Published in final edited form as:

Curr Opin Struct Biol. 2015 June ; 32: 102–112. doi:10.1016/j.sbi.2015.03.008.

Relating sequence encoded information to form and function of intrinsically disordered proteins

Rahul K. Das, Kiersten M. Ruff, and Rohit V. Pappu*

Department of Biomedical Engineering and Center for Biological Systems Engineering,
Washington University in St. Louis, One Brookings Drive, Campus Box 1097, St. Louis, MO
63130, USA

Abstract

Intrinsically disordered proteins (IDPs) showcase the importance of conformational plasticity and heterogeneity in protein function. We summarize recent advances that connect information encoded in IDP sequences to their conformational properties and functions. We focus on insights obtained through a combination of atomistic simulations and biophysical measurements that are synthesized into a coherent framework using polymer physics theories.

Introduction

Protein domains are modular building blocks of macromolecular complexes and interaction networks [1]. The concept of domains can be generalized to include sequence regions that fail to fold as autonomous units [2]. These intrinsically disordered regions / proteins, referred to collectively hereafter as IDPs, are distinct from structured domains. Their sequences encode an intrinsic *inability* to fold into singular well-defined three-dimensional structures [3–7] although some IDPs do fold into well-ordered structures in the context of functional complexes. IDPs are implicated in important cellular processes that include cell division [8,9], cell signaling [3,10], intracellular transport [11,12], bacterial translocation [13], cell mechanics [14,15], protein degradation [16,17], posttranscriptional regulation [18], and cell cycle control [19].

IDPs can be classified into distinct conformational classes based on their amino acid compositions [20–41]. We summarize recent results that have identified composition-to-conformation relationships (CCRs) through studies of archetypal IDPs. CCRs enable the assignments of conformational descriptors and inferences regarding the amplitudes of conformational fluctuations of IDPs. These insights are relevant because amino acid

*Corresponding author: pappu@wustl.edu, Phone: 1-314-935-7958.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of interest

None declared.

compositions are often well conserved among orthologs of IDPs even if their sequences are poorly conserved [42,43].

Compositional classes of IDPs

Amino acid compositions of IDPs are characterized by distinct biases [5]. They are deficient in canonical hydrophobic residues and enriched in polar and charged residues. Accordingly, IDPs fall into three distinct compositional classes that reflect the fraction of charged versus polar residues. The distinct classes are *polar tracts*, *polyampholytes*, and *polyelectrolytes* [41] (see Figure 1). Polar tracts are deficient in charged, hydrophobic, and proline residues. They are enriched in polar amino acids such as Asn, Gly, Gln, His, Ser, and Thr. Polyampholytes and polyelectrolytes can either be weak or strong depending on the fraction of charged residues or FCR that is quantified as the sum of f_+ and f_- (see Figure 2). The latter two parameters quantify the fraction of positive and negatively charged residues in an IDP sequence. Polyelectrolytes have an excess of one type of charge, *i.e.*, $f_+ > f_-$ or vice versa. Polyampholytes have roughly equivalent fractions of opposite charges, *i.e.*, $f_+ \approx f_-$. The designation of weak versus strong polyampholytes / polyelectrolytes is governed by the value of FCR. In strong polyampholytes / polyelectrolytes, the high FCR values encode an intrinsic tendency for populating expanded coil-like conformations because charged residues prefer to be solvated in aqueous milieus.

A formal language for describing conformational preferences of IDPs

Ensembles of conformations as opposed to singular representative structures are appropriate for describing IDPs. The balance between solvent-mediated intra-chain attractions versus repulsions determines the types of conformations that make up the ensemble that is thermodynamically accessible to an IDP sequence. When attractions dominate, the conformations in the ensemble are, on average, compact and spherical, *i.e.*, *globular*. Conversely, if intra-chain repulsions dominate over attractions or, stated differently, chain solvation is preferred over desolvation, then the conformations are, on average, expanded, prolate ellipsoidal, and *coil-like*. An intermediate scenario results if the strengths of intra-chain solvent mediated repulsions are counterbalanced by equivalent attractive interactions. Under such circumstances, the ensembles are characterized by maximal conformational heterogeneity and compact, semi-compact, expanded, and chimeric conformations become thermodynamically accessible [41]. Typical heteropolymeric IDP sequences can sample conformations that are chimeras of globules, coils, rods, and semi-compact hairpins. The preference is governed by the region-specific amino acid compositions along the linear sequence.

Polymer physics theories provide access to formal descriptors of conformational ensembles for heterogeneous systems such as IDPs and these have been reviewed recently [41,44]. Analytical relationships predict the scaling of parameters such as radii of gyration, mean end-to-end distances, and hydrodynamic radii as functions of chain length, amino acid composition, and intrinsic stiffness. Analytical relations are also available to relate the scaling of inter-residue distances to the linear sequence separation between residues [41]. Finally, one can also classify the sequence-specific conformational properties by quantifying

the amplitudes of conformational fluctuations [39]. All of these classifiers and descriptors rely on comparisons of measured or calculated values of conformational fluctuations to expectations from analytical theories for flexible polymers in different types of solvents. Figure 3 summarizes the typical workflow that leads from analysis of results from computer simulations or *in vitro* experiments to quantitative inferences regarding CCRs and / or sequence-toconformation relationships (SCRs).

Distinct compositional classes can be mapped to distinct conformational classes

Results from atomistic simulations obtained using explicit representations of solvent molecules [45,46] and studies based on fluorescence correlation spectroscopy [46,47] have shown that polyglycine chains, *i.e.*, polypeptide backbones sans sidechains, form collapsed globules in aqueous solvents. Dipole-dipole interactions are favored over the solvation of dipoles and this gives rise to the observed preference for globules [48]. In the language of polymer physics, the effective inter-residue interaction coefficient quantifies the energetic balance of chain-chain and chain-solvent interactions. For homopolymers, this coefficient is negative in a poor solvent, zero in an indifferent solvent, and positive in a good solvent [49,50]. The overall implication of the poor solubility and preference of polypeptide backbones for globules is that water is a poor solvent for polypeptide backbones. The intrinsic preference of polypeptide backbones for globules and poor solubility in aqueous solvents is retained for other polyamides such as polyglutamine [51,52] and polar tracts such as glycine-serine copolypeptides [45], and sequences that are enriched in Gln / Asn [53]. Collapsed globules are also preferred for sequences for which FCR < 0.25 and the magnitude of net charge per residue or NCPR (see Figure 3), is less than 0.25 [21,38,40].

Charged sidechains can modulate the intrinsic tendency of polypeptide backbones to form collapsed globules. Essentially, the sidechains can act as modulators of solvent quality thus altering the sign and magnitude of the effective inter-residue interaction coefficient. As the FCR crosses a threshold value, the favorable solvation of charged sidechains combined with electrostatic repulsions in polyelectrolytes and / or the screening of electrostatic repulsions by attractions in certain categories of polyampholytes will result in a preference for either expanded conformations. Sequences with |NCPR| and FCR values larger than a threshold value of 0.25 prefer expanded coil-like structures [21,22,36,38,40]. These inferences have been obtained from a combination of atomistic simulations based on the ABSINTH implicit solvation model and forcefield paradigm [24,54,55], fluorescence correlation spectroscopy [40,47,51], time-resolved fluorescence measurements [53], single molecule Förster resonance energy transfer experiments [20–24], single molecule force spectroscopy [44], pulse field gradient nuclear magnetic resonance experiments [36], measurements of paramagnetic relaxation enhancements [56], and small-angle x-ray scattering measurements [57].

Diagram-of-states summarizing composition-to-conformation relationships

A diagram-of-states summarizes our current understanding of CCRs for IDPs. This diagram is shown in Figure 4. It depicts four distinct conformational classes designated as R1, R2,

R3, and R4, respectively. Polar tracts and weak polyampholytes / polyelectrolytes are globule formers that make up region R1. Strong polyampholytes belong to region R3 and these form either coils or hairpins depending on the combination of FCR values and charge patterning (see below). Sequences from region R2 have intermediate compositional biases and their conformations are likely to be chimeras of globules and coils. IDPs that undergo folding upon binding predominantly populate region R2. This highlights the role played by context dependent interactions as determinants of conformational transitions for sequences drawn from R2 [58]. It is worth emphasizing that the boundary between R1 and R2 is rather *ad hoc*. The placement of this boundary reflects the limited “titration” of CCRs for sequences drawn from these two regions. Region R4 spans two areas, one each for acid versus base rich polyelectrolytes, respectively. For these sequences, the combination of electrostatic repulsions between charged sidechains and the favorable solvation free energies of these sidechains gives rise to semi-flexible worm-like chain conformations.

The diagram-of-states classification shown in Figure 4 is valid for IDP sequences that have at least thirty residues, low overall hydropathy, and low proline contents. The physical principles underlying the conformational properties of weak polyampholytes and polyelectrolytes suggest that the conformational transitions are likely to be continuous functions of FCR and NCPR [59,60]. If this expectation is borne out for longer sequences with low FCR and NCPR values or sequences with equivalent fractions of charged and polar residues, then the composition range spanned by R2 will be larger than what is shown in Figure 4. Unpublished results suggest that the classification of CCRs derived from the diagram-of-states, particularly the assignment of a sequence to region R1 or R2, might only be valid for IDP sequences within a certain length range and proline contents that fall below a reasonably low threshold.

Statistics for different regions of the diagram-of-states

The DISPROT database is an inventory of *bona fide* IDP sequences [61]. Analysis of the compositional biases of sequences from this database reveals that at least 70% of known IDP sequences belong to regions R2 and R3. These sequences are symmetric polyampholytes ($f_+ \approx f_-$), asymmetric polyampholytes ($f_+ \neq f_-$), or weak polyelectrolytes. Based on their compositional biases, sequences corresponding to regions R2 and R3 are expected to adopt coil-like conformations, semi-compact hairpins, or conformations that are chimeras of coils and globules or coils and semi-compact hairpins. In addition, their ensembles are expected to display significant conformational heterogeneity [39] as characterized by spontaneous conformational fluctuations whose amplitudes are likely to be considerably larger than those of globular proteins. Regions R1, R2, and R3 together encompass at least 95% of the known sequences of IDPs.

Connecting CCRs to function

We present highlights from a growing body of data to demonstrate the functional implications of CCRs. The overall theme presented in this discussion is summarized in Figure 5. Long disordered linkers that belong either to the R2 or R3 region of the diagram-of-states can help localize proteins to the junction between the endoplasmic reticulum and

plasma membrane [62]. C-terminal disordered tails of *E.coli* single stranded DNA binding proteins belong to region R1 and these tails engender positive cooperativity in single stranded DNA binding. Cooperativity in single stranded DNA binding is abolished if the tails are eliminated or replaced with sequences drawn from the R3 region [63].

Sterile alpha motifs or SAMs are ubiquitous in eukaryotic proteomes. SAMs are modular 70-residue alpha-helical motifs that have an intrinsic ability to undergo open-ended polymerization and form left-handed helical supramolecular polymers. Among the many functions attributed to SAMs, their polymerization / depolymerization reactions correlate with transcription repression / derepression activities of gene silencing proteins. Polyhomeotic (Ph) is a *Drosophila* protein that is a member of the polycomb group of proteins. These are chromatin-associated gene silencing proteins that epigenetically regulate gene expression. The 88-residue intrinsically disordered linker that is directly N-terminal to the SAM domain hinders open-ended polymerization of Ph. With an FCR of 0.15 and NCPR of -0.08, this linker belongs to region R1 on the diagram-of-states [64]. The human ortholog of Ph is designated as Polyhomeotic homolog 3 or PHC3. The N-terminal intrinsically disordered 84-residue linker of PHC3 also controls the open-ended polymerization of the corresponding SAM domain. With an FCR of 0.38 and NCPR of 0.07, the disordered linker from PHC3 belongs to region R3 on the diagram of states. This alternative linker promotes the open-ended polymerization of PHC3. A chimera of the SAM domain from Ph and the linker from the human ortholog enhances transcriptional repression. Clearly, polymerization requires that linkers tethered to the SAM domain be drawn from region R3 as opposed to R1 [64]. The results also demonstrate the connections between distinct CCRs and different outcomes both in terms of SAM polymerization and the efficiency of transcription repression / derepression.

IDPs can function as entropic bristles and the conformational class that is encoded by the amino acid composition of the IDP governs the properties of brushes or bristles. Investigations to assess the impact of IDP entropic bristles as solubilizing tags have established that sequences of dehydrins, which belong to region R3, are more efficient than sequences drawn from region R1 at solubilizing reporter proteins to which the bristles are tethered [65]. This observation has been rationalized in terms of the increased FCR for optimal solubilizing tags.

The importance of the magnitude of NCPR has been established in the recombination-activation gene or RAG2. The sequence architecture of RAG2 is modular and comprises a 60-residue "acidic hinge" region that connects the beta propeller core domain to a pleckstrin homology domain [66]. The acidic hinge region is important for the function of RAG2, which involves preventing access to inappropriate repair mechanisms for DNA double-stranded breaks such as alternative non-homologous end joining. Key observations regarding the acidic hinge highlight the importance of NCPR over details of the primary sequence. Neutralization of charged residues within the 31-residue N-terminal region of the acidic hinge leads to increased alternative non-homologous end joining whereas scrambling of the sequence that maintains NCPR maintains the functionality of the wild type sequence. Similarly, human sequence variants of RAG2 that lead to changes in NCPR cause increased alternative non-homologous end joining and impaired genome stability [66].

FG nucleoporins or FG-Nups can have distinct compositional biases and these are distinguished by their FCR values. FG-Nups with low FCR values belong to region R1 of the diagram-of-states and these are designated as “cohesive” in contrast to sequences with higher FCR values that belong to regions R2 and R3 and are designated as being “repulsive” [67]. The two categories of sequences are proposed to play distinct roles as modulators of gating mechanisms in the nuclear pore complex.

Going beyond CCRs: Connecting sequence patterns to conformational properties

The diagram-of-states relies purely on the details of amino acid compositions and provides a zeroth order classification of relationships between IDP sequences and conformational classes. The documented CCRs raise an interesting question: Since the number of sequences that are compatible with a given amino acid composition is astronomically large, do all conceivable sequences encode similar conformational properties and impact function in similar ways? Of course, since IDPs serve as scaffolds for short linear motifs or SLiMs [4,68–70], it stands to reason that conserving the identities and positions of SLiMs will winnow down the number of functionally relevant sequence alternatives for a given amino acid composition. Are there other additional constraints that could have a direct impact on global conformational properties and hence on function?

Quantitative studies of DNA binding proteins identified a curious pattern of clustering of like-charged residues [71,72]. Recent systematic studies of charge patterning have revealed the importance of the linear segregation versus mixing of oppositely charged residues as determinants of conformational properties of polyampholytic IDPs [38,73]. The patterning of oppositely charged residues is quantified in terms of a parameter designated as κ (see Figure 6). This parameter is bounded, $0 \leq \kappa \leq 1$, and approaches zero if the oppositely charged residues are well mixed in the linear sequence and approaches unity if the oppositely charged residues are segregated [38]. The number of sequences $n(\kappa)$ that are conceivable for a given value of κ is governed by the combination of FCR and the constraints placed by the presence of conserved SLiMs. In general, $n(\kappa)$ is higher by orders of magnitude for low to intermediate κ values when compared to high κ values. This high sequence entropy provides a default explanation for the observed preponderance of naturally occurring sequences drawn from R2 and R3 for κ values in the range of 0.1 to 0.4 and a depletion of sequences with higher κ values [38]. It is noteworthy that κ also serves as a single parameter surrogate for the strengths of intra-chain electrostatic interactions that determine the overall conformational properties and the amplitudes of conformational fluctuations. Specifically, in sequences with lower κ values, intra-chain electrostatic repulsions are screened by electrostatic attractions and these sequences favor expanded, coil-like ensembles. In contrast, for sequences with higher κ values, intra-chain electrostatic attractions become dominant. In addition to global compaction, locally compact domains can form for sequences with intermediate κ values. Therefore, κ serves as a parameter to rationalize the boundaries between sequences that conserve overall conformational properties – and hence functions and phenotypes – versus sequences that yield altered

conformational ensembles and hence a loss or alteration of functions and phenotypes – see the summary in Figure 7.

Enabling *de novo* sequence design

The connection between a parameter like κ and conformational properties enables the use of *de novo* design as a tool for modulating sequence-to-conformation relationships or SCRs. This should be helpful for establishing the connections between changes to SCRs and functions / phenotypes controlled by polyampholytic sequences drawn from regions R2 and R3. A range of targets for such design efforts is readily available from the rich literature on IDPs with established functional roles for polyampholytic sequences [8,9,14,19,63,74–77]. Of course, the patterning of oppositely charged residues quantified by κ is not the only way to conceive of modulating SCRs. Implicit in the work that uncovered the importance of κ is the idea that changes to SCRs can be realized by changes to sequence patterns that directly modulate the sequence-encoded balance between solvent mediated intra-chain repulsions and attractions. If the underlying energy scales cross some threshold vis-à-vis thermal energy, then we can expect substantial changes to SCRs. Accordingly, the patterning concept can be generalized to consider the patterns of charged versus polar residues or charged versus aromatic residues. The latter might be of particular relevance given growing interest in polycation-pi interactions [78].

Direct impact of sequence patterns on IDP functions

PSC-CTR is the C-Terminal Region of the Posterior Sex Combs subunit of the Polycomb Repressive Complex 1 system in *Drosophila* [79]. These proteins are involved in mediating heritable gene silencing and PSC-CTR is responsible for modulating non-covalent effects on chromatin structure. Specifically, PSC-CTR is essential for the inhibition of chromatin remodeling. The sequences of PSC-CTRs are poorly conserved across orthologs. Systematic feature selection methods combined with DNA binding studies and assays to quantify the repression of chromatin remodeling helped identify sequence patterns that distinguish repressive PSC-CTRs from non-repressive ones. Non-repressive PSC-CTRs are distinguishable by the “*maximum contiguous negative charge*”, which refers to the presence of contiguous stretches with negative NCPR values. *De novo* sequence designs that redistribute the negative charge to lower the linear charge density or eliminate the contiguous stretch of negative charges convert a non-repressive PSC-CTRs to repressive ones. The study of Beh et al. [79] highlights the sequence encoding of the energy scales for electrostatics interactions. It also highlights the need to go beyond single value descriptors of sequence patterning such as κ . Instead, the vectorial NCPR profile across the length of the sequence (see Figure 7) is likely to be more informative for identifying local clusters of charge that are directly relevant for controlling functions. There is also a case to be made for going beyond the identification of conserved SLiMs to include the presence of clusters of like charges in functional annotations of IDPs. Such clusters might contribute either to attractive or repulsive long-range interactions that engender specificity of functions through disordered regions.

Conclusions

We have summarized recent insights that help connect the information encoded in IDP sequences to conformational properties and functions. Efforts to uncover synergies among CCRs, SCRs, and SLiMs [69] as determinants of conformational properties and functions of IDPs both *in vitro* and *in vivo* are just burgeoning and several questions remain open for investigation especially with regard to the *in vivo* implications of CCRs and SCRs. The impact of chain length on CCRs and SCRs remains unexplored. Many IDP sequences have high proline contents and a systematic investigation of this feature is warranted. It is conceivable that different polar sidechains will have differential effects on the conformational properties and solubility profiles of IDPs, *i.e.*, there is good reason to conjecture that Ser-rich sequences might behave differently than Gln-rich sequences. This conjecture has merits given published accounts of differences between Gln versus Asn rich disordered regions [80]. Targets for alternative splicing are enriched in transcripts for IDPs [18]. This opens the door to the possibility that posttranscriptional processing provides a route to regulate CCRs and SCRs for tissue-specific control and rewiring of protein interaction networks. Many of the common cellular posttranslational modifications involve either addition (Ser / Thr / Tyr phosphorylation, Gln / Asn deamidation, Tyr, Trp, or hydroxy amino acid sulfonation, and Tyr nitration) or neutralization of charges (Lys acetylation, Glu / Asp amidation, and Arg citrullination). N- and O-linked glycosylation can either add or neutralize charge depending on the sugar being added. These post-translational modifications can lead to a change in conformational class. They can also influence the sequence patterning of oppositely charged residues or the linear charge density within contiguous stretches of like charges. Therefore, altered sequence patterns within IDPs and their functional consequences are likely to be an emergent property of posttranslational modifications. Finally, the connection between the time scales for inter-conversions between distinct conformations and equilibrium descriptions of CCRs and SCRs remains under explored. Preliminary work has focused on the impact of sequence-specific contributions to internal friction [20,81–84]. Advances in nuclear magnetic resonance [85–89] and single molecule spectroscopies [90–92] combined with novel computational and theoretical methodologies [93–95] should pave the way for comprehensive characterization of IDP dynamics and assessing their impact on the dynamical regulation of cellular phenotypes [96,97]. Overall, it is clear that continued synergistic investigations must be brought to bear in order to build on the insights that have been forthcoming with regard to connecting information encoded in IDP sequences to their form and function.

Acknowledgments

We are grateful to M. Madan Babu, Martin Blackledge, Doug Barrick, Ashok Deniz, Julie Forman-Kay, Tyler Harmon, Alex Holehouse, Richard Kriwacki, Petra Levin, Timothy Lohman, Tanja Mittag, Anuradha Mittal, Michael Rosen, Benjamin Schuler, and Andrea Soranno for many insightful discussions over the past two years. This work was supported by grants from the US National Science Foundation (MCB 1121867) and US National Institutes of Health (5R01NS056114).

References and recommended reading

Papers of particular interest, published within the period of review have been highlighted as:

- of special interest
- of outstanding interest

1. Chothia C, Gough J, Vogel C, Teichmann SA. Evolution of the Protein Repertoire. *Science*. 2003; 300:1701–1703. [PubMed: 12805536]
2. Babu MM, Kriwacki RW, Pappu RV. Versatility from Protein Disorder. *Science*. 2012; 337:1460–1461. [PubMed: 22997313]
3. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews: Molecular Cell Biology*. 2015; 16:18–29. [PubMed: 25531225] . ● An updated account of the importance of IDPs in cell signaling and the control of cellular decisions and fates.
4. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, et al. Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews*. 2014; 114:6589–6631. [PubMed: 24773235] . ● A comprehensive review of informatics and physical considerations that have enabled the classification of motifs and IDPs.
5. Uversky VN. Natively unfolded proteins: A point where biology waits for physics. *Protein Science*. 2002; 11:739–756. [PubMed: 11910019]
6. Uversky VN. A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Science*. 2013; 22:693–724. [PubMed: 23553817]
7. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovi Z. Intrinsic Disorder and Protein Function. *Biochemistry*. 2002; 41:6573–6582. [PubMed: 12022860]
8. Buske PJ, Levin PA. Extreme C Terminus of Bacterial Cytoskeletal Protein FtsZ Plays Fundamental Role in Assembly Independent of Modulatory Proteins. *Journal of Biological Chemistry*. 2012; 287:10945–10957. [PubMed: 22298780]
9. Buske PJ, Levin PA. A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo. *Molecular Microbiology*. 2013; 89:249–263. [PubMed: 23692518] . ● This paper demonstrates the central importance of the disordered C-terminal linker in forming Z-rings that mediate bacterial cell division. The linker is a polyampholyte and it connects the core domain of FtsZ, which is a tubulin homolog, to the conserved C-terminal SLiM that mediates interactions with the network of FtsZ binding proteins.
10. Tantos A, Han KH, Tompa P. Intrinsic disorder in cell signaling and gene transcription. *Molecular and Cellular Endocrinology*. 2012; 348:457–465. [PubMed: 21782886]
11. Meinema AC, Laba JK, Hapsari RA, Otten R, Mulder FAA, Kralt A, van den Bogaart G, Lusk CP, Poolman B, Veenhoff LM. Long Unfolded Linkers Facilitate Membrane Protein Import Through the Nuclear Pore Complex. *Science*. 2011; 333:90–93. [PubMed: 21659568]
12. Meinema AC, Poolman B, Veenhoff LM. Quantitative Analysis of Membrane Protein Transport Across the Nuclear Pore Complex. *Traffic*. 2013; 14:487–501. [PubMed: 23357007] . ● An elegant study showing the importance of amino acid composition as a determinant of IDP function. The authors demonstrate that composition controls the Stokes radius of disordered linkers that play a role in facilitating nuclear import of substrates.
13. Housden NG, Hopper JTS, Lukyanova N, Rodriguez-Larrea D, Wojdyla JA, Klein A, Kaminska R, Bayley H, Saibil HR, Robinson CV, et al. Intrinsically Disordered Protein Threads Through the Bacterial Outer-Membrane Porin OmpF. *Science*. 2013; 340:1570–1574. [PubMed: 23812713] . ●● A multi-pronged investigation that highlights the role of the unstructured N-terminal domain of colicin E9 in mediating the formation of bacterial translocons. The IDR threads through two of the pores of the trimeric OmpF porin and does so spontaneously in a manner that remains unexplained. The sequence of the unstructured IDR belongs to the R1 region of the diagram-of-states raising intriguing questions about mechanisms.
14. Srinivasan N, Bhagawati M, Ananthanarayanan B, Kumar S. Stimuli-sensitive intrinsically disordered protein brushes. *Nature Communications*. 2014; 5:5145.. ● This study demonstrates the feasibility of using IDPs in engineering applications. The authors use sequences that mimic those of the the polyampholytic heavy subunits of neurofilament sidearms and graft these to surfaces to generate polymer brushes. They show that these well-mixed polyampholytes characterized by low κ values undergo dramatic conformational transitions in response to changes

in pH and solution conditions. This clearly highlights the possibility that solution conditions might have an impact that mimics the effect of increasing κ .

15. Guharoy M, Szabo B, Martos SC, Kosol S, Tompa P. Intrinsic Structural Disorder in Cytoskeletal Proteins. *Cytoskeleton*. 2013; 70:550–571. [PubMed: 23761374]
16. van der Lee R, Lang B, Kruse K, Gsponer J, de Groot NS, Huynen MA, Matouschek A, Fuxreiter M, Babu MM. Intrinsically Disordered Segments Affect Protein Half-Life in the Cell and during Evolution. *Cell Reports*. 2014; 8:1832–1844. [PubMed: 25220455]
17. Gsponer J, Futschik ME, Teichmann SA, Babu MM. Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation. *Science*. 2008; 322:1365–1368. [PubMed: 19039133]
18. Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, Bateman A, Babu MM. Tissue-Specific Splicing of Disordered Segments that Embed Binding Motifs Rewires Protein Interaction Networks. *Molecular Cell*. 2012; 46:871–883. [PubMed: 22749400]
19. Wang YF, Fisher JC, Mathew R, Ou L, Otieno S, Sublet J, Xiao LM, Chen JH, Roussel MF, Kriwacki RW. Intrinsic disorder mediates the diverse regulatory functions of the Cdk inhibitor p21. *Nature Chemical Biology*. 2011; 7:214–221. [PubMed: 21358637]
20. Borgia A, Wensley BG, Soranno A, Nettels D, Borgia MB, Hoffmann A, Pfeil SH, Lipman EA, Clarke J, Schuler B. Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nature Communications*. 2012; 3:1195.
21. Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, Schuler B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with singlemolecule spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:16155–16160. [PubMed: 22984159]. ●● An important paper that quantifies the impact of charge on the dimensions of archetypal IDPs and unfolded proteins in aqueous solutions. This work helps in identifying the connection between amino acid composition and the effective theta point for different protein sequences. The combination of innovative single molecule measurements, its comprehensive nature, and the use of updated adaptations of polymer physics theories make this a very important paper.
22. Mueller-Spaeth S, Soranno A, Hirschfeld V, Hofmann H, Rueegger S, Reymond L, Nettels D, Schuler B. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:14609–14614. [PubMed: 20639465]
23. Soranno A, Koenig I, Borgia MB, Hofmann H, Zosel F, Nettels D, Schuler B. Singlemolecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:4874–4879. [PubMed: 24639500]
24. Wuttke R, Hofmann H, Nettels D, Borgia MB, Mittal J, Best RB, Schuler B. Temperature-dependent solvation modulates the dimensions of disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:5213–5218. [PubMed: 24706910]
25. Guerry P, Mollica L, Blackledge M. Mapping Protein Conformational Energy Landscapes Using NMR and Molecular Simulation. *Chemphyschem*. 2013; 14:3046–3058. [PubMed: 23703956]
26. Jensen MR, Blackledge M. Testing the validity of ensemble descriptions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:E1557–E1558. [PubMed: 24639541]
27. Jensen MR, Ruigrok RWH, Blackledge M. Describing intrinsically disordered proteins at atomic resolution by NMR. *Current Opinion in Structural Biology*. 2013; 23:426–435. [PubMed: 23545493]
28. Ozenne V, Schneider R, Yao M, Huang J-r, Salmon L, Zweckstetter M, Jensen MR, Blackledge M. Mapping the Potential Energy Landscape of Intrinsically Disordered Proteins at Amino Acid Resolution. *Journal of the American Chemical Society*. 2012; 134:15138–15148. [PubMed: 22901047]
29. Parigi G, Rezaei-Ghaleh N, Giachetti A, Becker S, Fernandez C, Blackledge M, Griesinger C, Zweckstetter M, Luchinat C. Long-Range Correlated Dynamics in Intrinsically Disordered

- Proteins. *Journal of the American Chemical Society*. 2014; 136:16201–16209. [PubMed: 25331250]
30. Schwalbe M, Ozenne V, Bibow S, Jaremko M, Jaremko L, Gajda M, Jensen MR, Biernat J, Becker S, Mandelkow E, et al. Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and alpha-Synuclein in Solution from NMR and Small Angle Scattering. *Structure*. 2014; 22:238–249. [PubMed: 24361273]
 31. Jain N, Bhattacharya M, Mukhopadhyay S. Chain Collapse of an Amyloidogenic Intrinsically Disordered Protein. *Biophysical Journal*. 2011; 101:1720–1729. [PubMed: 21961598]
 32. Forman-Kay JD, Mittag T. From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins. *Structure*. 2013; 21:1492–1499. [PubMed: 24010708]
 33. Krzeminski M, Marsh JA, Neale C, Choy W-Y, Forman-Kay JD. Characterization of disordered proteins with ENSEMBLE. *Bioinformatics*. 2013; 29:398–399. [PubMed: 23233655]
 34. Liu B, Chia D, Csizmek V, Farber P, Forman-Kay JD, Gradinaru CC. The Effect of Intrachain Electrostatic Repulsion on Conformational Disorder and Dynamics of the Sic1 Protein. *Journal of Physical Chemistry B*. 2014; 118:4088–4097.
 35. Marsh JA, Dancheck B, Ragusa MJ, Allaire M, Forman-Kay JD, Peti W. Structural Diversity in Free and Bound States of Intrinsically Disordered Protein Phosphatase 1 Regulators. *Structure*. 2010; 18:1094–1103. [PubMed: 20826336]
 36. Marsh JA, Forman-Kay JD. Sequence Determinants of Compaction in Intrinsically Disordered Proteins. *Biophysical Journal*. 2010; 98:2383–2390. [PubMed: 20483348]
 37. Marsh JA, Forman-Kay JD. Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. *Proteins-Structure Function and Bioinformatics*. 2012; 80:556–572.
 38. Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:13392–13397. [PubMed: 23901099] . ●●
This paper that introduces the importance of charge patterning as a determinant of IDP conformations. It also introduces the diagram-of-states classification that forms the focus of the current review.
 39. Lyle N, Das RK, Pappu RV. A quantitative measure for protein conformational heterogeneity. *Journal of Chemical Physics*. 2013; 139:121907. [PubMed: 24089719]
 40. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:8183–8188. [PubMed: 20404210]
 41. Mao AH, Lyle N, Pappu RV. Describing sequence-ensemble relationships for intrinsically disordered proteins. *Biochemical Journal*. 2013; 449:307–318. [PubMed: 23240611]
 42. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. *Current Opinion in Structural Biology*. 2011; 21:441–446. [PubMed: 21482101]
 43. Moesa HA, Wakabayashi S, Nakai K, Patil A. Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification. *Molecular BioSystems*. 2012; 8:3262–3273. [PubMed: 23076520]
 44. Brucule M, Schuler B, Samori B. Single-Molecule Studies of Intrinsically Disordered Proteins. *Chemical Reviews*. 2014; 114:3281–3317. [PubMed: 24432838]
 45. Tran HT, Mao A, Pappu RV. Role of backbone - Solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *Journal of the American Chemical Society*. 2008; 130:7380–7392. [PubMed: 18481860]
 46. Holehouse AS, Garai K, Lyle N, Vitalis A, Pappu RV. Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation. *Journal of the American Chemical Society*. 2015; 137:2984–2995. [PubMed: 25664638]
 47. Teufel DP, Johnson CM, Lum JK, Neuweiler H. Backbone-Driven Collapse in Unfolded Protein Chains. *Journal of Molecular Biology*. 2011; 409:250–262. [PubMed: 21497607]

48. Karandur D, Wong K-Y, Pettitt BM. Solubility and Aggregation of Gly5 in Water. *The Journal of Physical Chemistry B*. 2014; 118:9565–9572. [PubMed: 25019618]
49. Rubinstein, M.; Colby, RH. *Polymer Physics*. Oxford and New York: Oxford University Press; 2003.
50. Sanchez IC. Phase Transition Behavior of the Isolated Polymer Chain. *Macromolecules*. 1979; 12:980–988.
51. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:16764–16769. [PubMed: 17075061]
52. Crick SL, Ruff KM, Garai K, Frieden C, Pappu RV. Unmasking the roles of N- and C-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:20075–20080. [PubMed: 24282292]
53. Mukhopadhyay S, Krishnan R, Lemke EA, Lindquist S, Deniz AA. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104:2649–2654. [PubMed: 17299036]
54. Vitalis A, Pappu RV. ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions. *Journal of Computational Chemistry*. 2009; 30:673–699. [PubMed: 18506808]
55. Radhakrishnan A, Vitalis A, Mao AH, Steffen AT, Pappu RV. Improved Atomistic Monte Carlo Simulations Demonstrate That Poly-L-Proline Adopts Heterogeneous Ensembles of Conformations of Semi-Rigid Segments Interrupted by Kinks. *Journal of Physical Chemistry B*. 2012; 116:6862–6871.
56. Xue Y, Skrynnikov NR. Motion of a Disordered Polypeptide Chain as Studied by Paramagnetic Relaxation Enhancements, ¹⁵N Relaxation, and Molecular Dynamics Simulations: How Fast Is Segmental Diffusion in Denatured Ubiquitin? *Journal of the American Chemical Society*. 2011; 133:14614–14628. [PubMed: 21819149]
57. Bernado P, Svergun DI. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Molecular BioSystems*. 2012; 8:151–167. [PubMed: 21947276]
58. Mittal A, Lyle N, Harmon TS, Pappu RV. Hamiltonian Switch Metropolis Monte Carlo Simulations for Improved Conformational Sampling of Intrinsically Disordered Regions Tethered to Ordered Domains of Proteins. *Journal of Chemical Theory and Computation*. 2014; 10:3550–3562. [PubMed: 25136274]
59. Dobrynin AV, Colby RH, Rubinstein M. Scaling theory of polyelectrolyte solutions. *Macromolecules*. 1995; 28:1859–1871.
60. Dobrynin AV, Rubinstein M. Flory theory of a polyampholyte chain. *Journal de Physique II*. 1995; 5:677–695.
61. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, et al. DisProt: the database of disordered proteins. *Nucleic Acids Research*. 2007; 35:D786–D793. [PubMed: 17145717]
62. Kralt A, Carretta M, Mari M, Reggiori F, Steen A, Poolman B, Veenhoff LM. Intrinsically Disordered Linker and Plasma Membrane-Binding Motif Sort Ist2 and Ssy1 to Junctions. *Traffic*. 2015; 16:135–147. [PubMed: 25409870]
63. Kozlov AG, Weiland E, Mittal A, Waldman V, Antony E, Fazio N, Pappu RV, Lohman TM. Intrinsically Disordered C-Terminal Tails of E. coli Single-Stranded DNA Binding Protein Regulate Cooperative Binding to Single-Stranded DNA. *Journal of Molecular Biology*. 2015; 427:763–774. [PubMed: 25562210]
64. Robinson AK, Leal BZ, Nanyes DR, Kaur Y, Ilangoan U, Schirf V, Hinck AP, Demeler B, Kim CA. Human Polyhomeotic Homolog 3 (PHC3) Sterile Alpha Motif (SAM) Linker Allows Open-Ended Polymerization of PHC3 SAM. *Biochemistry*. 2012; 51:5379–5386. [PubMed: 22724443]

65. Santner AA, Croy CH, Vasanwala FH, Uversky VN, Van Y-YJ, Dunker AK. Sweeping Away Protein Aggregation with Entropic Bristles: Intrinsically Disordered Protein Fusions Enhance Soluble Expression. *Biochemistry*. 2012; 51:7250–7262. [PubMed: 22924672]
66. Coussens MA, Wendland RL, Deriano L, Lindsay CR, Arnal SM, Roth DB. RAG2's Acidic Hinge Restricts Repair-Pathway Choice and Promotes Genomic Stability. *Cell Reports*. 2013; 4:870–878. [PubMed: 23994475]
67. Yamada J, Phillips JL, Patel S, Goldfien G, Calestagne-Morelli A, Huang H, Reza R, Acheson J, Krishnan VV, Newsam S, et al. A Bimodal Distribution of Two Distinct Categories of Intrinsically Disordered Structures with Separate Functions in FG Nucleoporins. *Molecular & Cellular Proteomics*. 2010; 9:2205–2224. [PubMed: 20368288]
68. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H, Gibson TJ. Attributes of short linear motifs. *Molecular Biosystems*. 2012; 8:268–281. [PubMed: 21909575]
69. Tompa P, Davey NE, Gibson TJ, Babu MM. A Million Peptide Motifs for the Molecular Biologist. *Molecular Cell*. 2014; 55:161–169. [PubMed: 25038412]
70. Ba ANN, Yeh BJ, van Dyk D, Davidson AR, Andrews BJ, Weiss EL, Moses AM. Proteome-Wide Discovery of Evolutionary Conserved Sequences in Disordered Regions. *Science Signaling*. 2012; 5
71. Potoyan DA, Papoian GA. Energy Landscape Analyses of Disordered Histone Tails Reveal Special Organization of Their Conformational Dynamics. *Journal of the American Chemical Society*. 2011; 133:7405–7415. [PubMed: 21517079]
72. Vuzman D, Levy Y. DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:21004–21009. [PubMed: 21078959]
73. Srivastava D, Muthukumar M. Sequence Dependence of Conformations of Polyampholytes. *Macromolecules*. 1996; 29:2324–2326.
74. Mitrea DM, Yoon MK, Ou L, Kriwacki RW. Disorder-function relationships for the cell cycle regulatory proteins p21 and p27. *Biological Chemistry*. 2012; 393:259–274. [PubMed: 23029651]
75. Bertagna A, Toptygin D, Brand L, Barrick D. The effects of conformational heterogeneity on the binding of the Notch intracellular domain to effector proteins: a case of biologically tuned disorder. *Biochemical Society Transactions*. 2008; 36:157–166. [PubMed: 18363556]
76. Johnson SE, Barrick D. Dissecting and Circumventing the Requirement for RAM in CSL-Dependent Notch Signaling. *Plos One*. 2012; 7:e39093. [PubMed: 22876274]
77. Lai J, Koh CH, Tjota M, Pieuchot L, Raman V, Chandrababu KB, Yang D, Wong L, Jedd G. Intrinsically disordered proteins aggregate at fungal cell-to-cell channels and regulate intercellular connectivity. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:15781–15786. [PubMed: 22955885]
78. Song J, Ng SC, Tompa P, Lee KAW, Chan HS. Polycation- π Interactions Are a Driving Force for Molecular Recognition by an Intrinsically Disordered Oncoprotein Family. *Plos Computational Biology*. 2013; 9:e1003239. [PubMed: 24086122]
79. Beh LY, Colwell LJ, Francis NJ. A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:E1063–E1071. [PubMed: 22517748] . ●● This paper captures the essence of the connections between sequence patterning and IDP functions. The focus on the evolution of coarse grain sequence patterns that defy ready recognition by naïve sequence comparisons makes this a very appealing read.
80. Halfmann R, Alberti S, Krishnan R, Lyle N, O'Donnell CW, King OD, Berger B, Pappu RV, Lindquist S. Opposing Effects of Glutamine and Asparagine Govern Prion Formation by Intrinsically Disordered Proteins. *Molecular Cell*. 2011; 43:72–84. [PubMed: 21726811]
81. Schulz JCF, Schmidt L, Best RB, Dzubiella J, Netz RR. Peptide Chain Dynamics in Light and Heavy Water: Zooming in on Internal Friction. *Journal of the American Chemical Society*. 2012; 134:6273–6279. [PubMed: 22414068]
82. Soranno A, Buchli B, Nettels D, Cheng RR, Mueller-Spaeth S, Pfeil SH, Hoffmann A, Lipman EA, Makarov DE, Schuler B. Quantifying internal friction in unfolded and intrinsically disordered

- proteins with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:17800–17806. [PubMed: 22492978]
83. de Sancho D, Sirur A, Best RB. Molecular origins of internal friction effects on protein-folding rates. *Nature Communications*. 2014; 5:4307.
84. Echeverria I, Makarov DE, Papoian GA. Concerted Dihedral Rotations Give Rise to Internal Friction in Unfolded Proteins. *Journal of the American Chemical Society*. 2014; 136:8708–8713. [PubMed: 24844314]
85. Silvers R, Sziegat F, Tachibana H, Segawa S-i, Whittaker S, Guenther UL, Gabel F, Huang J-r, Blackledge M, Wirmer-Bartoschek J, et al. Modulation of Structure and Dynamics by Disulfide Bond Formation in Unfolded States. *Journal of the American Chemical Society*. 2012; 134:6846–6854. [PubMed: 22414027]
86. Guerry P, Schneider R, Huang JR, Delaforge E, Maurin D, Ozenne V, Communie G, Mollica L, Jensen M, Blackledge M. Protein conformational dynamics and molecular recognition in folded and unfolded proteins by NMR. *European Biophysics Journal with Biophysics Letters*. 2013; 42:S61–S61.
87. Markwick PRL, Bouvignies G, Salmon L, McCammon JA, Nilges M, Blackledge M. Toward a Unified Representation of Protein Structural Dynamics in Solution. *Journal of the American Chemical Society*. 2009; 131:16968–16975. [PubMed: 19919148]
88. Mittag T, Kay LE, Forman-Kay JD. Protein dynamics and conformational disorder in molecular recognition. *Journal of Molecular Recognition*. 2010; 23:105–116. [PubMed: 19585546]
89. Andresen C, Helander S, Lemak A, Fares C, Csizmok V, Carlsson J, Penn LZ, Forman-Kay JD, Arrowsmith CH, Lundstrom P, et al. Transient structure and dynamics in the disordered c-Myc transactivation domain affect Bin1 binding. *Nucleic Acids Research*. 2012; 40:6353–6366. [PubMed: 22457068]
90. Polinkovsky ME, Gambin Y, Banerjee PR, Erickstad MJ, Groisman A, Deniz AA. Ultrafast cooling reveals microsecond-scale biomolecular dynamics. *Nature Communications*. 2014; 5:5737.
91. Kalinin S, Peulen T, Sindbert S, Rothwell PJ, Berger S, Restle T, Goody RS, Gohlke H, Seidel CAM. A toolkit and benchmark study for FRET-restrained highprecision structural modeling. *Nature Methods*. 2012; 9 1218-U1129.
92. Olofsson L, Felekyan S, Doumazane E, Scholler P, Fabre L, Zwier JM, Rondard P, Seidel CAM, Pin J-P, Margeat E. Fine tuning of sub-millisecond conformational dynamics controls metabotropic glutamate receptors agonist efficacy. *Nature Communications*. 2014; 5:5206.
93. Bolhuis PG, Chandler D, Dellago C, Geissler PL. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual Review of Physical Chemistry*. 2002; 53:291–318.
94. Borrero EE, Dellago C. Overcoming barriers in trajectory space: Mechanism and kinetics of rare events via Wang-Landau enhanced transition path sampling. *Journal of Chemical Physics*. 2010; 133:134112. [PubMed: 20942528]
95. Juraszek J, Vreede J, Bolhuis PG. Transition path sampling of protein conformational changes. *Chemical Physics*. 2012; 396:30–44.
96. Borchers W, Theillet F-X, Katzer A, Finzel A, Mishall KM, Powell AT, Wu H, Manieri W, Dieterich C, Selenko P, et al. Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nature Chemical Biology*. 2014; 10:1000–1002. [PubMed: 25362358]
97. Ferreon ACM, Ferreon JC, Wright PE, Deniz AA. Modulation of allostery by protein intrinsic disorder. *Nature*. 2013; 498:390–394. [PubMed: 23783631]

Highlights

- Compositions of IDPs fall into distinct classes
- Compositional classes encode preferences for distinct conformational classes
- Functions of disordered proteins are governed by composition-to-conformation relationships
- Sequence patterning of oppositely charged residues can alter conformational properties
- Charge patterning can directly impact IDP functions

Figure 2. Summary of the typical workflow used to extract quantitative CCRs and SCRs from computer simulations, *in vitro* biophysical experiments, or synergy between the two modes of investigation

N : Number of residues in the sequence

N_+, N_- : Number of positive, negatively charged residues

$$f_+ = \frac{N_+}{N}, f_- = \frac{N_-}{N}$$

$$\text{FCR} = (f_+ + f_-)$$

$$\text{NCPR} = (f_+ - f_-)$$

$$\text{Charge asymmetry } \sigma = \frac{(f_+ - f_-)^2}{(f_+ + f_-)}$$

Figure 3. Summary of readily calculated compositional parameters that help in quantitative assessments of CCRs for IDP sequences

Figure 4. Diagram-of-states classification depicting the distinct conformational classes for IDP sequences

Statistics for different regions (percentages) are from analysis of *bona fide* IDPs in DISPROT [61].

Figure 5. Illustrations of the impact of conserved versus altered CCRs on IDP functions

$$\delta_{\text{seq}} = \frac{\sum_{i=1}^{n_w} (\sigma_i - \sigma)^2}{n_w}; \kappa = \left(\frac{\delta_{\text{seq}}}{\delta_{\text{max}}} \right)$$

wt	WPPDRGHDKSDRDRERGYDKVDRERERDRERDRDRGYDKADREEGKERRHHRREE	$\kappa = 0.02$
sv1	WPPYDDRSRHERRHKYRRRRARKRHKGDRREEGEEDVEDEDGDRRRRRKDDDEGE	$\kappa = 0.43$
sv2	WPPGGEDDEDDDEEDDEGEDEDEDEAHYYKSHGRRRRKKRRKRRRHRRRRRV	$\kappa = 0.91$

Figure 6. Calculation of κ and using it to distinguish the sequences with different linear patterns of oppositely charged residues

The top row shows how κ is calculated. The overall charge asymmetry σ is determined by the amino acid composition (see Figure 2). Each sequence is divided into n_w sliding windows and the mean squared deviation δ helps quantify the deviation of the charge asymmetry across different sequence windows vis-à-vis the charge asymmetry encoded by the amino acid composition. The value of δ is calculated for all sequence variants that correspond to the amino acid composition and this is used to evaluate the value of κ , as shown, thus ensuring that κ is bounded between 0 and 1. As an illustration of the patterning that is quantified using κ , we show the sequence of the “polar rich domain” extracted from the sequence of the polyglutamine tract binding protein PQB-P1. The bottom two rows show two *de novo* designed sequences designated as sv1 and sv2 for sequence variants 1 and 2. These two sequences were derived from alterations to the linear sequence distribution of oppositely charged residues [38]. On each row, the values of κ are shown to the right.

Figure 7. Illustrating the impact of sequence patterns and their conservation / alteration on IDP functions