Grand Challenges

The Collaborative Computational Project Number 4 in Protein Crystallography (CCP4)

Over the next ten years, CCP4 will transform into a unified AI-enhanced, hybrid-structured biology platform that integrates multiple experimental modalities with cutting-edge computational methods

The Community

www.ccp4.ac.uk

CCP4 aims to produce and support a world-leading, integrated suite of programs that allows researchers to determine macromolecular structures by X-ray crystallography, and other biophysical techniques. Integrating cutting-edge approaches, experimental determination and detailed analysis into protein structure is key to the development of this suite. CCP4 is a community-based resource that supports the widest possible researcher community, embracing academic, not-for-profit, and for-profit research. CCP4 endeavours to play a key role in the education and training of scientists in experimental structural biology, and encourages the wide dissemination of new ideas, techniques and practice.

The Challenge

Overcoming the emerging complexities that obstruct the path from experimental data to trusted macromolecular model

Structural biology underpins modern bioscience and drug discovery. CCP4 enables macromolecular structure determination by using X-ray crystallography, which is crucial in the understanding of biological processes and the development of new therapeutics against diseases. However, the path from experimental data to macromolecular models faces growing complexity, due to the emergence of hybrid models that integrate data from a range of sources, including X-ray diffraction, computational models, cryo-EM, NMR, Molecular Dynamics, and biophysical / biochemical analyses. This generates novel type models that encompass the dynamic nature of macromolecules and subsequently, CCP4 is evolving from a crystallography-only platform to a truly integrative structural biology environment.

Key challenges ahead include the incorporation of AI-generated information, integrating various data sources like crystallography and AlphaFold predictions, to ensure seamless workflows. There are difficulties with data quality, as AI needs to enhance non-optimal datasets and manage imperfect data efficiently. Scalability and automation are crucial as modern facilities generate vast datasets requiring real-time processing. Trust, reproducibility, and validation are essential for regulatory compliance, demanding clear systems. Lastly, training and accessibility must improve to support both beginners and experts in the global structural biology community while addressing the complexities of large datasets. The large complex datasets generated from AI and complex methods, however, provide a challenge for training.

Community and workforce development will be crucial:

- Expanded training resources, including interactive tutorials, and cloud-based practice platforms.
- Cultivating the next generation of structural biologists with strong foundations in computation, statistics, and AI.
- Enable the community to understand and evaluate AI generated information, such as computational structures, hybrid models, and predictions.

The Solution

Transform into a unified AI-enhanced, hybrid platform that integrates experimental modalities with cutting-edge methods

In the next ten years, CCP4 will further the interoperability and cooperation with the wider research community by transforming into a unified AI-enhanced, hybrid-structural-biology platform that integrates multiple experimental modalities with cutting-edge computational methods. CCP4 will fully integrate the use of AI for modelling and validation, incorporating machine learning and intelligent model building. Seamless hybrid pipelines will combine frameworks, inform experiment design and accelerate structure solution. The creation of digital twins for biomolecular systems will real-time feedback from experiments to optimize data collection, together with a focus on managing data and metadata for interoperability, utilizing large language models to capture important textual information. Data sharing and benchmarking to validate AI-driven algorithms is essential, and will assist in the expansion of training resources, such as interactive tutorials and cloud-based practice platforms.

The Outcome

Recognised as the global benchmark for crystallographic software by expanding into a hub for integrative structural biology

Accelerating discovery in biomedicine, biotechnology, and fundamental biology, CCP4 will maintain the role as the global benchmark for crystallographic software by expanding into a hub for integrative structural biology. CCP4 will be supporting both academic discovery and industrial drug development; enable faster, more reliable, and reproducible structure solutions and secure the UK's and CCP4's global leadership in structural biology innovation.

CoSeC

Grand Challenges

More Information

CoSeC
www.CoSeC.ac.uk
CoSeC@stfc.ac.uk

CCP4
www.ccp4.ac.uk
ccp4@ccp4.ac.uk