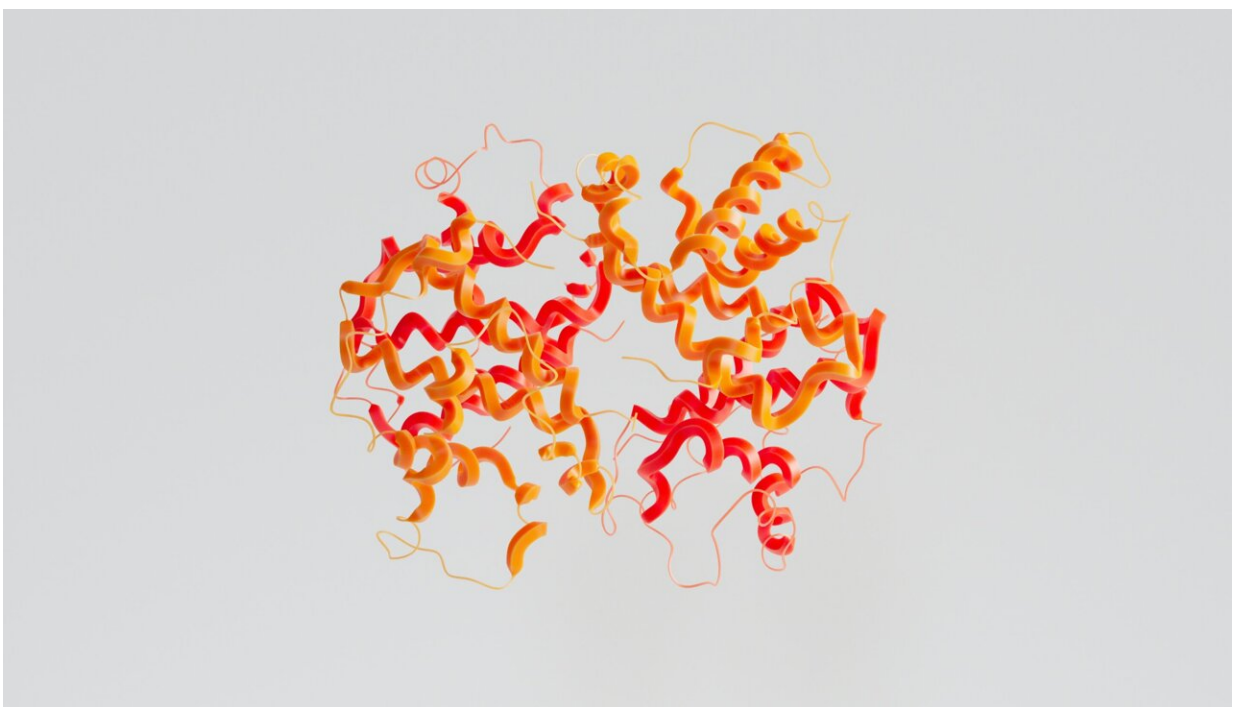


Biologists publish new guidelines to facilitate data sharing of research on disordered proteins

July 14 2023, by Dorota Badowska and Oana Stroe



For decades, structural biologists have been working on cracking the molecular 3D structures of proteins to understand their function. But what if a protein doesn't have a fixed structure? For molecules that keep changing their shape all the time, both research and sharing the findings

within the scientific community can be complicated. EMBL scientists have contributed to new guidelines that will make the data sharing part more efficient. The research is published in the journal *Nature Methods*.

Essentially, proteins are strings of amino acids, many of which fold like origami into a 3D structure. However, some proteins "prefer" to remain as a wobbly string similar to cooked spaghetti (ignoring the fact that spaghetti is mainly made of carbs). In fact, around a third of all known proteins are either completely or partially spaghetti-like.

This, however, doesn't mean they don't serve a function. Quite the contrary. This added flexibility gives proteins various abilities, such as adapting their own shape to the shape of other molecules. This way, they can interact with more diverse molecules, and thereby take part in a larger number of cellular processes than a protein with a rigid structure could.

Understanding unstructured proteins—also known as "intrinsically disordered proteins"—is important, because they are involved in many disease processes, such as cancer, neurodegeneration, and viral infection.

Making protein data meaningful

Scientific data, including that related to disordered proteins, are most useful to the community when they can be reanalyzed and integrated with other datasets to explore new research questions. To enable this, data should be accurately described and openly accessible. This is usually achieved by submitting data to public data resources, such as the ones managed by EMBL-EBI. Some of the most used protein data resources include UniProt for protein sequences and Protein Data Bank in Europe (PDBe) for protein structures.

The scientific community has already produced a wide range of

guidelines to ensure scientists include useful information alongside their [research data](#). Now, for the first time, EMBL and collaborators have developed such guidelines for disordered protein data.

Called "Minimum Information About a Disorder Experiment," or MIADE, this set of guidelines is aimed at anyone working on disordered proteins, to help them share their data in a useful manner. This open and shared framework is set to help protein scientists increase protein data mining and interoperability.

"Besides defining the minimum amount of information about an experiment needed to make the results meaningful for other scientists, we also define how to report this information," said Bálint Mészáros, former postdoctoral researcher in the Gibson Group at EMBL Heidelberg and a first author of the paper. "In essence, we develop a common language that can be used by the community to make communication unambiguous."

Tackling data loss

"It's very frustrating when you read a paper that describes great science, but you can't make full sense of the data because something really important is missing," explained Sandra Orchard, EMBL-EBI Team Leader for Protein Function Content. "Most of the time, the additional information exists, but the authors overlook the need to share it. It sounds silly, but one of the biggest data losses happens because submitters don't say what species the protein they are working on is from."

As the community adopts MIADE, more data should start getting through to public databases. This will allow researchers across the world to access information on related proteins and families of proteins they are interested in and compare their data with those of other labs.

MIADE should "tidy up" disordered protein research and make it more understandable for new people entering the field.

The structural characteristics of intrinsically disordered protein systems can be studied using various experimental techniques, including small angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS). SASBDB, the database for SAXS and SANS, is maintained and curated by the EMBL Hamburg's SAXS Team, which contributed to developing the MIADE guidelines.

"It's essential that scientific results are shared; otherwise they might end up as 'undiscovered-discoveries,'" said Cy Jeffries, Staff Scientist in the SAXS Team at EMBL Hamburg and co-author of the guidelines. "It was fantastic to work together with a diverse community of scientists, software engineers, programmers, and data resource managers. MIADE is a step towards ensuring scientists and data resources can communicate much more easily using a baseline set of terms and ideas that we (and computers) can all recognize."

MIADE will also help enable using [artificial intelligence](#) for new discoveries on disordered proteins. The availability of vast, standardized data is crucial for training machine learning and artificial intelligence tools. With sufficient training data, researchers could develop machine learning tools to help predict new disordered proteins, interpret the effects of protein modifications, identify interacting regions, and much more.

A community effort

The MIADE guidelines provide a systematic framework to share experimental definitions that, besides SASBDB, will also benefit many other databanks, such as BMRB (for Nuclear Magnetic Resonance, NMR), PCDDDB (for circular dichroism spectral data) and Protein

Ensemble Database (PED). This is also important for forwarding and contextualizing experimental data to "higher up" bioinformatic resources like DisProt and other [protein](#) structural knowledge bases, like those developed at the PDBe.

The MIADE guidelines were developed by scientists from more than 20 institutions in 11 countries. The work was led by the Institute of Cancer Research in London, U.K.

More information: Bálint Mészáros et al, Minimum information guidelines for experiments structurally characterizing intrinsically disordered protein regions, *Nature Methods* (2023). [DOI: 10.1038/s41592-023-01915-x](#)

Provided by European Molecular Biology Laboratory

Citation: Biologists publish new guidelines to facilitate data sharing of research on disordered proteins (2023, July 14) retrieved 29 April 2024 from <https://phys.org/news/2023-07-biologists-publish-guidelines-disordered-proteins.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--