Biophysical *Journal*
Computational Tool

Biophysical Society

CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins

Alex S. Holehouse,[1,*] Rahul K. Das,[1] James N. Ahad,[1] Mary O. G. Richardson,[1] and Rohit V. Pappu[1,*]
[1]Department of Biomedical Engineering and Center for Biological Systems Engineering, Washington University in St. Louis, St. Louis, Missouri

ABSTRACT Intrinsically disordered proteins and regions (IDPs) represent a large class of proteins that are defined by conformational heterogeneity and lack of persistent tertiary/secondary structure. IDPs play important roles in a range of biological functions, and their dysregulation is central to numerous diseases, including neurodegeneration and cancer. The conformational ensembles of IDPs are encoded by their amino acid sequences. Here, we present two computational tools that are designed to enable rapid and high-throughput analyses of a wide range of physicochemical properties encoded by IDP sequences. The first, CIDER, is a user-friendly webserver that enables rapid analysis of IDP sequences. The second, localCIDER, is a high-performance software package that enables a wide range of analyses relevant to IDP sequences. In addition to introducing the two packages, we demonstrate the utility of these resources using examples where sequence analysis offers biophysical insights.

Intrinsically disordered proteins and regions (collectively referred to as IDPs hereafter) make up ~30% of eukaryotic proteomes (1). They are associated with a variety of functions, including transcriptional regulation, cell signaling (2), chaperone activity (3), regulation of bacterial homeostasis and lifecycles, viral infectivity, and subcellular organization in eukaryotic cells (4). IDPs are also associated with a wide range of diseases, including neurodegeneration and cancer (5). Sequence-encoded conformational heterogeneity is a defining feature of IDPs. Properties of conformational ensembles are quantified in terms of average size, shape, local secondary structural preference, pattern of inter-residue distances, and amplitude of conformational fluctuations. Heuristics extracted from biophysical studies can be used to classify sequence-ensemble relationships of IDPs. These relationships are governed by the amino acid compositions and sequence patterns within IDPs. Recent studies have shown that sequence-ensemble relationships of IDPs contribute directly to their biological functions (6,7).

IDP sequences show poor conservation across orthologs (8). However, there is growing evidence that coarse-grained sequence features are well conserved in IDPs. These coarse-grained sequence properties, which can be readily deduced through analysis of primary sequences, determine the conformational properties of IDPs. The precise sequence-ensemble relationship is governed by their amino acid compositions and sequence patterns (7). Sequences encode the patterns of long-range interactions (9), secondary structural preferences (10), and fluctuations about well-defined conformational elements that characterize IDP ensembles. Accordingly, the ensembles of many IDPs can be partitioned into distinct conformational classes, and the relationships between sequence and conformational classes can be identified using a set of quantitative heuristics that are derived from amino acid sequences (7). The volume of sequence information is growing exponentially, and hence, it should be possible to uncover the evolution of sequence-ensemble-function relationships across disordered proteomes.

Low overall hydrophobicity is a defining feature of many IDP sequences. In a two-parameter space defined by the mean hydrophobicity, H, and mean net charge, q, Uversky et al. argued that a single empirical line delineates putative IDPs and autonomously foldable proteins (11). Studies focused on sequences that lie on the IDP side of this empirical line showed that there are distinct sub-classes among IDPs themselves. For example, the net charge per residue (NCPR) of an IDP contributes directly as a determinant of overall global dimensions (12–14). Polyelectrolytic IDPs with NCPR below a threshold value of 0.25 adopt compact globular ensembles, whereas sequences that lie above this threshold adopt well-solvated expanded coils and even stiff rod-like conformations. The degree of conformational heterogeneity within IDP ensembles can be decoupled from the overall size, shape, and local conformational preferences.

Submitted October 18, 2016, and accepted for publication November 29, 2016.

*Correspondence: alex.holehouse@wustl.edu or pappu@wustl.edu

Editor: David Eliezer.

http://dx.doi.org/10.1016/j.bpj.2016.11.3200

For example, sequences that predominantly favor collapsed globules can sample vastly different globular conformations and have higher conformational heterogeneity than highly charged polyelectrolytes that sample predominantly rod-like conformations (15). The importance of charged residues as one of the main determinants of conformational properties of IDPs was further underscored in work that showed that the fraction of charged residues (FCR) and the linear patterning of positively charged and negatively charged residues contribute directly to the size, shape, and amplitude of conformational fluctuations of polyampholytic IDPs (16).

Using the fractions of positively charged and negatively charged residues, viz., f_+ and f_-, respectively, IDP sequences can be partitioned into one of five different conformational classes. This predictive, albeit heuristic diagram of states, shown in Fig. 1 *a*, provides a simple way to classify IDPs and generate expectations regarding conformational properties (7,16). Assuming fixed charge states, IDP sequences of low overall hydrophobicity and low overall proline content (<15%) can be partitioned into one of five classes, viz., R1–R5. Additionally, the sequence patterning of oppositely charged residues contributes directly to the global compaction or expansion of IDPs, and this patterning is quantified by a parameter κ (16). Here, $0 \leq \kappa \leq 1$; low values of κ correspond to sequences—for a fixed amino acid composition—wherein the oppositely charged residues are well-mixed within the linear sequence. In contrast, large values of κ correspond to sequences where the oppositely charged residues are segregated into blocks of like charge (see Fig. 1 *b*). In addition to charged residues, the fraction of proline residues and the intrinsic propensities of individual residues to adopt polyproline II (PPII) conformations are thought to play an important role in driving local conformational transitions and global compaction/expansion (12,17). Finally, recent studies have focused on the sequence complexity of IDPs due to growing interest in drivers of the formation of membraneless organelles (4).

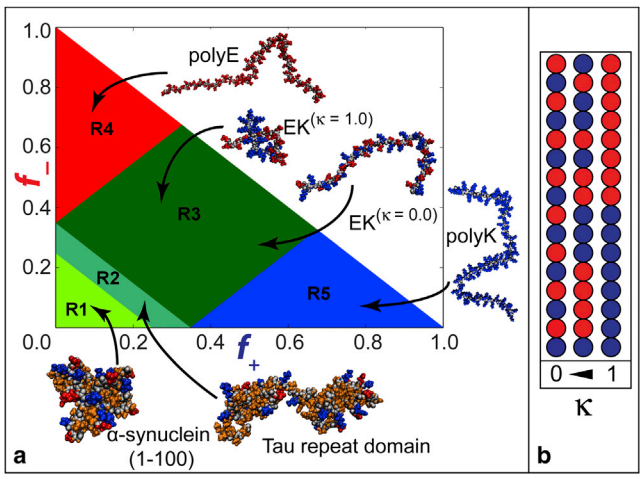

FIGURE 1 (*a*) Diagram of states annotated with representative conformations for specific IDPs that correspond to each of the five regions. (*b*) Schematic depiction of the implication of changing κ values. Here, red and blue circles represent negatively charged residues and positively charged residues, respectively.

Our goal is to enable efficient annotation of various sequence features of IDPs and to facilitate the rapid design of sequences that enable direct investigation of sequence-ensemble-function relationships. Accordingly, we have introduced a pair of tools to annotate IDP sequences by their expected sequence-ensemble relationships. Classification of Intrinsically Disordered Ensemble Relationships (CIDER) is a web server that provides instantaneous access to a range of properties that are derivable from the primary sequence of IDPs. This includes NCPR, FCR, κ values, hydrophobicity, compositional bias, and diagram-of-states classification. localCIDER is a locally installable software package for the high-throughput analysis of disordered sequences, and it includes a wider range of IDP-specific sequence-analysis routines.

The remainder of this report is organized is follows. First, we introduce CIDER and localCIDER and describe the relevant use-cases for each of the tools. We then describe the various analyses that can be performed using localCIDER. Finally, we outline several examples where sequence analysis via CIDER/localCIDER has been used to generate predictive insights regarding sequence-ensemble relationships of IDPs.

Overview of CIDER and localCIDER

CIDER is a user-friendly, modern web server that enables rapid analysis of IDP sequences to generate expectations based on prior observations regarding sequence-ensemble relationships. It is freely accessible via http://pappulab.wustl.edu/CIDER. Full documentation and a user guide are available at http://pappulab.wustl.edu/CIDER/help/. The web server takes unformatted or FASTA-formatted sequences as inputs. It uses an intelligent formatting algorithm to strip out non-alphabetic characters. The analysis performed by CIDER is synthesized in terms of a series of sequence-specific parameters and plots that quantify the information accessed from the sequence information that is input by the user. A sampling of the analysis that is provided by CIDER is shown in Fig. 2. CIDER makes all the calculated sequence parameters available in downloadable text format. Multiple sequences can be analyzed and visualized simultaneously.

Unlike CIDER, localCIDER is a standalone software package that was developed to be a high-performance toolkit for the programmatic analysis of IDP sequences. It combines a wide array of sequence-analysis routines with built-in plotting functions to create a single, all-encompassing framework for the analysis of IDP sequences. Installation information and documentation are available via http://pappulab.github.io/localCIDER/. The decision to create a standalone web server and a locally deployable

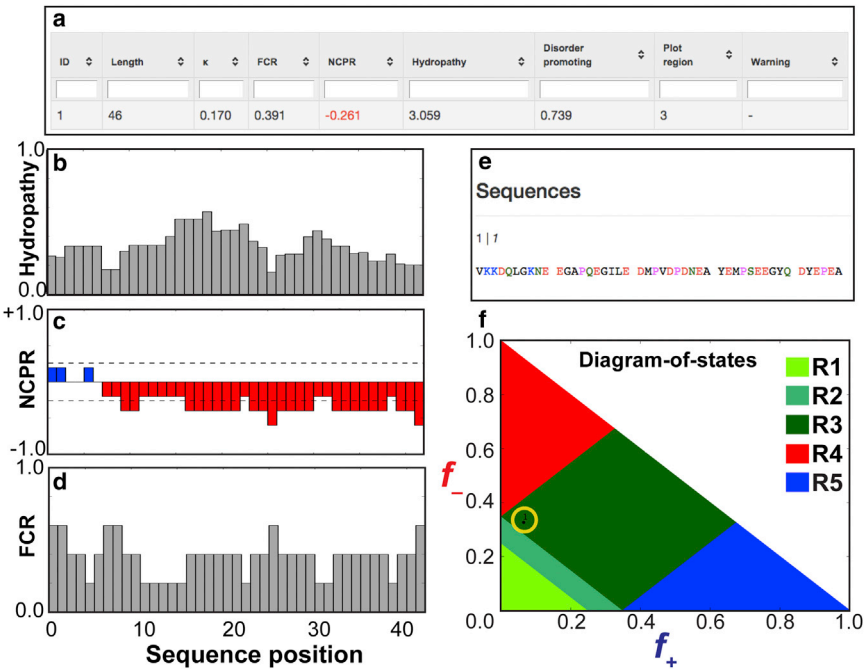

FIGURE 2 Overview of a subset of the output generated by CIDER. (*a*) Overview of the parameters that are calculated. When multiple sequences are analyzed, each column is sortable. (*b–d*) A sliding-window approach is used to show the linear hydrophobicity, NCPR, and FCR, respectively. (*e*) The physicochemically colored sequence. Here, black denotes hydrophobic residues, green denotes polar residues, and blue and red denote positive and negatively charged residues, respectively. (*f*) A sequence-annotated diagram of states.

software package was motivated by the fact that these two tools serve very different needs. A web server is ideal for quick, user-friendly access to summary statistics, since this does not require any time or resource investments from the user. Web servers, however, introduce the complication of network latency, as well as providing a single point of failure when one seeks high-throughput sequence analysis. localCIDER is an easily installable package that mitigates these issues and allows the deployment of powerful sequence-analysis pipelines.

Analyses available within localCIDER

The localCIDER package implements a wide array of customizable analysis routines for the study of IDP sequences. There are two main classes of analysis in localCIDER. Sliding windows can be used to analyze local sequence features, thereby generating position-specific descriptions of various physicochemical properties encoded by the IDP sequence. The sizes of sliding windows can be set to any value, which allows the analysis to be performed on any length scale (see Fig. 3 *a*). In addition, the user can quantify global descriptors that are computed as averages over the entire sequence. These include a range of parameters such as hydrophobicity, NCPR, FCR, κ, diagram-of-states classification, and average PPII propensities (17–20). The local and global enrichment of particular classes of amino acids is readily visualized and quantified (see Fig. 3 *b*). The linear sequence complexity can be calculated using one of three possible complexity measures, viz., Wootton-Federhen complexity (21), linguistic complexity (22), and Lempel-Ziv-Welch complexity (23). Many of these analysis routines allow the specification of user-defined adjustable parameters.

The parameter κ, which quantifies the patterning of oppositely charged residues, is calculated using a newly developed deterministic algorithm with $O(1)$ complexity. If phosphosites are known a priori, these can be passed in as inputs and the distribution of κ values associated with various possible phosphorylation states can be calculated automatically, providing insight into how sub-stoichiometric phosphorylation would influence κ. Recently, we introduced a binary patterning parameter, $\boldsymbol{\Omega}$, that quantifies the linear mixing/segregation of prolines and charged residues vis-à-vis all other residues. This is of particular relevance for IDRs with high proline and low charge contents (24). In addition to calculating $\boldsymbol{\Omega}$, one can generalize the calculation of patterning parameters to any arbitrary binary sequence-patterning parameter. In this approach, one collects one set of residues into one group and all others into the second group. This allows one to investigate the mixing/segregation of any pair of residue types that are grouped into two categories. Examples include hydrophobic patterning, whereby all residues are grouped into hydrophobic or non-hydrophobic sets, disorder- or order-promoting residues, or neutral polar residues or all other residues. Similarly, analysis of ternary patterning, where residues are assigned to one of three groups, is also possible. Finally, input sequences can be converted into a reduced alphabet using either a set of pre-defined reduced alphabets (25) or by passing in a user-defined reduced alphabet mapping. A reduced alphabet representation may

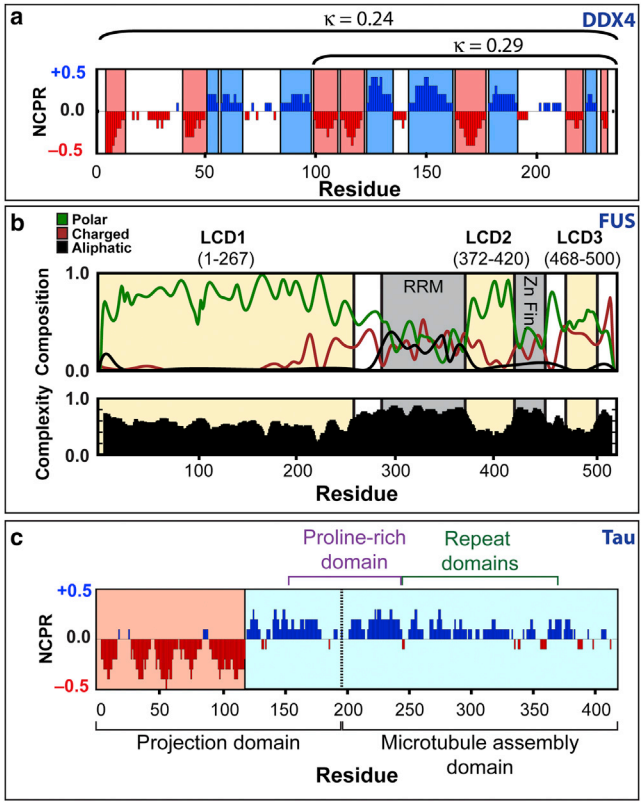

FIGURE 3 Three examples of linear sequence analysis performed by localCIDER. (*a*) Charge patterning in the protein DDX4, identifying the local region with a high net positive or negative charge and showing the full-sequence and C-terminal κ values. (*b*) Illustration of the local sequence composition and complexity of the protein FUS. RNA recognition motifs (*RRM*) and Zinc-finger domains (*Zn-Fin*) are annotated based on published structural information. (*c*) Charge distribution in the 4R-441 isoform, with various domains and regions annotated.

be convenient for creating more coarse-grained sequence representations for further analysis, sequence clustering, and sequence comparison.

In addition to the various numerical analysis routines described above, all the linear analysis routines can directly generate pre-formatted PDF or PNG figures. In addition, diagram-of-state annotations and charge-hydropathy plots can be generated with an arbitrary number of different sequences on the same plot. The utility of an analysis package such as localCIDER comes from the ability to combine local and global sequence analysis with additional classification tools and statistical methods to enable rapid, customized high-throughput analysis pipelines (26).

Discussion

In this section, we discuss examples of how sequence analysis can be used to uncover inferences regarding the biophysical properties of IDPs. These inferences serve as ideal starting points for the development of testable hypotheses. Pak et al. used localCIDER to identify local clusters of high negative charge in the disordered region of the Nephrin intracellular domain that drives phase separation via complex coacervation (26). The tools within localCIDER were coupled to a statistical analysis framework to identify the amino acid types that most strongly influenced phase separation. Similarly, work by Nott et al. (which pre-dates the release of localCIDER) identified clusters of charged residues in the N-terminal IDR of DDX4. These clusters are required to drive phase separation (27) (see Fig. 3 *a*). Given that many IDPs contain local clusters of charged residues, and that the patterning of charged residues has been shown to play a role in determining both conformation and function (16,28–30), we expect there to be many more examples where the distribution of charged residues has a major impact on the sequence-ensemble-function relationships of IDPs.

Amino acid compositions of IDPs play central roles in determining their conformational properties (7,11–16,28,31). localCIDER enables rapid, proteome-wide investigations of compositional biases and the evolutionary preferences within IDPs. We analyzed the complete set of IDPs from 16 model organisms to ask how general compositional biases in IDPs vary across diverse proteomes. For the higher eukaryotes (chordates), we found highly similar sequence properties, whereas lower eukaryotes displayed greater variety. The disordered proteome of *Dictyostelium discoideum* showed a substantial deficiency of charged residues and enrichment in polar residues (notably Asn and Gln) when compared to other species. This result is in accord with the findings of Malinovska et al. (32). Conversely, the disordered proteome of *Plasmodium falciparum* is enriched in strong polyampholytic IDPs, with almost 50% of IDPs falling into R3 on the diagram of states. These results and additional analyses are explored in greater detail in the Supporting Material. The complete analysis of 203,944 disordered fragments took just over 2 h on a desktop computer, showcasing the high-throughput nature of localCIDER. In a similar vein, proteome-scale analysis has also been performed using localCIDER (33) or ideas captured in localCIDER (34).

Fig. 3 shows three examples of the types of linear sequence analysis that localCIDER facilitates. In Fig. 3 *a*, the charge patterning associated with the N-terminal IDP from DDX4 identified by Nott et al. is re-examined (27). Fig. 3 *b* illustrates the complexity and composition associated with the protein FUS, which is known to drive liquid-liquid phase separation in vitro and in vivo (35). In addition to the well-characterized N-terminal low-complexity domain (LCD), which we refer to as LCD1, we highlight two shorter LCDs toward the C-terminus (LCD2 and LCD3). To our knowledge, these regions remain largely unexplored, and it is conceivable that they contribute to modulating the driving forces for phase separation. Finally, Fig. 3 *c* illustrates the charge distribution across the tau protein (4R-441 isoform). The N-terminal 120 residues encompass

a high density of acidic residues, whereas the remainder of the sequence is highly basic. The delineation of positively charged and negatively charged residues does not overlap with other known sequence annotations. This charge distribution is expected to have an impact on how tau associates with other charged biopolymers, such as heparin (36), due to an effective macro-dipole across the sequence.

Conclusions

The importance of sequence features in determining the conformational behavior of IDPs has been demonstrated in different sequences via different approaches. Here, we introduce a pair of computational tools to analyze, describe, and interpret IDP sequences directly. The combination of CIDER and localCIDER offer simple approaches to generate biophysically meaningful insights from analysis of physicochemical properties encoded in IDP sequences.

SUPPORTING MATERIAL

Supporting Materials and Methods, seventeen figures, and two tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(16)34269-2.

AUTHOR CONTRIBUTIONS

Conceptualization, A.S.H., R.K.D., and R.V.P.; Software Engineering, A.S.H., J.N.A., and M.O.G.R.; Analysis, A.S.H.; Writing, A.S.H. and R.V.P.

ACKNOWLEDGMENTS

We thank Anu Mittal, Kiersten Ruff, Max Staller, Ammon Posey, and Tyler Harmon for many useful and insightful discussions.

The National Science Foundation funded this work through grants MCB 1121867 and MCB 1614766 to R.V.P. Washington University's Center for Biological Systems Engineering provided partial support through a Lattig Fellowship to A.S.H.

SUPPORTING CITATIONS

References (37–45) appear in the Supporting Material.

REFERENCES

1. Pancsa, R., and P. Tompa. 2012. Structural disorder in eukaryotes. *PLoS One.* 7:e34687.
2. Wright, P. E., and H. J. Dyson. 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16:18–29.
3. Kovacs, D., and P. Tompa. 2012. Diverse functional manifestations of intrinsic structural disorder in molecular chaperones. *Biochem. Soc. Trans.* 40:963–968.
4. Mitrea, D. M., and R. W. Kriwacki. 2016. Phase separation in biology; functional organization of a higher order. *Cell Commun. Signal.* 14:1.
5. Uversky, V. N., C. J. Oldfield, and A. K. Dunker. 2008. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* 37:215–246.
6. van der Lee, R., M. Buljan, …, M. M. Babu. 2014. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114:6589–6631.
7. Das, R. K., K. M. Ruff, and R. V. Pappu. 2015. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 32:102–112.
8. Chen, J. W., P. Romero, …, A. K. Dunker. 2006. Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *J. Proteome Res.* 5:879–887.
9. Salmon, L., G. Nodet, …, M. Blackledge. 2010. NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* 132:8407–8418.
10. Borcherds, W., F.-X. Theillet, …, G. W. Daughdrill. 2014. Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nat. Chem. Biol.* 10:1000–1002.
11. Uversky, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11:739–756.
12. Marsh, J. A., and J. D. Forman-Kay. 2010. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* 98:2383–2390.
13. Mao, A. H., S. L. Crick, …, R. V. Pappu. 2010. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA.* 107:8183–8188.
14. Müller-Späth, S., A. Soranno, …, B. Schuler. 2010. From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA.* 107:14609–14614.
15. Lyle, N., R. K. Das, and R. V. Pappu. 2013. A quantitative measure for protein conformational heterogeneity. *J. Chem. Phys.* 139:121907.
16. Das, R. K., and R. V. Pappu. 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA.* 110:13392–13397.
17. Tomasso, M. E., M. J. Tarver, …, S. T. Whitten. 2016. Hydrodynamic radii of intrinsically disordered proteins determined from experimental polyproline II propensities. *PLOS Comput. Biol.* 12:e1004686.
18. Elam, W. A., T. P. Schrank, …, V. J. Hilser. 2013. Evolutionary conservation of the polyproline II conformation surrounding intrinsically disordered phosphorylation sites. *Protein Sci.* 22:405–417.
19. Rucker, A. L., C. T. Pager, …, T. P. Creamer. 2003. Host-guest scale of left-handed polyproline II helix formation. *Proteins.* 53:68–75.
20. Shi, Z., K. Chen, …, N. R. Kallenbach. 2005. Polyproline II propensities from GGXGG peptides reveal an anticorrelation with β-sheet scales. *Proc. Natl. Acad. Sci. USA.* 102:17964–17968.
21. Wootton, J. C., and S. Federhen. 1993. Statistics of local complexity in amino acid sequences and sequence databases. *Comput. Chem.* 17:149–163.
22. Troyanskaya, O. G., O. Arbell, …, A. Bolshoy. 2002. Sequence complexity profiles of prokaryotic genomic sequences: a fast algorithm for calculating linguistic complexity. *Bioinformatics.* 18:679–688.
23. Lempel, A., and J. Ziv. 1976. On the complexity of finite sequences. *IEEE Trans. Inf. Theory.* 22:75–81.
24. Martin, E. W., A. S. Holehouse, …, R. V. Pappu, T. Mittag. 2016. Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation. *J. Am. Chem. Soc.* 138:15323–15335.
25. Murphy, L. R., A. Wallqvist, and R. M. Levy. 2000. Simplified amino acid alphabets for protein fold recognition and implications for folding. *Protein Eng.* 13:149–152.
26. Pak, C. W., M. Kosno, …, M. K. Rosen. 2016. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell.* 63:72–85.
27. Nott, T. J., E. Petsalaki, …, A. J. Baldwin. 2015. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell.* 57:936–947.

28. Das, R. K., Y. Huang, …, R. V. Pappu. 2016. Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling. *Proc. Natl. Acad. Sci. USA.* 113:5616–5621.
29. Beh, L. Y., L. J. Colwell, and N. J. Francis. 2012. A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence. *Proc. Natl. Acad. Sci. USA.* 109:E1063–E1071.
30. Banjade, S., Q. Wu, …, M. K. Rosen. 2015. Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proc. Natl. Acad. Sci. USA.* 112:E6426–E6435.
31. Zerze, G. H., R. B. Best, and J. Mittal. 2015. Sequence- and temperature-dependent properties of unfolded and disordered proteins from atomistic simulations. *J. Phys. Chem. B.* 119:14622–14630.
32. Malinovska, L., S. Palm, …, S. Alberti. 2015. *Dictyostelium discoideum* has a highly Q/N-rich proteome and shows an unusual resilience to protein aggregation. *Proc. Natl. Acad. Sci. USA.* 112:E2620–E2629.
33. Vincent, M., and S. Schnell. 2016. A collection of intrinsic disorder characterizations from eukaryotic proteomes. *Sci. Data.* 3:160045.
34. Necci, M., D. Piovesan, and S. C. Tosatto. 2016. Large-scale analysis of intrinsic disorder flavors and associated functions in the protein sequence universe. *Protein Sci.* 25:2164–2174.
35. Patel, A., H. O. Lee, …, S. Alberti. 2015. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell.* 162:1066–1077.
36. Goedert, M., R. Jakes, …, R. Crowther. 1996. Assembly of microtubule-associated protein tau into Alzheimer-like filaments induced by sulphated glycosaminoglycans. *Nature.* 380:550–553.
37. Potenza, E., T. Di Domenico, …, S. C. Tosatto. 2015. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res.* 43:D315–D320.
38. UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* 43:D204–D212.
39. Sickmeier, M., J. A. Hamilton, …, A. K. Dunker. 2007. DisProt: the database of disordered proteins. *Nucleic Acids Res.* 35:D786–D793.
40. Dosztányi, Z., V. Csizmok, …, I. Simon. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.* 21:3433–3434.
41. Day, C. L., C. Smits, …, M. G. Hinds. 2008. Structure of the BH3 domains from the p53-inducible BH3-only proteins Noxa and Puma in complex with Mcl-1. *J. Mol. Biol.* 380:958–971.
42. Rogers, J. M., A. Steward, and J. Clarke. 2013. Folding and binding of an intrinsically disordered protein: fast, but not "diffusion-limited". *J. Am. Chem. Soc.* 135:1415–1422.
43. Burra, P. V., L. Kalmar, and P. Tompa. 2010. Reduction in structural disorder and functional complexity in the thermal adaptation of prokaryotes. *PLoS One.* 5:e12069.
44. Dobrynin, A. V., and M. Rubinstein. 1995. Flory theory of a polyampholyte chain. *J. Phys. II.* 5:677–695.
45. Harmon, T. S., M. D. Crabtree, …, R. V. Pappu. 2016. GADIS: Algorithm for designing sequences to achieve target secondary structure profiles of intrinsically disordered proteins. *Protein Eng. Des. Sel.* 29:339–346.