



# Computational Analysis of Crystallization Additives for the Identification of New Allosteric Sites

Jade Fogha, Julien Diharce, Alan Obled, Samia Aci-Seche, Pascal Bonnet

## ► To cite this version:

Jade Fogha, Julien Diharce, Alan Obled, Samia Aci-Seche, Pascal Bonnet. Computational Analysis of Crystallization Additives for the Identification of New Allosteric Sites. ACS Omega, 2020, 5 (5), pp.2114-2122. 10.1021/acsomega.9b02697 . hal-02904687

**HAL Id: hal-02904687**

**<https://hal.science/hal-02904687>**

Submitted on 12 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Computational Analysis of Crystallization Additives for the Identification of New Allosteric Sites

Jade Fogha<sup>□‡</sup>, Julien Diharce<sup>□‡†</sup>, Alan Obled<sup>□±</sup>, Samia Aci-Sèche<sup>□</sup>, Pascal Bonnet<sup>□\*</sup>

<sup>□</sup> Institut de Chimie Organique et Analytique, Université d'Orléans, UMR CNRS 7311, BP6759, 45067 Orléans, Cedex 2, France.

Jade Fogha: <https://orcid.org/0000-0001-5512-544X>

Julien Diharce: <https://orcid.org/0000-0002-6400-927X>

Samia Aci-Sèche: <http://orcid.org/0000-0003-2627-7924>

Pascal Bonnet: <http://orcid.org/0000-0001-6485-138X>

\*Corresponding author: Phone: +33 (0)238 41 70 42. E-mail: [pascal.bonnet@univ-orleans.fr](mailto:pascal.bonnet@univ-orleans.fr)

†Current address: Biologie intégrée du globule rouge, UMR\_S 1134 Inserm - Université Paris 7, Paris Diderot, Institut National de la transfusion Sanguine, 6 rue Alexandre Cabanel 75015 Paris Cedex 15.

± Current address: School of chemistry, University of St Andrews, North Haugh, St Andrews, Fife, KY16 9ST, United Kingdom.

‡ Those authors have contributed equally to the work

---

Authors declare no competing financial interests

---

**KEYWORDS.** *Allostery, Crystallization additives, Protein kinase, Nuclear receptor.*

---

**ABSTRACT:** Allosteric effect can modulate the biological activity of a protein. Thus, the discovery of new allosteric sites is very attractive for designing new modulators or inhibitors. Here, we propose an innovative way to identify allosteric sites, based on crystallization additives (CA), used to stabilize proteins during the crystallization process. Density and clustering analysis of CA, applied on protein kinase and nuclear receptor families, revealed that CA are not randomly distributed around protein structures, but they tend to aggregate near common sites. All orthosteric and allosteric cavities described in the literature are retrieved from the analysis of CA distribution. In addition, new sites were identified, which could be associated to putative allosteric sites. We proposed an efficient and easy way to use structural information of CA to identify allosteric sites. This method could assist medicinal chemists for the design of new allosteric compounds targeting cavities of new drug targets.

---

Proteins are fundamental entities in organism, controlling normal living cells and disorder processes. These biological entities are divided into families according to their amino acid sequences, three-dimensional (3D) structural motifs and primary function. Proteins perform their biological function through interactions with other proteins, nucleic acids or small ligands. The interactions between a protein and a ligand are often located in a well-defined active site or cavity of the protein, the so-called orthosteric site. These interactions are directly associated to protein function modulation. However, small ligands can also bind to other protein sites, called allosteric sites, distant from the active site. This binding could induce a conformational change on protein structure, resulting in an increase or decrease of its intrinsic activity.<sup>1,2</sup>

The term ‘allostery’ was introduced in 1961<sup>3,4</sup> even if Christian Bohr has already described the process as the ‘Bohr effect’ in the early 20<sup>th</sup> century related to hemoglobin conformational change. Since then, allostery has progressively evolved

to a unified concept<sup>5</sup> associated with its main property: the conformational change. It is an integrant part of the protein dynamics and may be present in every protein in the living world.<sup>6–8</sup> Not surprisingly, allostery has raised a great interest in pharmaceutical research especially in identifying allosteric sites in protein and/or developing allosteric drugs. The latter could present some advantages compared to drugs targeting orthosteric sites, such as a greater specificity, fewer side effects and an easier up- and down-regulation of proteins.<sup>9</sup> In some cases, this interest for allosteric approaches in drug discovery have led to successful results. Indeed, in 2004, Cinacalcet was the first allosteric drug approved by the Food and Drug Administration (FDA). This positive allosteric modulator targets the Calcium-Sensing Receptor belonging to GPCR family for the treatment of hyperparathyroidism.<sup>10</sup> Interest in allostery is well described in the protein kinase (PK) family, a major therapeutic target due to its implication in several diseases such as cancer.<sup>11,12</sup> Most of protein kinase

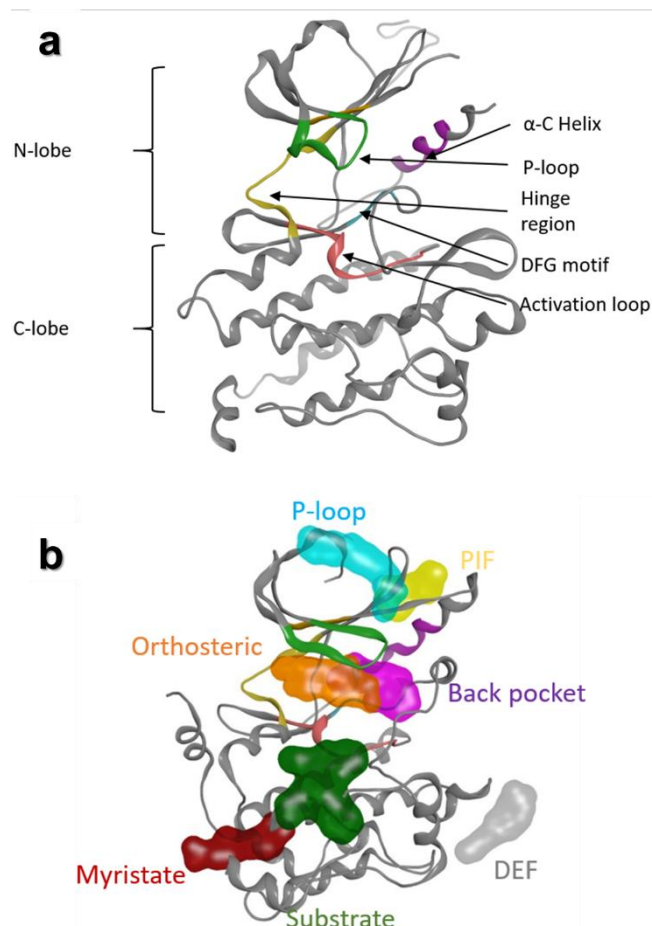
inhibitors approved by the FDA or under clinical trials are targeting the orthosteric adenosine triphosphate (ATP) binding site.<sup>13</sup> However, several allosteric sites have been identified in PKs such as ABL, CK2 $\alpha$ , FLT3 or MEK.<sup>14–19</sup> Several allosteric kinase inhibitors have been already approved by the FDA such as trametinib, cobimetinib and bimetinib in 2013, 2015 and 2018 respectively, for the treatment of patients with metastatic melanoma involving a BRAF V600E or V600K mutation.<sup>12,17</sup>

Nowadays, some databases<sup>20</sup> and benchmarks<sup>21</sup> are available for helping in the identification of allosteric cavities through computational approaches. Those approaches developed or adapted specifically for this objective are normal mode analysis, Gaussian network mode<sup>22,23</sup> and binding leverage approach<sup>24</sup> and are based on the calculation of protein cavity volumes. A computational mapping protocol, the Multiply Copy Simultaneous Search (MCSS) was also published in 1996. In this methodology, thousands of ligands are minimized around a protein structure, in order to identify the main binding sites.<sup>25</sup> In this paperwork, we proposed a novel computational approach to identify allosteric cavities in a protein family based on the presence of experimental crystallization additives (CA). Initially, those molecules are present together with ions, buffers and solvent in order to facilitate the crystallization process of proteins or protein-ligand complexes.<sup>26</sup> Interestingly, those molecules are not always randomly distributed around the structures, but seem located in protein hotspots, especially near the binding cavities.<sup>27</sup> Even if they cannot be directly used in FBDD projects, their binding reveals some key points on the interaction of drug-like ligand or fragments.<sup>28</sup> Some previous experimental mapping studies (Multiple Solvent Crystal Structures, MSCS) on crystalline proteins have demonstrated the ability of those additives to bind into interesting regions of protein surface<sup>25,29,30</sup> Thus, we decided to evaluate the relationship between the sites where CA are located and the known orthosteric and allosteric sites in a protein family. Here, we focused on two protein families that have been substantially crystallized: the PK and the nuclear receptor (NR) families. In a first step, starting from a dataset built from several databases, CA distribution is evaluated within 3D structure of PKs and NRs, aligned on a unique reference protein, to determine their location sites. Then, we assessed the ability of those CA to be collocated with known allosteric ligands (AL) already identified in PK and NR families through a clustering approach. Interestingly, we identified that the strong presence of CA in cavities of experimentally-determined protein structures corresponds to known orthosteric and allosteric sites. Therefore, this study suggests a novel approach to identify allosteric sites.

**PK family: a model for allostery.** PKs constitute one of the most studied protein families and their large involvements in several diseases (cancer, inflammation, Alzheimer's disease, etc.) lead to the search of novel therapeutic drugs, targeting the catalytic binding sites or allosteric cavities. In this study, we focused on PKs, since this protein family exhibits well-defined orthosteric and allosteric cavities. Indeed, PKs belong to the transferase superfamily, catalyzing the phosphate transfer to a protein substrate. Through the catalytic mechanisms and regulation of PKs, the ATP molecule located in the orthosteric site, bound to the hinge region (Figure 1, orange), will transfer the  $\gamma$  phosphate group to a protein substrate. PK inhibitors designed to target this orthosteric site and to compete with ATP are classified as type I and bind the active

conformation of the protein kinase. In addition to this major active site, several other cavities have been described during the past years, distributed all around the kinase structure.<sup>16,17</sup> In this study, 18 PDB structures representing various groups and subgroups of PK family will be considered as reference for the definition of those allosteric sites:

- The so-called back pocket (Figure 1, purple) is close to the orthosteric site. This back pocket concerned type II inhibitors, which bind to orthosteric and back pocket, in the inactive conformation of the kinase and type III inhibitors, which bind exclusively to the back pocket.<sup>31</sup> Here, back pocket is represented by PDB IDs 3O96 (AKT1), 3LW0 (IGF1R), 1S9J (MEK), 4LMN (MEK), 3EQC (MEK), 4ITH (RIPK1) and 4ZJI (PAK1).
- The myristate pocket (Figure 1, red) is located at the C-terminal lobe of ABL protein kinase and is targeted by type IV inhibitors. This pocket is associated to a sub-family of PKs, ABL (PDB IDs 3MS9, 3K5V and 3PYY).
- One additional pocket of interest is the substrate pocket (Figure 1, green), in which the protein that will be phosphorylated by the PK usually binds. Type V inhibitors target this site and PDB IDs 3JVR (CHK1) and 3F9N (CHK1) contain an allosteric inhibitor bound in this pocket.
- The DEF (Docking site for ERK, FFXF) pocket facilitates substrate recognition<sup>32</sup> as present in the PDB IDs 4E6C (MAPK14) and 3O2M (JNK1).
- The last cavities are located at the N-terminal part of PKs: the PIF (PDK1 interacting fragment) pocket<sup>33</sup> illustrated in PDB IDs 3PXF (CDK2) and 3HRF (PDK1) and the P-loop pocket, exemplified in PDB IDs 3H30 (CK2a1) and 4CFE (AMPK $\alpha$ 1/2).

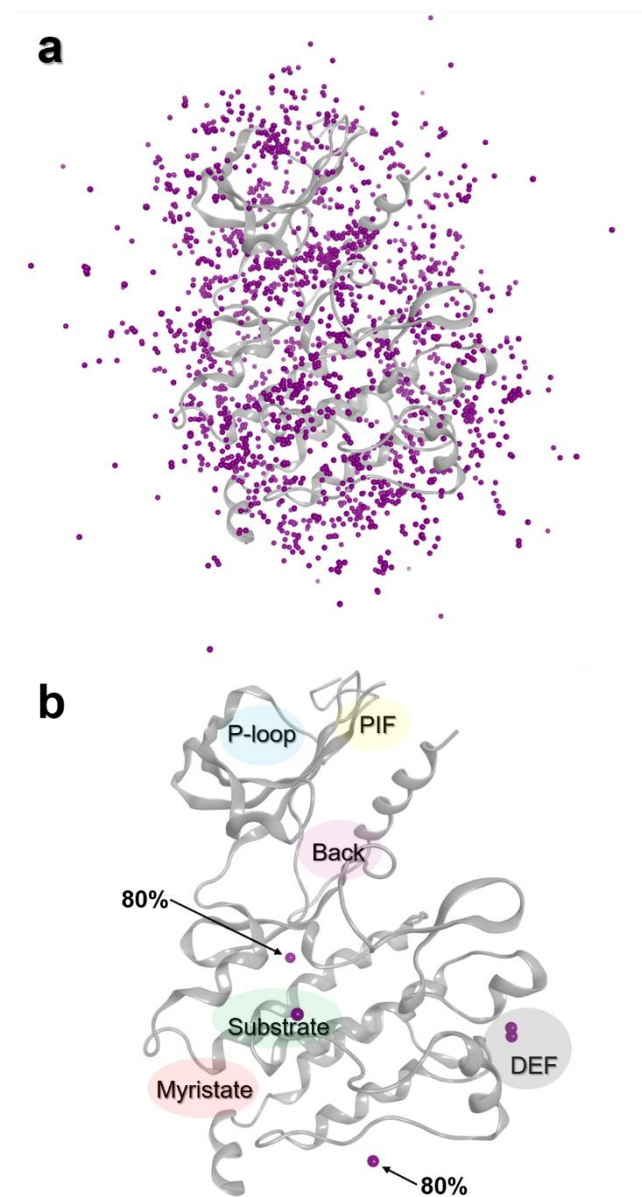


**Figure 1.** Structural properties of protein kinase family. (a) Structural common motifs in protein kinases, characterizing the active site. (b) Representation of cavities described in the literature (Orange: orthosteric cavity, purple: back pocket, cyan: P-loop pocket, yellow: PIF pocket, grey: DEF pocket, red: myristate pocket and green: substrate pocket).

**CA in kinase family.** Generally, 3D structures of complexes may also contain small molecules such as conventional ligands and/or CA. The latter are involved in the crystallography process and their coordinates are obtained from the electron density mapping like for protein and ligand atoms. In most *in silico* studies such as docking, 3D structural information of CA are ignored during calculations. In this study, we focused on the spatial positions of CA in several crystallographic structures of PKs to identify allosteric sites. 845 diverse structures of protein-ligand complexes were retrieved from the PDB and a total of 2459 CA molecules (CA dataset) are present in these complexes. Among these CA, we mainly noticed polar compounds like ethylene glycol or glycerol but also sugars ( $\beta$ -octylglucoside) which present amphipathic properties (an exhaustive list is provided in Table S1). For each CA, we only considered their geometric center (centroid) in order to simplify the structural analysis. Distribution of CA centroids was evaluated around and inside PK structures (Figure 2). An analysis of the aligned protein structures shows that CA could be placed in three different positions: (1) far away from the protein surface, (2) at the solvent exposed protein surface or (3) in cavities inside the protein structure (Figure 2,

panel a). We analyzed the density of CA distribution around the several aligned structures. The distribution of the CA centroids revealed that only two main cavities having more than 80% density of CA centroids are both surface exposed (details in the methodology section, Supp. Inf.). The first one is in the C-terminal part of the PK and did not correspond to an allosteric pocket already identified in the literature. The second one points toward the  $\alpha$ -E helix (Figure 2, panel b) and is located near the “peptide” pocket. In 2004, Heo *et al.* published the crystal structure of JNK1 (PDB ID 1UKI), a PK that binds to a peptide of the JIP1 protein and this interaction plays a role in the JNK1 phosphorylation activity. The interface formed by this protein-protein interaction is also an allosteric site.<sup>16,34</sup> By lowering the density to 50%, CA are aggregated in three other sites. Two centroids are very close together and are located in the DEF pocket. The last centroid is found near the substrate pocket.

Thus, in the case of PKs, it appears that CA are not always distributed arbitrarily but seem to be attracted to some conserved areas on the protein structure, such as the surface or deep cavities. This observation is in agreement with a very recent paper, proposing that fragments and CA often bind in the same way as drug-like ligands in four proteins (BACE2, CLK2, TYR1 and CAH2).<sup>27</sup> Moreover, in some cases these attractive regions in PKs could be related to allosteric sites such as pocket, substrate and DEF pockets.



**Figure 2.** Distribution of CA centroids around a PK (PDB ID 1K5V). (a) Representation of all CA centroids. (b) Position of centroids with density greater than 80% (centroids indicated by an arrow) and 50% (all purple spheres).

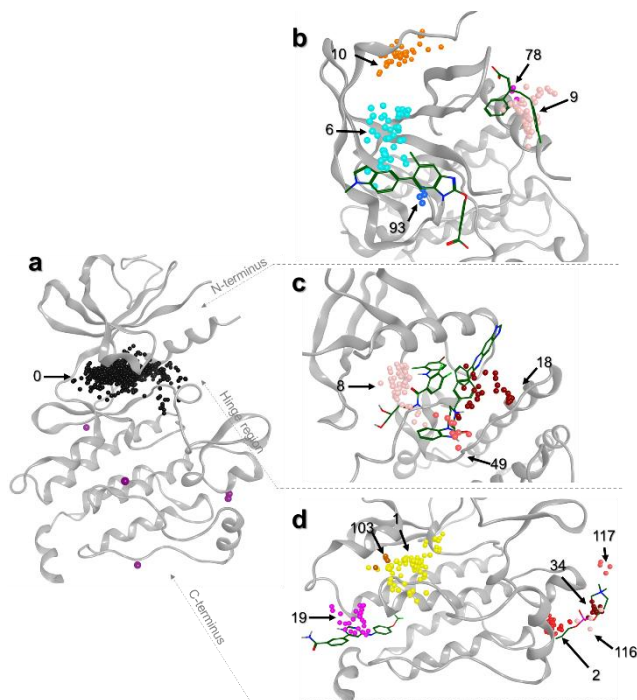
**CA: a kinase allosteric identifier.** In proteins, orthosteric ligands (OL) and AL bind to well-defined orthosteric and allosteric sites respectively. Based on the density of centroids, our preliminary results suggest an existence of preferential CA sites. A second analysis was performed to determine whether there is a possible relation between the position of CA molecules on the PK surface and the various sites occupied by OL and AL. A ligand dataset was built from 864 diverse structures of PK-ligand complexes containing exclusively OL and AL (1049 ligands). It is important to note that we have always considered complexes that have at least one AL, and in some cases an OL was also present in the crystal structures such as in PDB ID 4AN2 where ATP and cobimetinib are bound to MEK1 crystal structure. This dataset was joined to the CA

dataset, guided by an alignment of all PKs to the same reference structure PDB ID 1ATP. By focusing on CA, AL and OL centroids, two clustering analysis were performed using a density-based algorithm (DBSCAN)<sup>35</sup>, first on the CA dataset, and then on the combined CA and ligand datasets (Figure S1). Parameters of the clustering was controlled by changing gradually the minimal number of points in the cluster (minpts from 1 to 4) and the distance between two points in the cluster ( $\epsilon$  from 0.5 to 3 Å). According to the results, we observed three different types of clusters: i) some clusters contain only centroids of CA from the CA dataset or ii) some clusters contain only centroids of ligands (AL and OL) from the ligand dataset (those two types of cluster are called homogenous clusters) and iii) others clusters contain centroids from both datasets (called heterogeneous clusters). Heterogeneous clusters indicate that CA probably occupy the same site than ligands (AL and OL). Heterogeneous cluster IDs are depending on the clustering parameters and three parameter pairs [minpts -  $\epsilon$ ] have the most populated clusters ([1 - 2], [2 - 2] and [3 - 2]) (Figure S3). The 18 reference PK complexes containing AL were used in order to optimize the parameters. Three parameter pairs ([1 - 2], [2 - 2], [3 - 2]) give at least 8 heterogeneous clusters containing both reference ligands and CA. The parameter pair [3 - 2] provides the largest common cluster since there is a large number of CA and ligands in the same sites.

To compare the results obtained from the density analysis and the clustering method, an unsupervised clustering was carried out first on CA dataset with the parameters [3 - 2]. Under these conditions, about 33% of the CA cannot be clustered. The remaining CA are grouped into 118 clusters, classified by their population, and the most populated cluster is located in the substrate pocket, a site that has been already identified from density analysis with a density threshold of 50%.

Adding the ligand dataset (AL and OL) to the CA dataset in the clustering method provides also 118 different clusters constituted by homogeneous and heterogeneous clusters. 814 compounds (i.e. 23%), mainly AL, are identified as singletons. We observed that all the known pockets in PK are found in the 20 densest clusters, amongst 117 clusters in total (See Figure S4). Not surprisingly, OL are in the most populated cluster (~31% of all compounds) and are grouped with a small number of CA in this heterogeneous cluster (cluster 0, Figure 3, panel a; Figure S4). Thus, the ATP binding site is also a cavity that can accommodate CA, which is consistent with the study of Drwal et al.<sup>27</sup> The next most populated (Figure S4) heterogeneous clusters present interesting results. Indeed, we found that almost all the heterogeneous clusters define an allosteric site described in the literature and experimentally identified in crystal structures. Again, substrate pocket appears as an important attractive region for CA since this pocket contains centroids of cluster 1, the second most populated cluster (Figure 3, panel d; Table S2). Between the two reference ligands present in this pocket, only the centroid of allosteric CHEK1 inhibitor (PDB ID 3F9N) is present in cluster 1. The other allosteric CHEK1 inhibitor (PDB ID 3JVR) partially occupies this pocket and its centroid was not detected in the clusters. We also observed that cluster 103, close to the cluster 1, is located in the same substrate pocket. Thus, many clusters can point toward a same allosteric site.





**Figure 3.** Representation of heterogeneous clusters on a kinase structure (PDB ID 1K5V). (a) Orthosteric site represented by cluster 0 in black. Density points are still present in purple. (b) Heterogeneous clusters displayed on N-terminal region of the protein kinase. Reference ligands bound to AMPK (left, PDB ID 4CFE) and PDK1 (right, PDB ID 3HRF) are represented in stick. (c) Heterogeneous clusters displayed on the hinge region. Reference ligands bound to MEK1 (left, PDB ID 3EQC) and AKT1 (right, PDB ID 3O96) are represented in stick. (d) Heterogeneous clusters displayed on C-terminal part. Reference ligands bound to ABL1 (left, PDB ID 3K5V) and MAPK14 (right, PDB ID 4E6C) are represented in stick. Cluster IDs are indicated and centroids are color-coded based on their cluster ID.

Another example concerns the large DEF pocket, which presents four different cluster IDs (2, 34, 116, 117). In this case, the reference ligands are correctly retrieved in cluster 34 and 117 (Table S2). It is important to note that these reference ligands are bound to kinases classified in two different PK subgroups: MAPK14 (p38 $\alpha$ ) and JNK. There are major structural differences in the large DEF pocket between the two crystal structures (PDB ID 4E6C and PDB ID 3O2M), which explain the different binding mode observed for the two allosteric inhibitors (reference ligands). These substrate and DEF pockets were also identified above from the density analysis. Moreover, the clustering analysis allows the detection of additional allosteric pockets. As shown in Figure 3 panel d, cluster 19 indicates the position of the myristate pocket since this cluster is located at the same position than the reference ligand (myristate), crystallized in the ABL subgroup (PDB ID 3K5V). However, other AL of ABL selected as reference ligands (PDB IDs 3MS9 and 3PYY) were not classified and considered as singletons. This can be explained by the fact that the three ligands are not fully superimposed. Ligand co-crystallized in PDB ID 3K5V has a centroid located at 3.7 and 5.3 Å from ligand centroids of PDB IDs 3MS9 and 3PYY respectively.

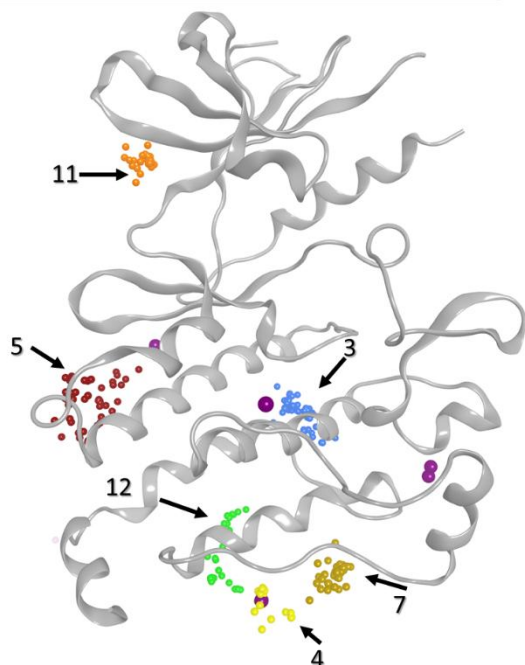
On the N-terminal lobe of the protein kinase (Figure 3, panel b), clusters 6 and 93 are correctly positioned on the P-loop pocket and clusters 78 and 9 on the PIF pocket. In fact, refer-

ence ligands are included in the different clusters corresponding to known allosteric sites (Table S2). In addition to the PIF and P-loop pockets, we also identified cluster 10, which did not correspond to any reference ligand. This cluster highlighted a new site, which was described very recently in the literature, and therefore not yet included in our database.<sup>36</sup> This pocket is occupied by ligands involved in an allosteric mechanism of Aurora kinase inhibition, and could be a new site of interest for other PKs.

In the case of the back pocket, the cavity volume is large due to structural variations present in different PKs and due to different chemical structures of type II inhibitors. In our study, we detected three different clusters in the back pocket (Figure 3, panel c). Among our reference ligands, the AL of AKT1 (PDB ID 3O96) is detected by the clustering method (cluster 18). This ligand binds in a rather different position than AL of MEK, RIPK or PAK, other ligands binding to the back pocket. We also identified cluster 49 in close proximity to cluster 18. AL of AKT occupies a space defined by the two clusters. Hence, those clusters seem to correctly define the allosteric pocket of AKT subgroup. The last cluster 8 is also involved in the back pocket and contains AL of MAPK group according to the presence of a ligand recently discovered as allosteric inhibitor of the ERK5 and MAPK7 proteins.<sup>37</sup> Unfortunately, the MEK reference ligands (MAPK subgroup - PDB IDs 1S9J, 4LMN, 3EQC) were not found in this cluster probably due to a bias induced by the consideration of centroid instead of the whole ligands. The back pocket is in the close vicinity of the orthosteric site. During the clustering step, the centroids of the OL are not correctly distinguished from centroids of the AL located in the back pocket. For this reason, AL of MEK (PDB IDs 1S9J, 4LMN, 3EQC), RIPK1 (4ITH) and PAK (4ZJI), used as references, were classified in the orthosteric cluster (cluster 0). To avoid this problem, OL were removed using a pharmacophore search defined near the hinge region. Clustering of datasets without OL provides more meaningful results since AL present in the back pocket were mainly grouped in cluster 0 and AL of AKT (PDB ID 3O96) in cluster 19 (Table S2). The other allosteric sites were not modified and for DEF pocket for example, reference AL were correctly identified in clusters 37 and 120.

Homogeneous clusters, formed by CA molecules, represent the largest number of centroids in a cluster (Table S2, Figure S4). They concerned CA that cannot be clustered with ligands (AL nor OL). According to the analysis of heterogeneous clusters, the different allosteric sites were found in clusters containing more than 25 centroids. Beyond this threshold, all known allosteric pockets were identified by the clustering method, meaning that every allosteric sites are retrieved in the 20<sup>th</sup> first denser clusters (Figure S4). For homogeneous cluster analysis, those clusters (number  $\geq 25$  centroids) were also identified in allosteric sites in PKs (Figure 4). The homogeneous cluster 5, located near the  $\alpha$ -E helix, represents the peptide pocket, identified above from density analysis. Cluster 11 is similar to the heterogeneous cluster 10, which defined a cavity already identified in the literature as an allosteric site.<sup>36</sup> Clusters 3, 4, 7 and 12 revealed three other sites at the C-terminal part of PKs. The first one, cluster 3, is located at the end of the  $\alpha$ -E helix and the second, cluster 12, is located at the C-terminal lobe of PKs. The third site, formed by clusters 4 and 7, is also at the bottom of PKs and interestingly, this site was identified as an important attractive cavity for CA in PKs using the density analysis. Those three sites are described for

the first time by our approach and were not described in our reference structures nor in the literature. They could potentially be considered as novel binding sites of interest

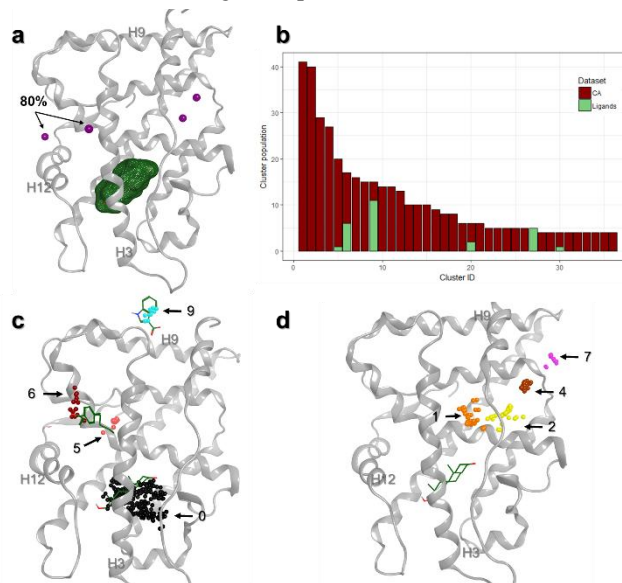


**Figure 4.** Representation of homogeneous clusters on a protein kinase (PDB ID 1K5V). Cluster IDs are indicated by an arrow.

Finally, we evaluated the overlap of CA and OL on one side and the overlap of CA and AL on the other side. In fact, the CA distribution, via density analysis, revealed that CA are preferentially grouped in close proximity to DEF and substrate pockets. The clustering analysis goes further and suggests that CA can be located in the orthosteric site and in the six known allosteric sites, defined by the reference structures. Thus, considering CA positions can help to determine orthosteric cavity but more importantly allosteric sites in PKs. Moreover, both density and clustering analysis highlighted some new sites, for which allosteric properties were not yet studied but could be targeted using *de novo* approaches for the design of novel allosteric ligands.

**Application of the CA approach on NR family.** Considering the interesting results obtained for the PK family, our approach was extended to the NRs. These proteins play an important role in biological events such as cell growth or development and as a pathological regulator in many diseases. NRs share 3D structural motifs: the N-terminal Activation Function 1 domain (AF1, or A/B domain), the DNA-binding domain including two Zn fingers (DBD, also called the C-domain), the nuclear localization region (D domain) and the C-terminal ligand-binding domain (LBD, referred to the E domain).<sup>38</sup> In general, NRs interact with ligands in an orthosteric pocket inside the LBD and this interaction results to the cofactor binding regulation and a gene transcription regulation. However, some cavities, away from the orthosteric pocket, have been identified in several studies.<sup>39–41</sup> Although these studies suggest that those sites are putative allosteric spots, no conclusion has been drawn on the nature of those cavities.<sup>42</sup> Even if the nature of those sites is not yet fully

characterized, we applied our approach on NRs to detect the known putative allosteric sites and to identify new ones. Using the same protocol as for PKs, 591 crystal structures were extracted for the CA dataset, and 450 crystal structures for the ligand dataset (containing also AL and OL). Density analysis with a threshold of 80% showed that CA have the tendency to interact in the AF2 co-regulator site,<sup>42</sup> a putative allosteric site near the helix H12 (Figure 5, panel a).



**Figure 5.** Allosteric sites identification on NR. (a) Superimposition of density points on the reference structure (PDB ID 2PIP) with the orthosteric site represented in green. All purple spheres represent a region with 50% density of presence and only 2 spheres have a density greater than 80%. (b) Clustering results obtained with parameter [3 - 2]. (c) Some heterogeneous clusters superimposed on the reference structure indicate the position of defined allosteric ligands (PDB IDs 2PIP, top and 2PIN, middle) and orthosteric ligand (PDB ID 2PIP, bottom) represented in stick. (d) Some homogeneous clusters superimposed on the reference structure show undefined allosteric sites. Orthosteric ligand (PDB ID 2PIP) is represented in stick. Cluster ID and some protein helices are indicated.

This site was also detected using the clustering analysis. In fact, as shown in Figure 5, panel c, heterogeneous clusters 5 and 6 are located in the same pocket than the allosteric reference ligand which is usually associated with a thyroid hormone receptor (PDB ID 2PIN).<sup>43</sup> A surface called binding function 3 or BF3<sup>44</sup> have been identified as a putative allosteric site in NR like androgen receptor.<sup>45</sup> This site was exclusively found in cluster 9 using the clustering analysis. Thus, these first results confirm the existence of these two putative allosteric sites in NRs. Moreover, density analysis at 50% threshold also revealed an additional cavity (Figure 5, panel a) filled by homogeneous clusters 1, 2 and 4 constituted only by CAs (Figures 5, panels b, d). The homogeneous cluster 7 also points a surface-exposed cavity. These two cavities are still untargeted since no AL reference ligand was identified in NRs.

Regarding the NRs, the combination of both methods (density and clustering) showed that the CA are located into validated known sites described in the literature. Consequently, our approach is not limited to PKs but can be successfully extended to other protein superfamilies like NRs. Considering the

size of those superfamilies a further study could be performed on each different sub-family or sub-group.

Nowadays, allostery is a fundamental concept in protein regulation, and reveal a great interest in order to modulate the activity of a biological target in a context of drug discovery. Here, we proposed a novel and efficient computational way to detect allosteric sites, using crystallization additives (CA), instead of protein cavity volume, and their overlap with orthosteric and allosteric ligands. CA are often present in 3D structures but generally unexploited in computational methods. We put forward that CA are not randomly distributed around protein structures but seem to be attracted by hotspots. We demonstrate, with unsupervised classification and density analysis, that CA tend to be attracted by orthosteric and allosteric sites of Protein Kinases and Nuclear Receptors. Indeed, all those cavities of interest have been detected by one or both of those methods. This leads to the conclusion that location of CA in crystal structures can be used to identify putative allosteric sites. Finally, this method has been effectively applied on two protein families and could be applied on other therapeutic targets for which allosteric cavities are still unexplored. Thus, our method could be used for designing new allosteric drugs for the treatment of human diseases.

## ASSOCIATED CONTENT

### Supporting Information

This material is available free of charge via the internet. Details of the methodology, dataset composition, general process, density calculation and unsupervised classification are provided. Supplementary figures and table regarding the process, datasets, CA considered and calculation on clusters are also given.

## ACKNOWLEDGMENT

The authors wish to thank the Région Centre Val de Loire for financial support (project Drug4Arth). P.B. is supported by LABEX SynOrg (ANR-11-LABX-0029).

## ABBREVIATIONS

NR, Nuclear Receptor; PK, Protein Kinase; CA, Crystallization additives; OL, orthosteric ligands; AL, allosteric ligands; PIF, PDK1 interacting fragment; DEF, Docking site for ERK FXF; FDA, Food and Drug Administration; AF, Activation function; BF, binding function 3; DBD, DNA binding domain; LBD, ligand binding domain; ATP, adenosine triphosphate.

## REFERENCES

- (1) Wenthur, C. J.; Gentry, P. R.; Mathews, T. P.; Lindsley, C. W. Drugs for Allosteric Sites on Receptors. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 165–184. <https://doi.org/10.1146/annurev-pharmtox-010611-134525>.
- (2) Guo, J.; Zhou, H. X. Protein Allostery and Conformational Dynamics. *Chemical Reviews*. **2016**, 6503–6515. <https://doi.org/10.1021/acs.chemrev.5b00590>.
- (3) Jacob, F.; Monod, J. ‘Genetic Regulatory Mechanisms in the-Synthesis. *J Mol Biol* **1961**, *3*, 318–356.
- (4) Monod, J.; Jacob, F. General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **1961**, *26*, 389–401. <https://doi.org/10.1101/SQB.1961.026.01.048>.
- (5) Tsai, C.-J.; Nussinov, R. A Unified View of How Allostery Works. *PLoS Comput. Biol.* **2014**, *10*, e1003394. <https://doi.org/10.1371/journal.pcbi.1003394>.
- (6) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508*, 331–339. <https://doi.org/10.1038/nature13001>.
- (7) Suplatov, D.; Švedas, V. Study of Functional and Allosteric Sites in Protein Superfamilies. *Acta Naturae*, **2015**, *7* (4 (27)).
- (8) Wrabl, J. O.; Gu, J.; Liu, T.; Schrank, T. P.; Whitten, S. T.; Hilser, V. J. The Role of Protein Conformational Fluctuations in Allostery, Function, and Evolution. *Biophys. Chem.* **2011**, *159*, 129–141. <https://doi.org/10.1016/j.bpc.2011.05.020>.
- (9) Peracchi, A.; Mozzarelli, A. Exploring and Exploiting Allostery: Models, Evolution, and Drug Targeting. *Biochimica et Biophysica Acta - Proteins and Proteomics*. Elsevier August 1, 2011, pp 922–933. <https://doi.org/10.1016/j.bbapap.2010.10.008>.
- (10) Wild, C.; Cunningham, K. A.; Zhou, J. Allosteric Modulation of G Protein-Coupled Receptors: An Emerging Approach of Drug Discovery. *Austin J. Pharmacol. Ther.* **2014**, *2*.
- (11) Chaikuad, A.; Diharce, J.; Schröder, M.; Foucourt, A.; Leblond, B.; Casagrande, A.-S.; Désiré, L.; Bonnet, P.; Knapp, S.; Besson, T. An Unusual Binding Model of the Methyl 9-Anilinothiazolo[5,4-*f*] Quinazoline-2-Carbimides (EHT 1610



- and EHT 5372) Confers High Selectivity for Dual-Specificity Tyrosine Phosphorylation-Regulated Kinases. *J. Med. Chem.* **2016**, *59*, 10315–10321. <https://doi.org/10.1021/acs.jmedchem.6b01083>.
- (12) Fabbro, D.; Cowan-Jacob, S. W.; Moebitz, H. Ten Things You Should Know about Protein Kinases: IUPHAR Review 14. *Br. J. Pharmacol.* **2015**, *172*, 2675–2700. <https://doi.org/10.1111/bph.13096>.
  - (13) Carles, F.; Bourg, S.; Meyer, C.; Bonnet, P. PKIDB: A Curated, Annotated and Updated Database of Protein Kinase Inhibitors in Clinical Trials. *Molecules* **2018**, *23*, 908. <https://doi.org/10.3390/molecules23040908>.
  - (14) Endicott, J. A.; Noble, M. E. M.; Johnson, L. N. The Structural Basis for Control of Eukaryotic Protein Kinases. *Annu. Rev. Biochem.* **2012**, *81*, 587–613. <https://doi.org/10.1146/annurev-biochem-052410-090317>.
  - (15) Fang, Z.; Grü, C.; Rauh, D.; Grütter, C.; Rauh, D. Strategies for the Selective Regulation of Kinases with Allosteric Modulators: Exploiting Exclusive Structural Features. *ACS Chem. Biol.* **2013**, *8*, 58–70. <https://doi.org/10.1021/cb300663j>.
  - (16) Mobitz, H.; Jahnke, W.; Cowan-Jacob, S. Expanding the Opportunities for Modulating Kinase Targets with Allosteric Approaches. *Curr. Top. Med. Chem.* **2017**, *17*, 59–70. <https://doi.org/10.2174/1568026616666160719165314>.
  - (17) Wu, P.; Nielsen, T. E.; Clausen, M. H. FDA-Approved Small-Molecule Kinase Inhibitors. *Trends Pharmacol. Sci.* **2015**, *36*, 422–439. <https://doi.org/10.1016/j.tips.2015.04.005>.
  - (18) Rivat, C.; Sar, C.; Mechaly, I.; Leyris, J.-P.; Diouloufet, L.; Sonrier, C.; Philipson, Y.; Lucas, O.; Mallié, S.; Jouvenel, A.; et al. Inhibition of Neuronal FLT3 Receptor Tyrosine Kinase Alleviates Peripheral Neuropathic Pain in Mice. *Nat. Commun.* **2018**, *9*, 1042. <https://doi.org/10.1038/s41467-018-03496-2>.
  - (19) Bestgen, B.; Krimm, I.; Kufareva, I.; Kamal, A. A. M.; Seetoh, W.-G.; Abell, C.; Hartmann, R. W.; Abagyan, R.; Cochet, C.; Le Borgne, M.; et al. 2-Aminothiazole Derivatives as Selective Allosteric Modulators of the Protein Kinase CK2. 1. Identification of an Allosteric Binding Site. *J. Med. Chem.* **2019**, *62*, 1803–1816. <https://doi.org/10.1021/acs.jmedchem.8b01766>.
  - (20) Huang, Z.; Zhu, L.; Cao, Y.; Wu, G.; Liu, X.; Chen, Y.; Wang, Q.; Shi, T.; Zhao, Y.; Wang, Y.; et al. ASD: A Comprehensive Database of Allosteric Proteins and Modulators. *Nucleic Acids Res.* **2011**, *39* (Database), D663–D669. <https://doi.org/10.1093/nar/gkq1022>.
  - (21) Huang, W.; Wang, G.; Shen, Q.; Liu, X.; Lu, S.; Geng, L.; Huang, Z.; Zhang, J. ASBench: Benchmarking Sets for Allosteric Discovery: Fig. 1. *Bioinformatics* **2015**, *31*, 2598–2600. <https://doi.org/10.1093/bioinformatics/btv169>.
  - (22) Xu, Y.; Wang, S.; Hu, Q.; Gao, S.; Ma, X.; Zhang, W.; Shen, Y.; Chen, F.; Lai, L.; Pei, J. CavityPlus: A Web Server for Protein Cavity Detection with Pharmacophore Modelling, Allosteric Site Identification and Covalent Ligand Binding Ability Prediction. *Nucleic Acids Res.* **2018**. <https://doi.org/10.1093/nar/gky380>.
  - (23) Ma, X.; Meng, H.; Lai, L. Motions of Allosteric and Orthosteric Ligand-Binding Sites in Proteins Are Highly Correlated. *J. Chem. Inf. Model.* **2016**, *56*, 1725–1733. <https://doi.org/10.1021/acs.jcim.6b00039>.
  - (24) Goncarencu, A.; Mitternacht, S.; Yong, T.; Eisenhaber, B.; Eisenhaber, F.; Berezovsky, I. N. SPACER: Server for Predicting Allosteric Communication and Effects of Regulation. *Nucleic Acids Res.* **2013**, *41* (Web Server issue), W266–72. <https://doi.org/10.1093/nar/gkt460>.
  - (25) Allen, K. N.; Bellamacina, C. R.; Ding, X.; Jeffery, C. J.; Mattos, C.; Petsko, G. A.; Ringe, D. An Experimental Approach to

- Mapping the Binding Surfaces of Crystalline Proteins<sup>†</sup>. *J. Phys. Chem.* **1996**, *100*, 2605–2611.  
<https://doi.org/10.1021/jp952516o>.
- (26) Vera, L.; Czarny, B.; Georgiadis, D.; Dive, V.; Stura, E. A. Practical Use of Glycerol in Protein Crystallization. *Cryst. Growth Des.* **2011**, *11*, 2755–2762.  
<https://doi.org/10.1021/cg101364m>.
- (27) Drwal, M. N.; Jacquemard, C.; Perez, C.; Desaphy, J.; Kellenberger, E. Do Fragments and Crystallization Additives Bind Similarly to Drug-like Ligands? *J. Chem. Inf. Model.* **2017**, *57*, 1197–1209.  
<https://doi.org/10.1021/acs.jcim.6b00769>.
- (28) Drwal, M. N.; Bret, G.; Perez, C.; Jacquemard, C.; Desaphy, J.; Kellenberger, E. Structural Insights on Fragment Binding Mode Conservation. *J. Med. Chem.* **2018**, *61*, 5963–5973.  
<https://doi.org/10.1021/acs.jmedchem.8b00256>.
- (29) English, A. C.; Groom, C. R.; Hubbard, R. E. Experimental and Computational Mapping of the Binding Surface of a Crystalline Protein. *Protein Eng. Des. Sel.* **2001**, *14*, 47–59.  
<https://doi.org/10.1093/protein/14.1.47>.
- (30) Liepinsh, E.; Otting, G. Organic Solvents Identify Specific Ligand Binding Sites on Protein Surfaces. *Nat. Biotechnol.* **1997**, *5*, 264–268.
- (31) Bosc, N.; Wroblowski, B.; Aci-Sèche, S.; Meyer, C.; Bonnet, P. A Proteomic Analysis of Human Kinome: Insight into Discriminant Conformation-Dependent Residues. *ACS Chem. Biol.* **2015**, *10*, 2827–2840.  
<https://doi.org/10.1021/acschembio.5b00555>.
- (32) Tzarum, N.; Komornik, N.; Ben Chetrit, D.; Engelberg, D.; Livnah, O. DEF Pocket in P38 $\alpha$  Facilitates Substrate Selectivity and Mediates Autophosphorylation. *J. Biol. Chem.* **2013**, *288*, 19537–19547.  
<https://doi.org/10.1074/jbc.M113.464511>.
- (33) Sadowsky, J. D.; Burlingame, M. A.; Wolan, D. W.; McClendon, C. L.; Jacobson, M. P.; Wells, J. A. Turning a Protein Kinase on or off from a Single Allosteric Site via Disulfide Trapping. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 6056–6061.  
<https://doi.org/10.1073/pnas.1102376108>.
- (34) Heo, Y.-S.; Kim, S.-K.; Seo, C. Il; Kim, Y. K.; Sung, B.-J.; Lee, H. S.; Lee, J. Il; Park, S.-Y.; Kim, J. H.; Hwang, K. Y.; et al. Structural Basis for the Selective Inhibition of JNK1 by the Scaffolding Protein JIP1 and SP600125. *EMBO J.* **2004**, *23*, 2185–2195.  
<https://doi.org/10.1038/sj.emboj.7600212>.
- (35) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters a Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. AAAI Press 1996, pp 226–231.
- (36) McIntyre, P. J.; Collins, P. M.; Vrzal, L.; Birchall, K.; Arnold, L. H.; Mpamhanga, C.; Coombs, P. J.; Burgess, S. G.; Richards, M. W.; Winter, A.; et al. Characterization of Three Druggable Hot-Spots in the Aurora-A/TPX2 Interaction Using Biochemical, Biophysical, and Fragment-Based Approaches. *ACS Chem. Biol.* **2017**, *12*, 2906–2914.  
<https://doi.org/10.1021/acschembio.7b00537>.
- (37) Chen, L. S.; Redkar, S.; Bearss, D.; Wierda, W. G.; Gandhi, V. Pim Kinase Inhibitor, SGI-1776, Induces Apoptosis in Chronic Lymphocytic Leukemia Cells. *Blood* **2009**, *114*, 4150–4157.  
<https://doi.org/10.1182/blood-2009-03-212852>.
- (38) Billas, I.; Moras, D. Allosteric Controls of Nuclear Receptor Function in the Regulation of Transcription. *Allosteric Interact. Biol. Regul. Part II* **2013**, *425*, 2317–2329.  
<https://doi.org/10.1016/j.jmb.2013.03.017>.
- (39) Hughes, T. S.; Giri, P. K.; de Vera, I. M. S.; Marciano, D. P.; Kuruvilla, D. S.; Shin, Y.; Blayo, A.-L.; Kamenecka, T. M.; Burris, T. P.; Griffin, P. R.; et al. An Alternate Binding Site for PPAR $\gamma$  Ligands. *Nat. Commun.* **2014**, *5*, 3571.
- (40) Burris, T.; Solt, L.; Wang, Y.; Crumbley, C.; Banerjee, S.; Griffett, K.; Lundasen, T.;

Hughes, T.; Kojetin, D. *Nuclear Receptors and Their Selective Pharmacologic Modulators*; 2013; Vol. 65.

- (41) Moore, T. W.; Mayne, C. G.; Katzenellenbogen, J. A. Minireview: Not Picking Pockets: Nuclear Receptor Alternate-Site Modulators (NRAMs). *Mol. Endocrinol. Baltim. Md* **2010**, *24*, 683–695. <https://doi.org/10.1210/me.2009-0362>.
- (42) Changeux, J.-P.; Christopoulos, A. Allosteric Modulation as a Unifying Mechanism for Receptor Function and Regulation. *Cell* **2016**, *166*, 1084–1102. <https://doi.org/10.1016/j.cell.2016.08.015>.
- (43) Estébanez-Perpiñá, E.; Arnold, L. A.; Jouravel, N.; Togashi, M.; Blethrow, J.; Mar, E.; Nguyen, P.; Phillips, K. J.; Baxter, J. D.; Webb, P.; et al. Structural Insight into the Mode of Action of a Direct Inhibitor of Coregulator Binding to the Thyroid Hormone Receptor. *Mol. Endocrinol.* **2007**, *21*, 2919–2928. <https://doi.org/10.1210/me.2007-0174>.
- (44) Buzón, V.; Carbó, L. R.; Estruch, S. B.; Fletterick, R. J.; Estébanez-Perpiñá, E. A Conserved Surface on the Ligand Binding Domain of Nuclear Receptors for Allosteric Control. *Mol. Cell. Endocrinol.* **2012**, *348*, 394–402. <https://doi.org/10.1016/J.MCE.2011.08.012>.
- (45) Estébanez-Perpiñá, E.; Arnold, L. A.; Arnold, A. A.; Nguyen, P.; Rodrigues, E. D.; Mar, E.; Bateman, R.; Pallai, P.; Shokat, K. M.; Baxter, J. D.; et al. A Surface on the Androgen Receptor That Allosterically Regulates Coactivator Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16074–16079. <https://doi.org/10.1073/pnas.0708036104>.

# Graphical Abstract

