

Minimum information guidelines for experiments structurally characterizing intrinsically disordered protein regions

Received: 12 July 2022

Accepted: 18 May 2023

Published online: 3 July 2023

 Check for updates

Bálint Mészáros^{1,2,4}, András Hatos^{2,25,26,27,28,29}, Nicolas Palopoli^{3,29}, Federica Quaglia^{2,4,29}, Edoardo Salladini^{2,29}, Kim Van Roey^{5,6,29}, Haribabu Arthanari^{7,8}, Zsuzsanna Dosztányi⁹, Isabella C. Felli¹⁰, Patrick D. Fischer^{7,8}, Jeffrey C. Hoch¹¹, Cy M. Jeffries¹², Sonia Longhi¹³, Emiliano Maiani^{14,15}, Sandra Orchard¹⁶, Rita Pancsa¹⁷, Elena Papaleo^{14,18}, Roberta Pierattelli¹⁰, Damiano Piovesan², Iva Pritisnac^{19,20}, Luigi Tenorio², Thibault Viennet^{7,8}, Peter Tompa^{17,21,22}, Wim Vranken^{5,6,22}, Silvio C. E. Tosatto² & Norman E. Davey²³ ✉

An unambiguous description of an experiment, and the subsequent biological observation, is vital for accurate data interpretation. Minimum information guidelines define the fundamental complement of data that can support an unambiguous conclusion based on experimental observations. We present the Minimum Information About Disorder Experiments (MIADEx) guidelines to define the parameters required for the wider scientific community to understand the findings of an experiment studying the structural properties of intrinsically disordered regions (IDRs). MIADEx guidelines provide recommendations for data producers to describe the results of their experiments at source, for curators to annotate experimental data to community resources and for database developers maintaining community resources to disseminate the data. The MIADEx guidelines will improve the interpretability of experimental results for data consumers, facilitate direct data submission, simplify data curation, improve data exchange among repositories and standardize the dissemination of the key metadata on an IDR experiment by IDR data sources.

The intrinsically disordered protein (IDP) field is generating increasingly large amounts of biophysical data on the structural properties of IDRs^{1–3}. The complexity of IDP-related data continues to increase, and in recent years there has been a noticeable growth in the number of analyses describing complex structural properties, conditional disorder and disorder–function relationships^{4–8}. Whereas a decade ago most IDP papers characterized disorder as a binary state, now many papers contain comprehensive analyses describing multiple conditional states using several complementary experimental methods^{9,10}. Moreover, the improved experimental tools now enable the

investigation of increasingly complex IDRs, IDPs and multi-domain proteins. A key responsibility of the IDP community is the development of minimum information guidelines to improve the description, interpretation, storage and dissemination of data generated in the rapidly evolving IDP field¹¹. In this document, we introduce the MIADEx guidelines for the definition and interpretation of experimental results from IDP experiments.

Minimum information guidelines define the fundamental unit of information for the unambiguous definition of experimental metadata to the level required for the key results of an experiment to be

A full list of affiliations appears at the end of the paper. ✉ e-mail: norman.davey@icr.ac.uk

comprehended by the wider scientific community¹². The role of minimum information guidelines is to minimize data loss by preserving essential data and removing ambiguity while avoiding redundancy. There are several requirements for a functional minimum information guideline. First, the core information conveyed by the experiment should be unequivocally defined. This should include not only the observation itself, but also any information that would change understanding or confidence in the biological or physical relevance of the observation. Second, adhering to the guidelines should be as effortless as possible, to enable their widespread adoption; that is, the guidelines should avoid any excessive burden in the description of an experiment while capturing the most important information to fulfill the first requirement. Third, the guidelines should be equally applicable to all IDR analysis methods so that the experimental metadata are comparable across all sources of primary data, regardless of the experimental approach. To fulfill these criteria, the MIADE guidelines recommend an unambiguous description of the protein and the construct of the region(s) being studied at amino acid resolution, other components of the sample, the experimental approach and the interpretation of the results. Importantly, any information about the experimental protocols, sample components or sequence properties that might affect the interpretation of the results is an essential part of the unambiguous description of the experimental results.

Minimum information guidelines are a compromise between the necessary depth of information to unambiguously describe an IDP experiment and the reporting burden on researchers producing the metadata. MIADE-compliant data records should allow users to quickly assess an IDP experiment and the associated data, and point to the source data for the complete experimental context, but do not require annotation to a level of detail that allows the experiment to be reproduced. Therefore, unless their definition is essential to unambiguously interpret the results of the experiment, descriptions of several aspects of the experimental setup are not required by the MIADE guidelines; for example, complete descriptions of the experimental constructs, the sample and the experimental protocol are not necessary. In addition, minimum guidelines focus on the description of a single experiment and therefore cannot define how multiple experiments should be integrated to describe more complex features of proteins, such as conformational ensembles. Furthermore, minimum information guidelines are abstract recommendations that do not specify the technical details of the structured data types that are guideline compliant. In this document, we provide examples of data that adhere to MIADE recommendations in multiple use cases, including details on updates that allow MIADE-compliant data to be stored in the DisProt IDP database¹. However, the technical specification of data storage is defined by exchange formats used to standardize and store compliant data, and therefore it is outside the scope of this document.

The MIADE guidelines provide a community consensus created by experimentalists, curators and data scientists on the minimum information required to appropriately describe metadata on experimentally and computationally derived structural state(s) of IDPs or IDRs. The aim is to increase the accuracy, accessibility and usability of published IDP data, to comply with FAIR (findability, accessibility, interoperability and reusability) data principles¹³, to support rapid and systematic curation of IDP data in public databases and to improve interchange of IDP data between resources. We believe that these guidelines will provide an important roadmap for the thousands of data producers, curators and database developers in the IDP field and will increase the utility of published IDP data for the larger biological community.

Where should MIADE be applied?

The vast majority of IDP experiments yield information about the structure or function of IDPs. Functional IDP studies most commonly analyze their interactions with other molecules. Because the Minimum Information about a Molecular Interaction experiment (MIMIX) guidelines¹⁴,

on which the MIADE guidelines have been modeled, already cover the molecular interaction aspects of these experiments, MIADE focuses only on the description of the structural aspects of the studied IDPs.

Experimental data can follow many paths to the final data consumer (Fig. 1a). At each point in the flow of data, valuable information can be lost, misinterpreted or misrepresented. After data production, the primary data are analyzed by field-specific specialists (typically the research group that conducted the experiment) who interpret these complex experimental results to provide a biological observation. These specialists will author a publication that describes the new observations, and, ideally, they will directly submit the findings to a core IDP data resource. Currently, much of the data in the IDP field passes into a branch where biocurators interpret the description of the experiments and observations in the publication and then annotate the information into manually curated resources. The role of MIADE is to provide general recommendations that can be applied at each potential point of data loss to maximize the precision with which information is transferred.

The MIADE guidelines should be applied to free-text descriptions when reporting on the experiment, to data extraction from the primary literature and to structured metadata for dissemination. Therefore, the MIADE guidelines provide a recommendation to unambiguously describe structural information on IDRs inferred from experimental or computational analysis, intended for: (1) researchers authoring an article on the structural state(s) of an IDR; (2) researchers who want to submit such data to an IDP resource directly, for example before peer-reviewed publication of the data; (3) biocurators who want to define or curate data on the structural state(s) of an IDR within an IDP resource; (4) database developers who want to disseminate IDR structural state data; and (5) data users who need to achieve full comprehension, requiring the meaning and origin of each piece of data to be clear (Table 1).

What information is required by the MIADE guidelines

Both the biological and the methodological contexts are required to understand and compare experimental data. Consequently, the MIADE guidelines recommend the clear definition of four components for reporting on IDP structural experiments: the protein region that was studied; the structural state of that region, as inferred from the experiment; the experimental or computational approach applied; and the data source. Each region of a protein for which a structural state was inferred from an experiment should be described separately. The exact application of the guidelines is use-case specific; however, when possible, stable identifiers of external resources should be referenced, for example, UniProt for protein definitions¹⁵, ECO (Evidence and Conclusion Ontology) for experimental definitions¹⁶ and IDPO (Intrinsically Disordered Proteins Ontology) for structural state definitions (<https://disprot.org/ontology>).

The MIADE checklist: minimizing ambiguity in the definition of an experiment

The following information is required to create a MIADE-compliant description of an experiment characterizing the structural properties of an IDR:

Protein region. Definition of the region for which a structural state was experimentally determined or computationally predicted. If several regions of a protein were inferred to be disordered, each region should be defined separately. The definition should be unambiguous and concise, and should leave no doubt about the identity of the protein that contains the region. The source organism and isoform should always be specified. If the sequence is synthetic and not mappable to an existing protein, this should be stated explicitly. The experimental sequence of the protein region being studied should always be defined.

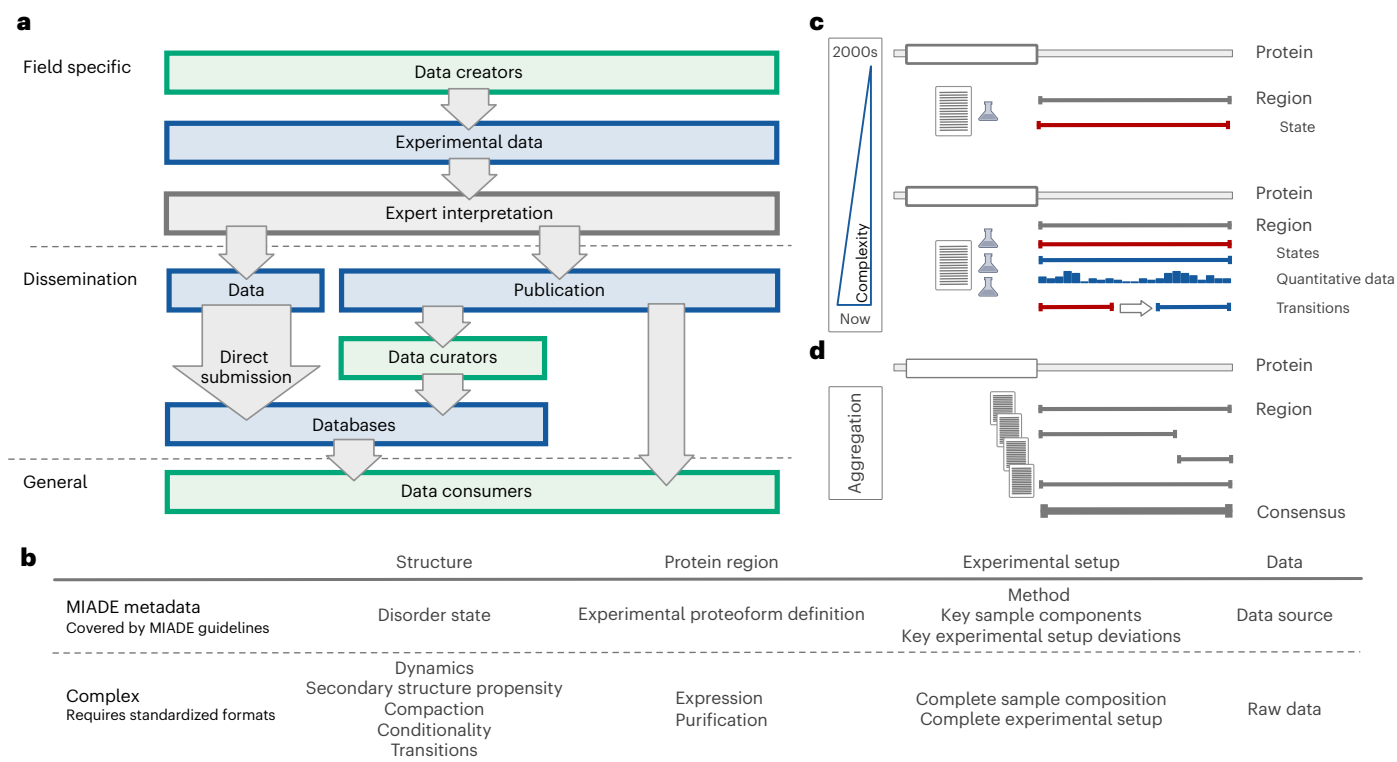


Fig. 1 | Overview of data flow and data types in the IDP field. a, Scheme of data flow from primary data capture by the experimentalist to data dissemination to the end consumer. **b**, Definition of the scope of the MIAGE guidelines and the requirements of a comprehensive standard for IDP data. **c**, Representation

of the evolution of complexity of cutting-edge experimental IDP papers. **d**, Representation of the requirement for data aggregation across analyses to build high-confidence consensus data on a region.

Similarly, any tags, labels, post-translational modifications or mutations present in the sample should be described. Each region should be characterized by:

- Definition of the source protein from which the region was derived:
 - The common name for the source molecule. Both the protein name and gene name should be added whenever possible. Ideally, this should be the official name provided by a nomenclature committee such as the HGNC symbol from the HUGO Gene Nomenclature Committee for human genes¹⁷. In cases in which the field-specific name is used, and it differs from the official name, the official name should be mentioned in the first definition of the molecule. *Example: mitotic checkpoint serine/threonine-protein kinase BUB1 β (BUBR1, also known as BUB1B).*
 - Scientific name, common name or NCBI taxonomy ID of the species of origin for the source protein (or free text for chemical synthesis, unknown and in silico origins). *Example: budding yeast (Saccharomyces cerevisiae strain ATCC 204508/S288c, NCBI Taxon ID: 559292).*
 - Accession or identifier for the source protein in a reference database. If an isoform of a protein was used in the experiment, the accession or identifier specifically identifying that isoform should be used whenever possible. The version number of the protein sequence in the database can be added to further reduce ambiguity. *Example: UniProt: P13569 (P13569-2 if isoform 2 was used).*
- Definition of the protein region(s) for which a structural state was determined:
 - Start and stop positions of the region: the position of the first and last residue of the region, based on (1) the sequence as described in the database annotating the

source protein from which the region was derived (that is, positions should refer to the natural sequence, and should not consider added purification and solubility tags), or (2) in the case of a sequence that is not mappable to a natural sequence, the sequence provided by the data producer. *Example: residues 708–831 of BUBR1.*

- The amino acid sequence of the experimental construct encoding the region(s), in IUPAC one-letter codes¹⁸.
- Definition of the experimental molecule (any tags in the construct that have been removed before the sample has been studied can be ignored), including any alterations and additions to the defined protein region:
 - Tags and labels that are present in the experimental construct. *Example: C-terminal 6 \times His tag.*
 - Experimental proteoform, including mutations, insertions, deletions and post-translational modifications. *Example: phosphorylation of BUBR1 at Ser21.*

Structural state. Structural state of the construct or a region(s) within the construct, as defined by the experimental data or as inferred by the experimentalist.

- Classically, structural states in IDP experiments are defined on the basis of a binary ‘order’ and ‘disorder’ description; however, as more complex structural properties are now being experimentally defined, the structural properties of the region and subregions should be defined at the highest resolution possible. The position of a structurally distinct subregion of a construct, such as the observation of partially populated secondary structural elements, should be defined explicitly, as described for the protein region definition. If the boundaries of the structure state elements within a construct are not clear, this should be stated. When possible, the corresponding term and term ID for that structural state in the

Table 1 | Cases in which the MIADE guidelines should be applied to improve data interpretability and minimize the loss of key data

Storing experimental metadata	<ul style="list-style-type: none"> • Allows storage of high-level metadata • Allows the integration and comparison of data from distinct experiments and experimental approaches
Direct submission of IDR data before publication	<ul style="list-style-type: none"> • Promotes early data capture by providing a standard with a low barrier for data entry to submit experimental results before publication to an IDP database directly • Facilitates collection of IDR data in light of increasing data management and open-science efforts • Increases data available for community blind testing of computational IDP tools
Defining key findings about IDRs in a publication	<ul style="list-style-type: none"> • Defines the requirements for unambiguous description of the specialist interpretation of primary source data • Increases the clarity of the paper and simplifies all downstream data interpretation • Allows the reported structural IDR data to be rapidly captured in an IDP resource, where they are readily available to the community
Curation of IDRs from a publication	<ul style="list-style-type: none"> • Provides clear guidelines for unambiguous and efficient definition of the fundamental details and results of the experiment(s), to facilitate data curation by non-specialists
Transfer of IDR data	<ul style="list-style-type: none"> • Standardized transfer of the key metadata on an IDR experiment to, between and from IDR data repositories • Promotes and/or facilitates the implementation of FAIR principles in the IDR community

IDPO controlled vocabulary should be given. If the observed structural property is not widely known by a general readership, for example, describing more complex attributes than a binary order and disorder definition, such as dynamics, secondary structure propensity or compaction, then the property should be clearly defined. *Example: disorder (IDPO:00076)*.

Experimental and computational approaches. Definition of the experiment or computational approach used to determine the structural state of the region. Each experimental setup should be described separately. For studies that derive structural information from the integration of data from several experiments, each individual experimental observation should be expressed in a MIADE-compliant manner. The following parameters should be included in the experiment description:

- The experimental or computational methods used to determine the structural state of the region. If possible, this should be annotated with the corresponding term and term ID for that experimental method in the ECO controlled vocabulary. The name of the computational or experimental method(s) used to define the structural state of the protein region(s) should be defined to the most detailed level possible. If relevant, any software used in the post-processing of experimental data, or to define the structural state directly, should be defined, including the software version. *Example: far-UV circular dichroism (ECO:0006179)*.
- The scientific name, common name or NCBI taxonomy ID of the host organism in which the experiment was performed (or free text for in vitro, unknown, in vivo or in silico experimental environments); further specification of the cell line or tissue is recommended. Special care should be taken in defining experimental details for in-cell or cell-extract studies. *Example: in vitro*.
- Any experimental deviation that could alter the interpretation of the results and any condition that could impact the results should be clearly described. These deviations are generally method specific: for example, in vitro experimental parameters (for example,

pH; pressure; protein concentrations; temperature; buffer; salt; and additional components, including other proteins), computational parameters (for example, non-default options), Molecular Dynamics (MD) simulation parameters (for example, the force field used) and integrative structural study parameters (for example, experimental sources and integration approach). See the next section and Table 2 for details. *Example: experiment was performed at 4 °C.*

- Any additional components in the sample that could alter the interpretation of the results. This attribute is important to clearly capture structural changes induced by binding partners. However, it also includes other components such as reducing agents, cofactors and crowding agents which may trigger a structural change on the protein of interest. Each component should be defined unambiguously, and if possible, include the concentration of the sample components and refer to external databases including a definition of the molecule (for example, Uniprot or ChEMBL). Additional protein components should be defined to the same level of detail as the experimental region being studied. See next section and Table 2 for details. *Example: experiment was performed in the presence of 10 g l⁻¹ polyethylene glycol 400 (PEG400) (ChEMBL:1201478)*.

If data are stored in a database, transferred between resources or defined in the absence of a paper, it is important to also include the source of the data.

Data source. A reference to where the data were originally described.

- If the data are published in a paper, the following information should be provided:
 - Publication database and identifier. *Example: PMID: 35055108.*
- If the data are directly submitted to a data resource, the following information should be provided:
 - The name of the data resource
 - The accession number of the record holding the data in that resource
 - The data creator who submitted the data
 - Contact details for the data creator

Key factors that can influence the interpretation of structural IDR data

Numerous factors connected to the protein region, protein construct or the experimental setup can influence the structural state of the protein region being studied and, consequently, our confidence in the biological relevance of the observed structure (Table 2)^{19,20}. These factors can be technical perturbations, to allow experimental measurements to be collected (for example, changes in temperature or pH), or perturbations related to the biological question under investigation (for example, proteoforms with a PTM or disease-relevant mutation, or the presence of an interacting partner). In these cases, any description of the structural state is meaningful only when the relevant factors that influence the observed state are specified. Although the minimum information requires the protein region and the experimental method to be defined, it is up to the discretion of the authors to report deviations from the established protocol, sample or sequence that could alter the interpretation of the results. Consequently, an explicit statement by an author will simplify the task of the curator or reader to make a judgment on the importance of a given deviation. In complex cases, the meaningful description of the inferred structural states can include several pieces of information that go beyond the specification of the protein region and the experimental method applied. In Table 2, we provide pointers on which factors might be considered important deviations on the basis of known biological cases of conditional protein disorder and common experimental perturbations.

Table 2 | Key factors that can influence the interpretation of structural IDR data

Deviation type	Deviation description	Example	Ref.
Deviations from the canonical protein sequence	Definition of the construct used in the experiment		
Post-translational modifications (PTMs) (Covalent modification of a residue side chain)	PTMs can change the physicochemical properties of a sequence and thereby alter the structural state, compaction or dynamics of an IDR. The structures of several IDRs have been shown to be modulated by the addition or removal of a PTM. Studies aimed at investigating these mechanisms will characterize modified proteoforms to understand the structural changes.	4E-BP2 folds into a four-stranded beta structure upon phosphorylation of residues T37 and T46.	10
Substitutions, insertions and deletions (Replacement, addition or removal of residues of the canonical protein sequence)	Substitutions, insertions and deletions can affect local and global physicochemical properties of a region (for example, the charge, hydrophobicity, interaction capacity and size), potentially affecting the structural properties of a protein. Studies altering the protein sequence can enable the testing of the effect of indels, polymorphisms or disease variants, certain PTMs (such as phosphomimetics) or isoforms (by addition or removal of an exon).	A p.F82K substitution in ferricytochrome c induces localized unfolding of a distal site in the ferric state.	29
Tags and labels (Covalent attachment of an entity that enables analysis, identification, purification or solubility of the protein)	Tags and labels can have a measurable influence on the dynamics and stability of the protein they are attached to. The addition of tags is almost always a technical necessity, and its aim is not to measure a biological phenomenon. Tags play three major roles in IDR experiments: (1) for purification (for example, FLAG tag), (2) for solubility (for example, maltose-binding protein), (3) for experimental readout (for example, fluorescent tags for fluorescent microscopy or paramagnetic tags for NMR).	The addition of a His tag influences myoglobin short time scale (picoseconds) dynamics.	30
Proteolytic cleavage (Cleavage of the protein chain induced by a protease)	Cleavage can disrupt both local structural elements and long-range contacts by increasing the distance between residue pairs. Cleavage also introduces new N and C termini in the protein chain, changing the polarity, solubility and interaction capacity of regions. Many proteins, especially extracellular proteins, are known to undergo cleavage, often in many subsequent steps. Cleavage products can be created in response to signaling events and often have very different biological activity, interaction capacity and structural states.	Cleavage of the disordered osteopontin removes long-range intramolecular interactions, changing the structural state and the accessibility of the integrin-binding site.	31
Experimental parameters	Parameters of the experimental setup for a sample		
pH (pH of the sample)	The pH can affect the strength of ionic and hydrogen bonds and so can modulate the structural state of a protein ³² . Experimental parameters are often tweaked to find the optimal experimental parameters for the study of a specific protein, sometimes resulting in the use of non-physiological pH. Furthermore, comparison of a physiological state with a non-physiological pH state can be used to probe the structural properties of the region of interest, for example, forcing the complete unfolding of a construct with harsh experimental conditions to allow comparison to a 'ground state.'	NhaA, a sodium proton antiporter of the inner membrane of <i>Escherichia coli</i> , is activated at pH values between 6 and 7, with a maximal activity at pH 8.5, and is inactivated by acidic pH.	33,34
Temperature (Temperature of the sample)	The temperature has an explicit role in determining the strength of entropic terms in the Gibbs free energy that controls the stability of protein structures and complexes. Thus, changing the temperature can drastically change the stability of folded proteins and dynamics of IDRs. Changing the temperature of a protein sample in an experiment can serve to explore its folding or unfolding kinetics, stability and oligomerization. For calorimetric techniques, such as differential scanning calorimetry, temperature regulation is what provides the measurable signal. For certain experiments, such as NMR, changing the temperature is performed for technical reasons to improve the signal-to-noise ratio.	Hp26 becomes active with increased temperature in a two-step mechanism that first activates the protein and then unfolds it.	35,36
Pressure (Hydrostatic pressure of the protein sample)	High hydrostatic pressure (HHP) can induce unfolding by breakage of intramolecular interactions and exposure of cavities allowing binding of water. HHP is used to study the structure of partially structured intermediate transition states and the monomeric forms of oligomeric and aggregated proteins.	The 1D ¹ H NMR spectra support the proposed molten-globule state of Arc repressor under high pressure; moreover, the ¹ H NMR spectra at a pressure range of 3.5–5 kbar are substantially different from those of the native state (1 bar, 20 °C) and the fully denatured state (1 bar, 70 °C).	37,38
Force (Mechanical force applied to the protein)	Opposing forces applied to different parts of the protein can mechanically unfold the structure (either partially or completely), converting mechanical signals into biochemical ones. The most typical information provided are the number of steps in which a protein unfolds (reflecting the number of domains or intermediate structural states) and the force required for unfolding. For proteins undergoing force-induced unfolding in biological settings, these measurements explore their biological function. Atomic force microscopy and high-speed force spectroscopy are used to assess the stability and the folding and unfolding kinetics of proteins.	Mechanical unfolding of TTN-1 and twitchin of <i>Caenorhabditis elegans</i> affects the auto-inhibitory region and the catalytic core of the protein.	39

Table 2 (continued) | Key factors that can influence the interpretation of structural IDR data

Deviation type	Deviation description	Example	Ref.
Redox potential (Redox potential of the sample)	The redox potential affects the behavior of residues, especially that of cysteine. Under oxidizing conditions, cysteines can form disulfide bridges; under reducing conditions, they can coordinate cations. Redox potential parameters are often tweaked to find the optimal experimental parameters for the study of a specific protein. Various cellular compartments have drastically different redox potentials (for example, the extracellular space is oxidizing, whereas the cytoplasm is reducing); thus, changing the redox potential in a sample can model various compartments or the transport between them.	The nuclear export signal (NES) of Yap1 is masked by a structured domain held together by disulfide bridges in the oxidized state. In reducing conditions, the domain unfolds, and the NES becomes exposed and functional.	40
Light (Irradiating the protein with visible, UV or infrared light)	Many light-sensitive proteins contain additional chromophores that can undergo structural changes (most often <i>cis-trans</i> isomerization) that consequently alter the structure and/or dynamics of the protein that they are embedded in. Light-induced folding or unfolding of photosensitive proteins as a response to light is studied by altering these conditions.	Light-induced unfolding of the water-soluble photoactive yellow protein (PYP) allows it to become functionally active and bind partners.	41
Protein concentration (Concentration of the protein being tested in the sample)	Increased protein concentration can promote aggregation, liquid-to-liquid phase separation and liquid-to-solid phase transition. Consequently, the structural state of an IDR can be concentration dependent. The solubility limit defines the concentration in which molecules are miscible in solution. If the protein concentration increases beyond that limit, the macromolecule–macromolecule interactions are energetically more favorable than the macromolecule–solute interactions.	Several phase-separation drivers (that is, FUS and hnRNP1) can undergo percolation or liquid–liquid phase separation in a concentration-dependent manner.	42
Protein source (Details of the protein purification)	An important element of the experimental setup is how the protein was generated, because its prior history may have a significant effect on its structural state by determining the exact proteoform, including post-translational modifications, partial proteolysis and so on. Best practice is to check the final proteoform used in the structural studies, either by mass spectrometry or, if possible, by the structural experimental method itself (such as NMR structure determination).	Important information includes the cell type in which the protein was expressed (for example, <i>E. coli</i> , yeast, insect cells (for example, SF9) or human cells (for example HEK-293); not the source genome where the protein is encoded), the method of extraction (for example, by sonication) and subsequent purification, especially if it included an intermittent heat treatment and the application of agents for solubilization and/or denaturation (for example, tween-20 and urea), protease inhibitors and/or reducing agents.	
Computational parameters (Details of the parameters used in computational processing)	Complex processing of experimental data is commonly required for data interpretation in the IDP field. Any software used and computational parameters that can influence the results should be described.	The interpretation of the results of raw data post-processing, residue-specific intrinsic disorder prediction, molecular dynamics and integrative structural studies all rely heavily on the software and parameterization that are used.	
Experimental sample components	Components added to the sample that are required for technical aspects of the experiment		
Crowding agents (Addition of crowding agents to a sample to mimic the molecular concentrations found in cells)	Quinary interactions can have a strong effect on both the structural properties and interactions of a protein. Consequently, proteins behave differently in different contexts: for example, in the cell, in high concentrations of crowding agents and in a buffer. Few experiments have been performed to probe the effect of crowding on structure and interactions; however, the limited data available have suggested that the contribution can be significant, and that it is largely protein specific. Biophysical measurements taken <i>in vitro</i> may not reflect the actual dynamics in the cellular milieu; consequently, the crowding agents are added to partially mitigate biases introduced by the non-physiological conditions.	Experiments studying the effects of a range of crowding agents at different concentrations on IDRs from PUMA, Ash1, E1A and p53 reveal that the induced structural changes depend on both protein sequence and the crowding agent used.	19
Solubility agents (High ionic strength, amino acids, organic solvent)	Solubility agents (or hydrotropic agents) are typically small molecules that have both a hydrophobic and a hydrophilic region, and can increase the solubility of proteins by shielding their local hydrophobic regions from the solvent. Molecules added to a sample in a structural analysis to improve the solubility of the protein to be studied may alter its structural state.	Ionic strength and glycerol are used to mirror protein charges or increased repulsions, respectively. These two experimental components were both used to keep proteins stable in solution.	43
Folding/unfolding agents (Small molecules, organic solvents, high salt or non-ionic detergents)	Folding and unfolding agents constitute a diverse set of molecules used in the structural characterization of an IDR. They are used to modulate the structural state of a protein by shifting it towards either a folded or unfolded state. This is then used as a reference state with known properties that can be compared with other states, helping understand a structural property of the region under investigation.	Several cosolvents were used to perturb protein's stability: guanidine hydrochloride (GdnHCl) and urea are used to denature or partially unfold proteins, whereas hexafluoroisopropanol (HFIP) and trifluoroethanol (TFE) induce secondary structure formation.	44,45

Table 2 (continued) | Key factors that can influence the interpretation of structural IDR data

Deviation type	Deviation description	Example	Ref.
Preservatives (Protease inhibitors, chelating agents and sodium azide)	Protease inhibitors, chelating agents and sodium azide are often used to improve the overall stability of samples (for example, against proteolysis) and might have an impact on protein's behavior.		
Biological background (Cell lysate, cell extract or in-cell sample)	IDPs are increasingly being investigated in biological backgrounds rather than in vitro. For example, isotopically labeled samples can be specifically studied by NMR in cell lysates, cell extract (nuclear/cytoplasmic extract) or even in cells or organelles. Fluorescently labeled proteins can also be studied in cells.	A range of cell lines, cell extracts and organelles have been used to characterize IDPs in their microenvironments. However, specific information on the amount of sample inside cells, the potential manipulation of cells with genetic engineering or drugs, for example, should be defined. Proper controls for intracellular pH and crowding should be provided for these data to be comparable.	46,47
Biological sample components	Components added to the sample that are directly related to the biological hypothesis being tested		
Binding partners (Known or predicted binding partners or ligands)	Binding an interaction partner including ions, small molecules, proteins, nucleic acids or lipids/membranes can modulate the dynamics, compaction or secondary and tertiary structure of an IDR. Many disordered regions will form distinct conformations in the presence of a specific binding partner. These conformational changes can be drastic, shifting the protein from disordered to highly ordered, or to partially ordered with residual large amounts of disorder. In all cases they result in a shift in the sampled conformations.		
	• Proteins	In isolation, p27 is disordered with nascent secondary structure. Upon binding to Cdk2–cyclin A complex, p27 becomes ordered.	48
	• Nucleic acids	In isolation, HMG-1 is intrinsically disordered; however, upon binding to DNA the protein becomes ordered and adopts a well-defined conformation in the minor groove.	49
	• Lipids or membranes	The intrinsically disordered N-terminal region of Hsp12 adopts a folded conformation comprising four α -helices upon micelle binding.	50
	• Small molecules	The dynamic KIX domain of the coactivator CBP/p300 can be stabilized by the addition of a small molecule.	51
Cofactors (Metal ions, iron-sulfur (Fe-S) clusters or organic cofactors (vitamins and their derivatives or fatty acids))	Cofactors acting as cell state signals can heavily modulate the behavior of an IDP. These observations can include folding and unfolding in the presence or absence of specific metal ions, protein aggregation by negatively charged cofactors compensating positively charged repeat regions or induction of liquid–liquid phase separation.	The calcium-binding Repeat-in-ToXin (RTX) of <i>Bordetella pertussis</i> adenylate cyclase toxin (CyaA) is disordered in the absence of calcium but folds upon calcium binding. This region acts as a switch integrating the differing calcium concentrations between the extracellular and intracellular environment.	52

Example use cases

There are several use cases for MIADE (Table 1). However, in practice, there are two major distinct applications: (1) creating an unambiguous description of an experiment in free text, and (2) encoding the fundamental unit of metadata for an experiment in a standardized format. In this section, we will give examples of how MIADE can be applied in each of these cases.

MIADE for authors

A key step in data capture is the unambiguous description of the specialist interpretation of the primary data. Consequently, an accurate and unequivocal definition of the experimental observation in the text of an article that adheres to the MIADE guidelines will simplify all downstream data interpretation. Defining an experiment in free text requires detail that allows the experiment to be fully reproduced. Consequently, most articles describe the experimental detail at a level

of granularity that far exceeds the requirements of a MIADE-compliant entry. However, a comprehensive description of an experiment's design and results does not mean that the data are accessible to the wider biological community. A common issue among non-specialist readers and curators is that the data are described in a manner that is highly technical, requires extensive knowledge of the experimental method or uses field-specific jargon. Furthermore, important details are often not apparent because they are in materials and methods sections, supplementary materials or even a previously published paper. Consequently, the MIADE guidelines recommend an explicit and unambiguous description of the experimental design, the proteins under analysis and the interpretation of the results.

Consideration should be given to the fact that the description should be understandable to the wider biological community, and the key data should be explicitly stated. This will improve the clarity of the document and allow rapid annotation by curators for community

resources. In many cases, writing engaging and readable scientific prose and writing unequivocal descriptions of complex experiments are conflicting goals. However, in any case where such conflicts occur, substance should take precedence over style. For example, the definition of a protein as ‘Budding yeast (*Saccharomyces cerevisiae* strain ATCC 204508 / S288c (TaxID: 559292)) spindle assembly checkpoint component MAD3 (UniProt: P47074)’ may be awkward in comparison to ‘yeast MAD3.’ However, it removes ambiguity from the protein definition. By following the examples in the checklist and understanding that a reader may not be familiar with terminology related to IDRs and IDR experiments, data can be presented in a manner that is both accurate and globally accessible.

MIADe implementation in DisProt

An important aspect to represent experimentally determined structural states of IDPs and IDRs in a standard format is the use of stable external identifiers and controlled vocabularies (CV) to unambiguously describe the captured data. In the future, IDP-specific exchange formats should be developed to define these attributes for experimental metadata; however, for the moment it is useful to consider how DisProt stores MIADe-compliant data.

DisProt is a manually curated resource of IDRs and IDPs in the literature, and it relies on both professional and community curation. All DisProt entries correspond to a specific UniProt entry (or one of its isoforms) and describe the structural state(s) of the region(s) of the protein. When available, information on the presence of transitions between states, interactions and functions is also curated. The annotation of structural states and transitions makes use of specific IDPO terms (<https://disprot.org/ontology>). As part of the development of the MIADe guidelines, we have updated the DisProt database and curation framework to allow the annotation of MIADe-compliant entries¹. An improved construct definition was required to encode tags, labels, mutations and modifications, and the experimental setup definition was updated to allow complex experimental samples to be described. Importantly, these additions will allow DisProt curators to annotate the observations of complex experiments that define conditional multi-state IDRs, which are becoming increasingly common in the literature.

Proteoform definition. The DisProt resource already included an unambiguous definition of the protein or protein isoforms (using UniProt accession numbers) and its regions by mapping to the UniProt sequence. The updated implementation can now define non-canonical and modified proteoforms. The MIADe integration allows deviations from the wild-type UniProt-defined protein sequence to be encoded. Furthermore, the complete sequence of the experimental construct can now be annotated if available. Annotatable construct alterations include tags and labels (using the PSI-MI ontology (<https://www.ebi.ac.uk/ols/ontologies/mod>)²¹), mutations (using the HGVS nomenclature (<https://varnomen.hgvs.org/>)) and PTMs and non-standard amino acids (using the PSI-MOD ontology (<https://www.ebi.ac.uk/ols/ontologies/mod>)²²).

Experimental conditions definition. DisProt uses the Evidence and Conclusion Ontology (ECO, <https://www.evidenceontology.org/>)²³ to annotate experimental methods. In addition, the DisProt database can now store a range of experimental parameters that can influence our understanding of the biological relevance of an experimental observation, that is pH, temperature, pressure, ionic strength and oxidation–reduction potential. The parameter can be quantified in cases where this information is available. All parameters are defined in the NCI Thesaurus OBO Edition controlled vocabulary (<https://ncit.nci.nih.gov/ncitbrowser/>) and their units in the Units of Measurement Ontology (<https://biportal.bioontology.org/ontologies/UO>). Deviations from the expected value in the experiment parameter (for example, within normal range, increased, decreased, not specified or

not relevant) can also be added. All information is annotated with the text description taken directly from the scientific article and curators’ statements can be added to further clarify annotation.

Experimental components definition. The DisProt database can now describe experimental sample components, such as lipids, nucleic acids, small molecules, metal ions or proteins present during the characterization of the structural state of an IDR. The concentration of the components and a cross-reference to the specific database, that is CheBI²⁴, ENA²⁵, RNAcentral²⁶ or UniProt¹⁵, can also be added. Similar to the other MIADe fields, a text description can be added into the corresponding Statement field.

A representative list of DisProt use cases highlighting novel information covered by the addition of fields from the MIADe update is provided in Table 3.

Case studies

Although MIADe captures only the core structural inferences derived from structural experiments on IDRs, it can be applied to the description of experimental data with a very wide range of complexity in terms of experimental design and studied system. In the following section, we demonstrate how MIADe-compliant information can be created using extracts from three papers that serve as examples of good practice. These experiments are accompanied by a MIADe-compliant entry in the DisProt resource (Table 3). We chose these papers to provide a set of examples of increasing complexity that represented several of the key issues tackled by the MIADe guidelines. A wide range of techniques are used to characterize the structural properties of IDRs; however, for simplicity, both owing to the available literature and the wider understanding of the experimental approach, all examples describe nuclear magnetic resonance (NMR) experiments. We highlight the three key areas covered by the MIADe guidelines from each paper: the definition of the protein construct used; the deviation from the wild-type proteoform (including mutations, post-translational modifications, tags, labels and dyes); and the definition of the experimental setup, including the environmental conditions and sample compositions that might have relevance for the structural state.

The first paper describes the disordered structural state of human calpastatin (CAST), an inhibitor of calpain, the Ca²⁺-activated cysteine protease²⁷. The authors unambiguously define two protein constructs that they used by referencing the common name of the protein and source organism, together with a UniProt accession (“¹⁵N-labeled and ¹³C-labeled full-length hCSD1 [corresponding to A137–K277 of human calpastatin, SwissProt entry P20810]” and “C-terminal half of calpastatin (position in whole calpastatin P204–K277)”). The constructs are defined by providing residue numbers in reference to the UniProt entry. However, the wording “C-terminal half of calpastatin” could be misleading, because the construct under investigation is the C-terminal region of the first domain of calpastatin. In addition, when providing UniProt residue start and stop numbering, the authors erroneously state that the construct is P204–K277 rather than P203–K277. This example highlights a common problem that stems from the custom of providing relative residue position within a region of interest or domain when defining constructs, instead of absolute residue position in reference to the full sequence. The authors clearly define the experimental method with different types of NMR experiments, including heteronuclear single quantum coherence (HSQC), calculation of the secondary chemical shift and ³J_{H_NH_α scalar coupling constants determined with 3D HNCA-based exclusive correlation spectroscopy (E.COSY). For these experiments, the relevant environmental conditions are temperature and pH, which the authors define in the materials and methods sections (“HSQC spectra collected at 298 K and at pH 4.3, 5.23, and 6.17 for hCSD1(67–141) as well as pH 3.85, 5.53, 6.07, and 7.25 for hCSD1. The temperature dependence of the same type of resonances was measured at 280, 300, and 320 K in aqueous solution for}

Table 3 | Extra data curated by DisProt for MIADe-compliant annotation for the case study examples

Protein definition	Observed region (for which structural observations are made)	Construct region (present in the experiment)	DisProt Identifier	MIADe field	Relevant parameters controlled in the experiment	Experimental method	Details	Structural state	Ref.
Calpastatin, UniProt: P20810	137–277	Same as observed region	DPO0196	r011 - Experimental condition	pH, temperature	NMR	pH = 3.85–7.25 T = 298 K	Disordered	27
	204–277	Same as observed region	DPO0196	r011 - Experimental condition	pH, temperature	NMR	pH = 4.3–6.17 T = 280–320 K	Disordered	27
eIF4E-binding protein 2, UniProt: Q13542	1–120	Same as observed region	DPO1293	–	–	–	–	Disordered	10
	1–120	Same as observed region	DPO1293	r007 - Construct alteration	Protein modification	NMR	Phosphorylated Ser65 (phospho-Ser65); phospho-Thr70; phospho-Ser83	Disordered	
	19–62	1–120	DPO1293	r007 - Construct alteration	Protein modification	NMR	phospho-Thr37	Molten globule	
	19–62	1–120	DPO1293	r007 - Construct alteration	Protein modification	NMR	phospho-Thr46	Molten globule	
	19–62	1–120	DPO1293	r007 - Construct alteration	Protein modification	NMR	phospho-Thr37; phospho-Thr46	Ordered	
	19–62	1–120	DPO1293	r007 - Construct alteration	Protein modification	NMR	phospho-Thr37; phospho-Thr46; phospho-Ser65; phospho-Thr70; phospho-Ser83	Ordered	
Cellular tumor antigen p53, UniProt: P04637	1–61	1–393	DPO0086	r077; r078 - Construct alteration	Labels and dyes	NMR	¹⁵ N label position: 1–61		
				r077 - Experimental components	Salt concentration		[NaCl] = 150–500 mM		
				r081; r082 - Construct alteration	Protein mutation	NMR	Insertion, p.Asp61_Glu62ins GlySerCysPhe AsnGlyThr; Substitutions, p.Met133Leu, p.Val203Ala, p.Asn239Tyr, p.Asn268Asp	Disordered	28

hCSD1(67–141)”; the authors use relative numbering inside the domain being studied instead of the absolute numbering in the full-length UniProt sequence, which would be 203–277). Using these setups, the authors then determine that both constructs are essentially disordered and that this observation is largely independent of temperature and pH in the ranges explored. The manuscript also includes more refined observations about the structural properties of the protein, such as: “subdomains A and B, two characteristic binding and functional sites of the inhibitor, have some helical character” or “restricted motions on a subnanosecond time scale indicated by larger than average J(O) values are observed for G13-M17, K68-L72, S101-C105, and S128-V132. These residues of restricted mobility also present some residual local structural features highlighted both by secondary chemical shifts, SCS, and by their hydrophobicity pattern.”

The second paper details experiments performed on eukaryotic translation initiation factor 4E-binding protein 2 (EIF4EBP2), an interacting partner of eukaryotic translation initiation factor 4E (eIF4E)¹⁰. The authors define the protein construct as the full-length

human protein by referencing its common name (4E-BP2). The HUGO Gene Nomenclature Committee (HGNC) gene name is *EIF4EBP2*, and no unambiguous identifier is provided; however, the naming is specific enough to unambiguously identify the protein being studied, given that the protein has no known alternative isoforms. In addition, throughout the paper, the authors reference several key residues in the protein (such as T37, T46, S65, T70 and S83), which readers and curators can use as a basis to confirm whether they map to the correct UniProt sequence. As opposed to the previous example in which conditions were changed, in this case, measurements were performed on distinct proteoforms of the protein. The main structural conclusion of the paper is that the structural state of EIF4EBP2 is dependent on its phosphorylation state. The HSQC NMR spectrum shows that “non-phosphorylated 4E-BP2 has intense peaks with narrow ¹HN chemical shift dispersion characteristic of IDPs [...] However, wild-type 4E-BP2 uniformly phosphorylated at T37, T46, S65, T70 and S83 shows widespread downfield and upfield chemical shifts for residues spanning T19–R62, suggesting folding upon phosphorylation.” Using partial

Table 4 | Representative set of databases for the submission of IDR experimental metadata and data

Data type	Database	Website	Submission Process	Ref.
IDR experiment metadata	DisProt: database of disordered regions manually curated from the literature	https://www.disprot.org/	Deposition, curation post-publication	1
NMR	BMRB (Biological Magnetic Resonance Bank): database of data on NMR spectroscopic investigations of biological macromolecules and metabolites	https://bmrbo.io/	Deposition	53
Circular dichroism	PCDDDB (Protein Circular Dichroism Data Bank): database of circular dichroism (CD) and synchrotron radiation CD (SRCD) spectral data and their associated experimental metadata	https://pcddb.cryst.bbk.ac.uk/	Deposition	54
Small-angle X-ray and neutron scattering (SAXS and SANS)	SASBDB (Small-Angle Scattering Biological Data Bank): database of small-angle scattering (SAS) experimental data and models	https://www.sasbdb.org/	Deposition	55
Protein ensembles	PED (The Protein Ensemble Database): database for the deposition of structural ensembles	https://proteinensemble.org/	Curation post-publication	56

phosphorylation, the authors then disentangle the individual contribution of each phosphorylation to the induced folding, stating: “No significant change in global dispersion was observed for 4E-BP2 phosphorylated only at S65/T70/S83, demonstrating that it remains disordered, while phosphorylating T37 and T46 (pT37pT46) induces a 4E-BP2 fold identical to phosphorylated wild type. Interestingly, when phosphorylated individually, pT37 or pT46 result in a partly folded state, with some chemical shift changes indicative of ordered structure (pT37). [...] Thus, phosphorylation of both T37 and T46 is necessary and sufficient for phosphorylation-induced folding of 4E-BP2.” The authors also measure the structural effect of binding to eIF4E and find that the interaction induces partial folding of the phosphorylated 4E-BP2: “The spectrum of pT37pT46 in isolated and eIF4E-bound states demonstrate an order-to-disorder transition upon eIF4E binding. [...] pT37pT46 undergoes an order-to-disorder transition upon binding to eIF4E.” Therefore, both phosphorylation and the presence of a binding partner can induce a structural transition of EIF4EBP2 through different mechanisms, and therefore the inference that EIF4EBP2 is disordered is dependent on the exact proteoform as well as the presence of other proteins. In addition to the structural state, the authors also directly address the connection between phosphorylation and the interaction capacity: “non-phosphorylated or minimally phosphorylated 4E-BPs interact tightly with eIF4E, while the binding of highly phosphorylated 4E-BPs is much weaker and can be outcompeted by eIF4G.” Although this piece of information is key to understanding the biological regulatory role of EIF4EBP2, it cannot be captured in the structural-state-focused framework of MIADE, and should be encoded as additional information in interactomics databases.

In the third example, the authors study the human cellular tumor antigen p53 (TP53), focusing on the structural features of the disordered N-terminal region²⁸. The authors clearly define the protein being studied by stating it is human TP53. In addition, they also provide an overview figure that contains the UniProt region boundaries of various p53 regions and domains that are used in the constructs. In contrast to the previous examples, the main construct used in this study is not a full-length protein or an isolated protein region, but a chimeric protein consisting of an isotopically labeled N-terminal and a non-labeled C-terminal region. The authors use a split intein splicing to produce the isotopically labeled disordered N-terminal region and fused to the unlabeled central C-terminal regions (“we utilized intein splicing to segmentally label the NTAD within tetrameric p53 [...] NTAD (residues 1–61) labeled with an NMR-active isotope (¹⁵N), while residues 62–393 remained unlabeled and NMR invisible”). As a result of this technique, the final construct has a short insertion where the intein was located, the position of which was carefully chosen: “The intein splice site was

selected as D61/E62, a site that is distant in the amino acid sequence from interaction sites or well-folded domains. Careful selection of the splice site is important, since the Npu DnaE intein system inserts nonnative residues (GSCFNGT in the p53 constructs used here) at the splice site.” This construct enables the assessment of the structural state of the disordered NTAD in the context of the full-length tetrameric TP53 by NMR HSQC spectra. For technical reasons, the authors further introduced mutations to the sequence outside the disordered regions being studied: “To improve expression levels, stabilizing mutations (M133L/V203A/N239Y/N268D) were introduced into the DNA-binding domain.” The definition of the environmental conditions covers the temperature and salt concentrations, with all other parameters in the normal range of similar NMR measurements: “unless otherwise stated, all spectra were recorded at 25 °C for samples in NMR buffer” and “salt titrations for p53(1–312) and p53(1–61) were carried out with protein concentrations of 150 μM. The initial titration point had a NaCl concentration of 150 mM, and NaCl from a 5-M concentrated stock was added to this sample at 50-mM increments up to 500 mM NaCl.” Apart from unambiguously defining the protein construct, the proteoform, the techniques and the environmental conditions, the main conclusion about the structural state is also clearly stated as: “the HSQC spectrum of the NTAD-p53 tetramer shows that the NTAD remains dynamically disordered in the full-length protein.”

MIADE-compliant metadata capture at source

To date, direct submission of data to community resources is underused by the IDP community. IDP resources should improve their capacity to receive data pre-publication, including the possibility to embargo data until the time of final publication (similar to the PDB model) and develop tools and resources that simplify MIADE-compliant reporting. Furthermore, the IDP community should enforce the deposition of experimental data and metadata as a required component of the publication process. The ideal situation would include the pre-publication submission of primary source data directly to the corresponding field-specific resource (Table 4). Subsequently, a reference to primary source data and MIADE-compliant experimental metadata should then be submitted to a community resource such as DisProt or IDEAL^{1,2}. This benefits the databases, as the efficiency of data collection and verification is increased. This in turn benefits the IDP community and wider biological community, as more and more precise data, linked to related primary data in field-specific databases, are readily available. Currently, several databases allow pre- or post-publication submission of data related to IDR experiments, each with their own submission process and data formats (Table 4). However, the proportion of data created that is captured by these resources varies widely, and no

resource is successful in capturing all data produced that fall within its scope. To facilitate data capture, as part of this work, the DisProt resource has added a MIAD-compliant form for the submission of metadata from experiments structurally characterizing IDRs (<https://disprot.org/biocuration>).

Discussion

Over the past 10 years, the development of new and improved methods and technologies to study IDPs has increased the complexity of the experiments characterizing the structural properties of IDRs (Fig. 1c). However, this revolution has not been reflected by advances in the data standardization of the field. Consequently, at all levels, improvements in the description, curation, storage and dissemination of the fundamental data from these analyses are needed. Guidelines for unambiguous definition of the key information from an experiment simplify data capture, minimize key data loss, standardize data transfer and maximize data use. The argument against standardized reporting guidelines has always been the unbalanced burden placed on the reporter. However, the advantages far outweigh the effort, allowing relevant data to be easily identified, recovered and reused, leading to improved data management, minimized data loss and simplification of data sharing within and between groups. Method-independent metadata also allow data to be aggregated and to be analyzed in subsets based on data quality (Fig. 1d). Furthermore, data aggregation across complementary methods simplifies cross-validation of data, permitting quality to be defined by consensus. Finally, improved data management and upgrades to data-deposition processes will improve data transfer to community resources, accelerating the open-science efforts of the IDP field.

Data capture should have the flexibility to cover old, new and future experimental approaches. The MIAD guidelines store observations together with details on the experiment to allow data to be reinterpreted in the future. While adding experimental parameters and sample components can add considerably to the curation burden, they also allow for more nuanced observations to be captured. As IDP experiments become increasingly complex by studying the modulatory effects of proteoforms, concentrations, conditions and binding partners, it is imperative that these rich data on the context of the studied protein regions are captured wherever they are needed to faithfully interpret the reported observations. These details can describe observations beyond binary order and disorder structure definition, to quantitative measures that include dynamics, secondary structure propensity and compaction. To capture every relevant detail, the MIAD guidelines will need to evolve over time on the basis of community requirements. Controlled vocabularies and ontologies are a key component of this evolution. These definitions standardize the meaning of the terms used to describe IDP data, allowing the complete, unambiguous annotation of an IDP experiment and its results. Ontologies such as the Intrinsically Disordered Proteins Ontology (IDPO) and the Evidence and Conclusion Ontology (ECO) will need to continually add terms as required to include novel experimental approaches, computational methods, non-binary structural classifications (that is more detailed than order and disorder, including dynamics, secondary structure propensity and compaction), structural transition definitions and conditionality.

The MIAD guidelines are only an initial step towards standardized and lossless IDP data representation within the biological community. Three key developments are still required: standardized exchange formats for reporting IDP metadata and raw data, simplified pre- and post-publication data-deposition mechanisms for the IDP data repositories, and a community-wide agreement to deposit data. The diversity of the methodologies and data in the IDP community has proved to be a barrier to data collection, and MIAD will allow the key data to be collected and aggregated across the field. In parallel, each experimental approach in the field can develop method-specific storage and exchange formats and standards for raw data. However, given

the parallel requirements across many of these approaches, efforts should be made to collaborate and reuse structured data formats when possible. These exchange formats should hold experimental data at a range of detail from a MIAD-compliant definition to a description of the experiment and results that would allow the experiment to be reproduced (Fig. 1b). Ultimately, the interpretation of raw experimental data will evolve as analysis methods improve. Consequently, the best long-term strategy to safeguard the knowledge accumulated by the IDP community is the standardized deposition of raw and processed experimental measurements in addition to interpreted structural observations derived from the data. Enforcing data deposition is a complex process; however, pressure at the point of publication by journals and reviewers can drive compliance.

We see this document as one of the initial steps to open the discussion to standardize the controls, experimental parameters and vocabulary for each method used by the IDP community. We advocate for the importance of a clear and unambiguous description of an IDR experiment, and we hope this document will encourage each experimental community to extend the guidelines to specify and enforce the reporting of the important information for their experimental methods. It is important that data producers, curators and database developers in the IDP field are conscious of the expanding interest in IDRs by the wider biological community. The growing understanding of the functional significance of IDPs by researchers outside the IDP field has increased the importance of making high quality and understandable IDP data accessible to the wider community such as cell biologists studying the function of IDRs, computational biologists developing tools to analyze IDRs and curators transferring IDR data into community resources.

References

1. Quaglia, F. et al. DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res.* **50**, D480–D487 (2021).
2. Fukuchi, S. et al. IDEAL in 2014 illustrates interaction networks composed of intrinsically disordered proteins and their binding partners. *Nucleic Acids Res.* **42**, D320–D325 (2014).
3. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
A highly influential review describing the structure–function relationship in the context of IDRs.
4. Davey, N. E. The functional importance of structure in unstructured protein regions. *Curr. Opin. Struct. Biol.* **56**, 155–163 (2019).
5. Schadt, E. et al. DIBS: a repository of disordered binding sites mediating interactions with ordered proteins. *Bioinformatics* **34**, 535–537 (2018).
6. Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **579**, 3346–3354 (2005).
7. Jakob, U., Kriwacki, R. & Uversky, V. N. Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chem. Rev.* **114**, 6779–6805 (2014).
8. Mitrea, D. M. & Kriwacki, R. W. Regulated unfolding of proteins in signaling. *FEBS Lett.* **587**, 1081–1088 (2013).
An excellent review introducing concepts related to conditional IDRs and their functional role.
9. Borgia, A. et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61–66 (2018).
10. Bah, A. et al. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**, 106–109 (2015).
A landmark study on the phospho-dependent folding of IDP EIF4EBP2, providing an exemplar for future studies on conditional IDRs.
11. Davey, N. E. et al. An intrinsically disordered proteins community for ELIXIR. *F1000Res* **8**, ELIXIR-1753 (2019).

12. Taylor, C. F. et al. The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* **25**, 887–893 (2007).
13. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
14. Orchard, S. et al. The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat. Biotechnol.* **25**, 894–898 (2007).
- Minimum information standards for the molecular interaction experiment that provided the framework for the development of the MIAPE guidelines.**
15. UniProt Consortium. UniProt: the universal protein knowledge-base in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
16. Giglio, M. et al. ECO, the Evidence & Conclusion Ontology: community standard for evidence information. *Nucleic Acids Res.* **47**, D1186–D1194 (2019).
17. Bruford, E. A. et al. Guidelines for human gene nomenclature. *Nat. Genet.* **52**, 754–758 (2020).
18. A one-letter notation for amino acid sequences. *Pure Appl. Chem.* **31**, 641–645 (1972).
19. Moses, D. et al. Revealing the hidden sensitivity of intrinsically disordered proteins to their chemical environment. *J. Phys. Chem. Lett.* **11**, 10131–10136 (2020).
20. Speer, S. L. et al. The intracellular environment affects protein-protein interactions. *Proc. Natl Acad. Sci. USA* **118**, e2019918118 (2021).
21. Sivade Dumousseau, M. et al. Encompassing new use cases — level 3.0 of the HUPO-PSI format for molecular interactions. *BMC Bioinf.* **19**, 134 (2018).
22. Montecchi-Palazzi, L. et al. The PSI-MOD community standard for representation of protein modification data. *Nat. Biotechnol.* **26**, 864–866 (2008).
23. Nadendla, S. et al. ECO: the Evidence and Conclusion Ontology, an update for 2022. *Nucleic Acids Res.* **50**, D1515–D1521 (2022).
24. Hastings, J. et al. ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, D1214–D1219 (2016).
25. Leinonen, R. et al. The European Nucleotide Archive. *Nucleic Acids Res.* **39**, D28–D31 (2011).
26. RNAcentral Consortium. RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res.* **49**, D212–D220 (2021).
27. Kiss, R., Kovács, D., Tompa, P. & Perczel, A. Local structural preferences of calpastatin, the intrinsically unstructured protein inhibitor of calpain. *Biochemistry* **47**, 6936–6945 (2008).
28. Krois, A. S., Dyson, H. J. & Wright, P. E. Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. *Proc. Natl Acad. Sci. USA* **115**, E11302–E11310 (2018).
29. Lalli, D., Rosa, C., Allegrozzi, M. & Turano, P. Distal unfolding of ferricytochrome c induced by the F82K mutation. *Int. J. Mol. Sci.* **21**, 2134 (2020).
30. Thielges, M. C., Chung, J. K., Axup, J. Y. & Fayer, M. D. Influence of histidine tag attachment on picosecond protein dynamics. *Biochemistry* **50**, 5799–5805 (2011).
31. Yamaguchi, Y. et al. NMR characterization of intramolecular interaction of osteopontin, an intrinsically disordered protein with cryptic integrin-binding motifs. *Biochem. Biophys. Res. Commun.* **393**, 487–491 (2010).
32. Scharnagl, C., Reif, M. & Friedrich, J. Stability of proteins: temperature, pressure and the role of the solvent. *Biochim. Biophys. Acta* **1749**, 187–213 (2005).
33. Padan, E. et al. NhaA of *Escherichia coli*, as a model of a pH-regulated Na⁺/H⁺ antiporter. *Biochim. Biophys. Acta* **1658**, 2–13 (2004).
34. Padan, E., Bibi, E., Ito, M. & Krulwich, T. A. Alkaline pH homeostasis in bacteria: new insights. *Biochim. Biophys. Acta* **1717**, 67–88 (2005).
35. Haslbeck, M. et al. Hsp26: a temperature-regulated chaperone. *EMBO J.* **18**, 6744–6751 (1999).
36. Stromer, T., Fischer, E., Richter, K., Haslbeck, M. & Buchner, J. Analysis of the regulation of the molecular chaperone Hsp26 by temperature-induced dissociation: the N-terminal domain is important for oligomer assembly and the binding of unfolding proteins. *J. Biol. Chem.* **279**, 11222–11228 (2004).
37. Silva, J. L., Foguel, D. & Royer, C. A. Pressure provides new insights into protein folding, dynamics and structure. *Trends Biochem. Sci.* **26**, 612–618 (2001).
38. Peng, X., Jonas, J. & Silva, J. L. Molten-globule conformation of Arc repressor monomers determined by high-pressure ¹H NMR spectroscopy. *Proc. Natl Acad. Sci. USA* **90**, 1776–1780 (1993).
39. Greene, D. N. et al. Single-molecule force spectroscopy reveals a stepwise unfolding of *Caenorhabditis elegans* giant protein kinase domains. *Biophys. J.* **95**, 1360–1370 (2008).
40. Wood, M. J., Storz, G. & Tjandra, N. Structural basis for redox regulation of Yap1 transcription factor localization. *Nature* **430**, 917–921 (2004).
41. Rubinstenn, G. et al. Structural and dynamic changes of photoactive yellow protein during its photocycle in solution. *Nat. Struct. Biol.* **5**, 568–570 (1998).
42. Kar, M., Posey, A. E., Dar, F., Hyman, A. A. & Pappu, R. V. Glycine-rich peptides from FUS have an intrinsic ability to self-assemble into fibers and networked fibrils. *Biochemistry* **60**, 3213–3222 (2021).
43. Vagenende, V., Yap, M. G. S. & Trout, B. L. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry* **48**, 11084–11096 (2009).
44. Ribeiro, S., Ebbinghaus, S. & Marcos, J. C. Protein folding and quinary interactions: creating cellular organisation through functional disorder. *FEBS Lett.* **592**, 3040–3053 (2018).
45. Wingfield, P. T. Use of protein folding reagents. *Curr. Protoc. Protein Sci.* **Appendix 3**, Appendix 3A (2001).
46. Theillet, F.-X. et al. Structural disorder of monomeric α -synuclein persists in mammalian cells. *Nature* **530**, 45–50 (2016).
47. Fonseca-Ornelas, L. et al. Altered conformation of α -synuclein drives dysfunction of synaptic vesicles in a synaptosomal model of Parkinson's disease. *Cell Rep.* **36**, 109333 (2021).
48. Lacy, E. R. et al. p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat. Struct. Mol. Biol.* **11**, 358–364 (2004).
49. Huth, J. R. et al. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.* **4**, 657–665 (1997).
50. Léger, A. et al. PiP2 favors an α -helical structure of non-recombinant Hsp12 of *Saccharomyces cerevisiae*. *Protein Expr. Purif.* **181**, 105830 (2021).
51. Wang, N. et al. Ordering a dynamic protein via a small-molecule stabilizer. *J. Am. Chem. Soc.* **135**, 3363–3366 (2013).
52. O'Brien, D. P. et al. Calcium-dependent disorder-to-order transitions are central to the secretion and folding of the CyaA toxin of *Bordetella pertussis*, the causative agent of whooping cough. *Toxicon* **149**, 37–44 (2018).
53. Ulrich, E. L. et al. BioMagResBank. *Nucleic Acids Res.* **36**, D402–D408 (2008).
54. Ramalli, S. G., Miles, A. J., Janes, R. W. & Wallace, B. A. The PCDDb (protein circular dichroism data bank): a bioinformatics resource for protein characterisations and methods development. *J. Mol. Biol.* **434**, 167441 (2022).
55. Kikhney, A. G., Borges, C. R., Molodenskiy, D. S., Jeffries, C. M. & Svergun, D. I. SASBDB: towards an automatically curated and validated repository for biological scattering data. *Protein Sci.* **29**, 66–75 (2020).

56. Lazar, T. et al. PED in 2021: a major update of the protein ensemble database for intrinsically disordered proteins. *Nucleic Acids Res.* **49**, D404–D411 (2021).

Acknowledgements

The authors would like to thank J. Forman-Kay for her feedback on the MIADE guidelines and the manuscript. This work was funded by ELIXIR, the research infrastructure for life-science data; a Cancer Research UK Senior Cancer Research Fellowship (C68484/A28159 to N.E.D.); a Carlsberg Foundation Distinguished Fellowship (CF18-0314); Danmarks Grundforskningsfond (DNRF125); National Research, Development and Innovation (NRDI) Fund Young Researchers' Excellence Programme research grant (project FK128133 and FK142285 to R.P.); the European Union's H-2020 MSCA-RISE programme (grant agreement No. 778247 'IDPfun'); Fondazione CR Firenze; Fondazione Umberto Veronesi; the Italian Ministry of University and Research (to E.M. and R.P.), PRIN 2017 under grant agreement no. 2017483NH8; EC H2020-WIDESPREAD-2020-5 Twinning grant (no. 952334 PhasAGE, to P.T. and S.C.E.T.); the VUB Strategic Research Program on Microfluidics (SRP51) at Vrije Universiteit Brussel (VUB, Brussels, Belgium, to P.T.) and a US National Institutes of Health grant (GM109046 to J.C.H.).

Author contributions

All authors contributed to the development of the MIADE guidelines. N.E.D., B.M., A.H., N.P., F.Q., E.S. and K.V.R. were the MIADE workgroup and defined the details of the MIADE guidelines. N.E.D. and B.M. wrote

the document with the help of the MIADE workgroup. All authors edited and refined the final manuscript. N.E.D. coordinated the project.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Norman E. Davey.

Peer review information *Nature Methods* thanks Jane Dyson, Xavier Salvatella, and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editor: Arunima Singh, in collaboration with the *Nature Methods* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature America, Inc. 2023

¹Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. ²Department of Biomedical Sciences, University of Padova, Padova, Italy. ³Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes - CONICET, Bernal, Buenos Aires, Argentina. ⁴Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council (CNR-IBIOM), Bari, Italy. ⁵Interuniversity Institute of Bioinformatics in Brussels, ULB-VUB, Brussels, Belgium. ⁶Structural Biology Brussels, Vrije Universiteit Brussel, Brussels, Belgium. ⁷Harvard Medical School (HMS), Boston, MA, USA. ⁸Department of Cancer Biology, Dana-Farber Cancer Institute (DFCI), Boston, MA, USA. ⁹Department of Biochemistry, Eötvös Loránd University, Budapest, Hungary. ¹⁰Department of Chemistry 'Ugo Schiff' and Magnetic Resonance Center, University of Florence, Sesto Fiorentino (Florence), Italy. ¹¹Department of Molecular Biology and Biophysics, UConn Health, Farmington, CT, USA. ¹²European Molecular Biology Laboratory (EMBL), Hamburg Unit, c/o Deutsches Elektronen-Synchrotron, Hamburg, Germany. ¹³Laboratory Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, Aix Marseille University and Centre National de la Recherche Scientifique (CNRS), Marseille, France. ¹⁴Cancer Structural Biology, Danish Cancer Society Research Center, Copenhagen, Denmark. ¹⁵UniCamillus - Saint Camillus International University of Health and Medical Sciences, Rome, Italy. ¹⁶European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Hinxton, UK. ¹⁷Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary. ¹⁸Cancer Systems Biology, Section for Bioinformatics, Department of Health and Technology, Technical University of Denmark, Lyngby, Denmark. ¹⁹Hospital for Sick Children, Toronto, Ontario, Canada. ²⁰Medical University of Graz, Graz, Austria. ²¹VIB-VUB Center for Structural Biology, Brussels, Belgium. ²²Structural Biology Brussels, Department of Bioengineering Sciences, Vrije Universiteit Brussel, Brussels, Belgium. ²³Division Of Cancer Biology, Institute of Cancer Research, Chester Beatty Laboratories, Chelsea, London, UK. ²⁴Present address: Department of Structural Biology and Center for Data Driven Discovery, St Jude Children's Research Hospital, Memphis, TN, USA. ²⁵Present address: Department of Oncology, Lausanne University Hospital, Lausanne, Switzerland. ²⁶Present address: Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. ²⁷Present address: Swiss Institute of Bioinformatics, Lausanne, Switzerland. ²⁸Present address: Swiss Cancer Center Leman, Lausanne, Switzerland. ²⁹These authors contributed equally: András Hatos, Nicolas Palopoli, Federica Quaglia, Edoardo Salladini, Kim Van Roey. ✉e-mail: norman.davey@icr.ac.uk