Research Article




# Prediction of Protein Allosteric Signalling Pathways and Functional Residues Through Paths of Optimised Propensity

**Nan Wu** [1], **Sophia N. Yaliraki** [1] and **Mauricio Barahona** [2*]

*1 - **Department of Chemistry Imperial College London,** United Kingdom*
*2 - **Department of Mathematics,** Imperial College London, United Kingdom*

***Correspondence to Mauricio Barahona:*** *m.barahona@imperial.ac.uk (M. Barahona)@CMPHImperial (M. Barahona)*
https://doi.org/10.1016/j.jmb.2022.167749
***Edited by Igor Berezovsky***

## Abstract

Allostery commonly refers to the mechanism that regulates protein activity through the binding of a molecule at a different, usually distal, site from the orthosteric site. The omnipresence of allosteric regulation in nature and its potential for drug design and screening render the study of allostery invaluable. Nevertheless, challenges remain as few computational methods are available to effectively predict allosteric sites, identify signalling pathways involved in allostery, or to aid with the design of suitable molecules targeting such sites. Recently, bond-to-bond propensity analysis has been shown successful at identifying allosteric sites for a large and diverse group of proteins from knowledge of the orthosteric sites and its ligands alone by using network analysis applied to energy-weighted atomistic protein graphs. To address the identification of signalling pathways, we propose here a method to compute and score paths of optimised propensity that link the orthosteric site with the identified allosteric sites, and identifies crucial residues that contribute to those paths. We showcase the approach with three well-studied allosteric proteins: h-Ras, caspase-1, and 3-phosphoinositide-dependent kinase-1 (PDK1). Key residues in both orthosteric and allosteric sites were identified and showed agreement with experimental results, and pivotal signalling residues along the pathway were also revealed, thus providing alternative targets for drug design. By using the computed path scores, we were also able to differentiate the activity of different allosteric modulators.

© **2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).**

## Introduction

Allostery, first introduced as a term over 50 years ago,[1] is a crucial mechanism in controlling and modulating biological and cellular processes.[2] The classic allosteric effect refers to the modification of the activity at the orthosteric site of an enzyme, due to the binding of a ligand at another, distant site (the allosteric site). The effect of this distal allosteric binding is transmitted through the structure and dynamics of the biological macromolecule to affect the active site.[3] The discovery of single-domain allosteric proteins[4] and innate dynamic behaviour of enzymes[5] indicates that allostery is a pervasive phenomenon. All proteins could be allosteric,[6] and allosteric regulation of proteins could potentially expand the space of druggable targets.

Controlled up- and down-regulation of protein activity by allosteric modulation is achieved through changes in the conformation and/or dynamics of the protein upon ligand binding events at the allosteric site.[7] Such induced changes can enhance or reduce the binding affinity of natural substrates at the orthosteric site.[8] Allosteric drugs

0022-2836/© 2022 The Authors. Published by Elsevier Ltd.This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

would thus present advantages, since controlled regulation is hard to attain with traditional orthosteric drugs, which interfere directly with binding at the active site. Furthermore, allosteric targets could also improve drug specificity: the existence of multiple potential allosteric sites on a single protein could provide higher specificity,[9] compared to targeting orthosteric sites, which are well conserved across families of proteins with similar functions.[10] Yet, despite these potential advantages, the underlying mechanisms of allostery are poorly understood[11] and a key challenge is to find *bona fide* allosteric sites (or allosteric modifying residues) as drug targets.

Both experimental and computational methods have been developed to identify putative allosteric sites. Experimental methods, including tethering,[12,13] nuclear magnetic resonance (NMR) techniques[14,15] and high-throughput screening combined with X-ray crystallography,[16,17] have successfully found druggable allosteric sites. Yet these methods suffer from laborious screening of huge compound libraries and serendipitous discovery. Alternatively, computational methodologies for predicting allosteric sites can take advantage of ever-increasing availability of protein structure data,[18] including for allosteric proteins (AlloSteric Database (ASD),[19] ASBench[20] and CASBench[21] benchmark sets), and rising computational power (for reviews see Refs.[22,23]). Several computational methods have focused on understanding communication pathways for the identification of allosteric sites using, e.g., molecular dynamics (MD) simulations,[24–26] measures based on normal mode analysis (NMA) of elastic network models (PARS,[27,28] AlloPred[29] and AllositePro[30]) as well as machine learning approaches exploiting allosteric site features.[31,32] Although these methods have achieved some success in predicting allosteric sites, their prediction accuracy is moderate. Furthermore, only MD simulations keep the atomistic detail of structures, although at the cost of substantial computational power,[33] whereas most of the other methods circumvent this limitation through the use of coarse-grained representations of proteins (usually at the level of residues) at the cost of losing resolution.[34] Different from the general methods discussed above, Guarnera and Berezovsky introduced a structure-based statistical mechanical model of allostery (SBSMMA) relying on calculations of changes in the Gibbs free energy upon perturbation.[35] Fogha *et al.* analysed the density and clustering of crystallisation additives to find putative allosteric sites.[36] Moreover, Wang *et al.* applied a perturbation propagation probability algorithm (Ohm) that acts on an unweighted residue-level graph based on distance-cutoffs to predict hotspots.[37]

Bond-to-bond propensities[38] is based on the analysis of high-resolution (atomistic) energy-weighted protein graphs constructed from structural data, and was shown to achieve high prediction accuracy while staying computationally efficient. Recently, the method has been further benchmarked on two large allosteric protein databases, ASBench (118 structures) and CASBench (314 structures), and achieved over 89.8% and 98.1% accuracy, respectively.[39,40] Bond-to-bond propensities quantify how the fluctuations applied to a group of bonds ('source') affect any other bond in the protein, and summarises this impact probabilistically through the *quantile score* (QS), a propensity score computed using quantile regression against the distance to the source. This procedure enables both short- and long-range couplings to be measured, a key feature of allosteric signalling. Importantly, the complexity of protein structures is retained by the fully atomistic graph,[41,42] yet advances in algorithmic matrix theory[43,44] make calculation of propensities efficient for large systems. These features make bond-to-bond propensity analysis a cost-effective computational method to predict allosteric sites using high-resolution structures.

It has been argued that allostery can be understood in terms of the propagation of allosteric signals through pre-existing pathways within the protein,[45] and that understanding such allosteric pathways is essential to determine sites and residues that are strongly coupled to the orthosteric site so as to make allosteric prediction more robust.[46] Although much effort has been devoted to understanding protein dynamics and allostery via solution NMR,[47] our focus here is on available computational methods. Demerdash *et al.* applied machine learning based on structural and network features to predict residues involved in allosteric signalling pathways which then allow the determination of allosteric sites.[48] MCPath uses Monte Carlo to predict potential functional residues and allosteric pathways.[49] Botello-Smith *et al.* obtained residue-level protein graphs from MD simulations and useed current-flow betweenness to identify key residues for signal transmission.[50] Ohm, specifically mentions predicted allosteric pathways and important residues identified within the paths, but the method relies on coarse-grained input.[37] Several of these methods have been validated with experimental results in the literature, but no single method can at once inform functional residues at orthosteric and allosteric sites together with key residues along the allosteric signalling pathways. Moreover, no attempt is made to combine allosteric site prediction and ranking of these sites at the atomistic level.

The idea of allosteric communication pathways is an intrinsic component of bond-to-bond propensity analysis, as its computation is based on the concept of signals being transmitted from the allosteric to the orthosteric site across the atomistic network. Here, we present a method that uses bond-to-bond propensities to compute and

assign a quantitative score to signalling pathways that connect the residues of the orthosteric and allosteric sites through paths of optimised propensity. For a full description, see Methods. Briefly, starting with bond-to-bond propensity analysis on atomistic graphs, all non-covalent interactions in the protein are scored with QS that reflect how strongly the bond is coupled to the 'source' (i.e., the orthosteric site). Once the atomistic computation of the QS is performed, we add non-covalent interactions, in decreasing order from high to low QS, until the shortest path(s) between the source and a given residue in the allosteric site is found, always meeting the criterion that there is no more than one consecutive step along the backbone. This requirement is inspired by the biophysical idea that allosteric signal transmission is akin to anisotropic thermal diffusion, for which heat flow is mainly via non-covalent interactions such that only the surrounding atoms (i.e., the neighbouring residues) are heated.[51] Each *propensity optimised path* (POP) thus obtained is then scored based on its sequence of propensities (see Eq (7)), thus allowing us to determine the crucial pairs of residues connecting the orthosteric and allosteric sites. In addition, to evaluate key signalling residues along the paths, we consider both residue participation in POPs and the effect that removal from the protein graph of the residue under study has on POP scores and path lengths (i.e., akin to the effect of computational mutagenesis).

We exemplify the method for allosteric signalling pathway detection and identification of key functional and signalling residues on three important enzymes: h-Ras, caspase-1, and 3-phosphoinositide-dependent kinase-1 (PDK1). In h-Ras and caspase-1, we predict correctly allosteric signalling pathways and crucial residues in agreement with experimental mutational studies. For PDK1, not only do we reveal the key signalling pathways and residues but also exploit the pathway scores to compare the effect of different ligands bound to the allosteric site, both qualitatively (inhibition *vs.* activation) and quantitatively (level of activity).

## Results

### Uncovering allosteric communication pathways and key signalling residues in h-Ras

The enzyme h-Ras, also known as transforming protein p21, is a GTPase involved in cell division regulation.[52] Calcium acetate has been shown to be an allosteric activator of h-Ras.[53] We started with a bond-to-bond propensity analysis of this structure (PDB ID: 3K8Y) using as 'source' its natural substrate, phosphoaminophosphonic acid-guanylate ester (GNP), and obtained the quantile scores of all non-covalent interactions, including hydrogen bonds, salt bridges and hydrophobic interactions (see also Ref. [42]). The orthosteric and allosteric sites were defined as all the residues interacting with, respectively, GNP (the substrate) and calcium acetate (the activator). We then obtained and scored all the propensity optimised paths between all the residues in the orthosteric and allosteric sites. See Methods for a detailed description of the approach. The computed POP scores (7) are collated in Figure 1.

The highest path score for h-Ras is 0.959 (see Figure 1a)), and was obtained for paths starting from 5 residues in the orthosteric site (Ser17, Ala18, Gln61, HOH172, HOH175) and reaching only three residues in the allosteric site (Arg97, Val109, Met111). To assess whether these sets of residues have been found to be essential for allosteric signalling, we compared our results to a network of hydrogen bonds connecting the allosteric and orthosteric sites, as proposed by Buhrman *et al.*[53] Within this network, the catalytic (HOH175) and bridging (HOH189) water molecules together with the key catalytic residue Gln61 are essential for the intrinsic hydrolysis in h-Ras, whereas Arg97 and Val109 are responsible for the binding of the acetate ion at the allosteric site. Our results show that four of these five residues are involved in allosteric pathways with the highest path score and HOH189 is involved in the pathway with the second highest score (0.956), which agree with the hydrogen-bonding network. Taking a closer look at the POP with the highest score from Ser17, Ala18 and HOH172 to Arg97, we see that they all converge into Gln61, and the final five steps are actually the POP from Gln61 to Arg97 with sequence Gln61-Gly60-HOH367-Tyr96-Arg97 (coloured in blue in Figure 1(b)). This sequence, where three of our propensity optimised paths converge, is indeed a section of the hydrogen-bonding network proposed by Buhrman *et al.*,[53] further supporting the finding of Gln61 as a key catalytic residue. Interestingly, the other three POPs with the highest score (the paths from Gln61 and HOH175 to Met111, and the path from HOH175 to Val109) do not involve the hydrogen-bonding network and reveal a divergent signalling pathway also starting at Gln61. Indeed, these three paths all start with (Gln61)-HOH175-Thr35-Mg168-Ser17-Lys16 and then pass through $\beta$4 and $\beta$5 strands (coloured in red and orange in Figure 1(b)-(c)). The appearance of the water molecule HOH175 and the magnesium ion Mg168 in these additional high score paths highlights their known importance for h-Ras catalytic processes.[54]

It is illustrative to compare our results to those obtained with other cutting-edge computational methods developed to quantify the coupling between orthosteric and allosteric sites based on other principles. Specifically, we analysed h-Ras with AlloSigMA 2, a recently developed tool based on SBSMMA.[55] By DOWN-mutating the allosteric residues (substituting them by small Ala/Gly-like residues) and quantifying the effect on the orthos-

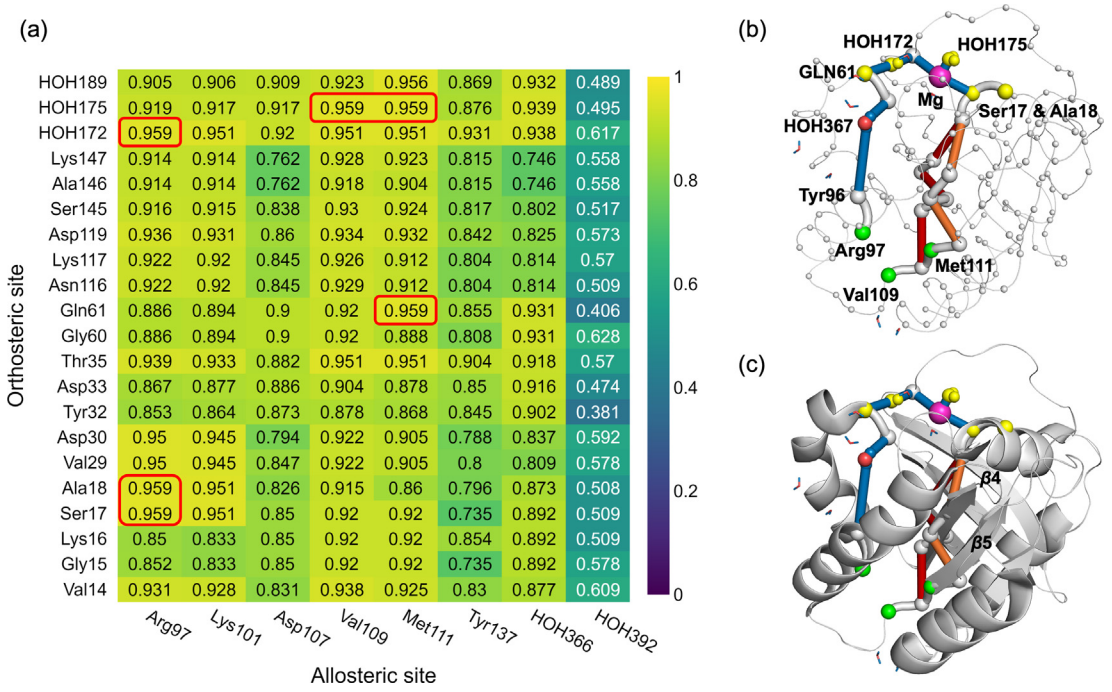



**Figure 1. Computation of propensity optimised paths between the orthosteric and allosteric site residues for activator-bound h-Ras (PDB ID: 3K8Y).** (a) Scores of propensity optimised paths for all residue pairs, with the ighest path score of 0.959 indicated with a red frame. (b) Propensity optimised paths with maximum score of 0.959 for h-Ras. The orthosteric residues (yellow spheres) communicate with allosteric residues (green spheres) via three distinct pathways (coloured in blue, red and orange). In each of the paths, the grey links are steps on the backbone (i.e., involving neighbouring residues) and the coloured connections are non-covalent interactions. (c) Superposition of the three paths onto the structure of h-Ras. The blue path passes through the hydrogen-bond network whereas the other two paths (red, orange) go through the $\beta$4 and $\beta$5 strands.

teric site using the Allosteric Signalling Map (ASM), Val109 and Met111 are indicated as key allosteric residues, in agreement with our results (Figure 1). On the other hand, ASM does not highlight the importance of the allosteric residue Arg97, which was found to be responsible for the binding of the acetate ion at the allosteric site. In addition, our method explicitly provides the intra-protein signalling pathway linking orthosteric and allosteric residues, thus allowing further understanding of structural information and access to other relevant residues and interactions that could be targeted. Detailed information can be found in Supplementary Information Part 1 and Figure S1.

It is also worth noting that when analysing h-Ras without the activator calcium acetate (PDB ID: 3LBN), we find substantially altered high-scoring POPs connecting the orthosteric and allosteric sites. In particular, the top six POPs for 3LBN, which connect HOH202, Lys16 and Ser17 in the orthosteric site to Val109 and Lys101 in the allosteric site, traverse the Switch II region, as well as the $\alpha$1 and $\beta$1, $\beta$2 and $\beta$3 strands. Hence, some of the most prominent allosteric signalling paths in h-Ras are only activated upon binding of the activator ligand. Yet, key paths involving Gln61 still include water molecules (HOH227 in 3LBN), thus underscoring their crucial role in h-Ras allostery (detailed paths and scores for 3LBN can be found in Supplementary Information Part 2 and Figure S2).

Returning to h-Ras with the bound activator (PDB ID: 3K8Y), we then identified key residues along the computed allosteric signalling paths by computing two measures of residue importance (see Methods). First, we obtained the residue participation frequency in the computed POPs. In total, there are 153 residues in the protein (excluding the orthosteric and allosteric sites) and 68 (out of 153) are involved in the POPs connecting the 21 residues in the orthosteric site

Table 1 Top five signalling residues within the POPs (excluding orthosteric and allosteric sites), according to frequency of participation in the POPs computed for h-Ras (PDB ID: 3K8Y).

| Residue | Residue frequency |
|---|---|
| Tyr96 | 46 |
| Ile100 | 43 |
| HOH367 | 42 |
| Leu133 | 34 |
| Ile84 | 29 |

Table 2 Top five signalling residues within the POPs (excluding orthosteric and allosteric sites), according to z-scores computed from the change in path score and path length upon residue removal for h-Ras (PDB ID: 3K8Y).

| Residue | z-score (path score) | Residue | z-score (path length) |
|---|---|---|---|
| Mg168 | 11.13 | Mg168 | 8.22 |
| Val112 | 1.55 | HOH367 | 5.18 |
| Ser106 | 1.39 | Tyr96 | 5.08 |
| Ser136 | 1.36 | Ser136 | 2.41 |
| Glu62 | 1.22 | Thr58 | 1.46 |

with the 8 residues in the allosteric site. On average, each of those 68 residues participates in 10 POPs but some residues have much higher participation frequency in those paths. The top five residues according to their POP participation frequency are shown in Table 1. Tyr96, Ile100 and HOH367 are the most involved residues in allosteric signalling pathways. Of these, HOH367 and Tyr96 lie on the shortest path from Gln61 to the allosteric site within the hydrogen-bonding network,[53] and Ile100 forms a hydrogen bond with Tyr96 of high propensity (QS = 0.91), indicating the importance of Tyr96. The other two residues do not feature in the hydrogen-bonding network and might hint at other unknown pathways.

As a second measure of residue importance, we computed the impact that the computational removal of a residue has on pathway scores and lengths. To do this, we mimic *alanine scanning mutagenesis* computationally by removing from the atomistic graph all the weak interactions associated with each residue of h-Ras, one residue at a time. This approach has been shown to be an efficient and robust computational means to mimicking the effect of alanisation, under the assumption that no large structural rearrangements occur.[42,56] The POP scores and lengths were then recomputed in the modified graph, and we obtained z-scores for each residue (see Methods). Table 2 shows the top five residues with the highest z-scores with respect to path scores and path lengths. The magnesium ion (Mg168), which is known to be a key to facilitating GNP binding,[54] appears as the most influential according to both scores. The z-scores of the POP scores are low and do not uncover any other residues in the hydrogen bond network. This is not surprising since POPs involve bonds with high QS, and the removal on one residue only reroutes the paths through other residues with high QS, with little change in the POP scores. Therefore, only key bonds essential in maintaining a high POP score, in this case the electrostatic interactions with Mg ion, are found. The z-scores for path lengths, on the other hand, reveal more structural features and indicate residues whose removal leads to longer paths. The bonds and local structure associated with such residues are responsible for connecting two sites with the shortest paths. Again, we find Mg168 again as the top residue, followed by HOH367 and Tyr96, consistent with the hydrogen bond network in Ref. [53]

## Key allosteric residues in caspase-1

Caspase-1 is a well-studied allosteric protein, which plays an important role in apoptotic processes in the cell. [57] In Ref. [42], bond-to-bond propensity was used to uncover the allosteric site using the crystal structure (PDB ID: 2HBQ) of wild-type human caspase-1 in complex with the orthosteric ligand z-VAD-FMK, Here, we take this analysis further to find out functionally important residues in connecting the orthosteric and allosteric sites. Following the procedure described above, we found and scored propensity optimised paths for each residue pair between the orthosteric and allosteric sites. These scores are compiled in Figure 2. Note that caspase-1 is a tetramer formed by two identical heterodimers, and the propensity calculation is indeed conducted on the complete tetramer, but we show the path scores are presented for half of the structure (chains A and B) since the scores for the other half (chains C and D) are the same.

Figure 2 reveals a strong connection between the orthosteric site (formed in this case by 13 residues) and the allosteric site (9 residues), as indicated by the high POP scores. Cys285, to which z-VAD-FMK is covalently attached, is the residue in the orthosteric site that is most strongly to the allosteric site residues (with most path scores $\geqslant$0.99). Residues Val338, Ser339, and Trp340 are also strongly coupled to the allosteric site. Focusing on the residues of the allosteric site, we found that Arg286, Glu390 and Arg391 connect strongly to Cys285 (with path scores equal to 1.0), noting that the path from Cys285 to Arg391 contains Arg286 and Glu390 as intermediate residues. Hence, Arg286 and Glu390 are crucial in the orthosteric-allosteric coupling, in agreement with experimental results that show them to be part of a hydrogen bond network connecting orthosteric and allosteric sites, and, additionally, the fact that the interaction between these two residues is disrupted upon allosteric inhibitor binding.[58]

To explore further the importance of these two residues, we repeated the analysis for the structures of two experimentally resolved caspase-1 mutants: Arg286Ala (PDB ID: 2HBR)

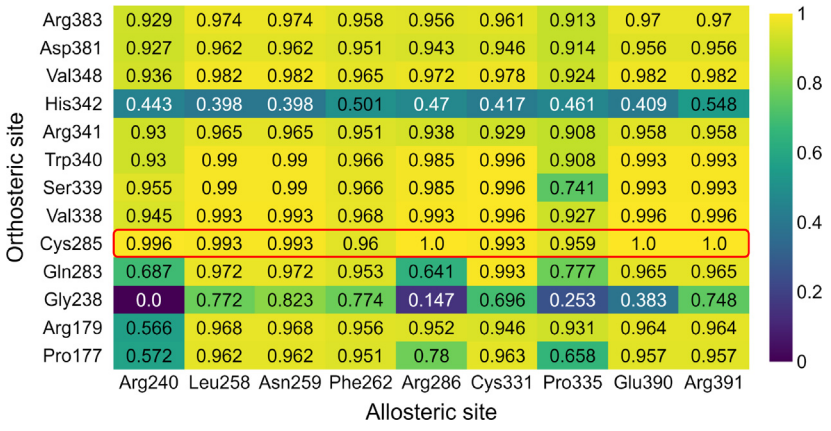



**Figure 2. Scores of propensity optimised paths between the residues of the orthosteric and allosteric sites for wild type caspase-1 (PDB ID: 2HBQ).** Highlighted (red frame) is the strong connection from Cys285, a key residue at the orthosteric site, to the allosteric residues.

and Glu390Ala PDB ID: 2HBY). The results are presented in Supplementary Information Part 3 and Figure S3. Upon Arg286Ala and Glu390Ala mutations, the path scores from Cys285 to Arg286 remain unchanged at 1.0, whereas the POP score from Cys285 to Glu390 drops from 1.0 (the top allosteric signalling pathway in wild-type caspase-1, 1/117 paths) to 0.88 and 0.937, respectively, rendering this pathway out of the top 10 signalling pathways for the mutants (78/117 paths for Arg286Ala and 48/117 paths for Glu390Ala). This indicates that Arg286 and Glu 390 are both key signalling residues that induce the loss of the strong orthosteric-allosteric connection in the observed experimental structures.

We also computed our measures of residue importance (residue frequency, and the two z-scores) for the wild type caspase-1 (PDB ID: 2HBQ), as described above. The residues with top 10 residue frequencies in all POPs were: Asn337, Ile282, His237, Asn263, Ser332, Asp336, Phe330, Ala329, Cys244 and Thr180. The results for both z-scores are shown in Table 3. Experimental evidence has shown that the Ser332Ala mutant has a 3.7-fold reduction in catalytic efficiency[58] and this residue is identified as the residue with fifth highest path participation frequency and the third highest z-score on the disruption of path lengths. In addition, the Asn337Ala mutant has a ~2-fold reduction in catalytic efficiency[58] and we found that this residue has the top path participation frequency and the sixth highest z-score of path scores. To the best of our knowledge, none of the other identified residues have been studied experimentally in the literature.

To make our measures more specifically targeted on the high scoring POPs, we considered paths with POP score equal to and above 0.99. The residue pairs involved in such high scoring POPs, shown in Table 4, are Asn337 (six times), Ser332 (four times), Asp336 (once) and Ala284 (once). Of these, Asn337 and Ser332 have appeared as important residues above, and have been shown experimentally to have a lower catalytic efficiency upon alanisation.[58] Therefore, considering high-scoring pathways and their corresponding intermediate residues can provide further insight for the determination of key signalling residues.

## Allosteric communication pathways in PDK1 under various allosteric modulators

In the previous examples, we have shown that computing and scoring propensity optimised paths

Table 3 Top 10 key signalling residues within all allosteric pathways according to z-scores based on change in path score and path length for wild type caspase-1 (PDB ID: 2HBQ).

| Residue | z-score (path score) | Residue | z-score (path length) |
|---|---|---|---|
| Ser289 | 12.52 | Asp336 | 10.71 |
| Ile282 | 3.71 | Ser289 | 6.57 |
| Cys244 | 3.40 | Ser332 | 3.91 |
| Ile243 | 3.04 | Ser136 | 3.03 |
| Ile239 | 2.86 | Thr58 | 2.73 |
| Asn337 | 2.38 | Asp108 | 2.44 |
| Ser236 | 2.24 | His94 | 2.44 |
| His237 | 2.01 | Gln129 | 2.14 |
| Gly287 | 1.72 | Glu126 | 2.14 |
| Phe349 | 1.71 | Val125 | 1.85 |

Table 4 Orthosteric and allosteric residues involved in paths with score $\geqslant 0.99$ for wild type caspase-1 (PDB ID: 2HBQ).

| Orthosteric residue(s) | Allosteric residue(s) |
|---|---|
| Cys285 | Val124, Val127, Arg131, Phe147, Thr148 |
| Val338, Ser339, Trp340 | Leu258, Asn259, Arg286, Cys331, Glu390, Arg391 |

allows us to identify crucial orthosteric, allosteric and signalling residues involved in allosteric signalling pathways. We now investigate whether the paths and scores can be used to understand and quantify the allosteric activity of different allosteric ligands.

To pursue this line of enquiry, we considered 3-phosphoinositide-dependent kinase-1 (PDK1), a member of the AGC kinase family.[59] PDK1 is constitutively active owing to the auto-phosphorylation at Ser241 (Sep241 upon phosphorylation), and it plays a critical role in phosphorylating and activating more than 23 AGC related kinases. It is also indicated in tumourigenesis.[60] Indeed, PDK1 is important in regulating the intracellular PI3K-AKT pathway, where its activation is closely related to human cancers.[61] Therefore, PDK1 modulators are attractive as anti-cancer agents.

There are three ligand binding sites on PDK1: the catalytic ATP binding site, the substrate binding site, and the PDK1 Interacting Fragment (PIF) binding site.[62] Targeting the ATP binding site (orthosteric site) with competitive inhibitors generally leads to low selectivity, as this site is conserved in over 500 protein kinases.[63] The PIF pocket is not only employed as a binding site for downstream substrate kinases, but also exploited for stimulation of PDK1 activity.[62] Hence, the PIF pocket behaves as an allosteric site and has spawned the development of non-competitive PDK1 modulators. An important example is PIFtide, a 24-amino acid polypeptide derived from PDK1 substrate PRK2, which binds to the PIF pocket and enhances 7-fold the activity of PDK1.[64]

Here, we used the crystal structure of PDK1 complexed with PIFtide (PDB ID: 4RRV, with ATP removed) for propensity analysis, where we defined the orthosteric site as the residues responsible for ATP binding, and the allosteric site as all the residues interacting with PIFtide. The results of our propensity optimised path computations are summarised in Figure 3, where we also highlight the the top 10 scoring POPs.

The key functional residues in the orthosteric and allosteric sites involved in the high-scoring POPs can be seen in Figure 3. The crystal structure solved by Rettenmaier *et al.* illustrates how the allosteric signal can be transmitted to the ATP binding site.[65] PIFtide binds to Arg131 which results in the movement of the helix $\alpha$C and leads to a change in Glu130. The ensuing hydrogen bonding between Glu130 and Lys111 then conducts the allosteric signal to the ATP binding site. Reassuringly, the highest scoring path (1.0) in our analysis was precisely Arg131–Glu130–Lys111, which is in agreement with the experimental results in Ref.[65] Alanine scanning mutagenesis conducted experimentally by Rettenmaier *et al.* determined binding energy hotspots on PIFtide and highlighted five residues (Phe14, Asp16, Phe17, Asp18, Tyr19) as the most contributing residues.[65] We found that Arg131 interacts with four of those five residues, whereas Thr148, Gln150 and Leu155 interact with two of them. This indicates that the key allosteric residues Arg131, Thr148 and Leu155, which were found amongst the paths with highest propensity (Table 5), match the experimental evidence, thus confirming their role in activation signalling.

To further test this hypothesis, we analysed the structures of PDK1 complexed with three different allosteric ligands: a strong activator J30 (PDB ID: 3OTU), a weak activator 2A2 (PDB ID: 3ORZ), and an inhibitor 1F8 (PDB ID: 3ORX). Using the same orthosteric and allosteric residues for the PDK1-PIFtide complex (except that Lys76 is not present in 3OTU), all the paths were computed and scored for each of the structures and the results are shown in Figure 4.

Comparing the scores for PDK1 bound to the strong activator J30 (Figure 4(a)) to the scores obtained for the PIFtide bound PDK1 (Figure 3), we found broad agreement: seven of the top ten scoring paths are the same and the key residues identified are generally the same. Indeed, both J30 and PIFtide are strong allosteric activators of PDK1 and the POP results confirm this. Figure 4




**Figure 3.** Scores of propensity optimised paths between the residues of the orthosteric and allosteric sites for PIFtide bound PDK1 (PDB ID: 4RRV) with top 10 path scores highlighted.

Table 5 Orthosteric and allosteric residues involved in the top 10 allosteric signalling paths predicted with path score $\geqslant$0.99 for PIFtide bound PDK1 (PDB ID: 4RRV).

| Orthosteric residue(s) | Allosteric residue(s) |
|---|---|
| Ser94 | Val124, Val127, Arg131, Phe147, Thr148 |
| Lys111 | Val124, Val127, Arg131, Leu155, Thr148 |

(b) shows the results for the weak activator 2A2. We found that although the top ten scoring POPs are different, the key residues highlighted are similar to those found with strong activators. On the other hand, when the inhibitor 1F8 is bound, the patterns of the signalling paths are strongly modified, as seen in Figure 4(c). Note how all the top ten scoring paths concentrate in one region, and the allosteric residues Lys76, Val124, Val127 and Arg131 do not appear as functionally important.

Combining the analyses of the structures bound to activators (J30, 2A2) and to the inhibitor (1F8) with the key allosteric residues in the PDK1-PIFtide structure, it emerges that interaction with Arg131 is essential for PDK1 activation, since only the inhibitor 1F8 does not make any interaction with Arg131. More specifically, Figure 5 shows that J30 forms a hydrogen bond with Arg131, whereas 1F8 does not interact with Arg131 at all. This structural analysis suggests that Lys115, Ile119, Val124 and Gln150 only play a role in positioning the ligand, but it is the hydrogen bond (circled in red) formed between the amide oxygen atom in J30 and the nitrogen bound hydrogen atom in Arg131 that initiates the allosteric signalling. Few inhibitors (other than 1F8) have been discovered for PDK1. An example is the alkaloid derivatives discovered by Bobkova *et al.*, which showed selectivity towards PDK1 and inhibition. Although no crystal structure was obtained for those compounds, *in silico* docking indicated that no binding with Arg131 is present, owing to a lack of a carboxyl moiety.[66] This again supports the argument that interaction with Arg131 could be indispensable for PDK1 activation, and avoidance of this interaction through careful design of molecules could result in an allosteric inhibitor.

The previous results show that our path analysis in PDK1 allows us to identify a qualitative difference between activating and inhibiting allosteric ligands. Our computations can also be used to quantify the effect of different ligands, and we present here some initial work in this direction. As discussed above, PDK1 remains active owing to the auto-phosphorylation of Ser241 into Sep241, and allosteric signals can also be transmitted from the allosteric site through Sep241 to control the protein activity.[65] Therefore, we examined the connection of the allosteric site to Sep241, which specifically keeps PDK1 active.

To investigate this issue, we calculated the POP scores from the modulator-bound allosteric residues to Sep241 for 11 PDB structures of

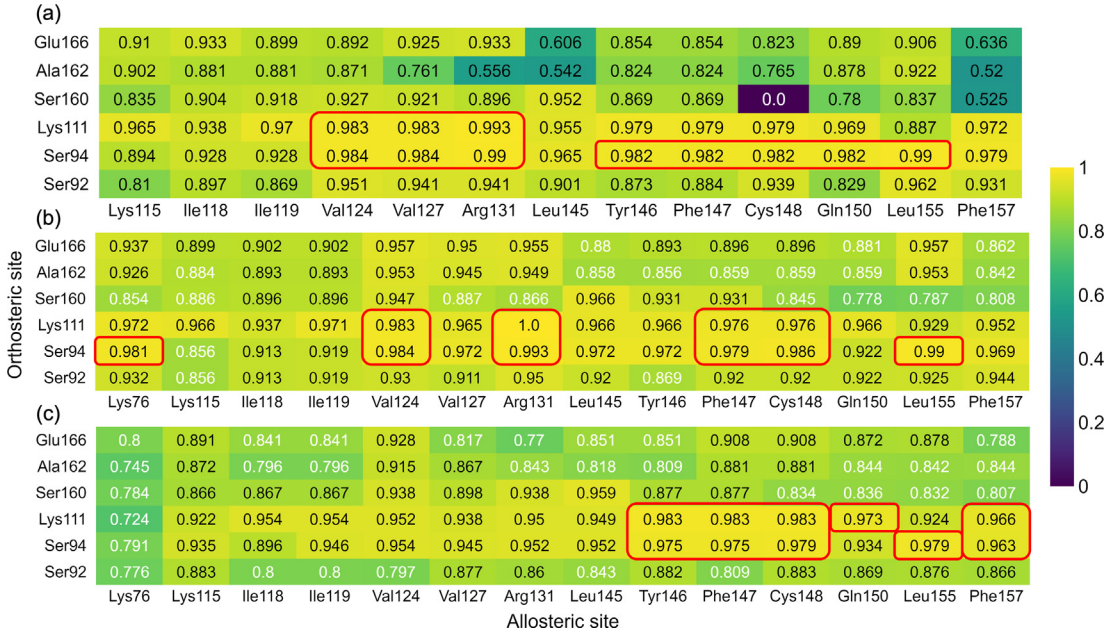



**Figure 4.** Scores of propensity optimised paths between the residues of the orthosteric and allosteric sites for three PDK1 structures bound to different activators and inhibitors with top 10 path scores highlighted. (a) Strong activator (J30) bound PDK1 (PDB ID: 3OTU). (b) Weak activator (2A2) bound PDK1 (PDB ID: 3ORZ). (c) Inhibitor (1F8) bound PDK1 (PDB ID: 3ORX). Note that Lys76 does not appear in the PDB structure 3OTU.

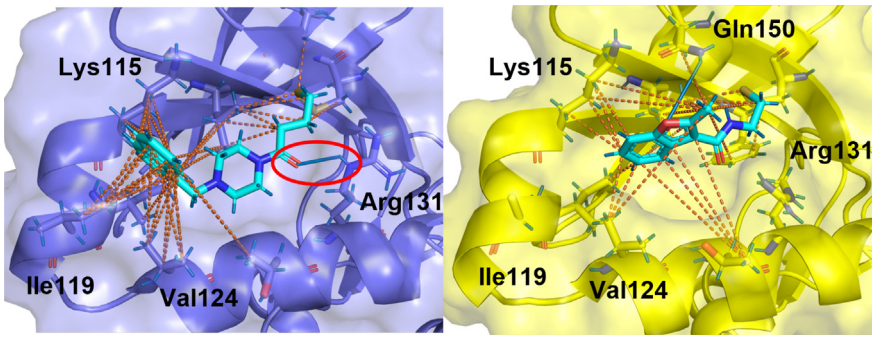



**Figure 5. Details on ligands–PIF pocket interactions.** (a) Non-covalent interactions formed between the strong allosteric activator J30 with the PIF pocket. The orange dotted lines are hydrophobic interactions identified with BagPype.[67] (b) Non-covalent interactions formed between the allosteric inhibitor 1F8 with the PIF pocket. Lys115, Val124 and Gln150 are for positioning of the ligand and no hydrogen bond is observed with Arg131.

PDK1 bound with different allosteric modulators (mostly activators). These include the 4 structures analysed above in Figs. 3 and 4 together with 7 additional experimentally resolved structures. The average POP scores from allosteric residues to Sep241 for all these structures are summarised in Table 6. We found that the average POP score from the allosteric site to Sep241 is highest (0.97) for 1F8, the only inhibitor in the set, whereas all the activators have scores lower than 0.86. Within the activators, we found that the POP scores are ordered as follows:

- Table 6: PS210 > RS2 > J30 > PIFtide > RS1 ≫ PS182 > PS171 > PS114 > PS48 ≫ 2A2

which compare well with experimental results measuring the activating strength of the ligands:

- Ref.[65]: PS210 > RS2 > PIFtide > RS1
- Ref.[68]: PS210 > PS182 > PS48
- Ref.[69]: J30 > 2A2
- Ref.[70]: PS114 > PS171

The activating strengths of the allosteric activators are all correctly ranked based on the average path scores, except for PS114 and PS171. However, we note that PS114 is not completely resolved in the PDB structure (4A06), which could explain the mismatch with experimental data.

Furthermore, the detailed structural information of the POPs reveals differences between the inhibitor and the activators. In particular, the inhibitor (3ORX) has high POP scores in two specific pathways (i.e., those connecting Val127 and Leu155 to Sep241), whereas the allosteric activators have low POP scores for those pathways. This and other distinct patterns (see Figure S4) can provide further information to potentially distinguish the relevant communication pathways for inhibitor and activators. Although a fuller comparison is not possible due to the lack of available experimental data, our results show promise to link the structural patterns and magnitude of POP scores with the type and strength of the allosteric modulation. For details of the POPs scores of the PDK1 set, see Supplementary Information Part 4 and Figure S4.

## Discussion

Using protein structural data and applying bond-to-bond propensity analysis to atomistic protein

Table 6 **Details of 11 PDK1 allosteric modulators.** The modulators are ordered according to the average path scores from the allosteric site residues to Sep241 for 11 allosteric ligands bound to PDK1 structures.

| PDB ID | Allosteric ligand | Ligand type | Average path score to Sep241 |
|---|---|---|---|
| 3ORX | 1F8 | Inhibitor | 0.97 |
| 4AW1 | PS210 | Activator | 0.86 |
| 4RQV | RS2 | Activator | 0.85 |
| 3OTU | J30 | Activator | 0.84 |
| 4RRV | PIFtide | Activator | 0.83 |
| 4RQK | RS1 | Activator | 0.81 |
| 4AW0 | PS182 | Activator | 0.77 |
| 4A07 | PS171 | Activator | 0.75 |
| 4A06 | PS114 | Activator | 0.73 |
| 3HRF | PS48 | Activator | 0.71 |
| 3ORZ | 2A2 | Activator | 0.60 |

graphs, we are able to quantify how strongly coupled a bond is to the selected source through a propensity quantile score. Adding bonds of increasing propensity quantile score to a simplified residue graph, we compute and score propensity optimised paths between the residues of the orthosteric and allosteric sites. The path scores allow us to identify important residues at both ends of the path. We can also find key intermediate (signalling) residues along the path using the residue participation frequency in the paths as well as measures to evaluate the impact of residue removal in path score and length.

We have demonstrated this computational approach with three allosteric proteins: h-Ras, caspase-1 and PDK1. Analysis of h-Ras correctly finds the key catalytic residue Gln61, and identifies Mg168, HOH367 and Tyr96 as the crucial signalling residues based on POP participation frequencies and z-scores of residue removal. These findings are consistent with a hydrogen bond network proposed in the literature.[53] Regarding caspase-1, focusing on the covalently bound Cys285 as the starting point of allosteric signalling pathways, we find that the highest path scores are to Arg286 and Glu390, two functionally important allosteric residues, and subsequent pathway analysis on Arg286Ala and Glu390Ala mutants demonstrates the loss of connection between the orthosteric site and Glu390 upon mutations. The analysis of high scoring POPs highlights the crucial signalling residues Asn337 and Ser332 which, together with Arg286 and Glu390, are in agreement with the hydrogen bond network tested experimentally. Lastly, in PDK1, we find the correct allosteric signalling pathway, Arg131–Glu130–Lys111, and suggest that interaction with Arg131 as the essential element for allosteric activation, which also allows to distinguish activators and inhibitors. We also ranked allosteric activators of PDK1 by scoring the paths from the allosteric site to the key activating residue Sep241, and found good agreement with experimental results. In summary, the analysis introduced here constructs short paths of high propensity based on bond-to-bond propensities to identify functionally important orthosteric, allosteric and signalling residues, and was applied to distinguish allosteric activators from inhibitors and to rank allosteric modulators.

Further ongoing work includes the robust computation of multiple pathways to allow for a probabilistic evaluation of paths between 'source' and 'target'. Beyond strict optimality, the distribution of scores of an ensemble of optimised paths would provide complementary insight into the robustness of the allosteric signalling between the residue pairs. In this respect, concepts such as effective distance,[71] residue betweenness and centrality[50] could be interesting for the optimisation and scoring of multiple pathways. We also note that bond-to-bond propensity analysis itself has been shown to be able to predict allosteric sites. Hence, combining both analyses, we can first predict the allosteric site, rank the sites based on pathway scores and then identify the key binding residues. The key residues could then act as the starting point for the design of allosteric drug molecules that specifically interact with the residues highlighted. Since the residues would be distributed within the predicted allosteric site, small fragments bound to such residues can be selected first and pieced together to form the final drug molecule, along the lines of fragment-based drug design.[72] Moreover, our approach could be combined with ligand-site interaction data to guide effective molecule design for a given target on a protein with the aid of supervised machine learning. Through validation and testing against data on pathways and functional residues from different allosteric proteins, the approach could be further optimised and developed with the aim to guide the analysis of unknown allosteric mechanisms for proteins and the identification of new structural targets.

## Materials and Methods

### Protein structure and sites definition

The structures of caspase-1, h-Ras and PDK1 in PDB format shown in Table 7 were retrieved from the RCSB Protein Data Bank (PDB),[18] and further processed to remove irrelevant solvent molecules for crystallisation and ions such as chloride ions. The orthosteric and allosteric sites are defined as the interacting residues with the corresponding ligands based on protein graphs constructed in Section *Construction of the atomistic protein graph*.

### Construction of the atomistic protein graph

An energy-weighted atomistic graph is constructed from the 3-dimensional coordinates of the atoms of a protein. The atoms are represented as nodes and the bonds and interactions (covalent and non-covalent) as edges. Edges are weighted based on the interaction energies calculated with relevant interatomic potentials: covalent bonds by bond-dissociation energies;[75] hydrogen bonds and salt bridges by the modified Mayo potential;[76,77] hydrophobic interactions by a hydrophobic potential of mean force;[78] electrostatic interactions by the OPLS (optimized potentials for liquid simulations) potential functions.[50] Details can be found in Refs.[38,41,42]

To construct the atomistic protein graphs, we used the software package Biochemical atomistic graph construction in Python for proteins (BagPype).[67] The process starts from the cartesian coordinates of atoms from a PDB file. Missing hydrogen atoms are added using Reduce.[79] Covalent bonds are identified using standard covalent bond length cutoffs. All weak interactions are

Table 7 Details of protein structures used.

| Protein | Description | PDB ID |
|---|---|---|
| Caspase-1 | Wild type | 2HBQ[73] |
| | Arg286Ala | 2HBR[73] |
| | Glu390Ala | 2HBY[73] |
| H-Ras | Wild type, with calcium acetate | 3K8Y[53] |
| | Wild type, without calcium acetate | 3LBN[53] |
| PDK1 | Allosteric activator | 3HRF,[74] 3ORZ, 3OTU,[69] 4A06, 4A07,[70] 4AW0, 4AW1,[68] 4RQK, 4RQV, 4RRV[65] |
| | Allosteric inhibitor | 3ORX[69] |

detected and weighted based on the above-mentioned potentials with a 9 Å distance cutoff for hydrophobic interactions, and a 0.01 kcal/mol energy cutoff for hydrogen bonds. The constructed energy-weighted atomistic graph is shown in Figure 6.

## Bond-to-bond propensity analysis

Bond-to-bond propensity analysis was introduced and extended in Refs.[38,80]. The analysis quantifies how a perturbation at the bonds of the source is transmitted to all other bonds of the protein via a graph-theoretical measure defined in edge-space. Consider a graph with $n$ nodes and $m$ edges. The $m \times m$ edge-to-edge transfer matrix $M$ was first introduced in Ref.[81] to examine nonlocal edge-coupling in graphs, and was then utilised to analyse atomistic protein graphs in Ref.[38]. $M$ can be interpreted as a Green function (see Refs.[81,38]), such that the effect of a perturbation at edge $i$ on edge $j$ is described by the element $M_{ji}$ of the matrix

$$M = \frac{1}{2} W B^T L^{\dagger} B. \tag{1}$$

Here $B$ is the standard $n \times m$ incidence matrix linking the $n$ nodes and $m$ edges; $W = \mathrm{diag}(w_{ij})$ is an $m$ Œ$m$ diagonal matrix containing all the edge interaction energies; and the $n \times n$ matrix $L^{\dagger}$ is the pseudo-inverse of the energy-weighted Laplacian matrix $L$, which governs the diffusion dynamics on the graph.[82,83]

We can then define the bond propensity, which is the effect of a perturbation from bonds $b'$ (within the source), belonging to the source, on bond $b$ of the protein:

$$\Pi_b^{\mathrm{raw}} = \sum_{b' \in \mathrm{source}} |M_{bb'}| \tag{2}$$

The raw propensity is normalised by the sum of all the bond propensities of the protein:

$$\Pi_b = \frac{\Pi_b^{\mathrm{raw}}}{\sum_b \Pi_b^{\mathrm{raw}}} \tag{3}$$

For full details of the derivations, see Ref.[38].

## Quantile regression (QR) and computation of the quantile scores

The propensity of a bond decreases naturally as the distance of the bond to the source increases. Hence, we aim to detect bonds with significantly large bond-to-bond propensity relative to other bonds at a similar distance from the source. To achieve this, we apply conditional quantile regression (QR)[84] with distance as the independent variable. Specifically, let $d$ be the minimum Euclidean distance from a point in space to any of the bonds of the source residues, calculated based on the midpoint of the bonds. Given that the bond-to-bond propensities, $\Pi_b$, decay exponentially with the distance from the source, the following QR minimisation problem is solved for a given quantile $p$ to obtain the estimated parameters of the model:

$$\left(\hat{\beta}_0^{\mathrm{prot}}(p), \hat{\beta}_1^{\mathrm{prot}}(p)\right) = \underset{(\beta_0,\beta_1)}{\arg\min} \sum_{b}^{\mathrm{protein}} \rho_p(\log(\Pi_b) - (\beta_0 + \beta_1 d)), \tag{4}$$

where $\rho_p(\cdot)$ is the tilted absolute value function:

$$\rho_p(y) = |y(p - \mathbb{1}(y < 0))|, \tag{5}$$

and $\mathbb{1}(\cdot)$ is the indicator function.

Let $d_b$ be the distance between bond $b$ and the source, i.e., the minimum Euclidean distance between $b$ and any of the bonds of the source residues. From the estimated model for the protein (4), the quantile score (QS) of bond $b$ with propensity $\Pi_b$ at distance $d_b$ is quantile $p_b$ given by:

$$p_b = \underset{p \in [0,1]}{\arg\min} \left| \log(\Pi_b) - \left(\hat{\beta}_0^{\mathrm{prot}}(p) + \hat{\beta}_1^{\mathrm{prot}}(p)\, d_b\right)\right| \tag{6}$$

Clearly, the QS of all the bonds in the protein is in the range between 0 and 1. The higher the QS of a bond, the stronger the bond couples to the source bonds. For more details see Ref.[38].

## Computation of propensity optimised paths between residue pairs

Based on the energy-weighted atomistic protein graph described above and the QS (6) obtained through bond-to-bond propensity analysis for all weak interactions, the graph is simplified to the residue level by considering each residue as a node and creating weighted edges between

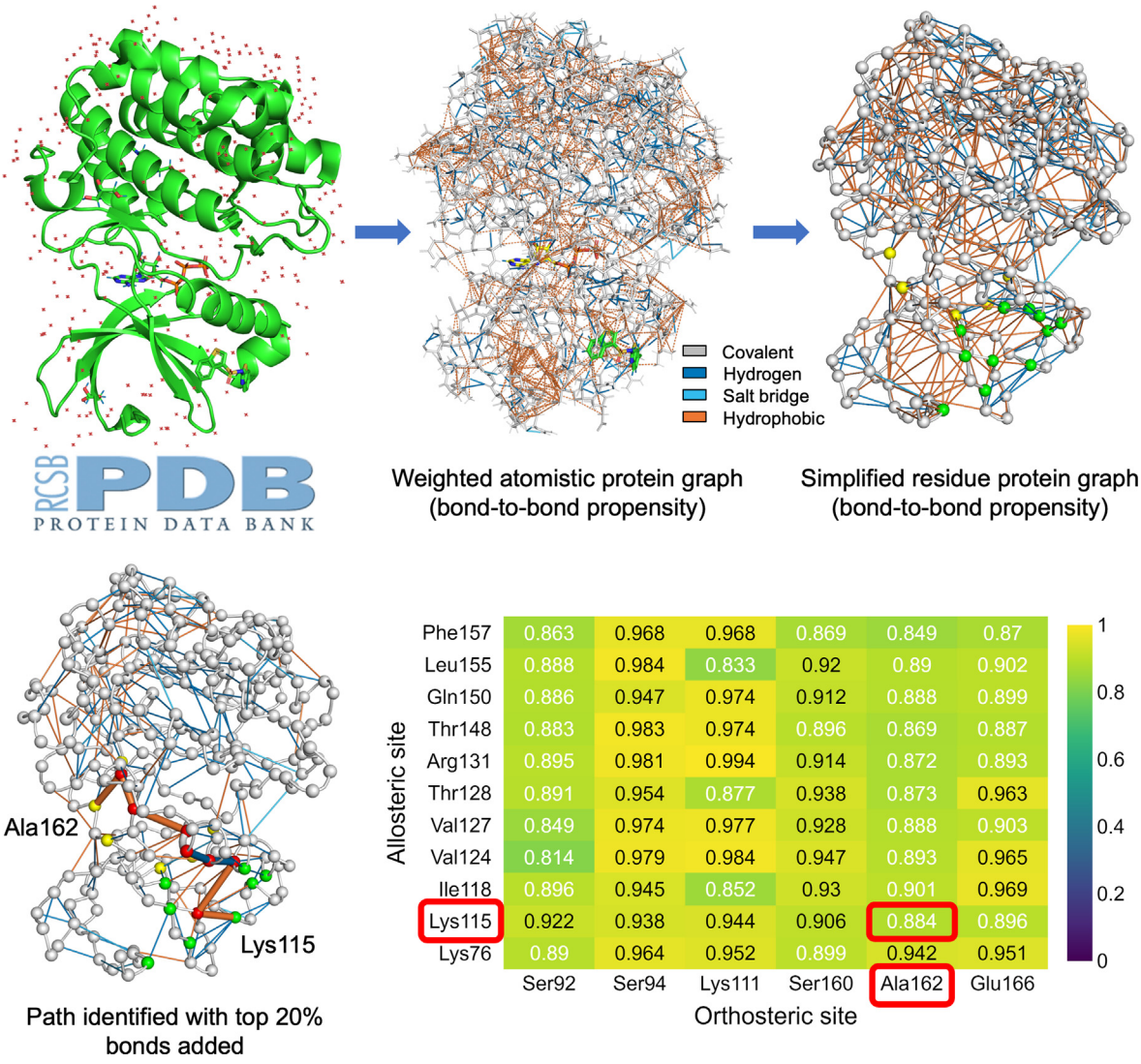



**Figure 6. Workflow of pathway computation using PDK1 (PDK ID: 4RQK)[65] as an example.** The X-ray crystallographic protein structure in PDB format is retrieved and processed as described in Section *Protein structure and sites definition*. An energy-weighted atomistic protein graph is constructed with weak bonds weighted applying bond-to-bond propensity analysis. The atomistic protein graph is then simplified to the residue level to keep the highest scoring bonds between any residue pair. The orthosteric site residues are coloured in yellow and allosteric site residues in green. Five percent of the bonds are added at a time starting from bonds with QS = 1.0 until a path meeting the criteria is identified. For instance, 20% of the bonds were added in order to find a path between Ala162 and Lys115. The computed path was highlighted in bold with the intermediate residues in red. The paths of each residue pair between the orthosteric and allosteric sites were computed and summarised in a heatmap.

residues by selecting the bond with the highest QS between any residue pair and using the QS as its weight. The backbone of the protein is maintained by adding an edge between two neighbouring residues. To compute the pathway between two residues, edges are added in order of decreasing QS. Starting from the top 5%, we attempt to find the shortest path between any given residue pair with the constraint of allowing no more than one consecutive step along the protein backbone. This allows us to restrain the path through the weak interactions within the protein and still take adjacent residues into consideration. A suitable biophysical analogy for this process is the transmission of allosteric signals as anisotropic thermal diffusion, for which heat flow is mainly via non-covalent interactions since only the surrounding atoms are heated, hence including the neighbouring residues.[51] If no path is found based on this criterion, the next 5% of the edges (of decreasing QS) is added, and we repeat the protocol until a path is found. The path is then scored

by taking the geometric mean of the QS of all weak interactions $b$ in the path, which we term the propensity optimised path (POP) score:

$$\Sigma_{\mathrm{POP}} = \sqrt[k]{p_{b_1} p_{b_2} \cdots p_{b_k}}, \tag{7}$$

where the POP has $k$ weak interactions $b_i,\ i = 1, \ldots, k$. If there are multiple paths with the same length, the one with the highest path score is selected as the POP for the residue pair. Note that there is no QS associated with a backbone hop; hence backbone hops are thus ignored in the calculation of the path score (7).

Figure 6 illustrates the full workflow of computing the pathways starting from the initial structural data.

### Determination of key signalling residues

To understand which residues are crucial for allosteric signalling in the paths, we employ two measures.

*Residues with high participation in propensity optimised paths:* The most direct measure is to count the number of appearances of each residue in all the orthosteric-allosteric POPs (i.e, participation residue frequencies). The residues that appear more often are considered important for allosteric signalling.

*Residues with high influence in pathway scores and length upon removal:* Another measure, adapted from del Sol *et al.*,[85] is the change in score and length of the propensity optimised paths under residue removal. This workflow is repeated by removing the weak interactions of each residue in the protein and computing the change in both the average path length and score of all the computed propensity optimised paths between the orthosteric and allosteric sites:

$$\Delta X_r = \left| X - X_{r,rem} \right| \tag{8}$$

where $X$ is the original average path length (or score) and $X_{r,rem}$ is the average path length (or score) when weak interactions involving residue $r$ are removed. Finally, a z-score is used to evaluate the importance of residue $r$.

$$z_r = \frac{\Delta X_r - \overline{\Delta X}}{\sigma} \tag{9}$$

## Data Availability

All data presented in this study are available at figshare with DOI: 10.6084/m9.figshare.20182166.

## Author Contributions

N.W., M.B. and S.N.Y. conceived the study. N.W. performed the computations and created the figures. All authors analysed the data and wrote the manuscript.

## Materials & Correspondence

Requests for data and code to m.barahona@imperial.ac.uk.

## CRediT authorship contribution statement

**Nan Wu:** Conceptualisation, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – Original draft preparation, Writing – Reviewing and Editing, Visualisation. **Sophia N. Yaliraki:** Conceptualisation, Methodology, Writing – Reviewing and Editing. **Mauricio Barahona:** Methodology, Writing – Reviewing and Editing, Supervision.

### DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

We acknowledge helpful discussions with Florian Song and Léonie Strömich. This work was funded by a President's PhD Scholarship, Imperial College London, to N.W. Funding from EPSRC award EP/N014529/1 supporting the EPSRC Centre for Mathematics of Precision Healthcare at Imperial is gratefully acknowledged by S.N.Y.and M.B.

## Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jmb.2022.167749.

*Received 12 April 2022;*
*Accepted 11 July 2022;*
*Available online 14 July 2022*

***Keywords***:
graph theory;
atomistic graph representation;
allosteric signalling pathway

1 z-VAD-FMK corresponds to the inhibitor carbobenzoxy-valyl-alanyl-aspartyl-[O-methyl]- fluoromethylketone.

## References

1. Monod, J., Jacob, F., (1961). General conclusions: Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 389–401. URL http://symposium.cshlp.org/content/26/389.short..

2. Cui, Q., Karplus, M., (2008). Allostery and cooperativity revisited. Protein science: a publication of the. *Protein Society* **17**, 1295–1307. URL https://www.ncbi.nlm.nih.gov/pubmed/18560010.
3. He, X., Duan, N., Lu, S. & Zhang, J. (2019). Characteristics of allosteric proteins, sites, and modulators.
4. Volkman, B.F., Lipson, D., Wemmer, D.E., Kern, D., (2001). Two-state allosteric behavior in a single-domain signaling protein. *Science* **291**, 2429 LP–2433. http://science.sciencemag.org/content/291/5512/2429.abstract.
5. Henzler-Wildman, K., Kern, D., (2007). Dynamic personalities of proteins. *Nature* **450**, 964. https://doi.org/10.1038/nature06522. URL http://10.0.4.14/nature06522.
6. Gunasekaran, K., Ma, B., Nussinov, R., (2004). Is allostery an intrinsic property of all dynamic proteins? Proteins: Structure. *Function, Bioinformat.* **57**, 433–443. https://doi.org/10.1002/prot.20232.
7. Kern, D., Zuiderweg, E.R.P., (2003). The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **13**, 748–757. URL https://www.sciencedirect.com/science/article/pii/S0959440X03001660.
8. Peracchi, A., Mozzarelli, A., (2011). Exploring and exploiting allostery: Models, evolution, and drug targeting. *Biochim. Biophys. Acta (BBA) - Proteins Proteomics* **1814**, 922–933. http://www.sciencedirect.com/science/article/pii/S1570963910002827.
9. Nussinov, R., Tsai, C.-J., (2013). Allostery in disease and in drug discovery. *Cell* **153**, 293–305. https://doi.org/10.1016/j.cell.2013.03.034.
10. Christopoulos, A., May, L.T., Avlani, V.A., Sexton, P.M., (2004). G-protein-coupled receptor allosterism: the promise and the problem(s). *Biochem. Soc. Trans.* **32**, 873–877. https://doi.org/10.1042/BST0320873.
11. Wodak, S.J. et al, (2019). Allostery in its many disguises: From theory to applications. *Structure* **27**, 566–578. URL https://www.sciencedirect.com/science/article/pii/S0969212619300036.
12. Hardy, J.A., Wells, J.A., (2004). Searching for new allosteric sites in enzymes. *Curr. Opin. Struct. Biol.* **14**, 706–715. URL https://www.sciencedirect.com/science/article/pii/S0959440X0400185X.
13. Erlanson, D.A., Wells, J.A., Braisted, A.C., (2004). Tethering: Fragment-based drug discovery. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 199–223. https://doi.org/10.1146/annurev.biophys.33.110502.140409.
14. Selvaratnam, R., Chowdhury, S., VanSchouwen, B., Melacini, G., (2011). Mapping allostery through the covariance analysis of nmr chemical shifts. *Proc. Nat. Acad. Sci.* **108**, 6133 LP–6138. URL http://www.pnas.org/content/108/15/6133.abstract.
15. Oyen, D., Wechselberger, R., Srinivasan, V., Steyaert, J., Barlow, J.N., (2013). Mechanistic analysis of allosteric and non-allosteric effects arising from nanobody binding to two epitopes of the dihydrofolate reductase of escherichia coli. *Biochim. Biophys. Acta (BBA) - Proteins Proteom.* **1834**, 2147–2157. http://www.sciencedirect.com/science/article/pii/S1570963913002823.
16. Rath, V.L. et al, (2000). Human liver glycogen phosphorylase inhibitors bind at a new allosteric site. *Chem. Biol.* **7**, 677–682. URL http://www.sciencedirect.com/science/article/pii/S1074552100000041.
17. Wright, S.W. et al, (2002). Anilinoquinazoline inhibitors of fructose 1,6-bisphosphatase bind at a novel allosteric site: synthesis, in vitro characterization, and x-ray crystallography. *J. Med. Chem.* **45**, 3865–3877. URL https://doi.org/10.1021/jm010496a.
18. Burley, S.K. et al, (2021). Rcsb protein data bank: powerful new tools for exploring 3d structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **49**, D437–D451. https://doi.org/10.1093/nar/gkaa1038.
19. Liu, X. et al, (2020). Unraveling allosteric landscapes of allosterome with asd. *Nucleic Acids Res.* **48**, D394–D401. https://doi.org/10.1093/nar/gkz958.
20. Huang, W. et al, (2015). Asbench: benchmarking sets for allosteric discovery. *Bioinformatics* **31**, 2598–2600. https://doi.org/10.1093/bioinformatics/btv169.
21. Zlobin, A., Suplatov, D., Kopylov, K., Švedas, V., (2019). Casbench: A benchmarking set of proteins with annotated catalytic and allosteric sites in their structures. *Acta Naturae* **11**, 74–80. URL https://pubmed.ncbi.nlm.nih.gov/31024751.
22. Lu, S., He, X., Ni, D., Zhang, J., (2019). Allosteric modulator discovery: From serendipity to structure-based design. *J. Med. Chem.* **62**, 6405–6421. https://doi.org/10.1021/acs.jmedchem.8b01749.
23. Greener, J.G., Sternberg, M.J.E., (2018). Structure-based prediction of protein allostery. *Curr. Opin. Struct. Biol.* **50**, 1–8. URL http://www.sciencedirect.com/science/article/pii/S0959440X17301252.
24. Ghosh, A., Vishveshwara, S., (2007). A study of communication pathways in methionyl- trna synthetase by molecular dynamics simulations and structure network analysis. *Proc. Nat. Acad. Sci. USA* **104**, 15711–15716. https://pubmed.ncbi.nlm.nih.gov/17898174.
25. Shukla, D., Meng, Y., Roux, B., Pande, V.S., (2014). Activation pathway of src kinase reveals intermediate states as targets for drug design. *Nature Commun.* **5**, 3397. https://doi.org/10.1038/ncomms4397.
26. van Gunsteren, W.F. et al, (2006). Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed.* **45**, 4064–4092. https://doi.org/10.1002/anie.200502655.
27. Panjkovich, A., Daura, X., (2012). Exploiting protein flexibility to predict the location of allosteric sites. *BMC Bioinformat.* **13**, 273. https://doi.org/10.1186/1471-2105-13-273.
28. Panjkovich, A., Daura, X., (2014). Pars: a web server for the prediction of protein allosteric and regulatory sites. *Bioinformatics* **30**, 1314–1315. https://doi.org/10.1093/bioinformatics/btu002.
29. Greener, J.G., Sternberg, M.J.E., (2015). Allopred: prediction of allosteric pockets on proteins using normal mode perturbation analysis. *BMC Bioinformat.* **16**, 335. https://doi.org/10.1186/s12859-015-0771-1.
30. Song, K. et al, (2017). Improved method for the identification and validation of allosteric sites. *J. Chem. Inf. Model.* **57**, 2358–2363. https://doi.org/10.1021/acs.jcim.7b00014.
31. Chen, A.S.-Y. et al, (2016). A random forest model for predicting allosteric and functional sites on proteins. *Mol. Informat.* **35**, 125–135. https://doi.org/10.1002/minf.201500108.

32. Huang, W. et al, (2013). Allosite: a method for predicting allosteric sites. *Bioinformatics* **29**, 2357–2359. https://doi.org/10.1093/bioinformatics/btt399.
33. Hollingsworth, S.A., Dror, R.O., (2018). Molecular dynamics simulation for all. *Neuron* **99**, 1129–1143. URL http://www.sciencedirect.com/science/article/pii/S0896627318306846.
34. Collier, G., Ortiz, V., (2013). Emerging computational approaches for the study of protein allostery. *Arch. Biochem. Biophys.* **538**, 6–15. URL http://www.sciencedirect.com/science/article/pii/S0003986113002324.
35. Tee, W.-V., Guarnera, E., Berezovsky, I.N., (2018). Reversing allosteric communication: From detecting allosteric sites to inducing and tuning targeted allosteric response. *PLOS Comput. Biol.* **14**, e1006228. https://doi.org/10.1371/journal.pcbi.1006228.
36. Fogha, J., Diharce, J., Obled, A., Aci-Sèche, S., Bonnet, P., (2020). Computational analysis of crystallization additives for the identification of new allosteric sites. *ACS Omega* **5**, 2114–2122. https://doi.org/10.1021/acsomega.9b02697.
37. Wang, J. et al, (2020). Mapping allosteric communications within individual proteins. *Nature Commun.* **11**, 3862. https://doi.org/10.1038/s41467-020-17618-2.
38. Amor, B.R., Schaub, M.T., Yaliraki, S.N., Barahona, M., (2016). Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. *Nature Commun.* **7**, 1–13. https://doi.org/10.1038/ncomms12477.
39. Mersmann, S. et al, (2021). Proteinlens: a web-based application for the analysis of allosteric signalling on atomistic graphs of biomolecules. *Nucleic Acids Res.*. https://doi.org/10.1093/nar/gkab350.
40. Wu, N., Strömich, L., Yaliraki, S.N., (2021). Prediction of allosteric sites and signaling: Insights from benchmarking datasets. *Patterns*. https://doi.org/10.1016/j.patter.2021.100408.
41. Delmotte, A., Tate, E.W., Yaliraki, S.N., Barahona, M., (2011). Protein multi-scale organization through graph partitioning and robustness analysis: application to the myosin–myosin light chain interaction. *Phys. Biol.* **8**, 55010. https://doi.org/10.1088/1478-3975/8/5/055010.
42. Amor, B., Yaliraki, S.N., Woscholski, R., Barahona, M., (2014). Uncovering allosteric pathways in caspase-1 using markov transient analysis and multiscale community detection. *Mol. BioSyst.* **10**, 2247–2258. https://doi.org/10.1039/C4MB00088A.
43. Kelner, J. A., Orecchia, L., Sidford, A. & Zhu, Z. A. (2013). A simple, combinatorial algorithm for solving sdd systems in nearly-linear time. 911–920 (Association for Computing Machinery, 2013). URL: https://doi.org/10.1145/2488608.2488724.
44. Spielman, D. A. & Teng, S.-H. (2004). Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. 81–90 (Association for Computing Machinery, 2004). URL: https://doi.org/10.1145/1007352.1007372.
45. del Sol, A., Tsai, C.-J., Ma, B. & Nussinov, R. (2009). The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure* (London, England: 1993) 17, 1042–1050. https://www.ncbi.nlm.nih.gov/pubmed/196790 84.
46. Dokholyan, N.V., (2016). Controlling allosteric networks in proteins. *Chem. Rev.* **116**, 6463–6487. https://doi.org/10.1021/acs.chemrev.5b00544.
47. East, K.W. et al, (2020). Nmr and computational methods for molecular resolution of allosteric pathways in enzyme complexes. *Biophys. Rev.* **12**, 155–174. URL https://pubmed.ncbi.nlm.nih.gov/31838649.
48. Demerdash, O.N.A., Daily, M.D., Mitchell, J.C., (2009). Structure-based predictive models for allosteric hot spots. *PLOS Comput. Biol.* **5**, e1000531. https://doi.org/10.1371/journal.pcbi.1000531.
49. Kaya, C., Armutlulu, A., Ekesan, S., Haliloglu, T., (2013). Mcpath: Monte carlo path generation approach to predict likely allosteric pathways and functional residues. *Nucleic Acids Res.* **41**, W249–W255. https://doi.org/10.1093/nar/gkt284.
50. Botello-Smith, W.M., Luo, Y., (2019). Robust determination of protein allosteric signaling pathways. *J. Chem. Theory Comput.* **15**, 2116–2126. https://doi.org/10.1021/acs.jctc.8b01197. URL https://doi.org/10.1021/acs.jctc.8b01197.
51. Ota, N., Agard, D.A., (2005). Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *J. Mol. Biol.* **351**, 345–354. URL https://www.sciencedirect.com/science/article/pii/S002228360500598X.
52. McCormick, F., (1995). Ras-related proteins in signal transduction and growth control. *Mol. Reprod. Dev.* **42**, 500–506. https://doi.org/10.1002/mrd.1080420419.
53. Buhrman, G., Holzapfel, G., Fetics, S., Mattos, C., (2010). Allosteric modulation of ras positions q61 for a direct role in catalysis. *Proc. Nat. Acad. Sci.* **107**, 4931–4936.
54. Huang, L., Hofer, F., Martin, G.S., Kim, S.-H., (1998). Structural basis for the interaction of ras with raigds. *Nature Struct. Biol.* **5**, 422–426. https://doi.org/10.1038/nsb0698-422.
55. Tan, Z.W., Guarnera, E., Tee, W.-V., Berezovsky, I.N., (2020). AlloSigMA 2: paving the way to designing allosteric effectors and to exploring allosteric effects of mutations. *Nucleic Acids Res.* **48**, W116–W124. https://doi.org/10.1093/nar/gkaa338.
56. Peach, R. (2018). Exploring protein dynamics using graph theory and single-molecule spectroscopy. http://hdl.handle.net/10044/1/68684.
57. Wilson, K.P. et al, (1994). Structure and mechanism of interleukin-1 beta converting enzyme. *Nature* **370**, 270–275. https://doi.org/10.1038/370270a0.
58. Datta, D., Scheer, J.M., Romanowski, M.J., Wells, J.A., (2008). An allosteric circuit in caspase-1. *J. Mol. Biol.* **381**, 1157–1167. URL https://www.sciencedirect.com/science/article/pii/S0022283608007420.
59. Mora, A., Komander, D., van Aalten, D.M., Alessi, D.R., (2004). Pdk1, the master regulator of agc kinase signal transduction. *Seminars Cell Develop. Biol.* **15**, 161–170. URL https://www.sciencedirect.com/science/article/pii/S1084952103001277.
60. Raimondi, C., Falasca, M., (2011). Targeting pdk1 in cancer. *Current Med. Chem.* **18**, 2763–2769.
61. Hennessy, B.T., Smith, D.L., Ram, P.T., Lu, Y., Mills, G.B., (2005). Exploiting the PI3K/AKT Pathway for Cancer Drug Discovery. *Nat. Rev. Drug Discovery* **4**, 988–1004. https://doi.org/10.1038/nrd1902.
62. Schulze, J. et al, (2016). Bidirectional Allosteric Communication between the ATP-Binding Site and the Regulatory PIF Pocket in PDK1 Protein Kinase. *Cell Chem.*

*Biol.* **23**, 1193–1205. https://doi.org/10.1016/j.chembiol.2016.06.017.
63. Bogoyevitch, M.A., Fairlie, D., (2007). P.A new paradigm for protein kinase inhibition: blocking phosphorylation without directly targeting atp binding. *Drug Discovery Today* **12**, 622–633. URL https://www.sciencedirect.com/science/article/pii/S1359644607002553.
64. Engel, M. et al, (2006). Allosteric activation of the protein kinase pdk1 with low molecular weight compounds. *EMBO J.* **25**, 5469–5480. URL https://www.embopress.org/doi/abs/10.1038/sj.emboj.7601416.
65. Rettenmaier, T.J. et al, (2014). A small-molecule mimic of a peptide docking motif inhibits the protein kinase pdk1. *Proc. Nat. Acad. Sci.* **111**, 18590 LP–18595. URL http://www.pnas.org/content/111/52/18590.abstract.
66. Bobkova, E.V. et al, (2010). Discovery of PDK1 Kinase Inhibitors with a Novel Mechanism of Action by Ultrahigh Throughput Screening. *J. Biol. Chem.* **285**, 18838–18846. https://doi.org/10.1074/jbc.M109.089946.
67. Song, F., Barahona, M. & Yaliraki, S. N. (2022). BagPype: A Python package for the construction of atomistic, energy-weighted graphs from biomolecular structures. https://doi.org/10.5281/zenodo.6326081.
68. Busschots, K. et al, (2012). Substrate-selective inhibition of protein kinase pdk1 by small compounds that bind to the pif-pocket allosteric docking site. *Chem. Biol.* **19**, 1152–1163. URL https://www.sciencedirect.com/science/article/pii/S1074552112002700.
69. Sadowsky, J.D. et al, (2011). Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proc. Nat. Acad. Sci.* **108**, 6056 LP–6061. URL http://www.pnas.org/content/108/15/6056.abstract.
70. Lopez-Garcia, L. et al, (2011). Allosteric regulation of protein kinase pkc by the n-terminal c1 domain and small compounds to the pif-pocket. *Chem. Biol.* **18**, 1463–1473. URL https://www.sciencedirect.com/science/article/pii/S107455211100319X.
71. Brockmann, D., Helbing, D., (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342. URL https://www.science.org/doi/abs/10.1126/science.1245200.
72. de Souza Neto, L.R. et al, (2020). In silico strategies to support fragment-to-lead optimization in drug discovery. *Front. Chem.* **8**, 93. URL https://www.frontiersin.org/article/10.3389/fchem.2020.00093.
73. Scheer, J.M., Romanowski, M.J., Wells, J.A., (2006). A common allosteric site and mechanism in caspases. *Proc. Nat. Acad. Sci.* **103**, 7595 LP–7600. URL http://www.pnas.org/content/103/20/7595.abstract.
74. Hindie, V. et al, (2009). Structure and allosteric effects of low-molecular-weight activators on the protein kinase pdk1. *Nat. Chem. Biol.* **5**, 758–764. https://doi.org/10.1038/nchembio.208.
75. Huheey, J.E., Keiter, E.A., Keiter, R.L., (1993). Inorganic chemistry: principles of structure and reactivity. HarperCollins College Publishers.
76. Dahiyat, B.I., Gordon, D.B., Mayo, S.L., (1997). Automated design of the surface positions of protein helices. *Protein Sci.* **6**, 1333–1337. https://doi.org/10.1002/pro.5560060622.
77. Mayo, S. L., Olafson, B. D. & III, W.A. G. (1990). Dreiding: A generic force field for molecular simulations. *J. Phys. Chem.; (USA)* 94.
78. Lin, M.S., Fawzi, N.L., Head-Gordon, T., (2007). Hydrophobic potential of mean force as a solvation function for protein structure prediction. *Structure* **15**, 727–740. https://doi.org/10.1016/j.str.2007.05.004.
79. Word, J., Lovell, S.C., Richardson, J.S., Richardson, D.C., (1999). Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation11 edited by j. thornton. *J. Mol. Biol.* **285**, 1735–1747. URL http://www.sciencedirect.com/science/article/pii/S0022283698924019.
80. Hodges, M., Barahona, M., Yaliraki, S.N., (2018). Allostery and cooperativity in multimeric proteins: bond-to-bond propensities in atcase. *Sci. Reports* **8**, 1–14. https://doi.org/10.1038/s41598-018-27992-z.
81. Schaub, M.T., Lehmann, J., Yaliraki, S.N., Barahona, M., (2014). Structure of complex networks: Quantifying edge-to-edge relations by failure-induced flow redistribution. *Network Sci.* **2**, 66–89. https://doi.org/10.1017/nws.2014.4.
82. Biggs, N., (1974). Algebraic Graph Theory. Cambridge Mathematical Library. Cambridge University Press.
83. Lambiotte, R., Delvenne, J.-C., Barahona, M., (2014). Random walks, markov processes and the multiscale modular organization of complex networks. *IEEE Trans. Network Sci. Eng.* **1**, 76–90.
84. Koenker, R., Hallock, K.F., (2001). Quantile regression. *J. Econ. Perspect.* **15**, 143–156. https://doi.org/10.1257/jep.15.4.143.
85. del Sol, A., Fujihashi, H., Amoros, D., Nussinov, R., (2006). Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol.* **2** 2006.0019–2006.0019, URL https://pubmed.ncbi.nlm.nih.gov/16738564.