

Making fluorescence-based integrative structures and associated kinetic information accessible



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 standard⁹, 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 dictionary that 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-state¹⁰ (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.

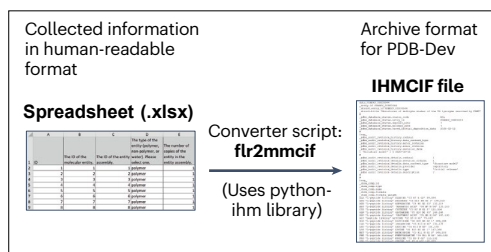
2. Software developers:

- 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.

a Dictionaries and their corresponding information content

Information	Dictionary
General information; Molecular system (supports single state)	PDBx /mmCIF
Integrative/hybrid modeling (supports multi-state)	IHM CIF
Fluorescence-related information	flrCIF

Tools for PDB-Dev deposition



b Collected information for fluorescence experiments

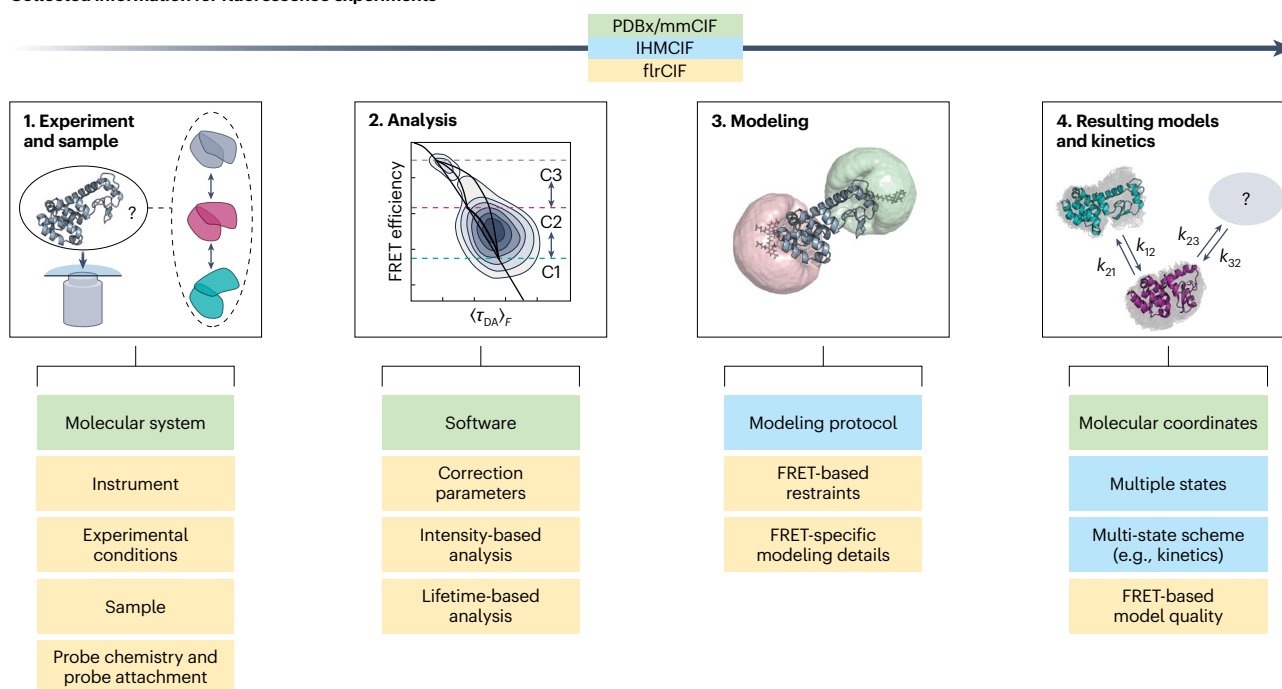


Fig. 1 | Information in dictionary-compliant file for a FRET-assisted integrative structure model. **a**, Information content of the different dictionaries supporting deposition of FRET-based structure models in PDB-Dev. PDBx/mmCIF describes general information about the entry, such as description of the molecular system and atomic coordinates as well as metadata regarding authors, software and citations. IHM CIF 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 **b**. The flr2mmcif converter script uses a spreadsheet to collect information in a human-readable format and converts this to a PDB-Dev-compatible IHM CIF 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 IHM CIF dictionaries for the description of fluorescence-related information. The green boxes show definitions in PDBx/mmCIF; blue boxes, definitions in IHM CIF; yellow boxes, definitions in flrCIF.

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 IHM CIF extends the applicability beyond fluores-

cence to various techniques used in integrative modeling studies and facilitates the collection and archiving of structural dynamics information.

Overall, flrCIF contributes toward dynamic structural biology by enabling standardized deposition, archiving and dissemination of information on biomolecules in their functional context – from static structures toward dynamic structural networks and the

connection of structures with energy landscapes and real time motions.

Reporting Summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

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>.

Christian A. Hanke¹, **John D. Westbrook**^{2,3,8}, **Benjamin M. Webb**⁴, **Thomas-O. Peulen**^{1,4,7}, **Catherine L. Lawson**², **Andrej Sali**⁴, **Helen M. Berman**^{2,5,6}, **Claus A. M. Seidel**¹ ✉ & **Brinda Vallat**^{2,3} ✉

¹Molecular Physical Chemistry, Heinrich Heine University Düsseldorf, Düsseldorf, Germany.

²Research Collaboratory for Structural Bioinformatics Protein Data Bank and the Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. ³Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA. ⁴Department 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.

⁵Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ, USA.

⁶Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA.

⁷Present address: Physical Chemistry, TU Dortmund University, Dortmund, Germany.

⁸Deceased: John D. Westbrook

✉ e-mail: cseidel@hhu.de;

brinda.vallat@rcsb.org

Published online: 30 September 2024

References

1. Lerner, E. et al. *eLife* **10**, e60416 (2021).
2. Agam, G. et al. *Nat. Methods* **20**, 523–535 (2023).
3. Peulen, T. O. et al. *eLife* **12**, e79565 (2023).
4. Rout, M. P. & Sali, A. *Cell* **177**, 1384–1403 (2019).
5. Sali, A. et al. *Structure* **23**, 1156–1167 (2015).
6. Berman, H. M. et al. *Structure* **27**, 1745–1759 (2019).
7. Vallat, B., Webb, B., Westbrook, J. D., Sali, A. & Berman, H. M. *Structure* **26**, 894–904.e2 (2018).
8. Vallat, B. et al. *Acta Crystallogr. D Struct. Biol.* **77**, 1486–1496 (2021).
9. Westbrook, J. D. et al. *J. Mol. Biol.* **434**, 167599 (2022).
10. Vallat, B. et al. *J. Mol. Biol.* **436**, 168546 (2024).

Acknowledgements

We thank M. Gabanyi for 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 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.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-024-02428-x>.

Peer review information *Nature Methods* thanks Don Lamb and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|---|
| n/a | Confirmed |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

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 file generation workflow, which can be used to prepare dictionary-compliant mmCIF files including fluorescence information is publicly available at <https://github.com/Fluorescence-Tools/flr2mmCIF>.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Use the terms *sex* (biological attribute) and *gender* (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data exclusions

Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Replication

Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.

Randomization

Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.

Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|--|
| n/a | Involvement in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involvement in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |