value of epsilon in used and consequently coordinates that are not adjacent in the grid map, are considered as different pockets. However, when increasing the value of epsilon, the maximum distance between two coordinates to be considered the same cluster becomes larger and consequently for expanding clusters becomes larger and consequently broader pockets are detected.

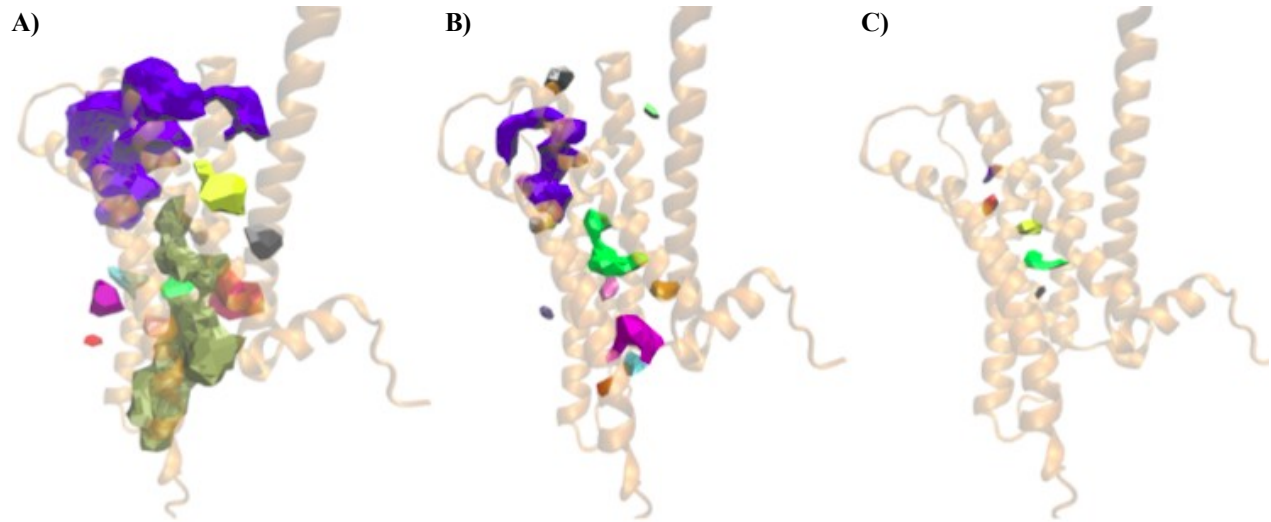

Figure S4: Representation of how isovalue parameter effects the detection of pockets. Three different isovalue (1, 3 and 5) have been used to run the pipeline over he beta-2 adrenergic receptor (PDB: 6PS5 and GPCRmd trajectory file ID: 15443) and the resulting pockets considered for the characterization step are shown in figures A, B and C respectively. Each coordinate point detected is colored according to the pocket it belongs. Isovalue parameter allow to obtain more or less conserved cavities during the trajectory. In figure A low conserved coordinates are detected by the pipeline. However, in figure C only very conserved pockets are detected.