STRUCTURAL BIOLOGY
ISSN 2059-7983

Decision making in macromolecular crystallography: how to be a productive structural biologist

Helen M. Ginn,[a,b,c]* Nicholas M. Pearce[d]* and Clemens Vonrhein[e]*

[a]Center for Free-Electron Laser Science CFEL, Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany, [b]Institute for Nanostructure and Solid State Physics, University of Hamburg, Hamburg, Germany, [c]The Hamburg Centre for Ultrafast Imaging, Hamburg, Germany, [d]Department of Cell and Molecular Biology, SciLifeLab, Uppsala University, SE-75237 Uppsala, Sweden, and [e]Global Phasing Ltd, 9 Journey Campus, Castle Park, Cambridge CB3 0AX, United Kingdom. *Correspondence e-mail: helen.ginn@desy.de, nicholas.pearce@icm.uu.se, vonrhein@globalphasing.com

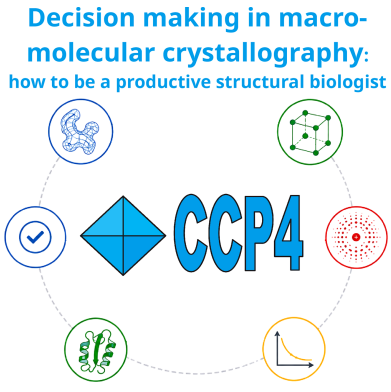
Keywords: CCP4 Study Weekend; decision making; macromolecular crystallography; automation.

The 2024 CCP4 Study Weekend was held from the 3rd to 5th of January 2024 as a hybrid event at the East Midlands Conference Centre in Nottingham, with the title 'Decision making in MX – how to be a productive structural biologist'. The latest advances in macromolecular crystallography now often involve large multi-crystal and multi-dimensional experiments, and the more that we can delegate to automated pipelines, the more attention we can direct towards the novel aspects of an experiment. Sometimes those automated tools work, sometimes they fail without indicating exactly why or how, and sometimes they do not yet exist, but in any case, a user of these tools can easily be confronted with a massive amount of results and data that needs assessment and interpretation. We wanted to ask some simple questions: when to use which automated tool, how to interpret its output and when manual intervention is necessary. Speakers were asked not just to demonstrate their methods and software, but to share the decisions that go into using them well: when to trust, when to question and when to examine more carefully. It is our pleasure to present this special issue of articles arising from the meeting.

The meeting opened on the evening of the 3rd of January, immediately following the Diamond MX User Meeting, with a keynote lecture by Patrick Reinke (DESY, Hamburg, Germany). Drawing on the experience of running large-scale fragment- and drug-screening campaigns against SARS-CoV-2 main protease and other targets at PETRA III, he laid out the practical decisions that arise at every step of an ambitious crystallographic project. These included: which crystals to mount, which fragments to pursue, which hits to follow up, which existing tools to exploit and which to develop further. This was an overview into how to rapidly develop the workflows needed to translate hundreds of datasets into trustworthy biological insight at short notice.

The session continued with a round-table discussion under the title 'Crystallography is dead! Long live crystallography!' chaired by Elspeth Garman (University of Oxford, UK), with David Brown (Servier, France/University of Kent, UK), Ashwin Chari (Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany) and Arwen Pearson (Universität Hamburg/HARBOR, Germany) as panellists. The panel and audience discussed the future of the field in the era of ever more powerful and accurate structure-prediction methods. The conclusion was emphatically the second half of the title: the need for experimental crystallography is, if anything, sharper now, given that structural predictions allow us to ask more precise questions about ligand binding, dynamics and post-translational modifications which uncover unexpected biology. These are questions that only experimental data can answer with a high level of confidence.

The second day began with the traditional 'What's New in CCP4' session and continued with the first themed session of the meeting, on planning and execution of a diffraction experiment. Elspeth Garman (University of Oxford, UK) opened the session by reminding us that any macromolecular X-ray diffraction experiment must begin with an understanding of radiation damage, and that the most reliable strategy is to minimize damage from the outset. She discussed how absorbed dose can be estimated and monitored using *RADDOSE-3D* and how site-specific structural damage can be identified retrospectively, including when the original diffraction images are no longer accessible.



Graeme Winter (Diamond Light Source, UK) then argued for matching the experiment to the source: selecting the appropriate experimental setup and mode of data collection from the diverse set of beamlines available at synchrotrons around the world, including room-temperature, rotation crystallography, serial synchrotron crystallography (SSX) and long-wavelength experiments. The associated paper from his group (Thompson *et al.*, 2025) describes enhanced intensity-based clustering of isomorphous multi-crystal datasets within *DIALS* and *xia2.multiplex*, addressing a recurring decision point – which datasets from what crystals should be combined – when small differences in unit cell or in subtle ligand occupancy must be resolved within a larger ensemble.

Rasmus Fogh (Global Phasing, Cambridge, UK) closed the session by demonstrating how strategy computation and data collection can be automated at synchrotron beamlines: the recurring decisions about how many sweeps to acquire, at which orientations, with which transmissions and with what redundancy are converted into a workflow that derives and executes them on the fly. The accompanying paper (Fogh *et al.*, 2026) describes how the workflow takes the crystal symmetry, mounting orientation and intended experimental resolution as input. It then builds a multi-sweep strategy in which different orientations average out differences in Lorentz enhancement and multiplicity, derives optimal transmissions from the measured beam flux density, and supports native and phasing experiments in either user-driven or unattended mode through its integration with the *MXCuBE* beamline-control system.

The midday session was devoted to structure solution and model building in the era of accurate structure prediction. Dorothee Liebschner (Lawrence Berkeley National Laboratory, USA) discussed how *AlphaFold* models have completely changed the practice of molecular replacement in macromolecular crystallography. She covered the features for initial model placement in the *Phenix* suite, and how the decisions a crystallographer now makes in early-stage structure solution have shifted by using a high-confidence prediction rather than a homology model or an *ab initio* search model. The corresponding contribution from the *Phenix* team (Moriarty *et al.*, 2026) is on a different subject – small-molecule ligand restraints – and describes a library of about 37 000 entries derived by quantum-mechanical optimization of components from the PDB Chemical Component Dictionary and validated against the Cambridge Structural Database intended to support both crystallographic refinement and molecular-mechanics simulations.

The strides made in protein structure prediction algorithms such as *AlphaFold* risk leaving other molecules not yet covered by those methods behind. Kamel El Omari (Diamond Light Source, UK) offered this counterpoint in a talk titled *What AlphaFold can't do*: he showed how anomalous data, particularly the long-wavelength data routinely collected on beamline I23 at DLS, can be exploited to identify and locate light atoms and metal ions whose chemical identities cannot be inferred from a predicted model. The accompanying paper (El Omari *et al.*, 2004) describes how anomalous Fourier maps and refinement of f'' values from native data complement *AlphaFold*2 models with experimentally determined identities of metal ions and other key chemical species, an essential corrective in an era when the chemical content of a deposited model can too readily be assumed rather than verified.

Paul Emsley (MRC Laboratory of Molecular Biology, Cambridge, UK) closed the session with an account of efficiency strategies in *Coot* or, in his preferred formulation, of being lazy. He demonstrated how good defaults, targeted shortcuts and the new web-deployable *Moorhen* implementation streamline interactive model building of loops, ligands, alternative conformations, nonstandard linkages and metals: namely, the parts of a model where decisions by the scientist are still indispensable.

The afternoon session focused on the difficult features of a model, such as subtle or unusual structural elements where decisions about restraints and chemistry are frequently critical to the quality of the final result. Tristan Croll (Altos Labs, Cambridge, UK) opened with a retrospective marking five years of *ISOLDE*, drawing together the lessons learned about interactive, physics-based fitting of models into low-resolution maps. In this resolution range, local geometry, restraints and the modeller's judgement interact most tightly. He outlined the prospective decisions to be made in an environment where predicted models are increasingly used as the routine starting point for refinement.

Lucy Schofield (University of York, UK) presented work on the accurate modelling of glycans in PDB entries, illustrating how the stereochemical complexity and conformational preferences of pyranose rings make carbohydrate models prone to errors, and describing the software tools available to build and validate them correctly. The accompanying review (Schofield *et al.*, 2024) broadens that perspective to the full range of post-translational modifications represented in the PDB, such as phosphorylation, glycosylation, ubiquitination and the interconversion of chemical groups. She reviewed their structural importance, their prevalence in deposited entries and the recurring difficulties of building them accurately.

Garib Murshudov (MRC Laboratory of Molecular Biology, Cambridge, UK) closed the session by examining the relationship between reciprocal-space refinement and real-space model analysis, using metal-containing ligands as an illustrative case and making the case for restraints derived from observed coordination geometries rather than from chemical intuition alone. The accompanying paper by Babai *et al.* (2024) introduces *MetalCoord*, a method for extracting metal-coordination geometries from the Crystallography Open Database and using them to derive restraints for refinement in *REFMAC*5 and *Servalcat*, with the resulting updates contributed back into the CCP4 monomer library.

The third day opened with a session on partial data and partial occupancies, *i.e.* on how to extract reliable information when the signal is weak, the population is heterogeneous or the data arise from partial contributions distributed across many crystals or many states. Arwen Pearson (Universität Hamburg/HARBOR, Germany) gave a talk on serial data

collection and processing under the title *Same but different*: the same X-rays, the same proteins, but datasets composed of small slices of reciprocal space from many different crystals, each requiring careful treatment at every processing stage. The accompanying paper by von Stetten & Pearson (2026) addresses a particular failure mode of serial data collection: electron-density maps that are biased by model phases. They introduce the perturbed-model real-space difference density (PMRDD) test, in which deliberate errors are introduced into an existing model and the difference density at those perturbed sites is checked. When the difference density recovers the perturbations, enough data have been collected; when it does not, more crystals are needed.

Elke De Zitter (Institut de Biologie Structurale, Grenoble, France) described *Xtrapol*8, a tool for identifying and modelling low-occupancy states that has found application in time-resolved, ligand-soaking and radiation-damage studies, with comparisons to alternative methods. Nick Pearce (Linköping University, Sweden) closed the session with multi-data-set and multi-state refinement: 'everything, everywhere, all at once'. He discussed how conformational heterogeneity, partial occupancy and disorder could be modelled jointly across many related crystals rather than locally within each, recasting the question of 'did I see binding?' into the more rigorous 'what fraction of states is consistent with the data I have?'.

The next session moved on to experimental interactions: ligands, everywhere, all at once. Judit Debreczeni (AstraZeneca, Cambridge, UK) described the decision-making framework of an industrial drug-discovery setting, where speed, accuracy and throughput must be reconciled across hundreds of structures: the workflows, decision criteria and selection of tools differ substantially from those of an academic project, even though the underlying crystallography is identical.

Ed Daniel (University of Oulu, Finland) followed with a systematic account of laboratory information management for macromolecular crystallography (Daniel *et al.*, 2024), illustrated through the *IceBear* LIMS, which captures information from crystallization through to PDB deposition and supports metadata exchange between a home laboratory and the synchrotrons it sends crystals to; this constitutes a major component of FAIR data-management practice in the modern drug-discovery and structure-determination workflow.

Oliver Smart (Global Phasing, Cambridge, UK) closed the session with an examination of the decisions a structural biologist must make when fitting a ligand into experimental electron density: the importance of an accurate description of compound chemistry and conformation, and the role of integrated, real-time validation feedback in supporting well founded modelling choices (Smart *et al.*, 2026).

The closing session, *Structural Analysis: Climbing the Data Mountain*, returned to the fundamental question underlying every crystallographic decision: did my experiment work? Helen Ginn (DESY, Hamburg, Germany) showed how her *RoPE* software can be used to assess models in conformational space, allowing scientists to examine a collection of related crystallographic datasets and ask not only whether each model is consistent with its data, but whether the ensemble of models actually supports the dynamic narrative one is seeking to establish.

Briony Yorke (University of Leeds, UK) presented an application of this to serial UV/X-ray pump–probe crystallography of crystallin proteins, demonstrating how time-resolved methods can be applied to the kinds of technically demanding crystallographic experiments that lie outside any standard automated pipeline, and walking through the experimental and analytical decisions specific to this style of work.

Ashwin Chari closed the meeting with a demonstration of what good crystals on a high-quality beamline can yield when processed through advanced workflows: ultrahigh-resolution datasets in which the interpretation of difference density and anisotropic displacement parameters reveals finely detailed features that are otherwise inaccessible such as protonation states, alternative conformations, tightly bound solvent molecules and ions. At such resolutions the act of modelling itself becomes a decision-intensive discipline, in which what is in the data and what one chooses to model are not the same question.

We are very grateful to the staff at CCP4 for ensuring the smooth operation of a hybrid event. Our thanks go in particular to Karen McIntyre (CCP4) for the administrative work that made the meeting possible, to Stuart Eyres, Jonathan Oldfield and Gassan Ahmed for the recordings and the website that ensured the talks reached the wider community, and to Ville Uski for organizing the What's New in CCP4 session. We thank the panellists who took part in our opening discussion, and the speakers, poster presenters, tutorial leaders and participants, both in person and online, whose contributions made up the rest of the programme. Finally, we thank CCP4 Working Group 2 and the CCP4 Core Team for inviting us to be the scientific organisers of this meeting, and for trusting us with the title and the programme.

Across the talks and the contributions to this Special Issue, the same underlying answer to our central question recurred throughout: productivity in modern macromolecular crystallography is not a matter of automating more, nor of performing experiments by routine. It is a matter of making better decisions about which experiment to do, which method to apply, which features in a model to trust and which questions to ask of the data. It is about having software, and a community, that supports those decisions rather than obscures them. We hope the articles collected here will help readers, new and experienced, to make those decisions well.

References

Babai, K. H., Long, F., Malý, M., Yamashita, K. & Murshudov, G. N. (2024). *Acta Cryst.* D**80**, 821–833.

Daniel, E., Wierenga, R. K. & Lehtiö, L. (2024). *Acta Cryst.* D**80**, 580–587.

El Omari, K., Forsyth, I., Duman, R., Orr, C. M., Mykhaylyk, V., Mancini, E. J. & Wagner, A. (2024). *Acta Cryst.* D**80**, 713–721.

Fogh, R. H., Keller, P., Flensburg, C., Vonrhein, C., Paciorek, W., Carter, L. & Bricogne, G. (2026). *Acta Cryst.* D**82**, 314–329.
Moriarty, N. W., Case, D. A., Liebschner, D. & Adams, P. D. (2026). *Acta Cryst.* D**82**, 216–226.
Schofield, L. C., Dialpuri, J. S., Murshudov, G. N. & Agirre, J. (2024). *Acta Cryst.* D**80**, 647–660.
Thompson, A. J., Beilsten-Edmands, J., Tam, C., Sanchez-Weatherby, J., Sandy, J., Mikolajek, H., Axford, D., Jaho, S., Hough, M. A. & Winter, G. (2025). *Acta Cryst.* D**81**, 278–290.
Smart, O. S., Sharff, A., Flensburg, C., Vonrhein, C. & Bricogne, G. (2026). *Acta Cryst.* D**82**, https://doi.org/10.1107/ S2059798326007047.
von Stetten, D. & Pearson, A. R. (2026). *Acta Cryst.* D**82**, 227–236.