Making fluorescence-based integrative structures and associated kinetic information accessible

Authors: Christian A. Hanke[1], John D. Westbrook[2,3,#], Benjamin M. Webb[4], Thomas-O. Peulen[1,4,5] Catherine L. Lawson[2], Andrej Sali[4], Helen M. Berman[2,6,7], Claus A.M. Seidel[1,*], Brinda Vallat[2,3,*]

[1]Molecular Physical Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

[2]Research Collaboratory for Structural Bioinformatics Protein Data Bank and the Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

[3]Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA

[4]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, University of California at San Francisco, San Francisco, California, USA

[5]Current address: Rudolf-Virchow-Zentrum - Center for Integrative and Translational Bioimaging, 97080 Würzburg, Germany

[6]Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

[7]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA

*Corresponding authors: brinda.vallat@rcsb.org, cseidel@hhu.de

[#]Deceased

Fluorescence- and, in particular, Förster Resonance Energy Transfer (FRET) - experiments provide rich insights into biomolecular systems, including structural and dynamic information, and are increasingly used for integrative structure modeling of biomolecules [1-3] (see Supplementary Note 1 for an overview).

In integrative structure modeling approaches, data from different experimental and theoretical techniques are combined to maximally utilize method-specific advantages and complementary information (Supplementary Note 2) [4]. Making these results publicly available following the FAIR (Findable, Accessible, Interoperable, Reusable) principles - for example, in structural biology model repositories (Supplementary Note 3) - is crucial for advancing science [5, 6]. The prototype PDB-Dev repository (https://pdb-dev.wwpdb.org/) [7, 8], which was recently unified with the PDB archive, offers a framework for deposition, curation, validation, archiving, and dissemination of integrative structure models. PDB-Dev uses the Protein Data Bank Exchange / macromolecular Crystallographic Information Framework (PDBx/mmCIF) data standard [9], along with an extension developed for representing Integrative and Hybrid Models (IHMCIF; https://github.com/ihmwg/IHMCIF) [7, 10] (see Supplementary Note 4 for details on the

dictionaries). IHMCIF is supported by the python-ihm library (https://github.com/ihmwg/python-ihm), which can be used to read and write IHMCIF-compliant files for PDB-Dev, and provides an application programming interface (API) for other software.

To support the deposition and validation of fluorescence-based integrative structure models in PDB-Dev, we present flrCIF (https://github.com/ihmwg/flrCIF), an extension of PDBx/mmCIF and IHMCIF. flrCIF was developed as a method-specific extension dictionarythat allows seamless interoperation of fluorescence and FRET data with PDB and PDB-Dev data. Together with PDBx/mmCIF and IHMCIF, flrCIF enables deposition of structure models, restraint data, references to experimental raw data in other repositories, and additional metadata that support FAIR principles (Fig. 1a, see Supplementary Note 5 for data definitions).

In the current version, flrCIF covers four key workflow aspects from the fluorescence experiment to the structure model (Fig. 1b and Supplementary Note 6): (1) experiment and sample description including setup, conditions, and fluorescent probes with their attachment; (2) analysis workflow with correction parameters and reference measurements; (3) structure modeling procedure with FRET-derived distance restraints and modeling software settings and parameters; and (4) assessment of structure models on the basis of input experimental data.

While the current version of flrCIF is focused on FRET experiments, the dictionary is extensible and can be further expanded to include descriptions of other fluorescence experiments, such as fluorescence quenching assays or fluorescence anisotropy measurements used to probe the local dynamics, as well as other analysis approaches (see Supplementary Note 6 for extension plans).

In addition to creating fluorescence-specific definitions in flrCIF, we also extended the IHMCIF dictionary with generic definitions applicable to other probe-based methods. Importantly, a model in IHMCIF can be multi-state [10] (see Supplementary Note 7 for details on molecular kinetics and multi-state models). This allowed us to introduce descriptions of complex kinetic schemes connecting multiple states, exchange kinetics, and intrastate dynamics characterizing flexibility. To quantify the kinetics, we included characteristic observables such as equilibrium constants, population fractions, relaxation times, and transition rate constants (see Supplementary Note 7 for details and an example). The deposition of fluorescence-aided structure models in PDB-Dev is outlined in Fig. 1a. PDB-Dev deposition requires an mmCIF file compliant with flrCIF, as well as the parent PDBx/mmCIF and IHMCIF dictionaries. To facilitate the preparation of such a file, we extended the python-ihm library to support the new definitions in flrCIF (Supplementary Note 8).

To facilitate the submission of experimental information that may be conveniently collected in a spreadsheet (see Fig. 1 for content categories with more details in Supplementary Note 8), we also created a converter script (flr2mmcif), which uses the python-ihm library, to convert the data captured in a spreadsheet to flrCIF and IHMCIF-compliant files for submission to PDB-Dev (https://github.com/Fluorescence-Tools/flr2mmcif).

The development of flrCIF offers possibilities for different interconnected user groups and applications:

1. Researchers and data users:
    - For scientists performing integrative structure modeling using fluorescence data, flrCIF provides a standardized framework and tools for deposition of fluorescence-assisted models in PDB-Dev (examples in Supplementary Note 9).

- Data consumers benefit from standardized storage of fluorescence-based structure models and information, facilitating access and reuse in further studies.
- Both groups profit from links to metadata, protocols and experimental data in other repositories through existing definitions in PDBx/mmCIF and IHMCIF.

2. Software developers:
   - flrCIF and the extended python-ihm library provide a convenient framework for handling fluorescence-related information in software applications.
   - Access to structure models alongside experimental data supports the development and validation of new computational modeling and analysis approaches.
3. Broader scientific applications:
   - flrCIF establishes data definitions adaptable to other experimental methods, such as electron paramagnetic resonance spectroscopy, particularly descriptions for probe chemistry and attachment.
   - The inclusion of complex kinetic networks and intra-state dynamics in IHMCIF extends the applicability beyond fluorescence to various techniques used in integrative modeling studies and facilitates the collection and archiving of structural dynamics information.

Overall, flrCIF contributes towards dynamic structural biology by enabling standardized deposition, archiving, and dissemination of information on biomolecules in their functional context - from static structures towards dynamic structural networks and the connection of structures with energy landscapes and real time motions.

Code availability

The flrCIF dictionary is publicly available at https://github.com/ihmwg/flrCIF. The updated python-ihm library, which aids in software access to the relevant fluorescence-related content and for creating mmCIF files for PDB-Dev, is publicly available at https://github.com/ihmwg/python-ihm. The spreadsheet-based tool, which can be used to prepare dictionary-compliant mmCIF files that include fluorescence information, is publicly available at https://github.com/Fluorescence-Tools/flr2mmcif.

Acknowledgments

We thank M. Gabanyi for helpful discussions during the development of flrCIF. We are grateful to M. Gabanyi and S. Burley for supporting and hosting a stay of C.A.M.S. at the Institute for Quantitative Biomedicine at Rutgers.
Development of IHMCIF dictionary and PDB-Dev have been supported by United States National Science Foundation awards DBI-2112966, DBI-2112967, DBI-1756248, DBI-1756250, and DBI-1519158 to H.M.B., B.V., and A.S. A.S. was also funded by NIH / NIGMS grants R01GM083960 and P41GM109824. H.M.B. and B.V. acknowledge the subcontract from Heinrich Heine University, Düsseldorf based on European Research Council Advanced Grant 2014 hybridFRET (671208) for funding the development of flrCIF. This research was funded by the Deutsche

Forschungsgemeinschaft (DFG, German Research Foundation) CRC 1208 (ID 267205415, project A08) and SE 1195/17-1 as well as the European Research Council through the Advanced Grant 2014 hybridFRET (671208) to C.A.M.S.

Author contributions

C.A.H. implemented flrCIF definitions in python-ihm, created the flr2mmcif tool, and wrote the initial draft of the manuscript. B.V. and C.A.H. designed the categories for flrCIF. B.V. and J.D.W. ensured compatibility of flrCIF with IHMCIF and PDBx/mmCIF. B.M.W. assisted in implementing flrCIF in python-ihm. T.O.P., C.L.L, and A.S. were involved in discussions on design decisions for flrCIF. B.V., T.O.P. and C.A.M.S. developed and wrote the basic description of categories necessary to describe fluorescence experiments. B.V. implemented the flrCIF dictionary. C.A.H., B.V., and C.A.M.S. wrote the manuscript with contributions from all authors. C.A.M.S., C.L.L, and H.M.B. initiated the project. B.V., C.A.M.S., and H.M.B. planned and supervised the project.

Competing interests

The authors declare no competing interests.

References

1. Lerner, E., et al., *FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices.* eLife, 2021. **10**: p. e60416.
2. Agam, G., et al., *Reliability and accuracy of single-molecule FRET studies for characterization of structural dynamics and distances in proteins.* Nature Methods, 2023. **20**(4): p. 523-535.
3. Peulen, T.O., et al., *Integrative dynamic structural biology unveils conformers essential for the oligomerization of a large GTPase.* eLife, 2023. **12**: p. e79565.
4. Rout, M.P. and A. Sali, *Principles for integrative structural biology studies.* Cell, 2019. **177**(6): p. 1384-1403.
5. Sali, A., et al., *Outcome of the first wwPDB Hybrid/Integrative Methods Task Force workshop.* Structure, 2015. **23**(7): p. 1156-1167.
6. Berman, H.M., et al., *Federating structural models and data: Outcomes from a workshop on archiving integrative structures.* Structure, 2019. **27**(12): p. 1745-1759.
7. Vallat, B., et al., *Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules.* Structure, 2018. **26**(6): p. 894-904.
8. Vallat, B., et al., *New system for archiving integrative structures.* Acta Crystallographica Section D-Structural Biology, 2021. **77**: p. 1486-1496.
9. Westbrook, J.D., et al., *PDBx/mmCIF Ecosystem: Foundational semantic tools for structural biology.* Journal of Molecular Biology, 2022. **434**(11): p. 167599.
10. Vallat, B., et al., *IHMCIF: An extension of the PDBx/mmCIF data standard for integrative structure determination methods.* Journal of Molecular Biology, 2024: p. 168546.

Figures

Figure 1

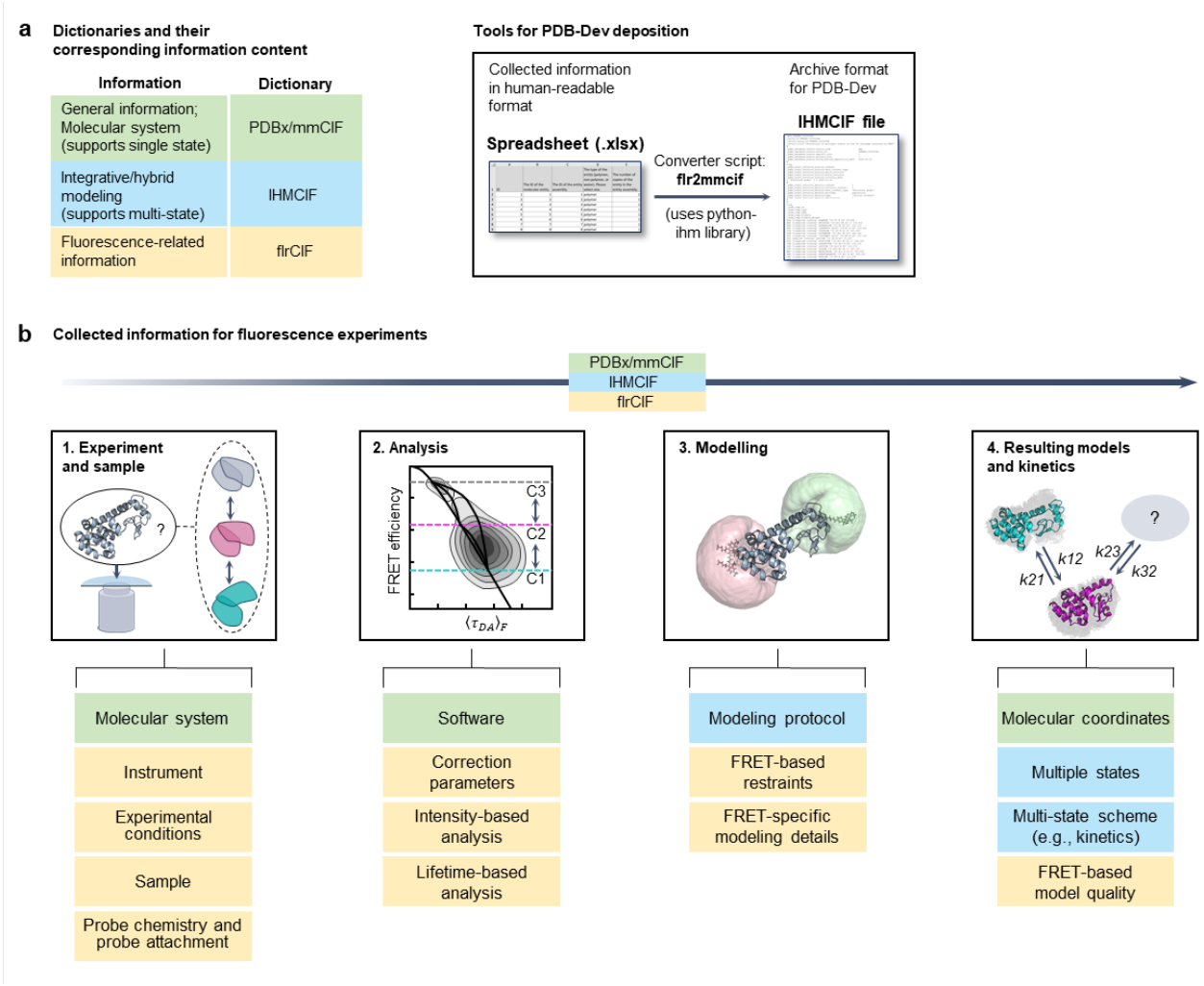


Figure legends

Figure 1: Information in dictionary-compliant file for a FRET-assisted integrative structure model. a) Information content of the different dictionaries supporting deposition of FRET-based structure models in PDB-Dev. PDBx/mmCIF describes general information about the entry, such as description of the molecular system and atomic coordinates as well as metadata regarding authors, software, and citations. IHMCIF contains definitions related to integrative modeling and descriptions of multi-scale, multi-state, and ordered models and collections of models. flrCIF provides definitions for fluorescence-experiment related information as shown in panel b. The flr2mmcif converter script uses a spreadsheet to collect information in a human readable format and converts this to a PDB-Dev compatible IHMCIF file using the python-ihm library (Supplementary Note 8). b) Data definitions for fluorescence and FRET-assisted structure

modeling. Components of the flrCIF dictionary complement the PDBx/mmCIF and IHMCIF dictionaries for the description of fluorescence-related information. The green boxes show definitions in PDBx/mmCIF, the blue boxes show definitions in IHMCIF, and the yellow boxes show definitions in flrCIF.

Making fluorescence-based integrative structures and associated kinetic information accessible

Authors: Christian A. Hanke[1], John D. Westbrook[2,3,#], Benjamin M. Webb[4], Thomas-O. Peulen[1,4,5] Catherine L. Lawson[2], Andrej Sali[4], Helen M. Berman[2,6,7], Claus A.M. Seidel[1,*], Brinda Vallat[2,3,*]

[1]Molecular Physical Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
[2]Research Collaboratory for Structural Bioinformatics Protein Data Bank and the Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[3]Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA
[4]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, University of California at San Francisco, San Francisco, California, USA
[5]Current address: Rudolf-Virchow-Zentrum - Center for Integrative and Translational Bioimaging, 97080 Würzburg, Germany
[6]Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[7]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA
*Corresponding authors: brinda.vallat@rcsb.org, cseidel@hhu.de
[#]Deceased

Supplementary Notes

Contents

Supplementary Note 1: Fluorescence and FRET experiments 2
Supplementary Note 2: Integrative structure modeling 3
Supplementary Note 3: Challenges to making results of fluorescence spectroscopy studies publicly available 4
Supplementary Note 4: PDBx/mmCIF and IHMCIF dictionaries 5
Supplementary Note 5: Perspectives on data definitions 6
Supplementary Note 6: flrCIF dictionary 7
flrCIF dictionary content and structure 7
Extending the flrCIF dictionary beyond its current state 7
Supplementary Note 7: Kinetic schemes 9
Molecular kinetics 9
Multi-state systems 9
Description of complex kinetic schemes in IHMCIF 10
Example 11
Supplementary Note 8: Software support for flrCIF 13
Extension of the python-ihm software library 13
Converter script to enable user-friendly preparation of flrCIF compliant files for PDB-Dev 13
Supplementary Note 9: Fluorescence-assisted structure models in PDB-Dev 16
Supplementary references 18

Supplementary Note 1: Fluorescence and FRET experiments

Fluorescence measurements are increasingly applied for quantitative studies of biomolecules, as they allow us to investigate large, complex biomolecular systems in solution and even in live cells. Comparative advantages of fluorescence experiments include the time resolution in the nanosecond range, which avoids structural averaging; the possibility to study single molecules, which allows us to resolve multiple states; and the ability to capture dynamics, which allows us to extract exchange pathways, equilibrium constants, and relaxation times.

In particular Förster Resonance Energy Transfer (FRET) measurements, where energy transfer between a donor and an acceptor probe can be employed to obtain inter-dye distances in multi-state systems, are increasingly used for integrative structure modeling of biomolecules and for revealing their kinetics [1-8]. Furthermore, FRET measurements are well suited to provide data on dynamic systems, where a distribution of structure models yields a better depiction of the system than a single model [9]. Notably, FRET has the ability to resolve fast dynamic motions of more than 5 Å [7].

Supplementary Note 2: Integrative structure modeling

Integrative structure modeling approaches generate structure models of biomacromolecules and their complexes by combining data from multiple experimental and theoretical methods [10-12]. These models and their corresponding experimental data have varied applications in biology and medicine.

Experimental data sources for integrative modeling include data from X-ray crystallography, Nuclear Magnetic Resonance (NMR) and Electron Paramagnetic Resonance (EPR) spectroscopy, three-dimensional electron microscopy (3DEM), small-angle X-ray and neutron scattering (SAXS and SANS), chemical crosslinking mass spectrometry (Crosslinking-MS), and fluorescence spectroscopy. These can be combined with a number of theoretical input data, such as from homology modeling or *ab initio* structure prediction methods (AlphaFold [13] and RoseTTAFold [14]) or computational simulations. The input data can be converted into spatial restraints and incorporated into the modeling workflow in different ways, e.g. as distance restraints or volumes to which the model is fit.

The resulting models can be rather complex with respect to different aspects [11]: The models can be multi-scale, i.e. different parts of the model could be represented at different levels of granularity (e.g. atomistic, spheres, or Gaussian volumes). Furthermore, depending on the input data, the resulting model can be multi-state, if multiple discrete states are required to collectively satisfy the input information, or to describe a set of states ordered in time. Uncertainty in the input information can be described by a collection of models, where each model satisfies the input data within a threshold.

Using the IHMCIF [15] and PDBx/mmCIF [16] dictionaries, the structure models obtained from integrative modeling approaches can be deposited in PDB-Dev together with important metadata describing the input data as well as the modeling approach.

Fluorescence experiments, and in particular Förster Resonance Energy Transfer (FRET) measurements, are well suited to provide both structural and kinetic information in large, complex, and dynamic systems. Therefore, they are increasingly used to complement other experimental methods in integrative structure modeling approaches of biomolecules and for revealing their kinetics [5, 7, 17-19].

Supplementary Note 3: Challenges to making results of fluorescence spectroscopy studies publicly available

Integrative structure modeling computes structure models of macromolecules and their complexes by combining data from multiple experimental and theoretical methods [10]. Integrative models and their corresponding experimental data have varied applications in biology and medicine. Therefore, making this information publicly available following the FAIR (Findable, Accessible, Interoperable, Reusable) principles [20] is crucial for advancing science [21-23]. Structure models from integrative approaches can be deposited in the prototype archiving system PDB-Dev (https://pdb-dev.wwpdb.org).

For the established structure determination techniques (X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and three-dimensional electron microscopy (3DEM)), it is possible to deposit structure models in the Protein Data Bank (PDB) and experimental data to their respective repositories (X-ray data in PDB, NMR data in the Biological Magnetic Resonance Data Bank (BMRB) and 3DEM maps in the Electron Microscopy Data Bank (EMDB)).

For other biophysical methods whose data are often used in integrative structure modeling, work is in progress within the respective scientific communities to archive their experimental data [23]. For example, chemical crosslinking mass spectrometry data can be deposited in ProteomeXchange and small-angle X-ray and neutron scattering (SAXS and SANS) data and models are archived in in the Small Angle Scattering Biological Data Bank (SASBDB) [24, 25]. In contrast, a comparable data repository for fluorescence spectroscopy does not yet exist. In addition, community-accepted data standards for fluorescence data are currently being developed [5].

While structural insights from fluorescence experiments in early publications were often depicted only as hand-drawn sketches, development of computational modeling approaches allows us to generate structure models based on experimental results. However, such structure models are of limited use for further studies if they are not archived and made publicly available. This archival and dissemination requires creation of data and metadata standards so that the results of fluorescence investigations can be Findable, Accessible, Interoperable, and Reusable (FAIR) [20] and also satisfy the data sharing requirements set by many journals and funding agencies.

flrCIF provides the data standard for archiving and disseminating structure models in PDB-Dev together with relevant metadata derived from fluorescence and FRET experiments. In this way, flrCIF is a step towards making fluorescence-assisted structure models as well as fluorescence data FAIR.

Supplementary Note 4: PDBx/mmCIF and IHMCIF dictionaries

The Protein Data Bank (PDB) uses the **P**rotein **D**ata **B**ank E**x**change / **m**acro**m**olecular **C**rystallographic **I**nformation **F**ramework (PDBx/mmCIF) data standard to archive molecular structures and corresponding metadata from X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy experiments [16]. The PDBx/mmCIF dictionary provides descriptions of small molecules, macromolecules and complexes, as well as metadata, such as information on authors, citations, software, revision history, and release status of the entry. The file format defined by the PDBx/mmCIF dictionary is the current archival standard for structures in the PDB and has a number of advantages compared to its predecessor, the legacy PDB file format. One of the major advantages is that the PDBx/mmCIF file format allows for representation of large biomolecular complexes in a single file, while the PDB file format was limited in the number of polymer chains and atoms that can be included [16].

Structure models that are determined by combining experimental data from various sources using integrative modeling approaches can currently be deposited in PDB-Dev, a prototype archiving system for **I**ntegrative/**H**ybrid **M**odels [22, 26, 27]. To this end, PDB-Dev uses the IHMCIF dictionary [15, 26] (https://github.com/ihmwg/IHMCIF), which extends the PDBx/mmCIF dictionary to represent data and metadata focused on integrative modeling [27, 28].

The IHMCIF dictionary provides an expanded set of definitions to describe the complex representation of macromolecular structures obtained from integrative modeling, i.e. multi-state, multi-scale, and ordered models and collections of models. Moreover, descriptions of modeling workflows, starting structure models, input spatial restraints derived from different biophysical techniques, and related data in external resources are included in IHMCIF.

Supplementary Note 5: Perspectives on data definitions

Note that from the perspective of a structure model repository (e.g. PDB-Dev), data refers to information describing the structure model, such as the molecular system, atomic coordinates, the chemical connectivity of the atoms, and restraints used for generating the model. Metadata refers to additional information, such as reference molecule information, accession codes for related data in external resources, modeling methodology, authors, citations, or software used. On the other hand, for an experimental data repository, data would refer to the raw or processed experimental data, while metadata could refer to information on the experimental setup or procedures in addition to the metadata described above. A typical example for data assignment of a fluorescence spectroscopy experiment is described in detail by Lerner et al. [5] in Figures 7 and 8.

Supplementary Note 6: flrCIF dictionary

flrCIF dictionary content and structure

The flrCIF dictionary can be found on GitHub: https://github.com/ihmwg/flrCIF. An overview of the categories and the content can also be found at:
https://mmcif.wwpdb.org/dictionaries/mmcif_ihm_flr_ext.dic/Index/

When creating the categories in flrCIF, we tried to keep the definitions and required information generally applicable and avoid enforcing arbitrary standards. This approach promotes FAIR data delivery, facilitates extensibility, and allows for adapting existing software tools to support the dictionary.

Therefore, we first evaluated on the basis of several examples, what information is required to describe the path from a FRET experiment to the resulting structure model.
The requirements for many categories were obvious, such as the need to describe which fluorescent probes were used (including structural information if available), how and where these probes are attached to the biomolecule, and whether this required mutations or modifications of the biomolecule.
We created the requirements for the input as flexible as possible by keeping details non-mandatory where possible. For the specific implementation of flrCIF categories into the dictionary, we followed standard practices used by IHMCIF and PDBx/mmCIF based on the underlying Dictionary Definition Language (DDL2) [29].

The flrCIF dictionary extends the PDBx/mmCIF and IHMCIF dictionaries by providing definitions for fluorescence-specific information in four main categories describing the pathway from the fluorescence experiment to the structure model (Figure 1 in the main text): Experiment and sample, analysis workflow, fluorescence-specific information on the modeling procedure, and fluorescence-specific information on resulting models. For details on the content of the categories, see also Table S8.1 and Table S8.2 in Supplementary Note 8.

In addition to the above-mentioned fluorescence-specific information, we also extended the IHMCIF dictionary to describe multi-state schemes as mentioned earlier. These definitions enable the description of complex multi-state graphs, kinetic rates for the exchange between states, and relaxation rates (see Supplementary Note 7 for details).

Extending the flrCIF dictionary beyond its current state

While flrCIF already supports the deposition of FRET-assisted structure models in PDB-Dev, it can be further extended in several ways:
1. The current version of flrCIF focuses on FRET experiments, as these are able to directly provide inter-probe distances, which can be used for integrative structure modeling approaches. However, there are other fluorescence experiments or analysis approaches that yield input for integrative modeling in terms of information on the structure or on the dynamics. This could include be fluorescence quenching experiments, fluorescence anisotropy measurements to probe the local flexibility of the biomolecule, or probing the solvent exposure by spectral polarity, or different kinds of analyses, such as fluorescence correlation spectroscopy (FCS). Due to its extensible design, additional types of experiments or analysis methods together with the respective results can be added to flrCIF.
2. In the current version of flrCIF some categories, such as the description of the instrument, are only represented as textual description without further characterization. This is due to the complexities involved in describing custom-built instrument setups that consist of a large number of components including the beam path. Therefore, we decided that, at this stage, a

textual description is the most feasible option. In future, we plan to extend these categories with additional subcategories allowing for a better description. This would, on the one hand, make retrieving and searching for specific data items easier, and on the other hand give guidance as to which information can and should be deposited.
3. FRET efficiency-derived inter-dye distances are inter-probe distances, which are not easily converted to distances on the biomolecule. At the moment, flrCIF contains definitions to describe such fluorescence-specific modeling information using accessible volume (AV) calculations with the FPS software [2, 30]. flrCIF can be easily extended to include other modeling approaches, such as the Nano-Positioning System (NPS) [31, 32] or rotamer libraries [33, 34].

We plan to discuss possible options with members of the fluorescence community in order to find a consensus that can be adopted. This could, for example, be achieved via fret.community (https://fret.community/) , which is open to everybody, and represents a central hub offering a discussion platform.
Accordingly, community involvement is crucial for further developing flrCIF according to required features beyond the currently available basic description, as this will increase the adoption of these standards [23].

Supplementary Note 7: Kinetic schemes

Molecular kinetics

Structure models of biomolecular systems provide snapshots of the local and global structure of at least temporarily stable states, in the best case with atomistic resolution. The dynamics of these systems, i.e their temporal evolution, provide additional information on the energy landscape and motions that determine the transitions between states. From a formal perspective, molecular kinetics describes the temporal evolution of chemical reactions on the molecular level of chemical and biomolecular systems by analyzing the rates. Rate laws (kinetic models) were established to describe and analyze the concentration dependence of reactions for elucidating their reaction mechanism to recover the underlying rate constants. Dynamic systems have multiple states, which can differ in composition, bond configuration and/ or conformation. The states can be ordered within a mechanistic reaction scheme that also forms the basis for the corresponding kinetic scheme for describing the transitions between states.

Multi-state systems

Within IHMCIF [15], where a structure perspective is taken, biomolecular systems with multiple states are described as "multi-state" models. In this context, the term "ensemble" is also frequently used in different communities. However, the understanding of the term "ensemble" has specific meanings in different scientific communities and is still under debate. To avoid the potential ambiguity of the term "ensemble", IHMCIF uses "multi-state model" and "collection of models" to describe related concepts.
Considering a state with a single structure, IHMCIF allows one to specify a "collection of models, where each one is consistent with given input information within an acceptable threshold. The variability among the models in the collection helps in assessing the uncertainty of modeling and the completeness of input data" [15].
On the other hand, Reference [15] suggests that a multi-state model is used when a "set of multiple states can be used to describe a system that exists in a mixture of multiple structural and/or compositional states that collectively satisfy the input information" [15]. This description could be used in scenarios where experimental data could resolve multiple distinct states (e.g. enzymes [3, 6]) in an experiment, but also when multiple conformational states are required to collectively satisfy the input data (e.g. the intrinsically disordered human protein histone H1 [35] or multidomain proteins with unstructured segments such as lipase-specific foldase (LIF) [36]). Moreover, one can also think of a mixture of the two concepts (e.g. a collection of models describing the equilibrium between a folded and unstructured protein).

From a kinetic perspective, exchange between multiple states of a biomolecular system can relate to different scenarios as shown in Figure S7.1. Based on Lerner et al., [5], conformational dynamics can be typically defined as:
(1) transitions/exchanges between distinct structure states separated by a barrier with an activation energy, typically defined as larger than the thermal energy, k_BT (example A in Figure S7.1). Here, k_B is Boltzmann's constant and T is the absolute temperature.
Or (2) conformational fluctuations within states (intrastate dynamics), defined by the shape of the potential wells between activation barriers (Kinetic state K2 in example B in Figure S7.1).
Notably, the distinction between dynamics within a conformational state and dynamics of transitions between different conformational states depends on the definition of an activation barrier for different modes of structural dynamics.
From a structural perspective, however, the kinetic properties of biomolecular systems do not play a role for the definitions of multi-state systems in IHMCIF.

Currently, IHMCIF supports the kinetic description of systems where structure and kinetic state are equivalent (Figure S7.1 Example A). In the future, we will extend the dictionary to describe more complex systems with different activation energies that result in systems with inter-state and intra-state dynamics (Figure S7.1 Example B).

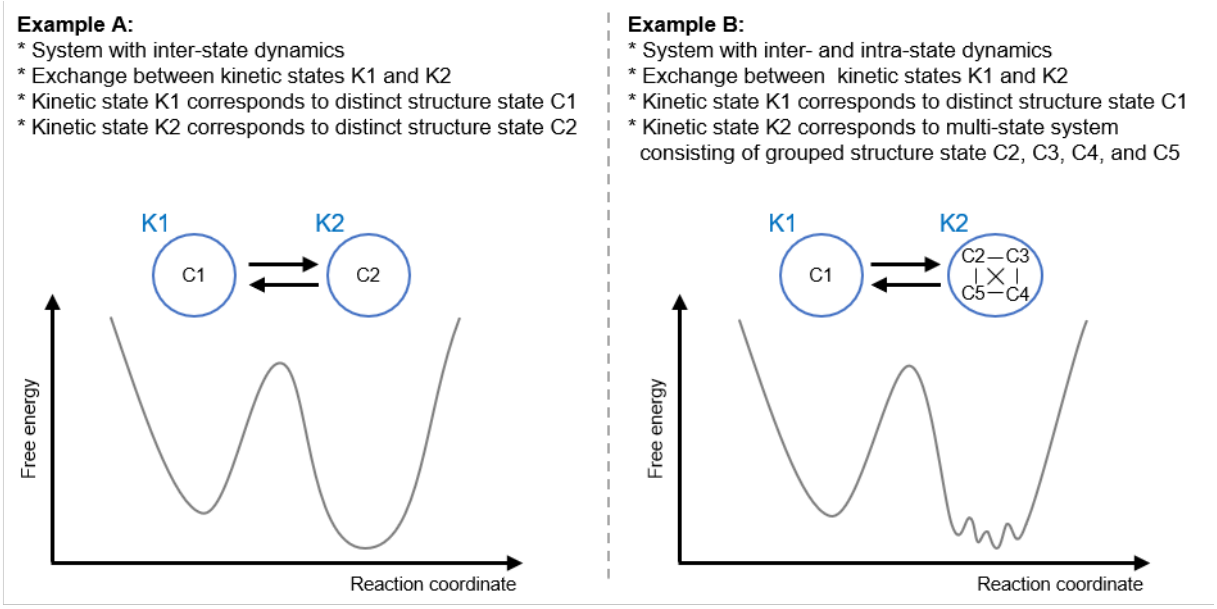


Figure S7.1: Examples for systems with exchange between kinetic states K1 and K2 together with their respective schematic energy landscapes. Example A shows a system with inter-state dynamics, where the kinetic states K1 and K2 directly correspond to the structure states C1 and C2, respectively. Example B shows a system with inter- and intra-state dynamics. Here, kinetic state K1 corresponds to structure state C1, but kinetic state K2 is made up of four structure states C2, C3, C4, and C5, which are collectively required to describe kinetic state K2.

Description of complex kinetic schemes in IHMCIF

The dynamics and kinetics in a biomolecular system can be characterized by fluorescence experiments [5] or other biophysical techniques. Depending on the time scales of the exchanges between conformational states, the population of the states, the underlying energy landscape, and the applied experimental approach, different levels of detail on the kinetics can be resolved [5].

In order to describe a biomolecular system with differently complex kinetics in IHMCIF, we implemented the description of a wide range of potential observables that could be extended for other use-cases. On the one hand, it is possible to describe the kinetics in the system using relaxation times without being able to assign them to the exchange between specific kinetic states. On the other hand, it is possible to quantify the exchange between specific states using transition rate constants, equilibrium constants, or relaxation times. In both cases, information on which measurements were used to derive the respective kinetic rates or relaxation times can be added.

The kinetic schemes themselves are implemented in IHMCIF in the form of multi-state schemes, which allow users to describe the general case of exchanges between states as directed graphs with states identified as nodes, and the connectivity between the states represented as directed edges (see example in Figure S7.2).

Since the information on multi-state schemes is not exclusive to fluorescence experiments, we generalized the description to be method-agnostic and added them to the IHMCIF dictionary instead of flrCIF. The description of multi-state schemes builds upon already existing

definitions in IHMCIF for multi-state representations. Here, it is also possible to assign each state a population fraction.
Multiple kinetic schemes can be defined for example when using different conditions (e.g. temperature, buffer conditions, presence of binding partners).

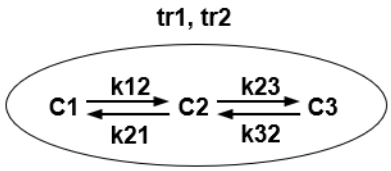
Example

Figure S7.2 shows a simple kinetic scheme containing three states C1, C2, and C3. These states can be defined using the definitions in IHMCIF together with their population fractions.
Figure S7.2a shows the underlying kinetic scheme, characterized by the three states C1, C2, C3 with exchanges between C1 and C2, C2 and C1, C2 and C3, and C3 and C2 with the respective transition rate constants k12, k21, k23, and k32. Furthermore, the system could be described with two relaxation times tr1 and tr2.
Figure S7.2b shows the transition rate matrix, from which the eigenvalues can be derived. These eigenvalues are related to the relaxation times tr1 and tr2 as shown in [3]:

$$\frac{1}{tr\{1,2\}} = \frac{1}{2} \cdot (k12 + k21 + k23 + k32) \pm \sqrt{(k12 + k21 + k23 + k32)^2 - 4 \cdot (k12 \cdot k23 + k12 \cdot k32 + k12 \cdot k32)} \qquad \text{(eq.1)}$$

Figure S7.2c shows the descriptions of the kinetic schemes in two ways: 1. Using the relaxation times, 2. Using the transition rate constants. Here, the columns describe pair-wise exchanges between states together with the corresponding transition rate constant.
Figure S7.2d shows an excerpt from a corresponding IHMCIF compliant file for the description of the two kinetic schemes with the corresponding observables. Highlighting colors are used to help indicate the respective identifiers (IDs) between the different categories of IHMCIF.

a



b

$$K = \begin{pmatrix} -k12 & k21 & 0 \\ k12 & -(k21 + k23) & k32 \\ 0 & k23 & -k32 \end{pmatrix}$$

c

Scheme 1: Relaxation times for the scheme

Begin State	C1	C2	C2	C3
End State	C2	C1	C3	C2
Relaxation time	tr1 = 120 ms, tr2 = 70 ms			

Scheme 2: Transition rate constants

	k12	k21	k23	k32
Begin State	C1	C2	C2	C3
End State	C2	C1	C3	C2
Kinetic transition rate constant	20 s^{-1}	300 s^{-1}	150 s^{-1}	30 s^{-1}

d

Corresponding mmcif file for the example

```
loop_
_ihm_multi_state_scheme.id
_ihm_multi_state_scheme.name
_ihm_multi_state_scheme.details
1 'Relaxation times' 'Relaxation times for the complete scheme from FCS'
2 'Transition rate constants' 'Transition rate constants between states'
#
#
loop_
_ihm_multi_state_scheme_connectivity.id
_ihm_multi_state_scheme_connectivity.scheme_id
_ihm_multi_state_scheme_connectivity.begin_state_id
_ihm_multi_state_scheme_connectivity.end_state_id
_ihm_multi_state_scheme_connectivity.details
1 1 C1 C2 .
2 1 C2 C1 .
3 1 C2 C3 .
4 1 C3 C2 .
5 2 C1 C2 k12
6 2 C2 C1 k21
7 2 C2 C3 k23
8 2 C3 C2 k32
#
loop_
_ihm_relaxation_time.id
_ihm_relaxation_time.value
_ihm_relaxation_time.unit
1 120 milliseconds
2 70 milliseconds
#
#
loop_
_ihm_relaxation_time_multi_state_scheme.id
_ihm_relaxation_time_multi_state_scheme.relaxation_time_id
_ihm_relaxation_time_multi_state_scheme.scheme_id
_ihm_relaxation_time_multi_state_scheme.scheme_connectivity_id
1 1 1 .
2 2 1 .
#
#
loop_
_ihm_kinetic_rate.id
_ihm_kinetic_rate.transition_rate_constant
_ihm_kinetic_rate.scheme_connectivity_id
1 20 5
2 300 6
3 150 7
4 30 8
#
```

Figure S7.2: Example of a kinetic scheme containing three states together with different observables. **(a)** The exemple kinetic scheme contains three states C1, C2, and C3, with exchanges between the states in a linear scheme described by the matrix K. The kinetics in the system can be solved on a system level and described by two relaxation times tr1 and tr2. The exchanges between the specific states can be quantified with the transition rate constants k12, k21, k23, and k32. **(b)** Transition rate matrix describing the kinetic scheme. **(c)** Overview over two possible schemes that could be described with information of which states could exchange as well as which observables could be derived. **(d)** Excerpt from an IHMCIF compliant file for the description of the two kinetic schemes with the corresponding observables. The definition of the states is omitted here. Highlighting colors are used to visualize the respective relationships of IDs in the categories.

Supplementary Note 8: Software support for flrCIF

Although it is, in principle, possible to manually prepare files that comply with the flrCIF, IHMCIF, and PDBx/mmCIF dictionaries (Figure 1a), this approach is very laborious and error-prone due to the large number of categories and the required references and connections between them. Therefore, the definitions from flrCIF are implemented in the python-ihm software library. For convenience, we provide a converter script, which allows one to generate a PDBx/mmCIF, IHMCIF, and flrCIF compliant file for deposition in PDB-Dev by filling a user-friendly spreadsheet template.

Extension of the python-ihm software library

The python-ihm software library [28] (https://github.com/ihmwg/python-ihm) enables writing and reading of data files compliant with the PDBx/mmCIF and IHMCIF dictionaries. In python-ihm, the data and metadata definitions required for representing integrative models are implemented as Python classes. In order to extend python-ihm to support fluorescence experiments, we added additional classes to the package based on the definitions in the flrCIF dictionary as well as the kinetic scheme definitions in the IHMCIF dictionary. Other software can use the python-ihm library to read and write IHMCIF dictionary-compliant files.

Converter script to enable user-friendly preparation of flrCIF compliant files for PDB-Dev

To facilitate the deposition of fluorescence-assisted structure models in PDB-Dev, we created a Python script that employs the python-ihm package to generate data files compliant with the IHMCIF and flrCIF dictionaries (https://github.com/Fluorescence-Tools/flr2mmcif). The Python script requires the user to fill a spreadsheet template with information corresponding to data and metadata defined in the PDBx/mmCIF, the IHMCIF, and the flrCIF dictionaries. This includes information on: primary citation, macromolecules studied, related datasets in external data resources, software used for modeling, modeling protocol, multi-state modeling if applicable, models submitted, fluorescence experiment setup, fluorescent labels, correction parameters, resulting distance restraints, model quality metrics, and potential kinetic schemes. The Python script uses the filled spreadsheet template together with the provided structure models to generate a combined data file in mmCIF format, which can be submitted to PDB-Dev (Figure 1).

The GitHub repository (https://github.com/Fluorescence-Tools/flr2mmcif) contains the Python converter script, the template spreadsheet, and an example folder with a short example. Note that it might be necessary to download the spreadsheet files in raw format in order for them to work.

An overview over the categories collected in the spreadsheet template is given in Table S8.1. An overview over fluorescence-specific information collected in the spreadsheet template in Tab "FLR" is given in Table S8.2.

Table S8.1: Information collected in the spreadsheet for flr2mmcif. The information is collected in different tabs of the spreadsheet. IDs in the tabs are used to refer to entries in other tabs.

	Excel Tabs for collecting information		
1.	**Citation** (Title, Authors, Journal, Year, DOI, ...)		
2.	**Entity**: Entities in the system		
	2.1.	Type of entity (polymer, non-polymer, water)	
	2.2.	Number of copies of the entity in the entity assembly	
	2.3.	Source of the entity	
	2.4.	If a polymer, give type and sequence. If not a polymer, give the chemical component ID (http://www.wwpdb.org/data/ccd), name, and formula	
3.	**Dataset** - Multiple datasets can be added to dataset groups		
	3.1.	Type of data (ihm_dataset_list, e.g. NMR data, SAS data, ...)	
		3.2.1.	Data deposited in a database? If so, where?
		3.2.2.	If not deposited in a database, deposited in a repository (e.g. Zenodo): DOI and URL
4.	**External files** - files within the datasets defined previously (file format, content of the file)		
5.	**Software** - software used e.g. for analyses or modeling (Name, classification, description, location where to find the software, e.g. URL)		
6.	**Instance** (AsymUnit)		
	6.1.	Details on the instance (entity, chain ID, sequence ID for start and end of the instance)	
	6.2.	Model representation	
		6.2.1.	How was the object modeled (atomistic, sphere, Gaussian, ...)?
		6.2.2.	Was the object rigid or flexible?
		6.2.3.	How was the starting model obtained (experimental model, *ab initio* model, integrative model, comparative model, other)? Chain ID of the starting model and sequence offset.
		6.2.4.	Corresponding dataset and external files
7.	If a **comparative model** (from homology modeling) was used as a starting model, details can be given (asym ID and sequence IDs for model and template, sequence identity)		
8.	**Modeling protocol** (steps of the modeling protocol, number of models at the beginning and the end of the step, did the modeling involve multi-scale, multi-state, or ordered models, or collection of models)		
9.	**Modeling post process** - Post-processing steps after the modeling, e.g. clustering.		
10.	**Multi-state modeling** - if multiple states were modeled, information on the states (names, population fractions, type of states (e.g. structural conformations) can be given. States can be grouped. Models can be assigned to the states.		
11.	**Multi-state scheme** - e.g. kinetic schemes - described by connectivities between states		
	11.1.	Connectivity between states (start state, end state)	
	11.2.	Quantifying the exchange between states within the multi-state scheme	
		11.2.1.	Relaxation time either for the entire scheme or assigned to a specific connectivity between states (relaxation time, unit, amplitude)
		11.2.2.	Kinetic rate for a specific connectivity between states (transition rate constant, equilibrium constant)
12.	**Models** - Information about the models to be deposited (corresponding state, representative of a a collection, modeling protocol, ...). Models can be grouped.		
13.	**Collection of models** - Information if a collection of models is deposited (how many models are in the collection? How many models from the collection are deposited, ...)		
14.	**Reference measurements** - Reference measurements for fluorescence lifetime experiments. Similar information to the FLR information (Table S8.2)		
15.	**FLR** - Fluorescence-specific information (see Table S8.2)		
16.	**FLR FPS MPP group** - Information for modeling in the FPS software when using the mean probe position approach. This is not recommended, but possible to use.		
17.	**FLR FPS global parameters** - Global parameters used in the FPS software		
18.	**FLR FRET Model distances** - Distances between probes for different probe pairs for each of the deposited models. From this, the deviation in the distance w.r.t. the input value can be calculated.		
19.	**FLR FRET Model quality** - The quality of the deposited models based on the FRET data. Often given as χ^2 value.		

Table S8.2: Information categories of fluorescence experiments from flrCIF collected in the spreadsheet for flr2mmcif (Tab "FLR"). Depending on the category, additional details are collected.

Collected information (FLR tab)		
Experiment and sample		
1.	**Instrument specification.** Components (lasers, optical elements, detectors, ...) and beam path: Free textual description of the parts	
2.	**Instrument settings.** Excitation wavelengths, laser power, observation volume, spectral detection ranges, ...: Free textual description	
3.	**Experimental conditions** (e.g. temperature, buffers, ...): Free textual description	
4.	**Fluorescent probes on the sample**	
	4.1.	How many fluorescent probes were used?
	4.2.	Probe type. Which fluorescent probes were used?
	4.3.	Attachment of the probe. Extrinsic or intrinsic probe (e.g. tryptophan)?
	4.4	For extrinsic probe: How was the probe attached?
	4.5.	Chemical information on the probes (SMILES, INCHI code, etc.)
	4.6.	Location. Where were the probes attached? (entity, residue, atom)
	4.7.	Nature of residues. Were the residues to which the probes were attached modified or mutated? If so, details can be provided.
	4.8.	Specificity of labeling. Was the labeling ambiguous?
5.	**Förster radius** for FRET experiments	
6.	**Additional information (raw and metadata).** For each of the results of a measurement for a sample, additional information such as corresponding datasets or external files can be provided.	
Analysis workflow		
7.	**Analysis.** What kind of analysis was performed?	
	7.1.	Intensity-based analysis: The report of several correction parameters is required. The ones currently implemented in flrCIF follow the definitions from Hellenkamp et al. (2018) [37]
	7.2.	Lifetime-based analysis: Information about reference measurements (e.g. Donor- or Acceptor-only measurements) should be provided as well as the employed fit model.
8.	**FRET distance restraints.** List of FRET-based distance restraints that were used in the structure modeling approach together with corresponding errors.	
	8.1.	Assignment. In case of multiple states, the same FRET pair could yield multiple distances
Fluorescence-specific information on modeling procedure		
9.	**Dye simulation type.** FRET efficiency-derived inter-dye distances are inter-probe distances, which are not easily converted to distances on the biomolecule. One approach to tackle this issue is the use of accessible volume calculations, where the label is implicitly described using the length of the linker, the width of the linker, and the probe radius [30]. Other approaches might include additional information into these accessible volumes [5, 38]. At the moment, flrCIF contains a description of the used Accessible Volume (AV) parameters, if the FPS (FRET positioning and screening) program [2, 30] is used. Note: The definitions for the FPS software were made due to familiarity with the software. flrCIF can however be easily extended to support parameters for other software as well.	

Supplementary Note 9: Fluorescence-assisted structure models in PDB-Dev

At present, there are eight FRET-assisted entries in PDB-Dev (Table S9.1). While this number is small compared to the total number of entries (188 in June 2024), FRET-assisted entries make up more than 50% of all multi-state entries in PDB-Dev (Figure S9.2). This contribution demonstrates the potential of fluorescence experiments in overcoming the single-state view by revealing insights into dynamic and multi-state systems, which are highly relevant for understanding biological function [9].

Table S9.1: Entries in PDB-Dev that employ FRET data in integrative modeling.

PDB-Dev ID	**Biomolecular system**	**Multi-state model**
PDBDEV_00000004 [a]	K63-linked Diubiquitin	Yes
PDBDEV_00000019	RNA four-way junction	Yes
PDBDEV_00000032 [a]	HCN Channel voltage sensor	Yes
PDBDEV_00000044 [b]	T4 lysozyme	Yes
PDBDEV_00000082	Alpha-synuclein monomer	No
PDBDEV_00000088 [b]	Human guanylate binding protein 1 (hGBP1)	Yes
PDBDEV_00000161	PSD-95	Yes
PDBDEV_00000164	PSD-95	Yes

[a] Does not use flrCIF due to being deposited before flrCIF was established.
[b] Includes information on kinetics

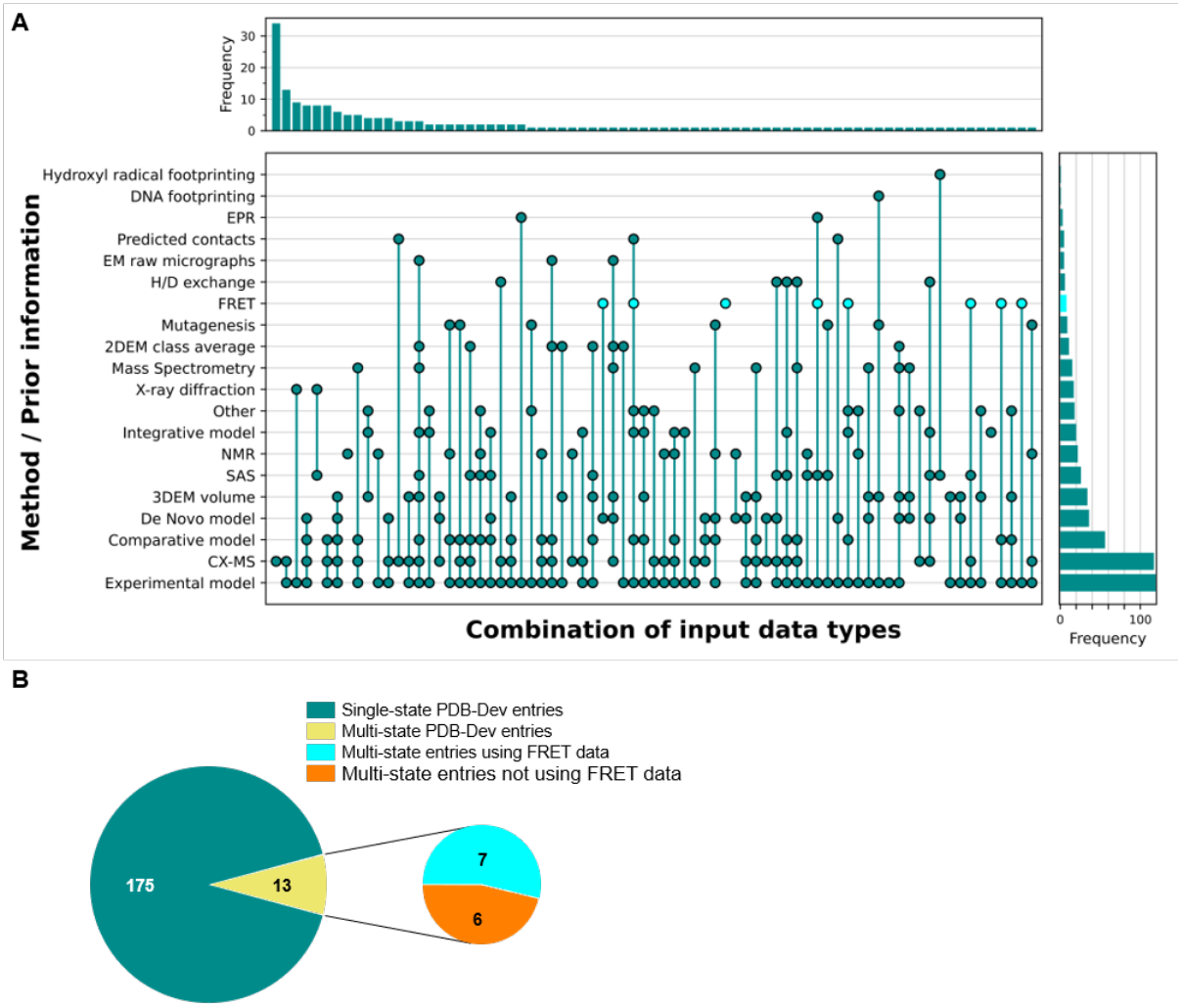


Figure S9.2: Combination of input data types in PDB-Dev entries and number of multi-state entries in PDB-Dev using FRET. **A:** Combination of input data types of PDB-Dev entries. The plot indicates which input data types are combined for entries in PDB-Dev. The histogram on top shows the number of entries with a specific combination. The histogram on the right shows the number of entries where a specific input data type was used. FRET data as input data type is highlighted in cyan. **B:** Pie diagram showing how many entries in PDB-Dev currently (June 2024) are single-state and multi-state entries, i.e. containing information on more than one structural state. More than 50% of the deposited multi-state entries use FRET as input data. Data was obtained from https://pdb-dev.wwpdb.org/.

Supplementary references

1. Hellenkamp, B., et al., *Multidomain structure and correlated dynamics determined by self-consistent FRET networks.* Nature Methods, 2017. **14**(2): p. 174-180. DOI: 10.1038/Nmeth.4081.
2. Dimura, M., et al., *Automated and optimally FRET-assisted structural modeling.* Nature Communications, 2020. **11**(1): p. 5394. DOI: 10.1038/s41467-020-19023-1.
3. Sanabria, H., et al., *Resolving dynamics and function of transient states in single enzyme molecules.* Nature Communications, 2020. **11**(1). DOI: 10.1038/s41467-020-14886-w.
4. Chen, J.X., et al., *The structural heterogeneity of alpha-synuclein is governed by several distinct subpopulations with interconversion times slower than milliseconds.* Structure, 2021. **29**(9): p. 1048-1064. DOI: 10.1016/j.str.2021.05.002.
5. Lerner, E., et al., *FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices.* eLife, 2021. **10**: p. e60416. DOI: 10.7554/eLife.60416.
6. Hamilton, G.L., et al., *Fuzzy supertertiary interactions within PSD-95 enable ligand binding.* eLife, 2022. **11**: p. e77242. DOI: 10.7554/eLife.77242.
7. Agam, G., et al., *Reliability and accuracy of single-molecule FRET studies for characterization of structural dynamics and distances in proteins.* Nature Methods, 2023. **20**(4): p. 523-535. DOI: 10.1038/s41592-023-01807-0.
8. Peulen, T.O., et al., *Integrative dynamic structural biology unveils conformers essential for the oligomerization of a large GTPase.* eLife, 2023. **12**: p. e79565. DOI: 10.7554/eLife.79565.
9. Lane, T.J., *Protein structure prediction has reached the single-structure frontier.* Nature Methods, 2023. **20**: p. 170-173. DOI: 10.1038/s41592-022-01760-4.
10. Sali, A., *From integrative structural biology to cell biology.* Journal of Biological Chemistry, 2021. **296**: p. 100743. DOI: 10.1016/j.jbc.2021.100743.
11. Rout, M.P. and A. Sali, *Principles for integrative structural biology studies.* Cell, 2019. **177**(6): p. 1384-1403. DOI: 10.1016/j.cell.2019.05.016.
12. Braitbard, M., D. Schneidman-Duhovny, and N. Kalisman, *Integrative structure modeling: Overview and assessment.* Annual Review of Biochemistry, 2019. **88**: p. 113-135. DOI: 10.1146/annurev-biochem-013118-111429.
13. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589. DOI: 10.1038/s41586-021-03819-2.
14. Baek, M., et al., *Accurate prediction of protein structures and interactions using a three-track neural network.* Science, 2021. **373**(6557): p. 871-876. DOI: 10.1126/science.abj8754.
15. Vallat, B., et al., *IHMCIF: An extension of the PDBx/mmCIF data standard for integrative structure determination methods.* Journal of Molecular Biology, 2024: p. 168546. DOI: 10.1016/j.jmb.2024.168546.
16. Westbrook, J.D., et al., *PDBx/mmCIF Ecosystem: Foundational semantic tools for structural biology.* Journal of Molecular Biology, 2022. **434**(11): p. 167599. DOI: 10.1016/j.jmb.2022.167599.
17. Kilic, S., et al., *Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1α.* Nature Communications, 2018. **9**: p. 235. DOI: 10.1038/s41467-017-02619-5.
18. Lerner, E., et al., *Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer.* Science, 2018. **359**(6373): p. eaan1133. DOI: 10.1126/science.aan1133.
19. Opanasyuk, O., et al., *Unraveling multi-state molecular dynamics in single-molecule FRET experiments. II. Quantitative analysis of multi-state kinetic networks.* Journal of Chemical Physics, 2022. **157**(3): p. 031501. DOI: 10.1063/5.0095754.
20. Wilkinson, M.D., et al., *The FAIR Guiding Principles for scientific data management and stewardship.* Scientific Data, 2016. **3**: p. 160018. DOI: 10.1038/sdata.2016.18.

21. Sali, A., et al., *Outcome of the first wwPDB Hybrid/Integrative Methods Task Force workshop.* Structure, 2015. **23**(7): p. 1156-1167. DOI: 10.1016/j.str.2015.05.013.
22. Burley, S.K., et al., *PDB-Dev: a prototype system for depositing integrative/hybrid structural models.* Structure, 2017. **25**(9): p. 1317-1318. DOI: 10.1016/j.str.2017.08.001.
23. Berman, H.M., et al., *Federating structural models and data: Outcomes from a workshop on archiving integrative structures.* Structure, 2019. **27**(12): p. 1745-1759. DOI: 10.1016/j.str.2019.11.002.
24. Valentini, E., et al., *SASBDB, a repository for biological small-angle scattering data.* Nucleic Acids Research, 2015. **43**(D1): p. D357-D363. DOI: 10.1093/nar/gku1047.
25. Kikhney, A.G., et al., *SASBDB: Towards an automatically curated and validated repository for biological scattering data.* Protein Science, 2020. **29**(1): p. 66-75. DOI: 10.1002/pro.3731.
26. Vallat, B., et al., *Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules.* Structure, 2018. **26**(6): p. 894-904. DOI: 10.1016/j.str.2018.03.011.
27. Vallat, B., et al., *New system for archiving integrative structures.* Acta Crystallographica Section D-Structural Biology, 2021. **77**: p. 1486-1496. DOI: 10.1107/S2059798321010871.
28. Vallat, B., et al., *Archiving and disseminating integrative structure models.* Journal of Biomolecular NMR, 2019. **73**(6-7): p. 385-398. DOI: 10.1007/s10858-019-00264-2.
29. Westbrook, J.D., H.M. Berman, and S.R. Hall, *Specification of a relational dictionary definition language (DDL2).* International Tables for Crystallography, 2006. **G, ch. 2.6**: p. 61-70. DOI: 10.1107/97809553602060000732.
30. Kalinin, S., et al., *A toolkit and benchmark study for FRET-restrained high-precision structural modeling.* Nature Methods, 2012. **9**(12): p. 1218-1225. DOI: 10.1038/Nmeth.2222.
31. Muschielok, A., et al., *A nano-positioning system for macromolecular structural analysis.* Nature Methods, 2008. **5**(11): p. 965-971. DOI: 10.1038/nmeth.1259.
32. Eilert, T., et al., *Fast-NPS - A Markov Chain Monte Carlo-based analysis tool to obtain structural information from single-molecule FRET measurements.* Computer Physics Communications, 2017. **219**: p. 377-389. DOI: 10.1016/j.cpc.2017.05.027.
33. Klose, D., et al., *Resolving distance variations by single-molecule FRET and EPR spectroscopy using rotamer libraries.* Biophysical Journal, 2021. **120**(21): p. 4842-4858. DOI: 10.1016/j.bpj.2021.09.021.
34. Montepietra, D., et al., *FRETpredict: a Python package for FRET efficiency predictions using rotamer libraries.* Communications Biology, 2024. **7**(1). DOI: 10.1038/s42003-024-05910-6.
35. Borgia, A., et al., *Extreme disorder in an ultrahigh-affinity protein complex.* Nature, 2018. **555**(7694): p. 61-66. DOI: 10.1038/nature25762.
36. Dittrich, J., et al., *Resolution of Maximum Entropy Method-derived posterior conformational ensembles of a flexible system probed by FRET and molecular dynamics simulations.* Journal of Chemical Theory and Computation, 2023. **19**(8): p. 2389-2409. DOI: 10.1021/acs.jctc.2c01090.
37. Hellenkamp, B., et al., *Precision and accuracy of single-molecule FRET measurements-a multi-laboratory benchmark study.* Nature Methods, 2018. **15**(9): p. 669-676. DOI: 10.1038/s41592-018-0085-0.
38. Dimura, M., et al., *Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems.* Current Opinion in Structural Biology, 2016. **40**: p. 163-185. DOI: 10.1016/j.sbi.2016.11.012.