

LAB PROTOCOL

idpr: A package for profiling and analyzing Intrinsically Disordered Proteins in RWilliam M. McFadden^{1‡}, Judith L. Yanowitz^{1,2*}

1 Magee-Womens Research Institute, Pittsburgh, PA, United States of America, **2** Department of Obstetrics, Gynecology, and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States of America

‡ Current address: Department of Pediatrics, Laboratory of Biochemical Pharmacology, Emory University School of Medicine, Atlanta, GA, United States of America

* yanowitzjl@mwri.magee.edu



OPEN ACCESS

Citation: McFadden WM, Yanowitz JL (2022) *idpr*: A package for profiling and analyzing Intrinsically Disordered Proteins in R. PLoS ONE 17(4): e0266929. <https://doi.org/10.1371/journal.pone.0266929>

Editor: Eugene A. Permyakov, Russian Academy of Medical Sciences, RUSSIAN FEDERATION

Received: December 6, 2021

Accepted: March 29, 2022

Published: April 18, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0266929>

Copyright: © 2022 McFadden, Yanowitz. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: *idpr* is available for download at: <https://doi.org/10.18129/B9.bioc.idpr>.

Funding: This work was funded by National Institutes of Health, grant #R01GM127569 to Judith Yanowitz. <https://public.era.nih.gov/>

Abstract

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are proteins or protein-domains that do not have a single native structure, rather, they are a class of flexible peptides that can rapidly adopt multiple conformations. IDPs are quite abundant, and their dynamic characteristics provide unique advantages for various biological processes. The field of “unstructured biology” has emerged, in part, because of numerous computational studies that had identified the unique characteristics of IDPs and IDRs. The package ‘*idpr*’, short for Intrinsically Disordered Proteins in R, implements several R functions that match the established characteristics of IDPs to protein sequences of interest. This includes calculations of residue composition, charge-hydrophathy relationships, and predictions of intrinsic disorder. Additionally, *idpr* integrates several amino acid substitution matrices and calculators to supplement IDP-based workflows. Overall, *idpr* aims to integrate tools for the computational analysis of IDPs within R, facilitating the analysis of these important, yet under-characterized, proteins. The *idpr* package can be downloaded from Bioconductor (<https://bioconductor.org/packages/idpr/>).

Introduction

Intrinsically disordered proteins (IDPs) are proteins that lack a single, rigid structure under native conditions [1–4], challenging the long-held paradigm that structure leads to function. In addition to typical cellular processes, IDPs have been implicated in human diseases such as neurodegenerative disorders and various cancers [5–7]. IDPs contain one or more intrinsically disordered region (IDR), which are regions of proteins composed of thirty or more disordered residues. Bioinformatic studies have shown that one-third to one-half of eukaryotic proteomes are predicted IDPs [8–11]. Further, viral proteomes appear to be enriched in IDPs, exemplified with the most disordered proteome observed belonging to the *Avian carcinoma virus* with an average disorder composition of over 77% [9].

Due to their apparent abundance and relevance, research interest in IDPs has been increasing [12]. In this regard, there are many computational tools that predict the intrinsic disorder

[commonsplus/](#). The funders had and will not have a role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: he authors have declared that no competing interests exist.

within a protein sequence [13–15]. These tools utilize known differences between disordered and ordered proteins, such as the distinct compositional profile, evolutionary rate, and biochemical properties of IDPs and IDRs compared to proteins or protein-regions with compact, ordered structure [16–19]. Since IDPs have decreased levels of secondary and tertiary structures [1], the primary structure serves as the principal source of computational information for IDPs. Thus, most IDP prediction tools rely on the protein's sequence of amino acids, commonly represented as a character string of individual letters [13–15]. While several R packages analyze protein characteristics based on the amino acid sequence alone, to our knowledge, there is not been a package that is focused on the unique features of IDPs and IDRs.

The R package that we created borrows its acronym from “IDPR” or Intrinsically Disordered Protein Regions; *idpr* stands for “Intrinsically Disordered Proteins in R”. The goal of this R package is to integrate tools for IDP analysis, including amino acid composition, charge, and hydrophathy, using the R platform. Additional IDP analysis is facilitated by several amino acid substitution matrices that are IDP-specific [20–22] as well as linking to the suite of disorder predictions by IUPred2A [23, 24] retrieved by connection to their REST API. The *idpr* package can be found at <https://bioconductor.org/packages/idpr/>.

idpr aims to balance a workflow that automatically generates key visualizations for users of any skill level with a workflow that allows dynamic input and custom output for more-experienced users. The *ggplot2* package [25] is used to generate the visualizations, allowing users to access *ggplot* theme options and aesthetics for further customization. Additionally, *idpr* graphic functions give users the option to return calculations as values for downstream analysis. Overall, *idpr* aims to integrate multiple tools for the computational analysis of intrinsically disordered proteins within R.

Methods

A. Implementation

idpr is implemented as an open-source R [26] / Bioconductor [27] package under an LGPL-3 license. For integration with various packages, *idpr* functions accept protein sequences as character strings, vectors of individual amino acids, and XString objects from the Biostrings package [28]. Alternatively, functions can analyze sequences directly from .fasta files. Substitution matrices within this package can integrate with other R packages used for sequence analysis and multiple sequence alignments. Package dependencies include R version 4.1.3, *Biostrings* [28], *jsonlite* [29], and several tidyverse packages [30] including *ggplot2* [25].

idpr is a well-documented package with detailed user manuals and function descriptions, generated with *roxygen2* [31]. This package also includes six vignette documents (long-form documentation) that discusses the theories of IDPs with the functionality of the package. Versions of *idpr* can be installed through the *BiocManager* package manager [32] from Bioconductor Release ≥ 3.13 (bioconductor.org) with the following line of R code: `BiocManager::install("idpr")`. *idpr* version 1.5.11 was used for this publication and the workflow can be found in the supplementary materials or at [dx.doi.org/10.17504/protocols.io.kqdg3p241125/v1](https://doi.org/10.17504/protocols.io.kqdg3p241125/v1).

B. ‘idprofile’

To quickly generate the *idpr* profile for a protein of interest, a UniProt ID and the amino acid sequence are used to create multiple plots with a single command. `idprofile()` serves as a wrapping function for key graphing tools within *idpr*. These plots include: Charge-Hydrophathy Plot, Local Charge Plot, Local Scaled Hydrophathy Plot, Structural Tendency Plot, Compositional Profile Plot, and IUPred Plot, and FoldIndex Plot (Discussed Below). If a UniProt ID is

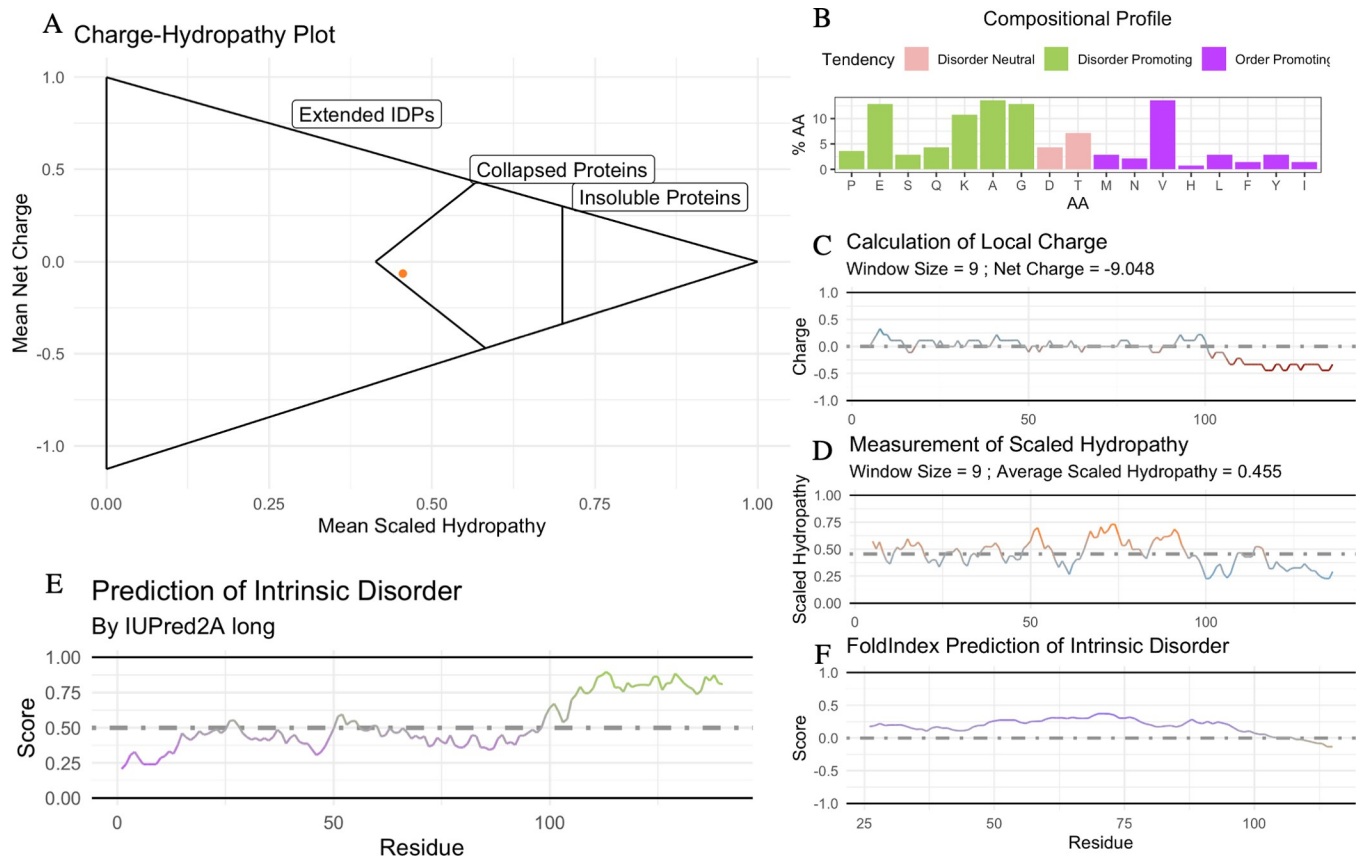


Fig 1. The idprofile of α -Synuclein, generated by *idpr*, returns IDR characteristics. (A) Charge-hydropathy plot of α -Synuclein (α Syn) predicts a collapsed protein. Method are described in [16]. Mean Scaled Hydropathy calculated with the Kyte and Doolittle measurement of hydropathy [33], scaled to Arg = 0.0 and Ile = 1.0. Mean Net Charge calculated with IPC_protein pKa values [34]. Cutoff equation is $\langle \text{Charge} \rangle = \pm 2.785 \langle \text{Hydropathy} \rangle \pm 1.151$ as described previously [18]. Proteins are considered insoluble when $\langle \text{Hydropathy} \rangle \geq 0.7$. (B) Structural tendency plot shows α Syn is enriched in disorder-promoting residues. Disorder-promoting residues (P, E, S, Q, K, A, and G) in green; order-promoting residues (M, N, V, H, L, F, Y, I, W, and C) in purple; disorder-neutral residues (D, T, and R) in pink [35]. (C) Local Charge Plot shows an acidic C-terminus. The local charge is the average of a 9 amino acid wide sliding window, calculated with the IPC_protein pKa values [34]. (D) Local Hydropathy Plot shows a C-terminus deficient in hydrophobic residues. The local hydropathy is the average of a 9 amino acid wide sliding window, calculated with the scaled Kyte and Doolittle measurement of hydropathy [33]. (E) IUPred2 predicts a C-terminal IDR in α Syn [23, 24]. Residues with a score 0.0–0.5 are predicted to be ordered, regions 0.5–1.0 are predicted to be disordered. (F) FoldIndex predicts a C-terminal IDR in α Syn [36]. Residues with a score 0.0 \rightarrow 1.0 are predicted to be ordered, regions -1.0–0.0 are predicted to be disordered.

<https://doi.org/10.1371/journal.pone.0266929.g001>

not included, the IUPred plot is skipped (Fig 1). Please refer to the supplementary workflow and package documentation for details on using *idprofile* and other *idpr* functions.

C. Charge and hydropathy

It has been previously shown that both extreme net charge and deficiency in hydropathy are characteristics of intrinsic disorder proteins [16]. Extended IDPs will occupy a unique area on the plots of both average net charge and mean scaled hydropathy, [16], meaning that the Charge-Hydropathy Plot can distinguish compact from extended proteins under native conditions (Fig 1A). One cannot, however, make a general rule about where IDPs on the spectrum from collapsed protein or an extended protein because IDPs can have the characteristics of either [16, 37].

Protein charges are calculated using the Henderson-Hasselbalch equation [38] with the IPC_protein pKa values [34] by default, although 15 additional pKa data sets are loaded into *idpr* for user preference. The Kyte and Doolittle measurement of hydropathy [33] are used,

scaled with Arg having a hydropathy of 0.0 and Ile having a hydropathy of 1.0. Local charges and local hydropathy are calculated using a sliding window to identify regions of interesting chemistry (Fig 1C and 1D). The sliding window is 9 residues by default but can be changed to any odd number. The resulting figure is similar to one that can be obtained by the ProtScale tool from ExPASy [39].

D. Structural tendency

IDPs as a class tends to have a different composition of amino acids, and therefore distinct overall chemistry, from that of ordered proteins [40]. The chemistry of the specific residues influences its tendency to favor an extended or a compact structure. Residues enriched in the amino acid sequences of IDPs are typically charged, flexible, hydrophilic, or small; whereas order-promoting residues, found in structured proteins, tend to be hydrophobic, aromatic, aliphatic, or disulfide bond. There are also disorder-neutral residues [18, 35]. The default values, described previously [35], are disorder-promoting residues: P, E, S, Q, K, A, and G; order-promoting residues: M, N, V, H, L, F, Y, I, W, and C; and disorder-neutral residues: D, T, and R. These are represented by the structural tendency plot (Fig 1B). Other definitions of order- and disorder-promoting have been published [11], so users can opt to manually specify residue definitions.

E. Disorder predictions

FoldIndex utilizes the described relationship of charge and hydropathy to identify unstructured regions of amino acid sequences [16, 36]. This method is implemented as part of many other prediction programs since it was described in 2005. Using a sliding window of size 51, a negative score (<0) indicates a region is predicted disordered; windows with a positive score (>0) are predicted as ordered [36]. Calculations are made with charge and hydropathy functions within *idpr* and uses IPC_protein pKa values [34] at pH 7.0 and the scaled Kyte and Doolittle measurement of hydropathy [33].

The IUPred2 algorithm calculates a score of intrinsic disorder based on a model of the estimated energy potential for each residue interactions [23]. The structure in protein comes from a network of intramolecular interactions between amino acids. In IDPs, the (lack of) structure comes from the increase interactions of the amino acids with the surrounding environment. This reduced number of interactions leads to the IDP lacking secondary and tertiary structure [41]. IUPred2 predictions are made on a scale of 0.0–1.0, with 0.5 being the dividing line between order and disorder. >0.5 predicting a disordered region; <0.5 predicting an ordered region [23, 24, 41] (Fig 1E). An additional prediction of intermolecular protein-protein interactions is performed with the ANCHOR2 program (Fig 2A), and another predictor of redox-sensitive disorder is performed with IUPred2A Redox (Fig 2B) [23, 24, 41]. A Uniprot ID is required to access the IUPred2A REST API, as well as an internet connection. Visit the IUPred2A website (<https://iupred2a.elte.hu/>) for terms of use, references, and additional information.

F. Visualizing discrete values

As mentioned above, specific amino acid residues are preferentially enriched in unstructured or ordered regions [35]. To visualize the location of assigned residue characteristics in the context of the amino acid sequence, *idpr* contains a way to visualize discrete values with a 'sequenceMap' (Fig 2C, S1B Fig in S1 File). This is not part of the *idprofile* function but is included within the package for additional investigation. The values visualized can be results from *idpr* or from any other source. This function can also visualize continuous values.

G. Substitution matrices for analyzing IDPs

Because there is less restraint to maintain a specific 3D structure IDPs and IDRs tend to evolve faster than ordered proteins [17, 43]. Therefore, IDPs tend to accept increased point mutations at disparate rates when compared to ordered proteins [21].

Currently, PAM and BLOSUM are the most amino acid substitution matrices [44, 45], which are integrated into many web-based tools including NCBI-BLAST+ and EMBOSS [46, 47]. However, customization of the matrices is often desired but is not possible with these online programs. That said, BLOSUM and PAM matrices can both be used with alignment programs in R when loaded via the Biostrings Package [28]. However, for the analysis of IDPs, PAM and BLOSUM matrices are not ideals since they are derived from ordered proteins or favor residue substitutions common among structured proteins [20–22]. To circumvent this pitfall in *idpr*, EDSSMat [20], Disorder [21], and DUNMat [22] which provide IDP-derived substitution matrices have been incorporated for use in alignments.

Results and discussion

A. Example 1 - α -Synuclein

To highlight the use of the *idpr* package, α -Synuclein (α Syn; UniProt ID: P37840), is used in an example analysis. α Syn is an IDP, experimentally validated using various methods [48–52].

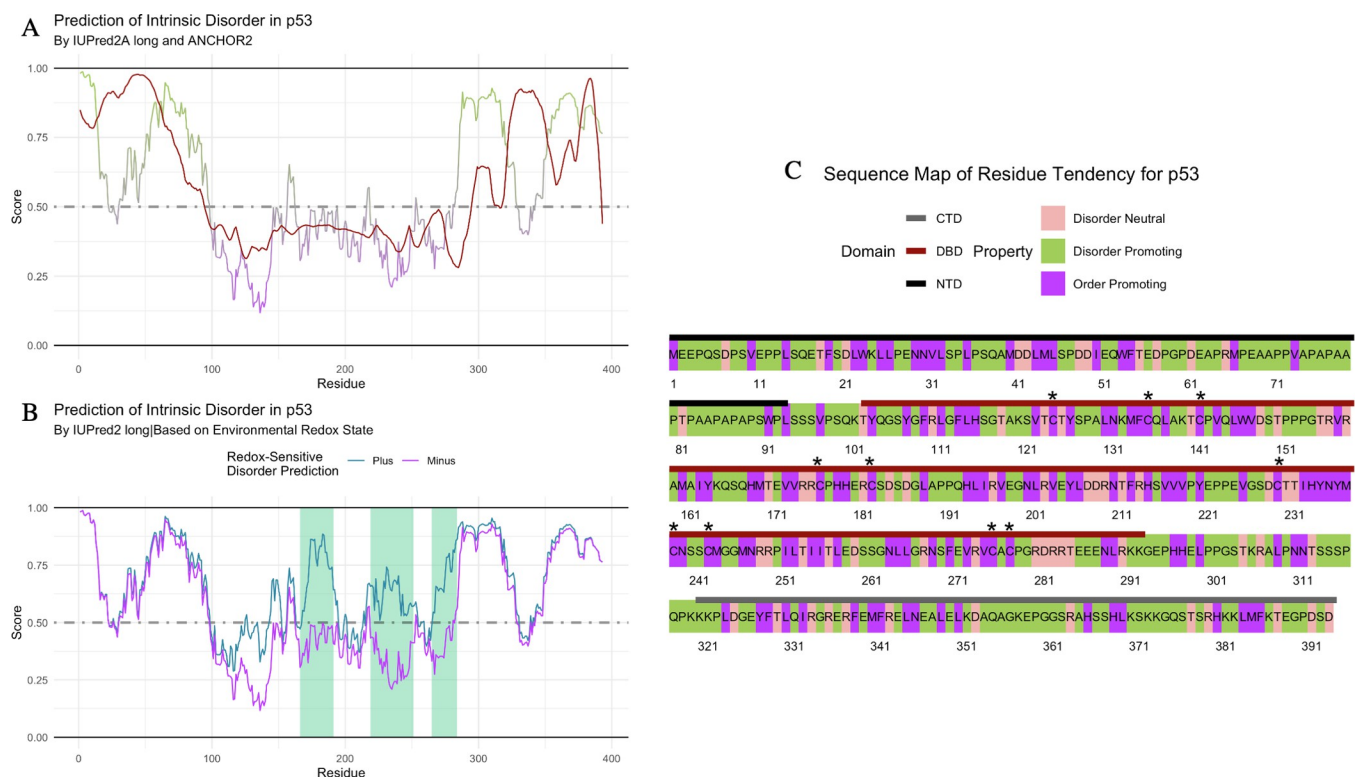


Fig 2. Disorder predictions for p53 domains recapitulate environmental sensitivities. (A) IUPred2A predicts multiple IDRs that promote protein-protein interactions in p53 [23, 24]. Residues with an IUPred2 score (green and purple line) of 0.0–0.5 are predicted to be ordered and residues 0.5–1.0 are predicted to be disordered. Residues with an ANCHOR2 score (red line) greater than 0.5 are predicted to be IDRs and protein-binding domains. (B) IUPred2A Redox predicts several oxidation-sensitive regions in p53 [23, 24]. Redox-plus (reducing environment) predictions are shown in blue, Redox-minus (oxidizing environment) predictions are shown in purple. Regions predicted as “Redox Sensitive” are highlighted in light green. Residues with an IUPred score of 0.0–0.5 are predicted to be ordered and residues 0.5–1.0 are predicted to be disordered. (C) Sequence map of structural tendency for each residue highlights the composition of p53 domains. N-terminal Domain (NTD) annotated by the black bar, DNA-Binding Domain (DBD) annotated by the red bar, C-terminal Domain (CTD) annotated by the grey bar. Conserved Cys residues (C124, C135, C141, C176, C182, C229, C238, C242, C275, C277) annotated [42]. Disorder-promoting residues (P, E, S, Q, K, A, and G) highlighted in green; order-promoting residues (M, N, V, H, L, F, Y, I, W, and C) in purple; disorder-neutral residues (D, T, and R) in pink [35].

<https://doi.org/10.1371/journal.pone.0266929.g002>

This protein has been extensively studied and is heavily implicated in Parkinson's Disease pathology [53, 54]. The idprofile of α Syn returns IDP characteristics (Fig 1). The Charge-Hydrophathy plot shows that α Syn appears to be a collapsed protein, rather than an extended IDP (Fig 1A). This is in line with previous reported data showing regions of α Syn are shielded from the cytoplasm under native conditions [49]. The structural tendency plot shows that α Syn is enriched in disorder-promoting residues, mostly represented by Glu, Lys, Ala, and Gly (Fig 1B). Interestingly, α Syn lacks Cys and Trp, both of which are the most order-promoting residues [35], in addition to lacking Arg, a positively charged and order-neutral residue. The local charges of the protein are mostly neutral, apart from a negatively charged C-terminal region (Fig 1C). In conjunction with the local charge, the C-terminal region is deficient in hydrophobic residues, as shown by the local scaled hydrophathy (Fig 1D). In fact, it has been reported that residues 104–140 of α Syn are more extended than the N-terminal portion of the protein [51]. The Charge-Hydrophathy plot of α Syn residues 104–140 returns an extended IDR, while residues 1–103 returns a collapsed protein with a more neutral charge (S1A Fig in S1 File). This is in line with the IUPred2 and FoldIndex predictions of intrinsic disorder for α Syn, which shows the C-terminal region predicted as disordered (Fig 1E and 1F, S1B Fig in S1 File). There are known point mutations in α Syn that are associated with familial Parkinson's Disease: A30P, E46K, H50Q, G51D, and A53T [55–57]. While these mutations are located in the more compact region of the protein, most mutations occur in disorder-promoting residues, with the exception of H50Q (S1B Fig in S1 File). Overall, the idprofile is useful for identifying biochemical features related to IDRs within a protein of interest.

B. Example 2—p53

Another well characterized IDP is the cellular tumor antigen p53 (UniProt ID: P04637). p53 has been studied extensively since it is mutated in over 50% of human cancers [58, 59]. It is an experimentally validated IDP [60–62] that acts as a protein hub, interacting with many different partners [3, 63]. The idprofile of p53 shows characteristics of a protein with several IDRs (S2 Fig in S1 File). The C-terminal domain (CTD) of p53 has been highly studied due to its ability to reversibly form various secondary structures depending on the specific binding partner studied [3, 60, 63]. For example, residues 377–388 gain an α -helical structure when interacting with S100 calcium-binding protein B, while in the same region, residues 379–387, form a β -strand when interacting with Sirtuin [3, 60, 63]. To predict such regions, ANCHOR2 scores, produced by IUPred2A, predict domains that are disordered and are protein-binding regions which may undergo a gain-of-structure when bound [23, 24]. For p53, ANCHOR2 predicts binding in multiple IDRs, including the CTD and recapitulates the known disorder-to-order transition of this domain mentioned above (Fig 2A).

There are several evolutionarily conserved cysteines within p53, most of which are within the central DNA-binding domain (DBD) [42, 64]. Further, p53 has reported roles in redox regulation [42, 65, 66]. To this point, IUPred2 contains a context-dependent predictor of disorder distinguishing between reducing (plus) or oxidizing (minus) environments that can be used to predict redox-sensitive IDRs that may experience induced folding [23, 24]. IUPred2 Redox predicts that p53 has multiple regions of redox sensitivity in the DBD (Fig 2B). This is in line with the known impact of redox conditions on the DBD that influences the structure—and the subsequent function—of p53, consistent with published literature [64, 65, 67, 68]. The structural tendency of each residue in p53 is highlighted in a sequence map with domains [60] and conserved Cys residues [42] annotated (Fig 2C). This p53 analysis exemplifies the use of functions within *idpr* that are not automatically generated using the idprofile wrapping function.

C. Example 3 –mouse GCNA

The germ cell nuclear acidic protein (GCNA) is required for male fertility and has roles in repairing DNA-protein crosslinks [69–71]. GCNA has orthologs from single-celled protists to mammals. In most species, the N-terminal half of GCNA is disordered and the C-terminal half contains an Sprt-Like metalloprotease domain, zinc finger, and HMG box. While there appears to be occasional losses of either the protease, HMG box, or zinc finger, all GCNA orthologs contain the IDR [70]. The IDR of GCNA is enriched in acidic residues, which contributes to the disordered nature. Interestingly, the mouse GCNA lacks all of the structured domains and was previously predicted to be entirely disordered by IUPRED [70, 72].

The idprofile of mouse GCNA (UniProt ID: A0A1D9BZF0) displays that of an unstructured protein (Fig 3A). There is a long stretch of acidic residues, with glutamic acid (E) being the most abundant residue in the amino acid sequence (Fig 3B and 3C). Further, there is a significant enrichment of disorder promoting residues in mouse GCNA (Fig 3B), aligning with the previously reported amino acid composition of GCNA being similar to that of Disprot, a

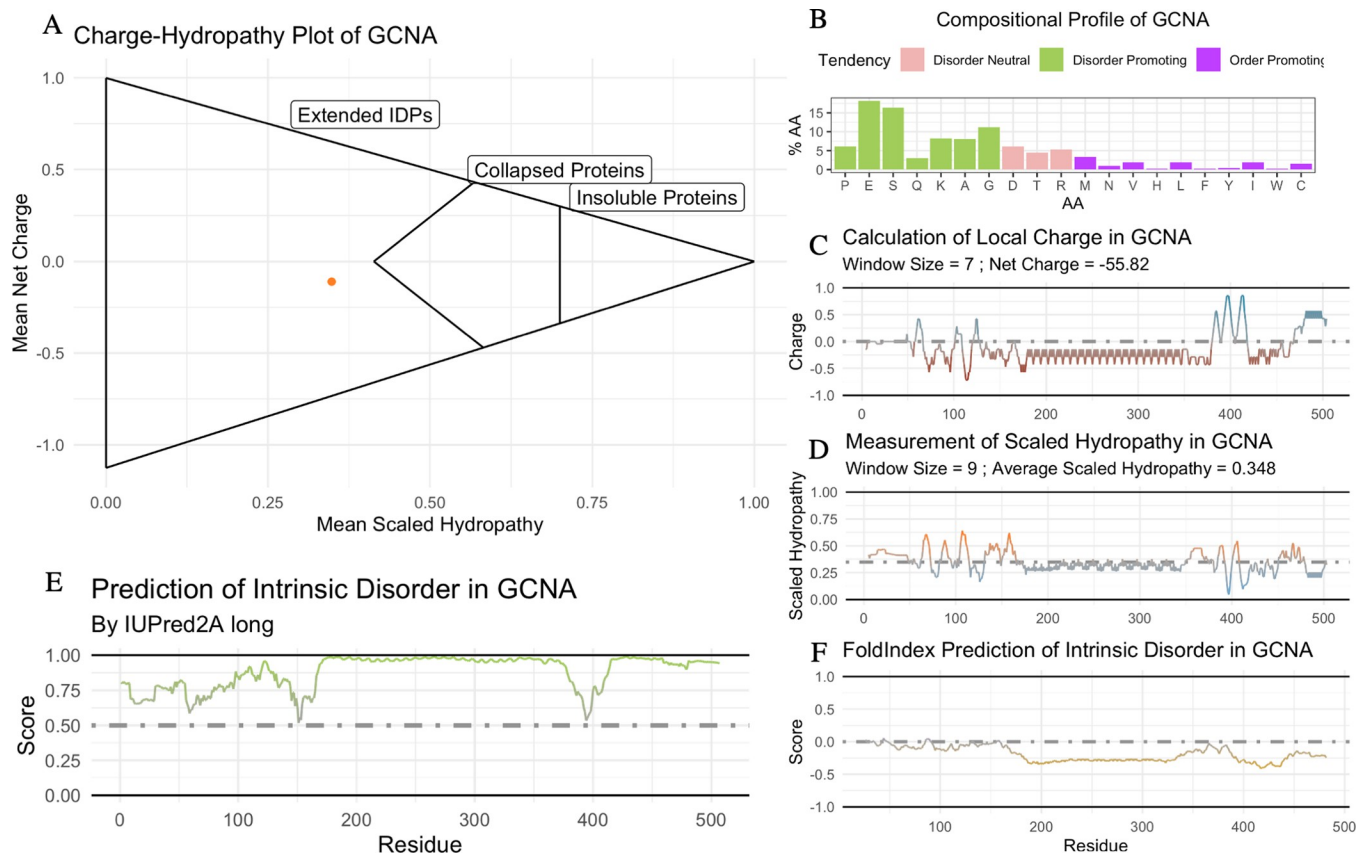


Fig 3. The idprofile of mouse GCNA shows prediction of an entirely disordered IDP. (A) Charge-hydropathy plot of GCNA predicts a disordered protein. Method are described in [16]. Mean Scaled Hydropathy calculated with the Kyte and Doolittle measurement of hydropathy [33], scaled to Arg = 0.0 and Ile = 1.0. Mean Net Charge calculated with IPC_protein pKa values [34]. Cutoff equation is $\langle \text{Charge} \rangle = \pm 2.785 \langle \text{Hydropathy} \rangle \pm 1.151$ as described previously [18]. Proteins are considered insoluble when $\langle \text{Hydropathy} \rangle \geq 0.7$. (B) Structural tendency plot shows GCNA is enriched in disorder-promoting residues. Disorder-promoting residues (P, E, S, Q, K, A, and G) in green; order-promoting residues (M, N, V, H, L, F, Y, I, W, and C) in purple; disorder-neutral residues (D, T, and R) in pink [35]. (C) Local Charge Plot shows an acidic C-terminus. The local charge is the average of a 7 amino acid wide sliding window, calculated with the IPC_protein pKa values [34]. (D) Local Hydropathy Plot shows a C-terminus deficient in hydrophobic residues. The local hydropathy is the average of a 9 amino acid wide sliding window, calculated with the scaled Kyte and Doolittle measurement of hydropathy [33]. (E) IUPred2 predicts a C-terminal IDR in GCNA [23, 24]. Residues with a score 0.0–0.5 are predicted to be ordered, regions 0.5–1.0 are predicted to be disordered. (F) FoldIndex predicts a C-terminal IDR in GCNA [36]. Residues with a score 0.0–+1.0 are predicted to be ordered, regions -1.0–0.0 are predicted to be disordered.

<https://doi.org/10.1371/journal.pone.0266929.g003>

database of intrinsically disordered proteins [70, 73]. There are very few hydrophobic residues, and the protein has an average scaled hydropathy of 0.348 (Fig 3D). Both the extreme acidity of GCNA and the enrichment of disorder-promoting soluble residues contribute to the entire peptide being predicted as disordered from N- to C-terminus (Fig 3E and 3F). This replicates previously reported predictions of mouse GCNA being disordered [70].

Conclusion

We have created an integrated R package that combines disorder prediction tools, hydropathy, and amino acid composition to facilitate the characterization of IDPs. The presence of charge repulsion and hydrophobic deficiencies are hallmark characteristics of an IDP or IDR [18]. The *idpr* package contains distinct, customizable methods for calculating charge and hydropathy for a protein sequence of interest. The output is a visually accessible, graphical readout of critical parameter for IDP analysis. We have validated the use of this tool with α -Synuclein, p53, and GCNA.

A significant portion of the eukaryotic proteome is thought to contain IDRs, but our understanding of these domains is still lacking. In some cases, these domains serve as bridges between two structured domains [74]. In others, like p53, the IDR attains different structure with unique protein partners [3, 75]. Yet in others, the IDRs support liquid-liquid phase separation [76]. In most cases, the role of the IDR is unknown. By providing an integrative tool for characterization of these domains, we envision *idpr* as platform upon which to find commonalities between IDPs and all for sub-division of these protein families.

Supporting information

S1 File. This contains a list of abbreviations and S1 and S2 Figs.
(DOCX)

S2 File. The code used to generate all graphics presented in this manuscript.
(PDF)

Acknowledgments

We would like to acknowledge the support of Dr. Michael Buszczak during the development of the package. Additionally, we would like to thank Dr. Miguel Briño-Enríquez and the members of the Briño-Enríquez and Yanowitz labs for feedback during package development. For providing essential details in making the package, we would like to acknowledge the book *R Packages* by Hadley Wickham (O'Reilly) ©2015 Hadley Wickham, ISBN: 978-1-491-91059-7.

Author Contributions

Conceptualization: William M. McFadden.

Funding acquisition: Judith L. Yanowitz.

Methodology: William M. McFadden.

Project administration: Judith L. Yanowitz.

Software: William M. McFadden.

Supervision: Judith L. Yanowitz.

Visualization: William M. McFadden.

Writing – review & editing: Judith L. Yanowitz.

References

1. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*. 2001; 19(11):26–59. [https://doi.org/10.1016/s1093-3263\(00\)00138-8](https://doi.org/10.1016/s1093-3263(00)00138-8) PMID: 11381529
2. Tompa P. Intrinsically unstructured proteins. *Trends in biochemical sciences*. 2002; 27(10):527–33. [https://doi.org/10.1016/s0968-0004\(02\)02169-2](https://doi.org/10.1016/s0968-0004(02)02169-2) PMID: 12368089
3. Uversky VN. Intrinsically disordered proteins from A to Z. *The International Journal of Biochemistry & Cell Biology*. 2011; 43(8):1090–103. <https://doi.org/10.1016/j.biocel.2011.04.001> PMID: 21501695
4. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*. 1999; 293(22):321–31. <https://doi.org/10.1006/jmbi.1999.3110> PMID: 10550212
5. Uversky VN, Oldfield CJ, Midic U, Xie H, Xue B, Vucetic S, et al. Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics*. 2009; 10(11):S7. <https://doi.org/10.1186/1471-2164-10-S1-S7> PMID: 19594884
6. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys*. 2008; 37:215–46. <https://doi.org/10.1146/annurev.biophys.37.032807.125924> PMID: 18573080
7. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic Disorder and Functional Proteomics. *Biophysical Journal*. 2007; 92(55):1439–56.
8. Pancsa R, Tompa P. Structural disorder in eukaryotes. *PLoS One*. 2012; 7(44):e34687–e. <https://doi.org/10.1371/journal.pone.0034687> PMID: 22496841
9. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *Journal of Biomolecular Structure and Dynamics*. 2012; 30(22):137–49. <https://doi.org/10.1080/07391102.2012.675145> PMID: 22702725
10. Kulkarni P, Uversky VN. Intrinsically Disordered Proteins: The Dark Horse of the Dark Proteome. *PROTEOMICS*. 2018; 18(21–22):1800061. <https://doi.org/10.1002/pmic.201800061> PMID: 30218496
11. Niklas KJ, Dunker AK, Yruela I. The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *Journal of Experimental Botany*. 2018; 69(77):1437–46. <https://doi.org/10.1093/jxb/erx493> PMID: 29394379
12. Kurgan L, Radivojac P, Sussman JL, Dunker AK. On the Importance of Computational Biology and Bioinformatics to the Origins and Rapid Progression of the Intrinsically Disordered Proteins Field. *Biocomputing 2020*. 2019:149–58.
13. Li J, Feng Y, Wang X, Li J, Liu W, Rong L, et al. An Overview of Predictors for Intrinsically Disordered Proteins over 2010–2014. *Int J Mol Sci*. 2015; 16(10):23446–62. <https://doi.org/10.3390/ijms161023446> PMID: 26426014
14. Meng F, Uversky V, Kurgan L. Computational prediction of intrinsic disorder in proteins. *Current protocols in protein science*. 2017; 88(11):2.16.01–14. <https://doi.org/10.1002/cpps.28> PMID: 28369666
15. Liu Y, Wang X, Liu B. A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. *Briefings in Bioinformatics*. 2017; 20(11):330–46.
16. Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins: Structure, Function, and Bioinformatics*. 2000; 41(33):415–27. [https://doi.org/10.1002/1097-0134\(20001115\)41:3<415::aid-prot130>3.0.co;2-7](https://doi.org/10.1002/1097-0134(20001115)41:3<415::aid-prot130>3.0.co;2-7) PMID: 11025552
17. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, et al. Evolutionary rate heterogeneity in proteins with long disordered regions. *Journal of molecular evolution*. 2002; 55(11):104. <https://doi.org/10.1007/s00239-001-2309-6> PMID: 12165847
18. Uversky VN. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Frontiers in Physics*. 2019; 7(10).
19. Forcelloni S, Giansanti A. Evolutionary Forces and Codon Bias in Different Flavors of Intrinsic Disorder in the Human Proteome. *Journal of Molecular Evolution*. 2020; 88(22):164–78. <https://doi.org/10.1007/s00239-019-09921-4> PMID: 31820049
20. Trivedi R, Nagarajaram HA. Amino acid substitution scoring matrices specific to intrinsically disordered regions in proteins. *Scientific Reports*. 2019; 9(11):16380. <https://doi.org/10.1038/s41598-019-52532-8> PMID: 31704957

21. Brown CJ, Johnson AK, Daughdrill GW. Comparing Models of Evolution for Ordered and Disordered Proteins. *Molecular Biology and Evolution*. 2009; 27[33]:609–21. <https://doi.org/10.1093/molbev/msp277> PMID: 19923193
22. Radivojac P, Obradovic Z, Brown CJ, Dunker AK. Improving sequence alignments for intrinsically disordered proteins. *Biocomputing 2002: World Scientific*; 2001. p. 589–600.
23. Mészáros B, Erdős G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic acids research*. 2018; 46(W1):W329–W37. <https://doi.org/10.1093/nar/gky384> PMID: 29860432
24. Erdős G, Dosztányi Z. Analyzing Protein Disorder with IUPred2A. *Current Protocols in Bioinformatics*. 2020; 70(11):e99. <https://doi.org/10.1002/cpbi.99> PMID: 32237272
25. Wickham H. *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag; 2016.
26. R Core Team. *R: A language and environment for statistical computing*. 3.6.1 ed. Vienna, Austria: R Foundation for Statistical Computing.; 2019.
27. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*. 2004; 5(10):R80. <https://doi.org/10.1186/gb-2004-5-10-r80> PMID: 15461798
28. Pagès H, Aboyou P, Gentleman R, DebRoy S. *Biostrings: Efficient manipulation of biological strings*. R package version 2.60.1. 2021.
29. Ooms J. The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv [Preprint] arXiv:14032805*. 2014.
30. Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, et al. Welcome to the Tidyverse. *Journal of open source software*. 2019; 4(43):1686.
31. Wickham H, Danenberg P, Csárdi G, Eugster M. *roxygen2: In-Line Documentation for R*. R Package version 7.1.1. 2020.
32. Morgan M. *BiocManager: access the Bioconductor project package repository*. R package version 1.30.16. 2021.
33. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*. 1982; 157(11):105–32. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0) PMID: 7108955
34. Kozłowski LP. IPC—Isoelectric Point Calculator. *Biology Direct*. 2016; 11(11):55. <https://doi.org/10.1186/s13062-016-0159-9> PMID: 27769290
35. Uversky VN. Unusual biophysics of intrinsically disordered proteins. *Biochimica et Biophysica Acta (BBA)—Proteins and Proteomics*. 2013; 1834(55):932–51. <https://doi.org/10.1016/j.bbapap.2012.12.008> PMID: 23269364
36. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, et al. FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*. 2005; 21(16):3435–8. <https://doi.org/10.1093/bioinformatics/bti537> PMID: 15955783
37. Uversky VN. Paradoxes and wonders of intrinsic disorder: Complexity of simplicity. *Intrinsically Disordered Proteins*. 2016; 4(11):e1135015. <https://doi.org/10.1080/21690707.2015.1135015> PMID: 28232895
38. Po HN, Senozan N. The Henderson-Hasselbalch equation: its history and limitations. *Journal of Chemical Education*. 2001; 78(11):1499.
39. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook*: Springer; 2005. p. 571–607.
40. Theillet F-X, Kalmar L, Tompa P, Han K-H, Selenko P, Dunker AK, et al. The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically disordered proteins*. 2013; 1(11):e24360–e.
41. Dosztányi Z. Prediction of protein disorder based on IUPred. *Protein Sci*. 2018; 27(11):331–40. <https://doi.org/10.1002/pro.3334> PMID: 29076577
42. Sun Y, Oberley LW. Redox regulation of transcriptional activators. *Free Radical Biology and Medicine*. 1996; 21(33):335–48. [https://doi.org/10.1016/0891-5849\(96\)00109-8](https://doi.org/10.1016/0891-5849(96)00109-8) PMID: 8855444
43. Franzosa EA, Xia Y. Structural Determinants of Protein Evolution Are Context-Sensitive at the Residue Level. *Molecular Biology and Evolution*. 2009; 26(10):2387–95. <https://doi.org/10.1093/molbev/msp146> PMID: 19597162
44. Dayhoff M, Schwartz R, Orcutt B. 22 a model of evolutionary change in proteins. *Atlas of protein sequence and structure*. 5: National Biomedical Research Foundation Silver Spring MD; 1978. p. 345–52.

45. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A*. 1992; 89(22):10915–9. <https://doi.org/10.1073/pnas.89.22.10915> PMID: 1438297
46. Johnson M, Zaretskaya I, Raytselis Y, Merezhuik Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Research*. 2008; 36(suppl_2):W5–W9. <https://doi.org/10.1093/nar/gkn201> PMID: 18440982
47. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*. 2019; 47(W1):W636–W41. <https://doi.org/10.1093/nar/gkz268> PMID: 30976793
48. Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry*. 1996; 35(43):13709–15. <https://doi.org/10.1021/bi961799n> PMID: 8901511
49. Theillet F-X, Binolfi A, Bekei B, Martorana A, Rose HM, Stuiver M, et al. Structural disorder of monomeric α -synuclein persists in mammalian cells. *Nature*. 2016; 530(7588):45–50. <https://doi.org/10.1038/nature16531> PMID: 26808899
50. Fauvet B, Mbefo MK, Fares M-B, Desobry C, Michael S, Ardah MT, et al. α -Synuclein in central nervous system and from erythrocytes, mammalian cells, and *Escherichia coli* exists predominantly as disordered monomer. *J Biol Chem*. 2012; 287(19):15345–64. <https://doi.org/10.1074/jbc.M111.318949> PMID: 22315227
51. Eliezer D, Kutluay E, Bussell R Jr, Browne G. Conformational properties of α -synuclein in its free and lipid-associated states. *Journal of molecular biology*. 2001; 307(44):1061–73. <https://doi.org/10.1006/jmbi.2001.4538> PMID: 11286556
52. Kim D-H, Lee J, Mok KH, Lee JH, Han K-H. Salient Features of Monomeric Alpha-Synuclein Revealed by NMR Spectroscopy. *Biomolecules*. 2020; 10(33):428. <https://doi.org/10.3390/biom10030428> PMID: 32164323
53. Lücking CB, Brice A. Alpha-synuclein and Parkinson's disease. *Cellular and Molecular Life Sciences CMLS*. 2000; 57(13):1894–908. <https://doi.org/10.1007/PL00000671> PMID: 11215516
54. Xu L, Pu J. Alpha-Synuclein in Parkinson's Disease: From Pathogenetic Dysfunction to Potential Clinical Application. *Parkinson's Disease*. 2016; 2016:1720621. <https://doi.org/10.1155/2016/1720621> PMID: 27610264
55. Kasten M, Klein C. The many faces of alpha-synuclein mutations. *Movement Disorders*. 2013; 28(66):697–701. <https://doi.org/10.1002/mds.25499> PMID: 23674458
56. Flagmeier P, Meisl G, Vendruscolo M, Knowles TPJ, Dobson CM, Buell AK, et al. Mutations associated with familial Parkinson's disease alter the initiation and amplification steps of α -synuclein aggregation. *Proceedings of the National Academy of Sciences*. 2016; 113(37):10328. <https://doi.org/10.1073/pnas.1604645113> PMID: 27573854
57. Fujioka S, Ogaki K, Tacik PM, Uitti RJ, Ross OA, Wszolek ZK. Update on novel familial forms of Parkinson's disease and multiple system atrophy. *Parkinsonism & Related Disorders*. 2014; 20:S29–S34. [https://doi.org/10.1016/S1353-8020\(13\)70010-5](https://doi.org/10.1016/S1353-8020(13)70010-5) PMID: 24262183
58. Hollstein M, Sidransky D, Vogelstein B, Harris CC. p53 mutations in human cancers. *Science*. 1991; 253(5015):49–53. <https://doi.org/10.1126/science.1905840> PMID: 1905840
59. Surget S, Khoury MP, Bourdon J-C. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. *Oncotargets and therapy*. 2014; 7:57.
60. Kannan S, Lane DP, Verma CS. Long range recognition and selection in IDPs: the interactions of the C-terminus of p53. *Scientific reports*. 2016; 6:23750. <https://doi.org/10.1038/srep23750> PMID: 27030593
61. Vise PD, Baral B, Latos AJ, Daughdrill GW. NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain. *Nucleic acids research*. 2005; 33(77):2061–77. <https://doi.org/10.1093/nar/gki336> PMID: 15824059
62. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, et al. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci U S A*. 2008; 105(15):5762–7. <https://doi.org/10.1073/pnas.0801353105> PMID: 18391200
63. Sullivan KD, Galbraith MD, Andrysiak Z, Espinosa JM. Mechanisms of transcriptional regulation by p53. *Cell Death & Differentiation*. 2018; 25(11):133–43.
64. Buzek J, Latonen L, Kurki S, Peltonen K, Laiho M. Redox state of tumor suppressor p53 regulates its sequence-specific DNA binding in DNA-damaged cells by cysteine 277. *Nucleic acids research*. 2002; 30(11):2340–8. <https://doi.org/10.1093/nar/30.11.2340> PMID: 12034820
65. Eriksson SE, Ceder S, Bykov VJN, Wiman KG. p53 as a hub in cellular redox regulation and therapeutic target in cancer. *Journal of Molecular Cell Biology*. 2019; 11(44):330–41. <https://doi.org/10.1093/jmcb/mjz005> PMID: 30892598

66. Maillet A, Pervaiz S. Redox Regulation of p53, Redox Effectors Regulated by p53: A Subtle Balance. *Antioxidants & Redox Signaling*. 2011; 16(11):1285–94.
67. Hainaut P, Milner J. Redox Modulation of p53 Conformation and Sequence-specific DNA Binding &in Vitro&in Vitro. *Cancer Research*. 1993; 53(19):4469. PMID: [8402615](https://pubmed.ncbi.nlm.nih.gov/8402615/)
68. Rainwater R, Parks D, Anderson ME, Tegtmeyer P, Mann K. Role of cysteine residues in regulation of p53 function. *Molecular and Cellular Biology*. 1995; 15(77):3892. <https://doi.org/10.1128/MCB.15.7.3892> PMID: [7791795](https://pubmed.ncbi.nlm.nih.gov/7791795/)
69. Bhargava V, Goldstein CD, Russell L, Xu L, Ahmed M, Li W, et al. GCNA Preserves Genome Integrity and Fertility Across Species. *Developmental Cell*. 2020; 52(11):38–52.e10. <https://doi.org/10.1016/j.devcel.2019.11.007> PMID: [31839537](https://pubmed.ncbi.nlm.nih.gov/31839537/)
70. Carmell MA, Dokshin GA, Skaletsky H, Hu Y-C, van Wolfswinkel JC, Igarashi KJ, et al. A widely employed germ cell marker is an ancient disordered protein with reproductive functions in diverse eukaryotes. *eLife*. 2016; 5:e19993. <https://doi.org/10.7554/eLife.19993> PMID: [27718356](https://pubmed.ncbi.nlm.nih.gov/27718356/)
71. Borgermann N, Ackermann L, Schwertman P, Hendriks IA, Thijssen K, Liu JC, et al. SUMOylation promotes protective responses to DNA-protein crosslinks. *The EMBO Journal*. 2019; 38(8):e101496. <https://doi.org/10.15252/embj.2019101496> PMID: [30914427](https://pubmed.ncbi.nlm.nih.gov/30914427/)
72. Dosztányi Z, Csizmek V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*. 2005; 21(16):3433–4. <https://doi.org/10.1093/bioinformatics/bti541> PMID: [15955779](https://pubmed.ncbi.nlm.nih.gov/15955779/)
73. Quaglia F, Mészáros B, Salladini E, Hatos A, Pancsa R, Chemes LB, et al. DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Research*. 2021; 50(D1):D480–D7.
74. Basile W, Salvatore M, Bassot C, Elofsson A. Why do eukaryotic proteins contain more intrinsically disordered regions? *PLOS Computational Biology*. 2019; 15(77):e1007186. <https://doi.org/10.1371/journal.pcbi.1007186> PMID: [31329574](https://pubmed.ncbi.nlm.nih.gov/31329574/)
75. Warren C, Shechter D. Fly Fishing for Histones: Catch and Release by Histone Chaperone Intrinsically Disordered Regions and Acidic Stretches. *J Mol Biol*. 2017; 429(16):2401–26. <https://doi.org/10.1016/j.jmb.2017.06.005> PMID: [28610839](https://pubmed.ncbi.nlm.nih.gov/28610839/)
76. McCarty J, Delaney KT, Danielsen SPO, Fredrickson GH, Shea J-E. Complete Phase Diagram for Liquid–Liquid Phase Separation of Intrinsically Disordered Proteins. *The Journal of Physical Chemistry Letters*. 2019; 10(8):1644–52. <https://doi.org/10.1021/acs.jpclett.9b00099> PMID: [30873835](https://pubmed.ncbi.nlm.nih.gov/30873835/)