

Article

Site-Specific Phosphorylation of RTK KIT Kinase Insert Domain: Interactome Landscape Perspectives

Julie Ledoux  and Luba Tchertanov * 

Centre Borelli, CNRS, ENS Paris-Saclay, Université Paris-Saclay, 4 Avenue Des Sciences,
F-91190 Gif-sur-Yvette, France

* Correspondence: luba.tchertanov@ens-paris-saclay.fr

Abstract: The kinase insert domain (KID) of RTK KIT is a key recruitment region for downstream signalling proteins (DSPs). KID, as a multisite phosphorylation region, provides alternative recognition sites for DSPs and activates them by binding a phosphotyrosine (pY) to their SH2 domains. Significant steric, biochemical, and biophysical requirements must be fulfilled by each pair of interacting proteins as the adaptation of their configurations is mandatory for the selective activation of DSPs. The accurate 3D atomistic models obtained by modelling and molecular dynamics (MD) simulations of phosphorylated KID (p-KID) have been delivered to describe KID INTERACTOME. By taking phosphorylated KID^{pY721} and the N-terminal SH2 domain of phosphatidylinositol 3-kinase (PI3K), a physiological partner of KID, we showed the two proteins are intrinsically disordered. Using 3D models of both proteins, we probe alternative orientations of KID^{pY721} relative to the SH2 binding pocket using automatic docking (HADDOCK) and intuitive user-guided docking. This modelling yields to two possible models of the functionally related non-covalent complex KID^{pY721}/SH2, where one can be regarded as the first precursor to probe PI3K activation via KIT KID. We suggest that such generation of a KID/SH2 complex is best suited for future studies of the post-transduction effects of RTK KIT.

Keywords: receptor tyrosine kinase (RTK) KIT; kinase insert domain; KID; site-specific phosphorylation; signalling protein PI3K; SH2 domain; post-transduction process; PTP; molecular dynamics and folding; conformational plasticity; intrinsically disordered region; IDR; transient states; protein–protein (PP) binding



Citation: Ledoux, J.; Tchertanov, L. Site-Specific Phosphorylation of RTK KIT Kinase Insert Domain: Interactome Landscape Perspectives. *Kinases Phosphatases* **2023**, *1*, 39–71. <https://doi.org/10.3390/kinasesphosphatases1010005>

Academic Editors: Mauro Salvi and Christian Borgo

Received: 28 December 2022

Revised: 9 February 2023

Accepted: 13 February 2023

Published: 15 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human cell-to-cell communication is monitored by signals from its environment, to which the cell must respond appropriately. Cell membrane receptors with tyrosine kinase activity (RTKs) frequently promote the external–internal exchange of information [1–3]. Once activated by its specific molecular stimuli, such as growth factors, these RTKs phosphorylate their downstream cytoplasmic substrates—adaptors, signalling, and scaffolding proteins. Reversible protein phosphorylation provides a central regulatory mechanism in cells and represents a crucial step of post-transduction processes (PTPs) [4]. PTPs involve many intrinsically disordered (ID) proteins that most contain phosphorylation sites [5–8].

KIT is a double champion among RTKs regarding the number of ID regions and site-specific phosphotyrosines. As was reported early, the cytoplasmic domain of RTK KIT contains a tyrosine kinase domain (TKD) crowned by at least four inherently coupled ID regions/domains—juxtamembrane region (JMR), kinase insert domain (KID), activation (A-) loop, and C-terminal tail [9,10].

Each KIT ID region is a key regulatory element contributing to KIT activation and/or mediating protein–protein interactions (PPI) through their functional phosphotyrosine residues. Two tyrosines, Y568 and Y570, identified *in vivo*, and two additional sites, Y547 and Y553, detected *in vitro* [11], provide JMR bi-functional activity, a regulatory role in the KIT activation/deactivation process and the recruitment of adaptors, signalling, and

scaffolding proteins [12,13]. KID of KIT likely only participates in such recruitment through protein partners' (PP) selective recognition and binding [14]. Multiple functional phosphorylation sites of KID, three tyrosine (Y703, Y721, and Y730) and two serine (S741 and S746) residues provide alternative binding sites for intracellular PPs [15,16]. Phosphorylation of Y703 supplies the binding site for the SH2 (Src Homology 2) domain of Grb2 (growth factor receptor-bound protein 2), an adaptor protein initiating the MAPK (mitogen-activated protein kinase) pathway [17,18]. Phosphorylated Y721 and Y730 are the recognition sites of phosphatidylinositol 3-kinase (PI3K) and phospholipase C (PLC γ), respectively [19]. The function of Y747 has not been described empirically yet. Considering its abundant intra-KID contacts, the 'organising role' of tyrosine Y747 in stabilising the KID structure was formerly attributed [20]. Phosphorylated serine residues, S741 and S746, bind protein kinase C (PKC) and contribute to the retro-control of PKC activity under receptor stimulation. The other phosphotyrosines in KIT are Y823 and Y900 from the TKD C-lobe, and Y936 from the C-terminal tail [21,22].

The high population of phosphorylation sites in RTK KIT cytoplasmic domain (in particular, eight tyrosines) and their location in ID regions furnish extraordinarily sophisticated problems to study its post-transduction processes. To our great satisfaction, we found that RTK KIT is a modular protein, similar to many proteins of the human proteome [23]. Consequently, one part of the post-transduction processes may be studied by using KIT individual domains, regarded in certain approximations as independent from the rest of the protein. This approach provides a promising route for using such domains as independent units in research and the biotechnology of large multidomain proteins, particularly for studying PPI [24].

Many issues remain despite the apparent simplification of multidomain protein studies when employing a per-domain approach for research (e.g., reduced molecular size).

The phosphorylation and binding study to signalling proteins of each KIT domain having multiple functional phosphorylation sites presents a great challenge [3,25]. First, the most archetypical question is: how does phosphorylation influence the structure and conformation of KIT intrinsically disordered (ID) domains/regions? It was reported that PTPs induced significant changes in IDP structural and dynamical properties by affecting their energy landscapes [26–28]. PTPs cause a broad spectrum of effects, from local stabilisation or destabilisation of the secondary structures to global disorder-to-order transformations. Such structural and conformational events may cause a complete change in the protein folding, varying from an intrinsically disordered to a well-folded structure or even a sporadic switch between monomeric and oligomeric states [29].

Secondly, for ID regions containing more than one phosphorylation site, the question arises about the required number of phosphorylated sites for protein(s) binding. Focusing on the multi-site KIT KID, we are still unsure if single-site tyrosine phosphorylation is sufficient to create a signalling protein scaffold (a one-to-one process). It can be suggested that protein binding to multisite KID is a more collective multithreaded pre-process: one to many, or many to one, or many to many. For example, a protein binding may induce the required conditions for another KID tyrosine phosphorylation followed by binding of another protein specific to the newly created scaffold or a partner binding requires phosphorylation of two or more tyrosine sites at KID. To inspect these hypotheses, it is necessary to consider many cases of phosphorylation events, described by the factorial function. Only for KID possessing three phosphotyrosines, the number of combinations analysed is seven.

The most sophisticated enigma, a bona fide Pandora's box, is understanding the phosphate binding order as a time-dependent sequence of the PTP events. Decoding the intrinsic kinetics of multisite phosphorylation is key to understanding how multiple phosphorylated sites within a domain can collectively shape the transcriptional response. Characterisation of such phosphorylation kinetic should be based on biochemical and/or biophysical techniques, which may be applied to complex systems, including multiple phosphorylation and kinase cascades [30,31].

To enrich the reliable knowledge of KIT KID post-transduction processes, we first apply a systematic *in silico* approach to map the phosphorylation effects on the multisite KID possessing three functional tyrosines—Y703, Y721, and Y730. Such study should help to understand the phosphorylation-induced effects on KID and deliver phosphorylated structures as suitable targets to probe signalling proteins binding. The structural and conformational changes induced by phosphorylation of the tyrosine residues (all possible combinations are considered) in the 80-residue KID of RTK KIT were investigated using 3D modelling and an extended molecular dynamics (MD) simulation. Second, we analysed the empirically determined (X-ray) structure retrieved from the Protein Data Bank (PDB) [32] and referenced it as 2IUH. This structure represents a complex of intracellular signal transducer PI3K p85 N-terminal SH2 domain (SH2) bound to the phosphopeptide TNEYMDMK (p-pep, residues 718–725) of KIT KID with the phosphorylated tyrosine pY721 [33]. Study of the structural and dynamical properties (by extended MD simulation) of the molecular complex p-pep/SH2 and its free-ligand SH2 derivative, highlighted the impact of p-pep binding on the PI3K SH2 domain and provided the optimised protein partner for KID docking. Finally, docking of the phosphorylated KID into the PI3K SH2 domain was performed, and the resulting *de novo* models of the molecular complex were compared to the unique empirically determined structure 2IUH. The *de novo* KID/SH2 model is postulated as the first comprehensive precursor of RTK KIT signalling complex. This model adds new information about the well-studied RTK KIT, offers insights into its less studied post-transduction events, and highlights RTK KIT interactions and signalling pathways through KID. RTK KIT interactions with its downstream proteins represent the prominent perspective for developing novel therapeutic agents.

2. Results

2.1. KIT KID *In Silico* Phosphorylation

2.1.1. Modelling of the Phosphorylated KIT KID

For modelling the alternatively phosphorylated KID entities, we used the cleaved KID from RTK KIT, previously studied as an unphosphorylated (native and inactive) species [10,34]. To get ahead a bit, we showed that KID from the inactive state of KIT is very similar in the active state (unpublished data, partially presented in Supplementary Information, Figure S1). This similarity approves the use of KID extracted from RTK KIT in an inactive state, published in [10,20], as an initial model for probing phosphorylation of KID tyrosine sites. Seven 3D models of phosphorylated KID (Figure 1) were generated from a KID conformation taken at $t = 2 \mu\text{s}$ of the membrane-embedded full-length inactive KIT cytoplasmic domain MD simulation [10].

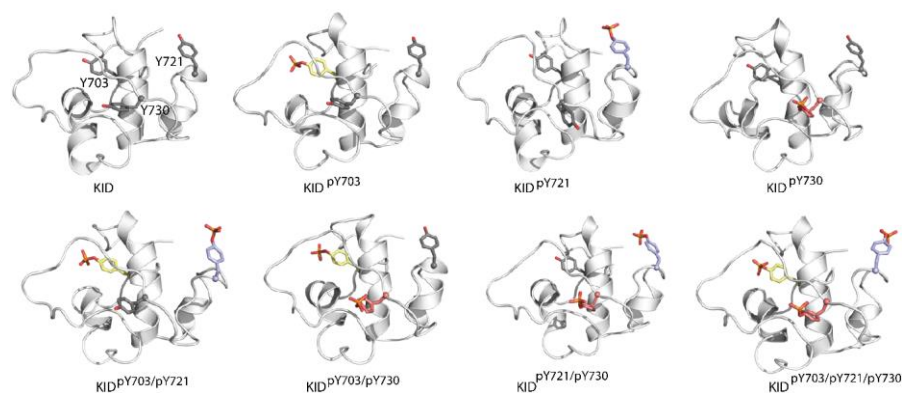


Figure 1. 3D models of phosphorylated KID. All models were generated from KID conformation taken at $t = 2 \mu\text{s}$ of MD simulation of the full-length cytoplasmic domain of KIT in an inactive state [10] by phosphorylation of tyrosine residues with phosphate group $-\text{O}-\text{PO}_3^{2-}$. Protein is shown as a cartoon, and phosphorylated tyrosine residues Y703, Y721, and Y730 are shown as yellow, light blue, and pink sticks, respectively.

2.1.2. General Characterisation of KIDs MD Simulations

Each optimised and well-equilibrated phosphorylated KID (p-KID) model was studied by extended (2- μ s) MD simulation running three times in strictly identical conditions using random initial velocity. First, the generated MD trajectories were characterised by conventional methods using commonly used descriptors and compared between them and the native KID.

The root means standard deviations (RMSDs) calculated on MD conformations from three replicas respective to the same initial structure of each p-KID presents good convergence (Figure S2). Compared to the ample variations of RMSDs in native KID, p-KID entities are slightly reduced. Except smoothers curves for KID^{pY721}, their profiles indicate significant conformational transitions, as observed in unphosphorylated KID [34].

2.1.3. Structural and Dynamical Features of p-KIDs

Assignment of secondary structures (SS) with DSSP reveals a helical fold in all p-KIDs, similar to native KID species (Figure 2A and Figure S3). Except for α H1-helix varying only in length, the other helices are transient and reversibly converted between folded and unfolded structures (α -helix; 3_{10} -helix; turn/bend), a phenomenon typical for IDPs. Depending on the unfolded residues fraction, the number and length of transient helices vary significantly within a MD trajectory of the same KID species and between them. Whether phosphorylated or not, KID contains from two to six helices formed by 20–35% amino acids (aas). The most folded helical structure is observed in KID^{pY721} (35%) and KID^{pY721/pY730} (32%), while the less folded in KID^{pY703} (20%). The other phosphorylated KIDs showed a folding in the 23–25% range, which differs from native (unphosphorylated) KIDs (29%). The ratio of $\alpha/3_{10}$ -helices, varying from two to five times in KIDs, except KID^{pY703} (by a factor of 21), indicates the apparent predominance of α -helices in all KID studied. Furthermore, the doubly phosphorylated KID^{pY721/pY730} shows two β -strands (D723-V728 aas) appearing in 45% of the simulation time instead of its 30% life-time in mono-phosphorylated KID^{pY721}. In another doubly phosphorylated KID, KID^{pY703/pY730}, a pair of short strands is observed in the D737-R740 fragment for 20% of the simulation time.

These observations illustrate the most apparent phosphorylation-induced effects on KID folding (secondary structures) in KIT entities possessing (i) the phosphorylation of pY721 either as single (KID^{pY721}) or double (KID^{pY721/pY730}) site; and (ii) the phosphorylation of pY730 in doubly phosphorylated KID^{pY703/pY730} and KID^{pY721/pY730}. In both cases, phosphorylation either increases KID overall folding, as evidenced in (i), or promotes additional folding in only its immediate environment, manifested in (ii), while the overall folding is diminished.

While a significant portion of p-KID residues is not folded in regular structures, they all exhibit high flexibility, as observed in unphosphorylated KID [10,20,34]. To rationally compare the phosphorylation-induced effects on KID flexibility, the root means square fluctuations (RMSFs) were calculated on each p-KID concatenated replicas after fitting on the best-conserved α H1-helix portion (Y703-L706) at $t = 0 \mu$ s. This approach characterises p-KID RMSF variations respective to the most conserved KID structural element.

The RMSF curves and tubes defined on the average conformations of KID and p-KIDs demonstrate (i) reduced RMSF values of N- and C-terminals residues in all p-KIDs relative to KID; (ii) partial re-distribution of the most and least fluctuating fragments along KID sequence (Figure 2B,C; Table S1).

For example, the highly fluctuating fragment near D716 in native KID shows reduced RMSF values in all p-KIDs. In contrast, the moderate fluctuations of KID segment G727-V728 significantly increased in KID^{pY703} (RMSF values up to 8 Å). The minimally fluctuating regions between p-KIDs and relative to native KID vary little in their sequence position and RMSF values.

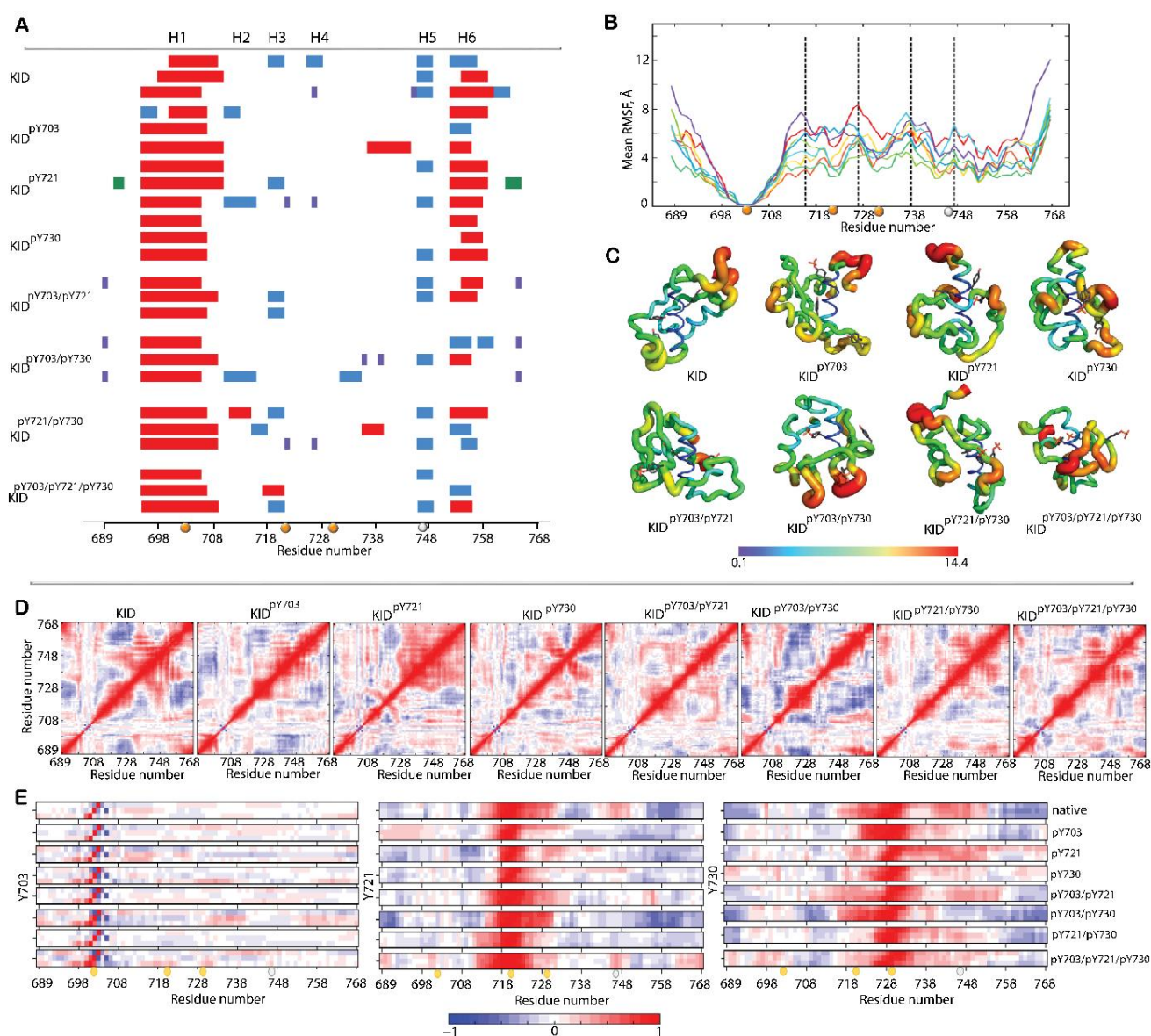


Figure 2. Structural and dynamical properties of unphosphorylated and phosphorylated KID. (A) The α H- (red), 3_{10} -helices (blue), and β -strands (violet and green) were assigned by DSSP on average conformation from each MD trajectory for every KID. The helical structures are labelled from H1 to H6; (B) RMSFs were computed on the C α atoms of MD conformations after fitting on initial conformation (at $t = 0$ ns) and alignment on the best-conserved portion of α H1-helix (Y703-L706). Different KID entities are distinguished by colour: KID (lilac), KID^{pY703} (red), KID^{pY721} (orange), KID^{pY730} (yellow), KID^{pY703/pY721} (lime), KID^{pY703/pY730} (green), KID^{pY721/pY730} (turquoise), and KID^{pY703/pY721/pY730} (blue); (C) the RMSFs of the average conformations for KIDs are presented as tubes. The tube size is proportional to the residue atomic fluctuations computed on the backbone atoms. Red–blue gradient shows the RMSF values, from large (>14 Å, in red) to small; (D) dynamical inter-residue cross-correlation maps computed for all C α -atom pairs of MD conformations of concatenated trajectories of each KID; (E) cross-correlations zoomed on the phosphotyrosines and their immediate environment after fitting on the α H1-helix. The blue–red gradient shows the correlations from -1 (blue) to 1 (red). The tyrosine positions are shown as balls: phosphotyrosines Y703, Y721, and Y730 in yellow, and Y747 in grey.

In general, KID^{pY703} and KID^{pY721/pY730} are the most flexible p-KIDs, while KID^{pY703/pY721} and KID^{pY703/pY730} are more ‘rigid’ entities. After a detailed comparison of all studied

KIDs, we noted that the minimally fluctuating fragments (RMSF value ≤ 3 Å) are M722-M724 (in KID^{pY703/pY721}, KID^{pY703/pY730} and KID^{pY721/pY730}), V732 (in KID^{pY703/pY730}), R743-G745 (in KID^{pY730} and KID^{pY703/pY721}), and T753-P754 (in KID, KID^{pY721}, KID^{pY730} and KID^{pY703/pY721/pY730}). These minimally fluctuating segments of p-KID regions are involved in non-covalent intramolecular interactions contributing to stabilising their **tertiary** structures, as observed in native KID [34].

The maximally fluctuating fragments (RMSF value ≥ 6 Å) are D716 (KID), P726-V728 (KID^{pY703}, KID^{pY730} and KID^{pY703/pY721/pY730}), K738-R739 (KID, KID^{pY721}, KID^{pY730}, KID^{pY721/pY730}, KID^{pY703/pY721/pY730}), and I748 (KID^{pY721/pY730}). The fluctuations of those regions are increased by at least 2 Å compared to native KID, suggesting an enhancement of conformational flexibility upon phosphorylation.

These results evidenced the various degrees of flexibility of p-KIDs, containing either minimally or maximally fluctuating fragments relative to α H1-helix, taken as a reference. Curiously, these fragments do not contain tyrosine residues.

The inherent dynamics of the intrinsically disordered KIDs were analysed with the cross-correlation matrix computed for all C α -atom pairs of each KID. The C α -C α pairwise cross-correlation maps demonstrate highly coupled motions between KID structural fragments, even vastly distant (Figure 2D). The maps are highly different for KIDs studied, showing either the block-like patterns composed of clearly delimited correlated segments (KID, KID^{pY703} and KID^{pY703/pY730}) or very smooth patterns (KID^{pY721}, KID^{pY730}, KID^{pY721/pY730} and KID^{pY703/pY721/pY730}). Focusing on the phosphotyrosines, we note that pY703 motion is correlated with only a minimal number of residues (2–3 amino acids, aas) close to Y703 at KID sequence. These correlations are positive and negative with preceding and following residues, respectively. This suggests that Y703 acts as a rotation centre (kerner) for α H1 helix. Apart from this region, Y703 does not reveal any strong correlation with the rest of the KID residues.

In contrast, Y721 and Y730 show strong correlations with many residues close to these sites (strong positive correlations) or distant residues (strong or moderate negative correlations). Surprisingly, the number of neighbouring (sequence position) correlated residues with Y721 is decreased on mono-phosphorylated KID or moderately shifted when phosphorylated with pY730. Concerning Y730, the correlated region is larger than Y703 and Y721 (up to 50% of residues). In KID^{pY703}, Y721 correlates positively with well-defined proximal residues in terms of sequence position when positive correlations are more diffused in other KIDs. This pattern is clearly reduced in KIDs phosphorylated at Y721.

2.1.4. Shape of p-KIDs and Their Stabilisation by H-Bonds

P-KIDs size described by the radius of gyration (Rg) is nearly similar in most studied proteins (Figure 3A). Statistically, the Rg shows a Gaussian unimodal distribution except for the bimodal curve for KID with three phosphorylated sites – KID^{pY703/pY721/pY730}. The most probable Rg values vary in a narrow range, from 11.7 to 13.5 Å, with a mean value (mv) of 12.5 Å for all p-KIDs. Only Rg mv for KID^{pY703} and KID^{pY703/pY721/pY730} (main peak) is slightly decreased (12.2–12.3 Å). p-KID Rg values correspond well to those observed in native KID. Similar to native KID, p-KID conformations are stabilised by many H-bonds maintaining their globular-like shape (Figure 3B,C), as in unphosphorylated KID [34].

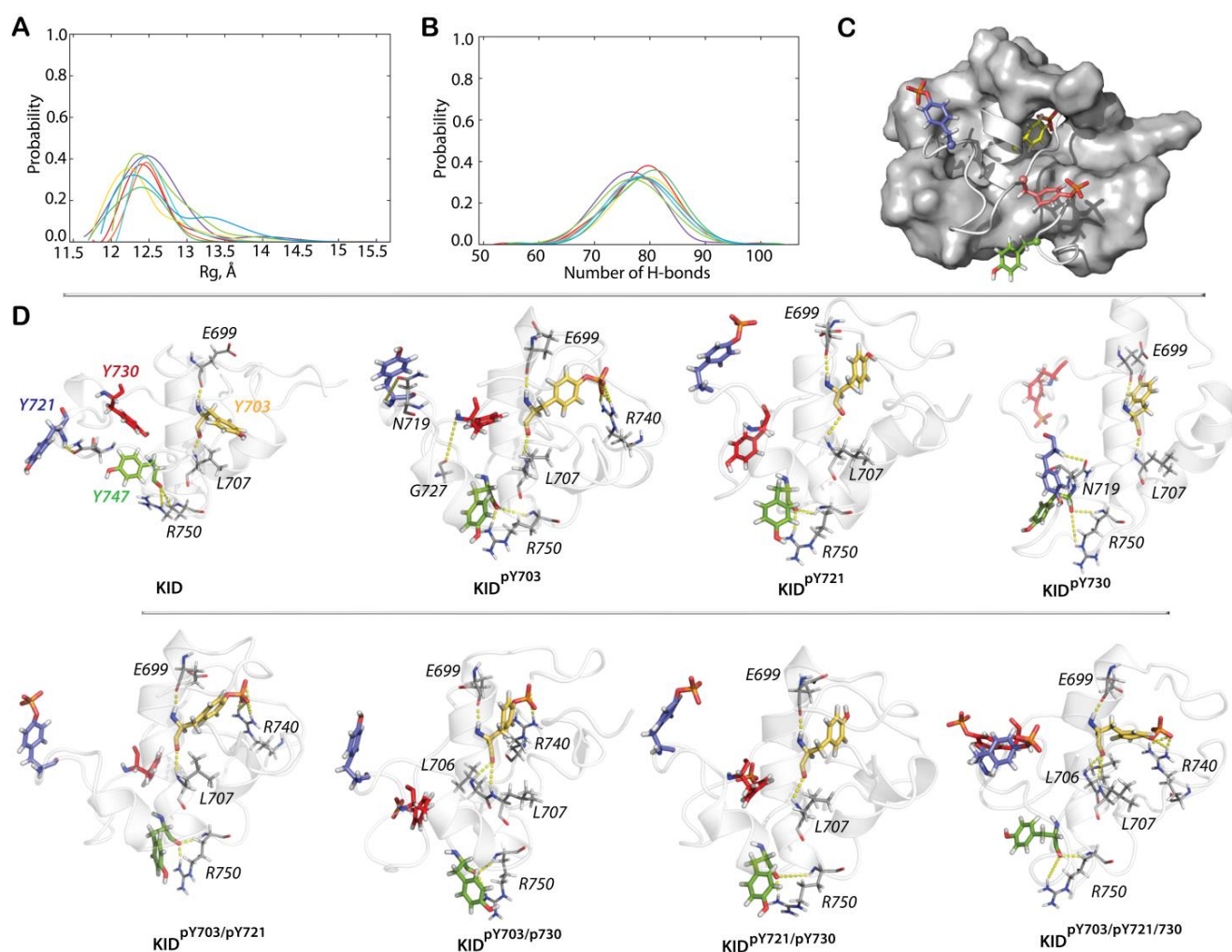


Figure 3. KIDs shape and its stabilisation by H-bonds. Radii of gyration (Rg) (A) and number of H-bonds stabilising KIDs conformations (B) that maintain their globular-like shape (C). Different KID entities are distinguished by colour: KID (lilac), KID^{pY703} (red), KID^{pY721} (orange), KID^{pY730} (yellow), KID^{pY703/pY721} (lime), KID^{pY703/pY730} (green), KID^{pY721/pY730} (turquoise), and KID^{pY703/pY721/pY730} (blue). In (C), protein is shown as a surface cartoon, and tyrosine residues as coloured sticks: Y703 in yellow, Y721 in blue, Y730 in red, and Y747 in green; (D) H-bonds involving tyrosine residues. H-bonds were calculated for the contacts D-H...A with distance D...A ≤ 3.6 Å and pseudo-valent angle DHA $\geq 120^\circ$ observed with an occurrence $\geq 30\%$. In (C,D), proteins are shown as cartoons, tyrosine residues as large coloured sticks and contacting residues as thick grey sticks. Dashed yellow lines display H-bonds. The numbering of tyrosine residues is shown on KID.

In all p-KIDs, the phosphotyrosines are located on the protein surface, and their phosphorylated sidechain is properly exposed to the solvent or proteins. However, pY703 is systematically involved in H-bonds by its phosphate oxygen atoms acting as acceptors. Indeed, two oxygen atoms of pY703 form the salt bridge with R740 in mono-phosphorylated (KID^{pY703}), bi-phosphorylated (KID^{pY703/pY721} and KID^{pY703/pY730}), and three-phosphorylated entities (Figure 3D). We defined relevant H-bonds if their occurrence was more than 30% (soft criterion). This cut-off suggests that many p-KID conformations with pY703 may have an abundant number of alternative orientations disfavouring pY703 intra-KID H-bonds. Other phosphotyrosines sidechains, however, hold all their oxygen atoms available for post-transduction events.

2.1.5. P-KIDs Solvent Accessibility

Analysis of the total solvent accessibility surface (SASA) of unphosphorylated KID and its phosphorylated derivatives shows their unimodal distribution, differing merely by the peak height (85–100%) (Figure 4A). The SASA calculated for each phosphotyrosine individually in each p-KID entity systematically showed two peaks at proximity to 160 and 200 Å². The Y703 smaller value of SASA (at 162 Å²) is observed in KID^{pY721}, KID^{pY730} and KID^{pY721/pY730}, while the second peak (at 205 Å²) in KID^{pY703}, KID^{pY703/pY721}, KID^{pY703/pY730} and KID^{pY703/pY721/pY730}. The Y721 smaller SASA value (at 160–162 Å²) is observed in KID^{pY703}, KID^{pY730} and KID^{pY703/pY730}, the second peak (at 203 Å²) in KID^{pY721}, KID^{pY703/pY721}, KID^{pY721/pY730} and KID^{pY703/pY721/pY730}. For Y730, the smaller SASA values is observed in KID^{pY703}, KID^{pY721} and KID^{pY703/pY721}, while the second peak (at 200 Å²) is observed in KID^{pY730}, KID^{pY703/pY730}, KID^{pY721/pY730} and KID^{pY703/pY721/pY730}.

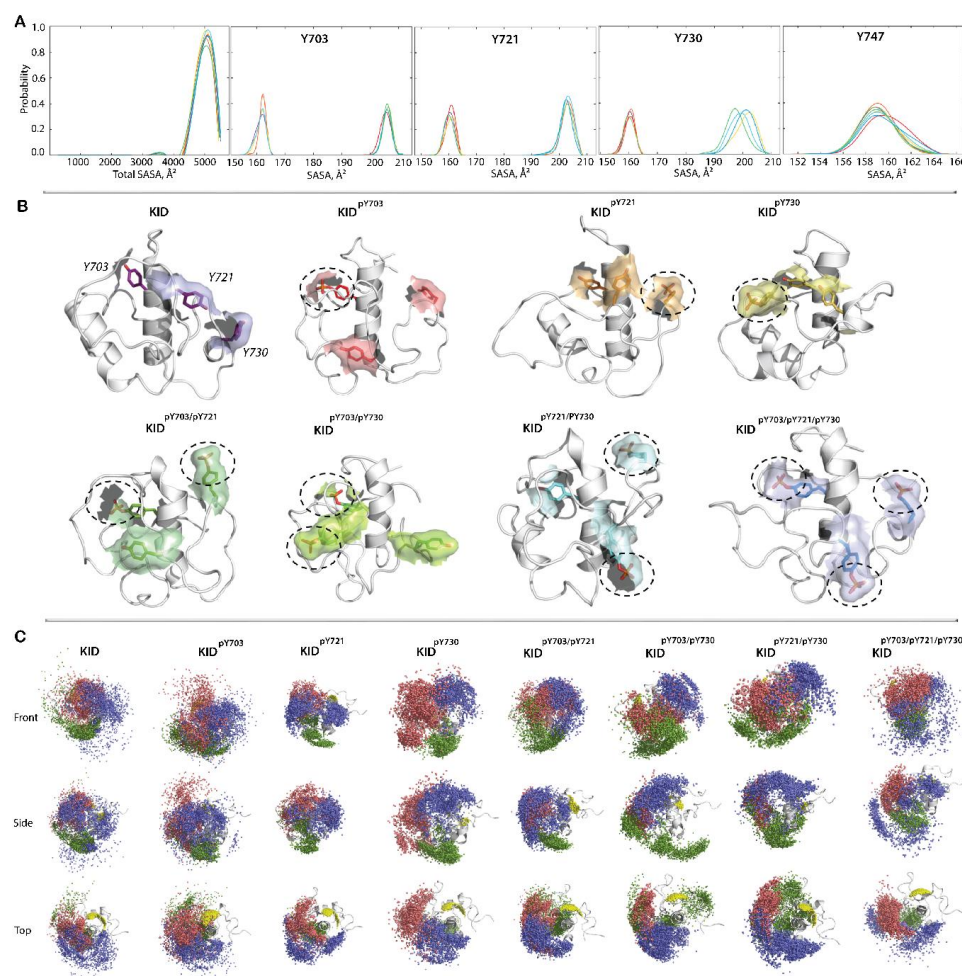


Figure 4. The solvent accessibility surface of phosphotyrosines and their spatial distribution. (A) The SASA probability distribution of total KID and each of its tyrosine residues. Different KID entities are distinguished by colour: KID (lilac), KID^{pY703} (red), KID^{pY721} (orange), KID^{pY730} (yellow), KID^{pY703/pY721} (lime), KID^{pY703/pY730} (green), KID^{pY721/pY730} (turquoise), and KID^{pY703/pY721/pY730} (blue); (B) the solvent/protein contact surface is traced by the van de Waals surfaces of water molecules in contact with the protein. Dashed ovals contour the contact surfaces of the phosphotyrosine in each p-KID. The numbering of tyrosine residues is shown in KID; (C) spatial distribution of KID tyrosines (after a fitting of conformations taken each 2 ns from the concatenated trajectories on Y703-L706, $t = 0$ μs) are presented by the oxygen atom of hydroxyl-group or phosphorus atoms for unphosphorylated and phosphorylated tyrosine residues, respectively. Three orthogonal projections with atoms of KID, Y703, Y721, pY730, and Y747, respectively, coloured in yellow, blue, red, and green.

The main conclusions of these observations are as follow: (i) the global SASA is practically identical in all p-KIDs studied, with slight variations in the distribution height, a minimum (82%) in KID^{pY703/pY721} and a maximum (98%) in KID^{pY703/pY721/pY730}; (ii) the SASA of each tyrosine shows two peaks; (iii) each smaller peak, corresponding to unphosphorylated tyrosines, is equivalent to native KIDs, suggesting unchanged availability to the solvent of those tyrosine; (iv) the second peak of each tyrosine, and with the most significant SASA value, is composed only of its phosphorylated derivatives; (v) the 40 Å² increment between two SASA peaks is consistent for all p-KIDs studied, and independent from the number of phosphorylated sites; (vi) the SASA of unphosphorylated Y747 reveals a near-equivalent unimodal distribution in all proteins studied.

The tyrosines spatial distributions are represented on the average conformation by the unphosphorylated tyrosines OH group oxygen atom, and the phosphorylated tyrosines phosphorus atom. Such distributions are significantly extended and show an oblate spherical sector shape subdivided into two regions for the same tyrosine (Figure 4C). Their frequent overlapping (i) clearly demonstrates that the tyrosine residues SASA are divergent in terms of surface area, spatial position, and orientation; and (ii) suggests a remarkable diversity of MD conformations where the phosphotyrosines exposition to the solvent is either adequate or limited by steric constraints induced by their neighbouring residues' environment.

To illustrate such cases, we represented the accessible surfaces for the tyrosine residues traced out by the van der Waals surfaces of the water molecules atoms when in contact with the protein. This alternative approach to estimating the possible contact surface is coherent with the SASA method. Traced on each phosphorylated tyrosine, the contact surface displays the tyrosine residues sidechain systematic accessibility to the solvent, surprisingly whether phosphorylated or not, except for unphosphorylated Y703 in native KID and in doubly phosphorylated KID^{pY721/yY730}, where their accessibility is constrained, and correlating with its small SASA value (at 162 Å²) (Figure 4A,B). Once Y703 is phosphorylated, its surface is accessible to the solvent, similarly to the other KID phosphotyrosines. Such availability of each phosphotyrosine is necessary for the recognition of KID linear motifs by protein partners, phosphate transfer, and signal transduction events.

2.2. RTK KIT Signalling Protein PI3K

Tyrosine phosphorylation controls RTK KIT cell signalling through the recruitment and activation of proteins involved in downstream signalling pathways. These events are mediated through pY binding of SH2 (Src Homology 2) and/or PTB (phosphotyrosine-binding) domains of signalling proteins [2,35,36]. SH2 and PTB domains typically recognise a pY residue within a specific amino acid sequence context. In particular, SH2 domain of human phosphatidylinositol 3 kinase (PI3K) has been reported to bind preferentially to pYφXφ (where φ is a residue with a hydrophobic side chain and X any residue) [37]. Therefore, the specificity of SH2 domains is conferred to some extent by binding various pY-containing motifs.

Our modelling of phosphorylated KID generated an exhaustive number of targets and opened a route for the description of a crucial part of KIT INTERACTOME—a set of macromolecular complexes formed by KID with its cellular protein partners.

2.2.1. Available Crystallographic Structures Related to KID Binding: Analyses and Hypothesis

The Protein Data Bank (PDB) [33] was searched for KIT signalling proteins focusing on KID-specific partners with empirically determined structures (benchmarks). We identified a co-crystallized molecular complex (PDB ID: 2IUH, 2.0 Å resolution) [33] composed of a fragment of signalling protein PI3K and a phosphopeptide TNEYMDMK (p-pep) of KIT KID (residues T718-K725), including the phosphorylated tyrosine pY721. To develop the first PP complex, we chose PI3K, an intracellular signal transducer enzyme specific to KID of KIT [35]. It contains a catalytic subunit of 110 kDa (p110) and a small regulatory subunit

of 85 kDa (p85). The PI3K p85 subunit contains two SH2 domains, an N- and a C-terminal SH2 domains, that bind to two closely spaced pYXXM motifs (pY is the phosphorylated tyrosine; X any amino acid; M Met). The p85 is phosphorylated during the binding of PI3K to RTK KIT and detached from p110, which then phosphorylates the plasma membrane phosphatidylinositol 4,5-bisphosphate (PIP₂) [38].

The structure of the molecular complex 2IUH includes p85 (N-term) sequence G321-D440 with a single point mutation Q330N. From here onwards, this domain will be referenced as SH2 for simplicity. It represents an archetypical structure of an SH2 domain consisting of a central antiparallel β -sheet formed by three or four β -strands (β 1- β 4) flanked by two α -helices (α H1 and α H2) (Figure 5A,B).

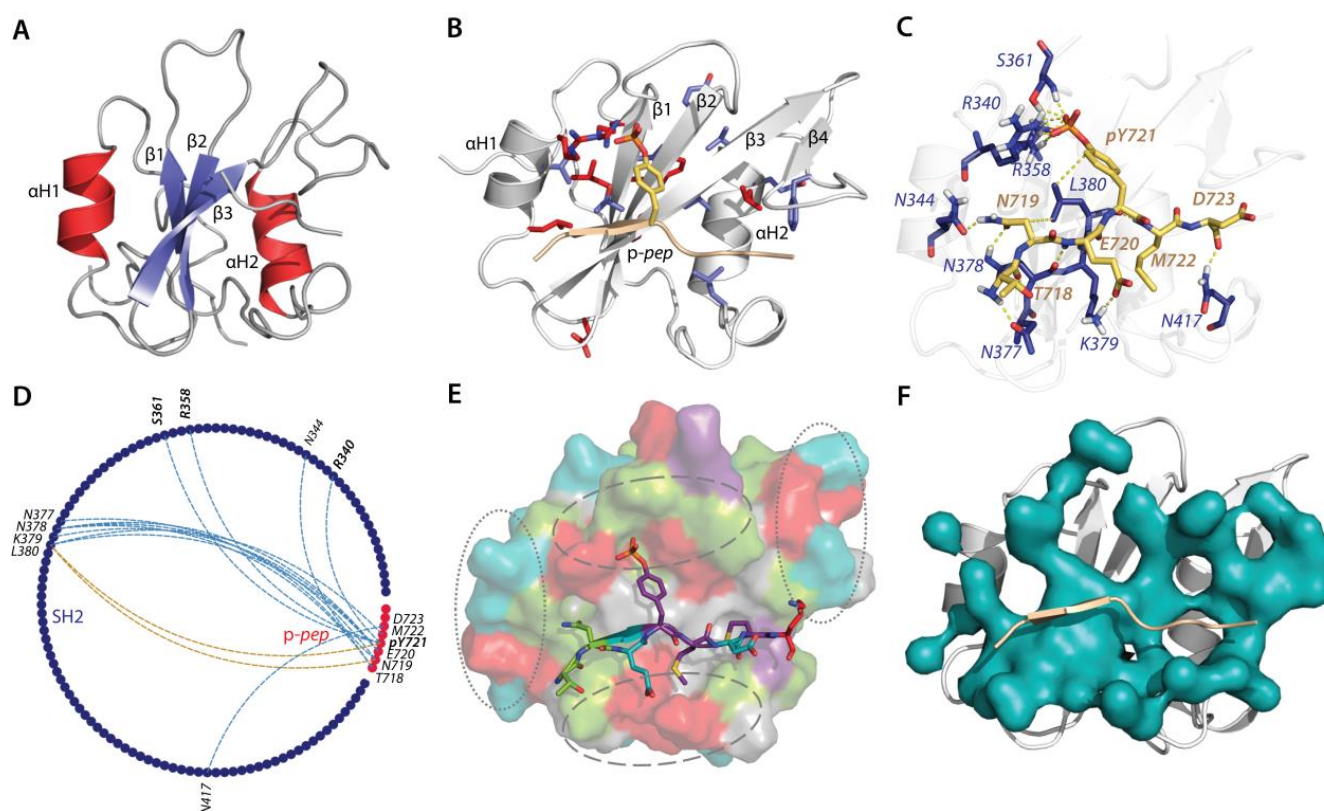


Figure 5. The crystallographic structure of the SH2 domain of PI3K and its characterisation. (A) The general view on structure of free-ligand SH2 domain of p85 (PDB ID: 1PIC). Protein is shown as a cartoon: α -helices, β -strands, and coils are displayed in red, blue, and grey, respectively; (B) structure of the co-crystallized molecular complex composed of SH2 domain fragment of PI3K and a phosphopeptide TNEYMDMK (*p-pep*) of KIT KID. SH2 and *p-pep* are shown as a cartoon in grey and beige, respectively. The highly conserved (identical) residues and residues similar by physicochemical properties among SH2 domains are in red and blue, respectively; (C) the non-covalent interactions in the molecular complex *p-pep*/SH2; (D) H-bonds (in blue) and hydrophobic contact (in beige) stabilising *p-pep* (in red) and SH2 domain of PI3K (in dark blue) showed as a string diagram; (E) surface representation of *p-pep*/SH2 binding pocket with *p-pep* (in sticks), stained by physicochemical property of amino acids: the positively and negatively charged are in red and blue, respectively, polar in green, amphiphilic in purple, and hydrophilic in grey. Two pairs of areas delimited by pointed and dashed lines are formed by residues displaying similar physicochemical properties; (F) definition of SH2 binding pocket by Fpocket.

KID *p-pep* is located on the surface of the SH2 binding pocket formed by the central β -sheet and α -helices residues (Figure 5B,E). The β -sheet is positively charged and polar residues combine two quasi-symmetrical pocket surfaces in respect to the *p-pep* main axis. In addition, the polar positively and negatively charged residues from α -helices

surround the N- and C-terminals of *p-pep*. Such a surface is greatly favourable to multiple non-covalent interactions with each *p-pep* residue without exception, encouraging the sandwich-like position of *p-pep* and the SH2 domain (Figure 5C,D).

First, the pY721 of *p-pep* is engaged in hydrogen (H-) bonds and ionic bonding with R340, R358, and S361 of the SH2 domain, forming salt bridges. Second, its nearest negatively charged residues, E720 and D723, form H-bonds with K379 and N417. The polar N719 contributes to an H-bond pattern with its homologs, N344 and N378, acting as an acceptor and donor, respectively. The H-bond pattern of *p-pep* is completed by a main chain interaction involving T718. Finally, the bifurcate van der Waals contact formed by KID M722 adds to the tight *p-pep* binding to SH2. The peptide position in the SH2 binding pocket and the H-bond pattern maintaining it correspond to the canonical mode of SH2 binding [39].

The overall binding pocket of SH2, defined with Fpocket [40], is more significant than its surface occupied by *p-pep* (Figure 5F). As so, it can accommodate a larger KID fragment or retain *p-pep* in an alternative orientation concerning the X-ray pose. Following these suggestions, we emphasise *p-pep* features. It should be noted that the short *p-pep* (TNEYMDMK) contains three residues preceding pY721 and four residues following it. These residues exhibit similar biophysical properties making *p-pep* almost symmetric concerning its recognition properties. Moreover, in a solvent, the polypeptide *p-pep* per se is a highly flexible entity. While the SH2 binding pocket demonstrates symmetry in its surface residues—two pairs of areas with similar biophysical properties (Figure 5E)—we suggested that (i) *p-pep*/SH2 complex in solution may differ from its conformation in the solid state; (ii) *p-pep* may maintain an alternative orientation in the SH2 binding pocket concerning the X-ray pose; and (iii) KIT KID docking into PI3K SH2 may also produce alternative solutions. Such suggestions prompt us to study *p-pep*/SH2 complex in solution.

2.2.2. Molecular Dynamics Simulations of *p-pep*/SH2 Molecular Complex

To investigate the dynamical properties of *p-pep*/SH2 complex, we used extended (2 μ s) MD simulation. As three replicas of simulations bore remarkable similarity in RMSD and RMSF profiles and values (Figure 6B–D), further analysis was performed on the concatenated trajectories for the time-dependent or time-independent metrics.

The stable and low RMSD values (1.5–2.0 Å) observed in the SH2 domain during all replicas are typical for excellent simulation convergence. The higher RMSF values are preserved for residues located in the loop between β 1 and β 2 strands, a part of the SH2 binding pocket accommodating pY721 from *p-pep* in structure 2IUH. *P-pep* residues from the N- and C-terminus fluctuate significantly. Such increased flexibility of SH2 and *p-pep* facilitates the two entities mutual adaptation.

The folding (2D) and tertiary (3D) structure of the *p-pep*/SH2 complex is very similar between the MD replicas and during a replica (Figure 6A,E). Congruent to empirical structure 2IUH, the SH2 domain MD conformations consist of the central ‘core’ of β -strands, β 1 (G353–R358), β 2 (Y368–R373), and β 3 (N378–F384), organised in an anti-parallel sheet, and two highly conserved α -helices, α H1 and α H2. The central axis of these α -helices is nearly parallel. As observed in the crystallographic structure 2IUH, the SH2 domain MD conformations contain a well-conserved second antiparallel sheet formed by two short β -strands, β 4 and β 5.

Apart from these well-organised and highly conserved structural units during the MD simulation, we localised in the SH2 domain two intrinsically disordered regions (IDRs)—*F1* and *F2*. As evidenced by transient structures reversibly converting between the folded and unfolded states (3^{10} helices; β -strand; turn; coil), these IDRs, one positioned between β 1 and β 2 stands (*F1*) and the other between α H2-helix and β 5 (*F2*), manifest a structural disorder.

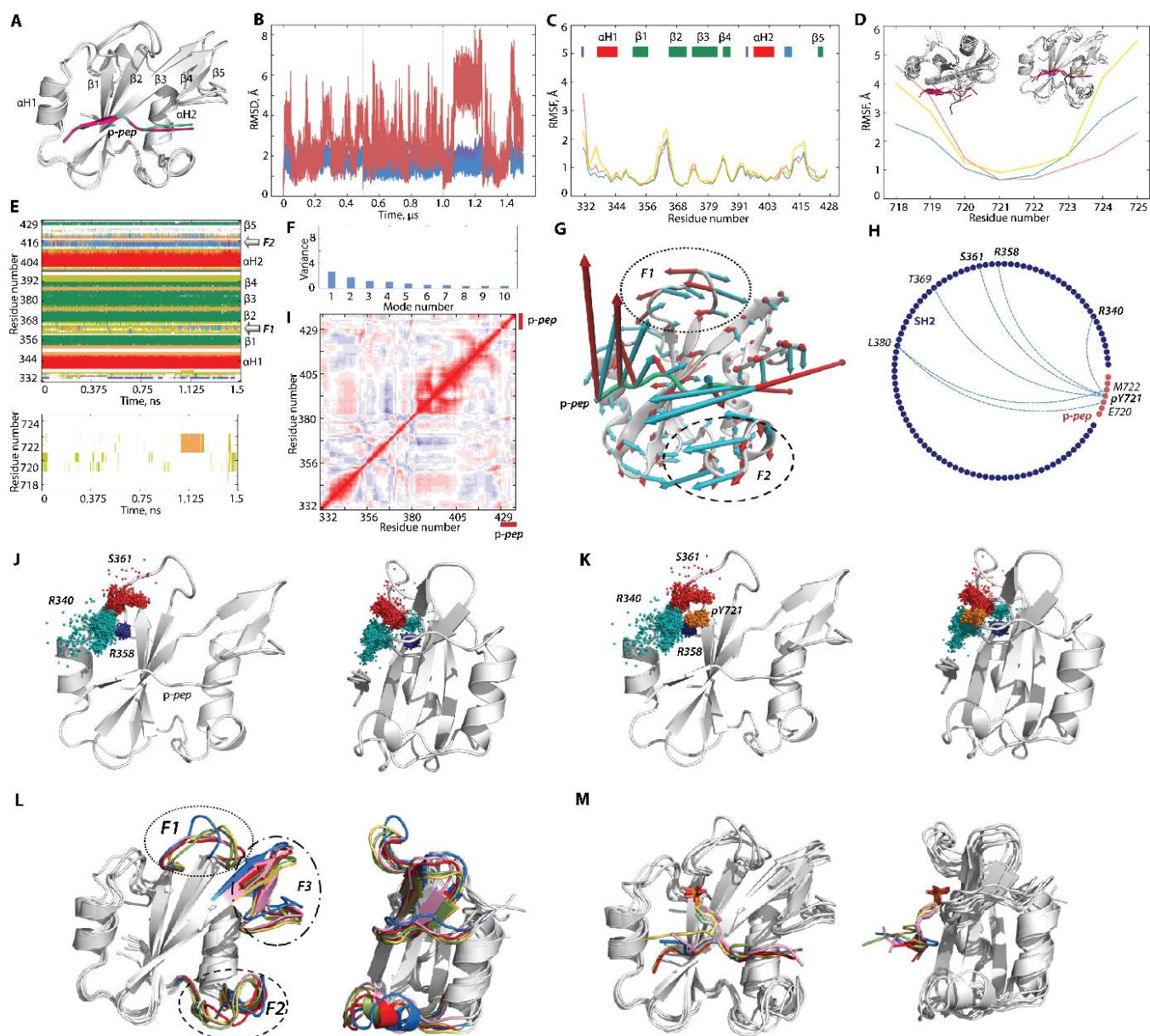


Figure 6. Structural and dynamical properties of p-pep/SH2 complex. (A) Superimposition of the equilibrated p-pep/SH2 complex at $t = 0 \mu\text{s}$ (p-pep in pink) on the X-ray structure 2IUH (p-pep in cyan); (B) RMSDs are computed on the C α atoms after fitting on the ‘core’ β -sheet (G353-R358, Y368-R373, N378-F384) of the initial conformation (at $t = 0 \text{ ns}$). The whole complex is coloured in purple, p-pep in red, and the SH2 domain in blue; (C) each MD trajectory RMSFs were computed on the C α atoms of the SH2 domain after least squares fitting on the ‘core’ β -sheet of the initial conformation (at $t = 0 \text{ ns}$) identical for each trajectory. The folding of the SH2 mean conformation is shown in the insert; (D) RMSFs computed on the C α atoms of p-pep after fitting on the initial conformation (at $t = 0 \text{ ns}$) identical for each trajectory. The insert shows two conformations illustrating the large RMSF values for its N- and C-terminals; (B–D) MD replicas 1–3 are distinguished by colour (red, yellow, and blue); (E) time-dependent evolution of each residue secondary structure as assigned by DSSP method for SH2 domain (top) and p-pep (bottom): α -helices are in red, 3^{10} -helices in blue, β -strand in green, turn in orange, and bend in dark yellow (replica 2); (F) SH2 PCA modes calculated for the concatenated MD trajectory (replicas 1–3) after least squares fitting of the MD conformations on the ‘core’ β -sheet (αC -atoms) of the initial conformation ($t = 0 \text{ ns}$). The bar plot gives the eigenvalue spectra in descending order for the first 10 modes; (G) atomic components in the two first PCA modes of the SH2 domain are drawn as red (first mode) and blue (second mode) arrows projected onto the

average structure. Only motion with an amplitude ≥ 4 Å is shown. SH2 is in grey, and *p-pep* is in green; (H) H-bonds (blue dashed lines) stabilising *p-pep* (in red) and the SH2 domain of PI3K (in blue) shown as a string diagram. Only the contacts with an occurrence $\geq 50\%$ were taken into consideration; (I) dynamical inter-residue cross-correlation maps computed for all C α -atom pairs of SH2 MD conformations of concatenated trajectories after fitting on the 'core' structure β -sheet (α C-atoms) of the initial conformation ($t = 0$ ns). The blue–red gradient shows the correlations from -1 (blue) to 1 (red); (J,K) the distribution of the representative atoms from residues R340, R358, and S361, which act as H-bond and/or salt bridge donor/acceptor centres in non-covalent interaction with pY721 of *p-pep* (data were taken every 500 ps). Representative atoms are: OG, the oxygen atom of S361 side chain; CZ, the carbon atom at two amino groups of R340 and R358. OG and CZ atoms are defined on the structural formula of serine and arginine in Figure S4; P is the phosphorus atom of pY721; (L) superimposition of the SH2 domain's representative conformations of each cluster shown in two orthogonal projections. Clustering was based on the RMSD values (cut-off 0.75 Å) after least squares fitting on the β -sheet core (G353–R358, Y368–R373, N378–F384). Protein is displayed as a grey cartoon with three IDRs, *F1*, *F2*, and *F3*, distinguished by colour corresponding to the respective cluster and delimited by pointed (*F1*), dashed (*F2*), and dotted (*F3*) lines; (M) conformation of *p-pep* in the representative conformations from C1–C4 clusters *p-pep*/SH2 complex and its orientation in respect to SH2 binding pocket. Two orthogonal projections are shown.

The cross-correlation matrix computed on the concatenated trajectories of SH2 shows a very smooth pattern and partially reflecting the expected coupling motion between the 'core' structure β -stands, and the relative rigidity of SH2 during the simulation. (Figure 6I).

The principal component analysis (PCA) demonstrates that the first six modes describe 80% of the motions' variance of *p-pep* complex, with the first and second modes explaining only 30 and 20% of the movement, respectively (Figure 6F). The principal contributor to these modes is *p-pep* because the largest displacements of its N- and C-terminals, in mutually perpendicular directions, reveal a significant degree of dynamical disorder (Figure 6G).

Nevertheless, the central part of *p-pep* conserved its position in the binding pocket of SH2 and demonstrated only a slight motion. Such *p-pep* location is stabilised by highly conserved salt bridges formed by pY721 with R340, R358, and S361 of the SH2 domain, as was observed in the X-ray structure 2IUH. Other multiple H-bonds and van der Waals contacts, stabilising *p-pep* in 2IUH, disappeared during MD simulation, and only the bifurcated H-bond E720...L380...M722 is maintained with occurrence $\geq 50\%$ (Figure 6H). Interactions with M724 of the pYXXM recognition motif were only transitory. The second significant contributor to the PCA modes is SH2 IDRs, *F1* and *F2*, showing a pendulum-like movement with a relatively large amplitude orthogonally to the α -helices axis.

Focusing on the SH2 domain residues contributing to pY721 binding in structure 2IUH and during MD simulation, we generated the spatial distribution of their representative atoms. The representative atoms from residues R340, R358, and S361, acting as H-bond and salt bridge donor/acceptor centres in non-covalent interaction with pY721 of *p-pep* atoms are: OG, the oxygen atom of S361 side chain; CZ, the carbon atom at two amino groups of R340 and R358; (defined on structural formula of serine and arginine in Figure S5); P is the phosphorus atom of pY721.

The oxygen (OG) atom of the appreciably fluctuating S361 residue and the carbon (CZ) atoms from the low-fluctuating R340 showed similar enlarged distributions around the pY721 well-packed areas (Figure 6J,K). From the other side, the CZ carbon atoms of the lowly fluctuating R358 manifest compact spatial distributions. We suggested that, in favourable steric conditions, arginine long side chain broadly explores its conformational space. Its outstanding conformational flexibility (multiple torsional degrees of freedom) generated many different rotamers.

As we aimed to define the principal factors governing *p-pep* recognition and binding by SH2 of PI3K, we first studied the conformational features of R340, R358, S361, and pY721.

The frequently used descriptors characterising rotation around each bond showed that the arginine residue conceivably exhibits a significant number of rotamers (Figure S5A). The arginine rotamers are described in the IUPAC conformational terms *trans* (*t*) and *gauche* (*g*) (<https://goldbook.iupac.org/> accessed on 27 December 2022) and a population. According to the torsion angles values, in the p-*pep*/SH2 complex, the *g-ttt* rotamer represents 56% of all R340 conformations. The other R340 conformations are fully heterogeneous and display a reciprocal similarity with occurrence less than 10%. R358 rotamers are regrouped into two clusters, C1, containing 85% of all MD conformations in *g+ttg+* configuration, and C2, which comprises 14% of *g+g+tt* rotamers. Residue S361 having only one torsional degree of freedom, shows two significant conformers, *g-* (60%) and *g+* (37%). The pY721 MD conformers are regrouped in three clusters, C1 (36%), C2 (35%), and C3 (21%), containing *g-g-g+t*, *g-g-g+g-*, and *g-g-g+g+*, respectively. Such results are explained by the local environment surrounding R340, R358, and S361. Despite salt bridges formed with the pY721 phosphate group, the majority of R340 neighbours are polar or identically charged residues that create repulsion forces favouring the flexibility of its sidechain; R358 neighbours have small hydrophobic or cyclic sidechains, consequently limiting the explored rotamers by attractive forces; S361's neighbours are polar or charged residues with donor sidechains for the majority. Stabilisation and destabilisation of non-covalent interactions between S361 and its surroundings may explain the resulting rotamers.

This conformational (or local) disorder of R340, R358, and S361 is another contributor to the inherent disorder of the SH2 domain. As S361 and R358 are located on the IDR **F1**, they appear as three types of disorder: (i) the backbone folding/unfolding (structural events); (ii) displacement/rotation (dynamical events); and (iii) inherent (local) rotational disorder. The high population of R358 rotamer (85%, *g+ttg+*) indicates a modest degree of dynamical and local disorders, also evidenced by the compact spatial distribution of the lowly fluctuating R358 CZ carbon atoms (Figure 6J,K). The highly fluctuating residue S361 is involved in structural, dynamical, and local disorders, resulting in its sparse distribution around the pY721 well-packed area. The lowly fluctuating residue R340 contributes to merely local rotational disorder, demonstrating a large conformational space.

Secondly, SH2 MD conformations of the p-*pep*/SH2 complex were clustered based on the RMSD values (cut-off 0.75 Å), calculated after the least squares fitting on the β -sheet 'core'. Five clusters, C1-C5, reveal various populations, the higher (C1, 37%), low (C2, 17%), and relatively lesser (C3-C4, 8% and C5, 4%) (Figure S6A). These clusters' representative conformations display noticeable structural and conformational disparities detected in (i) both IDRs, **F1** and **F2**, deriving from their unstable folding and large movements, and (ii) the β -sheet formed with β 4 and β 5 strands ending with a coiled C-term (Figure 6L). Consequently, the RMSD-based clustering of SH2 conformations from the p-*pep*/SH2 complex is identified as the third SH2 IDR, **F3**.

2.2.3. Characterisation of the Free-Ligand SH2 Domain of PI3K

To determine the impact of p-*pep* binding into the PI3K SH2 domain and to define the correct target for KIT KID docking, we examined the free-ligand (without p-*pep*) SH2 domain using extended (2 μ s) MD simulation. Like the p-*pep*/SH2 complex, three replicas of the free-ligand SH2 domain showed remarkable similarity of their RMSD and RMSF profiles and values (Figure 7C and Figure S7). Further analysis was performed either on one trajectory for characterisation of the time-dependent metrics or concatenated trajectories for the computing of time-independent statistical measures.

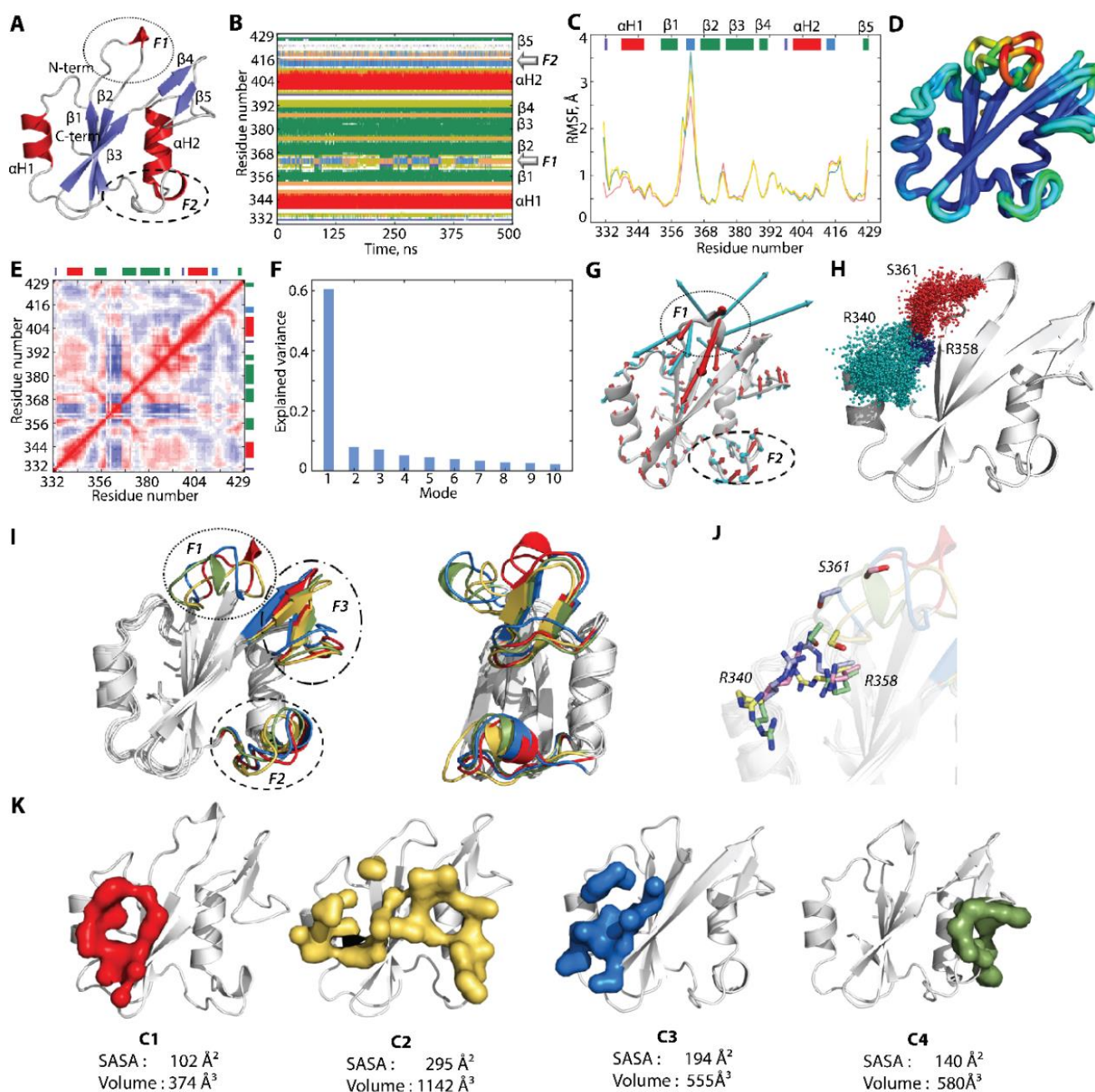


Figure 7. Structural and dynamical properties of the free-ligand SH2 domain from p85 α PI3K. (A) The mean conformations were calculated on the concatenated trajectories 1–3. Protein is shown as a cartoon: α -helices, β -strands, and coils are displayed in red, blue, and grey, respectively. Areas **F1** and **F2**, the most variable in 2D folding are delimited by pointed and dashed ellipses; (B) the secondary structure time-dependent evolution of each SH2 residue as assigned by the DSSP method: α -helices are in red, 3^{10} -helices in blue, β -strand in green, turn in orange, and bend in dark yellow (replica 2); (C) RMSFs computed for the $C\alpha$ atoms of the free-ligand SH2 domain conformations of each replica after fitting on the initial conformation ($t = 0$ ns), identical for trajectories 1–3. The 2D folding of the SH2 mean conformation is shown in the insert; (D) the average conformation of the SH2 domain is presented as tubes. The tube size is proportional to the residue atomic fluctuations computed on the backbone atoms. The red–blue gradient shows the RMSF values, from large (>3.7 Å, in red) to small; (E) dynamical inter-residue cross-correlation maps, computed for all $C\alpha$ -atom pairs of MD conformations of concatenated trajectories of SH2 after fitting on the ‘core’ β -sheet structure of the initial conformation ($t = 0$ ns). The blue–red gradient shows the correlations from -1 (blue) to 1 (red);

(F) PCA modes calculated for the concatenated MD trajectories (1–3) of free-ligand SH2 after least squares fitting on the ‘core’ β -sheet (α C-atoms) of the initial conformation ($t = 0$ ns). The bar plot gives the eigenvalue spectra in descending order for the first 10 modes; (G) atomic components in the two first PCA modes are drawn as red (first mode) and blue (second mode) arrows projected onto the average structure of the SH2 domain. Only motion with an amplitude ≥ 4 Å is shown. Two areas, **F1** and **F2**, manifested the most significant displacement are delimited by pointed and dashed ellipses; (H) distribution of the representative atoms from residues R340, R358, and S361, which act as H-bond and/or salt bridge donor/acceptor centres in non-covalent interaction with pY721 of p-pep (data were taken every 500 ps). Representative atoms are: OG, the oxygen atom of S361 side chain; CZ, the carbon atom at two amino groups of R340 and R358. OG and CZ atoms are defined on the structural formula of serine and arginine in Figure S4; (I) superimposition of the representative conformations from the free-ligand SH2 clusters are shown in two orthogonal projections. Protein is displayed as a grey cartoon with three IDRs, **F1**, **F2**, and **F3**, distinguished by colour corresponding to the respective cluster and delimited by pointed (**F1**), dashed (**F2**), and dotted (**F3**) lines; (J) orientation of the side chains of R340, R358, and S361 in the representative conformations from C1–C4 clusters of the free-ligand SH2 domain; (K) SH2 binding pockets (Fpocket) defined for each representative conformation from clusters C1–C4. Only pockets around the binding pocket observed in X-ray structure 2IUH are shown with its SASA and volume values.

The stable and low RMSD values (1.5–2.5 Å) observed during all replicas are typical of excellent simulation convergence. The RMSF profile of the free-ligand SH2 is akin to pep/SH2 complex. However, the highest RMSF values observed for residues located in the loop between $\beta 1$ and $\beta 2$ strands are twice as much (Figure 7C,D).

The folding (2D) and tertiary (3D) structure of the free-ligand SH2 domain is analogous to the p-pep/SH2 complex (Figures 6 and 7A). Identical to the p-pep/SH2 complex, we localised in the free-ligand SH2 domain three intrinsically disordered regions (IDRs)—**F1**, **F2**, and **F3**—manifesting: (i) a structural disorder, evidenced by transient structures reversibly converting between 3^{10} -helix; β -strand; turn; coil; (ii) a dynamical disorder apparent as the highest fluctuations; and (iii) a degree of such SH2 disorder is increased compared to the p-pep/SH2 complex (Figure 6C,E and Figure 7A–D).

The cross-correlation matrix computed on the concatenated trajectories shows more contrast and a better interpretable pattern than in the p-pep/SH2 complex. It reflects clearly (i) the coupling motion between the β -strands within a ‘core’ β -sheet structure; (ii) the weak anti-correlation between two IDRs, **F1** and **F2**; and (iii) the per block correlation (positive or negative) of each IDR, **F1**, **F2**, and **F3**, with the other SH2 structural fragments (Figure 7E).

PCA demonstrates that the first four modes describe a great majority of the variance of SH2 motion (80%), with the first mode explaining 60% (Figure 7F). The major contributor to the first mode is IDR **F1**, which showed a significant amplitude movement in the direction of the binding pocket of SH2 (Figure 7G).

The second mode shows a nearly perpendicular direction to the first vector and confirms the **F1** reciprocating motion concerning the pocket is accommodating KID p-pep in the p-pep/SH2 complex. Such flexibility facilitates binding pocket evolution upon the recognition process of p-pep. As seen from the PCA, the collective motion of the free-ligand SH2 domain differs markedly from that of the p-pep/SH2 complex. In the complex, we observed large amplitude motions of all IDRs, **F1**, **F2**, and **F3**: SH2 **F2** and **F3** motions are significantly reduced, while **F1** exhibits ample motions towards the SH2 binding pocket. The direction of **F1** movement in the free-ligand SH2 is completely different from the p-pep/SH2 complex’s.

To estimate the other p-pep binding effects, we generated the spatial distributions of the representative atoms from residues R340, R358, and S361 (for details, see Section 2.2.2) of the free-ligand SH2 domain. The oxygen (OG) atom of the highly fluctuating S361 residue showed a large distribution with a mushroom cap shape. In contrast, the low-fluctuating R358 CZ carbon atoms are concentrated in a minimal area (Figure 7H). On another side, the

R340 CZ carbon atom, belonging to the lowly fluctuating α H1-helix, manifests the most extensive spatial distribution.

Considering the capacity of arginine and serine to rotational conformational flexibility originating from their torsional degrees of freedom, we used the descriptors defined above (see Section 2.2.2). According to these descriptors, R340 in the two most populated clusters represents only 45% of all conformations, where 34 and 11% are the *ttg+t* and *tttt* rotamers, respectively (Figure S5B). Its other conformations are fully heterogeneous, showing occurrences of less than 10%. R358 rotamers are regrouped into four main clusters, each having a population of more than 10%, and comprises 83% of all MD conformations. The most populated clusters are composed of *g+ttg+* (36%), *tttt* (20%), and *g+tg-t* (16%) rotamers. Residue S361 having only one torsional degree of freedom, shows two major conformers, *g-* (45%) and *g+* (41%), while *g-* presents only 14%.

This analysis displayed that the rotational conformational flexibility of residues R340, R358, and S361 is increased in the free-ligand SH2 domain concerning *p-pep*/SH2 complex. However, the residue-related specificity is conserved. Indeed, R340 is more locally disordered in both cases, as well as only 56% (in the complex) and 45% (in the free-ligand SH2) of its rotamers regrouped in the cluster(s). In comparison, 99 and 83% of R358 rotamers, and 97 and 100% of S361 rotamers, form the well-defined clusters in the *p-pep*/SH2 complex and the free-ligand SH2, respectively. This suggests that the pY721 phosphate group may be the main factor for the limitations of rotamer diversity of R340, R358, and S361.

Free-ligand SH2 conformations regrouped using the RMSD criteria (cut-off 0.75 Å) reveal four clusters of various populations, the highest (C1, 74%), low (C2, 18%), and relatively minor (C3 and C4, 3%) (Figure 6B). Each cluster representative conformation superimposition exhibits the main differences between these conformations: (i) two IDRs, *F1* and *F2*, derived from their reversibly transiting fold and significant conformational motion, and (ii) the β -sheet formed by β 4- and β 5-strands exhibiting variability in length completed by their C-term collective movements (Figure 7G,I). Again, the RMSD-based clustering of the free-ligand SH2 conformations is identified as the third IDR, *F3*, likely to the *p-pep*/SH2 complex.

As S361 and R358 are located in IDR *F1*, they appear with two types of disorder: (i) the backbone reversible folding/unfolding co-occurring with displacement/rotation motion, and (ii) the side chain rotational disorder (local), where R340 contributes to only local rotational disorder demonstrating a large conformational space (Figure 7H,J).

Finally, the pockets found in the representative conformations of C1-C4 clusters (only pockets consistent with the binding pocket observed in the X-ray structure were considered) differ in their (i) location in SH2 domain, (ii) volume, and (iii) SASA values (Figure 7K). The largest (SASA of 295 Å²) and most voluminous (1142 Å³) pocket is observed in the conformation from C2. Its position on the SH2 surface is remarkably coherent with the X-ray structure's 2IUH. The SASA value (295 Å²) is also close to that observed in 2IUH (215 Å²). Nevertheless, the volume (1142 Å³) is three times smaller than in 2IUH (3312 Å³). The pocket of conformations from cluster C2 comprises two functional residues, R340 and R358, on its surface. Pockets found on the representative conformations from clusters C1 and C3 are localised close to the α H1-helix and characterised by a significantly reduced SASA and volume (102 Å² and 374 Å³ in cluster C1, and 194 Å² and 555 Å³ in C3). Moreover, both reduced pockets include only R340. The pocket defined on conformation from cluster C4 is located in proximity with α H2-helix. Its metrics, SASA (140 Å²), and volume (580 Å³) are also diminished, and the pocket's surface does not contain any functional residues.

Of the SH2 eight residues interacting with the *p-pep* of KID in the crystal, which portion is part of the surface pockets found in the representative conformation of each cluster? Despite the significant difference between the pockets located on the conformations of clusters C1 and C2, seven out of eight residues are localised on their pocket surface. The pocket defined on the representative conformation of cluster C3 contains only four residues from eight, and zero in C4.

Considering the pocket's size, the residues forming these pockets, and the juxtaposition of SH2 residues with those interacting with *p-pep* of KID, it seems that cluster C2 representative conformation represents an appropriate target for KID recognition/binding. Nevertheless, the large population (74%) of conformations in cluster C1 and the IDR *F1* close to the binding site may favour a better adaptation of the SH2 domain conformation to accommodate KIT KID. The SH2 binding pocket adaptability, evidenced by differences between the peptide complexes formed by *p-pep* with KIT and PDGFR, was reported in [33].

2.3. Protein–Protein Docking: Recognition and Binding of KIT KID and the PI3K SH2 Domain

The reported above MD simulation analysis of the empirical structure of 2IUH, KIT KID *p-pep*, and the PI3K SH2 domain, delivers a solid foundation to learn an essential step of RTK KIT post-transduction events—recognition and binding of an SH2 domain by KID.

2.3.1. Can the Crystallographic Structure of the Complex Formed by *p-pep* of KID and the PI3K SH2 Domain Be Reproduced by Docking?

Prior to studies of PI3K SH2 domain recognition/binding by KIT KID, we performed a bench test to investigate if docking can reproduce the empirically determined *p-pep*/SH2 complex. To begin with, the docking trials were conducted using the structure 2IUH as a benchmark set. The *p-pep*/SH2 complex was separated into unbound protein SH2 (X-ray Model, **M1**) and free *p-pep* peptide, and these isolated entities were docked with High Ambiguity Driven DOCKing (HADDOCK) [41]. Unlike other docking approaches, based on the combination of energetics and shape complementarity of studied proteins, HADDOCK uses biophysical interactions data—in our case, the H-bond contacts between *p-pep* and the SH2 domain—to drive the docking process. The *p-pep* residues were considered as ‘active centres’, while ‘passive’ residues were defined as all SH2 residues positioned within 4 Å from residues interacting with *p-pep* in structure 2IUH.

From 1000 decoys generated by rigid-body docking, 200 solutions were refined (semi-rigid docking). Analysis of these solutions by using the fraction of common contacts (FCC) clustering [42] produced four clusters, C1–C4, with different populations (Figure 8A).

Each cluster, the most populated C1 (57%) and the sparsely populated C2 (10%), C3 (8%) and C4 (5%), contains *p-pep* complex models showing differently oriented *p-pep*. Nevertheless, the great majority of them are occupying the SH2 binding pocket. In the binding pocket, the major *p-pep* position is similarly observed in structure 2IUH. However, *p-pep* orientation shows that its N-terminal is located close to either α H1-helix, similarly to the crystallographic structure 2IUH or in the opposite direction, in proximity to the α H2-helix. *P-pep* is rarely oriented along the β -strands direction (perpendicular to the major *p-pep* position). Curiously, *p-pep* orientation with its N-term located at the α H1-helix is observed in the four best solutions for the C1, C3, and C4 clusters, when only in C2 is the N-term localised at the α H2 helix (Figure 8B).

P-pep in two major orientations in the SH2 binding pocket is stabilised by the different networks formed by non-covalent interactions—H-bonds, salt bridges, and van der Waals contacts (Figure 8C,D). The non-covalent interactions pattern stabilising the *p-pep*/SH2 complex is much more crowded for the best docking solutions, which are composed in cluster C1 in respect of that of C2. Specific interactions, the salt bridges formed by pY721 with R340, R358, and S361, and H-bonds N719...N365, D723...N378 and K725...N344, are maintaining the complex with *p-pep* in two opposite orientations, and consequently, are conserved for the major *p-pep* location. The non-covalent interaction patterns observed in the representative conformation of cluster C1 and X-ray structure 2IUH are quasi-identical (recognition of pYXXM linear motif); the difference consists of some additional contacts in the ‘docked’ complex.

Consequently, docking trials showed that HADDOCK reproduces the benchmark structure. This protocol was further applied to the KIT KID^{pY721} de novo model of KIT (ligand) docking onto the PI3K SH2 domain (target) issued (i) from the empiric structure

2IUH (**Model 1, M1**) and (ii) the representative conformation of the free-ligand SH2 from cluster C1 (**Model 2, M2**).

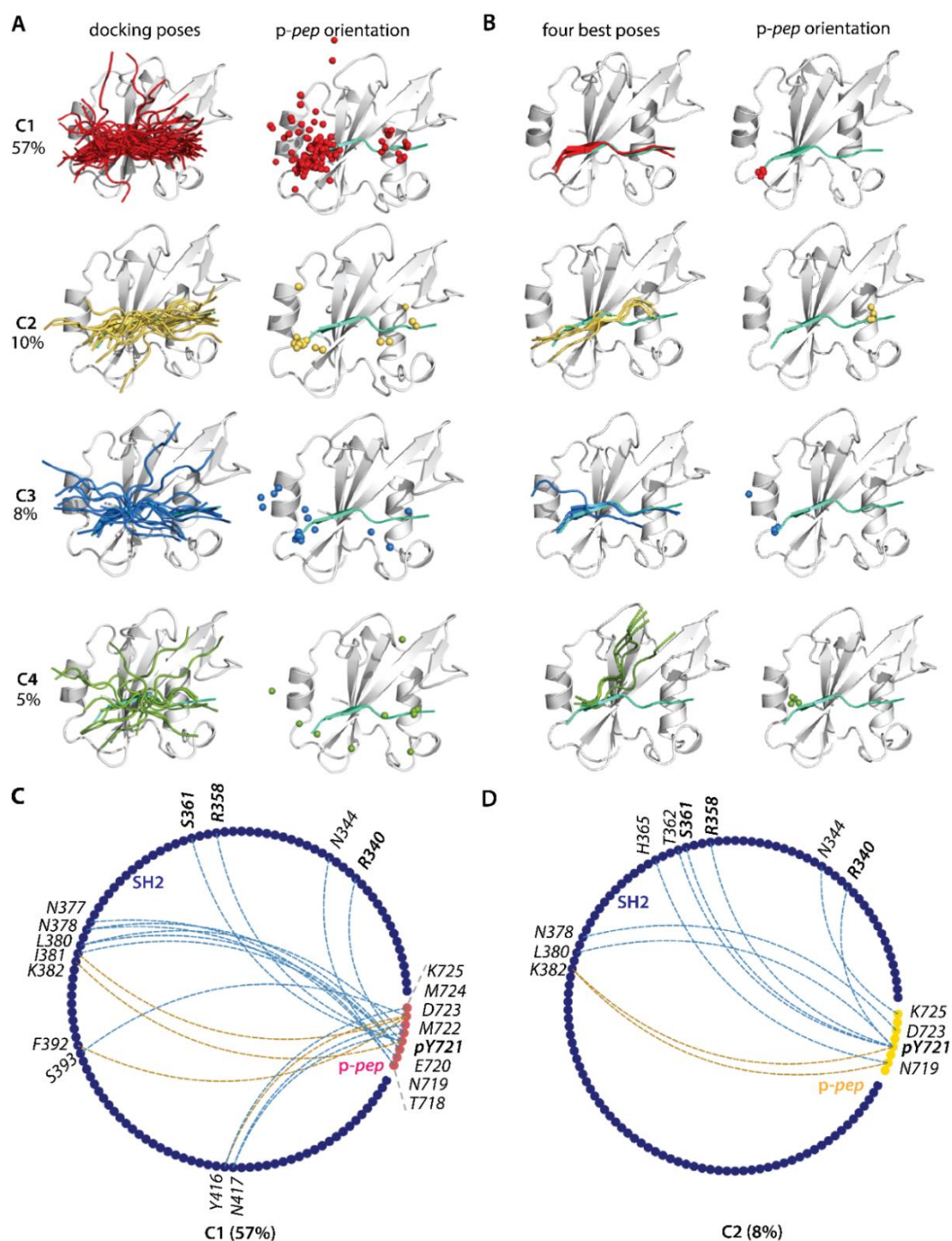


Figure 8. Computational docking of *p-pep* (ligand) onto SH2 (target) performed with HADDOCK using an information-driven method (benchmark). (**A**) Docking poses distributed into four clusters (left) showed *p-pep* orientation (right); (**B**) superimposition of the top four solutions (left) and *p-pep* orientation (right); (**A,B**) the target is a grey cartoon; the ligand is coloured per cluster. The N-terminal of *p-pep* is presented as a ball; (**C,D**) H-bonds (in blue) and hydrophobic contact (in beige) stabilising *p-pep* (in red/yellow) and the SH2 domain of PI3K (in blue) showed as a string diagram for the representative complex from cluster C1 (**C**) and C2 (**D**).

2.3.2. Docking of KID into the PI3K SH2 Domain

The bench test encouraging outcomes prompts us first to dock KIT KID^{P^{Y721}} into the SH2 domain from 2IUH (**M1**). KID^{P^{Y721}} docking into **Model 1** (1000 decoys generated by rigid-body docking) produced KID/SH2 (**M1**) complexes. The majority of retained 200 solutions (60%) were regrouped using the RMSD criteria (cut-off 0.6 Å) in seven clusters (C1–C7) with relatively low population varying from 22 to 5%, and regrouping 70% (cut-off 5%) of all solutions (Figure S8A).

The docking poses show opposed positions of KID^{P^{Y721}} (and its *p-pep*) concerning the SH2 binding pocket, but the *p-pep* major position is similar to what is observed in structure 2IUH. Nevertheless, KID^{P^{Y721}} orientation viewed by its *p-pep* is quite divergent, even within the same cluster, showing a circular-distributed orientation of KID *p-pep* around the SH2 (**M1**) binding site. After refinement of 200 docking solutions, the KID *p-pep* position in the SH2 (**M1**) binding cavity corresponds better to the X-ray structure. However, its orientation corresponds either to the structure 2IUH (C4) or appears as an opposite arrangement (C1) (Figure S8B).

The docking solutions obtained upon the docking of KID^{P^{Y721}} into the SH2 domain from 2IUH (**M1**) are discouraging. These ambiguous results may be due either to an incorrect initial structure of the one or two anchored partners, SH2 (**M1**) or/and KID^{P^{Y721}}, or to the docking algorithm.

First, we suggested that either: (i) using the 2IUH SH2 structure (**M1**) as a KID target is not a good idea; or (ii) the *p-pep* orientation in the SH2 binding site from the X-ray structure may correspond to a wrong solution due to the *p-pep*'s small length and binding site symmetry of residues showing similar biophysical properties.

Trying to explore our hypothesis, primarily a 'bad target choice', we performed a docking (always with HADDOCK) of KID^{P^{Y721}} into free-ligand SH2's most populated cluster C1 representative conformation (**M2**). Surprisingly, the docking produced similar results as those obtained for KID^{P^{Y721}} docking into **M1** (Figure S9).

The non-covalent interaction patterns in the KID^{P^{Y721}}/SH2 complex models obtained by KID docking into the two models of SH2 (**M1** and **M2**) differ (Figure 9). Moreover, this difference is the most meaningful compared to the *p-pep*/SH2 complex in solid state or water solution. Only the salt bridges formed by pY721 with residues R340, R358, and S361 are the non-covalent bonding's most conserved motifs, and were observed in the docking trials clusters C1 and C2 (benchmark) and cluster C2 of KID^{P^{Y721}}/SH2 (**M2**).

This motif is observed in the *p-pep*/SH2 complex in a solid state (structure 2IUH), in solution (MD conformations), in benchmark docking results, and in the **M2** KID^{P^{Y721}}/SH2 complex. This highly conserved motif of the non-covalent bonding stabilising KID^{P^{Y721}} *p-pep* and full-length KID^{P^{Y721}} would be helpful for constrained docking or dynamics simulation. Contrary to the benchmark (*p-pep* docking to 2IUH SH2 domain), pYXXM KID linear motif recognition with M724 stabilisation with SH2 is missing in the docking solutions of both **M1** and **M2**.

The buried surface area, which measures the size of the interface in a protein–protein complex [43], was calculated for complexes *p-pep*/SH2 and KID^{P^{Y721}}/SH2. Comparing only *p-pep*/SH2, the mean buried surface area (MBSA) of the complex obtained by the re-docking of the X-ray structure 2IUH (benchmark trials) is globally greater in the first six clusters (1–6). In contrast, the other clusters' MBSA values are almost identical within the standard deviations (Figure 9C). KID^{P^{Y721}}/SH2 MBSA values from the first two clusters (1–2) are nearly similar between models **M1** and **M2**, while in cluster 3 its values are more significantly different. Indeed, **M2**'s MBSA is greater than **M1**'s. This result is not unexpected because the total amount of surface area buried within its fold is tightly coupled to the overall flexibility of a protein [44]. Moreover, as KID is an intrinsically disordered protein, its vast conformational landscape makes the identification of favourable states for SH2 recognition and docking challenging.

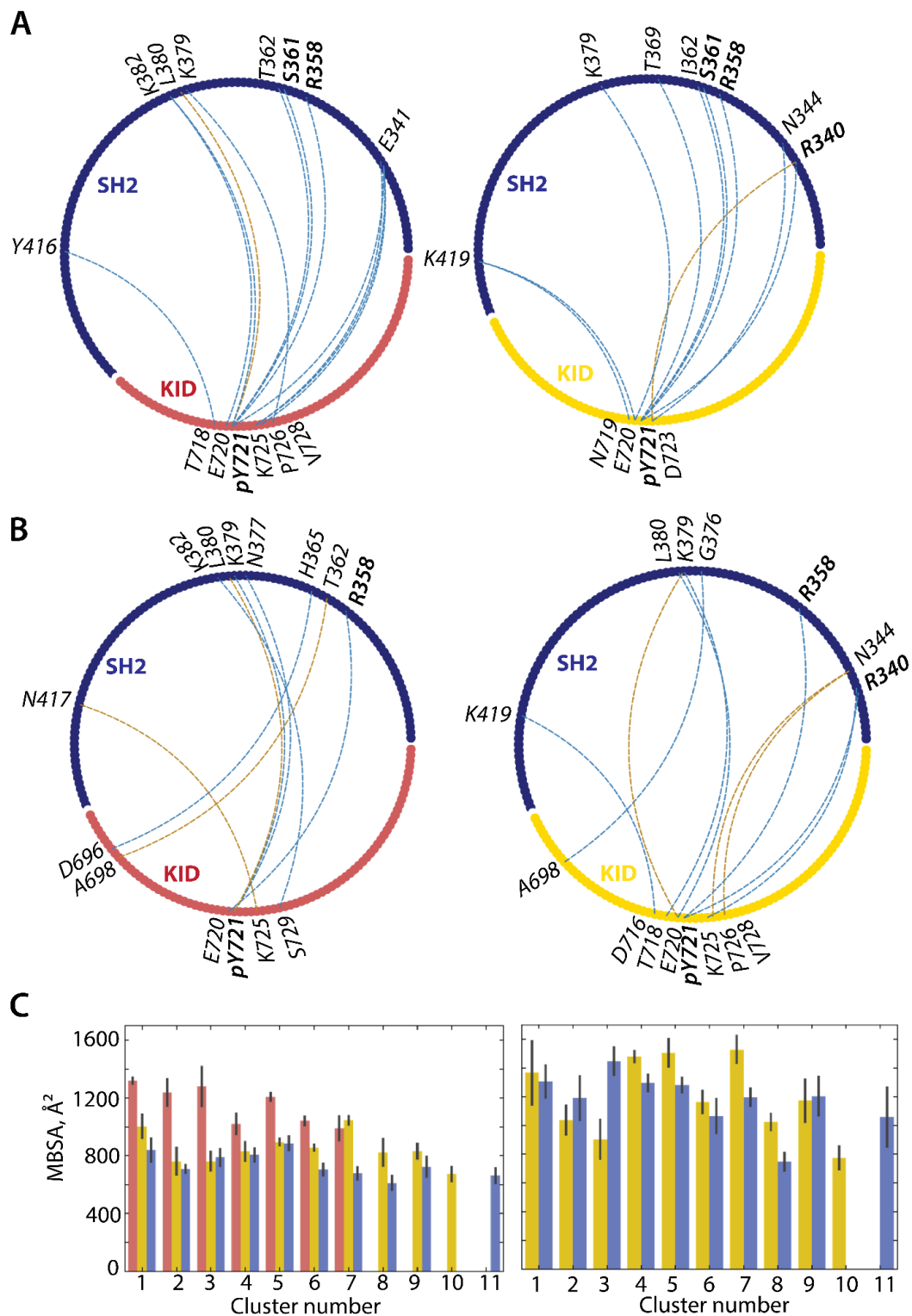


Figure 9. Analysis of KIT KID^{Y721}/PI3K SH2 models obtained by computational protein–protein docking. (A,B) H-bonds (in blue) and hydrophobic contacts (in beige) stabilising KID (red/yellow area) and PI3K SH2 domain (dark blue area) in KID^{Y721}/SH2 (M1) (A), and KID^{Y721}/SH2 (M2) (B) docking solutions for the representative conformations from clusters C1 (left) and C2 (right) are shown as a string diagram; (C) p-pep buried surface area calculated for complexes p-pep/SH2 (left) and KID^{Y721}/SH2 (right). Models of p-pep/SH2 are in red, KID^{Y721}/SH2 (M1) are in yellow, and KID^{Y721}/SH2 (M2) are in blue. Protein–protein docking was performed with HADDOCK using an information-driven method.

As the protein–protein binding ambiguous results may be also produced by the method applied, it is worthwhile to note that HADDOCK quantitative measures of binding modes—number of clusters, score and population—do not allow comparisons between docking solutions reflecting conformational and orientational characteristics, even though a simple superposition of the found solutions showed that a reference (X-ray) pose was observed.

2.3.3. Intuitive User-Guided Modelling of Molecular Complex KID^{pY721}/SH2

To resolve the KID^{pY721}/SH2 complex building problem and to take into account KID's inherent flexibility, we applied an alternative approach previously used for the construction of the protein–protein complex formed by vitamin K epoxide reductase (VKOR) and its redox protein, the protein disulfide isomerase (PDI) [45]. KID^{pY721}/SH2 complex 3D models were built using the KID *p-pep*/SH2 crystallographic structure as a reference for the KID^{pY721} initial positioning relative to the SH2 domain.

To be the most objective in KID^{pY721}/SH2 complex modelling, the structure 2IUH was not used as a template because of (i) the suggested alternative KID^{pY721} position/orientation in the SH2 binding pocket, (ii) KID and free-ligand SH2 high structural/conformational variability, and (iii) a considerable difference between *p-pep* and KID^{pY721} size.

For modelling the KID^{pY721}/SH2 complex and to bring the two proteins as close as possible, a conformation of KID^{pY721} having an excellent TNEYMDMK fragment similarity to the empirical structure 2IUH *p-pep* (minimal RMSD values) was chosen as the initial structure. As for the initial SH2 model, the most populated cluster C1 representative conformation from free-ligand SH2 was chosen and positioned at KID^{pY721} so that (i) the distance between KID pY721 phosphorus atom and SH2 CZ R340, R358 Z atoms and S361 OG atom was at least 10 Å; and (ii) KID^{pY721} was alternatively placed above the middle of SH2 binding pocket, with the TNEYMDMK fragment oriented (a) similarly to *p-pep* in structure 2IUH and (b) in the opposite direction, as evidenced by HADDOCK docking.

The obtained proto-models of KID^{pY721}/SH2 complex, **CM1** and **CM2**, were explored using the Gaussian accelerated molecular dynamics (GaMD) simulation [46,47] with restraints applied to the distances between the phosphorus atom (KID^{pY721}) and nitrogen/oxygen atoms of R340, R358, and S361 (SH2) (see Methods). Restraint distances were gradually diminished during a stepped 350 ns GaMD simulation, then removed entirely (Figure 10A).

After constraints relaxation, the inter-protein distances between the phosphorus atom from KID pY721 and SH2 nitrogen atoms from R340 and R358 are well conserved until the end of the MD simulation (500 ns) in both models. However, the contact between the KID pY721 phosphorus atom and the SH2 S361 oxygen atom shows large variations in two models. Curiously, this distance fell regularly to the value observed before all constraints were relaxed, and even lower.

The RMSD curves and values of **CM1** and **CM2** models, as well as each protein that composes these models, display similar profiles (Figure S10A). The RMSD of SH2 is absolutely stable in both models, and the principal contributor to the increase of RMSD values is KID^{pY721}. As expected, KID^{pY721} showed an overall decrease in fluctuations, except for the N/C-ends, compared to free KID^{pY721} (Figure S10C).

The SH2 folding in complex KID^{pY721}/SH2 is globally well-conserved in both models, except for a coiled β 1 and β 2 stands linker, recognised previously as IDR *F1*. However, the character of the *F1* reversible folding in the complex most resembles the *p-pep*/SH2 complex compared to the free-ligand SH2 (Figure 6E, Figure 7B, and Figure S10B). KID^{pY721} folding in both models, **CM1** and **CM2**, shows intrinsic disorder in a large portion of protein, except a stable α H1-helix, similar to free KID^{pY721}. Nevertheless, some distinguishable differences are likely complex-specific. In that manner, KID^{pY721} showed that (i) in **CM1**, the small β -sheet (T718–V728) is well conserved during simulation with and without constraints, while in **CM2**, it was transformed in the systematically unfolded state (turn; coil) in the unconstrained GaMD simulation; (ii) in **CM1**, the reversible H5 (3¹⁰-helix; turn; coil) is

folded as a well-stable α H5-helix in the range of 230–500 ns, while in **CM2**, this fragment shows the highly reversible structure. Focusing on the secondary structures' evolution of KID *p-pep* during GaMD, we noted that after removing the restraints, its folding is drastically decreased in **CM1** (random coil) and increased in **CM2**.

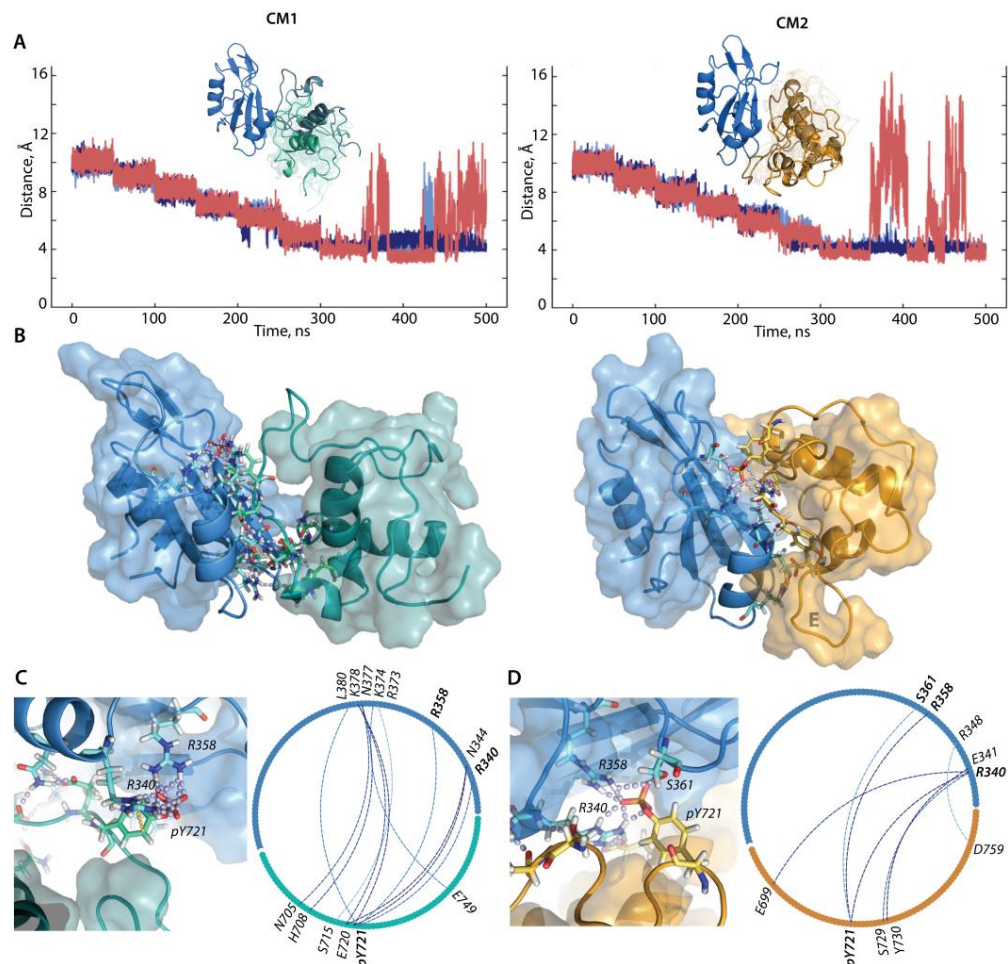


Figure 10. Intuitive modelling and GaMD simulation of molecular complex KID^{P^{Y721}}/SH2, represented by two alternative models in which KID TNEYMDMK peptide is oriented similarly to *p-pep* in structure 2IUH (**CM1**, left) and the opposite direction (**CM2**, right). (A) Distance variations between the phosphorus atom (KID^{P^{Y721}}) and nitrogen/oxygen atoms of R340 (blue light), R358 (dark blue), and S361 (SH2) (red). Each model of non-covalent complex and its time-dependent evolution (snapshots taken at $t = 0, 75, 125, 175, 225, 275, 325$, and 500 ns) is shown as insert. In light colour, the starting structure, and in darker colour, the final conformation. Intermediate conformations are shown as a transparent cartoon; (B) 3D models of complex KID^{P^{Y721}}/SH2 at 500 ns of GaMD simulation; (C,D) inter-protein H-bonds stabilising KID (teal/orange) and PI3K SH2 domain (blue) in **CM1** (C) and **CM2** (D) are shown as 3D structure fragment (left) and a string diagram (right) displaying the high ($\geq 80\%$, in dark blue) and low ($\leq 50\%$, in blue light) probability of these contacts. Proteins are shown as surfaced cartoons, KID in teal (**CM1**) and orange (**CM2**), and SH2 in blue. The interface residues are shown as sticks, non-covalent contacts as dashed lines, lilac for H-bond and yellow for van der Waals contacts.

Intermolecular contacts formed during the unconstrained simulations showed that KID^{P^{Y721}} and SH2 in the **CM1** model are linked by multiple H-bonds (11 contacts). With high probability (with occurrence $\geq 80\%$, six H bonds), or low probability (with occurrence between 30 and 50%, five H bonds), they together form a very large protein–protein interaction interface (Figure 10B,C).

As expected, the major contributor to the **CM1** interface contact network is two salt bridges formed by pY721 (KID^{pY721}) with R340 and R358 (SH2). These salt bridge interactions pulled together the disordered region of KID^{pY721} and two SH2 regions—the dynamically disordered linker *F1* (through the interaction with R358), and the SH2 α H1-helix (by interaction with R340). Surprisingly, these salt bridges are low probability events, and contact pY721...S361 is nearly disappeared (probability $\leq 30\%$). The salt bridge interactions are completed by H-bonding of neighbour residues to pY721, E720, and T718, each acting as a bifurcated centre. As shown, E720 interacts with N344 and K379, and T718 bounds N344 and N377. The H-bonds formed by KID^{pY721} residues N705 and H708 (from the stable α H1-helix) with N377 and K379 (from the β -sheet core of SH2), produce the second interface area of contacts, spatially separated from those formed by pY721 and its closest residues.

The **CM2** KID^{pY721}/SH2 interface is formed by a limited number of H-bonds (seven contacts) and substantially restricted the number of residues contributing to binding (R340–R361 from SH2 and E699–Y730 from KID) as well as structural elements involved (Figure 10B,D). First, KID pY721 interacts with R340, R358 (the high probability contacts), and S361 (the low probability contact) of the SH2 domain. This principal binding interactions motif between KID^{pY721} and SH2 is completed by H-bonds between (i) R340 (α H1-helix of SH2) with E699 (α H1-helix of KID^{pY721}) and S729 (disordered H3 of KID^{pY721}), making R340 a trifurcate centre, and (ii) E341 (α H1-helix of SH2) with Y730 (disordered H3 of KID^{pY721}). These contacts showed that in **CM2**, the recognition between two proteins, KID^{pY721} and SH2, is maintained by strong and stable interactions formed by two salts bridges and crosswise H-bonds involving the limited protein regions.

We also note the reduced size, estimated by the radius of gyration (R_g), of the KID^{pY721}/SH2 complex represented by the **CM2** model with respect to **CM1** (Figure S10D).

3. Discussion

We attempt to describe the RTK KIT INTERACTOME by modelling a set of macromolecular complexes formed by KIT with its cellular proteins partners (PPs) involved in signal transduction. To initiate this modelling, we built a 3D model of the first molecular complex formed by the most phosphotyrosine-rich kinase insert domain (KID) of RTK KIT with a phosphatidylinositol 3 kinase (PI3K) SH2 domain binding preferentially to KID.

Many interacting protein partners, RTK KIT and PI3K, are intrinsically disordered proteins (IDPs) with a modular structure. Indeed, as we previously reported, the multidomain RTK KIT comprises the sub-domains (JMR, TKD, KID, and C-term), whether structurally well-ordered or intrinsically and extrinsically disordered [9,10,34]. Similarly, PI3K from the non-receptor tyrosine kinases Src-family, constitutes a typical example of modular architecture: an amino-terminal SH3 and SH2 domains, flanking a kinase domain by intra-molecular SH3-binding and SH2-binding sites [24,48].

The modular architecture of protein structures is advantageous for a more efficient execution of their functional activity (e.g., allosteric regulation of protein–protein interactions, involved in cell signalling). Some parts of such interaction interfaces participate in the information transfer (inter-protein communication), while other interaction regions appear to contribute only binding affinity (switching) [49–51].

It is well known that ~40–60% of the human proteome appears to be composed of protein domains/regions that are intrinsically disordered [8,52,53]. IDPs are paradigmatic challenges because: they are disordered in their inactive state [52–54], can fold partially or fully upon the biological effectors binding [8,55], can bind selectively diverse partners [56,57], and exhibit allosteric regulation without a well-defined quaternary or even tertiary structure [58,59].

These three fundamental properties—modularity, intrinsic disorder, and allostery—provide proteins with a finely regulated molecular mechanism that illustrates how nature can govern cellular signalling [7,8,26,58,60,61].

Each module of multidomain proteins studied, KIT and PI3K, may possess structural, dynamic, and functional independence, as was evidenced for KIT KID [34] and SH2 [62]. To prepare the structural support for KID/SH2 complexes building, we first focused on each partner, the phosphorylated KID and SH2 domain.

As the phosphate is an effector, its covalent binding to KID may promote significant effects on this intrinsically disordered domain. It was earlier reported that multisite phosphorylation gradually tunes the affinity of the inhibitor SIC1 for the disordered cyclin-dependent kinase CDC4 [63], and a similar mechanism had earlier been suggested for the beta-catenin/E-cadherin complex [64].

Our systematic study of phosphorylation-induced effects on KIT KID showed that the one-site, two-site, and three-site phosphorylation considerably affected the KID structure. We evidenced it as (i) an increase in the overall p-KID folding or additional folding in the phosphorylated tyrosines' immediate environment, and (ii) an alternation of p-KID's flexibility. However, regardless of the phosphorylation site position and the number of phosphorylated tyrosine residues, all p-KIDs are intrinsically disordered entities holding the inherent coupled dynamics specific for each p-KID. Focusing on the phosphorylation sites, their motions are also site-specific. In particular, Y703 motion is correlated with only a minimal number of the sequence adjacent residues, positively and negatively with preceding and following residues, respectively. In contrast, Y721 and Y730 show strong correlations with many residues located close to Y721 or Y730 (strong positive correlations) or distant residues (strong or moderate negative correlations). Despite evident structural and dynamical alterations in p-KIDs, their globular-like shape, maintained by an extended H-bonds network pattern, is universally conserved. The solvent-exposed phosphotyrosine residues' position in all p-KIDs—a primary determinant of local residue flexibility [44]—and their sidechains' extended spatial distribution allow each phosphate group availability for post-transduction events with high probability.

This modelling of the phosphorylated KID generates an exhaustive number of well-characterised targets, which opened a route for describing KID INTERACTOME.

The tyrosine phosphorylation of RTK KIT generates the necessary conditions to recruit and activate downstream signalling proteins through pY binding to their SH2 domains. Consequently, we focused on phosphatidylinositol-3 (PI3K) kinase N-terminal SH2 domain, which preferentially binds to KIT KID via the phosphorylated tyrosine pY721 [65,66].

To deliver the appropriate molecular entity that specifically binds KID^{pY721}, the available crystallographic structure 2IUH—containing the PI3K N-terminal SH2 domainco-crystallized with the KIT KID phosphorylated peptide TNEYMDMK (p-pep)—was used as an initial cornerstone.

The archetypal structure of the SH2 domain is well conserved in a solid state and an aqueous solution in both SH2 forms, bound to p-pep and free-ligand entity. P-pep localisation, corresponding to the canonical mode of SH2 bonding [39], is maintained in the crystal by multiple non-covalent interactions involving all p-pep residues without exception. In an aqueous solution, the number of p-pep contacts with SH2 is significantly decreased. However, in both cases, pY721 of p-pep is engaged in multi-branching hydrogen and ionic bonding with R340, R358, and S361 of the SH2 domain, forming salt bridges. In a solvent, the formation of salt bridges is mainly due to entropy, usually accompanied by unfavourable ΔH contributions due to the desolvation of interacting ions upon association [67]. Due to many ionisable side chains of amino acids present in a protein, the pH in which it is placed is crucial for its stability. However, interactions between SH2 and KID M724 of the recognition linear motif pYXXM were not conserved during simulation.

This decrease of non-covalent contacts in solution is related to (i) the three intrinsically disordered regions (IDRs *F1*, *F2*, and *F3*) of SH2, which manifest either a structural disorder, evidenced by transient structures reversibly converting between 3_{10} -helix; β -strand; turn; coil, and/or a dynamical (conformational) disorder, and (ii) the high flexibility of p-pep N- and C-terminals. In the free-ligand SH2, the degree of such disorder is considerably increased compared to the p-pep/SH2 complex. We also established that, in this state,

SH2 coupled motion with the other SH2 structural fragments is increased: between (i) the β -strands of a 'core' β -sheet structure, (ii) between IDRs, and (iii) between each IDR. In free-ligand SH2, residues R340, R358, and S361, contributing to *p-pep* binding in the complex, appear in different types of disorder: R358 and S361 as located on the IDR showing backbone reversible folding/unfolding co-occurring with displacement/rotation motion and the sidechain rotational disorder (local), while R340 contributes to only local rotational disorder demonstrating a large conformational space.

These detailed characterisations of KID and SH2, the structural and functional domains of protein partners in cell signalling, KIT and PI3K, conceives a solid foundation for building their molecular complex. Prior to the docking of these domains, the bench test performed on the empirically determined structure 2IUH indicates a good reproducibility of the crystallographic solution. Such result was obtained with High Ambiguity Driven protein–protein DOCKing (HADDOCK) [41], which uses biophysical interactions data—in our case, the H-bond contacts between *p-pep* and SH2 domain—to drive the docking process.

The benchmark structure successful reproducibility encouraged the HADDOCK docking of KID into the SH2 domain, represented by the SH2 structure from 2IUH and the most probable MD conformation of the free-ligand SH2.

Independently from the SH2 structure, the obtained docking solutions show different positions of KID^{pY721} (and its *p-pep*) with respect to the SH2 binding pocket. Although the major *p-pep* docking position in the SH2 binding pocket matches its position in the structure 2IUH, KID^{pY721} shows a circular-distributed orientation of its *p-pep* around the SH2 binding site. These discouraging docking solutions may be attributed to (i) HADDOCK, which uses the quantitative measures of binding modes, such as the number of clusters, score, and population that do not allow comparisons between solutions reflecting conformational and orientational characteristics, (ii) the incorrectness of the X-ray solution, or (iii) the basic difference in the binding of peptide and protein to a target.

We suggest that *p-pep* can be located and oriented differently compared to its crystallographic structure pose. On one side, the *p-pep* pseudo symmetry concerning pY721, is issued from the biophysical properties' similarity and this fragment's high inherent conformational flexibility in a solution. On the other side, the double pseudo symmetry of the SH2 binding pocket is due to the similar biophysical properties of the pocket surface residues. Moreover, it is well known that peptides exhibiting either extended conformation or adopting β -turn or α -helix as a motif for target recognition can be completely buried in cavities, making multiple high-affinity interactions with a target [68], while the interaction interface between the two proteins is limited and involves significantly fewer amino acids from each partner, usually defined as 'hot-spot' residues and making the largest contributions to complex formation [69–71].

We suggested that to build the molecular complex composed of two intrinsically disordered proteins, which couple folding and binding possesses [72], an alternative strategy may be used. The empirically resolved structure 2IUH presenting a PI3K SH2 domain co-crystallised with a *p-pep* fragment of KIT KID, and the high conservation of the pY721 binding with residues R340, R358, and S361 in the solid state and water solution can be used for the intuitive user-guided building of the p-KID/SH2 complex as reference supports. Taking into account the high conformational variability (structural instability and flexibility) of both proteins, KID and SH2, in solution, the MD conformations of KID^{pY721} and free-ligand SH2 are the most appropriate starting structures in such a study.

A direct use of the X-ray 2IUH structure is not a dogma for modelling the KID^{pY721}/SH2 complex, but it can serve as a reference for the initial positioning of KID^{pY721} concerning SH2.

Two models of the KID^{pY721}/SH2 complex with KID^{pY721} positioned in front of the SH2 cavity, but alternatively oriented, either coherent to the *p-pep* orientation in structure 2IUH (**CM1**) or oppositely oriented (**CM2**), were further explored by accelerated (GaMD) simulations. The preliminary results showed that in both probed models, the proteins are bonded by a combination of salt bridges and hydrogen bonds, when approaching different domains from both proteins. These inter-domain interactions create a small binding cleft,

including a few residues in model **CM2**, while in **CM1** the inter-protein interface represents an area twice larger on each protein and spans widely spaced amino acids in protein sequences and structures.

Although a very compact and regular **CM2** interface model as well as an increased helical folding of the KID T718–723 fragment and the reduced size of the KID/SH2 complex are very attractive arguments for the choice of this model as functionally related, there are still doubts in such a conclusion.

For an objective assignment of the functionally related model, we are now engaged in extended unconstrained MD simulations (2–3 μ s) of both models to generate necessary and sufficient data for detailed comparative analysis of their structural, dynamical, and recognition properties as well as an accurate estimation of the binding energy.

Returning to the interacting proteins, RTK KIT and PI3K, regarded as modular disordered proteins, it seems that molecular modelling and molecular dynamics simulations provide powerful tools for the exploration of such type of proteins and their complexes. Such study will be most effective when analysed in close conjunction with experiments on a protein function, which would play an essential role in validating and improving the modelling and simulations.

4. Materials and Methods

4.1. 3D Modelling

P-KID models. The initial 3D model of KID (sequence F699–D768, Uniprot P10721) was taken at 2 μ s of MD simulation of the inactive KIT model reported in [10]. Tyrosine residues were phosphorylated by adding the phosphate group $-O-PO_3^{2-}$ with PyMOL Builder module [73,74].

The active KIT^{p568/p570}/2Mg²⁺/ATP. The initial 3D model of the activated KIT (sequence I516–R946, Uniprot P10721) was constructed using the crystallographic structure 1PKG (resolution 2.9 Å) [75] after 2Mg²⁺/ATP docking with Autodock Vina [76] into the kinase domain, and the model of the inactive KIT completed with transmembrane domain and C-terminal tail, as reported in [10].

The PI3K regulatory subunit p85 α N-terminal SH2 (N-SH2) domain. The template structure of p85 α (N)-SH2 domain complexed with a KID^{pY721} peptide [33] was retrieved from the PDB (PDB: 2IUH, resolution 2.0 Å) [32] and used to model the free-ligand (N)-SH2 domain (sequence E332–S429, Uniprot P27986), further referenced as SH2 for simplicity.

Five thousand models of (i) p-KID – KID^{pY703}, KID^{pY721}, KID^{pY730}, KID^{pY703/pY721}, KID^{pY703/pY730}, KID^{pY721/pY730}, KID^{pY703/pY721/pY730}; (ii) active KIT^{p568/p570}/2Mg²⁺/ATP; and (iii) SH2 were generated with Modeller 10.1 [77]. The best models were assessed using the DOPE score [78] and Procheck [79].

KID^{pY721}/SH2 complex. Molecular complex of KID^{pY721} (model) with SH2 (model, cluster C1 representative conformation) was modelled using the structure of the SH2 domain complexed with a KID^{pY721} peptide [33] as a reference for the initial KID positioning with respect to the SH2 domain. A conformation of KID^{pY721} having an excellent similarity of its TNEYMDMK fragment with p-pep from the structure 2IUH was chosen as the initial structure of KID^{pY721}. KID^{pY721} was placed in two orientations of its TNEYMDMK fragment with respect to SH2, with p-pep N-terminal at the α H1- and α H2 helix, respectively. In both cases, KID^{pY721} was positioned in front of the SH2 binding pocket. The initial distance between the p-atom from pY721 of KID and CZ and OG atoms of R340, R358, and S361 of SH2 in each built complex was at least 10 Å.

4.2. Molecular Dynamic Simulation

System set-up. The systems were prepared with the LEAP module of AMBER 20 (<http://ambermd.org/AmberTools.php>; accessed on 17 June 2021), using the ff19SB (phosaa19SB and phosaa19SB) all-atom force fields [80,81] for phosphorylated KID and active KIT. The latter was inserted in a phosphatidylcholine (POPC) lipid bilayer using Charm-Gui membrane and prepared with the additional lipid17 and ATP force fields [82]. Then, (i) hydrogen (H)

atoms were added, (ii) covalent bond orders were assigned, (iii) protonation states of amino acids were assigned based on their solution for pKa values at neutral pH, (iv) histidine residues were considered neutral and protonated on ϵ -nitrogen atoms, and (v) Na⁺ counterions were added to neutralise the protein charge. All systems studied were solvated with explicit TIP3P water molecules in a periodic octahedron box with at least a 12 Å distance between the protein and the boundary of the water box.

Equilibration. The set-up of the systems was performed with the SANDER module of AMBER 20. First, p-KID and free-ligand SH2 were minimised using the steepest descent and conjugate gradient algorithms as follows: (i) 10,000 minimisation steps where the water molecules have fixed protein atoms, (ii) 10,000 minimisation steps where the protein backbone is fixed to allow protein side chains to relax, and (iii) 10,000 minimisation steps without any constraint on the system. After relaxation, the systems were gradually heated from 10 to 310 K at a constant volume using the Berendsen thermostat [83] while restraining the solute C α -atoms by 10 kcal/mol/Å². After that, the systems were equilibrated for 100 ps at constant volume (NVT), and a further 100 ps at constant pressure (NPT) maintained by the Monte Carlo method [84]. A 100 ps NPT run achieved final system equilibration to assure that the water box of the simulated system had reached the appropriate density. Active KIT was equilibrated using the same protocol as in [10].

Molecular dynamics simulations. All MD trajectories were produced with the PMEMD module of AMBER 20 (GPU-accelerated versions) and the supercomputer JEAN ZAY at IDRIS (<http://www.idris.fr/jean-zay/>, accessed on 27 December 2022).

Three trajectory replicas of (i) 2 μ s for each p-KID equilibrated system and active KIT and (ii) 500 ns for SH2 were generated with an integration time step of 2 fs. The particle mesh Ewald (PME) method, with a cut-off of 10 Å, was used to treat long-range electrostatic interactions at every time step and bonds involving hydrogen atoms were constrained with the SHAKE algorithm [85]. The van der Waals interactions were modelled using a 6–12 Lennard-Jones potential. The initial velocities were reassigned according to the Maxwell–Boltzmann distribution. Coordinates were recorded every 1 ps.

In all p-KID equilibration and production steps, the N- and C-terminal C α -atoms were restrained to a 10 ± 0.2 Å interatomic distance with a weight of 20 kcal/mol to mimic their native measure typical of a full-length KIT kinase domain [20].

GaMD simulations. To estimate the parameters needed for the Gaussian accelerated molecular dynamics (GaMD) simulation [46,47], 50 ns of cMD trajectories (one for each model CM1 and CM2) were used with the following distance constraints: 10 ± 0.2 Å with a weight of 20 kcal/mol. Then, the 500 ns GaMD trajectories of the relaxed systems were generated, using as starting conformations the cMD conformations of the respective forms taken at $t = 50$ ns. Every 50 ns, the interatomic distance constraints between P of KID^{pY721}, CZ, and OG of R340, R358, and S361 of SH2 were reduced by 1 Å (with the same standard deviation) from 10 to 4 Å. The boosting was applied of both total and dihedral potential energies. The boosting energy threshold was set as the maximal total potential energy calculated during the cMD. The coordinates were recorded every 1 ps.

4.3. Data Analysis

All standard analyses of all protein trajectories were performed using the CPPTRAJ 4.25.6 program [86] of AmberTools20. Analysis of the protein MD conformations (every 10 ps) was realised after least squares fitting on a reference structure to remove rigid-body motions.

(1) The RMSD and RMSF values and cross-correlations were calculated for the C α -atoms using the initial model (at $t = 0$ ns) as a reference.

(2) Secondary structural propensities for all residues were calculated using the define secondary structure of proteins (DSSP) method [87].

(3) SH2 clustering analysis was performed on the productive simulation time of each MD trajectory using an ensemble-based approach [88]. The algorithm extracts representative MD conformations from a trajectory by clustering the recorded snapshots according

to their C α -atom RMSDs. The procedure for each trajectory can be described as follows: (i) a reference structure is randomly chosen in the MD conformational ensemble, and all conformations within an arbitrary cut-off r are removed from the ensemble; this step is repeated until no conformation remains in the ensemble, providing a set of reference structures at a distance of at least r ; (ii) the MD conformations are grouped into n reference clusters, based on their RMSDs from each reference structure. The tested cut-offs were 0.5, 0.75, 1, 1.5, and 2 Å. The analysis was performed every 10 ps after SH2 fitting on the β -core (G353-R358, Y7368-R373, N378-F384) C α -atoms at $t = 0$ ns.

(4) The H-bonds between the donor (D) and acceptor (A) atoms N, O, and S, satisfying the following geometrical parameters: $d(D-A) \leq 3.6$ Å, DHA angle $\geq 120^\circ$, were monitored. Van der Waals contacts were considered for all residues, with the side chains carbon atoms within 3.6 Å of each other.

(5) The principal components analysis (PCA) modes were calculated for all non-hydrogen atoms after least squares fitting on the average conformation calculated on the concatenated data [89]. The eigenvectors were visualised with the NMWiz module for VMD [90].

(6) SH2 dynamic correlations were calculated after the fitting of MD conformations on its β -core structure (G353-R358, Y7368-R373, N378-F384) C α -atoms at $t = 0$ ns.

(7) Binding pockets were found and analysed using the Fpocket program [40].

(8) Molecular docking of KID^{PY721} into SH2 was performed with HADDOCK 2.2 webserver [91] using (i) active residues R340, N344, R358, S361, N377-L380, and N417 for SH2, and (ii) T718-K725 for KID^{PY721}. Passive residues were assigned within 4 Å of the active residues. One thousand complexes were generated, and the two hundred best docking solutions clustered using the FCC method (fraction of common contacts), and a cut-off of 0.6 Å. Clusters with a minimum size of four were kept for the refining step.

4.4. Visualisation and Figure Preparation

Visual inspection of the conformations and figure preparation were performed with PyMOL [74]. The VMD 1.9.3 program [90] was used to prepare the protein MD animations. To visualise the motions along the principal components, the Normal Mode Wizard (NMWiz) plugin [92,93], which is distributed with the VMD program, was used. The three-dimensional representations of the free energy surface were plotted using Matlab (US, © 1994–2021 The MathWorks, Inc., Natick, MA, USA).

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/kinasesphosphatases1010005/s1>, Figure S1: KID of RTK KIT in its active and inactive states; Figure S2: Conventional MD simulations of pKID; Figure S3: Secondary structure (SS) for unphosphorylated and phosphorylated KID; Figure S4: The representative atoms of arginine and serine; Figure S5: Conformational features of the PI3K SH2 domain; Figure S6: Conformational features of the PI3K SH2 domain in p-pep/SH2 complex and free-ligand SH2; Figure S7: Conventional MD simulations of the free-ligand SH2 domain; Figure S8: The protein–protein docking (HADDOCK) of KID onto SH2 (X-ray structure); Figure S9: The protein–protein docking (HADDOCK) of KID onto SH2 (MD conformation); Figure S10: GaMD simulations of CM1 and CM2 models of KID/SH2 molecular complex; Table S1: KID conformational flexibility described by the RMSF values.

Author Contributions: Conceptualisation, L.T.; methodology, J.L. and L.T.; software, J.L.; validation, J.L. and L.T.; formal analysis, J.L.; investigation, J.L.; resources, J.L.; data curation, J.L.; writing—original draft preparation, L.T.; writing—review and editing, J.L. and L.T.; visualisation, J.L.; L.T.; supervision, L.T.; project administration, L.T.; funding acquisition, L.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, FRANCE (scholarship J.L.)

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The numerical model simulations upon which this study is based are too large to archive or to transfer. Instead, we provide all the information needed to replicate the simulations. The models' coordinates are available from L. Tchertanov at ENS Paris-Saclay.

Acknowledgments: This research was supported by Centre National de la Recherche Scientifique (CNRS), Institut Farman and Ecole Normale Supérieure (ENS) Paris-Saclay. The authors were granted access to high-performance computing (HPC) resources at the French National Computing Centre CINES (DARI A0070710973) by GENCI (Grand Equipement National de Calcul Intensif). Calculations were performed on the Jean Zay cluster at IDRIS (101063).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **2000**, *103*, 211–225. [[CrossRef](#)] [[PubMed](#)]
- Schlessinger, J. Receptor tyrosine kinases: Legacy of the first two decades. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a008912. [[CrossRef](#)] [[PubMed](#)]
- Lemmon, M.A.; Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **2010**, *141*, 1117–1134. [[CrossRef](#)] [[PubMed](#)]
- Olsen, J.V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127*, 635–648. [[CrossRef](#)]
- Dunker, A.K.; Brown, C.J.; Lawson, J.D.; Iakoucheva, L.M.; Obradović, Z. Intrinsic disorder and protein function. *Biochemistry* **2002**, *41*, 6573–6582. [[CrossRef](#)]
- Dunker, A.K.; Cortese, M.S.; Romero, P.; Iakoucheva, L.M.; Uversky, V.N. Flexible nets—The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **2005**, *272*, 5129–5148. [[CrossRef](#)]
- Bondos, S.E.; Dunker, A.K.; Uversky, V.N. Intrinsically disordered proteins play diverse roles in cell signaling. *Cell Commun. Signal.* **2022**, *20*, 20. [[CrossRef](#)]
- Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29. [[CrossRef](#)]
- Ledoux, J.; Tchertanov, L. Receptor Tyrosine Kinase KIT: A New Look for an Old Receptor. In *Bioinformatics and Biomedical Engineering, Proceedings of the 9th International Work-Conference, IWBBIO 2022, Maspalomas, Gran Canaria, Spain, 27–30 June 2022*; Springer International Publishing: Cham, Switzerland, 2022; pp. 133–137.
- Ledoux, J.; Trouvé, A.; Tchertanov, L. The Inherent Coupling of Intrinsically Disordered Regions in the Multidomain Receptor Tyrosine Kinase KIT. *Int. J. Mol. Sci.* **2022**, *23*, 1589. [[CrossRef](#)]
- DiNitto, J.P.; Deshmukh, G.D.; Zhang, Y.; Jacques, S.L.; Coli, R.; Worrall, J.W.; Diehl, W.; English, J.M.; Wu, J.C. Function of activation loop tyrosine phosphorylation in the mechanism of c-Kit auto-activation and its implication in sunitinib resistance. *J. Biochem.* **2010**, *147*, 601–609. [[CrossRef](#)]
- Binns, K.L.; Taylor, P.P.; Sicheri, F.; Pawson, T.; Holland, S.J. Phosphorylation of tyrosine residues in the kinase domain and juxtamembrane region regulates the biological and catalytic activities of eph receptors. *Mol. Cell. Biol.* **2000**, *20*, 4791–4805. [[CrossRef](#)] [[PubMed](#)]
- Chan, P.M.; Ilangumaran, S.; La Rose, J.; Chakrabarty, A.; Rottapel, R. Autoinhibition of the kit receptor tyrosine kinase by the cytosolic juxtamembrane region. *Mol. Cell. Biol.* **2003**, *23*, 3067–3078. [[CrossRef](#)] [[PubMed](#)]
- Locascio, L.E.; Donoghue, D.J. KIDs rule: Regulatory phosphorylation of RTKs. *Trends Biochem. Sci.* **2013**, *38*, 75–84. [[CrossRef](#)]
- Amit, I.; Wides, R.; Yarden, Y. Evolvable signaling networks of receptor tyrosine kinases: Relevance of robustness to malignancy and to cancer therapy. *Mol. Syst. Biol.* **2007**, *3*, 151. [[CrossRef](#)]
- Edling, C.E.; Hallberg, B. c-Kit—a hematopoietic cell essential receptor tyrosine kinase. *Int. J. Biochem. Cell Biol.* **2007**, *39*, 1995–1998. [[CrossRef](#)] [[PubMed](#)]
- Thatcher, J.D. The Ras-MAPK Signal Transduction Pathway. *Sci. Signal.* **2010**, *3*, tr1. [[CrossRef](#)]
- Shapiro, P. Ras-MAP kinase signaling pathways and control of cell proliferation: Relevance to cancer therapy. *Crit. Rev. Clin. Lab. Sci.* **2002**, *39*, 285–330. [[CrossRef](#)]
- Yee, N.S.; Hsiao, C.W.; Serve, H.; Vosseller, K.; Besmer, P. Mechanism of down-regulation of c-kit receptor. Roles of receptor tyrosine kinase, phosphatidylinositol 3'-kinase, and protein kinase C. *J. Biol. Chem.* **1994**, *269*, 31991–31998. [[CrossRef](#)]
- Inizan, F.; Hanna, M.; Stolyarchuk, M.; Chauvot de Beauchêne, I.; Tchertanov, L. The First 3D Model of the Full-Length KIT Cytoplasmic Domain Reveals a New Look for an Old Receptor. *Sci. Rep.* **2020**, *10*, 5401. [[CrossRef](#)]
- Blume-Jensen, P.; Wernstedt, C.; Heldin, C.H.; Rönnstrand, L. Identification of the major phosphorylation sites for protein kinase C in kit/stem cell factor receptor in vitro and in intact cells. *J. Biol. Chem.* **1995**, *270*, 14192–14200. [[CrossRef](#)]
- Lennartsson, J.; Jelacic, T.; Linnekin, D.; Shivakrupa, R. Normal and oncogenic forms of the receptor tyrosine kinase kit. *Stem Cells* **2005**, *23*, 16–43. [[CrossRef](#)] [[PubMed](#)]
- Hleap, J.S.; Blouin, C. The Semantics of the Modular Architecture of Protein Structures. *Curr. Protein Pept. Sci.* **2016**, *17*, 62–71. [[CrossRef](#)] [[PubMed](#)]

24. del Sol, A.; Araúzo-Bravo, M.J.; Amoros, D.; Nussinov, R. Modular architecture of protein structures and allosteric communications: Potential implications for signaling proteins and regulatory linkages. *Genome Biol.* **2007**, *8*, R92. [[CrossRef](#)] [[PubMed](#)]
25. Volinsky, N.; Kholodenko, B.N. Complexity of receptor tyrosine kinase signal processing. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a009043. [[CrossRef](#)]
26. Bah, A.; Forman-Kay, J.D. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J. Biol. Chem.* **2016**, *291*, 6696–6705. [[CrossRef](#)]
27. Collins, M.O.; Yu, L.; Campuzano, I.; Grant, S.G.; Choudhary, J.S. Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder. *Mol. Cell. Proteom. MCP* **2008**, *7*, 1331–1348. [[CrossRef](#)] [[PubMed](#)]
28. Iakoucheva, L.M.; Radivojac, P.; Brown, C.J.; O'Connor, T.R.; Sikes, J.G.; Obradovic, Z.; Dunker, A.K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **2004**, *32*, 1037–1049. [[CrossRef](#)] [[PubMed](#)]
29. Sen, S.; Udgaonkar, J.B. Binding-induced folding under unfolding conditions: Switching between induced fit and conformational selection mechanisms. *J. Biol. Chem.* **2019**, *294*, 16942–16952. [[CrossRef](#)]
30. Waudby, C.A.; Alvarez-Teijeiro, S.; Josue Ruiz, E.; Suppinger, S.; Pinotsis, N.; Brown, P.R.; Behrens, A.; Christodoulou, J.; Mylona, A. An intrinsic temporal order of c-JUN N-terminal phosphorylation regulates its activity by orchestrating co-factor recruitment. *Nat. Commun.* **2022**, *13*, 6133. [[CrossRef](#)]
31. Salazar, C.; Höfer, T. Kinetic models of phosphorylation cycles: A systematic approach using the rapid-equilibrium approximation for protein–protein interactions. *Biosystems* **2006**, *83*, 195–206. [[CrossRef](#)]
32. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)] [[PubMed](#)]
33. Nolte, R.T.; Eck, M.J.; Schlessinger, J.; Shoelson, S.E.; Harrison, S.C. Crystal structure of the PI 3-kinase p85 amino-terminal SH2 domain and its phosphopeptide complexes. *Nat. Struct. Biol.* **1996**, *3*, 364–374. [[CrossRef](#)] [[PubMed](#)]
34. Ledoux, J.; Trouvé, A.; Tchertanov, L. Folding and Intrinsic Disorder of the Receptor Tyrosine Kinase KIT Insert Domain Seen by Conventional Molecular Dynamics Simulations. *Int. J. Mol. Sci.* **2021**, *22*, 7375. [[CrossRef](#)] [[PubMed](#)]
35. Lennartsson, J.; Rönstrand, L. Stem cell factor receptor/c-Kit: From basic science to clinical implications. *Physiol. Rev.* **2012**, *92*, 1619–1649. [[CrossRef](#)] [[PubMed](#)]
36. Schlessinger, J.; Lemmon, M.A. SH2 and PTB domains in tyrosine kinase signaling. *Sci. STKE Signal Transduct. Knowl. Environ.* **2003**, *2003*, Re12. [[CrossRef](#)]
37. Songyang, Z.; Shoelson, S.E.; Chaudhuri, M.; Gish, G.; Pawson, T.; Haser, W.G.; King, F.; Roberts, T.; Ratnofsky, S.; Lechleider, R.J.; et al. SH2 domains recognize specific phosphopeptide sequences. *Cell* **1993**, *72*, 767–778. [[CrossRef](#)]
38. Backer, J.M.; Myers, M.G., Jr.; Shoelson, S.E.; Chin, D.J.; Sun, X.J.; Miralpeix, M.; Hu, P.; Margolis, B.; Skolnik, E.Y.; Schlessinger, J.; et al. Phosphatidylinositol 3'-kinase is activated by association with IRS-1 during insulin stimulation. *EMBO J.* **1992**, *11*, 3469–3479. [[CrossRef](#)]
39. Wagner, M.J.; Stacey, M.M.; Liu, B.A.; Pawson, T. Molecular mechanisms of SH2- and PTB-domain-containing proteins in receptor tyrosine kinase signaling. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a008987. [[CrossRef](#)]
40. Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinform.* **2009**, *10*, 168. [[CrossRef](#)]
41. van Zundert, G.C.P.; Rodrigues, J.; Trellet, M.; Schmitz, C.; Kastiris, P.L.; Karaca, E.; Melquiond, A.S.J.; van Dijk, M.; de Vries, S.J.; Bonvin, A. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720–725. [[CrossRef](#)]
42. Takemura, K.; Kitao, A. More efficient screening of protein-protein complex model structures for reducing the number of candidates. *Biophys. Phys.* **2019**, *16*, 295–303. [[CrossRef](#)] [[PubMed](#)]
43. Chakravarty, D.; Guharoy, M.; Robert, C.H.; Chakrabarti, P.; Janin, J. Reassessing buried surface areas in protein-protein complexes. *Protein Sci. A Publ. Protein Soc.* **2013**, *22*, 1453–1457. [[CrossRef](#)] [[PubMed](#)]
44. Marsh, J.A. Buried and Accessible Surface Area Control Intrinsic Protein Flexibility. *J. Mol. Biol.* **2013**, *425*, 3250–3263. [[CrossRef](#)] [[PubMed](#)]
45. Stolyarchuk, M.; Ledoux, J.; Maignant, E.; Trouvé, A.; Tchertanov, L. Identification of the Primary Factors Determining the Specificity of Human VKORC1 Recognition by Thioredoxin-Fold Proteins. *Int. J. Mol. Sci.* **2021**, *22*, 802. [[CrossRef](#)]
46. Miao, Y.; Feher, V.A.; McCammon, J.A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584–3595. [[CrossRef](#)]
47. Miao, Y.; McCammon, J.A. Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications. *Annu. Rep. Comput. Chem.* **2017**, *13*, 231–278. [[CrossRef](#)]
48. Lim, W.A. The modular logic of signaling proteins: Building allosteric switches from simple binding domains. *Curr. Opin. Struct. Biol.* **2002**, *12*, 61–68. [[CrossRef](#)] [[PubMed](#)]
49. Buck, E.; Iyengar, R. Organization and Functions of Interacting Domains for Signaling by Protein-Protein Interactions. *Sci. STKE* **2003**, *2003*, re14. [[CrossRef](#)]
50. Dueber, J.E.; Yeh, B.J.; Bhattacharyya, R.P.; Lim, W.A. Rewiring cell signaling: The logic and plasticity of eukaryotic protein circuitry. *Curr. Opin. Struct. Biol.* **2004**, *14*, 690–699. [[CrossRef](#)]

51. Dueber, J.E.; Yeh, B.J.; Chak, K.; Lim, W.A. Reprogramming control of an allosteric signaling switch through modular recombination. *Science* **2003**, *301*, 1904–1908. [\[CrossRef\]](#)
52. Berlow, R.B.; Dyson, H.J.; Wright, P.E. Expanding the Paradigm: Intrinsically Disordered Proteins and Allosteric Regulation. *J. Mol. Biol.* **2018**, *430*, 2309–2320. [\[CrossRef\]](#)
53. Tompa, P.; Davey, N.E.; Gibson, T.J.; Babu, M.M. A million peptide motifs for the molecular biologist. *Mol. Cell* **2014**, *55*, 161–169. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Uversky, V.N. A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Sci. A Publ. Protein Soc.* **2013**, *22*, 693–724. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Wright, P.E.; Dyson, H.J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321–331. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Demarest, S.J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H.J.; Evans, R.M.; Wright, P.E. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **2002**, *415*, 549–553. [\[CrossRef\]](#)
57. Waters, L.; Yue, B.; Veverka, V.; Renshaw, P.; Bramham, J.; Matsuda, S.; Frenkiel, T.; Kelly, G.; Muskett, F.; Carr, M.; et al. Structural Diversity in p160/CREB-binding Protein Coactivator Complexes. *J. Biol. Chem.* **2006**, *281*, 14787–14795. [\[CrossRef\]](#)
58. Hilser, V.J.; Thompson, E.B. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8311–8315. [\[CrossRef\]](#)
59. Motlagh, H.N.; Wrabl, J.O.; Li, J.; Hilser, V.J. The ensemble nature of allostery. *Nature* **2014**, *508*, 331–339. [\[CrossRef\]](#)
60. Dunker, A.K.; Babu, M.M.; Barbar, E.; Blackledge, M.; Bondos, S.E.; Dosztányi, Z.; Dyson, H.J.; Forman-Kay, J.; Fuxreiter, M.; Gsponer, J.; et al. What's in a name? Why these proteins are intrinsically disordered: Why these proteins are intrinsically disordered. *Intrinsically Disord. Proteins* **2013**, *1*, e24157. [\[CrossRef\]](#)
61. Panda, A.; Tuller, T. Exploring Potential Signals of Selection for Disordered Residues in Prokaryotic and Eukaryotic Proteins. *Genom. Proteom. Bioinform.* **2020**, *18*, 549–564. [\[CrossRef\]](#)
62. Wandless, T.J. SH2 domains: A question of independence. *Curr. Biol. CB* **1996**, *6*, 125–127. [\[CrossRef\]](#)
63. Mittag, T.; Orlicky, S.; Choy, W.-Y.; Tang, X.; Lin, H.; Sicheri, F.; Kay, L.E.; Tyers, M.; Forman-Kay, J.D. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17772–17777. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Huber, A.H.; Weis, W.I. The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. *Cell* **2001**, *105*, 391–402. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Lev, S.; Givol, D.; Yarden, Y. Interkinase domain of kit contains the binding site for phosphatidylinositol 3' kinase. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 678–682. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Vajravelu, B.N.; Hong, K.U.; Al-Maqtari, T.; Cao, P.; Keith, M.C.L.; Wysoczynski, M.; Zhao, J.; Moore Iv, J.B.; Bolli, R. C-Kit Promotes Growth and Migration of Human Cardiac Progenitor Cells via the PI3K-AKT and MEK-ERK Pathways. *PLoS ONE* **2015**, *10*, e0140798. [\[CrossRef\]](#)
67. Marcus, Y.; Hefter, G. Ion Pairing. *Chem. Rev.* **2006**, *106*, 4585–4621. [\[CrossRef\]](#)
68. Stanfield, R.L.; Wilson, I.A. Protein-peptide interactions. *Curr. Opin. Struct. Biol.* **1995**, *5*, 103–113. [\[CrossRef\]](#)
69. Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. Principles of protein-protein interactions: What are the preferred ways for proteins to interact? *Chem. Rev.* **2008**, *108*, 1225–1244. [\[CrossRef\]](#)
70. Keskin, O.; Haliloglu, T.; Ma, B.Y.; Nussinov, R. Protein-protein interactions: Structurally conserved residues at protein-protein interfaces. *Biophys. J.* **2004**, *86*, 267A.
71. Keskin, O.; Ma, B.; Nussinov, R. Hot regions in protein-protein interactions: The organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.* **2005**, *345*, 1281–1294. [\[CrossRef\]](#)
72. Gianni, S.; Dogan, J.; Jemth, P. Coupled binding and folding of intrinsically disordered proteins: What can we learn from kinetics? *Curr. Opin. Struct. Biol.* **2016**, *36*, 18–24. [\[CrossRef\]](#) [\[PubMed\]](#)
73. DeLano, W.L. Unraveling hot spots in binding interfaces: Progress and challenges. *Curr. Opin. Struct. Biol.* **2002**, *12*, 14–20. [\[CrossRef\]](#) [\[PubMed\]](#)
74. DeLano, W.L. The case for open-source software in drug discovery. *Drug Discov. Today* **2005**, *10*, 213–217. [\[CrossRef\]](#)
75. Mol, C.D.; Lim, K.B.; Sridhar, V.; Zou, H.; Chien, E.Y.; Sang, B.C.; Nowakowski, J.; Kassel, D.B.; Cronin, C.N.; McRee, D.E. Structure of a c-kit product complex reveals the basis for kinase transactivation. *J. Biol. Chem.* **2003**, *278*, 31461–31464. [\[CrossRef\]](#)
76. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinform.* **2016**, *54*, 5.6.1–5.6.37. [\[CrossRef\]](#)
78. Shen, M.Y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci. A Publ. Protein Soc.* **2006**, *15*, 2507–2524. [\[CrossRef\]](#)
79. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291. [\[CrossRef\]](#)
80. Case, D.A.; Cheatham, T.E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688. [\[CrossRef\]](#)

81. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [[CrossRef](#)]
82. Meagher, K.L.; Redman, L.T.; Carlson, H.A. Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **2003**, *24*, 1016–1025. [[CrossRef](#)]
83. Berendsen, H.J.C.; Postma, J.P.M.; Gunsteren, W.F.v.; DiNola, A.; Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690. [[CrossRef](#)]
84. Duane, S.; Kennedy, A.D.; Pendleton, B.J.; Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **1987**, *195*, 216–222. [[CrossRef](#)]
85. Andersen, H.C. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24–34. [[CrossRef](#)]
86. Roe, D.R.; Cheatham, T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095. [[CrossRef](#)]
87. Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure—Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577–2637. [[CrossRef](#)]
88. Lyman, E.; Zuckerman, D.M. Ensemble-based convergence analysis of biomolecular trajectories. *Biophys. J.* **2006**, *91*, 164–172. [[CrossRef](#)]
89. Bahar, I.; Lezon, T.R.; Bakan, A.; Shrivastava, I.H. Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chem. Rev.* **2010**, *110*, 1463–1497. [[CrossRef](#)]
90. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38, 27–38. [[CrossRef](#)] [[PubMed](#)]
91. Dominguez, C.; Boelens, R.; Bonvin, A.M.J.J. HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737. [[CrossRef](#)]
92. Bakan, A.; Dutta, A.; Mao, W.; Liu, Y.; Chennubhotla, C.; Lezon, T.R.; Bahar, I. Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics* **2014**, *30*, 2681–2683. [[CrossRef](#)] [[PubMed](#)]
93. Bakan, A.; Meireles, L.M.; Bahar, I. ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics* **2011**, *27*, 1575–1577. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.