Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Orchestration of signaling by structural disorder in class 1 cytokine receptors

Pernille Seiffert
Kobenhavns Universitet
Katrine Bugge
Kobenhavns Universitet
Mads Nygaard
Kobenhavns Universitet
Gitte W. Haxholm
Kobenhavns Universitet
Jacob H. Martinsen
Kobenhavns Universitet
Martin N. Pedersen
Kobenhavns Universitet
Lise Arleth
Kobenhavns Universitet
Wouter Boomsma
Kobenhavns Universitet
Birthe B. Kragelund (bbk@bio.ku.dk)
Kobenhavns Universitet https://orcid.org/0000-0002-7454-1761

Research

Keywords: IDRs, IDPs, signaling, NMR, SAXS, SLiM, disorder, structural biology, CIDER, IDDomainSpotter, cytokine receptors

Posted Date: May 5th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-23462/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

---

Version of Record: A version of this preprint was published on August 24th, 2020. See the published version at https://doi.org/10.1186/s12964-020-00626-6.

Abstract

Background

Class 1 cytokine receptors (C1CRs) are single-pass transmembrane proteins responsible for transmitting signals between the outside and the inside of cells. Remarkably, they orchestrate key biological processes such as proliferation, differentiation, immunity and growth through long disordered intracellular domains (ICDs), but without having intrinsic kinase activity. Despite these key roles, their characteristics remain rudimentarily understood.

Methods

The current paper asks the question of why disorder has evolved to govern signaling by the C1CRs by reviewing the literature in combination with new sequence and biophysical analyses of chain properties across the family.

Results

We uncover that the C1CR-ICDs are fully disordered and brimming with SLiMs. Many of these SLiMs are overlapping, jointly signifying a complex regulation of interactions, including network rewiring by isoforms. The C1CRs-ICDs have unique properties that distinguish them from most IDPs and we forward the perception that the C1CR-ICDs are far from simple strings with constitutively bound kinases. Rather, they carry both organizational and operational features left uncovered within their disorder, including mechanisms and complexities of regulatory functions.

Conclusions

Critically, the understanding of the fascinating ability of these long, completely disordered chains to orchestrate complex cellular signaling pathways is still in its infancy, and we urge a perceptional shift away from a simplistic view towards uncovering their full functionalities.

Background

The sequencing of the human genome and key observations from earlier research [1,2], spurred the recognition of proteins and protein regions functioning without three-dimensional folds. These intrinsically disordered proteins (IDPs) and regions (IDRs, collectively here referred to as IDPs), constitute around 30–40% of the human proteome [3] and perform key cellular and highly regulated processes such as transcription, translation and signaling [4–7]. IDPs show distinct sequence characteristics with higher frequencies of Pro, Glu, Ser, Gln, Lys, Ala, and Gly, and lower frequencies of Val, Leu, Ile, Trp and Cys. Hence, their sequences have particular properties of low hydrophobicity and high charge [8], resulting in

their intrinsic inability to fold into a single, well-defined structure. Instead they take on ensembles of almost isoenergetic states, although these are far from random. For example, they may harbor lowly populated secondary structures that are relevant to binding [9] and to tuning the disordered ensemble [10], and alteration in the populations of these elements by e.g. mutations may lead to promotion of pathological states [11] Interactions by IDPs are often mediated by short linear motifs (SLiMs), which are short sequence stretches, 2–12 residues long and usually recognized by patterns of conserved residues within an otherwise sparsely conserved sequence stretch [12]. SLiMs can overlap, and due to the structural plasticity of disordered regions, transiently exposed motifs and binding interfaces thus provide a trait of multispecificity to an otherwise simple binding site. Furthermore, flanking residues outside a SLiM and post translational modifications (PTMs) can tune affinity and add regulatory properties assisting in the spatiotemporal orchestration of multiple binding events [13]. Indeed, PTMs are frequent in IDPs, in particular phosphorylations [14] and ubiquitylation [15], impacting functionalities and regulatory potential in several different ways. Thus, PTMs allow IDPs to function in rheostatic regulation [16] which are graded quantitative responses, and may also drive the formation or disruption of membrane-less organelles through processes known as liquid-liquid phase separation (LLPS) [17,18]. Finally, the disordered nature of IDPs allows them to exploit diverse binding mechanisms by which they can fold-upon-binding [19–21], but also form complexes in which structural disorder is maintained to different degrees [22]. Here, the most extreme case being the formation of a completely disordered complex of functional relevance and extreme affinity [23].

The class 1 cytokine receptors

Structural disorder also exists in membrane proteins [24]. Besides being located in longer, disordered loops connecting transmembrane helices, such as in the sodium proton exchangers [25] and β*1*-adrenergic receptor [26], disorder prevails preferentially on the intracellular side [27], with membrane proteins having disordered N- and C-terminal tails of various length [28]. In fact, ~ 10% of the human membranome have disordered intracellular domains of > 100 residues classifying them as long disordered regions [28] and ~ 40% have disordered domains of > 30 resides [27]. An important family of membrane proteins with long disordered tails is the class 1 cytokine receptors (C1CRs) [29–31]. This family constitutes 40 members, which have been divided into five different groups based on their structural properties [30]. They all share having a tripartite structure characterized by a folded, extracellular domain (ECD) of various sizes and complexities, a single transmembrane helix (TMD), and an intracellular domain (ICD), also of various length. Recently, the proportions of the different domains were exemplified by the a three-dimensional structural model of the prolactin receptor (PRLR); the first full-length structure of any cytokine receptor [32], **Fig. 1a**. C1CRs are characterized by the presence of two conserved cysteine bridges in the membrane distal fibronectin type III domain (D1) of the ECD, a WSxWS motif in the membrane proximal domain (D2) of the ECD, and two conserved sequence motifs in their ICD, Box1 and Box2 [29,33]. Of the 40 receptors, 29 have an ICD (**Fig. 1c**), which all lack globular domains and intrinsic kinase activity. Signaling therefore critically depends on associated kinases such as Janus kinase 1–3 (JAK1, JAK2, JAK3) and tyrosine kinase 2 (TYK2) [34], Src kinases and mitogen activated

protein kinases (MAPKs) [35]. Box1 is a proline-rich SLiM, onto which JAK are proposed to be constitutively bound [36–38]. Box2 is a less conserved region consisting of a sequence of hydrophobic residues, followed by negatively and then positively charged residues [39,40]. The function of Box2 is unclear, but it is suggested to be important for efficient binding and activation of JAK1/2/3 and TYK2, possibly in cooperation with the region between Box1 and Box2 [36]. Recently, structures of complexes between JAK1/2 and TYK2 and a fragment of a cytokine receptor ICD were solved, namely of the erythropoietin receptor (EPOR) [41] (**Fig. 1b**), and of two class II cytokine receptors ICDs from the interferon-λ receptor 1 (IFNLR1) [42] and interferon-α receptor chain 1 (IFNAR-1) [43], respectively. These complexes have revealed a common mode of interaction, where Box1 makes hydrophobic contacts to the FERM-domain of JAK1/2;TYK2 and a Glu of Box2 intercalates into the phospho-tyrosine binding pocket of the non-canonical SH2 domain of JAK1/2;TYK2, **Fig. 1b**. Src kinases are also suggested to bind to this region [44–47] most likely via a SH3-SH2 interaction.

C1CRs act in homo- and heteromeric dimers and oligomers, in some of which they share common receptor chains (IL-6Rβ (gp130), βc (IL-3Rβ) and γc (IL-2Rγ)). By binding the cytokine on the extracellular side, multiple common and receptor specific signaling events are initiated intracellularly with the main common pathways being activation of JAK/STAT, Src and MAPK signaling (for reviews on signaling, see e.g. [34,35]). The receptors of the homodimeric group 1 (sometimes referred to as Type 1), constituting the EPOR, the growth hormone receptor (GHR), the thrombopoietin receptor (TPOR) and the PRLR, are considered to be structurally the simplest. This group has served as paradigmatic models of the family, and studies on the GHR have suggested that hormone binding leads to conformational changes in the receptor, including realignment and separation of the lower transmembrane domains, ultimately resulting in *trans*-phosphorylation of JAK2 and phosphorylation of the ICDs to initiate signaling [48,49].

Isoforms of C1CRs

Alternative splicing has been suggested to fine-tune the functions of IDPs [50,51] and in accordance with this, functionally relevant alternative splicing has been shown to occur primarily in IDPs [50–53]. Furthermore, a large-scale computational analysis of alternative exons in human genes showed that alternatively spliced IDPs were highly enriched in interaction sites and regions modified by PTMs [50]. This has been suggested to provide a mechanism to rewire interaction networks by including or eliminating specific interaction motifs or binding sites [51]. Alternative splicing is further suggested to be important for maintaining tissue identity, as tissue-specific exons were shown to contain disordered regions enriched in PTMs and binding motifs that were central parts of tissue-specific interaction networks [50]. Many of the C1CRs also exist in different isoforms, but there are currently no reports available on their properties or numbers, or how their characteristics differ. For the PRLR, there are to date nine isoforms identified in humans, with four differing in their ECD [54–57], out of which one is an ECD soluble isoform [58]. The remaining five isoforms result from alternative mRNA splicing, and differ solely in length and amino acid composition of the ICD. These ICD isoforms constitute the predominant long form (LF, 622 residues), the intermediate form (IF, 349 residues), the short form 1-a (SF1a, 376 residues), the short form 1-b (SF1b, 288 residues) and the short form 1c (SF-1c, 309 residues), **Fig. 1c**. Compared to

the LF, the IF, SF1a, SF1b and SF1c contain 13, 39, 3 and 24 unique C-terminal residues, respectively. The LF, IF, and SF1a all contain Box1 and Box2 motifs, while SF1b only contains Box1 [59–61]. Since activation of the JAK2/STAT5 pathway requires C-terminal tyrosine phosphorylation in the PRLR-ICD [62], this pathway is only activated by the LF due to lack of the pYxxQ docking site motif in the others [59,61,63,64]. In contrast, the MAPK pathway is activated by the LF, SF1a, and SF1b [63,65,66], while the PI3K pathway is activated by both the LF [67,68] as well as the short isoform in mice [63]. Heterodimerization of the LF with either of the shorter isoforms has been shown to inhibit activation of the JAK2/STAT5 pathway [59,60]. The shortest isoform, PRLR-SF1b, has been shown to negatively regulate PRLR signaling [62]. Hence, isoform ratios needs to be delicately balanced, and an increased ratio of short- to long forms has implications in breast- [69] and prostate [70] cancers. Lastly, since the different PRLR isoforms are expressed to different extents in different cell types and under different conditions [59–61], alternative splicing adds layers of regulation to PRLR signaling.

Structural biology of C1CRs

Many C1CR members have been subjected to structural investigations, but remarkably, > 95% of the available structural information comes from the ECDs alone or in complexes with cytokine ligands. Furthermore, most ECD structures have been solved by X-ray crystallography, not providing information on dynamical parts. The understanding of the structure and function of single-pass TMDs is generally lacking behind [71]. For the C1CRs, structures of the TMDs have been solved almost exclusively on group 1 members, i.e. PRLR [32], EPOR [72,73] and GHR [74] using detergent solubilized peptides and nuclear magnetic resonance (NMR) spectroscopy in both monomeric and dimeric states. However, monomeric structures have also been solved of the common receptor βc in bicelle membrane mimetics [75]. Still, no structural information is available for almost one fourth of the receptors, including the TPOR, interleukin (IL)-31Rβ, IL-31Rα, IL-12Rβ1, IL-12Rβ2, IL-27R, IL-9Rα, IL-11Rα, and the oncostatin-M-specific receptor (OSMR). Furthermore, only six structures are available of shorter fragments of ICDs in complex with signaling molecules, as exemplified by the juxtamembrane domain of the EPOR in complex with the FERM-SH2 domains of JAK2 [41], a 12-residue phospho-peptide from GHR-ICD in complex with SOCS2 [76], an 11-residue phospho-peptide of IL-4R-ICD with the phospho-tyrosine binding domain of the insulin receptor substrate (IRS)-1 [77], and an 8-residue peptide from IL-5Rα-ICD with the PDZ domain of syntenin [78], **Fig. 1b**. Drawn to scale on a representative model ICD (**Fig. 1b**), these binding sites take up only a minor part of the ICD, and leave a large area completely unexplored; not to mention the > 25 receptors for which no structures of any part of the ICD are available.

Using a combination of small-angle X-ray scattering (SAXS), NMR and circular dichroism (CD) spectroscopy we recently performed an extensive characterization of the ICDs of GHR and PRLR, showing that the entire ICDs of the LFs of the human PRLR (PRLR-LF-ICD) and the human GHR (GHR-LF-ICD) are disordered [32,79]. Moreover, the PRLR-LF-ICD interacts specifically with hallmark lipids of the inner membrane leaflet through three lipid interaction domains (LIDs), one overlapping with Box1 and Box2. The most membrane proximal of these was also identified in the GHR-LF-ICD suggesting similar roles in signaling. The only other reports on structural characterization of unbound ICDs of cytokine receptors

available in the literature are on small synthetic peptides. One example is a 17-residue peptide from the IL-2Rβ-ICD covering a sorting signal ($P^{285}SKFFSQL^{292}$) and forming a type I β-turn [80]. Another example is a short peptide from PRLR-ICD containing the Box1 sequence ($I^{243}FPPVPGP^{250}$). The latter work suggested the presence of cis-trans isomerization of Pro248, the third proline of Box1 [81], supported by a suggested interaction with cyclophilin A (CypA), which co-immunoprecipitated with the first 76 intracellular residues of PRLR [82]. Besides for EPOR, PRLR and GHR, and the short 5–10 residue peptides from the IL-4R and IL-5Rα, the ICDs of the remaining 26 class 1 cytokine receptors of the family have not been studied at the atomic level, and only the PRLR and GHR have been studied in the complete forms. This leaves a critical knowledge void inhibiting the understating of how C1CRs signal.

The present paper sets focus on the ICDs of the C1CRs and their structural disorder, and asks why disorder has evolved to manage versatility and fidelity in their signaling. We predict disorder in all human C1CRs and list all known isoforms that differ in the ICD. We analyze their primary structures, globally and locally, and, sequentially and experimentally analyze chain behavior identifying shared and unique characteristics across the family. We show that their sequences are brimming with SLiMs, conferring multispecificity to the chains. Instead of considering the ICDs as passive scaffolds for kinases, we put forward a more complex view of active orchestration via organizational and operational features left uncovered within their disorder.

Materials And Methods

Proteins - expression and purification – Human PRLR-ICD$^{236–396}$ was prepared as described in [79]. PRLR-SF1b-ICD was produced as a GST-tagged fusion protein containing a thrombin cleavage site. PRLR-SF1b-ICD was purified essentially following the procedure for PRLR-LF-ICD [79] with the following modifications: After inoculation of cells from overnight cultures in LB-medium, cells were left growing at 37 °C until OD_{600} = 0.8. The cells were subsequently centrifuged gently for 15 min at 2500 x g at 4 °C, the supernatant discarded, and the cells gently resuspended by swirling in 500 mL ^{15}N- or ^{13}C- and ^{15}N-labeled M9-minimal media (1.5 g KH_2PO4, 3.75 g Na_2HPO_4•$2H_2O$, 0.5 g NaCl, 1 mM $MgSO_4$, 0.5 ml M2 trace solution, 2 g ^{13}C α-D-Glucose, 0.5 g $^{15}NH_4Cl$) with 100 µg/ml Amp. The cell suspension was transferred back into the 5 L Erlenmeyer flask and left growing at 37 °C for 45 min after which protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 hours. The cells were harvested by centrifugation (20 min, 5,000 x g, 4 °C) and stored at -20 °C until thawed on ice and resuspended in 40 mL sonication buffer; 1x PBS (1.4 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4, 10 x stock), 0.1% (v/v) Triton X-100 and one complete EDTA-free protease inhibitor cocktail tablet (Roche Diagnostics GmbH). The cells were sonicated on ice, using an UP400S Ultrasonic Processor, 4 × 30 s with 30 s rest between rounds at 90% amplitude. The cell extracts were centrifuged (25 min, 20,000 x g, 4 °C), the pellets discarded and the supernatants used for purification. The glutathione column (Glutathione Sepharose 4 Fast Flow, GE Healthcare) was prepared by washing with 20 column volumes (CVs) 1xPBS buffer pH 8 and the supernatant from sonication was incubated with the column material for 1 hour at room temperature (RT) under gentle agitation. The column was washed

with 20 CVs 1xPBS buffer pH 8 and GST-PRLR-SF1b-ICD was eluted using 10 CVs elution buffer (50 mM Tris-HCl pH 8, 10 mM reduced glutathione). The sample was dialyzed overnight at 4°C under stirring against 1 L thrombin cleavage buffer (20 mM Tris-HCl, 150 mM NaCl pH 8.4) using a 6000–8000 MWCO dialysis membrane. The sample was added 100 units thrombin (GE Healthcare) and incubated for 2 hours at RT under gentle agitation. After cleavage, the sample was concentrated using centrifugal filters (3000 MWCO, Millipore) and applied to an analytical size exclusion column (Superdex 75 10/300 GL, GE Healthcare) equilibrated with 2 CVs of 50 mM Na_2HPO_4/NaH_2PO_4 pH 7.5, 150 mM NaCl and 0.1 mM DTT. The flowrate was 0.5 ml/min and PRLR-SF1b-ICD was eluted over 1.5 CV.

Lipids - Small unilamellar vesicles (SUVs) containing POPC/POPS and POPC were prepared as in [79].

CD spectroscopy - Far-UV CD spectra were recorded on a Jasco-810 spectropolarimeter from 250 nm to 190 nm with a scan speed of 20 nm/min, bandwidth 1 nm, 2 s response time at 25 °C in a 1 mm quartz cuvette. Protein concentration was 19 µM in 10 mM NaH_2PO_4-NaOH, pH 7.4, 1 mM TCEP. The spectra were averaged over 10 scans with the corresponding spectrum of the buffer subtracted. The resulting spectra were smoothed using a fast Fourier transform, removing the highest frequencies in the spectrum.

NMR spectroscopy - For backbone assignment, ^{13}C-^{15}N-labeled PRLR-SF1b-ICD was concentrated to 500 µM in 20 mM Na_2HPO_4/NaH_2PO_4 pH 7.3 or 7 M urea pH 7.3 (native and denatured conditions, respectively). The samples were added 10% (v/v) D_2O, 5 mM tris(2-carboxyethyl)phosphine (TCEP), and 0.5 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) for referencing in a total volume of 350 µl. The pH was adjusted to 7.3 if needed and the samples were transferred to 5 mm Shigemi tubes. All backbone spectra were recorded at 5 °C on Varian INOVA 750- or 800-MHz (^1H) spectrometers and backbone assignment accomplished from analyses of ^1H-^{15}N-HSQC[83], HNCACB[84], CBCA(CO)NH [85], and HNCO [86]spectra. Free induction decays were transformed and visualized using NMRPipe [87] and analyzed using CcpNmr Analysis [88]. Assignments were done manually. Transient secondary structure elements were identified by secondary chemical shifts (SCSs). These were calculated by subtracting the C^{α} and C' chemical shifts for each residue in 7 M urea from those obtained in 20 mM Na_2HPO_4/NaH_2PO_4. The amount of transient α-helices was assessed as described [89]. Two series of ^1H-^{15}N-HSQC [83] spectra were recorded on ^{15}N-PRLR-SF1b-ICD at 5°C to analyze the relaxation times. The spectra for PRLR-SF1b-ICD were recorded with delay times of 10-1000 ms (T1) and 10–250 ms (T2) with two triplicates in each series. The relaxation decays were fitted to single exponentials and the relaxation times were determined using the CcpNmr Analysis software [88]. For CypA interaction studies, samples containing 100 µM ^{15}N-labelled PRLR-SF1b-ICD or PRLR-LF-ICD with and without 85 µM human CypA (Sigma Aldrich) were prepared in 20 mM Na_2HPO_4/NaH_2PO_4 pH 7.3, 10% (v/v) D_2O, 1 mM TCEP, 0.5 mM DSS. To accommodate slight differences in conditions of the different protein batches, all batches were thoroughly dialyzed against the same buffer before mixing the samples. ^1H-^{15}N-HSQC spectra were recorded on each sample at 5 °C and chemical shift changes were analyzed as combined amide chemical shift changes by $\delta\Delta_{NH}=((\delta\Delta_H)^2+(0.154*\delta\Delta_N)^2)^{1/2}$ [90].

Small angle X-ray scattering – SAXS data were collected at the EMBL beamline P12 at Petra III in Hamburg, Germany [91]. SAXS data on PRLR-LF-ICD and GHR-LF-ICD (20 mM Na_2HPO_4/NaH_2PO_4 (pH 7.3), 10 times molar excess DTT) were collected at the PETRA III, P12 beamline (DESY synchrotron, Hamburg), following standard procedures. A series of different concentrations between 1–6 mg/mL were measured for each protein added 0, 75, 150 and 300 mM NaCl. The SAXS curves for PRLR-ICD-LF and GHR-ICD-LF were analyzed using the form-factor for a Gaussian random coil [92], together with a scaling factor for correcting the protein concentration and a constant to model the background. The fitting was done with the *minimize* method in optimization library in the *scipy* package for python 3 using the L-BFGS-B algorithm. Uncertainties in the fitting parameters were estimated from the diagonal of the inverse Hessian. Pair distance distribution functions were calculated using the BayesApp [93,94] and using the GenApp server (https://genapp.rocks (visited on April 15, 2020)). In the case of PRLR-ICD-LF, a few data points where omitted at low-Q based on the Guinier analysis, but data from 7.5E-3 $Å^{-1}$ and up was used in all cases. The Guinier analysis was also used to truncate the data for GHR-ICD so that only data above 8.4E-3 $Å^{-1}$ was used. Due to the lower S/N, data above 0.3 $Å^{-1}$ was also omitted from the analysis and the Lagrange multiplier was fixed to 1E14, in order to ensure a stable solution with respect to data range and input parameters.

To calculate a reference *Rg* for PRLR-LF-ICD and GHR-LF-ICD, we used the power-law in Eq. 1, where R_0 is a constant related to the persistence length of the polymer, N is the number of amino acids and v is related to the nature of the polymer.

$$Rg = R_0 N^{v} \qquad (1)$$

For R_0 and v, we used the experimental parameters from Kohn and co-workers [95], determined based on chemically unfolded proteins, i.e. R_0 = 1.927 Å and v = 0.598, where the latter value is very close to the theoretical value of 0.588 of self-avoiding polymer chains.

Disorder predictions, sequence alignments and bioinformatics - The original list of C1CRs was manually curated from the list of receptors in [30,31], after which isoforms were extracted from Uniprot [96]. Disorder predictions for each of the sequences were conducted as described previously [13], and the sequences were analyzed by IDDomainSpotter [97], CIDER [98] and Weblogo3.0 [99] using standard setting and TMHMM was used to predict borders between domains [100]. The amino acid propensities for each of the sequences were calculated relative to an ordered reference statistic, as (C(aa) - C_ref(aa)) / C_ref(aa), where C(aa) is the frequency for the individual residue and C_ref(aa) is the frequency of the residue in the reference set. We calculated our ordered reference frequencies, C_ref(aa), from entries in the MobiDB database [101], using the curated *DB* set and selecting all protein regions labeled as 'S'. The same set was used to calculate the frequencies for the Disorder category in Fig. 2b, this time selecting regions labeled as 'D'. To validate the sensitivity to this choice, we also calculated the frequencies from the larger *Derived* MobiDB set, with qualitatively similar results. Finally, we verified the impact of homology reduction on these frequencies by reweighting each protein by the number of other entries in its UniRef50 class, but again observed similar results, and homology reduction was therefore omitted for the

remaining calculations. The error bars in Fig. 2b indicate the 90% confidence interval of the estimated frequencies, calculated using a per-protein bootstrapping procedure with 1000 iterations [102].

SLiMs and phosphorylation sites were predicted in the longest isoforms of group 1 receptors (PRLR-FL, GHR-FL, EPOR-FL and TPOR1). SLiMs were predicted using the ELM resource [103] and filtered based on taxonomic context (*Homo sapiens*) and cellular compartment (cytosol) with a probability cut-off of 100. Modification sites were excluded from analysis. The iGPS 1.0 algorithm [104] was used for prediction of *in vivo* phosphorylation sites. Only instances above a medium threshold [104] was included. Confirmed SliMs and phosphorylation sites were manually curated.

Results

Disorder in C1CRs

The ICDs of the C1CRs have been very sparsely studied structurally, likely because of their high expected abundance of disorder. Moreover, many of the receptors exist in several different alternatively spliced versions, some of which differ in the ICDs. To provide an overview of the ICD isoforms, we conducted a survey of known isoforms of the entire family (Fig. 1c and Table 1). We followed the grouping of the receptors made based on structure and evolution [30,31], but excluded receptors absent in humans and/or for which no ICD could be annotated. Furthermore, IL-2Ra and IL-15Ra were excluded as they lack the structural hallmarks of the family and likely belong to a separate family [105]. This left us with a total of 29 different receptors, distributed with four members in group I (single-chain homodimers), ten members in group II (the gp130 family, not counting receptors binding ancestral cytokines), two members in group III (soluble a-chains, leaving out four receptors without ICDs (IL-27Rb, ciliary neurotrophic factor receptor subunit a (cntfRa), cytokine receptor-like factor 1 (CRLF1), and IL-12Rb), six members in group IV (long-tailed receptor chains) and seven members in group V (short-tailed receptor chains), Table 1. The 29 receptors provided a total of 54 ICD isoforms distributed across all groups. Approximately 40% of the receptors (12 receptors; leukemia inhibitory factor receptor (LIFR), OSMR, IL-6Rb (gp130), IL-27R, IL-6Ra, IL-2Rb, IL21R, IL4Ra, bc (IL-3Rb), CRLF2, gc (IL-2Rg), IL-13Ra1, IL-13Ra2, and IL3Ra) only had one ICD isoform, and three of these were the common receptors IL-6Rb (gp130), bc (IL-3Rb) and gc (IL-2Rg). For the remaining 16 receptors, up to five different isoforms could be identified and nine of the receptors with ICDs > 200 residues had a short isoform < 50 residues. A total of 16 isoforms had unique sequences, typically > 10 residues with an average length of 25±16 residues, and the longest unique sequence of 67 residues belonged to the LEPR isoform C. The average length of longest isoform ICDs were188 residues (group I), 210 residues (group 2), 82 residues (group 3), 335 residues (group 4) and 65 residues (group 5) (Table 1). IL-4Ra had the longest ICD of 575 residues (Table 1). Thus, the C1CR-ICDs are generally long and some have isoforms with unique sequences of considerable length.

The characteristic compositional bias of IDPs makes it possible to predict the degree of disorder in proteins computationally [106]. Almost 10 years ago, computational predictions of disorder was done for five of the C1CRs [107], but disorder predictors have since improved in quality and reproducibility [108],

and no study has examined the entire family in unison. We therefore predicted the disorder profiles of the longest ICD isoforms of all 29 C1CRs as well as the propensity of regions to undergo folding-upon-binding using the ANCHOR scores [109] (Fig. 2a and SI Fig. S1). From these predictions, we observed that the ICDs of the entire family have high scores (>0.5) for disorder along their complete sequence and none were predicted to harbor folded domains. Furthermore, almost all receptors had lower disorder scores in the juxtamembrane 20-50 residues, a region overlapping with the JAK1/2/3;TYK2 binding sites. Along the chains, regions of lower disorder propensity were observed, which at the same time were paralleled with high ANCHOR scores. Such signatures suggest the region to be prone to folding-upon-binding and thus constitutes a potential binding site [109]. Indeed, the dip with the lowest disorder score in PRLR-LF-ICD occurred around residue 610, which corresponds to the region of tyrosine phosphorylation by JAK2 (Y580/Y614), and docking site for STAT5 (YLDP). Comparing the profiles across the group 1 C1CRs revealed a similar pattern of disorder along the first 150-200 residues, although the extent of each of the regions with higher/lower disorder would vary (Fig. 2a). Similar group specific profiles of some similarity in the first half of the ICDs were seen for groups 2, 3, and 4, but not for group 5 (SI Fig. S1); an observation likely reflecting their shorter ICDs. Finally, we compared the disorder profiles for the five different isoforms of the PRLR-ICD (SI Fig. S2). Despite the change in sequence, all the ICD isoforms were predicted to be disordered and with almost identical disorder profiles. This is consistent with the general observation that sequence may change in a family of proteins while the disorder-order profile persist [107,110]. In summary, the predicted disorder profiles support that the ICDs of all the C1CRs are disordered (Fig. 2b), and highlight common disorder profiles with a distribution of binding sites prone to folding-upon-binding.

The ICDs of the C1CRs have compositional biases distinguishing them from other IDPs

To address if the C1CRs-ICDs have physiochemical properties that distinguish them from other IDPs, we compared the amino acid content of the entire family (Fig. 2b) as well as the individual groups (SI Fig. S3) to those of folded proteins and IDPs (for details see methods). The analysis revealed that the C1CRs-ICDs indeed have global sequence compositions that stand out from other IDPs in three ways: First, some amino acids are depleted in the C1CRs-ICDs, namely Met, Arg, Ala and Lys, and are less frequent than in general in IDPs and in folded proteins. Second, Cys, Trp, Leu and Val are significantly more frequent in the C1CRs-ICDs than in other IDPs, and are as frequent as in folded proteins (except Val, which is less frequent than in folded proteins). Third, Pro is highly enriched in the C1CRs-ICDs, and is even more frequent than in both folded proteins and IDPs in general. These differences are remarkable, but the role of these global compositional biases in C1CR functionality remains to be understood. The depletion in positively charged amino acids could be related to prevention of detrimental interactions with the negatively charged inner membrane leaflet to which the C1CRs-ICDs are tethered through their TMDs, or with other negatively charged molecules. The enrichment in Cys, Trp, Leu, Val and the over-enrichment in Pro contradicts common observations for IDPs, as often, when found in IDPs, they are part of SLiMs.

Indeed, the saturation with SLiMs along the C1CRs-ICD chains, as highlighted in Fig. 3 (see below), suggest similar enrichment in binding sites and can points towards large interactomes. Pros are known to preserve disorder in regions of IDPs with residual structural propensities [111], and hence could counter-balance the effects of the enrichment in hydrophobic residues. Furthermore, the chemistry of Pro causes rigidification of the backbone and consequently conformational expansion, as well as the formation of polyPro type II (PPII) helices by Pro-rich motifs [111]. Finally, several SLiMs and modification sites are Pro-based, including binding sites for JAKs and SH3s, and MAPK modification sites, which may increase the relative content of Pro in C1CR-ICDs.

Although sequence identity is often low among related IDPs, the sequence characteristics important for function are typically conserved, whether it being specific SLiMs, global conformational characteristics or specific functional domains [97]. Thus, regions of specific residue biases can be taken to represent domains of different chemical and structural properties, which may contribute differently to the function of the C1CRs-ICDs. To identify putative functional domains of specific physio-chemical properties across the C1CR-ICDs we submitted the sequences to IDDomainSpotter [97]. IDDomainSpotter reveals distinct conformational biases in regions of long IDP sequences by calculating the fractions of specified residues in a sliding window of 15 residues, meaning, that for each residue k, the fraction of the specified residues between k-7 and and k+7 is given. Here, we have analyzed e.g. the charge composition by setting Lys and Arg as positive contributions (+RK) and Glu and Asp (-DE) as negative contributors (net charge). Hence, a given residue k within the sliding window counts as +1 if the position is a Lys or Arg, -1 if the position is Glu or Asp, and 0 for any other residue.

The IDDomainSpotter analysis of the C1CR-ICDs revealed shared profiles for certain residues across the receptors (Fig. 2c), suggesting functional importance. First, they all shared a region of 10-20 residues in the region immediately following the TMD with a positive net charge, followed by a region of ~50-60 residues (only ~20 residues for the shorter TPOR) with a negative net charge (Fig. 2c, green). We denominate these regions as the positive domain (PD) and the negative domain (ND), respectively. For the ICDs of PRLR and GHR, the PD has been shown to specifically interact with negatively charged lipids of the inner leaflet of the membrane [79]. However, the role of the ND is not understood. The negative charges may be relevant for membrane repulsion, or for compaction with the PD when not membrane bound. Alternatively, it could provide negatively charged flanking regions for specific SLiMs, such as for Box2 binding to SH2 domains or Pro-rich motifs binding to SH3 domains, as recently supported by experiments [112]. Second, ~100 residues and onwards from the TMD, the net charge was close to equally balanced along the chains. Another shared property is the almost equal distribution of the unusually abundant Cys and Pro throughout the chains (Fig. 2c, orange and purple). This could suggest that their abundance is related to global conformational properties, rather than e.g. interaction sites or PPIIs. The ICDs of group 1 further shared a pattern of depletion and enrichment (20-50 residues) of the hydrophobic branched amino acids Ile, Leu and Val throughout their chain (Fig. 2c, red). Such hydrophobic side chains are usually less abundant in IDRs because of the energetic penalty of solvation, and hence in IDRs they are often primarily located in e.g. SLiMs, as mentioned above, but could be related to maintaining extended b-structures of relevance to binding. Finally, clustering of Phe, Tyr and Gly

(+FYG) was analyzed, as IDRs enriched in these residues may be involved in liquid-liquid phase separation [113,114], but these were found to be low throughout the chains.

The patterns observed for the group 1 receptors is overall shared across the C1CR-ICDs, with the exception of some noteworthy variations in charge composition. Generally, all the C1CR-ICDs harbor a PD (sometimes in a shorter version), followed by an ND of various length, with subsequent close to net neutral charge along the chains. However, the short ICDs of group 3 only harbor a PD, and lack an ND. Furthermore, the ICDs of IL-31Rb (OSMR) (group 2) and IL-4Rα (group 4) lack regions of substantial net charge throughout their ICDs, including the PD; a trait that may be related to their association with the IL-31Ra and IL-13Ra1, respectively.

Short linear motifs allow expansion of the interactome

In the disorder profiles and the IDDomainSpotter analysis we observed distinct patterns, which may suggest the presence of multiple binding sites [115] (Fig. 2). For all receptors, the first region of low disorder propensity corresponds to the juxtamembrane region containing the most conserved and well-known motifs, Box1 and Box2, involved in JAK/TYK binding. As a Pro-rich motif, Box1 represents one of the most abundant SLiMs in the eukaryotic proteome [116]. The polyPro scaffold inherently provides a conformational bias towards PPII formation [117], which creates a structural predisposition that may drive an interaction via reduced entropic penalty of complex formation [118]. Except for group 5, which lacks canonical Box1 and Box2 motifs, all receptors harbor the conserved PXP motif in Box1 (SI Fig. S4), known to interact with the JAK-FERM domain [41,119]. However, in most receptors, Box1 is further extended to PXXPXP (consensus for group 1: jjPXjPXP, where j is any hydrophobic residue), which thereby accommodates both the minimal SH3 binding motif (PXXP) [120] and the FERM binding motif (PXP) in one combined SLiM (Table 2). This enables competitive binding to Src and JAK kinases [45,121].

In the membrane distal region, SLiMs constituting docking sites for various signaling proteins have been mapped experimentally. These SLiMs are predominantly activated by phosphorylation to recruit SH2-containing proteins, such as those in STAT and SOCS proteins [122,123]. Each group of receptors is known to preferentially recruit a specific STAT for activation [124] and each group therefore contains a specific subset of phospho-tyrosine motifs. Hence, group 1 harbors the STAT5 consensus motif, pYXXL [125], whereas the group 2 harbors the STAT3 consensus motif pYXXQ [126]. In addition to distinct down-stream signaling related SLiMs, many of the experimentally known SLiMs are also related to endocytosis, trafficking and degradation. Some of these are frequent and are well-described motifs experimentally, such as the dileucine-motifs (i.e. [D/E]XXX[L/I] and [D/E]XXLL) seen in LIFR [127] and IL-2Rb [128], and the tyrosine-based motifs (i.e. YXXf), promoting clathrin-dependent endocytosis and internalization, first identified in the TPOR (YRRL) and later in the G-CSFR [129,130]. Additionally, phosphorylation dependent degrons (i.e. DSGXXS) [131–133], promoting ubiquitin-dependent proteasomal degradation are also well characterized in both GHR and PRLR [131–133]. The motifs are summarized in Table 2.

For the longest ICD-isoforms of group 1, we subsequently predicted SLiMs and phosphorylation sites using the Eukaryotic Linear Motifs server (ELM) [103] and the iGPS [104], respectively (see methods), and mapped these to the sequences, marking those already experimentally confirmed (Fig. 3b). We made three important observations. First, the predicted SLiMs as well as their flanking regions, are rich in amino acids that promote extended structures, such as Pro, Val and Leu [134], in accordance with the structures adapted in the bound states [76,135,136], and the compositional analysis made above. Second, from the distribution of SLiMs along the chains, it was evident that clusters of overlapping SLiMs are frequent and distributed across the ICD sequences, interleaved with stretches depleted in SLiMs (Fig. 3b). Clusters of overlapping SLiMs can be considered as scaffolding hot spots where multiple binding events can take place in a controlled manner, largely determined by binding competition, i.e. affinity, concentration and PTMs. However, IDPs may also accommodate simultaneous binding of several partners and thereby orchestrate signaling by bringing relevant proteins into close proximity [137]. As evident from Table 2, similarities between the tyrosine-based motifs are pronounced. Consequently, STAT and SOCS binding motifs may overlap, but also with phosphatase binding sites, such as e.g. for SHP2 [138–140], as well as with tyrosine-based internalization motifs. Thus, regulation of signaling fate by discrimination and availability via compressed motifs appear widespread in C1CRs and is critically linked to properties of disorder. Until recently, C1CR-regions with overlapping motifs have exclusively been characterized in the membrane distal regions. However, accumulating evidence suggest that also the membrane proximal regions contain SLiM-clusters. In GHR, the membrane proximal ~60 residues region contains a LID with an unknown function [79], a ubiquitin dependent degron whereto the E3-ligase, bTrCP docks and promote GHR downregulation [131], as well as JAK2 [48] and Src kinase (LYN) binding sites [45] (Fig. 3c). JAK2 and LYN are the primary kinases in GHR signaling, controlling the activation of JAK2/STAT5 and MAPK pathways, respectively [45]. Both are known to be constitutively associated with the receptor ICD [45,141] and their relative activation of pathways can be perturbed by mutations in the ECD affecting TMD alignment [45]. However, the molecular details of how the change in TMD alignment is associated with pathway selectivity are still unknown, but may be controlled by binding competition of JAK2 and LYN and even further affected by membrane interaction. Similarly, GHR downregulation by bTrCP may likewise by binding competition driven. Thus, this region represents one of the essential composite SLiM-clusters in GHR with hitherto unexplored implications for the regulation GHR signaling.

Typically, multiple binding events in IDPs are often regulated by phosphorylations and can be characterized as binary on/off switches. However, accumulating evidence have revealed that phosphorylations can generate much more complex responses (reviewed in [142]), and multisite phosphorylation can additionally generate sensitive threshold responses as well as graded responses. The third important observation we made was that in the C1CR group 1 receptors, several phosphorylation sites were predicted, but only a small subset of these were well-characterized experimentally. Hence, much remains to be understood in terms of regulation and the many modification sites open the possibility that the C1CR-ICDs can have rheostat regulatory potential [16]. In this way, successive phosphorylations may additively increase (or decrease) binding affinity enabling graded responses, or they may modulate the conformational ensemble, with impact on signaling output.

Importantly, multisite phosphorylations, which are functionally relevant for IDPs, remains to be addressed in the C1CRs.

In summary, interactions by C1CR-ICDs are primarily mediated by SLiMs, creating docking and modification sites for several accessory signaling proteins. Furthermore, clusters of overlapping SLiMs dominate the C1CR-ICDs, which together with the structural plasticity provided by the disorder properties, impart a unique condensed and versatile signaling scaffold, enabling establishment of large interactomes, whose content is controlled by the available pool and concentrations of interaction partners. The spatio-temporal orchestration of signaling therefore rely on availability of the binding partner, affinities and kinetics and altogether eventually determine signaling fate.

Comparison of two PRLR isoforms – maintaining binding and dynamics

In order to investigate different C1CR-ICD isoforms at the molecular level, we expressed and biophysically characterized PRLR-SF1b (residues 259-288) and compared it with PRLR-LF-ICD. The sequence of PRLR-SF1b-ICD is much shorter (32 versus 386 residues) and differs in the last three C-terminal residues. Thus, K286, G287 and K288 of PRLR-LF-ICD are substituted by V286, T287, and P288 in the short isoform, with P288 constituting the new C-terminus. Apart from the loss of multiple interaction sites by being shorter, including the loss of Box2, the chemical change from net positively charged to uncharged with more hydrophobic residues may influence the structural preferences as well as the interactome, including membrane binding. The far-UV CD spectrum of PRLR-SF1b-ICD showed no indications of substantially populated secondary structure, with little negative ellipticity in the 222-210 nm range, and large negative ellipticity around 200 nm (Fig. 4a), which is a characteristic signature of a IDPs. The largely disordered nature of PRLR-SF1b-ICD was further corroborated by a low dispersion in the ^{1}H-dimension of the ^{1}H-^{15}N-HSQC NMR spectrum (Fig. 4b), indicating a homogeneity of the chemical environment of the backbone amides. One way to detect transient secondary structures is to measure C^{α} NMR chemical shifts to derive the secondary chemical shifts (SCSs) (see methods). Generally, consecutive positive SCSs indicate a-helical structure, whereas negative SCSs indicates extended structure [143,144]. The SCSs of the C^{α} of PRLR-SF1b-ICD (Fig. 4c) were overall slightly negative, with positive values only for A281 and H282. These constitute the N-terminal of the first transient a-helix, TH1 previously identified in the PRLR-LF-ICD (Fig. 4d). However, where PRLR-LF-ICD had positive SCSs for F279-T287, the PRLR-SF1b-ICD only had positive SCSs for A281 and H282. For residues N-terminal to TH1, the negative SCS signature remained similar for the two isoforms (Fig. 4e). This difference strongly indicates that the three unique C-terminal residues V286, T287, and P288 in PRLR-SF1b-ICD distinctly lowered the a-helical content compared to PRLR-LF-ICD, while not changing the structural properties of the remainder of the sequence. However, similar to changes in structural propensities, changes in dynamics can affect how an IDP is presented to a binding partner. We therefore compared NMR parameters that report on fast-time scale backbone dynamics (ns-ps) measured on PRLR-SF1b-ICD to those of the first 100 residues of PRLR-LF-ICD [32] (Fig.

S5). The average relaxation rates of PRLR-SF1b-ICD were similar but slightly lower than for PRLR-LF-ICD [32], and from a comparison to a segmental-motion model of a completely unstructured chain, similar fast timescale dynamics was seen over the length of the peptide (SI Fig. S5b, red line) [145]. Last, a plot of R_2/R_1 revealed no clear indication of non-random structure throughout PRLR-SF1b-ICD (Fig. 4f). Thus, but isoforms are disordered and have fast dynamics similar to a completely unstructured chain in accordance with the disorder predictions. Remarkably, TH1, observed in the PRLR-LF-ICD was eliminated in the PRLR-SF1b-ICD, demonstrating that two isoforms differing only in three residues may have different structural propensities.

PRLR-SF1b-ICD includes part of the first LID, and may thus interact with the membrane, or, as two positively charged residues are substituted, membrane interaction could be abolished. To address this, ^{1}H-^{15}N-HSQC spectra were recorded of PRLR-SF1b-ICD in the presence and absence of small unilamellar vesicles (SUVs) consisting of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine/1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-L-serine (POPC/POPS) and chemical shift changes and intensity ratios were compared (Fig. 5A,B). The peaks corresponding to residues K259-F268 disappeared upon addition of POPC/POPS SUVs and chemical shift changes were observed for residues I276-L283 (Fig. 5a), indicating interaction with the SUVs. Furthermore, the low intensity ratio across the entire sequence suggested that the NMR observables for all of PRLR-SF1b-ICD were affected (Fig 5b). Thus, the lipid interaction observed for PRLR-LF-ICD [79] was preserved in PRLR-SF1b-ICD, despite structural and chemical changes. This is in accordance with the lack of induced α-helicity in PRLR-LF-ICD upon lipid binding [79].

To investigate if the *cis/trans* propensity of the prolines differed between PRLR-LF-ICD and PRLR-SF1b-ICD, we utilized the backbone chemical shifts and sequences of both PRLR-LF-ICD and PRLR-SF1b-ICD to calculate the probability for *cis* conformation [146]. Even in IDPs, prolines are only in *cis* conformation in ~5% of all proline peptide bonds [147,148] and for PRLR, only P269 was observed to have a slightly above average probability of being in *cis* for both isoforms with an absolute *cis* proline probability of 8-10% (SI Table S1). In contrast to previous suggestions of cis/trans probabilities in an 8-mer Box1 peptide focusing on P272 (I267-P274) [81], our results suggested a very low probability of P272 cis conformation of 0.01 (1%) for both isoforms. This is in line with recent work demonstrating that very short peptides strongly overestimate the amount of *cis*-Pro, supporting previous overestimation of the cis probability in the PRLR Box1 peptide [148]. Box1 was previously suggested to interact with CypA [149]. To investigate if PRLR-LF-ICD and PRLR-SF1b-ICD would interact with CypA, ^{1}H-^{15}N-HSQC spectra in the presence and absence of unlabeled CypA were recorded and chemical shift differences calculated (Fig. 5c,e). Weak NMR peaks originating from the lowly populated *cis* state of P269 were present in the spectra, and as CypA may be able to recognize either states, we investigated both peak sets. No changes were observed for PRLR-LF-ICD nor PRLR-SF1b-ICD, neither in the chemical shifts nor in the intensity ratios (Fig. 5d,f), suggesting that CypA are not interacting with the PRLR-ICD isoforms. Thus, Pro *cis/trans* isomerization appears not to be dominant in the free state of the PRLR, which is also reflected in the absence of CypA interaction.

Network rewiring by isoforms

From the comparison of the two isoforms of PRLR, we only observed small differences in dynamics with preserved lipid binding, although structural differences between the two were prominent. However, a major reduction in the interactome will take place when removing 90% of the ICD, which removed numerous SLiMs and phosphorylation sites, including docking sites for STAT and SOCS. On the other hand, ICD isoforms with longer stretches of unique sequence may have gained additional interactions sites. For the C1CRs with different ICD-isoforms, we predicted which common binding-SLiMs were gained or lost, disregarding potential phosphorylation sites and receptor-unique SLiMs (Table S2). The unique sequences were found to carry distinct SLiMs. In the case of group 1, a 14-3-3 binding SLiM has previously been identified in PRLR isoform 1 [150]; a SLiM originally discovered active in cytokine receptors in the IL-9R [151], but predictions suggested the presence of an additional 14-3-3 motif. Compared to the PRLR-LF-ICD, however, the short forms did not include the STAT-docking site or the original 14-3-3 binding SLiM. Instead, a different 14-3-3 binding SLiM was present in the unique sequences (SF1a and SF1c) (Fig. 6a). For TPOR, isoform 2 had two 14-3-3 SLiMs, which were absent in isoform 1, while the unique sequences of GHR and EPOR isoforms, which do not have any known 14-3-3 binding SLiMs, were much shorter and without any predictable common SLiMs. Thus, for PRLR, the preservation of the 14-3-3 SLiM suggest a key regulatory function, one of which may be to attenuate receptor signaling, as suggested [150]. Relevantly, the LFs of EPOR, GHR and PRLR all had a phosphorylation dependent degron, interacting with the SkpSCF-betaTrCP1 complex or the Skp1_Cullin-Fbox, leading to ubiquitylation and degradation, as shown experimentally for PRLR and GHR, where it negatively regulates receptor stability [131,132]. These were not identified in any of the shorter isoforms, which also have been seen to be stabilized on the membrane [152], perhaps because of the lack of associated proliferative signaling and hence lack of need for immediate down regulation.

For other receptors, e.g. the IL-31Ra (GLMR) and LEPR isoforms, we found that unique sequences introduced PDZ binding SLiMs at the new C-termini (Fig. 6b). Furthermore, when present, each IL-31Ra (GLMR) receptor isoform had a unique PDZ binding SLiM, allocating the isoforms to interact with different classes (Class 1,2 and 3) of PDZ domains (Fig. 6b) [153,154]. The same was true for the LEPR (Fig. 6c). In fact, the introduction of a PDZ SLiM in the C-terminus in one isoform, absent in the other, was observed for several receptors including G-CSFR, IL-7Ra, IL-9Ra, and GM-CSFRa (Table S2). Why these isoforms need PDZ binding motifs is not clear, but several scaffolding proteins with specialized subcellular localization and tissue specificity exist, known to contain multiple PDZ domains by which they orchestrate supramolecular complexes. Binding of the IL-5Ra-ICD to a PDZ domain from syntenin (Fig. 1b) [78] supports the involvement of further scaffolding proteins for formation of larger signaling complexes. PDZ containing protein may be of relevance to the C1CRs and could engage proteins from the NHERF and PSD-95 families [155], which also scaffold kinases as Fyn [156]. Alternatively, E3-ligases belonging to the MARCHs family coordinate binding via PDZ domains and are relevant for ubiquitylation of proteins in the intracellular membranes [157]. However, besides the complex between the IL-5Ra-ICD and the PDZ domain from syntenin, complexes of class 1 cytokine receptors with PDZ domains remain to

be experimentally explored. Finally, for all receptors with isoforms, the longest isoform carries the interaction with STATs, either STAT5 or STAT3 or both, but, additionally also carry a binding motif for TNF receptor-associated factor (TRAF)-2 or TRAF-6, none of which are found in other, shorter isoforms. In a few cases, the STAT and/or TRAF motifs are maintained in the second longest isoform, and sometimes a shift between STAT5 and STAT3 or between TRAF-2 and TRAF-6 occurs.

Thus, for the C1CRs, the disorder predictions and experimental characterization of selected representatives have suggested that the isoforms maintain structural disorder, and their presence suggests several mechanisms by which disorder orchestrates signaling. The first is the complete removal of a large part of the ICD, eliminating SLiMs typically for STAT activation, TRAF-interaction and downregulation by degradation via degron activation. In this way the shorter isoforms act as negative regulators, or decoy receptors, of signaling, as seen for the short forms of the PRLR and GHR [59,152]. However, these isoforms still maintain binding capacity as seen from for membrane binding of Sf1b above. The second mechanism by which isoforms orchestrate signaling is via rewiring of the interactome to access completely new networks, exemplified by the addition and removal of binding sites for e.g. 14-3-3 proteins and PDZ domains. This allows for different signaling profiles dependent on expression profiles of the C1CR-ICD isoforms. However, more studies into the network rewiring of the C1CRs are warranted, and the analysis made here provides a starting point.

The conformational ensemble of C1CR-ICDs

IDPs are functional without taking on a single, well-defined tertiary structure. Yet, they cannot adequately be described as simple statistical coil chains equally populating all possible conformations allowed by their backbone torsion angles. Instead, IDPs display varying degrees of compaction and elongations, and contain transient, short- and long-range structural organizations. Hence, the disorder of the C1CR-ICDs not only infer flexibility and high accessibility of binding sites, but certain chain dimensions and spatial organizations may influence the organization of the signaling complexes and orchestration of protein interactions, and in the end, signaling outcome. Currently, the conformation and dimensions of IDPs cannot be quantitatively predicted from sequence [158,159]. Nonetheless, the balance between chain-chain and chain-solvent interactions that determines the conformational preference is related to specific sequence features that influence the conformational ensembles in predictable ways [98,158,160,161]. One set of these relates to global compositional sequence features (i.e. parameters that are independent on the sequence order) and the fraction of charged residues and the net charge per residue are particularly important [160,162]. In addition, features relating to sequence patterning, especially the patterning of oppositely charged residues and expansion promoting residues, influence compaction [159]. However, the current difficulties in consistently predicting the conformational ensemble of all IDPs reflects that these behaviors are encoded in sequence features yet to be unraveled.

IDPs have been classified into five compositional groups in a diagram of states [162] based on their fraction of positively charged residues (f^+) and fraction of negatively charged residues (f^-). These two

global parameters are combined into two measures underlying a diagram of states: the fraction of charged residues (FCR = $f^+ + f^-$) and the net charge per residue (NPCR = $f^+ - f^-$). An explanation of the relation between these parameters and the properties of the chain is given in the supplemental data. Of 879 IDRs longer than 15 residues found in DisProt, CIDER [98] classified 40% as belonging to R1, 35% to R2, 22% to R3, and 3% to either R4 or R5 [158]. For each C1CR-ICD, the sequence of isoform 1 was submitted to CIDER [98], except for LEPR and G-CSFR, for which isoform B and 3, respectively, were selected as these were the longest isoforms (see Table 1). For GM-SCFRα, both isoform 1 and 2 were analyzed because they differed in more than 50% of their C-terminal sequences (see Table 1). The C1CR-ICDs generally fell close to the boundary between R1 and R2, with most belonging to R1 (61%) (Fig. 2a), suggesting a preference for compact, but still dynamic, heterogenous conformational ensembles [158]. Nonetheless, in particular for sequences belonging to R2, their overall charge neutrality means that their conformational preference cannot be predicted from global composition alone [158,160]. Furthermore, it should be noted that the boundary between R1 and R2 has been determined *ad hoc*, and has been suggested to be positioned at lower FCR for longer sequences [158,160]. Furthermore, for ICDs > 100 residues or with a high proline fraction (> 0.15), no qualitative prediction of the conformation can be made for sequences of R1, as these tend to have more extended conformations than their scores predict.

Since almost all the C1CR-ICDs are long IDPs of R1 and R2, the conformational preferences cannot be predicted from global composition alone but may also be influenced by e.g. sequence patterning. Particular the patterning, or mixing, of oppositely charged residues is important, as well as expansion driving- and aromatic residues. The parameter κ reports how well positively and negatively charged residues are segregated across the sequence and is normalized between 0 and 1, with κ close to zero representing sequences with evenly distributed charges, while sequences with κ close to 1 have highly segregated charges. It has been shown that as κ approaches 1, the conformational ensemble becomes more compact [162]. However, since κ is calculated by normalizing to the most segregated sequence within the given composition, a specific κ value will not have the same meaning for two sequences with different FCR and |NCPR| values. Furthermore, for long IDRs, such as most of the C1CR-ICDs, κ is calculated only within a window of 5 and 6 residues, ignoring long-range effects. κ is most informative for sequences with an FCR above 0.25 and NCPR between -0.1 and +0.1, for which a κ below 0.12 is considered low and a κ above 0.25 is high. Especially for polyampholytic sequences with an FCR beyond 0.4, charge patterning is predicted to have a major impact on the conformation [161]. There is one such example, namely the GM-CSFRa, which has an FCR of 0.41, an NCPR of -0.07 and a κ of 0.25, suggesting chain compaction.

The position of the far majority (94%) of the C1CR-ICDs in R1 and R2 is a consequence of their low net charges. Their FCR values are in the mediocre range of 0.1<FCR<0.3 [161], while at the same time, their NCPR is close to 0, demonstrating that they are near-symmetrical polyampholytes. For the C1CR groups with long ICDs (1, 2 and 4), the group average FCRs (0.21; 0.23; 0.19) and NCPRs (-0.06; -0.06; -0.05) are remarkably similar, suggesting that charge properties are a conserved trait. The similarity of these parameters also allows us to compare their κ values more directly, going from a group average of 0.20 for

group 1, 0.18 for group 2 to 0.22 for group 4. This is consistent with the IDDomainSpotter analysis presented earlier (Fig. 2c). Here we found that almost all of the C1CR-ICDs harbored a PD immediately following their TMD, succeeded by an ND, and with net charge neutral regions for the remainder of the chain. Together, this suggest that the influence of the global charges and the charge patterning on the conformational ensembles are consistent throughout group 1, 2 and 4, except for IL-31Rb (OSMR) (group 2) and IL-4Rα (group 4). As mentioned, the shorter ICDs of group 3 and 5 result in somewhat different global charge properties.

The Ω parameter both describes the patterning of the charged residues as well as of proline. Like for κ, Ω is normalized between 0 and 1, with Ω close to zero representing sequences with evenly distributed charges and prolines, while sequences with Ω close to 1 have highly segregated charges and prolines [163]. It has been shown that when Ω approaches 1, the preference for expanded conformations increases [163]. A high fraction of Pro (> ~15%) may cause more expanded conformations as Pro prefers to be solvated and promotes stiffness. Five of the C1CR-ICDs had a high fraction of Pro: IL-27R, IL-6Rα, IL-11Rα, βc, IL-13Rα2 (Fig. 7a, top). The amino acid fraction and IDDomainSpotter analysis (Fig. 2) revealed that the Pros are unusually abundant in the C1CR-ICDs and close to equally distributed. From the CIDER analysis we found that Ω, like κ, is similar for many of the C1CR-ICDs, but is lower for the shorter sequences (Fig. 7a). This could simply be a consequence of the Pro-rich Box1 sequences, leading to relatively higher scores in the shorter sequences.

To summarize, the theoretical analysis of the C1CR-ICD sequence parameters known to influence compaction suggest that many of them may be similarly biased towards a specific degree of extension or compaction of their conformational ensembles, but that this degree cannot be predicted from sequence. Hence, to determine this bias, we experimentally investigated the degree of compaction and its responsiveness to salt by SAXS using two long, representative C1CR-ICDs, namely that of PRLR-LF-ICD (Fig. 7b,c) and GHR-LF-ICD (SI Fig. S6). The SAXS profiles of the PRLR-LF-ICD were consistent with those expected for fully disordered proteins, and was fitted to an *Rg* of 57.3 ± 1.4 Å in 20 mM phosphate buffer. The predicted *Rg* of the PRLR-LF-ICD for a fully random coil state is, according to Kohn et al. [95], 65 Å, suggesting that the PRLR-LF-ICD in isolation populate a slightly compacted ensemble. The pair distance distribution function (*P(r)*), which is a histogram of distance distributions within the protein, peaks at ~45 Å and has a D_{max} of ~200 Å. Increasing the concentration of salt to 300 mM did not significantly affect the fitted *Rg* nor the *P(r)* distribution (Fig. 7b,c), suggesting that the global degree of compaction of PRLR-ICD is not sensitive to salt, as otherwise often observed for more charged IDPs [164], perhaps related to the high content of Pro and branched amino acids. The same trends were observed from SAXS on the GHR-LF-ICD, having a similarly slightly compacted ensemble that was insensitive to salt (SI Fig. S6). Hence, the ICDs across the C1CR family having similar global charge properties and patterning (Fig. 7a), may populate similarly compacted ensembles, although this remains to be experimentally more broadly verified.

Versatile and controlled orchestration of signaling by unique structural disorder in C1CRs

It is remarkable that the entire family of C1CRs, differentiating into > 50 isoforms, are all predicted to be disordered in their entire ICD sequence. Nonetheless, the disordered ICDs are critically understudied, leaving us with a naive and too simplistic schematic view of the ICDs as passive strings of varying lengths with kinases constitutively attached. In the present paper we have highlighted the properties linked to disorder responsible for controlling the diverse signaling by C1CRs (Table 2) and asked: Why has disorder been selected for governing intracellular C1CR signaling? Their complete disordered nature stands in contrast to the majority of other types of single-pass transmembrane receptors such as the receptor tyrosine kinases, where intracellular signaling is mainly governed by intrinsic kinase activity. We have here shown that the long disordered ICDs of C1CRs are brimmed with clusters of multifunctional SLiMs throughout their length, suggesting that one explanation is the signaling versatility and scaffolding capacity of this type of ICD. Furthermore, we have outlined that overlapping SLiMs are prevalent in the C1CR-ICDs, hinting that the disordered ICDs further allow for complex regulation of diverse signaling through competition and regulation of interactions with a plethora of different binding partners through multispecificity. Thereby, activation becomes dependent on the coupled equilibria and kinetics of two (or more) binding events. Indeed, the ability of a distinct region in the disordered ICD to bind to many different proteins is facilitated by structural adaptation and folding-upon-binding [165,166]. Additionally, the C1CR-ICDs, are hot spots for multiple phosphorylation events of which only a few a well-characterized as binary on-off switches. This directly – or indirectly – affect affinities and additionally expands the number of states accessible by the chain at any time. We have further suggested that an additional layer to this regulation is added by the existence of different C1CR-ICD isoforms, in which entire groups of SLiMs can be eliminated and new ones added, a feature that is much easier for IDPs to successfully obtain during evolution compared to folded proteins. By controlled expression of the isoforms, a complete rewiring of the interaction network can be done. Hence, the full disordered nature of the C1CR-ICDs allows for a fascinatingly versatile and complex interaction hub.

Can such signal complexity be facilitated through a simple string with kinases attached? Our sequence analysis and experimental studies have revealed biases in the C1CR-ICDs that differentiate them from being simple statistical coils. They have conserved distinct compositional biases that differentiate them from other IDPs; biases that are distributed throughout their chains, including the presence of disordered domains of specific physiochemical properties. This suggests that these compositional biases are representing shared functionalities yet to be characterized. Our experimental SAXS data on the long ICDs of the archetypical receptors GHR and PRLR revealed that they are slightly more compacted than expected for a fully random coil (~57 Å versus 65 Å for PRLR-LF-ICD). This indeed suggests an inherent conformational bias based on the conservation of certain sequence properties maintained across the family. Importantly, however, it should be kept in mind that in the cell, the C1CR-ICDs are most likely never completely void of interactions at any point. Previous characterizations of PRLR-LF-ICD and GHR-LF-ICD have revealed the presence of LIDs, further suggesting distinct organizational features at the membrane

interface where also many kinases are tethered. Additionally, we found that the same sets of SLiMs are placed differently in the C1CR-ICDs with variable distances, which may provide an additional tuning of the signaling outcome, both via the length of their disordered spacers as well as the properties of these [167]. Thus, SLiM organization within the chain may imprint different affinities in different complexes despite their exploitation of identical SLiMs, providing an additional layer of the spatio-temporal orchestration of signaling.

Conclusion

In conclusion, we suggest that the C1CR-ICDs are far from simple strings with constitutively bound kinases. Rather, they carry both organizational and operational features left uncovered in their disorder, but of key importance for understanding orchestration of signaling. How these features operate in the higher-order oligomers of the C1CRs, bringing ICDs from several chains in close proximity, increases the dimension of future studies. For example, the mere volume taken up by more chains may allow them to generate higher order ensembles of specific properties. In such disordered reaction chambers, lower affinity interactions may be boosted and even shared between chains adding features, mechanisms and complexities to regulations, yet to be discovered. Taken together, it is clear that the understanding of the fascinating ability of these long, completely disordered chains to orchestrate complex signaling pathways are still in its infancy.

Abbreviations

C1CR, class 1 cytokine receptors; CD, circular dichroism; ECD, extracellular domain; FCR, fraction charged residues; ICD, intracellular domain; IDP, intrinsically disordered protein; LID, lipid interaction domain; NCPR, net charge per residue; NMR, nuclear magnetic resonance; PPII, polyproline II; PTM, post-translational modification; SAXS, small angle X-ray scattering; SCS, secondary chemical shift; SLiM, short linear motif; SUV, small unilamellar vesicles; TMD, transmembrane domains;

Declarations

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

PS, KB and BBK conceived the study and designed the manuscript. PS did the SLiM, PTM and signaling analyses. KB did the CIDER and IDDomainSpotter analysis, and acquired and analyzed the SAXS data, the latter in collaboration with MHP and LA. JHM did the disorder analyses. GWH and MN purified PRLR isoforms and did the NMR analyses, and WB did computational analyses. KB, PS, and JHM did most figures and finalized the data. PS, KB and BBK wrote the manuscript. All authors have read and approved the submitted version.

Acknowledgements

We would like to acknowledge all the important work done on decomposing the regions, domains and residues of the C1CR-ICDs essential to signaling and apologize to those whose work we did not cite. We thank Johan G. Olsen for valuable discussions and Mikkel M. for support. The synchrotron SAXS data was collected at beamline P12 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). We would like to thank Melissa Graewert for assistance during the beamtime.

Funding

This paper is a contribution from REPIN – rethinking protein interactions funded by the Novo Nordisk Foundation Challenge program (grant #NNF18OC0033926). The project was also funded by the Novo Nordisk Foundation SYNERGY program (grant#NNF15OC0016670).

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its additional information files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Wright PE, Dyson HJ. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J Mol Biol. Academic Press; 1999;293:321–31.
2. Sigler PB. Acid blobs and negative noodles. Nature. 1988;333:210–2.
3. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. J Mol Biol. 2004;337:635–45.
4. Kumar R, Thompson E. Role of Phosphorylation in the Modulation of the Glucocorticoid Receptor's Intrinsically Disordered Domain. Biomolecules. 2019;9:95.

5. Mitrea DM, Yoon M-K, Ou L, Kriwacki RW. Disorder-function relationships for the cell cycle regulatory proteins p21 and p27. Biol Chem. 2012;393:259–74.
6. Follis AV, Galea CA, Kriwacki RW. Intrinsic protein flexibility in regulation of cell proliferation: Advantages for signaling and opportunities for novel therapeutics. Adv Exp Med Biol. 2012;
7. Staby L, O'Shea C, Willemoës M, Theisen F, Kragelund BB, Skriver K. Eukaryotic transcription factors: paradigms of protein intrinsic disorder. Biochem J. 2017;474:2509–32.
8. Uversky VN, Dunker AK. Understanding protein non-folding. Biochim. Biophys. Acta - Proteins Proteomics. Elsevier; 2010. p. 1231–64.
9. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, et al. Analysis of Molecular Recognition Features (MoRFs). J Mol Biol. 2006;362:1043–59.
10. Lee S-H, Kim D-H, J. Han J, Cha E-J, Lim J-E, Cho Y-J, et al. Understanding Pre-Structured Motifs (PreSMos) in Intrinsically Unfolded Proteins. Curr Protein Pept Sci. Bentham Science Publishers Ltd.; 2012;13:34–54.
11. Chhabra Y, Wong HY, Nikolajsen LF, Steinocher H, Papadopulos A, Tunny KA, et al. A growth hormone receptor SNP promotes lung cancer by impairment of SOCS2-mediated degradation. Oncogene. 2018;37:489–501.
12. Krystkowiak I, Davey NE. SLiMSearch: a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions. Nucleic Acids Res. 2017;45.
13. Prestel A, Wichmann N, Martins JM, Marabini R, Kassem N, Broendum SS, et al. The PCNA interaction motifs revisited: thinking outside the PIP-box. Cell Mol Life Sci. Springer; 2019;76:4923–43.
14. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, et al. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 2004;32:1037–49.
15. Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, et al. Identification, analysis, and prediction of protein ubiquitination sites. Proteins Struct Funct Bioinforma. 2010;78:365–80.
16. Lee CW, Ferreon JC, Ferreon ACM, Arai M, Wright PE. Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. Proc Natl Acad Sci. National Academy of Sciences; 2010;107:19290–5.
17. Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, et al. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science (80-). American Association for the Advancement of Science; 2020;367:694–9.
18. Kim TH, Tsang B, Vernon RM, Sonenberg N, Kay LE, Forman-Kay JD. Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. Science (80-). American Association for the Advancement of Science; 2019;365:825–9.
19. Arai M, Sugase K, Dyson HJ, Wright PE. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. Proc Natl Acad Sci U S A. National Academy of Sciences; 2015;112:9614–9.

20. Yang J, Gao M, Xiong J, Su Z, Huang Y. Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding. Protein Sci. Blackwell Publishing Ltd; 2019;28:1952–65.
21. Shammas SL, Crabtree MD, Dahal L, Wicky BIM, Clarke J. Insights into coupled folding and binding mechanisms from kinetic studies. J. Biol. Chem. American Society for Biochemistry and Molecular Biology Inc.; 2016. p. 6689–95.
22. Olsen JG, Teilum K, Kragelund BB. Behaviour of intrinsically disordered proteins in protein – protein complexes with an emphasis on fuzziness. Cell Mol Life Sci. Springer International Publishing; 2017;74:3175–83.
23. Borgia A, Borgia MB, Bugge K, Kissling VM, Heidarsson PO, Fernandes CB, et al. Extreme disorder in an ultrahigh-affinity protein complex. Nature. Nature Publishing Group; 2018;555:61–6.
24. Kjaergaard M, Kragelund BB. Functions of intrinsic disorder in transmembrane proteins. Cell Mol Life Sci. Birkhauser Verlag AG; 2017;74:3205–24.
25. Hendus-Altenburger R, Kragelund BB, Pedersen SF. Structural Dynamics and Regulation of the Mammalian SLC9A Family of Na+/H+ Exchangers. Curr Top Membr. Academic Press Inc.; 2014;73:69–148.
26. Warne T, Moukhametzianov R, Baker JG, Nehmé R, Edwards PC, Leslie AGW, et al. The structural basis for agonist and partial agonist action on a β1-adrenergic receptor. Nature. Nature Publishing Group; 2011;469:241–5.
27. Minezaki Y, Homma K, Nishikawa K. Intrinsically Disordered Regions of Human Plasma Membrane Proteins Preferentially Occur in the Cytoplasmic Segment. J Mol Biol. Academic Press; 2007;368:902–13.
28. Kassem N, Kassem MM, Pedersen SF, Pedersen PA, Kragelund BB. Yeast recombinant production of intact human membrane proteins with long intrinsically disordered intracellular regions for structural studies. Biochim Biophys Acta - Biomembr. Elsevier BV; 2020;1862:183272.
29. Brooks AJ, Dehkhoda F, Kragelund BB. Cytokine Receptors. Princ Endocrinol Horm Action. 2016;
30. Boulay J-L, O'Shea JJ, Paul WE. Molecular Phylogeny within Type I Cytokines and Their Cognate Receptors. Immunity. Cell Press; 2003;19:159–63.
31. Liongue C, Ward AC. Evolution of Class I cytokine receptors. BMC Evol Biol. 2007;7:120.
32. Bugge K, Papaleo E, Haxholm GW, Hopper JTS, Robinson C V., Olsen JG, et al. A combined computational and structural model of the full-length human prolactin receptor. Nat Commun. Nature Publishing Group; 2016;7:11578.
33. Shields DC, Harmon DL, Nunez F, Whitehead AS. The evolution of haematopoietic cytokine/receptor complexes. Cytokine. 1995;7:679–88.
34. Brooks AJ, Waters MJ. The growth hormone receptor: mechanism of activation and clinical implications. Nat Rev Endocrinol. Nature Publishing Group; 2010;6:515–25.
35. Baker SJ, Rane SG, Reddy EP. Hematopoietic cytokine receptor signaling. Oncogene. Nature Publishing Group; 2007;26:6724–37.

36. Lebrun JJ, Ali S, Ullrich A, Kelly P a. Proline-rich sequence-mediated Jak2 association to the prolactin receptor is required but not sufficient for signal transduction. J. Biol. Chem. 1995. p. 10664–70.
37. Rui H, Kirken R a, Farrar WL. Activation of receptor-associated tyrosine kinase JAK2 by prolactin. J Biol Chem. 1994;269:5364–8.
38. Pezet A, Buteau H, Kelly PA, Edery M. The last proline of Box 1 is essential for association with JAK2 and functional activation of the prolactin receptor. Mol Cell Endocrinol. 1997;129:199–208.
39. Bagley CJ, Woodcock JM, Stomski FC, Lopez AF. The Structural and Functional Basis of Cytokine Receptor Activation: Lessons From the Common β Subunit of the Granulocyte-Macrophage Colony-Stimulating Factor, Interleukin-3 (IL-3), and IL-5 Receptors. Blood. American Society of Hematology; 1997;89:1471–82.
40. Murakami M, Narazaki M, Hibi M, Yawata H, Yasukawa K, Hamaguchi M, et al. Critical cytoplasmic region of the interleukin 6 signal transducer gp130 is conserved in the cytokine receptor family. Proc Natl Acad Sci U S A. National Academy of Sciences; 1991;88:11349–53.
41. Ferrao RD, Wallweber HJA, Lupardus PJ. Receptor-mediated dimerization of JAK2 FERM domains is required for JAK2 activation. Elife. 2018;7:1–21.
42. Ferrao R, Wallweber HJA, Ho H, Tam C, Franke Y, Quinn J, et al. The Structural Basis for Class II Cytokine Receptor Recognition by JAK1. Structure. Elsevier Ltd; 2016;24:897–905.
43. Wallweber HJA, Tam C, Franke Y, Starovasnik MA, Lupardus PJ. Structural basis of recognition of interferon-α receptor by tyrosine kinase 2. Nat Struct Mol Biol. 2014;21:443–8.
44. Tilbrook P, Palmer G, Bittorf T, McCarthy D, Wright M, Sarna M, et al. Maturation of Erythroid Cells and Erythroleukemia Development Are Affected by the Kinase Activity of Lyn. CANCER Res. 2001;61:2453–8.
45. Rowlinson SW, Yoshizato H, Barclay JL, Brooks AJ, Behncken SN, Kerr LM, et al. An agonist-induced conformational change in the growth hormone receptor determines the choice of signalling pathway. Nat Cell Biol. 2008;10:740–7.
46. Lannutti BJ, Drachman JG. Lyn tyrosine kinase regulates thrombopoietin-induced proliferation of hematopoietic cell lines and primary megakaryocytic progenitors. Blood. American Society of Hematology; 2004;103:3736–43.
47. Clevenger C V., Medaglia M V. The protein tyrosine kinase P59fyn is associated with prolactin (PRL) receptor and is activated by PRL stimulation of T-lymphocytes. Mol Endocrinol. 1994;8:674–81.
48. Brooks AJ, Dai W, O'Mara ML, Abankwa D, Chhabra Y, Pelekanos R a, et al. Mechanism of Activation of Protein Kinase JAK2 by the Growth Hormone Receptor. Science (80-). 2014;344:1249783–1249783.
49. Brown RJ, Adams JJ, Pelekanos RA, Wan Y, McKinstry WJ, Palethorpe K, et al. Model for growth hormone receptor activation based on subunit rotation within a receptor dimer. Nat Struct Mol Biol. 2005;12:814–21.
50. Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, Bateman A, et al. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Mol

Cell. 2012;46:871–83.

51. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, et al. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proc Natl Acad Sci. 2006;103:8390–5.
52. Light S, Sagit R, Ekman D, Elofsson A. Long indels are disordered: A study of disorder and indels in homologous eukaryotic proteins. Biochim Biophys Acta - Proteins Proteomics. Elsevier; 2013;1834:890–7.
53. Light S, Sagit R, Sachenkova O, Ekman D, Elofsson A. Protein expansion is primarily due to indels in intrinsically disordered regions. Mol Biol Evol. 2013;30:2645–53.
54. Goffin V, Bogorad RL, Touraine P. Identification of Gain-of-Function Variants of the Human Prolactin Receptor. Methods Enzymol. 1st ed. Elsevier Inc.; 2010. p. 329–55.
55. Bogorad RL, Courtillot C, Mestayer C, Bernichtein S, Harutyunyan L, Jomain J-B, et al. Identification of a gain-of-function mutation of the prolactin receptor in women with benign breast tumors. Proc Natl Acad Sci. 2008;105:14533–8.
56. Kline JB, Rycyzyn M a, Clevenger C V. Characterization of a novel and functional human prolactin receptor isoform (deltaS1PRLr) containing only one extracellular fibronectin-like domain. Mol Endocrinol. 2002;16:2310–22.
57. Tan D, Huang K, Ueda E. S2 deletion variants of human PRL receptors demonstrate that extracellular domain conformation can alter conformation of the intracellular signaling domain. Biochemistry. 2008;47:479–89.
58. Kline JB, Clevenger C V. Identification and Characterization of the Prolactin-binding Protein in Human Serum and Milk. J Biol Chem. 2001;276:24760–6.
59. Trott JF, Hovey RC, Koduri S, Vonderhaar BK. Alternative splicing to exon 11 of human prolactin receptor gene results in multiple isoforms including a secreted prolactin-binding protein. J Mol Endocrinol. 2003;30:31–47.
60. Hu Z-Z, Meng J, Dufau ML. Isolation and Characterization of Two Novel Forms of the Human Prolactin Receptor Generated by Alternative Splicing of a Newly Identified Exon 11. J Biol Chem. 2001;276:41086–94.
61. Kline JB, Roehrs H, Clevenger C V. Functional characterization of the intermediate isoform of the human prolactin receptor. J Biol Chem. 1999;274:35461–8.
62. Pezet A, Ferrag F, Kelly PA, Edery M. Tyrosine Docking Sites of the Rat Prolactin Receptor Required for Association and Activation of Stat5. J Biol Chem. 1997;272:25043–50.
63. Bachelot A, Bouilly J, Liu Y, Rebourcet D, Leux C, Kuttenn F, et al. Sequence variation analysis of the prolactin receptor C-terminal region in women with premature ovarian failure. Fertil Steril. 2010;94:2772–5.
64. Lesueur L, Edery M, Ali S, Paly J, Kelly PA, Djiane J. Comparison of long and short forms of the prolactin receptor on prolactin-induced milk protein gene transcription. Proc Natl Acad Sci U S A. 1991;88:824–8.

65. Das R, Vonderhaar BK. Transduction of prolactin's (PRL) growth signal through both long and short forms of the PRL receptor. Mol Endocrinol. 1995;9:1750–9.
66. Huang K, Ueda E, Chen Y, Walker AM. Paradigm-shifters: phosphorylated prolactin and short prolactin receptors. J Mammary Gland Biol Neoplasia. 2008;13:69–79.
67. Yamauchi T, Kaburagi Y, Ueki K, Tsuji Y, Stark GR, Kerr IM, et al. Growth Hormone and Prolactin Stimulate Tyrosine Phosphorylation of Insulin Receptor Substrate-1, -2, and -3, Their Association with p85 Phosphatidylinositol 3-Kinase (PI3-kinase), and Concomitantly PI3-kinase Activation via JAK2 Kinase. J Biol Chem. 1998;273:15719–26.
68. Amaral MEC, Cunha DA, Anhê GF, Ueno M, Carneiro EM, Velloso LA, et al. Participation of prolactin receptors and phosphatidylinositol 3-kinase and MAP kinase pathways in the increase in pancreatic islet mass and sensitivity to glucose during pregnancy. J Endocrinol. 2004;183:469–76.
69. Meng J, Tsai-Morris C-H, Dufau ML. Human prolactin receptor variants in breast cancer: low ratio of short forms to the long-form human prolactin receptor associated with mammary carcinoma. Cancer Res. Cancer Res; 2004;64:5677–82.
70. Huang K, Walker AM. Long term increased expression of the short form 1b prolactin receptor in PC-3 human prostate cancer cells decreases cell growth and migration, and causes multiple changes in gene expression consistent with reduced invasive capacity. Prostate. Prostate; 2010;70:37–47.
71. Bugge K, Lindorff-Larsen K, Kragelund BB. Understanding single-pass transmembrane receptor signaling from a structural viewpoint - what are we missing? FEBS J. 2016;283:4424–51.
72. Li Q, Wong YL, Huang Q, Kang C. Structural Insight into the Transmembrane Domain and the Juxtamembrane Region of the Erythropoietin Receptor in Micelles. Biophys J. Biophysical Society; 2014;107:2325–36.
73. Li Q, Wong YL, Lee MY, Li Y, Kang C. Solution structure of the transmembrane domain of the mouse erythropoietin receptor in detergent micelles. Sci Rep. Nature Publishing Group; 2015;5:1–10.
74. Bocharov E V., Lesovoy DM, Bocharova O V., Urban AS, Pavlov K V., Volynsky PE, et al. Structural basis of the signal transduction via transmembrane domain of the human growth hormone receptor. Biochim Biophys Acta - Gen Subj. Elsevier B.V.; 2018;1862:1410–20.
75. Schmidt T, Ye F, Situ AJ, An W, Ginsberg MH, Ulmer TS. A conserved ectodomain-transmembrane domain linker motif tunes the allosteric regulation of cell surface receptors. J Biol Chem. American Society for Biochemistry and Molecular Biology Inc.; 2016;291:17536–46.
76. Kung WW, Ramachandran S, Makukhin N, Bruno E, Ciulli A. Structural insights into substrate recognition by the SOCS2 E3 ubiquitin ligase. Nat Commun. Springer US; 2019;10:1–14.
77. Zhou MM, Huang B, Olejniczak ET, Meadows RP, Shuker SB, Miyazaki M, et al. Structural basis for IL-4 receptor phosphopeptide recognition by the IRS-1 PTB domain. Nat Struct Biol. Nature Publishing Group; 1996;3:388–93.
78. Kang BS, Cooper DR, Jelen F, Devedjiev Y, Derewenda U, Dauter Z, et al. PDZ tandem of human syntenin: Crystal structure and functional properties. Structure. Cell Press; 2003;11:459–68.

79. Haxholm GW, Nikolajsen LF, Olsen JG, Fredsted J, Larsen FH, Goffin V, et al. Intrinsically disordered cytoplasmic domains of two cytokine receptors mediate conserved interactions with membranes. Biochem J. 2015;468:495–506.
80. Subtil A, Delepierre M, Dautry-Varsat A. An α-helical signal in the cytosolic domain of the interleukin 2 receptor β chain mediates sorting towards degradation after endocytosis. J Cell Biol. The Rockefeller University Press; 1997;136:583–95.
81. O'Neal KD, Chari M V, Mcdonald CH, Cook RG, Yu-Lee LY, Morrisett JD, et al. Multiple cis-trans conformers of the prolactin receptor proline-rich motif (PRM) peptide detected by reverse-phase HPLC, CD and NMR spectroscopy. Biochem J. 1996;315 (Pt 3:833–44.
82. Syed F, Rycyzyn M a, Westgate L, Clevenger C V. A novel and functional interaction between cyclophilin A and prolactin receptor. Endocrine. 2003;20:83–90.
83. Kay L, Keifer P, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc. 1992;114:10663–5.
84. Wittekind M, Mueller L. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. J Magn Reson Ser B. 1993;101:201–5.
85. Grzesiek S, Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. J Am Chem Soc. 1992;114:6291–3.
86. Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson. 1990;89:496–514.
87. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6:277–93.
88. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, et al. The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins Struct Funct Bioinforma. 2005;59:687–96.
89. Zhang H, Neal S, Wishart DS. RefDB: A database of uniformly referenced protein chemical shifts. J Biomol NMR. 2003;25:173–95.
90. Mulder F a, Schipper D, Bott R, Boelens R. Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. J Mol Biol. 1999;292:111–23.
91. Blanchet CE, Spilotros A, Schwemmer F, Graewert MA, Kikhney A, Jeffries CM, et al. Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). J Appl Crystallogr. International Union of Crystallography; 2015;48:431–43.
92. Debye P. Molecular-weight Determination by Light Scattering. J Phys Colloid Chem. 1947;51:18–32.
93. Hansen S. Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering. J Appl Crystallogr. 2000;33:1415–21.

94. Hansen S. *BayesApp*: a web site for indirect transformation of small-angle scattering data. J Appl Crystallogr. International Union of Crystallography (IUCr); 2012;45:566–7.
95. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, et al. Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci U S A. National Academy of Sciences; 2004;101:12491–6.
96. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47:D506–15.
97. Millard PS, Bugge K, Marabini R, Boomsma W, Burow M, Kragelund BB. IDDomainSpotter: Compositional bias reveals domains in long disordered protein regions—Insights from transcription factors. Protein Sci. Blackwell Publishing Ltd; 2020;29:169–83.
98. Holehouse AS, Das RK, Ahad JN, Richardson MOG, Pappu R V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophys J. Biophysical Society; 2017;112:16–21.
99. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188–90.
100. Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. Proceedings Int Conf Intell Syst Mol Biol. 1998;6:175–82.
101. Piovesan D, Tabaro F, Paladin L, Necci M, Mičetí I, Mičetí M, et al. MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. Nucleic Acids Res. 2018;46:471–6.
102. Cheng Y, LeGall T, Oldfield CJ, Dunker AK, Uversky VN. Abundance of intrinsic disorder in protein associated with cardiovascular disease. Biochemistry. American Chemical Society; 2006;45:10448–60.
103. Gouw M, Michael S, Amano-S ´ Anchez HS´, Kumar M, Andr´ A, Zeke A, et al. The eukaryotic linear motif resource-2018 update. Nucleic Acids Res. 2018;46.
104. Song C, Ye M, Liu Z, Cheng H, Jiang X, Han G, et al. Systematic Analysis of Protein Phosphorylation Networks From Phosphoproteomic Data. Mol Cell Proteomics. MCP Papers in Press; 2012;11:1070–83.
105. Giri JG, Kumaki S, Ahdieh M, Friend DJ, Loomis A, Shanebeck K, et al. Identification and cloning of a novel IL-15 binding protein that is structurally related to the alpha chain of the IL-2 receptor. EMBO J. Wiley; 1995;14:3654–63.
106. Mészáros B, Erdös G, Dosztányi Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 2018;46:W329–37.
107. Sigalov AB, Uversky VN. Differential occurrence of protein intrinsic disorder in the cytoplasmic signaling domains of cell receptors. Self Nonself. 2011;2:55–72.
108. Nielsen JT, Mulder FAA. Quality and bias of protein disorder predictors. Sci Rep. Nature Publishing Group; 2019;9:1–11.

109. Mészáros B, Simon I, Dosztányi Z. Prediction of Protein Binding Regions in Disordered Proteins. Casadio R, editor. PLoS Comput Biol. Public Library of Science; 2009;5:e1000376.
110. Christensen LF, Staby L, Bugge K, O'Shea C, Kragelund BB, Skriver K. Evolutionary conservation of the intrinsic disorder-based Radical-Induced Cell Death1 hub interactome. Sci Rep. Nature Research; 2019;9:1–15.
111. Theillet F-X, Kalmar L, Tompa P, Han K-H, Selenko P, Dunker AK, et al. The alphabet of intrinsic disorder. Intrinsically Disord Proteins. Informa UK Limited; 2013;1:e24360.
112. Gorelik M, Davidson AR. Distinct peptide binding specificities of Src homology 3 (SH3) protein domains can be determined by modulation of local energetics across the binding interface. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2012;287:9168–77.
113. Lin Y, Currie SL, Rosen MK. Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. J Biol Chem. American Society for Biochemistry and Molecular Biology Inc.; 2017;292:19110–20.
114. Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell. Cell Press; 2018;174:688-699.e16.
115. Dosztanyi Z, Meszaros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics. 2009;25:2745–6.
116. Rubin GM, Yandell MD, Wortman JR, Miklos GLG, Nelson CR, Hariharan IK, et al. Comparative genomics of the Eukaryotes. Science (80-). 2000;287:2204–15.
117. Brown AM, Zondlo NJ. A Propensity Scale for Type II Polyproline Helices (PPII): Aromatic Amino Acids in Proline-Rich Sequences Strongly Disfavor PPII Due to Proline − Aromatic Interactions. Biochemistry. 2012;51:5041–51.
118. Ferreon JC, Hilser VJ. Thermodynamics of Binding to SH3 Domains: The Energetic Impact of Polyproline II (P II) Helix Formation †. Biochemistry. 2004;43:7787–97.
119. Ferrao R, Lupardus PJ. The Janus Kinase (JAK) FERM and SH2 Domains: Bringing Specificity to JAK–Receptor Interactions. Front Endocrinol (Lausanne). Frontiers; 2017;8:71.
120. Saksela K, Permi P. SH3 domain ligand binding: What's the consensus and where's the specificity? FEBS Lett. 2012;586:2609–14.
121. Fresno Vara JA, Cáceres MA, Silva A, Martín-Pérez J. Src family kinases are required for prolactin induction of cell proliferation. Mol Biol Cell. 2001;12:2171–83.
122. Floss DM, Mrotzek S, Klöcker T, Schröder J, Grötzinger J, Rose-John S, et al. Identification of canonical tyrosine-dependent and non-canonical tyrosine-independent STAT3 activation sites in the intracellular domain of the interleukin 23 receptor. J Biol Chem. 2013;288:19386–400.
123. Hörtner M, Nielsch U, Mayr LM, Heinrich PC, Haan S. A new high affinity binding site for suppressor of cytokine signaling-3 on the erythropoietin receptor. Eur J Biochem. 2002;269:2516–26.

124. Morris R, Kershaw NJ, Babon JJ. The molecular details of cytokine signaling via the JAK/STAT pathway. Protein Sci. Blackwell Publishing Ltd; 2018;27:1984–2009.
125. Klingmüller U, Bergelson S, Hsiao JG, Lodish HF. Multiple tyrosine residues in the cytosolic domain of the erythropoietin receptor promote activation of STAT5. Proc Natl Acad Sci U S A. National Academy of Sciences; 1996;93:8324–8.
126. Stahl N, Farruggella TJ, Boulton TG, Zhong Z, Darnell JE, Yancopoulos GD. Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. Science (80-). American Association for the Advancement of Science; 1995;267:1349–53.
127. Thiel S, Behrmann I, Timmermann A, Dahmen H, Müller-Newen G, Schaper F, et al. Identification of a Leu-Ile internalization motif within the cytoplasmic domain of the leukaemia inhibitory factor receptor. Biochem J. 1999;339:15–9.
128. Amano Y, Yoshino K, Kojima K, Takeshita T. A hydrophobic amino acid cluster inserted into the C-terminus of a recycling cell surface receptor functions as an endosomal sorting signal. Biochem Biophys Res Commun. Elsevier Inc.; 2013;441:164–8.
129. Hunter MG, McLemore M, Link DC, Loveland M, Copelan A, Avalos BR. Divergent pathways in COS-7 cells: Mediate defective internalization and intracellular routing of truncated G-CSFR forms in SCN/AML. PLoS One. 2008;3:1–8.
130. Hitchcock IS, Chen MM, King JR, Kaushansky K. YRRL motifs in the cytoplasmic domain of the thrombopoietin receptor regulate receptor internalization and degradation. Blood. 2008;112:2222–31.
131. Da Silva Almeida AC, Strous GJ, van Rossum AGSH. βTrCP Controls GH Receptor Degradation via Two Different Motifs. Mol Endocrinol. The Endocrine Society; 2012;26:165–77.
132. Li Y, Suresh Kumar KG, Tang W, Spiegelman VS, Fuchs SY. Negative Regulation of Prolactin Receptor Stability and Signaling Mediated by SCFβ-TrCP E3 Ubiquitin Ligase. Mol Cell Biol. American Society for Microbiology; 2004;24:4038–48.
133. Meyer L, Deau BD, Forejtníková H, Duménil D, Margottin-Goguet F, Lacombe C, et al. β-Trcp mediates ubiquitination and degradation of the erythropoietin receptor and controls cell proliferation. Blood. 2007;109:5215–22.
134. Hörtner M, Nielsch U, Mayr LM, Johnston JA, Heinrich PC, Haan S. Suppressor of Cytokine Signaling-3 Is Recruited to the Activated Granulocyte-Colony Stimulating Factor Receptor and Modulates its Signal Transduction. J Immunol. The American Association of Immunologists; 2002;169:1219–27.
135. Bergamin E, Wu J, Hubbard SR. Structural Basis for Phosphotyrosine Recognition by Suppressor of Cytokine Signaling-3. Structure. Elsevier; 2006;14:1285–92.
136. Kershaw NJ, Murphy JM, Liau NPD, Varghese LN, Laktyushin A, Whitlock EL, et al. SOCS3 binds specific receptor-JAK complexes to control cytokine signaling by direct kinase inhibition. Nat Struct Mol Biol. Nature Publishing Group; 2013;20:469–76.
137. Berk AJ. Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. Oncogene. Nature Publishing Group; 2005. p. 7673–85.

138. Bjørbæk C, Lavery HJ, Bates SH, Olson RK, Davis SM, Flier JS, et al. SOCS3 mediates feedback inhibition of the leptin receptor via Tyr985. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2000;275:40649–57.
139. Nicholson SE, De Souza D, Fabri LJ, Corbin J, Willson TA, Zhang JG, et al. Suppressor of cytokine signaling-3 preferentially binds to the SHP-2-binding site on the shared cytokine receptor subunit gp130. Proc Natl Acad Sci U S A. National Academy of Sciences; 2000;97:6493–8.
140. Sasaki A, Yasukawa H, Shouda T, Kitamura T, Dikic I, Yoshimura A. CIS3/SOCS-3 suppresses erythropoietin (EPO) signaling by binding the EPO receptor and JAK2. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2000;275:29338–47.
141. Argetsinger LS, Campbell GS, Yang X, Witthuhn BA, Ihle JN, Carter-su C. Identification of JAK2 As a Growth Hormone Receptor-Associated Tyrosine Kinase. Cell. 1993;74:237–44.
142. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. Nature Publishing Group; 2015;16:18–29.
143. Nielsen JT, Mulder FAA. Potenci: Prediction of temperature, neighbor and ph-corrected chemical shifts for intrinsically disordered proteins. J Biomol NMR. Springer Netherlands; 2018;70:141–65.
144. Kjaergaard M, Brander S, Poulsen FM. Random coil chemical shift for intrinsically disordered proteins: Effects of temperature and pH. J Biomol NMR. Springer; 2011;49:139–49.
145. Schwalbe H, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, et al. Structural and Dynamical Properties of a Denatured Protein. Heteronuclear 3D NMR Experiments and Theoretical Simulations of Lysozyme in 8 M Urea. Biochemistry. 1997;36:8977–91.
146. Shen Y, Bax A. Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. J Biomol NMR. 2010;46:199–204.
147. Murrali MG, Piai A, Bermel W, Felli IC, Pierattelli R. Proline Fingerprint in Intrinsically Disordered Proteins. ChemBioChem. Wiley-VCH Verlag; 2018;19:1625–9.
148. Alderson TR, Lee JH, Charlier C, Ying J, Bax A. Propensity for *cis* -Proline Formation in Unfolded Proteins. ChemBioChem. Wiley-VCH Verlag; 2018;19:37–42.
149. Zheng J, Koblinski JE, Dutson L V., Feeney YB, Clevenger C V. Prolyl isomerase cyclophilin A regulation of Janus-activated kinase 2 and the progression of human breast cancer. Cancer Res. American Association for Cancer Research; 2008;68:7769–78.
150. Olayioye MA, Guthridge MA, Stomski FC, Lopez AF, Visvader JE, Lindeman GJ. Threonine 391 Phosphorylation of the Human Prolactin Receptor Mediates a Novel Interaction with 14-3-3 Proteins. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2003;278:32929–35.
151. Sliva D, Gu M, Zhu YX, Chen J, Tsai S, Du X, et al. 14-3-3zeta interacts with the alpha-chain of human interleukin 9 receptor. Biochem J. Portland Press Ltd; 2000;345 Pt 3:741–7.
152. Amit T, Bergman T, Dastot F, Youdim MBH, Amselem S, Hochberg Z. A Membrane-Fixed, Truncated Isoform of the Human Growth Hormone Receptor1. J Clin Endocrinol Metab. The Endocrine Society; 1997;82:3813–7.

153. Bezprozvanny I, Maximov A. Classification of PDZ domains. FEBS Lett. John Wiley & Sons, Ltd; 2001;509:457–62.
154. Sheng M, Sala C. PDZ Domains and the Organization of Supramolecular Complexes. Annu Rev Neurosci. Annual Reviews; 2001;24:1–29.
155. Cunningham R, Biswas R, Steplock D, Shenolikar S, Weinman E. Role of NHERF and scaffolding proteins in proximal tubule transport. Urol Res. Springer; 2010;38:257–62.
156. Kalia L V., Salter MW. Interactions between Src family protein tyrosine kinases and PSD-95. Neuropharmacology. Pergamon; 2003;45:720–8.
157. Bauer J, Bakke O, Morth JP. Overview of the membrane-associated RING-CH (MARCH) E3 ligase family. N Biotechnol. Elsevier B.V.; 2017;38:7–15.
158. Ruff KM. Predicting Conformational Properties of Intrinsically Disordered Proteins from Sequence. Kragelund BB, Skriver K, editors. Intrinsically Disord Proteins Methods Protoc Methods Mol Biol. Springer Science+Business Media, LLC, part of Springer Nature. In press; 2020;2141.
159. Marsh JA, Forman-Kay JD. Sequence Determinants of Compaction in Intrinsically Disordered Proteins. Biophys J. 2010;98:2383–90.
160. Das RK, Ruff KM, Pappu R V. Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr Opin Struct Biol. Elsevier Ltd; 2015;32:102–12.
161. Ginell GM, Holehouse AS. Analyzing the Sequences of Intrinsically Disordered Regions with CIDER and localCIDER. Kragelund BB, Skriver K, editors. Intrinsically Disord Proteins Methods Protoc Methods Mol Biol. Springer Science+Business Media, LLC, part of Springer Nature. In press; 2020;2141.
162. Das RK, Pappu R V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci. 2013;110:13392–7.
163. Martin EW, Holehouse AS, Grace CR, Hughes A, Pappu R V., Mittag T. Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. J Am Chem Soc. American Chemical Society; 2016;138:15323–35.
164. Müller-Späth S, Soranno A, Hirschfeld V, Hofmann H, Rüegger S, Reymond L, et al. Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc Natl Acad Sci U S A. National Academy of Sciences; 2010;107:14609–14.
165. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics. 2008;9:S1.
166. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol. 2002;12:54–60.
167. Sørensen CS, Kjaergaard M. Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics. Proc Natl Acad Sci U S A. 2019;116:23124–31.
168. Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, et al. IBS: an illustrator for the presentation and visualization of biological sequences: Fig. 1. Bioinformatics. 2015;31:3359–61.

Tables

Due to technical limitations, the tables are only available as a download in the supplemental files section.

Figures

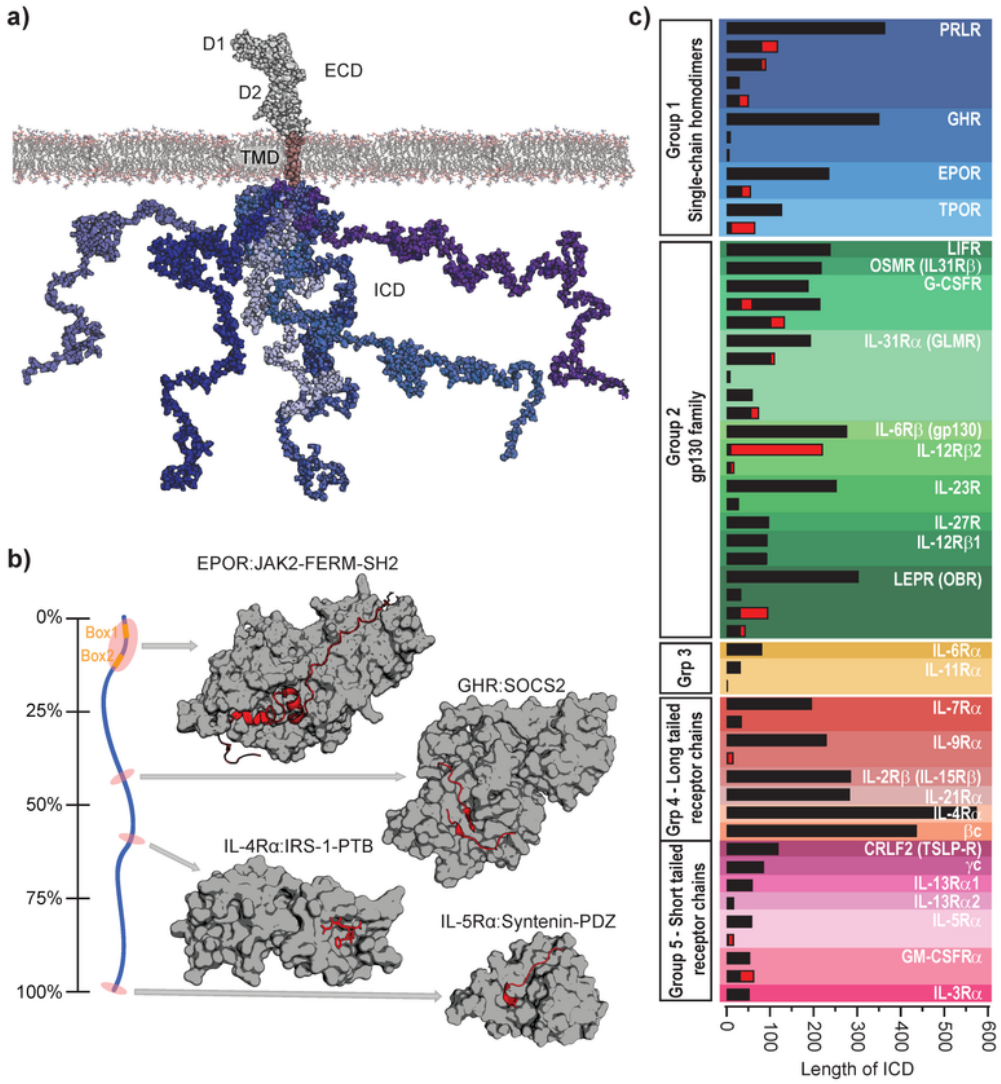

Figure 1

The C1CR family - structures and ICD isoforms. a) Structural model of the full-length human PRLR [32], with the ECD in light grey, the TMD in pink and a total of six ICD chains shown in blue shades to illustrate its disordered conformational ensemble. b) Left: A representative model of a long C1CR-ICD shown in blue and length given in % and with Box1 and Box2 highlighted in orange. In red shade are highlighted the approximate extent and position of small fragments of different C1CR-ICDs solved in complex with signaling proteins. Right: Three-dimensional structures of signaling proteins (grey surfaces) in complex with C1CR-ICD fragments (red cartoon and sticks) illustrated by the FERM-SH2 domain of JAK2 in complex with EPOR-ICD279-334 (pdb code 6E2Q) (top), SOCS2 in complex with GHR-ICD591-603 (pdb code 6I5J, 6I5N) (middle top), the PTB domain of IRS-1 in complex with IL-4R490-500 (pdb code 1IRS) (middle bottom) and the PDZ domain from syntenin in complex with IL-5R417-420 (pdb code 1OBZ) (bottom). c) A total of 29 receptors of the C1CR family with an ICD and their isoforms. The length of the ICDs are to scale and red indicates the length of unique sequences differing from the long-form. The receptors are divided into 5 groups according to [30].

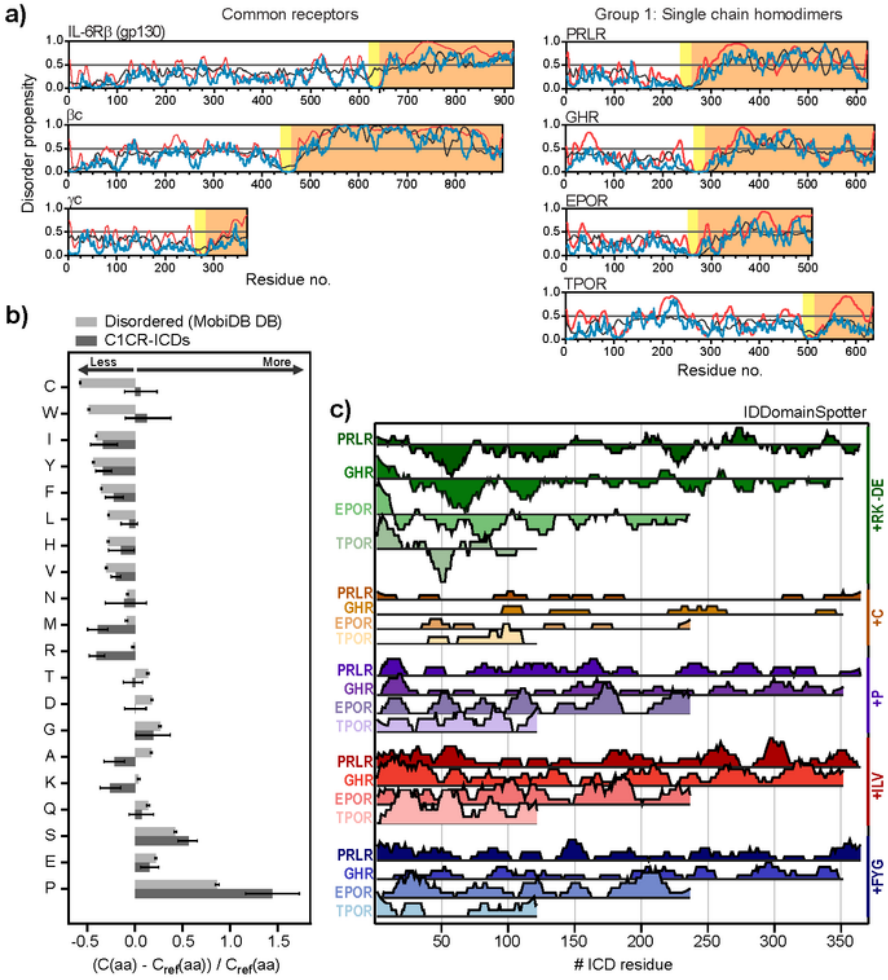

Figure 2

Intrinsic disorder and compositional bias of C1CR-ICDs. a) Disorder prediction for C1CR common receptors and group 1. The disorder propensity, ranging from 0 to 1 was predicted using IUPred2A (blue), ANCHOR (black) and Pondr-fit VSL2 (red) and is plotted as a function of residue number. The boundaries between the ECD (white background), TMD (yellow background) and ICD (orange background) were predicted using TMHMM v. 2.0. The sequence numbering includes the signal peptide. b) Fractional

differences in composition between the C1CR-ICDs (dark grey) or a set of IDPs (light grey), and a set of folded proteins calculated for each amino acid type (see methods for details). Negative values denote that the amino acid is less frequent than in folded proteins, and positive values denote that the amino acid is more frequent than in folded proteins. The arrows indicate the directions of “more” abundant than in folded proteins, and “less” abundant than in folded proteins. The error bars indicate the 90% confidence interval of the estimated frequencies, calculated using a per-protein bootstrapping procedure with 1000 iterations [102]. c) IDDomainSpotter profiles of C1CR the group 1 ICDs. Profiles display scores for +Arg,Lys-Asp,Glu (green), +Cys (brown), +Pro (purple), + Ile,Leu,Val (red) and +Phe,Tyr,Gly (blue) calculated over 15 residue windows for each of the ICDs.

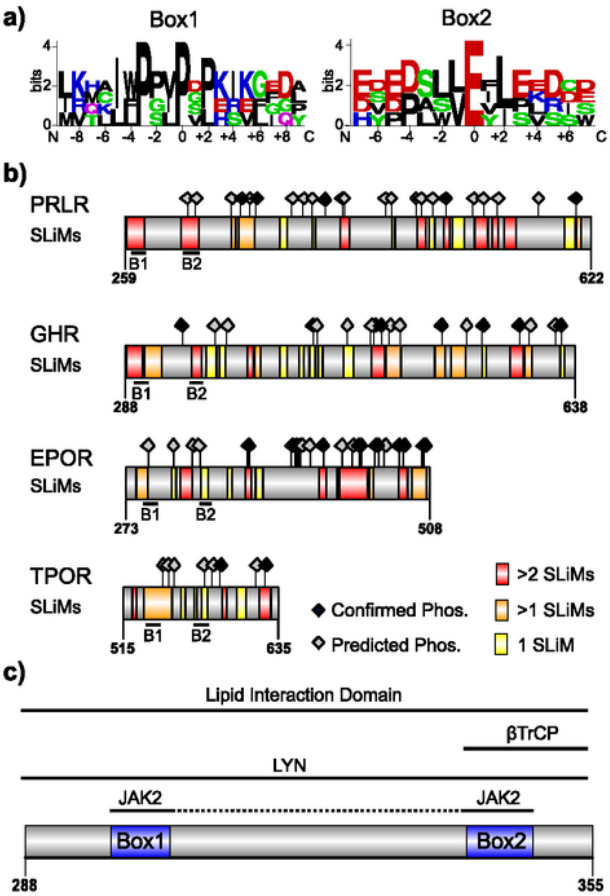

Figure 3

Short linear motifs form the basis for signaling choreography. a) Sequence logo of the amino acid distribution in C1CR group 1 Box1 and Box2 motifs. Letter-height indicates the relative abundance of the given amino acid at the specific position. Red indicates an acidic residue, blue - a basic residue, green – a polar residue, black – a hydrophobic residue and pink – a neutral residue [99]. b) SLiM and phosphorylation site mapping in C1CR group 1 ICDs. Regions with SLiMs are highlighted and colored

according to the amount of overlap. Black indicates regions with no overlap. Yellow indicates regions with two overlapping SLiMs, whereas red indicates regions with more than two overlapping SLiMs. Predicted phosphorylation sites are indicated by grey diamonds [104] and confirmed and well-characterized sites by black diamonds. c) Zoom on the membrane proximal 67-residue region of GHR-LF-ICD containing overlapping binding sites. Box1 and Box2 are highlighted in blue. Proteins were illustrated in IBS 1.0.3 [168].

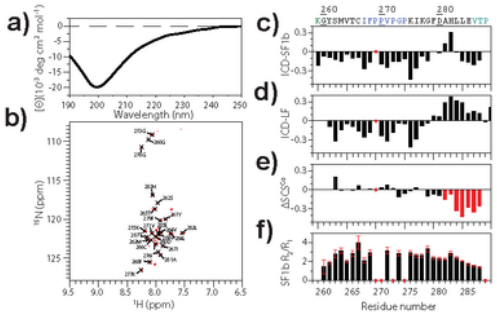

Figure 4

Comparison of molecular details of PRLR-ICD isoforms. a) Far-UV CD spectrum and b) 1H,15N-HSQC spectrum of PRLR-SF1b-ICD. NMR secondary chemical shifts (SCSs) of Cα for c) PRLR-SF1b-ICD and d) PRLR-LF-ICD. e) Differences in Cα SCSs between PRLR-LF-ICD and PRLR-SF1b-ICD, red indicates lower Cα SCSs in PRLR-SF1b-ICD compared to PRLR-LF-ICD. f) Ratios of R1 and R2 relaxation rates of PRLR-SF1b-ICD. Error bars represent errors from the fit, and red circles highlight prolines and/or unassigned residues.

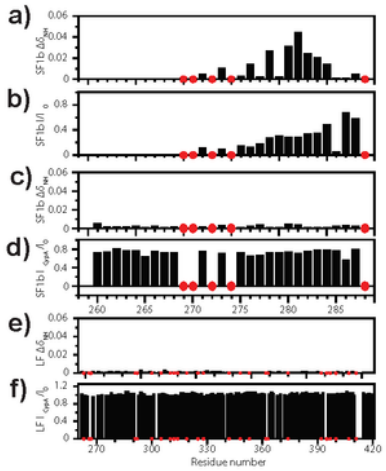

Figure 5

Interaction preferences persist in isoforms of PRLR-ICD. a) Chemical shift perturbations of PRLR-SF1b-ICD upon addition of POPC/POPS SUVs. b) Intensity ratios of PRLR-SF1b-ICD between unbound and in the presence of POPC/POPS SUVs. Chemical shift differences and intensity ratios in the absence and presence of CypA for PRLR-SF1b-ICD C) and D) and PRLR-LF-ICD E) and F), respectively. Red circles highlight prolines and/or unassigned residues.

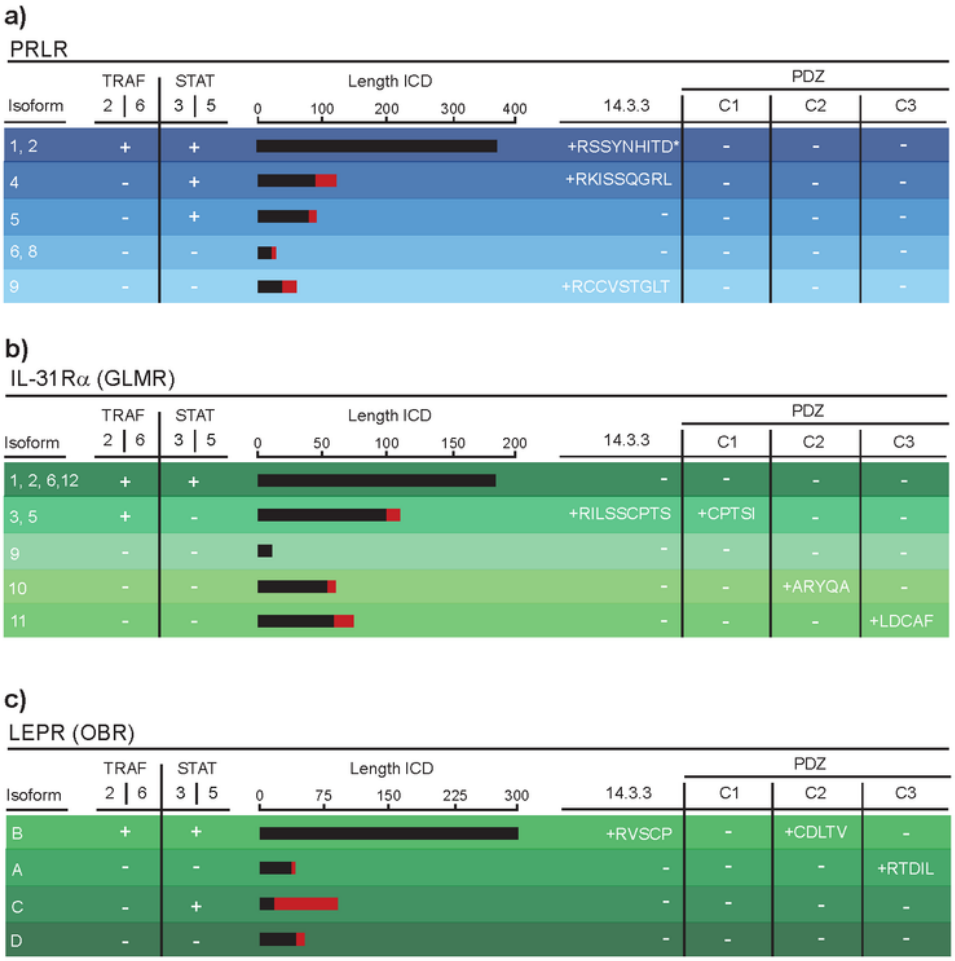

Figure 6

Network rewiring by disorder. Predicted and selected binding SLiMs in the isoforms of a) PRLR, b) IL-31R (GLMR), and c) LEPR (OBR). Introduction of potential new binding SLiMs in the unique sequences are indicated by the sequence of the SLiMs, and the presence of SLiMs in the LF of the isoforms is indicated with a +. When a SLiM is not present, this is indicated by a -. The length of the ICD is indicated by the scale bars on top and red lines illustrate the length of the unique sequence. For other C1CR isoforms, see Table S2.

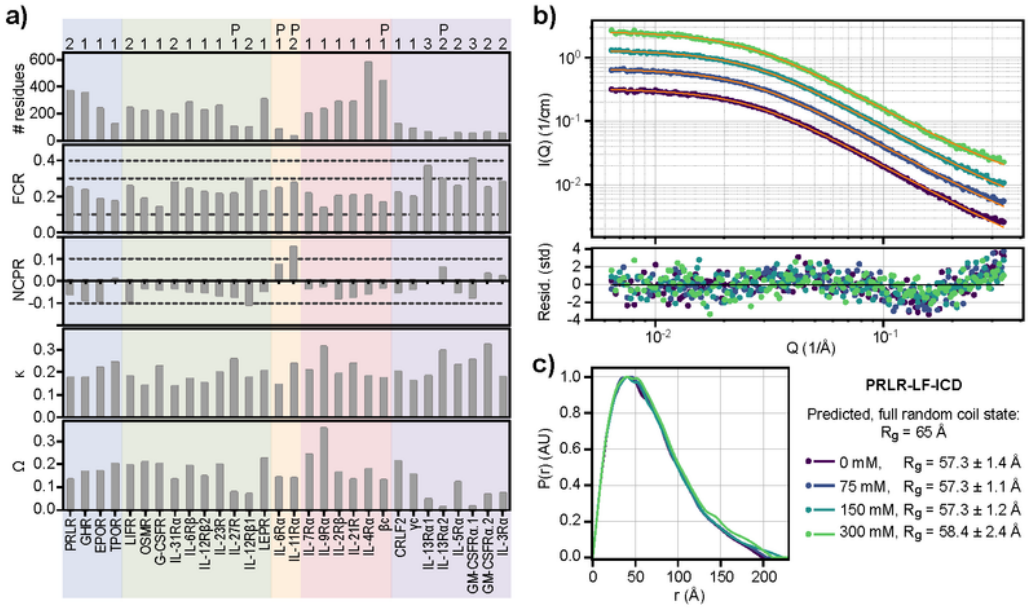

Figure 7

Conformational properties of C1CR-ICDs. a) Plot of conformational parameters provided by CIDER analyses of the longest isoforms of the C1CR-ICDs. For GM-SCFRα, both isoform 1 and 2 were analyzed as they differed > 50%. Each receptor group is highlighted with different colors; 1 – blue, 2 – green, 3 - yellow, 4 – red, and 5 -purple. When sequences have a Pro warning, i.e. a Pro fraction >0.15, the ICD is highlighted with a P on top. The numbers at the top are the compositional region assigned by CIDER. # residues are the numbers of residues in the ICD, FCR is the fraction of charged residues (FCR = f+ + f-), NPCR is the net charge per residue (NPCR = f+ - f-), κ is a measure of the charge mixing across the sequence (normalized between 0 and 1) and Ω additionally takes Pro mixing into account (normalized between 0 and 1). The dotted lines at 0.4, 0.3, and 0.1 in the FCR plot represent the borders for very highly charged (above), highly charged (above) and highly depleted (lower) in charges, respectively [161]. The area between the dotted lines at 0.1 and -0.1 in the NCPR plot represents the range occupied by 70% of human IDPs [161]. b) Experimental SAXS data (spheres) and Gaussian random coil model fit (lines) for PRLR-LF-ICD at different concentrations of NaCl (see color key), with the residuals from the fit shown below in units of standard deviations. c) Pair distance distribution functions for the PRLR-LF-ICD SAXS data. The color coding is the same as in b).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.docx
- Table2.docx
- SeiffertBuggeSupplemental.pdf