ACS Chemical Neuroscience

University of Sussex

Subscriber access provided by University of Sussex Library

Article

Understanding the Differential Selectivity of Arrestins toward the Phosphorylation State of the Receptor

Ozge Sensoy, Irina Sousa Moreira, and Giulia Morra

ACS Chem. Neurosci., **Just Accepted Manuscript** • DOI: 10.1021/acschemneuro.6b00073 • Publication Date (Web): 12 Jul 2016

ACS Publications

ACS Chemical Neuroscience is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Understanding the Differential Selectivity of Arrestins toward the Phosphorylation State of the Receptor

Ozge Sensoy[1,2], Irina S. Moreira[3.4], Giulia Morra[2,5*]

1 Istanbul Medipol University, The School of Engineering and Natural Sciences, 34810, Istanbul, Turkey

2 Weill-Cornell Medical College, Department of Physiology and Biophysics, 1300 York Ave, New York, NY 10065

3 CNC - Center for Neuroscience and Cell Biology; Rua Larga, FMUC, Polo I, 1ºandar, Universidade de Coimbra, 3004-517, Coimbra, Portugal.

4 Bijvoet Center for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, Utrecht, 3584CH, The Netherlands

5 ICRM-CNR Istituto di Chimica del Riconoscimento Molecolare, Consiglio Nazionale delle Ricerche, Via Mario Bianco 9, 20131 Milano, Italia

ABSTRACT

Proteins in the arrestin family exhibit a conserved structural fold that nevertheless allows for significant differences in their selectivity for G-protein-coupled-receptors (GPCRs) and their phosphorylation states. To reveal the mechanism of activation that prepares arrestin for selective interaction with GPCRs, and to understand the basis for these differences, we used unbiased Molecular Dynamics simulations to compare the structural and dynamic properties of wild type Arr1 (Arr1-WT), Arr3 (Arr3-WT), and a constitutively active Arr1 mutant, Arr1-R175E, characterized by a perturbation of the phosphate recognition region called “polar core”. We find that in our simulations the mutant evolves towards a conformation that resembles the known pre-activated structures of an Arr1 splice-variant, and the structurally similar phosphopeptide-bound Arr2-WT, while this does not happen for Arr1-WT. Hence we propose an activation allosteric mechanism connecting the perturbation of the polar core to a global conformational change, including the relative reorientation of N and C domains, and the emergence of electrostatic properties of putative binding surfaces. The underlying local structural changes are interpreted as markers of the evolution of an arrestin structure towards an active-like conformation. Similar

activation related changes occur in Arr3-WT in the absence of any perturbation of the polar core, suggesting that this system could spontaneously visit pre-activated states in solution.
This hypothesis is proposed to explain the lower selectivity of Arr3 towards non-phosphorylated receptors. Moreover, by elucidating the allosteric mechanism underlying activation we identify functionally critical regions on arrestin structure that can be targeted with drugs or chemical tools for functional modulation.

KEYWORDS:

Arrestin/GPCR coupling; functional selectivity; arrestin pre-activated state; Molecular Dynamics simulations

INTRODUCTION

The arrestins are a family of four proteins involved in signaling mechanisms of G protein coupled receptors (GPCRs). They mediate GPCR desensitization and internalization processes that terminate signaling by coupling to phosphorylated and activated GPCRs [1] [2]. Moreover, a recent surge of data has revealed that arrestins also have a complementary action as scaffolding units to initiate G-protein independent signaling pathways[3-5] involved in many physiological as well as pathological processes in cardiovascular system [6] and brain functions [7]. Consequently, these proteins have become promising targets in treatment of crucial diseases such as Alzheimer [8] and Parkinson [9]. Importantly the activation of G-protein independent signaling pathways can be selected by specific GPCR agonists that stabilize receptor-arrestin versus receptor-G protein complexes [10]. More generally the availability of multiple signaling pathways is a critical aspect of GPCR directed drug design, as it may bring about unwanted side effects depending on the type of the receptor [11]. In this perspective, the rising interest in understanding the interaction mechanism of a receptor with a certain signaling effector –either G-protein or arrestin- is part of the continuing effort to modulate GPCR-mediated pharmacology by blocking the signaling pathway that cause harmful effects, hence paving the way for developing safer drugs. To this end, it is fundamental to illuminate the molecular basis of arrestin function by elucidating the structural determinants underlying the formation of a

complex with a GPCR and, in particular, the activation mechanism, which prepares each member within the arrestin family for its selective interaction with a cognate (non)-phosphorylated receptor.

The arrestins share a high degree of sequence similarity (identity over 75%) and a conserved fold composed of two β-sheet sandwich domains (N- and C-) connected by a hinge region (Figure 1). A proposed mechanism for the formation of a high affinity complex between arrestin and a phosphorylated-activated receptor [12-15], supported by recent structural evidence [16], suggests a "two-step" process involving two different binding sites, namely the N domain for phosphate recognition, and the N-C region for high affinity binding of the activated receptor. According to this model, arrestin first interacts with the phosphorylated sites of the receptor, which causes the release of structural constraints responsible for stabilizing the inactive conformation, and then undergoes a change that exposes structural elements engaged in a high-affinity recognition. Among these constraints, the *polar core* and the *triple element* - both located in the N-domain – have been described in the literature (Figure 1). The *polar core* is composed of five charged residues that are closely packed within the interior of the protein, and the *triple element* is formed by a set of hydrophobic residues located on the first and the last β−strand of the protein and includes as well an α-helix located on the back of the N-terminal fold [14, 15] (See Figure 1 for corresponding residues). In the inactive conformation, both the *polar core* and the *triple element* participate in stabilizing the C-terminal tail of the protein that is folded back on the N-domain. Interaction with the phosphorylated receptor is thought to cause the release of the C-terminal tail, by disrupting the intricate charge balance within the *polar core.* Indeed, destabilization of *triple element* or *polar core* regions by mutagenesis produces constitutively active phenotypes of the arrestins that do not require the receptor to be phosphorylated either for activation or the formation of the high-affinity arrestin-receptor complexes [14, 17].

The crystal structures of Arr-receptor complexes [16, 18] show that the N-C inter-domain arrangement of the arrestin molecule is overall conserved upon complex formation, without the clamshell like motions of the two domains that were originally hypothesized [19]. The conservation of arrestin structure upon activation is also supported by DEER measurements [20, 21] and by crystal structures of pre-activated arrestin molecules, namely of the splice variant of Arr1, p44 (PDBID: 4J2Q) [22] and the phosphopeptide-bound Arr2 (PDBID:4JQI) [23]. The most notable structural rearrangement in these structures is the relative rotation of the C-domain with respect to the N-domain by approximately 20 degrees [22]. Other local rearrangements observed in the pre-activated and in the

complex-engaged structures involve the inter-domain surface loops, as originally hypothesized by a scanning mutagenesis study of Arr1 [24]. These are the *finger loop* [18 25], the *C-loop*, the *lariat loop*-including the *gate loop* region, and the *middle loop*, all related to the binding interface (See Figure 1 for a detailed representation of the regions).

Together, these findings suggest that the GPCR segment that is phosphorylated upon activation is the essential trigger for the transition of arrestin to the active state, and activation amounts to exposing the interface loops to engage in binding the receptor, through a small relative rotation of the C- and N-domains. Although this is proposed as a general mechanism, evidence shows that the binding affinity of different arrestin subtypes for GPCRs varies not only depending on the receptor type but also on the receptor phosphorylation state [5]. In particular, whereas Arr1 has high affinity only for phosphorylated-activated rhodopsin, Arr3 interacts with several different GPCRs and exhibits much lower preference for the phosphorylated receptor over the non-phosphorylated one [17 26-28]. The difference in selectivity toward the phosphorylation state of the GPCRs might be related to a modulation of the activation mechanism, possibly sustained by a structural difference, as suggested by crystallographic evidence [28]: In Arr3, strand XIV (residues 256-262) deviates from the ordered β−sheet observed in the other subtypes (highlighted in brown in the right panel of Figure 1) and this region is proven to modulate the selectivity of Arr3 towards non phosphorylated receptors[28]. Moreover, Kim et al. [22] recently proposed that a weakened hydrogen bond network in the vicinity of the N-C domain interface might favor interaction of Arr3 and Arr2 with unphosphorylated receptors, in contrast to Arr1. Along the same line, in Arr3 the C terminal tail can populate multiple conformations upon GPCR interaction and this deviates from the behavior of the other subtypes [21 5, 29-31].

To address the question on the differences in the activation mechanism of arrestin subtypes and propose a dynamic model at the molecular level, we have carried out comparative studies of the two members of the family that seem to exhibit the largest differences in terms of phosphorylation selectivity. Starting from molecular dynamics (MD) simulations of wild type Arr1 (Arr1-WT) and Arr3 (Arr3-WT), we have evaluated functionally relevant structural and dynamic properties guided by the activation related changes observed in the crystal structures of pre-activated p44 [22], phosphopeptide-bound Arr2 [23] and rhodopsin-bound Arr1[18]. In addition, we have modeled and simulated the Arr1-R175E mutant, known to be constitutively active, starting from the Arr1-WT crystal structure [26], and compared it to Arr1-WT and Arr3-WT to detect rearrangements compatible

with activation, under the hypothesis that the R175E mutation can induce an activated state in Arr1 that is similar to the physiological receptor-activated state. In our MD simulations we found structural elements that discriminate an inactive state (populated in our Arr1-WT simulation) from a pre-activated state (achieved in Arr1-R175E), including regions of the molecules that had not been considered before as determinants for activation, and used them to characterize the progression of arrestin constructs towards active-like conformations. Strikingly, some of the same structural changes are observed in the Arr3-WT simulations in the absence of any perturbation. The model emerging from the MD analysis proposes that the activation of Arr1-R175E is indeed connected to the perturbation of the *polar core* [14]. In contrast, we do not observe changes in the *polar core* region of Arr3-WT, in spite of the presence of the other activation-related changes. We propose that this difference might be related to the markedly reduced selectivity of Arr3-WT, relative to Arr1-WT, towards phosphorylated GPCRs when forming a high-affinity complex.

RESULTS AND DISCUSSION

Comparative studies of molecular dynamics were carried out as described in Methods for three arrestin constructs: Arr1-WT, Arr1-R175E, and Arr3-WT. For each of them, two replicas of unbiased MD simulations (at least 500 ns each) were produced. The resulting trajectories were analyzed as described in Methods to identify and compare the molecular rearrangements occurring in each construct.

A key descriptor of the rearrangements in the molecular structures of the simulated constructs is the relative mobility of C-domain and N-domain, as a rigid rotation of the C-domain by about 20 degrees around an axis intersecting the domain interface is the most prominent structural transition revealed by the recent crystal structures of pre-activated arrestins [18, 22, 23]. If this rearrangement is required for activation, we hypothesized that the relative motion of the domains might be prevented by a particular structural constraint that needs to be released for the transition to occur. Therefore, we investigated the dynamics of the N- and C-domains in Arr1-WT and Arr1-R175E for signs of any differences in the inter-domain mobility resulting from the activating mutation, and used the same analysis of the Arr3-WT trajectories.

1. The rearrangement of the C-domain in the transition towards an active-like conformation

1a. Comparison of Rigid-body Mobility

The average internal dynamics of the domains and their relative rigid-like mobility were first analyzed by considering the distance fluctuations of pairs of Cα atoms (see Methods, section 4). The distance fluctuations matrices for Arr1-WT, Arr1-R175E, and Arr3-WT are shown in Figures 2A, 2B, 2C, respectively, with a color code that ranges from blue (the lowest values, which represent the highest coordination and the lowest relative mobility), to red (the highest values, representing the highest mobility). Each of the three arrestin constructs exhibits two matrix blocks characterized by low values, corresponding, respectively, to the N-domain (residues 9-180 in Arr1-WT and 6-175 in Arr3-WT), and the C-domain (residues 181-386 in Arr1-WT and 176-408 in Arr3-WT). Within each rigid block, single red stripes of high fluctuation values indicate locally mobile regions such as loop tips moving with respect to the rest of the protein. The linker connecting N- and C-domains, around residue 180-200, is coordinated to both blocks in Arr1-WT. Correspondingly the inter-domain distance fluctuations are comparable to the intra-domain ones, suggesting an overall rigidity of the system as a whole (a set of high-fluctuation values for residues 157-162 indicates a locally high mobility of an N-domain loop far away from the domain interface). In comparison to Arr1-WT, Arr1-R175E (Figure 2B) exhibits a less coordinated linker region and higher inter-domain decoupling and mobility, visible as higher distance fluctuations between the two domain blocks (compare dashed rectangles in Figure 2A, B). This is paralleled by an increase of mobility of C-domain interface residues (285-320, including the *C-loop*, the *lariat loop-gate loop*), and in the N-domain (residues around 70, *finger loop*). Notably, increased mobility is also found at the C-terminal tail.

In Arr3-WT, the distance fluctuations between the N- and C-domain and in the inter-domain linker (Fig. 2C) suggest a very similar pattern to what is found in Arr1-R175E. Interestingly, besides the mobility peaks at the interface residues 280-315 (see for comparison the corresponding region in Arr1-R175E), the core of the C domain (around residue 260) appears to be less rigid, whereas the terminal residues of the C-terminal tail are coordinated to the N domain, despite the strong mobility of the preceding loop. Overall, the distance fluctuation analysis highlights more significant motion of the C-domain relative to the N-domain in both Arr1-R175E and Arr3-WT, than in Arr1-WT.

This higher inter-domain mobility appears to be coupled in both constructs to increased dynamics of selected loop regions facing the domain interface, and to an increased internal flexibility of the C-domain in the case of Arr3-WT only.

1b. Rotation of the C-domain

The rotational motion of the C-domain with respect to the N-domain, around an axis defined for each construct as described in Methods (see Rotation analysis) was monitored along the trajectory. The time-dependent rotation angles of three reference residues around the axis, relative to the starting structure, are shown in Figure 2 (panels E-G) for all simulated systems. An increase in the angle value corresponds to a clockwise rotation around the axis, if observed from the C-domain towards the N-domain (see arrow direction in Figure 2D). In Arr1-WT, the angle with respect to the starting structure for residues P212, S346 and A200 fluctuates around zero (Figure 2E), indicating no net rotation along the trajectory. In contrast, in the Arr1-R175E construct, where inter-domain mobility is higher than in Arr1-WT, the net rotation angle of the reference residues along the same axis shows a steadily increase by more than 10^0 for all residues, starting at ~170 ns (Figure 2F and SI-1E). A further increase is then observed at ~400ns. In Arr3-WT, on the other hand, the corresponding residues S331, S197 and E186 report a net increase of the rotation angle relative to the starting structure. Instead of a monotonic increase, this profile shows however a strongly fluctuating behavior between high and low values. The absence of net rotation for Arr1-WT, the increase of rotation angle for Arr1-R175E and the dynamic fluctuation between high and low values in Arr3-WT is also observed in the replicas (see SI-1), although in the latter system the rotation angle is lower in the second trajectory, suggesting that Arr3-WT can sample wider portions of the conformational space. As a consequence, in Arr3-WT, the distribution of rotational states is markedly shifted towards positive values relative to Arr1-WT in one replica, whereas the non-rotated state is sampled in the other replica (see SI-1G). This leads us to conclude that the mobility of Arr3-WT is higher than that of Arr1 constructs, and that the two inter-domain arrangements - rotated and not rotated- might be populated under the same conditions. This finding agrees with the experimental evidence of a higher flexibility of Arr3 relative to the other subtypes [28].

2. Local structural changes underlying the observed inter-domain rotations

The relation between the domain rearrangements observed in the MD trajectories and local structural changes was investigated, focusing on a series of structural elements that have been proposed to play a mechanistic role in activation and formation of complexes with GPCRs, such as the *C-loop*, the *finger loop*, the *lariat loop* –including the *gate loop* region identified in Figure 1. Other protein regions, namely the *aromatic core* and *short-helix* were identified here as new relevant elements in this class.

2.a The C-loop rearrangement exposes the residue important for interaction with the receptor

The *C-loop*, consisting of residues 249-254 in Arr1-WT (244-249 in Arr3-WT), is located in the crest between the N- and C-domains, facing the *middle loop* and the *finger loop* (see Figure 1, yellow, green and red color, respectively). Comparison of the crystal structures reveals that in the inactive state it is in contact with the *finger* and *middle loop* residues, but in the pre-activated state it is more exposed, as shown by the solvent accessible surface area (SASA) of the representative residue of the loop, Y250 (F245 in Arr3-WT)(see Table I). Also, it is involved in the interface with ICL2 in the rhodopsin-Arr1 complex [18]. The value of the %SASA relative to the standard value for the residue for Y250 (F245) was calculated from the MD trajectories of the different constructs, and compared to the reference values in the crystal structures of inactive and pre-activated arrestin (see Table I for average %SASA of the representative residue Y250/F245). In agreement with experiments, the *C-loop* is more exposed in both Arr1-R175E and Arr3-WT than in Arr1-WT (see Table I).

2.b An aromatic core region modulates the dynamics of the finger loop

The flexible *finger loop* (see Figure 1) adopts two conformations in the inactive state of the protein as evidenced by the crystal structure of Arr1-WT (PDB-ID 3UGX[36]). However, this heterogeneity disappears upon activation and the loop primarily adopts an upward position, where the β-strands at the tips of the loop are elongated[22,23]. This conformation is stabilized by a network of H-bond interactions formed between *middle* (Q137 and K141) and *finger loop (*E70) residues, and in the rhodopsin-Arr1 complex it interacts with TM7 and the N terminus of helix 8 of the receptor [18]. However, Q137 is replaced by E135 in Arr3-WT and the H-bond network seen in Arr1-WT is disrupted. In addition, A135 in Arr1-WT is replaced by G133 in Arr3-WT, which further increases

the flexibility to the loop. The *finger loop* arrangement was monitored in the MD trajectories by measuring the distance between the Cα atoms of one representative residue (D73 in Arr1-WT and D70 in Arr3-WT) and one reference residue on the N-domain, namely I173 in Arr1-WT and V168 in Arr3-WT, in analogy to ref. [20]. Indeed, in Arr3-WT, this distance transiently takes values close to the crystal structures of pre-activated proteins, whereas in Arr1-R175E it stays closer to the values of the inactive state, and more so in Arr1-WT (Figure SI-2B). The simulations allow us to get insight into the mechanistic determinants of the *finger loop* arrangement, and identify a conserved hydrophobic region, composed of aromatic residues located at the ends of the loop: F65, F67 and F79 (F62, F64 and F76 in Arr3-WT), and the tip of the *C-loop*, Y250 (F245 in Arr3-WT). These four residues, termed here the *aromatic core*, form a packed cluster in the structure of the inactive form that is disrupted in the pre-activated structures [23, 32]. In our MD trajectories we measured the distance between the Cα atoms of F79 and Y250 in Arr1-WT (F76-Y245 in Arr3-WT), to report on the evolution of the *aromatic core*. The increase of this distance, upon core disruption, is indeed coupled to the displacement of the *finger loop* from the N-domain in Arr3-WT (see SI-2C). Interestingly, the *aromatic core* is also stably disrupted in Arr1-R175E yet the *finger loop* is only occasionally displaced from the bent orientation toward N-domain.

2.c The short helix near the aromatic core unfolds

Another region that shows structural difference between inactive and pre-activated crystal structures is the short helical turn, here referred to as *short-helix* (cyan in Figure 3) comprising residues 318-322 in Arr1-WT (313-317 in Arr3-WT) and positioned near the *aromatic core*. In the inactive state these residues adopt a helical conformation stabilized by both *i-i+4* backbone hydrogen bonding interaction at K318 (K313 in Arr3-WT) and L322 (M317 in Arr3-WT), as well as side chain-side chain interactions between K318 and the D246 (N241 in Ar3-WT) located on the β-strand XI of the protein. These two stabilizing interactions appear to be disrupted upon activation, as the *short-helix* is found in the unfolded state in crystal structures of pre-activated arrestins [22, 23]. In the MD trajectories we monitored the unraveling of helix through the change both in %sasa of K318 (K313 in Arr3-WT, see Table 1) as well as the distance between the Cα atoms of the N and the C terminal residues of the short helix (see SI-3) whose exposure and distance increase upon

unfolding, respectively. The helix unfolds in both Arr3-WT replicas and in one of the Arr1-R175E trajectories, but not in Arr1-WT, where it maintains a helical conformation during the course of the simulation. Interestingly, K318 has been shown to be important for Arr1-rhodopsin interaction [24] and it constitutes part of the binding interface in the crystal structure of the complex[18].

2.d The polar core region and the C-terminal tail rearrange in Arr1-R175E but not in Arr3-WT

The rearrangement of the *gate loop* is another change revealed in the crystal structures of pre-activated Arr. This region, known as the functional part of the previously described *lariat loop* (see Figure 1), consists of residues D296-N305 in Arr1-WT and D291-300 in Arr3-WT. In the inactive form, two of the *gate loop* residues, D296 and D303, (D291 and D298 in Arr3-WT), participate in the *polar core* along with D30 and R382 (D27 and R393 in Arr3-WT) by interacting with residue, R175 (R170 in Arr3-WT)[14, 15, 33] (see Figure 4). Even in the absence of this interaction, the loop orientation is maintained in the crystal structure of the R175E mutant [34]. In contrast, in pre-activated crystal structures, D303 (D298 in Arr3-WT) is flipped away from the *polar core* towards the N-terminus of the *lariat loop*, named here *ARG loop*, and interacts with the conserved R288 residue (R283 in Arr3-WT). In contrast to the R175E crystal structure, but in agreement with the pre-activated conformations, D303 leaves the *polar core* in the MD trajectories of Arr1-R175E, due to the repulsive interaction upon charge reversal at residue 175 (see Figure 4). Hence, the distance between D303 and R288 decreases in Arr1-R175E (~ 1.1 nm) relative to Arr1-WT (~1.5 nm), where D303 fluctuates around the starting position. As a consequence, in Arr1-R175E the *polar core* is more exposed to the solvent than in Arr1-WT, as shown by the large changes in %SASA for R175 in the constitutively active mutant (see Table 1).

Notably, in Arr3-WT trajectory the *polar core* is fairly unperturbed. Here, the D298-R283 distance increases in both Arr3-WT trajectories, as D298 (corresponding to D303 in Arr1-WT) maintains its interaction with residue, R170, while the R283 residue (corresponding to R288 in Arr1-WT) moves towards the rotating C-domain (see Figure 4). Importantly, this difference reflects the position of the C-terminal tail in the two subtype constructs. While the interaction between the C-terminal β-strand XX and the N-terminal β-strand I is on average maintained in all systems, the unstructured tail of the C-terminus (residues 382-386 in Arr1-WT, and 404-408 in Arr3-WT) is completely detached from the *polar core* in both replicas of the Arr1-R175E simulation, at around 400 and

250ns respectively (Fig4 and SI 3), while this does not occur in Arr3-WT even during the course of an extended 1 μs trajectory.

3. A model for the activation mechanism of arrestin

3.a Activation Related Changes (ARCs) define steps in the activation process

The rearrangements involving the structural motifs that emerge from the comparison of the inactive and pre-activated crystal structures of Arr1 and 2 are reproduced in our simulations, where they are also accompanied by other changes. We therefore can assess the progress in the activation-related conformational transitions by monitoring a set of Activation Related Changes (ARCs) (see Table 1). They are: first, the degree of exposure of the *C-loop* (buried vs. exposed – classified according to the value of %SASA of Y250/F245 in comparison to inactive vs. pre-activated crystal structures), the latter related to the state of the *aromatic core*; second, the conformational state of the *short helix* (folded vs. unfolded); third, state of the C-terminal tail (attached or displaced). By classifying the dynamic snapshots according to the occurrence of ARCs, we can distinguish between inactive- and pre-activated states: the inactive-like state is identified by a closed *C-loop* and a compact *aromatic core*, as well as a folded *short helix* and an attached C-terminal tail. In contrast, the pre-activated state is identified by an exposed *C-loop* and *aromatic core*, an unfolded *short helix,* and/or a displaced C-terminal tail, the latter identified by the exposure of residue R175 (R170 in Arr3-WT). The MD trajectories were then parsed according to the ARCs to differentiate between inactive and pre-activated ensembles at several time intervals. For Arr1-R175E, the inactive ensemble corresponds to interval (0, 150 ns (0,200 ns for the second replica), and the pre-activated one (150 ns-onward (200 ns-onward for the second replica). Since Arr1-WT does not exhibit any stable ARCs, the entire trajectories correspond to the inactive ensemble. In contrast, in Arr3-WT, ARCs occur at 150ns and 50ns in first and second replica respectively (see SI-3), hence the activated ensemble is defined for the intervals (150,500ns) and (50,500 ns), respectively.

The rotation angle of residue S346 (S331 in Arr3-WT) was evaluated on the inactive and pre-activated ensembles, to define the rotational state of the system. On average, the rotational state

differs between the ensembles, confirming that the rotated C-domain is indeed connected to the presence of the ARCs.

C-domain rotation is essential to adopting an activated conformation capable of recognizing the receptor through exposure of surface loops. More in detail, we observe that in Arr3-WT the *short helix* unfolding, following the *C-loop* detachment, is concomitant to the net increase of C-domain rotation (compare Figure 2G, SI-1 and SI-3). On the other hand, in Arr1-R175E the C-domain rotation is significantly enhanced after the C-terminal tail displacement in both replicas (compare Figure 2F, SI-1 and SI-3)**.** Yet, only upon helix unfolding the rotation increase becomes monotonic. Therefore, the unraveling of the *short helix* emerges as plausible structural restraint that stabilizes the inactive conformation of the protein. Under physiological conditions, the tail displacement would be the consequence of the interaction of arrestin with phosphorylated receptor[20]. According to our findings, this local perturbation would in turn trigger C domain rotation and as a consequence induce the high affinity conformation.

The average ARCs values sampled in the pre-activated and inactive ensembles in our MD simulations are shown in Table I and are overall consistent with those measured in inactive and pre-activated crystal structures (see Table 1 for comparison). In particular, the difference between Arr1-WT and Arr1-R175E suggests that we observe progress towards activation in the constitutively active mutant. Moreover, the occurrence of similar structural changes in Arr1-R175E and Arr3-WT that are not observed in Arr1-WT, points to a higher propensity toward activation in Arr3-WT, possibly sustained by a different mechanism. In particular, the configuration of the ARCs, as well as the rotation of the C-domain are similar in Arr1-R175E and Arr3-WT, but the *polar core*, monitored by the SASAs value of R170 (see Table I) remains virtually unperturbed in Arr3-WT. Correspondingly, the C-terminal tail position reflects the perturbation of the *polar core*, as it is displaced in both Arr1-R175E simulations but not in Arr3-WT.

3.b Correlation among Activation Related Changes (ARCs)

To shed light on the correlation among the ARCs that may result in C-domain rotation and highlight mechanistic differences between the systems, Essential Dynamics analysis was used (see Methods) to describe the dominant collective motions occurring in the trajectories. Inspection of the residue-

based fluctuation components of the first two principal component analysis (PCA) eigenvectors of the three constructs, accounting for about 50% of the overall motion, highlights main differences in their flexibility. In the first PC of Arr1-R175E, the amplitude of collective motions involving the *gate* and the *ARG loop* (see section 2.d), as well as of *finger loop* (see section 2.a), is increased with respect to Arr1-WT. The mobility of the *helix* and *C-loop* (cyan and yellow, respectively) are enhanced in the second eigenvector (Figure 5A). The two extreme structures obtained by projecting the trajectory of Arr1-R175E along first and second eigenvector (Figure 5B) highlight the dominant motion of the *gate loop* and the *ARG loop* and on the other hand the unfolding of the *short helix* and the exposure of the *C-loop*. This leads us to hypothesize that the motion of the *gate loop* towards the C domain, upon perturbation of the *polar core,* is the main event together with the displacement of the *C-loop* and unfolding of the *short-helix*. The propagation of the perturbation between *gate loop* and *short helix* might be further facilitated by their physical connection through the linker across the N-C interface (residues 310-317). This hypothesis is supported by the evolution in time of the RMSD of the linker (SI-4), which increases immediately before the unfolding of the helix.

In Arr3-WT, the main contribution to the mobility expressed in the PCA is due to the long disordered C terminal loop (residues 350-394), while other peaks are observed at the *C-loop*, the *ARG loop*, *short helix* and *finger loop*. In contrast to Arr1-R175E, less motion is observed at the *gate loop* as expected. A significant part of the C domain, including the so called "distorted beta strand" comprising residues 256-262 in Arr3-WT and sequentially preceding the *C-loop*, shows high mobility in the first eigenvector, in contrast to Arr1-R175E (Figure 5A and 5C, brown). The coupling between this motion and the detachment of the *C-loop* is suggested in the extreme projections of the MD trajectory along the second essential eigenvector (see Figure 5D). In agreement with that, the timeline plot shows an RMSD transient increase in the distorted strand region right before the helix unfolds (compare Figure 2G and SI-4).

A significant peak is also found in *helix 1*, namely the α-helix that participates in the *triple element* region located on the N-domain (residues 101-109), which is also modulated in Arr1-R175E relative to Arr1-WT (grey, first eigenvector profile). Interestingly, an activation-increased flexibility of this region, which is connected to the *finger loop* and the *middle loop* at its N- and the C-terminal end, respectively, has been observed in a number of recent studies [16, 31].

The similarity in the nature of conformational changes and the ARCs involved in them, underscored by the differences observed in the behavior of some structural elements, suggests that while Arr1-

R175E and of Arr3-WT achieve similar pre-activated states in the simulations, the mechanisms can be different. In particular, the perturbation of the *polar core* appears to be essential to trigger the *gate loop* displacement in Arr1-R175E, thereby unfolding the *short helix* and releasing the structural constraints that lead to activation, including the C-terminal tail displacement. In contrast, the intrinsic flexibility of the distorted beta-strand in Arr3-WT appears to be sufficient to increase the mobility of the C domain, and lead to the exposure of the *C-loop*, disruption of the *aromatic core*, unfolding of the *short helix*, and eventually the domain rotation. Consistent with that, we do not observe exposure of the *polar core* in our Arr3-WT simulations, nor C-terminal tail displacement, even in the presence of the other ARCs. This difference between Arr1-WT and Arr3-WT might be connected to the different ability of Arr3 and Arr1 to interact with GPCRs in the absence of receptor phosphorylation.

Table 1: Comparison of ARCs in the simulated systems and the corresponding crystal structures. Standard errors are also reported.

Constructs	C-loop[a] <%SASA>		Short helix[b] <%SASA>		Polar core exposure[c] <%SASA>		<angle>[d] (deg)	
	I	A	I	A	I	A	I	A
p44 Arr1[e]	31.0	63.4	50.0	62.0	32.0	35.0	-	12.2
Arr2[e]	37.1	69.8	49.3	67.4	7.0	32.0	-	15.1
Arr1-R175E[e]	28.9		50.9		44.0		-	
Arr1-WT	43.4±4.1		56.1±0.2		8±2		-1.1±0.2	
Arr1-R175E	37.3±7.1	61.3±1.3	52.2±1.2	69.4±1.2	19.1±1.1	38.1±1	0.1±0.4	11.3±0.8
Arr3-WT	39.9±6.5	64.3±7.9	53.7±4.5	66.3±1.6	12.3±1	11.5±1.2	4.1 ±0.4	7.9±0.3

[a]The exposure of the *TYR-loop* is measured as the percent SASA of the representative TYR residue (Y250 in Arr1-WT, F244 in Arr2-WT, and F245 in Arr3-WT). [b]The state of the *helix* is given by average percent SASA of K318 in Arr1-WT/Arr1-R175E, K313 in Arr3-WT, and R312 in Arr2. [c]The state of the polar core, also indirectly reporting on the C-terminal tail position, is given by average percent SASA of residue R175 in Arr1-WT/Arr1-R175E, R170 in Arr3-WT and R169 in Arr2. [d]As shown in Fig.1, rotation angles around the axis defined in Methods are given as averages

sampled in the active-like ensemble for S331 (Arr3), S346 (Arr1), S330 (Arr2). I=Inactive, A=Active. [e]PDB ID-s for inactive and active splice variant Arr1 are 3UGU [35] and 4J2Q[22], respectively, for inactive and active Arr2 are 1JSY[36] and 4JQI[23], respectively and for Arr1-R175E is 4ZRG[34].

4. Surface properties of the pre-activated states.

The electrostatic potential over the surface of all three simulated constructs was calculated and compared to the crystal structures (Figure 6). In panel A the difference of electrostatic potential between pre-activated and inactive Arr1 PDB constructs is shown, and in panels B and C the difference of the average electrostatic potential between the pre-activated and inactive ensembles obtained from our MD simulations of Arr1-R175E/Arr1-WT and Arr3-WT, respectively (see previous section). In the case of Arr1-R175E, upon comparing the activated ensemble the mutant to the Arr1-WT ensemble, the bulk of N- and C-domains shows a more positive potential upon activation, possibly to facilitate the interaction with the negatively charged groups at the receptor tail. This is in agreement with the distribution of the electrostatic surface potential in the crystal structures of arrestins (Figure 6A). In contrast, the N-C interface region displays a different pattern with an increase of the negative electrostatic potential upon activation, particularly at the *finger loop*, which is one of the essential binding determinants to the receptor. Interestingly, the same behavior was found by Kang *et al*. in the recent crystal structure of Arr1-rhodopsin [18], which supports the hypothesis that the MD simulations are effectively sampling the pre-activated state. The increase of negative potential upon activation is especially important because of its complementarity to the increase of positive potential at the cytoplasmic surfaces of the receptors (same paper). In the case of Arr3-WT (Figure 6C) this modulation, obtained by considering the difference between active and inactive MD ensembles, is even more pronounced.

CONCLUDING REMARKS

The MD simulations of Arr1-WT, Arr1-R175E, and Arr3-WT constructs reported here provide for the first time a high-resolution structural and dynamical insight into local and global conformational changes occurring under activating conditions in Arr1 and for comparison in wild type Arr3. We hypothesize that those changes are associated with the transition from inactive to pre-activated states of the arrestins and propose a molecular mechanism of activation. In the context of structural changes revealed by comparisons of arrestin crystal structures in inactive and pre-activated states, the analysis of MD trajectories yields a set of interrelated discrete steps defining the transition (we termed here ARCs). This allows us to connect the observed structural rearrangements of Arr1-R175E, and Arr3-WT to known differences in modes and propensities for activation, from which a possible mechanism emerges at the molecular level.

According to our findings, activation can proceed when conformational changes are achieved along the N-C domain interface in local regions (identified in the *aromatic core* and in the *short helix*) that seem to prevent the C domain rotation acting as constraints. The proposed model identifies not only the set of interrelated discrete steps followed by the arrestins in the transition to the pre-activated state (the ARCs), but also suggests subtype specific differences in the ability to perform these transitions. According to our findings, in the case of Arr1 the emergence of the active-like structures mainly originates from the perturbation of the *polar core*, which in the wild type Arr1 is due to the interaction with the phosphorylated tail of the cognate GPCR and in the mutant R175E is mimicked by the charge reversal: this perturbation generates, together with the local C-terminal tail detachment, an allosteric relationship between the displacement of the *gate loop* and the rotation of the C-domain, which is associated with the exposure of the *C-loop*, On the other hand, Arr3-WT shows an intrinsic inter-domain flexibility that can spontaneously lead to a very similar set of ARCs and exposure of the same interface loops. The receptor binding interface proposed by the most recent structural evidence [16, 18] includes the *finger loop* on the N-domain side, in agreement with crystallographic data [25] and with the hydrogen exchange mapping in ref. [16]; on the C domain it involves the *C-loop* and the *short helix*. Remarkably, our results lead us to locate the main activation-related changes common to Arr1 and Arr3 in the same N-C interface region centered on the *finger* and *C-loop*. Moreover, the activation-related structural rearrangements are accompanied

by an increase of the positive electrostatic potential at this binding interface that is consistent with the potential calculated for the available X-ray structures.

In this respect, it is notable that while ARCs appear to be common to Arr1 and Arr3 (see Table I), the *gate loop* and the C-terminal tail show a markedly different behavior in the two systems, since in Arr3-WT, activation-related changes can occur without perturbation of the *polar core* and the concomitant displacement of tail. This observation suggests that the sequential character of the multistep binding process proposed for arrestins, namely binding of the phosphorylated tail of the receptor followed by high affinity binding to the N-C interface, might be differentially modulated depending on the subtype, and even that the involvement of the *polar core* might be less critical for Arr3. Indeed, whereas for Arr1-R175E, the connections between *polar core* perturbation and receptor binding clearly emerge from experiments, in the case of Arr3-WT, experimental data are less conclusive. Differently from Arr1, Arr3 can still bind to non-visual receptors when the phosphate sensing residues K10 and K11 are mutated to Alanine [5, 37], indicating that phosphate sensors and therefore the phosphorylation of the receptor, might play a different role in the formation of complexes.

The promiscuity of Arr3-WT and its reduced selectivity towards phosphorylated receptors have been hypothesized to depend upon an increased inter domain flexibility of this protein with respect to Arr1-WT, due to a less tight network of hydrogen bonds[22] and internal flexibility of the C domain, which might help in the formation of the interaction surface when binding to the GPCR [15, 38]. Our dynamic model is compatible with this structural hypothesis. Also, the other structural difference between Arr3 and the other subtypes highlighted in [28], namely the distorted beta strand that does not form a contiguous β -sheet in Arr3-WT, was shown to regulate the selectivity of Arr3, as swapping this strand between Arr3-WT and Arr2 resulted in reversing the selectivity properties of the two members [28] towards the phosphorylation state of the receptor. Along the same line, we observe an increased flexibility of this strand in Arr3 in comparison to Arr1. Moreover, by analyzing the collective dynamics of the protein we detect a correlation between the mobility of this strand on one hand and the fluctuation of the *C-loop* and the unfolding of the *short-helix* on the other hand, which is not present in Arr1-R175E. This leads us to identify the intrinsic flexibility of the distorted beta strand as an essential factor governing the release of the N-C constraints and the onset of C domain rotation in Arr3-WT, coupled to activation.

Due to the less inhibited character of activation related events we observed in the simulations of Arr3-WT, we propose that the pre-activated states of this arrestin subtype might be easily accessible in solution even in the absence of a phosphorylated receptor. Indeed, recent hydrogen-deuterium exchange experiments [31] highlight significant differences in the conformational dynamics of Arr2 and Arr3, which support this inference. These experiments showed that Arr2 with an activating mutation in the polar core exhibits dynamic perturbations of several regions on the C domain as well as on the N domain relative to the basal state, which is compatible with activation-related rearrangements. In contrast, the same mutation in Arr3 is not associated with significant changes in deuterium uptake compared to the basal state, except in the neighborhood of the mutation site, which is compatible with the hypothesis that the basal state might spontaneously adopt a similar pre-activated conformation.

In Arr1-R175E, the destabilization of the *polar core* by the mutation induces, besides the displacement of the *gate loop* and the exposure of the interface loops, also the detachment of the C-terminal tail, and this is reproduced in our simulations. In contrast, in Arr3-WT where the *polar core* is unperturbed, we do not observe any displacement of the tail, in spite of other dynamical modulations occurring in the same region, such as high flexibility at the N domain helix 1 in the nearby triple element region. We cannot predict whether the C-terminal tail displacement might occur at a later step in activation that is not captured in our simulations but we can speculate that Arr3 might bind to a receptor without displacing the tail whereas this transition might be not equally favorable for Arr1. On the other hand, one can assume that the C-terminal tail displacement will happen in Arr3 whenever the polar core is directly perturbed, in analogy to Arr1, when interacting with a phosphorylated receptor. Interestingly, a DEER spectroscopy study provided insight into the conformational distribution of the tail, in Arr3 upon binding a receptor, which turned out to include a population of arrangements rather close to the basal conformation [21]. Moreover, a very recent FRET paper focusing on the interaction between Arr3 and β2-AR [39] highlighted that the kinetics of the tail displacement upon receptor recruiting is slower than the complex formation. It is tempting to speculate that the structural modulation of the tail of Arr3 might occur when the protein interacts with receptors under different phosphorylation conditions, or possibly be involved in the activation of different cellular pathways and signaling events [40]. Along this line, phosphorylation has been suggested as a modulating factor governing the binding mode of Arr3 to β2-AR and has been

implicated in biased agonism [41]. Further investigations on Arr-receptor complexes will be required to address these questions.

METHODS

1. Molecular Constructs

The three arrestin constructs used in this study are Arr1-WT (PDBID: 1CF1, chain D [26]), Arr1-R175E, and Arr3-WT (PDBID: 3P2D, chain B [28]. Unresolved parts of the crystal structures were modeled with Swiss-Model [42, 43], and the loop regions were further refined with ModLoop [44, 45]. The R175 residue in Arr1-WT was mutated using Pymol [46], followed by MD equilibration (see below). The protonation states of residues in the physiological pH range were determined with the PROPKA methodology [47-49] that computes pKa values of ionizable residues by accounting for the effect of the protein environment.

2. Multiple Sequence alignment (MSA)

We performed mega Blast (all NCBI databases) using the Geneious Software version 5.4 [50] to search for Arr1-WT, 2 and 3 homologs, using a maximum e-value of 1^{-40} to sequences of similar length. The sequences were aligned using MUSCLE [51] with default parameters. We selected representative members among different species of vertebrates such as *Bos Taurus, Felis Catus, Homo Sapiens, Mus Musculus, Pan Troglodytes, Rattus Norvegicus and Xenopus Laevis*. The resulting alignment was used as an input in ConSurf Server [52] to calculate the conservation score for each amino-acid position.

3. MD Simulations:

All MD simulations, consisting of two independent replicas per system, were performed with the NAnoscale Molecular Dynamics (NAMD) Package[53] using the CHARMM27 force field with CMAP corrections [54], under constant pressure with Langevin Piston Period, and Langevin Piston Decay set to 200 fs and 50 fs, respectively. Constant temperature (310 K) was maintained with Langevin Dynamics. PME[55] was used to calculate long-range electrostatic contributions. Energy minimization was performed prior to MD production runs using the conjugate gradient algorithm, where the backbone atoms were initially fixed, and then harmonically constrained. Constraints were released gradually in four 100 ps-step MD simulations with decreasing force constants of 1, 0.5, 0.1 and 0.01 kcal/ (mol·Å^2), respectively. After the equilibration phase, the systems were simulated for at least 500 ns (we have 1 □s for one Arr1-wt and Arr3-wt replica, 500ns for the other simulations).

All MD simulations were performed with keeping all the bonds rigid using the SHAKE algorithm [56] with an integration-step of 2 fs. Outputs were saved every 20 ps. TIP3P was used to model water molecules in the system [57]. All the systems were neutralized in 0.10 M NaCl.

The stability of the systems was checked by calculating backbone root-mean-square-deviations (RMSD), with respect to the initial structures without including the highly mobile long loop region located at the opposite site of the concave receptor binding surface, whose C-terminus folds back on the N-domain. All subsequent analyses of the MD trajectories were carried out after removing the first 40 ns, when a relatively stable plateau was reached in the RMSD plots.

4. Analysis of rigid body mobility:

For each MD trajectory, the distance fluctuations map A is calculated over the trajectory as:

$$A_{ij} = \left\langle \left(d_{ij} - \left\langle d_{ij} \right\rangle \right)^2 \right\rangle \quad (1)$$

where d_{ij} is the (time-dependent) distance between the Cα atoms of amino acids i and j and the brackets indicate time-average over the simulation. Each matrix entry reports on fluctuation of the inter-residue distance in the corresponding residue pair. Lower distance fluctuation values correspond to a higher internal coordination between the residues (local rigidity). Matrix regions showing relatively low values identify protein sub-domains that move together (in coordination) while undergoing structural fluctuations [58].

5. Rotation Analysis:

All trajectories were aligned with respect to their initial conformation by superimposing their N-domains (i.e., residues 9-179 for Arr1-WT, and residues 6-176 for Arr3-WT). Rotation axes were defined in Arr1-WT by connecting the initial Cα atom coordinates of Cys138 and Ile321, and in Arr3-WT by connecting initial Cα atom coordinates of Cys126 and Ile306. These axes were chosen as representative directions of rotation after visual inspection of the superposition between inactive and pre-activated Arr2 proposed in ref[23].

These residue pairs coincide well in the superposition of the initial structures of the two constructs. The rotation angle values around the respective axes were calculated for residues E186, S197, S331

in Arr3-WT, and residues A200, P212 and S346 in Arr1-WT and Arr1-R175E along the MD trajectory, using an in-house built C code.

6. Binding Surface Properties

The electrostatic potential was calculated for the binding surfaces of the constructs. Continuum electrostatic potential calculations were carried out with DelPhi [59] using 2.5 grids/Å for scale, 0.001 kT/c for the convergence criterion; a 90% for the fill of the grid box; and the Coulombic method to set the potentials at the boundaries of the finite difference grid. The dielectric boundary was taken as the molecular surface defined by a 1.4 Å probe sphere for atoms with radii taken from the Parse van der Waals parameter set. The internal dielectric constant of the protein was set to 4 and the solvent to 80 [60, 61].

7. Essential Dynamics Analysis

The Essential Dynamics analysis[62] of the MD trajectories was carried out by applying the GROMACS modules *g_covar* and *g_anaeig* [63]; each frame of the trajectory was aligned on the Cα atoms of the starting conformation and the covariance matrix was calculated on Cα atoms. The analysis was restricted to the first 350 residues of Arr3-wt excluding the long loop region located at the opposite side of the receptor-binding surface, to facilitate comparison between the three arrestin constructs.

SUPPORTING INFORMATION

The Supporting Information is comprised of 5 Figures.

SI-1: **Inter-domain mobility and rotation**. Top, Distance fluctuation maps calculated for Cα-Cα distances along the MD trajectories for the second replicas of (A) Arr1-WT, (B) Arr1-R175E (B) and (C) Arr3-WT. N- and C-domains are identified by stripes with the same color code used in Figure 1. The dashed rectangles highlight the distance fluctuations in the N-C interface region. Bottom. Panels (D-F): Time dependent rotation of the selected residues in Arr1-WT (D), in Arr1-R175E (E) (Rotation values of Cα atoms of P212 S346 and A200 residues in red, green and blue, respectively) and Arr3-WT (F) (Rotation values of Cα atoms of S197, S331 and E186 residues in

red, green and blue respectively). Values in the time-line plots are averaged over 8-ns windows. (G) histogram showing the distribution of rotation angle of residue S346(S331) in both replicas of each system; **SI-2: : Finger loop dynamics.** (A) C-alpha atoms of the D73-I173 pair (D70-V168 in Arr3-wt), which are used to investigate the dynamics of the finger loop. They are shown in cyan on the inactive Arr1-WT (PDB ID: 1CF1) . N-domain is shown in purple, whereas the C-domain is shown in orange in new cartoon representation. The finger loop and the C-loop are shown in red and yellow respectively. Aromatic core residues (F65, F67, F79 and Y250) are shown in van der Waals representation. (B) Time-line plots showing the distance change of the finger loop with respect to a reference residue (I173 (V168 in Arr3-wt)) on the N-domain (B). Time-line plots show the change in the RMSD of the finger loop and the change in the distance between Cα atom of the F76-F245 pair in the first (left) and the second Arr3-wt replica (right);

SI-3: Top: Time-line plots show the distance change between the Cα atom of the first (K318 in Arr1-WT, K313 in Arr3-wt) and the terminal residue (M321 in Arr1-WT, L316 in Arr3-wt) of the short helix (top), in Arr1-wt (A), Arr1-mut (B), and Arr3-wt trajectories (C). Bottom, distance change between the Cα of residue R175 in Arr1-WT, R170 in Arr3-WT and the Cα of residue 392, which is part of the polar core, in Arr1-wt (D), Arr1-mut (E), and Arr3-wt trajectories (F); **SI-4**: Time line plots show the Cα RMSD change for the linker (residues 310-317) in Arr1-mut trajectories (top) and for the distorted strand (residues 256-262) in Arr3-wt trajectories (bottom).

SI-5: Structural representation of the inactive Arr1-WT (PDBID ID: 1CF1, chain D [26]) and Arr3-WT (PDBID: 3P2D, chain B [28]) proteins (same color code as Figure 1 in the main text). Bottom panel shows the Multiple Sequence Alignment (MSA) of the vertebrate phylogenetic family of Arr1-WT and Arr3-WT. Structural motifs shown to be important for activation are highlighted with the same color codes used in the structural representation above.

AUTHOR INFORMATION:

Corresponding Author

*Phone: +390228500031 Fax:+39 02 289.012.39 Email: giulia.morra@icrm.cnr.it

Author contributions: OS ran MD trajectories and analyzed results. ISM modeled the structures, performed sequence alignment and analyzed the electrostatic surface. GM carried out the dynamic and rotational analysis and supervised the research. GM, OS and ISM wrote the manuscript.

Funding. GM was supported by the CNR short-term mobility program and the CINECA grant BiaDop HP10CIQNJW for computational resources. ISM acknowledges support by FCT Investigator programme - IF/00578/2014 (co-financed by European Social Fund and Programa Operacional Potencial Humano) and by a Marie Skłodowska-Curie Individual Fellowship MSCA-IF-2015 [MEMBRANEPROT 659826]. ISM also acknowledges FEDER (Programa Operacional Factores de Competitividade - COMPETE 2020) and FCT–project: UID/NEU/04539/2013.

ACKNOWLEDGMENT

The authors are deeply grateful to Prof. Harel Weinstein from Weill Cornell Medical College (NYC, USA) for hosting the research and guiding their work with many inspiring discussions. They also thank Dr. Zhenlong Li and Michael Levine for the fruitful interaction. GM thanks Dr. Giorgio Colombo from ICRM CNR Milano for support and resources.

REFERENCES:

[1] Hausdorff, W. P., Caron, M. G., and Lefkowitz, R. J. (1990) Turning off the signal: desensitization of beta-adrenergic receptor function, *FASEB J. 4*, 2881-2889.

[2] Goodman, O. B., Jr., Krupnick, J. G., Santini, F., Gurevich, V. V., Penn, R. B., Gagnon, A. W., Keen, J. H., and Benovic, J. L. (1998) Role of arrestins in G-protein-coupled receptor endocytosis, *Adv. Pharmacol. 42*, 429-433.

[3] Lin, F. T., Krueger, K. M., Kendall, H. E., Daaka, Y., Fredericks, Z. L., Pitcher, J. A., and Lefkowitz, R. J. (1997) Clathrin-mediated endocytosis of the beta-adrenergic receptor is regulated by phosphorylation/dephosphorylation of beta-arrestin1, *J. Biol. Chem. 272*, 31051-31057.

[4] Shenoy, S. K., and Lefkowitz, R. J. (2011) beta-Arrestin-mediated receptor trafficking and signal transduction, *Trends Pharmacol. Sci. 32*, 521-533.

[5] Gimenez, L. E., Kook, S., Vishnivetskiy, S. A., Ahmed, M. R., Gurevich, E. V., and Gurevich, V. V. (2012) Role of Receptor-attached Phosphates in Binding of Visual and Non-visual Arrestins to G Protein-coupled Receptors, *J. Biol. Chem. 287*, 9028-9040.

[6] Lymperopoulos, A. (2012) Beta-arrestin biased agonism/antagonism at cardiovascular seven transmembrane-spanning receptors, *Curr. Pharm. Des. 18*, 192-198.

[7] Coureuil, M., Lecuyer, H., Scott, M. G., Boularan, C., Enslen, H., Soyer, M., Mikaty, G., Bourdoulous, S., Nassif, X., and Marullo, S. (2010) Meningococcus Hijacks a beta2-adrenoceptor/beta-Arrestin pathway to cross brain microvasculature endothelium, *Cell 143*, 1149-1160.

[8] Thathiah, A., Horre, K., Snellinx, A., Vandewyer, E., Huang, Y., Ciesielska, M., De Kloe, G., Munck, S., and De Strooper, B. (2013) beta-arrestin 2 regulates Abeta generation and gamma-secretase activity in Alzheimer's disease, *Nat. Methods 19*, 43-49.

[9] Bychkov, E. R., Gurevich, V. V., Joyce, J. N., Benovic, J. L., and Gurevich, E. V. (2008) Arrestins and two receptor kinases are upregulated in Parkinson's disease with dementia, *Neurobiol. Aging 29*, 379-396.

[10] Whalen, E. J., Rajagopal, S., and Lefkowitz, R. J. (2011) Therapeutic potential of beta-arrestin- and G protein-biased agonists, *Trends Mol. Med. 17*, 126-139.

[11] Gaidarov, I., Chen, X., Anthony, T., Maciejewski-Lenoir, D., Liaw, C., and Unett, D. J. (2013) Differential tissue and ligand-dependent signaling of GPR109A receptor: implications for anti-atherosclerotic therapeutic potential, *Cell Signal 25*, 2003-2016.

[12] Pierce, K. L., Premont, R. T., and Lefkowitz, R. J. (2002) Seven-transmembrane receptors, *Nat. Rev. Mol. Cell Biol. 3*, 639-650.

[13] Freedman, N. J., and Lefkowitz, R. J. (1996) Desensitization of G protein-coupled receptors, In *Recent Prog. Horm. Res., 51,* 319-353.

[14] Kovoor, A., Celver, J., Abdryashitov, R. I., Chavkin, C., and Gurevich, V. V. (1999) Targeted construction of phosphorylation-independent beta-arrestin mutants with constitutive activity in cells, *J. Biol. Chem. 274*, 6831-6834.

[15] Gurevich, V. V., Dion, S. B., Onorato, J. J., Ptasienski, J., Kim, C. M., Sterne-Marr, R., Hosey, M. M., and Benovic, J. L. (1995) Arrestin interactions with G protein-coupled receptors. Direct binding studies of wild type and mutant arrestins with rhodopsin, beta 2-adrenergic, and m2 muscarinic cholinergic receptors, *J. Biol. Chem. 270*, 720-731.

[16] Shukla, A. K., Westfield, G. H., Xiao, K., Reis, R. I., Huang, L.-Y., Tripathi-Shukla, P., Qian, J., Li, S., Blanc, A., Oleskie, A. N., Dosey, A. M., Su, M., Liang, C.-R., Gu, L.-L., Shan, J.-M., Chen, X., Hanna, R., Choi, M., Yao, X. J., Klink, B. U., Kahsai, A. W., Sidhu, S. S., Koide, S., Penczek, P. A., Kossiakoff, A. A., Woods, J., Virgil L, Kobilka, B. K., Skiniotis, G., and Lefkowitz, R. J. (2014) Visualization of arrestin recruitment by a G-protein-coupled receptor, *Nature 512*, 218-222.

[17] Vishnivetskiy, S. A., Schubert, C., Climaco, G. C., Gurevich, Y. V., Velez, M. G., and Gurevich, V. V. (2000) An additional phosphate-binding element in arrestin molecule. Implications for the mechanism of arrestin activation, *J. Biol. Chem. 275*, 41049-41057.

[18] Kang, Y., Zhou, X. E., Gao, X., He, Y., Liu, W., Ishchenko, A., Barty, A., White, T. A., Yefanov, O., Han, G. W., Xu, Q., de Waal, P. W., Ke, J., Tan, M. H. E., Zhang, C., Moeller, A., West, G. M., Pascal, B. D., Van Eps, N., Caro, L. N., Vishnivetskiy, S. A., Lee, R. J., Suino-Powell, K. M., Gu, X., Pal, K., Ma, J., Zhi, X., Boutet, S., Williams, G. J., Messerschmidt, M., Gati, C., Zatsepin, N. A., Wang, D., James, D., Basu, S., Roy-Chowdhury, S., Conrad, C. E., Coe, J., Liu, H., Lisova, S., Kupitz, C., Grotjohann, I., Fromme, R., Jiang, Y., Tan, M., Yang, H., Li, J., Wang, M., Zheng, Z., Li, D., Howe, N., Zhao, Y., Standfuss, J., Diederichs, K., Dong, Y., Potter, C. S., Carragher, B., Caffrey, M., Jiang, H., Chapman, H. N., Spence, J. C. H., Fromme, P., Weierstall, U., Ernst, O. P., Katritch, V., Gurevich, V. V., Griffin, P. R., Hubbell, W. L., Stevens, R. C., Cherezov, V., Melcher, K., and Xu, H. E. (2015) Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser, *Nature 523*, 561-567.

[19] Gurevich, V. V., and Gurevich, E. V. (2004) The molecular acrobatics of arrestin activation, *Trends Pharmacol. Sci. 25*, 105-111.
[20] Kim, M., Vishnivetskiy, S. A., Van Eps, N., Alexander, N. S., Cleghorn, W. M., Zhan, X., Hanson, S. M., Morizumi, T., Ernst, O. P., Meiler, J., Gurevich, V. V., and Hubbell, W. L. (2012) Conformation of receptor-bound visual arrestin, *Proc. Natl. Acad. Sci. U S A 109*, 18407-18412.
[21] Zhuo, Y., Vishnivetskiy, S. A., Zhan, X., Gurevich, V. V., and Klug, C. S. (2014) Identification of Receptor Binding-induced Conformational Changes in Non-visual Arrestins, *J. Biol. Chem. 289*, 20991-21002.
[22] Kim, Y., Hofmann, K., Ernst, O., Scheerer, P., Choe, H., and Sommer, M. (2013) Crystal structure of pre-activated arrestin p44, *Nature 497*, 142-146.
[23] Shukla, A. K., Manglik, A., Kruse, A. C., Xiao, K., Reis, R. I., Tseng, W.-C., Staus, D. P., Hilger, D., Uysal, S., Huang, L.-Y., Paduch, M., Tripathi-Shukla, P., Koide, A., Koide, S., Weis, W. I., Kossiakoff, A. A., Kobilka, B. K., and Lefkowitz, R. J. (2013) Structure of active b-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide, *Nature 497*, 137-141.
[24] Ostermaier, M. K., Peterhans, C., Jaussi, R., Deupi, X., and Standfuss, J. (2014) Functional map of arrestin-1 at single amino acid resolution, *Proc. Natl. Acad. Sci. U S A 111*, 1825-1830.
[25] Szczepek, M., re, F. B. e., Hofmann, K. P., Elgeti, M., Kazmin, R., Rose, A., Bartl, F. J., von Stetten, D., Heck, M., Sommer, M. E., Hildebrand, P. W., and Scheerer, P. (2014) Crystal structure of a common GPCR-binding interface for G protein and arrestin, *Nat. Commun.5*, 1-8.
[26] Hirsch, J. A., Schubert, C., Gurevich, V. V., and Sigler, P. B. (1999) The 2.8 A crystal structure of visual arrestin: a model for arrestin's regulation, *Cell 97*, 257-269.
[27] Granzin, J., Wilden, U., Choe, H. W., Labahn, J., Krafft, B., and Buldt, G. (1998) X-ray crystal structure of arrestin from bovine rod outer segments, *Nature 391*, 918-921.
[28] Zhan, X., Gimenez, L. E., Gurevich, V. V., and Spiller, B. W. (2011) Crystal Structure of Arrestin-3 Reveals the Basis of the Difference in Receptor Binding Between Two Non-visual Subtypes, *J. Mol. Biol. 406*, 467-478.
[29] Nobles, K. N., Guan, Z., Xiao, K., Oas, T. G., and Lefkowitz, R. J. (2007) The Active Conformation of β-Arrestin1: direct evidence for the phosphate sensor in the N-domain and conformaional differences in the active states of β arrestins 1 and 2., *J. Biol. Chem. 282*, 21370-21381.
[30] Shukla, A. K., Violin, J. D., Whalen, E. J., Gesty-Palmer, D., Shenoy, S. K., and Lefkowitz, R. J. (2008) Distinct conformational changes in beta-arrestin report biased agonism at seven-transmembrane receptors, *Proc. Natl. Acad. Sci. U S A 105*, 9988-9993.
[31] Yun, Y., Kim, D. K., Seo, M.-D., Kim, K.-M., and Chung, K. Y. (2014) Different conformational dynamics of β-arrestin1 and β-arrestin2 analyzed by hydrogen/deuterium exchange mass spectrometry, *Biochem. Biophys. Res. Commun. 457*, 50-57
[32] Hanson, S. M., and Gurevich, V. V. (2006) The Differential Engagement of Arrestin Surface Charges by the Various Functional Forms of the Receptor, *J. Biol. Chem. 281*, 3458-3462.
[33] Gurevich, V. V., and Benovic, J. L. (1997) Mechanism of Phosphorylation-Recognition by Visual Arrestin and the Transition of Arrestin into a High Affinity Binding State, *Mol. Pharmacol.. 51*, 161-169.
[34] Granzin, J., Stadler, A., Cousin, A., Schlesinger, R., and Batra-Safferling, R. (2015) Structural evidence for the role of polar core residue Arg175 in arrestin activation, *Sci. Rep. 5*, 15808.
[35] Granzin, J., Cousin, A., Weirauch, M., Schlesinger, R., Büldt, G., and Batra-Safferling, R. (2012) Crystal Structure of p44, a Constitutively Active Splice Variant of Visual Arrestin, *J. Mol. Biol. 416*, 611-618.

[36] Milano, S. K., Pace, H. C., Kim, Y. M., Brenner, C., and Benovic, J. L. (2002) Scaffolding functions of arrestin-2 revealed by crystal structure and mutagenesis, *Biochemistry 41*, 3321-3328.
[37] Gimenez, L. E., Vishnivetskiy, S. A., Baameur, F., and Gurevich, V. V. (2012) Manipulation of Very Few Receptor Discriminator Residues Greatly Enhances Receptor Specificity of Non-visual Arrestins, *J. Biol. Chem. 287*, 29495-29505.
[38] Vishnivetskiy, S. A., Hirsch, J. A., Velez, M. G., Gurevich, Y. V., and Gurevich, V. V. (2002) Transition of arrestin into the active receptor-binding state requires an extended interdomain hinge, *J. Biol. Chem. 277*, 43961-43967.
[39] Nuber, S., Zabel, U., Lorenz, K., Nuber, A., Milligan, G., Tobin, A. B., Lohse, M. J., and Hoffmann, C. (2016) β-Arrestin biosensors reveal a rapid, receptor-dependent activation/deactivation cycle, *Nature 531*, 661-664.
[40] Lee, M.-H., Appleton, K. M., Strungs, E. G., Kwon, J. Y., Morinelli, T. A., Peterson, Y. K., Laporte, S. A., and Luttrell, L. M. (2016) The conformational signature of β-arrestin2 predicts its trafficking and signalling functions, *Nature 531*, 665-668.
[41] Nobles, K. N., Xiao, K., Ahn, S., Shukla, A. K., Lam, C. M., Rajagopal, S., Strachan, R. T., Huang, T. Y., Bressler, E. A., Hara, M. R., Shenoy, S. K., Gygi, S. P., and Lefkowitz, R. J. (2011) Distinct Phosphorylation Sites on the 2-Adrenergic Receptor Establish a Barcode That Encodes Differential Functions of -Arrestin, *Sci. Signal. 4*, ra51.
[42] Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling, *Bioinformatics 22*, 195-201.
[43] Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009) The SWISS-MODEL Repository and associated resources, *Nucleic Acids Res. 37*, D387-D392.
[44] Fiser, A., Do, R. K. G., and Šali, A. (2000) Modeling of loops in protein structures, *Protein Sci. 9*, 1753-1773.
[45] Fiser, A., and Sali, A. (2003) ModLoop: automated modeling of loops in protein structures, *Bioinformatics 19*, 2500-2501.
[46] DeLano, W. L. (2002) The PyMOL molecular graphics system, DeLano Scientific, Palo Alto, CA, USA.
[47] Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008) Very fast prediction and rationalization of pKa values for protein–ligand complexes, *Proteins: Struct., Funct., Bioinf. 73*, 765-783.
[48] Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pKa values, *Proteins: Struct., Funct., Bioinf. 61*, 704-721.
[49] Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions, *J. Chem. Theory Comp.7*, 525-537.
[50] Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, *Bioinformatics 28*, 1647-1649.
[51] Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res. 32*, 1792-1797.
[52] Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids, *Nucleic Acids Res. 38*, W529-533.

[53] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J. Comput. Chem. 26*, 1781-1802.

[54] Mackerell, A. D., Feig, M., and Brooks, C. L. (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, *J. Comput. Chem. 25*, 1400-1415.

[55] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method, *J. Chem. Phys. 103*, 8577-8593.

[56] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys. 23*, 327-341.

[57] Mark, P., and Nilsson, L. (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K, *J. Phys. Chem. A 105*, 9954-9960.

[58] Morra, G., Potestio, R., Micheletti, C., and Colombo, G. (2012) Corresponding Functional Dynamics across the Hsp90 Chaperone Family: Insights from a Multiscale Analysis of MD Simulations, *Plos Comput. Biol. 8*, e1002433.

[59] Li, L., Li, C., Sarkar, S., Zhang, J., Witham, S., Zhang, Z., Wang, L., Smith, N., Petukh, M., and Alexov, E. (2012) DelPhi: a comprehensive suite for DelPhi software and associated resources, *BMC Biophysics 5*, 1-11.

[60] Lee, B., and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility, *J. Mol. Biol. 55*, 379 - 400.

[61] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J. Mol. Graph. 14*, 33-38.

[62] Amadei, A., Linssen, A. B. M., Berendsen, H. J. C. (1993) Essential dynamics of proteins., *Proteins: Struct. Funct. Genet. 17*, 412-425.

[63] Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *J. Chem. Theory Comp. 4*, 435-447.

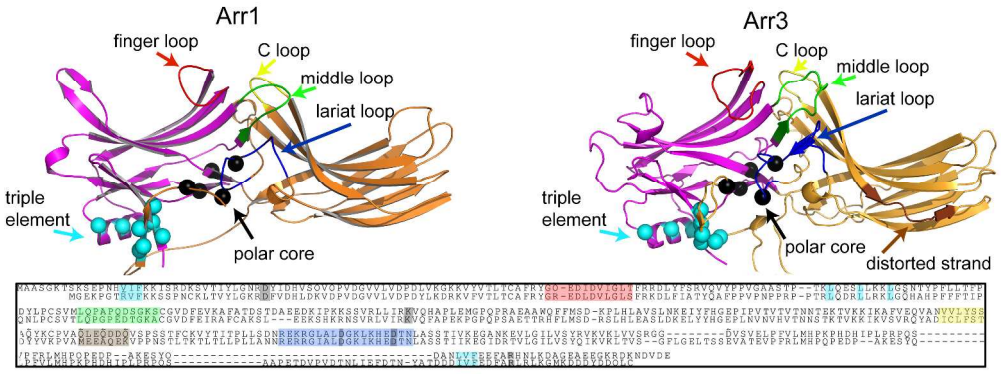


Figure 1: Structural representation of the inactive Arr1-WT (PDBID ID: 1CF1, chain D) and Arr3-WT (PDBID: 3P2D, chain B) proteins. The various segments are indicated by the colors: N-domain (in purple); C-domain (orange), the "finger loop" (red), the "middle loop" (green), the "lariat loop including the "gate loop" region (in blue), and the "C-loop" (yellow). Cα atoms of the residues that constitute the "polar core" and the "triple element region" are shown in black and cyan, respectively. Distorted β-strand on the C-domain of Arr3-WT is shown in brown on the right. Bottom panel shows the Sequence Alignment (MSA) of Arr1-WT and Arr3-WT for Homo sapiens and Bos Taurus. Structural motifs shown to be important for activation are highlighted with the same color codes used in the structural representation above.

427x170mm (300 x 300 DPI)

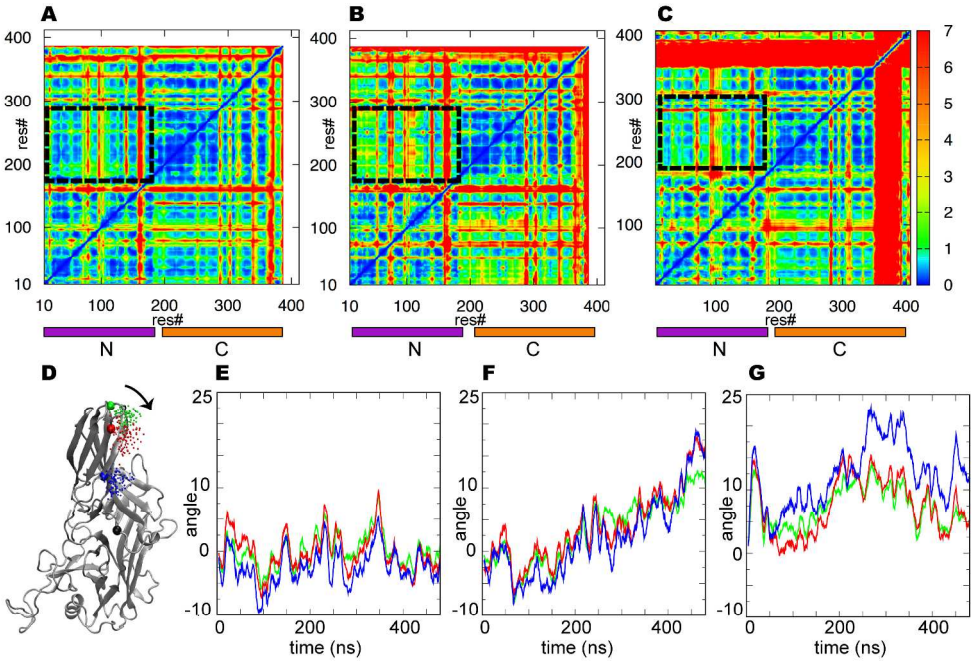


Figure 2: Top, Distance fluctuation maps calculated for Cα-Cα distances along the MD trajectories for (A) Arr1-WT, (B) Arr1-R175E (B) and (C) Arr3-WT. N- and C-domains are identified by stripes with the same color code used in Figure 1. The dashed rectangles highlight the distance fluctuations in the N-C interface region. Bottom, Panel D: View of Arr3-WT, with rotation axis perpendicular to the picture plane and passing through Cα of I321, depicted in black (I306 in Arr1-WT); rotation on this axis in the direction of the arrow corresponds to an increase in the rotation angle. Cα atoms used in quantifying the rotation along the MD trajectory are represented in blue, red, and green. Panels (E-G): Time dependent rotation of the selected residues in Arr1-WT (E), in Arr1-R175E (F) (Rotation values of Cα atoms of P212 S346 and A200 residues in red, green and blue, respectively) and Arr3-WT (G). (Rotation values of Cα atoms of S197, S331 and E186 residues in red, green and blue respectively). Values in the time-line plots are averaged over 8-ns windows.

305x217mm (300 x 300 DPI)

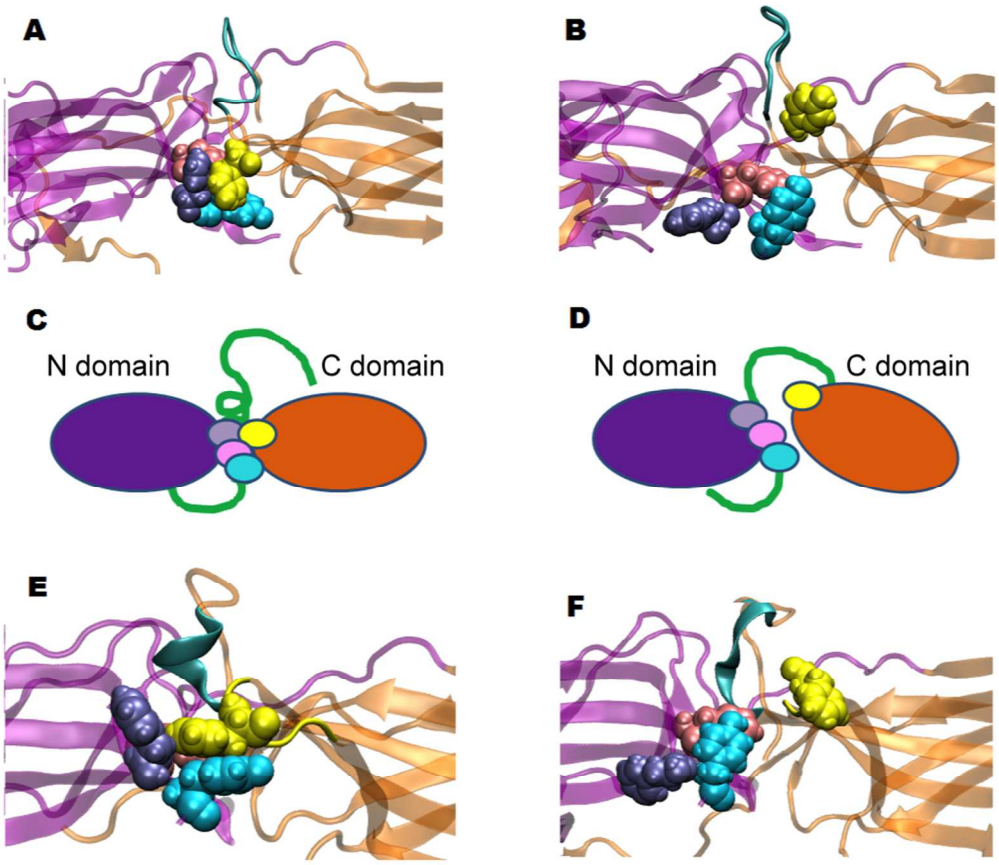


Figure 3: Repositioning of structural elements in a comparison of snapshots from the starting crystal structure (A, E) and the active-like conformations in the Arr1-R175E and in the Arr3-WT trajectory (B and F), showing the position and structure of the C-loop (yellow), the aromatic core and the short-helix (cyan) (for the description of the "active-like" conformations, see Section 3.a below). The repositioning of structural elements -going from the inactive to the active-like conformation- is also shown with cartoon representation (C, D). Color coding is the same as in Figure 1: the N domain is represented in purple, the C domain in orange. Aromatic core residues (F62, F64, F76 and F245) are shown in van der Waals representation with pink, cyan, purple and yellow color, respectively.

85x73mm (300 x 300 DPI)

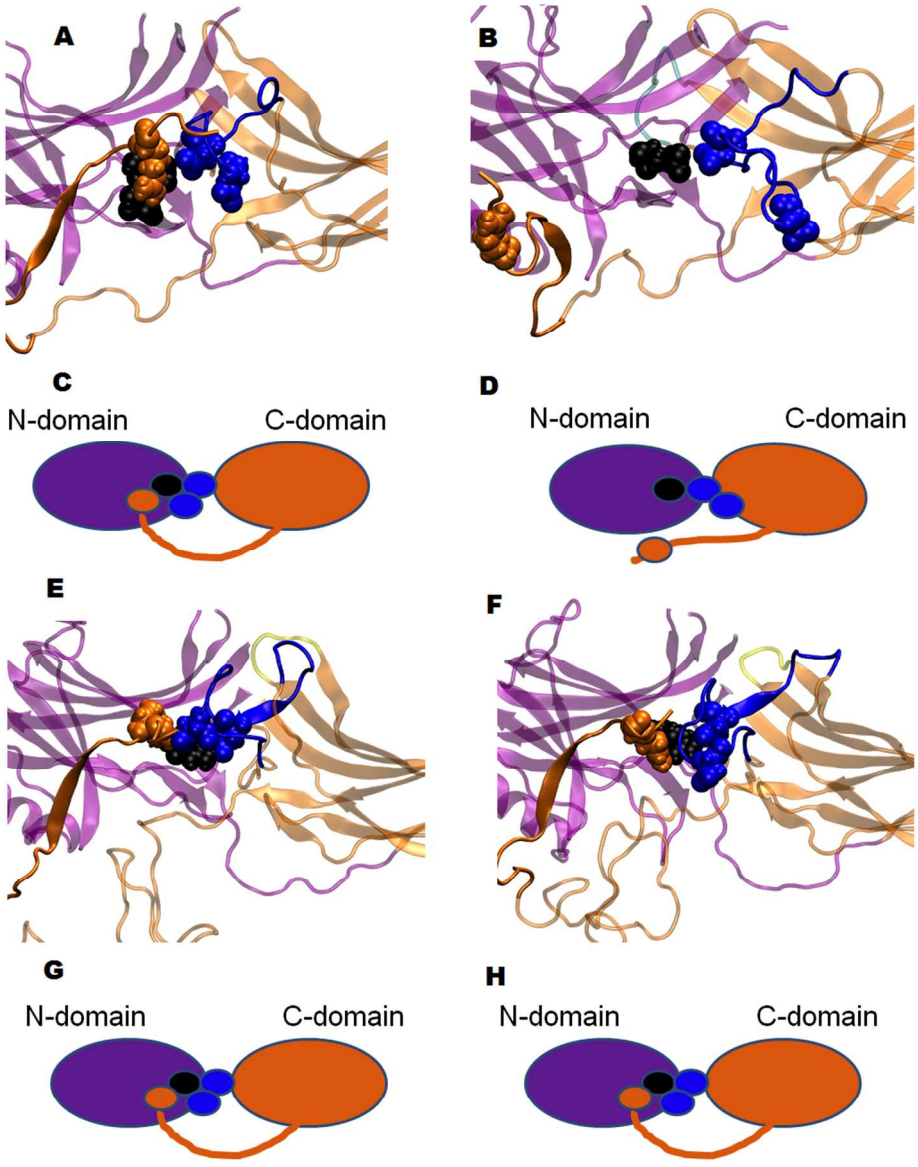


Figure 4: The polar core and the gate loop positions in the inactive state of Arr1-R175E and Arr3-WT trajectories (A and E), and see Section 3.a for description of inactive and active state of the trajectories), and snapshots of Arr1-R175E (B) and Arr3-WT (F) taken from the pre-activated region of the Arr1-R175E and Arr3-WT first replicas (see section 3) showing the state of polar core and gate loop. Residue R175 (R170 in Arr3-WT), D296 and D303 (D291 and D298 in Arr3-WT), and R382 (R398 in Arr3-WT) are shown in van der Waals representation in purple, blue, blue, and orange, respectively. The repositioning of structural elements -going from the inactive to the active-like conformation- is also shown with cartoon representation for Arr1-R175E and Arr3-WT in panels C, D and G, H, respectively. Color code of the rest: see Figure 1.

87x114mm (300 x 300 DPI)

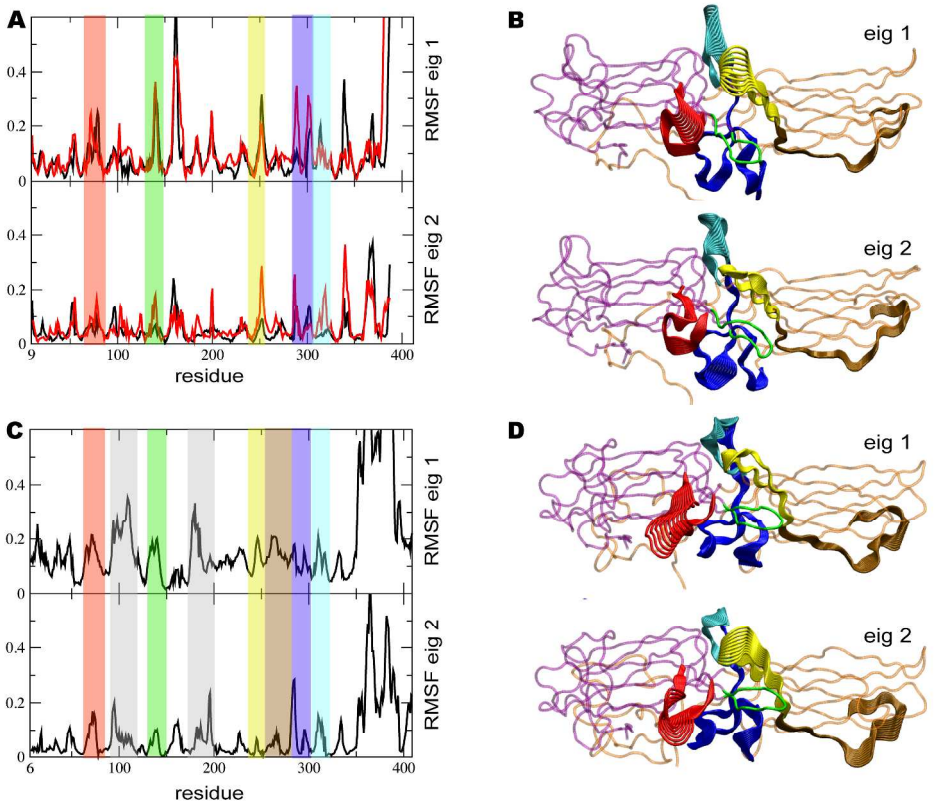


Figure 5. A) Profile of fluctuations along the first two essential eigenvectors in Arr1-WT (black) and Arr1-R175E (red). The covariance matrix is built considering the Cα atoms, (see Methods-Essential Dynamics). Specific regions are highlighted with color bars: the finger loop in red; the middle loop in green; the C-loop in yellow; the ARG and the gate loop (forming together the lariat loop) in blue; the short helix in cyan. B) Extreme structures along eigenvector 1 (top); Extreme structures along eigenvector 2 (bottom) for Arr1-R175E. Only the highlighted regions are explicitly shown with the same color code as in A). The middle loop is not shown for clarity. C) same as A) for Arr3-WT (black), with helix 1 region and N-C linker highlighted in grey and the "distorted beta strand" region in brown; ; D) same as B) for Arr3-WT.

320x307mm (300 x 300 DPI)

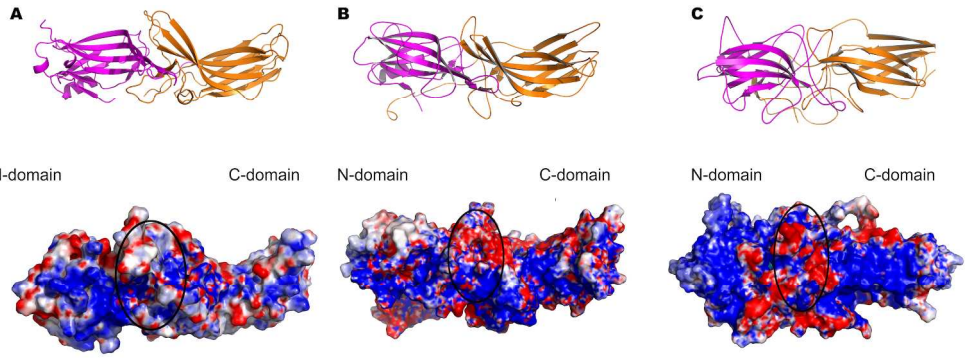
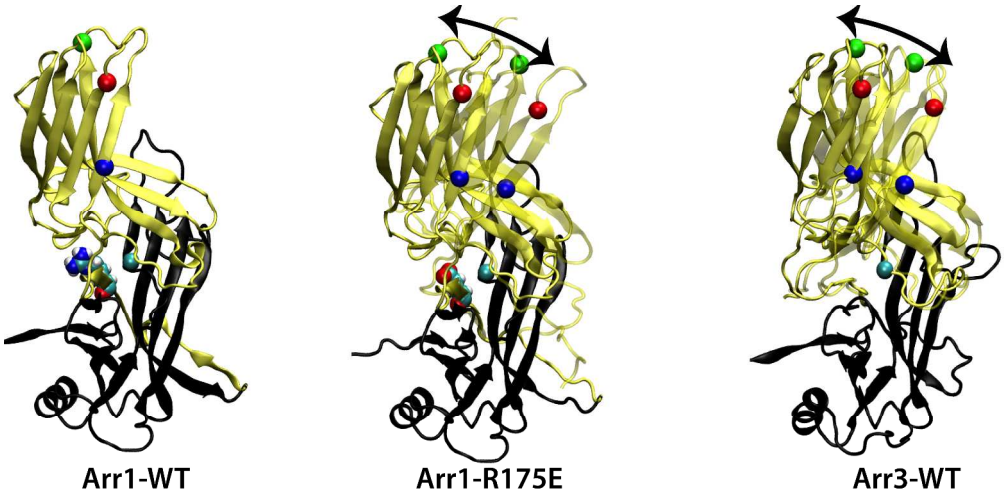


Figure 6: Electrostatics potential difference (red-white-blue) between active and inactive structures, calculated by the program DelPhi and averaged over the MD simulations (panel B and C) or from the PDB structures (panel A). Each protein is shown both with a cartoon representation (above) and with the surface representation (below) in the same orientation, the latter colored according to the electrostatic potential difference. . The N-C interface including the ARCs and the gate loop is highlighted in the oval. A. 4ZWF and 3UGU (Arr-WT active and inactive) difference; B. Arr1-R175E after C-domain rotation and Arr1-WT difference; C. Arr3-WT after and before C-domain rotation difference.

250x116mm (300 x 300 DPI)



281x144mm (300 x 300 DPI)