The role of protein loops and linkers in conformational dynamics and allostery

Elena Papaleo[1,2*], Giorgio Saladino[3], Matteo Lambrughi[4], Kresten Lindorff-Larsen[2], Francesco Luigi Gervasio[3], Ruth Nussinov[5,6]

[1] Computational Biology Laboratory, Unit of Statistics, Bioinformatics and Registry, Danish Cancer Society Research Center, Strandboulevarden 49, 2100, Copenhagen, Denmark

[2] Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Copenhagen, Denmark

[3] Department of Chemistry, University College London, London, U.K.

[4] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126, Milan, Italy

[5] Cancer and Inflammation Program, Leidos Biomedical Research, Inc. Frederick National Laboratory for Cancer Research, National Cancer Institute Frederick, MD 21702

[6] Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

* Corresponding author: Elena Papaleo, elenap@cancer.dk

Abstract

Proteins are dynamic entities that undergo a plethora of conformational changes that may take place on a wide range of length- and time-scales. These changes can be as small as the rotation of one or a few side chain dihedral angles or involve concerted motions in larger portions of the three-dimensional structure; both kinds of motions can be important for biological function and allostery. It is becoming increasingly evident that "connector regions" are important components of the dynamic personality of protein structures. These regions may either be disordered loops, i.e. poorly structured regions connecting secondary structural elements, or linkers that connect entire protein domains. Experimental and computational studies have, however, revealed that these regions are not mere connectors, and their role in allostery and conformational changes has been emerging in the last decades. Here we provide a detailed overview of the structural properties and classification of loops and linkers, and a discussion of the main computational methods employed to investigate their function and dynamical properties. We also describe their importance for protein dynamics and allostery using as examples key proteins in cellular biology and human diseases such as kinases, ubiquitinating enzymes and transcription factors.

Table of Contents

1. Introduction

Proteins are highly dynamic entities that may undergo different types of conformational changes, ranging from those involving single residues to large regions of their three-dimensional (3D) architecture. These conformational rearrangements can occur on different timescales and many of them are tightly related to the biological function of the protein and to its capability to interact with specific biological partners. Conformational changes or perturbed dynamical properties at sites that are spatially distant to a site where a modification or binding event occurs are fundamental components that relate allostery to protein function. It is thus important to study these motions and their complexity, by experiments, computational techniques or combinations of the two. Important components of protein architectures are not only the well-folded secondary structural elements but also the more disordered flexible regions such as loops or, in multi-domain proteins, linkers and their relative motions. This contribution will thus provide an overview of the structural properties and classification of protein loops and linkers, along with their relationship to allostery and conformational dynamics. In addition to illustrating

examples from both our own work and that of others we also include discussion on the main computational methods currently employed to investigate allostery. Our premise is that the structure–function-dynamics paradigm is biologically compelling and cannot be compromised. That paradigm posits that function is executed by distinct states which may be controlled by allostery.

1.1.   Multi-domain proteins and modular architectures

Domains are the basic functional and structural modules of proteins.[1,2] Most protein domains are autonomous folding units and each unit is often associated with a distinct function.[3] Protein domains can be defined portions of a polypeptide sequence, often in a single segment, that assume a stable 3D structure.[4] The majority of proteins are multi-domain: 2/3 of all the prokaryotic and more than 80% of the eukaryotic proteins include more than one domain.[5] Only a limited number of domain families exist in nature and thus the large number of domain combinations observed in proteomes suggests that domain shuffling is a major source of evolutionary innovation for new protein functions, together with domain duplication and recombination events.[2,6,7]

Signaling proteins are typically modular. Via their modular domains their scaffolds can interact with multiple partners to regulate the signaling pathways in space and time.[8] Signaling scaffolds direct the assembly of multiple proteins into larger complexes, which control the propagation of information in the cell. They offer ideal platforms protein engineering studies aiming to alter the signaling programs, for example, via recombination of libraries of signaling domains.[8] Shuffling of a catalytic domain with different regulatory domains can result in novel regulation or localization of the catalytic domain, leading to distinct changes in signaling behavior

and cellular phenotype. Multi-domain proteins are often at an advantage compared to single domain proteins, since they increase the effective local concentration of substrates (or products) along metabolic and signaling pathways.[5] This is expected to shorten the time and increase the specificity for cellular responses to environmental changes. The observation that catalytic units, which previously existed separately in simple organisms, have been linked covalently during evolution is likely to be related to these properties.[9] In addition, multi-domain proteins also enable more complex patterns of regulation.

Not only folded domains are discrete functional units, but also the linkers that connect the modular domains. Such structural elements should not be merely considered as flexible linkers that keep the domains together.[5] A deeper understanding of the functional role of linkers has, however, been hampered by the fact that they are often 'invisible' to X-ray crystallography, due to their intrinsic dynamical properties.[10] Nevertheless, multiple conformations attained by protein domains in different crystallographic structures help reveal the role of linkers in modulating the conformational changes occurring in different domains.[4] In many cases, it is also possible to identify 'supra-domains' i.e., combinations of two- or three-domain blocks that recur in different contexts and have a certain functional and spatial relationship. In several supra-domains, the geometry of the two blocks and the constraints on the domain-domain interfaces are crucial. Therefore, linkers and hinge regions between them are expected to play important roles.[11]

1.2. A heterogeneous repertoire of protein conformational changes

Proteins are dynamic entities, and their folded structures are mainly consolidated by non-covalent intramolecular interactions that can break and re-form, providing a high degree of inherent flexibility and plasticity. For many years we used to think in terms of static and rigid structures, in part because this is the view provided by X-ray crystallography. More recently, however, time-resolved crystallographic methods, nuclear magnetic resonance spectroscopy, a range of other biophysical techniques and computational methods have provided a finer-grained view of the multiple conformations involved in protein function, as well as the interconversion between these states which may occur on a broad range of timescales.[12–16]

Databases of protein movements and motions, such as the Database of Macromolecular Movements,[17] provide an impressive overview of hundreds of distinct protein motions; presumably even more types are possible but cannot easily be resolved structurally. The repertoire of protein conformational changes continues to increase. It encompasses cooperative movements of subunits as well as structural rearrangements that range from subtle changes in residue conformations to marked structural changes at the quaternary level.[18] Many of these changes can be triggered by distal sites in a cascade of events that occur throughout the protein architecture and are the key to understand protein allostery and its role in function. Disordered or poorly structured regions such as linkers and loops are important structural players for these mechanisms that provide the proper degree of structural flexibility and malleability.

1.3.   Allostery and conformational changes in proteins

Research during the last decades provided ample evidence that protein motions are not just random but related to biological functionality.[19–22] The current scenario is that

proteins feature a predisposition and intrinsic capability to undergo conformational changes of functional relevance. The relative population of these different pre-existing conformers can change after a binding event, and evolutionary pressure is likely to preserve such conformational transitions.[23–26]

This conformational scenario has stimulated new questions that are far from the old and established concepts, even if the hypothesis that conformational changes are the key to understand allostery was proposed more than 50 years ago.[27] Mechanisms such as induced fit[28] and conformational selection by two discrete 'open' and 'closed' states[29] continue to be discussed and studied; however, main questions are if and to what extent conformational changes are induced by a ligand or a substrate and the thermodynamic and kinetic effects that determine whether a ligand can drive a change in a protein structure if the protein structure itself were not predisposed to undergo that change. Is it likely that the substrate or ligand can stabilize *pre-existing* conformations of the protein – which may have low population in the unbound state – and where the conformational free energy change thus has a direct impact on ligand affinity. [23–26] The same questions hold when we aim to elucidate allosteric changes, i.e. functional conformational alterations that are driven by local phenomena such as ATP binding or hydrolysis,[30,31] cofactor binding,[32] post-translational modifications[33,34] and exert their perturbation effect over long distances in the structure from the allosteric site that is generally far away from the site where the functional consequences are read. In such cases it is important to understand how a predisposition towards structural changes helps to elicit these allosteric responses. Related is the identification of structure-encoded networks of interactions in cascading events, which could be achieved thanks to the possibility of identifying

certain key residues and interactions that modulate the propagation of a signal in many proteins.

Several earlier works have focused on common mechanisms that allow one to point out the pre-existence of protein conformations that resemble the bound or modified states of a protein even in the absence of the ligand triggering the structural rearrangements (for reviews see for example[35–39]). Protein motions are not random but finely defined and intimately linked to the 3D architecture,[40,41] suggesting that motions pre-exist, and proteins have evolved to achieve their intrinsic dynamics.[22,42] Finally, there is accumulating evidence to suggest that there are intra- and intermolecular pathways of communication that help ensure the propagation of perturbations to distal sites and trigger the allosteric responses.[15,43–45] This may occur through a cascade of collisions between residues that change their rotameric states during the allosteric propagation, as observed in the transition between the major and minor states of cyclophilin A.[46–48] In other cases, different classes of intra- and intermolecular interactions can break and new ones can form upon a structural perturbation at a distal site. Even the formation of transient nonnative interactions, such as hydrogen bonds, which partially compensate for the loss of native contacts, can allow a decrease of the energetic barrier for conformational transitions induced by an allosteric effect, as observed in NtrC^r.[39,45] Since proteins are tightly packed, atoms cannot move freely and independently, but they can vibrate. In this view, an energetic perturbation can be observed so that the modification or ligand binding result in local strain, and that this strain may dissipate throughout the structure in a nonhomogeneous way. The dissipation involves the propagation of changes in atomic interactions to relieve the strain and can occur through multiple major and minor pathways that will strictly depend on the protein topology and the population of

conformations in the ensemble. Perturbation at any site of the protein will thus reshape and shift in the distribution of the pre-existing conformational states.[39,49,50] Such a redistribution of states can be described and understood through statistical thermodynamics.[51,52] The states are separated by energetic barriers and the height of the energy barriers defines the time scale of the conformational exchange. Conformational changes in flexible regions of the proteins can have low barriers, which lead to a fast interconversion between the different states. This thermodynamic description implies that allostery can be expressed in terms of changes in both entropy and enthalpy and thus allostery can take place even in absence of evident conformational changes.[53–55] In this general scenario, to provide an atomic-level description and a full understanding of the complex mechanisms related to allosteric events and structural communications in proteins, we must also consider the contributions of the unfolded or partially disordered and often heterogeneous dynamics of a protein structure, which are the focus of this review.

2. Structural properties and classification

2.1. Loops

2.1.1. Loop classification and properties

Loops describe a diverse class of structures that include both well-defined turns and more disordered random-coil-like structures, that often connect the more regularly-folded secondary structures (α-helices and β-strands).[56] Loops are much more than mere connection elements between other secondary structural elements. The lengths of loops often exceed what one would expect if they were to serve merely as connectors.[57] Surface-exposed loops, for example, often play a crucial functional role since they have the potential to interact with solvent, ligands and other biomolecules.

Loop regions generally belong to the most flexible parts of a protein structure, though they may also be as rigid as α-helices and β-strands. They are also associated with a higher variability in terms of sequence composition, even in proteins that have conserved architectures and are homologs. They thus contribute by generating the required diversity and variability to acquire new or different functions to support diversification within different families of the same superfamily.[58] For example, enzyme evolution often involves sequence changes in loop regions.[56] Nevertheless, it is not possible to rule out the possibility of a variable level of structural heterogeneity, which may reflect limited flexibility, as well as the fact that they can accommodate short regions with a well-defined structure.

2.1.2. Time scales of loop dynamics

The determinants of loop plasticity and the associated time scales of motions are thus key elements for their biological function.[59] To answer fundamental open questions in biomolecular recognition related to mechanisms such as induced fit, conformational selection and population shift requires both an understanding of the conformational ensemble of loops as well as the associated time scales for their dynamics.[39,59] Loop motions can occur on broad range of time scales from few picoseconds to milliseconds or beyond.

In a recent work, Brüschweiler and co-workers[59] performed an extensive study of time scales of dynamics in 169 loops (in 38 different proteins) defined as regions that neither adopt a β-strand or α-helical conformation. They used molecular dynamics simulations to probe loop motions and a machine learning approach to classify the loops according to their overall flexibility. They divided loops into 'static' and 'dynamic', the former further divided into 'fast' and 'slow' according to correlation times that were smaller or larger than 10 ns (Figure 1). They also identified key

factors for loop dynamics, such as lengths, composition, hydrogen-bond patterns, atomic contacts and structural patterns in loop regions, such as turns, β-bridges, 3_{10} helices and bends. With these characteristics in hand, they developed a prediction algorithm for the timescales of loop dynamics (ToeLoop, Time scale of every Loop). The study also sheds light on amino acid propensities in 'fast' and 'slow' loops. Residues such as methionine, aspartate and lysine, have clear propensities for fast loops, whereas amino acids such as leucines, cysteines, tyrosines, isoleucines, phenylalanines, are important components of loops classified as static. Three polar or bulky residues (threonine, tryptophan and histidine) are representative of slow loops and this can find a rationale in the fact that they are able to participate in processes that involve the formation and breaking of hydrogen bonds, which can occur on timescales from tens to hundreds nanoseconds.[59] The authors also observed a correlation, albeit quite weak, with 'unfoldability' scales in databases of intrinsically disordered proteins[60] with residues in dynamic (or static) loops having a lower tendency to folding (or unfolding). The relatively modest correlation might, however, also suggest revisiting the 'unfoldability' scales in intrinsic disorder databases, and the relationship between propensities for various loop properties and intrinsic disorder.[61] Fast loops are generally shorter and with more negatively charged residues and lower hydrophobicity index, whereas static loops are characterized by larger lengths, more atomic contacts and increased hydrophobicity. Slow loops are more ambiguous to determine with respect to length and physico-chemical properties and they tend to fall between the two categories above. The statistical analysis provided by Gu et al.[59] does not necessarily establish a causal relationship between residue type and time scale of loop dynamics. It remains unclear, for example, whether methionine 'actively' accelerates loop dynamics or whether other evolutionary mechanisms result

in preferences of methionines in loops that were already intrinsically fast. Similar considerations hold for other residues. The composition of some loops is highly heterogeneous and can account for residues classified in either slow or fast loops. Despite the progress made by such studies it remains limited by the fact that with conventional MD, the authors have been able to only sample up to 500 ns. Further, while the average properties of the loops were validated by chemical shift calculations, the main experimental parameters used to probe loop dynamics are NMR relaxation order parameters (S^2) that are mostly sensitive to relatively fast dynamics occurring on a timescale shorter than that of the overall rotational motion (typically around 5-10 ns for small globular proteins). All loops that displayed correlation times slower than 10 ns but faster than 500 ns were classified as slow. One should also keep in mind that those loops identified as static, may undergo conformational dynamics on the micro/millisecond timescales or even slower, which is beyond the time scales probed by conventional MD simulations used in this pioneering study. Despite this, the work is important since it builds the foundation towards a better understanding of not only the structural heterogeneity of loops, but also the complex and heterogeneous dynamics of loops by combining simulations with experiments and machine learning approaches. With further increase of computational power, or with new applications in the field of enhanced sampling applied to MD (see Chapter 4.2), we will be able to cover larger time scales and thus examine even slower dynamics which we can then compare to a broader range of NMR measurements including paramagnetic relaxation enhancement, relaxation dispersion measurements and residual dipolar couplings.

Analyses of protein loops in terms of their time scales and conformational changes are crucial to unravel their functional roles, such as the sequence of binding events during

protein-protein or protein-ligand recognition processes, as well as the effects of mutations and modifications on protein ensembles and their functions. Fast loops are expected to be characterized by a relatively flat energy surface that can be easily re-shaped by binding partners or modifications. Slow loops have higher free-energy barriers between different conformational substates. Whether such higher barriers manifest reduced ability to be reshaped and lowered by the presence of binding partners is unclear, as well as whether the existence of minor populated states that resemble cofactor-bound conformations is important for their binding mechanism.

2.1.3 Relationship between local motions in loops and global protein dynamics

Loop motions are often thought not to be collective and instead confined to specific segments of the polypeptide backbone. If so, they can be attributed to local rather than global protein dynamics.[62] One way to analyze loop dynamics is to assume that conformational changes in loops are predominantly determined by local interactions. In a recent work using Principal Component Analysis (PCA) and Elastic Network Models (ENM), Bahar and coworkers demonstrated that local conformational changes at loop regions are not necessarily independent of the soft global modes of motions that are intrinsically embedded in the protein architecture, thus providing a mechanism for coupling between loop motions and motions in the remainder of the protein. Based on an analysis of more than 100 proteins they suggested that the highly collective soft modes can contribute more to rearrangements in loops along the directions stabilized than the high frequency modes, for example, by ligand binding.[62] Network theory applied to conformational ensembles of proteins also provided a link between local changes and global dynamics.[63] NMR studies similarly demonstrated a connection between different timescales of protein dynamics and the corresponding

amplitude of motions, so that motions on the ps- or ns- timescale facilitate motions occurring on larger scale with slower conformational changes.[20]

2.1.4. Triggering and triggered loops

A special class of loops that are important for catalysis are the so-called 'triggered loops'.[64] They have been identified in proteins that do not feature detectable structural or sequence homology. These classes of enzymes possess functional loops (i.e., the triggered loops), whose conformational changes can be triggered by a second smaller interacting loop (i.e. a triggering loop, Figure 2). The triggering loop is highly conserved within each enzyme family and is even more conserved than the triggered loop. Mutations that modify the interactions between the triggering loops and the target loops are able to alter the enzyme activity. The triggering loops are generally the ones that show little or no structural changes in crystallographic structures. They are generally rich in glycine residues and are able to 'communicate' flexibility to the triggered loops, i.e. the ones presenting the largest conformational changes. Nussinov and coworkers[64] showed examples of triggering loops in very diverse enzymes, such as β1,4-galactosyltransferase-I, enolases and lipases. For example, in β1,4-galactosyltransferase-I, the long functional loop undergoes a conformational displacement of more than 20 Å and this is facilitated by changes in the interaction with a shorter loop (Trp loop) that shows remarkably smaller conformational changes. Moreover, other loops in the surrounding region have coupled motions with the Trp loop and can contribute to modulate long range the triggered functional loop.[65] Not all enzyme loops are equally correlated and conserved in terms of loop-loop interactions, making it challenging to predict through the analysis of correlated motions the loops that can have a triggering function.[65] The authors suggested that triggering loops could lower the energy barriers for conformational changes of the functional loops

through loop-loop interactions such as hydrogen bonds or hydrophobic interactions. Altering selected loop-loop interactions may be a practical strategy to design new proteins with different dynamical patterns in the functional loops thus exerting effects on protein activity and ligand binding. Triggering loops are often elusive for X-ray crystallography and are thus not observed in different conformations in apo- and holo-structures of the same enzyme since they are subjected only to minor displacements from the average structure. Chloride-dependent α-amylases might provide a new example of triggering loops. The dynamics of α-amylase loop 7, which acts as a lid on the active site, can be modulated by a more solvent exposed loop (loop 8).[66–68] In this context, loop 8 would act as a triggering loop for the functional loop 7. Triggering loops can thus be a general property of proteins in diverse protein architectures and evolution might have adapted the same overall dynamic scaffold for different types of functions.

In other cases, it is not only the loop dynamics that matter. Structural rearrangements of the active site can be accompanied or facilitated by conformational changes of secondary structural elements in the proximity of the active sites, such as entire α-helices, which often precede or follow a flexible and disordered 'lid' region.[56] An example of this class is the cap domain of a haloalkane dehalogenase.[69]

2.1.5. Omega loops

Another special class of loops is the so-called 'omega' loops. They fold into a loop-shaped conformation where there is a small and specific distance between their end points (hinge points). The main chain of these loops, which is connected to the rest of the protein structure by the hinges, assumes a conformation that resembles a Greek omega, from which the name of these loops derives (Figure 3). They were discovered

in 1986 where a survey of more than 60 proteins identified 270 omega loops.[70] They are often associated with regulatory functions and biomolecular recognition.[70,71] Omega loops are defined by their length, the maximum distance between the $C\alpha$-$C\alpha$ atoms of each pair of residues in the loop, the absence of secondary structure and the distance between the two hinge residues at the extremity of the loop. The hinge distance is often found to be in the range of 3.7-10 Å and is shorter than two-thirds of the longest $C\alpha$-$C\alpha$ distance observed between the residues forming the omega loop. In few omega loops, turns or 3_{10} helices can be formed by few of the residues of the loop. Omega loops also show preferences for specific amino acids, such as glycine, proline, tyrosine, aspartate, serine and asparagine.[72,73]. Omega loops can often belong to allosterically regulated regions of proteins, such as the Tyr181 to Tyr188 omega loop of HIV-1 reverse transcriptase.[74] In cytochrome c, the loop between residues 40 and 57 acts as a cooperative unfolding/refolding unit and was classified as an omega loop.[75] Another example is loop 7 together with its acidic insertion in Cdc34-like E2 ubiquitin-conjugating enzymes,[76] whose role is discussed in Chapter 3.4.2.

2.1.6. Modulation of protein dynamics by shortening of loops

Using atomistic simulations Levi and coworkers recently investigated how shortening a loop region in four different proteins influences protein dynamics.[77,78] They analyzed different deletions in loop regions that are solvent-exposed and quite long and not necessarily expected to affect the structural integrity of the proteins. They could not identify a consensus in the effects on the different proteins. For some of them, such as AcP and Ubc7, loop deletions stabilized the native state of the protein and the effects become more pronounced with increase in the length of the deletion up to a certain threshold. The results on Ubc7, a Cdc34-like E2 ubiquitin conjugating enzyme (see Chapter 3.4.2.) fit the experimental observation that deletion of the entire

acidic loop provides a functional protein and only abolish the capability to be regulated by phosphorylation.[79] In contrast, in other proteins, such as SH3 domains, small deletions did not affect the dynamics, and deletion of six residues disturbs the native structure. The authors concluded that these differences can be due to several factors, such as the fold and topology of the protein structure, the protein size, and the networks of intermolecular interactions that are mediated by the loops. The magnitude of the net stabilization upon the reduction of loop length may depend on the increase in conformational entropy required to balance a reduction in enthalpy due to the loop deletion. Even if solvent exposed, some of those loops can populate states in which they directly or indirectly affect the network of intramolecular interactions, including long distances, and in principle they would also be able to increase the conformational entropy of other distal loops. One example is the acidic loop of E2 enzymes Cdc34 and Ubc7, where in solution the loop interacts with the area surrounding the active site, in the open and solvent exposed states, as well as in the closed states.[76,79] In the earlier study by Levi and coworkers, the truncation of six residues in the AcP loop increased the protein thermodynamic stability with major contributions related to changes in conformational entropy.[77] In SH3 domains the shortening of the loop did not result in changes in conformational dynamics, although stabilization could result from other mechanisms, such as loop length affecting the entropy of the unfolded state, in agreement with pervious results from Viguera and Serrano.[80]

2.2. Linkers

2.2.1 A general overview of linkers

The different modules of multi-domain proteins (see Chapter 1.1) are connected by short or long stretches of amino acids, which are often characterized by a certain degree of disorder. These are referred to as linkers. Early examples starting from the

1960s established a relationship between linker peptides and the functional dynamics that they enable in the protein, and since then much effort focused on understanding the basis of such motions and on defining protein regions that are involved.[4,81,82]

Protein linkers are not merely covalent connectors between different domains of the same protein. They i) contribute to cooperative modulation of inter-domain and protein-protein interactions,[81,83] for example acting as adaptors to fit and regulate different folded domains; ii) establish distal communication between different functional modules of multi-domain proteins;[4,81] iii) direct the correlated movements of domains acting as hinge elements (see Chapter 2.3.), iv) in spacers that maintain end-to-end distances between attached domains.[4] Linkers also contribute, together with domain shuffling, duplication and domain combinations, to generate structural/functional variability within the proteome.[84] Considering the limited number of domain families and architectures in nature, linkers may provide diverse collections of structural assemblies. At the same time, alterations in linker regions affect stability, oligomeric states, proteolytic resistance and solubility of single-chain proteins.[85]

Analyses of datasets of linkers provided by structural databases reveal lengths between 2 and 21 residues with an average length of 6.5 or 10 residues, depending on the dataset used for the study.[82,86] Such statistics might however be severely biased by the inherent difficulties in determining structures of highly flexible molecules. Solvent accessibility is related to linker length and the average hydrophobicity decreases with the increase in the length of the linker.[82] Proline is the main terminal residue of linkers followed by residues such as arginine, phenylanine, threonine, glutamate or glutamine. Thus, in general, preferable residues in linkers are uncharged

or charged polar residues even if, depending on the dataset used for the analyses, different results could be achieved.

It is generally difficult to define amino acid propensities in linker regions since this depends on their function, again attesting to the importance of these structural elements in modulating protein properties. In many cases linkers are glycine-rich and this residue is known to promote flexibility due to the absence of a β-carbon, which in turn allows glycine to access dihedral angles that are otherwise energetically forbidden. Different modules of a protein often need to act in a highly orchestrated manner, where linkers contribute to regulate the reciprocal interactions and functionalities. In these cases, linkers need to be provided a certain degree of flexibility and glycine-rich peptides are an optimal solution to provide hinge regions. Hinge properties will be discussed in Chapter 2.3 together with methods proposed to predict softness of a linker-related hinge motion.

2.2.2. Soft linkers

An important consequence of the flexibility allowed by soft peptide linkers is the ability of linked domains to move to and from spatial proximity. A classic example of this is represented by the diphtheria toxin where the entire 15 kDa "R" domain rotates by 180° from a detached open, dimeric swapped form to a closed, monomeric form through conformational changes occurring in a six-residue loop (Figure 4).[87]

Due to their ability to break and form contacts between adjacent domains, soft linkers often facilitate essential functional events for the protein. For example the NFκB glycine-rich hinge region is flexible enough to bring the p50 and SWI6/ANK domains into contact and these interactions are important to regulate the intracellular transport of the transcription factor.[88] The glycine linker of NFκB literally allows one terminus of the protein to 'fold back' on to other. The fold-back property of polyglycine has

been investigated through pulse-radiolysis experiments.[89] In this experiment, the electron donor and acceptor were separated by either proline or glycine linkers of 0-3 residues in length. The kinetic constant for the electron transfer between the donor and acceptor correlated with the length of the proline but not with the length of glycine bridges. This observation was interpreted to imply that a moiety attached to the glycine linker was able to transfer energy via direct collision rather than electron transfer through the linker backbone; that is, the glycine linker folded-back so that the electron donor could come into direct contact with the electron acceptor on the opposite side of the molecule. The concept of soft linker has been also used to engineer new linkers within proteins designing them as stretches of amino acids where at least four of every six residues were glycine (see for more details Chapter 2.3.).[90] However, a high linker flexibility can also be detrimental for single-chain protein stability, folding kinetics and function, where preferred orientation is advantageous.[85] In a recent study different results have been also obtained in linker design for engineered antibody fragments, suggesting that caution has to be taken in the design when the linker is required to enhance structural stability or maintain functionality of a construct.[91] Maximum stabilities were observed when randomized linker regions of more than 15 residues contained alternating alanine and glycine residues with alanine being the predominant component.[92] Flexible linkers also favor serine residues.[92] A typical case are the so-called GS linkers (e.g. Gly-Gly-Gly-Gly-Ser)$_n$. They are a suitable choice when certain movements or interactions are required for domain fusion. Flexible linkers can serve as passive connectors between domains or allow reciprocal motions; but they might also destabilize the protein and result in poor expression yields or loss of biological activity.[92]

2.2.3. Rigid linkers and molecular rulers

Rigid linkers can be applied to allow a fixed distance between domains and maintain their independent functions. One example concerns α-helix-forming linkers with the sequence (EAAAK)$_n$ which have been applied to the design of recombinant fusion proteins.[93] (EAAAK)$_n$ linkers display a mostly α-helical conformation that is stabilized by N- and C-terminal capping provided by Glu and Lys residues and their electrostatic interaction.[94] Their capability to separate domains was assessed by fluorescence resonance energy transfer (FRET) experiments.[95]

Rigid linkers are often rich in prolines. Proline is unable to donate hydrogen bonds or otherwise contribute to regular secondary structural elements. It can provide a rigid separation between the domains and prevent unfavorable interdomain contacts thanks to increased structural stiffness. Most prolines in proteins are in *trans* conformation and this helps to maintain rigid interdomain separation. Examples are (XP)$_n$ linkers where X can be Ala, Lys or Glu.[92] The main chain conformation in proximity to prolines is neighbor-dependent and there are cases where *cis-trans* isomerization of the proline is favored, thus making even a proline-rich linker flexible.[4] A linker with low proline content can adopt a polyproline type II helical conformation, such as the 14-residue SH2-kinase linker of Src-family kinases (Figure 4).[81] The polyproline helix of SH2-kinase linker is fundamental for the regulation of domain-domain interactions aimed at activating/inhibiting the kinase activity (see Chapter 3.3.). A similar mechanism has been suggested for the intra-polypeptide linkers of polyketide synthase enzymes (PKSs), which contain one or more proline residues.[81] Many linker regions between protein domains can have appropriate residue propensity to form α-helical coiled-coils, which are known to be suitable structural motifs to promote heterodimerization.[96–98]

Linkers that act as spacers are often rigid peptides and they are called molecular rulers since they serve a 'metric' function, for example keeping domains apart.[4,99–101] Molecular rulers often include stably folded α-helices in the linker region, not only polyprolines.[4] Rigid linkers with stiff structures can be generated either using sequences that promote helical structures or by multiple Pro repetitions. However, quantitative analyses of single-molecule Förster resonance energy transfer (FRET) using polyproline of different lengths as spacers between donor and acceptors carried out by three studies in 2005-2007[102–104] reported higher mean FRET efficiency than expected for polyproline stretches acting as rigid spacer. The authors suggested that flexibility of polyproline, i.e. existence of species containing *cis*-prolines could contribute to the observed effects. Overall, these studies pointed out that the exact stiffness and distance distribution of polyproline peptides remains difficult to determine.

Small angle X-ray scattering (SAXS) experiments have also demonstrated that short linkers (less than three residues) often cause multimerization, whereas longer linkers (more than five residues) allow monomeric chimeric proteins,[94] obeying the rule of molecular rulers. Chimeric proteins where the linker has a helical structure can assume a more elongated conformation than the ones where the linker is disordered and flexible, even though exceptions were identified with flexible and elongated linkers. Natural linkers can also adopt a variety of other secondary structures, not necessarily only helical or disordered conformations but also coils/bends and turns.[92] The concept of linker can be extended even further if one considers that many multimeric complexes recruit folded domains or proteins. These proteins - functioning as adaptors or scaffolds - can sometimes be viewed as linkers, as for example in the case of the cullin domains of E3 enzymes in the ubiquitination pathway. There the

cullin links the two arms of the Cullin-RING ubiquitin ligase (CRL) molecular machine (see Chapter 3.4.).

Mutations within stiff linkers do not necessaryly affect function.[105] Molecular rulers keep the other amino acids' interactions in a proper registry and the nature of the side chains that compose the linker may play a marginal role. Many molecular rulers include repeated rigid monomeric units with low sequence complexity. Nevertheless, in the case of molecular rulers with sequences of increased complexity, the side chains might play a crucial role in defining the nature of the molecular ruler itself. For example, if the stiff linker is an α-helix, the properties of residues at the N- and C-capping of the helix are crucial for the stability and structure of the helix.[106] Molecular rulers with functional roles have been found, especially in transcription factors, such as bacterial transcription elongation factors of the Gre families that promote elongation by stimulating a specific transcript cleavage activity of the RNA polymerase.[107]

2.2.4. Cleavable linkers and linker design

The linkers described above generally provide stable peptide sequences that are not cleaved *in vivo*. However, under some circumstances, cleavable linkers might be needed to release free functional domains *in vivo*. Cleavable linkers are challenging to design as the design should take advantage of the in *vivo* processes, so that they can be cleaved under specific conditions, such as the presence of reducing reagents or proteases.[92,108] One way is to exploit the reversible nature of the disulfide bond *in vivo* together with sites for protease cleavage to design cleavable disulfide linkers.[108–110] Other cleavable linkers can be designed so that they are sensitive to proteases only. [92] Thanks to structural genomics initiatives and the consequent increase in the number

of available structures, automated methods to identify domains and linkers have become an important tool in structural analyses.[4] In this context, the compilation and curation of linker databases[82] and the availability of tools for linker predictors or modelling[111,112] are fundamental not only to store available information on these structural elements but also to design new linkers and domain assemblies. Domain identification and linker design are important for gene fusion techniques[94] with the aim of increasing the expression of soluble proteins, facilitating protein purification, designing gene reporters, performing immunoassays, and engineering bifunctional enzymes. Selection of suitable linkers to join protein domains is not straightforward and is too often neglected in the design of fusion proteins.[92] Information about the compactness and global shape of a chimeric protein is necessary for optimization of linker design. For *de novo* linker design, geometric analysis and modelling are not sufficient and need to be accompanied by biophysical validation. This issue becomes even more critical if one considers that the linkers are not only connectors but can play an active role in distal structural communication between domains (see Chapter 2.2.5.).

2.2.5. Linkers as vehicles to propagate allosteric effects.

Recently, it has been suggested that linkers are not merely flexible and they not only serve to prevent interdomain steric effects or for spatial domain rearrangement. Flexibility on its own is unlikely to be sufficiently productive.[5] The dynamics of linkers can mediate the propagation of a perturbation that arises at a specific site of one domain, i.e. a ligand binding site or a post-translational modification site, to a distal site. The outcome can be a reorientation of a second domain or of part of the domain itself. Such a model thus suggests that linkers themselves can encode a series of successive preferred states, in which each state encodes a subsequent one. The high

flexibility of linkers and the existence of pre-encoded conformational substates of the linker, which are regulated by allosteric propagation, can allow lower barriers for conformational transitions in the connected domains that otherwise would be much slower (Figure 5).

In the free state of a multi-domain protein in solution, one may depict the conformational ensembles with linker regions that are fluctuating and domains that sample different conformations ($\beta 1$, $\beta 2$, $\beta 3$…). In order to have a functional form of the protein, the domains may need to be in a well-defined relative orientation (for example $\beta 1$), which will also depend on the linker conformation. The linker thus has to be in its 'functionally favored' conformation too. This state can have a very low population in the free ensemble and binding events can thus cooperatively increase the population of this state with respect to the populations of other states. The redistribution of the states in the ensemble will be reflected in a series of observed conformational changes of the linker leading to the favorable functional state $\beta 1$ for the association of the protein domains.[5,113] The barriers between these series of hierarchically populated states are lowered thanks to the linker motions achieving a faster time scale for the conformational changes of even entire protein domains. Different residues in different locations of the linkers can have an impact on its conformations and dynamics in many diverse ways, and their individual contributions can be assessed by for example mutational studies.

To avoid a scenario in which the conformational ensemble only rarely samples functionally-relevant linker states, evolution is likely to have selected the linker sequence so that it can successively follow the states that lead to the functional conformation. Residues at certain positions that hold the key for such structural changes are likely to be conserved even in linker regions that one would expect to

have a highly variable sequence.[5]

An important consequence of this model is that we would need to study not only conformational substates of individual domains but also conformational propensities of the linker regions themselves for a better understanding of structural transitions in proteins and their time scales. Moreover, from the standpoint of applications it also implies that in multi-domain proteins linker regions can be suitable candidates as targets for allosteric drug discovery.[114,115] In this view, linkers would be one more example of pre-encoded conformational states in a protein that can also be observed for other protein regions within folded domains.[23–25,39] Thus, from the theoretical point of view these considerations can be applied to allosteric propagation pathways in any part of a protein structure.

The function of a linker as a vehicle for allosteric propagation is especially important if we consider metabolic and cellular signaling pathways that require fast responses upon a stimulus. Indeed, signaling proteins are generally modular proteins[116–118] and to transmit a signal among components of the pathways the information needs to be communicated between different protein modules. We may thus consider these modules as signaling units and will need to understand how the information is transferred from one module to another. In multi-domain proteins, single modules are connected by linkers, which become one of the keys to understand allosteric propagation. They are crucial for an efficient and fast transfer of information, for control of protein function and they are also responsible for the coordination between the module that receives the signal (e.g. the binding site of an allosteric ligand) and the output module to which the signal is transmitted. In this view, linkers can encode successive conformational substates, which allow them to fulfil their function. These states pre-exist even in the inactive protein, but their relative populations change

during allosteric propagation in the linker following a signaling event.[5] To validate the model of allosteric linkers, efforts should be devoted to the study at the atom-level with experimental techniques and molecular simulations to accurately describe the conformational states of linkers in multi-domain proteins in the presence and absence of their allosteric modulators and their relative populations.

2.3.   Hinge motions

2.3.1. How to define a hinge

Movements in proteins can be complex and varied, and although a large repertoire of different conformational changes exists, some recurrent classes of motion can be identified.[4,119–124] When there are many contacts between domains of a protein, the movements are the result of a series of small and subtle conformational changes that affect the whole contact area. In other cases, only few contacts affect different domains and the conformational changes will depend on large but localized hinge motions.

How may one define a hinge? Hinges are regions that permit the rotation of parts of the protein, usually one or more domains, as rigid bodies around a screw axis. Hinge movements are similar to rotations around an articulated joint. A small number of residues can be directly involved in a hinge motion since even a single bond can provide sufficient rotational freedom.[125,126] Hinge motions can thus allow major conformational changes without altering the internal packing of the single domains, and thus provides a mechanism for larger changes with only modest differences in free energies.

Hinge motions are often localized in flexible regions such as loops (see Chapter 2.1.) and linkers (see Chapter 2.2.) between different domains of multi-domain proteins. They are also often associated with highly conserved sequences, attesting to their

importance for function.[81] Due to their flexibility, linker regions are generally optimally suited to act as hinges since steric constraints on the main-chain atoms are absent to allow the hinge motions.[126] Indeed, flexible linkers generally do not feature packing constraints and easily allow changes in the torsional angles of the polypeptide main chain, in turn permitting proper motions and rotations of the domains that are connected through the linker. In this way, the deformations associated with the motions are confined to the hinge and overall the domains maintain their structure. Hinge regions thus need a certain degree of 'softness' to allow for conformational changes. It was also suggested[127–129] that changes in the torsional angles of hinge regions have very low energy barriers and easily allow rotations involving a small number of residues in the flexible regions and do not require major effects such as local denaturation or unfolding. This is a necessary property to permit a sufficiently fast change in the relative orientation of the domains, giving rise to the description of multi-domain proteins as an ensemble of different interconverting conformations in solution.

As a consequence of the linker's softness and hinge motions, protein domains can move in proximity and tightly pack against each other, changing from open/extended to more compact and collapsed states. This is the case, for example, in lactoferrin (see Chapter 2.3.2) and diphteria toxin where hinge motions involve domain rotations and changes in packing of interaction interfaces, switching from open to closed states.[4,122] Hinge properties such as degree of unconstrained main chain have been extensively studied to classify and predict such structural scenarios. An early description was formulated by Schimmel and coworkers, who studied rotational freedom around main chain angles and predicted peptide softness by comparing angles of rotation in different polypeptide stretches with different amino acid composition.[130] They

estimated the so-called characteristic ratio, measuring the distance between the ends of the main-chain for peptides with different number and composition of amino acids before the polypeptide chain starts to deviate from a linear direction. In peptides characterized by random amino acid composition and distances lower than 40 residues, the rotational hindrance potential of individual residues forces the separation between the ends of the polypeptide to increase proportionally with the number of residues. For a chain that includes only prolines, the distance shows a proportional increase up to 100 residues since polyprolines have large steric constraints on the main-chain torsional angles, at least in the all-trans form. By contrast, polyglycine chains are more flexible and do not maintain direction even after a few residues. This is common in glycine-rich hinges that are special class of hinges and can mediate a complete fold back in the direction of the main chain progression (see also Chapter 2.2.2.). These results suggest that polypeptide chains are prone to changes in the rotational potential and their size is strongly dependent on these properties.

Overall, the structural properties of a region in terms of extended or more collapsed states, as well as a hinge potential are dependent on the $C\alpha$-C bond and the main chain angle of rotation, explaining the importance of subtle structural characteristics that limit rotational freedom and their potential impact on protein conformation. Figure 6 presents a geometric approach to define hinges. In the figure, two moieties M and M' are represented.[4] They exemplify protein domains that act as rigid bodies connected by a hinge region with a pivot point P, around which the motions occur. The hinge region allows changes in the hinge angle β, which is named 'latitude', and permits the rotation of M and M', which alters their relative positions. If the hinge were fully flexible, the distance between M and M' would change depending on β. If the moieties were considered extended, the distances would also change as a function

of the angle α, called 'longitude', and the twist angle γ. During the rotation, the distance distribution $<d>$ between the two moieties M and M' thus changes with the softness of the hinge region.[4]

According to the description above, a rigid body movement of a protein domain can be described by six degrees of freedom. The screw axis of such movements is a shifted rotational axis that has an optimal position when any residual translational vector is parallel to the rotational axis,[131] as defined by the theorem of Chasles. In a hinge motion the screw axis is located in proximity to the hinge region, allowing the identification of a hinge axis around which a rotation brings into register the M and M' moieties. This axis was defined by Wriggers and Schulten[132] as an 'effective rotation axis' Ω for hinge motions that is perpendicular to the distance PM' and PM and has been used in computational tools for hinge axis prediction.

2.3.2. An example of hinge motions: lactoferrin

An early example of hinge movements induced by ligand binding was identified in the closure mechanism of the two domains of lactoferrin. It provides a molecular basis to explain its activity as an iron transport protein.[122] Lactoferrin can be divided into two similar globular halves identified as the N and C-terminal lobes, which are composed by two domains each, N1, N2 and C1, C2 respectively. Each pair of domains within the same lobe is connected by a flexible linker (Figure 7). The iron binding sites are located at the interface between the two domains of each lobe. Different experimental structures of iron-free (apo) and iron-bound (holo) forms of lactoferrin have been solved and they showed that large conformational changes occur upon iron binding or release.[122,133–138] In the apo states of the lactoferrin the domains of each lobe are

separated by a cleft at their interface, defined as the "open" form, exposing the iron binding residues. The binding of iron induces conformational changes in the protein and the two domains in each lobe close into the "closed" state, filling the cleft between them and the iron binding site is sequestered from the solvent (Figure 7). Analysis of lactoferrin crystallographic structures pointed to rigid-body motions of the two domains that are made possible by specific hinge elements. Here, we provide the example of movements of the N-lobe for the sake of clarity, but similar hinge movements have been reported for the C-lobe.[138] The hinge region is located in the linker connecting the N1, N2 domains, behind the iron-binding site. During the hinge motion N1 and N2 move as rigid bodies with a rotation of 54° around a hinge axis that passes close to Thr90-His91 and Val250-Pro251. The hinge motions are driven by changes in the torsion angles of Thr90 and Val250 and allow for the formation of the complete iron binding site in the closed states of lactoferrin, formed by residues in the N1 and N2 domains and a carbonate ion (Figure 7).[126] The residues of the hinge region are highly accessible to the solvent and feature a low number of atomic contacts with the rest of the protein. They are thus free from steric constraints, and this can facilitate hinge motions. X-ray solution scattering experiments on lactoferrin and transferrin, as well as on the isolated N- and C-lobes, and upon binding or release of iron, showed alterations consistent with conformational changes associated with the opening or closing of the cleft between the two domains.[139] According to X-ray crystallography, apo-lactoferrin from different species can populate a diverse array of states, with the N- and the C-lobes in open conformations, or with the N-lobe in an open state and the C-lobe in a closed conformation.[122,133–138] These results suggest that in the absence of the metal lactoferrin is highly flexible. Thus,

even if SAXS experiments suggested that both lobes are mainly in the open state in the apo form,[139] it is more likely that apo-lactoferrin can populate minor states in the closed or partially closed conformations.[137] One suggestion is that the interaction with iron shifts the equilibrium towards the closed states through hinge motions and then the metal stabilizes this state thanks to favorable intermolecular interactions. The hinge motions and the related conformational changes observed in lactoferrin are likely to be a general and conserved mechanism for transferrins since the residues important for hinge motions and metal binding are highly conserved.[134,138] More broadly, they provide an example of hinges that modulate the transition between 'open' and 'closed' apo and holo states or vice versa. They have been pointed out in many other proteins, including lactate dehydrogenase, adenylate kinase, maltose binding protein and other perisplasmic binding proteins.[129,140–146] Hinge motions regulating changes from open to closed states and vice versa can also be observed within single domains thanks to conformational changes occurring in loop regions, as the example the acidic loop of Cdc34-like E2 enzymes (see Chapter 3.4.2.).

3. Examples of allosteric signaling where loops and linkers play important roles

3.1.   Enzyme function and regulation

Over the past few years, considerable interest centered on loops and their impact on enzyme function and catalysis. Many studies have focused on the role of surface and lid loops that can cover/modulate the active site of an enzyme and their functional role in substrate and cofactor binding, or in enzyme stability. Residues in loop regions

can be exploited to design enzymes with different substrate specificities, temperature or salt dependence, or even new or promiscuous catalysts. Point mutations can be introduced into loop regions, modulating not only intra- and intermolecular interactions but also backbone preferences with a high impact on enzyme function and specificity. Moreover, improving loop rigidity may improve enzyme thermostability, even if this relationship is not straightforward and unpredictable side effects can arise, often due to long-range structural effects.[56,147–149]

Conformational changes in loop regions are frequently observed to be an important component of enzyme mechanisms, with substrate or ligand binding associated with structural rearrangement of the protein. Many examples showed that diverse protein motions are critical for enzymatic function; however little is known about their precise role in catalysis (Figure 8).

The rigid-lid in triosephosphate isomerase, for example, allows motions important for enzyme activity.[150–153] An extremely flexible flap region acts as a gate for ligand entry and exit in HIV protease.[154–156] Enolases or aldolases feature loop movements that allow the catalytic residues to be oriented in the proper position for catalysis.[65,157–160] In a similar way, protein tyrosine phosphatases (PTPs)[161] are characterized by a WPD loop that includes a catalytic aspartate (Figure 8). This loop closes over the active site upon binding of the substrate and loop closure allows the correct positioning of functional residues around the ligand and protect the site from bulk solvent during catalysis.[162–164] In lipases, helical loops open or close the hydrophobic active site acting as lids.[165–171] A displacement of more than 20 Å in a long loop of β1,4-galactosyltransferase creates binding sites for different ligands.[65,172]

Different groups have used experimental and theoretical strategies to suggest that the dynamics of flexible loops play a role in the catalytic mechanism of the enzyme

dihydrofolate reductase (DHFR).[56,156,173–179] DHFR has several loops in the proximity of the active site (Figure 8). The so-called Met20 loop works as a "lid" which closes over the cofactor, thereby allowing DHFR to adopt closed or occluded conformational substates. The Met20 loop conformational transitions are allosterically accompanied by changes in the patterns of hydrogen bonds of other distal loops, which surround the catalytic sites.

In agreement with the importance of the Met20 loop dynamics for DHFR function, removal of the side-chain steric hindrance of central residues in the loop through glycine mutations resulted in a striking 500-fold decrease in the rate of hydride transfer.[180] Different experimental biophysical techniques have showed that in the apo enzyme the Met20 loop can fluctuate between the two conformations. Experimental and computational studies, including mutations of many residues of the protein, support a role of remote sites and a long-range dynamic network in DHFR in the enzyme mechanism.[46,47,173,175]

Embedded dynamic networks in enzymes that are related to their catalytic activity are likely to be a broad and general scenario.[46] During the catalytic cycle the enzyme has to pass through different conformational states, and these involve the environment of the active site.[36] Intrinsic flexibility of the molecular architecture, the capability to exploit regions such as loops, which are also the regions where mutations can more frequently occur during evolution, for a transition between these conformational states is an optimal choice selected by evolution for enzyme activity and function.

Flexibility in the proximity of the active site can also account for broad substrate specificity since the enzyme cavity can accommodate stereochemically diverse substrates.[56] Substrates or ligands can induce conformational changes in the surroundings of the active site or select different substates of the dynamic

conformational ensemble of an enzyme with different affinities. The notion that the conformation of loops in the proximity of the active site is a key determinant for substrate recognition is strengthen by the fact that experimental structures of apo- and holo-states of many enzymes only differ in the conformations of the loops in proximity of the active site.[56] For example, once the substrate is bound the solvent-exposed loops adopt a compact and ordered conformations since they can interact not only with different sites of the ligand/substrate but also with other residues of the protein itself, which they encounter during the conformational change. This is typical of enzymes undergoing opening to closure transitions. The closure of the loops around the ligand/substrate allows the substrate to be protected from the aqueous environment and from other reactive agents. It also protects or stabilizes reaction intermediates.

3.2.   Ligand and cofactor binding

An example of a linker that plays an important role in ligand binding comes from a recent NMR study on calmodulin (CaM), [181] which is one of the prototypical calcium sensing proteins. CaM has two small domains divided by a short and flexible linker that allows the protein to assume a wide range of extended and compact conformations (Figure 9). CaM conformational plasticity seems to be important for Ca^{2+} signaling within the cell and CaM needs to bind hundreds of different peptide sequences.

We are used to think almost intuitively that flexible linkers include low-complexity and poorly conserved sequences, whereas more rigid and structured protein segments can be more conserved in terms of primary sequence. CaM is one remarkable exception that challenges this view. The CaM linker is highly conserved and its length is invariant. With this observation in mind, Anthis and Clore examined the effects of

changing either the length or rigidity of the linker through a very elegant mutational approach on the transient association between the two domains of CaM in the free state. They showed that as the length of the linker increases the domains become less constrained and tumble more rapidly in NMR experiments. They then employed paramagnetic relaxation enhancement (PRE), which is the technique to unlock minor populated states of a highly dynamic protein where the distances of the minor species are shorter than the distances in the major one for the different CaM variants. With this approach, they demonstrated that the transient association is maximal for a linker that is only one residue longer than the wild-type and decreases for lengths longer or shorter. Their results are likely to account for a more general mechanism exploited by many different proteins.[85,182–184] In the absence of the ligand the transient and sparsely populated compact states of CaM correlate with the affinity of the protein for different target peptides suggesting that in the absence of the ligand the conformational dynamics mediated by the linker plays an important role in facilitating the binding of the target. This example attests to evolution finely tuning linkers' lengths and composition to fulfill a functional requirement for such an important protein in the cell.

3.3   Activation and inhibition of protein kinases

3.3.1 The structure of protein kinases

Protein kinases (PKs) are the typical examples of proteins whose loops, hinges and linkers play a major role in their allosteric regulation. PKs are present in both bacterial and eukaryotic cells, where they are the cornerstone of cellular signaling, catalyzing the transfer of a phosphate group from ATP to a wide range of substrates, including sugars and lipids.[185] Deregulation of PK activity can lead to several pathologies, from diabetes to cancer.[186,187] For this reason, kinase activity is generally highly controlled

and the resulting signaling cascades tightly regulated, with a high level of robustness and redundancy.[188] The catalytic domain of PKs (Figure 10) has a highly conserved sequence and fold; they all consist of two lobes, a smaller N-terminal one, with a predominance of β-strands, and a larger C-terminal one, mostly helical.[189] The ATP binding site is located in the cleft between the lobes, in the proximity of the hinge region connecting the two. This region is essential for the opening and closure of the kinase catalytic domain (CD), the so-called "hinge motion", which is essential for catalysis[189,190] and can be allosterically modulated.[191,192] Right above the hinge region, the flexible P-loop (the β1-β2 loop, also known as G-loop or Gly-rich loop), is crucial for the coordination of the ATP phosphates.[193] Two hydrophobic ''spines'' (the regulatory and the catalytic spine) connect the two lobes of PKs and dynamically link all the elements important for catalysis.[194]

The transition from an inactive to a catalytically active form involves complex conformational changes in at least three conserved structural motifs: the activation loop (A-loop), the Asp-Phe-Gly (DFG) motif and the αC-helix. In the inactive state the A-loop is folded onto itself mimicking the substrate. Its opening is required for full activation. This complex conformational change opens the ATP channel for cofactor binding and creates the main platform for substrate docking. The αC helix, the only helical element of the small lobe, might rotate assuming an inactive, "out" position. In the "in" active conformation it forms a hydrogen-bond with the β3 strand in the active form.[195,196] Similarly the conserved DFG motif may assume a flipped or "DFG-out" inactive conformation.[197] In some protein kinases, as those belonging to the AGC family, a smaller αB helix precedes the αC creating a cavity, the so-called PIF pocket, which is crucial for their allosteric regulation.[198]

3.3.2. Protein kinases regulation

Protein kinases exist in equilibrium between an active and one, or multiple, inactive states. Activation usually involves phosphorylation of the A-loop and protein-protein interactions inducing the active structure via an allosteric mechanism. While protein kinases share a high level of structural homology in their active state,[193] the inactive states are generally more heterogeneous and diverse,[194] and so is the regulatory mechanism that each family acquired through evolution.

In several members of the Src family of tyrosine kinases as well as in Abl, two Src-homology (SH) domains (SH3 and SH2) are responsible for locking the kinase into an inactive state. The SH2 binds to the C-lobe on the opposite side of the A-loop, while SH3 binds to the linker between the catalytic domain and the SH2.[199,200] The flexible linkers between the domains and the C-terminal tail of the kinase play a key role in the regulatory mechanism. A phosphotyrosine in the C-terminal tail of the kinase domain binds to the SH2, locking the conformation and rigidifying the linkers.[201,202] In this auto-inhibited conformation, the two domains act as a grip and suppress the hinge motion important for catalysis. Thus, activation of the kinase requires the disruption of the complex, e.g. through binding of (higher affinity) allosteric ligands to the SH2 or SH3 domains.[203,204] Interestingly, an allosteric connection between the ATP pocket of the catalytic domain and the SH2-SH3 domains has been reported. When the inhibitor imatinib binds to the ATP pocket it counterintuitively dislocates the SH2 and SH3 domains to form a dynamic open state.[205]

 In Abl, the SH2 domain not only has an auto-inhibitory role but it is also involved in its full activation.[200,206] X-ray structures have revealed a peculiar "top-hat" conformation, in which the SH2 domain sits on top of the kinase N-lobe,[199] that enhances allosterically the activity of the kinase. A similar allosteric activation by the SH2 domain has been observed in other kinases, including Fes[206] and Btk.[207] The

molecular mechanism of the activation of Abl by SH2 in the top-hat position has been elucidated by molecular simulations and mutagenesis. When sitting on top of the N-lobe, the SH2 changes the dynamics of the catalytic domain, redirecting the hinge motion, stabilizing the P-loop and the αC helix, while favoring the A-loop switch to the open conformation and strengthening the catalytically-important salt bridge between the β3 lysine and the αC helix glutamate.[192] Interestingly, the β3-αC loop appears to act as a switch, allowing the allosteric communication to take place. Allosteric effects resulting from the modulation of the dynamics of this area are also observed upon binding of cyclins to cyclin-dependent kinases,[208] in the dimerization of the EGF receptor,[209] and in some ACG kinases that are activated by the binding of a specific peptide to the PIF pocket.[198] In some members of the AGC PK family (e.g. PKB) the C-terminal tail, whose length can vary significantly, folds on to the catalytic domain and binds to the PIF-pocket (Figure 11). In the 3-phosphoinositide-dependent protein kinase 1 (PDK1), which lacks the long C-terminal tail, its role is instead played by substrate-derived docking peptides (PIFtide). In PDK1, the PIF-pocket also plays a role in the specific recognition of its substrates. [198]

In other kinases, the C-terminal tail has an auto-inhibitory role by obstructing the entrance to the substrate binding site.[210] Similarly to the C-terminal tail, the long juxtamembrane (JM) loop, present in most receptor tyrosine kinases (RTK) and connecting the trans-membrane helix to the kinase domain, can fold onto the N-lobe of the kinase and interact with the αC helix and the P- and A-loops, stabilizing the inactive conformation.[211] Interestingly, a similar mechanism has also been observed for the linker of the SH2 domain in other kinases, like Syk or Zap70.[212]

In a similar way, the FERM domain of FAK (Focal Adhesion Kinase) binds to the C-lobe and stabilizes the inactive structure by blocking the access to the catalytic site and

preventing phosphorylation of the A-loop.[213] Regulation of the kinase activity is also mediated by allosteric interactions in extracellular (EC) regions of RTKs. For instance, the activation of FGFR, which requires the receptor autophosphorylation, is mediated by the binding of FGF and heparin to the Ig-like EC domains, which in turn promotes receptor oligomerization.[214]

3.3.3. Allosteric modulators of kinases

In recent years there has been growing interest in the development of allosteric drugs due to their many advantages with respect to classic "orthosteric" drugs such as increased selectivity, decreased susceptibility to drug-resistance causing mutations and the possibility of targeting otherwise "undruggable" targets.[215] The efforts devoted to the development of allosteric kinase inhibitors have been at least in part successful.[216,217]

One of the most interesting allosteric inhibitors reported so far was "SSR128129E" (SSR) the first small-molecule allosteric inhibitor that acts extracellularly for receptor tyrosine kinases (RTKs).[218,219] By binding to an extracellular immunoglobulin-like domain of the fibroblast growth factor receptor (FGFR, a RTK) it inhibits FGF-induced signaling linked to FGFR internalization in an allosteric manner. A combination of crystallography studies, nuclear magnetic resonance, Fourier transform infrared spectroscopy, mutagenesis, molecular dynamics simulations and free energy calculations were used to elucidate its complex mode of action. SSR induces the opening of a "cryptic" binding pocket on the Ig-like domain, and by binding to it, it modulates the protein-protein and protein-membrane interactions. The flexible linkers connecting the Ig-like domains and the juxtamembrane segment play a non-negligible role in the mechanism. [219]

Interestingly not only allosteric inhibitors but also allosteric activators have been reported in PKs (Figure 11). A rationally developed low- molecular-weight compound was shown to bind to the PIF pocket of PDK1 inducing local conformational changes that in turn had an allosteric effect at the ATP binding site and the activation loop. The conformational changes induced by the small compounds triggered the activation of PDK1.[220] An altogether more worrying drug-induced allosteric activation effect has been reported in the B-RAF kinase, a therapeutic target for melanoma. ATP-competitive inhibitors binding to a RAF monomer, inhibited it, but resulted in the induction of dimerization and transactivation of the drug-free protomer (paradoxical activation).[221] By comparing the effect of inhibitors inducing the paradoxical activation with a new class of inhibitors that do not (paradox breakers) it was possible to identify that structural and dynamical differences at the level of the A-loop and of the αC-helix are involved in the dimerization and paradoxical activation.[222] The importance of the conformation of the αC-helix and of the adjacent "regulatory-spine" in the allosteric regulation of B-RAF was also shown in a crystal structures of the functionally asymmetric dimer[223] and of the monomeric "off" state of the kinase.[224]

3.4. Ubiquitination pathway

The ubiquitination pathway plays a central role in cellular biology and regulates the fate of substrates tagged with ubiquitin (Ub) or ubiquitin-like (Ubl) proteins. It not only allows to target proteins for proteasome degradation but it is also at the basis of many important regulatory and signaling processes, such as involvement in DNA repair, activation of protein kinases, endocytosis, autophagy and immune response.[225–230] The Ub pathway consists of three classes of enzymes acting in a cascade of multiple steps that result in targeting substrate proteins labeled with one or more Ub

or Ubl molecules. E1 (Ubl-activating enzymes), E2 (Ubl-conjugating enzymes) and E3 (Ubl ligases) are the key player in the Ubl cascade.[231] E1 initiates the pathway by recognizing and activating Ubls in an ATP-dependent reaction. E2 alone or in complex with E3 can conjugate Ubls and act in direct ligation of the target substrate. Several experimental and computational works demonstrated the importance of protein dynamics, allostery and conformational changes in the function and regulation of the enzymes of the Ubl cascade, where either loops or linkers play pivotal roles.[232–234]

3.4.1 Hinge motions in E3 complexes: the case of the cullin

In the ubiquitination pathway functionally-important hinge motions have been described for cullin proteins. Cullins are essential components in a large class of E3 ubiquitin ligase complexes, such as the cullin-RING E3 ligases (CRLs, Figure 12).[235–237] CRLs are macromolecular complexes generally composed by four proteins: cullin, an adaptor, RING-box (Brix) and substrate-binding proteins (SBP). CRLs can be described as two-arm machines, one arm that includes Brix, which in turn is the recruitment interface for the E2-Ub complex, and the other is where SBP is located. The two arms are separated by the cullin structure.[238–241] These two flexible arms mediate the function of CRLs so that the E2-Ub complex and the substrate can be brought to a distance close enough for the ubiquitin transfer to the substrate. These conformational changes occur thanks to allosterically modulated motions.[238–241] Cullin is a well-folded protein but due to its collocation in the E3 quaternary architecture is considered as having a linker scaffolding function, providing an example of a whole 3D structure acting as a linker. Two domains compose canonical cullins in humans: the N-terminal (NTD) that interacts with SBPs and adaptor proteins, and the C-terminal (CTD) that interacts with Rbx proteins. Cullins were

previously considered as mere spacers to separate the two flexible arms. MD simulations of different cullins showed that they undergo motions mediated by hinge elements in the NTD.[241] The hinge motions are associated with changes in conformations of the structure of the cullins that, in turn, alter the distance between the two arms, showing that they work as flexible scaffolds. These hinges rely on changes in the torsion angles around glycine residues that are highly conserved across different cullins. This flexibility is essential for the cullins to accommodate different SBPs, adaptor proteins and substrates. Thanks to the hinge movements in NTD, cullins allosterically modulate the distance between the E2 enzymes and the specific substrate proteins to ensure high efficiency of the ubiquitination. Allosteric regulation shifts the conformational ensemble of cullin NTD towards states more favored to accommodate specific substrates that have to be mono-ubiquitinated or poly-ubiquitinated. Also the CTD of cullins is flexible, as demonstrated by experiments and simulations.[241,242] It has been shown that NEDD8 can bind cullins and allosterically modulate its conformational changes, inducing conformational rearrangements and activating the CRL complex.[243,244] If a flexible linker is artificially introduced between the CTD and NTD of cullins, the ubiquitination process is abolished.[245] This result suggests that motions of cullins are finely regulated since an extreme rigidity or flexibility would affect activity.

3.4.2 An omega loop in Cdc34-like E2 enzymes is activated by post-translational modification at distal sites

Ubiquitination pathways consist of a cascade of three classes of enzymes i.e., E1 (ubiquitin-activating), E2 (ubiquitin-conjugating) and E3 (ubiquitinating ligase) enzymes. Critical steps in this process are either how ubiquitin can be recognized by E2 enzymes and how individual protein substrates are recruited by specific E2-E3

complexes.[231,246–248] We will illustrate examples of the two classes for which a major role is played by a conserved acidic insertion in a loop close to the catalytic site of Cdc34-like E2 enzymes.

E2 enzymes are crucial for the Ub cascade since they can regulate the topology of the poly-Ub chain, account in part for the specificity of the substrate and influence the processivity of the reaction.[231,246–248] E2 superfamily has been divided in 17 different families.[249] E2s are mostly multi-domain proteins and share the catalytic core domain where the catalytic cysteine is located as well as a conserved histidine-proline-asparagine motif.[250] Cdc34 belongs to family 3 of E2 enzymes, which shares an acidic insertion of 12-13 residues in a loop (often called loop 7) in proximity of the catalytic site across many different species.

The insertion turned out to be an ancestral and conserved motif for members of family 3 and is characterized by alteration of hydrophobic and acidic residues.[76] Loop 7 of Cdc34-like E2 enzymes has been classified as an omega loop (see also Chapter 2.1.5.) and it has been suggested to play a dual role during the Ub cascade.[76] It carries out a regulatory role that can be allosterically activated by post-translational phosphorylation at distal sites[79] and by interaction with the cognate E3.[251] This loop, including in Ubeg2g and Ubc7, can act as a lid that modulates the accessibility of the catalytic site and impairs Ub-charging activity until a conformational change toward an open state is promoted by phosphorylation that induces electrostatic repulsive effects.[79] Phosphorylation in the C-terminal part of the ubiquitin-conjugating domain[79] or in the proximity of the catalytic cleft[252] may have an impact on the dynamics of this loop .

In the open state (Figure 13), the catalytic cysteine is accessible for Ub, and L7 interacts through the hydrophobic residues with L4. L7 also plays a role in the steps

that follow Ub-charging in the E2 catalytic cleft.[76,79] Once displaced in the open conformation, loop 7 can provide an interface to interact with the positively charged face of Ub and to recruit the Cdc34 cognate Rbx domain of the E3 complex.[76] In line with these suggestions, in a recent study based on Protein Structure Network (PSN) approaches (see also Chapter 4.5.), we identified communication paths from the loop to the known interface for recruitment of E3 enzymes and showed that these paths are reinforced and more 'tunneled' in the open active states of the protein.[253] Interestingly, open states are likely to be sampled also in the wild type non-phosphorylated variant,[79] suggesting that they are an intrinsic property of the Cdc34-like E2 ensembles even if further studies are needed to better address this point.

The mechanisms elucidated by our computational studies on Cdc34 and other related enzymes were supported by experimental validation. NMR and mutational studies show that the interface of Rbx1 for recruitment of Cdc34 includes the residues predicted by our model, and supported by a recent experimental study on the role of the acidic insertion in Cdc34.[79,254–256]

3.5. Transcription factors and DNA-induced conformational changes

Transcription factors (TF) are often multi-domain proteins and multiple cases have been reported where linkers and loops have crucial roles.[257] For example, the glucocorticoid receptor (GR) structural dynamics relies on an important allosteric linker region. GR consists of N-terminal domain, a DNA-binding domain (DBD), a linker region, a ligand binding domain (LBD) and the C-terminal domain.[258] The binding of hormones, which act as agonists, occurs in the cytoplasm and induces a major allosteric change. The GR then homodimerizes and translocates to the nucleus

where the DBD can bind to specific DNA response elements (REs) to initiate transcription. DNA binding can induce a conformational change of the 'lever arm', which in turn alters the cofactor-binding site and modulates GR's activity. Allosteric conformational changes can also further alter GR surfaces that interact with different coactivators, cofactors or other transcription factors. Thus, in a complex array of dynamics, GR provides an example of how a linker can tightly modulate through allosteric effects binding with multiple partners and DNA.[5,259–261] Even a single base pair change in the RE leads to altered functional effects, which are also mediated by allosteric propagation.[262]

These general importance of TF dynamics is also exemplified by p53. Binding of the p53 DBD to the p53 REs initiates a signal that propagates to the p53 activation domain (p53AD). Here, the DNA acts as an allosteric effector and the signal propagates from the DBD to the activation domain through a flexible linker.[5] Different REs provide different atomic contacts with the DBD, which result in different pathways.[263–265] Moreover, DNA not only acts as allosteric effector through linker propagation in p53, but it might also induce effects within the p53 DBD domain itself regulating the conformation of distal loops that are likely to be important interfaces for recruitment of other binding partners.[264,266]

Thus linkers in multi-domain transcription factors can have crucial and general importance.[267] A DNA-binding protein needs not only to bind DNA with a certain affinity but it also has to scan with a reasonable speed the DNA sequence to identify the specific target site. In this context, multi-domain architectures connected by disordered loops are advantageous to accelerate the kinetics of search of the binding site along the DNA sequence.[268] Two (or more) domains can be tethered together, through the linker so that one domain is characterized by high affinity for the DNA

(i.e. high stability) and the other one has a lower affinity which allow for a search along the DNA sequence, providing a functional strategy for transcription factors.[268] Their action relies on a 'monkey-bar' mechanism in which the protein can form a bridged intermediate between two distant DNA regions.[268] The dynamics of these biomolecular machines is highly dependent not only on the affinity of the domain components for DNA but also on the length of the linker region that connects them. These mechanisms have been proposed by computational studies[269,270] and supported by experiments and mutational studies.[271] The authors also demonstrated that when the asymmetry in the properties of the two different domains (i.e. high vs low affinity for DNA) is avoided, the proper functionality of the protein is lost. .[271]

Another example of important structural communication between DNA binding loops and potential regions for cofactor recruitments are ARID/Bright DNA binding domains of different transcription factors including ARID3A and Dri.[272–275] We employed PSN (see Chapter 4.5.) and MD simulations (see Chapter 4.1.) to study the conformational ensemble on the microsecond time scale of this poorly explored class of TFs.[276] The definition of nodes and edges that occur with highest probability in the paths of long-range communication allowed identifying important residues in the dynamic ARID domain network. These edges and nodes are likely to be the most important component for transmitting the effects over long distances. We showed that structural communication to the DNA binding loops in ARID domains can pass through a subset of conserved hubs that are also known experimentally to affect protein function, stability and the interaction with DNA.[277] The effects of single mutations on the communication paths helped identify helix 5 as a central region for allosteric communication. We identified pre-existing communication paths in the unbound ensemble to the DNA binding loops that are strengthened by the interaction

with DNA. The region surrounding a tyrosine residue (tyrosine 119) of helix 5 has been proposed as hotspot for cofactor recruitment and it is likely to be allosterically regulated by DNA interaction through communication with the DNA-binding loop (Figure 14). This mechanism resembles the conformational modulation of cofactor recruitment interfaces in p53 DNA binding domain, suggesting that this can be a common mechanism for TFs' DNA binding domains.

4. Computational methods to study structure and dynamics of loops and linkers

4.1. Molecular dynamics simulations

Thanks to general advances in computer power, the development of dedicated hardware and advances in software and algorithms, atomistic molecular dynamics (MD) simulations are emerging as powerful approaches to characterize at the atom-level protein dynamics on timescales from picosecond to microsecond and even milliseconds in few cases.[14,16,278,279] Coarse-grained descriptions, where the physical model to describe the protein and the environment is further simplified, may also be used to describe general patterns in protein dynamics, for example when combined with elastic network models.[280,281] Most other computational approaches classify protein dynamics according to the magnitude of the fluctuations and the conformational change but they lose the details of the underlying time scale. In contrast, MD has the potential to describe directly the dynamical properties, though further efforts are still needed to examine the extent to which MD simulations can capture the timescales and mechanisms of conformational dynamics. The MD community itself has often had to compromise to study in a statistically significant

way conformational changes occurring on long time scales, relying for example on enhanced sampling approaches (see Chapter 4.2.).

4.2. Enhanced sampling atomistic simulations

The flexibility of highly dynamical elements, such as loops, hinges and linkers, can often result in their structure being difficult to observe with high-resolution structural biology techniques, such as X-ray crystallography or NMR. A typical example of this is the activation segment of protein kinases (discussed in Chapter 3.3.), which is missing in most deposited crystallographic structures and is mostly invisible in NMR spectra. Molecular dynamics simulations can be extremely helpful in these cases, recovering an atomic level description of the structure and dynamics of these elements and allowing investigating the complex underlying allosteric communication. MD is often based on a full-atom classical description of the system, whose dynamics is obtained by integrating Newton's equations of motion.[282] Complex potential functions involving a large set of carefully determined parameters (a so-called "force field") allow mimicking the physical evolution over time for the system under consideration. The efficacy of MD force fields in predicting the natural dynamics of biological systems is continuously tested as more accurate potentials are constantly being developed.[283–285]

Using long multi-μs MD simulations Shan and coworkers[286] were able to identify an allosteric pocket between the αFG and αGH loops in Src kinase, invisible in the X-ray structures, able to bind the tyrosine kinase inhibitors Dasatinib and PP1. Similarly, Sahun-Roncero and coworkers[191] were able to rationalize the allosteric effects and negative cooperativity of ligands binding to the choline kinase. The ligands were shown to affect the hinge motions of the protomers in the dimer. However, these successful examples were possible due to the intrinsically short time scales of the

events observed and in some cases the use of special purpose machines, like Anton.[287] More generally, conventional MD simulations are limited by the time-scales that can be directly accessed. While a typical atomistic MD simulation lasts up to a few microseconds, the conformational changes involved in allosteric regulation can take place on much longer time-scales. Different approaches have been developed in the last decades to overcome this so-called "time-scale problem". These methods, generally referred to as "enhanced sampling" methods, are numerous and discussing all of them is beyond the scope of this review. Here we will concentrate on the two categories that are often used to sample conformational changes linked to allostery: 1) those that run different replicas of the same system under different conditions (multiple replica methods) and 2) those that apply a bias potential to enhance the sampling of (and reconstruct the free energy as a function of) a set of collective variables (CVs) that describe the reaction coordinate (CV-based methods).

The first class of methods is exemplified by Replica Exchange MD (REMD).[288] In its most popular implementation, Parallel Tempering (PT),[289] several copies of the same system are simulated at different temperatures. Exchanges of conformations between replicas are attempted periodically and accepted with Metropolis-like criteria. As the systems at higher temperatures are able to sample a larger portion of the conformational landscape, the exchange progressively enhances the sampling of these regions at lower temperatures as well. PT has been extensively used to study large conformational rearrangements involved in folding.[290–294] Using a variant of PT, Kubitzki and de Groot were able to study the opening of the *E. coli* Adenylate Kinase (ADK).[295] In that study, PT is used to sample the rearrangements of two small domains, AMPbd and LID, constituting respectively the AMP and ATP binding sites of ADK, and to investigate the chain of events occurring during the transition. Recently, PT was

used by Palazzesi and coworkers to describe the allosteric communication in the KIX domain of the CREB binding protein.[296] Two conformational states are possible when KIX is associated with the transcription factor Mixed Lineage Leukemia (MLL), one of which is "invisible" to experimental techniques due to the short lifetime. The authors were able to characterize the minor state and rationalize the higher affinity of ligands binding to KIX in the presence of MLL.

One of the oldest CV-based methods, which are still frequently used, is umbrella sampling (US).[297] In US simulations, a harmonic "umbrella-like" potential is added to the system as a function of a few collective variables (CV), which are, in turn, functions of all the atomic coordinates. As PT, US has been extensively used to study large-scale allosteric conformational transitions. For instance in adenylate kinase US was used to show that a population shift mechanism was involved in the allosteric transition.[144] In dihydrofolate reductase, large scale functional rearrangements of the Met20 loop were observed.[298] US was also used to study the hinge motion of the ribose-binding protein (RBP).[299] RBP is formed by two homologous domains connected by a 3-strand hinge region. Binding of the substrate induces the closure of the two domains on the ribose. The reported results suggest that, while the open state is the most stable minimum in the apo structure, closed conformations are still relatively populated and, upon binding, a shift in population towards the closed structure occur. The authors were also able to establish that the higher stability of the apo open state is mainly due to the conformational entropy. Roux and coworkers also employed US to study the large conformational rearrangements of the activation loop of Src kinase (see also Chapter 3.3.).[300] The authors observed that, in the absence of phosphorylation, the activation segment is highly dynamical, sampling several conformations, thus not being restricted to the catalytically competent one. Upon phosphorylation, a narrower

region of the conformational space is explored, consistent with the kinase being locked in its catalytically competent state. US simulations of a kinase of the same family, Hck, also concluded that the closed A-loop conformation is the most stable in the absence of phosphorylation.[301]

Besides US, another CV-based method that has gained considerable momentum in recent years is Metadynamics.[302,303] This method and its variants are available for several MD simulation software thanks to the open-source PLUMED plugins.[304,305] In metadynamics, a history-dependent bias is progressively added to the system in the form of small Gaussians gradually disfavoring regions of the conformational space that have already been visited. In this way, Metadynamics simulations both act to enhance sampling by driving the system to sample new regions of the conformational landscape, and at the same time provides an imprint of the free energy landscape of those regions that have been explored. Metadynamics was used to study and reconstruct the free energy landscape associated with the large conformational transition of the A-loop of EGFR required for receptor activation and the allosteric effect of oncogenic mutations on the equilibrium between the open and closed states of the loop.[306] Metadynamics was also used to study the open-to-closed conformational transition of the A-loop in other bio-medically important kinases such as B-Raf.[307,308] In Abl these methods were used to understand the activation by the SH2 regulatory domain in the "top-hat" position.[192] Metadynamics has also been combined with RE to understand the mode of action of the allosteric FGFR inhibitor SSR (see also Chapter 3.3.).[219]

4.3. Elastic Network Models

Methods based on elastic network models (ENMs) have been extensively used to investigate flexibility in proteins,[41,280] helping to predict dynamics and get insights into the molecular mechanisms of function and allosteric regulation.[19,41] ENMs rely

on harmonic models that describe molecular systems, like proteins, as networks composed by nodes, corresponding to the positions of atoms from the 3D structure, connected by edges that are the interactions between nodes. In such models the nodes usually correspond to atoms or single residues (the position of the Cα atoms), but some models use different levels of resolution, e.g. a single node for representing multiple residues.[309] The existence of models with different level of resolution in the coarse-grained representation of the molecular system provides large scalability in ENM techniques. In ENMs, a contact is counted between a pair of atoms when a distance that is smaller than an empirically chosen cut-off separates them. The interactions are represented by springs, constituting harmonic restraints over the motions from the starting structure. Based on harmonic potential, ENM methods predict the collective motions starting from only the 3D structure of the protein and the associated network of native contacts. Such methods assume that the starting structure is a minimum energy conformation, e.g. variations from that structure result in an increase in energy and net force aimed to restore the equilibrium. With this approximate representation, ENMs can accurately predict motions and behavior only near the global energy minimum, such as for example starting from experimental structures. ENMs were used in combination with Normal Mode Analysis methods (NMAs) and approaches based on Principal Component Analysis (PCA).[310] Commonly employed ENM models are the so-called Gaussian network model (GNM),[311,312] anisotropic network model (ANM),[313] and torsional network models (TNM).[314] For example, ENM-based approaches can identify protein dynamics at the equilibrium, e.g. describe from the slowest modes the changes between different substates of a protein, in agreement with X-ray crystallography and show correspondence with collective motions described from MD simulations.[315,316] ENMs

permit to predict motions that occur over timescales of microseconds or slower that are considered to be important for the global dynamics of the proteins and in some cases proposed to be functionally relevant.[19] In this context, the ENM methods are useful tools to investigate global motions and predict the accessible conformational changes near the native state. ENMs have been used to get insights into a variety of motions and biological mechanisms associated with protein functions, where loops and linker regions or hinge motions play a crucial role.[19,62,317–322] They also predict the residues or protein regions most involved in the protein motions.[41] They can thus provide a suitable alternative to atomistic MD simulations when the study of slow timescale motions by MD becomes too computationally expensive, such as in very large macromolecular complexes. Since ENMs are simple approaches and depend on a small number of parameters they allow a fast analysis of complex biomolecular systems. An advantage of ENM models is that they can be applied to a wide range of biological systems, such as large protein complexes,[323] membrane proteins,[324–326] and ribosomes[327–330] thanks to their relative simplicity, scalability and their low computational cost. Web servers have also been developed to apply ENMs to investigate motions in proteins and biomolecules, such as the recent ANM 2.0 (http://anm.csb.pitt.edu) to cite an example.[331] It is however important to keep in mind some limitations in the application of ENMs to avoid misleading interpretations. For example, ENMs do not provide information about the specific time scale of the predicted motions. Moreover, it is generally assumed that ENMs are most useful to predict global motions and investigate long-range processes and allosteric effects, but may lack accuracy in the identification of local motions; this assumption, however, appears in contrast to the fact that the harmonic expansion is inherently most accurate locally. Additional limitations can arise from the definition of distance cut-offs for

harmonic contacts. Because of the harmonic approximation, ENMs can accurately predict only motions near the global energy minimum and are limited in predictions of conformational transitions to minor states, although recent applications are exploring this area.[332]

4.4. Methods based on higher-order statistics

To understand allosteric mechanisms it is essential to investigate conformational ensembles of proteins and quantitatively describe the populations of sub-states, as well as motions enabling the transitions between them.[39,114] As mentioned above, MD simulations are promising tools to describe protein conformational ensembles at the atomic level (see Chapters 4.1. and 4.2.). The possibility to execute one or more microsecond-length, or indeed millisecond, MD simulations leads to a large amount of data. Handling the analyses of such large amounts of MD trajectories and making the best out of them is challenging. In particular, the community now needs new or enhanced tools to efficiently and accurately detect conformational changes in protein ensembles. Visual inspection of simulation trajectories is still a crucial initial component, but becomes difficult as the amount of data increases. Thus, semi-automated analysis methods are often necessary to focus attention on key events in simulations, and also provide the starting point for more quantitative analyses. Analysis tools based on principal component analysis of MD trajectories have long been used to detect major conformational changes in proteins.[333] Despite their utility, such methods have shown limitations in their accuracy to identify populations and may not be sufficient to describe the complex conformational landscape of proteins.[334–337]

To overcome these issues, novel methods based on the so-called higher-order statistics have been recently proposed for the analysis of highly multidimensional

MD ensembles.[338–340] Higher-order statistics methods can reveal motions and characterize sub-states, including minor and transient structural populations that are challenging to unveil with alternative analysis tools. The main advantage of these methods is that they are able to identify also the individual components underlying the global dynamics, such as specific structural regions or subset of residues. Here we provide examples of some of the most recent such methods, their major advantages and limitations, as well as their application in the study of allostery and structural communication.

Shaw and co-workers developed TimeScapes, a package of analysis methods to structurally investigate micro-millisecond MD ensembles.[338] The TimeScapes approach is based on the observation that structural events in proteins are frequently associated with alterations in the contacts between residues. TimeScapes provides a time-dependent measure of the evolution of intramolecular contacts, indicated as the "activity" of the system, measuring broken and formed contacts between side-chains of pair of residues during MD simulations. It uses a coarse-grained model to describe the residues and employs a contact metric based on a method to estimate time-dependent contact graphs and identify significant alterations in the ensemble of conformations. The measure of activity permits one to monitor the global evolution of the protein during the MD trajectory, identifying periods of low activity or "basin" that correspond to local stable sub-states of the protein, separated by periods of high activity or "transitions". Moreover, the TimeScapes tools have the advantage of structurally localizing the individual components that underlie the conformational changes along time, where these are associated with the formation and breaking of contacts. Such analysis has higher sensitivity and describes detailed structural

information that is difficult to extract with methods based on Cartesian coordinates. Since the TimeScapes analysis depends on variation of contacts, biological events that do not significantly alter the proteins cannot be investigated in detail. One example of TimeScapes applications is provided by the study of the effects of inhibitory peptides on the conformational ensemble of a bacterial toxin, describing the molecular mechanisms associated with the inhibition process and the related alterations in protein dynamics.[341] In another application, it has been used to characterize the allosteric properties of kinesin motor domains associated with ATP hydrolysis.[342]

Recently Shaw and collaborators also proposed another statistical method called "simultaneous penalized likelihood estimation" (SIMPLE).[343] The SIMPLE method starts from a large number of time-series measurements, such as distances between atoms monitored during MD trajectory, and identifies the events corresponding to collective changes in a subset of specific measurements. SIMPLE describes which subset of input measurements are involved in each conformational change and the associated structural elements in the protein. The authors reported that SIMPLE has better performance and accuracy in detecting conformational populations and changes during protein simulations than other current methods used to analyze MD simulations, such as PCA.

Agrawal and collaborators proposed quasi-anharmonic analysis (QAA) a tool based on the usage of higher-order statistics of positional deviations obtained from MD ensembles.[337,339] QAA uses fourth-order statistics to investigate anharmonic fluctuations of atomic positions during MD simulations and organize the conformational landscape in a multi-level hierarchy of sub-states, which allows one to extract details related to structural changes. QAA has been used to

analyze MD simulations of ubiquitin, 0.5 µs of sampling time and T4 lysozyme, highlighting the presence of substates that resemble those in molecular recognition of multiple binding partners and even substrates in the unbound states of the proteins.[336]. QAA approaches have been recently extended to develop a tool to analyze long time-scale MD trajectories, called Higher-Order Statistics Toolbox for Molecular Dynamics, or HOST4MD, which is based on higher-order statistics and QAA.[339] HOST4MD was used to investigate allosteric motions in the enzyme adenylate kinase in its unbound forms, pointing out the dynamical coupling between sub-domains and structural transitions among open/closed states.

The higher-order statistic methods discussed here require long simulations in the time-scale of micro- millisecond to provide effective results since they suffer from limitations in sampling performed during MD simulations and approximations that are still present in the force fields.[14] These tools, however, can accurately analyze large and complex datasets as the one produced by MD simulations, in which there are many measured observables, permitting to efficiently capture the biologically relevant conformational changes. They provide an advantage in identifying the individual structural elements or even the individual residues involved in the conformational change, thus allowing dissecting the molecular mechanisms underlying the conformational changes and relating them to allosteric events and biological functions.

4.5. Protein Structure Networks

Methods inspired by graph theory and relying on the network paradigm have been used to describe dynamics and structural properties of proteins.[15,43,44,344–346] The weak intramolecular interactions in a protein can be collectively represented in the form of

a network, i.e. a Protein Structure Network (PSN), where the residues are the nodes of the network and they are connected by links (edges) that depend on their interaction strength or their energetic coupling. PSN can be derived in many different ways and recently many of them have been compared according to the way in which the edges are defined. PSNs are generally 'small worlds',[347,348] a property important for fast transmission of the conformational change over long distances. The residues can communicate through the shortest paths available and multiple paths have been observed between the same pair of distal sites with many nodes in common.[37,276] The capability of PSN methods to define paths of communication over long distances makes them a suitable technique for prediction of allosteric effects in protein structural ensembles. PSN methods can be, in principle, applied to conformational ensembles derived by NMR, MD, ENM or other coarse grained methods. PSN approaches can thus be used to provide a subset of paths of communication between distal residues that are likely to be the paths that allow the propagation of structural communication or even predict allosteric hot spots in a protein structure. Examples of these approaches are reported in Chapters 3.4 and 3.5.

It is important to couple these methods to experimental validation. Indeed, we cannot know a priori if the paths that have been calculated in a structural ensemble are really relevant for the experimentally observed protein dynamics.[15] Some examples are available supporting a relationship between calculated networks and the ones observed experimentally but they might still be limited to specific cases and additional studies are required.[46,48]

5. Combining experimental biophysics and simulations to study structural dynamics and correlated motions

Computational methods such as those described above have provided a wealth of information about the structural dynamics of proteins and their functional effects via allosteric regulation or long-range communication during functional cycles. Methods such as molecular dynamics or Monte Carlo simulations can provide an atomic level description of these processes and, in certain cases, the timescales and energetics involved. As discussed in Chapters 4.1. and 4.2, simulations are, however, inherently limited by the accuracy of the biophysical models (e.g. force fields and energy functions) as well as the timescales that can be sampled,[278]. There has been considerable progress in the accuracy of simulation methods and, combined e.g. with enhanced sampling procedures (Chapter 4.2.), one may study processes that occur on a wide range of timescales that include fast (ps-ns) local motions, loop motions on various timescales (Chapter 2.1.2.) as well as slower motions on the μs–ms, or beyond, that involve longer range correlations and cooperative dynamics. Despite this progress towards fully predictive computational approaches, experimental studies remain an integrated part of our understanding of structural dynamics and allosteric communication in proteins. Below we provide an overview of different kinds of approaches for combining experiments and simulations, followed by examples of how these approaches have been used to study protein motions.

5.1. Validating simulations

In the simplest - but important - cases, experiments can provide detailed and accurate benchmarks for validating computational studies. Comparisons between experimental NMR relaxation order parameters and MD simulations provide increasingly strong evidence for the ability of simulations to model accurately fast motions in the polypeptide backbone when the simulations are sufficiently long, a suitable force field is used,[349] and simulations and experiments are compared in an appropriate fashion.[350]

Similarly, comparisons between simulations and NMR residual dipolar couplings (RDCs), scalar couplings[285,351] or chemical shifts,[352] suggest that simulations can also model reasonably well the slower motions. Such comparisons between simulations and experiments lend credibility to our ability to model e.g. loop motions, which in turn provides opportunities for purely computational procedures such as those described above (see Chapter 2.1.2.).[59]

5.2. Improving biophysical models

Sometimes the comparison between simulation and experiments leads to differences beyond the experimental uncertainty. Once it has been ruled out that such differences are not due to limited sampling or inaccuracies in the way experiments and simulations are compared, the results may provide useful clues about remaining deficiencies in the models (force fields, energy functions, statistical potentials) used.[353,354] Importantly, such comparisons may subsequently be used as targets for force field optimization where the aim is to change the parameters in the energy function so as to improve agreement with experiments.[355–357] As in any such optimization procedure it is important to separate the data used for optimization from that used in subsequent testing and validation of the resulting models. A particularly promising strategy is targeting directly biophysical data, often from NMR spectroscopy, measured on proteins or peptides. While such measurements do not provide restraints for all parts of the energy functions, they provide extremely useful supplements that probe and restrain the particular cooperative and long-range interactions that separate the properties of macromolecules from those of small-molecule analogs.

5.3. Restraining molecular simulations

An alternative approach for using experimental data to modify simulations is to integrate the experimental measurements directly into the generation of conformational ensembles. Numerous methods for doing so exist, and a comprehensive overview is beyond the scope of this review. At the most general level, the different methods can be separated into two classes. In the first, one generates a large set of conformations by purely computational means without using experimental data, and then subsequently selects a subset of these structures to generate an ensemble that agrees with experiments.[358] In other approaches, these two steps are merged so that the experimental data are used directly during the process of generating of the ensemble.[359] While the two approaches differ in certain aspects they can often be thought of as two different strategies for obtaining the same goal. In particular, in a Bayesian framework, the two methods can be thought of as combining a set of prior information (typically a transferable force field) with new information (in the form of the experimental data on the system) to generate a posterior distribution (the final conformational ensemble).[360–363] In this way, the approach can both conceptually and formally be regarded as modifying an existing, transferable energy function using experimental data to generate a new, system-specific energy function. Thus, in this way the experimental data is used to remove some of the uncertainty and potential errors associated with using non-perfect energy functions.[340] Recent developments have strengthened considerably our fundamental understanding of the theory behind such approaches, and they can in turn provide a mechanistic framework for understanding allostery.[38]

5.4 Targeting experimentally observed processes

The three areas described above represent well-defined approaches in which experimental data can be used to validate, optimize or modify molecular simulations.

Such approaches require, however, quantitative data that can be compared with simulations. Often, a more informal approach is taken to combine experiments and simulations. In particular, one often has less easily quantifiable data that probe conformational dynamics or structural changes. These might include crystal structures of a protein in different states, but where the relative populations or energetics involved are not known. Mutational studies can provide hints towards long-range allosteric networks where e.g. a mutation is shown to have a quantified effect distal to the mutation site, but where again the structural or dynamical details are neither probed nor known. In such cases, the experimental data can sometimes be used to design simulation studies that target directly the processes probed by the experiments, but where quantitative validation or modifications are difficult.

5.5. Specific examples of combining experiments and simulations

Below we describe a few specific examples that illustrate the approaches outlined above, with a focus on how simulations and experiments together can be used to understand the molecular details and structural and dynamical origins of allostery and long-range communication in proteins.

NMR relaxation methods can be used to probe local order parameters, S^2, that report on the amplitudes of backbone and side chain motions in proteins. Such measurements provide valuable data for validating molecular force fields and have in particular been useful in highlighting how earlier force fields resulted in erroneously large amplitude loop motions at the relevant timescales.[349] It is, however, much more difficult to probe correlated motions, in particular over longer length scales, using NMR relaxation measurements. In an elegant study, Mayer and co-workers combined mutagenesis and NMR measurements to probe correlated dynamics in the IgG-binding B1 domain of protein G (Protein G).[364] They measured local backbone

dynamics in wild type Protein G as well as in ten mutants in which position 53 near the C-terminus was substituted for a range of different amino acids. The different mutations led to varying changes of the backbone dynamics, and by examining whether these mutation-induced changes of S^2 at two sites were correlated they were able to provide a "correlation map" of the dynamics in Protein G. The results reveal both local and long-range correlated changes in S^2 values and suggest that at least some of these report on correlated motions within the protein. However, such experiments can only provide indirect evidence for the correlated movements of atoms in a protein. To gain more direct atomic-level insight as well as provide a structural interpretation of the experimental data, Lange et al performed molecular simulations of wild type protein G and calculated both S^2 values and a map of the correlated motions.[365] The simulations revealed quantitative agreement with the experimental S^2 values for the wild type protein. Analysis of the correlated motions suggested that these are mostly local, with only much weaker correlations between spatially distant amino acid residues. While the overall picture is similar, this observation differs from that obtained from the mutational analysis.[364] As no simulations were performed of the mutant proteins, it is unclear whether the apparent wider spread of long-range correlations observed in experiments is an outcome of using mutations to probe correlations, or whether the simulations underrepresent such correlations.

The issue of correlated motions in the polypeptide backbone has also been explored using NMR RDCs and trans-hydrogen bond scalar couplings, again using protein G as a model system.[366] In that work, Bouvignies et al used a very large set of RDCs to generate an ensemble of peptide plane orientations that fit the data. Thus, in this work these experimental data were used directly to obtain the ensemble. The analysis

revealed an intriguing pattern of motions with alternating small- and large-amplitude motions along each beta-strand that was repeated across the entire beta-sheet in protein G. This pattern suggests correlated motions across the hydrogen bonds that connect the beta-strands, an observation supported further by comparing the ensemble to trans-hydrogen bond scalar couplings that were not used in fitting. When the motions of pairs of hydrogen-bonded peptide planes were assumed to move in a concerted and correlated fashion these independent scalar couplings were more accurately predicted.

The same kind of correlated, hydrogen-bond-mediated backbone fluctuations have also been studied more directly by MD simulations restrained with experimental RDCs. Using ubiquitin as a model system, Fenwick et al generated an ensemble of ubiquitin conformations that simultaneously represents the information encoded in a modern and accurate force field and at the same time satisfies the information in the experiments.[367] Analysis of the resulting conformational ensemble clearly reveals correlated motions not only locally; but also between residues that were spatially distant across the entire beta-sheet. While such long-range correlated motions are extremely difficult to detect directly from the experimental data, both this example and that of protein G above demonstrate how computational models restrained by experiments can help extract such information from experiments.

Allosteric effects often involve the ability of a protein to adopt two or more distinct and different conformations, whose relative stabilities depend e.g. on the binding of ligands or changes in the environment. Adenylate kinase is a multi-domain enzyme that converts ATP and AMP into two molecules of ADP. The enzyme has three domains and crystal structures reveal substantial domain reorganization between a central core domain and the two ATP and AMP binding domains when substrates

bind. By generating collections of structures and selecting ensembles that fit the experimental RDCs, Esteban-Martin et al provided strong evidence for correlated interdomain motions where the two domains appeared to close in a concerted fashion.[368]

Crystal structures may also be used as experimental data to study mechanisms of conformational exchange. Again using adenylate kinase as an example, Wang et al used a multi-state structure-based model to describe the conformational free energy landscape.[369] In such an approach, a system-specific energy function is constructed so as to have distinct minima at the known crystallographically-resolved states. Combined with an additional physical description of protein electrostatics, this model allowed the authors to observe transitions between the various states and thus to describe the mechanism of exchange that could not be obtained directly from crystal structures.

As structure-based models are based on the assumption that state-specific contacts dominate the shape of the entire free energy landscape it is desirable to be able to portray conformational exchanges by a more physics-based description. The functional effect of the structural dynamics in the enzyme cyclophilin A has recently been studied using a range of techniques. NMR experiments revealed extensive motions in this small enzyme with an exchange between at least two distinct conformations that occur on the timescale of catalysis both in the presence and absence of substrates.[370] A detailed analysis of the electron density from room-temperature X-ray diffraction experiments provided hints to a minor, transiently populated conformation.[371] Based on the structural difference between the two states the authors designed a mutation that substantially perturbed the conformational exchange process. Intriguingly, despite the mutational site being ~15Å away from the

central catalytic residue, the perturbed dynamics causes a several-hundred-fold drop in the catalytic efficiency demonstrating the long-range nature and functional importance of protein motions and conformational changes. To explore this conformational exchange process and the associated free energy landscape, Papaleo et al used all-atom, enhanced sampling metadynamics simulations with.[48] The simulations used as starting point the experimental information about the regions in which the motions were expected. In particular, the authors used the key conformational differences between the two states, observed in the electron density maps, as collective variables to drive the simulations. The resulting physics-derived free energy landscapes revealed a very good agreement with the experimental structural information, and the relative populations of the two states - the wild-type protein and the mutant where the residue substitution perturbed the residue network. Analyses of the conformational ensembles suggested a communication pathway that transmits the structural information from the mutation site to the active-site residue; this pathway was also observed in an independent analysis of the experimental data,[46] again indicating that simulations and experiments can converge to the same answers.

As described above X-ray diffraction data can be used to study not only the dominant structure of a protein, but also to reveal information about excursions away from this state. In the context of allosteric effects, diffuse (non-Bragg) diffraction is particularly appealing as it has the potential to provide information about correlated motions in proteins.[372] An early application of this technique to Staphylococcal nuclease implied liquid-like motions with correlations up to distances of 10Å.[373] Early simulations suggested that it might be difficult to sample these correlated motions,[374] but with increased computing power and more efficient algorithms for simulations it has recently become possible to compare MD simulations and diffuse scattering

experiments.[375] While the simulations resulted in good agreement with the overall properties, less good agreement was observed with the anisotropic motions probed by experiments. This both suggests areas for improvement and points to the fact that additional information might be extracted by restraining the simulations with diffuse X-ray scattering data.[372]

The examples described above provide a glimpse of the many ways through which experiments can be used to study allostery and correlated motions, and how a direct integration of experiments and simulations provides a powerful approach to extract information and provide structural models. As experimental and computational approaches continue to improve we expect that such integrative approaches may help reveal new principles and provide quantitative descriptions of protein allostery. A recent intriguing example has been provided by NMR spectroscopy that demonstrated very clearly how allosteric signaling may occur in the apparent absence of conformational changes, and how changes in protein dynamics may provide a powerful driving force for transmitting signals across proteins.[376,377] However, noteworthy, even though it may seem that allostery took place through sheer dynamics because a structural comparison between the active and inactive states does not detect a conformational change, it does not mean that there is no such change.[55] In general, the lack of observable conformational change may be due for example to crystallization conditions and crystal effects, because one of the states is disordered, due to structural comparisons disregard the quaternary protein structure, overlooking synergy effects among allosteric effectors including DNA and proteins and graded incremental switches,[378] too short molecular dynamics simulations and more. New and improved combinations of experiments and simulations are needed to understand

and fully describe the origins, modes and functional consequences of the expanding view of allostery.

6.   Concluding remarks

Even when it comes to the function of mostly disordered regions, such as loops or linkers, unrelated proteins that are characterized by different sequences and folds can share common themes. These regions undergo many conformational changes in the free and in the bound states and by mediating conformational and dynamic changes they often exert their effects over long distances. It is challenging to unravel - in atomistic detail - these complex conformational changes and motions, which often occur over a diverse range of scales in time and space. Integrating computational and experimental techniques may help reach this aim, and systematic studies may help to identify common properties. Today it is clear that loops and linkers do not act merely as flexible connectors between distinct folded domains or secondary structural elements of a protein; instead they have an active – and critical - role in protein function and allostery. As such they provide powerful biological design principle.[379] Recall the age-old paradigm of 'not observed does not imply that it is not there'. Even if subtle and not visually observed for all the reasons discussed above, allostery induces conformational and dynamic changes which control function, and loops and linkers – as well as chain termini not discussed here - are key components designed, optimized and retained by evolution to mediate and modulate its action.

Author Information

Corresponding author

* Tel. 0045-35257711, email: elenap@cancer.dk

References

(1) Chothia, C.; Gough, J.; Vogel, C.; Teichmann, S. a. Evolution of the Protein Repertoire. *Science* **2003**, *300* (5626), 1701–1703.

(2) Vogel, C.; Bashton, M.; Kerrison, N. D.; Chothia, C.; Teichmann, S. a. Structure, Function and Evolution of Multidomain Proteins. *Curr. Opin. Struct. Biol.* **2004**, *14* (2), 208–216.

(3) Bhattacharyya, R. P.; Rem, A.; Yeh, B. J.; Lim, W. A. Domains , Motifs , and Scaffolds : The Role of Modular Interactions in the Evolution and Wiring of Cell Signaling Circuits. *Annu Rev Biochem.* **2006**, *75*, 655–680.

(4) Wriggers, W.; Chakravarty, S.; Jennings, P. a. Control of Protein Functional Dynamics by Peptide Linkers. *Biopolym. - Pept. Sci. Sect.* **2005**, *80* (6), 736–746.

(5) Ma, B.; Tsai, C.-J.; Haliloğlu, T.; Nussinov, R. Dynamic Allostery: Linkers Are Not Merely Flexible. *Structure* **2011**, *19* (7), 907–917.

(6)   Bashton, M.; Chothia, C. The Generation of New Protein Functions by the Combination of Domains. *Structure* **2007**, *15* (1), 85–99.

(7)   Pawson, T. Assembly of Cell Regulatory Systems Through Protein Interaction Domains. *Science (80-.).* **2003**, *300* (5618), 445–452.

(8)   Peisajovich, S. G.; Garbarino, J. E.; Wei, P.; Lim, W. A. Rapid Diversification of Cell Signaling Phenotypes by Modular Domain Recombination. *Science (80-.).* **2010**, *328* (5976), 368–372.

(9)   Marcotte, E. M.; Pellegrini, M.; Ng, H. L.; Rice, D. W.; Yeates, T. O.; Eisenberg, D. Detecting Protein Function and Protein-Protein Interactions from Genome Sequences. *Science* **1999**, *285* (5428), 751–753.

(10)   Fukuchi, S.; Homma, K.; Minezaki, Y.; Nishikawa, K. Intrinsically Disordered Loops Inserted into the Structural Domains of Human Proteins. *J. Mol. Biol.* **2006**, *355* (4), 845–857.

(11)   Vogel, C.; Berzuini, C.; Bashton, M.; Gough, J.; Teichmann, S. A. Supra-Domains: Evolutionary Units Larger than Single Protein Domains. *J. Mol. Biol.* **2004**, *336* (3), 809–823.

(12)   van den Bedem, H.; Fraser, J. S. Integrative, Dynamic Structural Biology at Atomic Resolution—it's about Time. *Nat. Methods* **2015**, *12* (4), 307–318.

(13)   Palmer, A. G. Enzyme Dynamics from NMR Spectroscopy. *Acc. Chem. Res.* **2015**, *48* (2), 457–465.

(14)   Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **2012**, *41*, 429–452.

(15)   Papaleo, E. Integrating Atomistic Molecular Dynamics Simulations, Experiments, and Network Analysis to Study Protein Dynamics: Strength in Unity. *Front. Mol. Biosci.* **2015**, *2*, 28.

(16)   Orozco, M. A Theoretical View of Protein Dynamics. *Chem. Soc. Rev.* **2014**, *43* (14), 5051–5066.

(17)   Flores, S. The Database of Macromolecular Motions: New Features Added at the Decade Mark. *Nucleic Acids Res.* **2006**, *34* (90001), D296–D301.

(18)   Gerstein, M.; Echols, N. Exploring the Range of Protein Flexibility, from a Structural Proteomics Perspective. *Curr. Opin. Chem. Biol.* **2004**, *8* (1), 14–19.

(19)   Bahar, I.; Chennubhotla, C.; Tobi, D. Intrinsic Dynamics of Enzymes in the Unbound State and Relation to Allosteric Regulation. *Curr. Opin. Struct. Biol.* **2007**, *17* (6), 633–640.

(20)   Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. A Hierarchy of Timescales in Protein Dynamics Is Linked to Enzyme Catalysis. *Nature* **2007**, *450* (7171), 913–916.

(21)   Karplus, M.; Kuriyan, J. Molecular Dynamics and Protein Function. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6679–6685.

(22)   Haliloglu, T.; Bahar, I. Adaptability of Protein Structures to Enable Functional Interactions and Evolutionary Implications. *Curr. Opin. Struct. Biol.* **2015**, *35*, 17–23.

(23)   Tsai, C. J.; Kumar, S.; Ma, B.; Nussinov, R. Folding Funnels, Binding Funnels, and Protein Function. *Protein Sci.* **1999**, *8* (6), 1181–1190.

(24)   Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R. Folding Funnels and Binding Mechanisms. *Protein Eng.* **1999**, *12* (9), 713–720.

(25)   Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. Folding and Binding Cascades: Dynamic Landscapes and Population Shifts. *Protein Sci.* **2000**, *9* (1), 10–19.

(26)   Ma, B.; Shatsky, M.; Wolfson, H. J.; Nussinov, R. Multiple Diverse Ligands Binding at a Single Protein Site: A Matter of Pre-Existing Populations. *Protein Sci.* **2002**, *11* (2), 184–197.

(27)   Monod, J.; Changeaux, J. P.; Jacob, F. Allosteric Proteins and Cellular Control Systems. *J. Mol. Biol.* **1963**, *6*, 306–329.

(28)   Koshland, D. E. The Key–Lock Theory and the Induced Fit Theory. *Angew. Chemie Int. Ed. English* **1995**, *33* (2324), 2375–2378.

(29)   Monod, J.; Wyman, J.; Changeux, J. P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12*, 88–118.

(30)   Lu, S.; Huang, W.; Wang, Q.; Shen, Q.; Li, S.; Nussinov, R.; Zhang, J. The Structural Basis of ATP as an Allosteric Modulator. *PLoS Comput. Biol.* **2014**, *10* (9), e1003831.

(31)   Lu, S.; Deng, R.; Jiang, H.; Song, H.; Li, S.; Shen, Q.; Huang, W.; Nussinov, R.; Yu, J.; Zhang, J. The Mechanism of ATP-Dependent Allosteric Protection of Akt Kinase Phosphorylation. *Structure* **2015**, *23* (9), 1725–1734.

(32)   Malmstrom, R. D.; Kornev, A. P.; Taylor, S. S.; Amaro, R. E. Allostery through the Computational Microscope: cAMP Activation of a Canonical Signalling Domain. *Nat. Commun.* **2015**, *6*, 7588.

(33)   Nussinov, R.; Ma, B.; Tsai, C.-J.; Csermely, P. Allosteric Conformational Barcodes Direct Signaling in the Cell. *Structure* **2013**, *21* (9), 1509–1521.

(34)   Nussinov, R.; Tsai, C.-J.; Xin, F.; Radivojac, P. Allosteric Post-Translational Modification Codes. *Trends Biochem. Sci.* **2012**, *37* (10), 447–455.

(35)   Csermely, P.; Palotai, R.; Nussinov, R. Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events. *Trends Biochem. Sci.* **2010**, *35* (10), 539–546.

(36)   Ma, B.; Nussinov, R. Enzyme Dynamics Point to Stepwise Conformational Selection in Catalysis. *Curr. Opin. Chem. Biol.* **2010**, *14* (5), 652–659.

(37)   del Sol, A.; Tsai, C.-J.; Ma, B.; Nussinov, R. The Origin of Allosteric Functional Modulation: Multiple Pre-Existing Pathways. *Structure* **2009**, *17* (8), 1042–1050.

(38)   Fenwick, R. B.; Esteban-Martín, S.; Salvatella, X. Understanding Biomolecular Motion, Recognition, and Allostery by Use of Conformational Ensembles. *Eur. Biophys. J.* **2011**, *40* (12), 1339–1355.

(39)   Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5* (11), 789–796.

(40)   Meireles, L.; Gur, M.; Bakan, A.; Bahar, I. Pre-Existing Soft Modes of Motion Uniquely Defined by Native Contact Topology Facilitate Ligand Binding to Proteins. *Protein Sci.* **2011**, *20* (10), 1645–1658.

(41)   Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. Global Dynamics of Proteins: Bridging between Structure and Function. *Annu. Rev. Biophys.* **2010**, *39*, 23–42.

(42)   Liberles, D. A.; Teichmann, S. A.; Bahar, I.; Bastolla, U.; Bloom, J.; Bornberg-Bauer, E.; Colwell, L. J.; de Koning, A. P. J.; Dokholyan, N. V; Echave, J.et al. The Interface of Protein Structure, Protein Biophysics, and Molecular Evolution. *Protein Sci.* **2012**, *21* (6), 769–785.

(43)   Csermely, P.; Korcsmáros, T.; Kiss, H. J. M.; London, G.; Nussinov, R. Structure and Dynamics of Molecular Networks: A Novel Paradigm of Drug Discovery: A Comprehensive Review. *Pharmacol. Ther.* **2013**, *138* (3), 333–408.

(44)   Di Paola, L.; De Ruvo, M.; Paci, P.; Santoni, D.; Giuliani, A. Protein Contact Networks: An Emerging Paradigm in Chemistry. *Chem. Rev.* **2013**, *113* (3), 1598–1613.

(45)   Whitley, M. J.; Lee, A. L. Frameworks for Understanding Long-Range Intra-Protein Communication. *Curr. Protein Pept. Sci.* **2009**, *10* (2), 116–127.

(46)   van den Bedem, H.; Bhabha, G.; Yang, K.; Wright, P. E.; Fraser, J. S. Automated Identification of Functional Dynamic Contact Networks from X-Ray Crystallography. *Nat. Methods* **2013**, *10* (9), 896–902.

(47)   Bhabha, G.; Biel, J. T.; Fraser, J. S. Keep on Moving: Discovering and Perturbing the Conformational Dynamics of Enzymes. *Acc. Chem. Res.* **2015**, *48* (2), 423–430.

(48) Papaleo, E.; Sutto, L.; Gervasio, F. L.; Lindorff-Larsen, K. Conformational Changes and Free Energies in a Proline Isomerase. *J. Chem. Theory Comput.* **2014**, *10* (9), 4169–4174.

(49) Cui, Q. Allostery and Cooperativity Revisited. *Protein Sci.* **2008**, 1295–1307.

(50) Goodey, N. M.; Benkovic, S. J. Allosteric Regulation and Catalysis Emerge via a Common Route. *Nat. Chem. Biol.* **2008**, *4* (8), 474–482.

(51) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes of Proteins and Motions. *Science (80-.).* **1991**, *254*, 1598–1603.

(52) Hilser, V. J.; García-Moreno E, B.; Oas, T. G.; Kapp, G.; Whitten, S. T. A Statistical Thermodynamic Model of the Protein Ensemble. *Chem. Rev.* **2006**, *106* (5), 1545–1558.

(53) Cooper, a.; Dryden, D. T. F. Allostery without Conformational Change - A Plausible Model. *Eur. Biophys. J.* **1984**, *11* (2), 103–109.

(54) Tsai, C. J.; del Sol, A.; Nussinov, R. Allostery: Absence of a Change in Shape Does Not Imply That Allostery Is Not at Play. *J. Mol. Biol.* **2008**, *378* (1), 1–11.

(55) Nussinov, R.; Tsai, C.-J. Allostery without a Conformational Change? Revisiting the Paradigm. *Curr. Opin. Struct. Biol.* **2015**, *30*, 17–24.

(56) Nestl, B. M.; Hauer, B. Engineering of Flexible Loops in Enzymes. *ACS Catal.* **2014**, No. 4, 3201–3211.

(57) Fiser, A.; Do, R. K.; Sali, A. Modeling of Loops in Protein Structures. *Protein Sci.* **2000**, *9* (9), 1753–1773.

(58) Furnham, N.; Sillitoe, I.; Holliday, G. L.; Cuff, A. L.; Laskowski, R. A.; Orengo, C. A.; Thornton, J. M. Exploring the Evolution of Novel Enzyme Functions within Structurally Defined Protein Superfamilies. *PLoS Comput. Biol.* **2012**, *8* (3), e1002403.

(59) Gu, Y.; Li, D.-W.; Brüschweiler, R. Decoding the Mobility and Time Scales of Protein Loops. *J. Chem. Theory Comput.* **2015**, *11* (3), 1308–1314.

(60) Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V. N. et al. DisProt: The Database of Disordered Proteins. *Nucleic Acids Res.* **2007**, *35* (Database issue), D786–D793.

(61) Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B. Protein disorder prediction: Implications for structural proteomics. *Structure* **2003,** *11,* 1453-9.

(62) Kurkcuoglu, Z.; Bakan, A.; Kocaman, D.; Bahar, I.; Doruker, P. Coupling between Catalytic Loop Motions and Enzyme Global Dynamics. *PLoS Comput. Biol.* **2012**, *8* (9), e1002705.

(63) Vuillon, L.; Lesieur, C. From Local to Global Changes in Proteins: A Network View. *Curr. Opin. Struct. Biol.* **2015**, *31*, 1–8.

(64) Gunasekaran, K.; Ma, B.; Nussinov, R. Triggering Loops and Enzyme Function: Identification of Loops That Trigger and Modulate Movements. *J. Mol. Biol.* **2003**, *332* (1), 143–159.

(65) Gunasekaran, K.; Nussinov, R. Modulating Functional Loop Movements: The Role of Highly Conserved Residues in the Correlated Loop Motions. *ChemBioChem* **2004**, *5* (2), 224–230.

(66) Pasi, M.; Riccardi, L.; Fantucci, P.; De Gioia, L.; Papaleo, E. Dynamic Properties of a Psychrophilic Alpha-Amylase in Comparison with a Mesophilic Homologue. *J. Phys. Chem. B* **2009**, *113* (41), 13585–13595.

(67) Papaleo, E.; Pasi, M.; Tiberti, M.; De Gioia, L. Molecular Dynamics of Mesophilic-like Mutants of a Cold-Adapted Enzyme: Insights into Distal Effects Induced by the Mutations. *PLoS One* **2011**, *6* (9), e24214.

(68) Fraccalvieri, D.; Tiberti, M.; Pandini, A.; Bonati, L.; Papaleo, E. Functional Annotation of the Mesophilic-like Character of Mutants in a Cold-Adapted Enzyme by Self-Organising Map Analysis of Their Molecular Dynamics. *Mol. Biosyst.* **2012**, *8* (10), 2680–2691.

(69) Pries, F.; van den Wijngaard, A. J.; Bos, R.; Pentenga, M.; Janssen, D. B. The Role of

Spontaneous Cap Domain Mutations in Haloalkane Dehalogenase Specificity and Evolution. *J. Biol. Chem.* **1994**, *269* (26), 17490–17494.

(70)   Leszczynski, J. F.; Rose, G. D. Loops in Globular Proteins: A Novel Category of Secondary Structure. *Science* **1986**, *234* (4778), 849–855.

(71)   Fetrow, J. S. Omega Loops: Nonregular Secondary Structures Significant in Protein Function and Stability. *FASEB J.* **1995**, *9* (9), 708–717.

(72)   Espadaler, J.; Querol, E.; Aviles, F. X.; Oliva, B. Identification of Function-Associated Loop Motifs and Application to Protein Function Prediction. *Bioinformatics* **2006**, *22* (18), 2237–2243.

(73)   Ring, C. S.; Kneller, D. G.; Langridge, R.; Cohen, F. E. Taxonomy and Conformational Analysis of Loops in Proteins. *J. Mol. Biol.* **1992**, *224* (3), 685–699.

(74)   Mager, P. P.; Walther, H. A Hydrophilic Omega-Loop (Tyr181 to Tyr188) in the Nonsubstrate Binding Area of HIV-1 Reverse Transcriptase. *Drug Des. Discov.* **1996**, *14* (3), 225–239.

(75)   Krishna, M. M. G.; Lin, Y.; Rumbley, J. N.; Walter Englander, S. Cooperative Omega Loops in Cytochrome c: Role in Folding and Function. *J. Mol. Biol.* **2003**, *331* (1), 29–36.

(76)   Papaleo, E.; Casiraghi, N.; Arrigoni, A.; Vanoni, M.; Coccetti, P.; Gioia, L. De. Loop 7 of E2 Enzymes : An Ancestral Conserved Functional Motif Involved in the E2-Mediated Steps of the Ubiquitination Cascade. *PLoS One* **2012**, *7* (7).

(77)   Dagan, S.; Hagai, T.; Gavrilov, Y.; Kapon, R.; Levy, Y.; Reich, Z. Stabilization of a Protein Conferred by an Increase in Folded State Entropy. *Proc. Natl. Acad. Sci.* **2013**, *110* (26), 10628–10633.

(78)   Gavrilov, Y.; Dagan, S.; Levy, Y. Shortening a Loop Can Increase Protein Native State Entropy. *Proteins Struct. Funct. Bioinforma.* **2015**, *83*(12):2137-46.

(79)   Papaleo, E.; Ranzani, V.; Tripodi, F.; Vitriolo, A.; Cirulli, C.; Fantucci, P.; Alberghina, L.; Vanoni, M.; Gioia, L. De; Coccetti, P. An Acidic Loop and Cognate Phosphorylation Sites Define a Molecular Switch That Modulates Ubiquitin Charging Activity in Cdc34-Like Enzymes. *PLoS Comput. Biol.* **2011**, *7* (5).

(80)   Viguera, A. R.; Serrano, L. Loop Length, Intramolecular Diffusion and Protein Folding. *Nat. Struct. Biol.* **1997**, *4* (11), 939–946.

(81)   Gokhale, R. S.; Khosla, C. Role of Linkers in Communication between Protein Modules. *Curr. Opin. Chem. Biol.* **2000**, *4* (1), 22–27.

(82)   George, R. a; Heringa, J. An Analysis of Protein Domain Linkers: Their Classification and Role in Protein Folding. *Protein Eng.* **2002**, *15* (11), 871–879.

(83)   Reddy Chichili, V. P.; Kumar, V.; Sivaraman, J. Linkers in the Structural Biology of Protein-Protein Interactions. *Protein Sci.* **2013**, *22* (2), 153–167.

(84)   Apic, G.; Huber, W.; Teichmann, S. a. Multi-Domain Protein Families and Domain Pairs: Comparison with Known Structures and a Random Model of Domain Recombination. *J. Struct. Funct. Genomics* **2003**, *4* (2-3), 67–78.

(85)   Robinson, C. R.; Sauer, R. T. Optimizing the Stability of Single-Chain Proteins by Linker Length and Composition Mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (11), 5929–5934.

(86)   Argos, P. An Investigation of Oligopeptides Linking Domains in Protein Tertiary Structures and Possible Candidates for General Gene Fusion. *J. Mol. Biol.* **1990**, *211* (4), 943–958.

(87)   Bennett, M. J.; Choe, S.; Eisenberg, D. Domain Swapping: Entangling Alliances between Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (8), 3127–3131.

(88)   Henkel, T.; Zabel, U.; van Zee, K.; Müller, J. M.; Fanning, E.; Baeuerle, P. A. Intramolecular Masking of the Nuclear Location Signal and Dimerization Domain in the Precursor for the p50 NF-Kappa B Subunit. *Cell* **1992**, *68* (6), 1121–1133.

(89)   Bobrowski, K.; Wierzchowski, K. L.; Holcman, J.; Ciurak, M. Intramolecular Electron Transfer in Peptides Containing Methionine, Tryptophan and Tyrosine: A Pulse Radiolysis Study. *Int. J. Radiat. Biol.* **1990**, *57* (5), 919–932.

(90)   Mancebo, R.; Lo, P. C.; Mount, S. M. Structure and Expression of the Drosophila Melanogaster Gene for the U1 Small Nuclear Ribonucleoprotein Particle 70K Protein. *Mol. Cell. Biol.* **1990**, *10* (6), 2492–2502.

(91)   Klement, M.; Liu, C.; Loo, B. L. W.; Choo, A. B.-H.; Ow, D. S.-W.; Lee, D.-Y. Effect of Linker Flexibility and Length on the Functionality of a Cytotoxic Engineered Antibody Fragment. *J. Biotechnol.* **2015**, *199*, 90–97.

(92)   Chen, X.; Zaro, J. L.; Shen, W. C. Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug Deliv. Rev.* **2013**, *65* (10), 1357–1369.

(93)   Amet, N.; Lee, H.-F.; Shen, W.-C. Insertion of the Designed Helical Linker Led to Increased Expression of Tf-Based Fusion Proteins. *Pharm. Res.* **2009**, *26* (3), 523–528.

(94)   Arai, R.; Wriggers, W.; Nishikawa, Y.; Nagamune, T.; Fujisawa, T. Conformations of Variably Linked Chimeric Proteins Evaluated by Synchrotron X-Ray Small-Angle Scattering. *Proteins Struct. Funct. Genet.* **2004**, *57* (4), 829–838.

(95)   Arai, R.; Ueda, H.; Kitayama, A.; Kamiya, N.; Nagamune, T. Design of the Linkers Which Effectively Separate Domains of a Bifunctional Fusion Protein. *Protein Eng.* **2001**, *14* (8), 529–532.

(96)   Lupas, A. Coiled Coils: New Structures and New Functions. *Trends Biochem. Sci.* **1996**, *21* (10), 375–382.

(97)   Parry, D. A. D.; Fraser, R. D. B.; Squire, J. M. Fifty Years of Coiled-Coils and Alpha-Helical Bundles: A Close Relationship between Sequence and Structure. *J. Struct. Biol.* **2008**, *163* (3), 258–269.

(98)   Parry, D. A. D.; Smith, T. A. A Different Conformation for Linker L12 in IF Molecules in the Molecular and Filamentous Forms: An Hypothesis. *J. Struct. Biol.* **2010**, *170* (2), 364–368.

(99)   Johnson, J. L.; Cusack, B.; Hughes, T. F.; McCullough, E. H.; Fauq, A.; Romanovskis, P.; Spatola, A. F.; Rosenberry, T. L. Inhibitors Tethered near the Acetylcholinesterase Active Site Serve as Molecular Rulers of the Peripheral and Acylation Sites. *J. Biol. Chem.* **2003**, *278* (40), 38948–38955.

(100)   Profit, A. A.; Lee, T. R.; Niu, J.; Lawrence, D. S. Molecular Rulers: An Assessment of Distance and Spatial Relationships of Src Tyrosine Kinase Sh2 and Active Site Regions. *J. Biol. Chem.* **2001**, *276* (12), 9446–9451.

(101)   Arora, P. S.; Ansari, A. Z.; Best, T. P.; Ptashne, M.; Dervan, P. B. Design of Artificial Transcriptional Activators with Rigid Poly-L-Proline Linkers. *J. Am. Chem. Soc.* **2002**, *124* (44), 13067–13071.

(102)   Best, R. B.; Hofmann, H.; Nettels, D.; Schuler, B. Quantitative Interpretation of FRET Experiments via Molecular Simulation: Force Field and Validation. *Biophys. J.* **2015**, *108* (11), 2721–2731.

(103)   Watkins, L. P.; Chang, H.; Yang, H. Quantitative Single-Molecule Conformational Distributions: A Case Study with Poly-(L-Proline). *J. Phys. Chem. A* **2006**, *110* (15), 5191–5203.

(104)   Best, R. B.; Merchant, K. A.; Gopich, I. V.; Schuler, B.; Bax, A.; Eaton, W. A. Effect of Flexibility and Cis Residues in Single-Molecule FRET Studies of Polyproline. *Proc. Natl. Acad. Sci.* **2007**, *104* (48), 18964–18969.

(105)   Bottema, C. D.; Ketterling, R. P.; Ii, S.; Yoon, H. S.; Phillips, J. A.; Sommer, S. S. Missense Mutations and Evolutionary Conservation of Amino Acids: Evidence That Many of the Amino Acids in Factor IX Function as "Spacer" Elements. *Am. J. Hum. Genet.* **1991**, *49* (4), 820–838.

(106)   Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. Helix Propensities of the Amino Acids Measured in Alanine-Based Peptides without Helix-Stabilizing Side-Chain Interactions.

Protein Sci. **1994**, *3* (5), 843–852.

(107) Opalka, N.; Chlenov, M.; Chacon, P.; Rice, W. J.; Wriggers, W.; Darst, S. A. Structure and Function of the Transcription Elongation Factor GreB Bound to Bacterial RNA Polymerase. *Cell* **2003**, *114* (3), 335–345.

(108) Leriche, G.; Chisholm, L.; Wagner, A. Cleavable Linkers in Chemical Biology. *Bioorg. Med. Chem.* **2012**, *20* (2), 571–582.

(109) Chen, X.; Bai, Y.; Zaro, J. L.; Shen, W.-C. Design of an in Vivo Cleavable Disulfide Linker in Recombinant Fusion Proteins. *Biotechniques* **2010**, *49* (1), 513–518.

(110) Ren, C.; Song, Z.; Zheng, W.; Chen, X.; Wang, L.; Kong, D.; Yang, Z. Disulfide Bond as a Cleavable Linker for Molecular Self-Assembly and Hydrogelation. *Chem. Commun. (Camb).* **2011**, *47* (5), 1619–1621.

(111) Crasto, C. J.; Feng, J. a. LINKER: A Program to Generate Linker Sequences for Fusion Proteins. *Protein Eng.* **2000**, *13* (5), 309–312.

(112) Samant, V.; Hulgeri, A.; Valencia, A.; Tendulkar, A. V. Accurate Demarcation of Protein Domain Linkers Based on Structural Analysis of Linker Probable Region. **2012**, 1–19.

(113) Zhu, L.; Chen, L.; Zhou, X.-M.; Zhang, Y.-Y.; Zhang, Y.-J.; Zhao, J.; Ji, S.-R.; Wu, J.-W.; Wu, Y. Structural Insights into the Architecture and Allostery of Full-Length AMP-Activated Protein Kinase. *Structure* **2011**, *19* (4), 515–522.

(114) Kar, G.; Keskin, O.; Gursoy, A.; Nussinov, R. Allostery and Population Shift in Drug Discovery. *Curr. Opin. Pharmacol.* **2010**, *10* (6), 715–722.

(115) Nussinov, R.; Tsai, C.-J. Allostery in Disease and in Drug Discovery. *Cell* **2013**, *153* (2), 293–305.

(116) Pawson, T. Dynamic Control of Signaling by Modular Adaptor Proteins. *Curr. Opin. Cell Biol.* **2007**, *19* (2), 112–116.

(117) Pawson, T. Organization of Cell-Regulatory Systems through Modular-Protein-Interaction Domains. *Philos. Trans. A. Math. Phys. Eng. Sci.* **2003**, *361* (1807), 1251–1262.

(118) Mayer, B. J. The Discovery of Modular Binding Domains: Building Blocks of Cell Signalling. *Nat. Rev. Mol. Cell Biol.* **2015**.

(119) Gerstein, M.; Echols, N. Exploring the Range of Protein Flexibility, from a Structural Proteomics Perspective. *Curr. Opin. Chem. Biol.* **2004**, *8* (1), 14–19.

(120) Echols, N.; Milburn, D.; Gerstein, M. MolMovDB: Analysis and Visualization of Conformational Change and Structural Flexibility. *Nucleic Acids Res.* **2003**, *31* (1), 478–482.

(121) Krebs, W. G.; Gerstein, M. The Morph Server: A Standardized System for Analyzing and Visualizing Macromolecular Motions in a Database Framework. *Nucleic Acids Res.* **2000**, *28* (8), 1665–1675.

(122) Gerstein, M.; Anderson, B. F.; Norris, G. E.; Baker, E. N.; Lesk, A. M.; Chothia, C. Domain Closure in Lactoferrin. Two Hinges Produce a See-Saw Motion between Alternative Close-Packed Interfaces. *J. Mol. Biol.* **1993**, *234* (2), 357–372.

(123) Hayward, S. Structural Principles Governing Domain Motions in Proteins. *Proteins* **1999**, *36* (4), 425–435.

(124) Qi, G.; Lee, R.; Hayward, S. A Comprehensive and Non-Redundant Database of Protein Domain Movements. *Bioinformatics* **2005**, *21* (12), 2832–2838.

(125) Lesk, A. M.; Chothia, C. Mechanisms of Domain Closure in Proteins. *J. Mol. Biol.* **1984**, *174* (1), 175–191.

(126) Gerstein, M.; Lesk, A. M.; Chothia, C. Structural Mechanisms for Domain Movements in Proteins. *Biochemistry* **1994**, *33* (22), 6739–6749.

(127) Dobson, C. M. Protein Conformation. Hinge-Bending and Folding. *Nature* **1990**, *348* (6298),

198–199.

(128) Faber, H. R.; Matthews, B. W. A Mutant T4 Lysozyme Displays Five Different Crystal Conformations. *Nature* **1990**, *348* (6298), 263–266.

(129) Kumar, S.; Ma, B.; Tsai, C. J.; Wolfson, H.; Nussinov, R. Folding Funnels and Conformational Transitions via Hinge-Bending Motions. *Cell Biochem. Biophys.* **1999**, *31* (2), 141–164.

(130) Schimmel, P. R.; Flory, P. J. Conformational Energy and Configurational Statistics of Poly-L-Proline. *Proc. Natl. Acad. Sci. U. S. A.* **1967**, *58* (1), 52–59.

(131) Babcock, M. S.; Pednault, E. P.; Olson, W. K. Nucleic Acid Structure Analysis. Mathematics for Local Cartesian and Helical Structure Parameters That Are Truly Comparable between Structures. *J. Mol. Biol.* **1994**, *237* (1), 125–156.

(132) Wriggers, W.; Schulten, K. Protein Domain Movements: Detection of Rigid Domains and Visualization of Hinges in Comparisons of Atomic Coordinates. *Proteins* **1997**, *29* (1), 1–14.

(133) Hirose, M. The Structural Mechanism for Iron Uptake and Release by Transferrins. *Biosci. Biotechnol. Biochem.* **2014**, *64* (7), 1328–1336.

(134) Mizutani, K.; Toyoda, M.; Mikami, B. X-Ray Structures of Transferrins and Related Proteins. *Biochim. Biophys. Acta* **2012**, *1820* (3), 203–211.

(135) Mizutani, K.; Mikami, B.; Hirose, M. Domain Closure Mechanism in Transferrins: New Viewpoints about the Hinge Structure and Motion as Deduced from High Resolution Crystal Structures of Ovotransferrin N-Lobe. *J. Mol. Biol.* **2001**, *309* (4), 937–947.

(136) Baker, E. N.; Baker, H. M. Molecular Structure, Binding Properties and Dynamics of Lactoferrin. *Cell. Mol. Life Sci.* **2005**, *62* (22), 2531–2539.

(137) Baker, E.; Baker, H. A Structural Framework for Understanding the Multifunctional Character of Lactoferrin. *Biochimie* **2009**, *91* (1), 3–10.

(138) Baker, H. M.; Baker, E. N. A Structural Perspective on Lactoferrin Function 1 1 This Article Is Part of a Special Issue Entitled Lactoferrin and Has Undergone the Journal's Usual Peer Review Process. *Biochem. Cell Biol.* **2012**, *90* (3), 320–328.

(139) Grossmann, J. G.; Neu, M.; Pantos, E.; Schwab, F. J.; Evans, R. W.; Townes-Andrews, E.; Lindley, P. F.; Appel, H.; Thies, W. G.; Hasnain, S. S. X-Ray Solution Scattering Reveals Conformational Changes upon Iron Uptake in Lactoferrin, Serum and Ovo-Transferrins. *J. Mol. Biol.* **1992**, *225* (3), 811–819.

(140) Millet, O.; Hudson, R. P.; Kay, L. E. The Energetic Cost of Domain Reorientation in Maltose-Binding Protein as Studied by NMR and Fluorescence Spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (22), 12700–12705.

(141) Tang, C.; Schwieters, C. D.; Clore, G. M. Open-to-Closed Transition in Apo Maltose-Binding Protein Observed by Paramagnetic NMR. *Nature* **2007**, *449* (7165), 1078–1082.

(142) Wang, Y.; Tang, C.; Wang, E.; Wang, J. Exploration of Multi-State Conformational Dynamics and Underlying Global Functional Landscape of Maltose Binding Protein. *PLoS Comput. Biol.* **2012**, *8* (4), e1002471.

(143) Pontiggia, F.; Zen, A.; Micheletti, C. Small- and Large-Scale Conformational Changes of Adenylate Kinase: A Molecular Dynamics Study of the Subdomain Motion and Mechanics. *Biophys. J.* **2008**, *95* (12), 5901–5912.

(144) Arora, K.; Brooks, C. L. Large-Scale Allosteric Conformational Transitions of Adenylate Kinase Appear to Involve a Population-Shift Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (47), 18496–18501.

(145) Li, D.; Liu, M. S.; Ji, B. Mapping the Dynamics Landscape of Conformational Transitions in Enzyme: The Adenylate Kinase Case. *Biophys. J.* **2015**, *109* (3), 647–660.

(146) Callender, R.; Dyer, R. B. The Dynamical Nature of Enzymatic Catalysis. *Acc. Chem. Res.* **2015**, *48* (2), 407–413.

(147) Murphy, P. M.; Bolduc, J. M.; Gallaher, J. L.; Stoddard, B. L.; Baker, D. Alteration of Enzyme Specificity by Computational Loop Remodeling and Design. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (23), 9215–9220.

(148) Park, H.-S.; Nam, S.-H.; Lee, J. K.; Yoon, C. N.; Mannervik, B.; Benkovic, S. J.; Kim, H.-S. Design and Evolution of New Catalytic Activity with an Existing Protein Scaffold. *Science* **2006**, *311* (5760), 535–538.

(149) Aharoni, A.; Gaidukov, L.; Khersonsky, O.; McQ Gould, S.; Roodveldt, C.; Tawfik, D. S. The "Evolvability" of Promiscuous Protein Functions. *Nat. Genet.* **2005**, *37* (1), 73–76.

(150) Joseph, D.; Petsko, G. A.; Karplus, M. Anatomy of a Conformational Change: Hinged "Lid" Motion of the Triosephosphate Isomerase Loop. *Science* **1990**, *249* (4975), 1425–1428.

(151) Wade, R. C.; Davis, M. E.; Luty, B. A.; Madura, J. D.; McCammon, J. A. Gating of the Active Site of Triose Phosphate Isomerase: Brownian Dynamics Simulations of Flexible Peptide Loops in the Enzyme. *Biophys. J.* **1993**, *64* (1), 9–15.

(152) Kempf, J. G.; Jung, J.-Y.; Ragain, C.; Sampson, N. S.; Loria, J. P. Dynamic Requirements for a Functional Protein Hinge. *J. Mol. Biol.* **2007**, *368* (1), 131–149.

(153) Wang, Y.; Berlow, R. B.; Loria, J. P. Role of Loop−Loop Interactions in Coordinating Motions and Enzymatic Function in Triosephosphate Isomerase. *Biochemistry* **2009**, *48* (21), 4548–4556.

(154) Torchia, D. A. NMR Studies of Dynamic Biomolecular Conformational Ensembles. *Prog. Nucl. Magn. Reson. Spectrosc.* **2015**, *84-85*, 14–32.

(155) Louis, J. M.; Ishima, R.; Torchia, D. A.; Weber, I. T. HIV-1 Protease: Structure, Dynamics, and Inhibition. *Adv. Pharmacol.* **2007**, *55*, 261–298.

(156) Lee, J.; Goodey, N. M. Catalytic Contributions from Remote Regions of Enzyme Structure. *Chem. Rev.* **2011**, *111* (12), 7595–7624.

(157) Zgiby, S.; Plater, A. R.; Bates, M. A.; Thomson, G. J.; Berry, A. A Functional Role for a Flexible Loop Containing Glu182 in the Class II Fructose-1,6-Bisphosphate Aldolase from Escherichia Coli. *J. Mol. Biol.* **2002**, *315* (2), 131–140.

(158) Pegan, S. D.; Rukseree, K.; Capodagli, G. C.; Baker, E. A.; Krasnykh, O.; Franzblau, S. G.; Mesecar, A. D. Active Site Loop Dynamics of a Class IIa Fructose 1,6-Bisphosphate Aldolase from Mycobacterium Tuberculosis. *Biochemistry* **2013**, *52* (5), 912–925.

(159) Blaszczyk, J.; Lu, Z.; Li, Y.; Yan, H.; Ji, X. Crystallographic and Molecular Dynamics Simulation Analysis of Escherichia Coli Dihydroneopterin Aldolase. *Cell Biosci.* **2014**, *4* (1), 52.

(160) Hakobyan, D.; Nazaryan, K. Molecular Dynamics Study of Interaction and Substrate Channeling between Neuron-Specific Enolase and B-Type Phosphoglycerate Mutase. *Proteins* **2010**, *78* (7), 1691–1704.

(161) Gulerez, I. E.; Gehring, K. X-Ray Crystallography and NMR as Tools for the Study of Protein Tyrosine Phosphatases. *Methods* **2014**, *65* (2), 175–183.

(162) Kamerlin, S. C. L.; Rucker, R.; Boresch, S. A Molecular Dynamics Study of WPD-Loop Flexibility in PTP1B. *Biochem. Biophys. Res. Commun.* **2007**, *356* (4), 1011–1016.

(163) Hu, X.; Stebbins, C. E. Dynamics of the WPD Loop of the Yersinia Protein Tyrosine Phosphatase. *Biophys. J.* **2006**, *91* (3), 948–956.

(164) Zhang, Z.-Y. Protein-Tyrosine Phosphatases: Biological Function, Structural Characteristics, and Mechanism of Catalysis. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33* (1), 1–52.

(165) Trodler, P.; Schmid, R. D.; Pleiss, J. Modeling of Solvent-Dependent Conformational Transitions in Burkholderia Cepacia Lipase. *BMC Struct. Biol.* **2009**, *9*, 38.

(166) Jiang, Y.; Li, L.; Zhang, H.; Feng, W.; Tan, T. Lid Closure Mechanism of Yarrowia Lipolytica Lipase in Methanol Investigated by Molecular Dynamics Simulation. *J. Chem. Inf. Model.* **2014**, *54* (7), 2033–2041.

(167) Barbe, S.; Cortés, J.; Siméon, T.; Monsan, P.; Remaud-Siméon, M.; André, I. A Mixed Molecular Modeling-Robotics Approach to Investigate Lipase Large Molecular Motions. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (8), 2517–2529.

(168) Zisis, T.; Freddolino, P. L.; Turunen, P.; van Teeseling, M. C. F.; Rowan, A. E.; Blank, K. G. Interfacial Activation of Candida Antarctica Lipase B: Combined Evidence from Experiment and Simulation. *Biochemistry* **2015**, *54* (38), 5969–5979.

(169) Bordes, F.; Barbe, S.; Escalier, P.; Mourey, L.; André, I.; Marty, A.; Tranier, S. Exploring the Conformational States and Rearrangements of Yarrowia Lipolytica Lipase. *Biophys. J.* **2010**, *99* (7), 2225–2234.

(170) Rehm, S.; Trodler, P.; Pleiss, J. Solvent-Induced Lid Opening in Lipases: A Molecular Dynamics Study. *Protein Sci.* **2010**, *19* (11), 2122–2130.

(171) Ganjalikhany, M. R.; Ranjbar, B.; Taghavi, A. H.; Tohidi Moghadam, T. Functional Motions of Candida Antarctica Lipase B: A Survey through Open-Close Conformations. *PLoS One* **2012**, *7* (7), e40327.

(172) Gunasekaran, K.; Ma, B.; Ramakrishnan, B.; Qasba, P. K.; Nussinov, R. Interdependence of Backbone Flexibility, Residue Conservation, and Enzyme Function: A Case Study on beta1,4-Galactosyltransferase-I. *Biochemistry* **2003**, *42* (13), 3674–3687.

(173) Chen, J.; Dima, R. I.; Thirumalai, D. Allosteric Communication in Dihydrofolate Reductase: Signaling Network and Pathways for Closed to Occluded Transition and Back. *J. Mol. Biol.* **2007**, *374* (1), 250–266.

(174) Francis, K.; Stojkovic, V.; Kohen, A. Preservation of Protein Dynamics in Dihydrofolate Reductase Evolution. *J. Biol. Chem.* **2013**, *288* (50), 35961–35968.

(175) Boehr, D. D.; Schnell, J. R.; McElheny, D.; Bae, S.-H.; Duggan, B. M.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. A Distal Mutation Perturbs Dynamic Amino Acid Networks in Dihydrofolate Reductase. *Biochemistry* **2013**, *52* (27), 4605–4619.

(176) Venkitakrishnan, R. P.; Zaborowski, E.; McElheny, D.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. Conformational Changes in the Active Site Loops of Dihydrofolate Reductase during the Catalytic Cycle. *Biochemistry* **2004**, *43* (51), 16046–16055.

(177) Tuttle, L. M.; Dyson, H. J.; Wright, P. E. Side Chain Conformational Averaging in Human Dihydrofolate Reductase. *Biochemistry* **2014**, *53* (7), 1134–1145.

(178) Klinman, J. P.; Kohen, A. Evolutionary Aspects of Enzyme Dynamics. *J. Biol. Chem.* **2014**, *289* (44), 30205–30212.

(179) Kohen, A. Role of Dynamics in Enzyme Catalysis: Substantial versus Semantic Controversies. *Acc. Chem. Res.* **2015**, *48* (2), 466–473.

(180) Li, L.; Falzone, C. J.; Wright, P. E.; Benkovic, S. J. Functional Role of a Mobile Loop of Escherichia Coli Dihydrofolate Reductase in Transition-State Stabilization. *Biochemistry* **1992**, *31* (34), 7826–7833.

(181) Anthis, N. J.; Clore, G. M. The Length of the Calmodulin Linker Determines the Extent of Transient Interdomain Association and Target Affinity. *J. Am. Chem. Soc.* **2013**, *135* (26), 9648–9651.

(182) Krishnamurthy, V. M.; Semetey, V.; Bracher, P. J.; Shen, N.; Whitesides, G. M. Dependence of Effective Molarity on Linker Length for an Intramolecular Protein-Ligand System. *J. Am. Chem. Soc.* **2007**, *129* (5), 1312–1320.

(183) Shewmake, T. A.; Solis, F. J.; Gillies, R. J.; Caplan, M. R. Effects of Linker Length and Flexibility on Multivalent Targeting. *Biomacromolecules* **2008**, *9* (11), 3057–3064.

(184) Bobrovnik, S. A. The Influence of Rigid or Flexible Linkage between Two Ligands on the Effective Affinity and Avidity for Reversible Interactions with Bivalent Receptors. *J. Mol. Recognit. 20* (4), 253–262.

(185) Cohen, P. The Origins of Protein Phosphorylation. *Nat. Cell Biol.* **2002**, *4* (5), E127–E130.

(186) Blume-Jensen, P.; Hunter, T. Oncogenic Kinase Signalling. *Nature* **2001**, *411* (6835), 355–365.

(187) Lahiry, P.; Torkamani, A.; Schork, N. J.; Hegele, R. A. Kinase Mutations in Human Disease: Interpreting Genotype-Phenotype Relationships. *Nat. Rev. Genet.* **2010**, *11* (1), 60–74.

(188) Fabbro, D.; Cowan-Jacob, S. W.; Moebitz, H. Ten Things You Should Know about Protein Kinases: IUPHAR Review 14. *Br. J. Pharmacol.* **2015**, *172* (11), 2675–2700.

(189) Huse, M.; Kuriyan, J. The Conformational Plasticity of Protein Kinases. *Cell* **2002**, *109* (3), 275–282.

(190) Cox, S.; Radzio-Andzelm, E.; Taylor, S. S. Domain Movements in Protein Kinases. *Curr. Opin. Struct. Biol.* **1994**, *4* (6), 893–901.

(191) Sahún-Roncero, M.; Rubio-Ruiz, B.; Saladino, G.; Conejo-García, A.; Espinosa, A.; Velázquez-Campoy, A.; Gervasio, F. L.; Entrena, A.; Hurtado-Guerrero, R. The Mechanism of Allosteric Coupling in Choline Kinase α1 Revealed by the Action of a Rationally Designed Inhibitor. *Angew. Chem. Int. Ed. Engl.* **2013**, *52* (17), 4582–4586.

(192) Dölker, N.; Górna, M. W.; Sutto, L.; Torralba, A. S.; Superti-Furga, G.; Gervasio, F. L. The SH2 Domain Regulates c-Abl Kinase Activation by a Cyclin-like Mechanism and Remodulation of the Hinge Motion. *PLoS Comput. Biol.* **2014**, *10* (10), e1003863.

(193) Nolen, B.; Taylor, S.; Ghosh, G. Regulation of Protein Kinases; Controlling Activity through Activation Segment Conformation. *Mol. Cell* **2004**, *15* (5), 661–675.

(194) Taylor, S. S.; Kornev, A. P. Protein Kinases: Evolution of Dynamic Regulatory Proteins. *Trends Biochem. Sci.* **2011**, *36* (2), 65–77.

(195) Mobitz, H.; Fabbro, D. Conformational Bias: A Key Concept for Protein Kinase Inhibitors. *Eur. Pharm. Rev.* **2012**, *17*, 41–51.

(196) Xu, W.; Doshi, A.; Lei, M.; Eck, M. J.; Harrison, S. C. Crystal Structures of c-Src Reveal Features of Its Autoinhibitory Mechanism. *Mol. Cell* **1999**, *3* (5), 629–638.

(197) Shan, Y.; Seeliger, M. A.; Eastwood, M. P.; Frank, F.; Xu, H.; Jensen, M. Ø.; Dror, R. O.; Kuriyan, J.; Shaw, D. E. A Conserved Protonation-Dependent Switch Controls Drug Binding in the Abl Kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (1), 139–144.

(198) Biondi, R. M.; Cheung, P. C.; Casamayor, A.; Deak, M.; Currie, R. A.; Alessi, D. R. Identification of a Pocket in the PDK1 Kinase Domain That Interacts with PIF and the C-Terminal Residues of PKA. *EMBO J.* **2000**, *19* (5), 979–988.

(199) Nagar, B.; Hantschel, O.; Young, M. A.; Scheffzek, K.; Veach, D.; Bornmann, W.; Clarkson, B.; Superti-Furga, G.; Kuriyan, J. Structural Basis for the Autoinhibition of c-Abl Tyrosine Kinase. *Cell* **2003**, *112* (6), 859–871.

(200) Hantschel, O.; Nagar, B.; Guettler, S.; Kretzschmar, J.; Dorey, K.; Kuriyan, J.; Superti-Furga, G. A Myristoyl/phosphotyrosine Switch Regulates c-Abl. *Cell* **2003**, *112* (6), 845–857.

(201) Young, M. A.; Gonfloni, S.; Superti-Furga, G.; Roux, B.; Kuriyan, J. Dynamic Coupling between the SH2 and SH3 Domains of c-Src and Hck Underlies Their Inactivation by C-Terminal Tyrosine Phosphorylation. *Cell* **2001**, *105* (1), 115–126.

(202) Harrison, S. C. Variation on an Src-like Theme. *Cell* **2003**, *112* (6), 737–740.

(203) Moarefi, I.; LaFevre-Bernt, M.; Sicheri, F.; Huse, M.; Lee, C. H.; Kuriyan, J.; Miller, W. T. Activation of the Src-Family Tyrosine Kinase Hck by SH3 Domain Displacement. *Nature* **1997**, *385* (6617), 650–653.

(204) Cowan-Jacob, S. W.; Fendrich, G.; Manley, P. W.; Jahnke, W.; Fabbro, D.; Liebetanz, J.; Meyer, T. The Crystal Structure of a c-Src Complex in an Active Conformation Suggests Possible Steps in c-Src Activation. *Structure* **2005**, *13* (6), 861–871.

(205) Skora, L.; Mestan, J.; Fabbro, D.; Jahnke, W.; Grzesiek, S. NMR Reveals the Allosteric Opening and Closing of Abelson Tyrosine Kinase by ATP-Site and Myristoyl Pocket Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (47), E4437–E4445.

(206) Filippakopoulos, P.; Kofler, M.; Hantschel, O.; Gish, G. D.; Grebien, F.; Salah, E.; Neudecker, P.; Kay, L. E.; Turk, B. E.; Superti-Furga, G. et al. Structural Coupling of SH2-Kinase Domains Links Fes and Abl Substrate Recognition and Kinase Activation. *Cell* **2008**, *134* (5), 793–803.

(207) Marquez, J. A. Conformation of Full-Length Bruton Tyrosine Kinase (Btk) from Synchrotron X-Ray Solution Scattering. *EMBO J.* **2003**, *22* (18), 4616–4624.

(208) Jeffrey, P. D.; Russo, A. A.; Polyak, K.; Gibbs, E.; Hurwitz, J.; Massagué, J.; Pavletich, N. P. Mechanism of CDK Activation Revealed by the Structure of a cyclinA-CDK2Complex. *Nature* **1995**, *376* (6538), 313–320.

(209) Zhang, X.; Gureasko, J.; Shen, K.; Cole, P. A.; Kuriyan, J. An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell* **2006**, *125* (6), 1137–1149.

(210) Shewchuk, L. M.; Hassell, A. M.; Ellis, B.; Holmes, W. D.; Davis, R.; Horne, E. L.; Kadwell, S. H.; McKee, D. D.; Moore, J. T. Structure of the Tie2 RTK Domain: Self-Inhibition by the Nucleotide Binding Loop, Activation Loop, and C-Terminal Tail. *Structure* **2000**, *8* (11), 1105–1113.

(211) Hubbard, S. R. Juxtamembrane Autoinhibition in Receptor Tyrosine Kinases. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (6), 464–471.

(212) Yan, Q.; Barros, T.; Visperas, P. R.; Deindl, S.; Kadlecek, T. A.; Weiss, A.; Kuriyan, J. Structural Basis for Activation of ZAP-70 by Phosphorylation of the SH2-Kinase Linker. *Mol. Cell. Biol.* **2013**, *33* (11), 2188–2201.

(213) Frame, M. C.; Patel, H.; Serrels, B.; Lietha, D.; Eck, M. J. The FERM Domain: Organizing the Structure and Function of FAK. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (11), 802–814.

(214) Stauber, D. J.; DiGabriele, A. D.; Hendrickson, W. A. Structural Interactions of Fibroblast Growth Factor Receptor with Its Ligands. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (1), 49–54.

(215) Nussinov, R.; Tsai, C.-J. The Different Ways through Which Specificity Works in Orthosteric and Allosteric Drugs. *Curr. Pharm. Des.* **2012**, *18* (9), 1311–1316.

(216) Fang, Z.; Grütter, C.; Rauh, D. Strategies for the Selective Regulation of Kinases with Allosteric Modulators: Exploiting Exclusive Structural Features. *ACS Chem. Biol.* **2013**, *8* (1), 58–70.

(217) Gray, N. S.; Fabbro, D. Discovery of Allosteric BCR-ABL Inhibitors from Phenotypic Screen to Clinical Candidate. *Methods Enzymol.* **2014**, *548*, 173–188.

(218) Bono, F.; De Smet, F.; Herbert, C.; De Bock, K.; Georgiadou, M.; Fons, P.; Tjwa, M.; Alcouffe, C.; Ny, A.; Bianciotto, M. et al. Inhibition of Tumor Angiogenesis and Growth by a Small-Molecule Multi-FGF Receptor Blocker with Allosteric Properties. *Cancer Cell* **2013**, *23* (4), 477–488.

(219) Herbert, C.; Schieborr, U.; Saxena, K.; Juraszek, J.; De Smet, F.; Alcouffe, C.; Bianciotto, M.; Saladino, G.; Sibrac, D.; Kudlinzki, D. et al. Molecular Mechanism of SSR128129E, an Extracellularly Acting, Small-Molecule, Allosteric Inhibitor of FGF Receptor Signaling. *Cancer Cell* **2013**, *23* (4), 489–501.

(220) Hindie, V.; Stroba, A.; Zhang, H.; Lopez-Garcia, L. A.; Idrissova, L.; Zeuzem, S.; Hirschberg, D.; Schaeffer, F.; Jørgensen, T. J. D.; Engel, M. et al. Structure and Allosteric Effects of Low-Molecular-Weight Activators on the Protein Kinase PDK1. *Nat. Chem. Biol.* **2009**, *5* (10), 758–764.

(221) Poulikakos, P. I.; Zhang, C.; Bollag, G.; Shokat, K. M.; Rosen, N. RAF Inhibitors Transactivate RAF Dimers and ERK Signalling in Cells with Wild-Type BRAF. *Nature* **2010**, *464* (7287), 427–430.

(222) Arora, R.; Di Michele, M.; Stes, E.; Vandermarliere, E.; Martens, L.; Gevaert, K.; Van Heerde, E.; Linders, J. T. M.; Brehmer, D.; Jacoby, E. et al. Structural Investigation of B-Raf Paradox Breaker and Inducer Inhibitors. *J. Med. Chem.* **2015**, *58* (4), 1818–1831.

(223) Hu, J.; Stites, E. C.; Yu, H.; Germino, E. A.; Meharena, H. S.; Stork, P. J. S.; Kornev, A. P.; Taylor, S. S.; Shaw, A. S. Allosteric Activation of Functionally Asymmetric RAF Kinase Dimers. *Cell* **2013**, *154* (5), 1036–1046.

(224) Thevakumaran, N.; Lavoie, H.; Critton, D. A.; Tebben, A.; Marinier, A.; Sicheri, F.; Therrien, M. Crystal Structure of a BRAF Kinase Domain Monomer Explains Basis for Allosteric Regulation. *Nat. Struct. Mol. Biol.* **2014**, *22* (1), 37–43.

(225) Chen, J.-J.; Lin, F.; Qin, Z.-H. The Roles of the Proteasome Pathway in Signal Transduction and Neurodegenerative Diseases. *Neurosci. Bull.* **2008**, *24* (3), 183–194.

(226) Shaid, S.; Brandts, C. H.; Serve, H.; Dikic, I. Ubiquitination and Selective Autophagy. *Cell Death Differ.* **2013**, *20* (1), 21–30.

(227) Sun, L.; Chen, Z. J. The Novel Functions of Ubiquitination in Signaling. *Curr. Opin. Cell Biol.* **2004**, *16* (2), 119–126.

(228) Khaminets, A.; Behl, C.; Dikic, I. Ubiquitin-Dependent And Independent Signals In Selective Autophagy. *Trends Cell Biol.* **2015**, *xx*, 1–11.

(229) Passmore, L. A.; Barford, D. Getting into Position: The Catalytic Mechanisms of Protein Ubiquitylation. *Biochem. J.* **2004**, *379* (3), 513–525.

(230) Kleiger, G.; Mayor, T. Perilous Journey: A Tour of the Ubiquitin–proteasome System. *Trends Cell Biol.* **2014**, *24* (6), 352–359.

(231) Streich, F. C.; Lima, C. D. Structural and Functional Insights to Ubiquitin-like Protein Conjugation. *Annu. Rev. Biophys.* **2014**, *43*, 357–379.

(232) Nussinov, R.; Ma, B.; Tsai, C.-J. A Broad View of Scaffolding Suggests That Scaffolding Proteins Can Actively Control Regulation and Signaling of Multienzyme Complexes through Allostery. *Biochim. Biophys. Acta* **2013**, *1834* (5), 820–829.

(233) Liu, J.; Nussinov, R. The Role of Allostery in the Ubiquitin-Proteasome System. *Crit. Rev. Biochem. Mol. Biol.* **2013**, *48* (2), 89–97.

(234) Ganoth, A.; Tsfadia, Y.; Wiener, R. Ubiquitin: Molecular Modeling and Simulations. *J. Mol. Graph. Model.* **2013**, *46*, 29–40.

(235) Zimmerman, E. S.; Schulman, B. A.; Zheng, N. Structural Assembly of Cullin-RING Ubiquitin Ligase Complexes. *Curr. Opin. Struct. Biol.* **2010**, *20* (6), 714–721.

(236) Genschik, P.; Sumara, I.; Lechner, E. The Emerging Family of CULLIN3-RING Ubiquitin Ligases (CRL3s): Cellular Functions and Disease Implications. *EMBO J.* **2013**, *32* (17), 2307–2320.

(237) Hua, Z.; Vierstra, R. D. The Cullin-RING Ubiquitin-Protein Ligases. *Annu. Rev. Plant Biol.* **2011**, *62*, 299–334.

(238) Liu, J.; Nussinov, R. The Mechanism of Ubiquitination in the Cullin-RING E3 Ligase Machinery: Conformational Control of Substrate Orientation. *PLoS Comput. Biol.* **2009**, *5* (10), e1000527.

(239) Liu, J.; Nussinov, R. Molecular Dynamics Reveal the Essential Role of Linker Motions in the Function of Cullin–RING E3 Ligases. *J. Mol. Biol.* **2010**, *396* (5), 1508–1523.

(240) Liu, J.; Nussinov, R. Rbx1 Flexible Linker Facilitates Cullin-RING Ligase Function before Neddylation and after Deneddylation. *Biophys. J.* **2010**, *99* (3), 736–744.

(241) Liu, J.; Nussinov, R. Flexible Cullins in Cullin-RING E3 Ligases Allosterically Regulate Ubiquitination. *J. Biol. Chem.* **2011**, *286* (47), 40934–40942.

(242) Duda, D. M.; Scott, D. C.; Calabrese, M. F.; Zimmerman, E. S.; Zheng, N.; Schulman, B. A. Structural Regulation of Cullin-RING Ubiquitin Ligase Complexes. *Curr. Opin. Struct. Biol.* **2011**, *21* (2), 257–264.

(243) Duda, D. M.; Borg, L. A.; Scott, D. C.; Hunt, H. W.; Hammel, M.; Schulman, B. A. Structural Insights into NEDD8 Activation of Cullin-RING Ligases: Conformational Control of

Conjugation. *Cell* **2008**, *134* (6), 995–1006.

(244) Boh, B. K.; Smith, P. G.; Hagen, T. Neddylation-Induced Conformational Control Regulates Cullin RING Ligase Activity In Vivo. *J. Mol. Biol.* **2011**, *409* (2), 136–145.

(245) Zheng, N.; Schulman, B. A.; Song, L.; Miller, J. J.; Jeffrey, P. D.; Wang, P.; Chu, C.; Koepp, D. M.; Elledge, S. J.; Pagano, M. et al. Structure of the Cul1–Rbx1–Skp1–F boxSkp2 SCF Ubiquitin Ligase Complex. *Nature* **2002**, *416* (6882), 703–709.

(246) Ye, Y.; Rape, M. Building Ubiquitin Chains: E2 Enzymes at Work. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (11), 755–764.

(247) van Wijk, S. J. L.; Timmers, H. T. M. The Family of Ubiquitin-Conjugating Enzymes (E2s): Deciding between Life and Death of Proteins. *FASEB J.* **2010**, *24* (4), 981–993.

(248) Wenzel, D. M.; Stoll, K. E.; Klevit, R. E. E2s: Structurally Economical and Functionally Replete. *Biochem. J.* **2011**, *433* (1), 31–42.

(249) Michelle, C.; Vourc'h, P.; Mignon, L.; Andres, C. R. What Was the Set of Ubiquitin and Ubiquitin-like Conjugating Enzymes in the Eukaryote Common Ancestor? *J. Mol. Evol.* **2009**, *68* (6), 616–628.

(250) Cook, B. W.; Shaw, G. S. Architecture of the Catalytic HPN Motif Is Conserved in All E2 Conjugating Enzymes. *Biochem. J.* **2012**, *445* (2), 167–174.

(251) Das, R.; Mariano, J.; Tsai, Y. C.; Kalathur, R. C.; Kostova, Z.; Li, J.; Tarasov, S. G.; McFeeters, R. L.; Altieri, A. S.; Ji, X. et al. M. Allosteric Activation of E2-RING Finger-Mediated Ubiquitylation by a Structurally Defined Specific E2-Binding Region of gp78. *Mol. Cell* **2009**, *34* (6), 674–685.

(252) Valimberti, I.; Tiberti, M.; Lambrughi, M.; Sarcevic, B.; Papaleo, E. E2 Superfamily of Ubiquitin-Conjugating Enzymes: Constitutively Active or Activated through Phosphorylation in the Catalytic Cleft. *Sci. Rep.* **2015**, *5*, 14849.

(253) Papaleo, E.; Lindorff-larsen, K.; Gioia, L. De. Paths of Long-Range Communication in the E2 Enzymes of Family 3 : A Molecular Dynamics Investigation W. *Phys. Chem. Chem. Phys.* **2012**, 12515–12525.

(254) Spratt, D. E.; Wu, K.; Kovacev, J.; Pan, Z.-Q.; Shaw, G. S. Selective Recruitment of an E2~ubiquitin Complex by an E3 Ubiquitin Ligase. *J. Biol. Chem.* **2012**, *287* (21), 17374–17385.

(255) Lass, A.; Cocklin, R.; Scaglione, K. M.; Skowyra, M.; Korolev, S.; Goebl, M.; Skowyra, D. The Loop-Less tmCdc34 E2 Mutant Defective in Polyubiquitination in Vitro and in Vivo Supports Yeast Growth in a Manner Dependent on Ubp14 and Cka2. *Cell Div.* **2011**, *6* (1), 7.

(256) Ziemba, A.; Hill, S.; Sandoval, D.; Webb, K.; Bennett, E. J.; Kleiger, G. Multimodal Mechanism of Action for the Cdc34 Acidic Loop: A Case Study for Why Ubiquitin-Conjugating Enzymes Have Loops and Tails. *J. Biol. Chem.* **2013**, *288* (48), 34882–34896.

(257) Lefstin, J. A.; Yamamoto, K. R. Allosteric Effects of DNA on Transcriptional Regulators. *Nature* **1998**, *392* (6679), 885–888.

(258) Kumar, R.; Thompson, E. B. Gene Regulation by the Glucocorticoid Receptor: Structure:function Relationship. *J. Steroid Biochem. Mol. Biol.* **2005**, *94* (5), 383–394.

(259) Nussinov, R. The Spatial Structure of Cell Signaling Systems. *Phys. Biol.* **2013**, *10* (4), 045004.

(260) Nussinov, R.; Tsai, C.-J.; Ma, B. The Underappreciated Role of Allostery in the Cellular Network. *Annu. Rev. Biophys.* **2013**, *42* (1), 169–189.

(261) Meijsing, S. H. Mechanisms of Glucocorticoid-Regulated Gene Transcription. *Adv. Exp. Med. Biol.* **2015**, *872*, 59–81.

(262) Meijsing, S. H.; Pufall, M. A.; So, A. Y.; Bates, D. L.; Chen, L.; Yamamoto, K. R. DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity. *Science* **2009**, *324* (5925), 407–410.

(263) Pan, Y.; Nussinov, R. Cooperativity Dominates the Genomic Organization of p53-Response Elements: A Mechanistic View. *PLoS Comput. Biol.* **2009**, *5* (7), e1000448.

(264) Pan, Y.; Nussinov, R. Lysine120 Interactions with p53 Response Elements Can Allosterically Direct p53 Organization. *PLoS Comput. Biol.* **2010**, *6* (8), e1000878.

(265) Laptenko, O.; Prives, C. Transcriptional Regulation by p53: One Protein, Many Possibilities. *Cell Death Differ.* **2006**, *13* (6), 951–961.

(266) Lukman, S.; Lane, D. P.; Verma, C. S. Mapping the Structural and Dynamical Features of Multiple p53 DNA Binding Domains: Insights into Loop 1 Intrinsic Dynamics. *PLoS One* **2013**, *8* (11), e80221.

(267) Vuzman, D.; Levy, Y. Intrinsically Disordered Regions as Affinity Tuners in protein–DNA Interactions. *Mol. Biosyst.* **2012**, *8* (1), 47.

(268) Bhattacherjee, A.; Levy, Y. Search by Proteins for Their DNA Target Site: 1. The Effect of DNA Conformation on Protein Sliding. *Nucleic Acids Res.* **2014**, *42* (20), 12404–12414.

(269) Vuzman, D.; Polonsky, M.; Levy, Y. Facilitated DNA Search by Multidomain Transcription Factors: Cross Talk via a Flexible Linker. *Biophys. J.* **2010**, *99* (4), 1202–1211.

(270) Khazanov, N.; Levy, Y. Sliding of p53 along DNA Can Be Modulated by Its Oligomeric State and by Cross-Talks between Its Constituent Domains. *J. Mol. Biol.* **2011**, *408* (2), 335–355.

(271) Zandarashvili, L.; Vuzman, D.; Esadze, a.; Takayama, Y.; Sahu, D.; Levy, Y.; Iwahara, J. PNAS Plus: Asymmetrical Roles of Zinc Fingers in Dynamic DNA-Scanning Process by the Inducible Transcription Factor Egr-1. *Proc. Natl. Acad. Sci.* **2012**, *109* (26), E1724–E1732.

(272) Wilsker, D.; Probst, L.; Wain, H. M.; Maltais, L.; Tucker, P. W.; Moran, E. Nomenclature of the ARID Family of DNA-Binding Proteins. *Genomics* **2005**, *86* (2), 242–251.

(273) Kortschak, R. D.; Tucker, P. W.; Saint, R. ARID Proteins Come in from the Desert. *Trends Biochem. Sci.* **2000**, *25* (6), 294–299.

(274) Patsialou, A. DNA-Binding Properties of ARID Family Proteins. *Nucleic Acids Res.* **2005**, *33* (1), 66–80.

(275) Wilsker, D.; Patsialou, A.; Dallas, P. B.; Moran, E. ARID Proteins: A Diverse Family of DNA Binding Proteins Implicated in the Control of Cell Growth, Differentiation, and Development. *Cell Growth Differ.* **2002**, *13* (3), 95–106.

(276) Invernizzi, G.; Tiberti, M.; Lambrughi, M.; Lindorff-Larsen, K.; Papaleo, E. Communication Routes in ARID Domains between Distal Residues in Helix 5 and the DNA-Binding Loops. *PLoS Comput. Biol.* **2014**, *10* (9), e1003744.

(277) Nixon, J. C.; Rajaiya, J.; Webb, C. F. Mutations in the DNA-Binding Domain of the Transcription Factor Bright Act as Dominant Negative Proteins and Interfere with Immunoglobulin Transactivation. *J. Biol. Chem.* **2004**, *279* (50), 52465–52472.

(278) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the Accuracy of Physical Models Used in Protein-Folding Simulations: Quantitative Evidence from Long Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2014**, *24* (1), 98–105.

(279) Luitz, M.; Bomblies, R.; Ostermeir, K.; Zacharias, M. Exploring Biomolecular Dynamics and Interactions Using Advanced Sampling Methods. *J. Phys. Condens. Matter* **2015**, *27* (32), 323101.

(280) Fuglebakk, E.; Tiwari, S. P.; Reuter, N. Comparing the Intrinsic Dynamics of Multiple Protein Structures Using Elastic Network Models. *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850* (5), 911–922.

(281) Bahar, I.; Rader, A. J. Coarse-Grained Normal Mode Analysis in Structural Biology. *Curr. Opin. Struct. Biol.* **2005**, *15* (5), 586–592.

(282) Adcock, S. A.; McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev.* **2006**, *106* (5), 1589–1615.

(283)   Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8* (4), 1409–1414.

(284)   Best, R. B.; Buchete, N.-V.; Hummer, G. Are Current Molecular Dynamics Force Fields Too Helical? *Biophys. J.* **2008**, *95* (1), L07–L09.

(285)   Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS One* **2012**, *7* (2), e32131.

(286)   Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133* (24), 9181–9183.

(287)   Shaw, D. E.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P. et al. Millisecond-Scale Molecular Dynamics Simulations on Anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis - SC '09*; ACM Press: New York, New York, USA, 2009; p 1.

(288)   Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1-2), 141–151.

(289)   Hansmann, U. H. E. Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **1997**, *281* (1-3), 140–150.

(290)   Lei, H.; Wu, C.; Liu, H.; Duan, Y. Folding Free-Energy Landscape of Villin Headpiece Subdomain from Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (12), 4925–4930.

(291)   Zhou, R.; Berne, B. J.; Germain, R. The Free Energy Landscape for Beta Hairpin Folding in Explicit Water. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (26), 14931–14936.

(292)   Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C.-K.; Li, M. S. Free Energy Landscape and Folding Mechanism of a Beta-Hairpin in Explicit Water: A Replica Exchange Molecular Dynamics Study. *Proteins* **2005**, *61* (4), 795–808.

(293)   Pitera, J. W.; Swope, W. Understanding Folding and Design: Replica-Exchange Simulations of "Trp-Cage" Miniproteins. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (13), 7587–7592.

(294)   Sanbonmatsu, K. Y.; García, A. E. Structure of Met-Enkephalin in Explicit Aqueous Solution Using Replica Exchange Molecular Dynamics. *Proteins* **2002**, *46* (2), 225–234.

(295)   Kubitzki, M. B.; de Groot, B. L. The Atomistic Mechanism of Conformational Transition in Adenylate Kinase: A TEE-REX Molecular Dynamics Study. *Structure* **2008**, *16* (8), 1175–1182.

(296)   Palazzesi, F.; Barducci, A.; Tollinger, M.; Parrinello, M. The Allosteric Communication Pathways in KIX Domain of CBP. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (35), 14237–14242.

(297)   Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199.

(298)   Arora, K.; Brooks III, C. L. Functionally Important Conformations of the Met20 Loop in Dihydrofolate Reductase Are Populated by Rapid Thermal Fluctuations. *J. Am. Chem. Soc.* **2009**, *131* (15), 5642–5647.

(299)   Ravindranathan, K. P.; Gallicchio, E.; Levy, R. M. Conformational Equilibria and Free Energy Profiles for the Allosteric Transition of the Ribose-Binding Protein. *J. Mol. Biol.* **2005**, *353* (1), 196–210.

(300)   Meng, Y.; Roux, B. Locking the Active Conformation of c-Src Kinase through the Phosphorylation of the Activation Loop. *J. Mol. Biol.* **2014**, *426* (2), 423–435.

(301)   Banavali, N. K.; Roux, B. Flexibility and Charge Asymmetry in the Activation Loop of Src Tyrosine Kinases. *Proteins Struct. Funct. Bioinforma.* **2009**, *74* (2), 378–389.

(302) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Reports Prog. Phys.* **2008**, *71* (12), 126601.

(303) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562–12566.

(304) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A. et al. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180* (10), 1961–1972.

(305) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613.

(306) Sutto, L.; Gervasio, F. L. Effects of Oncogenic Mutations on the Conformational Free-Energy Landscape of EGFR Kinase. *Proc. Natl. Acad. Sci.* **2013**, *110* (26), 10616–10621.

(307) Marino, K. A.; Sutto, L.; Gervasio, F. L. The Effect of a Widespread Cancer-Causing Mutation on the Inactive to Active Dynamics of the B-Raf Kinase. *J. Am. Chem. Soc.* **2015**, *137* (16), 5280–5283.

(308) Bunney, T. D.; Wan, S.; Thiyagarajan, N.; Sutto, L.; Williams, S. V; Ashford, P.; Koss, H.; Knowles, M. A.; Gervasio, F. L.; Coveney, P. V et al. The Effect of Mutations on Drug Sensitivity and Kinase Activity of Fibroblast Growth Factor Receptors: A Combined Experimental and Theoretical Study. *EBioMedicine* **2015**, *2* (3), 194–204.

(309) Chennubhotla, C.; Bahar, I. Markov Propagation of Allosteric Effects in Biomolecular Systems: Application to GroEL-GroES. *Mol. Syst. Biol.* **2006**, *2*, 36.

(310) Yang, L.-W.; Eyal, E.; Bahar, I.; Kitao, A. Principal Component Analysis of Native Ensembles of Biomolecular Structures (PCA_NEST): Insights into Functional Dynamics. *Bioinformatics* **2009**, *25* (5), 606–614.

(311) Bahar, I.; Atilgan, A. R.; Erman, B. Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential. *Fold. Des.* **1997**, *2* (3), 173–181.

(312) Erman, B. The Gaussian Network Model: Precise Prediction of Residue Fluctuations and Application to Binding Problems. *Biophys. J.* **2006**, *91* (10), 3589–3599.

(313) Eyal, E.; Yang, L.-W.; Bahar, I. Anisotropic Network Model: Systematic Evaluation and a New Web Interface. *Bioinformatics* **2006**, *22* (21), 2619–2627.

(314) Mendez, R.; Bastolla, U. Torsional Network Model: Normal Modes in Torsion Angle Space Better Correlate with Conformation Changes in Proteins. *Phys. Rev. Lett.* **2010**, *104* (22), 228103.

(315) Bakan, A.; Bahar, I. The Intrinsic Dynamics of Enzymes Plays a Dominant Role in Determining the Structural Changes Induced upon Inhibitor Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (34), 14349–14354.

(316) Gur, M.; Zomot, E.; Bahar, I. Global Motions Exhibited by Proteins in Micro- to Milliseconds Simulations Concur with Anisotropic Network Model Predictions. *J. Chem. Phys.* **2013**, *139* (12), 121912.

(317) Williams, G. Elastic Network Model of Allosteric Regulation in Protein Kinase PDK1. *BMC Struct. Biol.* **2010**, *10*, 11.

(318) Kurkcuoglu, O.; Jernigan, R. L.; Doruker, P. Loop Motions of Triosephosphate Isomerase Observed with Elastic Networks. *Biochemistry* **2006**, *45* (4), 1173–1182.

(319) Kurkcuoglu, O.; Turgut, O. T.; Cansu, S.; Jernigan, R. L.; Doruker, P. Focused Functional Dynamics of Supramolecules by Use of a Mixed-Resolution Elastic Network Model. *Biophys. J.* **2009**, *97* (4), 1178–1187.

(320) Zheng, W.; Doniach, S. A Comparative Study of Motor-Protein Motions by Using a Simple Elastic-Network Model. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13253–13258.

(321) Zheng, W.; Liao, J.-C.; Brooks, B. R.; Doniach, S. Toward the Mechanism of Dynamical

Couplings and Translocation in Hepatitis C Virus NS3 Helicase Using Elastic Network Model. *Proteins* **2007**, *67* (4), 886–896.

(322) Song, G.; Jernigan, R. L. An Enhanced Elastic Network Model to Represent the Motions of Domain-Swapped Proteins. *Proteins* **2006**, *63* (1), 197–209.

(323) Yang, Z.; Májek, P.; Bahar, I. Allosteric Transitions of Supramolecular Systems Explored by Network Models: Application to Chaperonin GroEL. *PLoS Comput. Biol.* **2009**, *5* (4), e1000360.

(324) Okada, T.; Sugihara, M.; Bondar, A.-N.; Elstner, M.; Entel, P.; Buss, V. The Retinal Conformation and Its Environment in Rhodopsin in Light of a New 2.2 A Crystal Structure. *J. Mol. Biol.* **2004**, *342* (2), 571–583.

(325) Taly, A.; Delarue, M.; Grutter, T.; Nilges, M.; Le Novère, N.; Corringer, P.-J.; Changeux, J.-P. Normal Mode Analysis Suggests a Quaternary Twist Model for the Nicotinic Receptor Gating Mechanism. *Biophys. J.* **2005**, *88* (6), 3954–3965.

(326) 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* (3), 1463–1497.

(327) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular Dynamics Simulations of Large Macromolecular Complexes. *Curr. Opin. Struct. Biol.* **2015**, *31*, 64–74.

(328) Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. Global Ribosome Motions Revealed with Elastic Network Model. *J. Struct. Biol.* **2004**, *147* (3), 302–314.

(329) Burton, B.; Zimmermann, M. T.; Jernigan, R. L.; Wang, Y. A Computational Investigation on the Connection between Dynamics Properties of Ribosomal Proteins and Ribosome Assembly. *PLoS Comput. Biol.* **2012**, *8* (5), e1002530.

(330) Zhang, Z.; Sanbonmatsu, K. Y.; Voth, G. A. Key Intermolecular Interactions in the E. Coli 70S Ribosome Revealed by Coarse-Grained Analysis. *J. Am. Chem. Soc.* **2011**, *133* (42), 16828–16838.

(331) Eyal, E.; Lum, G.; Bahar, I. The Anisotropic Network Model Web Server at 2015 (ANM 2.0). *Bioinformatics* **2015**, *31* (9), 1487–1489.

(332) Das, A.; Gur, M.; Cheng, M. H.; Jo, S.; Bahar, I.; Roux, B. Exploring the Conformational Transitions of Biomolecular Systems Using a Simple Two-State Anisotropic Network Model. *PLoS Comput. Biol.* **2014**, *10* (4), e1003521.

(333) Daidone, I.; Amadei, A. Essential Dynamics: Foundation and Applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (5), 762–770.

(334) Lange, O. F.; Grubmüller, H. Can Principal Components Yield a Dimension Reduced Description of Protein Dynamics on Long Time Scales? *J. Phys. Chem. B* **2006**, *110* (45), 22842–22852.

(335) Lange, O. F.; Grubmüller, H. Full Correlation Analysis of Conformational Protein Dynamics. *Proteins* **2008**, *70* (4), 1294–1312.

(336) Savol, A. J.; Burger, V. M.; Agarwal, P. K.; Ramanathan, A.; Chennubhotla, C. S. QAARM: Quasi-Anharmonic Autoregressive Model Reveals Molecular Recognition Pathways in Ubiquitin. *Bioinformatics* **2011**, *27* (13), i52–i60.

(337) Ramanathan, A.; Savol, A. J.; Langmead, C. J.; Agarwal, P. K.; Chennubhotla, C. S. Discovering Conformational Sub-States Relevant to Protein Function. *PLoS One* **2011**, *6* (1), e15827.

(338) Wriggers, W.; Stafford, K. A.; Shan, Y.; Piana, S.; Maragakis, P.; Lindorff-Larsen, K.; Miller, P. J.; Gullingsrud, J.; Rendleman, C. A.; Eastwood, M. P. et al. Automated Event Detection and Activity Monitoring in Long Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2009**, *5* (10), 2595–2605.

(339) Ramanathan, A.; Savol, A. J.; Agarwal, P. K.; Chennubhotla, C. S. Event Detection and Sub-State Discovery from Biomolecular Simulations Using Higher-Order Statistics: Application to Enzyme Adenylate Kinase. *Proteins Struct. Funct. Bioinforma.* **2012**, *80* (11), 2536–2551.

(340) Tiberti, M.; Papaleo, E.; Bengtsen, T.; Boomsma, W.; Lindorff-Larsen, K. ENCORE: Software for Quantitative Ensemble Comparison. *PLoS Comput. Biol.* **2015**, *in press*.

(341) Swett, R. J.; Cisneros, G. A.; Feig, A. L. Disruption of Intrinsic Motions as a Mechanism for Enzyme Inhibition. *Biophys. J.* **2013**, *105* (2), 494–501.

(342) Scarabelli, G.; Grant, B. J. Mapping the Structural and Dynamical Features of Kinesin Motor Domains. *PLoS Comput. Biol.* **2013**, *9* (11), e1003329.

(343) Fan, Z.; Dror, R. O.; Mildorf, T. J.; Piana, S.; Shaw, D. E. Identifying Localized Changes in Large Systems: Change-Point Detection for Biomolecular Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (24), 7454–7459.

(344) Collier, G.; Ortiz, V. Emerging Computational Approaches for the Study of Protein Allostery. *Arch. Biochem. Biophys.* **2013**, *538* (1), 6–15.

(345) Di Paola, L.; Giuliani, A. Protein Contact Network Topology: A Natural Language for Allostery. *Curr. Opin. Struct. Biol.* **2015**, *31*, 43–48.

(346) Tiberti, M.; Invernizzi, G.; Lambrughi, M.; Inbar, Y.; Schreiber, G.; Papaleo, E. PyInteraph : A Framework for the Analysis of Interaction Networks in Structural Ensembles of Proteins. *J. Chem. Inf. Model* **2014,** *54* (5)*,* 1537-51.

(347) Vendruscolo, M.; Dokholyan, N. V; Paci, E.; Karplus, M. Small-World View of the Amino Acids That Play a Key Role in Protein Folding. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **2002**, *65* (6 Pt 1), 061910.

(348) Atilgan, A. R.; Akan, P.; Baysal, C. Small-World Communication of Residues and Significance for Protein Dynamics. *Biophys. J.* **2004**, *86* (1 Pt 1), 85–91.

(349) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712–725.

(350) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M. Ø.; Xu, H.; Trbovic, N.; Friesner, R. A. et al. Microsecond Molecular Dynamics Simulation Shows Effect of Slow Loop Dynamics on Backbone Amide Order Parameters of Proteins †. *J. Phys. Chem. B* **2008**, *112* (19), 6155–6158.

(351) Lange, O. F.; van der Spoel, D.; de Groot, B. L. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophys. J.* **2010**, *99* (2), 647–655.

(352) Robustelli, P.; Stafford, K. A.; Palmer, A. G. Interpreting Protein Structural Dynamics from NMR Chemical Shifts. *J. Am. Chem. Soc.* **2012**, *134* (14), 6365–6374.

(353) Best, R. B. Atomistic Molecular Simulations of Protein Folding. *Curr. Opin. Struct. Biol.* **2012**, *22* (1), 52–61.

(354) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the Accuracy of Physical Models Used in Protein-Folding Simulations: Quantitative Evidence from Long Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98–105.

(355) Norgaard, A. B.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Experimental Parameterization of an Energy Function for the Simulation of Unfolded Proteins. *Biophys. J.* **2008**, *94* (1), 182–192.

(356) Li, D.-W.; Brüschweiler, R. Iterative Optimization of Molecular Mechanics Force Fields from NMR Data of Full-Length Proteins. *J. Chem. Theory Comput.* **2011**, *7* (6), 1773–1782.

(357) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113* (26), 9004–9015.

(358) Różycki, B.; Kim, Y. C.; Hummer, G. SAXS Ensemble Refinement of ESCRT-III CHMP3 Conformational Transitions. *Structure* **2011**, *19* (1), 109–116.

(359) Vendruscolo, M. Determination of Conformationally Heterogeneous States of Proteins. *Curr. Opin. Struct. Biol.* **2007**, *17* (1), 15–20.

(360) Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining Experiments and Simulations Using the Maximum Entropy Principle. *PLoS Comput. Biol.* **2014**, *10* (2), e1003406.

(361) Olsson, S.; Vögeli, B. R.; Cavalli, A.; Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K.; Hamelryck, T. Probabilistic Determination of Native State Ensembles of Proteins. *J. Chem. Theory Comput.* **2014**, *10* (8), 3484–3491.

(362) Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M. Metainference: A Bayesian Inference Method for Heterogeneous Systems. *Science Adv.* **2015**, *2* (1), e1501177.

(363) Hummer, G.; Köfinger, J. Bayesian Ensemble Refinement by Replica Simulations and Reweighting. *J Chem Physics* **2015**, *143,* 243150.

(364) Mayer, K. L.; Earley, M. R.; Gupta, S.; Pichumani, K.; Regan, L.; Stone, M. J. Covariation of Backbone Motion throughout a Small Protein Domain. *Nat. Struct. Biol.* **2003**, *10* (11), 962–965.

(365) Lange, O. F.; Grubmüller, H.; de Groot, B. L. Molecular Dynamics Simulations of Protein G Challenge NMR-Derived Correlated Backbone Motions. *Angew. Chem. Int. Ed. Engl.* **2005**, *44* (22), 3394–3399.

(366) Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Bruschweiler, R.; Blackledge, M. Identification of Slow Correlated Motions in Proteins Using Residual Dipolar and Hydrogen-Bond Scalar Couplings. *Proc. Natl. Acad. Sci.* **2005**, *102* (39), 13885–13890.

(367) Fenwick, R. B.; Esteban-Martín, S.; Richter, B.; Lee, D.; Walter, K. F. A.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. Weak Long-Range Correlated Motions in a Surface Patch of Ubiquitin Involved in Molecular Recognition. *J. Am. Chem. Soc.* **2011**, *133* (27), 10336–10339.

(368) Esteban-Martín, S.; Fenwick, R. B.; Ådén, J.; Cossins, B.; Bertoncini, C. W.; Guallar, V.; Wolf-Watz, M.; Salvatella, X. Correlated Inter-Domain Motions in Adenylate Kinase. *PLoS Comput. Biol.* **2014**, *10* (7), e1003721.

(369) Wang, Y.; Gan, L.; Wang, E.; Wang, J. Exploring the Dynamic Functional Landscape of Adenylate Kinase Modulated by Substrates. *J. Chem. Theory Comput.* **2013**, *9* (1), 84–95.

(370) Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D. Intrinsic Dynamics of an Enzyme Underlies Catalysis. *Nature* **2005**, *438* (7064), 117–121.

(371) Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T. Hidden Alternative Structures of Proline Isomerase Essential for Catalysis. *Nature* **2009**, *462* (7273), 669–673.

(372) Wall, M. E.; Adams, P. D.; Fraser, J. S.; Sauter, N. K. Diffuse X-Ray Scattering to Model Protein Motions. *Structure* **2014**, *22* (2), 182–184.

(373) Wall, M. E.; Ealick, S. E.; Gruner, S. M. Three-Dimensional Diffuse X-Ray Scattering from Crystals of Staphylococcal Nuclease. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (12), 6180–6184.

(374) Clarage, J. B.; Romo, T.; Andrews, B. K.; Pettitt, B. M.; Phillips, G. N. A Sampling Problem in Molecular Dynamics Simulations of Macromolecules. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92* (8), 3288–3292.

(375) Wall, M. E.; Van Benschoten, A. H.; Sauter, N. K.; Adams, P. D.; Fraser, J. S.; Terwilliger, T. C. Conformational Dynamics of a Crystalline Protein from Microsecond-Scale Molecular Dynamics Simulations and Diffuse X-Ray Scattering. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (50), 17887–17892.

(376) Tzeng, S.-R.; Kalodimos, C. G. Protein Activity Regulation by Conformational Entropy. *Nature* **2012**, *488* (7410), 236–240.

(377) Tzeng, S.-R.; Kalodimos, C. G. Dynamic Activation of an Allosteric Regulatory Protein.

Nature **2009**, *462* (7271), 368–372.

(378) Nussinov, R.; Tsai, C. J. Free Energy Diagrams for Protein Function. *Chem. Biol.* **2014**, *21* (3), 311–318.

(379) Nussinov, R.; Tsai, C.-J. Tandem Domains with Tuned Interactions Are a Powerful Biological Design Principle. *PLOS Biol.* **2015**, *13* (11), e1002306.

(380) Bennett, M. J.; Choe, S.; Eisenberg, D. Domain Swapping: Entangling Alliances between Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (8), 3127–3131.

Biographies

Elena Papaleo received her Ph.D. in 2006 from University of Milano-Bicocca (Italy) and did her Post Doctoral works from 2007-2009 at the Department of Biotechnology and Biosciences of Milano-Bicocca University (Italy) in the group of Prof. Luca De Gioia. She was then appointed as Adjunct Professor in Computational Biology at the University of Milano-Bicocca from 2010 to 2012. Afterwards, she was senior Post Doctoral Researcher in the group of Prof. Kresten Lindorff-Larsen at the Department of Biology of University of Copenhagen (Denmark) from 2011 to 2015. She has been Visiting Researcher at many international institutes included the group of Prof. Salvador Ventura at the Institute of Biotechnology and Biomedicine (IBB, Barcelona, Spain) and the group of Prof. Francesco Luigi Gervasio at The Spanish National Cancer Research Center (CNIO, Madrid, Spain). In August 2015, she joined as Group Leader of the Computational Biology Laboratory at the Danish Cancer Society Research Center (Copenhagen, Denmark). She has authored more than 50 scientific papers as main or senior author and she is Academic Editor of PLoS One, Frontiers in Molecular Biosciences, PeerJ, and Journal of Molecular Graphics and Modeling. Her research focuses on molecular simulations and modelling integrated with biophysical data and network theory applied to the study of conformational changes and dynamics of many key players in cellular biology and disease, included transcription factors, ubiquitinating and deubiquitinating enzymes, post-translational modifications,

intrinsically disordered proteins, metallo-proteins, DNA damage/repair and autophagy.

Giorgio Saladino received his PhD in Physical Chemistry in 2010 from the Universita' degli Studi di Milano (Italy). In 2011, he joined the group of Computational Biophysics at the Spanish National Cancer Research Center (CNIO) in Madrid (Spain). In 2013, he joined the University College London (UCL) as a Research Assistant first in the Institute of Structural and Molecular Biology and later in the Department of Chemistry. He published several research article in high impact journals, as Nature Chemistry and Cancer Cell, including a review on drug binding published in Accounts of Chemical Research.

Matteo Lambrughi is currently Post Doctoral Researcher at the Department of Life Sciences at the University of Modena and Reggio Emilia, Italy. He received his Ph.D. in Industrial Biotechnology from the Department of Biology and Biotechnology at University of Milano Bicocca, Italy, in February 2015 under the supervision of Prof. Luca De Gioia and Dr. Elena Papaleo. During his Ph.D., he has been Visiting Researcher at the Structural Biology and NMR Laboratory at the University of Copenhagen, Denmark, in the group of Prof. Kaare Teilum under the supervision of Dr. Gaetano Invernizzi. His research interests focus on computational studies of intrinsically disordered domains and metal-binding proteins, as well as of structural effects of post-translational modifications. In particular, he has been integrating approaches from quantum chemistry and classical molecular dynamics simulations to explore biological problems with implications on human health, such as cancer, neurodegeneration and effects of tumorigenic agents.

Kresten Lindorff-Larsen was a graduate student at the University of Cambridge with Christopher Dobson and Michele Vendruscolo, where he developed and applied methods to integrate experiments and simulations to study protein structure, folding and dynamics. After receiving his Ph.D. in 2004 he was an EMBO fellow and assistant professor at the Institute of Molecular Biology, University of Copenhagen. Between 2007 and 2011 he was a staff scientist at D.E. Shaw Research, New York, U.S.A., where his research included studies of protein folding and improving molecular energy functions. In 2011 he returned to the University of Copenhagen as an associate professor in the Structural Biology and NMR Laboratory and Linderstrøm-Lang Centre for Protein Science at the Department of Biology. His lab studies the relationship between protein structure, dynamics and function, as well as how mutations may give rise to drug resistance and disease. In 2006 he received the Danish Independent Research Councils' Young Researchers' Award, in 2009 he was a co-recipient of the ACM Gordon Bell Prize, in 2011 he received a Hallas-Møller award from the Novo Nordisk Foundation and in 2012 he was awarded a Sapere Aude Starting Grant from the Danish Council for Independent research.

Francesco Luigi Gervasio is professor of Chemistry, professor of Structural and Molecular Biology and holds a Chair of bio-molecular modelling at University College London. He holds a Ph.D. degree in chemistry from the University of Firenze and has been a Post-Doc and an Oberassistant in the group of Michele Parrinello from 2002 to 2009, before joining the Spanish National Cancer Research Centre as group leader (2009-2013) and UCL as full professor in 2013. His research focuses on integration of computer simulations, free energy calculations, and NMR spectroscopy

to understand in details the dynamics of biomolecular systems, with a focus in the activation dynamics of oncogenic kinases. He developed methods to study rare events that occurs on time-scales that are not accessible to standard ab-initio and classical molecular dynamics simulations. In collaboration with Sanofi-Aventis and several academic partners, he has described the mode of action of one of the first allosteric inhibitors of a receptor tyrosine kinase; Fibroblast Growth Factor Receptor (FGFR). He also has a strong background in ab-initio molecular dynamics and multi-scale methods (QM/MM) which he applied to study charge transport in DNA and the catalytic properties of natural occurring alkaloids.

Ruth Nussinov received her Ph.D. in 1977 from Rutgers University and did post-doctoral work in the Structural Chemistry Department of the Weizmann Institute. Subsequently she was at the Chemistry Department at Berkeley, the Biochemistry Department at Harvard, and a visiting scientist at the NIH. In 1984 she joined the Department of Human Genetics, at the Medical School at Tel Aviv University. In 1985, she accepted a concurrent position at the National Cancer Institute of the NIH, Leidos Biomedical Research, where she is a Senior Principal Scientist and Principle Investigator heading the Computational Structural Biology Chapter at the NCI. She has authored over 500 scientific papers. She is the Editor-in-Chief in PLoS Computational Biology and Associate Editor and on the Editorial Boards of a number of journals. She is a frequent speaker in Domestic and International meetings, symposia and academic institutions, won several awards and elected fellow of several societies. Her National Cancer Institute website gives further details. https://ccr.cancer.gov/ruth-nussinov
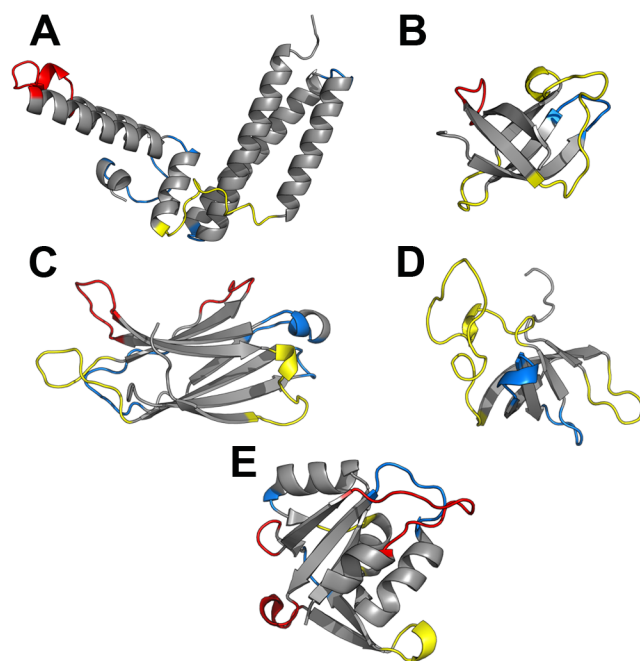
Figure 1 Time scales of loop dynamics in proteins. The loops are classified as proposed by Brüschweiler and coworkers[59] based on the Cα−Cβ bond vector S^2 order parameters calculated from Molecular Dynamics trajectories and compared to NMR chemical shifts. The time scales of loop dynamics have been divided accordingly to their flexibility as "static" over the length of the MD trajectory (500 ns) indicated with azure, "fast" loops with correlation times smaller than 10 ns (red) and "slow" loops with correlation times larger than 10 ns (yellow). In the figure we reported five representative proteins studied in the work: A) Hsc20 (PDB 1FPO) B) CspA (PDB 1MJC) C) C2 domain of synaptotagmin I **(**PDB 1RSY) D) S1 domain of RNase E (PDB 1SMX) E) Protein CC0527 **(**PDB 2O0Q).
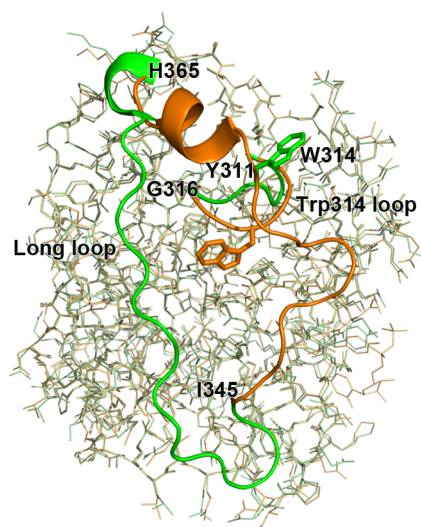
Figure 2. Triggering and triggered loops. The two crystal structures of the β1,4-galactosyltransferase-I (β4-Gal-T1) in the unbound (green PDB 1FGX) and substrate bound state (orange PDB 1NKH). Upon the substrate binding the functional long (Ile345-His365, the triggered loop) loop undergoes to a large conformational change of more than 20 Å, important for enzyme activity. The displacement of long loop is triggered and controlled by changes in the interaction with a shorter and highly conserved among β4-Gal-T1 family, Trp314 loop (Tyr311-Gly316 the triggering loop), that undergo to little conformational changes. The two loops are highlighted as cartoon, and the Trp314 is indicated with sticks.
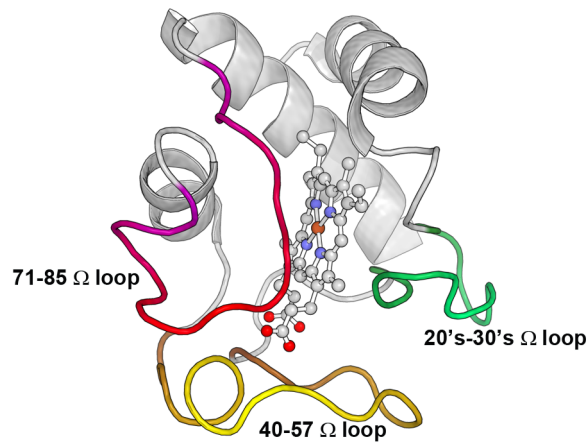
Figure 3 Omega loops are often involved in molecular recognition and regulatory functions. Furthermore they are frequently associated with allosterically regulated regions of the proteins. The Ω loops of cytochrome c (PDB 1HRC) are here showed: 20's-30's Ω loop (green) 40-57 Ω loop (yellow, residues Thr40-Ile57) and 71-85 Ω loop (red, Pro71-Ile85). It has been showed that the 40-57 Ω loop acts as a cooperative unfolding/refolding elements in the cytochrome c.[75] The heme is highlighted with ball and stick representation.
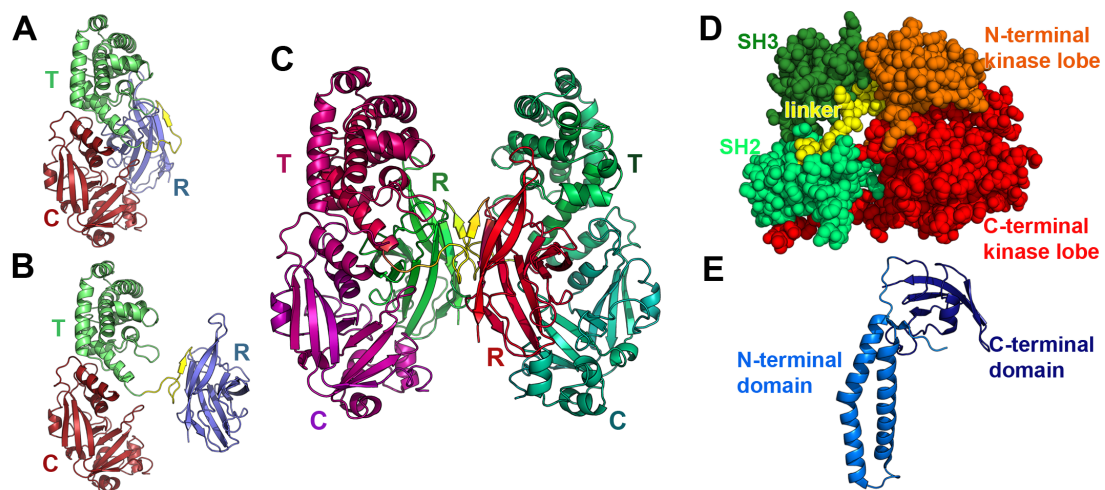
Figure 4 Soft and Rigid Linkers and molecular rulers. Left panel. Domain swapping of dimeric diphtheria toxin (DT). A) The DT monomer in the monomeric closed state (PDB 1MDT) and B) monomeric open state (PDB 1DDT). The DT domains are the catalytic (C, dark red), translocation (T, light green), and receptor domain (R, azure). B) The DT dimer in the dimeric state with two open monomers interconnected by their R domains (PDB 1DDT). The soft loop, highlighted in yellow, between the R and T domains of DT allow the transition from closed to peon forms and mediates the rotation and twists of the swapped R domain in the dimer.[380] Right panel, examples of rigid loops. D) Structure of Human Src protein-kinase (PDB 2SRC) shown with sphere representation. Src protein kinase comprises a catalytic domain, composed by the C-terminal (red) and N-terminal (orange) lobe, and regulatory domains SH2 (cyan) and SH3 (dark green). The polyproline type II helix of SH2 linker, shown in yellow, of Src family kinases is functionally important for the kinase activity. E) Structure of the bacterial transcript cleavage factor GreB (PDB 2P4V), represented with cartoon. The protein has a C-terminal (dark blue) and a N-terminal coiled-coil domain (azure). In

the bacterial transcription elongation factors of the Gre families the N-terminal domain acts as a molecular ruler to helps in determining the length of RNA cleaved.
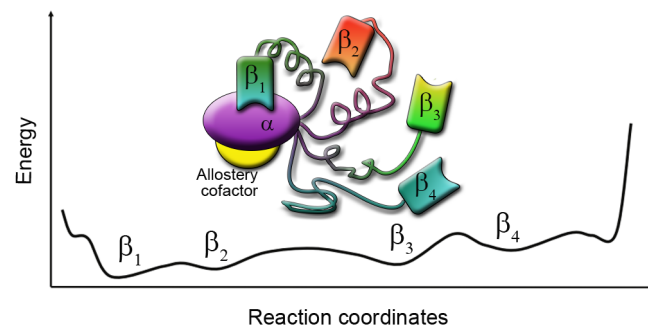

Figure 5. Successive pre-encoded conformational states of linker regions in multi-domain proteins. α and β indicate two domains connected by a linker. $\beta1$ to $\beta4$ are different conformations of the second domain that are allowed through conformational transitions in the linker region.
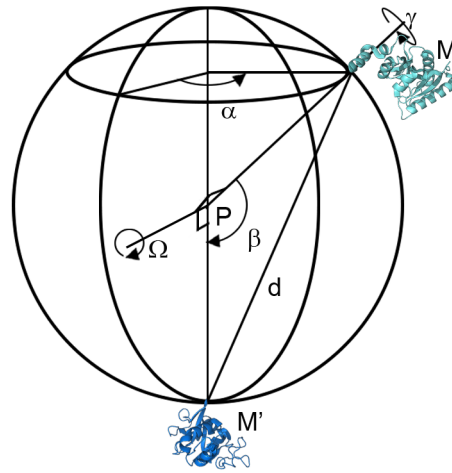
Figure 6. Geometrical representation of hinges. The hinge region is indicated as P, whereas the two domains of lactoferrin (PDB entry 1LFH) are shown as cartoon (N2 domain blue, N1 domain cyan) and here indicated as M and M', to exemplify domains that move as rigid bodies attached by a hinge. Ω indicates the effective rotation axis between M and M'. The geometrical parameters to describe the hinge motions are the distance d between M and M', the hinge angle β (named 'latitude'), the angle α(named 'longitude') and the twist angle γ.
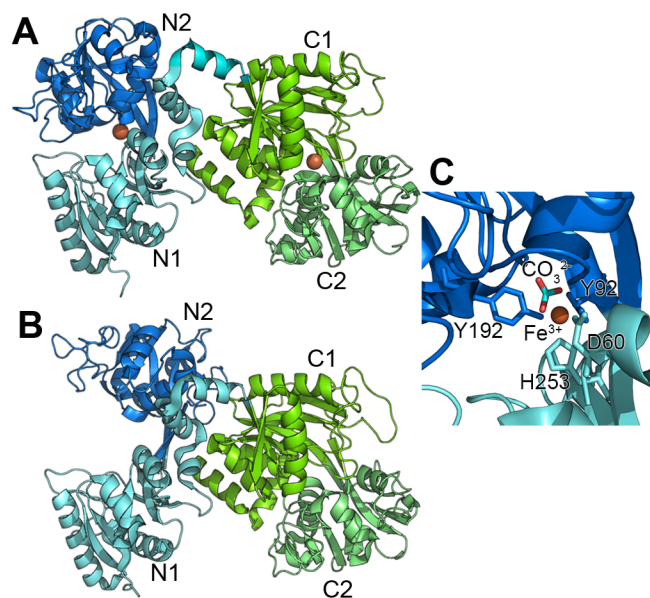
Figure 7. Hinge motions induced by iron binding in human lactoferrin. The crystallographic structures of the iron-bound (holo) (PDB entry 1LFG) and iron-unbound (apo) (PDB entry 1LFH) states are shown as cartoon in panel A and B respectively. The domains of lactoferrin are highlighted with different color: N1 cyan, N2 blue, C1 green, C2 green-cyan, while iron atoms are represented by red spheres. The two structures show that iron binding induces conformational changes in N1 and N2 domains of lactoferrin (similar motions have been reported also for the C domains), closing the cleft between them and sequestering the iron-binding site from the solvent. The hinge motions permit the formation of the complete iron-binding site of lactoferrin, shown in panel C. The binding site is composed by residues in the N1 domain, Asp60 and His253, N2 domain, Tyr92 and Tyr192 and a carbonate ion. The metal coordinates are represented as sticks.
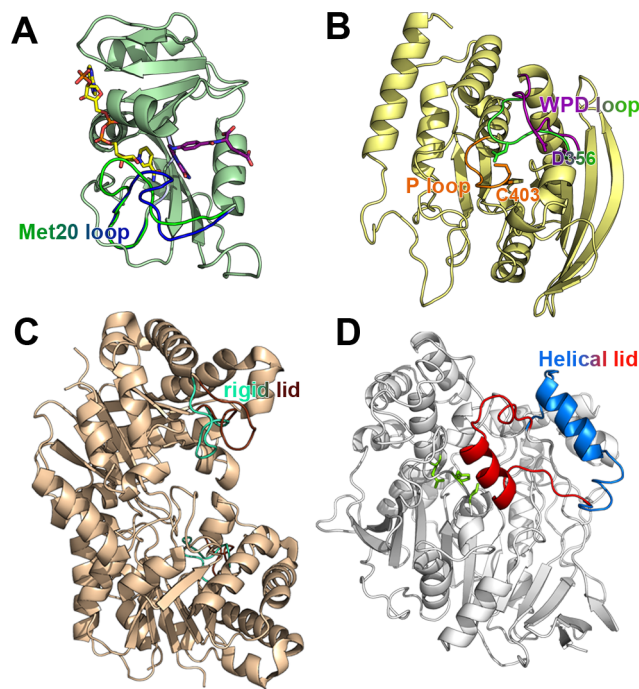
Figure 8. Examples of loops and linkers involved in allosteric regulation of enzymes. A) Dynamics in the dihydrofolate reductase (DHFR). The crystal structures of E. coli DHFR, shown as cartoon, point out that the active site loop (Met20 loop, residues 9-24) can assume occluded (1RX4, blue) and closed (1RX2, green) conformations. In the occluded conformation the Met20 loop extends towards the active site blocking the nicotinamide-binding site, while in the closed conformation it moves near the nicotinamide ring of NADPH. These motions of the Met20 loop are involved in the regulation of different phases of the catalytic cycle. The two cofactors, NADPH and folate, are represented in yellow and purple, respectively. B) The *Yersinia* protein tyrosine phosphatase (PTP) YopH is a highly active virulence factor. The crystal structure of the ligand-unbound (PDB 1YPT) and the ligand-bound YopH in complex with a phosphate anion (PDB 1LYV) are shown as light yellow cartoon.YopH has a highly conserved catalytic site called "P-loop", shown in orange, that includes the nucleophilic Cys403, highlighted with stick representation. As other

PTPs YopH has the WPD loop (residues 351-361) that comprises the catalytic Asp356, shown as stick. For clarity only the WDP loop is shown for the closed structure. The WPD loop has open and inactive conformations, shown in purple for the ligand-unbound structure, and can assume closed and enzymatically active conformations, shown in green for the ligand-bound structure. Such motions of WPD loop are important to correctly orient the catalytic Asp356 for catalysis.

C)Free (PDB 1YPI) and ligand bound (PDB 2YPI) crystallographic structures of yeast triosephosphate isomerase dimer. In the free triosephosphate isomerase the loop (residues 166-176) is in "open" state, shown in brown. In ligand bound state the loop moves for around 7 Å as a rigid lid towards the active site and assumes a "closed" conformation, shown in cyan. These motions of the rigid-lid near to active site are important for the catalytic mechanism of the enzyme.

D) The *Candida rugosa* lipase, as several other lipases, has a helical loop (residues 66-92) that is important for the enzyme function, working as a lid to open or close the hydrophobic active site. The crystal structures of *Candida rugosa* lipase for the open (PDB 1CRL) and closed form (PDB 1TRH) are shown as cartoon and the lid is highlighted in azure and red, respectively. For clarity only the lid is shown for the closed structure. The residue in the active site that are shown as green sticks.
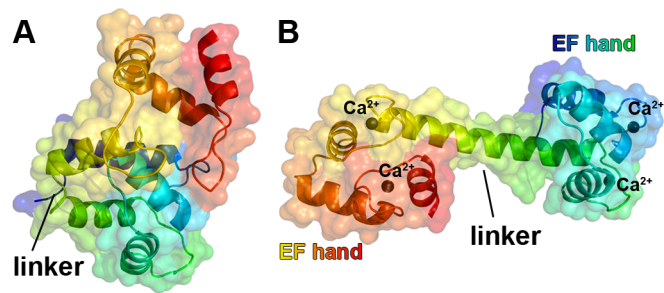
Figure 9. The calmodulin (Cam) has a linker important in cofactor binding. The crystal structure of Ca^{2+} unbound (A, PDB 1QX5) and Ca^{2+} bound (B, PDB 1EXR). The protein structures are shown as cartoon and surface and highlighted as rainbow gradient from N-terminal (blue) to C-terminal (red). The Ca^{2+} atoms are represented as black spheres. The Cam has two globular domains, each one contains a pair of Ca^{2+} binding motifs called EF-hands, connected by a flexible linker. The interaction with Ca^{2+} is associated with changes in the linker conformation and in the orientations of the helices in EF-hands, corresponding to transition from a "closed" (A) to a "open" Ca2+-bound form (B). These motions expose a hydrophobic patch between the domains that forms a binding site to interact and activate with multiple cofactor proteins. The linker is highly conserved and its flexibility is essential for the Cam functions.
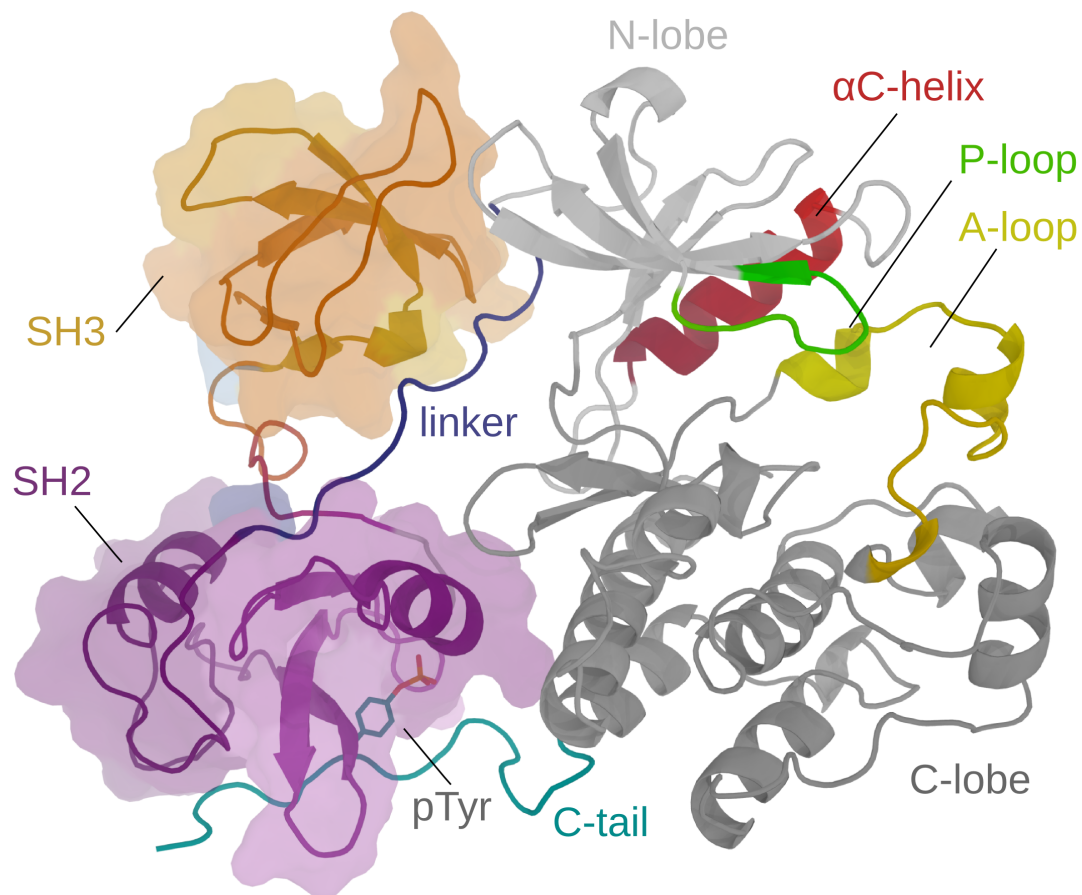
Figure 10. Structure of the Src protein kinase. The catalytic domain (on the right) shows the typical structural features of protein kinases: the αC-helix (red), the P-loop (or glycine-rich loop, green) and the activation loop (A-loop, yellow). The two lobes are highlighted in light grey (N-terminal lobe) and dark grey (C-terminal lobe). The long C-terminal tail typical of Src is shown in cyan with the phospho-tyrosine in sticks. The two domains SH2 and SH3 involved in the regulation of Src and other kinases are shown in purple and orange, respectively. The long flexible (and functionally relevant) linker between SH2 and the catalytic domain is shown in blue.
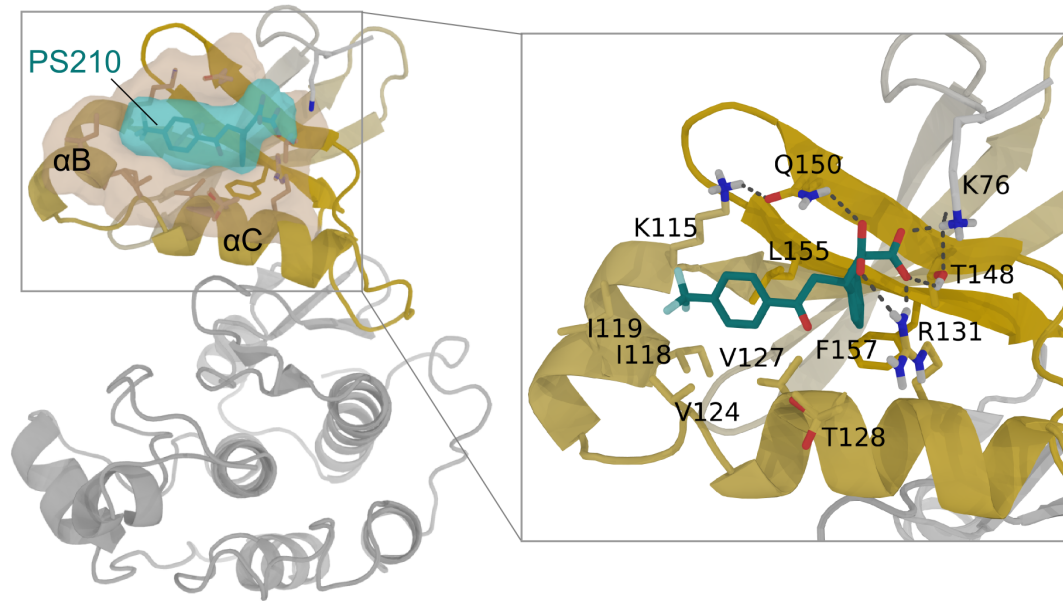
Figure 11. (left) the structure of the AGC kinase PDK1 bound to the allosteric activator PS210 (in cyan). PS210 is bound in the PIF pocket, a pocket formed by the αB and αC helices. (right) Residues forming the PIF pocket and interacting with PS210. Hydrogen bonds and polar contacts are shown with dotted lines.
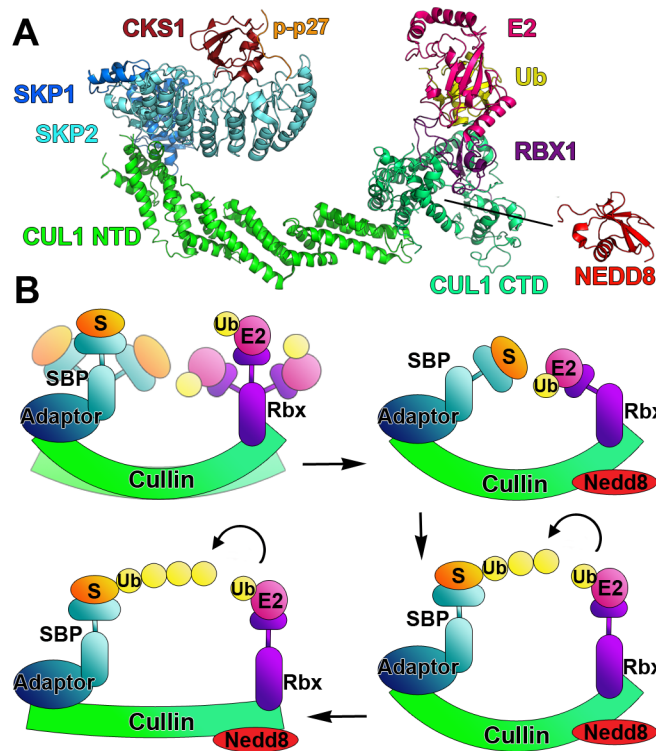
Figure 12. Cullin RING E3 ligases (CRLs) are multisubunit proteins composed by different proteins: Cullin (CUL), adaptor protein, substrate binding protein (SBP) and RING-box (RBX) proteins that binds the E2-Ubiquitin complex. We here shown a model of an assembled CLR with an E2-ubiquitin intermediate (A). The model is composed by superimposing the crystallographic structures of CUL1–RBX1–SKP1–SKP2$_{F-box}$, SKP1–SKP2–CKS1–p27 phosphopeptide and RING–UBCH5–E2 (PDB 1LDK, 2AST and 4AP4 respectively). The panel A shows the cullin N-terminal domain (green) and C-terminal domain (greencyan), adaptor protein SKP1 (azure), SKP2 (cyan), CKS1 (dark red), p27 phosphopeptide (orange), RBX1 (purple), E2 (magenta), ubiquitin (yellow), NEDD8 (red). NEDD8 binds to the cullins and allosterically activates CRL. B) Proposed schematic model for the CRL mechanism and the role of hinge region linker in cullins. S indicate the substrate, Ub ubiquitin, Upper left CRL is a flexible two-arm machine and both cullins, Rbx, SBP

are flexible and can populate wide range of different conformations. Different shifts into the populations of the ensemble of conformations available to CRLs promote conformations suitable to initiate the ubiquitination of the substrate from the E2 ubiquitin complex (Upper right), chain elongation (bottom right) and formation of longer ubiquitin chain thanks to population shifts in the cullins.[241]
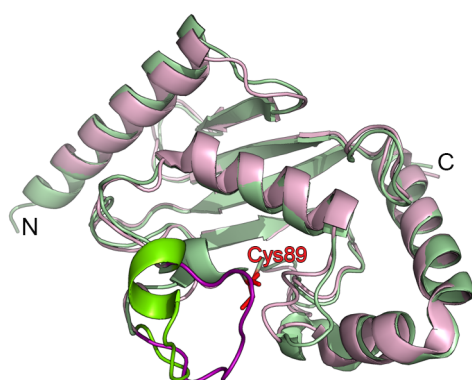
Figure 13. Open (green) and closed (magenta) states of the acidic loop of Cdc34-like E2 enzymes. The loop conformation can regulate the accessibility of the catalytic site (in red) and open states might be observed in the free state of the wild type enzyme but they are further enhanced by phosphorylation.
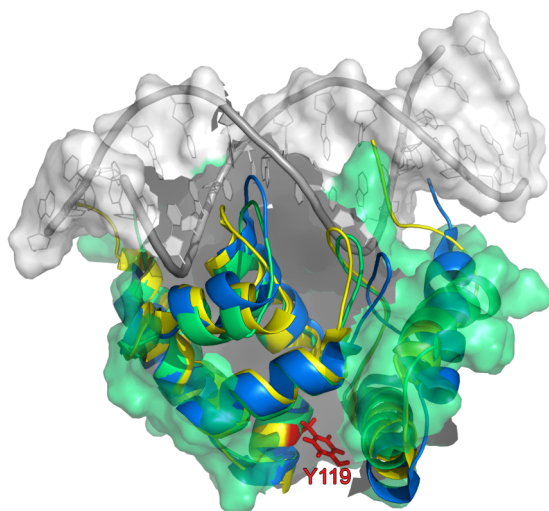
Figure 14. In ARID DNA-binding domains (showed as cartoons of different colors), the DNA binding loops are long-range coupled to a tyrosine residue (red stick) that is located in a potential region for cofactor binding. DNA is shown as a reference to orient the view.

TOC Graphics