STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Submission of structural biology data for review purposes

Edward N. Baker,[a]* Charles S. Bond,[b]* Elspeth F. Garman,[c]* Janet Newman,[d,e]* Randy J. Read[f]* and Mark J. van Raaij[g]*

[a]School of Biological Sciences, University of Auckland, Auckland, New Zealand, [b]University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia, [c]University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom, [d]Collaborative Crystallisation Centre (C3), CSIRO, 343 Royal Parade, Parkville, VIC 3052, Australia, [e]School of Biotechnology and Biomolecular Sciences, UNSW, Sydney, NSW 2052, Australia, [f]Cambridge Institute for Medical Research, University of Cambridge, The Keith Peters Building, Hills Road, Cambridge CB2 0XY, United Kingdom, and [g]Departamento de Estructura de Macromoleculas, Centro Nacional de Biotecnologia, Consejo Superior de Investigaciones Cientificas, E-28049, Madrid, Spain. *Correspondence e-mail: en.baker@auckland.ac.nz, charles.bond@uwa.edu.au, elspeth.garman@bioch.ox.ac.uk, janet.newman@csiro.au, rjr27@cam.ac.uk, mjvanraaij@cnb.csic.es

Keywords: structural biology; data; peer review.

The IUCr stable of journals are proponents of the FAIR and FACT principles of publishing: data should be Findable, Accessible, Interoperable and Reusable (Wilkinson *et al.*, 2016) and Fair, Accurate, Confidential and Transparent (van der Aalst *et al.*, 2017; Helliwell, 2019). This has been manifested by a number of leading contributions to the field of structural biology including publication standards for a variety of data types (Kroon-Batenburg & Helliwell, 2014; Adams *et al.*, 2019; Guss & McMahon, 2014) as well as mandatory deposition of coordinates, and subsequently structure factors, for publication of macromolecular crystal structures. This approach has largely satisfied the FAIR and FACT principles. However, a further step is required to ensure that the data deposited in a third-party database fully support the conclusions drawn from those data. This step requires the mandatory submission of data (*e.g.* coordinates and reduced diffraction data, discussed further below) to accompany the manuscript describing the work.

This action does not imply that the various databases [PDB (Burley *et al.*, 2019), EMDB (Lawson *et al.*, 2016), SASBDB (Kachala *et al.*, 2016), BMRB (Ulrich *et al.*, 2008) *etc.*] are not doing a good job of validating data submitted to them. Indeed, the validation reports produced by databases are an important pillar in the FAIR approach to structural biology. Nevertheless, it is not the role of the databases to test the extent to which the interpretations and conclusions of a manuscript incorporating structural biology research (possibly by multiple methods) are supported by the deposited data. This role belongs in the manuscript review cycle.

Authors may ask why submission of data needs to be made mandatory. After all, the Notes for Authors for *Acta Cryst. D*, *Acta Cryst. F* and *IUCrJ* all indicate that editors may request such data from the author. However, the current system causes delays in the review system when editors have to request data. The added inconvenience increases the possibility of poorly supported interpretations avoiding scrutiny and being published. In order to clarify the situation, and maintain the IUCr's position leading the field in data quality standards in structural biology, submission of deposited data for review is best practice.

Historically there has been reticence to provide data for review as this might provide a competitor with an advantage. This concern has reduced substantially over recent years as increasing numbers of macromolecular structures are released prior to publication, and complementary methods such as accurate and precise computational predictions are more widely used. Indeed, the increasing prevalence of preprint versions of papers has further expedited the release of non-peer-reviewed data. Data submitted to IUCr Journals will be mandated *for review purposes only*, and thus the ethical position of a reviewer receiving them is clearly defined.

The concept of data deposition can be extremely broad, so it is important that we clarify what is mandated and what is recommended. The non-negotiable starting point is that for any manuscript containing conventional structures determined by the most common techniques (crystallography, NMR, cryoEM, SAXS) the data that are deposited

with the relevant database to obtain the accession code and validation reports must be uploaded prior to editorial review. This set of standard files (which may include coordinate files, reduced diffraction data, restraint lists, electron-density maps, buffer-subtracted scattering curves, along with any validation reports) should already be immediately at hand to the submitting author. Additionally, depending on the manuscript in question there are cases where additional data should be provided. For example, many crystal structures may have indications of space-group ambiguity, twinning, anisotropy, alternate unit cells or other pathologies for which the review process would benefit from the submission of scaled but unmerged intensity data. We do not mandate submission for review of primary data as this is not practicable, however we continue to recommend that such data be made publicly available (Guss & McMahon, 2014).

To accompany the requirement for data, which will apply from the start of 2022, the journals *Acta Cryst. D*, *Acta Cryst. F* and *IUCrJ* are unifying their requirements for information to be included in the standard tables for various experiment types (colloquially Table 1 for crystal structures). These new requirements will be described in detail in the revised Notes for Authors for the journals.

References

Adams, P. D., Afonine, P. V., Baskaran, K., Berman, H. M., Berrisford, J., Bricogne, G., Brown, D. G., Burley, S. K., Chen, M., Feng, Z., Flensburg, C., Gutmanas, A., Hoch, J. C., Ikegawa, Y., Kengaku, Y., Krissinel, E., Kurisu, G., Liang, Y., Liebschner, D., Mak, L., Markley, J. L., Moriarty, N. W., Murshudov, G. N., Noble, M., Peisach, E., Persikova, I., Poon, B. K., Sobolev, O. V., Ulrich, E. L., Velankar, S., Vonrhein, C., Westbrook, J., Wojdyr, M., Yokochi, M. & Young, J. Y. (2019). *Acta Cryst.* D**75**, 451–454.

Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Costanzo, L. D., Christie, C., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., Liang, Y., Lowe, R., Peisach, E., Periskova, I., Randle, C., Rose, A., Sekharan, M., Shao, C., Tao, Y.., Valasatava, Y., Voigt, M., Westbrook, J., Young, J., Zardecki, C., Zhuravleva, M., Kurisu, G., Nakamura, H., Kengaku, Y., Cho, H., Sato, J., Kim, J. Y., Ikegawa, Y., Nakagawa, A., Yamashita, R., Kudou, T., Bekker, G.., Suzuki, H., Iwata, T., Yokochi, M., Kobayashi, N., Fujiwara, T., Velankar, S., Kleywegt, G. J., Anyango, S., Armstrong, D. R., Berrisford, J. M., Conroy, M. J., Dana, J. M., Deshpande, M., Gane, P., Gáborová, R., Gupta, D., Gutmanas, A., Koča, J., Mak, L., Mir, S., Mukhopadhyay, A., Nadzirin, N., Nair, S., Patwardhan, A., Paysan-Lafosse, T., Pravda, L., Salih, O., Sehnal, D., Varadi, M., Vařeková, R., Markley, J. L., Hoch, J. C., Romero, P. R., Baskaran, K., Maziuk, D., Ulrich, E. L., Wedell, J. R., Yao, H., Livny, M. & Ioannidis, Y. E. (2019). *Nucleic Acids Res.* **47**, D520–D528.

Guss, J. M. & McMahon, B. (2014). *Acta Cryst.* D**70**, 2520–2532.

Helliwell, J. R. (2019). *Struct. Dyn.* **6**, 054306.

Kachala, M., Westbrook, J. & Svergun, D. (2016). *J. Appl. Cryst.* **49**, 302–310.

Kroon-Batenburg, L. M. J. & Helliwell, J. R. (2014). *Acta Cryst.* D**70**, 2502–2509.

Lawson, C. L., Patwardhan, A., Baker, M. L., Hryc, C., Garcia, E. S., Hudson, B. P., Lagerstedt, I., Ludtke, S. J., Pintilie, G., Sala, R., Westbrook, J. D., Berman, H. M., Kleywegt, G. J. & Chiu, W. (2016). *Nucleic Acids Res.* **44**, D396–D403.

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H. & Markley, J. L. (2008). *Nucleic Acids Res.* **36**, D402–D408.

van der Aalst, W. M. P., Bichler, M. & Heinzl, A. (2017). *Bus. Inf. Syst. Eng.* **59**, 311–313.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B. (2016). *Sci. Data*, **3**, 160018.