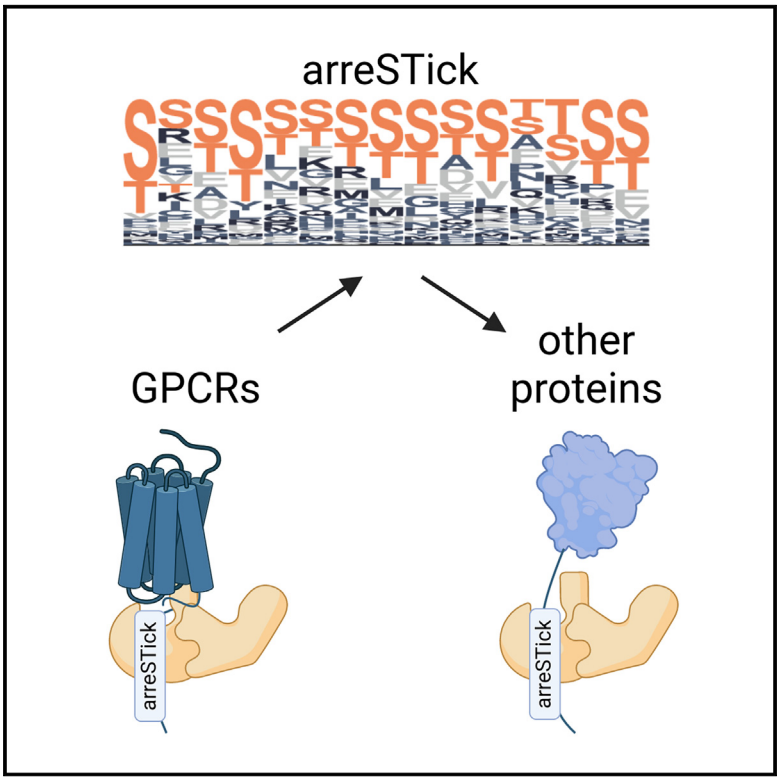
Cell Reports

Article

ArreSTick motif controls β-arrestin-binding stability and extends phosphorylation-dependent β-arrestin interactions to non-receptor proteins

Graphical abstract



Authors

András Dávid Tóth, Eszter Soltész-Katona, Katalin Kis, ..., Asuka Inoue, László Hunyady, Gábor Turu

Correspondence

hunyady.laszlo@ttk.hun-ren.hu (L.H.), turu.gabor@ttk.hun-ren.hu (G.T.)

In brief

Tóth et al. developed a machine learning-based approach to identify sequence motifs in GPCRs, termed arreSTick, that mediate stable β-arrestin binding. They found that arreSTick is also present in many other proteins, suggesting that β-arrestins may regulate non-GPCR proteins through a phosphorylation-dependent mechanism, similar to its function in GPCRs.

Highlights

- ArreSTick is a sequence pattern characteristic of stable GPCR-β-arrestin interaction
- Numerous other proteins possess the arreSTick pattern
- β-arrestin2 regulates HTSF1 via binding to its arreSTick motif

Tóth et al., 2024, Cell Reports *43*, 114241
May 28, 2024
https://doi.org/10.1016/j.celrep.2024.114241

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

ArreSTick motif controls β-arrestin-binding stability and extends phosphorylation-dependent β-arrestin interactions to non-receptor proteins

András Dávid Tóth,[1,2,7] Eszter Soltész-Katona,[1,3,7] Katalin Kis,[3] Viktor Guti,[3] Sharon Gilzer,[3] Susanne Prokop,[3] Roxána Boros,[3] Ádám Misák,[3] András Balla,[3,4] Péter Várnai,[3,4] Lilla Turiák,[5] András Ács,[5] László Drahos,[5] Asuka Inoue,[6] László Hunyady,[1,3,8,*] and Gábor Turu[1,3,8,9,*]

[1]Institute of Molecular Life Sciences, Centre of Excellence of the Hungarian Academy of Sciences, HUN-REN Research Centre for Natural Sciences, Magyar Tudósok krt. 2., 1117 Budapest, Hungary
[2]Department of Internal Medicine and Haematology, Semmelweis University, Szentkirályi street 46, 1088 Budapest, Hungary
[3]Department of Physiology, Semmelweis University, Tűzoltó street 37-47, 1094 Budapest, Hungary
[4]HUN-REN SE Hungarian Research Network Laboratory of Molecular Physiology, Budapest, Hungary
[5]Institute of Organic Chemistry, HUN-REN Research Centre for Natural Sciences, Magyar Tudósok krt. 2., 1117 Budapest, Hungary
[6]Molecular and Cellular Biochemistry, Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan
[7]These authors contributed equally
[8]These authors contributed equally
[9]Lead contact
*Correspondence: hunyady.laszlo@ttk.hun-ren.hu (L.H.), turu.gabor@ttk.hun-ren.hu (G.T.)
https://doi.org/10.1016/j.celrep.2024.114241

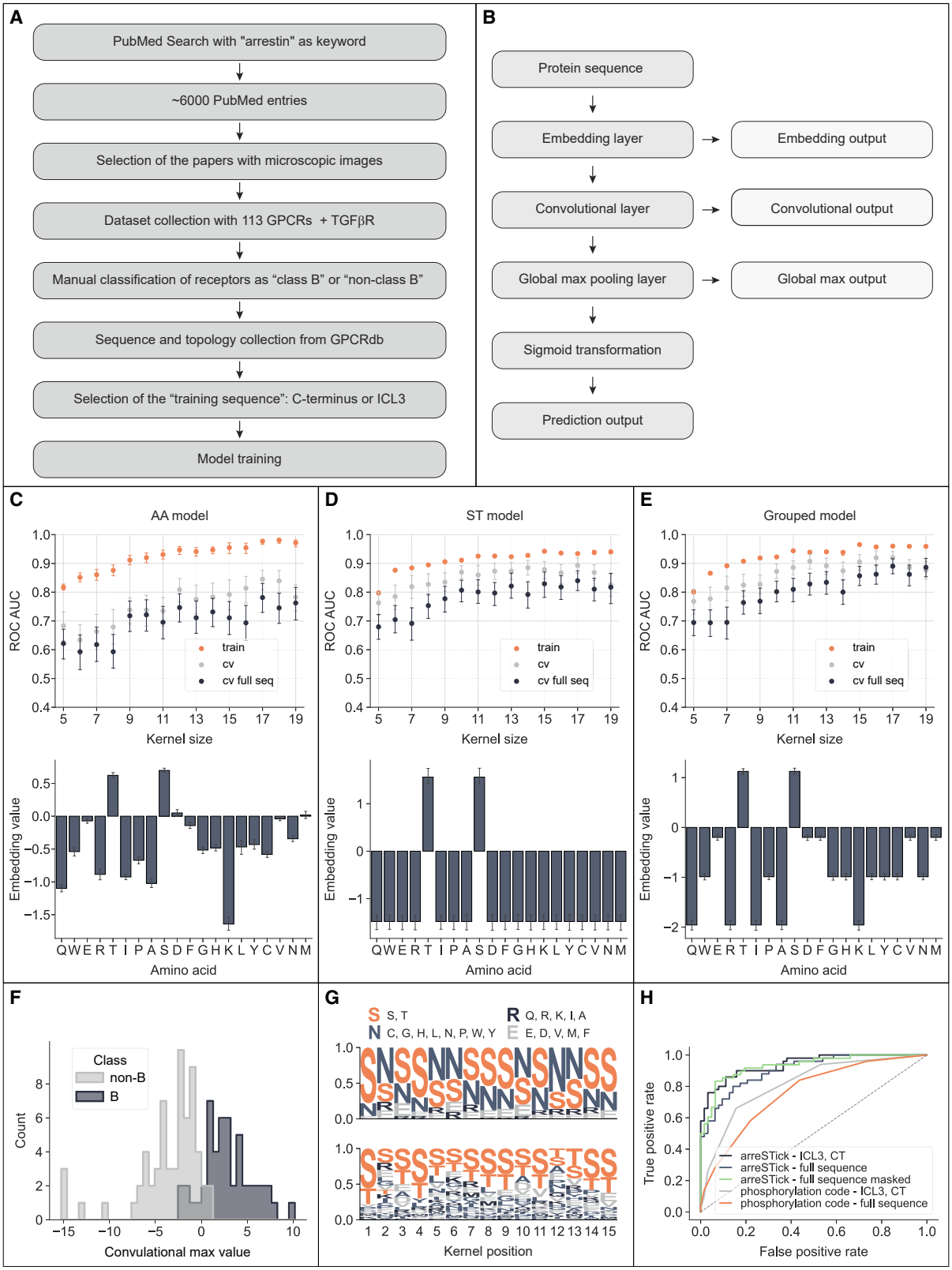
SUMMARY

The binding and function of β-arrestins are regulated by specific phosphorylation motifs present in G protein-coupled receptors (GPCRs). However, the exact arrangement of phosphorylated amino acids responsible for establishing a stable interaction remains unclear. We employ a 1D sequence convolution model trained on GPCRs with established β-arrestin-binding properties. With this approach, amino acid motifs characteristic of GPCRs that form stable interactions with β-arrestins can be identified, a pattern that we name "arreSTick." Intriguingly, the arreSTick pattern is also present in numerous non-receptor proteins. Using proximity biotinylation assay and mass spectrometry analysis, we demonstrate that the arreSTick motif controls the interaction between many non-receptor proteins and β-arrestin2. The HIV-1 Tat-specific factor 1 (HTSF1 or HTATSF1), a nuclear transcription factor, contains the arreSTick pattern, and its subcellular localization is influenced by β-arrestin2. Our findings unveil a broader role for β-arrestins in phosphorylation-dependent interactions, extending beyond GPCRs to encompass non-receptor proteins as well.

INTRODUCTION

G protein-coupled receptors (GPCRs) form one of the largest families of proteins in the human proteome, and they are prominent targets in human therapy.[1] Stimulation of the receptors leads to the activation of the canonical signal transduction pathways, followed by receptor phosphorylation and coupling to β-arrestin proteins.[2,3] β-arrestin binding results in the desensitization and internalization of GPCRs, and they act as important scaffold proteins as well. It initiates a broad range of signaling events, such as activation of mitogen-activated protein kinase (MAPK) signaling pathways, including ERK1/2, p38, and c-Jun N-terminal kinase-3, as well as that of c-Src family kinases, Akt kinase, PI3 kinase, and RhoA.[2,3]

Based on the duration and stability of their binding to β-arrestins, receptors have been traditionally classified into class A and class B groups. Class A receptors, including V1A vasopressin receptor, β2-adrenergic receptor (β2AR), and CB1 cannabinoid receptor, bind β-arrestin at the plasma membrane and release it quickly after internalization.[4–6] In contrast, class B receptors, such as AT1 angiotensin receptor (AT1R), V2 vasopressin receptor (V2R), and oxytocin receptor, have a stronger and more stable association with β-arrestins, an interaction that can also be found at intracellular vesicles after internalization.[4,7] This strong, long-lasting interaction leads to a more pronounced activation of β-arrestin-dependent signaling pathways.[8,9] This classification is still frequently utilized and conceptually valid; however, recent advancements have unveiled a wide spectrum in the strength and dynamics of β-arrestin recruitment and identified multiple mechanisms that regulate the assembly or disassembly of the receptor-β-arrestin complexes.[3] GPCRs interact with β-arrestins via their cytoplasmic side of the transmembrane regions (also referred to as core interaction) and their phosphorylated C terminus or third intracellular loop (ICL3),[10,11] with all interaction sites playing important roles in the signaling outcomes.[12–15] Sustained interaction is primarily determined by the presence of phosphorylated serine/threonine clusters on the receptor, which interact with a positively charged region of the N-domain of



(legend on next page)

β-arrestins.[11,16–18] Distinct receptors have varying phosphorylation motifs, and the exact pattern required for tight β-arrestin binding remains unclear. Zhou et al. identified long and short phosphorylation codes,[19] whose presence and number in a few GPCRs correlate with their ability to bind β-arrestins with high affinity. A common pattern in the short and long code is the presence of a PxxP motif, where P denotes a phosphorylated serine or threonine and x any other amino acid residue, and the code is extended at the start either with a shorter Px or longer Pxx sequence, respectively. Mayer et al. analyzed the importance of the rhodopsin C-terminal serines and threonines in the tight binding of the visual arrestin.[20] They found a pattern containing the PxxP motif similar to the short and long codes. In recent studies, a PxPP motif was identified as a sequence present in many GPCR C termini. This motif is important in inducing the active conformation of the β-arrestins; however, it does not seem to discriminate well between class A and B receptors.[18,21]

β-arrestins have been shown to directly interact with a broad range of non-GPCR proteins, and they act as crucial adaptors in various cellular pathways.[3] While the significance of specific GPCR phosphorylation patterns for stable interaction with β-arrestins is well recognized, it is not known whether such sequence patterns in non-receptor proteins may serve as β-arrestin recognition sites. Consequently, the positively charged region of the β-arrestin N-domain, which is responsible for receptor C terminus binding, is viewed as a site primarily intended for interactions with GPCRs.[3,22] However, if such sequence motifs are present in non-GPCR proteins as well, they may drive interaction with β-arrestins.

In this study, we established a convolutional neural network to identify motifs characteristic of receptors with stable β-arrestin binding using a dataset of 114 receptors. These receptors were categorized by their reported capacity for promoting class B-type β-arrestin translocation. Using only the sequence and the class information of the receptors, we were able to predict the published type of interaction with high accuracy. We also found that the seronine/threonine-rich pattern, which we term arreSTick, is present in many non-GPCR proteins and can bind to β-arrestin2 through its phosphate-binding residues. As a proof of concept, we have studied the effect of β-arrestin binding on the intracellular localization of the HTSF1 transcription factor, one of the intracellular proteins that contains the arreSTick motif. Our data show that HTSF1 interacts with both β-arrestins, and β-arrestin2 can regulate its intracellular localization via the arreSTick motif. The prediction algorithm is available on Zenodo, or it can be used online for arreSTick prediction at www.arreSTick.org.

RESULTS

Identification of the serine-threonine pattern responsible for class B-type β-arrestin binding

The phosphorylation of the C terminus or sometimes ICL3 of GPCRs is a main determinant of β-arrestin binding, and, most likely, the pattern of the phosphorylation (phosphorylation barcode) determines the stability of the receptor-β-arrestin interaction.[7] We hypothesized that the specific amino acid pattern in GPCRs required for the stable interaction can be predicted with machine learning algorithms using only sequence information. To test this hypothesis, we first constructed a comprehensive training dataset that categorizes GPCRs based on their ability to stably bind β-arrestins. The conventional class A/B categorization is the most widely used to describe such properties of receptors and is based on well-defined criteria and experimental readouts. Therefore, receptors that have been demonstrated to engage in class B-type β-arrestin binding are likely to contain sequence motifs that drive sustained interaction. By reviewing the literature, we created a receptor-β-arrestin-binding-class database consisting of 114 receptors (113 GPCRs + transforming growth factor β [TGFβ] receptor; Figure 1A; Table S1), which includes their β-arrestin-binding properties as class B or non-class B. To build a machine learning algorithm, we opted for convolutional neural networks since these are well suited for local pattern search in 2D (e.g., images) and 1D (e.g., sequence) data structures. Convolutional networks learn kernels (i.e., assign weights for each kernel position), which represent characteristic patterns. The analyzed sequences need to be represented (embedded) as vectors of numbers in one or more dimensions. At each point of the sequence, the dot product of the kernel and the equal-sized parts of the following sequence is calculated (Figure S1). These products are the values of the convolution at any particular point in the sequence. Higher values mean a better match with the pattern, and we can use the maximal value along the sequence to predict the class of the receptor and the best-matching position in the sequence (Figure 1B, see Figure S1 for an example convolution on a single receptor).

After the initial optimization steps, we opted for the simplest network structure consisting of a single convolutional layer and a single kernel, and the amino acids were represented as single floating-point numbers (Figure S1). For most of the receptors, we used the amino acid sequence of C termini (annotated using the gpcrdb.org API[23]) as training data. Since some of the receptors are known to bind β-arrestins through their ICL3, for those having

Figure 1. Convolutional model for prediction of the stable interaction between GPCRs and β-arrestins

(A) Flowchart of the training set creation process.

(B) Convolutional neural network structure used in the prediction models.

(C–E) CV results (upper panels) and the trained embedding values of individual amino acids (lower panels). Mean and 95% confidence intervals of 50 cross-validations are shown.

(F) Representative distribution of the global maximal values after passing the training receptor set into the grouped model. Distribution of the global maximal outputs for each receptor is shown for class B and non-B (class A or β-arrestin non-binding) receptors.

(G) The arreSTick pattern. A single grouped model was trained with all receptors in the training set, and the 15-amino-acid-long sequence (arreSTick) starting at the global maximal convolutional value was selected for each receptor. The logo with the grouped amino acid frequencies (upper panel) and with the individual amino acid frequencies (bottom panel) are shown. The groups are labeled as follows: S (positive embedding values), [S,T]; E (neutral embedding values), [E,D,V,M,F]; R (large negative embedding values), [Q,R,K,I,A]; N (slightly negative embedding values), all other amino acids.

(H) ROC curves with different CV strategies using the grouped model predictions or the total number of short and long phosphorylation codes.

very short C terminus and long ICL3 (over 80 amino acids), we used the sequence of the latter instead (Figure S2; Table S2). We utilized a cross-validation (CV) strategy by dividing the entire dataset in each step into random train and CV sets (91% and 9%), trained the model on the train set, and predicted the receptors in the CV sets, and then we repeated the process 50 times. In order to find the best filter size, we also used different kernel sizes between 5 and 19. During each round, the model was trained on the C-terminal and ICL3 regions, and the CV set was predicted based on their C-terminal and ICL regions or full sequences (Figures 1C–1E, upper panels). After running the CVs, we selected an optimal kernel length and trained the model multiple times using all receptor data and extracted the amino acid embedding values to evaluate the importance of different amino acids in the prediction (Figures 1C–1E, bottom panels). Initially, we embedded all amino acids in the model (AA model). As shown in Figure 1C, this model performed very well on the training set, especially using kernel lengths of at least 15 amino acids. However, the CV results were substantially lower than the training performance, suggesting that the model was overfitting on the training data. Nevertheless, only serines and threonines had large positive values in the embedding, and all other amino acids had negative or nearly zero embedding values (Figure 1C, lower panel). This is in agreement with the observation that phosphorylation of certain S/T amino acids is required for strong coupling of receptors to β-arrestins.[24] To investigate whether reducing the number of features could reduce overfitting, we grouped the amino acids and assigned the same embedding values within each group. First, we categorized the amino acids into S/T versus non-S/T amino acids (ST model"; Figure 1D). With this embedding strategy, there was less overfitting compared to the AA model. As an intermediate embedding strategy, we also grouped the amino acids into four groups (ST, EDVMF, QRKIA, and CGHLNPWY groups) based on their embedding values in the AA model, since amino acids with comparable values may have similar roles in the β-arrestin binding (grouped model). By applying this model, the CV receiver operating characteristic curve (ROC) AUC values went over 0.9 and near 0.9 when the C terminus and the full sequence were predicted, respectively (Figure 1E). For subsequent predictions, we opted for the grouped model with a kernel length of 15, as this model structure showed a good performance in the CVs.

The classification of receptors in the model is based on the maximal convoluted values. These maximal values effectively differentiate the class B receptors from non-class B receptors, with only a few exceptions (Figure 1F). The kernel in the trained model shows the importance of the individual positions within the sequence region. The sequence region within individual receptors that best matches with the kernel, when classified as class B, likely corresponds to the region that undergoes phosphorylation and binds to β-arrestin. Therefore, we named this receptor region arreSTick, referring to the sticky, phosphorylated S/T pattern in the sequence. Twenty example kernels are shown in Figures S3–S5 from different individual trainings. For the visualization of the pattern, we grouped the amino acids into the four groups according to the grouped model, and the group frequencies in the best-matching region for all GPCRs with class B-binding properties using a representative kernel are shown in Figure 1G upper panel. The individual amino acid frequencies based on the same group model in these regions are shown in Figure 1G lower panel. The position of the S/T amino acids resembles some of the previously reported patterns.[19] Namely, one could recognize both short (PxPxxP, e.g., positions 9–14) and long (PxxPxxP, e.g., positions 1–7) phosphorylation codes. However, the convolutional model performs better in CV than the model using only the number of these reported phosphorylation codes within a sequence, particularly when the full sequence is predicted (Figure 1H). Since β-arrestins bind to the unstructured C-terminal part of the receptors, we hypothesized that excluding the structured α-helical parts of the receptors during the prediction could improve the predictions. To implement the masking, we opted for the AlphaFold2 confidence score,[25,26] which indicates a structured region with high confidence, when exceeding a value of 70. As anticipated, masking improves our predictions (Figure 1H). To get a closer view of the prediction of individual receptors, we convoluted the full sequences of four receptors with class B-type β-arrestin binding properties, AT1R, V2R, ACK2 receptor (ACKR2), and complement C5a receptor 1 (C5AR1) (Figure 2). For each receptor prediction, the models were trained without including the predicted receptor in the training set. At each point of the sequence, the convolution values were calculated, and these values were sigmoid transformed, corresponding to the probability of class B-type-binding region at each point. In the cases of AT1R, V2R, and ACKR2, the predicted β-arrestin-binding regions, starting at the maximal probability values, overlap with the experimentally identified regions.[7,9,11,27,28] For C5AR1, the maximal convolutional value is found in a helical region, which is unlikely to serve as a phosphorylation site or participate in β-arrestin binding. However, the sequence with the highest probability, outside the areas with high AlphaFold2 confidence values, coincides with the experimentally identified β-arrestin binding site.[18] To evaluate whether arreSTick motifs generally represent GPCR-β-arrestin interaction sites, we have compiled a literature-based dataset containing GPCR mutations in the C-terminal region along with reported effects on β-arrestin subcellular trafficking (Table S3). We identified 21 receptors with published mutations demonstrated to disrupt class B-type β-arrestin binding. Remarkably, class-changing mutations in 19 out of 21 receptors affected the predicted arreSTick region, supporting the feasibility of our approach. Thereafter, we predicted the presence of arreSTick motifs for all GPCRs, and the identified regions are shown in Figure 3. Intriguingly, aminergic and muscarinergic receptors of the rhodopsin family of GPCRs mainly contain the arreSTick pattern in their ICL3 loop, while, in most other cases, it is located in the C-terminal region. To assess the conservation of the arreSTick motif, we calculated the evolutionary conservation of the amino acids within and outside of this motif in the ICL3 and C-terminal regions of GPCRs (Figure S6). Amino acids within arreSTick exhibit more conservation than those outside, underlying the important role of this pattern.

Phosphorylated arreSTick is sufficient for β-arrestin binding

We have previously demonstrated that AT1R does not require the active state of the receptor to bind β-arrestins;

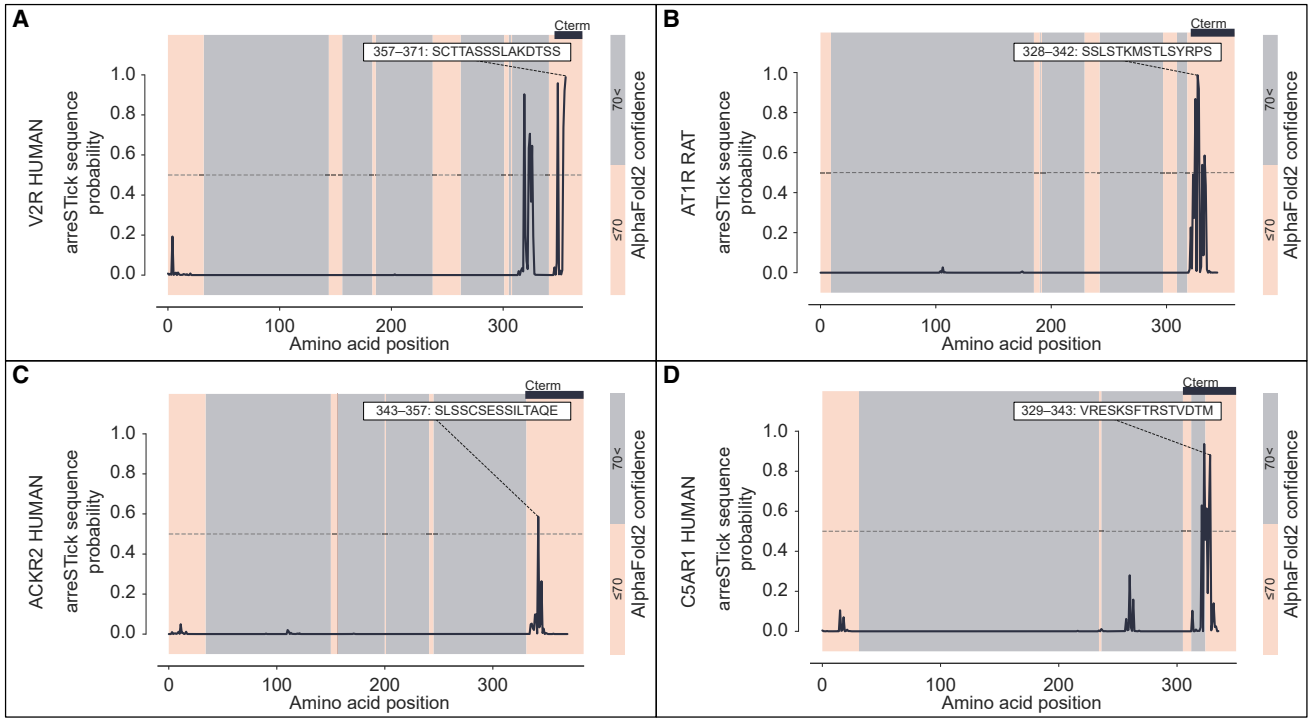


Figure 2. Convolution-based mapping of the arreSTick sequence for individual receptors

Full-length protein sequences of V2R (A), AT1R (B), ACKR2 (C), and C5AR1 (D) were passed into a trained grouped model with a kernel length of 15. Sigmoid transformation with the model's weights was carried out on the convolutional output to get probability values at each position in the protein sequence. The 15-amino-acid-long sequence regions (arreSTick motifs) with the maximal likelihood of β-arrestin binding are shown, and the AlphaFold2 model confidences (>70 or ≤70 corresponding to high and low confidence, respectively) are highlighted.

phosphorylation of the C terminus by protein kinase C (PKC) is sufficient to trigger this interaction.[17] Moreover, *in vitro* studies have shown that the phosphorylated C termini of different GPCRs can also bind to β-arrestins without the involvement of the receptor core.[16,18,21,29] Therefore, we sought to determine whether phosphorylated peptides alone could interact with β-arrestin2 (βarr2) in living cells. To experimentally investigate this, we designed a bioluminescence resonance energy transfer (BRET)-based setup, in which the interaction between RLuc8-labeled βarr2 and Venus-tagged C termini of GPCRs, without the seven transmembrane structures, was assessed. To facilitate the phosphorylation of the receptor termini, they were targeted to the plasma membrane, where GPCR-phosphorylating kinases are more abundant (Figures 4A–4C). The phosphorylation of the C-terminal peptide of AT1R (AT1R-Cterm) was induced by treatment with the PKC-activator phorbol 12-myristate 13-acetate (PMA), while coexpression of G protein-coupled receptor kinase 5 (GRK5) was applied for the V2R C terminus (V2R-Cterm). Both strategies are known to promote activation-independent phosphorylation of these receptors.[17,30,31] As shown in Figure 4D, the BRET signal between βarr2-RLuc8 and AT1R-Cterm-Venus increased after PMA stimulation. Angiotensin II (AngII) had no effect, as the construct lacks the transmembrane regions responsible for AngII binding. The binding between AT1R and βarr2 is stabilized by the interactions that we referred to as the "stability lock" in an earlier study.[17] A high-affinity binding is formed between amino acids K11 and K12 in the N-domain of βarr2 and the phosphorylated S/T residues in the receptor C terminus.[16,18,21] In contrast, the K11,12A (K2A)-mutant βarr2 is unable to establish this high-affinity interaction with GPCRs.[17,32,33] When we repeated the previous experiment using the phosphorylation-deficient (and arreSTick-motif-deficient) TSTS/A-mutant AT1R C terminus (Figure 4E), or the phosphate-binding-deficient K2A-mutant βarr2 (Figure 4F), PMA treatment did not affect the BRET signal. The coexpression of GRK5 led to an increase of the BRET ratio between wild-type (WT)-βarr2-RLuc8 and V2R-Cterm-Venus, which was not observed when K2A-βarr2-RLuc8 was expressed (Figure S7A). As expected, vasopressin stimulation induced no interaction since the receptor core is missing from the construct. Highly similar results were obtained when βarr1 was tested instead of βarr2 for both the AT1R and V2R C termini (Figures 4G–4I and S7B). These data show that phosphorylated GPCR C termini without the presence of other receptor regions can be sufficient for β-arrestin recruitment in living cells.

Cytoplasmic and nuclear proteins harboring the arreSTick pattern interact with β-arrestin2 through its cognate phosphate-binding residues

Our results imply that, if a protein contains an arreSTick pattern exposed to relevant kinases, the protein may also bind β-arrestins in a similar manner to GPCR C termini. Therefore, we investigated whether the arreSTick pattern is also present in proteins other than membrane receptors. We applied the arreSTick

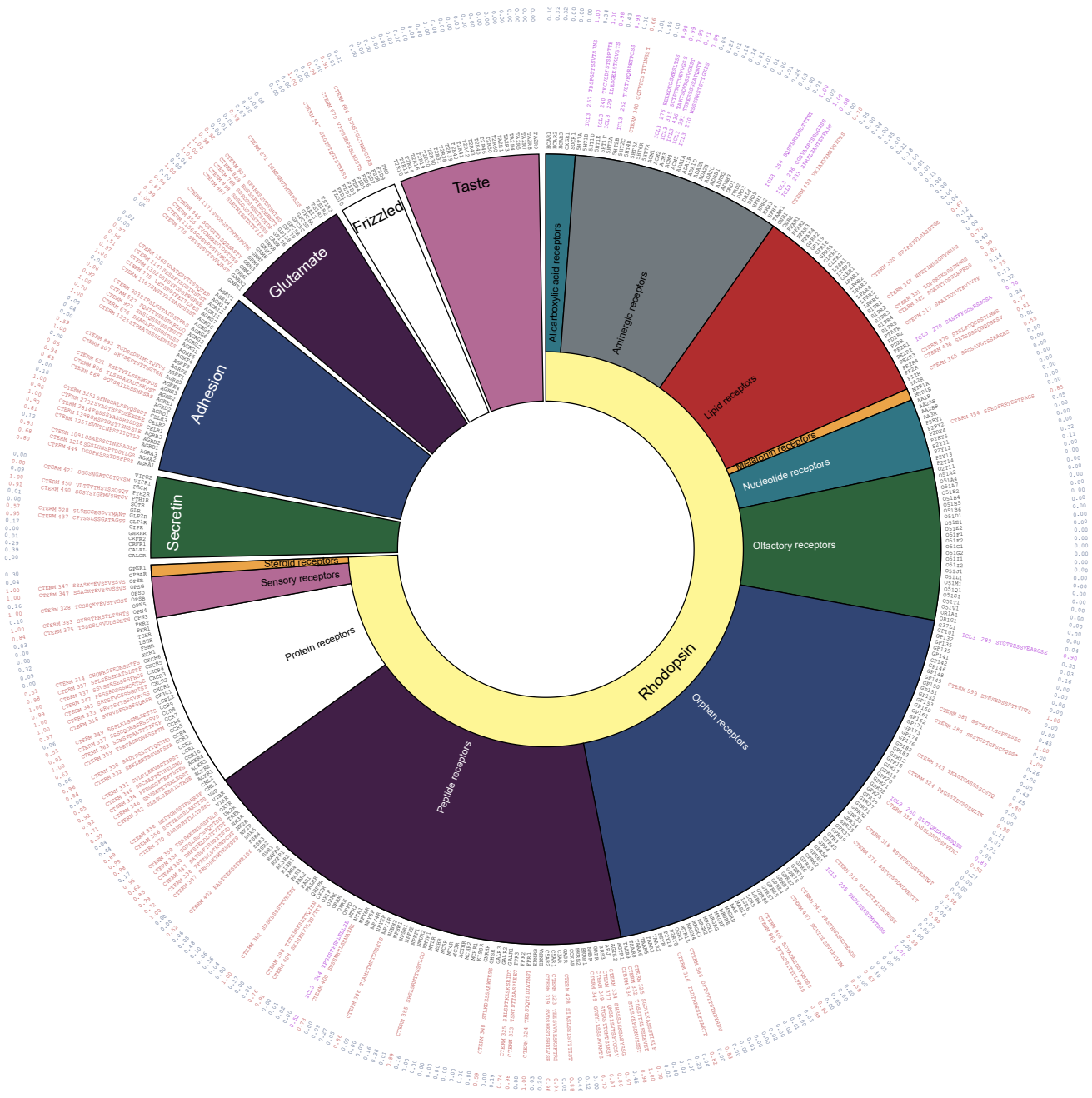


Figure 3. The prevalence of arreSTick in the human GPCRome

Prediction of the presence of arreSTick motif for human GPCRs. The sequence information of C termini and ICL3 regions of GPCRs was collected from gpcrdb.org. The numbers on the outer circle indicate the highest probability scores within each receptor sequence, derived from the C terminus or ICL3 regions. For receptors that possess a motif with at least 0.5 probability score, the sequence of the motif is presented. Receptors predicted to have an arreSTick pattern in ICL3 or C-terminal regions are highlighted in purple or red, respectively. Asterisk (*) indicates that both ICL3 and C terminus were predicted to contain the arreSTick pattern.

prediction to all human proteins with known cellular locations according to the Human Protein Atlas[34] (Figure 4J). We excluded the protein segments that are predicted to have well-defined structures with AlphaFold2, since these regions are unlikely to bind β-arrestins in a similar manner to the unstructured C termini of GPCRs. Remarkably, more than 30% of all proteins were predicted to contain the β-arrestin-binding arreSTick pattern, while more than 40% of the nuclear proteins possess this potential binding site. On the other hand, mitochondrial proteins only contain arreSTick motifs in around 15%. Interestingly, among

CellPress
OPEN ACCESS

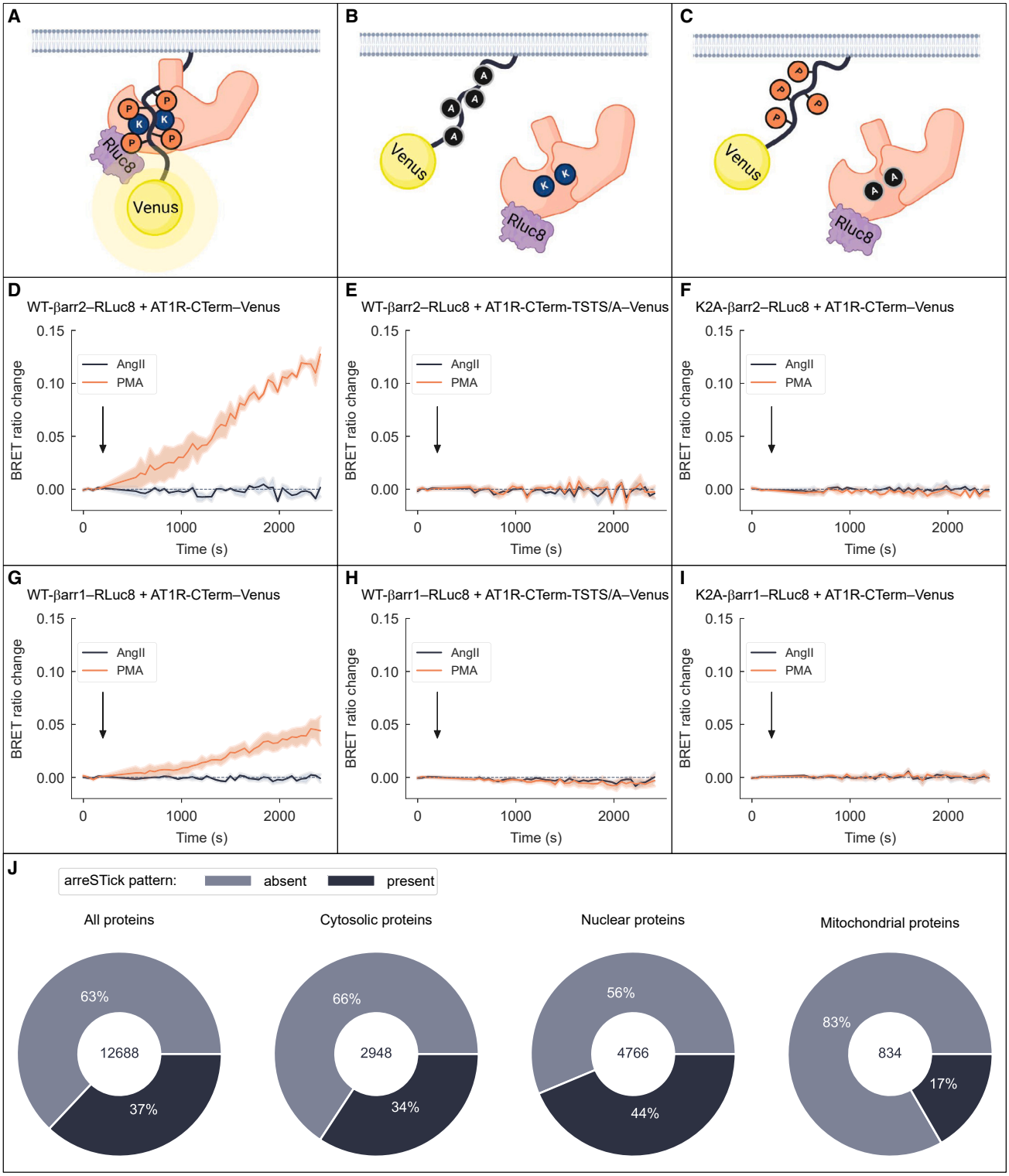


Figure 4. GPCR C termini with phosphorylated arreSTick pattern can recruit β-arrestin2 even in the absence of the GPCR core region

(A–C) Schematic representation of the BRET setups for interrogating the interaction of β-arrestins with receptor C termini.

(D and G) Kinetic BRET measurements between membrane-targeted AT1R-Cterm-Venus and WT-βarr2-RLuc8 (D) or WT-βarr1-RLuc8 (G). HEK293T cells were stimulated with vehicle, 100 nM AngII, or 100 nM PMA. BRET ratio changes reflect differences compared to control samples. The arrows indicate the time of treatment.

(legend continued on next page)

the well-documented non-GPCR interaction partners of β-arrestins, as reviewed by Peterson and Luttrell,[3] 51% feature an arreSTick motif (Table S4). The high number of arreSTick motifs in the human proteome raises the intriguing possibility that some non-receptor proteins, if phosphorylated, may also utilize a similar phosphorylation-dependent mechanism for interacting with β-arrestins as GPCRs do.

To investigate this experimentally, we carried out a proximity labeling assay[35] for the interrogation of the βarr2 interactome. We designed a biotin ligase-related assay format, which exploits the fast kinetics and high activity of the TurboID ligase enzyme. We hypothesized that, if non-receptor proteins interact with β-arrestins by phosphorylated S/T amino acids, they should preferentially bind to the wild-type βarr2 over the K2A-mutant. Therefore, we used TurboID-fused WT-βarr2 or K2A-βarr2 to carry out proximity labeling in HEK 293T cells. We verified that the fusion of TurboID does not alter the subcellular localization of βarr2 by anti-βarr2 immunostaining (Figure S8). To test the experimental setup, we first coexpressed AT1R-Cterm-Venus with WT-βarr2-TurboID or K2A-βarr2-TurboID (Figure 5A). We pulled down the biotinylated proteins with NeutrAvidin beads[17] and measured AT1R-Cterm-Venus fluorescence corresponding to the magnitude of the interaction between the C terminus and βarr2. PMA treatment increased the interaction with the AT1R-Cterm-Venus only in the case of the wild-type βarr2-TurboID, confirming that this setup is able to identify proteins that bind βarr2 through a phosphorylation-dependent mechanism.

Next, we applied this system to determine the entire phosphorylation-dependent interactome of βarr2 in HEK 293T cells (Figures 5B–5E). α1A-adrenergic receptor (α1AR), a GPCR that has no detectable β-arrestin binding,[17] was also coexpressed in these cells to activate a broad range of cytoplasmic kinases. Therefore, half of the samples were stimulated with an α1AR-selective agonist, A61603. After isolating the biotin-labeled proteins, the samples were analyzed with label-free quantitative mass spectrometry. Altogether, we detected 1,563 proteins across all samples (Table S5). We predicted the presence or absence of the arreSTick pattern in all eluted proteins (with the exclusion of their structured regions) and investigated how the presence of the arreSTick shaped the preference of the detected proteins toward WT-βarr2 with or without α1AR stimulation. The proteins without arreSTick regions had no preference for either of the βarr2s; however, the proteins containing this sequence had increased binding to the WT-βarr2 (Figure 5C). These results highlight the significance of the arreSTick sequence in facilitating protein binding to the positively charged N-domain region of βarr2. Interestingly, there was no difference between the stimulated and unstimulated samples, suggesting that the phosphorylation of arreSTick motifs for most proteins was not increased by α1AR stimulation. Next, we checked whether the proteins with previously described βarr2-binding patterns would also have preference to WT-βarr2. First, we compared those proteins that contain at least one phosphorylation code pattern[19] in their sequence versus the ones without such a feature. Proteins with either a short or a long code showed similar tendencies to the ones with an arreSTick pattern (Figure 5D), although the difference was less pronounced. Proteins containing a phosphocode slightly increased their preference for WT-βarr2 upon α1AR stimulation, suggesting that kinases activated by α1AR stimulation may phosphorylate parts of the phosphocodes recognized by the N-domain of βarr2. Proteins containing the PxPP sequence, a motif previously demonstrated to play a role in the activation of β-arrestin molecules,[18,21] also showed preference for WT-βarr2 (Figure 5E). These data suggest that non-receptor proteins may not only bind to but also participate in the activation of β-arrestins. The top 10 proteins containing the arreSTick pattern with the highest preference for WT-βarr2 are shown in Figure 5F. Next, we checked the cellular location of the WT-βarr2 preferring proteins containing the pattern according to the Human Protein Atlas.[34] Most of the proteins were nuclear, with a few with cytoplasmic and other locations (Figure 5G). These suggest that the arreSTick motif regulates the coupling of non-receptor proteins to βarr2 and that the interaction is dependent on the phosphoserine- and phosphothreonine-binding region of the β-arrestins, similarly to the mechanism well known for GPCRs.

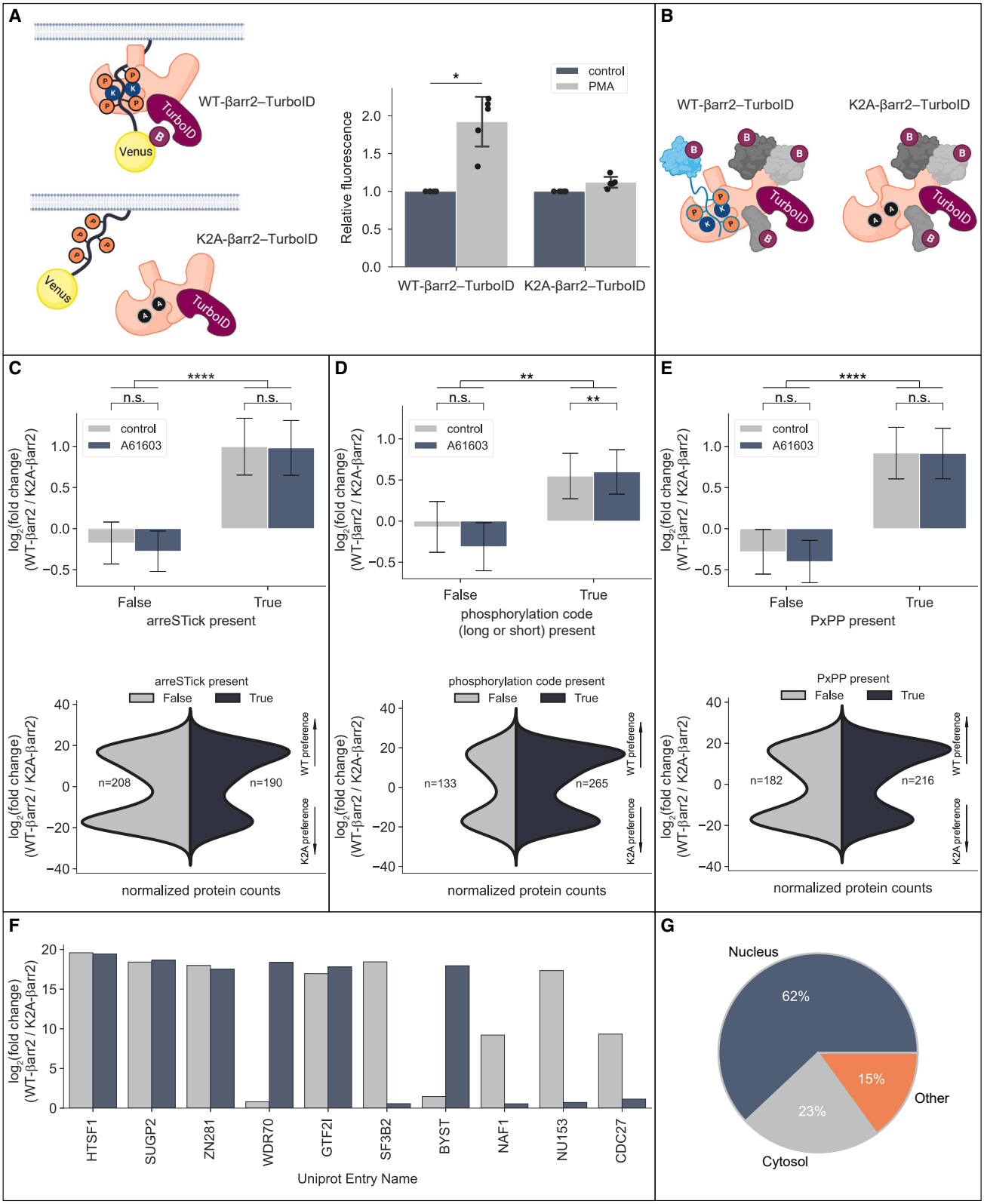
HIV Tat-specific factor 1 interacts with β-arrestin2 through its arreSTick region

Our mass spectrometry analysis revealed that the transcription factor HIV Tat-specific factor 1 (HTSF1 or HTATSF1) exhibited the highest preference for wild-type βarr2. A previous proteomic analysis identified HTSF1 among the proteins that immunoprecipitate with βarr2,[36] but the role of its phosphorylation in the interaction has not been studied. Our grouped prediction model predicts that HTSF1 contains an arreSTick pattern in the C-terminal part of HTSF1 (Figure 6A). This region has been previously reported to be phosphorylated.[37–39] To test the potential phosphorylation-dependent interaction of HTSF1 with both β-arrestin subtypes, we expressed wild-type or K2A-mutant Venus-labeled βarr1 or βarr2 proteins in HEK 293T cells and performed an immunoprecipitation assay. We were able to pull down endogenously expressed HTSF1 using both Venus-labeled βarr1 and βarr2, but not with the K2A mutants (Figure 6B). To further explore the specific interaction between HTSF1 and βarr2, we performed BRET measurements between βarr2-Rluc8 and HTSF1-Venus proteins. We opted for a titration BRET experiment to be able to distinguish the specific interaction from the non-specific energy transfer signal.[40] To verify the role of the

(E, F, H, and I) The effect of PKC stimulation is abolished if an arreSTick motif-disrupted receptor C terminus (AT1R-Cterm-TSTS/A-Venus, E and H) or phosphate-binding deficient β-arrestin mutants (K2A-βarr2-RLuc8, F, or K2A-βarr1-RLuc8, I) were applied.

(J) The arreSTick pattern is present in non-GPCR proteins. Protein intracellular localization data were downloaded from proteinatlas.org, and protein sequences were analyzed with the grouped model. The sequence regions with predicted secondary structures were masked before analysis.

In (D)–(I), data are mean ± SEM. One-sample t tests were performed to statistically test whether the average changes in BRET ratio after stimulation are significantly different from 0, $n = 3$ (D–F) or 4 (G–I) independent experiments performed in triplicate. (D) PMA, $^{**}p = 0.0078$; AngII, n.s., $p = 0.3368$; (E) PMA, n.s., $p = 0.8658$; AngII, n.s., $p = 0.0663$; (F) PMA, n.s., $p = 0.1328$; AngII, n.s., $p = 0.8449$; (G) PMA, $^{*}p = 0.0178$; AngII, n.s., $p = 0.4101$; (H) PMA, n.s., $p = 0.1723$; AngII, n.s., $p = 0.0603$; (I) PMA, n.s., $p = 0.7235$; AngII, n.s., $p = 0.8502$.



(legend on next page)

predicted arreSTick pattern in HTSF1, we mutated the S/T residues in this region to alanines (HTSF1-ST/AA, Figure 6C). As shown in Figure 6D, wild-type HTSF1 interacted with wild-type βarr2, reflected by the saturating BRET signal. In contrast, no BRET signal was observed when either the phosphorylation-deficient HTSF1-ST/AA-Venus or the phosphate-binding-deficient K2A-βarr2-Rluc8 mutant was expressed. These data show that the arreSTick region is involved in the interaction between HTSF1 and βarr2.

βarr2 regulates the intracellular location of HTSF1

βarr2, unlike βarr1, harbors a nuclear export signal,[41] raising a possible function of this interaction in the regulation of the intracellular location of the HTSF1 protein, similar to what was already reported for some other βarr2 partner proteins, such as Mdm2 and JNK3.[42,43] Therefore, we hypothesized that βarr2 may play a similar role in the case of HTSF1, and the nucleocytoplasmic transport may be dependent on the interaction between the arreSTick pattern and the positive charges in βarr2. To investigate this, we have coexpressed HTSF1-WT-mNeonGreen or HTSF1-ST/AA-mNeonGreen with either WT-βarr2 or K2A-βarr2 in βarr1/2-KO HEK 293A cells.[44] In cells expressing only the HTSF1-WT-mNeonGreen, HTSF1 localized mainly to the nucleus. When we coexpressed WT-βarr2, HTSF1 localization shifted toward the cytoplasm (Figure 6E). For the automated and unbiased quantification of the subcellular localization of HTSF1-mNeonGreen in each individual cell, we took advantage of ImageXpress high-throughput fluorescence microscopy and a deep learning-based cellular segmentation algorithm, Cellpose[45] (Figures 6F and S9A, and S9B). As shown in Figure 6F, HTSF1-WT-mNeonGreen cytoplasmic-to-nuclear fluorescence ratio was increased upon coexpression of WT-βarr2 with HTSF1-WT-mNeonGreen. However, this increased cytoplasmic localization was not detected with WT-βarr1, K2A-βarr2, or when HTSF1-ST/AA-mNeonGreen was overexpressed. Similarly, if we immunolabeled endogenous HTSF1 (Figure S9C), the cytoplasmic-to-nuclear fluorescence ratio increased when WT-βarr2 was coexpressed but not when βarr1 or K2A-βarr2 was coexpressed (Figure 6G). These data suggest that the interaction between the arreSTick motif and βarr2 modulates the intracellular location of HTSF1 and non-receptor proteins may undergo phosphorylation-dependent regulation by β-arrestins similar to GPCRs.

DISCUSSION

GPCRs bind to one of their main interaction partners, β-arrestins, with varying strength and dynamics, which substantially influences receptor trafficking and signaling.[46] Here, we applied convolutional neural networks to identify protein sequence requirements in receptors that facilitate stable interactions persisting after endosomal trafficking from the plasma membrane. To avoid overfitting on our relatively small set of training examples, we opted for a simpler network structure featuring only one hidden layer, one-dimensional embedding, and a single convolutional filter. Furthermore, we reduced the number of amino acid features by grouping amino acids based on their embedding values in the AA model. Despite its simplicity, this model achieved over 0.9 ROC AUC values in our CV strategy. A significant advantage of using a single kernel and convolutional layer is the improved interpretability of our model, which allows greater insight into the classification process. For instance, we can pinpoint the exact region based on which the model classifies the receptors, which likely corresponds to direct β-arrestin interaction sites. In accordance with that, the identified arreSTick motifs in receptors greatly overlap with experimentally determined β-arrestin-binding regions (Figure 2; Table S3). The simplistic structure also allows to explore the kernel for hints about which positions in the region are more important. While the exact kernels may differ in the same model structure when trained multiple times due to randomly initialized kernel weights, we observed only minor variations when the training was repeated on several occasions

Figure 5. Identification of phosphorylation-dependent β-arrestin2 partners with proximity biotinylation assay

(A) Proof-of-concept measurement scheme for the βarr2-TurboID-based system using the arreSTick-containing AT1R-Cterm peptide. In HEK293T cells, either WT-βarr2-TurboID or K2A-βarr2-TurboID was coexpressed with the membrane-targeted AT1R-Cterm-Venus. Upon binding of TurboID-labeled βarr2, AT1R-Cterm-Venus is biotinylated, enabling its pull-down using NeutrAvidin beads. Fluorescence of the pulled-down Venus-labeled C termini was determined. Interaction was induced by the PKC-activator PMA (100 nM). Data were normalized to the vehicle-treated conditions, expressed as mean ± SEM. Paired two-tailed t tests were performed on the raw data, $n = 5$ independent experiments performed in triplicate, WT-control vs. WT-PMA *, $p = 0.0457$; K2A-control vs. K2A-PMA n.s., $p = 0.2297$.

(B) Rationale of the proximity biotinylation assay-linked mass spectrometry experimental setup. The protein partners coupling through the phosphorylated arreSTick-K11/K12-βarr2 interaction (cyan colored) are expected to be overrepresented in the interactome of the WT-βarr2, while others (proteins represented with gray color) are expected to have no preference.

(C–E) Log2 fold difference between the proteins in the interactome of the WT-βarr2-TurboID and K2A-βarr2-TurboID. HEK293T cells were cotransfected with α1AR and wild-type or K2A-mutant βarr2-TurboID and were stimulated with vehicle or the α1AR agonist A61603 (1 μM) for 1 h. Proteins were grouped based on either the grouped model prediction (C), the presence of at least one phosphorylation code (D), or the presence of the PxPP motif (E). The upper panels show the average log2 fold difference of all proteins between WT-βarr2-TurboID and K2A-βarr2-TurboID samples, in control or stimulated samples. To test whether the presence of the pattern has significant effect, two-tailed Mann-Whitney tests were performed (C and E, ****$p < 0.0001$; D, **$p = 0.005$). The effect of stimulation was analyzed using Kruskal-Wallis test with Dunn's *post hoc* test (C, false control, $n = 880$ vs. false A61603, $n = 888$, n.s., $p > 0.9999$; true control, $n = 601$ vs. true A61603, $n = 601$, n.s., $p = 0.1484$. D, false control, $n = 593$ vs. false A61603, $n = 599$, n.s., $p > 0.9999$; true control, $n = 888$, vs. true A61603, $n = 890$, **$p = 0.0072$. E, false control, $n = 764$ vs. false A61603, $n = 773$, n.s., $p > 0.9999$; true control, $n = 717$ vs. true A61603, $n = 716$, n.s., $p = 0.0757$). The violin plots (bottom panels) show the distribution of the proteins with at least 2-fold difference between WT-βarr2-TurboID and K2A-βarr2-TurboID control, unstimulated samples. The width of violins is scaled by the number of observations in each group. The number of proteins in each group is indicated.

(F) Log2 fold difference of the 10 most differentially interacting proteins between the WT-βarr2-TurboID and K2A-βarr2-TurboID samples in any of the stimulated/unstimulated samples.

(G) Subcellular location of the proteins with arreSTick pattern and statistically significant preference for the WT-βarr2-TurboID.

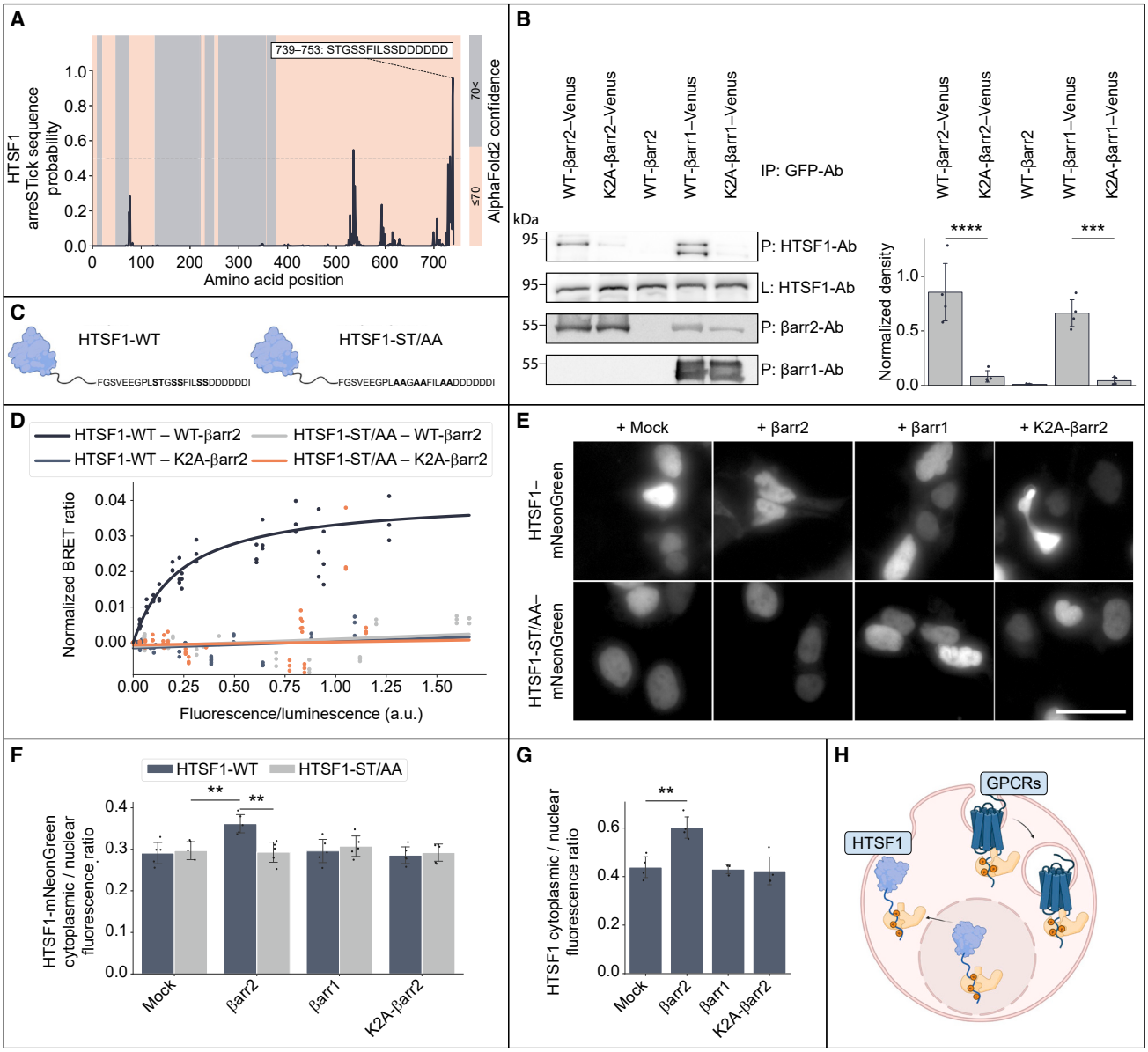


Figure 6. β-arrestin interacts with HTSF1 and determines its subcellular location

(A) Identification of the arreSTick pattern in HTSF1.

(B) Immunoprecipitation of endogenous HTSF1 with overexpressed WT-βarr2-Venus, K2A-βarr2-Venus, WT-βarr1-Venus, or K2A-βarr1-Venus in HEK293T cells. Non-labeled WT-βarr2 was used as a control (middle lane). The precipitation (P) was carried out with an anti-GFP antibody, and the blots were stained with an anti-HTSF1 antibody, anti-βarr2, or anti-βarr1 antibodies. Lysates (L) were stained for HTSF1 (second row). $n = 4$ independent experiments, without technical replicates; data are mean ± SEM; WT-βarr2-Venus vs. K2A-βarr2–Venus, $^{****}p < 0.0001$; WT-βarr1-Venus vs. K2A-βarr1-Venus, $^{***}p = 0.0003$ (one-way ANOVA with Tukey's *post hoc* test).

(C) Sequence of the arreSTick-containing C-terminal region of HTSF1 and the mutations in the HTSF1-ST/AA construct.

(D) BRET titration experiments with coexpressed Venus-labeled HTSF1 and RLuc8-labeled βarr2 in HEK293T cells. BRET ratios were normalized to that of cells transfected only with βarr2-Rluc8. BRET ratios of individual wells from three independent experiments performed in triplicate; a one-site specific binding curve was fitted on the HTSF1-WT + WT-βarr2 data. Since this equation resulted in ambiguous fits for the other conditions, simple linear regression was used in those cases.

(E) Representative confocal images show the subcellular localization of HTSF1-mNeonGreen in βarr1/2-KO HEK293A cells coexpressing the indicated constructs (see also Figure S9A). The images were gamma corrected with a value of 0.5 for better visualization of the cytoplasmic HTSF1 localization. Scale bar represents 35 μm.

(F and G) Quantification of the distribution of HTSF1 in βarr1/2-KO HEK293A cells. HTSF1 was either labeled with mNeonGreen (F) or the endogenous HTSF1 was immunostained (G), and cytoplasmic/nuclear fluorescence ratios were calculated. Data are mean ± SEM, analyzed by two-way ANOVA with Tukey's *post hoc*

(legend continued on next page)

(Figures S3–S5). The kernels suggest that some amino acid positions seem to have greater importance than others (Figure 1G), corresponding to positions that may be sterically available for binding to the positively charged amino acids on the N-domain of β-arrestins.[18,19] However, the position of the phosphorylated amino acids required for the strong interaction is not strictly determined. Since there are a number of possible positively charged amino acid partners in β-arrestins,[18] these variations may lead to slightly different β-arrestin binding and active conformations, consistent with the "barcode theory."[47–49] Variations of the phosphorylation-specific micro-locks may lead to distinct β-arrestin activations and signaling outcomes.[50,51]

Interestingly, certain amino acids seem to have a negative impact on the stable interaction (Figure 1C). Although it may be expected in the case of the positively charged arginine and lysine or the amino-group-containing glutamine, it might be surprising in the case of alanine. While the exact reason for this is unknown, it is possible that alanines interfere with the phosphorylation of nearby residues. Indeed, alanine is not among the preferred amino acids within the phosphorylation motif for GRKs.[52] It is also noteworthy that glutamate and aspartate, which are negatively charged and have been suggested as part of the phosphorylation code,[19] had a neutral impact on the classification within our binding motif (Figure 1C).

During multiple trainings, we found that the vast majority of receptors are consistently categorized into the same class even with slightly varying kernels. However, certain receptors were frequently categorized into different groups with distinct kernels, which resulted in their occasional misclassification. For example, the CB1 cannabinoid receptor, a known class A receptor,[5] was classified as class B in more than half of the cases, while the B2 bradykinin receptor, a class B receptor,[53] was often predicted as a class A receptor. Interestingly, our model consistently predicted arreSTick motifs in metabotropic glutamate receptors, despite some of these receptors being demonstrated in a study published after the construction of our database to bind weakly or not at all to β-arrestins upon glutamate activation.[54] Nonetheless, it is conceivable that phosphorylation and class B-type binding could occur even within this receptor class in certain cellular contexts, with other agonists or available GRKs. Indeed, receptors may exhibit variable β-arrestin binding affinity in response to different agonists, and it is possible that class B-type binding is only elicited by one ligand but not another.[55–58] One potential underlying mechanism is the different activation of receptor kinases of biased and nonbiased agonists,[47] as the ligand-specific G protein coupling profile significantly influences which GRKs are activated.[59] Additionally, the dissociation rate constants of agonists have been demonstrated to play a crucial role in the stability of receptor-β-arrestin interaction by regulating endosomal β-arrestin binding.[60]

In a previous study, we established that phosphorylation of the arreSTick motif in AT1R by PKC is sufficient to facilitate βarr2 binding to the inactive state of AT1R.[17] Upon PKC activity, βarr2 adopts an active conformation that is different from its conformation when bound to the active receptor-bound state, with significant functional implications, including the promotion of AT1R endocytosis. Here, we extended this observation by demonstrating that the receptor core is entirely dispensable for the PKC-induced β-arrestin binding to AT1R C terminus. Conversely, the stability lock,[17] formed between phosphorylated residues and conserved phosphate-binding lysine residues in the N-domain of β-arrestins, is crucial for the phosphorylation-dependent but core-independent β-arrestin recruitment.

While a primary role of β-arrestins is believed to be the regulation of GPCRs, many GPCR-independent functions have also been described.[61,62] β-arrestins are considered to bind GPCRs and signaling effector proteins via distinct interaction sites.[3] We hypothesized that, if other proteins contained the arreSTick pattern, they may also bind β-arrestins in a similar manner to GPCRs, suggesting the potential existence of a β-arrestin-dependent regulatory mechanism. Unexpectedly, we found that a substantial proportion of non-GPCR proteins possess patterns that may be capable of binding to β-arrestin with high affinity, given they are phosphorylated (Figure 4I). Indeed, the presence of the arreSTick pattern in non-GPCR proteins led to a preference for the wild-type βarr2 over the phosphate-binding deficient K2A mutant in the βarr2 interactome.

We further analyzed the binding of the protein with highest preference, HTSF1. HTSF1 has been initially recognized as protein regulating the gene expression of human immunodeficiency virus type 1 (HIV-1),[63–65] and it might also be involved in the formation of metastases.[66] The phosphorylation of the HTSF1 in the predicted arreSTick region has been previously reported.[37–39] In line with that, we found that the binding of β-arrestins to HTSF1 was dependent on both the arreSTick pattern and the K11 and K12 amino acids, suggesting an interaction similar to that observed with GPCRs. We demonstrated that the binding of βarr2 to HTSF1 induced its translocation from the nucleus to the cytoplasm, implying a potential role of βarr2 in the regulation of this transcription factor. This translocation appears analogous to the role of β-arrestins with GPCRs, wherein the phosphorylated protein is removed from its primary site of action and transported to another compartment (Figure 6H). In the case of GPCRs, the active phosphorylated receptors bind to β-arrestins and undergo internalization into intracellular vesicles while being desensitized during this process. For HTSF1, the nucleus is the primary site of action, and the phosphorylated HTSF1 might be transported into the cytoplasm when coupled to βarr2. Notably, βarr1 was not able to induce a similar translocation, consistent with the absence of a nuclear export signal on this protein.[42]

In summary, we established a method for the accurate prediction of the amino acid pattern required for stable interaction between phosphorylated GPCRs and β-arrestins. Such regions

test. (F) $n = 5$ independent experiments, without technical replicates; For HTSF1-WT-Mock vs. HTSF1-WT-βarr2, $^{**}p = 0.0037$; for HTSF1-WT-βarr2 vs. HTSF1-ST/AA-βarr2, $^{**}p = 0.0053$. (G) $n = 4$ independent experiments, without technical replicates; $^{**}p = 0.0032$.
(H) Model of the arreSTick function in protein-protein interactions involving βarr2. For GPCRs, βarr2 binding, stabilized by the arreSTick motif, leads to receptor internalization and compartment change. In the case of HTSF1, the binding to βarr2 through the arreSTick motif also leads to a compartmental shift, relocating HTSF1 from the nucleus to the cytoplasm.

are present not only in GPCRs but also in other proteins, in which they may provide interaction sites for β-arrestins. These findings suggest that the role of β-arrestins in regulating phosphorylated proteins may be more extensive than previously recognized.

Limitations of the study

Although our model identifies receptors with class B-type β-arrestin binding with high accuracy, prediction errors may occur; some amino acid features in the receptor sequence might not be captured or the training set itself may contain misclassified receptors, which interferes with the training process. Furthermore, it should be emphasized that the lack of an arreSTick motif in a receptor does not imply that it is unable to form phosphorylation-dependent but transient (class A-like) interaction with β-arrestins. To accurately capture the wide diversity of the strength and dynamics of GPCR-β-arrestin interactions with our machine learning approach, a larger and more complex training set would be required. Achieving that, however, necessitates more experimental data with consistent experimental readouts across studies and different receptor subtypes.

In our proximity labeling assays, the K2A mutant of βarr2 serves as a valuable tool to distinguish between βarr2 partners in the proteome that are dependent on the phosphorylated arreSTick pattern for the interaction and those that are not. However, this way we only detected proteins that are either constitutively phosphorylated at their arreSTick motifs in HEK 293 cells or whose phosphorylation can be induced by α1AR stimulation. For this reason, not all partners harboring arreSTick patterns are anticipated to bind βarr2 in our assay, but we expect the enrichment of true-positive binding partners in the WT-βarr2 interactome. On the other hand, arreSTick containing false-negative hits are also expected, due to the high basal biotinylation rate of TurboID enzyme,[67] which may result in an expanded list of protein partners of K2A-βarr2-TurboID.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell culture and transfection
- METHOD DETAILS
  - Database development
  - Convolutional neural network and protein predictions
  - Materials and plasmid constructs
  - Bioluminescence resonance energy transfer (BRET) measurements
  - Confocal microscopy
  - Immunoprecipitation and immunoblot analysis of HTSF1
  - Affinity purification using biotin ligase
  - Mass spectrometry
  - Mass spectrometry data analysis
  - Evolutionary conservation analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.114241.

ACKNOWLEDGMENTS

The technical assistance of Eszter Halász, Ilona Oláh, and Kata Szabolcsi is greatly appreciated. Figure drawings were created with BioRender.com. This work was supported by the Hungarian National Research, Development, and Innovation Fund (NKFI FK 138862, K 139231, and K 134357), the Hungarian National, Development & Innovation Office's TKP Fund (SEMTHREUM, TKP2021-EGA-29), Competitive Central Hungary Operational Programme VEKOP-2.3.2-16-2016-00002, and the János Bolyai Research Scholarship and János Bolyai Research Scholarship Plus of the Hungarian Academy of Sciences (BO/00807/21). Á.M. was supported by the Gedeon Richter Talentum Foundation in the framework of Gedeon Richter Excellence PhD Scholarship of Gedeon Richter. A.I. was funded by JP21H04791 and JP21H05113 from Japan Society for the Promotion of Science, JPMJFR215T and JPMJMS2023 from the Japan Science and Technology Agency, and JP22ama121038 and JP22zf0127007 from the Japan Agency for Medical Research and Development.

AUTHOR CONTRIBUTIONS

A.D.T., G.T., E.S.-K., and L.H. conceptualized the work. A.D.T. and S.P. collected the data for the GPCR-β-arrestin database. T.G. and Á.M. built the machine learning model. E.S.-K. and V.G. performed the proximity biotinylation and immunoprecipitation experiments. E.S.-K., K.K., G.T., and S.G. performed the microscopy experiments. R.B., E.S.-K., S.G., K.K., and A.D.T. performed the BRET measurements. Á.M. did the evolutionary conservation analysis. L.T., A.Á., and L.D. performed the mass spectrometry measurements. T.G. performed the mass spectrometry results analysis. A.I. provided the βarr1/2-knockout HEK 293A cells. All authors analyzed the results, wrote the manuscript, and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used OpenAI-ChatGPT and Grammarly services in order to enhance the language quality and correct grammatical errors in the manuscript. After using these tools/services, the authors thoroughly reviewed and edited the content and take full responsibility for the content of the publication.

Received: October 2, 2023
Revised: March 11, 2024
Accepted: May 1, 2024
Published: May 17, 2024

REFERENCES

1. Hauser, A.S., Attwood, M.M., Rask-Andersen, M., Schiöth, H.B., and Gloriam, D.E. (2017). Trends in GPCR drug discovery: new agents, targets and indications. Nat. Rev. Drug Discov. *16*, 829–842. https://doi.org/10.1038/nrd.2017.178.
2. Gurevich, V.V., and Gurevich, E.V. (2019). GPCR Signaling Regulation: The Role of GRKs and Arrestins. Front. Pharmacol. *10*, 125. https://doi.org/10.3389/fphar.2019.00125.
3. Peterson, Y.K., and Luttrell, L.M. (2017). The Diverse Roles of Arrestin Scaffolds in G Protein-Coupled Receptor Signaling. Pharmacol. Rev. *69*, 256–297. https://doi.org/10.1124/pr.116.013367.

4. Oakley, R.H., Laporte, S.A., Holt, J.A., Caron, M.G., and Barak, L.S. (2000). Differential Affinities of Visual Arrestin, βArrestin1, and βArrestin2 for G Protein-coupled Receptors Delineate Two Major Classes of Receptors. J. Biol. Chem. *275*, 17201–17210. https://doi.org/10.1074/jbc.M910348199.

5. Gyombolai, P., Boros, E., Hunyady, L., and Turu, G. (2013). Differential β-arrestin2 requirements for constitutive and agonist-induced internalization of the CB1 cannabinoid receptor. Mol. Cell. Endocrinol. *372*, 116–127. https://doi.org/10.1016/j.mce.2013.03.013.

6. Terrillon, S., Barberis, C., and Bouvier, M. (2004). Heterodimerization of V1a and V2 vasopressin receptors determines the interaction with beta-arrestin and their trafficking patterns. Proc. Natl. Acad. Sci. USA *101*, 1548–1553. https://doi.org/10.1073/pnas.0305322101.

7. Oakley, R.H., Laporte, S.A., Holt, J.A., Barak, L.S., and Caron, M.G. (2001). Molecular determinants underlying the formation of stable intracellular G protein-coupled receptor-beta-arrestin complexes after receptor endocytosis. J. Biol. Chem. *276*, 19452–19460. https://doi.org/10.1074/jbc.M101450200.

8. Tohgo, A., Choy, E.W., Gesty-Palmer, D., Pierce, K.L., Laporte, S., Oakley, R.H., Caron, M.G., Lefkowitz, R.J., and Luttrell, L.M. (2003). The stability of the G protein-coupled receptor-β-arrestin interaction determines the mechanism and functional consequence of ERK activation. J. Biol. Chem. *278*, 6258–6267. https://doi.org/10.1074/jbc.M212231200.

9. Wei, H., Ahn, S., Barnes, W.G., and Lefkowitz, R.J. (2004). Stable interaction between beta-arrestin 2 and angiotensin type 1A receptor is required for beta-arrestin 2-mediated activation of extracellular signal-regulated kinases 1 and 2. J. Biol. Chem. *279*, 48255–48261. https://doi.org/10.1074/jbc.M406205200.

10. 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., et al. (2014). Visualization of arrestin recruitment by a G-protein-coupled receptor. Nature *512*, 218–222. https://doi.org/10.1038/nature13430.

11. Maharana, J., Sano, F.K., Sarma, P., Yadav, M.K., Duan, L., Stepniewski, T.M., Chaturvedi, M., Ranjan, A., Singh, V., Saha, S., et al. (2024). Molecular insights into atypical modes of β-arrestin interaction with seven transmembrane receptors. Science *383*, 101–108. https://doi.org/10.1126/science.adj3347.

12. Yang, F., Yu, X., Liu, C., Qu, C.X., Gong, Z., Liu, H.D., Li, F.H., Wang, H.-M., He, D.F., Yi, F., et al. (2015). Phospho-selective mechanisms of arrestin conformations and functions revealed by unnatural amino acid incorporation and (19)F-NMR. Nat. Commun. *6*, 8202. https://doi.org/10.1038/ncomms9202.

13. Yang, F., Xiao, P., Qu, C.-X., Liu, Q., Wang, L.-Y., Liu, Z.-X., He, Q.-T., Liu, C., Xu, J.-Y., Li, R.-R., et al. (2018). Allosteric mechanisms underlie GPCR signaling to SH3-domain proteins through arrestin. Nat. Chem. Biol. *14*, 876–886. https://doi.org/10.1038/s41589-018-0115-3.

14. Liu, Q., He, Q.-T., Lyu, X., Yang, F., Zhu, Z.-L., Xiao, P., Yang, Z., Zhang, F., Yang, Z.-Y., Wang, X.-Y., et al. (2020). DeSiphering receptor core-induced and ligand-dependent conformational changes in arrestin via genetic encoded trimethylsilyl 1H-NMR probe. Nat. Commun. *11*, 4857. https://doi.org/10.1038/s41467-020-18433-5.

15. Kumari, P., Srivastava, A., Banerjee, R., Ghosh, E., Gupta, P., Ranjan, R., Chen, X., Gupta, B., Gupta, C., Jaiman, D., and Shukla, A.K. (2016). Functional competence of a partially engaged GPCR-β-arrestin complex. Nat. Commun. *7*, 13416. https://doi.org/10.1038/ncomms13416.

16. 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., et al. (2013). Structure of active β-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. Nature *497*, 137–141. https://doi.org/10.1038/nature12120.

17. Tóth, A.D., Prokop, S., Gyombolai, P., Várnai, P., Balla, A., Gurevich, V.V., Hunyady, L., and Turu, G. (2018). Heterologous phosphorylation–induced formation of a stability lock permits regulation of inactive receptors by β-arrestins. J. Biol. Chem. *293*, 876–892. https://doi.org/10.1074/jbc.M117.813139.

18. Maharana, J., Sarma, P., Yadav, M.K., Saha, S., Singh, V., Saha, S., Chami, M., Banerjee, R., and Shukla, A.K. (2023). Structural snapshots uncover a key phosphorylation motif in GPCRs driving β-arrestin activation. Mol. Cell *83*, 2091–2107.e7. https://doi.org/10.1016/j.molcel.2023.04.025.

19. Zhou, X.E., He, Y., de Waal, P.W., Gao, X., Kang, Y., Van Eps, N., Yin, Y., Pal, K., Goswami, D., White, T.A., et al. (2017). Identification of Phosphorylation Codes for Arrestin Recruitment by G Protein-Coupled Receptors. Cell *170*, 457–469.e13. https://doi.org/10.1016/j.cell.2017.07.002.

20. Mayer, D., Damberger, F.F., Samarasimhareddy, M., Feldmueller, M., Vuckovic, Z., Flock, T., Bauer, B., Mutt, E., Zosel, F., Allain, F.H.T., et al. (2019). Distinct G protein-coupled receptor phosphorylation motifs modulate arrestin affinity and activation and global conformation. Nat. Commun. *10*, 1261. https://doi.org/10.1038/s41467-019-09204-y.

21. Isaikina, P., Petrovic, I., Jakob, R.P., Sarma, P., Ranjan, A., Baruah, M., Panwalkar, V., Maier, T., Shukla, A.K., and Grzesiek, S. (2023). A key GPCR phosphorylation motif discovered in arrestin2•CCR5 phosphopeptide complexes. Mol. Cell *83*, 2108–2121.e7. https://doi.org/10.1016/j.molcel.2023.05.002.

22. Gurevich, V.V., and Gurevich, E.V. (2013). Structural determinants of arrestin functions. Prog. Mol. Biol. Transl. Sci. *118*, 57–92. https://doi.org/10.1016/B978-0-12-394440-5.00003-6.

23. Pándy-Szekeres, G., Caroli, J., Mamyrbekov, A., Kermani, A.A., Keserű, G.M., Kooistra, A.J., and Gloriam, D.E. (2023). GPCRdb in 2023: state-specific structure models using AlphaFold2 and new ligand resources. Nucleic Acids Res. *51*, D395–D402. https://doi.org/10.1093/nar/gkac1013.

24. Gurevich, V.V., and Gurevich, E.V. (2006). The structural basis of arrestin-mediated regulation of G-protein-coupled receptors. Pharmacol. Ther. *110*, 465–502. https://doi.org/10.1016/j.pharmthera.2005.09.008.

25. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

26. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. *50*, D439–D444. https://doi.org/10.1093/nar/gkab1061.

27. Pandey, S., Kumari, P., Baidya, M., Kise, R., Cao, Y., Dwivedi-Agnihotri, H., Banerjee, R., Li, X.X., Cui, C.S., Lee, J.D., et al. (2021). Intrinsic bias at non-canonical, β-arrestin-coupled seven transmembrane receptors. Mol. Cell *81*, 4605–4621.e11. https://doi.org/10.1016/j.molcel.2021.09.007.

28. Dwivedi-Agnihotri, H., Chaturvedi, M., Baidya, M., Stepniewski, T.M., Pandey, S., Maharana, J., Srivastava, A., Caengprasath, N., Hanyaloglu, A.C., Selent, J., and Shukla, A.K. (2020). Distinct phosphorylation sites in a prototypical GPCR differently orchestrate β-arrestin interaction, trafficking, and signaling. Sci. Adv. *6*, eabb8368. https://doi.org/10.1126/sciadv.abb8368.

29. Kumari, P., Srivastava, A., Ghosh, E., Ranjan, R., Dogra, S., Yadav, P.N., and Shukla, A.K. (2017). Core engagement with β-arrestin is dispensable for agonist-induced vasopressin receptor endocytosis and ERK activation. Mol. Biol. Cell *28*, 1003–1010. https://doi.org/10.1091/mbc.E16-12-0818.

30. Li, L., Homan, K.T., Vishnivetskiy, S.A., Manglik, A., Tesmer, J.J.G., Gurevich, V.V., and Gurevich, E.V. (2015). G Protein-coupled Receptor Kinases of the GRK4 Protein Subfamily Phosphorylate Inactive G Protein-coupled Receptors (GPCRs). J. Biol. Chem. *290*, 10775–10790. https://doi.org/10.1074/jbc.M115.644773.

31. Drube, J., Haider, R.S., Matthees, E.S.F., Reichel, M., Zeiner, J., Fritzwanker, S., Ziegler, C., Barz, S., Klement, L., Filor, J., et al. (2022). GPCR kinase knockout cells reveal the impact of individual GRKs on arrestin binding and GPCR regulation. Nat. Commun. *13*, 540. https://doi.org/10.1038/s41467-022-28152-8.

32. 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. https://doi.org/10.1074/jbc.M111.311803.

33. 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. https://doi.org/10.1074/jbc.M007159200.

34. Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. Science *356*, eaal3321. https://doi.org/10.1126/science.aal3321.

35. Branon, T.C., Bosch, J.A., Sanchez, A.D., Udeshi, N.D., Svinkina, T., Carr, S.A., Feldman, J.L., Perrimon, N., and Ting, A.Y. (2018). Efficient proximity labeling in living cells and organisms with TurboID. Nat. Biotechnol. *36*, 880–887. https://doi.org/10.1038/nbt.4201.

36. Xiao, K., McClatchy, D.B., Shukla, A.K., Zhao, Y., Chen, M., Shenoy, S.K., Yates, J.R., and Lefkowitz, R.J. (2007). Functional specialization of beta-arrestin interactions revealed by proteomic analysis. Proc. Natl. Acad. Sci. USA *104*, 12011–12016. https://doi.org/10.1073/pnas.0704849104.

37. Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. *40*, D261–D270. https://doi.org/10.1093/nar/gkr1122.

38. Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci. Signal. *3*, ra3. https://doi.org/10.1126/scisignal.2000475.

39. Ruse, C.I., McClatchy, D.B., Lu, B., Cociorva, D., Motoyama, A., Park, S.K., and Yates, J.R. (2008). Motif-specific sampling of phosphoproteomes. J. Proteome Res. *7*, 2140–2150. https://doi.org/10.1021/pr800147u.

40. Marullo, S., and Bouvier, M. (2007). Resonance energy transfer approaches in molecular pharmacology and beyond. Trends Pharmacol. Sci. *28*, 362–365. https://doi.org/10.1016/j.tips.2007.06.007.

41. Scott, M.G.H., Le Rouzic, E., Périanin, A., Pierotti, V., Enslen, H., Benichou, S., Marullo, S., and Benmerah, A. (2002). Differential nucleocytoplasmic shuttling of beta-arrestins. Characterization of a leucine-rich nuclear export signal in beta-arrestin2. J. Biol. Chem. *277*, 37693–37701. https://doi.org/10.1074/jbc.M207552200.

42. Wang, P., Wu, Y., Ge, X., Ma, L., and Pei, G. (2003). Subcellular localization of beta-arrestins is determined by their intact N domain and the nuclear export signal at the C terminus. J. Biol. Chem. *278*, 11648–11653. https://doi.org/10.1074/jbc.M208109200.

43. McDonald, P.H., Chow, C.W., Miller, W.E., Laporte, S.A., Field, M.E., Lin, F.T., Davis, R.J., and Lefkowitz, R.J. (2000). Beta-arrestin 2: a receptor-regulated MAPK scaffold for the activation of JNK3. Science *290*, 1574–1577. https://doi.org/10.1126/science.290.5496.1574.

44. O'Hayre, M., Eichel, K., Avino, S., Zhao, X., Steffen, D.J., Feng, X., Kawakami, K., Aoki, J., Messer, K., Sunahara, R., et al. (2017). Genetic evidence that β-arrestins are dispensable for the initiation of β2-adrenergic receptor signaling to ERK. Sci. Signal. *10*, eaal3395. https://doi.org/10.1126/scisignal.aal3395.

45. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods *18*, 100–106. https://doi.org/10.1038/s41592-020-01018-x.

46. Oakley, R.H., Laporte, S.A., Holt, J.A., Barak, L.S., and Caron, M.G. (1999). Association of β-Arrestin with G Protein-coupled Receptors during Clathrin-mediated Endocytosis Dictates the Profile of Receptor Resensitization. J. Biol. Chem. *274*, 32248–32257. https://doi.org/10.1074/jbc.274.45.32248.

47. 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., et al. (2011). Distinct Phosphorylation Sites on the β_2-Adrenergic Receptor Establish a Barcode That Encodes Differential Functions of β-Arrestin. Sci. Signal. *4*, ra51. https://doi.org/10.1126/scisignal.2001707.

48. Butcher, A.J., Prihandoko, R., Kong, K.C., McWilliams, P., Edwards, J.M., Bottrill, A., Mistry, S., and Tobin, A.B. (2011). Differential G-protein-coupled receptor phosphorylation provides evidence for a signaling bar code. J. Biol. Chem. *286*, 11506–11518. https://doi.org/10.1074/jbc.M110.154526.

49. Latorraca, N.R., Masureel, M., Hollingsworth, S.A., Heydenreich, F.M., Suomivuori, C.-M., Brinton, C., Townshend, R.J.L., Bouvier, M., Kobilka, B.K., and Dror, R.O. (2020). How GPCR Phosphorylation Patterns Orchestrate Arrestin-Mediated Signaling. Cell *183*, 1813–1825.e18. https://doi.org/10.1016/j.cell.2020.11.014.

50. He, Q.-T., Xiao, P., Huang, S.-M., Jia, Y.-L., Zhu, Z.-L., Lin, J.-Y., Yang, F., Tao, X.-N., Zhao, R.-J., Gao, F.-Y., et al. (2021). Structural studies of phosphorylation-dependent interactions between the V2R receptor and arrestin-2. Nat. Commun. *12*, 2396. https://doi.org/10.1038/s41467-021-22731-x.

51. Sente, A., Peer, R., Srivastava, A., Baidya, M., Lesk, A.M., Balaji, S., Shukla, A.K., Babu, M.M., and Flock, T. (2018). Molecular mechanism of modulating arrestin conformation by GPCR phosphorylation. Nat. Struct. Mol. Biol. *25*, 538–545. https://doi.org/10.1038/s41594-018-0071-3.

52. Johnson, J.L., Yaron, T.M., Huntsman, E.M., Kerelsky, A., Song, J., Regev, A., Lin, T.-Y., Liberatore, K., Cizin, D.M., Cohen, B.M., et al. (2023). An atlas of substrate specificities for the human serine/threonine kinome. Nature *613*, 759–766. https://doi.org/10.1038/s41586-022-05575-3.

53. Simaan, M., Bédard-Goulet, S., Fessart, D., Gratton, J.-P., and Laporte, S.A. (2005). Dissociation of beta-arrestin from internalized bradykinin B2 receptor is necessary for receptor recycling and resensitization. Cell. Signal. *17*, 1074–1083. https://doi.org/10.1016/j.cellsig.2004.12.001.

54. Abreu, N., Acosta-Ruiz, A., Xiang, G., and Levitz, J. (2021). Mechanisms of differential desensitization of metabotropic glutamate receptors. Cell Rep. *35*, 109050. https://doi.org/10.1016/j.celrep.2021.109050.

55. Zidar, D.A., Violin, J.D., Whalen, E.J., and Lefkowitz, R.J. (2009). Selective engagement of G protein coupled receptor kinases (GRKs) encodes distinct functions of biased ligands. Proc. Natl. Acad. Sci. USA *106*, 9649–9654. https://doi.org/10.1073/pnas.0904361106.

56. Rajagopal, S., Kim, J., Ahn, S., Craig, S., Lam, C.M., Gerard, N.P., Gerard, C., and Lefkowitz, R.J. (2010). Beta-arrestin- but not G protein-mediated signaling by the "decoy" receptor CXCR7. Proc. Natl. Acad. Sci. USA *107*, 628–632. https://doi.org/10.1073/pnas.0912852107.

57. Lee, D.K., Ferguson, S.S.G., George, S.R., and O'Dowd, B.F. (2010). The fate of the internalized apelin receptor is determined by different isoforms of apelin mediating differential interaction with beta-arrestin. Biochem. Biophys. Res. Commun. *395*, 185–189. https://doi.org/10.1016/j.bbrc.2010.03.151.

58. Zhu, Y., Watson, J., Chen, M., Shen, D.R., Yarde, M., Agler, M., Burford, N., Alt, A., Jayachandra, S., Cvijic, M.E., et al. (2014). Integrating High-Content Analysis into a Multiplexed Screening Approach to Identify and Characterize GPCR Agonists. J. Biomol. Screen *19*, 1079–1089. https://doi.org/10.1177/1087057114533146.

59. Kawakami, K., Yanagawa, M., Hiratsuka, S., Yoshida, M., Ono, Y., Hiroshima, M., Ueda, M., Aoki, J., Sako, Y., and Inoue, A. (2022). Heterotrimeric Gq proteins act as a switch for GRK5/6 selectivity underlying β-arrestin transducer bias. Nat. Commun. *13*, 487. https://doi.org/10.1038/s41467-022-28056-7.

60. Tóth, A.D., Szalai, B., Kovács, O.T., Garger, D., Prokop, S., Balla, A., Inoue, A., Várnai, P., Turu, G., and Hunyady, L. (2023). Receptor endocytosis orchestrates the spatiotemporal bias of β-arrestin signaling. Preprint at bioRxiv *27*, 538587. https://doi.org/10.1101/2023.04.27.538587.

61. Gurevich, V.V., and Gurevich, E.V. (2014). Overview of different mechanisms of arrestin-mediated signaling. Curr. Protoc. Pharmacol. *67*, 2.10.1–2.10.9. https://doi.org/10.1002/0471141755.ph0210s67.

62. Ma, L., and Pei, G. (2007). Beta-arrestin signaling and regulation of transcription. J. Cell Sci. *120*, 213–218. https://doi.org/10.1242/jcs.03338.

63. Zhou, Q., and Sharp, P.A. (1996). Tat-SF1: cofactor for stimulation of transcriptional elongation by HIV-1 Tat. Science *274*, 605–610. https://doi.org/10.1126/science.274.5287.605.

64. Miller, H.B., Saunders, K.O., Tomaras, G.D., and Garcia-Blanco, M.A. (2009). Tat-SF1 is not required for Tat transactivation but does regulate the relative levels of unspliced and spliced HIV-1 RNAs. PLoS One *4*, e5710. https://doi.org/10.1371/journal.pone.0005710.

65. Hulver, M.J., Trautman, J.P., Goodwin, A.P., Roszczenko, S.K., Fogarty, K.H., and Miller, H.B. (2020). Human Tat-specific factor 1 binds the HIV-1 genome and selectively transports HIV-1 RNAs. Mol. Biol. Rep. *47*, 1759–1772. https://doi.org/10.1007/s11033-020-05267-z.

66. Chang, Y.-C., Chan, M.-H., Li, C.-H., Yang, C.-J., Tseng, Y.-W., Tsai, H.-F., Chiou, J., and Hsiao, M. (2021). Metabolic protein phosphoglycerate kinase 1 confers lung cancer migration by directly binding HIV Tat specific factor 1. Cell Death Dis. *7*, 135. https://doi.org/10.1038/s41420-021-00520-1.

67. May, D.G., Scott, K.L., Campos, A.R., and Roux, K.J. (2020). Comparative Application of BioID and TurboID for Protein-Proximity Biotinylation. Cells *9*, 1070. https://doi.org/10.3390/cells9051070.

68. Qureshi, H.Y., Ahmad, R., and Zafarullah, M. (2008). High-efficiency transfection of nucleic acids by the modified calcium phosphate precipitation method in chondrocytes. Anal. Biochem. *382*, 138–140. https://doi.org/10.1016/j.ab.2008.07.027.

69. Tóth, A.D., Garger, D., Prokop, S., Soltész-Katona, E., Várnai, P., Balla, A., Turu, G., and Hunyady, L. (2021). A general method for quantifying ligand binding to unmodified receptors using Gaussia luciferase. J. Biol. Chem. *296*, 100366. https://doi.org/10.1016/j.jbc.2021.100366.

70. Tareen, A., and Kinney, J.B. (2020). Logomaker: beautiful sequence logos in Python. Bioinformatics *36*, 2272–2274. https://doi.org/10.1093/bioinformatics/btz921.

71. Shimoyama, Y. (2022). pyCirclize: Circular Visualization in Python.

72. Turu, G., Soltész-Katona, E., Tóth, A.D., Juhász, C., Cserző, M., Misák, Á., Balla, A., Caron, M.G., and Hunyady, L. (2021). Biased Coupling to β-Arrestin of Two Common Variants of the CB2 Cannabinoid Receptor. Front. Endocrinol. *12*, 714561. https://doi.org/10.3389/fendo.2021.714561.

73. Gulyás, G., Radvánszki, G., Matuska, R., Balla, A., Hunyady, L., Balla, T., and Várnai, P. (2017). Plasma membrane phosphatidylinositol 4-phosphate and 4,5-bisphosphate determine the distribution and function of K-Ras4B but not H-Ras proteins. J. Biol. Chem. *292*, 18862–18877. https://doi.org/10.1074/jbc.M117.806679.

74. Szalai, B., Hoffmann, P., Prokop, S., Erdélyi, L., Várnai, P., and Hunyady, L. (2014). Improved Methodical Approach for Quantitative BRET Analysis of G Protein Coupled Receptor Dimerization. PLoS One *9*, e109503. https://doi.org/10.1371/journal.pone.0109503.

75. Saha, S., Ranjan, A., Godara, M., and Shukla, A.K. (2022). In-cellulo chemical cross-linking to visualize protein-protein interactions. Methods Cell Biol. *169*, 295–307. https://doi.org/10.1016/bs.mcb.2021.12.024.

76. Turiák, L., Ozohanics, O., Tóth, G., Ács, A., Révész, Á., Vékey, K., Telekes, A., and Drahos, L. (2019). High sensitivity proteomics of prostate cancer tissue microarrays to discriminate between healthy and cancerous tissue. J. Proteonomics *197*, 82–91. https://doi.org/10.1016/j.jprot.2018.11.009.

77. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372. https://doi.org/10.1038/nbt.1511.

78. Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics *13*, 2513–2526. https://doi.org/10.1074/mcp.M113.031591.

79. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745. https://doi.org/10.1093/nar/gkv1189.

80. UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.

81. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539. https://doi.org/10.1038/msb.2011.75.

82. Schneider, T.D., and Stephens, R.M. (1990). Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. *18*, 6097–6100. https://doi.org/10.1093/nar/18.20.6097.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
rabbit anti-β-arrestin2	Cell Signaling Technology	Cat#3857S; RRID: AB_2258681
rabbit anti-β-arrestin1	Cell Signaling Technology	Cat#30036S; RRID: AB_2798985
HRP-conjugated anti-rabbit	Cell Signaling Technology	Cat#7074S; RRID: AB_2099233
HRP-conjugated anti-mouse	Cell Signaling Technology	Cat#7076S; RRID: AB_330924
anti-HTATSF1	Proteintech	Cat#20805-I-AP; RRID: AB_10695767
Alexa 488-conjugated secondary goat anti-rabbit antibody	Invitrogen	Cat#A11034; RRID: AB_2576217
Bacterial and virus strains		
Supercompetent cells	home-made	N/A
Chemicals, peptides, and recombinant proteins		
Biotin	SERVA Electrophoresis GmbH	Cat#15060.03
High Capacity NeutrAvidin-Agarose Resin	Thermo Scientific	Cat#29204
GFP-Trap Magnetic Agarose resin	Chromotek	Cat#gtma-20
Phorbol 12-myristate 13-acetate (PMA)	Sigma-Aldrich	Cat#P1585-5MG
Poly-L-lysine	Sigma-Aldrich	Cat#P8920-100ML
Lipofectamine 2000	Invitrogen	Cat#52887
DMEM	VWR	Cat#392-0416
Penicillin/Streptomycin	Lonza	Cat#DE17-602E
Fetal bovine serum	Biosera	Cat#FB-1090/500
SDS	SERVA	Cat#20760.03
Plasmocin	InvivoGen	Cat#MPP-44-02
Coelenterazine h	Regis Technologies	Cat#1-361304-200
Arginine vasopressin	Sigma-Aldrich	Cat#V9879-1MG
Angiotensin II	Sigma-Aldrich	Cat#A9525-5x1MG
A61603	Tocris	Cat#1052
DAPI	Sigma-Aldrich	Cat#A2547-2ML
cOmplete Protease Inhibitor	Roche	Cat#11836145001
Phosphatase Inhibitor Mixture 3	Sigma-Aldrich	Cat#P0044-1ML
Sodium deoxycholate (SOC)	Sigma-Aldrich	Cat#D6750-25G
Sodium pyrophosphate	Sigma-Aldrich	Cat#S-6422
Sodium orthovanadate	Sigma-Aldrich	Cat#S6508-10G
Sodium fluoride	Sigma-Aldrich	Cat#S7920-100G
β-Glycerophosphate	Sigma-Aldrich	Cat#G9422-100G
Triton X-100	Sigma-Aldrich	Cat#T8787-100ML
Iodoacetamide (IAA)	Fluka Chemie GmbH	Cat#57670
1,4-Dithiothreitol (DTT)	Roche Diagnostics	Cat#10 197 777 001
Trypsin Gold, Mass Spectrometry Grade	Promega	Cat#V5280
Disuccinimidyl suberate (DSS)	PIERCE	Cat#21555
Immobilon Western chemiluminescent HRP Substrate	Millipore	Cat#WBKLS0500
Bovine Serum Albumin (BSA)	Sigma-Aldrich	Cat#A3059-100G
Methanol	MOLAR Chemical	Cat#05730-101-340
Ethanol	MOLAR Chemical	Cat#02910-101-340
Paraformaldehyde	Polysciences	Cat#18814-20

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SOC	Invitrogen	Cat#15544-034
RapiGest	Waters	Cat#186001860
Deposited data		
Western blot original images	Mendeley	Mendeley Data: https://doi.org/10.17632/j7xhxfzdsx.1
Mass spectrometry data	MassIVE	MassIVE Data: ftp://massive.ucsd.edu/v06/MSV000092868/
Experimental models: Cell lines		
HEK 293T	ATCC	Cat#CRL-3216
HEK 293A parent	O'Hayre et al.[25]	N/A
HEK 293A βarr1/2-KO	O'Hayre et al.[25]	N/A
Recombinant DNA		
pEYFP-N1 wt-βarr2–Venus	Gyombolai et al.[5]	N/A
pEYFP-N1 K2A-βarr2–Venus	Tóth et al.[20]	N/A
pEYFP-N1 wt-βarr2–Rluc8	Turu et al.[67]	N/A
pEYFP-N1 K2A-βarr2–Rluc8	This paper	N/A
pEYFP-N1 wt-βarr1–Venus	Gyombolai et al.[5]	N/A
pEYFP-N1 K2A-βarr1–Venus	This paper	N/A
pEYFP-N1 wt-βarr1–Rluc8	This paper	N/A
pEYFP-N1 K2A-βarr1–Rluc8	This paper	N/A
pEGFP-N1 L10–mRFP	This paper	N/A
pEGFP-N1 AT1R-Cterm–Venus	This paper	N/A
pEGFP-N1 AT1R-Cterm-TSTS/A–Venus	This paper	N/A
pEGFP-N1 V2R-Cterm–Venus	This paper	N/A
pEYFP-N1 wt-βarr2–TurboID	This paper	N/A
pEYFP-N1 K2A-βarr2–TurboID	This paper	N/A
pcDNA3.1+ adrenergic receptor alpha 1A	cDNA Resource Center	Cat##AR0A1A0001
pCMV6 HTATSF1	Origene	Cat#SC114990
pEYFP-N1 GRK5–FLAG	This paper	N/A
pCMV5 βarr1 (rat)	Dr. Stephen S. G. Ferguson	N/A
pCMV5 βarr2 (rat)	Dr. Stephen S. G. Ferguson	N/A
pEYFP-N1 HTSF1–mNeonGreen	This paper	N/A
pEYFP-N1 HTSF1-ST/AA–mNeonGreen	This paper	N/A
pEYFP-N1 HTSF1–Venus	This paper	N/A
pEYFP-N1 HTSF1-ST/AA–Venus	This paper	N/A
Software and algorithms		
Byonic	https://proteinmetrics.com/byonic/	v3.5.0, Protein Metrics Inc
GraphPad Prism 9	https://www.graphpad.com	GraphPad Prizm Software
MaxQuant	https://www.maxquant.org/	software version 1.6.7
Compass Data Analysis software 4.3	Bruker Daltonik GmbH	Version 4.3
Python code	This paper	Zenodo: https://doi.org/10.5281/zenodo.10804718
Cell analysis python code	This paper	Zenodo: https://doi.org/10.5281/zenodo.10072720
Other		
96-well white flat bottom plates	Greiner	Cat#655083
10 cm cell culture dish	Greiner	Cat#664160
μ-Slide 8 Well Ibidi plate	Ibidi GmbH	Cat#80826
PVDF transfer membrane	Thermo Scientific	Cat#88518

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Gábor Turu (turu.gabor@ttk.hun-ren.hu).

Materials availability

Plasmids generated in this study are available from the lead contact upon request without restrictions.

Data and code availability

- Original western blot images have been deposited at Mendeley and are publicly available as of the date of publication. The DOI is listed in the key resources table. Microscopy data reported in this paper will be shared by the lead contact upon request. Mass spectrometry original data have been deposited at MassIVE with accession number MSV000092868 and will be available as of the date of publication.
- The model code is shared at Zenodo and is publicly available as of the date of publication. The DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell culture and transfection

HEK 293T cells were from American Type Culture Collection (ATCC CRL-3216 Manassas, VA). HEK 293A parent and βarr1/2-KO cells were described earlier.[44] The cells were cultured in DMEM supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin in 5% CO_2 at 37°C. Cells were treated with plasmocin (25 μg/mL) for two weeks before the experiments. For BRET measurements, cells were transfected in suspension using Lipofectamine 2000 according to the manufacturer's protocol and plated on white poly-L-lysine-coated 96-well plates. For co-precipitation and confocal microscopy experiments, the calcium phosphate precipitation method was used for cell transfection as described previously.[68,69] The cells were plated on poly-L-lysine-coated 10 cm plates or on μ-Slide 8 Well Ibidi (Grafelfing, DE) plates, and the medium was replaced with fresh DMEM after 6–7 h. Cells were regularly tested for mycoplasma contamination.

METHOD DETAILS

Database development

We conducted a PubMed search using the keyword "arrestin" and reviewed approximately 6000 entries. Studies containing high-resolution microscopic images of β-arrestin subcellular recruitment were examined. Specifically, images of cells before and at least 10 min after agonist stimulation were included in the analysis. Receptors were classified as class B if they were capable of inducing endosomal trafficking of βarr1 or βarr2 either in their basal state or upon stimulation with any of their agonists, indicating their potential to form sustained interactions. Differences in affinities for βarr1 and βarr2 were not considered in the categorization process. Receptors that did not promote endosomal βarr1 or βarr2 recruitment under any conditions (resulting in either no trafficking or only plasmalemmal trafficking) were categorized as non-B. The sequences of these categorized receptors formed the training set. To construct the database of mutant receptors, we reviewed the literature for serine/threonine mutations or C-terminal truncations of receptors classified as class B. We determined whether the missense mutation or truncation involved the predicted arreSTick region (the motif with the highest convolutional score; in case of equal scores, the proximal one), and whether it caused a change in class from B to non-B.

Convolutional neural network and protein predictions

The convolutional neural network model was implemented in Python 3 using the Tensorflow[2.6.1] library. The network structure is shown in Figure S1, and the code is available at Zenodo. During the training, we used either the sequence of the C terminus or the ICL3 loop of GPCRs as an input. We set a convergence threshold of 0.8 of the area under the receiver operating characteristic curve (ROC AUC) value on the training data. The training was repeated in each round until the threshold was reached. During cross-validation, the receptor dataset was randomly divided into a training group and cross-validation set (104 vs. 10 receptors). The model was trained on the training set, and the cross-validation set was predicted based on either their training sequence or full sequence. The cross-validation was repeated 50 times, and average AUC ROC values were plotted. To visualize amino acid frequencies within the arrestin-binding regions, we trained a single network with the "grouped model" and extracted 15 amino acids from the position of the global max value in the convoluted sequence of receptors labeled as class B in the training sequence. The amino acid frequencies were calculated for each position, and the Logomaker Python library was used for display.[70] For ROC AUC curve generation, we ran a cross-validation 50 times, and each time a random set of 10 receptors were predicted. The predictions were averaged, and ROC AUC values were calculated for the predicted training set, cross-validation set, and cross-validation set using full sequences. ROC curves

were calculated with different cross-validation strategies in the case of the grouped model predictions or the phosphorylation code[19] predictions. In the case of the grouped model, cross-validations were performed similarly as described above. In each round, the hold-out receptors were predicted using the training sequences, the full sequences, or the full masked sequences. The predicted class B probabilities were averaged for each receptor, and these values were used for the ROC curve plotting. In the case of the phosphorylation codes, the total number of the short and long codes in each receptor and receptor class were used to create the plot. For the short and long codes definition we used the following regex patterns, respectively: ''[S|T].[S|T][^P][^P][S|T|E|D]'' and ''[S|T].[S|T][^P][^P][S|T|E|D]''. To predict the probability of stable β-arrestin binding and the arreSTick motif sequence in all human GPCRs, we have collected receptor sequences, information, and topological data from the GPCRdb.org.[23] Circular representation of the receptors was done with the pyCirclize Python library.[71] Protein location data were collected from the Human Protein Atlas (proteinatlas.org).[34] Protein structure data for human proteome were collected from the AlphaFold2 Website: https://alphafold.ebi.ac.uk/,[25] and the sequence regions with over 70% model confidence were masked out (replaced with dummy amino acids) in proteins before the prediction.

Materials and plasmid constructs

Cell culture reagents were from Thermo Fisher Scientific (Waltham, MA) and Biosera (Cholet, FR). Cell culture dishes and plates were from Greiner (Kremsmunster, AT). Plasmocin was from InvivoGen (Tolouse, FR). Coelenterazine h was obtained from Regis Technologies (Morton Grove, IL). Biotin was from SERVA Electrophoresis GmbH (Heidelberg, DE). High Capacity NeutrAvidin-Agarose Resin was from Thermo Scientific (Waltham, MA), and GFP-Trap Magnetic Agarose resin was from Chromotek (Planegg-Martinsried, DE). Anti-β-arrestin2, anti-β-arrestin1, and HRP-conjugated anti-rabbit antibodies were from Cell Signaling Technology, Inc. (Beverly, MA, USA). Phorbol 12-myristate 13-acetate (PMA), angiotensin II (AngII), and arginine vasopressin (AVP) were from Sigma-Aldrich (St. Louis, MO). LC-MS grade solvents and urea were purchased from Merck (Darmstadt, DE). Mass spectrometry grade trypsin was obtained from Promega (Promega Corporation, Madison, WI, USA). Reagents used for enzymatic digestion (1,4-Dithiothreitol (DTT) and Iodoacetamide (IAA)) were purchased from Roche Diagnostics (Mannheim, DE) and Fluka Chemie GmbH (Buchs, CH).The plasmids encoding K2A-βarr2–Venus,[17] βarr2–Rluc8,[72] wt-βarr1–Venus and wt-βarr2–Venus[5] have been previously described. The pmNeonGreen-C1 plasmid was kindly provided by Dr. Balázs Enyedi. GRK5–YFP was kindly provided by Dr. Marc G. Caron. Untagged rat βarr1 and βarr2 were provided by Dr. Stephen S. G. Ferguson. K10A and K11A mutations were introduced to rat β-arrestin1 by precise gene fusion PCR to create K2A-βarr1–Venus. Wild-type and K2A-βarr1–RLuc8 and K2A-βarr2–RLuc8 were constructed by replacing Venus to RLuc8 with AgeI/NotI restriction digestion in wt- or K2A-βarr1–Venus, or K2A-βarr2–Venus, respectively. To generate TurboID-tagged wild-type and K2A-mutant rat β-arrestin2, the coding sequence of TurboID[35] was synthesized in gBlock gene fragment (IDT, Coralville, IA, USA), and it was cloned into wt-βarr2–Venus or into K2A-βarr2–Venus by replacing Venus using AgeI/NotI restriction enzymes.GRK5–FLAG was generated using annealed oligo cloning by replacing the YFP-encoding DNA sequence to that of the FLAG tag. HTSF1–mNeonGreen was produced by cloning the HTATSF1 from pCMV6-Entry (Origene, Rockville, MD, USA) vector into the pEYFP-N1 vector between AfeI/SalI restriction sites, then YFP was replaced with mNeonGreen using AgeI/NotI restriction enzymes. To create HTSF1–Venus, YFP was replaced by monomeric Venus (containing the A206K mutation, all Venus-tagged constructs used in this study harbored this monomerizing mutation). Alanine mutant form of HTSF1 (HTSF1-ST/AA: S739A, T740A, S742A, S743A, S747A, and S748A) was created by gBlock gene fragment (IDT, Coralville, IA, USA) synthesis, and it was cloned into pEYFP-N1 vector between BglII/AgeI restriction sites. After that, YFP was replaced by mNeonGreen or Venus with AgeI/NotI restriction enzymes. The L10–mRFP construct, containing the plasma membrane target sequence L10 (MGCVCSSNPENNNN, the first 10 amino acids of mouse Lck followed by polyglutamine linker), was created by replacing Venus by mRFP in L10–Venus construct.[73] To generate the AT1R-Cterm–Venus construct, the coding sequence of the C terminus of rat AT1a angiotensin receptor (residues 320–359, IPPKAKSHSSLSTKMSTLSYRPSDNMSSSAKKPASCFEVE) together with Venus from a Venus-tagged full-length receptor construct[5] was PCR-amplified, then it was in-frame fused with DPTRSRAQASNSGGG linker to the L10 sequence by replacing mRFP in L10–mRFP. A similar strategy was used for the two following receptor C terminus constructs. For AT1R-Cterm-TSTS/A–Venus, AT1R-TSTS/A–Venus was used as a template,[17] the sequence of AT1R-Cterm-TSTS/A: IPPKAKSHSSLSAKMAALAYRPSDNMSSSAKKPASCFEVE. For V2R-Cterm–Venus, the C terminus (residues 343–371, ARGRTPPSLGPQDESCTTASSSLAKDTSS) of the human Venus-tagged V2 vasopressin receptor[74] was fused together with Venus to the L10 sequence.

Bioluminescence resonance energy transfer (BRET) measurements

Transiently transfected HEK 293T cells were plated on poly-L-lysine-coated 96-well white-walled tissue culture plates, and the measurements on adherent cells were performed 24–28 h after transfection. Luminescence intensities were measured using a Thermo Scientific Varioskan Flash multimode plate reader at 37°C as described previously.[17] Briefly, before the measurements, we replaced the medium with a modified Kreb's-Ringer medium (120 mM NaCl, 10 mM glucose, 10 mM Na-HEPES, 4.7 mM KCl, 0.7 mM MgSO4, 1.2 mM $CaCl_2$, pH 7.4). We determined the expression of the Venus-tagged proteins by recording fluorescence intensity at 535 nm with excitation at 510 nm. After the addition of the luciferase substrate coelenterazine *h* (5 μM), we measured luminescence intensities using 530 nm and 480 nm filters. In the BRET titration experiments, luminescence intensity was measured without a filter as well in order to assess the expression of the donor-labeled construct. The BRET ratio was determined by dividing the luminescence intensities at 530 nm and 480 nm with each other (I_{530nm}/I_{480nm}). In the titration BRET experiments, BRET ratios were normalized to

those wells in which no Venus-tagged construct was expressed. To assess the interaction between plasma membrane-targeted receptor C-termini and β-arrestins, cells were transfected with Rluc8-tagged wild-type or K2A-mutant βarr1 or βarr2 (0.001 μg/well) and with Venus-tagged receptor C termini (0.05 μg/well). After measuring the baseline BRET ratio, cells were stimulated with 100 nM AngII, 100 nM PMA or 100 nM AVP, and the change of the BRET ratio (stimulated - vehicle-treated) was continuously determined. Kinetic measurements were performed in triplicate. For the titration BRET experiments, we transfected the cells with Rluc8-tagged wild-type or K2A-mutant βarr2 constructs (0.02 μg/well) and Venus-tagged HTSF1 plasmids in increasing concentrations (0–0.2 μg/well), we also added pcDNA3.1 to keep the total amount of transfected DNA constant (0.25 μg/well). Data from all wells are shown in the BRET titration experiments.

Confocal microscopy

βarr1/2-KO HEK 293A cells were cotransfected with L10–mRFP, HTSF1–mNeonGreen and either untagged wt-βarr2, wt-βarr1, K2A-βarr2 or pcDNA3.1 in suspension using calcium phosphate precipitation and plated immediately on poly-L-lysine-coated μ-Slide 8 well Ibidi plates. The medium was changed the next day, and 24 h after the transfection, the cells were fixed with paraformaldehyde (4%, 15 min), and the nuclei were stained with DAPI. We used a Zeiss LSM 710 confocal laser scanning microscope for obtaining representative images and ImageXpress confocal microscopy for the quantification of protein localization. For the latter, 49 images per well were obtained in three channels (L10–mRFP, HTSF1–mNeonGreen, and DAPI). L10-mRFP images were used for cell segmentation, the DAPI channel for nucleus segmentation, and the mNeonGreen channel was used for the HTSF1 fluorescence determination. Images were segmented using the Cellpose Python library (https://github.com/MouseLand/cellpose), and total mNeonGreen fluorescence in the cytoplasm (cell mask minus nuclear mask) was divided with nuclear fluorescence. The applied analysis code is available on GitHub (https://github.com/turugabor/cellAnalysis).

To assess the subcellular localization of untagged and TurboID-tagged βarr2 constructs, βarr1/2-KO HEK 293A cells were plated on poly-L-lysine-coated μ-Slide 8 well Ibidi plates, followed by transfection using the calcium phosphate precipitation method. The day after transfection, the cells were fixed with ice-cold methanol for 30 min. Thereafter, the cells were washed three times with PBS solution at room temperature. The samples were treated with PBS supplemented with 0.1% Triton X-100 (PBST) for 5 min, then washed with PBS three times. After a 30-min blocking step with 1% BSA containing PBST, the cells were treated with primary rabbit anti-β-arrestin2 antibodies (1:200, Cell Signaling, C16D9, catalog No. #3857) for 1 h at room temperature. The samples were washed three times with PBST. For detection, Alexa 488-conjugated secondary goat anti-rabbit antibody (1:500, Thermo Fisher Scientific, catalog No. A-11034) was applied for 1 h on room temperature, followed by three final washes with PBST. Fluorescence images were taken with a Zeiss LSM 710 confocal laser scanning microscope.

Immunoprecipitation and immunoblot analysis of HTSF1

HEK 293T cells were transfected in suspension with plasmids encoding Venus-tagged wild-type or K2A-mutant βarr1 or βarr2. 24 h after transfection, we placed the dishes to ice and washed them with ice-cold PBS (supplemented with 1.2 mM $CaCl_2$) solution. The washing step was repeated three times. Then the cells were lysed with lysis buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% Triton X-100) supplemented with cOmplete Protease Inhibitor mixture (Roche) and Phosphatase Inhibitor Mixture 3 (Sigma). For protein cross-linking, we added 1 mM disuccinimidyl suberate (DSS) to the lysate at 37°C for 15 min. After that, we quenched the reaction by adding Tris-containing buffer (50 mM Tris-HCl, 50 mM L-glycine, 1.2 mM $CaCl_2$, 1 mM $MgCl_2$, 50 mM NaCl, pH 7.4) to the samples at a ratio of 1:10 at 4°C for 15 min.[75] Samples were centrifuged at 20,800 × *g* for 10 min and the supernatants were incubated with 15 μL GFP-Trap Magnetic Agarose resin for 1 h at 4°C. After centrifugation, the lysates were collected and the beads were washed three times with a washing buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA). We eluted the proteins from the surface of the beads using Laemmli SDS sample buffer (2x) containing 10% mercaptoethanol, boiled it at 95°C for 5 min. Proteins were separated by SDS-polyacrylamide gel electrophoresis and were blotted onto PVDF membranes. A 1-h blocking step was performed using 5% fat-free milk powder in PBS with 0.05% Tween 20 (PBST) at room temperature. Thereafter, membranes were treated overnight at 4°C with primary antibodies, diluted 1:1000 in PBST, including mouse anti-HTSF1 (Proteintech, Cat#20805-I-AP), rabbit anti-β-arrestin2 (Cell Signaling Technology, Cat#3857S), and rabbit anti-β-arrestin1 (Cell Signaling Technology, Cat#300036S). After three 10-min washing steps with PBST, membranes were incubated with HRP-conjugated anti-rabbit (Cat#7074S Cell Signaling Technology) or HRP-conjugated anti-mouse (Cat#7076S Cell Signaling Technology) secondary antibodies, diluted 1:5000 in PBST, for 1 h at room temperature, followed by washing with PBST three times. Visualization was performed using Immobilon Western chemiluminescent HRP Substrate (Millipore), and the signal was detected with Azure c600 Western Blot Imaging System (Biosystems). The results were quantitatively evaluated with densitometry using ImageJ software.

Affinity purification using biotin ligase

HEK 293T cells were transfected in suspension with plasmids encoding βarr2–TurboID (biotin ligase) or K2A-βarr2–TurboID and untagged α1A-adrenergic receptor. 24h after transfection, cells were serum starved for 2–4 h, then 100 μM biotin was added for 1h, and cells were stimulated with A61603 (1 μM at 37°C) for 1 h to allow substantial biotinylation. Reactions were stopped by placing the dishes on ice and washing them with an ice-cold PBS solution. The washing step was repeated three times. Then the cells were lysed with 2% sodium deoxycholate (SOC) buffer (Sigma), supplemented with 0.025% RapiGest (Waters), cOmplete Protease Inhibitor mixture (Roche), and Phosphatase Inhibitor salts (1 mM sodium pyrophosphate, 2 mM sodium orthovanadate, 10 mM sodium

fluoride, 50 mM β-Glycerophosphate). Lysates were collected, sonicated for 45 s, and then centrifuged at 20,800 × *g* for 10 min. Supernatants were incubated with 100 μL of High Capacity NeutrAvidin-agarose resin for 1h at 4°C. The beads were washed three times with ice-cold supplemented SOC and once with PBS. We eluted all proteins from the surface of the beads in Laemmli SDS sample buffer (2x) containing biotin and 10% mercaptoethanol, boiled it at 95°C for 5 min. After centrifugation, the supernatant was transferred to a new tube. To remove SDS from the protein solution, the proteins were precipitated with 1 mL 100% ethanol (at 4°C for 24 h).

Mass spectrometry

Enzymatic Digestion: Precipitated, air-dried samples were digested in solution using trypsin as previously described with minor modifications.[76] In brief, precipitated pellets were dissolved in 30 μL 8 M urea in 50 mM ammonium bicarbonate. DTT was added at a final concentration of 5 mM and incubated at 37°C for 30 min. For alkylation, IAA was added at a final concentration of 10 mM and incubated in the dark at room temperature for 30 min. Samples were diluted 10-fold with 50 mM ammonium bicarbonate, and enzymatic digestion was performed with 1 μL 1 μg/μL trypsin overnight at 37°C. The reaction was quenched by the addition of 1 μL formic acid. Peptide clean-up and desalting were performed on Pierce C18 spin columns (Thermo Fisher Scientific, Waltham, MA).Nano LC-MS/MS: Mass spectrometry measurements were performed on a Maxis II Q-TOF (Bruker Daltonics, Bremen, Germany) equipped with a CaptiveSpray nanoBooster ion source coupled to an Ultimate 3000 nanoRSLC system (Dionex, Sunnyvale, CA, USA). Samples were dissolved in 2% AcN, 0.1% FA and injected onto an Acclaim PepMap100 C-18 trap column (5 μm, 100 μm × 20 mm, Thermo Scientific, Sunnyvale, CA, USA) for sample desalting. Peptides were separated on an ACQUITY UPLC M-Class Peptide BEH C18 column (130 Å, 1.7 μm, 75 μm × 250 mm, Waters, Milford, MA, USA) at 48°C applying gradient elution (4% B from 0 to 11 min, followed by a 120 min gradient to 50% B). Eluent A consisted of water +0.1% formic acid, while eluent B was acetonitrile +0.1% formic acid. MS spectra were recorded at 3 Hz, while the CID was performed at 16 Hz for abundant precursor ions and at 4 Hz for low-abundance ones. Sodium formate was used as an internal standard, and raw data were recalibrated by the Compass Data Analysis software 4.3 (Bruker Daltonik GmbH, Bremen, Germany). Protein identification and label-free quantitation: Proteins were identified by searching against the human SwissProt database (2019_06) using the Byonic (v3.5.0, Protein Metrics Inc, USA) software search engine. First, the combined LC-MS results were searched by Byonic (30 ppm peptide mass tolerance, 50 ppm fragment mass tolerance, 2 missed cleavages, carbamidomethylation of cysteines as fixed modification, deamidation (NQ), oxidation (M), acetyl (Protein N_Term), Glu->Pyro-Glu and Gln->Pyro-Glu as a variable modification) and proteins were identified using 1% FDR limit. This protein list was used for label-free quantitation (LFQ) using MaxQuant[77] (software version 1.6.7), applying its default parameters except modifications listed above, and each LC-MS/MS run was aligned using the "match between runs" feature (match time window 1.5 min, alignment time window 15 min). MaxQuant analysis searched only for those proteins that were identified previously by Byonic search (this makes false identification less likely), and 1% FDR was set at the protein identification level.

Mass spectrometry data analysis

Label-free quantification of the mass spectrometry data was carried out similarly to the MaxQuant LFQ method.[78] This method assumes that the majority of the proteome does not change between any two conditions (i.e., two β-arrestin protein partners). Peptide intensities were normalized by minimizing the sum of all squared logarithmic fold differences between samples. In the case of a protein interactome, in contrast to a full-cell proteome, some proteins might be completely absent from the sample because of the lack of interaction. Therefore, in further analysis, we calculated the protein expression ratios using all the protein peptides instead of only the paired peptides across sample pairs as previously described.[78] The missing peptide intensity values were replaced with zeros. The peptides from six experiments were pooled together, and the median log2 fold difference in intensities was taken as a difference to decrease the effect of the outliers on the average value. Differences between individual proteins in the wild-type- and K2A-mutant β-arrestin2 samples were statistically analyzed with the Wilcoxon test, and false discovery rate correction was performed with the Benjamini-Hochberg method. Protein preference to either of the β-arrestins was determined as a positive (wild-type preference) or negative (K2A preference) log2 fold difference.

Evolutionary conservation analysis

The sequences for the analysis were downloaded in 2023 July from the NCBI Reference Sequence Protein Database.[79] We performed searches with the NCBI "blastp" tool, using the sequences of human GPCRs taken from UniProt.[80] The maximum target sequence number was set to 1000. We restricted our BLAST search to animal species and discarded "partial" and "low quality" sequences. Multiple sequence alignments (MSAs) were performed using Clustal Omega.[81] The evolutionary conservation of an MSA column was calculated by[82]:

$$R_{seq} = S_{max} - S_{obs} = \log_2(21) - \left(- \sum_{n=1}^{21} p_n \log_2 p_n \right)$$

S_{max} is the maximal entropy $log_2(21) = 4.39$; 20 amino acids and gap symbol), S_{obs} is the observed entropy at a given column, p_n is the observed frequency of symbol n in the column. The R_{seq} values were multiplied with $1\text{-}p_{gap}$ to avoid assigning high conservation scores to columns with high gap frequency. The motif prediction was done only on the 3rd intracellular loops and the C-terminals of

human GPCRs. Annotations were taken from GPCRdb.[23] Amino acid conservation scores were computed individually, and from these, the median scores for each GPCR were calculated from the ICL3/C-terminal residues. ICL3 and C terminus sequences were defined based on human GPCRs. The calculation involved all amino acids or specifically S/T amino acids, both within and outside the arreSTick region. The conservation distributions were drawn with Seaborn's boxplot or histogram functions with default parameters.

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless otherwise stated, data analysis and plotting were done in Python, using Pandas, Numpy, Matplotlib, Seaborn, and Scipy libraries. No statistical methods were used to predetermine the sample size. Except for the mass spectrometry analysis, the sample size (N) in figure legends refers to the number of independent experiments (biological replicates), all data points were included in the statistical analyses. The experiments were not randomized, and the investigators were not blinded. Statistical analysis and curve fitting for one-site specific binding and linear regression equations in BRET titration experiments were conducted using GraphPad Prism 9 software. Data were statistically analyzed using either one sample t test, paired two-tailed t test, Kruskal-Wallis test with Dunn's post-hoc test, Wilcoxon matched-pairs signed rank test, two-way ANOVA with Tukey's post-hoc test, or three-way ANOVA with Bonferroni post-hoc test, as indicated in the corresponding figure legends. Unless otherwise stated, data are expressed as mean ± SEM. Statistical significance was attributed to effects with a *p*p-value below 0.05. Unbiased fluorescence image analysis was ensured by application of automated, machine learning-based algorithms. N always refers to the number of independent biological experiments and the number of technical replicates are also indicated in the figure legends.

ADDITIONAL RESOURCES

An application predicting arreSTick patterns within proteins is available at the arrestick.org website.