

On the allosteric puzzle and pocket crosstalk through computational means

Riccardo Aguti,^{1,2} Mattia Bernetti,^{1,2,*} Stefano Bosio,^{1,2} Sergio Decherchi,^{3,§} Andrea Cavalli^{1,2,§}

1 Computational & Chemical Biology, Fondazione Istituto Italiano di Tecnologia, Via Morego 30, I-16163 Genova, Italy

2 Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy

3 Data Science and Computation, Fondazione Istituto Italiano di Tecnologia, Via Morego 30, I-16163 Genova, Italy

* Author to whom correspondence should be addressed: mattia.bernetti@iit.it

§ SD and AC share co-last authorship

Abstract

Allostery is a constitutive, albeit often elusive, feature of biomolecular systems which heavily determines their functioning. Its mechanical, entropic, long-range, ligand, and environment-dependent nature creates far from trivial interplays between residues and, in general, the secondary structure of proteins. This intricate scenario is mirrored in computational terms as different notions of “correlation” among residues and pockets can lead to different conclusions and outcomes. In this contribution, we put on a common ground and challenge three computational approaches for the correlation estimation task and apply them to three diverse targets of pharmaceutical interest: the Adenosine A_{2A} Receptor, the Androgen Receptor and the EGFR kinase domain. Results show that partial results consensus can be attained, yet different notions lead to pointing the attention to different pockets and communications.

I. INTRODUCTION

To exert their function, biomolecular systems often require concerted, either global or local, conformational changes. Communications networks within these systems hence arise and represent constitutive features, which collectively go under the name of allosteric regulation.¹⁻³ Since its historical conception,^{4,5} the term allostery has evolved to indicate any type of long-distance signaling caused by a local perturbation at a distal site, including protein-protein interaction, covalent modifications, mutations, and the classic binding of small molecule effectors at sites other than the orthosteric one.²⁻⁵ The result of allosteric regulation is the modulation of the biomolecular activity.^{1,5,6} Despite conceptually straightforward, the mechanism through which allostery is accomplished can be highly complex. Indeed, several models have been proposed.²⁻⁵ Due to this state of affairs, allosteric pathways remain nowadays largely uncharacterized.⁷ Besides the remarkable implications for fundamental science, characterizing allosteric mechanisms can have practical outcomes of great significance. In drug discovery, for instance, potential advantages associated with the development of allosteric modulators are, among others: high selectivity, safer dosage, and reduced risk of toxicity or side effects.⁸⁻¹⁰ From an experimental point of view, allostery is rather challenging to investigate through the most commonly used methods, including X-ray crystallography, NMR, and mutagenesis.^{1,7,11} Additionally, studies to assess allosteric mechanisms are typically conducted to compare the structural dynamics of biomolecules in the presence and absence of the triggering effector promoting the allosteric event. This is also typically the case for computational studies,^{6-8,10,12} where

simulations and analyses of a biomolecular target are, for instance, done both in the presence and in the absence of ligands.^{13–15} The simulative outcomes are then compared to disentangle the relevant allosteric hotspots. In a dynamics-based view of allostery,^{2–4,6,16} structural fluctuations in the biomolecule can account for intramolecular communication between distinct binding sites,^{2,17} and can hold true even in situations where no ligand may be present (apo protein).^{2,6} This suggests that, in some cases, hallmarks of allostery can be an intrinsic feature of biomolecules,^{18–20} which could be appreciated also in the apo forms.

In this manuscript, we employ computational methods to study pocket intramolecular communication in different biomolecules in the absence of ligands. We compare the results of three different methods by applying them to the same, diverse set of relevant biomolecules that are reported to display allosteric behaviors. Specifically, we study three pharmaceutically relevant targets: the adenosine A_{2A} receptor (A_{2A}), the androgen receptor (AR), and the epidermal growth factor receptor (EGFR) kinase domain. A_{2A} is a transmembrane receptor involved in pathological conditions such as inflammation, cancer, and Parkinson's disease,^{21,22} AR is a nuclear receptor that plays a crucial role in the growth and progression of prostate cancer,²³ while the EGFR plays crucial roles in the signaling pathways controlling cell growth, differentiation, and survival.²⁴ The goal is to assess to what extent some information about allosteric communication can be retrieved from the apo simulations by the three methods, also comparing their outcomes. Our procedure is based on the combination of molecular dynamics (MD) simulations²⁵ with graph theory to achieve a more intuitive interpretation of the results. Figure 1 provides a schematic representation of the pipeline we followed.

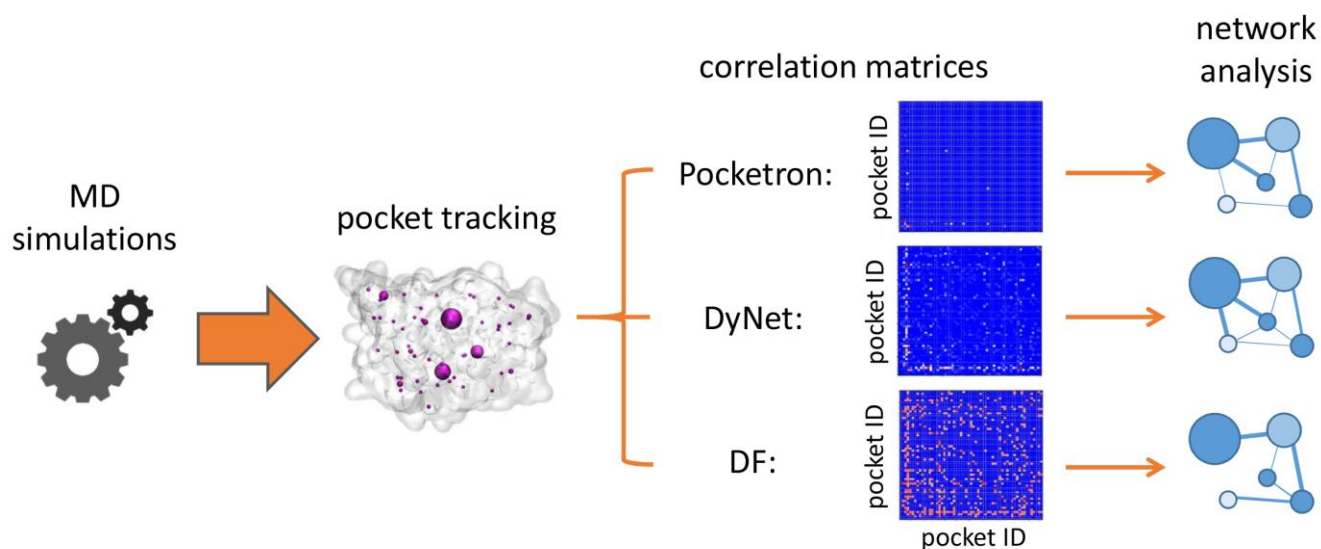


Figure 1. Schematic of the computational workflow followed in the present study. Pockets are detected and tracked via Pocketron on the MD trajectories. Pocketron¹³, Dynamic network analysis²⁶ and Distance fluctuation analysis¹⁵ are used to build proper correlation matrices, and lastly network analysis is done all the resulting graphs.

First, we perform μ s-long plain MD simulations. We then identify pockets on the MD-sampled structures using a pocket-tracking algorithm named Pocketron¹³. We further analyze correlations between the identified pockets through three different computational strategies, namely a pocket crosstalk analysis,¹³ a dynamic network analysis,²⁶ and a distance fluctuation analysis.¹⁵ Finally, to allow for a more intuitive interpretation of the results and an effective comparison among the three different approaches, we visualize their results in the form of network diagrams, where we visualize the highest-ranking communications.

II. METHODS

MD simulations

Molecular dynamics simulations were conducted on three distinct systems: the Adenosine A_{2A} Receptor (modelled as reported in a previous work¹³ using PDB IDs 3UZC and 4E1Y), the Androgen Receptor (PDB ID: 1T63), and the EGFR kinase domain (PDB ID: 2JIT). For the EGFR system, the MD trajectory was taken from Ref.²⁷. To prepare the systems, ligands were removed, while crystal water molecules and ions were retained. The systems were simulated using the ff99SB-ILDN force field,²⁸ with the TIP3P model²⁹ for water molecules and using a cubic box, resulting in systems of around 65000, 75000, 61000 atoms for A_{2A}, AR and EGFR, respectively. Na⁺ and Cl⁻ ions were added for neutralization. All systems were prepared with BiKi Life Sciences.³⁰ The systems were equilibrated in both the NVT, through three subsequent stages at 100, 200 and 300 K, and the NPT ensembles to achieve a temperature of 300K and pressure of 1 bar. The A_{2A} system, embedded in a POPC membrane, underwent a longer equilibration in the NPT ensemble (Figure S1). Specifically, this system was simulated in the NPT ensemble for 50 ns with restraints on the heavy atoms of the protein, followed by an additional 50 ns simulation without restraints to fully relax the system. Subsequently, molecular dynamic simulations were performed on all three systems for a total of 3 μ s for the A_{2A} Receptor and EGFR and 2.6 μ s for the Androgen Receptor. Simulations were conducted using GROMACS 2021.4³¹ with a time step of 2 fs. From the simulations of each system, 50000 frames were extracted (about 48000 for the Androgen Receptor) for pocket detection and tracking through the Pocketron algorithm¹³ available in BiKi Life Sciences.

Pocketron's pocket crosstalk analysis

Pocketron¹³ is an algorithm that allows tracking pockets on a molecular dynamics trajectory. The static (per frame) pocket detection algorithm at its basis is NanoShaper³² (freely available at <https://gitlab.iit.it/SDecherchi/nanoshaper/>) and uses the solvent excluded surface (SES) or Connolly-Richards surface as a basis. This surface is created by rolling a spherical probe over the van der Waals surface of the system. Pockets are identified by comparing the regions enclosed by the SESs obtained with two different probe radii (by default 1.4 and 3 Å). Retained pockets are those with volume values over 34.5 Å³, the equivalent volume of 3 water molecules. The algorithm tracks the atoms within pockets as they change over time in the trajectory. The algorithm tracks pockets through the assignment of unique pocket identifiers (pID) to each pocket identified in the first frame, as defined by their constituent atoms. These initial pockets are stored in a list, and for each subsequent frame, the algorithm recalculates the pockets and compares them to the stored ones. In instances where there is a discrepancy between the newly calculated pockets and those stored, the current pocket is added as a new entry in the stored pockets list and assigned a new pID. The algorithm monitors all atom-based events without needing any prior knowledge or structural alignment of the region being analyzed. It tracks the atoms "exchange" between different pockets, which is used as an indicator for "pocket crosstalk". This analysis includes "merging" and "splitting" events, which are calculated by comparing the sets of atoms within each pocket at different points in time. A "merge" event occurs when atoms that were in separate pockets in a previous frame now belong to the same pocket. A "split" event occurs when atoms that were in a single pocket in a previous frame now belong to multiple pockets. The outcome of the tracking process is a set of pockets ID, the associated atom indices, and a pair of split merge matrices, which represent the propensity in "exchanging" atoms. As in to obtain a symmetric and single split-merge matrix of pockets' we first symmetrize the split and merge matrix in isolation and then we get the average matrix.²⁷ This last matrix is used for the subsequent network analysis. Notably, Pocketron has

recently been applied to a structurally complex target (i.e., tubulin) identifying innovative pockets,³³ subsequently targeted in the context of an anticancer drug discovery program.³³

Generalized correlation-based dynamical network analysis:

Another method we challenged to study the dynamic motion and communication between residues in a MD simulation is a graph-based approach, named Dynamic Network (DyNet) analysis.²⁶ This network analysis is based on correlated protein motions: it allows modeling of how communication signals travel through the protein structure, thus possibly identifying allosteric pathways.

In this method, pair-correlations between motions of two amino acid residues are quantified through the so-called generalized mutual information, as defined by Lange and Grubmüller:³⁴

$$r_{MI}[x_i; x_j] = \left\{1 - e^{-2I[x_i; x_j]/d}\right\}^{1/2} \quad 6$$

where I is the mutual information²⁶ and d is the dimensionality of the variables x_i and x_j . The value of r_{MI} measures the correlation between the motions of two residues along a MD simulation.

To construct the protein graph, each node is associated with the alpha-carbon (C_α) of an amino acid residue in the primary sequence. Then, an adjacency matrix (A) is created, which defines the connections (edges) between the nodes.

The connections are determined using a distance cutoff (typically 3.5-5.5 Å) that corresponds to pairs of residues that are at chemical distance. The distances between heavy atoms of the two residues above the cutoff are filtered out. This first geometric cutoff is supplemented by a statistical time-percentage cutoff, which reflects how long two residues are in contact during the MD trajectories. For a fixed cutoff distance, the contacts are evaluated along the trajectories, and their existence is determined based on the percentage of frames (from 65% to 85%) in which these contacts are present.

The generalized mutual information can be used to weight the protein graph. The obtained graph, called dynamical weighted network, models the information exchange within a protein based on its dynamics. To create the dynamical weighted network, the non-zero entries of the adjacency matrix are filled with edge weights w_{ij} , which are derived as $-\log[r_{MI}[x_i, x_j]]$ for all pairs of residues i and j that are connected in the network. Conversely, if two nodes (i and j) are not connected by an edge, the corresponding entry in the matrix (A_{ij}) will be zero. In a graph weighted by w_{ij} coefficients, the adjacency matrix A_{ij} represents which pairs of residues exchange information (through correlated motions) and are therefore linked in the dynamical weighted network.

The dynamical weighted network contains information paths and critical nodes that are important for communication within the protein system. Given this architecture, the edge betweenness, EB,³⁵ is defined as the number of shortest pathways (SPs) that cross a given edge, measuring of how much this edge is responsible for connecting the other pairs of nodes in the network. The SPs within the protein communication network can be calculated using the Floyd-Warshall algorithm.³⁶ Therefore, the edges with the highest betweennesses are those conveying the highest amount of information exchange.

The EB can be used to find “communities,” or in other words, clusters. These communities can be found according to a clustering procedure.³⁷ In this work, the communities, instead, are directly derived from the residues pocket membership relation; in other words, there is a one-to-one correspondence between pockets and communities. Next, we still take advantage of coarse graining, where linked communities (linked pockets) are connected by edges that are weighted by the inter-communities EB (IEB), i.e., the sum of the EBs associated with the pairs of residues connecting the two communities. The IEBs indicate the strength of the communication flow between pairs

of communities and thus represent a simplified way to represent the communication associated to correlated protein motions within a protein.

Practical guidance to perform the dynamical network analysis is implemented in the form of Jupyter notebooks,²⁶ which we used in the present work. The analyses were performed by defining as nodes the C_α Carbons of the systems (305 for A_{2A}, 250 for AR and 289 for EGFR), using a distance cutoff of 5.5 Å along with a percentage existence of 75%, and a unique block comprising 50000 frames.

Distance fluctuation analysis

This analysis method, developed in the Colombo's lab,¹⁵ extended and adapted the approach proposed by Bahar and colleagues¹⁷ to the analysis of all-atom molecular dynamics simulation trajectories. The analysis of signal propagation is here based on an elastic network model, and it allows an understanding of the signal transduction events in proteins by evaluating the fluctuations of atoms during the dynamics. The method is based on the distance fluctuation, otherwise known as communication propensity, defined as the mean-square fluctuation of the inter-residue distance d_{ij} , computed between C_α atoms of residues i and j :

$$DF = \langle (d_{ij} - \bar{d}_{ij})^2 \rangle \quad 3$$

where \bar{d}_{ij} is the average value of the distance over the whole simulation. Once computed on each $C_\alpha - C_\alpha$ pair, the result is an $N \times N$ DF matrix where N is the number of C_α atoms in the structure. Low DF values can be obtained for residues on the same structure (e.g. alpha helix); those fluctuations are low but not interesting. Hence, as suggested by Colombo and coworkers,^{15,38} not only one detects low fluctuating pairs (under a given threshold) but also applies a minimum distance threshold on \bar{d}_{ij} to avoid detecting obvious correlations. This double filtration allows the detection of small fluctuations which are induced by residues that are not too near, rendering the results not trivial. Overall this requires to set two thresholds; an upper bound on DF and a lower bound on \bar{d}_{ij} . To estimate the DF upper bound we used, as previously suggested,¹⁵ the average DF value for consecutive amino acids along the sequence, calculated considering for each residue i the neighbors comprised between $i - 4$ and $i + 4$. Hence, the value is system dependent, as it reflects the system-specific local fluctuations; we obtained DF upper bound values of 0.06, 0.04, and 0.10 Å² for the A_{2A}, AR, and EGFR systems, respectively. For the second threshold, namely, the lower bound to \bar{d}_{ij} , we chose the value of 8 Å, to ensure filtering out all trivially correlated atomic distances.

Coarse graining of the correlation matrices and construction of the network diagrams

Comparing the results of the three different methods explored in this work is not trivial, as they provide results in different forms. Pocketron's crosstalk analysis produces an $M \times M$ matrix where the element ij indicates crosstalk probability between pockets i and j . This crosstalk probability is computed from the number of merging and splitting events occurring during the simulation. The DyNet method generates a weighted network with edge betweenness associated with pairs of nodes corresponding to C_α carbons. Finally, the DF method generates an $N \times N$ matrix of pairwise distance fluctuations between the N C_α carbons in the system. To normalize these representations to a common ground, we aggregated the information provided by the DyNet and DF approaches in a pocket-oriented fashion to highlight correlations between entire pockets as done in other works.^{10,37} Following the procedures reported in these works, we dissected the weighted network generated with DyNet into communities corresponding to the pockets identified by Pocketron. As a result, we obtained an $M \times M$ matrix, with M being the number of pockets, where the element ij is the sum of the edge betweenness associated with the pairs of residues

connecting pockets i and j . Analogously, the original DF matrix was aggregated into an $M \times M$ matrix, where the element ij is the average of the DFs between the residues comprised in pockets i and j .

A practical issue in the construction of the $M \times M$ matrices is that the pocket definition by the Pocketron algorithm allows for a residue to be part of two different pockets. This reflects the dynamical nature of pockets, which can undergo merge and split events, as detected by the algorithm. Coarse graining the information produced by the DyNet and DF methods into $M \times M$ matrices becomes thus non-straightforward. To address this issue, each time that pockets i and j displayed an overlap in residues, the computation was repeated twice, once assigning to pocket i all the residues in common, and once to pocket j . We then took the highest and lowest value in the case of DyNet and DF, respectively, corresponding to the highest correlation in both methods.

For convenience, as the DF approach rewards lower DF values, the values of the $M \times M$ matrix were converted into a more intuitive distance fluctuation score (DFS):

$$DFS = e^{-\beta DF} \quad 4$$

which leads to values in the range (0,1). To cover the full range [0,1] we further linearly rescaled the DFS result in this range. In this way, higher DFS values (closer to 1) indicate higher correlation. The parameter β plays the role of the inverse of the variance of the distribution (see Figure S2), and a value of 10 was selected to improve the discrimination of values closer to 1 by spreading the distribution towards lower DFS values. Upon coarse graining into pockets, we checked the dependence of the matrix entries on the distance threshold value. For the A_{2A} , AR, and EGFR systems we show in Figure S3, S4, and S5 the correlation matrices for thresholds values 6, 8, 10 and 15 Å.

For consistency, the $M \times M$ matrices obtained with Pocketron and DyNet were also normalized in the range [0,1]. The resulting matrices for the three systems are reported in Figure S6, S7, S8.

Finally, the correlations contained in the $M \times M$ matrices were displayed in the form of network diagrams (Figure 2, 4, and 6). In our representation, network nodes correspond to pocket identities (pIDs), with radius proportional to the average non-zero volume sampled for the identified pockets during the MD simulations and the color scale indicating pocket persistency, which is defined as the fraction of MD snapshots with non-zero volume. The elements ij of the correlation matrices are represented as edges connecting the nodes corresponding to the involved pockets, with thicker edges indicating higher correlation. For clarity, we displayed in the networks only the highest ranking (i.e. top 30) correlations for each system, so as to straightforwardly visualize the most relevant communications, and rescaled such edges to the same range for the three systems to improve discrimination between higher and lower correlations. Network diagrams were produced with the NetworkX³⁹ Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. The nodes were projected in a 2D space obtained through Multidimensional scaling on the Cartesian coordinates of the pockets' centres of mass. This allowed obtaining a 2D space that reflected as much as possible the distances of the pockets in the original 3D space. As pockets closeness is preserved the interpretation of the results is eased.

III. RESULTS

Molecular dynamics simulations were performed on three systems: the adenosine A_{2A} receptor (A_{2A}), the androgen receptor (AR), and the epidermal growth factor receptor (EGFR) kinase domain. 3 μ s-long trajectories were generated for each structures once stripped of the respective ligands. Pocketron¹³ was employed to detect cavities (i.e. pockets) parsing the MD trajectories generated for the three systems. This pocket identification process represented a shared baseline as the residues' composition was used as the common invariant for the subsequent analysis. Indeed, from this pocket definition we can derive, as discussed before, correlation matrices for

Pocketron,¹³ DyNet²⁶ and DF analysis,¹⁵ which all span the [0,1] domain, where 1 means maximal “correlation”. We obtained for each system such three matrices quantifying correlations between the identified pockets (Figure S6, S7, S8). When analyzing MD simulations, higher values were assigned i) by Pocketron, to pocket pairs that more often interacted via merge and split events, ii) by DyNet, to pocket pairs displaying higher edge betweenness, and iii) by the DF method, to pocket pairs displaying small DF.

As first preliminary check, we evaluated the convergence rate of the matrix entries values. To do that (see SI Figure S9, S10, S11) we took as reference the EGFR system. We computed the Frobenius norm⁴⁰ of the difference of the matrices at 1.0 μ s, 1.5 μ s, 2.0 μ s, 2.5 μ s, 3.0 μ s (total simulation time) and the matrix obtained at 0.5 μ s. We found the DyNet method to be the fastest converging one, as after 1.0 μ s the norm keeps on oscillating around a well-defined value. The DF and Pocketron methods both prove to converge even if they require a higher simulation time; convergence is approximately attained at 2.5 μ s. The time variations of the Pocketron matrix also take into account the rate of pockets emergence along time, whereas the other methods get advantage of the final Pocketron pockets’ set. Hence, the Pocketron result is also affected by the pockets’ appearance rate along the trajectory.

A comparison of the final results from the three methods highlighted a different sparseness of the correlation matrices. In particular, the DF method showed an overall homogeneous distribution of higher and lower signals, with regions of higher density of correlation, the DyNet approach returned less signals, mostly concentrated in specific regions of the matrices, lastly Pocketron resulted in very few signals. This is related to the way the different methods assign values to the correlation matrices. For instance, Pocketron identifies pocket communication only if there is an exchange of residues comprised in different pockets, which is a rather strict condition to assess the occurrence of an event, and thus resulted in remarkably sparse correlation matrices. For the DyNet method, the betweenness is computed only for pocket pairs with distances between residues that are below a pre-defined distance cutoff, in the order of 3.5-5.5 Å,^{26,37} thus reducing by construction the possibility of direct communication for pockets that were distant in the 3D space. Finally, DFs are directly computed for each pocket pair returning non-zero values, thus, although the matrices were usually filtered via a DF upper bound, they typically displayed a larger amount of non-null entries, since also pockets at larger distances can display visible signals.

A closer look into the results revealed similarities in the regions with denser signals. For instance, in the A_{2A} system (Figure S6), DyNet and Pocketron displayed many signals associated with pocket identifiers (pID) 3, 4, and 5, with the highest values for pairs (3,5). Analogously, although the DF analysis displayed signals that are more homogeneously distributed among all pocket pairs, a similar “L”-shaped pattern formed by pIDs 3, 4, and 5 was visible. Therefore, all the three methods consistently highlighted the significance of these regions of the A_{2A} receptor in displaying pocket correlations.

Concerning the AR (Figure S7), the signals identified by Pocketron match with the denser region of DyNet. However, discrepancies can also be observed, particularly in the signal associated with the pID pair (14,39), which was determined as significant by the DyNet analysis and null by Pocketron. This suggested that, while the receptor structural dynamics displayed a correlation between regions involved in the pocket formation, the identified pockets had a low probability of undergoing merging and splitting events by interchanging atoms, apart from a few cases such as (7,13), (8,9), and (13,21). The DF analysis showed an overall pattern that is similar to the one obtained with DyNet, indicating that the observed correlations result from low fluctuations in the distances between the corresponding pocket pairs.

Finally, the results for the EFGR (Figure S8) revealed an agreement between the pocket correlation observed with DyNet and Pocketron, which resulted in rather sparse matrices of correlation with few signals in consistent regions. Specifically, multiple high signals were associated with pID 1, with strong communication between pairs (1,7), (1,34), (1,44), and (1,57) and a medium-to-high value for pair (7,34). Conversely, the DF analysis revealed correlations throughout the whole pID space.

A more intuitive interpretation of the information contained in the correlation matrices can be achieved by taking advantage of basic graph theory by representing the data as networks of nodes and edges. In particular, the nodes represent the pockets identified in the target's structure, while the edges connecting the nodes represent the correlation between pocket pairs. In the next section we provide such insights; we now describe the representation conventions we adopted. In the subsequent graph plots the node radius is proportional to the average non-zero volume of the pocket during the simulation, while the color scale reports on the pocket persistency. The thickness of the edges between the nodes is proportional to the correlation value between the corresponding pockets, as identified via the three different approaches examined in this work. To make the representation clearer, in the networks only the highest (top 30) signals are shown, corresponding to the most relevant communications between pockets. Furthermore, to facilitate the interpretation of the pocket network, nodes are projected in a 2D space obtained through Multidimensional scaling on the Cartesian coordinates of the pockets' centers of mass, which allows for considerably preserving the distances of the original 3D space. We present such results now system-wise.

Adenosine A_{2A} Receptor (A_{2A})

Figure 2 shows the results for the A_{2A} receptor (see Figure S12 for a 3D representation). The largest pocket identified (pID 4) corresponds to the orthosteric site, which is located near the extracellular region and buried surrounded by the seven transmembrane helices. Along with the orthosteric site, pIDs 3, 5 and 6 pockets display the highest persistency, meaning that they were observed in most of the simulation.

The inspection of the resulting pocket correlation network can give us insight into how the protein's different regions are interconnected, unveiling pathways of communication that can possibly be involved in protein's function and dynamics.

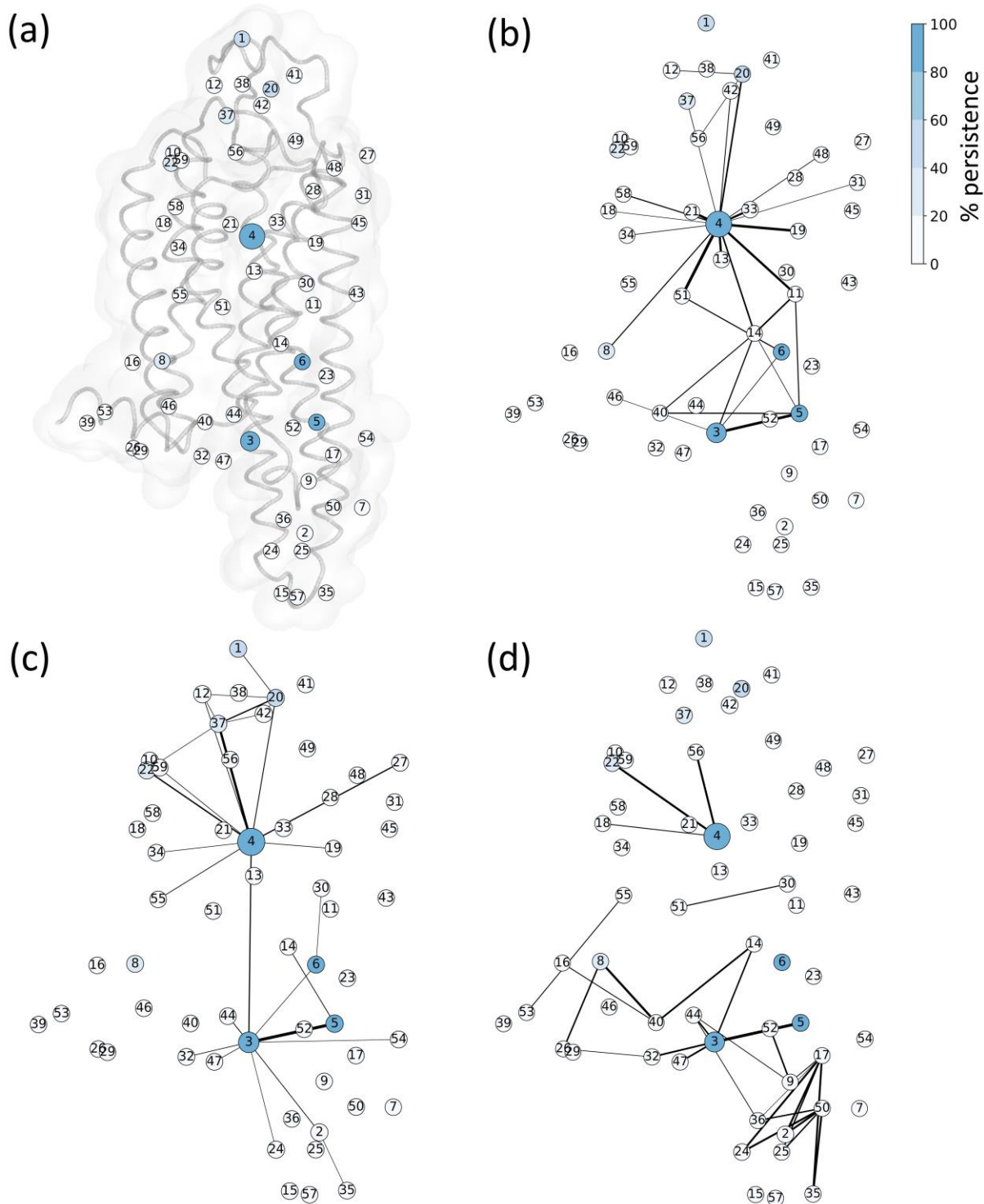


Figure 2. Network representation for the analyses on the A_{2A} system. The different nodes indicate the different pockets identified, with node size proportional to the average non-zero volume of the pockets, and color scale denoting pocket persistency along the MD simulation. In (a), the 2D node topology, obtained with multidimensional scaling on the Cartesian coordinates of pockets' center of mass, is superposed to the 3D structure of the A_{2A} receptor, in surface and tube representation, to provide the approximate location of the pockets in the 3D space. The edges connecting the nodes represent the top ranking (i.e. top 30) results from the (b) DyNet, (c) Pocketron, and (d) DF analyses of pocket communication. The edges in the three analyses are in the same scale and thicker edges indicate higher signals of the A_{2A} receptor,

in surface and tube representation, to provide the approximate location in the 3D space. The edges connecting the nodes represent the top ranking (i.e. top 30) results from (b) DyNet, (c) Pocketron, and (d) DF analyses. Thicker edges indicate higher correlation values.

According to the DyNet results (Figure 2(b)), the orthosteric site is highly connected with other surrounding pockets, showing the highest communication with pID 11, 13, 19, 33, 37, and 58. Similarly, the results from Pocketron (Figure 2(c)) also showed connections between the orthosteric site (pID 4) and its surrounding pockets, with a good overlap with the DyNet results. Most remarkably, also Pocketron identified the connection with pID 3. However, in remarkable contrast, the connection is direct in this case and was not mediated by intermediate pockets. Notably, the connection between pIDs 3 and 5 was also preserved.

Finally, the DF approach also identified connections involving the most persistent pockets 3, 4 and 5 (Figure 2(d)). However, neither direct nor indirect pathways were observed between pIDs 4 and 3. The obtained picture indicated a dense region of connections between the lower portion of the intracellular side of the receptor. Interestingly, the connection between pID 3 and 5 were also preserved in this case, as in DyNet and Pocketron. As already introduced, the DF prioritizes connections involving smaller distance fluctuations, i.e., where pocket residues move in a highly coordinated fashion. As a result, highly structured regions, such as, for instance, residues within an alpha helix structure, were more likely to return high signals. Therefore, a common practice is to filter out the system-specific local fluctuations and inspecting the distance fluctuations at increasing values of the distances. We reproduced this approach in our network representation, and observed the network obtained when focusing on inter-pocket residue distances larger than predefined cutoff values (Figure S15, S16, S17; note that only the highest ranking correlations are shown in the network representation. See Figure S3, S4, S5 for the complete correlation matrices). Interestingly for the case of the A_{2A} receptor, by filtering out DFs at distances $< 10 \text{ \AA}$ (i.e., focusing on DF associated with inter-pocket residue distances larger than 10 \AA) revealed several longer-range correlations between the orthosteric site and the surrounding pockets.

Interestingly, a pathway from pID 4 to pID 3 arose, passing through pIDs 14, which corresponds to one of the multiple pathways between pIDs 4 and 3 as identified by DyNet. Increasing the distance lower bound to 15 \AA , the connection between pID 4 and 3 was preserved and becomes direct, filtering out the intermediate pockets 13 and 14 that were connected by inter-pocket residue distances $< 15 \text{ \AA}$. As a further note, we highlighted how, while the choice of these distance cutoffs appears arbitrary, it was nevertheless reasonably related to the system size. For instance, while AR and EGFR have a maximum inter-pocket residue distance of 46 and 55 \AA , respectively, the A_{2A} receptor reaches 76 \AA , thus allowing for investigations of long-range correlations using larger distance cutoffs.

To further assess the significance and contextualize the results obtained with our analyses, we investigate the relation with respect to hotspots in the structures already reported in the literature. Concerning the A_{2A} receptor, previous studies^{41–43} suggested several non-orthosteric sites that can be suitable hotspots for allosteric modulation. Interestingly, we observed a high level of agreement between the location of these pockets and the ones identified through Pocketron (Figure 3).

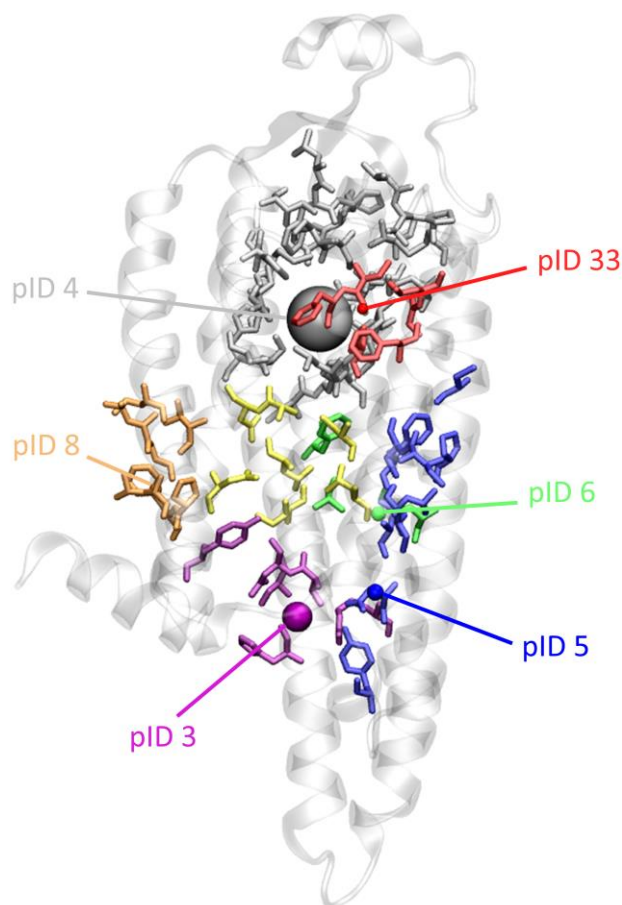


Figure 3. Comparison between relevant hotspots reported in the literature and pockets identified by Pocketron for the A_{2A} system. The different hotspots from the literature are shown in licorice with different colors, while the corresponding pockets identified with Pocketron are shown as spheres, with consistent colors. The spheres are positioned in the center of mass of the residues comprised in the pocket as identified by Pocketron, and sphere radius are proportional to the average non-zero volume of the pockets in the MD simulation. Color codes: orthosteric site in grey, the intracellular crevice in blue, G protein-coupling site in purple, the lipid interface in green, C-terminus cleft in orange, the extracellular cleft in red, and the sodium site in yellow.

Specifically, we identified pockets in the regions of the "G Protein-Coupling Site" (pID 3), the "Lipid interface" (pID 6), the "C-Terminus Cleft" (pID 8), and the "Extracellular Cleft" (pID 33).⁴² The position of the "Intracellular Crevice" (Figure 3, blue residues) is rather extended, as it can vary depending on the receptor's conformation (active, intermediate 1, intermediate 2, inactive). Interestingly, we were able to identify a small pocket (pID 5) in a subpart of this region. Consistently, the pockets located in the region of the "C-Terminus Cleft" (pID 8) and the "Extracellular Cleft" (pID 33), which are only visible in the active conformation, displayed a very small persistency in our analyses, although were found, by the study, to show some degree of correlation with the orthosteric site.

Another known hotspot of the receptor is the ion site (Figure 3, yellow residues), normally occupied by a sodium ion that can be displaced by allosteric drugs. It has been suggested that sodium binding triggers the deactivation of the system.^{43,44} In particular, the study of the movement and positioning of the sodium ion indicated that it is very stable when it is linked to the receptor's inactive form.⁴⁴ Moreover, the impact of a physiological saline concentration was tested in the simulations, and no variations in the outcomes were noticed. Additionally, none of

the sodium ions from the extracellular environment were able to pass through the narrow channel and enter the allosteric pocket. Conversely, when evaluating the active form of the receptor, research has demonstrated that the TM7 partially closes off the sodium ion binding pocket, rendering it incapable of accommodating ions.⁴⁴ Our MD simulations were performed starting from a crystal structure (PDB ID 3UZC) with the sodium in the ion pocket, which remained in place for the whole simulation. Nevertheless, the ion pocket was absent as an independent pocket in the final pocket list. For pocket identification, we used the standard Pocketron setup, where pockets need to display a suitable volume to host a probe size corresponding to about three water molecules in volume. Notably, in a previous study by some of us,¹³ where a pocket search was conducted on a 100 ns long simulation of the A_{2A} receptor, the ion pocket was correctly identified. Interestingly, in our MD simulation the ion pocket merged with the neighboring orthosteric site. As a result, the residues comprised in the ion pocket were included in the orthosteric pocket.

Androgen Receptor (AR)

The pocket search on the AR indicated the presence of three large pockets, namely pID 7, pID 8, and pID 13, as well as a medium-sized pocket pID 3, all denoted by a significant persistency over the MD simulation.

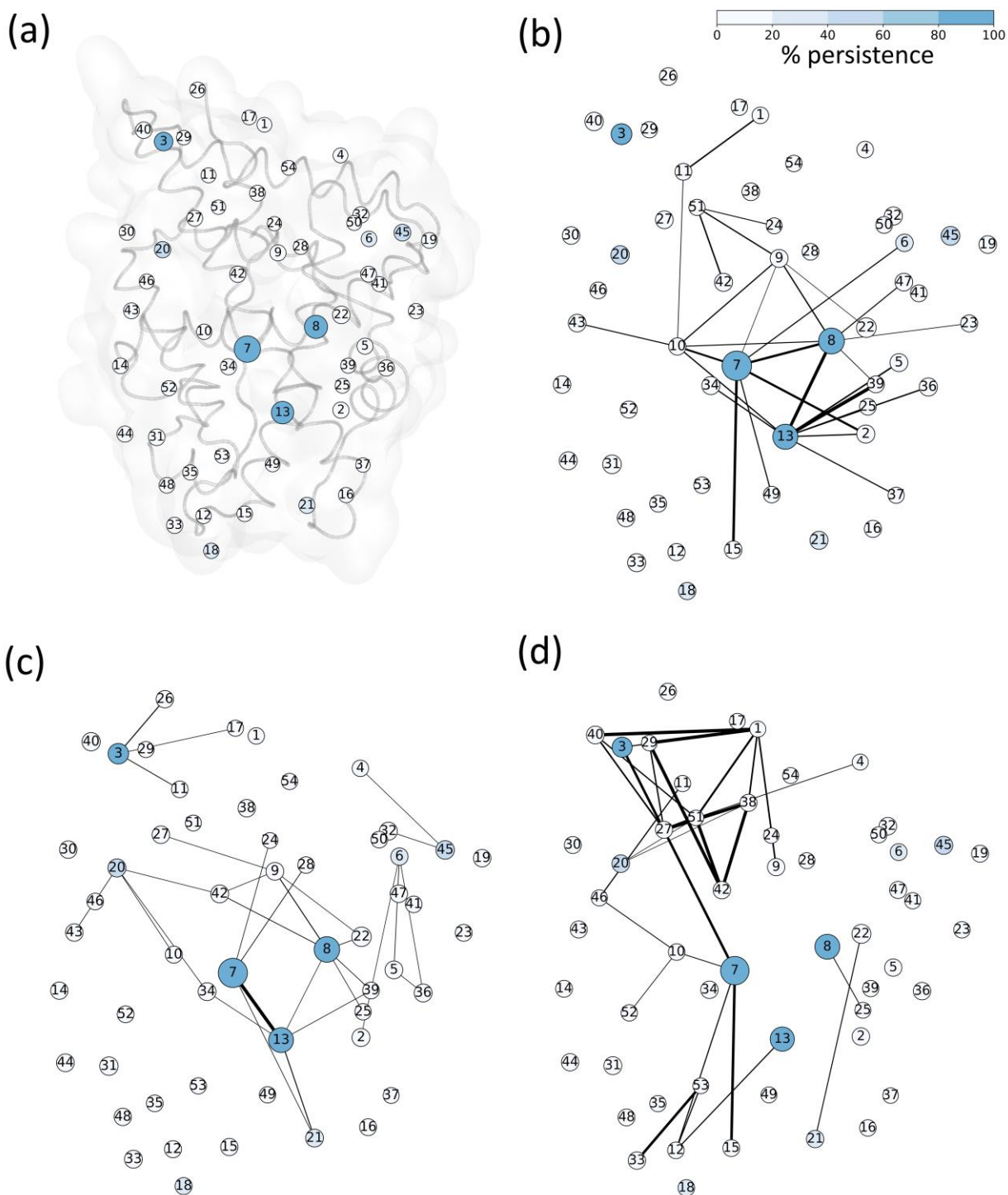


Figure 4. Network representation for the analyses on the AR system. The conventions used by panels (a)-(c) are consistent with Figure 2. Results from (b) DyNet, (c) Pocketron, and (d) DF analyses are shown.

The comparison of the results from the three approaches provided results that are more inhomogeneous (Figure 4, and Figure S13 for a 3D representation). The DyNet method (Figure 4(b)) indicated correlation of pID 13,

corresponding to the orthosteric pocket, with several surrounding pockets. The highest signals are found on the direct connection with pIDs 8 and 39. While there was no direct correlation with the high-volume and persistent pID 7, indirect connections through pIDs 2, 8, or 10 were identified. Furthermore, there was also a discernible path between pID 7 and pID1, passing through pIDs 10 and 11.

Conversely, the crosstalk network obtained with Pocketron (Figure 4(c)) displayed a more central role of the orthosteric site pID 13, which is connected to several surrounding pockets, including the larger and persistent pIDs 7 and 8, and other smaller and less persistent pockets. Regarding the separate, tough persistent, pID 3, there was only a local communication with surrounding pockets, such as pID 11, 17 and pID 26, while no connection with any of the biggest pockets was found. The difficulty in identifying communication with Pocketron may be due to the conformation of the system where the presence of helix 4 separated the two pIDs 13 and 7 from pID8, thus interrupting the possible interchange of pocket's residues.

Finally, the DF analysis (Figure 4(d)) suggested the presence of two key clusters with high correlation signals, one around pID 3 and 1, and one around the orthosteric pocket (pID 13), mainly connected by the edge between the relevant node 7 and node 27.

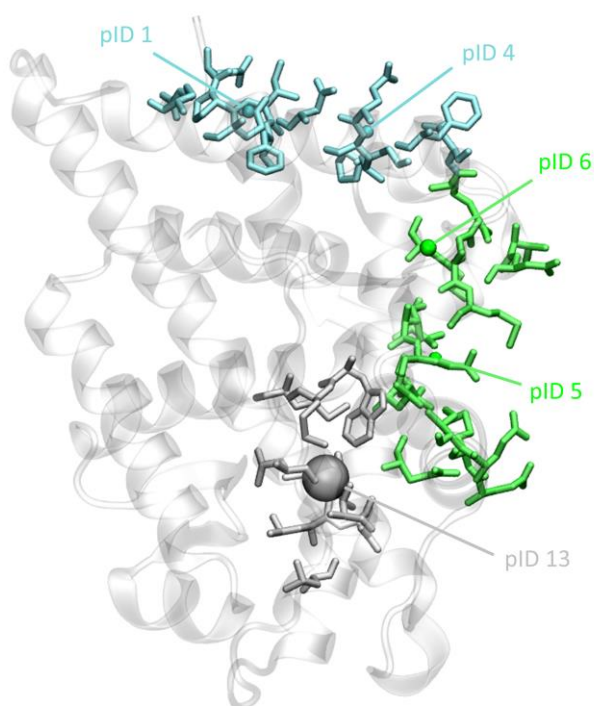


Figure 5. Comparison between relevant hotspots reported in the literature and pockets identified with Pocketron for the AR system. The different hotspots from the literature are shown in licorice with different colors, while the corresponding pockets identified with Pocketron are shown as spheres, with consistent colors. Color codes: orthosteric site in grey, BF3 site in cyan, and AF2 site in green.

The AR presents three main hotspots, consisting in the orthosteric binding pocket (Figure 5, grey residues), the BF-3 pocket (cyan residues), and the AF-2 pocket (green residues). The AF2 site is important for the binding of co-activator proteins, which triggers the transcription of genes regulated by AR. Some antagonists have been found to bind to this site, which prevents co-activator proteins from binding and activating AR.⁴⁵ The BF3 site is a recently discovered site where antagonists can bind to suppress AR activity. When an antagonist binds to any of these sites, it causes a change in the structure of the AF2 site, making it impossible for co-activator proteins to bind to AR due to the conformational changes induced by the movement of H12. Notably, besides the orthosteric

site, corresponding to pID 13, several pockets were identified from our simulations in these hotspot regions. These included pIDs 5 and 6, located in the AF-2 regions, which showed low persistence in our simulations. It is interesting to notice that, despite this known communication between the orthosteric site and the AF-2 region, no remarkable communication between the two region was observed in our analyses. Only the pocket crosstalk analysis with Pocketron (Figure 4(c)) displayed some communication between pID 13 and the small cluster comprising pockets 6, 5 and 36, passing through pID 39, however with relatively small edge intensities. Finally, concerning the BF3 pocket, shown in cyan and corresponding to pIDs 1 and 4 in our results, Pocketron displayed no relevant communication with other pockets, while DyNet showed a pathway starting from pID 1 to pID 5 through pID 11, 10, and the orthosteric pID 13. Finally, Pocketron correctly identified a non-negligible communication between pID 13 and pID 21, which is in the terminal part of H11 and undergoes a conformational change after binding of an agonist or antagonist, confirming the possibility of information exchange between the two regions.

EGFR Kinase Domain

The analyses on the EGFR system (Figure 6, and Figure S14 for a 3D representation) showed consistent results in identifying pID 1, which corresponded to the hinge region, as a central hub between two separate regions of the structure that were mainly involved in inter-pocket communication. The DyNet approach (Figure 6(b)) resulted in several connections between the hinge region (pID 1) and the closer pockets, such as pIDs 30, 44, 53, and 57, as well as the pockets formed by the A loop, such as pIDs 7, 8, 37 and 45. Moreover it showed a clear connection between pID 57 and pID 44. Comparing these results with those obtained using Pocketron (Figure 6(c)), we can see consistent connections, such as those connecting pID with pIDs 57, which is located between five beta-strands and the alpha C-helix, and 44, a deep groove located at the interface. The networks showed connections involving the pockets both with the hinge and with respect to each other. Additionally, an interesting signal is observed between the hinge (pID1) and pIDs 7 and 8, which are formed by the A-loop in this conformation, in both methods creating a communication network within that region. The DF analysis (Figure 6(d)) highlighted high correlations within two opposite regions of the system, providing overall information in agreement with what was observed from DyNet and Pocketron. Passing through the main persistent PID 1, the two distant regions of the EGFR, in both of which high communication is observed, had the possibility to communicate and transfer structural dynamics information.

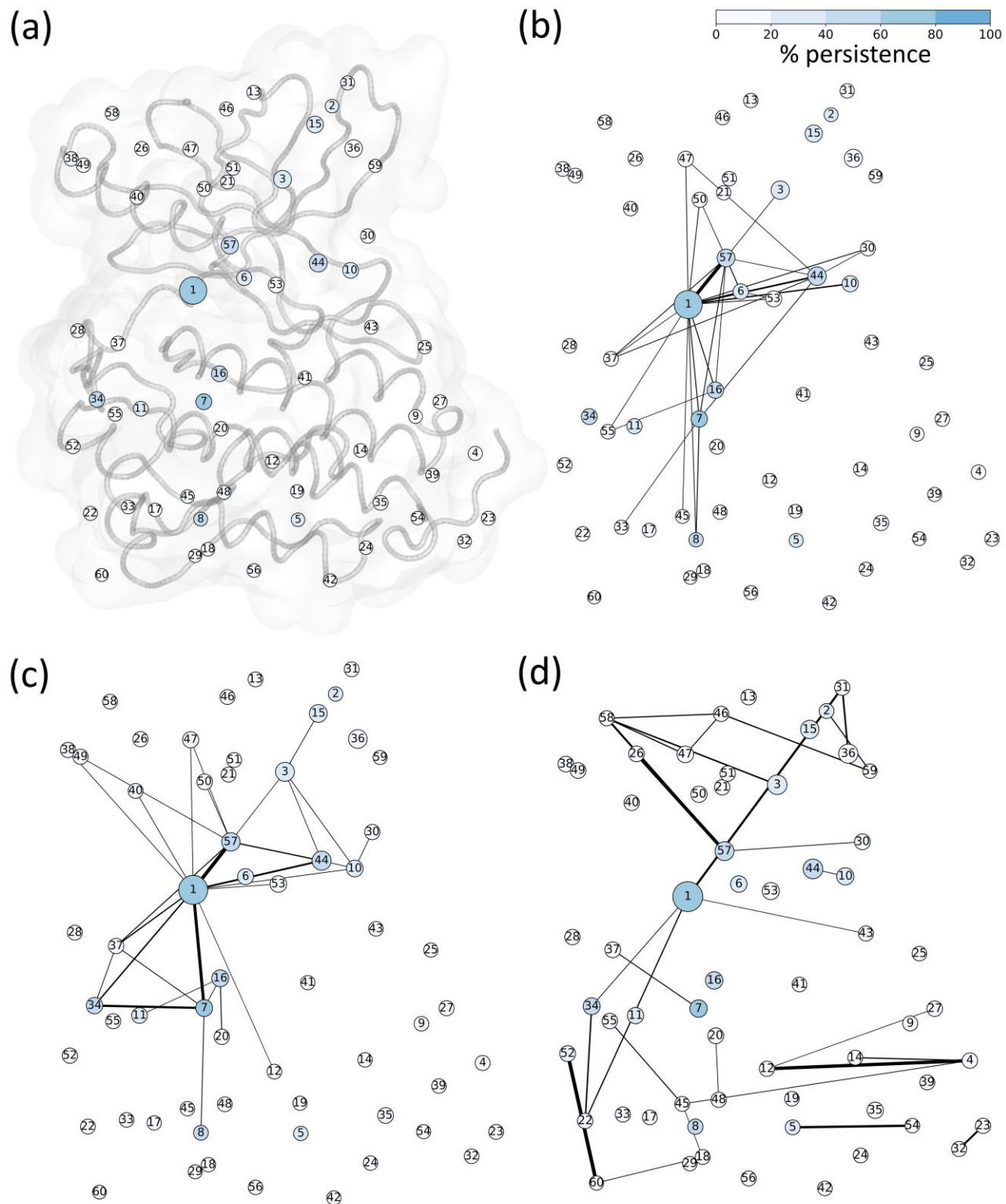


Figure 6. Network representation for the analyses on the EGFR system. The information provided by panels (a)-(c) is consistent with Figure 2. In brief, results from (b) DyNet, (c) Pocketron, and (d) DF analyses are shown.

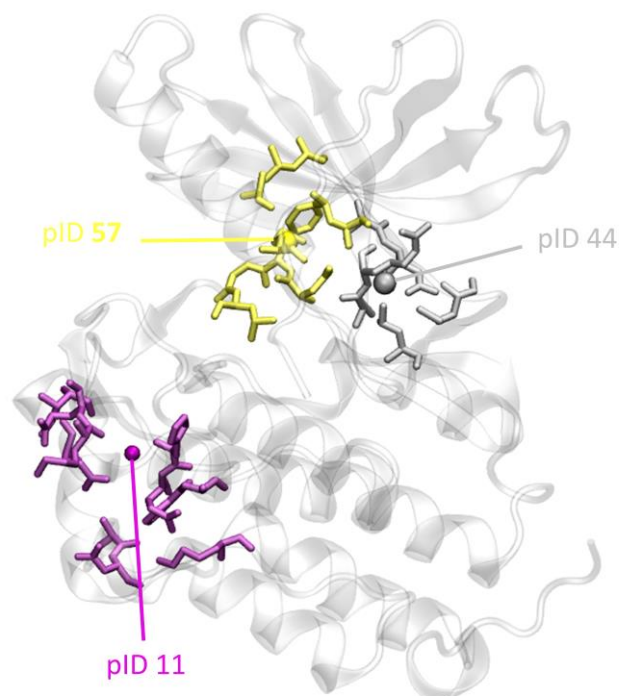


Figure 7. Comparison between relevant hotspots reported in the literature and pockets identified with Pocketron for the EGFR system. The different hotspots from the literature are shown in licorice with different colors, while the corresponding pockets identified with Pocketron are shown as spheres, with consistent colors. Color codes: orthosteric site in grey, MEK site in yellow, and pocket X in purple.

In the literature, it has been shown that there are a few main structures of the receptor mainly responsible for the activation and deactivation of the system. The structure presents a hinge allowing for flexibility and movement in this specific part of the structure. This region is typically composed of a stretch of amino acids that can bend or rotate without disrupting the overall structure. In close vicinity, it can be found a small helix called the α C-helix, 5 β -strands, a loop known as the A loop, and the orthosteric ATP pocket that includes the “DFG” motif made by the residues D855, F856, G857. These structures assume different conformations to activate and deactivate the receptor. A thorough study on the behavior of the EFGR kinase was conducted by Lu et al.,⁴⁶ which examined the changes in the apo structure caused by the presence of a ligand in the ATP binding pocket (Osimertinib) alone and in combination with a ligand in the allosteric MEK pocket (JBJ). The study revealed that the presence of Osimertinib resulted in the system assuming an α C-out conformation and an open A loop. On the other hand, the presence of the allosteric inhibitor JBJ led to an inactive state, closing the loop inside of the hinge that interacts with the out-rotated α C-helix and a 2 turn small helix formed in the loop (“Src-like” inactive state).

Figure 7 presents the location of critical regions in the EGFR structure, namely the ATP binding pocket (grey residues in the figure), the MEK pocket (red residues), and a potential allosteric site identified by Lu et al.⁴⁶ Our pocket detection revealed a strong correspondence for the first two known pockets, namely the ATP and MEK pockets, corresponding to pIDs 44 and 57, respectively. Additionally, the A-loop was identified as another important area of the structure, which corresponds to the region around pID 7, that mostly displayed in DyNet and Pocketron correlations with mainly pIDs 1 and 57 (directly or with small path). Therefore, our results confirmed the view reported in the literature, with the central hinge mediating the exchange of communication between the different regions of the structure.

IV. DISCUSSION AND CONCLUSIONS

Analyzing the communication patterns within a biomolecular structure is of utmost importance to understand its functioning and to potentially design allosteric binders able to modulate its activity. In this respect, computational methods are enabling tools to achieve an atom-level, holistic picture of the underlying mechanistic details. In this study, we combined molecular dynamics (MD) simulations in the μs timescale with some tools to identify correlations in the structural dynamics of three pharmaceutically relevant targets: the adenosine A_{2A} receptor (A_{2A}), the androgen receptor (AR), and the epidermal growth factor receptor (EGFR) kinase domain. The systems were all simulated in the absence of any co-factor/ligand. The protocol is based on two major steps: 1) the pockets detection/tracking and 2) the evaluation of the pocket crosstalk. We took advantage of Pocketron¹³ to address the first step. Then we performed the second step by employing three different techniques: post-processing²⁷ of Pocketron merge-split matrices, the Dynamic Network Analysis (DyNet), and distance fluctuation (DF) analysis methods. For all the methods we used the notion of pocket coming from the Pocketron analysis. The three correlation methods are based on different conceptual frameworks, therefore we aimed to assess to what extent they could provide similar, complementary, or different indications of pocket correlations.

Pocketron has shown that, in all the analyzed three cases, orthosteric pockets are placed in the top-3 largest pockets, such as pID 4 for the A_{2A} case, pID 13 for the AR case and pID 44 (ATP site) for the EGFR. Overall, the investigated approaches allowed, in most cases, to straightforwardly highlight the most relevant known sites; these typically emerged as persistent and highly connected pockets.

In pockets communication terms the degree of consistency of the tested methods proved to be system-dependent showing the partial orthogonality of the outcomes. This confirms the “many disguises” nature of allostery,⁶ albeit from the computational standpoint.

The analysis of the A_{2A} revealed that DyNet can detect strong communications surrounding the orthosteric site, where also Pocketron finds multiple correlations. Conversely, DF identifies fewer connections from the orthosteric site, and points to pockets in different regions compared to the other methods, particularly close to the intracellular portion. The work from the McCammon laboratory and collaborators⁴² identified five non-orthosteric pockets and all of them were identified by Pocketron (pID 3, 5, 6, 8, and 33).

In the Androgen Receptor case, similarly to the A_{2A} results, higher communication was associated with the orthosteric site (pID 13) in DynNet and Pocketron, with DF displaying one main connection and then a denser area of signals further away close to the BF3 site (pIDs 3 and 4). Analogously, the former methods identified communication with the AF2 allosteric site (pIDs 5 and 6), albeit with different intensities, while no relevant connections were observed through the DF approach.

For the EGFR kinase domain, DyNet and Pocketron’s results both highlighted several communications originating from the ATP site (pID 44). The MEK pocket (pID57), which is a known allosteric site,⁴⁷ showed a high degree of communication with the surroundings and especially the hinge region for all the three methods. Furthermore, Figure 6D (the DF analysis) revealed two main clusters of communications within the receptor, connected by the hinge, resulting in low distance fluctuations in regions not involved in major conformational changes. The study by Lu et al.⁴⁶ demonstrated a correlation between the MEK pocket (upon binding of the JBJ ligand) and the closing movement of the A-loop, which leads to the inactivation of the receptor. Their results identified a potential allosteric pocket in the lower region of the receptor called "Pocket X". Comparison of the residues that create this pocket revealed it corresponds to our pID 11, which is found, in the case of Pocketron, connected to the hinge and the ATP site through other intermediate pockets. For the first two methods, pID 7 has a higher persistence and a stronger, direct connection to the hinge (and thus the ATP site).

As we selected systems known to bear allosteric links, the results obtained from the application of the three different methods seem to confirm the presence of allosteric phenomena. We stress again that we focused only on

the top-ranking communication results to facilitate the interpretation and comparison of the complex interplays emerging from our networks. We cannot hence exclude additional interesting connections that may exist; such system specific analysis is out of the scope of the present study, which aims at providing an overall comparison of the three different approaches. The issue of identifying potential allosteric patterns is approached differently by the three methods. The DyNet approach, being based on the mutual information, represents a probabilistic viewpoint. Specifically, the correlation within biomolecular structures is quantified by the degree of statistical independence between distributions of observables (e.g. Cartesian coordinates, distances, etc.) suitable to describe structural motions. Conversely, the DF method quantifies the correlation in terms of the variance of the observables, namely distances; the variance is a parameter of a distribution not a distribution itself. Finally, in Pocketron, once the pockets in the structure are identified, the correlation stems from the dynamical sharing of residues observed between neighboring pockets during a MD simulation.

Interestingly, our results showed that Pocketron, despite not manipulating explicitly distributions, is able to achieve good consistency with the overall picture obtained from the mutual-information based DyNet approach.

Additionally, these two methods display partial consistency with the variance-based DF analysis. Notably, the DF approach prioritizes correlations corresponding to low fluctuations, i.e. low variance, of the distances; one can interpret this correlation measure as a synchronous motion between two pockets. Furthermore, differently than DyNet and Pocketron that assess correlation between adjacent pockets only, in DF the correlation is instead estimated for all pairs. As a result, the DF method provides a global description of the correlations within the structure. In this framework, the use of a distance threshold thus gains particular relevance in light of identifying genuinely long-range correlations.

Opposite to this global and synchronous spirit, the DyNet and Pocketron methods quantify local and possibly asynchronous correlations. Hence, long-range allosteric communication in these two methods is always a consequence of a path in a graph, whereas in DF direct links rule all the long-range interactions. Within this picture, our results showed, unexpectedly and interestingly, that Pocketron seems to be more consistent with the probabilistic approach, even if there is no explicit probabilistic formulation.

For all these reasons, while we envisage an exploitation of the DF method to identify regions that can be sensitive to long-range and synchronous modulation, DyNet and Pocketron appear more suitable to reconstruct signaling pathways happening through more general communication patterns.

The employed algorithms, as usual, present free hyper-parameters; hence studying the “persistence” of topological features⁴⁸ of graphs varying the hyper-parameters could be an interesting line of further investigation. Nevertheless, while the persistent homology solution is inspired by topological principles, one could instead envision a more physical approach: for the DF case in particular, as genuine allosteric interactions are inherently long range, here one could take advantage of the rate of decay of the electrostatic interactions to define a proper threshold. Further, the threshold definition can stem from secondary and tertiary structure knowledge of the system.

Although our approach for pocket identification and study of the correlations takes explicitly into account the dynamics of the system, it is nevertheless intrinsically limited by the conformational space accessible to the apo systems, the simulative timeframe, and the water-only buffer. To improve this last aspect for future works, one may envisage the doping of water through hydrophobic molecular probes.⁴⁹ This can improve the sampling of biomolecular conformations exposing so-called cryptic pockets; the sampled structures conformations could well feed the here introduced computational pipeline.

Supplementary material

The Supplementary Material includes correlation matrices obtained using the three methods for all the three systems, 3D views of the pocket networks, the effect of parameters used in the analyses, and further details about the MD simulations.

Acknowledgements

Giorgio Colombo and Stefano Serapian are gratefully acknowledged for sharing the code to conduct the distance fluctuation (DF) analysis and providing useful indications on its use. We gratefully acknowledge the referees for their useful suggestions.

Author Declarations

Conflict of interest

S.D. and A.C. are shares owners of BiKi Technologies s.r.l., a company that commercializes software solutions for medicinal chemistry.

Data availability

MD trajectories, preprocessed data and a Jupyter Notebook that can be used to perform our analyses and generate all the figures in the manuscript, along with input files for the MD simulations, are openly available and can be downloaded at: <https://doi.org/10.48557/TVYQBB>

Preprocessed data and the Jupyter Notebook only can also be found at: <https://github.com/mbernett/pocketcorrel> NanoShaper is freely available at <https://gitlab.iit.it/SDecherchi/nanoshaper/>. Pocketron is available in the commercial software “BiKi Life Sciences” (<https://www.bikitech.com>). For no profit institutions, the software can be obtained with a nominal fee.

References

- ¹ P.R. Arantes, A.C. Patel, and G. Palermo, “Emerging methods and applications to decrypt allostery in proteins and nucleic acids,” *J. Mol. Biol.*, 167518 (2022).
- ² L.G. Ahuja, S.S. Taylor, and A.P. Kornev, “Tuning the ‘violin’ of protein kinases: The role of dynamics-based allostery,” *IUBMB Life* **71**(6), 685–696 (2019).
- ³ J. Guo, and H.-X. Zhou, “Protein allostery and conformational dynamics,” *Chem. Rev.* **116**(11), 6503–6515 (2016).
- ⁴ A.P. Kornev, and S.S. Taylor, “Dynamics-driven allostery in protein kinases,” *Trends Biochem. Sci.* **40**(11), 628–647 (2015).
- ⁵ J. Liu, and R. Nussinov, “Allostery: an overview of its history, concepts, methods, and applications,” *PLoS Comput. Biol.* **12**(6), e1004966 (2016).
- ⁶ S.J. Wodak, E. Paci, N.V. Dokholyan, I.N. Berezovsky, A. Horovitz, J. Li, V.J. Hilser, I. Bahar, J. Karanicolas, and G. Stock, “Allostery in its many disguises: from theory to applications,” *Structure* **27**(4), 566–578 (2019).
- ⁷ K.W. East, E. Skeens, J.Y. Cui, H.B. Belato, B. Mitchell, R. Hsu, V.S. Batista, G. Palermo, and G.P. Lisi, “NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes,” *Biophys. Rev.* **12**(1), 155–174 (2020).
- ⁸ R.E. Amaro, “Toward understanding ‘the ways’ of allosteric drugs,” *ACS Cent. Sci.* **3**(9), 925–926 (2017).
- ⁹ R. Nussinov, and C.-J. Tsai, “Allostery in disease and in drug discovery,” *Cell* **153**(2), 293–305 (2013).
- ¹⁰ A. Triveri, C. Sanchez-Martin, L. Torielli, S.A. Serapian, F. Marchetti, G. D’Acerno, V. Pirota, M. Castelli, E. Moroni, M. Ferraro, P. Quadrelli, A. Rasola, and G. Colombo, “Protein Allostery and Ligand Design: Computational Design Meets Experiments to Discover Novel Chemical Probes,” *J. Mol. Biol.*, 167468 (2022).

- ¹¹ S. Gianni, and P. Jemth, “Allostery frustrates the experimentalist,” *J. Mol. Biol.*, 167934 (2022).
- ¹² S. Bowerman, and J. Wereszczynski, “Detecting allosteric networks using molecular dynamics simulation,” *Methods Enzymol.* **578**, 429–447 (2016).
- ¹³ G. La Sala, S. Decherchi, M. De Vivo, and W. Rocchia, “Allosteric communication networks in proteins revealed through pocket crosstalk analysis,” *ACS Cent. Sci.* **3**(9), 949–960 (2017).
- ¹⁴ I. Rivalta, M.M. Sultan, N.-S. Lee, G.A. Manley, J.P. Loria, and V.S. Batista, “Allosteric pathways in imidazole glycerol phosphate synthase,” *Proc. Natl. Acad. Sci.* **109**(22), E1428–E1436 (2012).
- ¹⁵ G. Morra, G. Verkhivker, and G. Colombo, “Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full-length dimer,” *PLoS Comput. Biol.* **5**(3), e1000323 (2009).
- ¹⁶ A. Cooper, and D. Dryden, “Allostery without conformational change,” *Eur. Biophys. J.* **11**(2), 103–109 (1984).
- ¹⁷ C. Chennubhotla, and I. Bahar, “Signal propagation in proteins and relation to equilibrium fluctuations,” *PLoS Comput. Biol.* **3**(9), e172 (2007).
- ¹⁸ A. Chatzigoulas, and Z. Cournia, “Rational design of allosteric modulators: Challenges and successes,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **11**(6), e1529 (2021).
- ¹⁹ K. Gunasekaran, B. Ma, and R. Nussinov, “Is allostery an intrinsic property of all dynamic proteins?,” *Proteins Struct. Funct. Bioinforma.* **57**(3), 433–443 (2004).
- ²⁰ Y. Zhang, P. Doruker, B. Kaynak, S. Zhang, J. Krieger, H. Li, and I. Bahar, “Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior,” *Curr. Opin. Struct. Biol.* **62**, 14–21 (2020).
- ²¹ A. Ohta, E. Gorelik, S.J. Prasad, F. Ronchese, D. Lukashev, M.K. Wong, X. Huang, S. Caldwell, K. Liu, and P. Smith, “A2A adenosine receptor protects tumors from antitumor T cells,” *Proc. Natl. Acad. Sci.* **103**(35), 13132–13137 (2006).
- ²² K. Sivak, A. Vasin, V. Egorov, V. Tsevtkov, N. Kuzmich, V. Savina, and O. Kiselev, “Adenosine A2A receptor as a drug target for treatment of sepsis,” *Mol. Biol.* **50**(2), 200–212 (2016).
- ²³ T. Matsumoto, M. Sakari, M. Okada, A. Yokoyama, S. Takahashi, A. Kouzmenko, and S. Kato, “The androgen receptor in health and disease,” *Annu Rev Physiol* **75**(1), 201–224 (2013).
- ²⁴ Y. Yuza, K.A. Glatt, J. Jiang, H. Greulich, Y. Minami, M.S. Woo, T. Shimamura, G.I. Shapiro, J.C. Lee, and H. Ji, “Allele-dependent variation in the relative cellular potency of distinct EGFR inhibitors,” *Cancer Biol. Ther.* **6**(5), 661–667 (2007).
- ²⁵ M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, “Role of molecular dynamics and related methods in drug discovery,” *J. Med. Chem.* **59**(9), 4035–4061 (2016).
- ²⁶ M.C. Melo, R.C. Bernardi, C. De La Fuente-Nunez, and Z. Luthey-Schulten, “Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories,” *J. Chem. Phys.* **153**(13), 134104 (2020).
- ²⁷ S. Majumdar, F. Di Palma, F. Spyraakis, S. Decherchi, and A. Cavalli, “Molecular dynamics and machine learning shed light on the flexibility-activity relationships in tyrosine kinase,” Submitted, (2023).
- ²⁸ K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, and D.E. Shaw, “Improved side-chain torsion potentials for the Amber ff99SB protein force field,” *Proteins Struct. Funct. Bioinforma.* **78**(8), 1950–1958 (2010).
- ²⁹ W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.* **79**(2), 926–935 (1983).
- ³⁰ S. Decherchi, G. Bottegoni, A. Spitaleri, W. Rocchia, and A. Cavalli, “BiKi Life Sciences: A New Suite for Molecular Dynamics and Related Methods in Drug Discovery,” *J. Chem. Inf. Model.* **58**(2), 219–224 (2018).
- ³¹ M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, and E. Lindahl, “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers,” *SoftwareX* **1**, 19–25 (2015).
- ³² S. Decherchi, and W. Rocchia, “A general and robust ray-casting-based algorithm for triangulating surfaces at the nanoscale,” *PloS One* **8**(4), e59744 (2013).
- ³³ T. Mühlethaler, D. Gioia, A.E. Prota, M.E. Sharpe, A. Cavalli, and M.O. Steinmetz, “Comprehensive analysis of binding sites in tubulin,” *Angew. Chem.* **133**(24), 13443–13454 (2021).

- ³⁴ O.F. Lange, and H. Grubmüller, “Generalized correlation for biomolecular dynamics,” *Proteins Struct. Funct. Bioinforma.* **62**(4), 1053–1061 (2006).
- ³⁵ M. Girvan, and M.E. Newman, “Community structure in social and biological networks,” *Proc. Natl. Acad. Sci.* **99**(12), 7821–7826 (2002).
- ³⁶ R.W. Floyd, “On ambiguity in phrase structure languages,” *Commun. ACM* **5**(10), 526 (1962).
- ³⁷ I. Rivalta, and V.S. Batista, in *Allostery* (Springer, 2021), pp. 137–151.
- ³⁸ C. Sanchez-Martin, E. Moroni, M. Ferraro, C. Laquatra, G. Cannino, I. Masgras, A. Negro, P. Quadrelli, A. Rasola, and G. Colombo, “Rational design of allosteric and selective inhibitors of the molecular chaperone TRAP1,” *Cell Rep.* **31**(3), 107531 (2020).
- ³⁹ A. Hagberg, P. Swart, and D. S Chult, *Exploring Network Structure, Dynamics, and Function Using NetworkX* (Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008).
- ⁴⁰ G.H. Golub, and C.F. Van Loan, “Matrix computations, 1983,” Johns Hopkins Baltim. MD, (1985).
- ⁴¹ D. Kozakov, L.E. Grove, D.R. Hall, T. Bohnuud, S.E. Mottarella, L. Luo, B. Xia, D. Beglov, and S. Vajda, “The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins,” *Nat. Protoc.* **10**(5), 733–755 (2015).
- ⁴² A.D. Caliman, Y. Miao, and J.A. McCammon, “Mapping the allosteric sites of the A2A adenosine receptor,” *Chem. Biol. Drug Des.* **91**(1), 5–16 (2018).
- ⁴³ W. Liu, E. Chun, A.A. Thompson, P. Chubukov, F. Xu, V. Katritch, G.W. Han, C.B. Roth, L.H. Heitman, and A.P. IJzerman, “Structural basis for allosteric regulation of GPCRs by sodium ions,” *Science* **337**(6091), 232–236 (2012).
- ⁴⁴ H. Gutiérrez-de-Terán, A. Massink, D. Rodríguez, W. Liu, G.W. Han, J.S. Joseph, I. Katritch, L.H. Heitman, L. Xia, and A.P. IJzerman, “The role of a sodium ion binding site in the allosteric modulation of the A2A adenosine G protein-coupled receptor,” *Structure* **21**(12), 2175–2185 (2013).
- ⁴⁵ P. Axerio-Cilies, N.A. Lack, M.R.S. Nayana, K.H. Chan, A. Yeung, E. Leblanc, E.S.T. Guns, P.S. Rennie, and A. Cherkasov, “Inhibitors of androgen receptor activation function-2 (AF2) site identified through virtual screening,” *J. Med. Chem.* **54**(18), 6197–6205 (2011).
- ⁴⁶ Y. Qiu, X. Yin, X. Li, Y. Wang, Q. Fu, R. Huang, and S. Lu, “Untangling dual-targeting therapeutic mechanism of epidermal growth factor receptor (EGFR) based on reversed allosteric communication,” *Pharmaceutics* **13**(5), 747 (2021).
- ⁴⁷ Z. Zhao, L. Xie, and P.E. Bourne, “Insights into the binding mode of MEK type-III inhibitors. A step towards discovering and designing allosteric kinase inhibitors across the human kinome,” *PLoS One* **12**(6), e0179936 (2017).
- ⁴⁸ N. Otter, M.A. Porter, U. Tillmann, P. Grindrod, and H.A. Harrington, “A roadmap for the computation of persistent homology,” *EPJ Data Sci.* **6**, 1–38 (2017).
- ⁴⁹ J.P. Arcon, L.A. Defelipe, C.P. Modenutti, E.D. López, D. Alvarez-Garcia, X. Barril, A.G. Turjanski, and M.A. Martí, “Molecular dynamics in mixed solvents reveals protein–ligand interactions, improves docking, and allows accurate binding free energy predictions,” *J. Chem. Inf. Model.* **57**(4), 846–863 (2017).