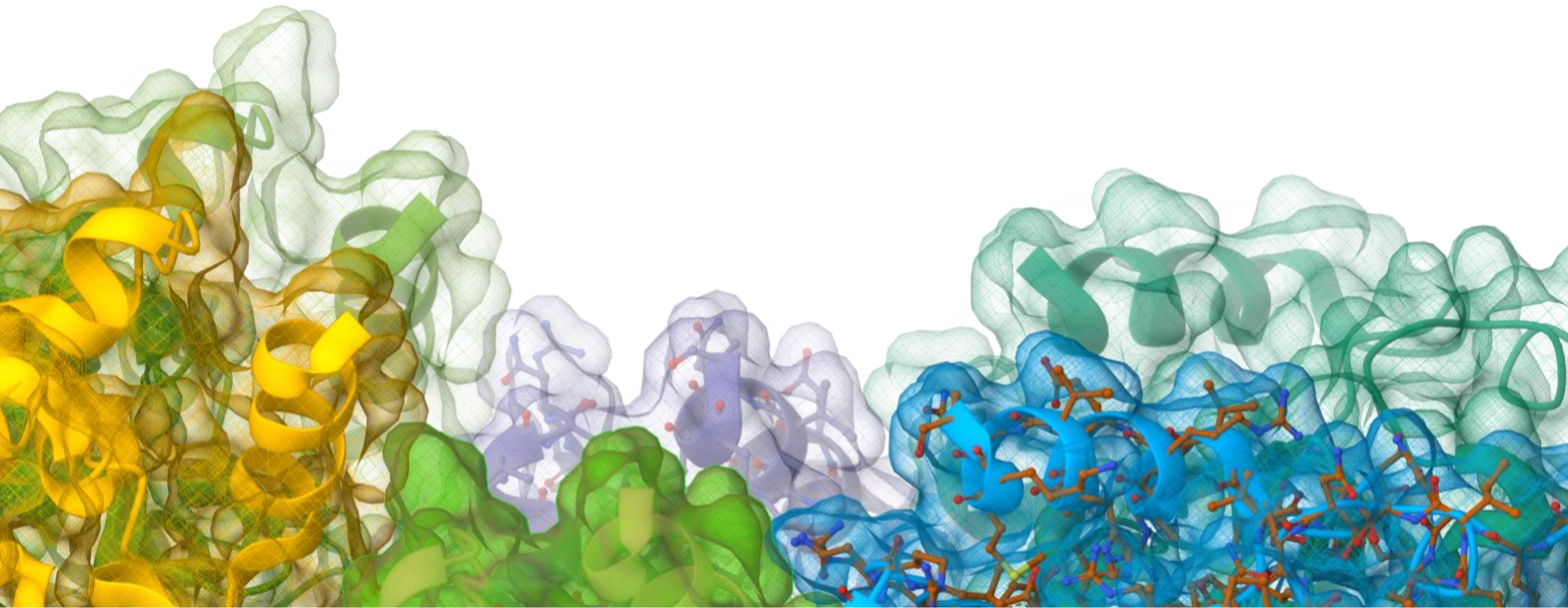
instruct ERIC Slovenia

Basics of Methodological Approaches in Structural Biology 2024

November 5–7, 2024
Ljubljana, Slovenia
National Institute of Chemistry

Book of Abstracts



Basics of Methodological Approaches in Structural Biology 2024
Book of Abstracts

Organizer:
Instruct ERIC Slovenia (Instruct.SI)

Issued by:
Department of Molecular Biology and Nanobiotechnology
National Institute of Chemistry, Ljubljana, Slovenia
Ljubljana, 2024

Editors:
Assoc. Prof. Dr. Marjetka Podobnik
Assoc. Prof. Dr. Miha Pavšič
Dr. Matic Kisovec
Jošt Hočevar
Maja Jamnik

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID 219521795
ISBN 978-961-7238-00-6 (PDF)

Workshop website:
https://instruct-eric.si/bmasb2024/

Book of abstracts:
https://instruct-eric.si/assets/docs/2024-bmasb2024-book_of_abstracts.pdf

Organising committee: Marjetka Podobnik[1], Miha Pavšič[2], Maja Jamnik[1], Matic Kisovec[1], Jošt Hočevar[2]

Scientific committee: Marjetka Podobnik[1], Miha Pavšič[2]

[1]*Department of Molecular Biology and Nanobiotechnology, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia*
[2]*University of Ljubljana, Faculty of Chemistry and Chemical Technology, Chair of Biochemistry, Večna pot 113, Ljubljana, Slovenia*

All lectures were held in the Great Lecture Hall at the National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana.

The members of the Instruct.Si consortium contributed to this workshop:

And with the financial support of:

Slovenian Research and Innovation Agency

Foreword

The members of the Instruct.SI consortium are delighted to organize the workshop **Basics of Methodological Approaches in Structural Biology**, which we have been eagerly anticipating for many years. We have observed that structural biology is rather modestly represented in most study programs in life sciences, with a few notable exceptions, despite the importance of structural data in interpreting the roles of biological macromolecules in normal and pathological processes, drug design, and similar areas.

Methodological approaches such as crystallography, nuclear magnetic resonance, and cryo-electron microscopy, together with other complementary methods, enable us to understand the structure of biological molecules or systems at the atomic level – how large molecules fold into stable yet dynamic structures, and how this relates to their mechanisms of action and, consequently, their biological functions.

These methods have been evolving for over 50 years, with crystallography advancing for over 100 years, and their development shows no signs of slowing down. From determining the 3D structures of isolated molecules or complexes until recently, we have reached a point where we can observe them at work in cells – this is cellular structural biology. Artificial intelligence now plays a significant role in this, as it has learned from already-determined 3D molecular structures to predict, with considerable accuracy, the structures of yet unexplored or even artificially created molecules, as well as interactions between them, thus entering the field of synthetic biology.

The goal of the workshop is to equip undergraduate, master's, and doctoral students, as well as other interested researchers from various generations, with knowledge that will enable them to engage in competent discussions with structural biologists, conduct independent research, and perhaps even inspire them to pursue a career in this field.

This is the initial workshop in what we hope will be a series of workshops on structural biology. For this reason, we will cover the basics and address the following questions:

- How does each method work?
- When should I use it?
- What types of samples can I measure?
- How should I prepare them?
- How do I know if the samples are suitable for analysis?
- What equipment do I need?
- Do we have it in Slovenia, and what if we don't?
- How do I perform measurements?
- What do raw measurement results look like?
- How do I process the data?
- How do I derive an atomic model of a biological molecule or complex from the data?

The workshop will last three days, with each day dedicated to a specific methodological approach – the first day to macromolecular crystallography, the second to nuclear magnetic resonance, and the third to cryo-electron microscopy.

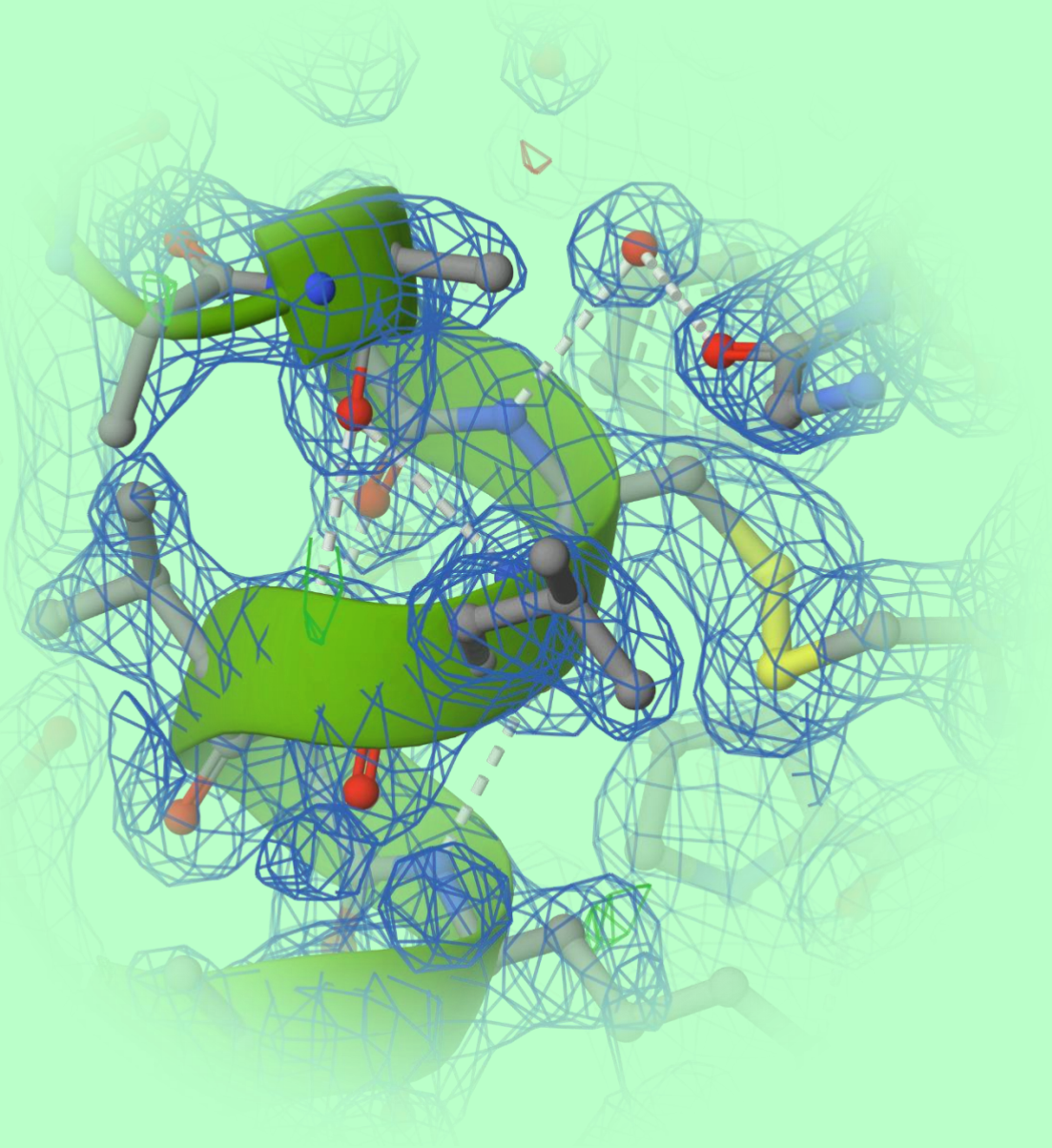
Together with you, we look forward to this workshop and wish all participating researchers, lecturers, and attendees a successful and enjoyable experience, along with plenty of new knowledge!

Workshop agenda

List of speakers in alphabetical order with hyperlinks to abstracts:

[1]*Department of Molecular Biology and Nanobiotechnology, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia*
[2]*Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia*
[3]*Jožef Stefan Institute, Department of Biochemistry and Molecular and Structural Biology, Jamova cesta 39, Ljubljana, Slovenia*
[4]*Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Jamova cesta 39, Ljubljana, Slovenia*
[5]*University of Ljubljana, Faculty of Chemistry and Chemical Technology, Chair of Biochemistry, Večna pot 113, Ljubljana, Slovenia*
[6]*University of Ljubljana, Biotechnical Faculty, Department of Biology, Večna pot 111, Ljubljana, Slovenia*

Day 1 (Tue, Nov 5):
Macromolecular Crystallography (MX)



Introduction to macromolecular crystallography

Dušan Turk[1,2]

[1]*Jožef Stefan Institute, Department of Biochemistry and Molecular and Structural Biology, Jamova cesta 39, Ljubljana, Slovenia*
[2]*Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Jamova cesta 39, Ljubljana, Slovenia*
dusan.turk@ijs.si

Macromolecular crystallography (MX) is a powerful method used to determine the three-dimensional structures of biological macromolecules, such as proteins, nucleic acids, and complex assemblies, at atomic resolution thereby providing detailed insights into their molecular architecture and function. It relies on the diffraction of X-rays by high-quality crystals of macromolecules. X-rays have much shorter wavelengths (on the order of 0.1 nm) compared to visible light, which allows them to interact with the electron clouds of atoms and reveal atomic-scale details. However, to obtain useful diffraction data, crystals are required rather than individual molecules. A single molecule cannot produce a strong diffraction pattern on its own due to low signal. Crystals, which contain many identical molecules arranged in a regular, repeating pattern, generate a coherent and interpretable diffraction pattern that is amplified and can be analyzed to extract structural information. The amplitude of the diffracted X-rays carries critical information about the positions of atoms in the macromolecule. The higher the intensity, the smaller the sample size required. This raises hopes that in near future the atomic structure of individual molecules will be resolved by X-rays produced by free electron lasers. However, due to the loss of phase information during diffraction, experimental or computational phase determination methods and computational refinement are employed to reconstruct the complete structure from the diffraction data.

The three-dimensional atomic-level structures obtained through MX reveal protein folding, molecular contacts, and ligand binding and therefore help us understand how biological macromolecules, such as enzymes, receptors, and antibodies, interact with other molecules and perform their biological functions. However, determining a structure alone isn't enough, as it doesn't reveal dynamic changes or interactions in real-time. Serial crystallography, which captures rapid, time-resolved snapshots of macromolecules, is emerging as a powerful tool to address this. It will enhance drug design, offering deeper insights into biological processes, disease mechanisms, and medical treatments.

Target selection and construct design

Ajda Taler-Verčič

University of Ljubljana, Faculty of Chemistry and Chemical Technology, Chair of Biochemistry, Večna pot 113, Ljubljana, Slovenia
ajda.taler-vercic@fkkt.uni-lj.si

Target selection and construct design are critical steps in structural biology, significantly influencing the quality and quantity of protein samples obtained for crystallography and other structural analyses. The day in the library can save a year in the laboratory. Small to medium-sized proteins are typically favored, as they are more amenable to crystallization. The design of protein constructs involves a careful integration of bioinformatics tools and laboratory techniques (such as limited proteolysis), which facilitate the prediction of suitable constructs that can enhance stability and solubility.

The incorporation of expression tags is essential for the effective production and purification of recombinant proteins. Common tags can enable isolation of protein by affinity chromatography (His-tag) or even improve protein solubility, stability, proper folding and more. Following purification, tag removal is often necessary to obtain the native protein for structural studies.

Additionally, the crystallization chaperone approach leverages specific binding partners or domains to facilitate crystal growth, while strategic mutations can be employed to improve protein properties or crystallization behavior, or to study enzymes.

Overall, the meticulous design of constructs, informed by both computational predictions and experimental insights, plays a pivotal role in advancing our understanding of protein structures and functions in the field of structural biology.

Protein production and purification

Nataša Lindič[1], Miha Pavšič[2], Ajda Taler-Verčič[2], Aleksandra Usenik[1]

[1]*Jožef Stefan Institute, Department of Biochemistry and Molecular and Structural Biology, Jamova cesta 39, Ljubljana, Slovenia*
[2]*University of Ljubljana, Faculty of Chemistry and Chemical Technology, Chair of Biochemistry, Večna pot 113, Ljubljana, Slovenia*
natasa.lindic@ijs.si, miha.pavsic@fkkt.uni-lj.si, ajda.taler-vercic@fkkt.uni-lj.si, aleksandra.usenik@ijs.si

Protein expression and purification are critical initial steps for successful macromolecular crystallography. Commonly employed expression systems are E. coli, yeast, insect and also mammalian cells. The choice depends on protein size and complexity, required post-translational modifications, sub-cellular localization, and other factors. Following construct design with the goal to improve crystallizability, an subsequent cloning, expression conditions, such as temperature, induction timing, and media composition, are optimized to enhance yield and maintain protein stability. Next, techniques like affinity chromatography, ion exchange, and size exclusion chromatography are used to isolate the protein and remove impurities. The purified protein must be monodisperse and of high purity, as these characteristics are essential for forming quality crystals for subsequent X-ray diffraction analysis.

Here, we provide specifics of the most commonly used expression systems in the light of macromolecular crystallography:

- Bacterial expression system is often the preferred choice for producing prokaryotic proteins due to its speed and cost-effectiveness. However, when it comes to eukaryotic proteins, this system has notable limitations. Specifically, bacterial systems do not provide post-translational modifications. Additionally, only the proteins with a limited number of disulfide bonds can be produced, by production in the periplasm or the use of specialized strains. Bacterial expression system is typically more suitable for low molecular weight proteins.
- Yeast (e.g. *Pichia pastoris*) are unicellular fungi and the simplest eukaryotic expression system. It shares most of the advantages of bacterial system as well as facilitates the correct folding of proteins from higher eukaryotes and enables proper post-translational modifications, including humanized N-glycosylation. Protozoan *Leishmania tarentolae* (LEXSY) offers a simple and low cost alternative to produce functional eukaryotic proteins *in vivo*. *In vitro* LEXSY is a rapid cell-free expression system mainly suitable for high-throughout small scale production of recombinant proteins.
- Insect cell expression system is used for producing proteins with complex post-translational modifications and proper folding, mimicking mammalian modifications. This makes them ideal for expressing eukaryotic proteins and large multiprotein complexes, including membrane proteins, that are difficult to produce in simpler systems like *E. coli*, ensuring higher structural fidelity for crystallography studies. Most often baculovirus-based gene delivery is utilized.
- Production of proteins in mammalian cells offers a native environment that reduces aggregation and is applicable when accurate post-translational modifications are essential for correct protein folding, stability, and function. It is especially suitable for the expression of complex, multi-subunit and membrane proteins (usually of mammalian origin). Two types of expression are used: transient expression, where plasmids are introduced into cells for rapid, scalable protein production, and stable expression, where gene(s) of interest is integrated into the cell's genome for consistent long-term production.

Choosing the appropriate expression system is essential for obtaining sufficient yield and quality of the target protein, directly influencing the success of macromolecular crystallography experiments.

Characterization of macromolecular samples

Aleksandra Usenik[1,2]

[1]*Jožef Stefan Institute, Department of Biochemistry and Molecular and Structural Biology, Jamova cesta 39, Ljubljana, Slovenia*
[2]*Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Jamova cesta 39, Ljubljana, Slovenia*
aleksandra.usenik@ijs.si

Macromolecular X-ray crystallography is a powerful method for determining the three-dimensional structures of macromolecules at atomic level resolution. As successful crystallization and high-resolution X-ray diffraction data collection are largely determined by the quality of the crystallized sample, prior characterization of macromolecular samples is a crucial experimental step. Key protein sample properties important for X-ray crystallography include: identity, purity, homogeneity, solubility, stability and functionality.

The identity and purity of protein samples are usually verified by methods, such as sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), size-exclusion chromatography coupled with multiangle light scattering (SEC-MALS), mass photometry (MP), mass spectrometry (MS) or N-terminal sequencing. The protein sample should be tipically more than 95% pure according to SDS-PAGE and Commassie blue staining assessment. Furthermore, the absorbance ratio (A_{260}/A_{280}) of around 0.6 indicates that the protein is relatively free of nucleic acid contamination.

Since heterogeneity and instability of the protein sample (e.g. presence of aggregates, oligomers, multiple conformational states, proteolytic degradation) can lead to poor-quality crystals or difficult interpretation of X-ray diffraction data, additional methods including native polyacrylamide gel electrophoresis (native PAGE), dynamic light scattering (DLS), isoelectric focusing (IEF) and circular dichroism (CD) should be used to ensure the sample is homogeneous in terms of size (monodispersity), charge and conformational state.

Adequate protein sample solubility, typically ranging from 5 to 50 mg/mL, is critical for growing high-quality crystals. This can be achieved through buffer screening, which may include a polyethylene glycol (PEG) solubility test or thermal stability assessments using methods such as differential scanning fluorimetry (DSF) or nano differential scanning fluorimetry (nanoDSF). Subsequent precrystallization test can help in determining the appropriate protein concentration for crystallization screening. Both DSF and nanoDSF are also frequently employed in heavy atom derivative screening for experimental phasing.

Crystallization, crystal harvesting, and preparation for data collection

Nataša Lindič

Jožef Stefan Institute, Department of Biochemistry and Molecular and Structural Biology, Jamova cesta 39, Ljubljana, Slovenia
natasa.lindic@ijs.si

Single, high-quality crystals are fundamental in macromolecular crystallography, enabling high-resolution structural determination of biomolecules and their complexes. These crystals form through self-organization of molecules into ordered, three-dimensional networks, held together by weak intermolecular interactions in a regular, repeating pattern. By providing X-ray diffraction patterns, they allow scientists to visualize macromolecular structures in atomic detail.

After successful protein expression and purification, a high-quality protein sample is used for crystallization. Crystallization involves a series of screening steps to identify conditions that promote the ordered arrangement of protein molecules in a supersaturated environment. To achieve supersaturation and induce crystallization, precise adjustments are made to variables such as protein concentration, pH, temperature, and precipitant type and concentration. Techniques like vapor diffusion, microbatch, free-interface diffusion, and microdialysis aid in optimizing these variables. When successful, crystallization produces well-ordered lattices that enable high-resolution diffraction, often better than 2.5 Å, which is crucial for a detailed atomic-level insight.

Protein crystals typically contain 50% solvent, which, while potentially reducing diffraction quality, helps to maintain proteins's structure and biochemical properties, such as ligand binding and enzymatic activity. This solvent environment allows for small molecules (e.g., substrates, inhibitors, drugs) to diffuse within the crystal, enabling X-ray analysis of complexes through soaking. In addition, co-crystallization can be used to incorporate ligands during the initial crystal formation process, providing an alternative method for studying ligand interactions within a structured lattice.

After formation, crystals are transferred to a cryoprotectant and flash-cooled to cryogenic temperatures to mitigate radiation damage during X-ray exposure. Optimizing cryogenic conditions and minimizing crystal mosaicity are essential for high-quality data collection. Cryogenic storage further facilitates transport and enables remote data collection at synchrotron beamlines where recent advances, such as microfocus beamlines and automated handling, have improved data quality, reduced radiation damage, and accelerated data acquisition, greatly supporting drug discovery and enhancing our understanding of protein function.

Diffraction data collection and processing

Matej Novak

Jožef Stefan Institute, Department of Biochemistry and Molecular and Structural Biology, Jamova cesta 39, Ljubljana, Slovenia
matej.novak@ijs.si

Protein crystals exposed to X-rays enable us to determine the protein structure. Crystals are signal amplifiers. They diffract incoming X-rays in distinct diffraction patterns, which contain information about their unit cell geometry and diffraction planes intensity. The combined information of a diffraction experiment enables us to reconstruct the atomic structure of the crystals.

Synchrotrons are the main facilities used nowadays for determining protein structures. Radiation produced by such particle accelerators is unparalleled in brightness and intensity, when compared to laboratory anodes. The high intensity and narrow focus are crucial for collection of high-quality diffraction data. In addition, they greatly reduce data collection times. There are many synchrotrons spread around the world. To be able to collect data, researchers have to submit proposals.

Data sets consist of many diffraction images, which have to be processed. There are a several programs available for this task (XDS, Dials, iMOSFLM, HKL3000, etc.). Data set processing determines the crystal lattice constants, diffraction indices and intensities of reflections. Processing also produces a report about quality of the diffraction data set. Once processing is complete, the diffraction data is exported to an ".mtz" file.

One of the biggest problems is retrieving the phase information from diffraction data. There are multiple methods to solve the phase problem like using derivative crystals with heavy metals – single/multiple isomorphous replacement (SIR/MIR), anomalous diffraction – single/multi-wavelength anomalous diffraction (SAD/MAD) or "borrowing" phases from a similar model - Molecular replacement (MR).

Successful completion of the phase solution process results in an initial electron map with substructures or a starting model, which has to be further refined through model building and refinement that follow.

Model building and refinement in macromolecular crystallography

Miha Pavšič

University of Ljubljana, Faculty of Chemistry and Chemical Technology, Chair of Biochemistry, Večna pot 113, Ljubljana, Slovenia
miha.pavsic@fkkt.uni-lj.si

Model building and refinement in macromolecular crystallography aim to produce an atomic model that fits observed data and adheres to stereochemical parameters. This process begins with an electron density map derived from diffraction data via experimental phasing or molecular replacement. In experimental phasing, phases are derived using heavy atoms or anomalous scatterers, while molecular replacement uses a homologous structure to generate approximate phases by positioning the search model within the crystal's asymmetric unit.

Building typically combines automated and manual approaches. Automated tools like ARP/wARP or Phenix AutoBuild place residues in high-confidence density regions, often yielding nearly complete models from high-resolution maps. Manual building, essential for complex regions such as regions with poor density and ligands, utilizes graphical programs allowing crystallographers to refine side-chain orientations and adjust the main chain for optimal map fit. Refinement iteratively enhances the model's fit using both automated and manual adjustments. Tools such as Phenix Refine and REFMAC minimize the difference between observed and calculated structure factors, while manual refinement enables corrections for side chains, poorly resolved regions, or ligand placement. During refinement, key metrics describing fit to the experimental data and stereochemical parameters are followed: R-factor, R-free, RMSD of bond distances, angles and impropers, clash score, and Ramachandran plot statistics. Refinement is considered complete when metrics stabilize, typically aiming for an R-free below 25% (depending on the resolution) and minimal clashes, yielding a model that accurately fits the density and satisfies stereochemical standards.

For the described steps in model building and refinement several programs are available, and most of them are bundled together with a unified user interface into two suites, CCP4 Suite and Phenix, and by default both use Coot for 3D graphical model depiction and manipulation. In recent time their cloud-based online counterparts became available: CCP4Cloud and Moorhen. The latter will be used at the hands-on part to manipulate and improve protein structural model to complement the introductory lecture.

Links:

CCP4 Suite (https://www.ccp4.ac.uk/)

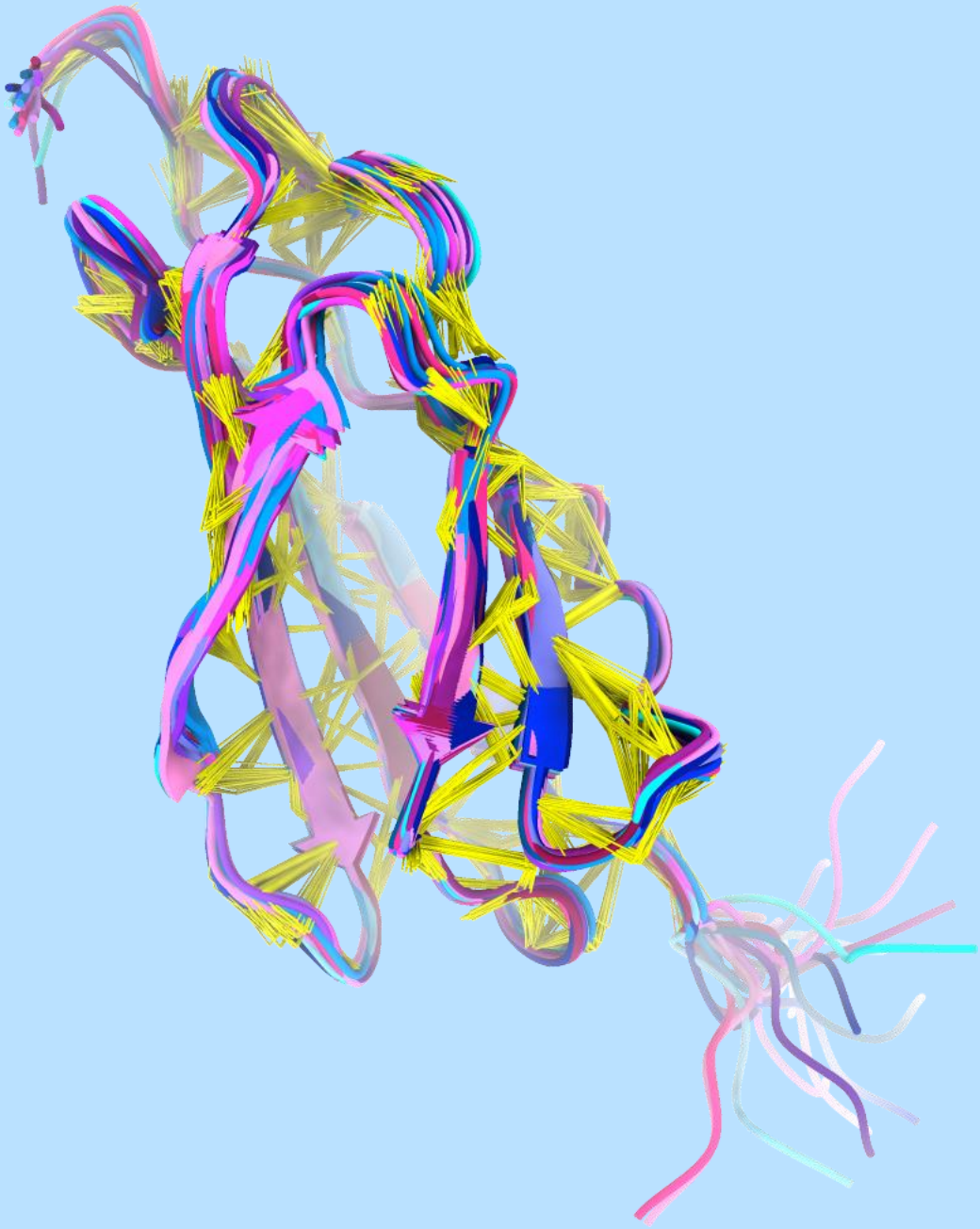
Phenix (https://phenix-online.org/)

Coot (https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/)

Moorhen (https://moorhen.org/)

Day 2 (Wed, Nov 6)

Nuclear Magnetic Resonance (NMR)



The basics of NMR

Vojč Kocman, Anita Kotar

Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia
vojc.kocman@ki.si, anita.kotar@ki.si

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful analytical technique widely employed in chemistry, biology, and materials science to investigate molecular structure, dynamics, and interactions.

NMR relies on the magnetic properties of atomic nuclei. Certain nuclei, such as ^{1}H, ^{13}C, and ^{15}N for biomolecules, possess an intrinsic magnetic moment that aligns with or against an external magnetic field when exposed to it. When these nuclei are perturbed by a radiofrequency pulse at a specific resonance frequency, they absorb energy and subsequently emit a signal as they relax back to their equilibrium state. This emitted signal, captured as free induction decay (FID), contains information about the chemical environment surrounding the nuclei. NMR spectra provide a detailed fingerprint of a molecule, with chemical shifts revealing the electronic environment of nuclei, coupling constants indicating spin-spin interactions between neighboring atoms, and peak intensities correlating to the number of nuclei present.

A critical advantage of NMR is its non-destructive nature, allowing for the analysis of samples in their native states. Additionally, it can provide dynamic information, such as reaction kinetics and molecular motion over a range of time scales. However, NMR's sensitivity is limited, requiring relatively large sample quantities (in mM range) or isotopic labeling for analysis. In this lecture, we also discuss details how to prepare the NMR samples, from concentration of macromolecules, (deuterated) solvent and composition of the buffer.

Introduction to 1D and 2D NMR spectra

Damjan Makuc

Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia
damjan.makuc@ki.si

NMR is an excellent tool for structural characterization of small molecules as well as biomolecules. Most commonly used NMR techniques are 1D ^{1}H and ^{13}C NMR spectra, due to hydrogen and carbon being the most abundant atoms in the structure of organic molecules. During this lecture, we will follow "live" NMR characterization of ferulic acid. The analysis of 1D NMR spectrum will include integration of signals, peak picking and multiplet analysis. Furthermore, 2D NMR experiments will allow us to fully characterize the model compound using homonuclear COSY (correlation spectroscopy) and TOCSY (total correlation spectroscopy) spectra to detect coupled protons in the same spin systems as well as heteronuclear ^{1}H-^{13}C HSQC (single quantum coherence spectroscopy) for detection of directly bonded protons and carbons, and HMBC (multiple bond correlation) that is suitable for determining long-range connectivities and can yield signals for those nuclei that are separated by 2–4 bonds. NOESY (Nuclear Overhauser Effect spectroscopy) shows through-space interactions within the molecule, rather than the through-bond interactions seen in the other methods. An advantage of this experiment is that it can lead to an understanding of a molecule's three-dimensional conformation. These through-space correlations are made via spin-lattice cross-relaxation. Dipole interactions between protons close in space generate NOE transfers that result in 2D correlation signals. It is sometimes necessary to run NMR experiments at temperatures significantly higher or lower than ambient. Such variable temperature (VT) experiments are used for a number of reasons, mainly to provide insights into the dynamic and kinetic behaviour of molecules, or to simplify spectra for compounds that are undergoing conformational exchange e.g. rotamers. The combination of all NMR experiment discussed above allows one to define the structure of a molecule in great detail.

Production of DNA/RNA and protein molecules suitable for NMR structural characterization

Maja Marušič

Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia
marusic.maja@ki.si

Structural studies of biological macromolecules by NMR spectroscopy require the preparation of pure, correctly folded, stable and often isotopically labelled macromolecules in milligram quantities. Isotopic labelling with ^{13}C, ^{15}N, ^{2}H and ^{19}F introduces NMR-active nuclei and requires detailed planning of the preparation of biological macromolecules to maximise yields and increase spectral resolution.

Nucleic acid preparation is possible by solid support chemical synthesis, which allows full control over the labelling or modification site, and by various enzymatic approaches and derivatives, where labelling is typically residue specific. Enzymatic methods include hairpin extension and palindrome-nicking-dependent amplification for DNA and in vitro transcription combined with ribozymes for RNA. An integration of approaches is also possible in the context of segmental labelling, which allows extension of nucleic acid size and labelling schemes using T4 DNA/RNA ligase in combination with RNaseH.

Isotopic labelling of proteins requires the use of minimal medium for growth of e.g. E. coli cells supplemented with a ^{13}C, ^{15}N and/or ^{2}H source. Various labelling schemes are used, such as uniform labelling for smaller proteins, labelling of methyl groups to exploit the TROSY effect for large proteins and macromolecular complexes, and various derivatives such as segmental labelling or the LEGO labelling approach. Fluorinated amino acid analogues are added to the growth media to introduce 19F nuclei, which are particularly useful for in-cell NMR studies. Other expression systems range from yeast, insect and mammalian cells to the cell-free system, which allows the production of proteins with optimal NMR properties.

Strategies for assignment of DNA/RNA NMR spectra

Anita Kotar

Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia
anita.kotar@ki.si

Assignment of NMR spectra involves associating each resonance in the spectrum with a specific nucleus in the studied molecule. For DNA and RNA, this is achieved through the principle of sequential walking, which links neighboring nucleotides. Starting with the first identified nucleotide (i), the next nucleotide (i+1) is identified by its connectivity with (i), for example H1'(i)-H8(i+1). This process is carried out by analyzing the NOESY spectrum, a key experiment in sequential assignment of DNA/RNA NMR spectra. The Nuclear Overhauser Effect (NOE) reveals interactions between nuclei that are spatially close (within ≤ 5 Å), rather than directly bonded. Sequential walking is repeated across the NOESY spectrum until all nucleotides are assigned. However, this process can be disrupted by structural motifs such as bulges, mismatches or loops (which provide valuable insights for structural modeling) and also by overlapping in NMR spectra. Signal overlap in NMR spectra, common due to the limited chemical diversity of DNA and RNA's four nucleotide building blocks, is particularly challenging for larger molecules. This limitation can be addressed using isotopic labeling strategies with ^{2}H, ^{13}C, and ^{15}N. These isotopes broaden the chemical shift range, reducing overlap and facilitating assignment. Accurate NMR assignment is bases for subsequent analyses, including structural determination. Due to time consuming process of NMR assignments there are several attempts for automatization. The algorithms like NMRFx incorporate primary sequences and predicted secondary structures into prediction of chemical shifts of studied molecule, which promise for accelerating the elucidation of nucleic acid structures and dynamics.

Links:

NMRFx (https://nmrfx.org/)

Strategies for assignment of protein NMR spectra

Anita Kotar

Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia
anita.kotar@ki.si

NMR protein assignment integrates advanced experimental techniques and computational tools to match resonance signals with nuclei, providing a foundation for structural analysis. NMR assignment of double-labeled (^{13}C,^{15}N) protein molecules relies on a number of specialized pulse sequences using triple-resonance experiments that correlate the signals of ^{1}H, ^{15}N, and ^{13}C nuclei of protein backbone.

2D ^{15}N-HSQC is a basic NMR experiment, which serves also as a "fingerprint" for the sample quality. Most of the signals in 2D ^{15}N-HSQC spectrum are represented by HN-N pairs corresponding to backbone amide groups. The triple resonance NMR experiments, where correlations between nuclear spins arise due to scalar coupling trough small number of covalent bonds, are showing the correlations between ^{1}HN, ^{15}N, $^{1}H\alpha$, $^{13}C\alpha$ and ^{13}CO atoms of the backbone and $^{13}C\beta$ in $^{1}H\beta$ atoms of the side chains. The most standard 3D experiments are HN(CO)CA, HNCA, CBCA(CO)NH and HNCACB to determine the connections between two neighboring, (i-1) and (i), amino acid residues. In addition, ^{13}C-detected heteronuclear experiments are critical for side-chain assignments. Manual assignment can take weeks or months to complete depending on the size of the protein, so there are several programs for automatic backbone assignment (for example CARA and MARS). They use chemical shifts from 3D assignment spectra and secondary structure prediction as its input. Advanced tools leveraging artificial intelligence, like AutoAssign, can complete backbone assignments for small proteins in seconds. Despite these advancements, manual validation remains crucial to ensure the accuracy and completeness of assignments.

Links:

CARA (http://cara.nmr.ch/doku.php/home)

MARS (https://www3.mpibpc.mpg.de/groups/zweckstetter/_links/software_mars.htm)

AutoAssign (https://montelionelab.chem.rpi.edu/index.php/our-software-2/)

NMR's time to shine: dynamics and conformational changes of molecules

Maja Marušič

Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia
marusic.maja@ki.si

Flexibility and changes in structure are required for macromolecules to perform their desired functions. Changes in structure are called dynamics and indicate that the structure of a molecule is in equilibrium and energy is required to move between different structural states. To reveal dynamic processes and higher energy structures, a myriad of NMR methods have been developed, covering timescales from picoseconds to hours, to elucidate these dynamics in macromolecules with atomic resolution.

Chemical and conformational exchange can affect the intensity, chemical shift and line width of the NMR signal. The appearance in the NMR spectrum depends mainly on the difference in chemical shifts of a given nucleus in the exchanging structures/chemical environments with respect to the exchange rate and the relative populations of the exchanging conformers. It is most straightforward to group exchange into three different regimes: slow, intermediate and fast, each of which requires the use of a specific NMR method and provides a specific description of the exchanging system.

As an example of dynamics in a slow regime, we discuss the development and use of the SOFAST real-time 2D NMR methods, which detect the folding or unfolding of proteins at atomic resolution. For real-time NMR, a non-equilibrium process is initiated by a change in conditions, such as a jump in pH, which initiates (re)folding, followed by the (dis)appearance of signals, and allows the cooperativity of the folding process to be determined by the lifetimes of individual amide protons.

Furthermore, we discuss the detection of so-called invisible states in the intermediate timescale regime in GT mismatches. Short-lived, high-energy enolic and anionic forms of GT base pairs evade detection of errors during nucleotide selection and are thus determinants of mutations during replication.

High-resolution structures of biomolecules constructed from NMR derived structural restraints

Maja Marušič

Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia
marusic.maja@ki.si

NMR structure determination is conducted through the utilisation of software that facilitates a variant of molecular dynamics called simulated annealing. In this process, an elevated temperature is employed to facilitate the molecule's transition through high-energy barriers that separate it from the global energy minimum state. The calculation of the structure is guided by the structural restraints, which are derived from NMR spectral data. These structural restraints include distances obtained on the basis of the intensity of the nuclear Overhauser effect, hydrogen bond distances, torsion angle constraints, chemical shifts and potentially long-range distances such as residual dipolar couplings, paramagnetic relaxation enhancements or incorporation of data obtained with various other techniques, for example small angle X-ray scattering.

The structures determined by NMR spectroscopy differ from those determined by other methods in that they are presented as an ensemble of structures, comprising a selection of structures with the lowestenergy and least violations of the structural restraints. Flexible regions are represented in such ensembles by a larger set of conformations. This may reflect a smaller number of data points or ambiguous data for a given part of the macromolecule. However, it may also indicate the presence of dynamic regions that are important for the functional outcomes.

Hands-on: NMR spectra recording demonstration and evaluating ligand-protein/DNA/RNA interactions

Maja Marušič, Anita Kotar

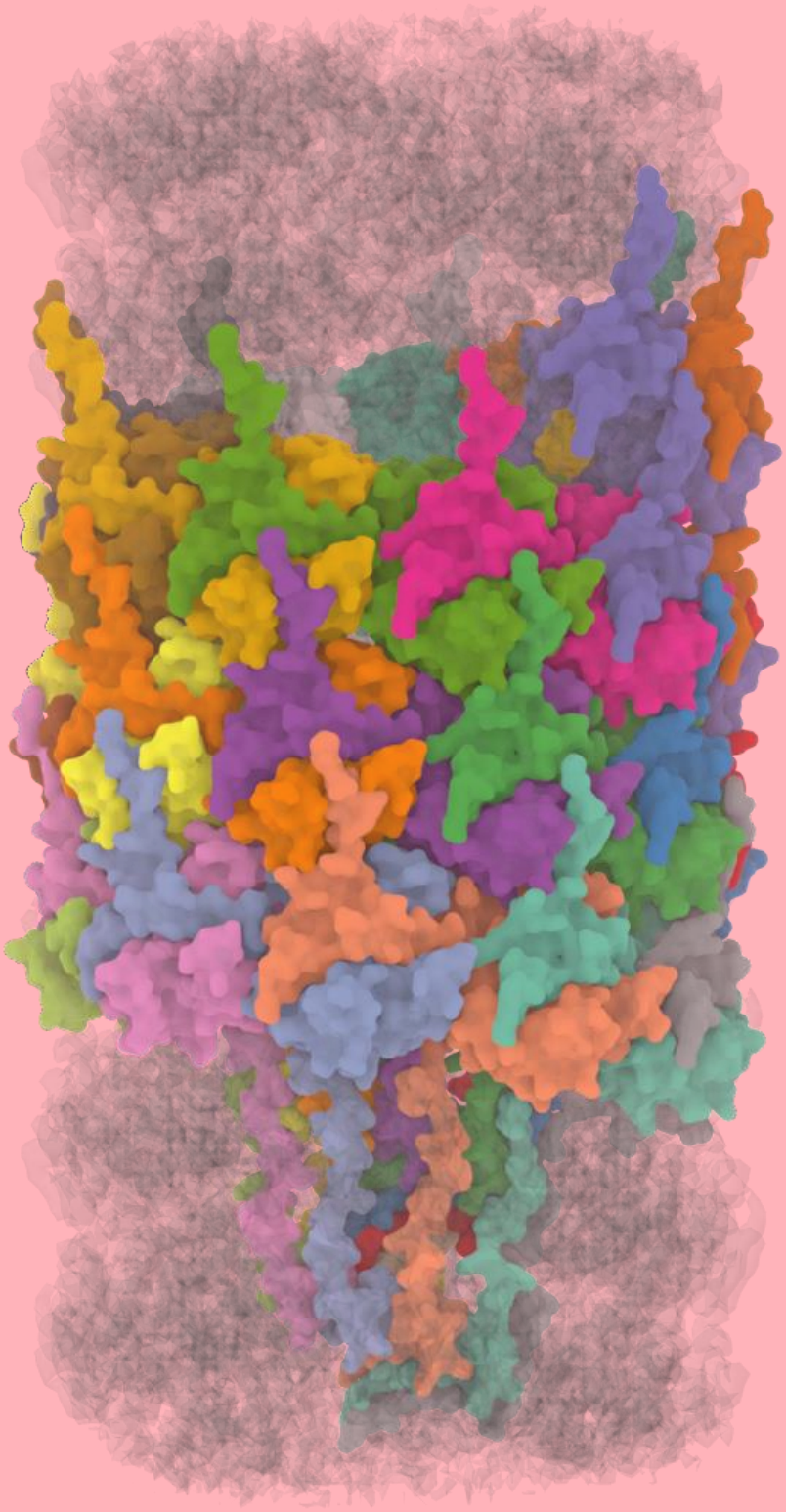
Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia
marusic.maja@ki.si, anita-kotar@ki.si

In the first part, we use remote access to the NMR spectrometer and record a 1D ^{1}H spectrum using the zgesgp pulse sequence on the 41-residue complex between RNA and PNA. The process begins with the insertion of the sample into the magnet, after which the magnet is prepared for recording. This entails establishing a lock, tuning and matching the probe, and homogenising the magnetic field through a shimming procedure. Subsequently, the length of the 90° pulse and the receiver gain are determined, which simultaneously provides insight into the efficiency of water suppression. The width of the spectrum is set based on the expected chemical shift of the resonances, and the number of scans is adjusted according to the sample concentration. The spectrum is recorded in a few minutes. The Fourier transform from the time to the frequency domain allows us to obtain a spectrum from which we can determine the type of macromolecules present in the sample, their purity and fold, as well as the presence of organic solvents and impurities.

In the second part, we analyse the interactions between two D-mannose-based glycomimetics and DC-SIGN protein receptor. DC-SIGN (Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin) is a type II transmembrane C-type lectin located specifically on dendritic cells, which acts as an adhesion molecule and binds highly glycosylated proteins present on the surface of several pathogens (HIV-1, Ebola virus, SARS, etc). We use saturation transfer difference (STD) NMR spectroscopy to investigate the binding propensity of monovalent glycomimetics to DC-SIGN at the atomic level of resolution. STD NMR is a ligand-based NMR technique for the study of protein–ligand interactions that requires only 1 nmol of protein (> 10 k Da) (high sensitivity) and allows characterisation of the structure-activity relationship without the need to isotopically label either the ligand or the protein. We demonstrate the process of analysing STD NMR spectra: 1) processing of NMR spectra (phasing, baseline corrections), 2) evaluation of STD effects and 3) analysis of the epitope mapping on the ligands.

Day 3 (Thu, Nov 7)

Cryo-Electron Microscopy (Cryo-EM)



Transmission electron microscopy – a look into structure from molecules to tissues

Nada Žnidaršič

University of Ljubljana, Biotechnical Faculty, Department of Biology, Večna pot 111, Ljubljana, Slovenia
nada.znidarsic@bf.uni-lj.si

Structure is inherently linked to function at all levels of biological organization, comprising molecules, cells, tissues and finally organisms. The key advantages of transmission electron microscopy (TEM) include the capacity for imaging with superior spatial resolution and the ability to resolve the cell ultrastructure simultaneously with the localization of selected molecules. Recent advances in 3D imaging across a range of size scales by electron tomography and volume electron microscopy have significantly contributed to improved understanding of biological systems. TEM is capable of atomic resolution, but the effective resolution of imaging is limited for biological samples and strongly depends on the specimen type and preparation procedures (ultrathin sections of cells/tissues or suspension of particles). Nevertheless, TEM enables us to image the ultrastructure of cell organelles, the structure of bacteria, viruses, macromolecular complexes and molecules. Preparation of biological samples for TEM is inherently complex because of the competing demands to image the sample in a vacuum environment of the microscope column and as close to its original state as possible. Additionally, the specimen has to be thin enough to permit transmission of electrons (generally less than 100 nm), the inherent contrast of biological samples is low and samples are susceptible to damage induced by electron beam. 3D structure of molecules is analyzed either by imaging the cryoimmobilized molecules in suspension in cryo-conditions (cryo-TEM) or by applying the negative staining technique (NS-TEM). Numerous variants of NS-TEM have been developed, but in general, molecules in suspension are surrounded, permeated, supported and embedded by a layer of heavy metal-containing solution, which generates high-contrast images of the molecules in TEM. Finally, advanced image analysis is critical to extract quantitative results from image data. The frontiers of imaging in life sciences are currently being pushed toward the integration and correlation of multiple imaging modalities across size scales to obtain a comprehensive insight into biological structures and processes.

Cryo-electron microscopy

Matic Kisovec

Department of Molecular Biology and Nanobiotechnology, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia
matic.kisovec@ki.si

Cryogenic electron microscopy (cryo-EM) is now an established method for determining the detailed 3D structures of biomolecules such as proteins, lipids, DNA and RNA. The samples are vitrified quickly, there is less radiation damage due to the low temperature and the sample is more stable. All this enables the observation of biomolecules in near-native conditions at atomic resolutions.

I will introduce you to the advantages of the 'cryo' and explain how this relates to the accelerating voltages of modern transmission electron microscopes. We will take a look at the basic workflow from sample in solution to final data analysis, which is the prerequisite for 3D model building. Three modalities of cryo-EM (Single Particle Analysis (SPA), cryo-electron tomography (cryo-ET) and microcrystal electron diffraction (MicroED) will be presented with examples. We will discuss the limitations and advantages of cryo-EM and new approaches to in situ and integrative structural biology, e.g. cryo-correlative light and electron microscopy (CLEM) and time-resolved cryo-EM.

Software development is an integral part of the so-called 'resolution revolution'. The 2017 Nobel Prize in Chemistry was awarded for advances in data analysis software and equally important advances in sample preparation and microscope hardware improvements.

Resolution and map generation will be compared between cryo-EM and X-ray crystallography and both methods will be compared with NMR with their advantages and disadvantages.

In the end we will look at the interactions between molecules and how these can be studied with cryo-EM. Cryo-EM is also used in drug discovery, and I will present some examples. Finally, we will take a look at the cryo-EM facility, which is part of the Centre for Molecular Interactions and Structural Biology (CMISB) at the National Institute of Chemistry.

Data analysis, structure determination I: Introduction to single particle analysis

Gašper Šolinc

Department of Molecular Biology and Nanobiotechnology, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia
gasper.solinc@ki.si

This lecture will explore the fundamental principles of single particle analysis (SPA), a powerful cryo-electron microscopy (cryo-EM) technique that allows us to determine high-resolution 3D structures of biomolecules from 2D images. First, we will discuss the basics of electron microscope imaging and how the interaction of electrons with samples generates 2D projections. We will then explore the principles of reconstructing 3D maps from 2D images, emphasizing the role of different orientations in obtaining accurate spatial information.

The core workflow will be presented step by step, starting from motion correction which compensates for sample drift, to contrast transfer function (CTF) estimation, particle picking, and the different classification steps. Each step will be discussed in terms of its purpose and impact on data quality and final resolution. We will conclude with 3D refinement explaining how it allows for the construction of a high-resolution model that represents the 3D structure of the studied biomolecular complex.

The most used software packages for analyses are Relion and CryoSPARC, in some cases, a combination of both is used. Participants will gain hands-on experience, with the latter, by following along with a demo dataset, allowing them to directly observe the effects of each processing step and understand the rationale behind this sequential workflow. By the end of the lecture, participants will be equipped with a basic understanding of SPA and the steps required to go from raw micrographs to a refined 3D map.

Links:

Relion (https://relion.readthedocs.io/)

CryoSPARC (https://cryosparc.com/)

Data analysis, structure determination II

Neža Koritnik, Andreja Kežar, Primož Bembič

Department of Molecular Biology and Nanobiotechnology, National Institute of Chemistry, Hajdrihova ulica 19, Ljubljana, Slovenia
neza.koritnik@ki.si, andreja.kezar@ki.si, primoz.bembic@ki.si

In the second part of cryo-EM data analysis and structure determination, we will address special cases, such as helical assemblies and macromolecular complexes with flexible components.

Helical symmetry is frequently observed in many natural complexes, including DNA, filamentous viruses, bacterial flagella, and cytoskeleton. To determine the structure of a helical filament, it is essential to establish the approximate helical parameters of its assembly, such as rise and twist. These parameters are calculated from the Fourier space (power spectrum of a helical projection image) using Fourier-Bessel indexing, which is often considered laborious, difficult, and error-prone. Recently, a software tool named Helixplorer has been developed to assist in determining these parameters from the power spectra. We will provide examples of data analysis and structure determination of helical particles.

In order to fulfill their biological function, macromolecules must be able to adapt to different interaction partners, which means that either small parts or entire subdomains must be flexible. In the case of cryo-EM, however, flexibility can lead to parts of the protein volume being less well defined or noisier when compared with more rigid parts of the macromolecule. Using p97 (a cellular multitool with over 40 known adapters) as an example, we will show how to address these challenges. Using a mask over the flexible N-terminal domain of p97 and with the help of 3D classification, 3D variability analysis and local refinement, we were able to obtain a better-defined N-terminal domain.

Links:

Helixplorer (https://rico.ibs.fr/helixplorer/helixplorer/)

CryoSparc guide (https://guide.cryosparc.com/)

CryoSparc tutorial video series (https://www.youtube.com/@structurabio/videos)