



**HAL**  
open science

## Structural Insights into the Intrinsically Disordered GPCR C-Terminal Region, Major Actor in Arrestin:GPCR Interaction

Myriam Guillien, Assia Mouhand, Aurelie Fournet, Amandine Gontier, Aleix Martí Navia, Tiago N Cordeiro, Frédéric Allemand, Aurélien Thureau, Jean-Louis Banères, Pau Bernadó, et al.

► **To cite this version:**

Myriam Guillien, Assia Mouhand, Aurelie Fournet, Amandine Gontier, Aleix Martí Navia, et al.. Structural Insights into the Intrinsically Disordered GPCR C-Terminal Region, Major Actor in Arrestin:GPCR Interaction. *Biomolecules*, 2022, 12 (5), pp.617. 10.3390/biom12050617 . hal-03748408

**HAL Id: hal-03748408**

**<https://hal.science/hal-03748408v1>**

Submitted on 9 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

# Structural Insights into the Intrinsically Disordered GPCR C-Terminal Region, Major Actor in Arrestin-GPCR Interaction

Myriam Guillien <sup>1</sup>, Assia Mouhand <sup>1</sup>, Aurélie Fournet <sup>1</sup>, Amandine Gontier <sup>1</sup>, Alex Martí Navia <sup>1</sup>, Tiago N. Cordeiro <sup>1,2</sup>, Frédéric Allemand <sup>1</sup>, Aurélien Thureau <sup>3</sup>, Jean-Louis Banères <sup>4</sup>, Pau Bernadó <sup>1</sup> and Nathalie Sibille <sup>1,\*</sup>

- <sup>1</sup> Centre de Biologie Structurale (CBS), Univ Montpellier, CNRS, INSERM, 34090 Montpellier, France; guillien@cbs.cnrs.fr (M.G.); mouhand@cbs.cnrs.fr (A.M.); fournet@cbs.cnrs.fr (A.F.); gontier@dynamic-biosensors.com (A.G.); elloto13@gmail.com (A.M.N.); tiago.cordeiro@itqb.unl.pt (T.N.C.); allemand@cbs.cnrs.fr (F.A.); bernado@cbs.cnrs.fr (P.B.); [nathalie.sibille@cbs.cnrs.fr](mailto:nathalie.sibille@cbs.cnrs.fr) (N.S.)
- <sup>2</sup> Instituto de Tecnologia Química e Biológica António Xavier (ITQB), Universidade NOVA de Lisboa, Av. da República, 2780-157 Oeiras, Portugal
- <sup>3</sup> Beamline SWING, Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91190 Gif-sur-Yvette, France; aurelien.thureau@synchrotron-soleil.fr (A.T.)
- <sup>4</sup> IBMM, UMR5247 CNRS, Pôle Chimie Balard Recherche, 1919 route de Mende, CEDEX 5, 34293 Montpellier, France; jean-louis.baneres@umontpellier.fr (J.-L.B.)
- \* Correspondence: [nathalie.sibille@cbs.cnrs.fr](mailto:nathalie.sibille@cbs.cnrs.fr); Tel.: +33-4-6741-7705; Fax: +33-4-6741-7913

**Abstract:** Arrestin-dependent pathways are a central component of G protein-coupled receptor (GPCRs) signaling. However, the molecular processes regulating arrestin binding are to be further illuminated, in particular with regard to the structural impact of GPCR C-terminal disordered regions. Here, we used an integrated biophysical strategy to describe the basal conformations of the C-terminal domains of three class A GPCRs, the vasopressin V2 receptor (V2R), the growth hormone secretagogue or ghrelin receptor type 1a (GHSR) and the  $\beta$ 2-adrenergic receptor ( $\beta$ 2AR). By doing so, we revealed the presence of transient secondary structures in these regions that are potentially involved in the interaction with arrestin. These secondary structure elements differ from those described in the literature in interaction with arrestin. This suggests a mechanism where the secondary structure conformational preferences in the C-terminal regions of GPCRs could be a central feature for optimizing arrestins recognition.

**Keywords:** GPCR; arrestin; intrinsically disordered proteins or regions (IDPs/IDRs); NMR

**Citation:** Guillien, M.; Mouhand, A.; Fournet, A.; Gontier, A.; Navia, A.M.; Cordeiro, T.N.; Allemand, F.; Thureau, A.; Banères, J.-L.; Bernadó, P.; Sibille, N. Structural Insights into the Intrinsically Disordered GPCR C-Terminal Region, Major Actor in Arrestin-GPCR Interaction.

*Biomolecules* **2022**, *12*, x.

<https://doi.org/10.3390/xxxxx>

Academic Editor(s): Vladimir N. Uversky

Received: date

Accepted: date

Published: date

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



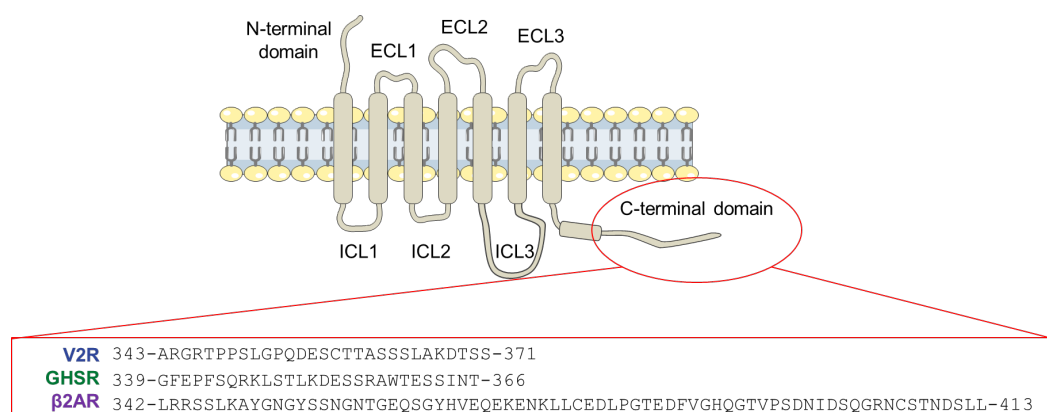
**Copyright:** © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

G protein-coupled receptors (GPCRs) are integral membrane proteins involved in signal transduction. They are central in the cellular response for a wide range of extracellular ligands, such as hormones, nucleotides, lipids, ions, photons, and neurotransmitters [1]. Their signaling outcome regulates a large number of biological functions and, therefore, their dysfunctions are linked to various pathologies [2]. Consequently, GPCRs are the target of a third of the current clinical drugs [3]. Their functional diversity comes from the existence of a large number of GPCRs (~800 members in humans), classified according to their sequence and phylogenetic analyses [4]. While they share a highly conserved core domain (7TM) composed of seven transmembrane helices connected by three extracellular (ECL) and intracellular (ICL) loops, their extracellular N- and intracellular C-termini, as well as the loops, are highly variable in both sequence and length (Figure 1) [5,6].

Upon extracellular ligand binding, GPCR conformational rearrangements allow G protein association and G protein dependent signaling initiation. Besides G proteins, GPCRs can trigger other signaling pathways by interacting with arrestin, i.e., desensiti-

zation, internalization, and receptor trafficking [7]. Arrestin interaction requires GPCR C-terminal domain (GPCR-Cter) phosphorylation by specific G protein-coupled receptor kinases (GRKs). The current model of GPCR:arrestin interaction, the so-called phospho-barcode model [8,9], states that, upon activation, GPCRs become phosphorylated at specific sites on their C-terminal domain, and this impacts their interaction with arrestin, in turn affecting its conformation and, ultimately, the intracellular signal triggered by arrestin. However, we are just beginning to grasp the mechanistic basis of this mode of functioning as experimental demonstrations are still scarce even for the best-studied GPCRs.



**Figure 1. Schematic representation of a GPCR.** GPCRs share a common core domain with seven transmembrane helices (7TM) linked by extracellular (ECL) and intracellular (ICL) loops. These loops, as well as the N- and C-terminal domains, are variable in length and sequence. The sequences of the three peptides used in this study are indicated just below.

GPCR:arrestin complex formation leads to arrestin activation through conformational changes (for a review see [10]). However, putative conformational changes of GPCR-Cter induced by GRK phosphorylation or arrestin binding are still poorly described. These C-terminal regions are predicted to behave as intrinsically disordered regions (IDRs) [11,12]. Intrinsically disordered proteins (IDPs) and IDRs are highly dynamic proteins/regions with a low content of transient secondary structures, which make their structural characterization difficult [13]. Atomic structures of GPCR-arrestin complex revealed that the C-terminus of the vasopressin 2 receptor (V2R) and rhodopsin are partially folded on arrestin surface [13–16]. This suggests a mechanism by which the IDR undergoes conformational changes upon post-translational modification and/or binding to its target [14]. Indeed, IDPs/IDRs most likely contain some pre-formed secondary structure elements that are often involved in the recognition of specific partners and modulate the affinity [15].

In order to better understand how the structural features of the GPCR C-terminal regions could impact on their functional role, we characterized the structure of the truncated C-terminal regions of three GPCRs, namely the vasopressin V2 receptor (V2R), the ghrelin receptor type 1a (GHSR) and the  $\beta$ 2-adrenergic receptor ( $\beta$ 2AR) (Figure 1). These three class-A receptors are important therapeutic targets [2] and are representative of the two different classes of arrestin binders [16]. The first class, which includes  $\beta$ 2AR and GHSR, forms a transient complex with arrestin that dissociates near the plasma membrane. Thus, arrestin does not internalize with the receptor. In contrast, the second class, which includes V2R, forms a more stable complex allowing the internalization of the whole assembly.

Here, we confirmed the disordered nature of these three GPCR-Cters and the presence of transient secondary structures by a set of biophysical tools: circular dichroism (CD), multi angle light scattering (MALS), and small angle X-ray scattering (SAXS). Then, we used nuclear magnetic resonance (NMR) to probe the conformational and dy-

namic preferences at residue level, using a set of complementary experiments, such as secondary chemical shifts (SCS), scalar (J) and residual dipolar couplings (RDCs), paramagnetic relaxation enhancement (PRE) and relaxation [17–20]. We show that the three C-terminal regions of the chosen GPCRs displayed different transient secondary structures, which could be involved in arrestin binding. These regions could act as short linear motifs (SLiMs), partner recognition IDP segments that are embedded in poorly conserved, disordered regions. By comparison with crystallographic structures of synthetic peptides in interaction with arrestin [21–23], our results suggest that structural changes in these putative SLiMs occur either after phosphorylation of the C-terminal region or upon arrestin binding.

## 2. Materials and Methods

### 2.1. Sample Preparation

The (human) cDNA coding for the C-terminus of V2R (343–371), GHSR (339–366) and  $\beta$ 2AR (342–413) were ordered to GeneArt® gene synthesis (Life Technologies, Courtaboeuf, France). They were cloned into the expression vector pETM33 for GHSR-Cter and V2R-Cter, while  $\beta$ 2AR-Cter was cloned into pET1a expression vector. Vectors were transfected into *E. Coli* BL21 DE3 strain (Fisher Scientific S.A.S., Illkirch, France) and (His)<sub>6</sub>-GST-tagged proteins were expressed in self-inducible medium N5052 or in minimal medium M9 by adding 2 mM IPTG when the OD<sub>600</sub> reached 0.6. Labeling with <sup>15</sup>N and/or <sup>13</sup>C was carried out by growing cells in the previous media with 99% <sup>15</sup>NH<sub>4</sub>Cl (CortecNet, Voisins-Le-Bretonneux, France) and/or 99% <sup>13</sup>C<sub>6</sub>-Glucose (Cambridge Isotope Laboratories Inc., Tewksbury, United States) as unique source of nitrogen and carbon, respectively. After harvesting, bacteria were resuspended and lysed by sonication in 20 mM Tris 20 pH 7.5, 300 mM NaCl and 2 mM dithiothreitol (DTT) (buffer A) supplemented with anti-protease (Complete® EDTA free tablet (Sigma-Aldrich Chimie S.A.R.L.-Roche, Saint Quentin Fallavier, France)). Cell debris was removed by centrifugation and the soluble fraction was loaded onto 5 mL HisTrap FF column (Cytiva Europe GmbH, Velizy Villacoublay, France). The resin was washed with buffer A containing 10 mM imidazole and proteins were eluted with a linear gradient of buffer A supplemented with 500 mM imidazole. The combined eluate was dialyzed over night at 4 °C in 1 L of buffer A with the 3C PreScission protease (ratio protease:protein 1:100 (w/w)) or TEV protease (ratio protease:protein 1:50 (w/w)) for the pETM33 and pET1a constructs, respectively. Proteases and His-GST tags were removed using gravity column containing 2.5 mL Ni-sepharose (GE Healthcare). Untagged proteins were concentrated using a 2 kDa Millipore concentrator and injected into a HiLoad 16/60 Superdex 75 exclusion column (GE Healthcare) with SEC buffer (50 mM Bis-Tris pH 6.7, 50 mM NaCl, 2 mM DTT). The purified proteins were then concentrated to 200–300  $\mu$ M using Millipore concentrator and aliquots were flash-frozen in 50 mM Bis-Tris pH 6.7, 50 mM NaCl, 1 mM EDTA (EthyleneDiamineTetraAcetic acid), 0.5 mM TCEP (Tris(2-CarboxyEthyl)Phosphine) and stored at -80 °C until use.

Concentration of V2R-Cter was determined by refractometry after gel filtration, while concentrations of  $\beta$ 2AR-Cter and GHSR-Cter were estimated using absorbance at 280 nm.

### 2.2. Bioinformatic Analyses

#### 2.2.1. Disorder Prediction

The disorder state of the three C-termini was determined using the Charge-Hydropathy plot computed by PONDR [24]. Disordered propensity of amino-acid were predicted using PONDR-FIT [25], PrDOS [26], SPOT-Disorder [27,28], Espritz-NMR [29], DisPro [30], and DISOPRED3 [31] predictors.

#### 2.2.2. Secondary Structure Prediction

Secondary structure predictions were computed with different servers: SOPMA [32], PSIPRED [33], JPRED4 [34], PSSpred [35], SPOT1D [36], SPIDER3 [37].

### 2.3. Size Exclusion Chromatography-Multi-Angle Light Scattering (SEC-MALS)

The experiments were performed at 25 °C using a Superdex 75 10/300 GL column (GE HealthCare) connected to a miniDAWN-TREOS light scattering detector and an Optilab T-rEX differential refractive index detector (Wyatt Technology, Santa Barbara, CA, USA). The column was equilibrated in 50 mM BisTris pH 6.7, 50 mM NaCl, 1 mM TCEP and 0.5 mM EDTA buffer filtered at 0.1 µM, and the SEC-MALS system was calibrated with a sample of Bovine Serum Albumin (BSA) at 1 mg/mL. Samples at 1.5 mM, 0.6 mM, and 0.7 mM were prepared for V2R-Cter, GHSR-Cter, and β2AR-Cter, respectively. For each GPCR-Cter, 40 µL of sample were injected at 0.5 mL/min. Data acquisition and analyses were performed using the ASTRA software (Wyatt).

### 2.4. Circular Dichroism (CD)

Far UV-spectra of the C-termini were recorded in a quartz cuvette (path length 0.1 cm) at 0.08 mg/mL in H<sub>2</sub>O at 20 °C using a Chirascan. The ellipticity was scanned from 190 to 260 nm with an increment of 0.5 nm, an integration time of 3 s, and a constant band-pass of 1 nm. Data were treated using Chirascan and, after subtraction of the buffer signal, were converted to mean residue ellipticity ( $[\theta]_{\text{MRW}}$ , mdeg.cm<sup>2</sup>.dmole<sup>-1</sup>) using equation (1) [38]:

$$[\theta]_{\text{MWR}} = [(\theta \times \text{Mw}) / (L \times C \times 10)] / (n-1) \quad (1)$$

where  $\theta$  is the ellipticity (mdeg), Mw is the molecular weight (g/mol), L is the cell length (cm), C is the protein concentration (mg/mL), and n is the number of peptide bonds.

### 2.5. Small-Angle X-ray Scattering (SAXS) Measurement and Analysis

Synchrotron radiation SAXS data were acquired for GPCR-Cters at the SWING beamline at the SOLEIL synchrotron (Saint-Aubin, France) [39] using an X-ray wavelength of 1.03 Å and a sample-to-detector distance of 1.99 m. Samples were measured at 15 °C and at two concentrations, 5 mg/mL and 10 mg/mL, for all GPCR-Cters, in 50 mM BisTris pH 6.7, 50 mM NaCl and 2 mM DTT buffer. Before exposure to X-rays, 45 µL of sample were injected into 3 mL Superdex 75 5/150 GL column (GE HealthCare) at 0.2 mL/min, pre-equilibrated into the same buffer as the samples. The intensity was measured as function of the magnitude of the scattering vector,  $s$ , using equation (2) [40]:

$$s = 4\pi \cdot \sin(\theta) / \lambda, \quad (2)$$

where  $\theta$  is the scattering angle and  $\lambda$  is the X-ray wavelength.

The scattering patterns of the buffer were recorded before the void volume of the column (1 mL). The scattering profiles measured covered a momentum transfer range of  $0.002 < s < 0.5 \text{ \AA}^{-1}$ . Data were processed using CHROMIX from ATSAS [41] software package to automatically select frames corresponding to buffer and sample, and performed buffer subtraction. The scaled and averaged SAXS curves were analyzed using Primus from ATSAS software package.

### 2.6. Nuclear Magnetic Resonance (NMR) Spectroscopy

All NMR experiments were performed on a Bruker Bruker Avance III 700 MHz spectrometer, except for the 3D assignment of β2AR-Cter performed on a 800 MHz, and for the <sup>3</sup>J<sub>H<sub>N</sub>H<sub>A</sub> of GHSR-Cter performed on a 500 MHz. The 700 MHz and 800 MHz spectrometers are equipped with a cryogenic triple-resonance (<sup>1</sup>H, <sup>15</sup>N, <sup>13</sup>C) probe and shielded z-gradients. All NMR experiments were recorded at 20 °C in a buffer (named NMR buffer) composed of 50 mM Bis-Tris pH 6.7, 150 mM NaCl, 1 mM EDTA, 0.5 mM TCEP, 5% D<sub>2</sub>O (Eurisotop), and 5 mM DSS-d6 (2,2-dimethyl-2-silapentane-5-sulfonate,</sub>

Sigma) as internal reference [42]. All experiments used the pulse sequences provided by Bruker Topspin 3.2. Squared cosine apodization was used in indirect dimensions, prior to zero-filling and Fourier transformation using TOPSPIN (version 4.0.6, Bruker) and data processing was performed using NMRFAM-SPARKY (version 1.414, [43]). For each NMR experiments, concentrations of GPCR-Cters were indicated in Table S1. For all NMR experiments, data were measured for all residues of C-terminus regions excepted proline residues, the residue A339 of  $\beta$ 2AR-Cter, and the first N-terminal residue.

### 2.6.1. Backbone Assignment

For the sequential assignment of the  $^{13}\text{C}/^{15}\text{N}$  GPCR C-terminus of V2R, GHSR and  $\beta$ 2AR, HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCA(CO)NH and HNCACB triple resonance 3D experiments were recorded.  $\text{H}_\text{N}$ , N, CO,  $\text{C}_\alpha$  and  $\text{C}_\beta$  nuclei of all residues were assigned, expected the first N-terminal residue, A339 for  $\beta$ 2AR-Cter and proline residues.

### 2.6.2. Secondary Chemical Shift (SCS)

$^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  chemical shifts were used to calculate Secondary Chemical Shift (SCS) by subtraction of experimental chemical shifts (from the 3D experiment) from random-coil chemical shift computed by POTENCI database [44,45]. SCS were calculated for all residues of C-terminal domains excepted proline residues, the N-terminal residue, A339 for  $\beta$ 2AR-Cter, and the last residue.

### 2.6.3. $^3\text{J}_{\text{HNHA}}$ Coupling

$^3\text{J}_{\text{HNHA}}$  scalar coupling measurements were obtained according to Vuister and Bax [46]. Briefly, HNHA experiments were recorded on  $^{15}\text{N}$ -labelled GPCR C-termini. Intensity of the cross-peak ( $S_{\text{cross}}$ ) and intensity of the corresponding diagonal peak ( $S_{\text{diag}}$ ) were extracted using Sparky. They were used to calculate the  $^3\text{J}_{\text{HNHA}}$  scalar coupling of each amino acid using the equation (3):

$$(S_{\text{cross}}/S_{\text{diag}}) = -\tan^2(2\pi \times ^3\text{J}_{\text{HNHA}} \times \xi) \quad (3)$$

where  $2\xi$  is the total evolution time for the homonuclear  $^3\text{J}_{\text{HNHA}}$  coupling, which has been set to 26.1 ms.

$^3\text{J}_{\text{HNHA}}$  scalar coupling were measured for all residues of C-terminal tails excepted proline residues and the first glycine residue. Random coil scalar coupling were predicted using RC\_3JHNHa server [47].

### 2.6.4. $^1\text{H}$ - $^{15}\text{N}$ Residual Dipolar Couplings (RDCs)

RDCs were obtained by recording 2D IPAP HSQC spectra [48] in isotropic and anisotropic media. The anisotropic media were obtained by adding a 5% (*w/v*) mixture of polyoxyethylene 5-lauryl ether (PEG/C12E5) (Sigma) and 1-hexanol (Sigma) in a molar ratio of 0.85 [49] or by adding ~ 20 mg/mL of filamentous phage Pf1 (Asla biotech) [50]. Spectra were recorded on  $^{15}\text{N}$ -labelled GPCR C-termini in alcohol and phage media.  $^1\text{D}_{\text{NH}}$  dipolar couplings were measured from the difference of doublet peak positions in the  $^{15}\text{N}$  dimension measured in the anisotropic (J + D) and isotropic (J) spectra.

### 2.6.5. $^{15}\text{N}$ Relaxation Experiments

Relaxation data were measured on  $^{15}\text{N}$ -labelled GPCR C-termini for all residues except proline residues and the two first N-terminal residues. Heteronuclear  $^{15}\text{N}\{^1\text{H}\}$ -NOE values were determined from two experiments with on- (saturated spectrum) and off-resonance  $^1\text{H}$  saturation (unsaturated spectrum) that were recorded in an interleaved manner. The saturation time by  $120^\circ$  pulses (~ 10 kHz) was set to 6 s and the recycle delay to 6 s. NOEs values were obtained from the ratio of intensities measured in the saturated (I) and unsaturated ( $I_0$ ) spectra. Longitudinal ( $R_1$ ) and transversal ( $R_2$ ) relaxation

rates were measured through acquisition of  $^{15}\text{N}$ -HSQC spectra with different relaxation delays: 10, 50, 100, 200, 400, 600, 800, 1000 ms for  $R_1$ , and 16, 32, 64, 96, 160, 240, 480, 640 ms for  $R_2$ . For each peak, the intensity was fitted to a single exponential decay using Sparky [43] to obtain the relaxation parameters. For all relaxation parameters, three residues at N- and C-termini were discarded from the calculation of average values due to their inherent higher flexibility.

#### 2.6.6. Paramagnetic Relaxation Enhancement (PRE)

The C378A or C406A variants of  $^{15}\text{N}$   $\beta 2\text{AR}$ -Cter were labeled on the remaining cysteine using 3-(2-Iodoacetamodi)-proxyl (Merck). Paramagnetic samples were recorded with a recycling delay of 2 s. Reference diamagnetic samples were recorded in the same conditions after the addition of 5 mM fresh ascorbic acid, pH 6.7, in the NMR tube. PRE were analyzed by measuring the peak intensity ratios ( $I_{\text{para}}/I_{\text{dia}}$ ) between two  $^{15}\text{N}$ -HSQC spectra of paramagnetic and diamagnetic samples. The theoretical profile expected for a strictly random coil polymer was calculated according to [51].

#### 2.6.7. Ensemble Calculations

Ensembles of explicit models were generated using Flexible-Meccano (FM) [52], which sequentially builds peptide planes based on amino acid specific conformational propensity and a simple volume exclusion term. To account for deviations from a random-coil description, different structure ensembles of 50,000 conformers were computed including user-defined local conformational propensities in different regions of the protein. Local conformational propensities were first localized using the consensus of all NMR data, and then were adjusted by comparing back-calculated and experimental  $^1\text{D}_{\text{HN}}$  RDCs to get the lowest  $\chi^2$  (for more details [53]). For  $\beta 2\text{AR}$ -Cter, a long-range contact of 15 Å between two regions affected by the probe, i.e., from residues 338 to 357 and from residues 367 to 386 (regions in grey in Figure S7b), was introduced to get a better agreement between back-calculated and experimental  $^1\text{D}_{\text{HN}}$  RDCs.

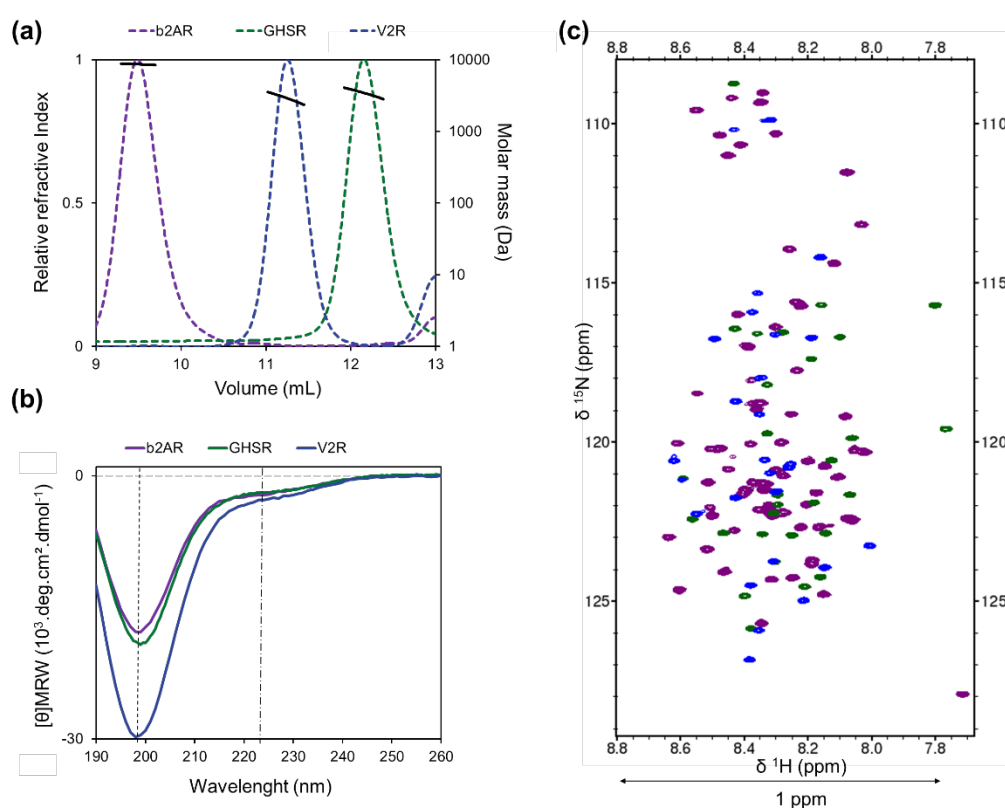
### 3. Results

#### 3.1. The Disordered C-Termini of V2R, GHSR, and $\beta 2\text{AR}$ Contain Transient Secondary Structures

As is the case for many C-terminal domains of GPCRs [54], the C-terminal regions of V2R, GHSR, and  $\beta 2\text{AR}$  are shown to be disordered (Figure 2) and are predicted to contain transient secondary structures by various computational tools (Figure 3). In fact, they are composed of more than 50% of disorder-promoting residues (Arg, Gly, Gln, Ser, Pro, Glu, and Lys) (Figure 3a) [55]. In the Charge-Hydrophathy plot [56] (Figure S1), V2R-Cter is at the disorder-order boundary, while  $\beta 2\text{AR}$ -Cter (pink) and GHSR-Cter (green) appeared in the cluster of disordered proteins (yellow). Sequence based disorder prediction by a set of six predictors (Table S2) showed an overall disordered state (values higher than 0.5) for the three C-termini (Figure 3b). More specifically, prediction by PrDOS, DISOPRED3, and Espritz-NMR showed less disordered regions from residues ~ 351 to 365 for V2R-Cter and from residues ~ 343 to 359 for GHSR-Cter (Figure 3b).  $\beta 2\text{AR}$ -Cter is also predicted to be less disordered from ~ 345 to 356 and from ~ 369 to 405 according to PrDOS, DISOPRED3, and Espritz-NMR predictors (Figure 3b). Interestingly, these regions are predicted to contain helical conformations by a set of distinct secondary structure predictors (Figure 3c, Table S3).

In order to characterize these domains experimentally, we expressed and purified the intrinsically disordered and soluble C-terminal regions of V2R (343–371), GHSR (339–355), and  $\beta 2\text{AR}$  (342–413). SEC-MALS analysis revealed a single elution peak for each protein at a volume that corresponded to the volume of standard proteins with a molecular mass greater than 13 kDa. However, the masses derived from a MALS analysis are 3.9 ( $\pm 3.2\%$ ), 3.4 ( $\pm 3.7\%$ ), and 8.6 ( $\pm 1.1\%$ ) kDa for V2R-Cter, GHSR-Cter, and

$\beta$ 2AR-Cter, respectively (Figure 2a). This is in agreement with the expected molecular weight of their monomeric forms (3.3, 3.6, and 8.2 kDa, respectively). Molecular masses were confirmed by mass spectrometry (MS) with 3.264, 3.672, and 8.175 kDa obtained for V2R-Cter, GHSR-Cter, and  $\beta$ 2AR-Cter, respectively. The behavior of V2R-Cter, GHSR-Cter, and  $\beta$ 2AR-Cter on a SEC column is typical of a disordered protein, with a smaller elution volume than expected for globular proteins of the same molecular mass [57]. In addition, far UV CD spectra showed a minimum around 198 nm, characteristic of unfolded proteins [58] (Figure 2b), and a negative shoulder around 222 nm, suggesting the presence of residual secondary structures [59] (Figure 2b). The Kratky plots extracted from SAXS data were typical of disordered regions with no clear maximum and a monotonic increase along the momentum transfer range (Figure S2) [18,60,61]. Additionally,  $^{15}\text{N}$ -HSQC spectra of the three C-terminal regions showed a reduced amide proton spectral dispersion (around 1 ppm) typical of a disordered protein (Figure 2c) [18,19].



**Figure 2. The C-terminal domain of V2R-Cter (blue), GHSR-Cter (green), and  $\beta$ 2AR-Cter (purple) are monomeric and disordered.** The color code is maintained for all panels. (a) SEC-MALS curves (colored dashed lines). Molar masses derived from MALS are indicated by thick black line (right axis); (b) far-UV circular dichroism (CD) spectra (colored lines) present a minimum around 200 nm (black dashed line) and a shoulder at 220 nm (black dashed and dotted line) characteristic of disordered proteins with transient secondary structure content; (c)  $^{15}\text{N}$ -HSQC spectra display a low proton spectral dispersion ( $\sim 1$  ppm) typical of IDPs. HSQCs were recorded on 300  $\mu\text{M}$  samples at 700 MHz and 20  $^{\circ}\text{C}$ , in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.

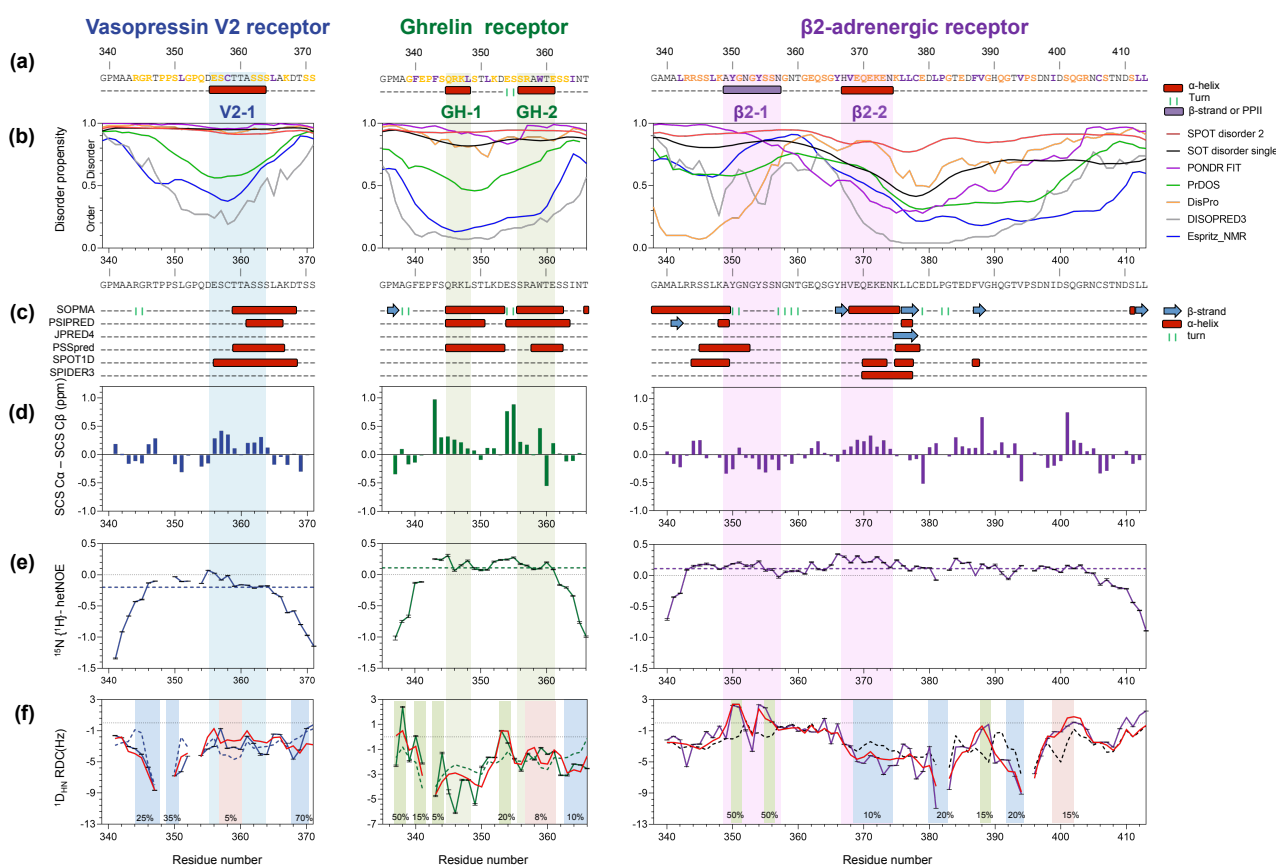
Altogether, bioinformatics analyses and experimental data confirm the disordered nature of the C-terminal regions of V2R, GHSR, and  $\beta$ 2AR [62]. Furthermore, these results suggest the presence of residual secondary structures.

### 3.2. Location of the Transient Secondary Structures in V2R, GHSR, and $\beta$ 2AR C-Termini

In order to localize the transient secondary structures of these three GPCR C-terminal regions, we performed a NMR study, as described in [53]. Before NMR in-



investigation, the backbone assignments were performed on the studied C-terminal domains of GPCRs (Figure S3) (BMRB accession codes, respectively: 51318, 51317, and 51316). Then, we used a consensus of four NMR parameters to highlight the secondary structure content. First,  $^{13}\text{C}$  secondary chemical shifts (SCS), which are highly sensitive to the backbone conformations, were computed (Figure 3d and S4b–g) [45,63]. Three-bond  $\text{H}_\text{N}\text{-H}_\alpha$  J-coupling constants ( $^3J_{\text{HNHA}}$ ), which are related to the  $\phi$  angles of the polypeptide chain, were compared to random coil scalar coupling [46,47]. Residual dipolar couplings (RDCs), which are related to the orientation of the amide backbone vector to the magnetic field [64], give information on the location and secondary structure type. These data were also compared to back-calculated RDCs on random-coil ensembles computed with Flexible-Meccano [52,65]. Finally, dynamic parameters, such as heteronuclear  $^{15}\text{N}\{^1\text{H}\}$ -NOEs, longitudinal ( $R_1$ ), and transverse ( $R_2$ ) relaxation rates, give information on local backbone mobility on the ps–ns timescale, while  $R_2$  is also sensitive to motions on the  $\mu\text{s}$  to ms timescale (chemical or conformational exchange processes). Thus,  $R_2/R_1$  ratio reveals slow conformational motions or conformational/chemical exchange. Depending on the NMR experiments, the GPCR-Cter sample concentrations varied, but did not affect the structure (Table S1, Figure S3).



**Figure 3. Bioinformatics predictions, secondary structure propensity, backbone dynamics, and RDC conformational profile of the disordered V2R-Cter (blue), GHSR-Cter (green), and  $\beta$ 2AR-Cter (purple).** (a) GPCR C-termini are composed of more than 50% disorder-promoting residues (in orange). Ordered-promoting residues are indicated in purple. Secondary structures obtained by the consensus of all NMR data are indicated under the sequence and highlighted according to their respective color code. Helix,  $\beta$ -strand, extended conformation ( $\beta$ -strand or polyproline helix 2, PPII), and turns are represented as red cylinder, blue arrow, purple box, and green bars, respectively; (b) disorder prediction by SPOT-Disorder 2 (red), SPOT-Disorder Single (black), PONDR-FIT (purple), PrDOS (green), DisPro (orange), DISOPRED 3 (grey), and Espritz-NMR (blue). The disorder/order threshold (0.5) is indicated in black line; (c) secondary structure prediction by SOPMA, PSIPRED, JPRED4, PSSpred, SPOT 1D, and SPIDER 3 web-servers are represented for helices (red cylinder), strands (blue arrow), and turns (green); (d) computed secondary

structure SCS  $C\alpha$  - SCS  $C\beta$  using random coil chemical shifts from POTENCI; (e) heteronuclear  $^{15}\text{N}\{^1\text{H}\}$ -NOEs are represented with colored curves and the average with dash colored lines; (f) back-calculated and experimental  $^1\text{D}_{\text{NH}}$  RDCs. Comparison of experimental RDC (colored according to the GPCR-Cter color code) and back-calculated values computed with FM on a random-coil ensemble (dash color line) and biased ensemble (red line) and populated as indicated in the figure for helical structure (highlighted in red), extended structure (highlighted in blue), or turn (highlighted in green). Spectra were recorded at 700 MHz and 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.

### 3.2.1. Vasopressin V2 Receptor C-Terminal Domain (V2R-Cter)

For V2R-Cter, a central helical region was identified from residues 356 to 364, called V2-1 from now on (Figure 3). Indeed,  $^{13}\text{C}$  SCS presented positive values in this region (Figure 3d). The presence of this helix is in agreement with disorder and secondary structure predictions (Figure 3b,c). Globally, in V2R-Cter, 32% of the residues displayed  $^3\text{J}_{\text{HNHA}}$  scalar coupling values below 6 Hz, normally assigned to helical conformations, and the rest were between 6 and 8 Hz, consistent with a random coil (RC) polypeptide chain (Figure S4i). Moreover, experimental  $^3\text{J}_{\text{HNHA}}$  scalar couplings for the V2-1 region were lower than the predicted ones (Figure S4h) [47], substantiating its helical conformation. Further evidence for transient secondary structure elements in V2-1 came from dynamical parameters. V2R-Cter showed low heteronuclear  $^{15}\text{N}\{^1\text{H}\}$ -NOE values as expected for a disordered polypeptide. However, in the V2-1 region, heteronuclear NOE values were above the average ( $-0.09 \pm 0.01$ ), suggesting an enhanced rigidity in the region (Figure 3e). Residues of V2-1 adopted larger  $R_2$  values than the average ( $2.80 \pm 0.05$  Hz), suggesting the presence of slow dynamic processes.  $R_1$  and  $R_2/R_1$  also displayed slightly higher values for V2-1 than their averages,  $1.75 \pm 0.03$  Hz and  $1.59 \pm 0.04$ , respectively (Figure S5). This transient helical secondary structure in V2-1 was also highlighted by  $^1\text{D}_{\text{NH}}$  RDC values. RDCs were measured in alcohol mixture ( $\text{C}_{12}\text{E}_5/\text{hexanol}$  media) and in filamentous bacteriophage Pf1. However, the homogeneity of the aligned sample was checked on the quadrupolar splitting of  $\text{D}_2\text{O}$  and was higher in alcohol medium (27 Hz) than in Pf1 medium. While RDC values were mainly negative, as observed in disordered proteins, the segment encompassing V2-1 displayed higher values than those expected for a random coil, which was in agreement with the presence of a helix in this region (Figure 3f).

### 3.2.2. Ghrelin Receptor C-Terminal Domain (GHSR-Cter)

For GHSR-Cter we identified two transient secondary structures: from residues 345 to 348 (called GH-1) and from residues 356 to 361 (called GH-2) (Figure 3). These regions were predicted to form helices (Figure 3c) and SCS showed positive values, suggesting helical conformations (Figure 3d). The SCS profile also presented high positive values for 354-355 that could be related to the presence of a turn as predicted by the secondary structure predictor SOPMA.  $^3\text{J}_{\text{HNHA}}$  scalar couplings confirmed that GHSR-Cter was mostly disordered (50%) with transient helical conformations (46%) and a small portion of extended conformation (4%) (Figure S4i). The comparison of experimental and random coil  $^3\text{J}_{\text{HNHA}}$  scalar couplings was in agreement with the presence of transient helices in GH-1 and GH-2 (Figure S4h). Additionally, in these two regions, heteronuclear NOE,  $R_1$ ,  $R_2$ , and  $R_2/R_1$  showed higher values than their averages,  $0.11 \pm 0.01$ ;  $2.11 \pm 0.04$  Hz;  $3.17 \pm 0.06$  Hz and  $1.49 \pm 0.04$ , respectively (Figure 3e and S5). This indicated less flexibility in these regions, suggesting some local structuration. When measuring RDCs, a small interaction between the alcohol mixture and GH-1 region (chemical shift differences  $> 0.01$  ppm between  $^{15}\text{N}$ -HSQC spectra) was observed (Figure S6); thus we analyzed RDCs measured in phage Pf1 medium (quadrupolar splitting of 18 Hz).  $^1\text{D}_{\text{NH}}$  RDCs exhibited lower (GH-1 region) and higher (GH-2 region) values than those expected from a random coil, suggesting an extended and a helical conformation in these regions, respectively (Figure 3f).

### 3.2.3. $\beta$ 2-Adrenergic Receptor C-Terminal Domain ( $\beta$ 2AR-Cter)

For  $\beta$ 2AR-Cter, we identified two regions forming secondary structures: from residues 349 to 357 (called  $\beta$ 2-1) and from 368 to 376 (called  $\beta$ 2-2). They were predicted to be more ordered and to contain helical secondary structures (Figure 3a,b). However, SCS presented negative values for  $\beta$ 2-1 and positive values for  $\beta$ 2-2, corresponding to an extended and a helical conformation, respectively (Figure 3d).  $^3J_{\text{HNHA}}$  scalar couplings were consistent with an overall unstructured protein (73% of the residues) with a small content of helical (25%) and extended conformations (1%) (Figure S4i). In  $\beta$ 2-2, experimental scalar couplings were lower than those predicted for a random coil, suggesting a helical conformation in this region. Note that glycines are not predicted by the RC\_3JHNHa server; thus scalar couplings in  $\beta$ 2-1 could not be properly compared to random-coil values [47] (Figure S4h). In  $\beta$ 2-1 and  $\beta$ 2-2, heteronuclear NOE and  $R_2$  values were above their average,  $0.11 \pm 0.01$  and  $3.44 \pm 0.01$  Hz, respectively. This indicated restricted flexibility in these regions (Figure 2e and S6). However, in  $\beta$ 2-2,  $R_1$  remained flat. Consequently,  $R_2/R_1$  displayed larger values than the average ( $1.83 \pm 0.03$ ), suggesting conformational fluctuations on the  $\mu\text{s}$ – $\text{ms}$  time-scale. RDC measurement in Pf1 medium showed a small interaction with  $\beta$ 2-2 (chemical shift differences  $> 0.01$  ppm) (Figure S6); thus we used RDC data extracted from the alcohol mixture (quadrupolar splitting of 31 Hz). Compared to random coil values computed with FM, experimental RDCs showed more positive values in  $\beta$ 2-1 and more negative values in  $\beta$ 2-2 (Figure 3f), suggesting a helical and an extended conformation, respectively. However, RDCs are not only sensitive to local structuration, such as SCS, but they can also probe long-range transient contacts [66,67]. Indeed, paramagnetic relaxation enhancement (PRE) data of the C406A variant, where the paramagnetic probe in C378 lies just after  $\beta$ 2-2 region, showed reduction of the intensity ratio in the N-terminal part of  $\beta$ 2AR-Cter, including  $\beta$ 2-1 region and its N-flanking region, revealing long-range contacts between the N-terminal part of  $\beta$ 2AR-Cter and  $\beta$ 2-2 C-flanking region (respectively, from residues 338 to 357 and 367 to 386). PRE affected regions are highlighted in grey in Figure S7b. These long-range interactions could affect the overall profile of RDCs, probably explaining the absence of consensus between SCS and RDC data (see below). The second PRE dataset, measured in the C378A variant, where the paramagnetic probe lies at the C-terminal region (C406), did not induce substantial reduction of the intensity ratio in  $\beta$ 2AR-Cter, indicating the overall disorder of the protein. Interestingly, a slight reduction of intensity in the N-terminal part  $\beta$ 2AR-Cter confirmed the presence of fuzzy long-range contacts between the N- and C-terminal parts of the protein (Figure S7c).

### 3.2.4. Conformational Ensemble of GPCR-Cters

To further illuminate the presence of transient secondary structures or turns in the three C-terminal domains of GPCRs, we built biased ensembles using Flexible Meccano (FM) (Figure 3f). Ensembles of 50,000 conformers were built for V2R-Cter, GHSR1a-Cter, and  $\beta$ 2AR-Cter, respectively, and were used to compute back-calculated RDCs (Figure 3f). In these ensembles, we added as constraints the transient secondary structures determined by the consensus analyses of all NMR data to improve the agreement between experimental and back-calculated RDCs. Then, the population of the local conformational propensities and turns were adjusted by monitoring the agreement using  $\chi^2$ . The quality of these ensembles was evaluated by comparing the back-calculated RDCs with the experimental ones and was optimized until obtaining the lowest possible  $\chi^2$  (near 1). RDCs were measured in alcohol mixture for V2R-Cter and  $\beta$ 2AR-Cter or in phage Pf1 medium for GHSR1a-Cter. For V2R-Cter, the best agreement ( $\chi^2 = 1.41$ ) was obtained when 5% of  $\alpha$ -helix was imposed in V2-1 and when poly-proline helices II (PPII) were added at 25%, 35%, and 70% in position 344–347, 349–350, and 368–370, respectively. For GHSR1a-Cter, the best ensemble ( $\chi^2 = 1.49$ ) was obtained by adding 8% of  $\alpha$ -helix in GH-2 and 20% of type 1  $\beta$ -turn that was predicted for 353 to 354. Additional secondary

structures were added (residues 337–338: 50% of type I  $\beta$ -turns, 340–341: 15% of type I  $\beta$ -turns; 343–344: 5%  $\gamma$ -turn; 363–366: 10% of  $\beta$ -strand). For  $\beta$ 2AR-Cter, the lowest  $\chi^2$  ( $\chi^2 = 1.36$ ) was obtained when the long-range contact (15 Å) identified with PRE between regions surrounding  $\beta$ 2-1 (from residues 338 to 357) and  $\beta$ 2-2 (from residues 367 to 386) was incorporated to the model (see details in Materials and Methods 2.6.7., and PRE affected regions are highlighted in grey in Figure S7b). These results substantiate the presence of transient long-range interactions in  $\beta$ 2AR-Cter. In  $\beta$ 2-1, two type I  $\beta$ -turns at 50% were added, and in  $\beta$ 2-2, a type II Poly-Proline helix (PPII) at 10% was imposed. This ensemble was also constrained by two other PPII in position 380–383 and 391–394, a type I  $\beta$ -turn in position 388–389 and a helix from residues 399 to 402.

In all C-terminal domains, the introduction of these secondary structures resulted in a better description of the experimental data, suggesting a more accurate description of the conformational ensembles.

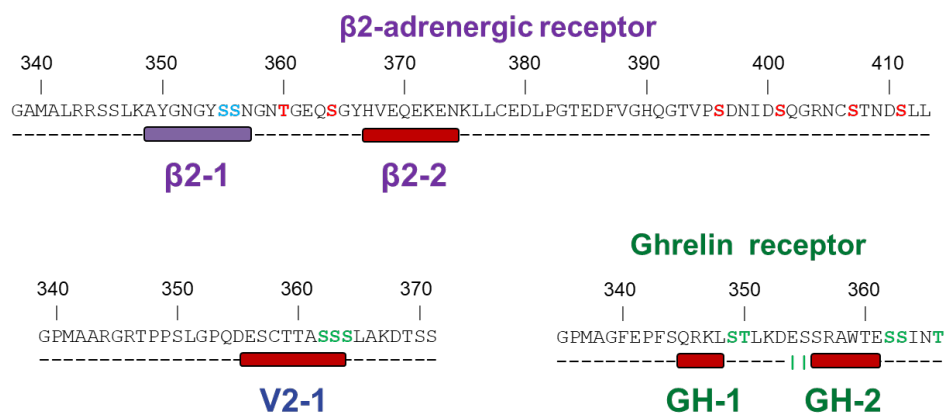
#### 4. Discussion and Conclusions

All GPCRs share a conserved and folded 7TM domain involved in the signal transmission. Conversely, the extracellular regions (N-terminal and C-terminal domains and loops) are rather diversified in length and sequence [4] and are involved in the functional properties of GPCRs, such as, respectively, ligand and partner binding. It is interesting to note that these extracellular regions are predicted to contain intrinsically disordered regions (IDRs) [5,6], which could play key roles in GPCR interaction (for review [68,69]). Indeed, IDRs explore an astronomical number of conformations in solution that we assume in fast equilibrium, and very often contain pre-formed secondary structure elements. In the majority of cases, these transient secondary structures, or short linear motifs (SLIMs), are involved in the binding process with their partner [15], which makes IDPs very well suited agents for signaling processes, such as arrestin:GPCR interaction. Thus, the characterization of the structural features of these extracellular regions of GPCRs is crucial to reveal the molecular basis of signaling and cell regulation [70]. For instance, it was reported that the truncated C-terminal domain of GPR50 (GPR50-Cter) translocates to the nucleus and directly regulates gene transcription; thus the cleavage of this Cter has an unconventional signaling mode of GPCRs [70]. Intriguingly, this GPR50 receptor has been deeply remodeled through evolution by the mutation of numerous residues and by the addition of a long C-terminal domain. Indeed, the GPR50 homolog found in lower vertebrates was lacking the Cter of the human GPR50 [71]. This illustrates how the study of the isolated C-terminal domain of GPCRs is of relevant importance per se for understanding the multitude of signaling pathways regulated by the C-terminal domains of GPCR.

With the hypothesis that the C-terminal domains of GPCRs contain partially structured elements involved in signaling pathways, we characterized the free state of the C-terminal domains from three commonly studied GPCRs, the vasopressin V2 receptor (V2R-Cter), the ghrelin receptor type 1a (GHSR-Cter) and the  $\beta$ 2-adrenergic receptor ( $\beta$ 2AR-Cter). These three class-A receptors are important therapeutic targets [2] and present different affinities for arrestins [16].

The structural characterization of these disordered C-termini was challenging, due to their inherent flexibility, and required the synergistic application of several biophysical tools. Firstly, by SAXS, MALS, and CD, we confirmed that V2R-Cter, GHSR-Cter, and  $\beta$ 2AR-Cter are IDRs containing secondary structures. Then, we used NMR in combination with computationally generated ensembles, to locate and identify these residual secondary structure elements. The C-terminus of V2R displayed a central helix from residues 356 to 364 (V2-1). The C-terminus of GHSR encompassed two helices from 345 to 348 (GH-1) and from residues 356 to 361 (GH-2). The C-terminus of  $\beta$ 2AR contained two structured regions: an extended conformation from 349 to 357 residues ( $\beta$ 2-1) and a helix from 368 to 376 ( $\beta$ 2-2) (Figure 4a). Moreover, the propensities of these secondary structures were low (~25% on average), which illustrate the high flexibility of

GPCR-Cters and their ability to adopt distinct conformations in solution, a general feature of IDPs/IDRs.



**Figure 4.** The C-terminal regions of β2AR-Cter (purple), GHSR-Cter (green), and V2R-Cter (blue) revealed different transient secondary structures, which could be involved in arrestin binding. GRK2 (red) and GRK6 (blue) phosphorylated sites according to [9] for β2AR, [72] for GHSR, and [73] for V2R are indicated in the sequence. Secondary structures obtained by NMR secondary structure consensus are indicated under the sequence, according to Figure 3.

V2R-Cter, GHSR-Cter, and β2AR-Cter are variable in length and in sequence (Figure 1). However, it is interesting to note that their residual secondary structures either encompass or are next to residues known to be phosphorylated by GRKs (Figure 4a). It is accepted that the phosphates of the C-terminal domains of GPCRs are important for the interaction with arrestin (for a review see [10]). This strongly suggests that the transient secondary structures that we have identified are directly involved in arrestin binding.

Interestingly, V2-1 region has been characterized in complex with arrestin in crystallographic and cryo-EM structures (Figure 5) [21,22,74]. These structures were obtained using a fully synthetic phospho-peptide of V2R C-terminal domain (V2Rpp), truncated or attached to other receptors (chimeric receptor) (Figure 5), and each of these complexes was stabilized with a Fab30 antibody. In these structures, the central region of V2Rpp adopts a β-strand that interacts with the N-domain of arrestin. Here, we show that in solution, this SLiM displays a helical structure instead of an extended conformation. This result suggests that a conformational change must occur upon binding to arrestin and/or phosphorylation, a feature that has been found in several SLiMs [75]. Moreover, this structural transition might be at the basis of the molecular regulation of GPCR:arrestin interaction. Indeed, β-strand formation has been proposed to serve as a general mechanism by which arrestins recognize the phosphorylated carboxy-terminal domains of receptors [21]. In the pre-structuration profile of GHSR-Cter and β2AR-Cter, only GH-2 and β2-2 are in helical conformations as found for the V2-1 region. Thus, we can hypothesize that these regions interact with arrestin with a similar mechanism to the one found for V2R, and that their basal conformation changes upon arrestin binding and/or phosphorylation.



**Data Availability Statement:** The  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  assignment of  $\beta 2\text{AR}$ , GHSR and V2R C-termini are available in the BioMagResBank (<http://www.bmrb.wisc.edu>) under the respective BMRB accession codes 51316, 51317 and 51318 (deposited on 9 February 2022).

**Acknowledgments:** The authors are grateful to Baptiste Legrand for CD experiment, Pierre Sanchez and Gilles Subra for MS support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bockaert, J.; Philippe Pin, J. Molecular Tinkering of G Protein-Coupled Receptors: An Evolutionary Success. *EMBO J.* **1999**, *18*, 1723–1729. <https://doi.org/10.1093/emboj/18.7.1723>.
2. Schöneberg, T.; Schulz, A.; Biebermann, H.; Hermsdorf, T.; Römler, H.; Sangkuhl, K. Mutant G-Protein-Coupled Receptors as a Cause of Human Diseases. *Pharmacol. Ther.* **2004**, *104*, 173–206. <https://doi.org/10.1016/j.pharmthera.2004.08.008>.
3. Hauser, A.S.; Attwood, M.M.; Rask-Andersen, M.; Schiöth, H.B.; Gloriam, D.E. Trends in GPCR Drug Discovery: New Agents, Targets and Indications. *Nat. Rev. Drug Discov.* **2017**, *16*, 829–842. <https://doi.org/10.1038/nrd.2017.178>.
4. Fredriksson, R. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol. Pharmacol.* **2003**, *63*, 1256–1272. <https://doi.org/10.1124/mol.63.6.1256>.
5. Lagerström, M.C.; Schiöth, H.B. Structural Diversity of G Protein-Coupled Receptors and Significance for Drug Discovery. *Nat. Rev. Drug Discov.* **2008**, *7*, 339–357. <https://doi.org/10.1038/nrd2518>.
6. Fonin, A.V.; Darling, A.L.; Kuznetsova, I.M.; Turoverov, K.K.; Uversky, V.N. Multi-Functionality of Proteins Involved in GPCR and G Protein Signaling: Making Sense of Structure-Function Continuum with Intrinsic Disorder-Based Proteoforms. *Cell Mol. Life Sci.* **2019**, *76*, 4461–4492. <https://doi.org/10.1007/s00018-019-03276-1>.
7. Gurevich, V.V.; Gurevich, E.V. GPCR Signaling Regulation: The Role of GRKs and Arrestins. *Front. Pharmacol.* **2019**, *10*, 125. <https://doi.org/10.3389/fphar.2019.00125>.
8. Ren, X.-R.; Reiter, E.; Ahn, S.; Kim, J.; Chen, W.; Lefkowitz, R.J. Different G Protein-Coupled Receptor Kinases Govern G Protein and Beta-Arrestin-Mediated Signaling of V2 Vasopressin Receptor. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1448–1453. <https://doi.org/10.1073/pnas.0409534102>.
9. 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. Distinct Phosphorylation Sites on the 2-Adrenergic Receptor Establish a Barcode That Encodes Differential Functions of -Arrestin. *Sci. Signal.* **2011**, *4*, ra51. <https://doi.org/10.1126/scisignal.2001707>.
10. Gurevich, V.V.; Gurevich, E.V. The Structural Basis of the Arrestin Binding to GPCRs. *Mol. Cell. Endocrinol.* **2019**, *484*, 34–41. <https://doi.org/10.1016/j.mce.2019.01.019>.
11. Venkatakrisnan, A.; Flock, T.; Prado, D.E.; Oates, M.E.; Gough, J.; Madan Babu, M. Structured and Disordered Facets of the GPCR Fold. *Curr. Opin. Struct. Biol.* **2014**, *27*, 129–137. <https://doi.org/10.1016/j.sbi.2014.08.002>.
12. Jaakola, V.-P.; Prilusky, J.; Sussman, J.L.; Goldman, A. G Protein-Coupled Receptors Show Unusual Patterns of Intrinsic Unfolding. *Protein Eng. Des. Sel.* **2005**, *18*, 103–110. <https://doi.org/10.1093/protein/gzi004>.
13. Uversky, V.N. What Does It Mean to Be Natively Unfolded? *Eur. J. Biochem.* **2002**, *269*, 2–12. <https://doi.org/10.1046/j.0014-2956.2001.02649.x>.
14. Wright, P.E.; Dyson, H.J. Intrinsically Disordered Proteins in Cellular Signalling and Regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29. <https://doi.org/10.1038/nrm3920>.
15. Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. Preformed Structural Elements Feature in Partner Recognition by Intrinsically Unstructured Proteins. *J. Mol. Biol.* **2004**, *338*, 1015–1026. <https://doi.org/10.1016/j.jmb.2004.03.017>.
16. Oakley, R.H.; Laporte, S.A.; Holt, J.A.; Caron, M.G.; Barak, L.S. Differential Affinities of Visual Arrestin, BARrestin1, and BARrestin2 for G Protein-Coupled Receptors Delineate Two Major Classes of Receptors. *J. Biol. Chem.* **2000**, *275*, 17201–17210. <https://doi.org/10.1074/jbc.M910348199>.
17. Dyson, H.J.; Wright, P.E. Unfolded Proteins and Protein Folding Studied by NMR. *Chem. Rev.* **2004**, *104*, 3607–3622. <https://doi.org/10.1021/cr030403s>.
18. Sibille, N.; Bernadó, P. Structural Characterization of Intrinsically Disordered Proteins by the Combined Use of NMR and SAXS. *Biochem. Soc. Trans.* **2012**, *40*, 955–962. <https://doi.org/10.1042/BST20120149>.
19. Delaforge, E.; Cordeiro, T.N.; Bernadó, P.; Sibille, N. Conformational Characterization of Intrinsically Disordered Proteins and Its Biological Significance. In *Modern Magnetic Resonance*; Webb, G.A., Ed.; Springer International Publishing: Cham, Switzerland, 2018; pp. 381–399, ISBN 978-3-319-28388-3.
20. Milles, S.; Salvi, N.; Blackledge, M.; Jensen, M.R. Characterization of Intrinsically Disordered Proteins and Their Dynamic Complexes: From in Vitro to Cell-like Environments. *Prog. Nucl. Magn. Reson. Spectrosc.* **2018**, *109*, 79–100. <https://doi.org/10.1016/j.pnmrs.2018.07.001>.
21. 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. Structure of Active  $\beta$ -Arrestin1 Bound to a G Protein-Coupled Receptor Phosphopeptide. *Nature* **2013**, *497*, 137–141. <https://doi.org/10.1038/nature12120>.

22. Staus, D.P.; Hu, H.; Robertson, M.J.; Kleinhenz, A.L.W.; Wingler, L.M.; Capel, W.D.; Latorraca, N.R.; Lefkowitz, R.J.; Skiniotis, G. Structure of the M2 Muscarinic Receptor- $\beta$ -Arrestin Complex in a Lipid Nanodisc. *Nature* **2020**, *579*, 297–302. <https://doi.org/10.1038/s41586-020-1954-0>.
23. Nguyen, A.H.; Thomsen, A.R.B.; Cahill, T.J.; Huang, R.; Huang, L.-Y.; Marcink, T.; Clarke, O.B.; Heissel, S.; Masoudi, A.; Ben-Hail, D.; et al. Structure of an Endosomal Signaling GPCR-G Protein- $\beta$ -Arrestin Megacomplex. *Nat. Struct. Mol. Biol.* **2019**, *26*, 1123–1131. <https://doi.org/10.1038/s41594-019-0330-y>.
24. Romero, P.; Obradovic, Z.; Li, X.; Garner, E.C.; Brown, C.J.; Dunker, A.K. Sequence Complexity of Disordered Protein. *Proteins* **2001**, *42*, 38–48. [https://doi.org/10.1002/1097-0134\(20010101\)42:1<38::aid-prot50>3.0.co;2-3](https://doi.org/10.1002/1097-0134(20010101)42:1<38::aid-prot50>3.0.co;2-3).
25. Xue, B.; Dunbrack, R.L.; Williams, R.W.; Dunker, A.K.; Uversky, V.N. PONDR-FIT: A Meta-Predictor of Intrinsically Disordered Amino Acids. *Biochim. Biophys. Acta (BBA) Proteins Proteom.* **2010**, *1804*, 996–1010. <https://doi.org/10.1016/j.bbapap.2010.01.011>.
26. Ishida, T.; Kinoshita, K. PrDOS: Prediction of Disordered Protein Regions from Amino Acid Sequence. *Nucleic Acids Res.* **2007**, *35*, W460–W464. <https://doi.org/10.1093/nar/gkm363>.
27. Hanson, J.; Paliwal, K.; Zhou, Y. Accurate Single-Sequence Prediction of Protein Intrinsic Disorder by an Ensemble of Deep Recurrent and Convolutional Architectures. *J. Chem. Inf. Model.* **2018**, *58*, 2369–2376. <https://doi.org/10.1021/acs.jcim.8b00636>.
28. Hanson, K.; Paliwal, K.; Litfin, T.; Zhou, Y. SPOT-Disorder2: Improved Protein Intrinsic Disorder Prediction by Ensembled Deep Learning—ScienceDirect. Available online: <https://www.sciencedirect.com/science/article/pii/S1672022920300243> (accessed on 29 June 2021).
29. Walsh, I.; Martin, A.J.M.; Di Domenico, T.; Tosatto, S.C.E. ESpritz: Accurate and Fast Prediction of Protein Disorder. *Bioinformatics* **2012**, *28*, 503–509. <https://doi.org/10.1093/bioinformatics/btr682>.
30. Cheng, J.; Sweredoski, M.J.; Baldi, P. Accurate Prediction of Protein Disordered Regions by Mining Protein Structure Data. *Data Min. Knowl. Disc.* **2005**, *11*, 213–222. <https://doi.org/10.1007/s10618-005-0001-y>.
31. Jones, D.T.; Cozzetto, D. DISOPRED3: Precise Disordered Region Predictions with Annotated Protein-Binding Activity. *Bioinformatics* **2015**, *31*, 857–863. <https://doi.org/10.1093/bioinformatics/btu744>.
32. Geourjon, C.; Deléage, G. SOPMA: Significant Improvements in Protein Secondary Structure Prediction by Consensus Prediction from Multiple Alignments. *Bioinformatics* **1995**, *11*, 681–684. <https://doi.org/10.1093/bioinformatics/11.6.681>.
33. Buchan, D.W.A.; Jones, D.T. The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Res.* **2019**, *47*, W402–W407. <https://doi.org/10.1093/nar/gkz297>.
34. Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G.J. JPred4: A Protein Secondary Structure Prediction Server. *Nucleic Acids Res.* **2015**, *43*, W389–W394. <https://doi.org/10.1093/nar/gkv332>.
35. Yan, R.; Xu, D.; Yang, J.; Walker, S.; Zhang, Y. A Comparative Assessment and Analysis of 20 Representative Sequence Alignment Methods for Protein Structure Prediction. *Sci. Rep.* **2013**, *3*, 2619. <https://doi.org/10.1038/srep02619>.
36. Hanson, J.; Paliwal, K.; Litfin, T.; Yang, Y.; Zhou, Y. Improving Prediction of Protein Secondary Structure, Backbone Angles, Solvent Accessibility and Contact Numbers by Using Predicted Contact Maps and an Ensemble of Recurrent and Residual Convolutional Neural Networks. *Bioinformatics* **2019**, *35*, 2403–2410. <https://doi.org/10.1093/bioinformatics/bty1006>.
37. Heffernan, R.; Yang, Y.; Paliwal, K.; Zhou, Y. Capturing Non-Local Interactions by Long Short-Term Memory Bidirectional Recurrent Neural Networks for Improving Prediction of Protein Secondary Structure, Backbone Angles, Contact Numbers and Solvent Accessibility. *Bioinformatics* **2017**, *33*, 2842–2849. <https://doi.org/10.1093/bioinformatics/btx218>.
38. Chemes, L.B.; Alonso, L.G.; Noval, M.G.; de Prat-Gay, G. Circular Dichroism Techniques for the Analysis of Intrinsically Disordered Proteins and Domains. In *Intrinsically Disordered Protein Analysis: Volume 1, Methods and Experimental Tools*; Uversky, V.N., Dunker, A.K., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, USA, 2012; pp. 387–404, ISBN 978-1-61779-927-3.
39. Thureau, A.; Roblin, P.; Pérez, J. BioSAXS on the SWING Beamline at Synchrotron SOLEIL. *J. Appl. Crystallogr.* **2021**, *54*, 1698–1710. <https://doi.org/10.1107/S1600576721008736>.
40. Blanchet, C.E.; Svergun, D.I. Small-Angle X-ray Scattering on Biological Macromolecules and Nanocomposites in Solution. *Annu. Rev. Phys. Chem.* **2013**, *64*, 37–54. <https://doi.org/10.1146/annurev-physchem-040412-110132>.
41. Manalastas-Cantos, K.; Konarev, P.V.; Hajizadeh, N.R.; Kikhney, A.G.; Petoukhov, M.V.; Molodenskiy, D.S.; Panjkovich, A.; Mertens, H.D.T.; Gruzinov, A.; Borges, C.; et al. ATSAS 3.0: Expanded Functionality and New Tools for Small-Angle Scattering Data Analysis. *J. Appl. Crystallogr.* **2021**, *54*, 343–355. <https://doi.org/10.1107/S1600576720013412>.
42. Wishart, D.S.; Sykes, B.D. [12] Chemical Shifts as a Tool for Structure Determination. In *Methods in Enzymology*; Elsevier: New York, NY, USA, 1994; Volume 239, pp. 363–392, ISBN 978-0-12-182140-1.
43. Lee, W.; Tonelli, M.; Markley, J.L. NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* **2015**, *31*, 1325–1327. <https://doi.org/10.1093/bioinformatics/btu830>.
44. Tamiola, K.; Acar, B.; Mulder, F.A.A. Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010**, *132*, 18000–18003. <https://doi.org/10.1021/ja105656t>.
45. Nielsen, J.T.; Mulder, F.A.A. POTENCI: Prediction of Temperature, Neighbor and PH-Corrected Chemical Shifts for Intrinsically Disordered Proteins. *J. Biomol. NMR* **2018**, *70*, 141–165. <https://doi.org/10.1007/s10858-018-0166-5>.
46. Vuister, G.W.; Bax, A. Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J(HNH.Alpha.) Coupling Constants in <sup>15</sup>N-Enriched Proteins. *J. Am. Chem. Soc.* **1993**, *115*, 7772–7777. <https://doi.org/10.1021/ja00070a024>.



47. Shen, Y.; Roche, J.; Grishaev, A.; Bax, A. Prediction of Nearest Neighbor Effects on Backbone Torsion Angles and NMR Scalar Coupling Constants in Disordered Proteins. *Protein Sci.* **2018**, *27*, 146–158. <https://doi.org/10.1002/pro.3292>.
48. Ottiger, M.; Delaglio, F.; Bax, A. Measurement Of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. *J. Magn. Reson.* **1998**, *131*, 373–378. <https://doi.org/10.1006/jmre.1998.1361>.
49. Rückert, M.; Otting, G. Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J. Am. Chem. Soc.* **2000**, *122*, 7793–7797. <https://doi.org/10.1021/ja001068h>.
50. Hansen, M.R.; Mueller, L.; Pardi, A. Tunable Alignment of Macromolecules by Filamentous Phage Yields Dipolar Coupling Interactions. *Nat. Struct. Biol.* **1998**, *5*, 1065–1074. <https://doi.org/10.1038/4176>.
51. Teilum, K.; Kragelund, B.B.; Poulsen, F.M. Transient Structure Formation in Unfolded Acyl-Coenzyme A-Binding Protein Observed by Site-Directed Spin Labelling. *J. Mol. Biol.* **2002**, *324*, 349–357. [https://doi.org/10.1016/s0022-2836\(02\)01039-2](https://doi.org/10.1016/s0022-2836(02)01039-2).
52. Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J.R.; Jensen, M.R.; Segard, S.; Bernado, P.; Charavay, C.; Blackledge, M. *Flexible-meccano*: A Tool for the Generation of Explicit Ensemble Descriptions of Intrinsically Disordered Proteins and Their Associated Experimental Observables. *Bioinformatics* **2012**, *28*, 1463–1470. <https://doi.org/10.1093/bioinformatics/bts172>.
53. Senicourt, L.; Le Maire, A.; Allemand, F.; Carvalho, J.E.; Guee, L.; Germain, P.; Schubert, M.; Bernadó, P.; Bourguet, W.; Sibille, N. Structural Insights into the Interaction of the Intrinsically Disordered Co-Activator TIF2 with Retinoic Acid Receptor Heterodimer (RXR/RAR). *J. Mol. Biol.* **2021**, *433*, 166899. <https://doi.org/10.1016/j.jmb.2021.166899>.
54. Tovo-Rodrigues, L.; Roux, A.; Hutz, M.H.; Rohde, L.A.; Woods, A.S. Functional Characterization of G-Protein-Coupled Receptors: A Bioinformatics Approach. *Neuroscience* **2014**, *277*, 764–779. <https://doi.org/10.1016/j.neuroscience.2014.06.049>.
55. Dunker, A.K.; Lawson, J.D.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Higgs, K.W.; et al. Intrinsically Disordered Protein. *J. Mol. Graph. Model.* **2001**, *19*, 26–59. [https://doi.org/10.1016/s1093-3263\(00\)00138-8](https://doi.org/10.1016/s1093-3263(00)00138-8).
56. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why Are “Natively Unfolded” Proteins Unstructured under Physiologic Conditions? *Proteins* **2000**, *41*, 415–427. [https://doi.org/10.1002/1097-0134\(20001115\)41:3<415::aid-prot130>3.0.co;2-7](https://doi.org/10.1002/1097-0134(20001115)41:3<415::aid-prot130>3.0.co;2-7).
57. Uversky, V.N. Size-Exclusion Chromatography in Structural Analysis of Intrinsically Disordered Proteins. In *Intrinsically Disordered Protein Analysis: Volume 2, Methods and Experimental Tools*; Uversky, V.N., Dunker, A.K., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2012; pp. 179–194, ISBN 978-1-4614-3704-8.
58. Woody, R.W. Circular Dichroism of Intrinsically Disordered Proteins. In *Instrumental Analysis of Intrinsically Disordered Proteins*; John Wiley & Sons, Ltd.: New York, NY, USA, 2010; pp. 303–321, ISBN 978-0-470-60261-4.
59. Woody, R.W. Theory of Circular Dichroism of Proteins. In *Circular Dichroism and the Conformational Analysis of Biomolecules*; Fasman, G.D., Ed.; Springer: Boston, MA, USA, 1996; pp. 25–67, ISBN 978-1-4757-2508-7.
60. Bernadó, P.; Svergun, D.I. Structural Analysis of Intrinsically Disordered Proteins by Small-Angle X-Ray Scattering. *Mol. Biosyst.* **2011**, *8*, 151–167. <https://doi.org/10.1039/C1MB05275F>.
61. Cordeiro, T.N.; Herranz-Trillo, F.; Urbanek, A.; Estaña, A.; Cortés, J.; Sibille, N.; Bernadó, P. Small-Angle Scattering Studies of Intrinsically Disordered Proteins and Their Complexes. *Curr. Opin. Struct. Biol.* **2017**, *42*, 15–23. <https://doi.org/10.1016/j.sbi.2016.10.011>.
62. Shimada, I.; Ueda, T.; Kofuku, Y.; Eddy, M.T.; Wüthrich, K. GPCR Drug Discovery: Integrating Solution NMR Data with Crystal and Cryo-EM Structures. *Nat. Rev. Drug Discov.* **2019**, *18*, 59–82. <https://doi.org/10.1038/nrd.2018.180>.
63. Wishart, D.S.; Sykes, B.D. The <sup>13</sup>C Chemical-Shift Index: A Simple Method for the Identification of Protein Secondary Structure Using <sup>13</sup>C Chemical-Shift Data. *J. Biomol. NMR* **1994**, *4*, 171–180. <https://doi.org/10.1007/BF00175245>.
64. Bax, A.; Kontaxis, G.; Tjandra, N. Dipolar Couplings in Macromolecular Structure Determination. In *Methods in Enzymology*; Elsevier: New York, NY, USA, 2001; Volume, 339, pp. 127–174, ISBN 978-0-12-182240-8.
65. Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R.W.H.; Blackledge, M. A Structural Model for Unfolded Proteins from Residual Dipolar Couplings and Small-Angle X-ray Scattering. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17002–17007. <https://doi.org/10.1073/pnas.0506202102>.
66. Bernadó, P.; Bertoncini, C.W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. Defining Long-Range Order and Local Disorder in Native  $\alpha$ -Synuclein Using Residual Dipolar Couplings. *J. Am. Chem. Soc.* **2005**, *127*, 17968–17969. <https://doi.org/10.1021/ja055538p>.
67. Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M.R.; Zweckstetter, M.; Blackledge, M. NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010**, *132*, 8407–8418. <https://doi.org/10.1021/ja101645g>.
68. Guillien, M.; le Maire, A.; Mouhand, A.; Bernadó, P.; Bourguet, W.; Banères, J.-L.; Sibille, N. IDPs and Their Complexes in GPCR and Nuclear Receptor Signaling. In *Progress in Molecular Biology and Translational Science*; Elsevier: New York, NY, USA, 2020; Volume 174, pp. 105–155, ISBN 978-0-12-822615-5.
69. Appadurai, R.; Uversky, V.N.; Srivastava, A. The Structural and Functional Diversity of Intrinsically Disordered Regions in Transmembrane Proteins. *J. Membr. Biol.* **2019**, *252*, 273–292. <https://doi.org/10.1007/s00232-019-00069-2>.
70. Ahmad, R.; Lahuna, O.; Sidibe, A.; Daulat, A.; Zhang, Q.; Luka, M.; Guillaume, J.-L.; Gallet, S.; Guillonnet, F.; Hamroune, J.; et al. GPR50-Ctail Cleavage and Nuclear Translocation: A New Signal Transduction Mode for G Protein-Coupled Receptors. *Cell. Mol. Life Sci.* **2020**, *77*, 5189–5205. <https://doi.org/10.1007/s00018-019-03440-7>.

71. Dufourny, L.; Levasseur, A.; Migaud, M.; Callebaut, I.; Pontarotti, P.; Malpaux, B.; Monget, P. GPR50 Is the Mammalian Ortholog of Mel1c: Evidence of Rapid Evolution in Mammals. *BMC Evol. Biol.* **2008**, *8*, 105. <https://doi.org/10.1186/1471-2148-8-105>.
72. Bouzo-Lorenzo, M.; Santo-Zas, I.; Lodeiro, M.; Nogueiras, R.; Casanueva, F.F.; Castro, M.; Pazos, Y.; Tobin, A.B.; Butcher, A.J.; Camiña, J.P. Distinct Phosphorylation Sites on the Ghrelin Receptor, GHSR1a, Establish a Code That Determines the Functions of  $\beta$ -Arrestins. *Sci. Rep.* **2016**, *6*, 22495. <https://doi.org/10.1038/srep22495>.
73. Nobles, K.N.; Guan, Z.; Xiao, K.; Oas, T.G.; Lefkowitz, R.J. The Active Conformation of Beta-Arrestin1: Direct Evidence for the Phosphate Sensor in the N-Domain and Conformational Differences in the Active States of Beta-Arrestins1 and -2. *J. Biol. Chem.* **2007**, *282*, 21370–21381. <https://doi.org/10.1074/jbc.M611483200>.
74. Lee, Y.; Warne, T.; Nehmé, R.; Pandey, S.; Dwivedi-Agnihotri, H.; Chaturvedi, M.; Edwards, P.C.; García-Nafria, J.; Leslie, A.G.W.; Shukla, A.K.; et al. Molecular Basis of  $\beta$ -Arrestin Coupling to Formoterol-Bound  $\beta$  1 -Adrenoceptor. *Nature* **2020**, *583*, 862–866. <https://doi.org/10.1038/s41586-020-2419-1>.
75. Tompa, P.; Schad, E.; Tantos, A.; Kalmar, L. Intrinsically Disordered Proteins: Emerging Interaction Specialists. *Curr. Opin. Struct. Biol.* **2015**, *35*, 49–59. <https://doi.org/10.1016/j.sbi.2015.08.009>.
76. 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. Distinct G Protein-Coupled Receptor Phosphorylation Motifs Modulate Arrestin Affinity and Activation and Global Conformation. *Nat. Commun.* **2019**, *10*, 1261. <https://doi.org/10.1038/s41467-019-09204-y>.
77. Bah, A.; Vernon, R.M.; Siddiqui, Z.; Krzeminski, M.; Muhandiram, R.; Zhao, C.; Sonenberg, N.; Kay, L.E.; Forman-Kay, J.D. Folding of an Intrinsically Disordered Protein by Phosphorylation as a Regulatory Switch. *Nature* **2015**, *519*, 106–109. <https://doi.org/10.1038/nature13999>.

# Structural insights into the intrinsically disordered GPCR C-terminal region, major actor in arrestin-GPCR interaction

Myriam Guillien <sup>1</sup>, Assia Mouhand <sup>1</sup>, Aurélie Fournet <sup>1</sup>, Amandine Gontier <sup>1</sup>, Aleix Martí Navia <sup>1</sup>, Tiago N. Cordeiro <sup>1,2</sup>, Frédéric Allemand <sup>1</sup>, Aurélien Thureau <sup>3</sup>, Jean-Louis Banères <sup>3</sup>, Pau Bernadó <sup>1</sup>, and Nathalie Sibille <sup>1,\*</sup>

<sup>1</sup> Centre de Biologie Structurale (CBS), Univ Montpellier, CNRS, INSERM, Montpellier, France

<sup>2</sup> Instituto de Tecnologia Química e Biológica António Xavier (ITQB), Universidade NOVA de Lisboa, Av. da República, 2780-157 Oeiras, Portugal.

<sup>3</sup> Beamline SWING, Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91190 Gif-sur-Yvette, France

<sup>4</sup> IBMM, UMR5247 CNRS, Pôle Chimie Balard Recherche, 1919 route de Mende, 34293 Montpellier cedex 5, France

\* Correspondence: [nathalie.sibille@cbs.cnrs.fr](mailto:nathalie.sibille@cbs.cnrs.fr); Tel.: +33 4 67 41 77 05, Fax: +33 4 67 41 79 13

Protein	3D	$^3J_{\text{HNHA}}$	relaxation	RDC alcohol	RDC Pf1	PRE
V2R	170	170	373	310	139	NT
GHSR	312	314	124	390	285	NT
$\beta$ 2AR	255	487	487	487	220	50

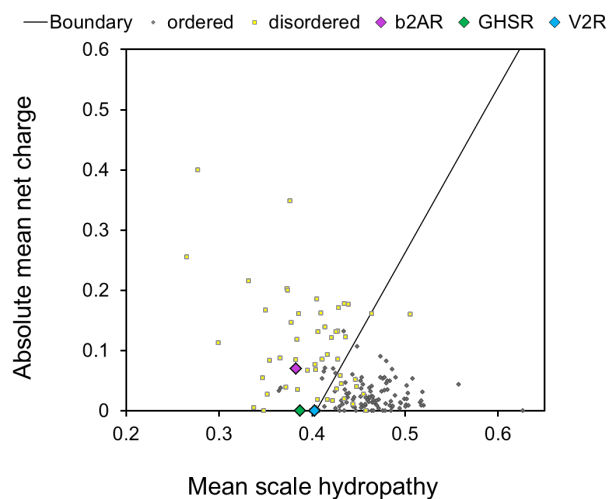
**Table S1: Concentration used for the NMR characterization of GPCR C-termini of V2R-Cter (blue), GHSR-Cter (green) and  $\beta$ 2AR-Cter (purple).** Concentration used in each experiment is indicated in  $\mu\text{M}$ . The letters NT mean that these experiments have not been carried out.

Disordered predictor	Attributes		Meta-predictor	Webserver
	Sequence alignment	Structural information		
Pr Dos	Psi-Blast	No	Yes	<a href="http://prdos.hgc.jp/cgi-bin/top.cgi">http://prdos.hgc.jp/cgi-bin/top.cgi</a>
PONDR FIT	No	No	Yes	<a href="http://original.disprot.org/metapredictor.php">http://original.disprot.org/metapredictor.php</a>
DISOPRED3	Psi-Blast	Psi-Pred	Yes	<a href="http://bioinf.cs.ucl.ac.uk/psipred/">http://bioinf.cs.ucl.ac.uk/psipred/</a>
SPOT-disorder 2	Psi-Blast + Hhblits	SPOT-1D	No	<a href="https://sparks-lab.org/server/spot-disorder2/">https://sparks-lab.org/server/spot-disorder2/</a>
SPOT-disorder single	No	No	No	<a href="https://sparks-lab.org/server/spot-disorder-single/">https://sparks-lab.org/server/spot-disorder-single/</a>
Espritz NMR	No	No	No	<a href="http://old.protein.bio.unipd.it/espritz/">http://old.protein.bio.unipd.it/espritz/</a>
DISpro	UniqueProt	Sspro + ACCpro	No	<a href="http://scratch.proteomics.ics.uci.edu/">http://scratch.proteomics.ics.uci.edu/</a>

**Table S2: Related to Figure 3. Comparison of the disorder predictors used in this study.** For each predictor (first column), disordered is predicted based on amino-acid sequence and can be completed with sequence alignment and/or secondary structure information computed with tools listed in the second and third column. Some of these servers use multiple predictors (meta-predictor) as indicated in the fourth column. Webservers are listed in the last column of the table.

Predictor	Sequence alignment	Predicted secondary structure	Webserver
SOPMA	BLAST + Clustal	Helix, Strand, Turn, Coil	<a href="https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html">https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html</a>
PSIPRED	PSI-BLAST	Helix, Strand, Coil	<a href="http://bioinf.cs.ucl.ac.uk/psipred/">http://bioinf.cs.ucl.ac.uk/psipred/</a>
JPRED4	PSI-BLAST + HMMER	Helix, Strand, Coil	<a href="https://www.compbio.dundee.ac.uk/jpred/">https://www.compbio.dundee.ac.uk/jpred/</a>
PSSpred	PSI-BLAST	Helix, Strand, Coil	<a href="https://zhanggroup.org/PSSpred/">https://zhanggroup.org/PSSpred/</a>
SPOT1D	PSI-BLAST + HHblits	Helix, Strand, Coil	<a href="https://sparks-lab.org/server/spot-1d/">https://sparks-lab.org/server/spot-1d/</a>
SPIDER3	PSI-BLAST + HHblits	Helix, Strand, Coil	<a href="https://sparks-lab.org/server/spider3/">https://sparks-lab.org/server/spider3/</a>

**Table S3: Related to Figure 3. Comparison of the secondary structure predictors used in this study.** Predictors (first column) use sequence alignments computed with tools listed in the second column, predicted secondary structures are listed in the third column. Webservers are reported in the last column of the table.

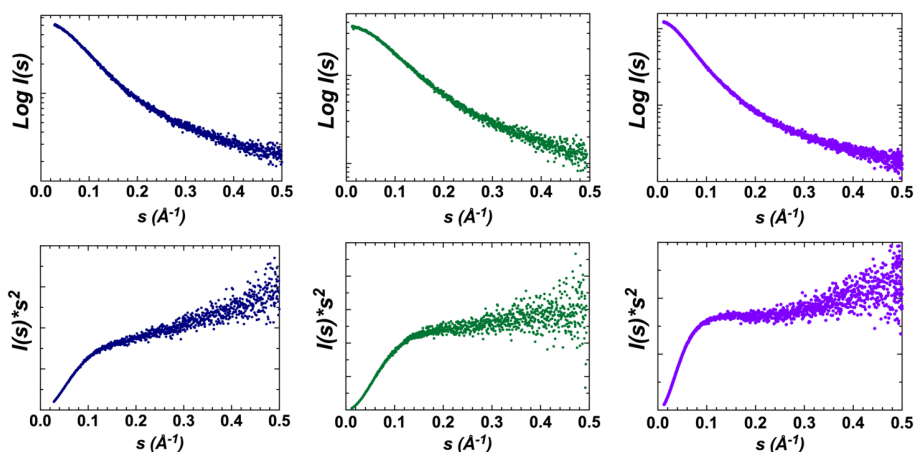


**Figure S1: Charge-Hydropathy plot [1].** 54 disordered proteins are shown in yellow circles while 105 ordered proteins are shown in grey circles. The boundary between order/disorder (dark line) was calculated with the equation:  $\overline{Hydropathy} = (\overline{net\ charge} + 1,151)/2,785$ . V2R-Cter (blue), GHSR-Cter (green) and  $\beta$ 2AR-Cter (pink) are represented with diamonds. This graph was generated using PONDR ([www.pondr.com](http://www.pondr.com)).

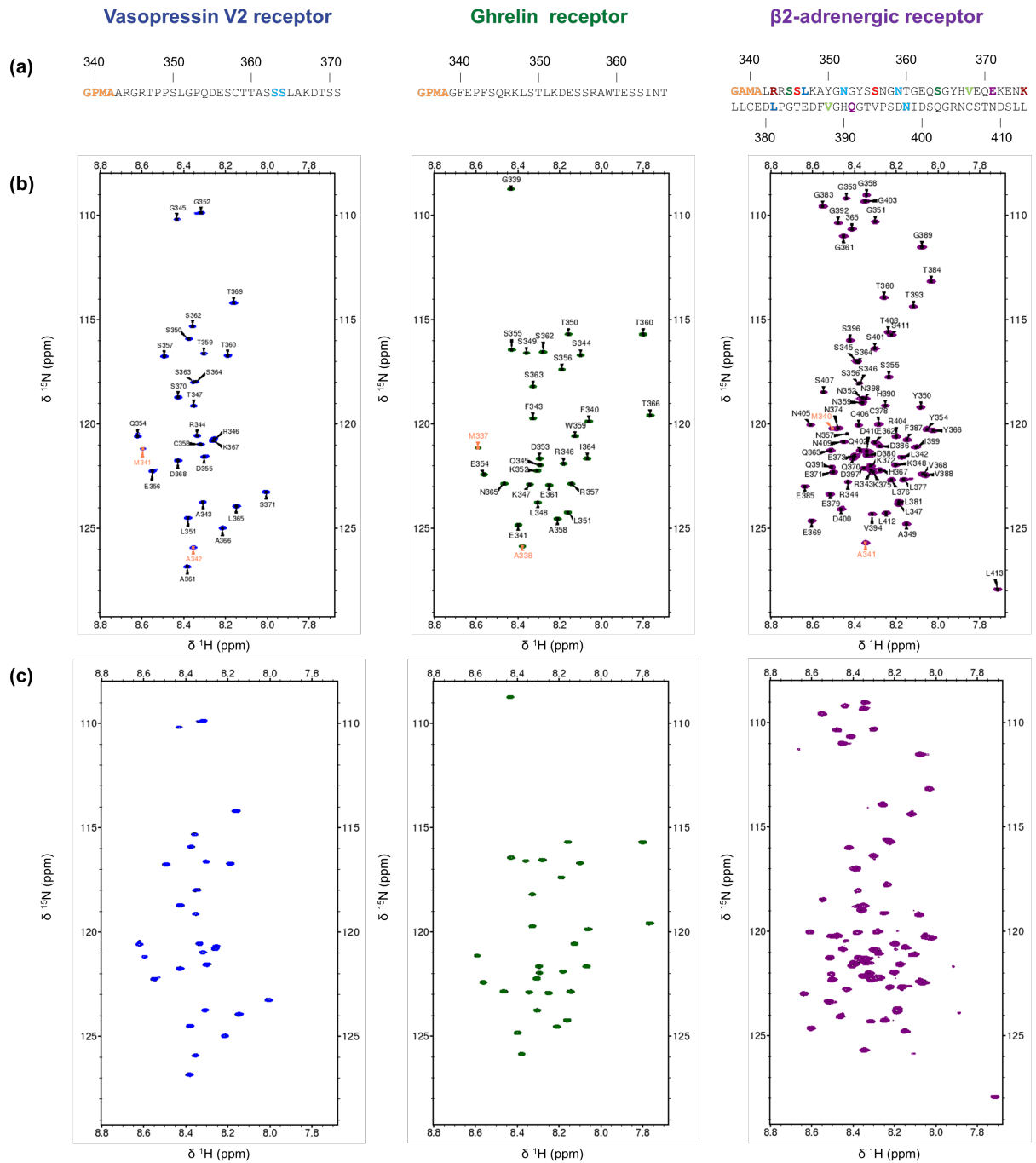
Vasopressin V2 receptor

Ghrelin receptor

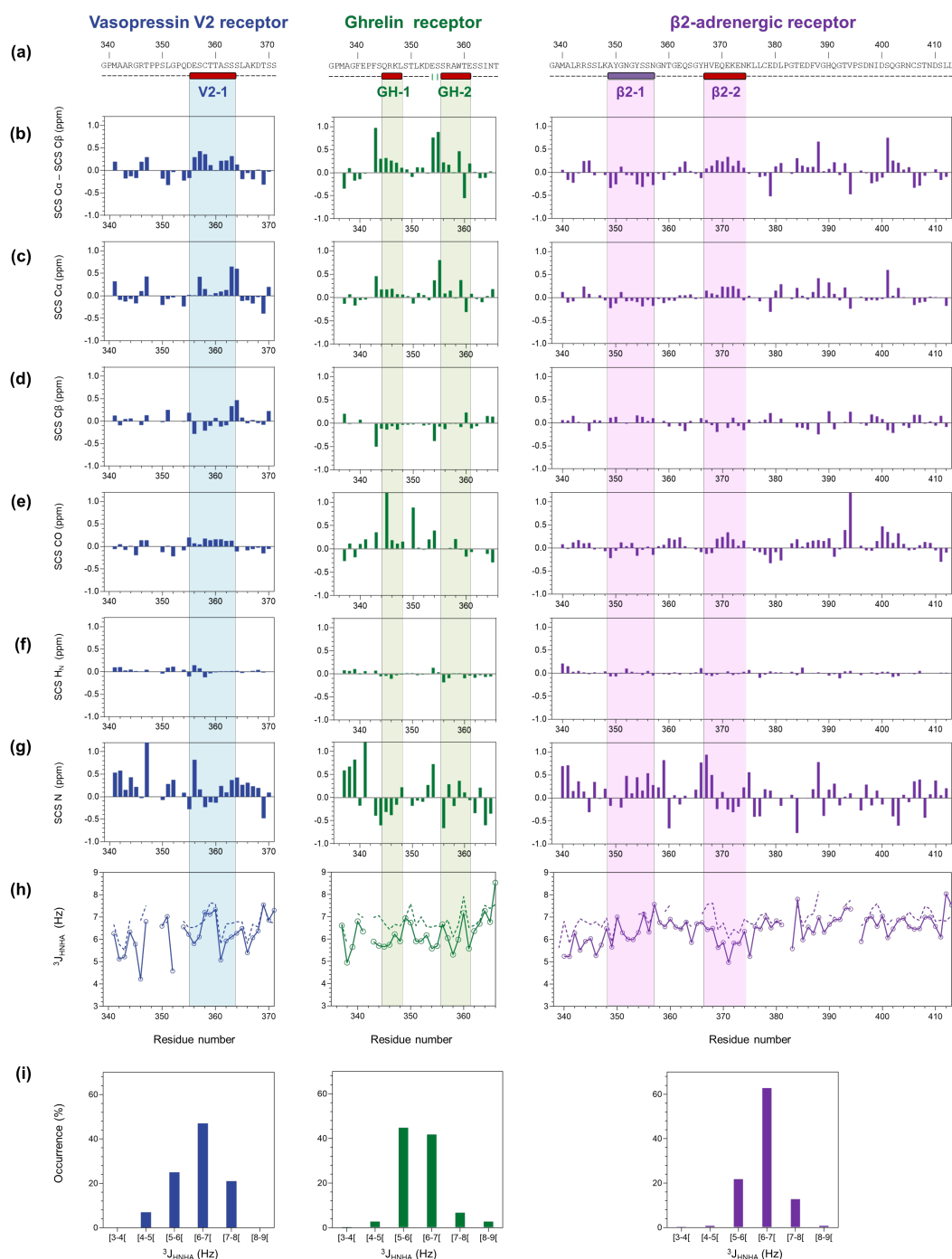
$\beta$ 2-adrenergic receptor



**Figure S2: Related to Figure 2. SAXS data of V2R-Cter (blue, left), GHSR-Cter (green, middle) and  $\beta$ 2AR-Cter (purple, right). Semi-logarithmic representation of the experimental SAXS curves versus scattering angle (upper panel) and Kratky plots (bottom panel). The Kratky plots of GPCR-Cters display a typical profile of IDP with no clear maximum and a monotone increase along the momentum transfer range [2–4].**

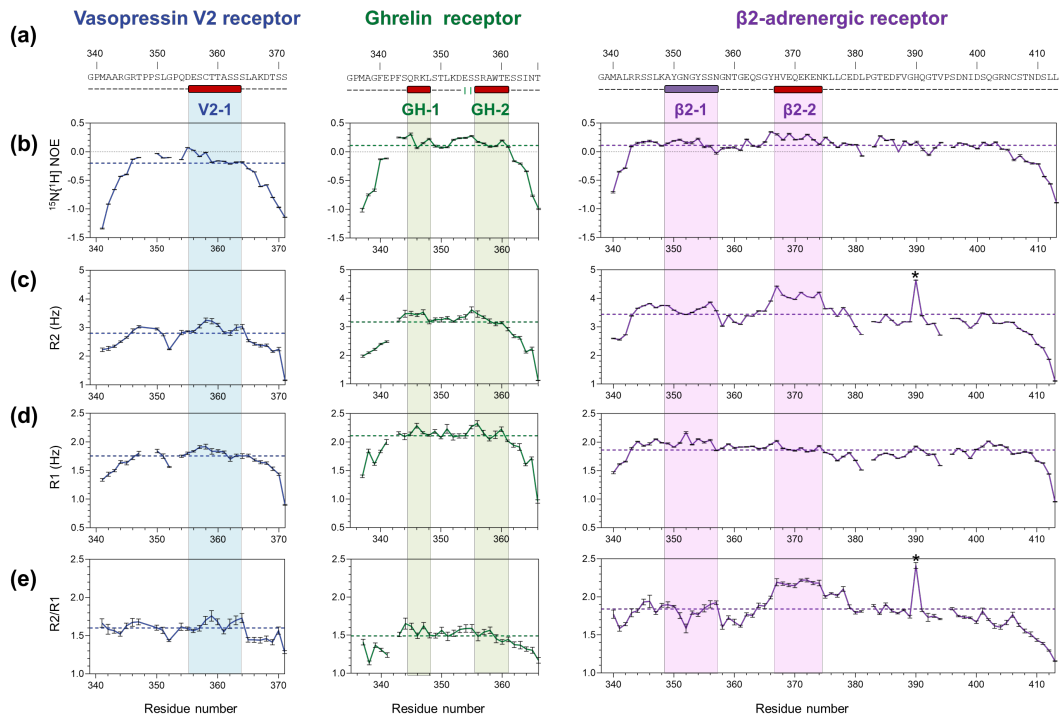


**Figure S3: Related to Figure 3. Assignments of V2R-Cter (blue, left), GHSR-Cter (green, middle) and β2AR-Cter (purple, right).** (a) Sequence of the GPCRs C-termini. Assigned residues are indicated in black while assignments of overlapping cross-peak are noticed in color. Amino acids coming from the protease cleavage site are shown in orange and set of overlapping residues in <sup>15</sup>N-HSQC are in the same color. (b) <sup>15</sup>N-HSQC were recorded on 300 μM samples at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer. Residue-specific assignments are shown on spectra. Amino acids coming from the protease cleavage site are shown in orange. (c) The concentration of GPCR-Cter does not affect their conformations. Overlay of <sup>15</sup>N-HSQC of V2R-Cter (blue, left), GHSR-Cter (green, middle) and β2AR-Cter (purple, right) recorded on samples at 300 μM (black) or 100 μM (colored) at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl bufer. No chemical shift changes were observed in NMR spectra at different concentrations and thus, the spectrum at 300 μM (black) is hidden by the spectrum at 100 μM (colored).

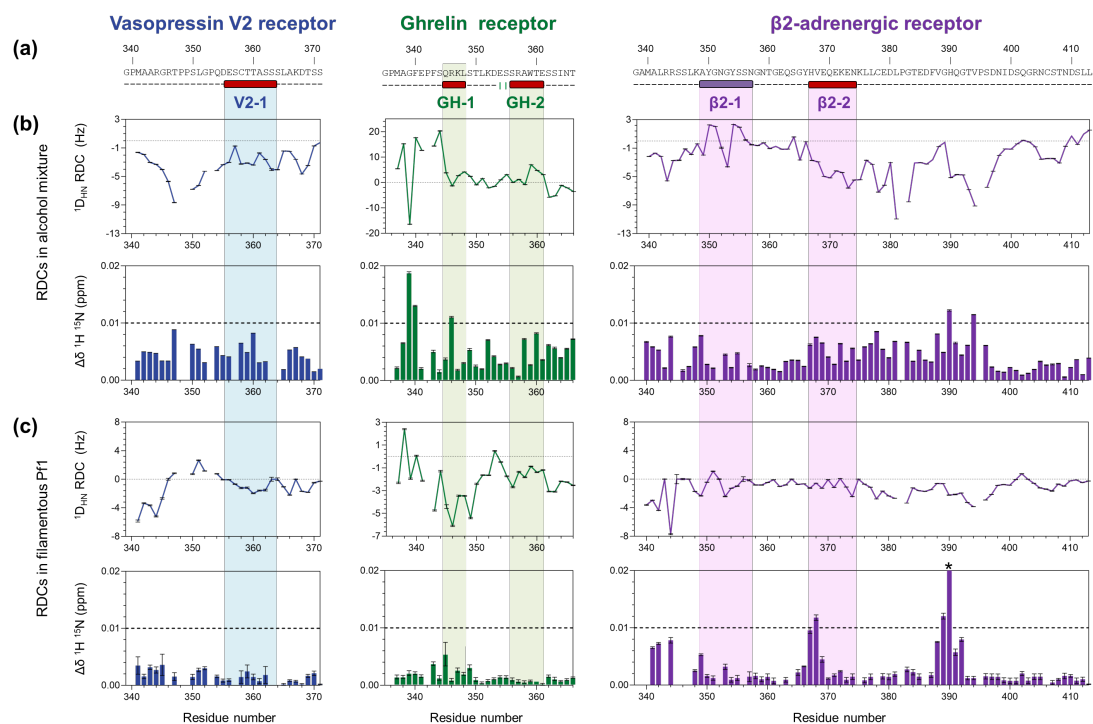


**Figure S4: Related to Figure 3. Secondary chemical shift (SCS) and  $^3J_{\text{HNHA}}$  scalar coupling measured for for V2R-Cter (blue, left), GHSR-Cter (green, middle) and  $\beta$ 2AR-Cter (purple, right). (a) Sequence and NMR secondary structure consensus are indicated according to Figure 3. (b-g) Secondary chemical shift (SCS) were calculated using random coil chemical shifts database from POTENCI. (h)  $^3J_{\text{HNHA}}$  scalar couplings (colored line) along the sequence are compared to predicted couplings (RC\_3JHNHa server) for a fully disorder chain (dash colored line). (i) Occurrences of  $^3J_{\text{HNHA}}$  for each Cters. The couplings are consistent with an overall unfolded protein (values from 6 to 8 Hz) with a low content of helical and extended conformation (values below 6 Hz and above 8 Hz respectively). SCS were extracted from 3D experiments recorded at 700 MHz (800 MHz for  $\beta$ 2AR-Cter), 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer and  $^3J_{\text{HNHA}}$  were recorded at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.**

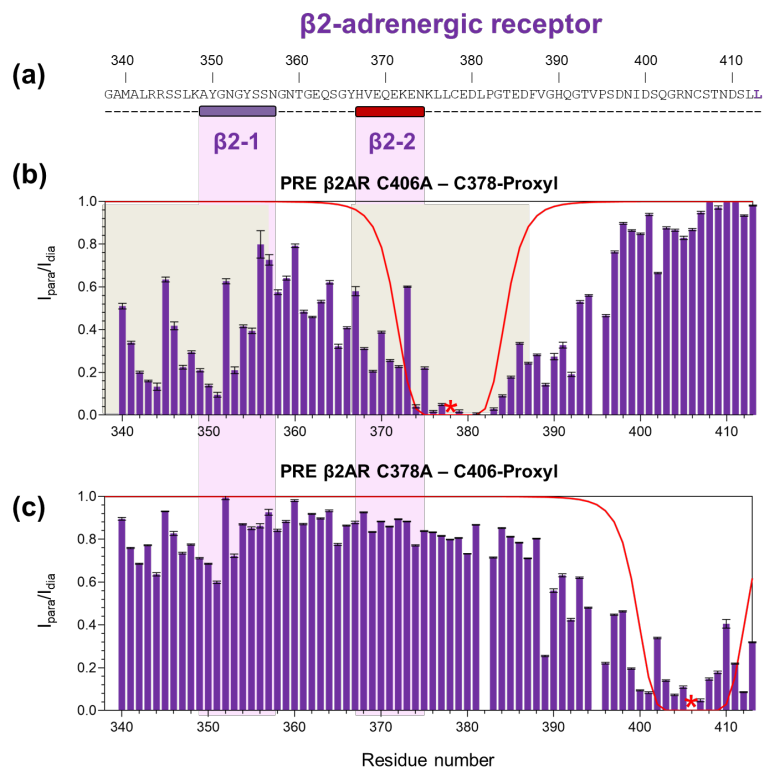




**Figure S5: Related to Figure 3e. Relaxation data of V2R-Cter (blue, left), GHSR-Cter (green, middle) and  $\beta$ 2AR-Cter (purple, right). (a) Sequence and NMR secondary structure consensus are indicated according to Figure 3. (b)  $^{15}\text{N}\{^1\text{H}\}$  heteronuclear NOE, (c) transversal relaxation rates R2, (d) longitudinal relaxation rates R1 and (e) R2/R1 ratio are shown in colored lines. Average values are indicated by dashed colored line. Stars indicate histidine 390 outlier due its extreme pH sensitivity. Relaxation data were recorded at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.**



**Figure S6: Related to Figure 3f.  $^1D_{NH}$  RDCs values of V2R-Cter (blue, left), GHSR-Cter (green, middle) and  $\beta$ 2AR-Cter (purple, right). (a) Sequence and NMR secondary structure consensus are indicated according to Figure 3. RDCs data measured in (b) alcohol mixture or in (c) filamentous phage Pf1. In each media,  $^1D_{NH}$  RDCs profile (top), chemical shift perturbations (bottom) are shown. Stars indicate histidine 390 outlier due to its extreme pH sensitivity. RDCs were measured at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.**



**Figure S7: Related to Figure 3. PRE data of  $\beta$ 2AR-Cter variants.** (a) Sequence and NMR secondary structure consensus are indicated according to Figure 3. (b) PRE data of C406A mutant labeled on C378 (red star). The C378 labeled with the paramagnetic probe, which just lies after  $\beta$ 2-2 region, induced a reduction of the intensity ratio in  $\beta$ 2-1 region, suggesting long-range interaction. PRE affected regions are highlighted in grey. (c) PRE data of C378A mutant labeled on C406 (red star). The C406 labeled with the paramagnetic probe, does not lead to reduction of intensity ratio. This region does not form long-range interaction with other regions. PREs were calculated from the ratio of intensity ( $I_{para}/I_{dia}$ ) of  $^{15}\text{N}$ -HSQC spectra measured for the paramagnetic ( $I_{para}$ ) sample and for the diamagnetic sample ( $I_{dia}$ ). PRE were recorded at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer on samples at 50  $\mu\text{M}$ . Theoretical PRE curves for a completely disordered protein are shown in red.

## REFERENCES

1. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why Are “Natively Unfolded” Proteins Unstructured under Physiologic Conditions? *Proteins* **2000**, *41*, 415–427, doi:10.1002/1097-0134(20001115)41:3<415::aid-prot130>3.0.co;2-7.
2. Sibille, N.; Bernadó, P. Structural Characterization of Intrinsically Disordered Proteins by the Combined Use of NMR and SAXS. *Biochem Soc Trans* **2012**, *40*, 955–962, doi:10.1042/BST20120149.
3. Cordeiro, T.N.; Herranz-Trillo, F.; Urbanek, A.; Estaña, A.; Cortés, J.; Sibille, N.; Bernadó, P. Small-Angle Scattering Studies of Intrinsically Disordered Proteins and Their Complexes. *Curr Opin Struct Biol* **2017**, *42*, 15–23, doi:10.1016/j.sbi.2016.10.011.
4. Cordeiro, T.N.; Herranz-Trillo, F.; Urbanek, A.; Estaña, A.; Cortés, J.; Sibille, N.; Bernadó, P. Structural Characterization of Highly Flexible Proteins by Small-Angle Scattering. *Adv Exp Med Biol* **2017**, *1009*, 107–129, doi:10.1007/978-981-10-6038-0\_7.

